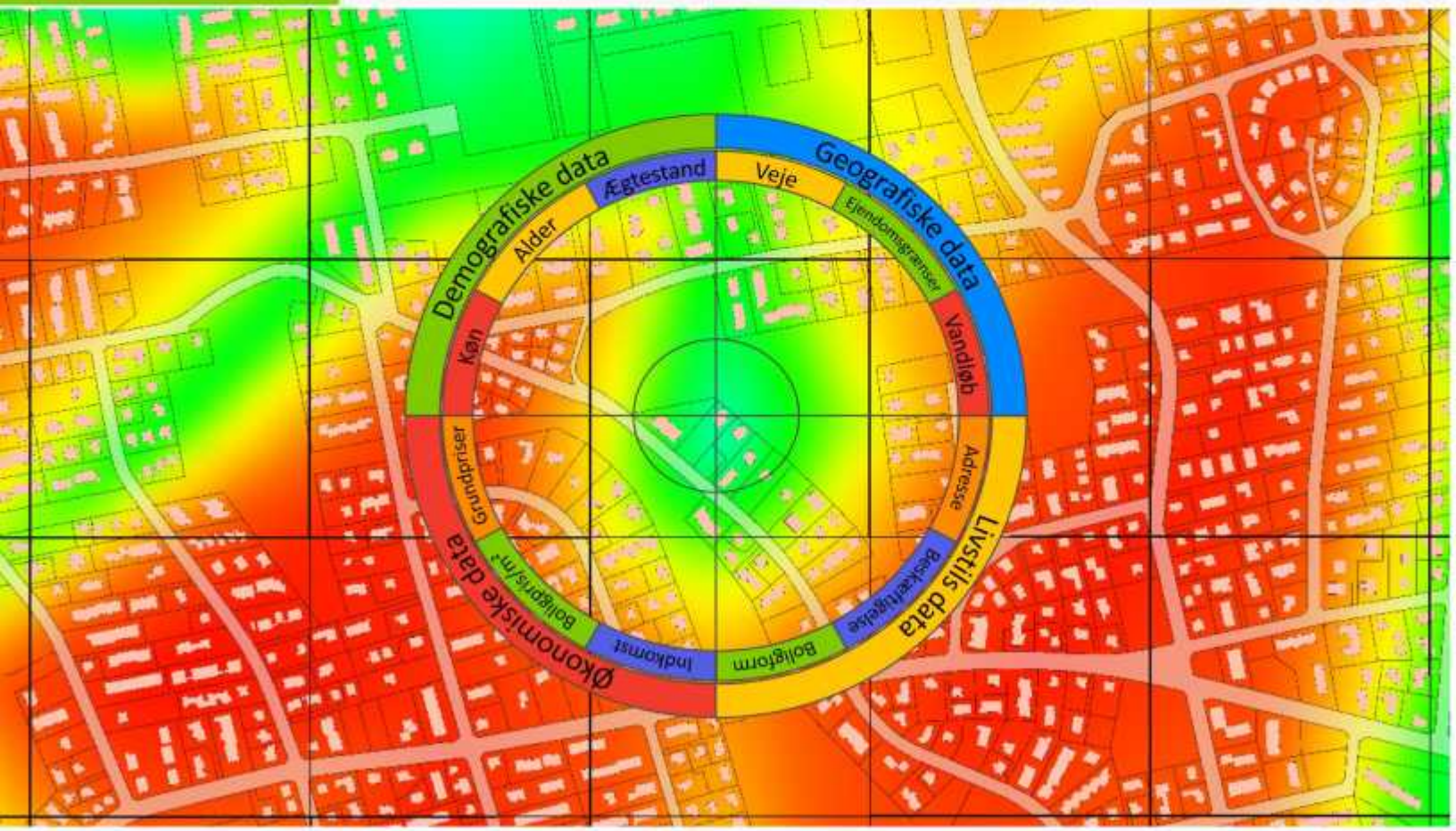
Værdier og information i geografiske data - nøglen til et beslutningsgrundlag

Aalborg Universitet, Institut for planlægning, Landispektøruddannelsen



Steen Petersen
Aalborg Universitet
2015

TITELBLAD

Titel:	Værdier og information i geografiske data – nøglen til et beslutningsgrundlag.
Tema:	Afgangsprojekt – Institut for planlægning.
Projektperiode:	1. juli 2014 – 16. januar 2015.
Oplagstal:	4
Samlet sideantal:	105
Bilagsantal:	9

Udarbejdet af:

_____

Steen Petersen

Vejleder: Lars Brodersen

1. Forord

Projektet beskriver min arbejdsmetodik, herunder de valgte afgrænsninger og den fremlægger mine overvejelser og konklusioner om brugen af data i geografisk analyse og visualisering af data i forbindelse med identificering af sammenhænge, tendenser og mønstre af tabeldata. Det skal dels ses i lyset af, at ved at omdanne tal til grafiske former, kan man gøre det lettere og enklere for en bruger, at forstå "historierne" som disse tal skjuler.

1 FORORD

Dette projekt er resultatet af mit afgangsprojekt, der er udarbejdet ved Aalborg Universitet, Landinspektørstudiet. Afgangsprojektet er udarbejdet i perioden 1. juli 2014 til 16. januar 2015. Projektet er skrevet i Helsingfors i Finland, hvor jeg til daglig bor og arbejder.

Projektet beskriver min arbejdsmetodik, herunder de valgte afgrænsninger og den fremlægger mine overvejelser og konklusioner, om brugen af data i geografisk analyse og visualisering af data, i forbindelse med identificering af sammenhænge, tendenser og mønstre af tabeldata. Det skal dels ses i lyset af, at ved at omdanne tal til grafiske former, kan man gøre det lettere og enklere for en bruger, at forstå "historierne", som disse tal skjuler.

Projektet er udsprunget af et ønske om, at dygtiggøre mig på et område inden for GIS-verdenen, som i de seneste år har fået større og større betydning på min arbejdsplads, og i de kommende år vil få en dominerende rolle i arbejdsgangen, i forbindelse med GIS-relaterede projekter til vores kunder.

Hensigten med projektet er at synliggøre de fordele, der kan være ved en effektiv anvendelse af geografisk analyse i forskellige situationer, ved at visualiseret og dermed definere et beslutningsgrundlag. Jeg håber at projektet kan fremkalde en debat om mulighederne ved anvendelsen af geografisk analyse, som en struktureret og integreret tilgang til en problemløsning, en beslutningsproces, samt planlægning af tjenesteydelser. Et supplerende formål med projektet er, at inspirere interesserede brugere af GIS og geografisk analyse, til at udnytte de muligheder, der allerede nu findes inden for dette område.

Jeg vil gerne rette en varm tak til Aalborg Universitet, der har givet mig lov til at færdiggøre min kandidatuddannelse, efter så mange års fravær fra universitetsmiljøet – TAK.

Kapitler, afsnit, figurer og tabeller er nummereret med tal i projektet, mens bilag er nummereret med bogstaver og tal. Der angives kilder ved figurerne, hvis disse ikke er udarbejdet eller modificeret af projektets forfatter.

Kildehenvisninger til bøger, publikationer og internetsider er vist som: [Nummer]. En fyldestgørende informationer om de brugte kilder, kan findes i litteraturlisten bagest i dette projekt.

Helsingfors, den 16. januar 2015

2. English Summary

By combining the analytical power of databases with spatial properties and options in short, companies can explore and analyze the relationship between spatial data and business data. Data is the primary form of information. It consists mainly of recordings of transactions or events that will be used for exchanges between people. Data has no sense unless you understand the context in which these have been collected. The goal in this project is to create a version of the truth for spatial data, that can be used throughout an organization to create accurate, meaningful and consistent information, that serve as a foundation for a decision-making. The conclusion in my project is, that values and information in geographical data is a key to form a basis of a decision. Spatial data represented in the form of objects, lines and polygons, is not unique representatives of a result, but the basis for making a decision. "Crow flock together", is a saying that means that we tend to associate with like-minded. Maybe that's just the way it is – after all.

2 ENGLISH SUMMARY

Data constitutes one of the primary forms of information. It essentially consists of recordings of transactions or events, which will be used for exchange between humans or even with machines. As such data does not carry meaning, unless we understand the context in which the data has been gathered. A word, a number or a symbol can be used to describe a business result. It is the context which gives the meaning and this meaning makes it informative.

Geographic data shares many of the same problems associated with other types of data, such as multiple source systems and spatial data formats, data quality issues and semantic inconsistencies. Location Intelligence technology can help with spatial data integration and quality improvement efforts.

The goal with this project is to examine the data-values and information “hidden” in Geographical data and see if there is a correlation between the car models and accommodation types in decision-making, to improve the possibility of finding a larger customer segment to a product, in this case a VW-Golf car and to show how geographical data can be used throughout an organization to create accurate, meaningful and consistent reports and maps, and serve as a foundation for advanced spatial analysis.

Advanced spatial analysis tools relate to familiar advanced analytics software for statistical analysis, data mining, real-time forecasting and business optimization, but they are modified to address the unique characteristics of spatial data and relationships. In this present project, the DIKW-hierarchy will be used as a theoretical reference model throughout the project to explain and set into relation, how the DIKW links the above mentioned values and information into a development process, where raw data, collected from different data vendors to knowledge based result through two different analytical models, Excel and SQL-analysis, will found the result for the basic for a decision making.

The overall goal of the data mining process in this project shows, how to extract information from different geographical and non-geographical datasets and transform them into an understandable structure, for further use and decision-making and conclusions. The term data mining used in this project is – to some extent - a wrong or inaccurate name or designation, because the goal in this project is the extraction of patterns and knowledge from large amounts of data, not the extraction itself!!!

The actual data mining task is the automatic or semi-automatic analysis of large quantities of geographical data, to extract previously unknown interesting patterns, such as groups of data records, cluster analysis, unusual records, anomaly detections

and dependencies. These patterns can then be seen as a kind of summary of the input data, and may be used for further analysis or, for example, in machine learning and predictive analytics. For example, the step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results.

The Analysis in this project is structured so, that I will make a summary of the different variation of facts obtained by sort and classify my data and display them in graphic form. For each method of analysis is an illustrated analysis plan. That is to say that I beforehand outline, *how* I build my analysis, to achieve a result that I think can be the basis for answering my hypothesis. The performing of the Analysis is based on the principle of input - process – output. The schematic procedure for the analysis plan will be followed by a detailed description of the analysis process, with comments, illustrations and results. To the extent that I find it necessary, I will referred to various theories and methods in order to consolidate my choices or decisions to take further action.

The conclusion in my project is, that values and information in geographical data is a key to form a basis of a decision. Spatial data represented in the form of objects, lines and polygons, is not unique representatives of a result, but the basis for making a decision. It is the authenticity of data that determines the outcome of the test of my hypothesis, and thus the hypothesis is true or false. The true relationship between housing and car types can be found only by having the correct and original data on how people who own a VW Golf car live. There through, and only through that, it is possible to make an analysis of the link between housing and car types. Everything else is fiction, nonsense or not comparable. And from that point of view, *my hypothesis is rejected*. There is no relationship between housing and car types in my experimental area - PAKILA, on the basis of the evidence I have collected and analyzed.

Reporting is one of the cornerstones of business intelligence, where a new phenomenon called "Dashboards" currently gaining in popularity against management reports. "Dashboard" is a well-known interactive graphical tool for business users to access key figures, and can be used to better understand what has happened, when and why. By adding these Dashboards properties containing mapping options – visualization - you give companies the opportunity to integrate spatial analysis to a wider audience. No other reporting format can match a maps ability to concentrate information and compare multiple variables of one problem.

It is only your imagination and not yet developed techniques in information technology, which limits the implementation of our profession. I can only come up with this call to my fellow students - think interdisciplinary !!

3. Indholdsfortegnelse

3 INDHOLDSFORTEGNELSE

4. Indledning

Projektets struktur og indhold bygger på datafangst, dataforædling og dataforbedring. Den information, der skabes på denne baggrund, gør det muligt at dreje kort- og tabeldata ind til handlingsrettet information, gennem analyser og visualisering af mønstre og tendenser fra disse kort- og tabeldata. Gennem informationsforædling søger man altså strukturer og mønstre i data, som passer til projektidentitet og anvendelsesmodeller, der gør det muligt at træffe beslutninger og foretage en aktivitet. Det er her at GIS kommer til sit rette med hensyn til visualisering af kortdata og mønstre samt trends i datadistributioner hen over fysiske områder. Det er så op til vores dømmeevne og evnen til at anvende og fortolke de fundne resultater af en analyse, på en kvalitativ måde, der gør, om vi er succesfulde med vores evne til at evaluere et resultat.

4 INDLEDNING

4.1 Introduktion til projektet.

Det er velkendt, at interessen for produkter varierer på tværs af forskellige grupper af kunder. Disse interesser er direkte relaterede til kunders demografiske karakteristika, såsom (boligtype, uddannelse, beskæftigelse, indkomst, alder og etniske baggrund). Enorme mængder data - Big Data med modifikationer - med oplysninger om kunder og kundeadfærd, er i dag delvis tilgængelige for den almene borger, fra en række private og offentlige kilder, foruden naturligvis firmaers egne kunderegistre. Ved at tillægge oplysninger om kunders livstil (interesser, meninger, aktiviteter), går vi et skridt videre i nedbrydningen og forståelsen af et kundesegment. ***Figur 1*** på ***side 11*** illustrerer de vigtigste elementer, der for eksempel har indflydelse på valget af en biltype. Information om kunders livsstil er med til at påvirke, hvad de køber og hvor de køber, samt at se et mønster i hvilke grupper af folk, der står for disse indkøb (køn og alder). Med viden om folks livsstil har man én af grundstenene den til den første forståelse af folks interesse, meninger, aktiviteter og indkøbsvaner. Tendensen med *at samle sig om noget* er baseret på erkendelsen om, at "Krage søger mage", det vil sige at vi gerne vil omgås ligesindede. Har du nogensinde lagt mærke til, at boliger og biler i et bestemt beboelseskvarter, normalt er ens i størrelse og værdi? Hvis du kunne se ind i folks hjem, ville du finde mange af de samme produkter. Naboer har også en tendens til at deltage i lignende fritids-, sociale og kulturelle aktiviteter. Ved hjælp af veludviklede informationssystemer og høj datakvalitet, får vi mulighed for at forudsige en forbrugers adfærd. [9]

Mens GIS og Business Intelligence traditionelt har fokuseret på forskellige behov inden for forretningslivet, har man i den senere tid oplevet et større og større sammenfald mellem disse to teknologier, hvor *beliggenhedselementet* er blevet en fællesnævner således, at analytikere er begyndt at udvide deres opfattelse af forretningsstrategier og fordele og mangfoldighederne, ved at integrere georelaterede data med forretningsoplysninger. Samtidig er forbrugernes stigende anvendelse og til en vis grad forståelse af geografisk information, på grund af den stadig større udbredelse af de små elektroniske hjælpemidler (smartphones, tablets med mere), med til at øge virksomhedernes bevidsthed om anvendelsen og udnyttelsen af beliggenhed som en kommerciel faktor ved kortlægningen og manipulering af folks indkøbsvaner. Som en konsekvens af ovenstående, kunne en fusionering af GIS og Business Intelligence-teknologier være med til at skabe resultater, der rakte endnu længere end blot visuel repræsentation af data på et kort, nemlig Location Intelligence.

Et centralt element ved brugen af et databaseret system og kunderegister, er muligheden for at kunne skabe "ny" information ved hjælp af analyse (beregning) og sam-

menstilling (fusion) med andre data, for eksempel sammenstilling af forskellige registerdata. I alle tilfælde er problemstillingen, at virkeligheden (og ofte også databasen) er for stor at overskue og forstå. Hvis menneskets hjerne og bevidsthed var tilstrækkelig stor, var der ingen grund til at udføre informationsforædling, som for eksempel statistik, idet man så faktisk kunne overskue, gennemskue og forstå hele virkeligheden på én gang. Men da menneskets hjerne er for lille hertil, er det nødvendigt at foretage *informationsforædling*, for eksempel i form af matematiske analyser, for at finde *strukturer* og *mønstre* i de udvalgte dele af virkeligheden, som gør det muligt at *beslutte* og eventuel også *handle*. [22]

Gennem informationsforædling søger man altså strukturer og mønstre i data, som passer til projektidentitet og anvendelsesmodel, det vil sige gør det muligt for brugeren at finde den søgte betydning, altså det svar som muliggør beslutning og eventuelt handling i relation til projektet. Sammenstilling handler oftest om homogenisering af heterogene datasæt, det vil sige modelharmonisering. Der opstilles en hypotese (teori) for den handling, som ønskes udført. Ofte kan det være tilstrækkeligt at visualisere simple attributsammenhænge, og dermed give brugere overblik over de rumlige og tidslige variationer over udvalgte egenskaber. Andre gange er det nødvendigt med komplekse matematiske analyser, for at nå den ønskede indsigt i strukturer og mønstre. Erkendelse af "ny" viden opstår, når brugeren relaterer informationen til sin foreliggende viden og erkender, at den er forskellig; enten som udvidet viden eller som erstatning af tidligere opfattelser. [22]

Dette leder os langsomt hen mod undersøgelsestemaet eller "problemstillingen" om man vil, og den undren omkring ovennævnte kundeadfærd og informationsforædling, som har vagt min nysgerrighed for at starte dette projekt, og dermed det, som jeg er interesseret i at undersøge. Undersøgelsestemaet, vil tage sit udgangspunkt i integrering og anvendelse af produktdata, samt integreringen af geografiske og demografiske data, i forbindelse med bestemmelsen af sammenhængene mellem personers valg af biltyper og tilhørende boligforhold i et udvalgt boligområde i den nordlige del af Helsinki kommune, kaldet PAKILA. Ved at anvende og sammenligne den information, der findes i de geografiske data for dette område, vil jeg ved hjælp af statistiske værktøjer, undersøge og analysere disse data (tabeldata), og se om disse kan være med til af danne grundlaget for en beslutning, til for eksempel en udvidelse eller sammenligning af et kundesegment. Som empirisk undersøgelsesgrundlag, har jeg tænkt mig at anvende ca. 200 *rigtige adresser* på registrerede Volkswagen Golf biler i PAKILA og se, om jeg med dem som referenceobjekter eller punkter, kan udlede en trend eller se et mønster i de tilsvarende geografiske data (boligtyper) for produktion og distribution af viden og information i disse data.

Rumlige statistiske værktøjer kan hjælpe mig med at udføre disse opgaver, og andre opgaver, som allerede er gjort med kort og kortlægning. Med rumlig statistik åbnes et nyt sæt af spørgsmål, for at få endnu bedre information og være endnu mere sikker i vores forståelse og beslutninger: Hvor sikker er jeg på, at det mønster jeg ser, ikke blot beror på en tilfældig hændelse? I hvilket omfang er værdien af en funktion afhænger af værdierne af omgivende funktioner? Hvor godt vil værdien af én attribut forudsige værdien af en anden? Hvad er tendenserne i dataene? Statistik beskriver eller opsummere store mængder data, der er nyttige i geografisk analyse, hvor jeg ofte beskæftiger sig med store datasæt. [15]

Figur 1: *De vigtigste elementer der har indflydelse på valget af biltype. De **grønne** elementer er livstil betingede, de **gule** elementer er økonomisk betingede og de **røde** elementer er demografisk betingede.*

Hovedvægten i projektets analytiske del, har sit omdrejningspunkt omkring anvendelsen af nogle få simple statistiske redskaber, til at få meningsfulde resultater, snarere end på den matematiske teori bag statistikken. Der er mange rumlige statistiske redskaber og metoder til rådighed inden for GIS-verdenen i dag, mere end der beskrives i dette projekt. Dem der beskrives (Korrelation, Forhold), er dem, der er meget udbredte og gælder inden for GIS-analyse på tværs af en række discipliner.

4.2 Geografiske data - nøglen til et beslutningsgrundlag

GIS er omdrejningspunktet i forbindelsen med at bygge bro mellem virksomhedernes interne og eksterne data-aktiver. Vektor- og rasterbaserede topografiske kort, luftfotos, satellitbilleder samt statistik og andre demografiske data, er meget værdifulde variabler i forretningsoplysninger hvis de anvendes i analyser, sammen med de interne data i en virksomhed. GIS gør det også muligt at se efter fysiske omgivelser og områder, hvor der for eksempel kunne være potentielle bilejere. Bybilledet fortæller jo ofte en historie. Er det et ældre villakvarter der er ”gemt lidt væk” i et parkområdelignende bydel, eller et nyt højhusområde på et tidligere industriområde, der har fået en ny renæssance.

Ved at kigge byplanlæggeren over skulderen eller se lidt på by strukturere i det hele taget (klynger af boliger i forhold til infrastruktur, såsom vejnet, jernbaner og metro stationer), kan der allerede tidligt tegnes et visuelt billede af potentielle søgeområder. Det er her at GIS kommer til sit rette med hensyn til visualisering af kortdata og mønstre samt trends i datadistributioner hen over fysiske områder. For at identificere tendenser, mønstre og relationer, er det derfor vigtigt at forstå karakteristika og fordelinger af data og objekter i en analyse. Selv inden man begynder at analysere og drage konklusioner, kan det første indtryk af en datafordeling (både visuelt og talmæssigt) give et fingerpeg om et muligt resultat (specielle høje eller lave koncentration af objekter på bestemte steder, butikker, vejkryds, institutioner).

Men – vi skal altid huske på, at GIS-værktøjer KUN er et hjælpemiddel til at træffe beslutninger. GIS er ikke et Orakel eller kilden til visdom, der kan give entydige svar på vores spørgsmål om bedste/dårligste beliggenhed af et objekt. Det er i sidste ende os selv, altså brugeren eller analytikeren, der må drage konklusionerne på basis af de data og informationer, som vi har ”fodret” vores GIS-systemer med. Resultaterne er et produkt af hvilke modeller og algoritmer, vi har valgt at anvende for at løse en opgave eller få svar på et spørgsmål eller undren. En fortsat udvikling af disse modeller og algoritmer og en stadig større integration mellem forskellige informationssystemer inde for GIS og Business Intelligence, er med til at give et mere nuanceret beslutningsgrundlag og – efter min mening – et mere *rigtigt billede* af en situation eller hændelse i det *virkelige liv*, men her er det igen beroende på ægtheden af de data, der er til rådighed på et givet tidspunkt, samt kvaliteten af disse, der har betydning for det endelige resultat.

4.3 Mit projekts idé.

Datagrundlaget for mit forsøgsområde forandres jo konstant – om end kun periferisk i den periode hvor mit projekt er blevet til. Folk flytter, bliver skilte eller dør, der bygges nye huse, og andre rives ned. Infrastrukturen forandres også over tid. Resultaterne af

min undersøgelse er kun et øjebliksbillede. I morgen ville samme undersøgelse give et andet resultat, - dog sandsynligvis meget nær det første, alt efter hvad det er jeg undersøger. Og sådan er det med alle undersøgelser inden for til eksempel videnskaber, erhvervslivet, undervisningsområdet og også GIS. Dette fænomen vil jeg diskutere senere i dette projekt i ***Kapitel 10*** om ***Perspektivering***.

Det er så op til vores dømmekraft og evnen til at anvende og fortolke de fundne resultater af en analyse, på en kvalitativ måde der beviser, om vi er succesfulde med vores evne til at evaluere et resultat. Den menneskelige faktor spiller altid ind og *skal* spille ind. Ved blindt at stole på resultaterne fra en beregning fra en GIS-analyse, kan det gå grueligt galt (beregning og visualisering af potentielle oversvømmelsesområder og deraf tegning af forsikring via forsikringsselskaber). Resultatet/resultaterne er kun vejledende, visuelt, numerisk eller grafiske. Det er på basis af disse resultater at vi skal forsøge at drager vores konklusioner og beslutninger – på bedste vis.

Endelig er det et sidste aspekt, der er vigtigt at huske på ved anvendelsen af GIS og GIS-data. Det er ikke bydende nødvendigt at have en dybere baggrundsviden om GIS som et redskab til at blotlægge information, som basis for et beslutningsgrundlag for fremtidige handlinger eller investeringer. Mit eksempel med Excel-analysen og den simple anvendelse og manipulation af boligdata og bildata i ***Kapitel 8*** om ***Analyse***, hvor fællesnævneren er adressen, viser at selv almindelige kontoruddannede mennesker, der til daglig arbejder med for eksempel Microsoft Office programmer, har de første basale redskaber lige ved hånden og med en smule fingersnilde, og lidt kreativ sans, kan hente mange og alsidige mængder information fra disse frie databaser, til brug for en første eller bredere understøttelse i en beslutningsproces. For eksempel har ESRI udviklet et kortlægningssystem, der mod en mindre betaling, der kan tilkobles Excel-programmet og derigennem kortlægge tabeldata, som et første og relativt simpel visuelt og grafisk grundlag og fingerpeg i en beslutningsproces og rapportering. [23]

4.4 Projektets struktur og indhold.

Projektets struktur og indhold bygger på datafangst, dataforædling og dataforbedring. Den information, der skabes på denne baggrund, gør det muligt at dreje kort- og tabeldata ind til handlingsrettet information, gennem analyser og visualisering af mønstre og tendenser fra disse kort- og tabeldata. Næsten hvert kapitel i projektet kører efter princippet om Input – Proces – Output. Derved opnås en forenkling af tankegangen med hensyn til besvarelsen af min hypotese. Denne findes i ***Kapitel 5*** om ***Problemanalyse***. Ved at benytte denne fremgangsmåde, bliver det i en tidlig fase i projektet illustreret, hvordan mit datamateriale vil blive behandlet og hvilke output (resultater),

den eller disse aktiviteter *kan* medfører. Der er således hele tiden fokus på hypotesen og det undersøgelsestema, der er gennemgående i dette projekt.

Figur 2: *Grundprincippet i de enkelte kapitlers opbygning. Til hvert input (data og/eller information) i en proces (beskrivelse/forklaring/forsøg), kommer der et output (resultat eller ny information)* [25]

Projektet består af 13 kapitler, hvoraf de 7 er hovedkapitler. Hovedkapitlerne i dette projekt er: Indledning, Problemanalyse, Teori og Metode, Empiriske materiale, Analyse samt en Konklusion. Efter konklusionen har jeg indlagt et kapitel om Perspektivering. Her ytre jeg mine egne meninger og overvejelser om, hvordan jeg ser anvendelsen af Location Intelligence, Business Intelligence og GIS, som problemløsere eller værktøj inden for informationsteknologien og som en slag brobygger, der efter min mening, kunne have afgørende betydning for en bæredygtig udvikling inden for dette område, der på sigt, stadig er i sin spæde begyndelse.

Det teoretiske kapitel i projektet beskriver relationerne mellem Data, Information, Viden og til tider Visdom i en hierarkisk ordning, der har været en del af sproget i informationsvidenskaben. DIKW-modellen (Data – Information – Knowledge - Wisdom) hierarkiet baserer sig på filtrering, reducering og transformering af data og er illustreret i form af en pyramide, der består af fire niveauer. Progressionen foregår i al sin enkelthed ved, at man bevæger dig gennem hvert niveau af pyramiden og når toppen.

Det empiriske kapitel i projektet beskriver det kort- og tabelmaterialet - input, der danner grundlag for dette projekts undersøgelsestema og hvordan det vil blive anvendt i kapitlet om analyse. Endvidere gives der en grundigere gennemgang af materialernes oprindeles, kvalitet, anvendelsesmuligheder og eventuelle fejlkilder, der måtte være tilknyttet eller forbundet med de anvendte data.

Den analytiske del af projektet - processering - og dermed metoderne til at lave udtræk fra kort- og tabeldata, har sit omdrejningspunkt omkring to metoder. Excel-manipulation af tabeldata med hjælp af Microsoft Excel, og SQL-analyse af tabeldata i MapInfo. Structured Query Language eller SQL er det mest udbredte programmeringssprog til relationelle databaser (tabeldata). Sproget dækker både definition af databaser og manipulation af data. SQL er et deklarativt programmeringssprog, hvilket vil sige, at man kun beskriver, *hvad* der skal ske, men ikke *hvordan* det sker. Altså - SQL bringer information på et sådant niveau eller stade, at det er muligt at udlede en viden om en begivenhed ud fra det, der er spurgt om. En tredje metode til analyse af mine data var

at anvende en Prædiktiv analyse, altså at forsøge at fremskrive en hændelse. Med Prædiktiv analyse er det mulighed at "lærer" programmet (i dette tilfælde MapInfo), hvordan vi klassificer et område, på grundlag af kriterier for indlæsning af data. Desværre har jeg ikke kunnet få programmet til at virke efter hensigten. Metoden *skulle* have undersøger de statistiske karakteristika for flere kvadratnet til indlæsning af data, der er fundet inden for et given "reference/trænings" område og derefter lokaliserer andre områder med lignende egenskaber. Ideen var at beregne et datasæt, der havde et mønster, der kunne genkendes andre steder i mit kort og dermed argumenterer for, at der var en korrelation mellem mine datasæt og dermed et mønster, som i sin tur kunne betragtes som en viden.

Konklusionen i projektet sammenfatter de forskellige teoretiske, empiriske og analytiske metoder der er anvendt, samt de resultater der er opnået, for at få svar på min hypotese. Foruden at afslutte en ring af undersøgelser, var hensigten med konklusionen også at starte en diskussion i kapitlet om perspektivering, omkring fremtidige problemområder inden for informationsteknologien med særlig henblik på GIS-områder, primært set i relation til mit projekt, men også et kik over hegnet og de mere almene trender og nuancer i de teknologier, der er omfattet af informationsvidenskab.

Endelig findes der en litteraturliste med relevant undersøgt litteratur og information fra bøger, artikler og hjemmesider, der har gjort det muligt – for mit vedkommende – at danne mig et indblik i og forståelse af det problemfelt, som jeg har ønsket at undersøge.

Bilagsinformation findes som sidste kapitel i dette projekt. I kapitlet med bilag, er der hensat sådanne informationer og illustrationer, der nok har relevans for projektet, men ikke har kunnet finde vej som selvstændige afsnit i kapitlerne som sådan. Bilagene dokumenterer således en ekstra information til læseren, der har en mere overordnet funktion til dette projekt.

4.5 Hvad kan vi så forvente os efter det her?

Som tidligere nævnt i indledningen, er der i dag en stigende tendens til at tillægge geografiske beliggenhedsdata til oplysninger om kunder og kortlægge (visualisere) disse og dermed få en helt ny indgangsvinkel og viden omkring registrerede kundedata. Geografisk analyse og visualisering af kundedata, giver en helt ny sammenhæng og forståelse af kundesegmenter, der ikke tidligere har kunnet lade sig illustrerer gennem tabeller og grafer. Jeg håber med dette projekt og de næste kapitler, at kunne give læseren en indføring i og ide om det, som jeg selv interessere mig for og via den skrevne tekst og illustrationer, forhåbentlig på en overbevisende måde, kan formidle og im-

plementerer denne vision og mission, så det bliver mere end bare en almindelig gang kaffebordssnak.

Med dette indledningskapitel har jeg nu sat fokus på de hovedemner, der er aktuelle for at kunne opstille min hypotese og det undersøgelsestema, som har haft min interesse og vakt min undren, og har været med til at danne grundlaget for hele dette projekt. Problemanalysen i næste kapitel har til formål at kaste lys over, hvilken teoretiske og analytiske undersøgelsesmetoder jeg vil anvende i min fremgangsmåder i dette projekt, for at kunne få svar på min hypotese. Problemanalysen er således en af grundstenen i projektet.

Dette kapitel sætter fokus på min hypotese og de forsøgsbeskrivelser og overvejelser, der har ført til den endelig udformning af dette projekt. Kapitlet vil vise hvordan min hypotese er blevet til, dels via overvejelser fra indledningen, men også rent metodemæssigt, set i lyset af de numeriske, tekniske og informationsmæssige redskaber, der har været til rådighed for dette projekt. Input - process - output principperne danner grundlaget for opbygningen af de enkelte kapitler og projektmodellen der anvendes, samt de overvejdelser jeg har gjort mig omkring disse udforminger i dette projekt, for at være i stand til at svare på min hypotese.

5. Problemanalyse

5 PROBLEMANALYSE

5.1 Indledning.

Kapitlet om Problemanalyse har til formål at vise, hvordan jeg opbygger mit projekt for at få svar på min hypotese. Jeg har længe gået og "GIS-agtigt" funderet på, om der er en sammenhæng mellem personer der ejer et bestemt bilmærke og deres boligforhold – altså, hvor (sted) og hvordan (boligtype) bor folk, der for eksempel ejer og/eller kører i en Volkswagen Golf, *herefter VW-Golf*, (er der en sammenhæng mellem bilmærke og boligtype?). Interessen for dette fænomen skal dels ses i lyset af at min kone, der arbejder i Volkswagen Finland, ofte har diskuteret forskellige marktingsstrategier og måder, hvorpå hendes firma kunne finde nye kundepotentialer for deres bilprodukter. VW-Golf er det mest solgte bilmærke i Finland og Volkswagen som bilmærke i det hele taget, har en markedsandel i Finland på ca. 12 % (2013). Og - dels skal interessen ses i en avisartikel i Jyllandsposten fra 2009, (*Hvilken bil er du?),* hvor lektor fra Markedshøjskolen i Oslo, Karl Fredrik Tangen, og bilanalytiker i TNS Gallup, Anders G. Hovde, giver deres bud på, hvilke typer købere, der står bag de forskellige bilmærker, der findes i trafikken i dag. ***Bilag B2***. Sammenhængen mellem bilmærke og boligforhold er måske ikke specielt interessant i en større sammenhæng, men princippet for at undersøge sådanne sammenhænge er interessante, fordi det bør kunne ekstrapoleres til mange andre temaer.

Endelig er der den geografiske interesse som af "naturlige" grunde falder tættere på min interessesfære, grundet uddannelse og arbejde, men også en generel interesse for geografiske fænomener og sammenhænge samt udviklinger på det historiske område i det hele taget – i mit tilfælde byudvikling. Ved at observere dagligdags hændelser og adfærd i og omkring det livsrum hvor vi trives, dannes der ofte et karakteristikum af bestemte typer af orden og systematiske foranstaltninger. Er det et ældre villakvarter der er "gemt lidt væk" i et parkområde-lignende bydel, eller et nyt højhusområde på et tidligere industriområde, der har fået en ny renæssance, eller er det bare billedet af biltyper på en parkeringsplads i et bestemt boligtype-område, (villakvarter, højhuskompleks). Mange gange findes svaret ikke i indsamlede data men i de naturlige omgivelser vi bevæger os rundt i til daglig.

5.2 Hypotesen.

"Hypotese: Jeg tror, at jeg ved hjælp af GIS-analyser kan påvise, at der er en sammenhæng A mellem personers valg af biltype og tilhørende boligforhold, ved at anvende og sammenligne den information, der findes i geografiske data".

Jeg har skrevet "sammenhæng A", fordi jeg ikke ved nok om statistik og trends, for at kunne skrive helt præcist, hvilken sammenhængstyper og -kvaliteter, der skal vises, for at resultatet er signifikant.

Resultatet af undersøgelsen af hypotesen forventer jeg kan bruges til at bevise, at det kun er *ægtheden* i data (kundedata og geografiske data), der kan give GIS den funktion og placering i forretningsverdenen, som den nye partner og ”arbejdshest” som den er fuldt berettiget til i dag, i formidlingen af information og viden, i det informations-samfund, med myriader af tilbud om teknologiløsninger, i forbindelse med optime-ringer og tilvejebringelse af nye forretningspotentialer.

For at kunne få svar på min hypotese, vil jeg gennemføre nogle forsøg – analyser – af en række relevante datasæt som jeg er kommet i besiddelse af, der indeholder infor-mation om biler og boliger. Som nævnt i Indledningen til dette kapitel har jeg en rund tid gået og funderet over, om der var en egentlig sammenhæng mellem menneskers valg af biltype og boligtype (parcelhus, rækkehuse eller højhus)? For at få svar på dette spørgsmål, har jeg tænkt mig at lave en række forsøg – analyser, for at se om der er hold i min hypotese. Forsøgene går i al sin enkelthed ud på at lave en sammenligning mellem forskellige boligtyper og bilmærket VW-Golf, og se om der er en korrelation mellem visse boligtyper og VW-Golf eller ej. Og – er det muligt på den baggrund, at lave en analyse af denne antagelse og videreudvikle min “hypotese” således, at denne eventuelle sammenhæng, kunne bruges, til at finde og fremskrive nye potentialer for kunder til et bestemt bilmærke?

