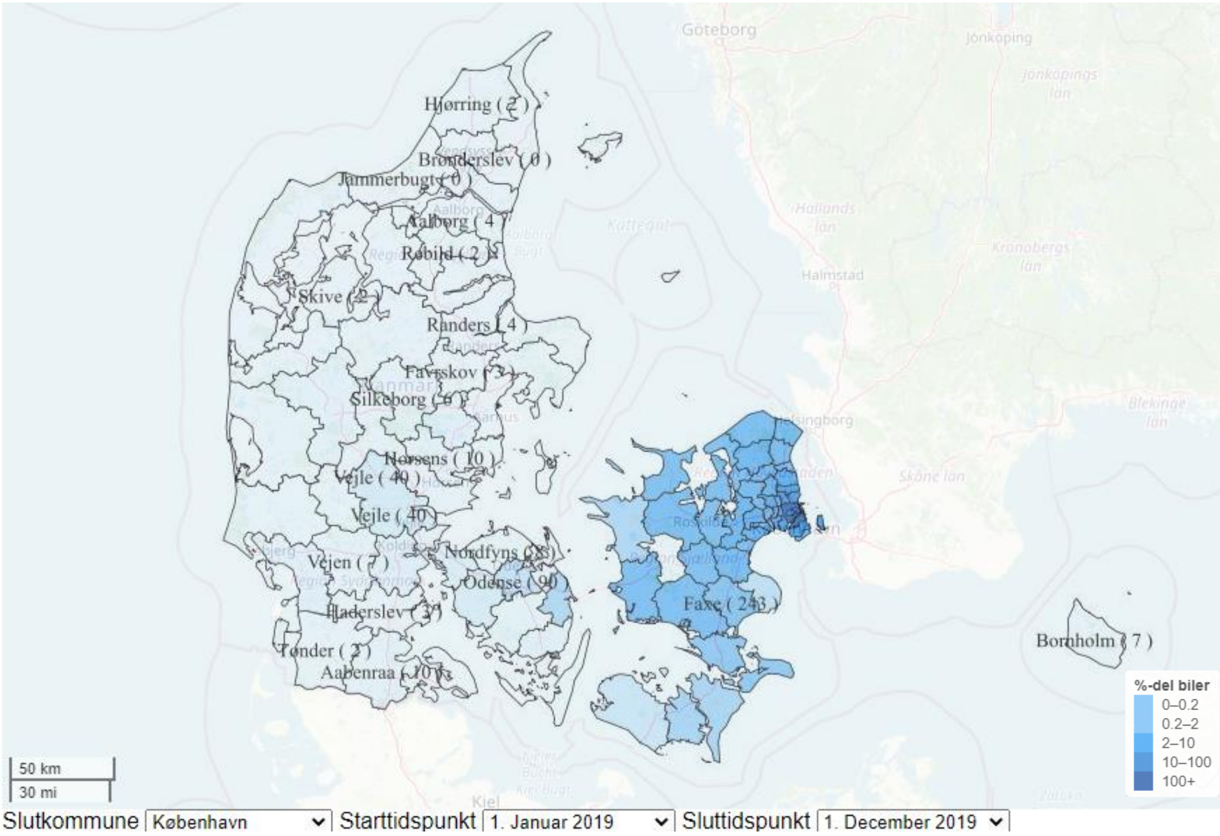
AALBORG UNIVERSITY

Cand. Tech. Surveying and Planning – Geoinformatics



Spatio-temporal pattern estimation in urban and traffic planning

- An exploratory analytical tool

Author:	*Student no.*	*Supervisor:*
Louise H. F. Kristiansen	20154598	Jamal Jokar Arsanjani

[Page intentionally left blank]

Title:

Spatio-temporal pattern estimation in urban and traffic planning

– An exploratory analytical tool

Project:

Master's Thesis (Cand.Tech.) in Surveying and Planning – Geoinformatics, Aalborg University, Copenhagen

Project period: February-June 2020

Student:

Louise Hejlskov Flygstrup Kristiansen

Supervisor: Jamal Jokar Arsanjani

Total number of pages: 70

Number of standard pages: 45
(400 words per page – Abstract, summaries, bibliography, appendices excluded)

Number of Appendices: 3

Hand-in date: 4th of June 2020

Abstract:

This dissertation proposes an exploratory analytical tool for examining spatio-temporal patterns in any area of interest for urban and traffic planning. By combining solely Open-Source technologies to create a user-friendly analytical tool, planners can easily adjust the tool to fit multiple purposes, such as keeping track of the improvements towards reaching UN's Sustainable Development Goals or finding congested areas in urban areas. The latter has been done throughout this study, using a large set of Origin-Destination pairs, collected from cars in Denmark during 2019.

For the case study, the morning and afternoon traffic internally in Copenhagen and Frederiksberg municipalities are compared, to examine whether the flow of cars is different in the morning than in the afternoon, as well as to examine if there are any areas that are more common as destinations. It seems that there are some areas, where more cars end their trip, both during the morning and afternoon peak hours, which might cause more congestion during these hours. Comparison with the existing infrastructure of public transport and bike lanes shows no coherence between poor infrastructure for transportation alternatives and common destinations.

Summary (DK)

Som led i specialeafviklingen på kandidatuddannelsen i *Surveying and Planning* med *Geoinformatik* som specialisering, er der i løbet af foråret 2020 blevet udarbejdet et analyseværktøj til brug af trafik- og byplanlæggere.

Analyseværktøjet er udviklet i samarbejde med COWI, som 1. opstillede et ønske om at få et lignende værktøj udarbejdet til fremtidige analyser i deres afdeling for trafik- og byplanlægning, og 2. stillede GPS-data til rådighed til brug i projektet. Data er indsamlet af bilister over hele landet, og samles i en database, som COWI har fået adgang til via Connected Cars database, under forudsætning om hemmeligholdelse og anonymisering af personhenførbare data. GPS-data brugt i dette projekt indeholder ikke hele ture, men blot start- og slutpunkterne (Origin-Destination).

Værktøjet er udviklet med en kombination af gratis, Open-Source teknologier, der tilsammen opstiller et værktøj, der let kan implementeres i forskellige sammenhænge. Værktøjet er let anvendeligt og brugervenligt, hvorfor det gør sig godt i mange planlægningsopgaver, ved at tilpasse til det data der er tilgængeligt til den givne opgave.

Med brugen af dette GPS-data er der udviklet en første prototype af et analyseværktøj, der aggregerer GPS-data i grids og zoner, herunder kommunezoner og zoner fra Landstrafikmodellen (LTM), der tilsammen giver to forskellige skalaer at lave analyser på. Værktøjet er udarbejdet ved brugen af Leaflet, der forbindes via et GeoServer-lag til databasen, hvorfra data trækkes baseret på brugerinputs. Brugeren kan vælge hvilke måneder, analysen ønskes set for, samt hvilket område (kommune eller LTM-zone). Det er desuden muligt at vælge tidspunkt på dagen, der ønskes analyseret, hvilket gøres ved at definere et lag i GeoServer, der aggregerer data inden for en given tidsramme.

Der er efterfølgende blevet analyseret på nogle af disse udtræk af data, hvor der har været fokus på at finde geografiske og tidsmæssige mønstre. Dette er gjort for at analysere, hvor der er størst risiko for klyngedannelse, hvilket kan give trafikmæssige problemer.

Analysen fokuserer primært på Københavns og Frederiksberg Kommune, da det er de områder med flest bilejere. Der ses en tendens til, at der både om morgenen og eftermiddagen kører mange biler ind i København. I morgenmyldretiden ses, at flere kommer længere væk fra og mere jævnt fordelt over de tættestbeliggende kommuner. Om eftermiddagen, derimod, er det især i sydgående retning fra Nordsjælland ind i København og sydfra langs Køge Bugt Motorvejen, hvorfor det primært er her, der vil kunne opstå problemer med kødannelse. Samtidig er der i begge tidsintervaller et stort antal interne ture i de to kommuner, og en nærmere analyse af de to kommuner er gennemført. Denne viser sammenhængen mellem hvor de fleste biler slutter deres ture i et OTM-zonelag. Dette

sammenlignes med den eksisterende infrastruktur for at analysere mulighederne for forbedring af situationen.

Der er efterfølgende blevet diskuteret, hvorvidt samt hvordan by- og trafikplanlægningsafdelingen i COWI, og by- og trafikplanlægning generelt, kan drage fordel af et sådant analyseværktøj. Derudover diskuteres de metodiske overvejelser, der ligger til grund for den valgte fremgangsmåde.

Problemfeltet blev afgrænset til en undersøgelse af, hvordan et værktøj kunne opstilles samt bruges i trafikplanlægningsmæssige sammenhænge til at udpege og analysere mønstre i trafikken efter tid og sted. På baggrund af det arbejde, der er udført gennem de seneste cirka fire måneder, kan det konkluderes, at et værktøj som det, der er udviklet i løbet af specialeperioden, kan gavne by- og trafikplanlægning i at udføre hurtige og lettilgængelige udtræk af O-D data, som kan bruges til diverse analyser af trafikken. Analyseværktøjet kan bruges som et udpegningsværktøj i de indledende faser af en analyse, for at udpege kritiske områder. Desuden kan det være med til at validere de modeller, der bruges i dag, såsom Landstrafikmodellen, til at forudsige det fremtidige flow af bilister. Derudover kan analyseværktøjet tilpasses til forskellige formål og forskellige datasæt, hvormed dette analyseværktøj kan ses som en skabelon til, hvordan analyser af data kan benyttes til at udpege mønstre i tid og sted. Opfølgning på FN's verdensmål er en af mange måder, hvorpå et sådant analyseværktøj kan være gavnligt.

Summary (EN)

During spring 2020, an analysis tool for the use of traffic and urban planners has been proposed as part of the dissertation at the master's program in *Surveying and Planning* with the specialization in *Geoinformatics*.
The exploratory analytical tool was developed in collaboration with COWI, who 1. proposed to have a similar tool prepared for future analyses in their Urban Planning and Transport department, and 2. made GPS data available for use in the project. Data was collected by cars across the country and is collected in a database that COWI has accessed through Connected Cars' database, on the condition of secrecy and anonymization of personally identifiable data. GPS data used in this project does not contain entire trips, but merely the starting and ending points (Origin-Destination).

The analytical tool has been developed with a combination of free, Open-Source technologies that together create a tool that can be easily implemented in different contexts. The tool is convenient and user-friendly, which is why it works well in many planning tasks, by adapting to the data available for the given task.

With the use of this GPS data, a first prototype of an analytical tool that aggregates GPS data in grids and zones has been developed, including municipal zones and zones from the Land Traffic Model (LTM), which together provide two different scales to make analyses. The tool is designed using Leaflet, which is connected via a GeoServer layer to the database, from which data is drawn based on user input. The user can choose which months the analysis is to be seen for, as well as which area (municipality or LTM zone). In addition, it is possible to choose the time of day to be analyzed, which is done by defining a layer in GeoServer that aggregates data within a given time frame.

Subsequently, some of these extracts have been analyzed, focusing on finding spatio-temporal patterns, to examine where the greatest possibility of cluster formation can be seen, which can cause traffic problems.

The analysis focuses primarily on the Copenhagen and Frederiksberg municipalities, as this is where the car ownerships is at its highest. There is a tendency for many cars to drive into Copenhagen in both the morning and the afternoon. In the morning rush hours, it seems that cars are entering Copenhagen from further away and that the origins of the trips are more evenly distributed in the adjacent municipalities. In the afternoon, more cars come from northern Zealand into Copenhagen and from south along the Køge Bay Motorway, which is why it is primarily here that problems with queue formation occurs. At the same time, there are numerous internal trips in both municipalities

in both time intervals, and a closer analysis of the two municipalities has been carried out. This shows the relation between where most cars end their trips in an OTM zone layer, which is then compared to the existing infrastructure to analyze the opportunities for improving the situation.

It has subsequently been discussed whether and how the Urban Planning and Transport department of COWI, and urban and traffic planning in general, can benefit from such analytical tool. In addition, the methodological considerations underlying the chosen method are discussed.

The problem area was limited to a study of how an exploratory analytical tool could be set up and used in traffic planning contexts to identify and analyze traffic patterns in time and space. Based on the work that has been done over the past four months, it can be concluded that a tool, such as the one developed during the thesis period, can benefit urban and traffic planning in carrying out fast and easily accessible extractions of OD data that can be used for various traffic analyses. The analytical tool can be used as a designation tool in the initial stages of an analysis, to identify critical areas. In addition, it can help validate the models used today, such as the Land Traffic model, to predict the future flow of motorists. In addition, the tool can be adapted for different purposes and different datasets, which can be seen as a template for how data analyses can be used to identify patterns in time and space. Following up on the UN's global goals is one of many ways in which such an analytical tool can be beneficial.

Preface

This report presents the work done in relation to the thesis of the Master's programme (Cand.Tech.) in *Surveying and Planning - Geoinformatics* at Aalborg University in Copenhagen, carried out by Louise Hejlskov Flygstrup Kristiansen. The study period was 4 months, beginning in February and lasting until June 4th, 2020. The learning objectives of the curriculum was the only limitations to the scope of the thesis. This report's focus is developing an exploratory analytical tool for COWI's Urban Planning and Transport department and carrying out analyses based on the tool.

Acknowledgements

The application details were set out by COWI employee Jonas Hammershøj, supported by Rasmus Guldberg Jensen and Peter Banke Ravn. I would like to thank all three of them for the support along the way, and for making it possible to do this collaboration for my master thesis and for helping me getting the data needed. I would also like to thank the team at Connected Cars for letting me use their data in my thesis. Without it, this study would not have been possible.

From COWI, I would also like to thank James Richardson, who assisted me in my SQL processes and whenever I ran into problems related to the database, as well as Aske Butze, who helped me in the process of creating the analytical tool.

Lastly, I would like to thank my supervisor, Jamal Jokar Arsanjani, who helped me finding the focus area and assisting me along the way and helping me stay motivated, when times were difficult.

Readers guide

The report contains hyperlinks and internal cross-references, which makes it suitable for electronic reading.

Throughout the report, figures and tables will be presented to support the text. If no credits are listed, the figure or table are made by the student.

The text contains in-text references including the author(s) and the year of issue. For a full reference refer to the bibliography by the end of the report. Figures and tables from other authors than the student will be referenced likewise. The methods used for referencing is the Harvard Method.

To the main report follows three appendices. Appendix A includes the search criteria for the literature review, as carried out in chapter 3.1 *Literature review*. Appendices B-C contains the full length of the code behind the analytical tool. When considered relevant for the understanding of the context when reading, code snippets is displayed alongside the text.