Som udgangspunkt for at kunne svarer på min hypotese, havde jeg tænkt mig at anvende ”rigtige” data om personer der ejede en VW-Golf. Det vil sig – adressedata på den enkelte bilejer, samt hvilken biltype (VW-Golf Hatchback eller VW-Golf Variant) denne person var indehaver af. Ved en simpel kombination af adressen på bilejeren og den boligtype han besidder, ville det være relativt enkelt at finde andre lignende karakteristika for tilsvarende bilejere og boligtyper. Desværre har det ikke været muligt at komme i besiddelse af disse ”rigtige” data, til alle mine input i mine analyser, fordi det efter en del af de forskellige dataleverandørens mening, vil komme for tæt på de enkelte personers ”privatsfære”. Dette har naturligvis ikke været hensigten med testen af min hypotese – slet ikke. Derfor har jeg været nødt til at anvende ”fiktive” data som input i en del af dette projekt. Det er specielt data omkring ejerforhold af biler, økonomiske data og til en vis grad demografiske data, der er blevet til, efter egen idé og kreativitet. Her vejer begrebet om integritet altså højere, forstået på den måde, at kvaliteten af at have en følelse af ærlighed og sandfærdighed i forhold til motiverne for egne handlinger, vejere mere end et ”sandfærdigt” resultat.

Det betyder helt enkelt, at det eller de resultater der bliver producerret i min analyse, for at tilfredsstille min undren og dermed hypotese, ikke har noget at gøre med virkeligheden. Data og information der anvendes i dette projekt som in- og output, får derfor kun status som elementer til manipulation, for at vise, hvordan forskellige analysemetoder kan anvendes til at ekstrahere information, numerisk, visuelt og grafisk som grundlag for en senere beslutningsproces.

5.3 Den overordnede forsøgsbeskrivelse.

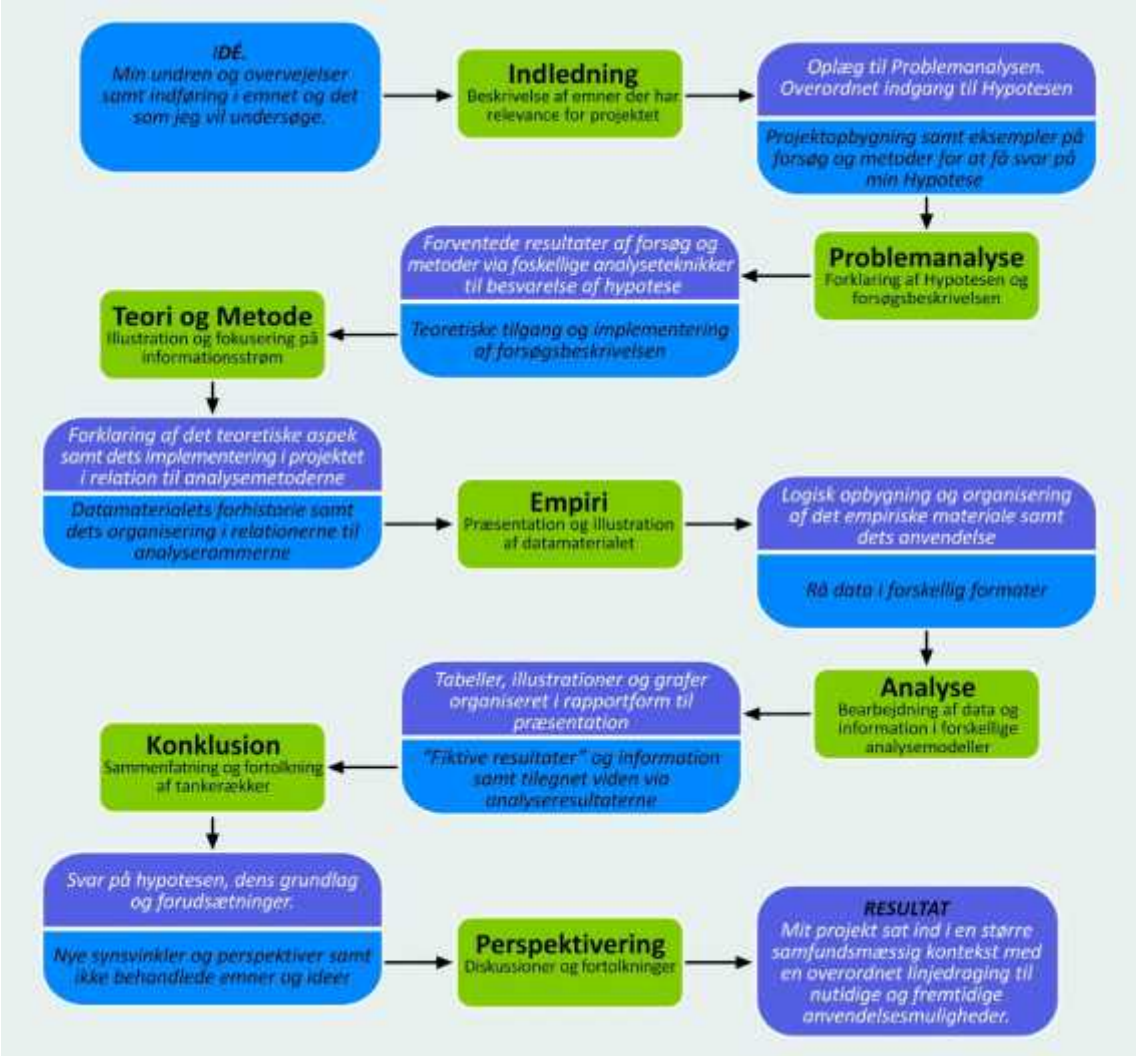
Jeg har valgt at sætte forsøgsbeskrivelsen helt frem i mit projekt, for at i et så tidligt stadie som muligt, at indføre læseren i det/de forsøg/analyser jeg har tænkt mig at foretage, for at kunne få svar på min hypotese. Forsøgsbeskrivelsen er derfor ikke en dybdegående og detaljeret beskrivelse og gennemgang af hver enkelt metode der anvendes i analyserne. Disse findes i ***Kapitel 8*** om ***Analyse***. Forsøgsbeskrivelsen har til formål, at vise og forklare læseren, hvordan min fremgangsmåde med input – proces – output filosofien foregår som beskrevet i indledningen, og dermed tydeliggøre samspillet mellem de enkelte processer og de tilstande (input/output data og information) der har betydning for besvarelsen af min hypotese.

Forsøgsbeskrivelsen vil i princippet blive ”opdelt” i to afsnit. Et afsnit, der indfører læseren i den overordnede tankegang og idé med projektet og hvordan det er bygget op, for at få svar på min hypotese, samt den mere specifikke forsøgsbeskrivelse, der kort fortæller om ideen bag de enkelte analysemetoder (Excel, SQL, og Prædiktiv analyse) og hvordan jeg har tænkt mig at sætte dem i spil med input-data, proces og output-data. Enkelt kan det siges således:

”Materialet i projektet er DATA. Data bliver forarbejdet ved hjælp af redskaber (Excel og MapInfo/Vertical Mapper) i mit tilfælde ANALYSER. Det forarbejde produkt er RESULTATET. Resultatet kan være et slutprodukt eller et delprodukt, der igen skal videreforarbejder (forædles) til et endnu finere produkt (specifikt resultat), der så igen kan være et slutprodukt eller et delprodukt for videre forarbejdning og så videre. Processen er iterativ indtil resultatet eller resultaterne, kan opstilles i forhold til hypotesen og bekræfte eller afkræfte mit undersøgelsesudsagn”.

En uddybende forklaring og beskrivelse af de anvendte metoder, teoretiske grundlag samt den endelige tilgang til analysen, fremgår i de afsnit hvor begreberne og de metodiske tilgange sættes i spil (grønne proces-bokse). ***Figur 3*** viser en illustrativ opbygning af specialet med input-proces-output metoden som drivkraft. En uddybende forklaring af hver enkel proces, findes som tekstafsnit efter denne illustration.

Strukturen i den overordnede forsøgsopbygning er den måde og metode, som jeg har anset for at være den mest dækkende og enkleste metode, at komme frem til et resultat, der kan give svar på min hypotese. Strukturen i den overordnede forsøgsbeskrivelse, har altså ikke noget at gøre med selve kvaliteten (sandheden) af de resultater, som jeg håber på at udtrække fra mine analysemetoder. Dog er indhold og emner der vil blive behandlet i projektet nøje overvejet i relation til det for eksempel undersøgelsestemaet (projektets titel) og det faktum, at jeg har måttet arbejde med "fiktive data" i en del af de input der har dannet grundlag for analyserne.



Figur 3: *Forsøgsopbygning illustreret via input-proces-output metoden. De* ***lyseblå bokse*** *illustrerer materialer for input. De* ***grønne bokse*** *er processen eller metoden der beskriver selve udførelsen. Endelig findes der et output –* ***violette bokse****. Disse output tjener igen som en del af det nye input det næste kapitel og så videre.* [25]

Projektet viser således mere hvilke rammer og tilgange jeg arbejder indenfor, for at kommer frem til et svar på min hypotese. Det er der ikke noget forkert i. Men - nu bliver projektet og anvendelsen af tabeldata som input mere en slags "leg" med data.

Ved at have anvendt originale eller ægte data om bilen, det vil sige den "rigtige" adresse hvor VW-Golf bilerne findes, ville jeg have været i stand til at lade data fortælle historien fra starten af min undersøgelses og dermed lade hele undersøgelsesprocessen gå i en slags selvsving. Hvert resultat af en beregningsproces ville have været til at føle på med det samme og dermed udlede en delkonklusion for hvert output, der var produceret, fordi de opnåede resultater havde en sandhedsværdi (ægthed).

Kapitel 4. Indledning
Indledningen indeholder en præsentation af nogle af de hovedemner, som har været med til at inspirerer og give mig ideen til dette projekt og de emner, som jeg har tænkt mig at bygge min rapport op omkring. Indledningen giver et lidt "større billede" af, hvordan GIS og for eksempel Business Intelligence har fundet en slags synergi inden for forretningsverdenen og hvilke fællesnævnere i de to IT-værktøjer, for eksempel inden for analyse og visualisering, der er med til at give overblik over data med en geografisk relation eller reference. Nogle teoretiske og metodiske overvejelser i informationsoverførsel sættes i relation til projektet og undersøgelsestemaet.

Kapitel 5. Problemanalyse
Dette kapitel sætter fokus på min hypotese og forsøgsbeskrivelser, samt de overvejelser der har ført til den endelig udformning af dette projekt. Afsnittet vil vise hvordan min hypotese er blevet til via dels overvejelser fra indledningen, men også rent metodemæssigt, set i lyset af de numeriske informationer og tekniske redskaber, der har været til rådighed for dette projekt. Forsøgsbeskrivelsen i dette afsnit bliver hold på et "overordnet plan", forstået på den måde, at den kort beskriver projektets opbygning via input – proces - output ideen og hvilke elementer der bidrager til de forskellige processer og tilstande. Der gives også en kort, men om end mere detaljeret beskrivelse af de to analysemetoder, der anvendes i dette projekt, for at bearbejde mine data til besvarelse af hypotesen. Den egentlige empiriske undersøgelsesbeskrivelse, design og implementering af datamaterialet findes i ***Kapitel 7*** om ***Empiriske materiale*** og ***Kapitel 8*** om ***Analyse.***

Kapitel 6. Teori og Metode
Som teoretisk referenceramme anvendes DIKW-hierarkiet med Russell Ackoff og hans teori omkring relationerne mellem Data, Information, Viden og Visdom. DIKW- hierarkiet anvendes i dette speciale til at fokuserer på, hvorledes det er muligt, at forstå grundlæggende ændringer i informationsstrømmen, som anvendelsen af data sætter i gang i forbindelse med implementering af hvad, hvor, hvem spørgsmål. DIKW- hierarkiet beskæftiger sig med centrale begreber som Data, Information og Viden, hvorfor disse begreber vil være gennemgående i dette speciale. Teorien sætter fokus på DIKW-hierarkiet grundlag for viden-transformation fra rå data til anvendelig information som

beslutningsgrundlag til at opnå en viden om en ting eller et produkt. Visdom (læs forståelse) som det 4. element i DIKW-hierarkiet, danner grundlag for den Prædiktive analyses ide i dette projekt. Som tidligere nævnt vil der kun blive givet en *beskrivelse* af hvordan denne analyse metode fungere, da det ikke har været muligt af udføre et praktisk forsøg med denne metode i MapInfo.

Metodeafsnittet beskriver specialets arbejdsidé, samt hvilke metoder der konkret anvendes til at besvare hypotesen. Der arbejdes med en empirisk undersøgelse, der har til hensigt, at konkretisere hypotesen og forsøgsbeskrivelsen i relation til forskellige indsamlede data på tabelform (produkt, geografi, demografi og økonomi). Teori og metode afsnittet barsler således ikke med noget egentlig resultat eller produkt, der kan videreføres som et nyt direkte input i Empirien. Det er mere de filosofiske overvejelser omkring DIKW-hierarkiet i relation til min hypotese, og anvendelsen af information generelt på forskellige niveauer i mine analyser og overvejelser om brugen af denne information, der skulle klargøres med dette afsnit.

Kapitel 7. Empiriske materiale
I dette afsnit sættes fokus på specialets undersøgelsesmateriale samt de angivne analyserammer. Den empiriske undersøgelse i dette projekt tager sit udgangspunkt i et boligområde i Helsingfors, nærmere betegnet PAKILA, med tilhørende geografiske data. Formålet med det Empiriske studie har været at indsamle, organisere og evaluere forskellige datatyper (geografiske, økonomiske og demografiske), for det boligområde, der danner den fysiske ramme om mine analyser.

Kapitel 8. Analysen
Specialets indsamlede empiriske grundlag analyseres i relation til de fremkomne fællesnævnere i forbindelse med dataforbedring, dataintegrering og datamanipulering. Selve analysen af de indsamlede data vil blive implementeret gennem to analysemetoder: Excel analyse og SQL-analyse. Endvidere diskuteres empirien i forhold til specialets teoretiske referenceramme samt problemstilling. Datakvalitet og datafangst vil blive sat i relation til denne diskussion. Fejlkilder i data og deres betydning igennem analyseprocessen vil også blive behandlet.

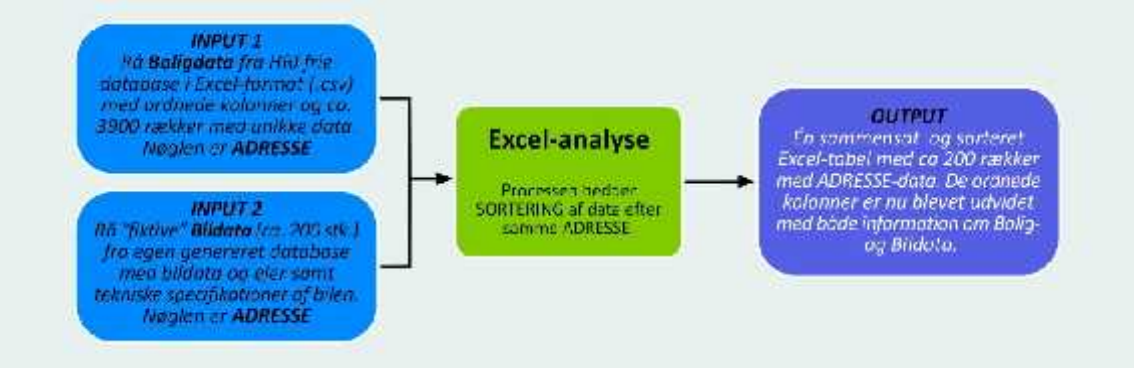
Kapitel 9 og 10. Konklusion og perspektivering
I dette afsnit fremkommer en samlet konklusion samt diskussion af undersøgelsens foreskrivende overvejelser, meninger og de evt. nyopståede perspektiver brugen af DIKW-hierarkiet som grundsten i besvarelsen af min hypotese. SQL-analyse og Prædiktiv analyse som metode til at opnå et resultat vil blive konkluderet og sat relation til DIKW-hierarkiet som et anvendeligt redskab til afdækning af et kundesegment i forsøgsområdet. I perspektiverings-afsnittet vil der blive diskuteret nutidens anvendelse af Hard- og software til brug ved Location Intelligence og den indflydelse dette kan/vil

have på fremtidens forståelse af Location Intelligence. Endelig vil emner, tanker og ideer som det ikke har været muligt at tage med i dette projekt, på grund af afgrænsninger, eller ikke direkte relevans til dette projekts undersøgelsesdesign, blive overfladisk behandlet og kommenteret for fremtidige anvendelsesmuligheder.

5.4 Den specifikke forsøgsbeskrivelse.

Som nævnt i indledningen har jeg valgt at fokusere mine analyser på 3 metoder. Den Prædiktive analysemetode, vil som så mange gange tidligere nævnt i dette projekt og kapitel, *ikke blive anvendt* i praksis, da analyse-forsøget i MapInfo ikke lykkedes. Det er ikke for at få et mere nuanceret billede eller resultat af mine forsøg, men kun for at vise hvordan jeg opfatter og tror, at relativt simple analysemetoder kan fortælle historien i data og dermed visualisere (numerisk, grafisk eller illustrativt) data i tabelform.

Den første analysemetode hedder **Excel-analysen**. Den bygger i al sin enkelthed på sammensætning og sortering af to tabeldata. Ved at kombinere to tabeller (Bolig- og Bildata), genereres en sammenhængende tabel med fælles adresse for de to produkter eller temaer. Tabellen med Boligdata indeholder mange interessante oplysninger om for eksempel, boligtyper (parcelhus, rækkehus og højhus), boligstørrelse (boligareal, etager). Sammenlagt findes der 112 egenskaber (kolonner) for en bolig i denne fil.



Figur 4: *Hovedprincippet i **Excel-analysen**. Analysemetoden anvender to datasæt Bolig- og Bildata. Disse kombineres til én samlet Excel-tabel med adressen som fællesnævner. Antallet af kolonner i Excel-tabellen er summen af Bolig- og Bildata. Output-data skal viderebehandles*

Ved at kombinere de to tabeller kan vi se, hvilke boligtyper der har flest/færrest VW-Golf biler og hvilke boligtyper, der har størst eller mindst boligareal og eller antal etager. Output resultatet danner grundlaget for et nyt input i en iterativ proces, der danner basis for et endeligt resultat til verificering af min hypotese. Excel-analysen producere for det meste numeriske data (tabeldata), men der findes også muligheder i Excel-programmellet, til at grafisk fremstiller en del af resultaterne via for eksempel

søjle-, lagkage- eller linjediagrammer. (***Figurerne 19 - 22*** er eksempler på sådanne grafiske fremstillinger).

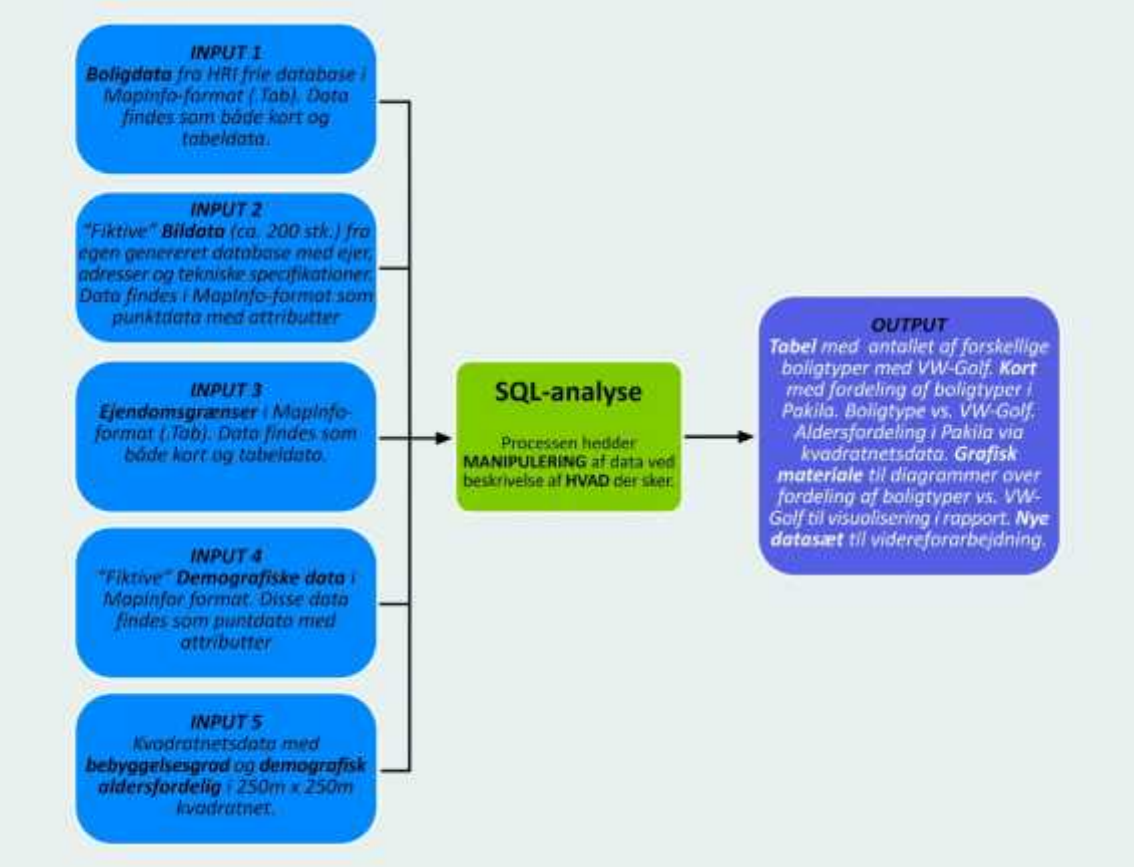
Selve databeskrivelsen og indholdet i det anvendte tabelmateriale, samt den detaljerede analyse og resultater, findes i ***Kapitlerne 7*** og ***8*** om ***Empiriske materiale*** og ***Analyse.***

Det er med udgangspunkt i ovennævnte tekst, at vi kan drage de første "simple konklusioner" med hensyn til at finde lignende karakteristika i vores tabelmateriale. Alle boligtyper der ikke har en VW-Golf, men som falder inden for samme kategori (Boligtype eller boligstørrelse) må antages at være en potentiel kunde. Dette lyder næsten for simpelt, men taget i betragtning af, at dette er en metode for personer der ikke arbejder med GIS og GIS-metoder til hverdag, er det en ekstra værdi og informations, der kan anvendes i et beslutningsgrundlag. Og - jeg har mit første resultat til at teste min hypotese. I ***Kapitel 8*** om ***Analysen***, vil dette blive uddybet og beskrevet mere systematisk. Ovennævnte ***Figur 4*** viser princippet i opbygningen af analysen og hvilke komponenter, der danner grundstenene i analysen.

Med **SQL-analysen** bevæger vi os ind i den forunderlige GIS-verden. Nu er det ikke kun "almindeligt kontorredskaber" der anvendes for at analysere vores data, men GIS-orienteret software og programmel, der er specielt udviklet til at håndtere et bredt spekter af dataindsamling, data-editering, data-analysering og datavisualisering samt kortlægning i mere eller mindre ønskværdig kvalitet.

Structured Query Language eller SQL er et programmeringssprog til relationelle databaser (tabeldata). SQL er et deklarativt programmeringssprog, hvilket vil sige, at man kun beskriver, *hvad* der skal ske, men *ikke hvordan* det sker. Altså - SQL bringer information på et sådant niveau eller stade, at det er muligt at udlede en viden om en begivenhed ud fra det, det er spurgt om.

Ideen med SQL-analysen er at vise, hvordan et GIS-system kan anvendes til at manipulere mine geografiske data, for at få svar på min hypotese. SQL-analysen, kan producere output resultater i numerisk, grafisk og illustrativ form. Her skal det allerede nu bemærkes, at kombinationen af forskellige tabeldata med ordnede (klassificerede) kolonner (biler, boliger, ejendomsgrænser, demografi og økonomi) er mange – rigtig mange. Der kan laves et utal af udtræk fra disse tabeller, alt efter hvad det er vi er på jagt efter. I dette tilfælde svaret på min hypotese.



Figur 5: *Hovedprincippet i* ***SQL-analysen****. Analysemetoden anvender flere datasæt såsom Boligdata, Bildata, Matrikeldata og Demografiske data i Kvadratnet. Trend- og økonomiske data har ikke kunnet finde vejen til denne analyse. Denne form for data beskrives under Empiriafsnittet.*

Tricket er at opbygge SQL-sætninger, der tilfredsstiller det vi søger efter. Dette er ikke en enkelt og ligetil proces. Der findes mange måder at ”spørge på”. Der findes ingen forkerte eller dumme måde at spørge på, men opbygningen og udformningen af SQL-sætninger *kan* lede til ikke ønskede eller mangelfulde resultater. SQL-analysen i dette projekt er en iterativ proces lige fra Input til proces til output. Som vist i ***Figur 5***, bruges der mange forskellige inputdata for at få et output. Alle inputdata anvendes ikke på samme tid men manipuleres successivt, alt efter hvilke oplysninger det er, som jeg er interesseret i at hente, for om muligt, at kunne få svar på hypotesen. Det skal naturligvis ikke forstås sådan, at jeg bliver ved med at foretage analyser via spørgsmål (SQL) indtil jeg får det svar, som *jeg vil have*, for at tilfredsstiller min hypotese.

Eksempel*:* Inputdata er tabellen med Bildata (ca. 200 stk.) indlæses i MapInfo. Data vises som objekter. Derefter indlæses tabellen med ejendomsgrænser (ca. 3.800 stk.) i kortform (baggrundsinformation). Ligeledes indlæses tabeldata med Boliger (ca. 3.920 stk.) som kortdata (illustreret som fysiske objekter). Derefter skrives en SQL-sætning *processen*, der ønsker at sortere alle bildata efter boligtyper (parcelhus, rækkehus og højhus). Resultatet (output) vises som et kort med ejendomsgrænser som baggrund

med tilhørende bygninger og VW-Golf objektet på den tilhørende adresse. Derefter laves en ny SQL-analyse, hvor de frembragte outputdata fra første SQL-undersøgelse manipuleres således, at for eksempel alle andre bygninger i PAKILA (ca. 3.920 stk.), der opfylder kriteriet om (parcelhuse på mellem 157m² – 224m²) = det spektrum hvor VW-Golf findes fra første analyse. Det nye output fra denne undersøgelse projiceres på kortet. Analysen viser, at det findes 367 steder i PAKILA, hvor der er parcelhuse på mellem 157m² – 224m².

Dette output er i sig selv ikke noget færdigt eller enestående resultat, men skal bare dokumentere, at ved at kombinere de forskellige tabeldatas unikke egenskaber og information, ses der lidt efter lidt et mønster (historien) i de anvendte data. Disse mønstre kan visualiseres via kort, grafer eller nye tabeller. Ved at graduere disse output data med forskellige symboler og farver alt efter værdi, og anvende et kort *med nedtonede farver* som baggrundsreference, kan dette delresultat sættes ind i en rapport eller PowerPoint og fremvises for chefen, kunden eller hvem der nu måtte have interesse. De "tørre" data begynder at fortælle historien. Nu er det bare op til fortælleren at gøre den så spændende som mulig. Det er her at GIS og Business Intelligence forenes – og derigennem er med til at give svar på min hypotese. I ***Kapitel 10*** om ***Perspektivering***, findes der en rapport, der dokumenterer denne synergi og fuldfører ***Figur 13*** intentioner om et grafisk og numerisk resultat af en SQL-analyse af mine data. Som tidligere nævnt under gennemgangen af Excel-analysen, forgår hele den detaljerede undersøgelse af mine data med SQL, i ***Kapitel 9*** om ***Analyse***. Her vil flere forsøg og aspekter, der har direkte relation til besvarelsen af hypotesen, blive testet og visualiseret.

Prædiktiv analyse er taget med i dette projekt i en beskrivende form, som en analyse-metode for at vise, hvordan fremskrivning af data kan hjælpe med til at danne et beslutningsgrundlag, via værdier og information gemt i Geografiske data og naturligvis hjælpe med at kaste lys over min hypotese og dermed besvare min undren.

Figur 6: *Hovedprincippet i **Prædiktiv-analyse**. Analysemetoden anvender flere datasæt – indlæringstabeller – med forskellige data omkring boligtyper og biler samt demografi.*

Prædiktiv analyse i MapInfo er en metode til såkaldt "smart" klassifikation. Metoden undersøger de statistiske karakteristika for flere kvadratnet til indlæsning af data, der

er fundet inden for et given "reference/trænings" område og derefter lokaliserer andre områder med lignende egenskaber. Før den Prædiktiv analyse i MapInfo kan køres, skal der først oprettes en "indlærings tabel". Dette er en MapInfo tabel med objekter i et udvalgt områder, der bruges som prøveområde.

5.5 Opsummering.

Med projektanalysen og de enkelte forsøgsbeskrivelser der vil blive anvendt i dette projekt til besvarelse af min hypotese, har jeg forsøgt at give den første indføring i, hvordan geografisk analyse og visualisering af tabeldata via Excel-regneark og et GIS-baserede program (MapInfo), er i stand til at finde ny sammenhænge og forståelse af datamaterialer, der ikke tidligere har kunnet lade sig illustrerer. Som tidligere nævnt i indledningen, er der i dag en stigende tendens til at tillægge geografiske beliggenheds-data til oplysninger om kunder og kortlægge - *visualisere* - disse og dermed få en helt ny indgangsvinkel omkring registrerede kundedata.

Forsøgsbeskrivelserne med den proces-orienterede undersøgelsesform (input-proces-output), skulle gerne dokumentere den simplicitet der i bund og grund råder i dette projekt, for at vise hvordan værdier og information i geografiske data kan være med til at danne grundlag for beslutninger og videre foranstaltninger i en forretningsverden. Efter at have gennemgået selve projektets forsøgsbeskrivelse og de elementer der danner grundlaget for at teste min hypotese, vil vi nu bevæge os over i de mere teoretiske aspekter i dette projekt, og kigge lidt nærmere på, hvordan data kan repræsenteres, opfattes og formidles i forbindelse med forandringer og forarbejdning af råmaterialet - DATA - op igennem DIKW-hierarkiet.

Det menneskelige aspekt i forholdet mellem afsender og modtager af information, samt tilgængeligheden af information og dermed formidling af viden, vil blive belyst og diskuteret. Det er informationsteori i denne betydning, der har interesse i forbindelse med forskning og undervisning, samt i forbindelse med udviklingen og informations-systemer, og den måde hvorpå jeg anvender og opfatter brugen af data i dette projekt, til at få svar på en undren.

6. Teori og Metode

Teori og Metode sætter fokus på DIKW-hierarkiets grundlag for viden-transformation, fra rå data til anvendelig information som et beslutningsgrundlag, til at opnå en viden om en ting eller et produkt. Præsentationen af relationerne mellem Data, Information, Viden og til tider Visdom i en hierarkisk ordning, har været en del af sproget i informationsvidenskaben i mange år. Teori og Metodeafsnittet har til formål at vise, hvordan jeg vil indsamle, identificere, filtrere, fokusere og bearbejde mine data til at finde en sammenhæng mellem to datatyper, samt et større kundesegment for et produkt og via DIKW-hierarkiet, legitimere min anvendelse af kort- og tabelmaterialet, til at opnå mit resultat. Analysemetoderne er baseret på Excel- og SQL-analyse.

6 TEORI OG METODE

6.1 Indledning.

Jo længere de vertikale og horisontale rækker af data er i en matrix (Excel-tabel), jo mere information indeholder tabellen og jo mere information kan vi trække ud af tabellen og derved danne os et overblik over forskellige sammenhænge og mønstre, der gemmer sig bag disse data (for det meste tal), altså vi kan opnå en hvis viden om de indsamlede data. Her er det naturligvis vigtigt at holde sig for øje, at opdeling, klassificering og manipulering af data er menneskelig. Det vil sige, at vi på forhånd opstiller nogle parametre og egenskaber (største og mindste værdier, positive eller negative værdier) som vi er interesseret af at undersøge og trække ud af denne tabel med data.

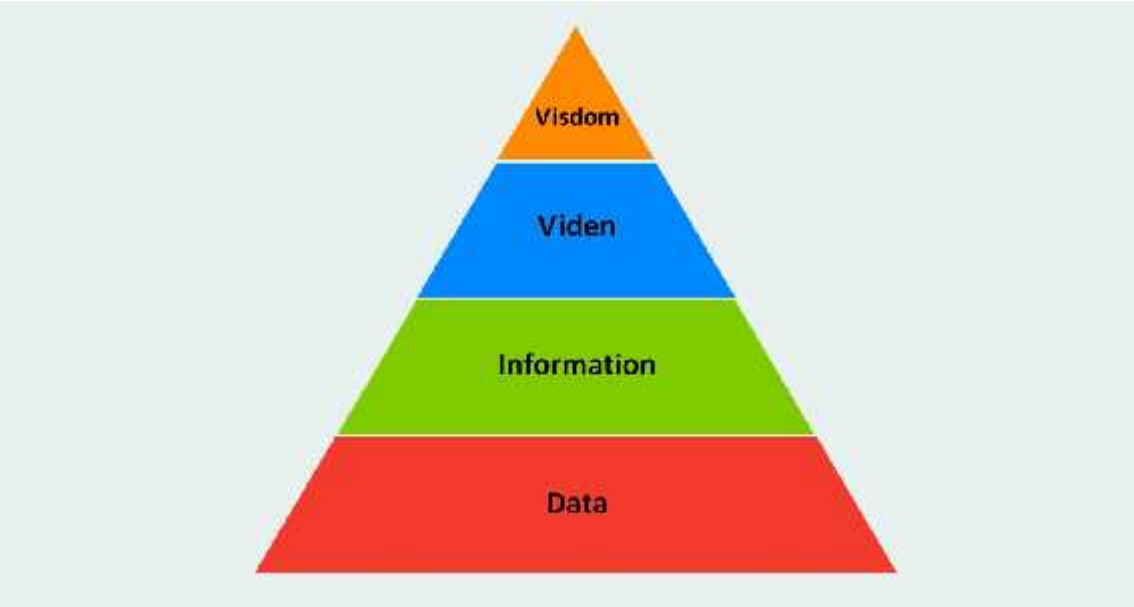
Præsentationen af relationerne mellem Data, Information, Viden og til tider Visdom i en hierarkisk ordning, har været en del af sproget i informationsvidenskaben i mange år. Selv om det er usikkert, hvornår og af hvem disse relationer blev først fremlagt, er det allestedsnærværende begreb af et hierarki, forankret i brugen af akronymet DIKW, som en forkortet gengivelse af Data-til-Information-til-Viden-til-Visdom transformationen[20]. Teori og metodeafsnittet har til formål at vise, hvordan jeg vil indsamle, identificere, filtrere, fokusere og bearbejde mine data til at finde en sammenhæng mellem biltyper og boligtyper i et boligområde, og via DIKW-hierarkiet legitimere min anvendelse af kort- og tabelmaterialet, til at opnå mit resultat. Analysemetoderne er baseret på Excel- og SQL-analyse.

6.2 Teori - Data, Information, Viden og Visdom.

En meget anvendt og også omdiskuteret model (hierarki) til belysning at informationsstrøm er DIKW-hierarkiet (Data – Information – Knowledge – Wisdom) = Data – Information – Viden – Visdom. Data, information, viden, visdom, hierarkiet baserer sig på filtrering, reducering og transformering og er illustreret i form af en pyramide – se ***Figur 7***, der bestående af fire niveauer [16]. Progressionen foregår i al sin enkelthed ved, at man bevæger dig op gennem hvert niveau af pyramiden og når toppen. Processen begynder ved at indsamle data på det første niveau, som derefter behandles for at danne information på det andet niveau. Når denne information er undersøgt eller reflekteret, tager den form af viden på det tredje niveau, og skabelsen af viden fører til erhvervelse af visdom på det fjerde eller øverste niveau.