Abbreviations

COVID-19 – Corona Virus Disease

CSS – Cascading Style Sheets

DB – Database

GeoJSON – Geographic JavaScript Object Notation

GDPR – General Data Protection Regulation

GIS – Geographical Information System

GPS – Global positioning system

HTML – Hypertext Mark-up Language

JS – JavaScript

JSON – JavaScript Object Notation

LTM – Landstrafikmodellen

O-D – Origin-Destination

OTM – Ørestadstrafikmodellen

QGIS – Quantum Geographical Information System

SDGs – Sustainable Development Goals

SHP – Shape (file)

SLD – Styled Layer Descriptor

SQL – Structured Query Language

SSH – Secure Shell

UN – United Nations

VM – Virtual Machine

WFS – Web Feature Service

WMS – Web Map Service

Table of Content

List of Figures

List of Tables

List of Code

1 Introduction

The world population keeps rising, and by 2050, UN expects an additional 2.5 billion people to live in urban areas around the globe, and that we will see more megacities. The growth and the speed in which the urban areas are developing will put a great pressure on the cities and planners to help solve some of the problems that will follow, including for transportation, housing, energy supply and much more (United Nations, 2018). This calls for planning of these areas more than ever.

Throughout this master thesis study, an exploratory analytical tool that can help examining the changes and patterns in speed and time, has been created. This study will undergo the steps of creating this tool that is to optimize the workflows of traffic and urban planners as much as possible, as well as being applicable to multiple purposes. Furthermore, an analysis based on the results from this tool will be executed, to illustrate how the results could be used in traffic planning.

There are multiple ways of planning the traffic infrastructure, in order to make the flow as good as possible for as many cars as possible. With the technological development, more and more devices now contain a GPS receiver (Schoier & Burroso, 2012), which can help planners change the cities better suited for the needs as seen from the GPS receivers. For the purpose of developing an analytical tool, a set of GPS points has been provided, covering all of Denmark. The data has been collected throughout a year and will be the base for the analysis. The GPS points are joined with a grid, from where information about origin and destination of trips can be called in a web map application. This makes the process of collecting big amounts of data faster and easier, even for traffic planners who do not have the expertise and competencies within handling or extracting big amounts of data from a database.

Based on a literature review, it seems that there has been no study regarding the development of an exploratory analytical tool that can be adjusted for multiple purposes. Hence this thesis will add a new way of working with spatio-temporal data, as an analytical tool like that will be examine a way of simplifying the processes of extracting information from a large set of O-D pairs to be used for various planning purposes. The case study will support the development and discussion of the tool and its utility.

Based on the results of aggregated data that can be extracted from the tool, an analysis of the flow of cars in Copenhagen and Frederiksberg municipalities is carried out, followed by an analysis of some of the existing infrastructure to help improve the flow of cars in an urban area like Copenhagen. The benefits and the relevance of the tool will be discussed, as well as how urban and traffic planners can benefit from it, e.g. in the follow-up on UN's Sustainable Development Goals.

This thesis is based on a request from COWI, who desired an interactive map portal solution for traffic planning to be created in order to use for O-D analyses in the future. In collaboration with them, the theme has been chosen to fit both their needs and requirements, as well as having to meet the requirements from the study board.

1.1 Problem Statement

Spatio-temporal patterns can be seen in most of our daily activities, changes in nature, climate, and much more. It can be detected on a world-wide scale or in a very local scale, even just inside a building, and the movements can either be detected as specific path or as a set of Origin-Destination pairs.

Various models and algorithms have been developed to predict and examine different patterns in time and space, most of them being applicable for a certain type of analyses, such as micro-mobility analysis or identification of relevant attractions. However, it seems that most these methods and analytical tools are specific for a certain type of spatio-temporal pattern detection, which either focus on estimating most likely trajectories between a set of O-D or purpose-specific algorithms that are not applicable to other types of data or useful in other case studies (refer to chapter 3 *State of the art*).

This study examines the development of an exploratory analytical tool that, based on user inputs, calls the database and visualize aggregated spatio-temporal data on a basemap. The analytical tool should be applicable to various purposes and case studies, by changing some parameters. By not black-boxing the composition of the used Open-Source software, the analytical tool in this study can be applied as a template of a tool that can be adapted to the purpose of need.

In order to validate the usefulness of the analytical tool, a case study will be conducted, focussing on O-D pairs collected from GPS in personal vehicles. Additionally, this study will discuss some other possible use cases of the analytical tool, and the usefulness of it in traffic and urban planning.

1.1.2 Research questions

In order to examine the problem statement at hand in depth, the following research questions has been used as a guideline. These will be answered throughout the study and will be further discussed and followed by conclusions born from the achieved results.

- Having a unique dataset, how does the spatio-temporal pattern of cars across Denmark look like? Are there any distinct patterns in time/space?
- How can an exploratory Origin-Destination tool using mobility coordinates be used in traffic and urban planning to indicate congested areas?

- How can an Origin-Destination analysis based on a large set of GPS data from cars benefit urban/traffic planning?

1.2 Research Design

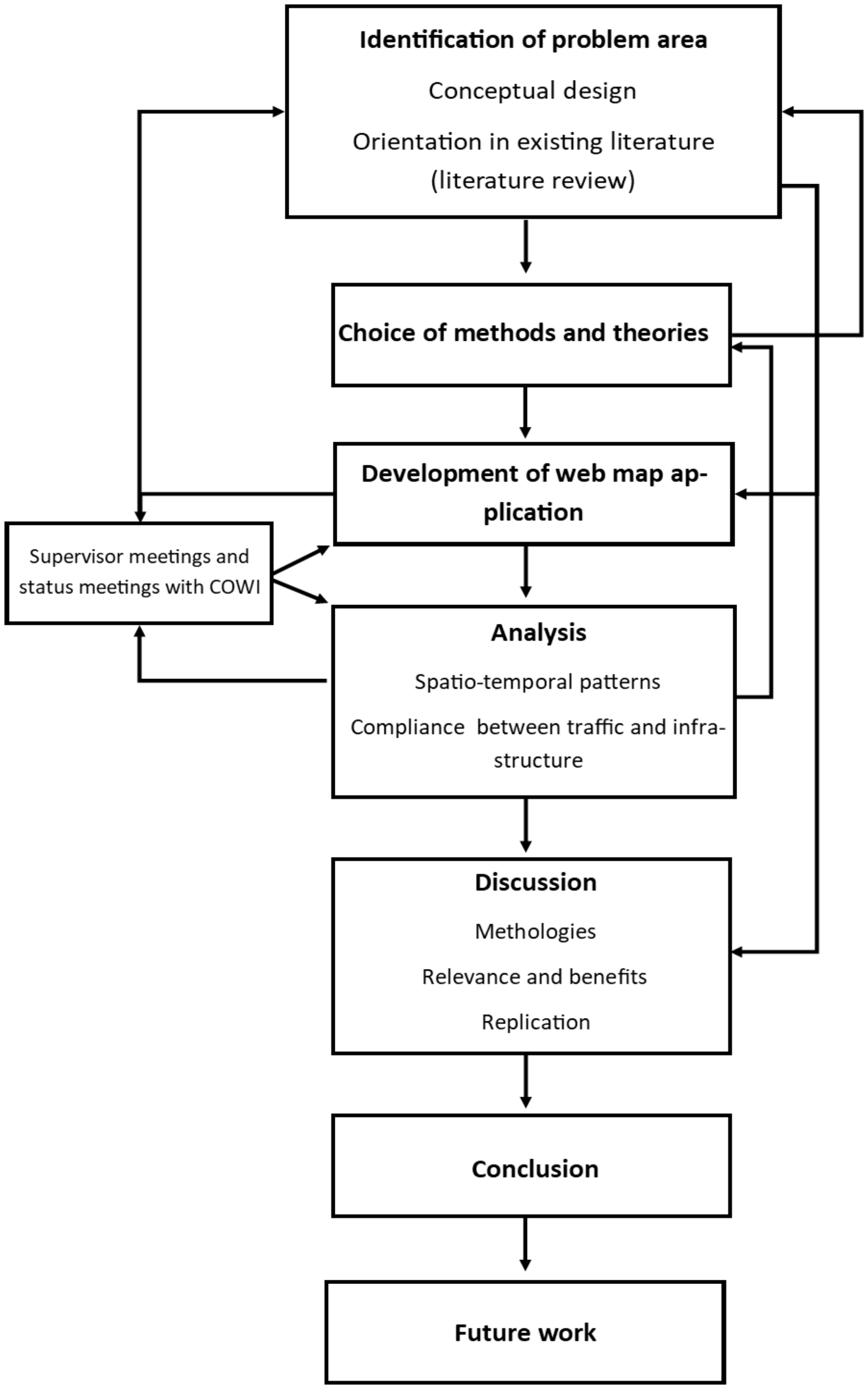


Figure 1: Research Design

Figure 1 shows the research design of the thesis. As the figure reflects, the process was not as linear, as the report reflects. This is due to the learning curve during the process, where a greater knowledge within one area led to researching further within this or closely related areas.

The process started by finding the initial scope of the problem area, which was done based on a preliminary meeting with some of COWI's traffic planners to talk about their inquiry of what type of tool and analysis they had in mind, combined with a supervisor meeting. This was followed up by a literature search, in order to examine what types of work has previously been done within this field.

The choice of methods and theory was based on the meetings and the literature review. When reading through theories and methodologies, more theories and methods were presented, which lead to a further research of these methods and theories.

With a basis of theory and methodologies presented, the development of the tool started. This, and the analysis was an ongoing process, which looped back and forth many times; from development of the tool, to meetings to agree on the following steps in order to carry out the analysis, to updating the methods and theories based on what has been done and used. Through most of the process, new literature was found and applied, in order to support the analyses and choices made along the way. This leads to the discussion of the chosen methods and the relevance of analyses and the tool made, which is done to validate the results of the study, as well as to see where there is place for improvement. This was examined as the last part of the thesis, after concluding on the work done.

2 Conceptual Design

This chapter will give a short introduction to what the initial thoughts for this study was. The contents of the conceptual design are based on meetings with Jonas Hammershøj, Peter Banke Ravn and Rasmus Guldborg Jensen from COWI's Urban Planning and Transport department.

The idea for this thesis initially occurred during my internship at COWI, who has a lot of analyses and work they wish to be done, using the huge amount of GPS data they have available. Luckily, there was one of these solutions that fit to the terms set by the study board.

What COWI proposed me to work with, is a tool which can calculate the results of an Origin-Destination pair of GPS points on a grid level. By loading the GPS points into grid cells, the processing time of the data would be minimized as well as data would be more anonymized, by not showing the exact start and end point of a trip. It should be possible to read in an area of interest, from which the results should give results. Furthermore, it should be possible to choose a period of time and a time interval to get the data extract for. In that way, the user could select an area, and see how the flow is in or out of that area at a given point in time, e.g. 8:00 AM on a Monday (either a specific or an average of all Mondays within the selected time period).

The resulting tool, created as an interactive web map application, should be useable in traffic planning, where the GPS data can bring some validity to traffic models. The aggregated GPS data can give an idea of how the flow of cars are in reality, whereas the models can only give a prediction of it as well as it is to be used in analyses of the traffic flow of a given area for various projects. It is meant to be a tool that decreases the processing time and working hours put into it, as it is a process that will be replicated multiple times.

As the time provided for the thesis was a limitation of fulfilling all the requirements, a first prototype was developed through the last months. The tool is limited to contain only two zone layers of different scale, as it was enough in order to carry out an analysis on the base of this. Furthermore, not all functionalities were done yet, such as being able to select the exact time periods down to 15 min on a specific day. The data is aggregated into larger time windows, which gives more data to rely the analysis on. However, it is possible to choose the area of interest as well as a time period, for which the user wants to see the aggregated data, but some more processing would make the tool even more user-friendly, which will be explained in more detail in chapter 9 *Future work*.

3 State of the art

This chapter will undergo the literature review that was carried out within the themes of this thesis. This is done to get an overview of what has already been done in the research field by spring 2020 and to see to what degree this thesis will add new knowledge to the area.

3.1 Literature review

For the literature review, Scopus was the database of choice, because it is an international database, and even though the focus is in a Danish context, it is relevant to see the general development within the area in a world-wide perspective. Furthermore, Scopus was chosen as it is the "world's largest abstract and citation database of peer-reviewed research literature." (Scopus, n.d.) and through Aalborg University's library, it was possible to gain access to most of the literature. There are many other databases that could have been chosen, and if any other database had been chosen, the same string of words would most likely result in other literature that would have a different focus.

The literature review was conducted in three steps; the first step was carried out in the beginning of the process, where the knowledge to the research field was limited, a second step that was carried out later, based on knowledge about the field and the literature from the first step, and a third part, which took place during the entire process of the study.

Through the first step, a few key words and synonyms were used, in order so search rather broad. Though, this was limited by publication year, subject area, language and source type, in order to find the most relevant research, as there was some *noise* in the found literature. Table 4 in Appendix A shows the search criteria of the first step of literature review carried out in the beginning of March.

With a bit more insight to the study area and a more focussed research area, the second step of the literature review was carried out in the same way as the first step. The search criteria for this step follows right after the one of the first step in appendix A, in Table 6.

The third and last step was conducted over and over, in order to find relevant literature within the field regarding methods, theories and inspiration for the development of the tool.

3.1.1 Results of literature review

When doing the first step of the literature search, including refinements (see appendix A) in March 2020, it gave 31 results. Of these, there were 28 journals and 3 books series. Each of these were categorized based on their relevance for the thesis. This was done by reading through the abstracts, introduction and conclusion of the 31 texts, to get an idea of their relevance.

In appendix A, Table 5, each text has been evaluated in regard to their relevancy. The same is done for the second step of the literature review, but with some words being changed as the study area was more focussed at the point of doing the second step of literature review.

As seen in appendix A, Table 5, only four texts from the first step were marked as "relevant"; what they have in common, and why they are related to this thesis, is because they focus on pattern estimations, travellers' behaviour, traffic models, and clustering of traffic (Rao, et al., 2018; Tang & Cheng, 2016; Rasmussen, et al., 2017; Schoier & Burroso, 2012). (Rao, et al., 2018) analyses the behaviour of different types of travellers and route choice models, to conclude that most travellers choose the same route to/from work every day. This article is relevant because it investigates travel behaviour and traffic models and uses these to estimate the future traffic flow. Besides basing analysis on estimation models, this research also differs in the level of detail, as (Rao, et al., 2018) analyse the traffic situation at link level, and the thesis is working on a larger scale. Similarly, (Rasmussen, et al., 2017) differs from this thesis in scale and level of detail, but also investigates traffic models and compares the reliability and computation time. This thesis will also compare the methods used with other traffic models, but other models are used for the comparison, and the comparison will not be as in-depth. This article also focusses on the Danish road network, why there are a greater similarity between this article and the study. However, the study by (Rasmussen, et al., 2017) examines the validation and calibration of a *Restricted Stochastic User Equilibrium with Threshold model* and validates the model on eastern Denmark. (Tang & Cheng, 2016) estimates travel patterns in Kunshan, China, based on Automatic License Plate Recognition to be used in real-world traffic management. They concluded the method to be very affective and accurate, but the method cannot be applied in this thesis, as is focuses on O-D GPS data. Finally, (Schoier & Burroso, 2012) has developed a density-based algorithm to help find clusters along different routes. By doing so, they wish to see if congestion affects the route choice of travellers. What can be said about these texts in general is that they either create or evaluated each their method or algorithm.

The second step of the literature review resulted in 14 texts based on the selection criteria that can be seen in appendix A, Table 6, 10 articles and 4 book chapters. The literature in this step of the literature search were focussed more on the spatio-temporal aspect. These texts where focussed more on examining and determining any possible patterns in time and space, for various focus areas. Some of these focussed on estimating the trajectories between the origin and the destination, to gain a more detailed understanding of the movements and patterns in an area *(Zhua, et al., 2020; Naghizade, et al., 2020; Song, et al., 2017; Song & Miller, 2015; Wu, et al., 2014; Millward, et al., 2013)*. As is seems that trajectory estimation based on O-D pairs seem to be covered sufficiently, this was chosen not to focus on for this study. *(Cerqueira, et al., 2018)* proposed a model for analyzing large GPS datasets including multiple aspects, and similarly did *(Jiang, et al., 2015)* create a visual tool for pattern estimation of taxicabs in Hangzhou. The tool created in their study enables the user to find statistics and graphs that allows for interpretation and analysis. The method proposed by *(Cerqueira, et al., 2018)* evaluates the quality of a GPS dataset by examining the data based on statistical indicators.

Based on this literature review, it is clear that some of the literature has worked with similar themes, as the one that was the idea for this study. It seems, based on the literature found, that there have been no recent studies focussing on creating an exploratory analytical tool that could be used for multiple purposes, such as optimizing the work flow of planners, who could easily aggregate a large dataset to determine spatio-temporal patterns in big cities as well as in the countryside.

The literature review hereby constitutes a scientific evidence that this thesis' focus is relevant as it will examine an area of research that has not been worked with, at least in the past ten years. Accordingly, this thesis will add new knowledge within the field of research, as it will focus on developing an exploratory analytical tool that can be used for multiple purposes. The study will additionally include a case study of spatio-temporal patterns based on O-D data from GPS points.

It is crucial to mention that this is just one way of doing the literature review and using other strings of words might result in finding other documents that are relevant for the area of research as well. The synonyms and word strings may be very different from researcher to researcher, and this literature review shows one way of combining words and criteria to find relevant literature. Replicating the literature review at a later point will result in more texts if something has been published. The search string was checked again in the end of May to ensure that no new literature had been published on the subject.

Because of the fewer criteria for the third step of the literature review, a great all-round knowledge to the area was obtained, but will not result in as much academic content, as well as some of the literature found might contain outdated theories, methods or knowledge. This requires one to be a bit more critic about the selection of literature found, but on the other hand it is not limited beforehand and hence might give results which would not have been obtained otherwise. The extended literature search included texts from both journals, articles and webpages. The results of the literature search are used to support the analyses and will also be present in chapters 4 *Background* and 5 *Development of the O-D tool*. As with the first step of the literature search, the criteria for the second step can be found in appendix A.

4 Background

This chapter will shortly explain some of the main themes of the study, in order to set a base for the study and examine some of the main aspects behind the development of a GPS-based O-D tool for urban and traffic planning.

The themes that will be described includes grid shape and size, GPS data and O-D analyses, which are relevant because they are the key features of the data used in the thesis. This will be followed by a short introduction to travel patterns and traffic models, where a few selected traffic models will be explained shortly.

4.1 Grids

According to Cambridge Dictionary a grid is *"a pattern or structure made from horizontal and vertical lines crossing each other to form squares"* or *"a pattern of squares with numbers or letters used to find places on a map"* (Cambridge Dictionary, u.d.) in geospatial contexts. However, grids are not necessarily squared; they can also be hexagons or equilateral triangles, as they can all be tessellated. This means that they can cover an area, without leaving any gaps, by repeatedly being placed edge to edge. However, each type of grids has each their assets and liabilities (esri, n.d. a).

Each type of grid has some advantages and disadvantages; triangular grids and hexagonal grids can hold more data than a squared grid (Zhai, et al., 2009; Strimas-Mackey, 2020). Both also work great on curved surfaces. Squared grids, on the other hand, are easier to work with, as a square grid is defined by the coordinates of the bottom-left corner of the grid, the cell size, and number of grid cells in each direction (Strimas-Mackey, 2020). In other words, it has a similar structure as the one for the Cartesian coordinate system (Zhai, et al., 2009).

For this study, squared grids were the one of choice, because of its uncomplicated processing. Furthermore, predefined squared grids with labelling that fits with data from Danmarks Statistik could be used. This makes it easier to add information about demographics at a later point, in order to compare the results of travel patterns with demographic aspects such as income or car ownership. These grid cells are 100 m x 100 m and cover entire Denmark.

The cell size was estimated to be fit for the purpose and the resolution of the data, because the size of the grid cells are not be too large, which would decrease the level of detail, nor are they too small, which ensure the processing time is not too long (esri, n.d. b).

4.2 GPS – Global Positioning System

The data used in this thesis is GPS (global positioning system) data, which is used to find a position on the Earth's surface.

By calculating the distance from at least four satellites, a position on the surface of the Earth can be found; the more satellites are in use, the more precise location can be found. For more on GPS data and satellites, refer to (Kyes, 2020; European Global Navigation Satellite Systems Agency, 2018; European Global Navigation Satellite Sytems Agency, 2020).

The GPS data used in this thesis will be presented in section 5.1 *Data description*, where both the accuracy and the temporal resolution will be mentioned.

4.3 O-D – Origin-Destination

Origin/Destination is a way of analysing patterns in an area over a period of time (SMATS, n.d.).

Through O-D analyses it is possible to analyse the movement in space from an originating point (O) to a destinating point (D). O-D pairs can be presented as geographic coordinates and contains data about a flow between two points or zones, but the routes in between the origin and destination are not included. These can be estimated using different models for this purpose. The O-D analyses are optimal for transport planning purposes (Lovelace & Leigh, n.d.).

Origin/Destination analyse has previously been a very time-consuming and expensive solution for traffic planning, but with new technology such as mobile devices with GPS tracking implemented, it is much easier and faster to collect big amounts of data for O-D analyses (Mehndiratta & Alvim, 2014). The same goes for the data used in this study (to be presented in section 5.1 *Data description*), because it is a large dataset of GPS points from all over Denmark for an entire year, which would not have been possible to collect through surveys or other types of data collection that has previously been used.

4.4 Travel patterns and behaviour

Travel patterns can be defined as multiple points having similar geometric properties, location or time stamp. They can be useful in examining an area, such as a city, to point out any areas of interest, which can either be derived from data, based on clusters of data points, or it can be defined by the user, though the latter is a rather subjective definition of interest areas (Guo, et al., 2012). Travel patterns can be seen as a result of the time-geographic place of cars.

4.1.1 Time geography

Spatio-temporal data includes information about both geography and the time of a dataset. The two types of data, though differing in dimensionality and direction, can be combined to show changes in georeferenced aspects over time. This could be anything from birds flying south for the winter and returning to the north during spring, to urban growth, climate change and much more. Spatio-temporal analyses has been practiced for many years, but it has only been since the 1990's that it has been made easier and better to store the spatio-temporal data, using technologies as GPS and GIS (Miller, 2017).

According to (Hägerstraand, 1970), the individual's movement is constrained by three main aggregation of constraints; capability constraints, coupling constraints, and authority constraints. Capability constraints describes the constraints of our needs, e.g. by the time we need to spend sleeping, eating, etc. which limits the individual in taking part in activities out of their reach. Coupling constraints deals with the time individuals spend together, where their paths meets in space and time, e.g. in school, at work, in traffic jams or in meetings (Hägerstraand, 1970). This can also be seen as clusters in traffic.

The final, the authority constraints, cover being limited in space. The limitations could either be for a short amount of time, or permanent. All three types of constraints interact and can be seen in everyday life (Hägerstraand, 1970).

The need for traffic management is rooted in the fact that all people since birth are located at some place at all time. The need for better traffic planning is a result of the total population of cars' need for infrastructure. Congestion can be seen as a result of a flow of life-paths that are distributed based on constraints and capabilities in time and space. By changing the constraints, it is possible to some extent to limit the number of population flow, which in this case study is the flow of cars (Hägerstraand, 1970).

4.4.2 Clusters

Simply put, clustering is a way of organizing data points into groups with similar data points and separate them from dissimilar data points, which can be put into other clusters. The similarity of data points can be based on a similarity in either concept or distance. E.g. will cars in a congested area be considered as a distance-based cluster (Anon., n.d.). In a time-geographic context, clusters can be seen as when multiple paths meet in time and space.

Clustering can be a great way of analysing a large number of O-D data, as two cars arriving at the same location, such as a station or airport, might not necessarily have the same GPS points. Accordingly, it can be helpful to group spatially close objects together in clusters, to find any potential meaningful places (Guo, et al., 2012).

4.5 Traffic models

In traffic and urban planning, models to predict the flow, accidents and patterns in traffic are a vital tool. It helps pointing out any potential areas that need change in the infrastructure. The range of traffic management models is wide, and should be chosen depending on the case, considering e.g.

the level of detail (Bellemans, et al., 2002). COWI's traffic planners often use either the Øresdunstrafikmodel (OTM) or the Landstrafikmodel (LTM), which also depends on the level of detail.

4.5.1 Landstrafikmodellen (LTM): a nation-wise transport model

Landstrafikmodellen (LTM) is a nation-wide traffic model that divides Denmark into 907 zones. It was created to improve the process of decision making for planning and investments in the infrastructure and can as an example be used to analyse how different factors such as demographics will influence the traffic. The LTM is suited for large scale traffic analyses and can also analyse the traffic to and from the surrounding countries, but it is not suited for smaller scale, where other models are needed – such as OTM for East Demark (DTU, 2017). Both LTM and OTM zones are smaller in urban areas, which ensures a more detailed overview and allows for more detailed analyses.

4.5.2 Ørestadstrafikmodellen (OTM): a local transport model

The Ørestadstrafikmodel, OTM, was created to help calculate the traffic demand on a weekday and in peak hours on a new stretch of city line in Copenhagen. Furthermore, it was created to calculate the flow in and out of an area distributed on different modes of transport, as well as to being able to estimate the consequences of amendments in negotiations about new rail lines or in capacity assessments (Skovgaard, 1996).

The model differs from previous traffic management models in the way that it is a four-step, passenger-based model; the model tries to figure out how a closed road, congestions, etc. might affect the choice of transport mode or the distribution of trips (Skovgaard, 1996).

OTM contains 4048 zones covering the eastern part of Denmark.

5 Development of the O-D tool

This chapter will go through the methods used throughout the thesis, in order to be able to answer the problem statement (refer to section 1.1 *Problem Statement*).

5.1 Data description

Firstly, before going into detail with the used methods, the data used for the study will be presented, as it might help understand some of the choices of methods. Furthermore, the data forms the base for this thesis, and accordingly, it is relevant to introduce it at this stage.

The data that this study is based on is GPS data collected from cars spread out over entire Denmark, throughout 2019, which can be seen in Figure 2. The data gets collected from personal vehicles only, and only some car brands from the Semler Group are represented. Though, it seems that the distribution of the car types of the dataset is quite similar to the one of all cars in Denmark, refer to Figure 3. By the end of 2019, the number of cars in the data set made up around 4 % of the total number of cars in Denmark.

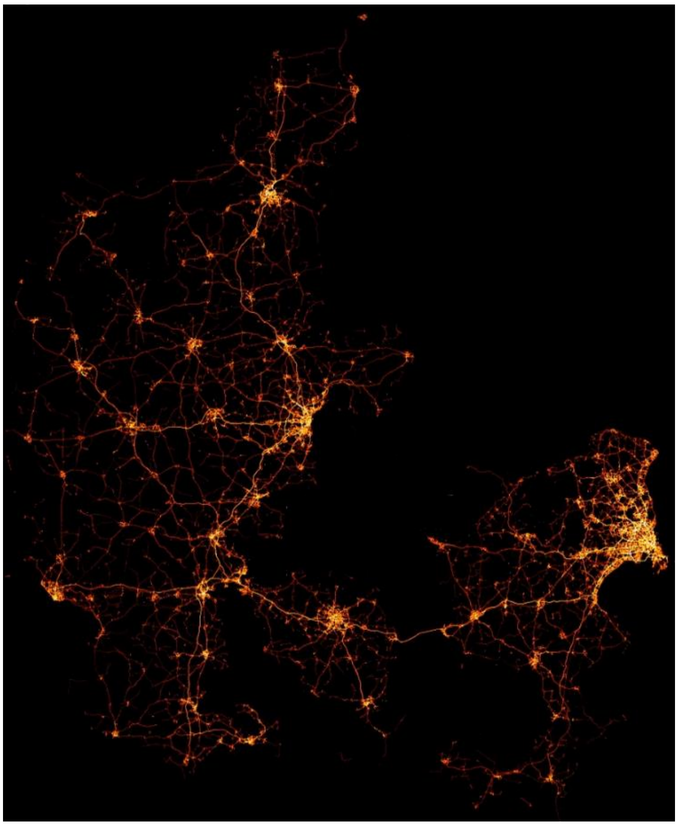


Figure 2: 15 minutes' worth of GPS data in the morning traffic is enough to light up most of Denmark's road network

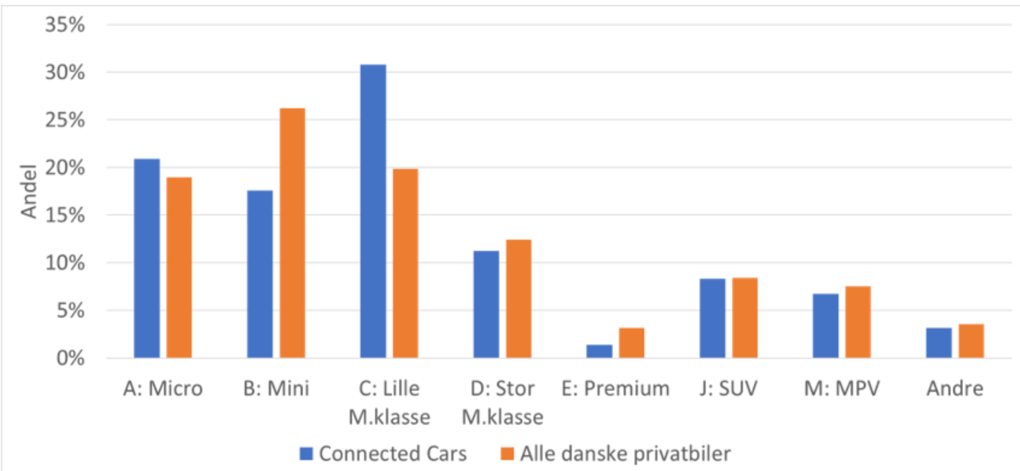


Figure 3: Distribution of types of cars in Denmark for all privately own cars and in the Connected Cars data set. Source: (COWI, 2019)

For cars from the Semler Group, newer than 2008, it is possible to get a device installed, that can connect to an app on a smart phone. Through the app, it is possible to keep track of mileage, fuel consumption and much more. It is also possible for the mechanic to call in for a check whenever there is an issue with the car (such as error code or alarms), and in that way tailor the service to each costumer (Connected Cars, n.d.). It is also in the app that the user accepts which data will be collected. There are different consent levels to accept, so not all possible data gets collected from all cars.