Data, der udgør bunden af hierarkiet, kan indsamles enten manuelt eller via automatiserede systemer og værdi føjes til disse input, ved at fortolke og konvertere dem til et brugbart og meningsfuld format - betegnet som Information. Når information anvendes i en bestemt situation og konverteres til ekspertise, bliver den defineret som

Viden. Visdom findes på toppen af DIKW-hierarkiet. Visdom adskiller sig fra data og information, som begge kan styrkes, og også fra viden, der kan deles. Visdom er mere abstrakt, da det er noget reelt og er typisk en ophobning af værdier, vurdering, tidligere erfaring eller fortolkninger. Ordet *forståelse* ville på mange måder måske være en mere rammende terminologi, i denne her sammenhæng. [3]



Figur 7**:** *Data-Information-Viden-Visdom hierarkiet.*

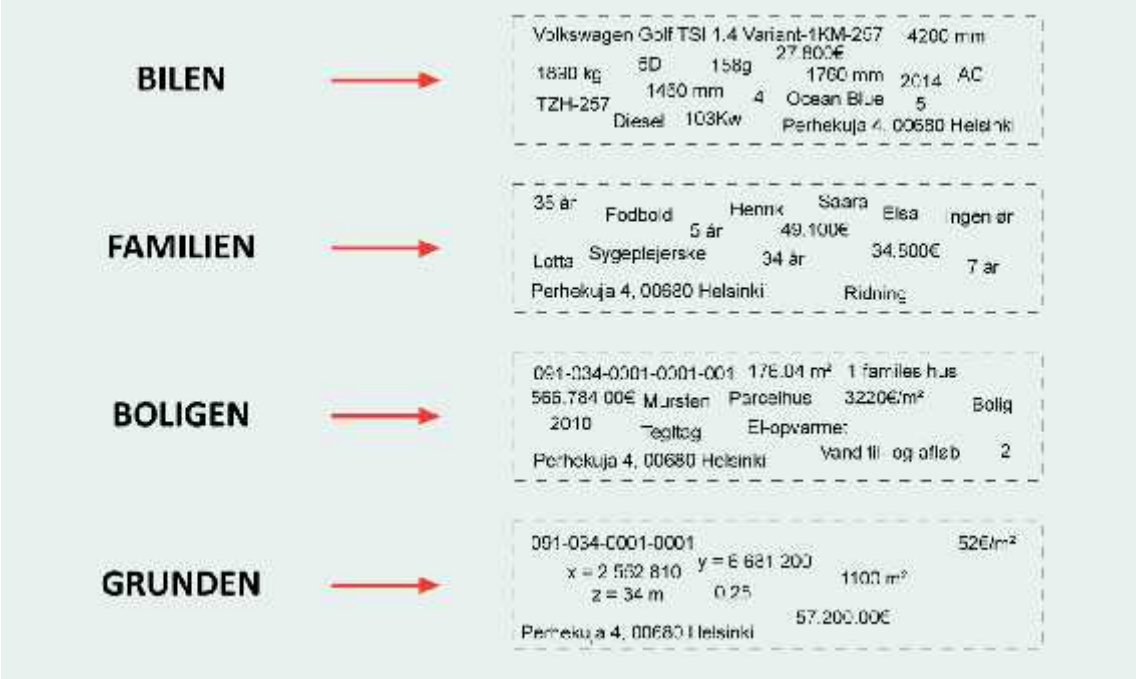
Præsentationen af relationer mellem data, information, viden og til tider visdom i en hierarkisk ordning har været en del af terminologien i informationsvidenskaben i mange år. Som den første der har beskrevet DIKW-filosofien som et hierarki, krediteres Thomas Stearns Eliot (1888 – 1965). T.S. Eliot var poet og dramatiker og modtog Nobels litteraturpris i 1948. Den person, der i dag nok har fået mest opmærksom for sin fortolkning af DIKW-hierarkiet, er Russel Ackoff. Ifølge Russell Ackoff, en systemteoretiker og professor i Organisatorisk Forandring, kan indholdet af det menneskelige sind inddeles i fem kategorier: [19]

1) **Data** = Symboler
2) **Information** = data der er blevet bearbejdet således at der i stand til at besvare spørgsmål om *hvem*, *hvad*, *hvor* og *hvornår*.
3) **Viden** = anvendelse af data og information til at besvarer spørgsmål som *hvordan*.
4) **Forståelse** = vurdering af *hvorfor.*
5) **Visdom** = evalueret forståelse, til et kig ind i fremtiden.

Ackoff indikerer, at de første fire kategorier (data, information, viden og forståelse) vedrører fortiden; de beskæftiger sig med hvad der har været, eller hvad der er kendt. Kun den femte kategori, visdom, beskæftiger sig med fremtiden, fordi det indeholder visioner og design. Med visdom kan folk skabe fremtiden, snarere end blot forstå

nutiden og fortiden. Men at opnå visdom er ikke let; folk skal bevæge sig successivt gennem de andre kategorier. En mere detaljeret redegørelse af Ackoff definitioner af de 4 begreber (Data, Information, Viden og Visdom) er som følger: [18]

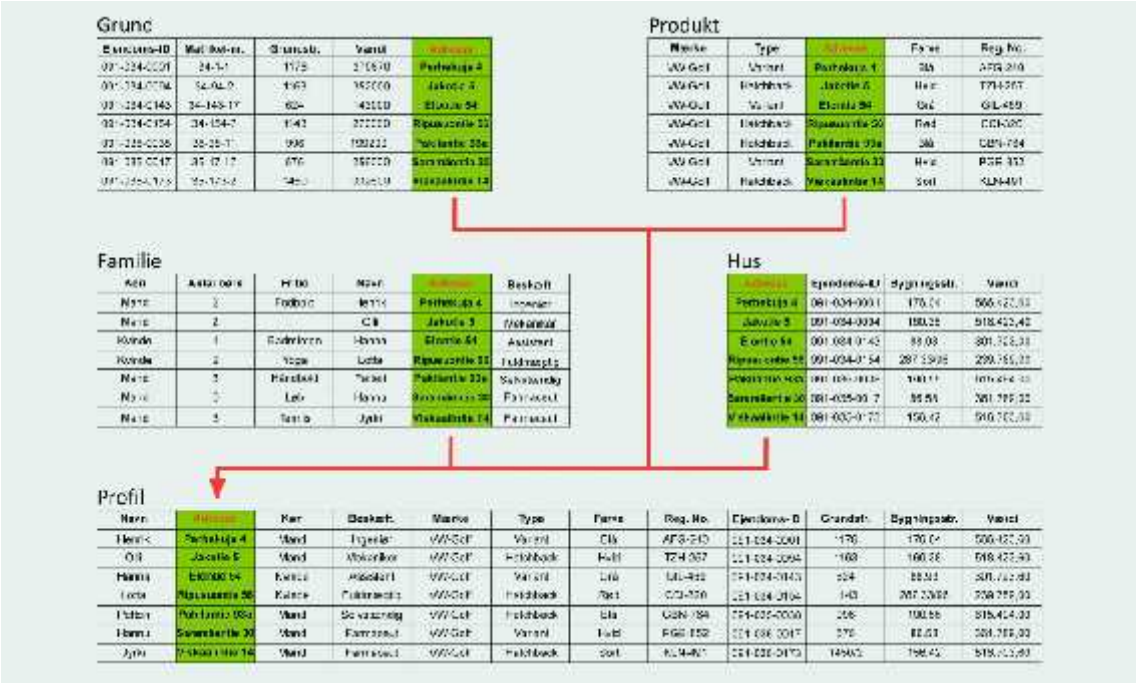
1) **Data** henfører til en samling af information, som typisk er resultatet af erfaring, observation, eksperimenter eller en mængde af præmisser. Data kan bestå af tal, ord eller billeder. Data eksisterer bare og har ingen betydning ud over dens eksistens (i sig selv). Det kan eksistere i nogen form, anvendelige eller ej. Det har ikke betydning i sig selv. ***Figur 8*** viser eksempler på data og datatyper, der er blevet indsamlet til mit projekt og min fortolkning af data, samt i hvilken form og format.



Figur 8: *Data er et element eller en begivenhed ud af sin sammenhæng uden nogen relation til andre ting.*

Der skelnes mellem to typer af data: Kvalitative data og kvantitative data. Kvalitative data er en datatype, der fortæller om de fænomener, der ikke kan tælles og måles. Dataene bruges i kvalitativ analyse og indsamles ofte ved et kvalitativt forskningsinterview. De er detaljerede og giver dybde og helhedsforståelse af specifikke forhold og står ofte for sig selv. Der kan ikke uden videre generaliseres ud fra kvalitative data. Dataene bidrager ofte til udvikling af teorier og hypoteser. Kvantitative data er data, der kan tælles eller måles. Data optræder samtidigt i en ensartet form og eventuelt i store mængder. De giver bredde og overblik. Et eksempel på kvantitative data er statistiske datasamlinger og registre.

2) **Information** er et ord, der i vid udstrækning har fortrængt det danske ord "oplysning". Det er et svært begreb, som der er mange modstridende teorier og forklaringer om. Information er data, der er blevet givet indhold ved hjælp af relationelle forbindelse. Dette "indhold" kan være nyttig, men behøver nødvendigvis ikke at være det. Nedennævnte ***Figur 9*** illustrere hvordan jeg har klassificeret og kombineret forskellige datasæt fra ***Figur 8***, sat dem i system i forhold til typer og kategori, ved blandt andet at spørge med *hvem, hvad, hvor* spørgsmål og derigennem tilvejebragt en informativ forbindelse (kontekst) til mine data. Fællesnævneren i dette tilfælde har været ***Adresse.***

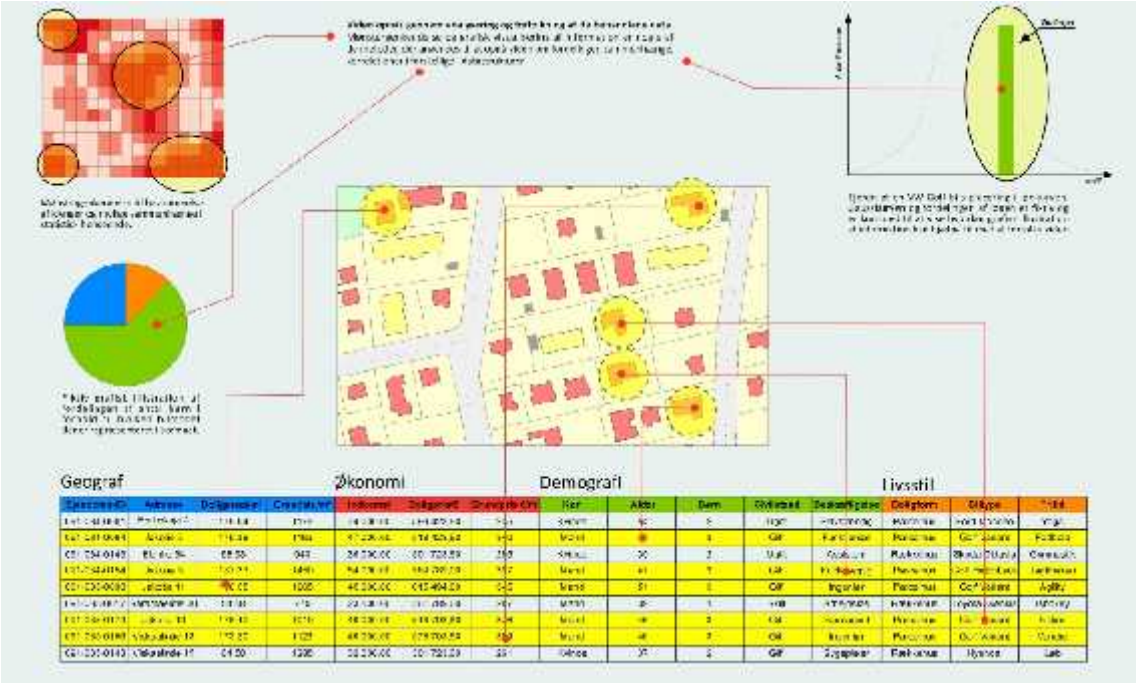


Figur 9: *Information repræsenteret ved relationer mellem data og eventuel anden information. Information opstår ved at tilvejebringe kontekst til data. Informationen bruges til at besvare hvem, hvad, hvor og hvornår spørgsmål. Samme opbygning for datastuktur findes i en relationel database.*

Vi behøver jo ikke at gå længere end til vores egen verden og GIS (Geografisk Informations System), der er en betegnelse for en type af software, hvormed man kan fremstille kort og analysere geografisk information - geodata. GIS er et IT-værktøj, der som et analyse- og visualiseringsredskab, kan give overblik over data med en geografisk relation. Når data kobles til digitale kort, bliver det muligt at afdække geografiske sammenhænge i data. Hidtil ukendte mønstre kan dermed fremtræde klarere og fremme kommunikationen, fordi alle kan forstå, hvad de ser på et kort (parkområder i forhold til bebyggelse, forskellige vejtyper, seværdigheder). [8]

Der er to hovedgrupper af teorier omkring information og forståelsen af denne: En objektiv gruppe og en subjektiv gruppe. I følge den objektive forståelse er information noget, der findes i verden, uanset om der er nogen, der betragter det eller ej. Information er f.eks. forskelle i stof og energi. I følge den subjektive forståelse er information noget, der kan informere nogen om noget. Gregory Bateson er kendt for at have sagt: "*Al modtagelse af information er nødvendigvis modtagelse af budskaber om forskel*". Med andre ord er information en forskel der gør en forskel. Alt kan potentielt være information, da alting potentielt kan besvare et eller andet spørgsmål for nogen. En træstub er informativ for den, der ved at tælle årringene udregner træets alder da det blev fældet. En træstub kan altså være information. Det følger videre, at ingen ting er information i sig selv, men er kun information i forhold til de spørgsmål, den kan bidrage til at besvare. At en ting er information er ikke noget givet, men noget der bygger på en vurdering, som kan vise sig at være forkert. Det som er i databaserne og på Internettet som i daglig tale kaldes information, er altså efter den subjektive opfattelse, kun potentiel information. Det som er information for nogle spørgsmål, for nogle brugere og for nogle domæner, er ikke nødvendigvis information i andre kontekster. En generel og ultrakort definition på begrebet "information" i både objektiv og subjektiv betydning er: "svar på et foreliggende spørgsmål". [5]

3) **Viden** er en hensigtsmæssig indsamling af oplysninger, således at det er hensigten at være nyttig. Viden er en deterministisk proces, altså at alle begivenheder har en årsag. Når nogen "husker" information, da har man samlet viden. Denne viden er nyttig, men den giver ikke mulighed for, i sig selv, en integration som ville udlede yderligere viden, i dette tilfælde forståelse, at forstå en bestemt sammenhæng; måde hvorpå man opfatter og forstår noget. Viden er en form af fortrolighed, bevidsthed eller forståelse af nogen eller noget, som fakta, information, beskrivelser, eller færdigheder, som er erhvervet gennem erfaring eller uddannelse ved at opfatte, opdage og lære. Viden kan henvise til en teoretisk eller praktisk forståelse af et emne. Viden kan være implicit (som med praktiske færdigheder eller ekspertise) eller eksplicitte (som med den teoretiske forståelse af et emne). [21] I ***Figur 10*** har jeg sammensat en række illustrationer, der skal vise hvorfra og hvordan jeg vil udtrækker information fra forskellige mønstre i redigerede tabeller, kort og grafiske fremvisninger, for at opnå en viden og derigennem en forståelse, der igen kan danne grundlag for en beslutning for en analysemetode (SQL eller Prædiktive), for at finde et segment eller en begivenhed.



Figur 10: *Viden er repræsenteret ved mønstre blandt data, informationer og eventuel anden viden. Disse mønstre udgør faktisk ingen viden indtil de er forstået.*

4) **Visdom**. Fænomenet visdom adskiller sig både fra viden, intelligens, evner og færdigheder. Kort sagt er **visdom** noget man ***er*****,** mens **viden** er noget man ***har***. Hvor intelligens er noget, der mere eller mindre bliver fastlagt i en tidlig alder, så siges det ofte, at en persons visdom kan vokse eller falde hele livet igennem. Visdom er på den måde hverken afhængig af arv eller opvækst. En uuddannet bonde i en fattig del af verden, kan således sagtens være mere vís end en distræt professor på et vestligt universitet, der omvendt må formodes at være mere vidende og (måske) mere intelligent. Visdom handler i personlig henseende især om, bevidst eller ubevidst, at kende sine egne begrænsninger. Således blev det sagt, at Sokrates (5. årh. f.Kr.) var den viseste mand i Athen, fordi ”han vidste, at han ikke vidste noget”, eller mere præcist fordi han vidste, at alt hans menneskelige viden strengt taget ikke var noget værd. Visdom omhandler fænomenet ”hvorfor”. Visdom indikerer en endnu dybere indholdsbestemt bevidsthed til viden. Visdom betyder, at du er i stand til at se og anvende "det store billede" af de faktiske omstændigheder, at besvare "hvorfor" spørgsmål. l. Visdom er evnen til at øge effektiviteten. Visdom tilfører værdi, som kræver den mentale funktionalitet, som vi kalder dømmekraft. Betydningen af de etiske og æstetiske værdier er uløseligt forbundet med personen og er unikke og personlige.

6.3 Implementering.

Formålet med DIKW-hierarkiet er således at formalisere en tankegang, nemlig - at gå fra data til evnen til at anvende dem på bedste vis. Som jeg ser det i mit projekt og generelt med anvendelsen af data inden for informationsteknologien, er der ingen problemer med at forstå og også anvende filosofien omkring DIKW. De forskellige datakilder (Excel-tabeller) jeg bruger i mit projekt, er organiseret i felter og kolonner. Der er ingen bindeled eller logik mellem de forskellige felter og kolonner, men når de forskellige tabeller sættes sammen, via en fællesnævner som for eksempel adresse, dannes der en information, der individuelt kan anvendes eller videreudvikles alt efter behov. Der kan endda blive afledt ny information, ved at integrere kolonner og felter med andre tabeller. Data og for den sags skyld også den deraf udledte information, er ikke falsk eller usand. De data jeg anvender, er baseret på identifikation, målinger, indsamlinger og filtreringer, dels fra offentlige instanser og private firmaer, men også udledt af egen tankegang, forundring og interesse for dette projekt og den problem-stilling jeg vil undersøge.

I dette projekt forsøger jeg ikke at finde den korteste afstand til et objekt, eller det bedste sted at placere en butik/virksomhed. Her er der tale om at finde en sammen-hæng mellem en biltype og boligtyper (geografiske elementer). Typen af data og deraf udledt information, bliver i nogen grad personificeret. I dette projekt søger jeg heller ikke efter nogen specifik viden til at afklare mit spørgsmål i min hypotese. Jeg forsøger at opnå en divergeret viden om mit kundesegment på basis af et produkt, således at jeg kan drage en konklusion med hensyn til om der er en relation mellem boligtype og produkt - (VW-Golf).

Projektet har allerede vist, at anvendelsen af for eksempel persondata og personers integritet har en afgørende indvirkning på dette projekts resultat. Når en adresse på et produkt kan lede til en identifikation af en person og dermed ”afslører” eller vise en sammenhæng i et mønster af hændelser, der går år ud over det, der var intentionen med fremskaffelse af en information. Dermed har DIKW-teorien vist, at kombinationer af forståelsen af forholdet mellem data, information og viden kan lede til handlinger og resultater, der måske ikke var ønskværdige i mit tilfælde, men for et firma der fore-tager samme undersøgelse, vil få et helt nyt og meget mere nuanceret billede af en kunde og det som kunden står for som person (livsstil).

Vi tale her om en såkaldt internalisering, altså en indarbejdelse af en anden persons værdier, personlighed og identitet til visualisering af flerdimensionelle, sammenhængs gennem ekspressionen, der er tilpasset en given situation ved hjælp af for eksempel diagrammer, kort eller animering. [6]

6.4 Kritik af DIKW-hierarkiet.

Der har været fremført en del kritik af DIKW-hierarkiet som model i forbindelse med præsentationen og synergien af relationerne mellem Data, Information, Viden og Visdom. DIKW-hierarkiet er ofte citeret, eller anvendt implicit i definitioner af data, information og viden i forvaltning af information og informationssystemer, men der har kun været begrænset diskussion af hierarkiet som sådan. Forskellige artikler og lærebøger om DIKW-hierarkiet viser, at der ikke er en enig opfattelse om de definitioner, der anvendes i og omkring modellen, og endnu mindre i beskrivelsen af de processer, hvorved elementerne i hierarkiet transformeres fra et lavere stadiet til dem der er over dem. [17]

Det har måske sin berettigelse, taget i betragtning, hvornår modellen er blevet "opfundet" (1934 – 1989) og den informationsalder vi lever i lige nu (2014). Kritikken går lige fra den illustrative udformning af modellen (trekant, størrelsen af de forskellige emner), til samspillet af relationerne mellem Data, Information, Viden og Visdom. Det som ser ud til at være det største "problem" for diverse kritikere er det, ***at der ikke er en synergi, logisk rækkefølge og sammenbinding mellem de forskellige led i DIKW-pyramiden,*** og også inden for hvilke videnskabsgrene dette hierarki anvendes. Informationsteknologien ser ingen større problemer i anvendelsen af modellen, medens naturvidenskaberne ikke helt kan takle den.

Det som er den afgørende faktorer for anvendelsen og forståelse af DIKW-hierarkiet, som lede-model for synergien mellem Data, Information, Viden og Visdom er dels tiden vi lever i lige nu, med hensyn til dataindsamling – Big Data - fra forskellige online og realtid kilder, såsom GPS og video, men i lige så høj grad i hvilken informationsvidenskab, DIKW-hierarkiet sættes i relation til, det vil sige, teoriopbygningen omkring, hvordan data kan repræsenteres, opfattes og formidles. Hermed menes:

- Den naturvidenskabelige og statistiske betydning af informationsteori (entropi - den samlede uorden eller tilfældighed i et system)
- Den datalogiske betydning af informationsteori (opbevaring, transmittering, kryptering og transformering)
- De menneskelige aspekter i forholdet mellem afsender og modtager (forskning og undervisning). [24].

Ackoff's model var primært rettet mod ledelse og forvaltning i organisationer og kun tilfældigvis, hvis overhovedet, til for eksempel til bibliografiske materialer og samlinger. [18]

Frické (2009) finder at hierarkiet er usundt og ikke metodologisk ønskværdigt. Der påvises en grundlæggende logisk fejl i DIKW-hierarkiet. Modellen har baggrund i foræl-

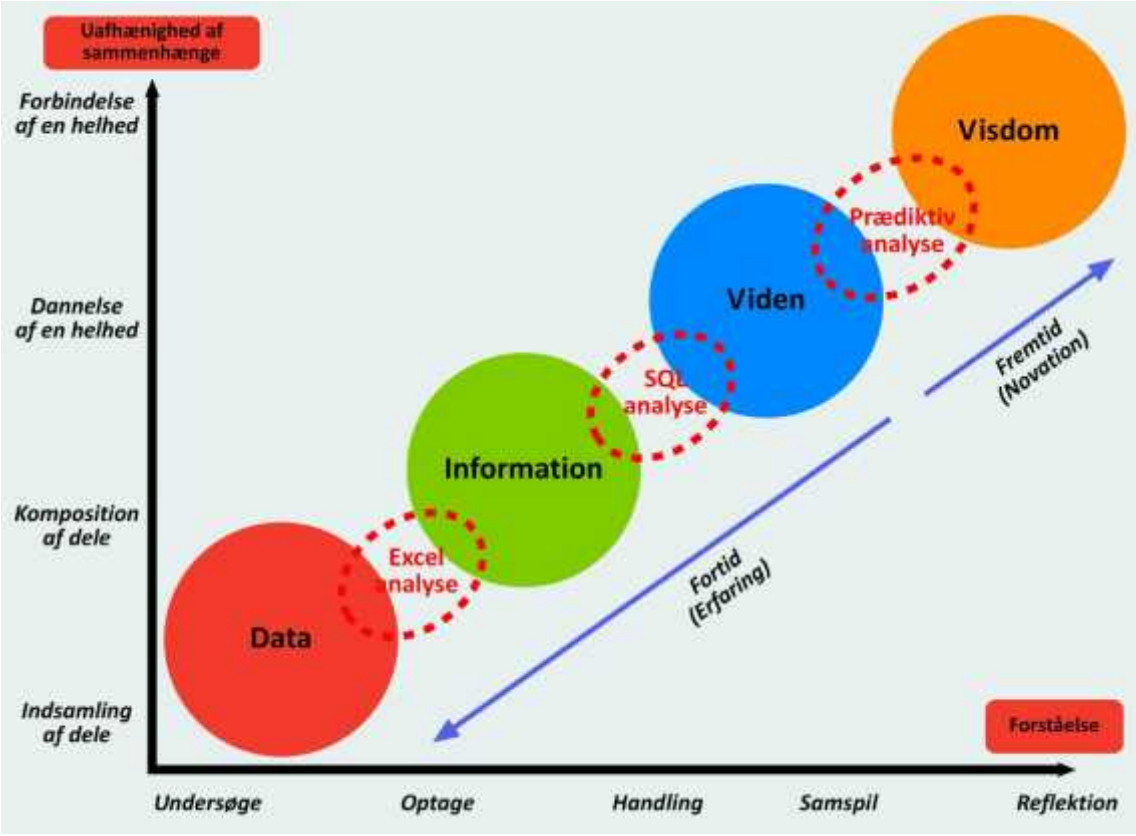
dede og uhensigtsmæssige filosofiske positioner fra operationalismen og induktionismen. [16]

6.5 Metode.

Som nævnt i indledningen og problemanalysen, var ideen med projektet at finde en sammenhæng mellem biltyper og boligtyper, ved at plotte ca. 200 "rigtige" adresser af VW-Golf bilejere ind på et kort med boliger og ejendomsgrænser. Derved ville der være muligt på baggrund af en simpel statistisk optælling at få et førstehåndsindtryk af, hvor disse bilejere boede (område i Helsingfors) og hvad de boede i (parcelhus, rækkehus, højhus), samt typen af grunden hvorpå boligen var placeret. Den førstehåndsinformation skulle så videreudvikles og tillægges ydereligere information (som vist i ***Figur 9***) om bilejeren efter at en "profil" af denne var blevet udarbejdet på basis af boligtype og grundejerforhold. So far so good. Desværre har det ikke været muligt at få tilgang til disse "rigtige" kundedata med den rigtige ejer af dette produkt, da data-producenten (*TRAFI**, se* ***Bilag B3***) har en formodet mistanke om, at dette ville kompromittere bilejerens identitet i en for afslørende grad til, at det ville være hensigtsmæssigt i mit projekt. Det er naturligvis beklageligt, men forståeligt. Det skal hertil siges, at private firmaer, der foretager kundeanalyser til eksempel bilfirmaer, har en vis adgang til disse personoplysninger. Derfor har jeg valgt at anvende såkaldte "fiktive data" om bilejerens adresse og hvilke VW-Golf bilmodeller de ejer. Excel-tabellen, der indeholder disse data, er derfor en tilfældighed af adresser og biltyper og har ingen relation til virkeligheden.

En mere nuanceret og beskrivende illustration af DIKW-hierarkiet er vist i ***Figur 11.*** [21] Via denne 2-dimesionelle illustration vil jeg vise, hvor de to analysemetoder – Excel og SQL-analyse, som jeg anvender til at udtrække mine data, har deres placering i DIKW-hierarkiet samt deres berøringsflader som jeg ser dem anvendt og implementeret i mit projekt. Hvorvidt data, information, viden og visdom er fortid eller fremtid, er ikke så interessant i denne sammenhæng.

Det, der illustrere hele DIKW-filosofien, er den måde hvorpå *forholdet mellem forståelse og sammenhænge har en relation*, efterhånden som vi bevæger os "op gennem relationerne" og hvordan intensiteten af parametrene (tekst under cirkelillustrationerne), der har betydning. Denne opbygning og illustration af DIKW giver en mindre hierarkisk præsentation af elementernes indbyrdes forhold, som på mange måder er ønskværdigt.



Figur 11: *DIKW-hierarkiet i en mere nuanceret og modificeret udgave, som jeg ser den anvendt og implementeret i mit projekt. De to aktivt anvendte analysemetoder er efter min mening placeret i DIKW-hierarkiet mellem data og information og mellem information og viden. Den beskrevne Prædiktive analyse findes efter min mening mellem viden og visdom (forståelse).*

Undersøge: At undersøge et emne eller en opgave gør det muligt for personen at opdage forskellige forhold, relationer, strukturer og/eller modeller om et emne.

Optage: Læse, se, lytte, og føle er gode metoder til at absorbere information. Men denne information bliver ikke til viden med det samme, når vi har absorberet den, fordi vi simpelthen har kopieret oplysninger fra en kilde, såsom en bog, til en anden kilde, vores hjerne.

Handling: En aktivitet der indebærer mental manipulering af information eller fysisk manipulering af et objekt. Den mentale manipulation af information adskiller sig fra for eksempel refleksion derved, at handling normalt indebærer en planlagt øvelse eller aktivitet, hvorimod refleksion er en aktivitet i mere ”fri format”, der normalt indebærer en masse "hvad og hvis'er".

Samspil: En form for handling, der opstår, når to eller flere objekter har en effekt på hinanden.

Refleksion: At tænke i en længere periode ved at knytte de seneste erfaringer til tidligere erfaringer, med henblik på at fremme et mere komplekst og indbyrdes forbundne mentale skema. Tankegangen indebærer udkig efter fællestræk, forskelle og

indbyrdes forbindelser ud over de overfladiske elementer. Målet er at udvikle færdigheder i tankning af en højere orden.

Dataanalyse er processen med at udvikle svar på spørgsmål ved undersøgelse og fortolkning af data. De grundlæggende trin i den analytiske proces består i at:

- Identificere problemer,
- Bestemmelse af tilgængeligheden af egnede data,
- Beslutter hvilke metoder der er egnede til at besvare spørgsmål af interesse, og
- Anvendelse af de metoder og evalueringer, der sammenfatter og formidle resultaterne. [11]

I dette projekt har jeg som tidligere nævnt, tænkt mig at anvende to analysemetoder for at få svar på mine spørgsmål med hensyn til udvælgelse af et kundesegment.

6.6 Excel-analyse.

Excel-analysen der udføres som den første analyse i dette projekt, skal mere ses om en slags simpel manipulation af data, før de bliver, eller kan fortolkes som, egentlig information. Denne form for analyse skal ses som en slags "førstehjælp" til at danne sig et bredere billede af hvor potentielle kunder kunne findes. Her er fokus direkte rettet på manipulation og sammenligning af rå data fra et register **(HRI's Boligregister)** der findes i tabeldataform som *.csv format (konvertibel format til Excel).

Denne første analyse af mine data fra boligregisteret skal vise, hvordan man med simple teknikker i Excel (**VLOOKUP** samt **Find and Replace**), kan udtrække et kundesegment og dermed en profil af bilejeren, ved at indsætte egne data (kundeadresser) i tabellen og lave en sammenligning. Denne form for udtræk af en førstehåndsinformation om bilejeren, kræver kun lidt viden om brugen af Excel-tabeller og lidt kreativ tankegang. Det er endvidere vær at bemærke, at denne form for analyse ikke kræver noget GIS-baseret system eller viden om GIS i det hele taget, men kan fortages på et hvilket som helst kontor eller institution. Denne form for analyse kan give et første indblik i et mønster eller trends med foreliggende data. En helt almindelig kontorfunktionær er altså i stand til at lave sin egen analyse ud fra frie data, ved en simpel manipulation.

6.7 SQL-analyse.

Structured Query Language eller **SQL** er det mest udbredte programmeringssprog til relationelle databaser (tabeldata). Sproget dækker både definition af databaser og manipulation af data. SQL er et deklarativt programmeringssprog, hvilket vil sige, at man kun beskriver, *hvad* der skal ske, men *ikke hvordan* det sker. Altså - SQL bringer

information på et sådant niveau eller stade, at det er muligt at udlede en viden on en begivenhed ud fra det, der er spurgt om. [12]

I en relationel database kan man oprette tabeller. Tabeller består af rækker af felter. De felter, i hver kolonne, der står i samme række, kaldes en post. Relationerne mellem tabellerne udgør den enkelte databases struktur. Det teoretiske grundlag for den relationelle datamodel er mængdelæren. I praksis er det dog de færreste systemer, der lever helt op til reglerne om mængder. For eksempel er det ofte muligt at have resultater af forespørgsler, der indeholder dubletter af rækker.

Data i en relationel tabel skal som minimum overholde følgende:

1. Hver rækker i en tabel skal have en unik nøgle. Dette kaldes entitetsintegritet
2. Alle rækker har det samme antal felter
3. Alle felter har en fast længde defineret ved datatypen.

Ejendoms-ID	**Kolonne**	*Adresse*	*Postnumer*	*Biltype*	*Boligtype*	*Kvad. m²*	*Boligværdi*
Række							

Figur 12: *Eksempel på en relationel tabel (Excel-tabel eller MapInfo-Browser) i en simpel opbygning med et relevant indhold af klasser som i dette projekt.*

Ved læsning af data bruges tre grundlæggende operationer.

1. Selektion: Find rækker, der opfylder en bestemt betingelse som for eksempel "alder = 42"
2. Projektion: Vælg et antal kolonner fra en tabel. Det kunne være fornavn og efternavn.
3. Krydsprodukt: Kombiner alle rækker i en tabel med alle rækker i en anden tabel.

Operationerne kan kombineres, så der kan laves forespørgsler i stil med:

"Kombiner persontabellen og adressetabellen. Find fornavn, efternavn og adresse på dem, der er 42 år gamle".

6.8 Prædiktiv Analyse.

Prædiktiv analyse i MapInfo er en metode til såkaldt "smart" klassifikation. Metoden undersøger de statistiske karakteristika for flere kvadratnet til indlæsning af data, der er fundet inden for et given "reference/trænings" område og derefter lokaliserer andre områder med lignende egenskaber. Hvis vi antager, ud fra en erfaring, at tilstedeværelsen af VW-Golf biler i et boligområde er stærkt relateret til tilstedeværelsen af bestemte typer af huse, eller andre elementer som størrelsen af ejendommen/grunden. Hvis vi kan etablere en sammenhæng mellem tilstedeværelsen af disse andre elementer og VW-Golf biler, så det bliver muligt at måle koncentrationen af de andre elementer, til at forudsige tilstedeværelsen af VW-Golf biler, hvilket er langt mindre kostbar, end at forsøge at tilfældigt finde VW-Golf biler i et område.

Med Prædiktiv analyse er det mulighed at "lærer" programmet, hvordan vi klassificer et område, på grundlag af kriterier for indlæsning af data. For eksempel, hvis jeg identificere flere områder, hvor VW-Golf er til stede, kan jeg ved hjælp af disse områder som modeller, lærer programmet at identificere andre områder med lignende karakteristika.

Før den Prædiktiv analyse i MapInfo kan køres, skal der først oprettes en "indlærings tabel." Dette er en MapInfo tabel med objekter i et udvalgt områder, der bruges som prøveområde. I alle ”indlæringstabeller”, skal der defineres mindst to forskellige klasser. For eksempel, hvis vi ønsker at identificere områder med sandsynlige høje frekvenser af VW-Golf biler baseret på demografiske oplysninger, skal vi definere både et område, hvor frekvensen er VW-Golf biler er kendt for at være høj, og et område, hvor frekvensen af VW-Golf biler er kendt for at være lav. For mere præcise resultater, kan der bruges mere end et område for hver klasse i ”indlæringstabellen”.

6.9 Afslutning.

I informationsalderen er det sommetider let at glemme, at de informations færdigheder der er udviklet i en tidligere æra, stadig gælder. De basale værktøjer har måske nok ændret sig, men nogle af tankegangene omkring behandling af information består uændret. At anvende information effektivt har altid krævet et vist sæt af færdigheder som inkluderer det at tænke på, hvilke informationer der forders, lokalisere informationen, evaluere, vælge og organisere den, og så bruge eller koncentrere sig om den.

Opbygningen af dette projekt har på flere måder samme stuktur som DIKW-hierarkiet. *Teori og Metode = Data*. Her findes mange forskellige indgange forklaringer til et emne eller en ting. *Empirien = Information*. Dette kapitel forklarer hvordan projektet har til hensigt at strukturere og implementere til rådighed stående data. *Analysen = Viden*.

Her transformeres al Datamateriale og struktureret information til viden via analyser og visualiseringer i forskellige afskygninger. *Konklusionen og Perspektivering = Visdom (forståelse)*. På baggrund af den indsamlede og fortolkede viden, kan jeg nu udstikke retningslinjer for fremtidige tiltag i forbindelse med besvarelsen af min hypotese og eventuelle sammenhænge mellem biltyper og boligtyper. Ved at reflektere over disse sammenhænge, kan dette forbindes til en helhed.

Og endelig – alt udvikles over tid og via erfaringerne lærer vi. Således også med DIKW-hierarkiet.

Med det teoretiske aspekt som et forklaringsled til informationsstrøm og segmentering af information, vil vi nu bevæge os over i den empiriske del af projektet, og gennemgå det analysemateriale, der danner grundlaget for de analyser jeg vil fortage for at få svar på min hypotese.

7. Empiriske materiale

Én af ideerne med dette projekt er, at undersøge de værdier og informationer der "gemmer" sig i geografiske data. På den baggrund og via forskellige analysemetoder af de indsamlede data, undersøges det, om det er muligt at finde og se mulige sammenhænge og korrelationer i den informationsmængden, der er til stede for dette projekt, og dermed få et nyt og bredere beslutningsgrundlag. Ved at tillægge, organisere, klassificere og manipulere værdier og informationer fra andre datakilder, kan der tegnes et nyt og måske mere nuanceret og informativt billede af den eller de personer, der ejer en VW-Golf. Kapitlet om Empiriske materiale, forsøger at klarlægge denne tankegang.

7 EMPIRISKE MATERIALE

7.1 Indledning.

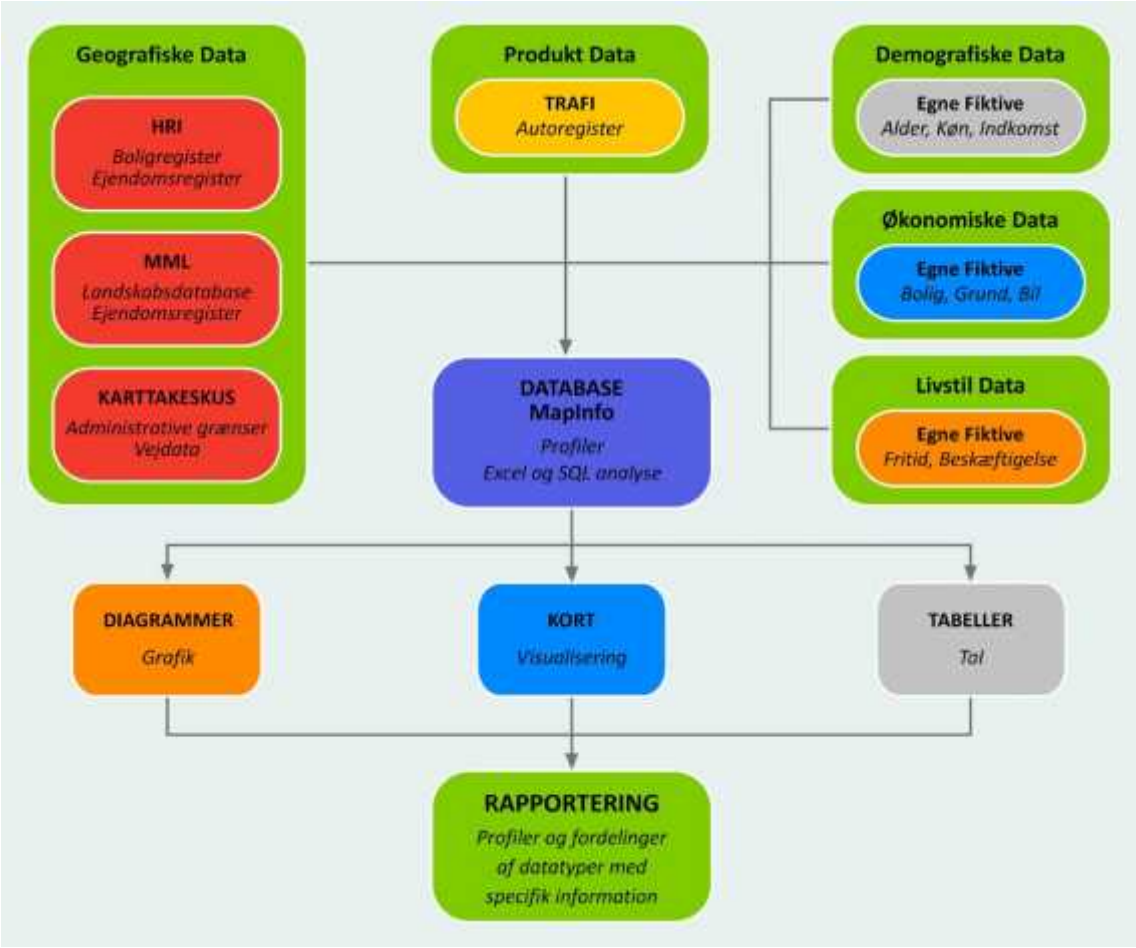
Dette afsnit er en gennemgang af det data- og informationsmateriale, der er anvendt i dette projekt og som danner grundlag for dette projektets analytiske del. Endvidere gives der en grundigere gennemgang af materialernes oprindeles, kvalitet, anvendelsesmuligheder og eventuelle fejlkilder, der måtte være tilknyttet eller forbundet med de anvendte data.

Datamaterialet er en kombination af frie data fra offentlige institutioner, data fra kortleverandører og andre relevante institutioner og firmaer, samt fiktive data. Tabelmaterialet, der danner det geografiske datagrundlag for dette projekt, er hentet fra Helsinki Kommune **(HRI)**, Den finske Landmålingsstyrelse **(MML)** og **Karttakeskus**. Produktdata (VW-Golf bilerne) er hentet fra Trafiksäkerhedsverkets **(TRAFI)** hjemmeside. Data der indeholder oplysninger om personer (demografi og økonomi) er fiktive. Via interviews med personer fra for eksempel Befolkningsregistercentralen **(VRK)**, har det været muligt at få et indblik i datastrukturen og indholdet af disse persondata. En dybere forklaring og beskrivelse af de nævnte dataleverandører med tilhørende hjemmesideadresse, findes i ***Bilag B1***. Data, der indeholder information om personer, er kun tilgængelige for firmaer eller andre institutioner, der laver statistiske analyser eller sammenkører registre for kunder, og som har en aftale om personoplysninger og integritet. Som privatperson og studerende med det foreliggende projekt in mente, har det ikke været muligt at få adgang til denne specifikke information. Denne ”problematik” vil blive diskuteret senere i dette afsnit og hvad det betyder på sigt i behandling og anvendelse af information i dette projekt.

7.2 Empiriens struktur og opbygning.

For at kunne få en så optimal udnyttelse af det indsamlede datamateriale, har jeg opbygget og struktureret min empiri som vist i ***Figur 13.*** Ved hjælp af rå og i nogle tilfælde ustrukturerede input data fra forskellige data leverandører – vist i de grønne rektangler – har jeg opbygget en *”profil relations-database”* i MapInfo – den violette rektangel – hvorfra der foretages Excel- analyse og SQL-analyse via søgninger (Query) af de indsamlede data. De forskellige resultater af mine analyser, det kan være del- eller færdige resultater, opdeles og organiseres i henholdsvis:

- *Tabeldata* (MapInfos’s egen Browser eller Excel-tabel)
- *Tematiske kort* for visualisering (”Heatmaps”, Choropleth kort eller punktspredningskort)
- *Statistik* (diagrammer)

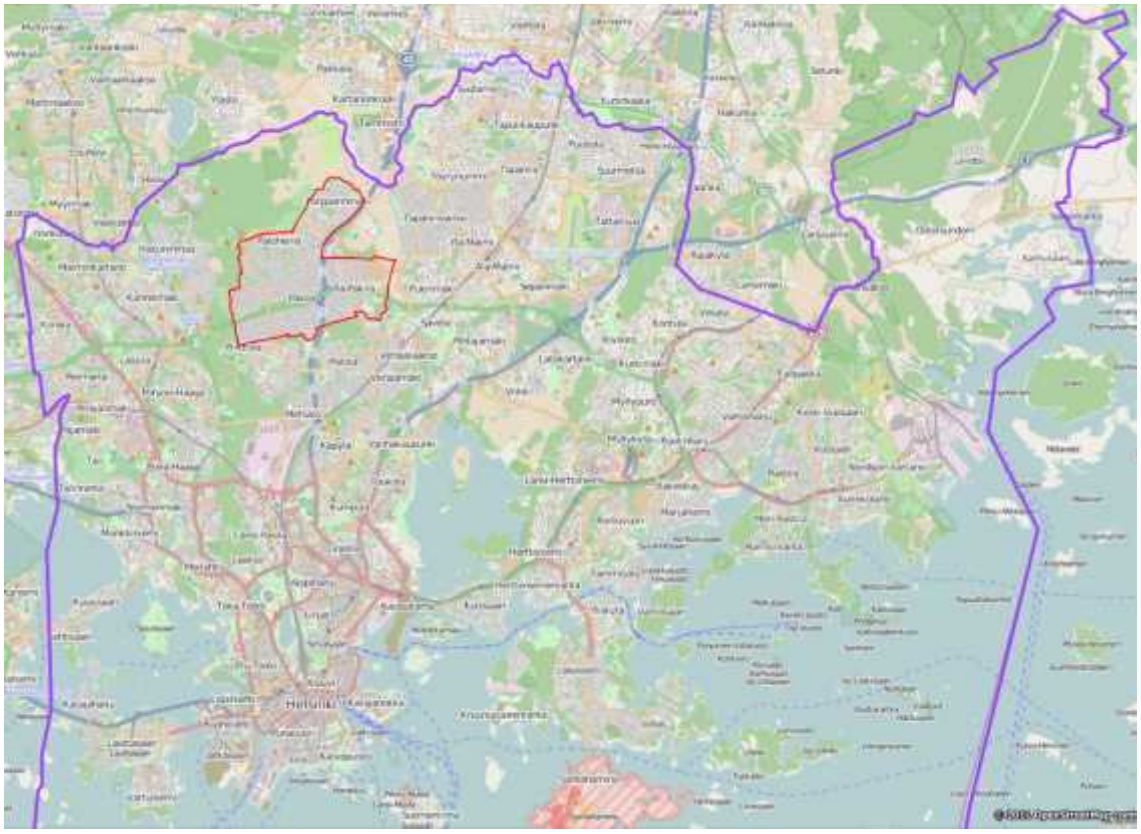


Figur 13: *Empiriens opbygning. Data fra forskellige leverandører og egne fiktive data genereres i en database i MapInfo for analyse. Resultaterne af forskellige analyser og manipuleringer findes som ny tabeller, kort eller diagrammer. På basis af de opnåede resultater, numerisk eller grafisk, udfærdiges en rapport med de ny potentielle kunder for et produkt.*

7.3 Empiriske undersøgelsesområde.

Det boligområde der danner rammen on den Empiriske undersøgelse, findes i den nordlige del af Helsinki Kommune. Området er ca. 7 km² stort og består for størstedelen af parcelhuse og rækkehuse. Der findes enkelte højhuse med maksimalt 3 etager, samt en del institutioner (skoler, dagpleje) og enkelte kommunaltekniske foranstaltninger. Der findes tillige et kolonihaveområde på ca. 1 km². PAKILA er fysisk placeret omkring to af de mest trafikerede færdselsåre i Helsinki. *Ring 1* og *Tusbyvägen*. *Ring 1* leder trafikken i hovedstadsregionen mod øst/vest og *Tusbyvägen* leder trafikken til og fra centrum af Helsinki. ***Figur 14*** viser Helsinki Kommunes kommunegrænse (tykke violette linje) og mit forsøgsområde – PAKILA – som en rød polygon i øverste venstre hjørne. Der findes andre tilsvarende områder i Helsinki kommune, der inde-

holder samme bygningsstruktur som PAKILA, men de er enten større eller mindre i område-omfang.



Figur 14: *Helsinki Kommune. Den røde polygon i øvre venstre hjørne af kortet viser hvor PAKILA er placeret. ©OpenStreetMap*

Der findes 5.338 bygninger i forsøgsområdet PAKILA. Det samlede antal boliger i undersøgelsesområdet, der er helårsbeboede, er ca. 3.778. Fordelingen af bygningerne er som følger:

- 2.980 Parcelhuse (1 eller 2 familier)
- 721 Rækkehuse (3 eller flere familier)
- 47 Højhuse (Bygninger med mere end 2 etager, antal beboere ukendt)
- 1.590 Andre bygninger (Institutioner, anden offentlig bygning, udhuse og garager)

Der findes ingen speciel grund til at jeg har valgt lige netop PAKILA som mit forsøgsområde. Det skal mere ses i lyset af at der findes en masse forskellig bebyggelse og boligtyper (private boliger) og ejendomsgrunde, der jo er det primære mål for min målgruppe. Ideen var i første række at arbejde med hele Hovedstadsregionen, men bland andet på grund af begrænsninger i Vertical Mapper i Mapinfo (antal observationer det er muligt at behandle), og størrelsen af det statistisk tabelmateriale, har jeg

valgt et mindre område, for at undgå afbrydelser samt opdeling og sammen-sætninger af forskellige beregninger og filer.

Figur 15: *Det boligområde der danner rammen on den Empiriske undersøgelse – kaldet PAKILA. Området er ca. 7 km² stort og består for størstedelen af parcelhuse og rækkehuse.*

Figur 15 og ***Figur 16*** viser mere detaljerede illustrationer af mit forsøgsområde. Der bor ca. 20.000 mennesker (2012) i forsøgsområdet. Det er ca. 2.860 personer/km². Den gennemsnitlige størrelse på boligerne er ca. 85.4 m². Den gennemsnitlige kvadratmeterpris for en bolig i PAKILA er ca. 3.150 €/m² og gennemsnitsprisen for en bolig i PAKILA er ca. 265.000.00€ = (1.975.000.00 Dkr.) i 2014 priser.

Der findes 5 folkeskoler med over og understadier (1-9 klasse). PAKILA er et sted hvor børnefamilier gerne vil bo og betragtes i almindelighed som et boligområde der ligger lidt over den gennemsnitlige levestandard i Helsinki kommune.

Figur 16: *Mere detaljeret udsnit af boligområdet fra PAKILA. De **røde** bygninger er parcelhuse, de **gule** bygninger er rækkehuse og de **orange** bygninger er højhuse. Bygninger der er vist med **gråt** er enten institutioner, andre offentlige bygninger, udhuse eller garager. De små **grønne** firkanter er geokodede adresser, hvor der findes en bolig og en ejer af en VW-Golf.*

7.4 Empiriske undersøgelsesmateriale.

Dataintegration er en proces med at kombinere data fra forskellige kilder og derigennem give brugeren en samlet visning af data. Dataintegration forekommer med stigende hyppighed, efterhånden som omfang og behovet for at dele eksisterende data, eksploderer både inden for forskningsverdenen (sammenligning af forskningsresultater), men også inden for erhvervslivet (fusionering af databaser). Under processen med dataintegration, bliver data fra flere kilder kombineret til et enkelt datasæt. (***Figur 9*** på **s. 32** under ***Teori og Metode*** er et eksempel på simpel dataintegration). Dataintegration er vigtig, fordi den forener oplysninger, der kan bruges til at træffe strategiske beslutninger, give hurtig og nem adgang til data, forbedre kvaliteten og omfanget af data, og reducerer fremtidige omkostninger til indsamling, lagring og behandling af disse. [10]

Med databerigelse eller forbedring af data, tilføjer man mere info fra andre interne eller eksterne datakilder til oplysninger, der allerede bliver anvendt. Denne proces øger den analytiske værdi til den eksisterende information. Et eksempel på databeri-

gelse, er at knytte de aktuelle kundeoplysninger i et firmas kundedatabase (register) med for eksempel købeadfærd og demografiske oplysninger fra andre kilder.

HRI Boligregister (**H**elsinki **R**egion **I**nfoshare) er et tabelmateriale, der indeholder oplysninger om Helsinki-regionen, og er åbne til borgere, virksomheder, universiteter, forskningsinstitutioner og lokale myndigheder samt den offentlige forvaltning. Oplysningerne er frit tilgængelige og kan anvendes uden nogen restriktioner eller andre begrænsninger. Det omtalte og anvendte datamaterialet i dette projekt, består af 112 felter (klasser) og 53.321 rækker (antal bygninger i Helsinki Kommune). Af de 112 felter med forskellige klasser, har jeg valg at medtage ca. 30 klasser, der har eller kunne have betydning for dette projekt. De vigtigste klasser er (ejendomsregister-nummer, adresse, boligstørrelse, boligtype, antal værelser og rum, matrikelnummer, bygnings årgang, kommune, koordinater, anden anvendelse). Dette materiale er meget omfattende og det "lever" i den forstand igennem hele analyseprocessen, forstået på den måde, at mange af de egenskaber og klasser der findes i materiale, kan anvendes i de forskellige analysemetoder, til at udvælge specifikke grupper og typer af boliger, når for eksempel resultater skal sammenliges.

Figur 17: *Udsnit af boligregisterdatabasen, fra HRI som den tager sig ud i Mapinfo. Søjlen til højre i illustrationen, viser data på én bygning i kortudsnittet. I alt findes der 112 felter med forskellige klasser.*

Grunddata (matriklen) kommer fra **Lantmäteriverket** (MML). Disse data indeholder matrikelnummer, antal bygninger, koordinater, grundstørrelse, kommune og kommunenummer). Disse data opdateres kontinuerligt. Persondata oplysninger er ikke til-

gængelig for dette materiale. Ved at kombinere matrikeldata med informationen fra HRI's Boligregister (matrikelnummer), kan adressen på hver matrikel udledes og anvendes i kombination med andre datakilder der måtte have en adresse (fællesnævner). De matrikulære data er ikke offentlige i den forstand, at de frit kan hentes fra MML's database. Karttakeskus har venligst udlånt disse data til mig for at anvende dem i dette projekt.

Produktdata (VW-Golf) er en sammensætning af to tabelmaterialer. **TRAFI's** frie data omfatter information om registrering, licenser og de tekniske specifikationer for personbiler i trafikken (køretøjsklasser M1 og M1G). Oplysningerne i biltrafikken vedligeholdes af TRAFI. Materialet dækker ikke de personlige data, der indgår i dataene på køretøjerne. TRAFI's bilregister indeholder den 30.08.2014; 2.609.480 rækker med data. Adressedata – som ikke var mulige at få udleveret fra TRAFI, er fiktive, medens de tekniske specifikationen på produktet, er originaldata fra TRAFI's egne frie data, der står til rådighed for alle. Produktdata indeholder blandt andet information om (*biltype, mærke, farve, karosserier, brændstof, motoreffekt, CO^2-udslip, førstegangsregistrering = alder, modelbetegnelse, sidepladser og kommune*).

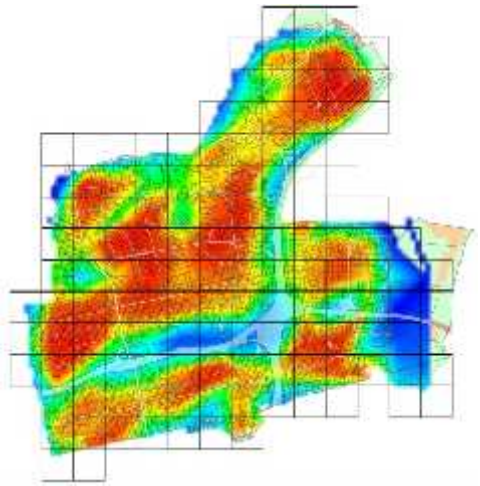
Vejdata og administrative data (postdistrikt) kommer fra **Karttakeskus**. Disse data benyttes fortrinsvis i projektet til geokodning af objekter, der har en adresse, men ikke nogen placering på kortet. Endelig har dataene også en visuel betydning for kortets layout. Vejdata er klassificeret efter (vejnavn, vejnummer på højre eller venstre sider af vejen, vejtype og kommunenummer). Der findes 6 postdistrikter i PAKILA. De er klassificeret ved (postnummer, uddelingssted, postdistriktets navn, kommune og kommunenummer). Postnummer områder fra Karttakeskus opdateres 4 gange om året. Alle rettelser til vejnettet kommer fra Finlands Kommuner, der indsender rettelser til blandt andet Karttakeskus. Postnumre og områder der indeholder postnumre (polygoner), opdateres via Post Finlands database. Denne er åben for alle borgere og firmaer.

Økonomiske data (værdidata af grund og ejendom) er fiktive data i den forstand, at de beror på en generel udledning fra statistiske data fra ejendomsmæglere og hjemmesider, med salg af fast ejendom. Jeg har foretaget en mildning af flere værdidata for hus- og grundpriser i PAKILA. Disse data er indlæst i forskellige Excel tabeller og sammensat med størrelsen og typen af de forskellige ejendoms- og matrikeldata, der figurerer i mine geografiske tabeldata. Priserne på de enkelte ejendomstyper er som følger:

- *Parcelhus med 1 familie* = 3.700 €/m^2
- *Parcelhus med 2 familier* = 3.500 €/m^2
- *Rækkehus med flere familier* = 3.500 €/m^2

- *Højhus med flere familier* = 3.700 €/m²
- Grundpris = 410 €/m²

Kvadratnetsdata med blandt andet demografiske oplysninger, kommer fra **Finlands Statistik** (Statistikcentralen). Dette materiale kan også hentes fra HRI-hjemmeside. Det er et kvadratnets-materiale, der dækker hele hovedstadsregionen med et kvadratnets vidde på 250m x 250m. Mit forsøgsområde indeholder 134 kvadratnets-felter.



***Figur 18*:** *Kvadratnets-materialet der dækker forsøgsområdet. Figuren viser den relative bebyggelsestæthed i PAKILA, hvor rød er 100 % og blå er 0 %. Bebygget område. Dette kvadratnet vil blive brugt til andre statistiske illustrationer og fremvisning af beregningsresultater, for at få et homogent billede og reference til alle beregninger der foretages med kvadratnettet.*

Kvadratnettets funktion i min rapport, vil for det meste blive anvendt til at påvise mønstre i mine data og datafordelingen i forsøgsområdet. Som det kan ses af illustrationen i ***Figur 18***, findes der visse klyngeområder i mit område af interesse. I MapInfo kan der genereres et eget kvadratnet med ønsket densitet i kvadratnettets størrelse. Dette er hensigtsmæssigt hvis jeg skal påvise eventuelle mønstre i et lille område med mange objekter. Et kvadratnet med lille vidde, for eksempel 50m x 50m ville give et mere detaljeret billede af en punktspredning og dermed være i stand til at fremhæve/nedtone et resultat, alt efter hvad der er behov for (datamanipulation). Dog skal vi huske på at, mindre kvadratnet, giver færre punkter per kvadrat og dermed et andet grundlag at evaluere et resultat ud fra.

7.5 Implementering af det empiriske undersøgelsesmateriale.

Projektets analyse kører på den måde, at jeg geokoder 205 produktdata (VW-Golf) med tilhørende adresser, på det kort som jeg har opbygget over PAKILA. Disse produktdata og deres placering på kortet, er fiktive og valgt helt tilfældigt. De tekniske specifikationer på bilen er originale data fra TRAFI's frie database. Sammen med den fiktive adresse og originale produktdata, der er integreret i ét datasæt (tabelmateriale), geokodes disse informationer ind på kortet. Derefter ser jeg på hvor punkterne "rammer" = bydel/boligtype/postnummer/anden region. De 205 punkter med attributter bliver samlet i en Excel-tabel, der nu også indeholder koordinater på hvert objekt.

Nu er jeg i stand til at manipulere og sortere mine produktdata data efter for eksempel:

- Hustype (*parcelhus, rækkehus, højhus eller noget andet*)
- Bydel (*centrum, opland*)
- Postnummer og distrikt (*centrum/opland*)
- Statistisk kvadratnet i 250m x 250m opdeling (*befolkningens aldersfordeling*)
- Statistisk kvadratnet i 250m x 250m opdeling (*boligfordeling = tætheden af boliger i et område*)
- Andre regionale opdelinger (*valgkredse*)

Derefter vil jeg lave et udtræk fra kortet med de boligtyper, der har flest objekter med VW-Golf produktet = karakteristika, og se om der findes lignende boligtyper andre steder i kortet, der kunne matcher denne søgning = (Mønstre). Her får jeg et første indtryk af, hvor og hvor mange andre steder i PAKILA, der har samme egenskaber som de oprindelige 205 punktdata, altså hvor mange andre boliger og potentielle "VW-Golf bilejere" opfylder det første krav, med hensyn til boligtype og størrelse.

Ovennævnte beskrivelse af en analyse af datamateriale er baseret på SQL-teknikker. I SQL analyse-afsnittet vil disse analyse metoder blive beskrevet grundigere, efterhånden som de forskellige analyser af datamateriale præsenteres.

De enkelte delresultater af de forskellige analyser, vil blive sammensat til en *"interaktiv rapportering"*, se ***Figur 13***, (resultattavle med grafik og tekst), som den for eksempel ville blive præsenteret for en kunde eller firma, der har bestilt en undersøgelse med de ny potentielle kunder, der findes for VW-Golf produktet i et undersøgelsesområde. Udformningen af den interaktive rapport findes i ***Kapitel 10*** om ***Perspektivering*** som ***Figur 35.***

Koordinatsystemet der anvendes i dette projekt hedder **ETRS-TM35FIN**. ETRS35FIN er et nyt finske kortprojektion system, der blev taget i brug i 2005. Det bruger den samme projektion som UTM zone 35 (central meridian: 27 ° East Greenwich), med ETRS89 som geodætisk datum. Betegnelsen ETRS-TMzn (hvor zn er zonenummer), er meget udbredt i Europa for UTM zoner. For landsdækkende kort over Finland bruges UTM projektion med zone 35 for Finland. FIN er blevet tilføjet for at fremhæve, at der er Finland. Koordinatsystemet i dette projekt har mindre betydning, dels fordi mit forsøgs område er så tilpas lille, at det ikke har nogen indflydelse på punktspredningen og det endelige resultat som sådan. At jeg har valg at anvende et "fælles" koordinatsystem som ETRS-TM35FIN skal ses i lyset af at de data jeg anvender fra forskellige leverandører, har forskellige koordinatsystemer. Alle data er transformeret via **FME** (se ***Bilag B3***) til dette fælles koordinatsystem. MapInfo understøtter dette koordinatsystem.

7.6 Afslutning.

Én af inspirationskilderne til dette projekt blev vakt for nogle år siden, da jeg læste en artikel i Jyllandsposten om biler. Her diskuterede to personer, Markedsanalytiker Anders G. Hovde fra TNS Gallup og lektor fra Markedshøjskolen i Oslo, Karl Fredrik Tangen, under artiklen *"Hvilken bil er du"?* hvilke typer af købere af en bil og dermed personer, der stod bag de mest almindelige bilmærker der findes i trafikken. Her er definitionen på en Volkswagen:

> *"De fleste kender til mærket Volkswagen - et mærke, der ifølge eksperterne, også er legitimt for folk med høj status. VW er tysk, noget som forbindes med præcision og præstige".*
>
> ***Golf:** Golf giver dig ingen sociale problemer og går an overalt. Bilen fordeler sig ligeligt mellem kvinder og mænd og udmærker sig som en såkaldt nummer to bil - en bil, som man gerne kører i til arbejde og praktiske gøremål. Er det derimod en Toyota, må han forklare kollegaerne, at det er en nummer to-bil – det behøver man ikke med en Golf".*

En fuldstændig liste og kommentar over de mest almindelige biltyper, findes i ***Bilag B2.***

Det overordnede mål for datamining processen i dette projekt vil viser, hvordan jeg er i stand til at udtrække oplysninger, fra mine forskellige geografiske og ikke-geografiske datasæt og transformere dem til en forståelig struktur, til videre brug og beslutningstagning og måske endda konklusioner. Udtrykket datamining er - til en vis grad - en forkert eller unøjagtig betegnelse, fordi målet i dette projekt er at: *udvinde mønstre og viden fra store mængder af data, og ikke udvinding i sig selv!*

De faktiske datamining opgaver er automatiske eller semi-automatiske analyser af store mængder af geografiske data, til udtrækning af hidtil ukendte og interessante mønstre, såsom grupper af dataposter – klynger og registre af anden udseende, samt afhængighedsforhold. Disse mønstre kan så betragtes som en slags sammenfatning af inputdata, og kan anvendes til yderligere analyse via SQL eller Prædiktiv analyse, til at opnå mere præcise forudsigelser af resultater.

Efter at have gennemgået det empiriske materiale, der danner grundlaget for den egentlige analyse af mine data, og dermed besvarelsen af min hypotese, vil vi nu bevæge os over i ***Kapitel 8* om *Analysen***, og begynde den egentlige organisering, forædling og beregning af data. Hensigten med dette kapitel er at vise, hvordan relativt enkle beregnings- og analysemetoder kan frembringe information og viden af tabeldata, der er relevante, for at kunne komme i betragtning som velegnet til en analyse og i sidste ende, forhåbentlig kan danne grundlag for besvarelsen af min hypotese.

8. Analysen

Projektets indsamlede empiriske grundlag danner baggrunden for den egentlige analyse. Data analyseres i relation til de fremkomne fællesnævnere i forbindelse med dataforbedring, datamining og datamanipulering. Der fortages en undersøgelse af de indsamlede data via regnearksprogrammet Microsoft Excel og GIS-programmet MapInfo/Vertical Mapper. Metoderne der vil blive anvendt i analysen af tabeldatamaterialet er Excel-analyse og SQL-analyse. Endvidere diskuteres de opnåede resultater i forhold til specialets teoretiske referenceramme samt problemstilling. Datakvalitet og datafangst vil også blive sat i relation i denne diskussion. Fejlkilder i data og deres betydning igennem analyseprocessen vil også blive behandlet.

8 ANALYSEN

8.1 Indledning.

Analyse er processen med at observere og nedbryde et komplekst emne eller stof i mindre dele, for at få en bedre forståelse af det. De analyser jeg vil foretage i dette kapitel, handler for en stor del om at vise, hvordan jeg med relativt simple søge- og manipulationsmetoder, er i stand til at omforme data i tabelform til information og viden, der kan bruges til at besvare min hypotese, og i sidste ende anvendes som et beslutningsgrundlag, for at foretage en handling – eller ikke. Resultaterne af de forskellige metoder er ikke entydige, forstået på den måde, at der er rum for fortolkning af resultatet, samt det faktum, at flere af mine grundlæggende data er baseret på fiktive observationer. [13]

Kapitlet om Analyse er opbygget således, at jeg laver en generel statistisk oversigt over de data, som jeg har indsamlet. Det vil sige bildata og boligdata. Her vil jeg lave en oversigt over de forskellige variationen i fakta der fremkommer, ved at sortere og klassificere mine data og vise dem på grafisk form. Til hver analysemetode er der illustreret en analyseplan. Det vil sige at jeg på forhånd skitsere, hvordan jeg har tænk mig at opbygge min analyse, for at opnå et resultat, som jeg tror, kan danne grundlag for besvarelsen af min hypotese. Analyseplanen bygger på princippet om *Input – Proces – Output* som blev skitseret i ***Kapitel 5*** under ***Problemanalysen*****.** Den skematiske fremgangsmåde for analyseplanen bliver fulgt op af en detaljeret beskrivelse af selve analyseforløbet, med kommentarer, illustrationer og resultater. I det omfang jeg finder det nødvendigt, vil jeg refereret til forskellige teorier og metoder for at underbygge mine valg eller beslutninger for at foretage en handling.

Fordi jeg ikke ved nok om statistik og den slags, så at jeg kan skrive helt præcist, hvilke typer af sammenhænge, forholde, intervaller samt kvaliteter af målinger, der skal foretages eller vises, for at resultatet er signifikant til besvarelse af min hypotese, vil jeg i kapitlet om Analyse ikke komme ind på nogen specifik beregningsmetoder og beregningseksempler, for at validere mine data, i forhold til kvalitet og resultat. Derimod vil jeg nævne og beskrive en del af de mest gængse metoder og filosofier inden for dette område, for at sætte mine egne metoder i relation til det jeg søger – nemlig sammenhængen mellem biltyper og boligtyper.

8.2 Generelt om GIS-analyse.

GIS-analyse er en proces til at se på geografiske mønstre i data og relationer mellem forskellige egenskaber. Metoderne kan være meget enkle, sommetider kan det bare være at lave et kort, der i sig selv er en analyse. Mere komplekse analyser kan invol-

vere modeller, der efterligner den virkelige verden ved at kombinere mange datalag. Ofte startes en analyse med at finde ud, hvilke oplysninger der er brug for. Dette udføres ofte i form af et spørgsmål. Hvor findes der flest VW-Golf Variant i et bestemt boligområde? Hvilke huse findes inden for 500 meter fra en detailhandel? At være så specifik som muligt om det spørgsmål, som vi forsøger at svare på, og illustrere dem i *Input-Proces-Output* skemaet, vil hjælpe med til at beslutte, hvordan en analyse tilrettelægges, hvilke metoder der skal anvendes, samt præsentationen af resultatet eller resultaterne. [7]

Andre faktorer, der har indflydelse på udførelsen af analysen er, hvad resultatet skal bruges til, og hvem der skal bruge det. Analyse kan være, at undersøge data på egen hånd for at få en bedre forståelse af, hvordan et sted udvikles (byudvikling), få svar på en hypotese (sammenhænge), optimering (bedste placering af en detailhandel), eller at præsentere en undersøgelse i den offentlige debat. Den type data og egenskaber, jeg arbejder med, har også været med til at bestemme, hvilken analysemetode(r) jeg tror, der egner sig bedst til at få et svar eller resultat af en undersøgelse og en hypotese. Omvendt, hvis jeg skal anvende en bestemt metode til at udlede en bestemt mængde information, har jeg måske behov for at tillægge ekstra data eller information til mine eksisterende data, for at opnå denne nye ekstra information eller viden, som jeg tror jeg kan have gavn af eller kan være nyttig i en beslutningsproces. Jeg er altså nødt til at vide, hvad jeg har af typer af funktioner og egenskaber i mine data, og hvad jeg har brug for at få eller skabe. Generering af nye data kan ganske enkelt betyde, at beregne nye værdier i data tabellen. [4]

Resultatet af en analyse kan vises som et kort, værdier i en tabel eller et diagram som beskrevet i ***Kapitel 7*** om ***Empiriske materiale***. Her er det vigtig at vide, hvilke oplysninger der skal medtages på et kort, en tabel eller en graf, og hvordan disse oplysninger præsenteres på den mest hensigtsmæssige og informative måde (tekst og teksttype, farver, symboler, betoning samt størrelsen af illustrationen), for at værdierne på bedste måde, præsentere et resultat. At se på resultaterne via de forskellige præsentationsmuligheder, kan også hjælpe med at se og vurdere, om oplysningerne er gyldige eller nyttige, eller om analysen skal gentages med ved hjælp af nye eller forskellige parametre, eller måske en hel ny eller anden analysemetode. GIS gør det relativt let at foretage disse ændringer og lave nye output. [1]

Geografiske karakteristika optræder og præsenteres på forskellige måder. Disse karakteristika kan være diskrete, kontinuerlig, eller opsummeret efter område. For diskrete steder og linjer, kan den faktiske placering blive udpeget. På et givet sted, er egenskaben enten til stede eller ikke. Matrikler eller ejendomsgrænser, der for eksempel er illustreret med forskellige farver efter værdi, størrelse eller beliggenhed i et område (polygon), er et godt eksempel på diskret karakteristika. [3] ***Se Figur 28.***

Kontinuerlige fænomener som nedbør eller temperatur kan findes eller måles overalt. Disse fænomener dækker hele det område der kortlægges, det vil sige, at der ikke er ”huller” eller mangler i dataene. Kontinuerlige data består af en række punkter, enten med en ensartet afstand eller uregelmæssigt afstand. GIS-systemet bruger disse punkter til at tildele værdier til området mellem punkterne, i en proces, der kaldes interpolation. For eksempel kan du oprette et kort over grundværdierne for et bestemt område - postdistrikt, bydel eller kommune - ved at interpolere midtpunkterne (centroids) af alle de matriklerne i et område, der har en værdi. Højdekurver i et topologisk kort eller en ”heatmap” over et boligområde med forskellige ejendomsværdier, er et godt eksempel på kontinuerlige fænomener. [14] ***Se Figur 33***.