Even though they have accepted the collection of data, it is still crucial to make sure that the data processing and the results of the analyses carried out through this thesis, does not collide with GDPR. In order to make sure that all requirements are met, no data of a single trip or car will be shown, and the data will only be handled on an online server, so no data will be stored locally. These two actions will be described in more detail in sections 5.2.2.1 *Virtual Machines* and 5.2.5 *pgAdmin*.

The temporal resolution of the GPS data is 30 seconds, unless they turn, break or there is any other change in movement; then it will track the position more frequently. Though, for this thesis, only the start and end point (origin and destination) of every trip was needed. The total amount of individual trips for the 2019 dataset is 4,374,002. Figure 4 and Table 1 show the distribution of data per month.

Table 1: Distribution of GPS data spread over the months of 2019

Month	Trips	Percentage
January	240012	5%
Febru-ary	283798	6%
March	302558	7%
April	297034	7%
May	428583	10%
June	289780	7%
July	362884	8%
August	452540	10%
Sep-tember	465529	11%
October	554038	13%
No-vember	381321	9%
Decem-ber	315925	7%
Total	**4374002**	**100%**

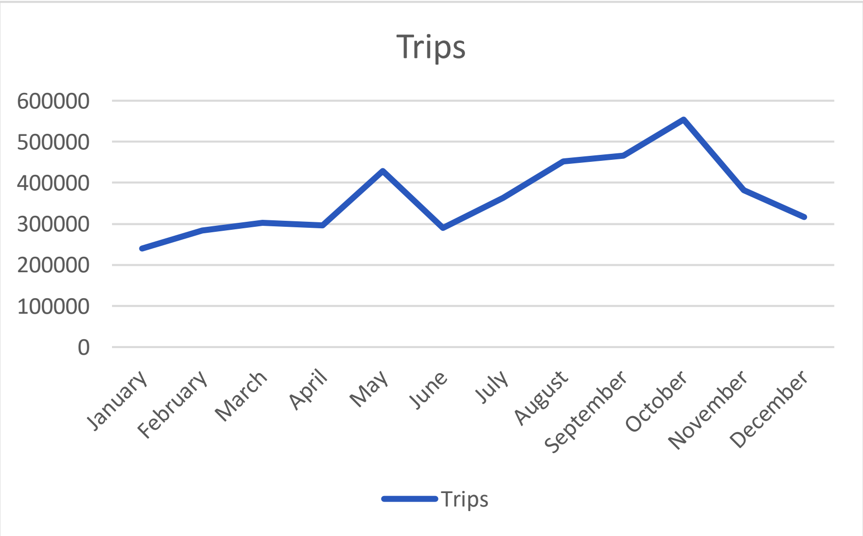


Figure 4: Graph showing the distribution of GPS data spread over the months of 2019

Besides having data about start and end point, the delivered data also contains information about the trip- and car id, the timestamps, the duration of the trip, and more. A full overview of the data can be seen in Table 2.

Table 2: Overview GPS data table; columns and data type

trip_id	**car_id**	**start_time**	**start_lon**	**start_lat**	**end_time**	**end_lon**	**end_lat**	**mil-age_km**	**drive_dura-tion_min**	**first_active_day**	**last_actie_day**
Var char	nu-meric	Timestamp without time zone	nu-meric	nu-meric	Timestamp without time zone	nu-meric	nu-meric	nu-meric	nu-meric	Timestamp without time zone	Time stamp without time zone

5.2 Software

In this section, the software used to carry out this thesis, will be described, as well as for what purpose it has been used. An overview of the software (and languages) can be seen in Figure 5.

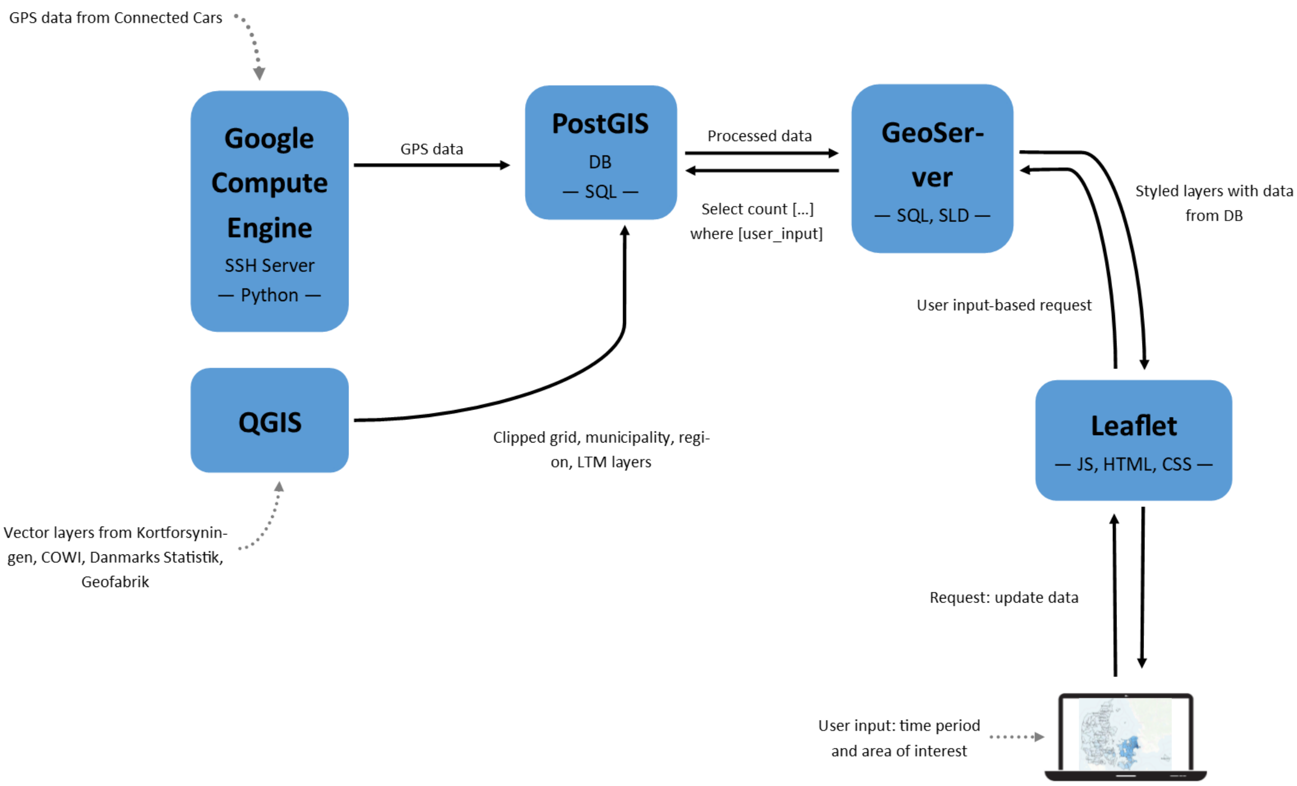


Figure 5: Overview of infrastructure of the used software and languages, and the GPS data's way through the system

5.2.1 Google Cloud Platform

Google offers a Cloud Platform, Google Cloud Platform (GCP), which is a package of the public cloud computing services that they offer. It allows the user to use the computing models to make virtual machines or resources alike (Rouse, 2019).

GCP contains both physical assets and virtual resources. Among the virtual resources are Virtual Machines, which will be explained in section 5.2.2.1 *Virtual Machines* below (Google, 2020). Among the services provided are Google Compute Engine, Google App Engine and Google Cloud Storage (Rouse, 2017).

The services can be reached either by using the Google Cloud Console, by command line interface, or by client libraries. The Console has a graphical user interface (GUI), from where projects can be created and open already existing projects (Google, 2020). As will be explained in more detail in the following section 5.2.2 *Google Compute Engine*, the command-line interface was the chosen method for the thesis; for extraction of the GPS data the Cloud Shell was used.

5.2.2 Google Compute Engine

As part of the package Compute is Google Compute Engine, which offers to work in Virtual Machines that run in Google's data centres. This gives the possibility of storing large amounts of data and a high-performance networking infrastructure. Some of the benefits of using Compute Engine are that there are different VMs which can be chosen amongst, depending on what the need and purpose is and that it can be integrated with the other Cloud services (Google, u.d.).

5.2.2.1 Virtual Machines

Virtual Machines (VM) consist of a host and guests; the host is the operating system on the user's own computer, and the guests are the operating systems running inside the VM. When working in a VM, it is possible to try out different software, apps and operating systems, without interfering with the software on your own computer, just by opening the VM program. This separation of operating systems is also called "*sandboxing*" as it keeps every process and app used in the VM here, without interfering with the computer (Hoffman, 2017).

Secure Shell (SSH) server

Within the Google Cloud Virtual Machine, there is a connection to an SSH server, which secure the data handling and exchanging between computers. Furthermore, it protects the privacy of identities, data and files (SSH.com, n.d. a), which is a plus when having to transfer data including positions, which is a form of personal data according to the GDPR Article 4, concerning Definitions (REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016). The SSH server can be connected through different clients, such as Chrome SSH Extension (SSH.com, n.d. b), which is the one used for this study.

Linux

The VM instance SSH connection opens a Linux instance. Linux is an operating system like Windows and Mac OS, which oversees managing the interaction between the hardware and software of a computer or mobile device (Linux.com, u.d.). Linux is used to retrieve the data and import it to the database, which was done in the console, as described above in section 5.2.1 *Google Cloud Platform*.

5.2.4 PostgreSQL

PostgreSQL is an open source object-relational database management system (ORDBMS), which handles and stores huge amounts of data, by using and extending the SQL language. PostgreSQL has many add-ons, such as PostGIS (Postgresql, n.d. a). PostgreSQL supports SQL and offers features such as complex queries and transactional integrity among other features, and it can be extended by adding new data types, functions or even languages (Postgresql, n.d. b).

5.2.5 pgAdmin

pgAdmin is an Open Source tool for working with PostgreSQL, as it has a user friendly Graphical User Interface (GUI) that simplifies the work with the database and allows even beginners to be able to interact with and get data from the database (pgAdmin, u.d.).

5.2.6 PostGIS

As mentioned above, PostGIS is a commonly used extension to PostgreSQL. It offers extra feature types, such as geometry, geography and raster to be loaded into the database, as well as functions, operators and index bindings (PostGIS.net, n.d. a; PostGIS.net, n.d. b). In this study it has been used to perform various spatial queries on the GPS data as well as vector layers, such as municipalities and LTM.

5.2.7 QGIS

The open-source Geographic Information System QGIS was used in the early stages of the development of the analytical tool. It is commonly used to visualize and analyse different data types with a geographic aspect to it. The most commonly used data types in GIS systems are vector and raster data, which suits different purposes (qgis.org, n.d. a).

In this thesis, vector data, in the form of GPS points, has been turned into raster data, by joining the GPS data with a grid, where they intersect. This causes the data to have a lower spatial resolution.

QGIS was used to clip the grid data to fit the boundaries of land, in order to decrease the processing time, by not processing empty grid cells. Furthermore, QGIS was used to join the grid layer with both the municipality and region boundaries from Kortforsyningen, to be able to extract data about the municipality of a grid cell, to see where the flow of car originates from. This will be explained more thoroughly in section 5.4.2 *Preparing grid layer* about preparation of data. Furthermore, it has been used to carry out additional analysis of the flow of cars in Copenhagen.

5.2.8 GeoServer

GeoServer is, similar to QGIS, an open source geospatial service that allows users to share data and create maps in a flexible way. GeoServer is Java-based and is built on a Java GIS tool, GeoTools. By using various extensions to GeoServer, it can generate many map and map functions, and can implement Web Map service (WMS) and Web Feature Service (WFS) standards (GeoServer.org, n.d. a).

GeoServer is used here to call the database and visualize the data on the analytical tool. When clicking an area or time period in the application, it calls the database again, which outputs the results for these parameters.

5.2.9 Visual Studio Code

Visual Studio Code is a source code editor that can be run from Windows, macOS and Linux. When installed, it comes with built-in support for JavaScript, TypeScript, and Node.js, but it can also support other languages such as Python or Java by using extensions for those languages. The software has a built-in Git command, which makes working with others an easy task (visualstudiocode.com, n.d.).

Visual Studio Code has been used to build the code for the analytical tool, using JavaScript, HTML and CSS. It was chosen because of its code support and as it is easily approached.

5.3 Languages

Just like section

5.2 Software described the software of this study, this section will shortly describe the programming languages used for the thesis, and for what purpose they were used. Figure 5 shows the connection between languages and software.

5.3.1 Python

Python is a programming language with high-level data structure. Because of its simple, yet effective approach, it is suited for beginners, as it is easy to learn. The programming language can be applied on most platforms to quickly develop applications and scripting, due to its elegant syntax (The Python Software Foundation, n.d.).

In the thesis, Python has been used to import the received GPS data into the PostgreSQL database. This was done in GCP's SSH server.

5.3.2 SQL

Structured Query Language (SQL) is a database language used to create, maintain and retrieve relational databases, such as Postgres. Relational databases consist of tables, in which data is stored in relations. Each table is made up of columns and rows, each of which can be manipulated through the use of SQL (MDN contributors, 2020 a).

In this thesis, SQL has been used to create a spatial join between the grid table and the table containing the GPS data, in order to get one table containing all the data of the two tables.

Furthermore, it has been used in a query calling the data in the database, to show on the map. When selecting an area, and/or time period, this is called by the use of SQL. An example of this could look like this:

```
SELECT count(*), start_kommune FROM test WHERE end_kommune='København' AND start_time BE-
TWEEN '2019-01-01 00:00:00' AND '2019-01-31 23:59:59' GROUP BY start_kommune
```

Code 1: SQL code example showing a way of extracting the count of where trips ending in Copenhagen during January originates.

5.3.3 JS

JavaScript (JS) is a light-weight compiled programming language that is often used for creating webpages, though it can be used for other purposes as well, due to its powerful scripting language. JavaScript runs on the client from where it is used to program how the webpage should react when interacting with it (MDN contributors, 2020 b). When JS is applied to an HTML document, it adds interactivity to the webpage, which could be used for games, animations, and much more) (MDN contributors, 2020 a).

JS is very user friendly, and can be fitted for many purposes, because of the wide variety of tools and libraries that can be added to the JS language (MDN contributors, 2020 b).

JS has been used to add interactivity to the analytical tool; update map output when an *update* button is clicked, select zone, etc. Furthermore, it is where the different layers are extracted from GeoJSON files and created as a vector layer in the web page.

5.3.3.1 Leaflet

Leaflet is an open source JavaScript library used to create interactive maps. A similar JS library exists, called OpenLayers. The main differences between the two are the possibilities and the ease of getting started; Leaflet is easier to get started with, as it does not require a lot to create a map, and simple functions can be done in a few lines of code. With OpenLayers, on the other hand, it takes more code, but it holds many more functionalities. Many of these functionalities can be used in Leaflet as well, using plugins. But if too many functions need a plugin, it is recommended to switch to using OpenLayers instead. OpenLayers also reads more data types than Leaflet, but the maintenance of this is less frequent than that of Leaflet, since it has more contributors (Tarasenko, 2019).