Opsummerede data repræsenterer tællingerne eller tætheden af individuelle karakteristika i et afgrænset område. Eksempler på funktioner opsummeret efter et område, kan for eksempel være antallet af bygninger i et postdistrikt, eller den samlede længde af et vejnet i en kommune. Værdien af dataene tilhører hele området og ikke et specifikt punkt inde i polygonen (postdistriktet eller kommunen). En del data, findes allerede opsummeret efter område. Det gælder især demografiske data, som består af samlede værdier (samlede befolkning, samlede antal husstande, og så videre) eller en procentsats baseret på en kategori. ***Se Figurerne 19 og 20.***

Hvis en egenskab allerede er tildelt en kode, der henfører den til et bestemt område eller polygon, er det kun et spørgsmål om at lave lidt statistik på vores tabeldata, som for eksempel at opsummere den samlede omsætning for alle virksomheder i hvert postdistrikt, som også har et postnummer og dermed en kode (attribut). Et andet eksempel han være den demografiske fordeling af aldersgrupper for personer i et kvadratnet. Disse informationer kan bruges i en kortlægning - visualisering - til af vise eller endda ”afsløre” mønstre i tabeldataene. Hvis egenskaberne ikke er tildelt en kode for de områder eller polygoner, som jeg ønsker at undersøge, analysere eller opsummere, kan jeg tillægge et nyt tomt datalag i mit GIS-system og detektere - tilføje - disse data til polygonen eller kvadratnettet og kalde det for eksempel *”aldersfordeling efter boligtype”*. ***Se Figur 31.***

Kategorier er grupper af samme ting eller emne. De hjælper os med at organisere og få mening ud af vores data. Alle funktioner med den samme værdi for en kategori, er ens på en eller anden måde, og forskellige fra funktioner med andre værdier for denne kategori. Veje kan kategorisere som; motorveje, hovedveje eller lokale veje. Den *kategoriske værdi* af en ting eller emne, kan repræsenteres ved hjælp af numeriske koder eller tekst. Tekstværdier er ofte forkortelser, for at spare plads i tabellen. Kategorier er kolonneopdelte i tabelmaterialer.

Rangorden er en rækkefølge, hvor et sæt af data eller egenskaber går forud for et eller flere andre sæt. Hvis der tillige er en præcis afstand mellem hver enkelt data(sæt), kaldes værdierne på skalaen for *numeriske variabler*. Numeriske variabler kan rubriceres på baggrund af reglerne for stokastiske eksperimenter. Hvis der ikke er præcise afstande mellem værdierne, kaldes værdierne på skalaen *ordinale*.

Antal og Mængde vise det samlede antal af en ting eller emne. Antallet er det faktiske antal funktioner på kortet. En mængde kan være noget målbar, der er forbundet med en funktion, såsom antallet af boliger i et defineret område, for eksempel PAKILA. Ved at anvende antal eller mængde, fås den faktiske værdi af de enkelte funktioner samt størrelsesorden i forhold til andre funktioner (forskellige boligtyper).

Forhold viser forholdet mellem to størrelser, og fremkommer ved at dividere en mængde med en anden, for hver funktion. For eksempel, at dividere antallet af mennesker i hvert område, med antallet af husstande, giver os det gennemsnitlige antal personer pr. husstand. Brug af forhold udligner forskelle mellem store og små områder eller områder, der har mange funktioner, og de, der har få, således at kortet eller den grafiske fremstilling, mere præcist viser fordelingen af funktionerne. To særlige forhold er proportioner og tæthed. Proportioner viser, hvad en del af en samlet hver værdi er. Proportioner er ofte præsenteret som procentdele (andelen multipliceret med 100). Tæthed viser fordelingen af funktioner eller værdier pr. arealenhed. For eksempel ved at dividere befolkningen i en kommune med dets areal i kvadrat kilometer, vil vi få en værdi for personer pr. kvadratkilometer. [14]

En vigtig del i en GIS-analyse, er at arbejde med og bearbejde tabeldata, der indeholder attributværdier og summariske statistikker. At vælge, at beregne og at opsummere, er de tre mest almindelige fremgangsmåder, i forbindelse med manipulering af tabeldata.

At **Vælge,** vil sige at arbejde med en delmængde, eller at tildele specifikke værdier til en del eller udvalgt mængder af funktioner. For eksempel, det at vælge boliger fra en mængde/tabel med bygninger. SQL er et godt eksempel på hvordan man vælger værdier.

Attribut = værdi

For at vælge bolig, skal du ville angive:

Anvendelse = BOLIG

Hvor Anvendelse er attributnavnet, og BOLIG er værdien. Typiske operationer er "er lig med", ”større end” (>), ”mindre end” (<), og ”ikke lig med” (<>).

Udtryk kan også sammensættes for at vælge funktioner, der opfylder en række kriterier. For eksempel, for at finde ejendomsgrunde større end 2 hektar, ville jeg angive:

Arealanvendelsen = Matrikel og størrelse > 2

Skal funktioner opfylder mindst en af flere værdier eller kriterier, bruges "eller". For eksempel, for at vælge ejendomsgrunde med boliger og grunde med industri, ville jeg skrive en SQL-sætning der hedder:

Arealanvendelsen = BOLIG eller Arealanvendelsen = INDUSTRI

Beregning af attributværdier udføres for at tildele nye værdier til funktioner i datatabellen. Først tilføjes et nyt felt i tabellen, og derefter tildeles de nye værdierne for attributterne til hver funktion. Værdierne kan tildeles med *Rangorden* (fra mindste til største værdi), eller tildeles som *Forhold*, for eksempel, boligens samlede størrelse divideres med antal lejligheder, for at få den ”rigtige” størrelse for boligarealet (*Parcelhus med 1 familie*; *Parcelhus med 2 familier*, *Rækkehus med flere familier* og så videre). Eller - jeg kan beregne den gennemsnitlige huspris pr. kvadratmeter, ved at multiplicere husets størrelse (beboelsesareal eller samlede areal) med en gennemsnitlig kvadratmeterpris for boliger i PAKILA. Jeg får en ny kolonne med information, basseret på del-information fra tabellen, men også indhentet information udefra. [14]

Opsummere vil sige at arbejde med tabeldataler, ved at opsummerer værdierne for specifikke attributter for at få statistik. I nogle tilfælde fås en enkelt værdi, såsom en hel eller en gennemsnitlig (middel) værdi og i andre tilfælde laves en ny tabel, der angiver visse værdier for hver type (kategori), plus en optælling af funktioner. Dette kaldes en frekvens. [15]

Almindeligvis betyder Bias en hældning eller fordom. Vi har en tendens til at opdage ting, der understøtter de hypoteser eller holdninger, vi i forvejen har. Da geografiske data repræsenterer den virkelige verdens fænomener og processer, kan visse karakteristika føre til fysisk skævhed eller Bias dataene. To almindelige former for partiskhed, er geografisk autokorrelation (gentagne mønstre) og korrelerede datalag.

Geografisk autokorrelation: Med geografiske data, er det ofte sådan, at tingene i nærheden af hinanden er mere ens end tingene langt fra hinanden (værdien af et hus i et bestemt område af en by, kriminaliteten i en bydel, vegetationen i en nationalpark eller temperaturfordelingen i et land). Dette fænomen er kendt som geografisk auto-

korrelation. Mange analyser baserer sig på data i stikprøver. Hvis nogle prøver er tæt på hinanden, og andre spredt ud, og vores data er rumligt autokorreleret, kan de prøver, der er tæt på hinanden, overbetone værdierne i disse områder, hvilket fører til forkerte resultater. [2]

Korrelation (eller "ko-relation", "sam-relation") er i statistik et mål for *sammenhængs graden* mellem et sæt af to variable/målinger. En høj korrelation betyder, at det ene sæt af variable kan forudsiges fra det andet og omvendt, eller at begge variable i en vis udstrækning, er et resultat af samme fælles årsag, eller at de evt. er kommet til at dele et fælles betydningsindhold (rent semantisk). Korrelation betyder således ikke nødvendigvis, at der er en direkte årsagssammenhæng mellem to variabler. For eksempel er *vægt* og *højde* to variabler omkring et menneske, der i en vis udstrækning er afhængige af hinanden – højere personer er ofte tungere end lavere personer. Men afhængigheden er ikke entydig. Personer med samme højde kan som bekendt godt have forskellig vægt. Ikke desto mindre er det i dette tilfælde tydeligt for enhver, at der *gennemsnitligt* kan iagttages en vis relation mellem højde og vægt blandt mennesker. Størrelsen af denne relation beregnes ved hjælp af en matematiske formel og ender med et slutresultat, kaldet en korrelationskoefficient (eller **"r"**), som varierer fra -1,00 til +1,00. Jo nærmere **r** er til yderpunkterne +1,00 eller -1,00 desto større eller tættere er sammenhængen mellem de to variabler. [15]

Regressionsanalyse er en gren af statistikken, der undersøger sammenhængen mellem en afhængig variabel og andre specificerede uafhængige variable. Lineær regressionsanalyse går ud på, at finde ligningen for den rette linje, der passer bedst til givne målepunkter. Man forsøger altså at opstille en matematisk sammenhæng, mellem en række observerede størrelser, ved at tage højde for den statistiske usikkerhed. Når modellen er fastlagt, kan man benytte den til at forudsige værdien af den afhængige variabel, ud fra andre værdier af baggrundsvariablene, og desuden kan modellen i sig selv give indsigt i de dybereliggende faktorer, bag variablenes sammenhæng. [15]

Den matematiske model af variablernes sammenhæng kaldes *regressionsligningen*. Den afhængige variabel modelleres som en stokastisk variabel på grund af den indbyggede usikkerhed angående dens værdi, givet værdierne af de andre uafhængige variable. En regressionsligning indeholder estimater (dvs. skøn), af en eller flere *regressionsparametre* (konstanter), som kvantitativt forbinder den afhængige og de uafhængige variable. Parametrene estimeres fra det givne datasæt. [15]

Brug af regressionsanalyse inkluderer forudsigelser, modellering af årsagsbestemte forhold samt test af videnskabelige hypoteser om sammenhæng mellem variable. Lineær regressionsanalyse bygger på antagelsen om, at sammenhænge mellem variabler kan beskrives lineært. Det betyder, at grafen for regressionsligningen vil være en

ret linje, hvis der kun er én baggrundsvariabel, eller en hyperplan, hvis der er flere baggrundsvariable. Forudsigelser uden for dataenes rækkevidde kaldes ekstrapolation og er mere risikable. Hvis man eksempelvis har data for en persons højde målt hver anden måned, vil det, at udregne den estimerede højde for hver måned ud fra modellen, være interpolation. Hvis man bruger modellen til at udregne højden nogle måneder frem i tiden, er der tale om ekstrapolation.

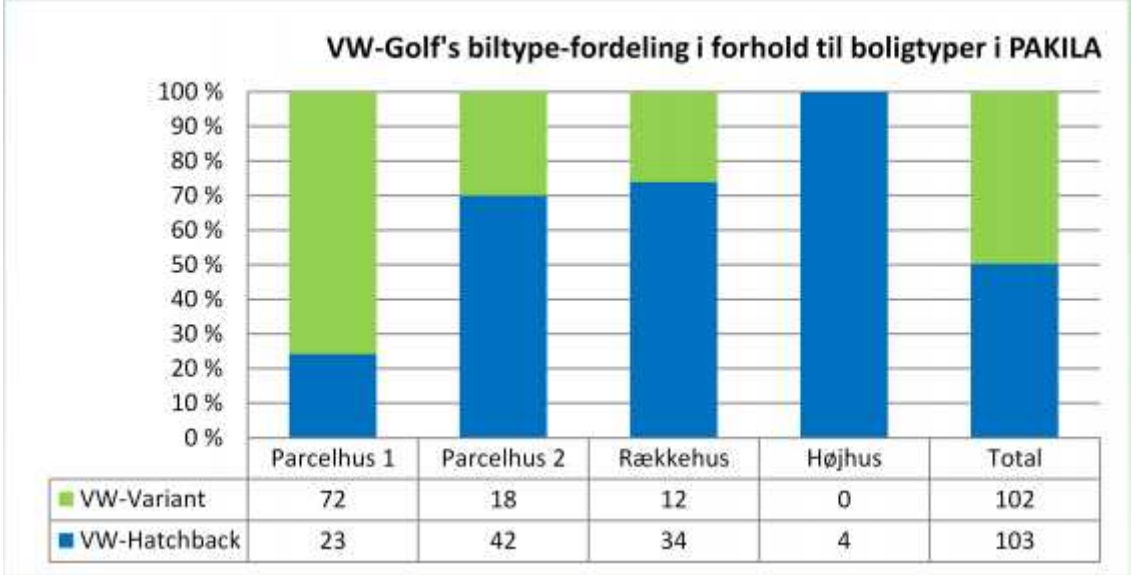
Mindste kvadraters metode benyttes blandt andet i regressionsanalyse, for eksempel til at finde den bedste rette linje, der beskriver en lineær sammenhæng mellem to dataset. Metoden minimerer her summen af kvadraterne på residualerne (de lodrette afstande mellem de enkelte punkter og den rette linje). Dette kan gøres grafisk ved at tegne punkterne fra et datasæt ind i et koordinatsystem og tegne en ret linje, som ligger nogenlunde der, hvor man tror, den bedste rette linje kan ligge. Herefter tegnes de lodrette afstande mellem punkterne og linjen. Disse punkter kvadreres så. Afstanden mellem punktet og linjen er den ene side i et kvadrat. Man kan så rykke rundt på linjen, ved at rotere den og parallelforskyde den, indtil det samlede areal af alle kvadraterne er mindst muligt. Heraf navnet ”Mindste kvadraters metode”. [15]

8.3 Fakta om mine egne data.

Dette afsnit er en lille forhistorie og illustration af de data, der anvendes i de to anvendte analyser i dette projekt. Jeg har valg at vise dem her, fortrinsvis på tabelform og derefter igennem de to analyse metoder og kontinuerlig referere til de viste figurer. Tabeller og figurer er fremkommet ved at organisere og manipulere mine indsamlede data - fortrinsvis fra HRI-tabeldataene - samt en del af mine egen fiktive data.

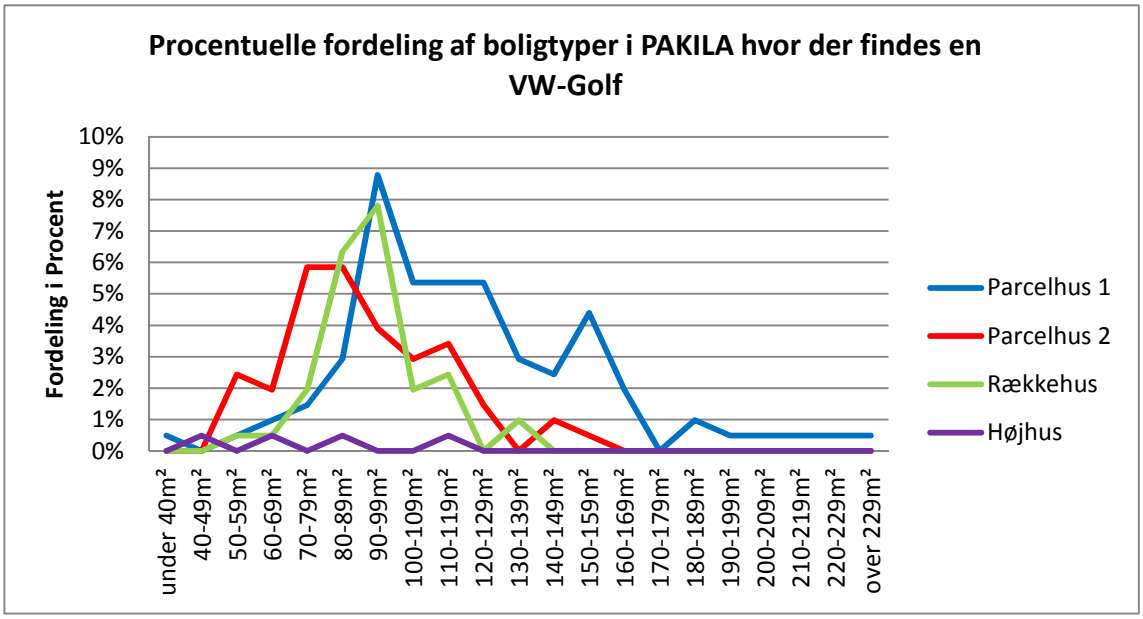
Den første illustration, ***Figur 19***, viser fordelingen af mine 205 VW-Golf biltyper (Variant og Hatchback) i PAKILA. Disse biler der er ”spredt” ud over forsøgsområdet på de 4 forskellige boligtyper der findes i PAKILA, som de er organiseret i HRI-tabellen fra Helsinki Kommune. Klassificeringen er derfor ikke manipuleret eller på nogen måder omformet, men afbilder den virkelige boligtype, som den findes i HRI-tabelmaterialet. Nominelt er opdelingen af mine biltyper således:

- *Parcelhus med 1 familie* = 72 VW-Golf Variant
- *Parcelhus med 1 familie* = 23 VW-Golf Hatchback
- *Parcelhus med 2 familier* = 18 VW-Golf Variant
- *Parcelhus med 2 familier* = 42 VW-Golf Hatchback
- *Rækkehus med flere familier* = 12 VW-Golf Variant
- *Rækkehus med flere familier* = 34 VW-Golf Hatchback
- *Højhus med flere familier* = 0 VW-Golf Variant
- *Højhus med flere familier* = 4 VW-Golf Hatchback



Figur 19: *Fordelingen af VW-Golf biltyper i mit forsøgsområde PAKILA i forhold til registrerede boligtyper.*

Ud fra den "visuelle" fremstilling af biltype-fordelingen i ***Figur 19*** ses det, at VW-Golf Variant for en stor dels vedkommende, findes i *Parcelhus med 1 familie*. VW-Golf Hatchback findes derimod næsten ligeligt fordelt på *Parcelhus med 2 familier* og *Rækkehus med flere familier*.

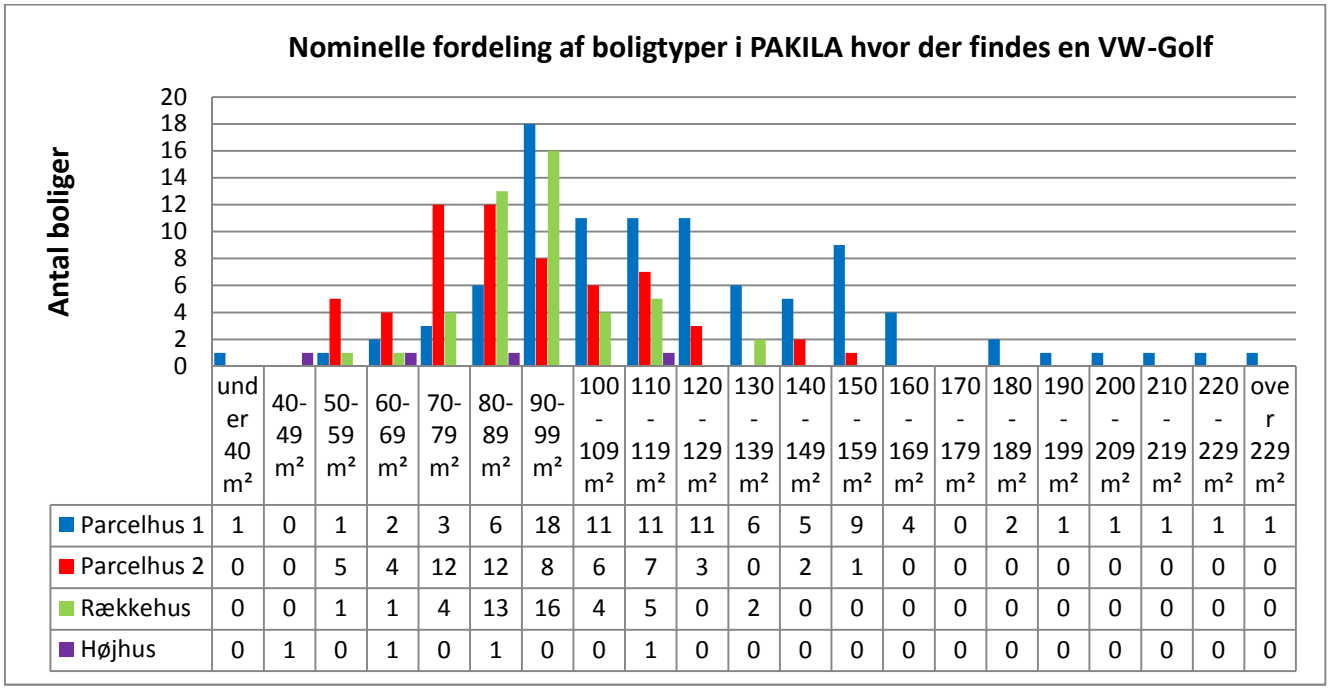


Figur 20: *Den procentuelle fordeling af 205 boliger i PAKILA i m². Kurverne illustrerer ganske tydeligt, hvor den største koncentration af "interessante m²" findes i hver boligtype.*

Figur 19 og ***Figur 20*** viser den nominelle og procentuelle fordeling af boligtyperne i PAKILA, hvor der findes en VW-Golf bil. Disse data er fremkommet ved at klassificere

størrelsen af de fire forskellige boligtyper i et interval på 10m² (*Forhold.*) Der findes ikke nogen speciel grund til at det er 10m² der er separator, men med en forholdsvis lille mængde boliger, og en ikke for stor spredning i antal kvadratmeter boligareal, viser denne opdeling ganske tydeligt, hvilke boligtyper der er mest repræsenterede i PAKILA.

Ud fra kurvebilledet i ***Figur 20*** og stolpediagrammet i ***Figur 21*** ses det ganske tydeligt, at der er en *Gaussisk fordeling* af størrelsen af de forskellige boligtyper. *Parcelhus med 1 familie* har en større spredning med hensyn til antal fordelte kvadratmeter. *Parcelhus med 2 familier* har næsten samme afbildning som *Parcelhus med 1 familie*. Kurven er bare "rykket ca. 20m²" mod venstre, der indikere, at familier der bor i *Parcelhus med 2 familier*, har færre kvadratmeter at rutte rundt med. Det kunne for eksempel også indikere, at der boede færre antal personer som gennemsnit i disse hustyper. Rækkehusene har en mere jævn *Gaussisk fordeling* omkring de 80 - 90m². ***Figur 21*** viser samme fordeling som ***Figur 20.*** Her er det den nominelle værdi, der vises som illustration for spredningen af kvadratmeter, på de forskellige boligtyper.



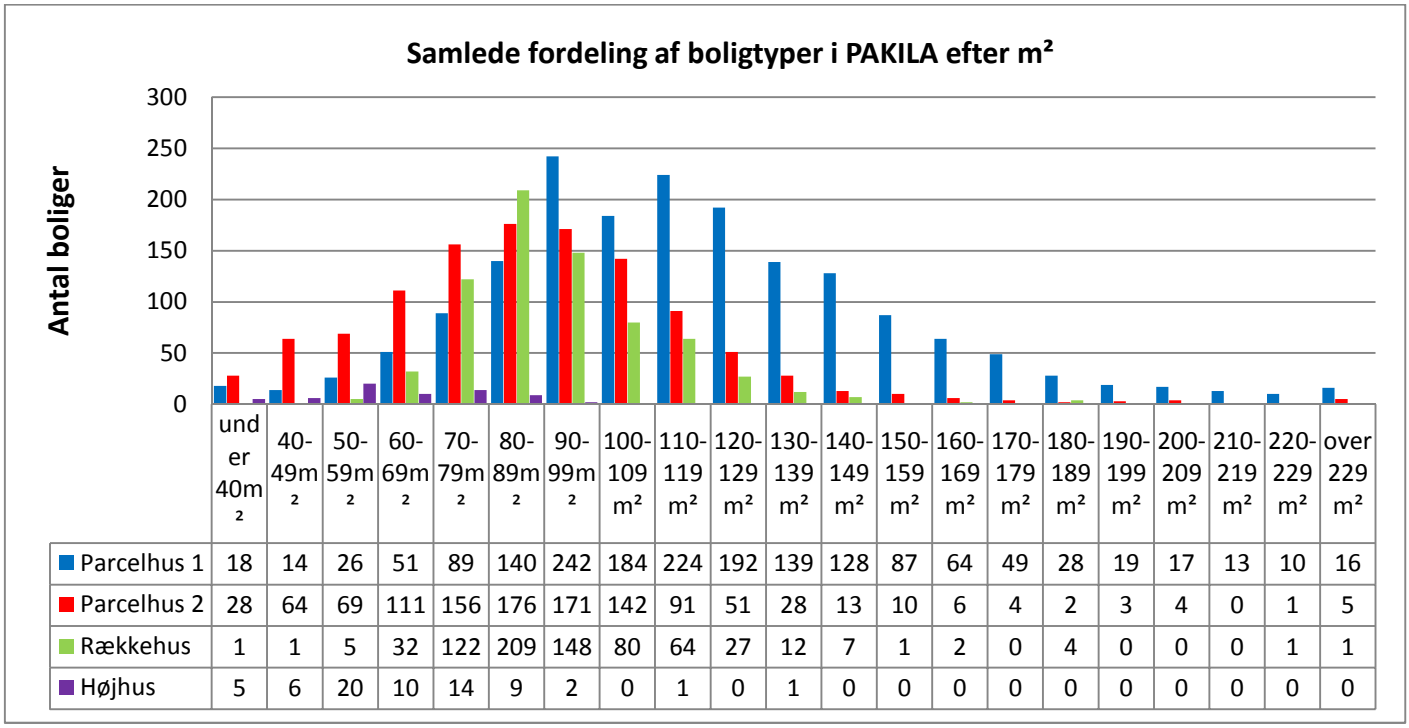
	under 40 m²	40-49 m²	50-59 m²	60-69 m²	70-79 m²	80-89 m²	90-99 m²	100-109 m²	110-119 m²	120-129 m²	130-139 m²	140-149 m²	150-159 m²	160-169 m²	170-179 m²	180-189 m²	190-199 m²	200-209 m²	210-219 m²	220-229 m²	over 229 m²
Parcelhus 1	1	0	1	2	3	6	18	11	11	11	6	5	9	4	0	2	1	1	1	1	1
Parcelhus 2	0	0	5	4	12	12	8	6	7	3	0	2	1	0	0	0	0	0	0	0	0
Rækkehus	0	0	1	1	4	13	16	4	5	0	2	0	0	0	0	0	0	0	0	0	0
Højhus	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Figur 21: *Fordelingen af 205 boliger i PAKILA efter boligstørrelse i m², og hvor der findes en VW-Golf Variant eller VW-Golf Hatchback.*

Figur 22 viser den samlede fordeling af kvadratmeterne i PAKILA, fordelt ud over de 4 forskellige boligtyper. Det er værd at bemærke, at fordelingen er ganske nær den "tilfældige" fordeling af boligtyper efter Biltype. Ved at se på ***Figur 21*** og ***Figur 22*** med ***Figur 19*** in mente, kan vi allerede her sige noget om sammenhængen mellem bolig-type, boligstørrelse og biltype. Endvidere kan vi ud fra de tre illustrationer se, at fokus på boligtype for VW-Golf Variant, skal koncentreres omkring *Parcelhus med 1 familie*, der

har en størrelse på 90 - 130m². Den gennemsnitlige størrelse på et parcelhus, hvor der findes en VW-Golf Variant er ca. 120m², medens VW-Golf Hatchback har sine tilhængere i *Parcelhus med 2 familier*, der har en boligstørrelse på 70 - 90m². Den gennemsnitlige størrelse for et *Parcelhus med 2 familier* og den lykkelige ejer af en VW-Golf er ca. 90m². *Rækkehus med flere familier*, der ejer en VW-Golf, bor i en bolig mellem 80 - 100m², der har en gennemsnitlig størrelse på ca. 93m². *Højhus med flere familier* der ejer en VW-Golf, er i klart mindretal. Her bor familier der ejer en VW-Golf meget spredt, fra 40 - 120m². Den gennemsnitlige størrelse for en højhuslejlighed er ca. 73m²

De samme gennemsnitstal for fordelingen af kvadratmeterne på de 4 boligtyper i hele PAKILA er: *Parcelhus med 1 familie* er 118m². *Parcelhus med 2 familier* er 88m². *Rækkehus med flere familier* har en gennemsnitlig størrelse på 93m² og endelig har alle højhusene, der findes i PAKILA en gennemsnitlig kvadratmeterstørrelse på 65m².



	under 40m²	40-49m²	50-59m²	60-69m²	70-79m²	80-89m²	90-99m²	100-109 m²	110-119 m²	120-129 m²	130-139 m²	140-149 m²	150-159 m²	160-169 m²	170-179 m²	180-189 m²	190-199 m²	200-209 m²	210-219 m²	220-229 m²	over 229 m²
Parcelhus 1	18	14	26	51	89	140	242	184	224	192	139	128	87	64	49	28	19	17	13	10	16
Parcelhus 2	28	64	69	111	156	176	171	142	91	51	28	13	10	6	4	2	3	4	0	1	5
Rækkehus	1	1	5	32	122	209	148	80	64	27	12	7	1	2	0	4	0	0	0	1	1
Højhus	5	6	20	10	14	9	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0

Figur 22: *Fordelingen af alle boliger - 3.670 - i PAKILA efter størrelse i m². Alle andre bygninger end beboelse er renset bort.*

8.4 Excel-analyse.

Min idé med Excel analysemetoden er dels at vise, hvordan "almindelige" mennesker uden nogen forudgående viden om GIS og GIS-systemer, kan drage fordel af den geografiske information, der er til rådighed i frie offentlige data og andre datakilder. Det kan være som ekstra eller supplerende information til egne data, for at underbyggen et beslutningsgrundlag og dermed forstærke grundlaget for beslutningsprocessen, der

kunne føre til en handling, investering eller måske forkastelse. Dels er ideen med Excel-analysen at se, om der findes en sammenhæng mellem mine Biltyper og Boligtyper, ved at se på de forskellige resultater, som jeg har lavet med mine tabeldata. Endelig er ideen med analysen at vise hvordan de første to trin i DIKW-hierarkiet hænger sammen og hvordan sammenhænge og mønstre i tabeldata kan kombineres, via enkle funktioner og metoder og derigennem være med til at udtrækker information og besvare spørgsmål som; *hvem, hvad hvor eller hvornår*. Eller mere enkelt - nemlig, hvordan data bliver til information, ved at fortolke og konvertere dem til brugbar og meningsfuld information.

Som nævnt i ***Kapitel 5*** under ***Problemanalyse***, foregår den praktiske udførelse af Excel-analysen ved, at sammensætte to tabeller med data og finde værdier i disse tabeller, der kan besvare min hypotese. Det sker ved primært at se om der er en korrelation (trend) mellem biltypen (VW-Golf Variant eller VW-Golf Hatchback) og boligtyper (Parcelhuse, Rækkehus eller Højhus). Sekundært om der i de fundne observationer, er en indikator i de beregnede data som for eksempel boligstørrelse, der kan frekventere søgningen til en bestemt bilmodel, boligtype eller boligstørrelse. Fællesnævneren for søgningen er adressen på bilejeren.

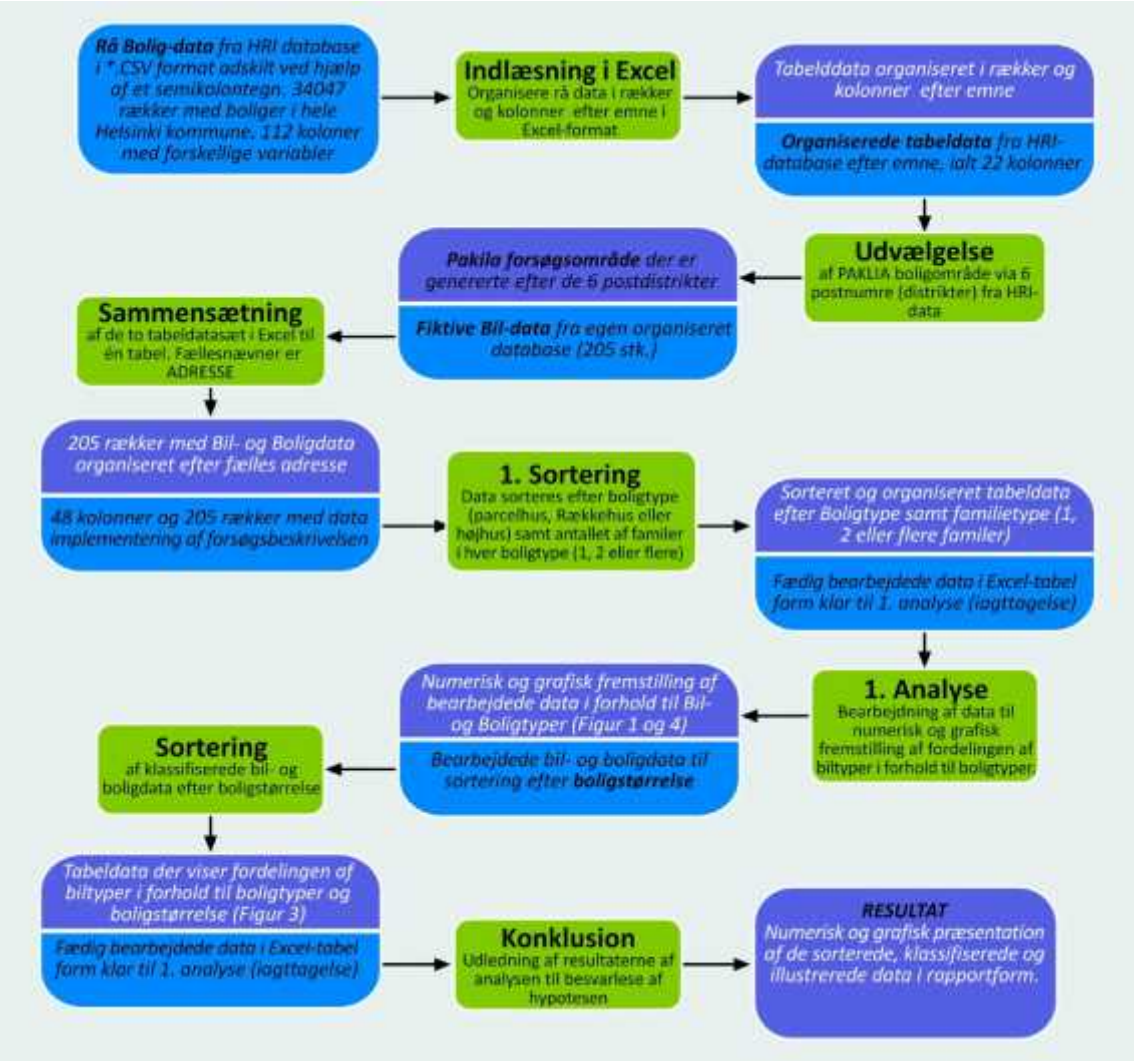
Jeg skal med det samme gøre opmærksom på, at jeg ikke er nogen speciel ”ørn” i anvendelsen af Excel programmet og manipulering af tabeldata i det hele taget, men ved hjælp af lidt kreativ tænkning og brugen af hjælpefunktionerne, der tilbydes i Excel, er det muligt at komme ganske langt i analysen af data og grafisk fremstilling af diverse histogrammer. På den måde er jeg et godt eksempel på et ”almindeligt” menneske.

I tekst, foretages søgningen på følgende måde:

Åbner HRI-tabel datamaterialet. Der findes 52.307 rækker med bygninger i hele HRI-Excel-tabellen og 112 kolonner med forskellige genskaber for disse bygninger. Af de 52.307 bygninger i HRI-tabellen findes der 34.047 boliger. Mit forsøgsområde består af 6 postdistrikter med postnumrene (00630, 00640, 00660, 00670, 00680 og 00690). Ved at sortere efter disse 6 postnumre i HRI-Excel tabellen, kan jeg skabe mit forsøgsområde - PAKILA. I PAKILA findes der 4.490 bygninger med forskellige funktioner, lige fra beboelse til offentlig service, handel og let industri. Af disse er 3.690 boliger fordelt over følgende boligtyper: (2.905 Parcelhuse, 717 Rækkehuse og 68 Højhuse).

Herefter åbnes min anden Excel-tabel, der indeholder data om mine VW-Golf biler. Der findes 26 kolonner med forskellige egenskaber der beskriver mine VW-Golf biler. Specifikationer på VW-Golf er hentet via TRAFI’s register for bilmodeller i Finland. Adres-

sen på ejeren af disse 205 VW-Golf er som tidligere nævnt fiktive. Adresserne er blevet genereret, ved at geokode de 205 VW-Golf biler ind på mit kortmateriale over PAKILA.



Figur 23: *Via denne processering af data i Excel, er jeg i stand til at lave en analyse af mine indsamlede data, der giver et første resultat til besvarelse af min hypotese*.

De to tabeller sammensættes via en funktion i Excel kaldet ***VLOOKUP***. (V står for Vertikal = søgning i kolonneform). Fællesnævner i de to tabeller er som sagt *ADRESSE*. Via denne ***VLOOKUP*** funktion i Excel, har jeg nu en ny tabel der indeholder alle mine oplysninger om boliger og VW-Golf biler i PAKILA.

HRI-tabellen indeholder som tidligere nævnt 112 kolonner med forskellige egenskab for alle bygningerne i Helsinki Kommune. Jeg har kun brug for ca. 20 kolonner med egenskaber, som for eksempel (boligtype, boligstørrelse, antal lejligheder, årgang og så videre). Den samlede tabel med de to egenskabsdata bliver renset ned, til at indeholde ca. 48 kolonner med egenskaber for bygninger og biler.

Herefter sorteres mine sammensatte 205 VW-Golf data med bil- og boligtype efter boligtyper i PAKILA. Fordelingen af boligtyper med en VW-Golf på ejendommen er som følger:

- 155 familier bor i Parcelhuse (1 og 2 familiers huse)
- 46 familier bor i Rækkehuse, og
- 4 familier bor i Højhuse.

I **afsnit 8.3** findes der en detaljeret gennemgang af indholdet af mine data, der har været underkastet en generel Excel-sortering og udvælgelse. ***Figur 19*** og ***Figur 21*** er et eksempel på den grafiske udledning *(output)* af data, jeg har været i stand til at foretage, ved at sammensætte mine to tabeller. Ved at se på de forskellige genererede diagrammer, er det muligt at lave nogle første konklusioner om fordelingen af Biltyper og Boligtyper i mit forsøgsområde i PAKILA. Der findes flest VW-Golf Variant biler i *Parcelhus med 1 familie* med en kvadratmeterstørrelse på 90 - 120m² med overvægt i 90-99m². Ved at foretage en søgning i min Excel-tabel med alle boliger i hele PAKILA ud fra kriteriet ***(Parcelhus med 1 familie, størrelse 90 - 119m²)*** findes der 649 boliger, der opfylder dette kriterie. Tilsvarende tal ved kun at søge på ***90 - 99m²,*** som jo er den største enkelte koncentration af parcelhuse i PAKILA, ville tallet blive 241 boliger.

VW-Golf Hatchback har jo sine største tilhængere i *Parcelhus med 2 familier* i kvadratmeterstørrelsen 70 - 89m² og i *Rækkehus med flere familier* i kvadratmeterstørrelsen 80 - 99m². Ved at foretage en søgning efter samme mønster som med *Parcelhus med 1 familie*, bliver resultaternes som følger:

I ***(Parcelhus med 2 familier, størrelse 70 - 89m²)*** findes der 332 boliger, der opfylder dette kriterie. Tilsvarende tal for antallet af boliger med en bestemt kvadratmeterstørrelse, for eksempel ***70 - 119m²***, ville give et resultat der viser, at der findes 736 boliger der opfylder dette kriterie.

Ved at fortage samme søgning for Rækkehuse bliver resultatet som følger:

(Rækkehus med flere familier, størrelse 90 - 99m²) findes der 151 boliger i PAKILA, der opfylder dette kriterie. Tilsvarende tal ved at søge på en bredere kvadratmeterstørrelse, for eksempel ***80 - 99m²,*** da der findes to søjler der klart skiller sig ud fra de andre for VW-Golf Hatchback og rækkehuse. Resultatet af søgningen i Excel-tabellen bliver, at der findes 362 boliger der opfylder dette kriterie.

Trenden er derfor, at personer der ejer en VW-Golf Variant bor i *Parcelhus med 1 familie* og personer der ejer en VW-Golf Hatchback bor i *Rækkehus med flere familier* og *Parcelhus med 2 familier*.

Ved at lave samme forsøg og anvende de samme søgekriterier i hele HRI-tabel data materialer, får jeg følgende tal:

I hele Helsinki Kommune findes der 4.257 *Parcelhus med 1 familie*, der har en størrelse på 90 - 119m². En søgning på 90 - 99m² giver et samlet antal boliger på 1.430 stk.

I Helsinki Kommune findes der 226 *Parcelhus med 2 familier* og boligstørrelse 70 - 89m². Ved at søge over et breder spekter ud fra min viden om, at der findes en større og mere jævn spredning for VW-Golf Hatchback i *Parcelhus med 2 familier (*70 - 119m²), finder jeg 929 boliger der opfylder dette kriterie.

For *Rækkehus med flere familier* og en boligstørrelse på 90 - 99m² findes der i hele Helsinki Kommune 803 boliger der opfylder dette kriterie. Hvis vi anvender samme søgekriterier som for PAKILA med en udvidelse af antal kvadratmeter bolig til 80 - 99m², findes der 1.886 boliger i hele Helsinki Kommune.

Højhuse i PAKILA har næsten ingen repræsentation på grund af at der kun findes 4 VW-Golf Hatchback biler i denne boligtype. Der er således ingen grund til at sammenligne Biltype og boligtype i PAKILA med Højhuse, der findes 1 bil i hver boligstørrelse. Kun kan vi sige ud fra de observerede data med Højhuse, at den største koncentration af størrelsen af lejligheder i PAKILA findes i intervallet 50 - 59m².

Excel-analysen viser, at det muligt at lave forskellige udtræk fra tabeldata og omforme denne information til grafisk og ny numerisk information der kan danne grundlag for en evaluering og i sidste ende som en beslutningsproces for en handling. Processen med at sammensætte mine to tabeldatabaser til én fuldstændig tabeldatabase, hvor jeg har alle elementer der er nødvendige for at lave en sammenligning, vises med al tydelighed i ***Figurerne 19, 20, 21 og 22***. På baggrund af den information jeg har genereret i ***Figur 19 og Figur 21***, er det muligt at se en fordeling og også sammenhæng mellem de to biltyper og respektive boligtyper. Dette er dog en ganske simpel metode, men med denne basisviden er det muligt at søge i den sammensatte tabel og finde andre ligheder eller nuancer, der kunne bruges i et markedsføringsøjemed eller som for mit vedkommende til at besvare min hypotese.

Hvis ideen var at lave en markedsføringskampagne for en VW-Golf Variant, ville personen i bilfirmaet kunne udskrive for eksempel: *649 adresser til personer i PAKILA i de 6 postdistrikter med postnumrene (00630, 00640, 00660, 00670, 00680 og 00690), der opfylder kriteriet for en VW-Golf Variant med hensyn til boligform, (Parcelhus med 1 familie, størrelse 90 - 119m²)* og sende markedsføringsmaterialet om dette produkt til potentielle kunder, der ifølge VW-Golf Variant profilen, ville opfylde en del af de kriterier som en VW-Golf Variant bilejer må formodes at have. Eller personen i bilfirmaet kunne sende disse 649 adresser til TRAFI, og bede om at få at vide, hvilke bilmodeller der fandtes på disse adresser. Derved kunne der laves en direkte markedsføringskampagne til for eksempel det eller de bilmærker, der var den største konkurrent til VW-Golf Variant. Denne form for markedsføring vil dog næppe finde sted i det virkelige business verden (Business Intelligence), men eksemplet er taget med for at vise

det rent praktiske, i anvendelsen af de geografiske data, der findes tilgængelige i dag – helt frit - og hvordan de kan anvendes rent praktisk. Endelig kunne en verificering fra TRAFI, med hensyn til, hvilke bilmodeller der fandtes på de fundne adresser, jo også fortælle, om der rent faktisk var en sammenhæng mellem boligtyper og biltyper.

Og kan det så bruges til noget? Nej – ikke som et entydigt resultat til besvarelsen af min hypotese, af den simple årsag, at mine bildata med adresser er fiktive. Og Ja - som metode til at finde potentielle kunder for VW-Golf biler i PAKILA eller i Helsinki Kommune eller i hele Finland for den sags skyld, er det én ud af mange metoder. Den er i hvert fald retningsgivende. Og – er der så en sammenhæng mellem folk der ejer en VW-Golf Variant eller VW-Golf Hatchback fordelt på boligtyper? Ja – det er der ifølge de tal, som jeg får ud af mine forskellige søgninger og sorteringer i min tabel, og vist som en grafisk præsentation. Ka' det bruges til noget, det her Excel-analyse? – Ja – helt bestemt! Microsoft Excel er et glimrende værktøj til at formidle et budskab og nedbryde data til mindste detaljeringsgrad gennem analyser af datamængder, via beregninger, opstilling af tabeller, skabelse af rapporter, som grundlag for at træffe beslutninger.

Til sidst vil jeg lige gøre opmærksom på at det er værd at bemærke, at for eksempel boligernes størrelse i mit forsøgsområde og i HRI-tabeldataene i det hele taget, kan illustreres på to måder:

- Den samlede størrelse af boligen, eller
- Den samlede størrelse af boligen for hver familie.

Dette fænomen er tilfældet i de boligtyper hvor der bor mere én familie. I Excel-tabellen har jeg derfor sat en ekstra kolonne ind, hvor alle boliger er divideret med antallet af familier. Derved opnås et mere korrekt billede af boligstørrelse i forhold til familien. Den gennemsnitlige boligstørrelse for en VW-Golf familie bliver derfor for hver enkelt boligtype: (boligernes samlede størrelse divideret med antallet af boliger).

8.5 SQL-analyse via MapInfo.

SQL-analyses primære formål er at vise, hvordan tabeldata kan visualiseres for at kunne anvendes i et beslutningsgrundlag og hvordan man kan søge i en tabeldatabase med forespørgsler - Query - for at tilfredsstille en nysgerrighed og eventuelt få svar på et spørgsmål. Det er selvfølgelig også at vise hvordan denne form for analyse, kan anvendes til at besvare min hypotese. SQL-analysen og omformning og visualisering af tabeldata til kortdata, har mange sider og fasetter. I min analyse vil jeg forsøge at vise hvordan, mine tabeldata, omformet til visuel information, kan være med til at danne grundlag for et beslutningsgrundlag, ud fra de resultater der findes i forskellige søgninger.

Der findes og er jo uendelig mange måder, at analysere data på. Derfor vil jeg ikke i afsnittet on SQL-analyse gå i nogen speciel lille detalje med et lille kortudsnit eller en lille graf og så videre og derigennem prøve at finde *det perfekte svar*, men mere prøve – med referencer til kortdata, at vise forskellige muligheder for at trække information og viden ud af kortdata, med støtte i tabeldata.

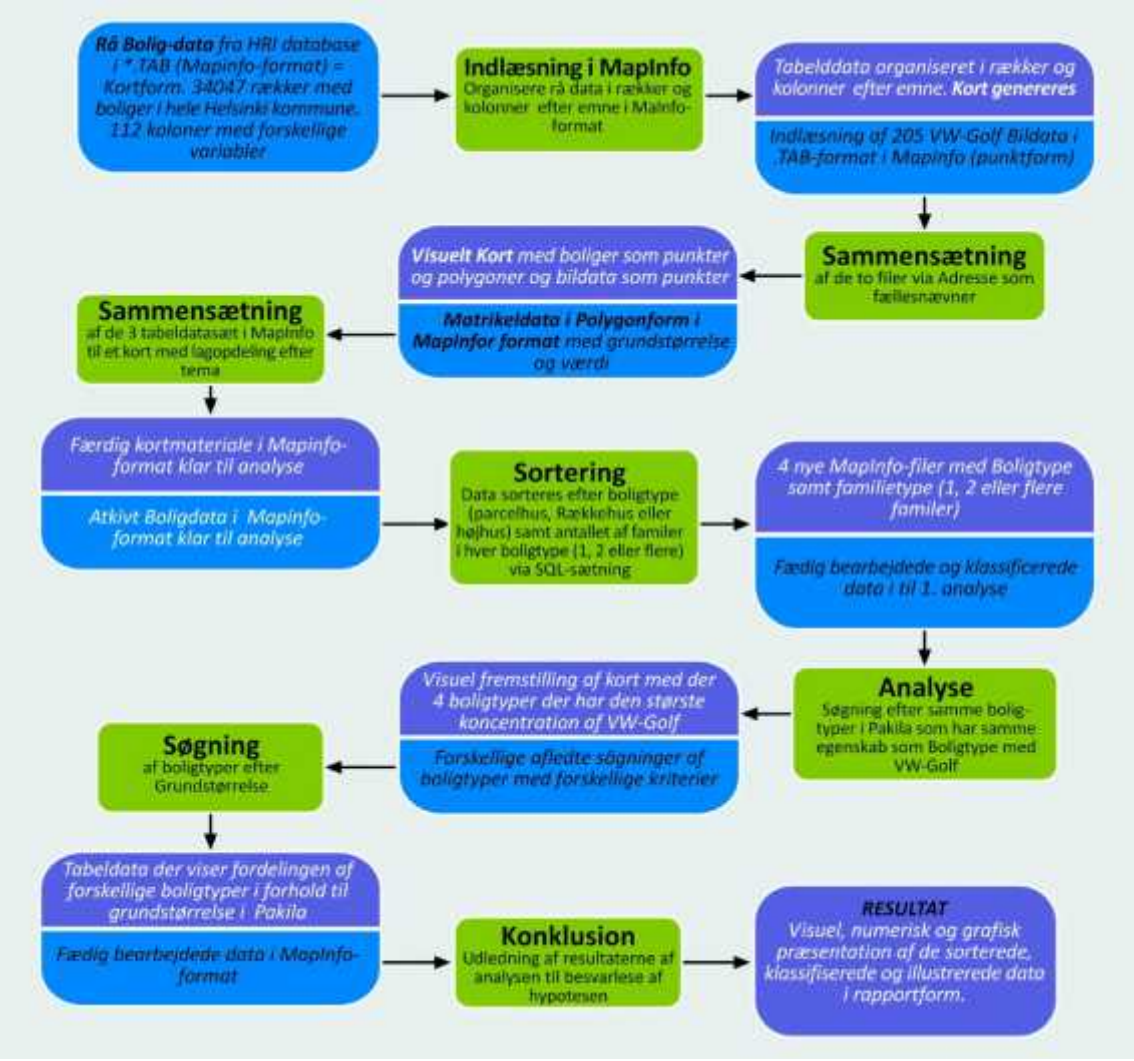
Ideen med geografisk analyse er blandt andet at illustrere mine resultater ved at projicere de fundne værdier fra en søgning med opstillede kriterier, ind på et kort og dermed visualisere mit resultat. Derved får jeg en mulighed for at, foruden den numeriske værdi af en analyse, også visuelt se hvor mine objekter findes i vores forsøgsområde. Her findes den største forskel jo nok, når vi sammenligner Excel-analyse og anvendelse i det hele taget, af grafisk analyse. Dette emne vil jeg uddybe i ***Kapitel 10*** om ***Perspektivering***. Den geografiske analyse og synsvinkel på data og information, er dog på ingen måde med til at forbedre analysen og dermed det endelige resultat af en undersøgelse, men den visuelle effekt kan være med til at bestemme en mere specifik eller helt anden metode, for at forbedre vores analyse (for eksempel optimering).

Endelig kan forskellige effekter i GIS-systemet ved anvendelsen af for eksempel farver, give en retningslinje om, hvor lignende områder findes, der har samme karakteristika som de detekterede. Sandsynligvis er resultatet af analysen den samme, men den visuelle effekt, er med til at se resultatet fra en anden synsvinkel og måske endda være med til at øge interessen eller inspirationen for, at prøve en anden metode, for at finde en anden indgangsvinkel til et resultat.

Som tidligere nævnt i ***Kapitel 4*** i ***Indledning***, er GIS med til at afsløre mønstre i vores tabeldata. Ved at gå ”baglæns” og bruge den visuelle effekt til at se mønstre og trends i tabeldataene, er det også muligt at lave en søgning direkte i tabeldataene, ved at se på forskellige værdier og deres spredning i selve tabellen. Derved kan Excel-analysen udvides via min information og viden fra mønstre i kortet, til at søge på for eksempel et specielt postnummer eller postdistrikt, vejsegment eller bydel, der på kortet har ”afsløret” en trend eller karakteristika.

Nedennævnte ***Figur 24*** er en *Input-Proces-Output* model for hvordan den ”almindelige” SQL-analyse forløber i mit projekt som jeg have tænkt mig den kunne foregå. Da der ikke findes nogen entydig proces med tilsvarende *Input* materialer og output resultater, skal denne grafiske fremstilling af mit SQL-analyse forløb ses, som den måde, *jeg tror* jeg får det bedste resultat ud af mine tabeldata, for at kunne besvare min hypotese. Længere fremme i dette afsnit har jeg taget en del eksempler med for at vise, hvordan udsnit fra mit kortmateriale med tilhørende objekter, kan danne grundlag for en beslutning om at foretage en ny eller anderledes søgemetode, eller hvordan det er muligt, at foretage en analyse af et søgningsresultat rent visuelt, på baggrund af for-

skellige objekt-fordelinger i kortet. Endvidere vil jeg ikke som i afsnittet om Excel-analysen, foretage mange forskellige søgninger med dertil fremkomne resultater (tal der fortæller hvor mange VW-Golf biler der findes til en bestemt boligtype), men mere vise de forskellige nuancer og muligheder der findes for en visuel repræsentation af et søgeresultat i kortmaterialet og hvordan disse data (visuelle objekter) kan danne grundlag for direkte eller indirekte analyse. Med direkte analyse menes, at for eksempel en objekt-fordeling i kortet, kan fortælle os en historie. Med indirekte analyse mener jeg; et resultat af en analyse og det dertil fundne visuelle resultat, der kan være med til at bestemme en ny eller anden analysemetode.



Figur 24: *Via denne processering af data i MapInfo, er jeg i stand til at lave en analyse af mine indsamlede data, der giver et første resultat til besvarelse af min hypotese*.

Den første proces-boks med "**Sammensætning af de to filer via adresse som fællesnævner"**, foregår i MapInfo ved en funktion der hedder ***"Append Rows to Table"***. Jeg skal ikke komme ind på den tekniske proces der ligger bag denne funktion, men bare nævne, at det er der svarer til **VLOOKUP-funktionen** i Excel, hvor to filer "parres" sammen til én samlet fil, med fælles egenskaber og hvor Adressen i begge tilfælde er fæl-

lesnævneren. ***Figur 25*** viser resultatet af denne sammensætning som en illustration (kort) med punktobjekter.

Figur 25: *Sammensætningen af de to hovedfiler i min Analyse, der danner grundlaget for udtræk af information og i sidste ende viden, der kan danne grundlag for besvarelsen af min hypotese.*

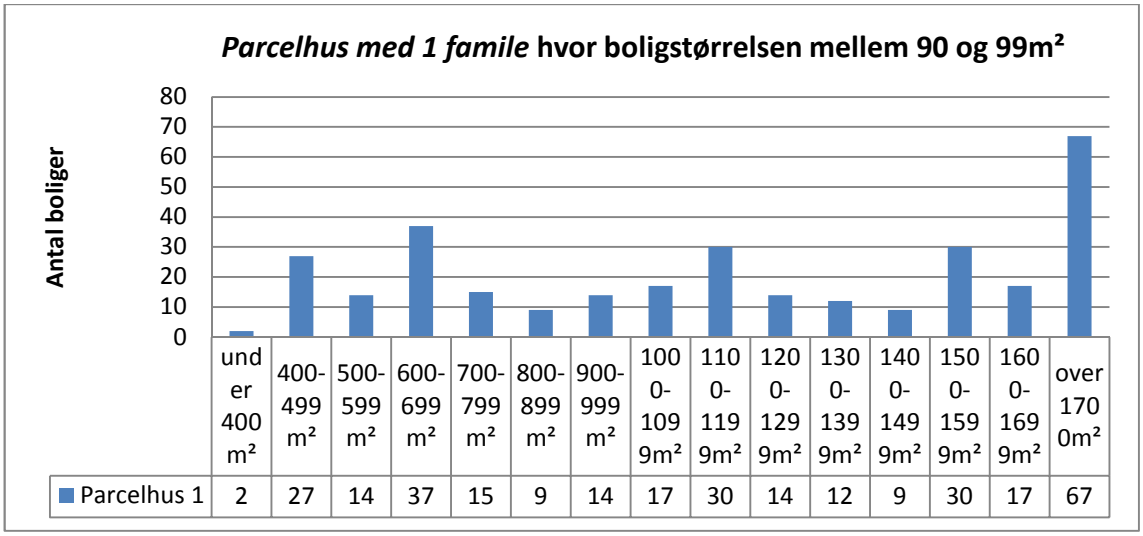
Min første analyse med SQL går ud på at anvende min forhåndsviden fra de generelle data fra ***afsnit 8.3***, der viser fordelingen af VW-Golf Variant og VW-Golf Hatchback på de forskellige boligtyper i PAKILA. Min SQL-sætning hedder:

Rigtige_storrelse_af_lejlighed between 90 and 99 and A_kode3 = "11"

På dansk: *find størrelsen af alle boliger med en kvadratmeterstørrelse mellem 90 - 99m² der hedder Parcelhus med 1 familie*.

Resultatet af søgningen viser, at der findes 314 objekter i mine korttabeldata der opfylder dette kriterie. ***Figur 26*** i nedennævnte diagram viser fordelingen af *Parcelhus med 1 familie*, hvor boligstørrelsen er mellem 90 - 99m². Derefter kan jeg plotte disse objekter ind på mit kort som punkter og se hvor i PAKILA de findes. ”På købet” har jeg også fået at vide, hvor store grunde der findes hvor mine *Parcelhus med 1 familie* findes. Denne information kunne jeg have udeladt, ved samtidig i min SQL-sætning af skrive:

Rigtige_storrelse_af_lejlighed between 90 and 99 and A_kode3 = "11" not Grundstorrele between under 400 and over 1700



Figur 26: *Fordelingen af den samlede grundstørrelse af et Parcelhus med 1 familie i PAKILA, hvor størrelsen af boligen er mellem 90 - 99m².*

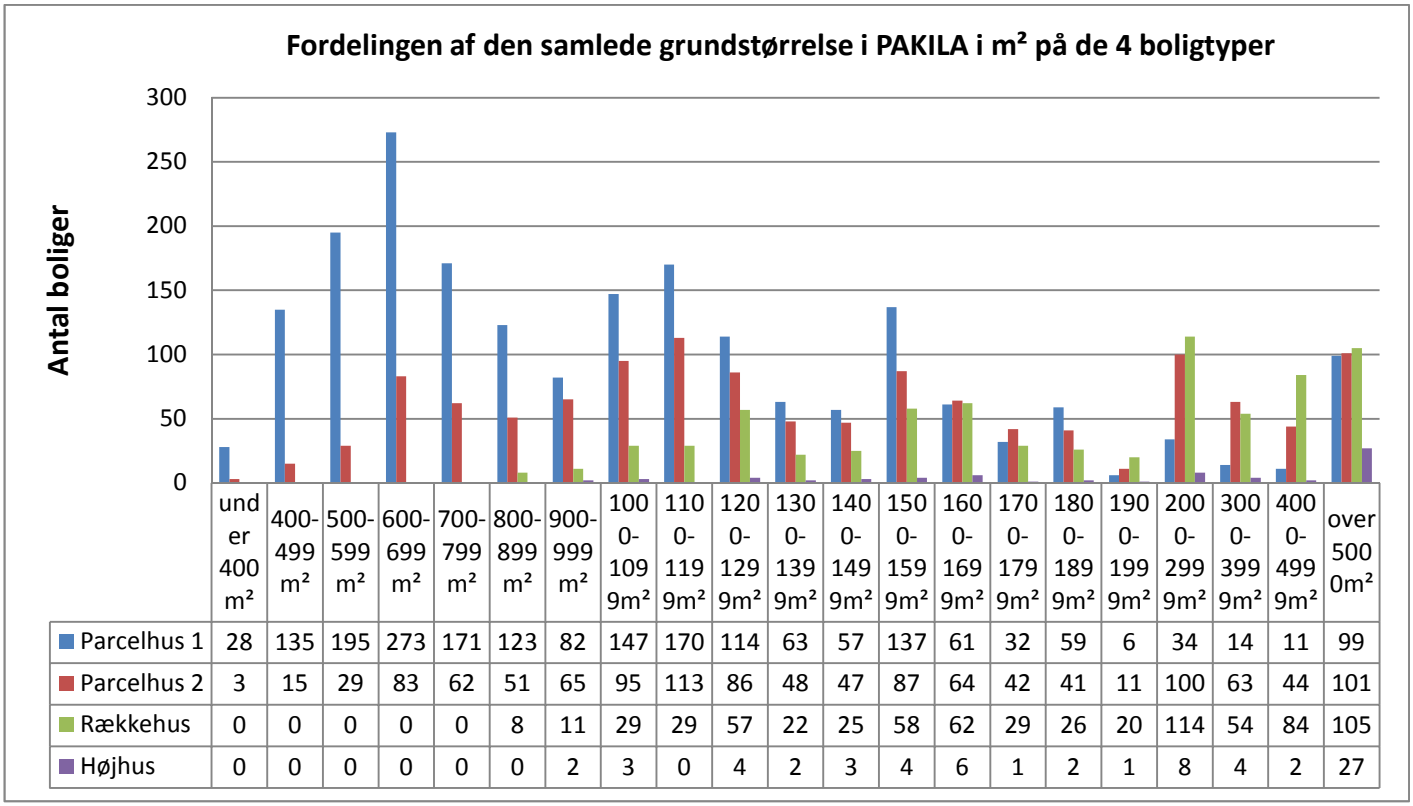
I ***Bilag B4*** findes en mere detaljeret forklaring og illustration af hvordan MapInfo's SQL-Query funktion virker, med opbygning og funktionsval. I resten af dette analyseafsnit, vil jeg ikke vise flere eksempler på mine søgninger for at finde forskellige boligtyper, der kunne fortælle om de matcher mine biltyper. Resultaterne af de forskellige SQL-søgninger, kan ses i ***Figur 29***, der viser hvor mine VW-Golf biler findes for hver boligtype, der har den største fordeling med udgangspunkt *i* ***Figur 21.***

Hvis vi ser på ***Figur 26*** og den spredning, der findes på *Parcelhus med 1 familie* i forhold til grundstørrelsen, ser vi at denne spredning er ganske jævn, dog med 600 - 699m² grundstørrelse og over 1.700m² som de to søjler der indikere en forhøjelse af en koncentration. Den information kan jeg overføre på min næste ***Figur 27***, der viser en fordeling af alle grundstørrelserne i PAKILA med 100m² interval op til 2.000m². Og – igen. Dette er et meget simpelt, ja næsten banalt eksempel på, hvordan information fra forskellige tabeller og information fra forskellige kolonner i tabeller, ved en simpel organisering, kan være med til at give en viden for en ny søgning eller en sammensætning af en SQL-sætning, der ville tilfredsstille et ønske. Ved at kigge på ***Figur 22*** og ***Figur 27***, der viser henholdsvis fordelingen af alle boligtyper i PAKILA på kvadratmeterstørrelsen pr. familie, samt fordelingen af de samlede grundstørrelser i PAKILA efter de 4 forskellige boligtyper, har jeg sammensat et kort i ***Figur 29,*** der viser hvor de 4 mest potentielle boligtyper, hvor der var mulighed for at finde eller markedsføre en VW-Golf Variant eller VW-Golf Hatchback. Disser boligtyper er:

- *Parcelhus med 1 familie*, boligstørrelse 90 - 99m² (234 boliger)

- *Parcelhus med 2 familier*, boligstørrelse 70 - 89m² (324 boliger)
- *Rækkehus med flere familier*, boligstørrelse 90 - 99m² (148 boliger)
- *Højhus med flere familier* (ingen boliger).
-

Jeg har ikke taget nogen højhuse med i min udvælgelse af boliger, da spredningen på højhuse er ens, det vil sige at der findes 4 højhuse, hvor der findes én VW-Golf Hatch-back.



	under 400 m²	400-499 m²	500-599 m²	600-699 m²	700-799 m²	800-899 m²	900-999 m²	1000-1099m²	1100-1199m²	1200-1299m²	1300-1399m²	1400-1499m²	1500-1599m²	1600-1699m²	1700-1799m²	1800-1899m²	1900-1999m²	2000-2999m²	3000-3999m²	4000-4999m²	over 5000m²
Parcelhus 1	28	135	195	273	171	123	82	147	170	114	63	57	137	61	32	59	6	34	14	11	99
Parcelhus 2	3	15	29	83	62	51	65	95	113	86	48	47	87	64	42	41	11	100	63	44	101
Rækkehus	0	0	0	0	0	8	11	29	29	57	22	25	58	62	29	26	20	114	54	84	105
Højhus	0	0	0	0	0	0	2	3	0	4	2	3	4	6	1	2	1	8	4	2	27

Figur 27: *Fordelingen af den samlede grundstørrelser i PAKILA efter de 4 forskellige boligtyper der er repræsenteret. Tabellen er ikke entydig og bygger på originale værdier fra HRI-tabeldatamaterialet.*

Men i det virkelige liv ville det jo nok se lidt mærkeligt ud, at begynde at vælge biltyper efter grundstørrelsen. Men som et sorteringselement for at finde den "eksakte profil" for en VW-Golf Variant bilejer kunne det måske være relevant som et kriterie. Og – det kunne være ét af de elementer, der er med til at vise, om der er sammenhæng mellem boligtyper og biltyper. Derfor har jeg lavet et såkaldt Choropleth kort over fordelingen af grundstørrelserne i PAKILA. Jeg kunne også have vist samme opdeling for mine boliger, men da disse er relativ små polygoner, ville effekten ikke have været så i øjnefaldende.

Farvesammensætninger af områder (læs tabeldata), kan være med til at danne karakteristika og gøre mine tabeldata "levende" og fortælle mig, hvor der kunne være

områder af interesse for en dybere efterforskning eller undersøgelse. Denne form for illustrationer har på det seneste vundet mere og mere indpas i medierne – trykte og digitale, som supplerende information til en artikel eller interview, der måske ikke har kunnet formidle et budskab, uden at illustrere meningen på en mere simpel måde. Således også her i mit projekt. Her har jeg igen helt enkelt opdelt mine grunde efter størrelser med farver. Under ***400m² til 1.000m²*** er vist med ***røde nuancer***, ***1.000m² til 1.700m²*** er vist med ***grønne nuancer*** og endelig er grundstørrelser fra ***1.700m² til over 5.000m²*** vist med ***blå nuancer***. Fra 2.000m² er der en intervalopdeling på 1.000m². Man skelner mellem tre forskellige opdelinger af farvekombinationer ved brugen af Choropleth kort:

- *Sekventiel* (fortløbende) hvor samme farvegruppe anvendes for hele kortet med forskellige nuancer,
- *Divergerende* (afvigende) hvor farverne adskiller to grupper af information (høj, lav eller god, dårlig), farvefordelingen er for eksempel rød til gul.
- *Kvalitativ*, hvor flere objekter skal klassificeres. Her anvendes en større farveskala med flere nuancer (rød, grøn og blå), som i tilfældet med min illustration i ***Figur 28.***

Figur 28: *Et Choropleth kort der viser fordelingen af grunde i PAKILA efter størrelse. Denne form for visualisering af data og information, kan være med til at give et fingerpeg om, hvor der kunne være interessante områder for nærmere undersøgelse og analyse = samme farvekoncentration i et større eller mindre område(klynger).* [4]

Som den kvikke læser måske allerede har bemærket, findes der ikke en indholdsfortegnelse, med opdelingen af de forskellige farvegrupper med tilhørende værdier. Det er imidlertid ikke nødvendigt i dette tilfælde, men som hovedregel, ej at forglemme. Af kortudsnittet kan vi se, at der findes flere områder der grænser op til hinanden, der har samme grundstørrelse.

Figur 29: *Fordelingen af boligtyper i PAKILA med de mest potentielle boligtyper for VW-Golf Variant og VW-Golf Hatchback. De **røde** objekter viser Parcelhuse med 1 familie, de **gule** objekter viser Parcelhus med 2 familier og de **blå** objekter viser Rækkehus med flere familier. Højhuse er ikke vist på dette kort.*

Om det er topologisk bestemt (bakket terræn hvor det er svært eller dyrt at bygge enkelthus), eller det er fordi der er en service i nærheden, der gør kvadratmeterprisen høj for grunden, og det derfor er mere hensigtsmæssigt, at bygge flere huse eller rækkehuse med mange familier, så kvadratmeterprisen bliver lavere, får stå hen i det uvis-

se. Her behøver vi mere information for at kunne drage en mere entydig konklusion. Alt dette er bare eksempler på hvordan det er muligt at fortolke temakortkort med data for at se og forstå sammenhænge. Det er ofte kun fantasien der sætter den endelige grænse for en fortolkning. Svarene kan være mange, ligesom med SQL-søgningerne.

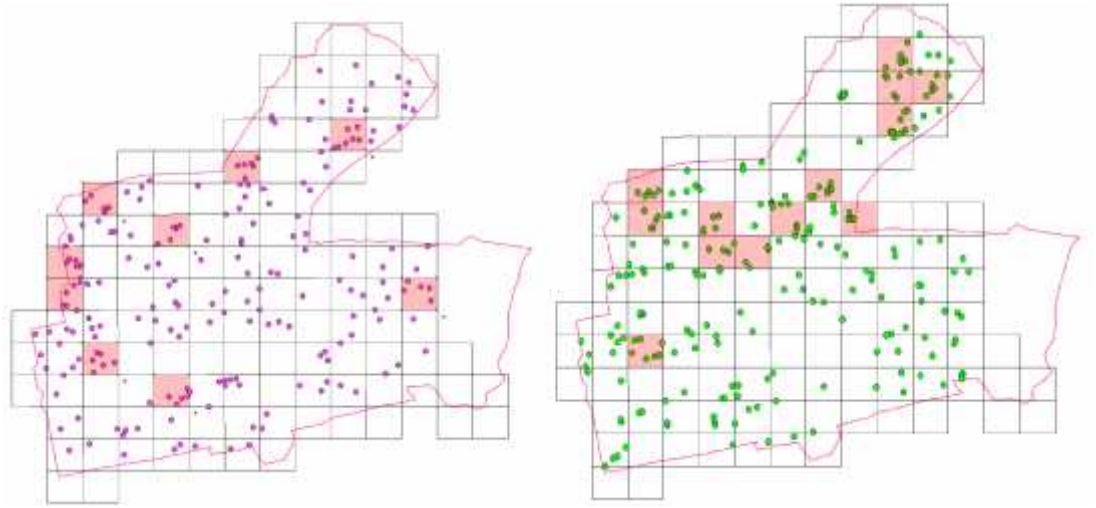
Figur 30: Sammensætning af to kortinformationer, der viser f*ordelingen af boligtyper og grundstørrelser i PAKILA, med de mest potentielle boligtyper for VW-Golf Variant og VW-Golf Hatchback.*

Vi skal huske på, at det at anvende kortet eller illustrationer, der fremkommer ved en analyse i et GIS-system, ofte ikke er det grundlag hvorpå der skal tages en beslutning. De illustrative data står altid i relation til et tabeldata eller anden nominel information. Forstået på den måde, at kortet nok viser forskellige områder med en aktivitet eller handling ud fra en klassifikation, men den nominelle værdi, (adresse, antal, retning eller sammenhæng), findes ikke i den visuelle information.

Selvfølgelig kan illustrationer bruges til at tage de første skridt i retning mod en beslutning eller handling, men som værdimåler er vi nødt til at finde tabellerne frem igen. Hvis vi skal tage en beslutning alene på farverne i et kort og klassificeringen af objekter via farver eller størrelser i en signaturforklaring, ville vi jo være nødt til at have en eller anden form for tal- eller referencemateriale, til at indikere den enkelte størrelse, af et element eller område, for at være i stand til at tage en beslutning. Tabeller og illustrationen står i et "interaktivt" forhold, hvor den ene information støtter den anden. Derfor er det ofte lidt trist og læse artikler om for eksempel *location Intelligence* og *Business Intelligence*, hvor visualiseringen af tabeldata bliver udråbt som den store frelser. Det er en sandhed med modifikationer, som jeg vil diskutere i ***Kapitel 10*** under ***Perspektivering***.

Som et sidste eksempel i min analyse, vil jeg vise hvordan det er muligt via fortolkning af resultatet af en søgning, at gå "baglæns", for at finde et resultat. Dette gøres ved at anvende illustrationerne der er fremkommet ved en analyse og "optage" den information, der findes på numeriske form i kortet og fører den tilbage i GIS-systemet som tabeldata.

Eksempel:
Mine Kvadratnets data fra Finlands Statistikcentral med aldersfordelingen vist i 250m x 250m kvadratnet for hele hovedstadsregionen i Finland.