In this study Leaflet has been used based on previous experience; earlier on during my master's degree, I have been taught and used Leaflet. Though, based on dialogs with colleagues from COWI, it got to my attention that COWI mostly use OpenLayers, and for that reason it would have been an easier solution when handing over the tool. Though, as I did not have as much experience with working with this and time was limited, I decided to switch to using Leaflet, which was accepted by COWI as well.

5.3.4 HTML

Hyper Text Mark-up Language, HTML, is a language used to create a web page, as it is a language that a browser can read and translate to a webpage that we can read and interact with (HTML.dk, n.d.). A fundamental aspect of websites is the function that allows the user to click on a web page and get to another, as they are inter-linked (MDN contributors, 2020).

HTML was used to define the structures of the tool, in which the data get visualized.

5.3.5 CSS

Cascading Style Sheets, CSS is a language describing the styles of an HTML file and is what will determine how the elements are displayed in the webpage (w3schhols, n.d.). In this study it is used to set the height and width of the map shown in the tool.

5.3.6 SLD

Styled Layer Descriptor, SLD, is an XML-based mark-up language used to style geographic data in GeoServer. It is a very powerful, but complex language (GeoServer, n.d. b). In this thesis SLD is used to create a style to apply to each of the layers in GeoServer. It is where the colour scale of the layers shown on the map is defined.

5.4 Implementation

Before getting into detail with the actual analyses, the data had to be prepared and loaded into the database. This section will go through the preliminary processing of the data.

5.4.1 Indexing

Because of the huge amount of data used for this thesis, there is a need for indexing. Indexing is a way of structuring data in a database that helps optimizing the processing time and performance of a database (tutorialspoint, n.d.).

The indexing was done in PostgreSQL. Different types of index performances can be done, depending on the type of query it is to be used for. The CREATE INDEX command creates a B-tree index. This is the one used for this thesis, because it is suitable for most common situations. Furthermore, indexes can either be single-column, multicolumn, or unique indexes. Here, multicolumn index was the approach of choice, because it was concluded that two columns where to be used most frequently in where clauses, and in order to address this the best way, a multicolumn index is the better choice (tutorialspoint, n.d.).

5.4.2 Preparing grid layer

In order to keep the GPS data anonymized and decrease the processing time at later steps of the analysis, data was aggregated into grids of 100 m x 100 m. As mentioned in section 4.1 *Grids*, the grid for this thesis is a 100m x 100m grid named as the one used in *Danmarks Statistik*. This grid layer was clipped to fit the coastline, and later joined with the municipality and region layers from Kortforsyningen (Styrelsen for Dataforsyning og Effektivisering, n.d.), which was done using QGIS. For the LTM layer that was also added in the analytical tool, the layers were later joined using PostGIS.

When this was done, the data was loaded into the database, which could be done in one of three ways; the first possible way to add data to the database, is to use QGIS' database manager tool (qgis.org, n.d. b). The other options are to use shp2pgsql or Postgres' COPY function, which works better for large datasets. Working with SSL-certificates might cause permission errors, in which case shp2pgsql, which is part of the Postgres package, can be used to convert .shp files to .sql files (Packt, n.d.). The latter was the chosen process that was done, because a huge amount of data was to be imported. After converting the data to a .sql file, it was loaded into the database, using psql's \copy function, after creating a table to copy the data into. For it to work, the SSL certificates had to be provided alongside with the database name, the user, the port and the host.

5.4.3 Retrieval of GPS data

The data was delivered as .gz.tar files in a Google Cloud environment. In order to be able to work with the data, it was loaded into the database, which was done from the SSH server through the Google Cloud's Virtual Instance.

In an SSH screen, the data was then cloned to another directory in order to not interfere with the original data. Here, it was extracted and loaded into the database. This was done using a python query that establishes a connection to the database including the required SSL certificated in order to get access to the DB.

From the SSH screen, a new table was created in the DB, and the GPS data was loaded into it. The total amount of 12 months' worth of GPS data summed up to 4,374,002 trips.

5.4.4 Spatial join of grids and GPS data

With the grid and the GPS data loaded into the database and indexed, they could then be joined. The output table needed to include the name of the starting and ending zone, e.g. the start and municipality, and a start and end time, as well as the geometry for the start and end zones.

First, the data was stored in two separate tables (one for the starting point, one for the ending point) and was then joined with the grid table. The trip IDs were used to make a full outer join of these two tables, to make one table, where both the start and end municipality and region was included. Though, the trip_id column was not included in the joined table, in order to ensure privacy, as trip IDs can be linked to individuals.

```
CREATE TABLE test AS
SELECT intersects_start.start_time, intersects_start.geom,  intersects_start.kommune AS start_kommune, intersects_start.region AS start_region, intersects_start.kn100mdk AS start_grid, intersects_end.geometry, intersects_end.kommune AS end_kommune, intersects_end.region AS end_region, intersects_end.kn100mdk AS end_grid
FROM intersects_start
FULL OUTER JOIN intersects_end
ON intersects_start.trip_id = intersects_end.trip_id
```

Code 2: SQL code creating the joined table

As can be seen in Code 2, the full outer join was the selected join type. This was in order to get all the data from both the table for start point and for end point. The data in common, which was needed in order to do the join, was the trip ID, as can be seen in the bottom of the code above.

There might be a quicker or more efficient way of getting the same output table, but it was not prioritized to spend more time on figuring out a way of doing so, when a table with the wanted parameters was already created. The same process was later done with the LTM layer, which was cut to fit the boundaries of Copenhagen and Frederiksberg municipalities in order to limit the processing time and focussing the analysis to Copenhagen and Frederiksberg only.

Each of the two data tables contains information about the zone in which a trip originates and ends and at what time. This is what is needed for the next step, that is to visualise the aggregated data on a map.

5.4.4 Calling the database

After the data tables were joined, it was read to be loaded onto the map. This was done with the help of GeoServer.

First thing was to connect to the DB by creating a data store, including the needed information in order to find the DB, and connect to it. Then, a style was created using SLD. This sets the colour scales for the map view with breakpoints. A snippet of the style code can be seen below, where the break points are defined and assigned a colour.

```
  […]
<Rule>
  <Title>&lt; 1</Title>
  <ogc:Filter>
    <ogc:PropertyIsLessThan>
     <ogc:PropertyName>tripcount</ogc:PropertyName>
     <ogc:Literal>1</ogc:Literal>
    </ogc:PropertyIsLessThan>
  </ogc:Filter>
  <PolygonSymbolizer>
     <Fill>
        <!-- CssParameters allowed are fill (the color) and fill-opacity -->
        <CssParameter name="fill">#ffffff</CssParameter>
        <CssParameter name="fill-opacity">0.0</CssParameter>
     </Fill>
  </PolygonSymbolizer>
</Rule>
<Rule>
  <Title>1 - 10</Title>
  <ogc:Filter>
    <ogc:PropertyIsBetween>
      <ogc:PropertyName>tripcount</ogc:PropertyName>
      <ogc:LowerBoundary>
        <ogc:Literal>1</ogc:Literal>
      </ogc:LowerBoundary>
      <ogc:UpperBoundary>
        <ogc:Literal>10</ogc:Literal>
      </ogc:UpperBoundary>
    </ogc:PropertyIsBetween>
  </ogc:Filter>
  <PolygonSymbolizer>
     <Fill>
        <!-- CssParameters allowed are fill (the color) and fill-opacity -->
        <CssParameter name="fill">#b9d7ed</CssParameter>
        <CssParameter name="fill-opacity">0.7</CssParameter>
     </Fill>
  </PolygonSymbolizer>
</Rule>
[…]
```

Code 3: Code snippet of the xml code defining the breakpoints for the style for the layers in GeoServer.

Now, when the connection to the database and the styles for the layer ready, the layer itself could be created. This was done using an SQL statement that created a temporary table. This table takes the start_zone, end_zone, start_time and end_time into consideration, as well as the count of data points in each start_zone ending in each end_zone. The zone can be any wanted zone, such as municipality, region, district, or other.

This layer can then be published and included in a Leaflet (or any other) map as a WMS layer, by inserting the link to the WMS and define which layer to show on the map.

```
var wmsLayer = L.tileLayer.wms('http://localhost:8080/geoserver/test/wms?', {
  layers: 'kommune_end',
   opacity: 0.75
}).addTo(mymap);

$('#startkom').change(UpdateViewParams);
$('#starttime').change(UpdateViewParams);
$('#endtime').change(UpdateViewParams);

function UpdateViewParams() {
```

```
  selectedParams = 'startkom:' + $('#startkom').val() + ';start-
time:' + $('#starttime').val() + ';endtime:' + $('#endtime').val();
  wmsLayer.setParams({ viewparams: selectedParams }, false);
}

UpdateViewParams();
```

Code 4: Snippet of JS code calling the WMS layer from the GeoServer and adding update *function to the layer.*

In the map application, a function giving the possibility to select start zone and the time interval is presented. When selecting the zone of interest and time period, an *update* function in the map layer sends a new request back to the GeoServer, where the results of the newly selected zone (and/or time period) will be sent back to the layer and visualized on the map. The request sent to the DB is the one defined in the layer segment in the GeoServer and is similar to the query described in section 5.3.2 *SQL*.

The HTML/JS files in full extent can be seen in appendix B. In appendix C an example of the SQL and SDL query to create the layer and style for the layer can be seen.

6 Integration of the O-D tool in traffic planning

This chapter will examine the results of the tool, showing some of the map outputs that can be extracted from the tool. Furthermore, an example of a use case based on these results will be presented, in which Copenhagen and Frederiksberg municipalities will be examined in more detailed.

The exploratory analytical tool is developed with a focus of being applicable to multiple purposes, by changing a few parameters. This could be the zones, in which the data is to be aggregated of the type of data. The following case study will examine the spatio-temporal patterns seen from GPS data, with a focus on finding any possible congested areas, to which planning of new infrastructure could be used. This is one way of using the GPS O-D data that was provided, and one way of using the analytical tool to explore patterns. Other use cases will be discussed in section 7.3.3 *Other use cases*.

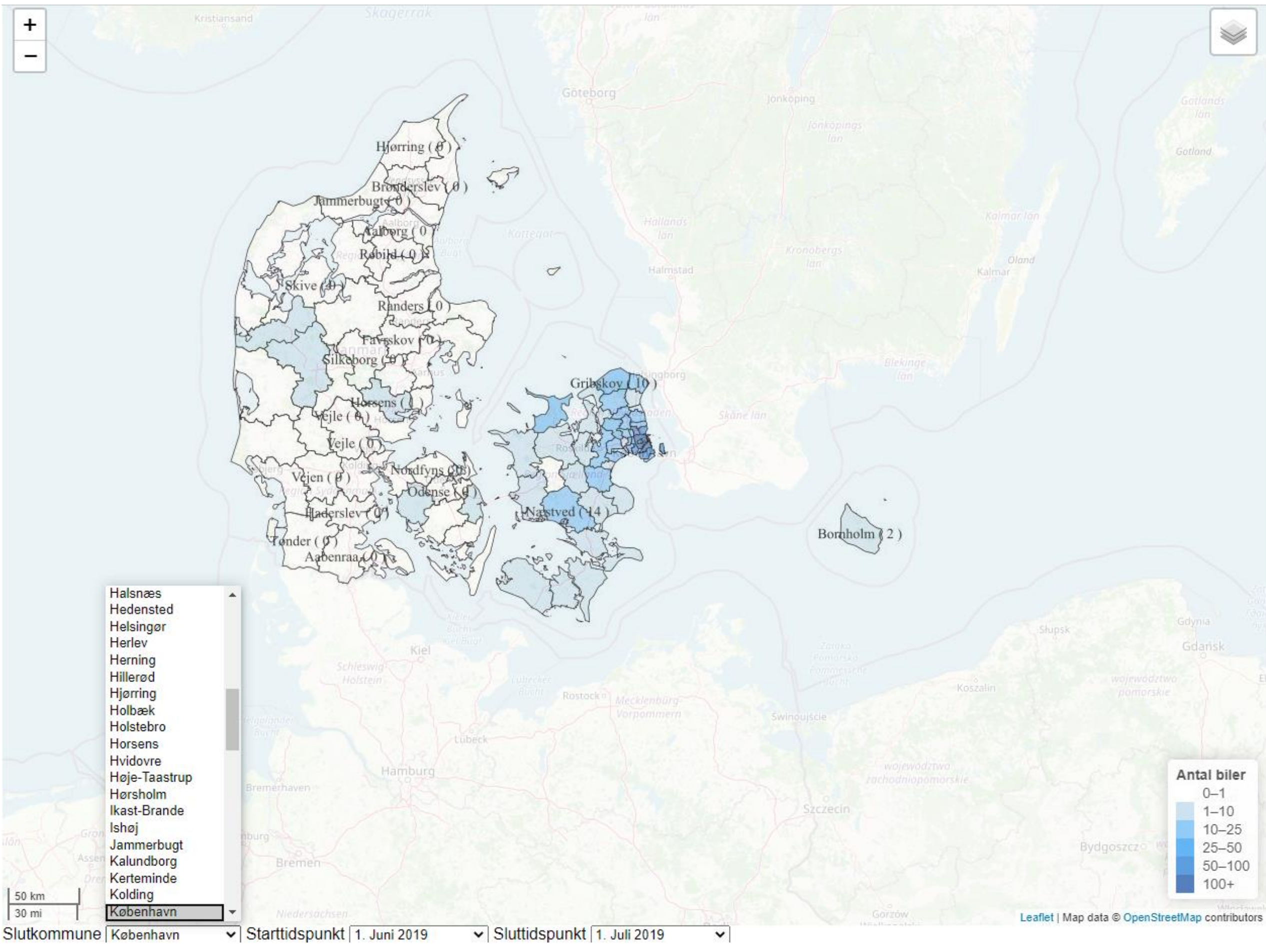


Figure 6: The exploratory analytical tool and its basic functionalities

The analytical tool and its basic functions can be seen in Figure 6. It is created to be a tool that is easy for urban and traffic planners to use in the process of a project, because the data is already pre-processed, and can be used again for similar analyses. For this study, the tool contains two levels of detail, in which the planner can examine an area; municipality and LTM, which both covers entire

Denmark. There are 98 municipalities covering Denmark, and 907 LTM zones covering Denmark. More zones could be pre-processed in the application, to make it even more flexible to more purposes, though for this dissertation it was not prioritized, as it was considered that the results carried out in the following, would not be of significant change, if other zones had been made. A full discussion of this can be seen in chapter 7 *Significance of the exploratory analytics tool*.

The focus for this thesis is Copenhagen municipality; both in municipality level to see how many incoming cars there are distributed over time, and on LTM level, in order to get a more detailed view of traffic flow patterns. The reason why Copenhagen is chosen, is because it is the highest populated municipality in Denmark (Danmarks Statistik, 2020 a), but it is also the municipality with the highest car ownership per person (Danmarks Statistik, 2020 b). Furthermore, it has the highest number of work places (Danmarks Statistik, 2020 c), which could either be reached by citizens of Copenhagen or from other municipalities. Accordingly, it was estimated that by focussing on Copenhagen, there should be more data to base the analysis on.

6.1 Spatio-temporal pattern of cars

Analysing the spatio-temporal patterns of traffic flow in the urban areas, the morning and afternoon peak hours of weekdays are often of interest, as these time periods are when the most cars are on the road, going to or from work, respectively. This is done by extracting the hour and the day of the week (DOW) parts of the start time of each trip in the SQL, as can be seen in Code 5:

```
select start_kommune, count(*) as tripcount
from lokn.test
where end_kommune = '%endkom%' and
start_time between '%starttime%' and '%endtime%' and extract(iso-
dow from start_time) < 5 and extract(hour from start_time) between 7 and 9
```

Code 5: SQL defining a WMS layer that gets called from the JS code as seen in Code 4, and updates based on user inputs for selected end_kommune, start_time *and* end_time.

end_kommune can be chosen from the analytical tool as well as the months of aggregated data that is of interest. The more months included in the analysis, the more data will support the choices made based on it. The area of interest could also be LTM zones, as that is the other zone layer that has been integrated in the map application. The different time slots over a day, such as morning or afternoon, is defined in different layers, and can be applied if wanted.

Fridays are excluded from these aggregations of data, because it often is excluded in traffic planning because the travel patterns often have been found to be different from the rest of the weekdays, either because people leave work earlier, go in to office later, commute longer in the afternoon, because they are visiting family, or for other reasons.

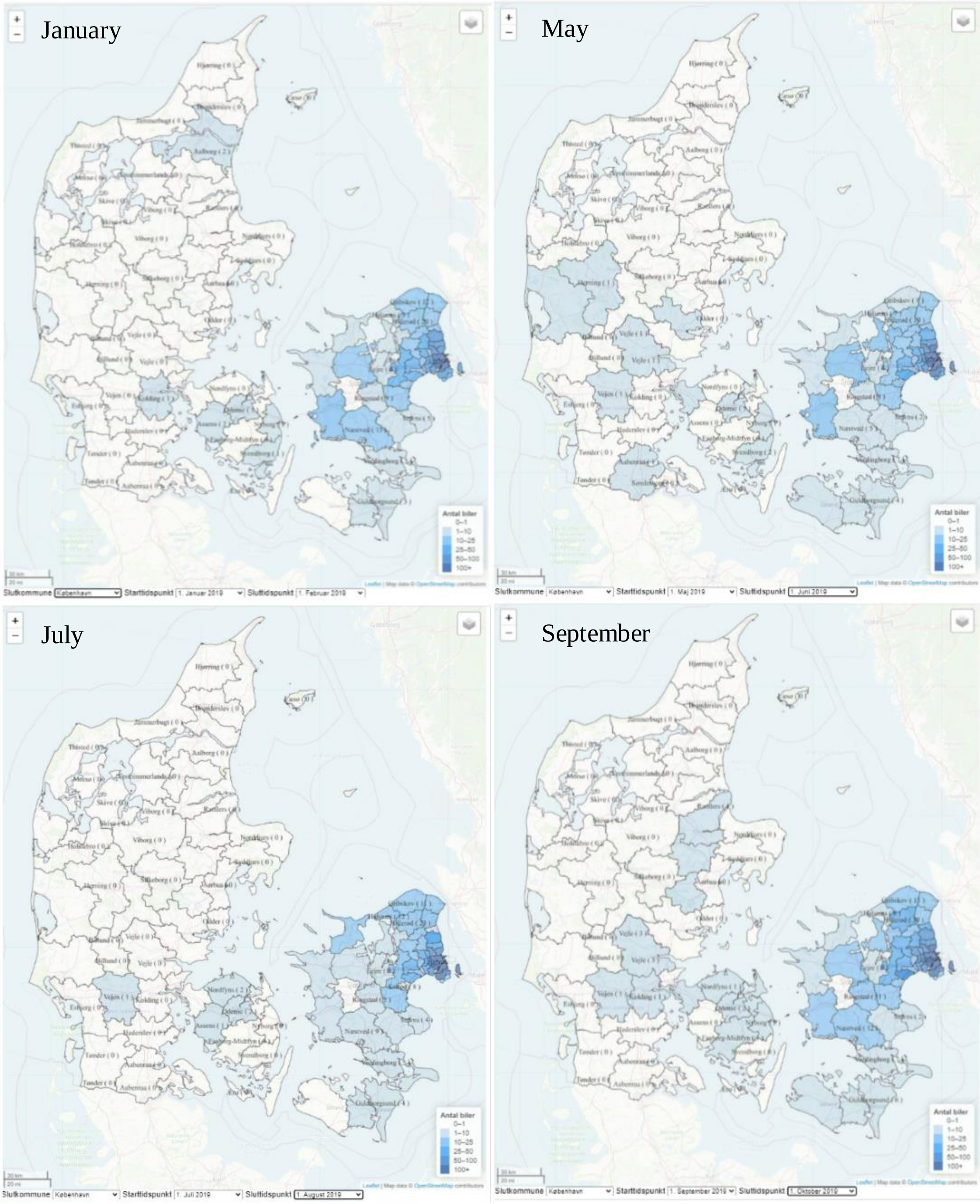


Figure 7: Seasonal changes in trips' origin during morning peak hours; January, May, July, and September.

Examining the GPS data that was provided for this study, within four different months, seasonal differences can be seen. Figure 7 shows how the distribution of trips to Copenhagen differ between the months January, May, July and September in the morning. In July, it seems that less cars enter Copenhagen in general, and less cars cross the Great Belt to get to Copenhagen compared to the other months. This difference is likely a result of the summer holidays, where less people will commute to work in the morning. Likewise, in January fewer cars comes from Jutland. The reason for that is not as clear as for July, it and would need further investigation of the patterns to understand the reason for that, but that will not be addressed in this study.

Examining the data for a month alone, a general trend can bee see no matter which municipality is selected; the further away another municipality is, the smaller the probability of cars coming from there is, and the closer it is, the more cars has a destination in the focus area.

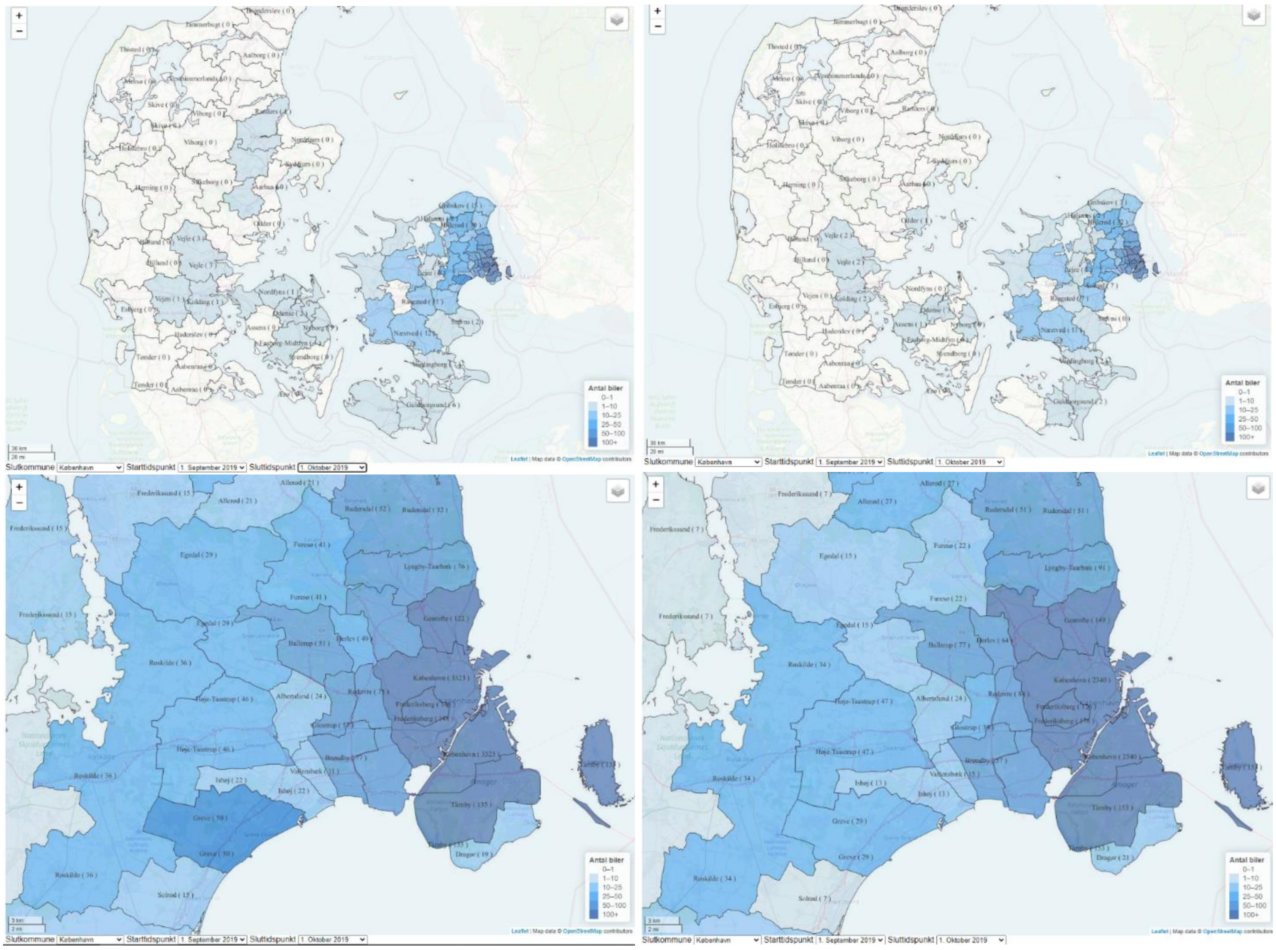


Figure 8: Distribution of originating municipalities with Copenhagen as destination, different zoom levels. Left: morning traffic (07:00-09:00), right: afternoon traffic (15:00-17:00) on weekdays (Friday excluded)

When comparing the incoming car flow to Copenhagen in the morning peak hours (07:00-09:00) and the afternoon peak hours (15:00-17:00) during September last year, it seems that cars are coming in from further away in the morning than in the afternoon of weekdays (Fridays not included), which can be seen in Figure 8. If we zoom in a bit more, it can be seen that the municipalities from where the cars are coming is more scattered in the morning, whereas the traffic seems to be more concentrated from some municipalities in the afternoon. Especially from north of Copenhagen, there seems to be coming a lot of cars in the afternoon, which might cause some congestion along Lyngbyvejen and Helsingørmotorvejen. When many cars are going into Copenhagen in a short time period, and many cars are driving internally in Copenhagen, this might cause a lot of congestion on the roads. Additionally, there are also the cars going to Frederiksberg, Tårnby and Dragør, who have to go through Copenhagen as well. This adds up to quite a lot of cars. Though, the total number of cars with Copenhagen municipality as destination is approximately the same in the morning and the afternoon, but the distribution of cars is different.

It seems that these patterns are a sign of a lot of people from most of the eastern part of Zealand is working in Copenhagen, but there is also a lot of people from Copenhagen who work outside of the municipality. However, it can also be seen that there are a lot of internal rides in Copenhagen municipality, which has the highest number of cars with Copenhagen as the destination. It makes sense when looking at the overall pattern of origins, where it generally can be seen that more cars are originating in the adjacent municipalities than the ones further away. But some of these trips maybe could have been done by bike or public transport, since the distances internally are rather short. If more people are commuting by car internally in the municipality, it will only worsen the situation of congestion. In order to get a better idea of the internal traffic, it could be relevant to look at the municipality in a smaller scale. For this, the LTM zones, among others, could be used. For this purpose, the LTM zones lying within the boundaries of Copenhagen and Frederiksberg municipalities have been cut out, to focus on the commuting trips inside the municipalities. Frederiksberg is included because it is surrounded by Copenhagen.

In Figure 9 we can see the internal O-D pairs of Copenhagen and Frederiksberg municipalities in more detail. The OTM zones make it possible to get an idea of where most cars are going at specific time periods. When comparing the two data extracts, a distinction between the morning and the afternoon traffic flow can be seen; in the morning, more cars destinates at Langelinje and around Refshaleøen. In the afternoon, there seem to be clusters around Frederikskaj, and Vesterbro in general contains many destinations of trips, in the afternoon as well as in the morning. Also, along the metro line to Vestamager, a lot of cars end their trip, especially in the middle, which is likely due to the shopping mall *Fields*. Frederiksberg seem to be the destination of rather many of the trips, but more are entering the area in the afternoon than in the morning. This could indicate a slightly high frequency of residents in Frederiksberg who works in Copenhagen or Frederiksberg use their car to get to work.

6.2 Usage of GPS based O-D tool in traffic planning

In section 6.1 *Spatio-temporal pattern of cars*, some examples of the analyses and aggregated data that can be extracted from the tool, is presented, where various patterns in time and space (and scale) can be distinguished. But in order to make sense of the results, they must be put into a context. For urban and traffic planners, it is of great relevance to analyse such patterns, in order to plan the city and its road segments in the best way.

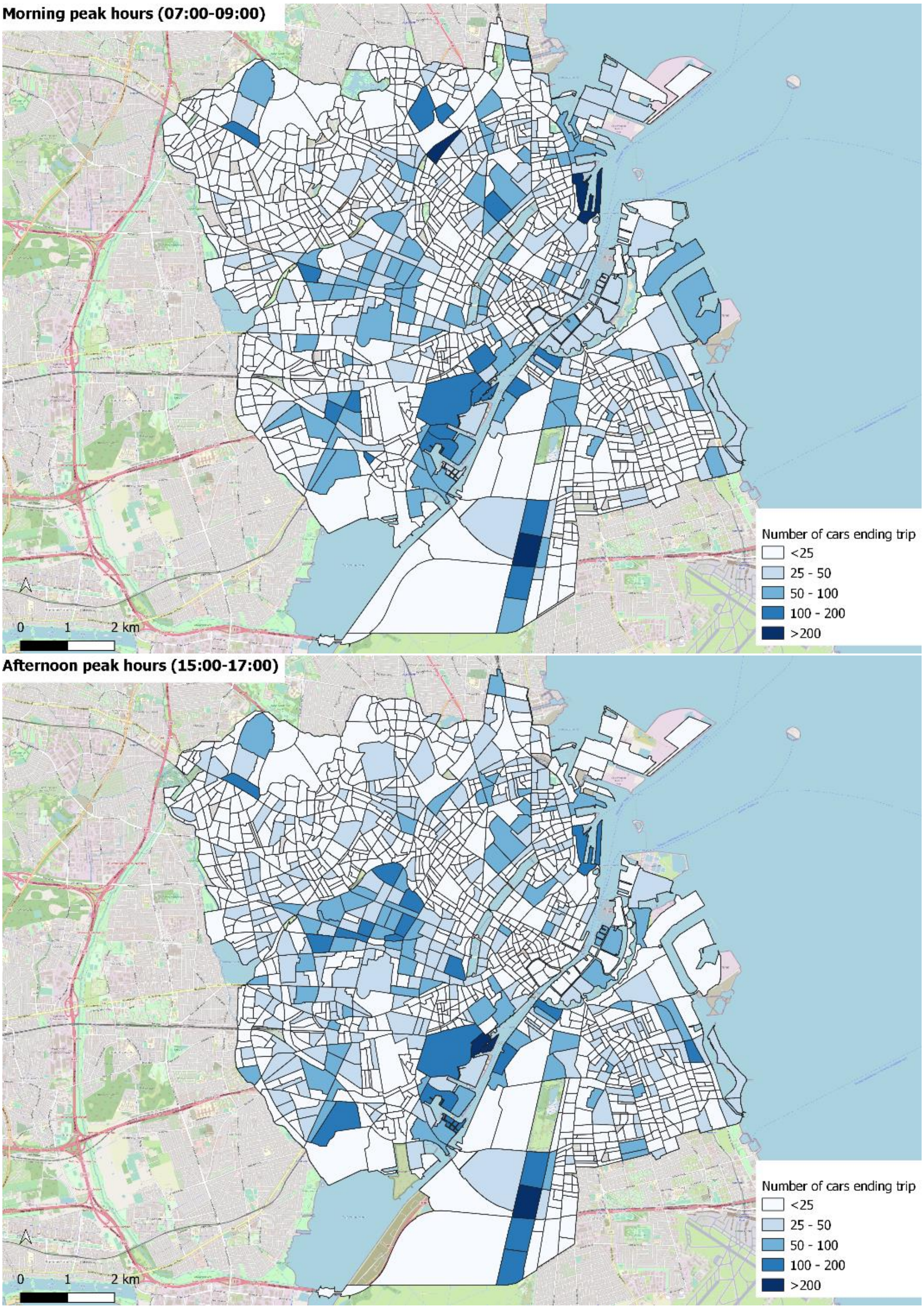


Figure 9: Destinations of trips internally in Copenhagen and Fredeiksberg on OTM level