Figur 31: *De to figurere viser fordelingen og de største koncentrationer af henholdsvis alle VW-Golf biler i PAKILA (røde objekter) og fordelingen af Parcelhus med 1 familie, hvor boligstørrelsen er mellem 90 - 99m². Oven på disse data er der lagt et kvadratnet på 250m x 250m indeholdende demografiske data (aldersfordeling).*

Ved at ligge dette kvadratnet ned over mine objekter, der viser hvor mine VW-Golf biler har deres ejer, kan jeg finde (se i kvadratnettet), den eller de aldersgrupper, der ejer en VW-Golf Variant eller VW-Golf Hatchback. Derefter kan jeg tillægge mine bolig-

data, for eksempel *Parcelhus med 1 familie* og boligstørrelse på 90-99m². Ved at kombinere de to objektdata med aldersgruppe fordelingen, kan jeg se, at der findes flest VW-Golf Variant ejere der er mellem 50 - 59 år der bor i *Parcelhus med 1 familie*.

Ved at "highlighte" de områder, hvor koncentrationen er størst af de to objekttyper, kan jeg "skabe" en tabel med data via disse informationer og gemme denne tabel eller vise den i en rapport og der for eksempel, farve de rækker eller kolonner i tabellen, der har specielt interesse. ***Figur 32*** viser sådan et udtræk fra mine visuelle data i ***Figur 31.***

INDEX	Antal Indb.	0-9 år	10-19 år	20-29 år	30-39 år	40-49 år	50-59 år	60-69 år	70-79 år	over 80 år	Parcal-1fam-90-99m²
16271	177	26	14	8	25	16	33	35	13	7	10
16754	121	10	22	5	10	22	21	24	3	4	8
15627	210	19	26	18	20	27	27	38	22	13	7
16432	238	30	45	16	18	39	39	33	16	2	7
16752	176	20	21	28	10	25	36	24	8	4	7
16753	105	4	22	8	17	16	19	13	2	4	6
16915	158	20	24	14	14	25	23	22	13	3	6
15622	286	42	42	16	24	47	51	44	17	3	5
15623	226	25	21	19	23	36	40	46	16	0	5
15947	251	24	52	21	22	39	49	36	7	1	5
15948	257	37	43	13	18	49	42	32	15	8	5
16110	182	19	32	16	19	24	34	24	8	6	5
16595	103	19	9	8	19	11	9	22	5	1	5
	2490	295	373	190	239	376	423	393	145	56	81

Figur 32: *En tabel med den demografiske aldersfordeling i de 13 områder hvor den største koncentration af Parcelhus med 1 familie hvor bolig, størrelsen er mellem 90 - 99m². Kolonnen med INDEX er nummeret på de **orange** oplyste kvadrater i **Figur 31** (**grønne** objekter).*

Endelig kan fortolkningen af tallene tabellen også fortælle mig, at folk der bor i et *Parcelhus med 1 familie* og en aldersgruppe på 50 - 59 år, sandsynligvis også har børn. Her kan jeg se af tabellen at den største gruppe af børn eller unge mennesker er mellem 10 og 19 år. Det passer jo meget fint sammen med forældre der er 50 - 59 år og ejer en VW-Golf Variant. (Plads til børn og deres aktiviteter)

Ideen for mit vedkommende med at anvende en "Heatmap" til at vise, hvor der findes større koncentrationer (fordeling) af boligtyper, der matche min profil for en VW-Golf-ejer, er ikke statistiskt at begynde at beregne, hvilke områder i PAKILA eller for eksempel Helsinki Kommune, der ville opfylde et specielt kriterie (korteste, længste afstand mellem samme karakteristika). Nej – det er igen som i Excel-analysen, at illustrere, hvordan man kan visualisere et resultat til anvendelse af et beslutningsgrundlag, eller til markdesføring i specielle områder. Dette kunne gøres ved at lægge en "administrativ grænse" (postdistrikt med nummer, bydel med reference nummer), ned over de områder der har størst eller mindst koncentration, og derefter hente adresserne fra mine tabeldata.

Figur 33: *"Heatmap" fra PAKILA hvor 3 boligtyper; (Parcelhus med 1 familie, Parcelhus med 2 familier og Rækkehus med flere familier), viser tætheden (density) af data. Heatmap er én måde at illustrere fordelingen og koncentrationen af data, for visuel fremvisning eller præsentation af et resultat.*

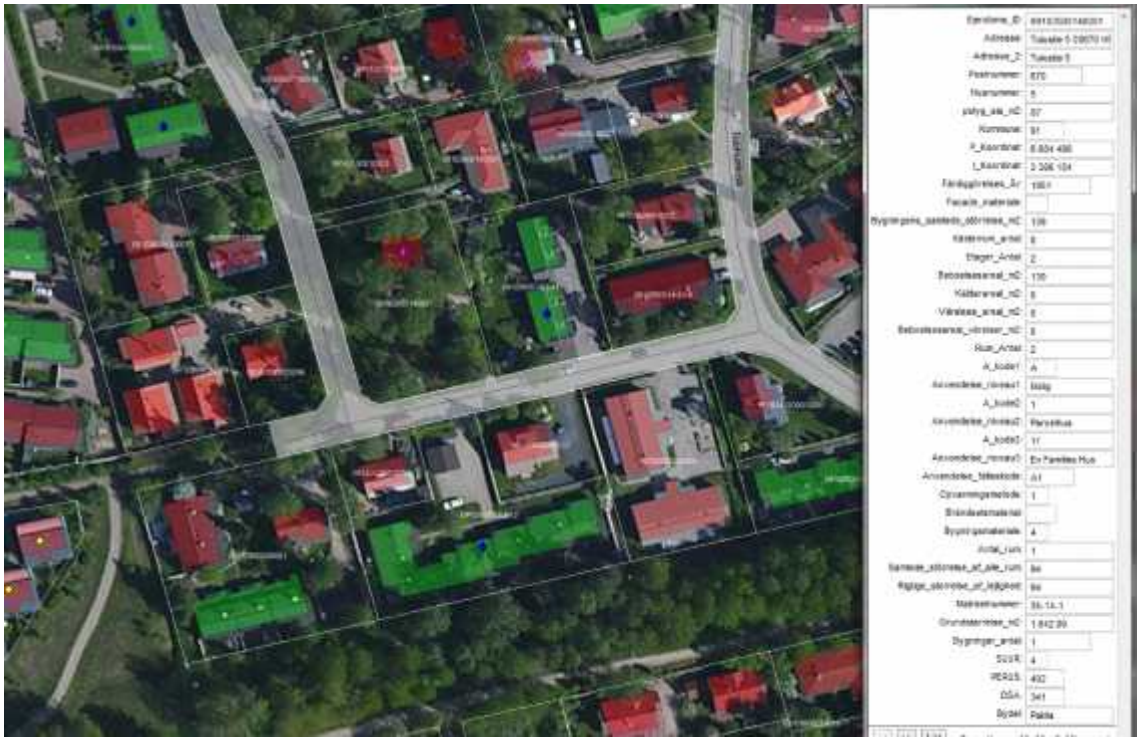
8.6 Afrunding.

I Analyseafsnittet har jeg forsøgt at give en oversigt over, hvordan forskellige beregnings- og illustrationsmetoder i Excel og SQL via MapInfo, kan illustrerer tabeldata og den information de indeholder, på en sådan måde, at det er muligt at drage konklusioner af de opnåede søgninger og resulterer for at besvare min hypotese. Som jeg flere gange har pointeret i forbindelse med analysen og arbejdet med tabeldata, findes der ikke noget rigtig eller for den sags skyld entydigt svar, når der barsles med et resul-tat. Ofte er der rum – endda meget, for fortolkning af et resultat. Her er der op til den enkelte bruger, at kunne skelne et resultat fra forskellige beregningsmetoder i analyseprogrammerne, og afgøre dets anvendelighed for en konkluderende beslutning, eller om resultatet er vejledende og behøver en finslibning for at kunne anvendes som resultat i en beslutningsproces.

En ting der ikke er nævnt endnu i analysen, men som har en har indflydelse på resultater generelt i enhver analyse eller undersøgelse, er kvaliteten af data. Her tænker jeg specielt på fejlkilder. Som en tommelfingerregel kan man gå ud fra, at jo tættere data kommer fra en offentlig institution, jo mere pålidelig er oprigtigheden af og dermed troværdigheden af data. De betyder ikke at data fra private firmer nødvendigvis har en

dårligere kvalitet, men tilblivelsen af data kan måske ikke spores tilbage til kilden så klart, som data fra offentlige institution der jo ofte har dokumentationspligt – for eksempel INSPIRE og de direktiver (INSPIRE-direktivet af 14. marts 2007) der følger med dette projekt.[26] Det som var at bemærke i mine data fra HRI-datamaterialet var, at der manglede adresser på adskillige boliger, ligesom størrelsen af flere af disse manglede. En del af denne mangel skyldes de facto, at boligerne var under opførsel, og derfor ikke registreret med adressen, men kun med Boligens ID og Matrikel. Ca. 1.9 % (73 boliger) var uden adresser eller anden ID i PAKILA.

En måde at "kontrollere" mine boligdata, for rigtigheden af bygningerne på de manglende adresser, er ved at anvende et ortofoto, der er findes for hele hovedstadsregionen plus de nærmeste omegns-kommuner.



Figur 34: *Ortofoto som baggrundsmateriale som verificering af, om der findes en bolig eller ikke på grunden foruden naturligvis det generelle visuelle effekt som ortofotoet levere, nemlig det originale udseende af et område for analyse, foruden den "stærke" WYSIWYG-effekt, der kan få mange beslutningstagere til at blive interesseret i et produkt eller idé. Tabellen til højre i illustrationen viser MapInfo's browser-window med information om én bygning.*

Ved at lægge disse ortofotos under mine boligdata i MapInfo, kan jeg hurtigt - rent visuelt - detektere de adresser med matrikelnumre, hvor der ikke findes en adresse, men en registrering af en bygning, matrikel eller anden aktivitet, der udløser en eller anden form for registrering.

En ting, der er værd at se efter og bemærke i forbindelse med anvendelse af data og information, til bestemmelse af en metode, handling eller implementering, er hvordan data bedst anvendes, for at give et så godt et svar som muligt på et spørgsmål eller en søgning. Hvis vi for eksempel kigger på en kolonne med data i vores bil- og boligtabel, og for eksempel vil anvende størrelsen af boligen som en indikator for standarten, der skal sættes i forhold til søgningen, findes der flere måder at udlede denne på. Skal vi for eksempel anvende middelværdien +/- 5 - 10m² som konstant eller skal vi anvende alle data på samme tid. Her er der ikke nogen entydig svar, ligesom der heller ikke er noget entydig resultat.

Kortet eller illustrationen om man vil, kan bruges som en referenceramme eller indikator til at vise, hvor en hændelse eller aktivitet kan eller vil foregå. Men som et grundlag for en beslutning er det ikke godt nok alene. Det ville jo være det samme som at anvende et matrikelkort uden skelpunktsnumre eller koordinater. Den illustrative værdi findes, men brugen er lig med nul. Illustrativ information er kun retningsgivende, ikke konkluderende.

Efter at have dvælet i analyser af mine tabeldata, er tiden kommet til at samle op på de mange skrevne linjer og gøre status. I Konklusionen vil jeg lave en opsummering og sammenfatning af det skreven i dette projekt, og sætte det op mod min hypotese og konkludere ud fra mine analyser, om der er en sammenhæng mellem valg af biltype og boligtype.

I dette kapitel fremkommer en samlet konklusion samt diskussion af undersøgelsens foreskrevne overvejelser, meninger og de eventuelt nyopståede perspektiver samt brugen af DIKW-hierarkiet som et af grundlagene for besvarelsen af min hypotese. Excel- og SQL-analyse som metode til at opnå et resultat, vil blive konkluderet og sat i relation til DIKW-hierarkiet som et anvendeligt redskab, i forbindelse med besvarelsen af min hypotese.

9. Konklusion

9 KONKLUSION

"Krage søger mage", er en talemåde der betyder, at vi har tilbøjelighed til at omgås ligesindede. Og sådan kunne man måske også sige om biltyper og boligtyper. Hvor der findes en biltype eller boligtype, kommer der en beslægtet. Naboen skal ikke tro, at jeg er anderledes, eller, hvis der findes sådan en bil eller bolig her, så skal jeg også have sådan en bil eller bygge sådan et hus her. I dette projekt har jeg forsøgt at vise, om der er en mulig sammenhæng mellem personers valg af biltype og tilhørende boligforhold, ved at anvende og sammenligne den information, der findes i geografiske data som en indikator. Svaret er selvfølgelig Ja og Nej.

Konklusionen er, at det er *ægtheden* af data, der bestemmer resultatet af testen af min hypotese, og dermed om hypotesen er sand eller falsk. Den sande sammenhæng mellem boligtyper og biltyper kan kun findes ved, at have de rigtige og originale data af, hvor folk der ejer en VW-Golf bor. Der igennem, og kun der igennem, er det muligt at lave en analyse af sammenhængen mellem boligtyper og biltyper. Alt andet er fiktivt, nonsens eller ikke sammenligneligt. Og ud fra den anskuelse, falder min hypotese. *Der er ingen sammenhæng mellem boligtyper og biltyper i mit forsøgsområde - PAKILA, på baggrund af det datamateriale, jeg har indsamlet og analyseret*. Integreringen af andre informationsfaktorer som, demografi og økonomi gør ingen forskel, så længe de ikke er ægte eller har direkte og sand relation til et givet objekt. Den demografiske aldersfordeling af befolkningen i PAKILA via et kvadratnet, viser hvor der findes koncentrationer af befolkningsgrupper, der kunne være potentielle kunder for et bilmærke, hvis jeg vel at mærke har den fornødne og sande baggrundsviden om en kundeprofil.

Konklusionen er, at der ikke er nogen sammenhæng mellem boligtyper og biltyper, før jeg får bekræftet, hvilke biltyper (VW-Golf eller helt andet bilmærke), der rent faktisk findes i de boligtyper, som jeg tror korrelere med en bestemt biltype. Dette kan betegnes som *"Ground Truth"*. (udtrykket *"Ground truth*" refererer til rigtigheden af klassificeringen af træningssæt og processen med at indsamle de korrekte objektive data, til brug i statistiske modeller, til at bevise eller modbevise en hypotese. Altså henter man referencepunkter eller karakteristika for at forsøgsområde, så man kan sammenligne resultatet med det man måler). TRAFI er naturligvis det sted hvor disse referencedata eller karakteristika findes, men denne information er ikke tilgængelige for privatpersoner som mig, som tidligere nævnt i projektet. Derfor kan jeg ikke bekræfte min hypotese, med mine fundne værdier for boligtyper (boligtyper i forhold til biltyper), og dermed sige, at der findes en sammenhæng mellem boligtyper og biltyper. Kort sagt, jeg har ikke noget at veje mine resultater i op imod. Jeg kunne selvfølgelig for sjov skyld, have spredt en masse forskellige biltyper ud over mit forsøgs

område og set, hvor mange af disse, der ville have ramt min boligtype. Men det ville jo ikke have be- eller afkræftet min hypotese, kun fiktivt vise, om der var en korrelation.

Konklusionen er, at brugen af frie data (Open Source Data) understøtter den aktive brug af data og gør det lettere at referere til offentlig information. At give gratis adgang til offentlige data er godt for virksomheder og andre interessegrupper. Det skaber nye markeder og støtter innovation og gør viden-overførsel mellem brugerne af informationsteknologi lettere. Åbne data hjælper med at finde metoder, hvor potentialet i digital information endnu ikke er realiseret endnu. HRI-tabeldatabasen fra Helsinki Kommune understøtter på eksemplarisk vis min Excel-analyse, og viser med al tydelighed, hvordan frie data, direkte kan anvendes af den "almindelige borger", som et supplerende data- og informationsmateriale, dels som referencemateriale i form af bestemmelsen af forskellige boligtyper med tilhørende adresser, dels som et grundlag og forudsætning for at kunne fortage en SQL-analyse, ved at *spørge direkte* til fordelingen af forskellige boligtyper i tabeldataform. HRI-tabeldatabasen kan ikke give svar på min hypotese, men er det andet ellement i hypotesens undersøgelsesobjekt – boligen. HRI-tabel-databasen indeholder alle data om boligen i mit forsøgsområde, der er nødvendige for at differentiere begrebet boligtype.

Konklusionen er, at data, der indeholder information om personer, kun er tilgængelige for firmaer eller andre institutioner der laver statistiske analyser eller sammenkører registre for kunder, og som har en aftale om personoplysninger og integritet. Som privatperson og studerende med det foreliggende projekt in mente, har det ikke været muligt at få adgang til denne information. Persondataloven handler om hvornår og hvordan personoplysninger må behandles. Personlige data eller data der kan lede til en "afsløring" af en persons identitet, er ikke omfattet af frie data.

Konklusionen er, at data er den primære form for information. Den består hovedsagelig af optagelser af transaktioner eller begivenheder, som vil blive anvendt til udveksling mellem mennesker. Data har ingen mening, medmindre man forstår den sammenhæng, hvori disse er blevet indsamlet. På den baggrund har jeg gjort den observation, at det ikke er kvaliteten af mine data der er med til at besvare min hypotese og give et "dårligt resultat". Kvaliteten af mine anvendte data er med til at give et *troværdigt* grundlag hvorpå analyserne foretages. De genererede kort- og illustrationer, der opstår via analyser, har en troværdig baggrund og afbilder en virkelighed af i dag, selv om dataene er fiktive. Ægtheden af data, altså den føromtalte sandhedsværdi i data, har ikke noget at gøre med kvalitet.

Konklusionen er, at værdier og information i geografiske data ér en nøgle til et beslutningsgrundlag. Geografiske data repræsenteret i form af objekter, linjer og polygoner, er ikke entydige repræsentanter for et resultat, men danner grundlaget for at lave en

beslutning. Heatmaps og Choropleth kort, som for eksempel er nedtonede eller mønstret i forhold til målingen af en statistiske variable, er kun nogle af de måder hvorpå geografiske data kan anvendes som vejledning, retningsgivende i en proces, der har til hensigt at barsle med et resultat, investering eller anden form for handlingsplan. Visuel repræsentation af et geografisk objekt som grundlag i en beslutningsproces, er efter min mening kun retningsgivende og burde aldrig danne det egentlige grundlag for en beslutning, som det så ofte prædikes og fremhæves i forbindelse med foreningen af Location Intelligence og Business Intelligence, som repræsentanter for fremtidens løsninger af al forretnings-aktivitet. Dertil er den menneskelige manipulation for stor. Visualisering af geografiske data i et meget lille område, giver sjældent mening. Her tænker jeg på mit eget forsøgsområde, hvor både Heatmap og Choropleth kort samt nuancering af objekter, der repræsenterer en værdi, er anvendt som indikator for en observation. Farvning af polygoner som område-repræsentation, er meget bedre i stor målestok, hvor den første indikation af en forskel, er at observere. På den baggrund kan der ”zoomes ind” på mere specifikke områder, med den førstehåndsinformation, som nuanceringer af kort-arealer via Heatmaps og Choropleth kort fremhæver og til tide ”afslører”. Når området er nede på vej- eller adresseniveau, er det objekter, som repræsentationsmiddel for en geografisk værdi det tager over. Og sådan er det også for mit projekt og de resultater jeg kommer frem til via forskellige analysemetoder.

Konklusionen er, at de to analysemetoder (Excel- og SQL-analyse) som jeg har valgt at ”aktivt” anvende for at manipulere mine data, for at finde strukturer og mønstre i tabeldata til besvarelsen af min hypotese, bearbejder mine data som jeg havde forventet. Med manipulering mener jeg (datamining, dataforædling, data processing og data fusion). Microsoft Excel er et glimrende værktøj til at formidle et budskab og nedbryde data til mindste detaljeringsgrad gennem analyser af datamængder, via beregninger, opstilling af tabeller, skabelse af rapporter, som grundlag for at træffe beslutninger. MapInfo SQL er til geovisualisering – brobyggeren mellem menneskelig og fysisk information formidlet via udforskning, syntese, præsentation og analyse, avancerede GIS-operationen samt kortlægning.

Konklusionen er, at *Excel til fulde opfylder mine ønsker og mål med min analyse* af tabeldata og metode som det formidlende bindele mellem data og information, som jeg have tænkt det, i ***Kapitel 6*** om ***Teori og Metode.*** Forventningen var at ”parringen” mellem to tabeldata via *VLOOUP*-funktionen, kunne vise den ”tekniske” sammenhæng mellem de fiktive bildata og originale boligdata. Den ”ægte” sammenhæng mellem de to tabeldata, og dermed svaret på min hypotese, findes som tidligere nævnt i dette kapitel - ikke. *Eneste* forskel på Excel og MapInfo er *den visuelle fremstilling* af et resultat og *manipuleringen af grafiske data* (punkt, linje og polygon). Excel besidder alle de andre funktioner, om end flere end MapInfo i forbindelse manipulering af data. En anden ”væsentlig” forskel i de to programmer, og det der gør det mere spændende at

arbejde med et GIS-system, er den visuelle baggrundseffekt, når man skal udfylde menuer i programmet. Her kan der være et kort eller et kvadratnet til stede som billede og en menu under denne der *interaktivt forandres* hver gang man indsætter en værdi i menuen. Derved kan man "se ind i beregningen" og dermed tilrettelægge sin manipulation efter "udseendet" af den forestående manipulering. Det kan man ikke i Excel, eller i hvert fald ikke i samme omfang, omend det nye sidste Excel-version (2013) er kommet rigtigt godt med, med hensyn til visualisering ved input af data i formler.

Konklusionen er, at MapInfo SQL-analysemetoden bringer information på et sådant niveau eller stade, at det er muligt at udlede en viden om en begivenhed ud fra det, der er spurgt om. SQL-analysen i MapInfo bringer således ikke ny tal til verden som det ikke var muligt ar beregne via Excel-analyse. SQL-analysen anvender og implementere de informationer (tabeldata) jeg har fundet i Excel-analysen og konvertere denne information til kort og derigennem viser, hvor en hændelse optræder og i hvilken form (Heatmap, Choropleth kort eller punktkort). Vores information bliver visualiseret og der opstår en forklaring eller viden omkring et mønster eller en hændelse fra numeriske data.

Konklusionen er, at den Prædiktiv analyse i MapInfo, den såkaldt "smart" klassifikation ikke virker rent teknisk. Før den Prædiktiv analyse i MapInfo kan køres, skal der først oprettes en indlærings-tabel. Dette er en MapInfo tabel med objekter i et udvalgt områder, der bruges som prøveområde. Det kvadratnet som skulle indeholde og beregne disse værdier, gav af en eller anden grund hele tiden værdien 0 (nul) i tabel-formen og kunne derfor ikke bruges. Jeg ved stadigvæk ikke hvad det er der mangler eller ikke passer i strukturen af de boligdata der skulle danne grundlaget for kvadratnettet. Og det overrasker mig i den grad!!! Prædiktiv analyse var jo tiltænkt som "bindeledet" mellem min indhentede viden fra min SQL-analyse (boligtyper), der skulle fortælle mig, hvor lignende områder (klynger) i PAKILA, kunne være potentielle bolig-typeområder, og dermed få en forståelse af hvordan sammensætningen og indholdet af disse boligområder, der matcher en VW-Golf biltype ville se ud – rent visuelt.

Konklusionen er, at DIKW-hierarkiet som informationsmodel, forklare en informationsstrøm og derigennem formaliseringen af en tankegang, ved at gå fra rå data til evnen til at eksponere information, fremkaldt ved informationsforædlingsmodellering (strukturer og mønstre i data). De forskellige datakilder (Excel-tabeller) jeg bruger i mit projekt, er organiseret i felter og kolonner. Der er ingen bindeled eller logik mellem de forskellige felter og kolonner, men når de forskellige tabeller sættes sammen, via en fællesnævner som for eksempel adresse, dannes der en information, der individuelt eller samlet (tabeldata eller klynger), kan anvendes eller videreudvikles til viden og i sidste ende en forståelse, ved for eksempel at spørge med SQL-sætninger eller fremskrive en hændelse med Prædiktiv analyse.

10. Perspektivering

Vi lever i en alder af informations-stimulering. Vi bliver bombarderet med flere data, end vi kan håndtere eller administrere på en effektiv måde. Rent faktisk kan det distrahere så meget, at vi ikke mere er i stand til at fokusere på selve løsningen af opgaven. Mellem de informationskilder (internettets, sociale medier etc.) og antallet af enheder, vi er tilsluttet (smartphones, tabletter og bærbare computere), er informationsstrømmen uendelig. Vi er nødt til at være fokuseret og strategisk forberedt på, hvordan vi i fremtiden håndterer og administrere information, så vi kan sætte det i sammenhæng og få indsigt eller visdom. I perspektiveringsafsnittet fremlæggers og diskuteres fremtidige problemområder inden for informationsteknologien, der efter min mening, kunne have afgørende betydning for en bæredygtig udvikling, inden for dette område.

10 PERSPEKTIVERING

10.1 Indledning.

Kære læser. Nu er vi ved at være nået til vejs ende i mit projekt og den historie som jeg har forsøgt at fortælle. For at runde dette projekt af på en værdig måde, vil jeg gerne bruge kapitlet om perspektivering, til at sætte fokus på nogle af de emner, som jeg har beskrevet – og ikke beskrevet - i dette projekt, ind i en større sammenhæng og perspektiv, med udgangspunkt i brugen af geografiske data.

Der har været rigtig mange ting, som jeg gerne ville have haft lejlighed til at dokumentere og diskuterer i dette projekt, i forbindelse med besvarelsen af min hypotese og de emneområder, som jeg har berørt i den forbindelse, dels for at vise, hvordan de hænger sammen og berører mange af nutidens trends og arbejdsmetoder, der er gældende inde for vores faglige område og GIS-verdenen i det hele taget og dels for at vise, hvordan jeg selv opfatter og ser det arbejdsmiljø, der til dags dato, har omgivet mig og været en del af min hverdag og på en måde været en - identitet - i de sidste 25 år. Perspektiveringen i dette projekt skal altså ikke kun opfattes som en selvstændig afsluttende del, for at runde af på en "pæn" måde, men mere ses som en videreføring af min indledning, samt fortolkningen af de teorier og analysemetoder, der gør sig gældende inden for området omkring informationsteknologi og som jeg anvender i mit projekt, det vil sige en del af Landinspektørens kompetenceområder, nemlig GIS-verdenen. Med perspektiveringen vil jeg også gerne sætte mit projekt ind i en større samfundsmæssig kontekst og trække nogle mere overordnede linjer fra projektet ud i en bredere sammenhæng, som jeg ser dem i dag.

Berøringsfladerne er mange, med stikord som Location Intelligence, Business Intelligence, Data Mining, Livsstil, Dashboard, Prædiktiv analyse, Big Data, er bolden givet op til en diskussion og fortolkning i det nærmest uendelige med fordele og ulemper, gode og dårlige sider ved anvendelsen af "beliggenhedselementet" som én af de nye fællesnævnere for fremtidens effektive anvendelse og integration af data i erhvervslivet, den offentlige sektor samt naturligvis den individuelle borger.

10.2 Location Intelligence – fup eller fakta.

Udtrykket Location Intelligence bruges ofte til at beskrive værktøjer og data, der anvendes til geografisk "kort" information. Location Intelligence er *"evnen til at organisere og forstå komplekse fænomener ved anvendelse af geografiske relationer, der er indbygget i alle oplysninger"*. Disse kortlægningsmetoder (læs GIS), kan forvandle store mængder data i farvekodede visuelle fremstillinger, der gør det let at se mønstre og tendenser og derved skabe forståelig "intelligens" (viden) omkring objekter. Location

Intelligence kan også bruges til at beskrive en integration af geografiske komponenter i Business Intelligence og de processer og værktøjer, som ofte er indarbejdet og gemt i rumlige databaser.

Udtrykket "Business Intelligence" henviser til *"teknologier, applikationer og processer der anvendes til indsamling, lagring, analyse samt adgang til data"*. Formålet med Business Intelligence er at hjælpe virksomhedens medarbejdere med at træffe bedre beslutninger. Business Intelligence er en kombination af processer, forretningslinjer og teknologier, der anvender rå-data, organisere dem til meningsfuld og handlingsrettet information. Ideen med Business Intelligence er at organiserer information, så den kommer ud til de rette personer på det rigtige tidspunkt, for at understøtte en forretningsmæssig beslutningsproces i en virksomhed.

Location Intelligence er interessant i den forbindelse, fordi det kan *opfattes* som et redskab der:

- Giver stedbaseret og virkelig information om forretningsdata,
- Giver det store overblik og muliggør analyser af komplekse og uensartede data,
- "Genkender" kundedata i forbindelse med identificering af sammenhænge, mangler, tendenser og mønstre,
- *Kan* gøre det muligt at visualisere forskellige situationer og dermed definere et beslutningsgrundlag,
- Sikre en struktureret og integreret tilgang til en beslutningsproces.

Et af de største problemer eller udfordringer med integrationen af Location Intelligence og Business Intelligence i erhvervslivet i dag, som et hjælpemiddel og medspiller i en forretningsproces, som jeg ser det, er, at den almindelige medarbejder i en virksomhed *ikke forstår at anvende* og tilgodese de muligheder, der gemmer sig i GIS-baserede hjælpemidler. Og - ganske få virksomheder i dag *har og ved simpel hen ikke*, at sådanne "hjælpemidler" findes lige om hjørnet, i form af forskellige programmer (ERSI, MapInfo, Geomedia og SAP). Det er ikke nok at *tale* om, hvordan man integrere Location og Business Intelligence i forretningslivet og handelsvirksomheder med fine ord og vending, hvis man ikke kan komme "ned på jorden" og vise og forklare almindelige brugere, hvordan det her Location og Business Intelligence rent faktisk hænger sammen i hverdagens arbejdsrutiner og for eksempel markedsføring, i forbindelse med optimeringer og opsøgningen af nye kundegrupper i forretningslivet. Her findes helt klart et forklaringsproblem.

Det er heller ikke nok at frigive offentlige data – for eksempel geografiske data – som ***Open Source Data***, og tro, at så kan alle bare hente og bruge data frit uden omkostninger. Tanken er okay, men implementeringen og informationen til virksomheder omkring, hvordan de kan anvende geodata som en ekstra *booster* til forretningsdata, har stadig lange udsigter. Mange firmaer vender efter en kort stund med egne erfaringer

inden for brugen af Frie Data, tilbage til den oprindelige dataleverandør, når projektioner, attributter og elementer i de frie data ikke mere matcher de oprindelige set ups fra den tidligere dataleverandørs leverancer af data. Frie data er gode at anvende, når man bare ved hvordan.

Her må jeg så også erkende, at man for eksempel i Danmark med hjemmesiden www.brugstedet.dk, har taget et første og langt skridt, i forbindelse med implementeringen af og udbredelsen af information omkring anvendelsen af Geografiske data i dagligdagen. I Finland findes der ikke nogen tilsvarende informationsorgan fra de offentlige instanser, der varetager denne form for integration og information. Ligesom i Danmark har Finland for nyligt (2-3 år), frigivet en stor del af de geografiske data der findes gemt i Geodatastyrelsen. Men mere er der ikke sket. Mine HRI-data har kun fundet vej til dette projekt, fordi jeg arbejder inden for dette område og til daglig følger "lidt med i", hvad der sker på den front.

Min Excel-analyser og brugen af HRI-tabeldatabasen fra Helsinki Kommune, illustrere ganske godt – syntes jeg selv – hvordan frie data med lidt kreativitet omtanke, kan anvendes som et supplerende sæt information i en beslutningsproces, og underbygge og forstærke denne og være med til at opnå mere viden, via frit tilgængelig information. Men – hvor mange forretningsdrivende i for eksempel Helsinki Kommune ved, at sådanne data er til rådighed, omkostningsfrit, og i filformater, der passer ind i næsten enhver erhvervsvirksomheds IT-system? Er det dataleverandøren, det offentlige, handelskamre, eller andre institutionelle organisationer, der burde informere, have ansvaret for eller markedsføre denne form information? *Location, Business* eller *Intelligence*, Hvad vejer mest?

Beliggenhed *ér* et kritisk komponent i næsten enhver virksomheds forretningsstrategi. Selv om en masse data har en stedlig dimension - det være sig kunder, butikker, lagre eller andre aktiver – bliver denne information sjældent udnyttet i traditionel Business Intelligence analyse. For at få maksimal udnyttelse fra den stadig stigende mængde af data, der bliver genereret i virksomheder i dag, er virksomhederne nødt til at anvende Location elementet fra deres kundedata, for at få en dybere indsigt i deres egen forretningsstrategi og dermed forbedre konkurrenceevnen og virksomhedens resultat. Men – hvem er det så, der er de potentielle målgrupper, der skal implementere dette "værktøj" ind i firmaer, institutioner eller private hjem?

Er det faktisk sådan, at bilforhandleren i Thisted, kan have fordel af at integrere et GIS-baseret kundesystem i firmaet, eller er det nonsens = kundepotentiale i forhold til investering af et GIS-system). Er det en fordel, at fiskeavleren i Varde har et Business Intelligence system, med et kunderegister der kan hjælpe med til at finde nye markeder for hans fisk, via et GIS-baseret program som MapInfo, der kan optimere logi-

stikken, gennem ruteplanlægning og tilføjelse af hans distributørnet med adresser, ved at plotte dem ind på Danmarkskortet med fine farve og se, hvor der findes andre aftagere af fisk af hans type. Eller hvordan med Grundfos. Har de et integreret Business Intelligence system kombineret med Location Intelligence, til at finde kundesegmenter og nye markeder til deres pumper via et fuldt fungerende GIS-system - Næppe. Hvor findes de, dem som skal bruge vores GIS-systemer?

Her er det igen værd at bemærke, at når der for eksempel findes udstillinger, konferencer og sammenkomster for GIS og relaterede produkter inden for dette område, er det ofte de samme mennesker (gamle studiekammerater, udviklingschefen for det konkurrerende firma eller medarbejdere i ens firma, der har fået en fridag), der år efter år besøger informationsdisken for de firmaer, der ud- og tilbyder GIS-produkter. Hvor findes den ”almindelige” bruger, der burde have disse oplysninger?

10.3 Analyse og rapportering – visualisering af information.

Mange virksomhedsledere er under pres for at opnå mere med færre ressourcer. De skal også tilegne sig en stigende mængde information vedrørende stadig mere komplekse forretningsmæssige problemer. Rapportering er en af grundpillerne i Business Intelligence, hvor et nyt fænomen der kaldes ”Dashboards” for tiden vinder frem i popularitet mod ledelsesrapporter. ”Dashboard” er et velkendt interaktivt grafisk værktøj for erhvervskunder til at få adgang til nøgletal, og kan bruges til at visualisere, og dermed bedre at forstå, hvad det er der sket, hvornår og hvorfor, når data fra forskellige filer eller baser integreres.

Ved at tillægge disse Dashboards egenskaber der indeholder kortlægningsmuligheder – visualisering – giver man virksomhederne mulighed for at integrere rumlig analyse, ud til et bredere publikum, og dermed flytte GIS- og Business Intelligence arbejdsredskaber fra en snæver kreds af specialister, ud til alle erhvervskunder. Ingen anden rapporteringsformat kan matche et korts evne til at koncentrere information og sammenligne flere variabler af ét problem. GIS er særlig velegnet til at forbedre ledelsesrapporter og "Dashboards" med interaktive kort og grafik, der tager mindre tid til at evaluere, og som er lettere at forstå. [6]

I ***Figur 35*** har jeg anvendt mine egne data fra min analyse (Excel og SQL), for at illustrere og eksemplificere, hvordan en rapport i Dashboard form kunne tage sig ud, ved at have ovennævnte tekst in mente. Der findes jo uendelig mange måder at fremstille en grafisk rapport på, for at få et budskab ud på den mest effektive måde, forstået på den måde, at også chefen forstår hvad vi taler om. Her er det igen ofte en hierarkisk opbygning af en rapport, der spiller ind. Chefen skal have nøgletallene, til at lave en beslutning. Mellemlederniveauet skal have de samme tal, sammen med en

baggrundsinformation og uddybning af detaljerne, og den "almindelige medarbejder" skal have det hele plus sourcedata, altså hvordan er det her i grunden blevet til, for at kunne forsætte med at implementere en handlingsplan, basseret på et beslutningsgrundlag fra de værdier og information, som det har været muligt at trække ud af tabeller, grafer og kort.