There are many different purposes to where this kind of exploratory tool could be relevant. Amongst these are finding service areas for households, or the other way around, finding the areas of costumers in order to focus branding and commercials, where to elaborate on the public transportation, based on where many cars seem to be commuting and there is a *gap* in the existing public transport network, estimation of carpooling possibilities, and much more. This case study focuses on an analysis of the possibilities of reducing the congestion of cars in Copenhagen, mainly with the focus on commuting cars internally in Copenhagen

The analytical tool can be used alone or combined with other methods or tools, for a more in-depth understanding of the case. In order to examine any coherence between the infrastructure and the most common places of destination in the two municipalities, QGIS was used. By means of the OTM zone layers, the distribution of cars over time and space in Copenhagen and Frederiksberg municipalities can be viewed in more detail, which will hopefully result in a better understanding of where most cars are going to or from.

As seen in section 6.1 *Spatio-temporal pattern of cars*, many cars are entering Copenhagen, both in the morning and the afternoon. In Figure 9, the distribution of cars on OTM level can be seen, which provides a better level of detail.

It might be difficult to get drivers onto other types of transportation, but some of the solution suggested here, might affect the number of cars from other municipalities as well. The reasons why so many people choose to go by car internally in the municipality is difficult to know from a

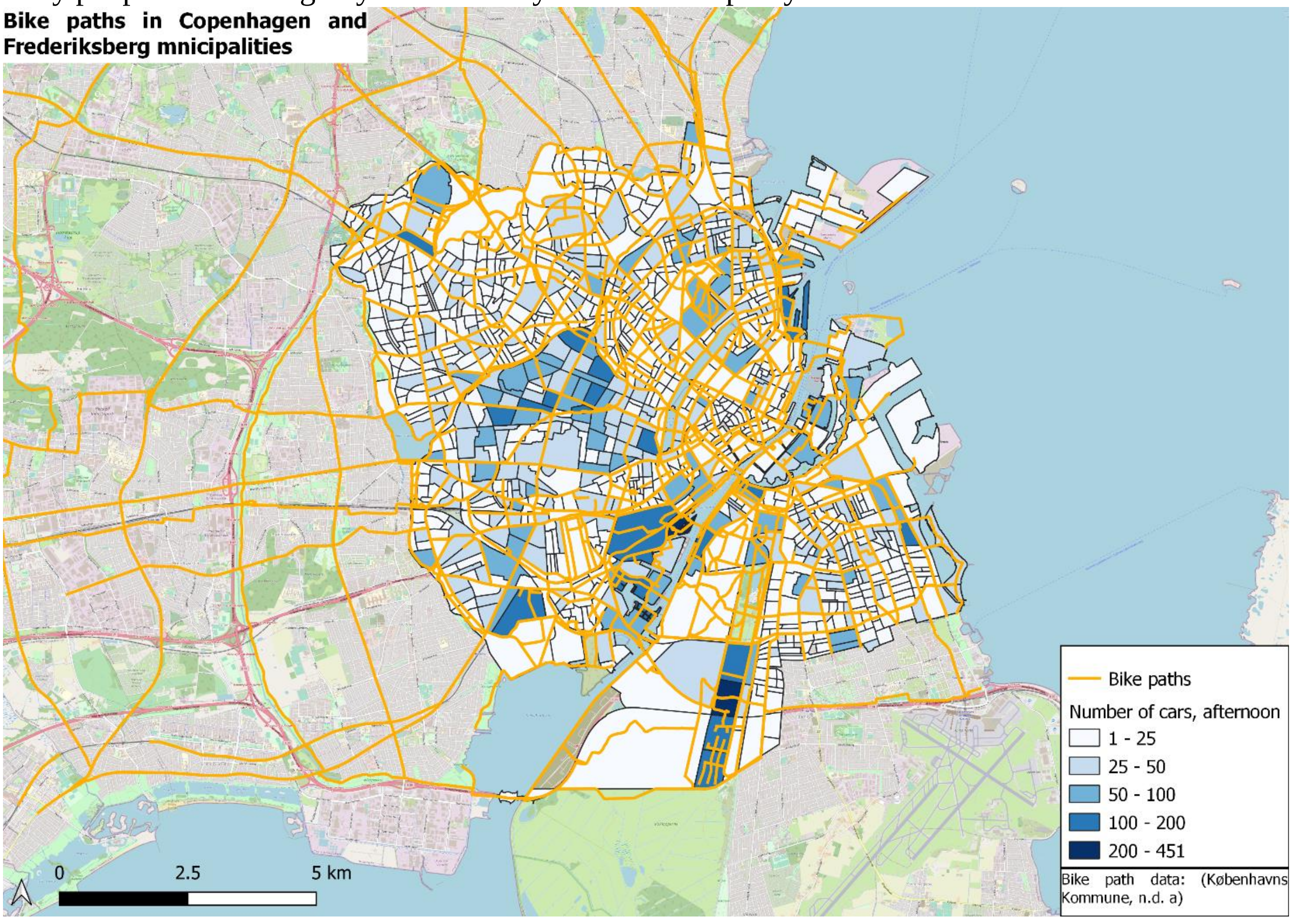


Figure 10: Bike lanes in Copenhagen and Frederiksberg municipalities, shown on top of the distribution of cars in the afternoon

quantitative analysis like this; it would require interviewing the commuters to get a better understanding of why they do not choose either public transport or the bike instead. This will not be addressed in this study due to time resources but including a qualitative aspect to the study would improve the understanding of the patterns that can be detected using the exploratory tool. This could be very helpful for planners to adjust their solutions for the needs and preferences of the commuters, in order to make sure the solution they choose will make the impact they wish to.

However, from the GPS data, the originating and destinating clusters could be compared to e.g. train stations and bike paths' locations throughout the municipalities. An overview of relatively the bike paths and the public transport network can be seen on top of the OTM layer in Figure 10 and Figure 11. There seem to be no clear compliance between a high number of cars destinating in an area and any lack of bike paths. Though, there might be other reasons to not choose the bike, even though the infrastructure is rather good.

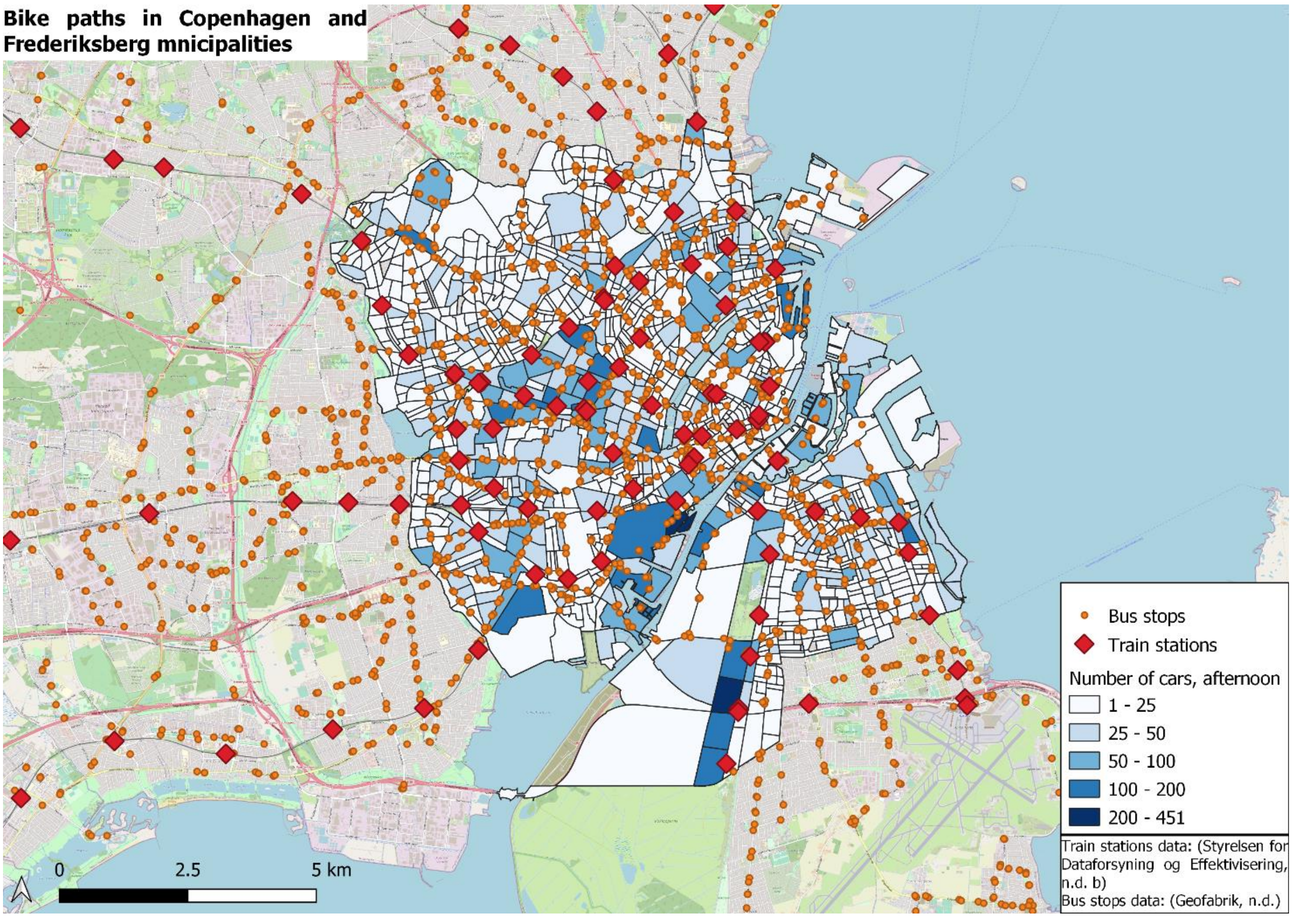


Figure 11: Train stations and bus stops in Copenhagen and Frederiksberg. Source: (Styrelsen for Dataforsyning og Effektivisering, n.d. b; Geofrabrik, n.d.)

Using public transport is another option to replace the cars. A similar analysis has been done with the train stations in Copenhagen. Here, it seems that the train stations are placed along the lines according to the *finger plan*, but in between there is a bit longer. Despite that, it seems that there are more destinations by car in close proximity to the stations, than other areas where the distance to the stations are longer, which especially can be seen in the Bispebjerg/Brønshøj-Husum area. In addition to the train stations, the bus stops in the two municipalities can also be seen in Figure 11. These are spread out even more regularly over the municipalities, and there are not that long distances from any of the bus stops to the places where most cars seem to terminate.

It seems that Copenhagen and Frederiksberg are both covered adequately when it comes to public transportation possibilities, and that this does not seem to be the reason for that many people to choose the car over other modes of transportation. A study carried out by the Alexandra Institute for Movia has analysed the mobility practices of commuters, which is the underlying reason of choosing one mode of transport over another. The analysis shows, that the main considerations that come into play are the individual's perception and ideas, competences and abilities, and physical aspects, products and technology. Additionally, aspects such as freedom, economy and our mental map plays a role in the choice of transport mode (Trafikselskabet Movia, 2017). This is summed up in Figure 12.

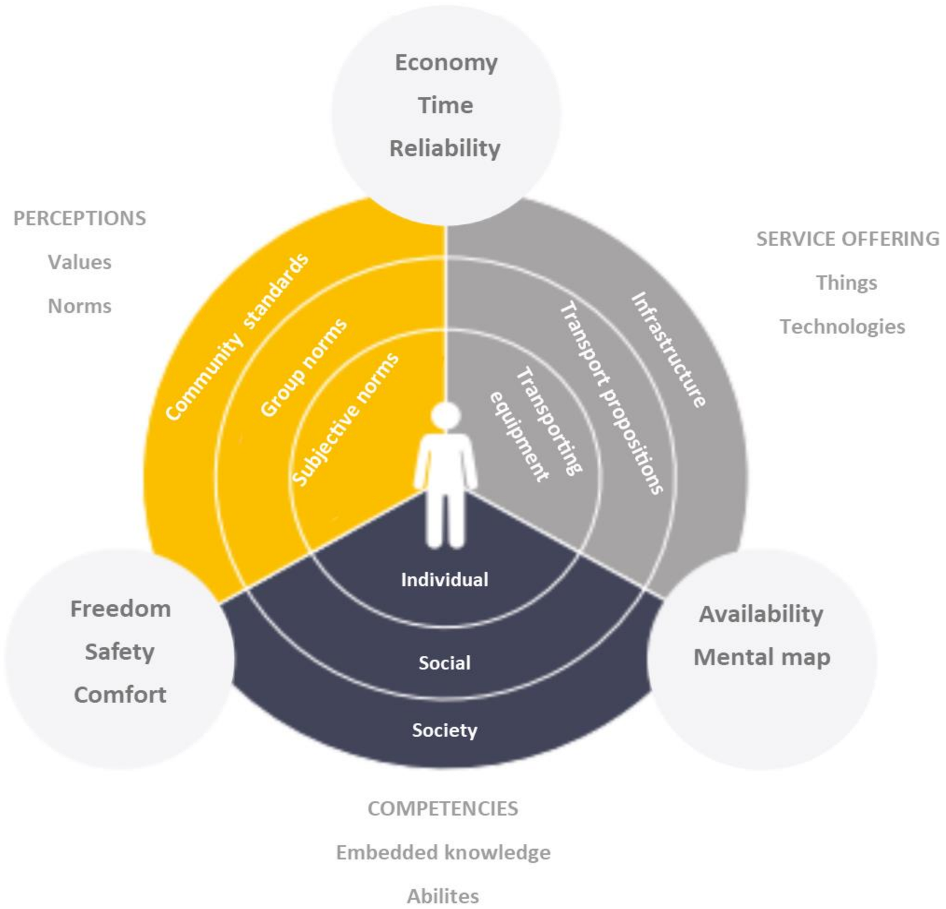


Figure 12: Considerations that play a role in choice of transport mode. The figure is a translated version of the original, Danish figure found in (Trafikselskabet Movia, 2017).