Rapporten kan laves interaktivt, eller manuel. Ved interaktiv rapportering, er det muligt at "scrolle" gennem rapporten, og pege på tal eller data i kortet, der så vises i tabellen og visa versa. Herved kan man se, hvad der sker hvis man øger eller formindsker et tal eller et område, og hvordan det visuelt i kortet, i grafen eller i tabellen, forøges, formindskes eller skifter farve. Integrationen er total!!

Figur 35: *Rapportering ved hjælp af en interaktiv rapport der sammensætter forskellige statistiske materialer fra forskellige filer og baser. Hver gang et emneområde aktiveres, skifter tabeller, grafer og kort farve eller værdi, i forhold til hvad det er der undersøges, som for eksempel (Bolig, Bydel, Parcelhus 2).*

Mulighederne er uendelige, og naturligvis også "manipulationen" af, hvordan data præsenteres. Det er ret imponerende og "dragende", forstået på den måde, at de mu-

ligheder der findes for opbygning af skellige rapporter med tilhørende visualisering og af grafer, kort, og tabeller, er uendelig og det er kun fantasien der sætter grænser. Hver graf, tabel eller kort, kan spares i forskellige filformater (PDF, TXT, JPG med mere) for igen at kunne bruges i en ny Excel-tabel, PowerPoint, eller individuelt i et tekstdokument.

Jeg har indlagt et link til en hjemmeside, der ganske godt illustrere brugen af interaktiv rapportering: http://www.ecrion.com/products/biserver/overview

Det er her at sammenhængen mellem Location Intelligence og Business Intelligence fødes og findes. En pegefinger i den forbindelse skal dog løftes: *Det er skaberen af den interaktive rapport der har ansvaret for, at de rette data, der giver den rette information, på de rette tidspunkt anvendes*. Det betyder at sammenstrikningen af de tabeller, grafik og kort der anvendes til at formidle budskabet, nøje skal overvejes. Her har den eller de medarbejdere, der er involveret i grundlæggelsen af denne informationsstrøm, en meget vigtig rolle i formidlingen af information og viden.

På grund af de enorme mængder af data, der genereres og lagres i virksomheder i dag, på ethvert tænkeligt trin i en forretningsaktiviteter, er det måske uundgåeligt, at fejl undertiden forekomme, der til en hvis del kan miskreditere nytten og troværdighed i informationen. Med mange forskellige datatyper i store klient databaser, kan nogle af disse fejl være vanskelige, ja endog umuligt at opdage i tabeldata-formater. Ofte er det kun, når disse data er trukket ind i en GIS-relateret proces som for eksempel geokodning, kortlægning eller rumligt analyse, at nogle af disse fejl afsløres.

Således var det også tilfældet med HRI-databasen med de ca. 57.000 rækker af boligdata fra Helsinki Kommune. Ved at sortere disse tabeldata efter for eksempel adressen på boliger, opdagede jeg, at der manglede der næsten 1.500 rækker med information, altså ca. 2.5 % af tilgængelige data med ”fejl” eller ukorrekte oplysninger.

For at forblive konkurrencedygtige, investere firmaer i informationsteknologi, forretnings- og færdiggørelsesprocesser bliver mere automatiseret og teknologier der måler effektiviteten, bliver mere og mere sofistikerede. Nettoresultatet af alt dette er, at virksomhederne, som tidligere nævnt, erhverver sig massive mængder af interne og eksterne data vedrørende kunder, forretningsmæssige aktiviteter og økonomiske resultater. Disse data – *Big Data med modifikationer* - er blevet en udfordring, fordi disse data kræver lagring, lokalisering og analyse, så de kan viderebringe information og viden til de personer der skal i sidste ende skal tage beslutninger om fremtidige strategier. Analyse af Big data henviser til processen med at indsamle, organisere og analysere store datasæt, til at opdage mønstre og andre nyttige oplysninger.

Ved at kombinere den analytiske effekt i databaser med de geografiske egenskaber og muligheder der findes i kort, kan virksomheder udforske og analysere forhold mellem geografiske data og virksomhedsdata. Mens Business Intelligence-værktøjer er ideelle til at analysere *hvem, hvad og hvornår* (kunde, produkt, tid), kan de samme værktøjer ikke leve op for at besvare spørgsmål i relation til *hvor* (sted), såsom *sammenhængen* mellem hvor kunderne bor, og hvor de gør deres indkøb. Det kan Location Intelligence.
Location Intelligence er omdrejningspunktet i forbindelsen med at bygge bro mellem virksomhedernes interne og eksterne data aktiver. Vektor- og rasterbaserede topografiske kort, luftfotos, satellitbilleder samt statistik og andre demografiske data, kan blive meget værdifulde variabler i forretningsoplysninger hvis de anvendes i analyser, sammen med de interne data i en virksomhed.

Location Intelligence udnytter også Prædiktiv analyse til at undersøge information fra datasæt, identificere mønstre og forudsige resultater og tendenser med et acceptabelt niveau af pålidelighed. Ved Prædiktiv analyse forstås den praksis at udtrække oplysninger fra eksisterende datasæt for at bestemme mønstre og forudsige fremtidige resultater og tendenser. Prædiktiv analyse fortæller os ikke, hvad der vil ske i fremtiden. Det forudsiger, *hvad der kan ske i fremtiden* med et acceptabelt grad af pålidelighed, og omfatter hvad-hvis scenarier og risikovurdering. Teknikken kan opdatere resultatet af en undersøgelse efterhånden som den bliver tilgængelige. Virksomheder bruger ofte denne fremgangsmåde til at forudsige kundeadfærd, nemlig om folk i bestemte områder er tilbøjelige til at købe et nyt produkt eller reagere positivt på en markedsføringskampagne (recency).

Sådan havde jeg også tænkt mig at lave et lille forsøg med mine indsamlede data for at vise min opfattelse af sammenhængen mellem viden og visdom, som beskrevet i ***Kapitel 6*** om ***Teori og Metode***. Desværre kunne jeg ikke få mit kvadratnet med de data, der var tiltænkt som referencer i den Prædiktive analyse i mit projekt, til at fungere i Vertical Mapper i MapInfo. På den baggrund, har jeg ikke været i stand til at fremvise nogle resultater fra en Prædiktiv analyse, som det ellers var tiltænkt i ***Kapitel 8*** om ***Analysen*****.** Resultatet af den Prædiktive analyse skulle have været dokumentationen for min påstand om, at den viden man har om forskellige mønstre i data – fordelingen af biler og boliger i et område - gør det muligt at vurdere potentialer forbundet med et bestemt sæt af betingelser og dermed opnå den ”højeste grad af viden” nemlig visdom (forståelse for et fænomen). Metoden skulle have undersøger de statistiske karakteristika for flere kvadratnet til indlæsning af data, der er fundet inden for et given "reference/trænings" område og derefter lokaliserer andre områder med lignende egenskaber.
Ideen var selvfølgelig at beregne et datasæt, der havde et mønster, der kunne genkendes andre steder i mit kort og dermed argumenterer for, at der var en korrelation (som beskrevet i ***afsnit 8.2***) mellem mine datasæt og dermed et mønster, som i sin tur kun-

ne betragtes som en viden. Dette har jeg ikke kunnet bevise, så jeg må derfor tage til takke med den litteratur der beskriver Prædiktiv analyse med relation til GIS, Location Intelligence og i sidste ende Business Intelligence og sige, at ja - sådan ville det have fungeret, og med den viden, ville jeg have opnået en forståelses af, hvad sammenhængen mellem forskellige mønstre i et område kunne fortælle om bil- og boligtypers kontinuitet.

Som DIKW-hierarkiet viser, indtager visdom den sidste plads i toppen af pyramiden. *Forståelse* ville måske være en mere rigtig terminologi at anvende i denne her forbindelse. Altså – at via fordelingen af mønstre i et område, ville jeg få en forståelse af, hvordan der kunne være eller ikke være nogen sammenhænge mellem forskellige mønstre i et område, baseret på fordelingen af biltyper og boligtyper. Uden at skulle begynde hele forklaringen om DIKW-hierarkiet én gang til, vil jeg dog understrege det, som jeg også allerede har gjort i ***Kapitel 6*** om ***Teori og Metode***:

"*at opdeling, klassificering og manipulering af data er menneskelig bestemt. Det vil sige, at vi på forhånd opstiller nogle parametre og egenskaber (største og mindste værdier, positive eller negative værdier) som vi er interesseret af at undersøge og trække ud af tabeller med data".* I morgen ville resultatet være anderledes, på grund af ændringer i tingenes tilstand (flytning, dødsfald, ny bil).

10.4 Afslutning.

Jeg håber at læseren har haft samme fornøjelse af at læse min rapport, som jeg har haft det med at skrive den. Det er ikke hverdagskost at lave sådan et arbejde, og der er helt sikkert plads til forbedringer og meninger om, hvordan en indgangsvinkel kunne havde været, under andre forudsætninger. Men det korte af det lange er, at grundstenene i min projekt, Data Information og Viden, har lige så mange fasetter og måder at blive opfattet på, som der er mennesker, der arbejder og interesserer sig for dette emne. I vores fagverden, Geokommunikation, er der naturligvis metoder og teorier til at analysere og forstå sammenhænge, der tilgodeser vores opfattelsesverden af tingenes tilstand, men som helhed, spænder Geokommunikation og anvendelsen af geografiske data sig over et utroligt breds spekter. Det er kun fantasien og endnu ikke udviklede teknikker inden for informationsteknologien, der sætter grænser for implementeringen af vores fagverden, men som "ældre studerende" og tilknyttet arbejdsmarkedet i mange år, kan jeg kun komme med denne opfordring til mine medstuderende - tænk tværfagligt!!

11. Litteratur og hjemmesider

Kildehenvisninger til bøger, publikationer og internetsider er vist som: [Nummer]. Litteraturlisten indeholder foruden henvisninger til bøger, der er anvendt til at udfærdige dette projekt, også henvisninger til hjemmesider. Disse er, lige som litteraturen, også vist med en parantes samt den hjemmeside-adresse (http:// eller www.) hvor informationen er fundet.

11 LITTERATUR OG HJEMMESIDER

[1] Nathan Yau. *Data Points – Visualization that means something*. John Wiley & Sons, 2013, ISBN: 978-1-118-46219-5.

[2] Andy Mitchell. *The ESRI Guide to GIS Analyzing, Modeling Suitability, Movement and Interaction*. Volume 3, ESRI Press, 2012.

[3] Alberto Cairo. *The Functional Art – An Introduction to information graphics and visualization*. New Riders, 2013, ISBN 13: 978-0-321-83473-7.

[4] Nathan Yan. *Visualize This – The Following Data Guide to Design, Visualization, and Statistics.* Wiley Publishing, Inc. 2011. ISBN: 978-0-470-94488-2.

[5] Isabel Meirelles. *Design for Information – An introduction to the histories, theories, and best practices behind effective information visualization*. Rockport Publishers, 2013, ISBN: 978-1-59253-806-5.

[6] Jason Lankow, Josh Ritchie, Ross Crooks. *Infographics – The power of visual storytelling*. John Wiley & Sons, Inc., 2012, ISBN: 978-1-118-31404-3.

[7] Christian Harder, Tim Ormsby, Tomas Balstrøm. *Understanding GIS – An ArcGIS Project Workbook*. Second Edition, ESRI Press, 2013, ISBN: 978-1-58948-346-0.

[8] Michael Law, Amy Collins. *Getting to know ArcGIS for Desktop*, Third Edition, ESRI Press, 2013, ISBN: 978-1-58948-308-8.

[9] Downtown and Business District Market Analyses. Tool to Create Economically Vibrandt Commercial Districts in small Cities. University of Wisconsin-Extention, 2014, http://fyi.uwex.edu/downtown-market-analysis

[10] John Krygier, Denis Wood. *Making Maps – A Visual Guide to Map Design for GIS*, Second Edition, The Guilford Press, 2011, ISBN: 978-1-60918-166-6.

[11] David Maguire, Victoria Kouyoumjian, Ross Smith. *The Business Benefits of GIS – An ROI Approach*. First Edition, ESRI Press, 2018, ISBN: 978-1-58948-200-5.

[12] Richard L. Church, Alan T. Murray. *Business Site Selection, Location Analysis, and GIS*. John Wiley & Sons, Inc. 2009, ISBN: 978-0-470-19106-4.

[13] David W. Allen. *GIS Tutorial 2: Spatial Analysis Workbook*, ESRI Press, 2011, ISBN: 978-1-58948-258-6.

[14] Andy Mitchell. *The ESRI Guide to GIS Analysis, Volume 1: Geographic Pattern & Relationships.* ESRI Press, 2009, ISBN: 978-1-87910-206-4.

[15] Andy Mitchell. *The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements & Statistics.* ESRI Press, 2009, ISBN: 978-1-58948-116-9.

[16] Martin Frické. *The Knowledge Pyramid: A Critique of the DIKW Hierarchy.* Journal of Information Science*, 2007, pp 1-13*.

[17] Anett Hoppe, Rudolf Seising, Andreas Nürnberger, Constanze Wezel. *Wisdom – the blurry top of human cognition in the DIKW-model?,* EUSFLAT Conf. 2011: 584-591.

[18] Jay H. Bernstein. *The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis.* Proceedings North American Symposium on Knowledge Organization, Vol2. Available at: http://dlist.sir.arizona.edu/2633/.

[19] Ackoff, R. L., *From Data to Wisdom*, Journal of Applies Systems Analysis, Volume 16, 1989.

[20] Gene Bellinger, Durval Castro, Anthony Mills. *Data, Information, Knowledge, and Wisdom.* Copyright © 2004 Gene Bellinger.

[21] Jonathan Hey. *The Data, Information, Knowledge, Wisdom chain: The Metaphorical link.* Published at *Intergovernmental Oceanographic Commission – Ocean Teacher: a training system for ocean data and information management* December 2004.

[22] Lars Brodersen, *Geokommunikation*. Copyright © 2007, Forlaget Tanke-gang a-s. pp. 542-543. ISBN 978-87-984-1135-2.

[23] GIS Lounge, *Excel mapping with ArcGIS Online*. 2014 http://www.gislounge.com/excel-mapping-with-arcgis-online/

[24] http://da.wikipedia.org/wiki/Informationsteori

[25] Lars Brodersen, *Informationsledelse med mindset-map*, juli 2014, Forlaget Projektkonsulenten, ISBN-13: 9788799755707

[26] http://inspiredanmark.dk/media/gst/76945/Temabeskrivelser%20for%20bilag%201%20og%202.pdf

I figuroversigten findes der en kronologisk oversigt over de figurer og illustationer, der er anvendt i dette projekt. Teksten efter hvert figurnummer er den første sætning, der er påbegyndt i hver figurtekst. I selve projektet findes den fulde længde af teksten til hver figur.

12. Figuroversigt

12 FIGUROVERSIGT

13. Bilag

I bilaget findes der dokumenter, tekstsider, tabeller og illustrationer, som jeg har vedlagt som et supplement, formidling eller mere detaljerede oplysninger, i forbindelse med udfærdigelsen af min dokumentation. Disse supplerende informationer har jeg ikke fundet hensigtsmæssigt at vise i selve projektet som separate afsnit, hvorfor disse informationer er at finde i bilaget, med tilhørende forklaring og referencer til hovedafsnittet.

13 BILAG

B1 Dataleverandører.

Helsinki Region Infoshare (HRI) *http://www.hri.fi/fi/*

Helsinki Region Infoshare (HRI) tillhandahåller information om Helsingforsregionen så att alla kan komma åt den snabbt och behändigt. Informationen öppnas för medborgare, företag, universitet, högskolor, forskningsanstalter samt kommunförvaltning och statsförvaltning. Informationen finns gratis tillgänglig och får användas fritt. Projektet tar i huvudsak fram statistik som mångsidigt beskriver olika stadsfenomen, såsom levnadsförhållanden, ekonomi och välfärd, sysselsättning, trafik och förflyttning. Många av de informationsmaterial som publiceras inom projektet bygger på geokodad information.

Projektet bygger upp en webbtjänst med vars hjälp man snabbt och behändigt kan ta sig fram till de öppna informationskällorna. Användarna kan sedan ladda ner data och använda dem för t.ex. beslutsfattande, för egna tillämpningar eller för att utveckla och bygga upp helt nya serviceformer. Tanken bakom projektet är att invånarna får ett bättre begrepp om sin regions utveckling ifall den offentliga informationen öppnas för allmänheten. Detta i sin tur ökar förutsättningarna för medborgaraktivitet. Öppen tillgång till information kan också föda nya serviceformer och affärsverksamhet i regionen samt främja forskning och utveckling.

- Ett täckande nätverk av organisationer som äger och sitter inne med basdata byggs upp i Helsingforsregionen. Medlemmarna producerar, upprätthåller, delar och utvecklar närverkets informationsmaterial i samarbete och med gemensamma spelregler.
- Nätverkets informationsmaterial öppnas för alla intresserade. De data som ingår kan också behändigt vidarehanteras och användas för datatekniska tillämpningar. Informationen är gratis och får vidarenyttjas fritt.
- En webbtjänst byggs upp som gör det lätt att hitta önskad information, ladda ner den och nyttja den. Webbtjänsten skall också uppmuntra informationsproducenter och –användare till närmare växelverkan sinsemellan.
- En verksamhetsmodell för öppen information piloteras, och dess verkningar för både dataprodenterna och –användarna utreds. Man lär sig genom att göra – och delar också med sig av sin kunskap.

Projektet finansieras av städerna Esbo, Helsingfors, Vanda och Grankulla samt Jubileumsfonden för Finlands självständighet Sitra. Dessutom har Finansministeriet beviljat kommunsamarbetsbidrag för projektet. För det operativa genomförandet svarar en styrgrupp med företrädare från projektets finansiärer, samt Helsingfors stads faktacentral och Forum Virium Helsinki. Sistnämnda bär i synnerhet ansvar för projektplaneringen, startandet av delprojekt och koordineringen.

Projektet genomförs åren 2010-2012. Efter att ha beretts sedan början av år 2010 kom det igång i juni 2010. En testversion av webbtjänsten publicerades under första hälften

av 2011. Därefter framskred projektet användarinitierat genom testning och inlärning. Nu år 2014, webbtjänsten bjuder på ett rikt utbud av information om Helsingforsregionen och dess olika delar.

Trafiksäkerhetsverket (TRAFI) *http://www.trafi.fi/sv*

Trafi utvecklar transportsystemets säkerhet, främjar trafikens miljövänlighet och svarar för de myndighetsuppgifter som anknyter till transportsystemet.

- Utfärdar certifikat och licenser, godkännanden och andra beslut samt utfärdar rättsregler som gäller verksamhetsområdet.
- Ansvarar för att prov och examina ordnas, för beskattning och registrering inom verksamhetsområdet och upprätthåller en tillförlitlig informationstjänst.
- Utövar tillsyn över transportmarknaden och över att regler och föreskrifter som gäller transportsystemet följs.
- Deltar i det internationella samarbetet.
- Sörjer för att transportsystemet fungerar också i exceptionella förhållanden och vid störningar i den normala verksamheten.
- Skapar förutsättningar för en innovativ utveckling av intelligent trafik.
- Informerar allmänheten om transportrelaterade val.

MaanmittausLaitos/Lantmäteriverket (MML) *http://www.maanmittauslaitos.fi/*

Lantmäteriverket producerar fakta om landet. Vi sörjer för lantmäteriförrättningar, fastighetsuppgifter och kartmaterial samt för lagfarter och inteckningar. Såväl privatpersoner och företag som hela samhället behöver denna information. Vår verksamhet omfattar hela landet från Mariehamn till Ivalo. Hos oss arbetar nästan 1 800 experter på 35 orter. Lantmäteriverket:

- /yder under jord- och skogsbruksministeriet
- ansvarar för lantmäteriförrättningar, kartmaterial, fastighetsuppgifter, lagfarter och inteckningar samt för ett riksomfattande sökregister över all geografisk information
- tillhandahåller information och tjänster om fastigheter, terräng och miljö med anpassning till privatpersoners, företags och samhälleliga behov
- De privata hushållen utgör lantmäteriverkets största kundgrupp. Övriga kunder är näringslivet, staten och kommunerna.

Väestörekisterikeskus/Befolkningsregistercentralen (VRK) *http://www.vrk.fi*
The Population Register Centre was founded in 1969 and it operates under the Ministry of Finance. The offices are located in Sörnäinen, Helsinki, at Lintulahdenkuja 4 and in Kokkola, at Rantakatu 16. The Centre employs some 120 persons.

The basic task of the Population Register Centre is to enable usage of the data contained in the Population Information System and the PRC's certified electronic services to support society's functions and information services and management. Through its activity, the PRC promotes the protection of privacy and personal data as well as information security and the development of and compliance with good data processing and data management practices.

Together with the local register offices, the Population Register Centre is the data controller for the Finnish Population Information System. The PRC maintains and develops the Population Information System, its data and data quality as well as certified electronic services. The PRC offers Population Information System information services and certificate services.

The Population Register Centre also develops and maintains the Public Sector Contact Directory (JULHA) and performs duties related to elections.

KARTTAKESKUS: *http://www.karttakeskus.fi/*
Karttakeskus is the largest IT and professional service company in Finland whose sole focus is the location intelligence, analytics and GIS. Map production and publishing is also included in the production line of Karttakeskus. Main topics covered are:

- Technology, data and management consulting
- System deliveries and integrations
- Solution and application (apps) design and development
- Location intelligence and analytics
- Visualization and analytics development
- HD and tactical route and resource optimization
- Database design
- Data maintenance, management and services (big data)
- Spatial data sets and services

- Service design and user-centered design for service
- Outsourcing services
- Printed and digital consumer products
- The net sales of Karttakeskus were over 11.9 million euros in 2011 with a staff of approximately 100 experts. Karttakeskus is wholly owned by Affecto Plc, listed in Helsinki, Finland

B2 Hvilken bil er du?

VOLKSWAGEN
De fleste kender til mærket Volkswagen - et mærke, der ifølge eksperterne, også er legitimt for folk med høj status. VW er tysk, noget som forbindes med præcision og præstige.
Golf: Golf giver dig ingen sociale problemer og går an overalt. Bilen fordeler sig ligeligt mellem kvinder og mænd og udmærker sig som en såkaldt nummer to bil - en bil, som man gerne kører i til arbejde og praktiske gøremål. Er det derimod en Toyota, må han forklare kollegaerne, at det er en nummer to-bil – det behøver man ikke med en Golf.
Passat: Type: Du er mere bilinteresseret end gennemsnittet. Lægger stor vægt på udstyr og komfort samt køreegenskaber. Du er et ordentligt middelklassemenneske i den øvre ende – formentligt ingeniør i en mellemlederstilling. Nummer to-bilen er gerne en Polo, men man kan også have en japaner.

FORD
Ford er for dem som ikke ønsker at stikke for meget ud. Et moderat mærke, der giver et anonymt udtryk. Folk betragter som regel en Ford som «bare en bil», mens Volkswagen er «bilen».
Mondeo: Er, Ifølge eksperterne, en fantastisk bil, som ligger på andenpladsen efter Passat. De fleste mennesker opfatter Mondeoen som diskret og grå. Den typiske ejer er en gennemsnitlig mand, som formentligt har studeret noget inden for økonomi eller samfundsvidenskab og nu sidder på kommunen som kontorchef.
Focus: Focus er Fords udgave af en Golf. Ejeren er formentligt uddannet inden for sundhed eller pædagogik og arbejder med skole, industri eller offentlig administration. Den typiske Focus-ejer er lærer og over 50 år.

TOYOTA
Toyota er det bilmærke, som folk overordnet set har det bedste indtryk af. Næsten alle tager en Toyota med i betragtningen, når de skal erhverve sig en ny bil. Toyota er typisk for den rationelle bilkøber, som ønsker en problemfri bil. Det forklarer, ifølge eksperterne, også den store andel af kvinder, som køber Toyota.
Avensis: Der er flere kvinder, der køber Avensis end nogen anden biltype. Avensis-ejere kan arbejde inden for alle brancher - dog med en svag overvægt af folk, der arbejder inden for skole- og sundhedssektoren. Er du Avensis-ejer og har en nummer to bil, så er den formentlig en billig brugtbil.
Auris (den nye Corolla): Ejerne er fornuftige og ekstremt optaget af kvalitet og driftssikkerhed. Andelen af kvinder er derfor, ifølge bileksperterne, stor. Corolla-ejere

befinder sig inden for alle brancher – færrest inden for medier og reklamer og flest inden for undervisning og sundhed.

AUDI
Audi er lig med status. Audi-køberen har lov til at være optaget af sin bil, men det er på den anden side ikke accepteret, at man viser sin velstand frem. Blærerøven har ikke en Audi - nærmere en BMW eller Mercedes.
A6: Du er typisk en voksen mand, som tjener godt. Du er bilinteresseret, teknisk opdateret og mere klar over, end de fleste, hvilke værdier der er knyttet til bilen. Med en Audi A6 udstråler du succes og er klar over, at du har stil. Den typiske A6-ejer er markedschef og finkulturel i fritiden.

VOLVO
Folk som kører i Volvo vægter sikkerhed højt – derfor er Volvo en typisk familiebil. Audi har teknikken, BMW har køreegenskaberne, Toyota har driftssikkerheden, og Volvo har sikkerheden.
V70: Klassisk familiebil, som manden i familien formentligt har valgt. Familiefaderen er typisk ingeniør eller sidder på en anden høj stilling inden for universitetsmiljøet. V70-ejeren er højt på strå, men har ikke brug for at vise det til andre.

NISSAN
Bileksperterne sammenligner Nissan-ejere med folk, der kører Toyota. Der er dog det særlige ved Nissan-ejerne, at de altid er på jagt efter gode tilbud – en såkaldt kræmmertype.
Qashqai: Qashqai er proletariatets Range Rover. Vil man gerne have en Quahqai, har man formentligt også kigget på Toyota Rav4 og Honda CRV. Ifølge eksperterne får man meget bil for pengene. Qashiqai-ejeren er ofte ejer af en mindre virksomhed og har et par ansatte under sig.

SKODA
De typiske Skoda-ejere ser på alle typer mærker, når de er på udkig efter en ny bil og anser Skoda som et godt køb.
Octavia: Du har en gennemsnitlig indkomst, er som regel mellemleder i et job uden bonus-forhold og vil i virkeligheden helst have en Audi A6, men økonomien er bare ikke til det. Skodaen hører ind under Volkswagen-koncernen, og det betyder tysk kvalitet, derfor vælger du den.

HONDA
Honda er en japaner med krudt i. Den er med i motorsport, og samtidig er Honda den bil, som står i flest garager verden over - hvis altså man ser bort fra græsslåmaskiner.
CR-V: CR-V-køberen er en illoyal shopper, som også ser på andre bilmærker. Det er en mandebil og ejes som oftest af ingeniører i en ledende stilling.

BMW
Er for bilentusiasten. BMW er teknisk avanceret, sporty og er ikke for den miljøbevidste type, der ønsker at udlede mindre CO2 - BMW-ejeren elsker sin bil, da den er effektiv og økonomisk at køre i.

5-serie: Det er den sporty direktør, der sidder bag rettet i en BMW i 5-serien. Hvis Audi A6 er marketingchefen, er BMW 5-serie for salgschefen. Det er vigtigt for dig som BMW 5- type, at du som bilen, har fart over feltet.

MAZDA
Du er ikke specielt bilinteresseret. Det vigtigste er, at du har en god og kvalitativ bil - du har heller ikke noget imod at spare lidt for at få råd til en Mazda.
Mazda 6: Den typiske ejer er familiefar, som er nødt til at indgå kompromisser for at få Mazda 6. Han arbejder i inden for den offentlige sektor, har ikke alverdens uddannelse og lønnen er heller ikke noget at prale af. Mazda-ejeren vil have en problemfri bil, som kører sikkert.

CITROËN
Citroen har sænket priserne i løbet af de seneste år. Dermed har bilfirmaet tiltrukket sig en del kunder, som ellers normalt ikke er de typiske Citroen-entusiaster. Her vægtes kvalitet ikke så højt.
C3: Ofte er C3-ejerne kvinder, som arbejder inden for sundhedssektoren. Pris og økonomi er to meget vigtige faktorer.

MERCEDES
De fleste Audi og BMW-ejere er lidt misundelige på Mercedes-bilens unikke og populære design. Mercedes er bilverdenens supermodel.
E-klasse: Den typiske køber er mand og ingeniør - ofte med en ledende stilling. Han er ekstremt bilinteresseret og foragter alle japanske bilmærker. Der findes kun Mercedes i hans verden.

ELBIL
Elbilen er til mennesker med en god moral, og som aktivt arbejder på at nedbringe CO2-udslippet. De arbejder typisk selv inden for miljøområdet. Endnu sælges der ikke så mange elbiler, men det mener eksperterne vil ændre sig, efterhånden som de mere funktionelle modeller kommer på markedet.

Kilde: Jyllands-Posten/ Motor 04.05.2009

B3 FME

FME Platform Overview
Safe Software's FME® is the only true spatial ETL (extract, transform and load) platform that helps you easily solve the complete spectrum of data interoperability challenges, including managing proprietary and evolving data formats, adapting to new schemas and lack of standards and difficulties accessing, restructuring, integrating and distributing data.

FME provides support for over 250 GIS, CAD, raster and database formats to let you seamlessly translate, transform, integrate and distribute your spatial data in hundreds of formats - while improving efficiency, reducing costs and lowering risks.

With the flexibility and power of the FME platform, you can quickly and easily:

- Enable spatial data migration / translation projects
- Exchange spatial data between CAD and GIS systems
- Distribute spatial data to multiple user communities
- Extend the lifetime of legacy systems through data interoperability
- Supply spatial data to web applications
- Federate data from disparate database systems
- Transform data from one data model to another

Address the Complete Spectrum of Spatial ETL Tasks
FME is the only solution available that can truly address the complete range of spatial ETL tasks you may need to perform when solving your data interoperability challenges:

Data Translation
Use FME to quickly translate data to/from over 250 GIS, CAD, raster and database formats, while viewing and inspecting these translations at any step in the process.

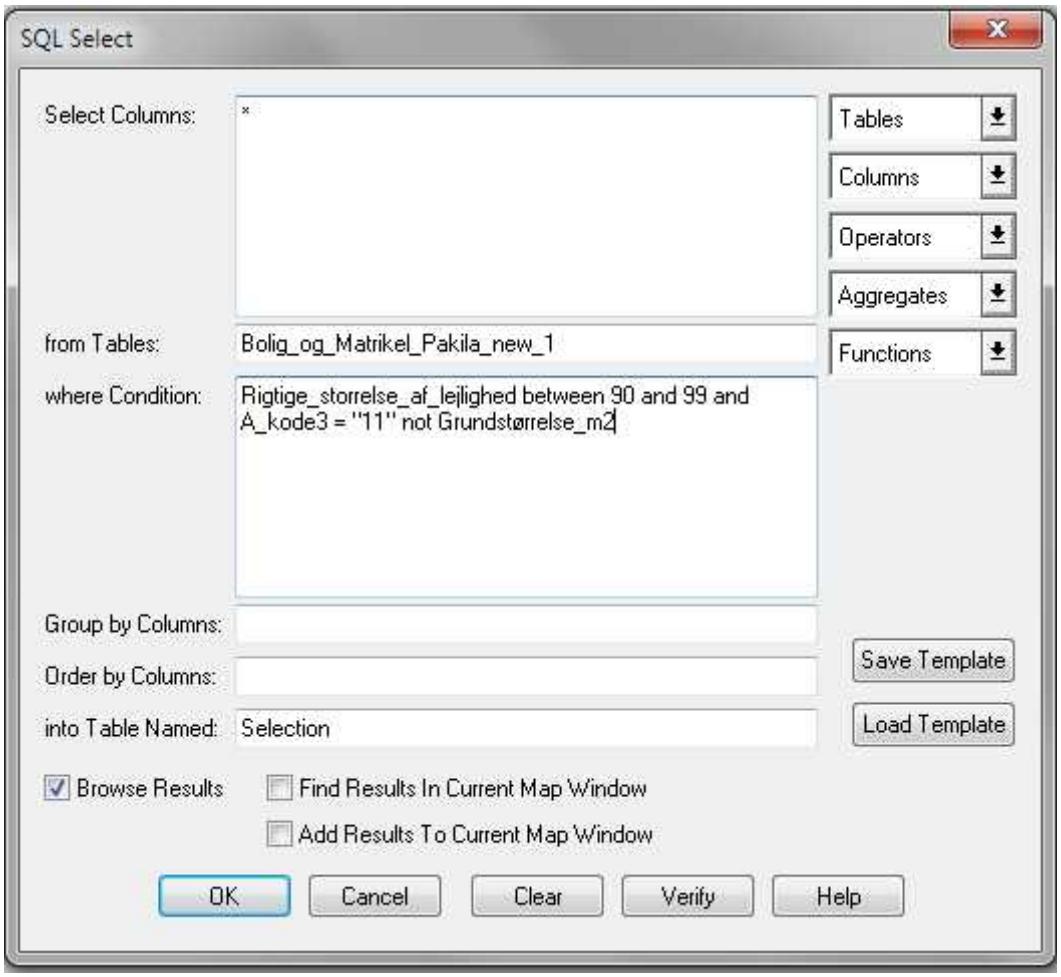
Data Transformation
Transform the geometry of your data using FME's gallery of hundreds of transformers and freely re-project your data into the coordinate system you prefer.

Data Integration
Integrate data from different formats and data structures with complete control using the advanced data transformation capabilities offered by FME.

Data Distribution
Enable large amounts of spatial data to be efficiently distributed over the Web with FME's flexible web services and powerful batch processing capabilities. End users can easily access the most current spatial data in the format and coordinate system of their choice - without increasing the burden on your GIS or IT teams.

http://mdl.library.utoronto.ca/guides-help/converting-gis-data-using-fme-quick-translator

B4 MapInfo SQL-Query Select Function



Denne illustration viser hovedmenuen i MapInfo's SQL-søgningsfunktion. Uden at skulle gå i detaljer med de enkelte funktioner, kan jeg dog berette hovedideen med opbygningen.

I højre side af menuen findes de funktioner der skal undersøges.

TABEL indeholder alle de tabeldata med eget navn, der kan eller skal indgå i en forespørgsel. (Boligtype.TAB, Biltype.TAB)

COLUMNS svarer til de føromtalte Kolonner i ***Kapitel 5*** om ***Teori og Metode, afsnit 5.4*** om SQL-analyse. Her findes de specifikke kolonner med egenskaber som (Adresse, Årgang, Bilmodel, Postnummer, Rigtig størrelse af lejlighed og så videre).

OPERATORS fortæller os om søgefunktionen (større end >, lig med =, Within), og så videre.

AGGREGATES er de funktioner der fortæller os hvilken beregning der skal foretages som nævnt i ***Kapitel 8*** om ***Analyse, afsnit 8.2*** om Generel GIS analyse. Avg. = Gennem-snit, Count. = tæller, Min = Mindste Max. = maksimale og så videre.

FUNKTIONS viser os færdige parametre for en bestemt beregning som for eksempel tilpasning af koordinater efter en transformation til et andet koordinatsystem. CentroidX og CentroidY er de færdige funktionen som man vælger, for at tilpasse punkterne X og Y på den nye plads efter en transformation.

I Venstre side af diagrammet findes de felter, hvori man kan skrive sin SQL-sætning. Her findes der også mulighed for at forudbestemme, hvad og hvordan man vil fortage sin søgning.

SELECT COLUMNS gør det muligt kun at søge i specifikke kolonner af interesse i samme eller flere tabeller. Derved opnås et mere specifikt resultat for sammenligning.

FROM TABLES er det felt, hvor de forskellige tabeller, der skal analyseres, vælges. Flere tabeller kan vælges på samme tid, alt efter søgningens intention.

WHERE CONDITION er det felt hvor selve SQL-sætningen skabes. Her er det kun fantasien der sætter grænser, foruden naturligvis sætningsbygningen, der skal indeholde de rette syntakster, for at have en beregningsværdi. Her skal man ofte rette til, før at man får meddelelsen, *"Syntax is correct"*. Derefter kan søgningen foretages og et resultat kan vises, enten som et kort eller en tabel. (Map or Browser).

De sidste tre felter **GROUP BY COLUMN, ORDER BY COLUMN** og **INTO TABLE NAMED** er mere funktioner for den endelige præsentation af søgningen på tabelform. Disse funktioner er derfor mere af sekunder betydning for en søgning.