The survey concludes that mainly the individual's mental map, being on time and economy are important factors in the decision-making of mode of transport. This supports the solution to increase prices for parking, decreasing the amount of parking spaces or introducing road-pricing. However, it is concluded in the report that economy has an influence on the choice of transport, but for car owners, the effect seems to be smaller compared to cyclists or users of public transport (Trafikselskabet Movia, 2017).

When the infrastructure seems to be adequate, it seems that it is the habits of the commuters that needs to be changed. This could be done by nudging or otherwise by political actions. Some of these could be higher pricing on the parking spaces or less possibilities for parking in the city. As well as for the bike paths and train stations, the parking zones has been compared to where the clusters of O-D pairs are located, to see if the possibilities of parking match the distribution of cars.

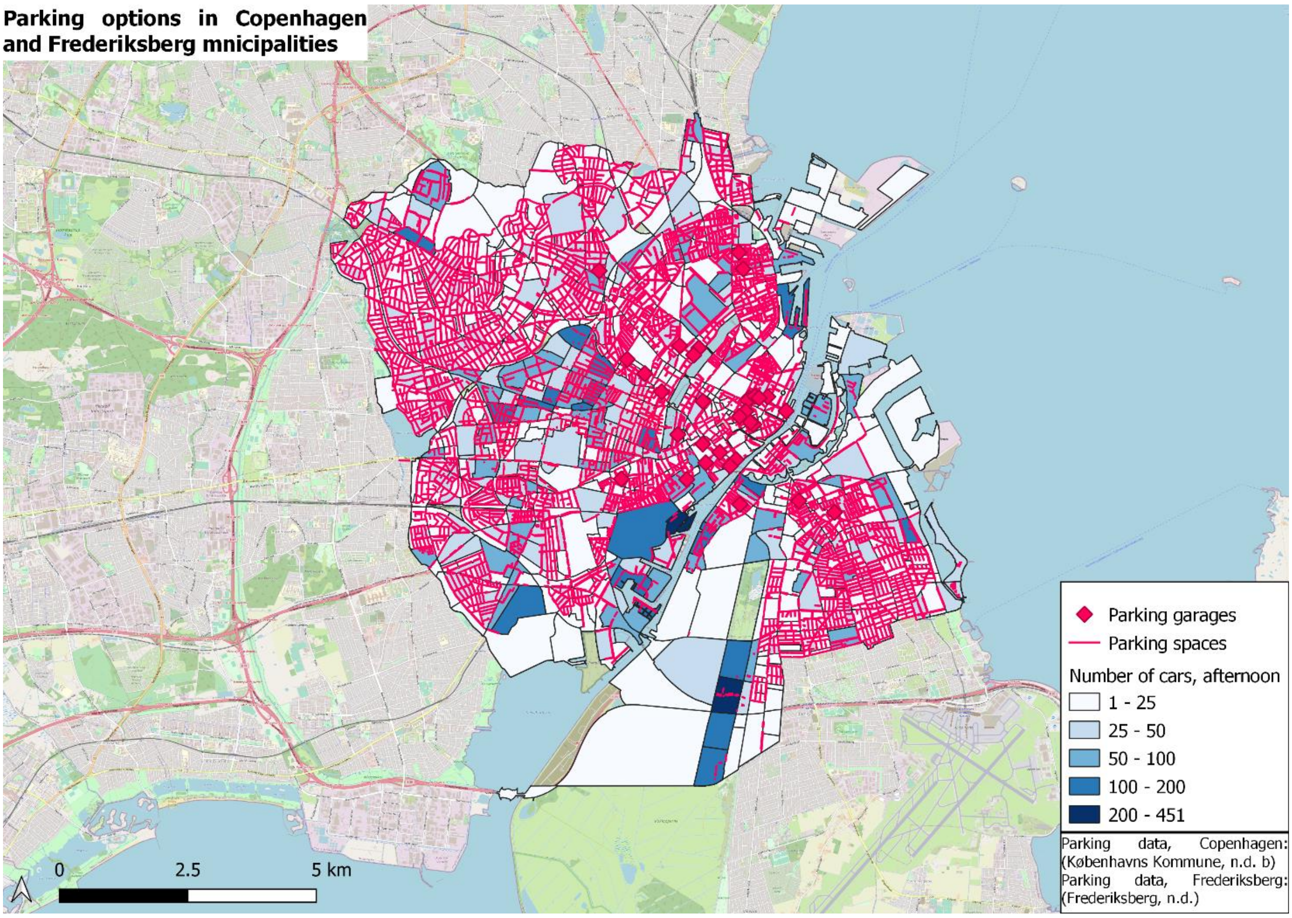


Figure 13: Parking spaces and garages in Copenhagen and Frederiksberg. Source: (Kommune, n.d.; Kommune, n.d.)

In Figure 13 it can be seen that there are quite a lot of parking zones in Copenhagen, and it seems that there are great options for parking, which might encourage more people to go by car. A solution that could help decrease the number of cars, would be to increase the parking taxes of public parking spaces or introduce road-pricing, as economy is one of the factors that affects the choice of transport, according to Movia (Trafikselskabet Movia, 2017).

Another solution that could help decrease the congestion is either to expand the existing infrastructure or in other ways improve it, e.g. by optimizing the signalling program to fit the temporal changes in traffic flow. Expansion of the road network is difficult in the cities as the fortification level is rather high, which is why optimization is the possibility to opt for within the cities. Where there are not as much built up areas, expansion is an option. However, when optimizing the infrastructure for cars or expanding the road network, it will avert the congestion problems that occurred before the improvement. But when creating better circumstances for travelling by car, there is a risk of this leading to more people choosing the car, and then it is back at the starting point, and there has to be found a new way of planning the way out of the problem. This is known as the *Induced Demand* that has been seen near various big cities around the world (Schneider, 2018).

7 Significance of the exploratory analytical tool

Based on the analyses carried out in the previous chapter, some of the results will be discussed in this chapter. This is done to validate the methods used and the analyses drawn from the tool.

7.1 Scale of analysis

As mentioned in the beginning of chapter 6 *Integration of the O-D tool in traffic planning*, more zone layers could have been included in the exploratory analytical tool. The layers chosen for this analysis is the municipality layer from Kortforsyningen, and a traffic model, LTM. Other layers from Kortforsyningen could also have been used, as they have vector layers for different administrative zones, which would be natural to include in a traffic or an urban planning project.

The layer of choice is dependent on the purpose of the analysis; some needs more detail and for that layers like the LTM, OTM or maybe the post code layer from Kortforsyningen. This could be analyses of the traffic flow from one part of a city to another, like it has been seen in the analysis above. For larger scale analyses, municipality or region layers would be relevant to use – this could be relevant for examining the flow of cars from one municipality to another. An example of this could be to see if there are many cars going into one municipality at the same time during peak hours – e.g. like seen for Copenhagen municipality in Figure 8. This could help traffic planners to see from which other municipalities are the biggest flows of traffic and where to develop the infrastructure. Another purpose could be to find an optimal location for a new industrial area, to place it somewhere that does not already have problems with a high number of incoming cars.

Even though layers with a smaller scale, such as post code zones might be of interest for some purposes, it was simply not considered to be relevant or add more value to the results of the analysis. There is a fine balance between the processing time and the detail of the data that should be considered, when choosing how to scale the analysis and the data properly.

When carrying out the analysis for Copenhagen and Frederiksberg municipalities, it would of course have given more detail to the results, and some patterns might have been more precise. On the other hand, other patterns would not have looked as distinct, as they do in this analysis, because the data would have been distributed at smaller zones. One analysis that would have been possible to perform with a smaller scale, would have been to estimate the paths between an origin-destination pair and the estimated mean velocity through that path. This could be used to conclude whether there are too many people driving above the speed limit in an area or to find areas with much congestion. Even though the scale of the start and end zone was smaller, the path between them could still be various.

Having larger zones gives a higher amount of data to base the analysis on, which might give a greater basis for the analysis, as the results will be based on more data, and hence might be closer to the actual traffic flow.

7.2 Benefits of the exploratory tool

The GPS data provided for this study could have been analysed in several ways, either by using a GIS, with a more geographic focus or in a more statistical way, looking at the data in matrices. As the data set contains more than 4 million data points, it is a lot of data to handle, which would have to be broken down into smaller bits. However, it was chosen to use a GIS and create an exploratory analytical tool, which handles the analysis of the data by aggregating it into groups, in order for planners to see if they can find any patterns in traffic flow.

This analytical tool can be useful in urban and traffic planning contexts, where various types of analyses and purposes are set for the data. Because the tool has grouped the GPS data points into grids, it can easily be applied to the map, either directly or through a join with a zone layer of any choice. Further development of the application could include creating a function that makes it even easier to add new layers to the map, as there is a bit pre-processing to do when adding a new layer. One way of optimizing and making the data more accurate, which was part of the initial thought, is to get the grids split percentwise based on the borders of the overlay. This would give more detailed and better distributed data and results. Having both municipality and LTM as overlays, which are of different scale and detail, limits the need of being able to include more layers. Besides, the difference might not be that big because of the amount of data and the size of the grid cells; if e.g. 20 % of a grid's data points belongs to another grid, this is most likely not a big part of the total number of cars, especially for larger zones. But this also depends on the time scale; for a large timescale, the impact of this would be larger, as more data would be included.

The analyses could have been carried out in other ways, but as done for this study, it has been developed to help urban and travel planners of COWI's Urban Planning and Transport department. The development of the analytical tool has been time consuming, and is not fully developed, as mentioned above (also refer to chapter 9 *Future work*), but after setting it up, multiple new analyses can be carried out much faster and cheaper. Since a similar analysis can be carried out using a few clicks, it does not take much time to get the results for a different area, and a specific period of time.

In order to create any estimations of future traffic flows using LTM, it requires the user to have completed a theoretical and a practical course at DTU, and the organization has to rent a license to the tool (DTU, 2019). For that reason, it has not been able to do a comparison of the output of an LTM analysis with an analysis based on the exploratory analytics tool created for this thesis.

However, this supports the need for an exploratory analytical tool like the one created here, where users do not need any licensing or courses. This can make the process of carrying out an analysis for an area much easier for planners, who do not know how to use the LTM, and hence usually has to wait for a colleague to do that part of the analysis.

Moreover, the analytical tool is a great way of creating a visual input and a basis for many analyses, where the processing of the data would be rather repetitive. By having a tool, which has preprocessed the data, and accordingly only needs a few clicks to get the wanted output, a lot of time could be saved, which could be beneficial for the team. The tool could either be used as a standalone tool, or in combination with other methods, as was shown in section 6.2 *Usage of GPS based O-D tool in traffic planning*.

7.2.1 Relevance of the exploratory analytics tool

The analytical tool is created based on a wish from COWI, as stated in chapter 2 *Conceptual Design*. Accordingly, it was estimated that a similar tool was essential for the processes of the planners of the Urban Planning and Transport department. Though, they do already work with other methods and tools for analysing the traffic situation of an area. Some of which are most typically used are the Ørestadstrafikmodellen (OTM) and the Landstrafikmodellen (LTM), as was shortly introduced in section 4.5 *Traffic models*. LTM estimates the future flow based on previous traffic counting. The model is regularly fixed and tested against actual traffic counting, to validate and calibrate the output of the model. This results in projections of future traffic flow that are as close to reality as possible. Besides estimating the future flow based on previous traffic data, the model also bases the outputs on projections of the population development, BNP, development in workplaces, the total road network, costs of travelling and public transport possibilities for the relevant year. In the next step, the population is grouped based on age, gender, zone of residence, affiliation with the labour market, followed by an estimation of the need for different types of transport depending on purpose of the trip and the number of trips. It is safe to say that the LTM takes many parameters into consideration when estimating the future flow of cars (or other types of transport). In that sense, the LTM might be more exact than the exploratory analytical tool created for this thesis. It also includes buses and lorries on the roads (Rich & Nielsen, 2018), which might have an effect on the flow of private cars, whereas the GPS data provided for this study only include private cars. This study focusses less on implementing many factors for estimation of a spatio-temporal change in the proposed tool, and instead focusses more on creating a multi-purpose tool that is scalable and applicable for various projects.

This case study is based on GPS data, which secures a rather high accuracy of the data, and the analysis is based on historical data. But as COWI has access the real-time data, which is used for the trafikinfo.dk, in which data is processed and visualized on the map within a very short amount of time, new data could be loaded into the database and included in the analyses as an ongoing process. This would ensure that the output of the tool would be more up to date and closer to the traffic situation at a given time.

The exploratory analytical tool created as part of this thesis will probably not be used instead of the existing methods, as they are well-established and based on multiple parameters, which make them rather trustworthy, and it was not wised upon to replace any of these methods entirely. But the application can assist the planners in the processing of the data, to get an overview of the flow, as it is an easily applied method to be used for screening an area's traffic flow. Moreover, it can be used to validate the results of the other methods, such as LTM, as they are validated occasionally, in order to calibrate them to fit the actual flow of traffic better.

7.3 Discussion

In the process of developing the tool, some functionalities had been added in the code, but are not used as it is now, due to the limitations in time (refer to chapter 9 *Future work* for suggestions on development of the analytical tool). It was decided that the most basic functionalities were of highest priority with the aim of being able to extract and analyse some data. These functionalities can be seen in Figure 6 in chapter 6 *Integration of the O-D tool in traffic planning*. Other functionalities were not necessary, but they will make the analytical tool more flexible and adds a better user interface. One of these functionalities that is added, but not fully functional, is interactivity to the layers. In the top-left corner of the tool, various layers can be projected onto the map. These layers have a mouseover function that highlights the individual zones, and a click function that zooms in to the selected area (see Figure 14). This could be applied to the layers created in GeoServer, however this would require them to be converted to WFS layers instead. By adding the click function, the area of interest could be selected by clicking the zone instead of choosing it from a drop-down menu. This is especially helpful for vector layers containing many zones, such as the OTM or LTM zone layers, which results in a long list of zone names, to which the spatial relation is not always obvious.

Moreover, more possibilities in relation to the time period could be added, so it is not just the months that could be chosen directly from the tool, but also time of day or predefined periods in time, such as holidays, or other. This could be in a drop-down menu as here, or by a combination of a drop-down menu, where time of day is chosen and a calendar from where the date(s) is chosen, or it could be multiple sliders, as in (Jiang, et al., 2015), where they included a slider for month, day and hour. The way the time periods is chosen now is by creating different layers in the GeoServer that aggregates the GPS data in the specified time intervals such as morning or afternoon traffic. In the user-interface it is only the months that can be chosen. This requires a bit more work to change the WMS layer that is being called in the JS code, which does make the tool less user-friendly. However, with the time remaining, it was chosen not to spend too much time figuring out a way of calling both types of temporal data, and instead implement a solution, where the hours of interest are called in the layer, as this uses a simple SQL query.

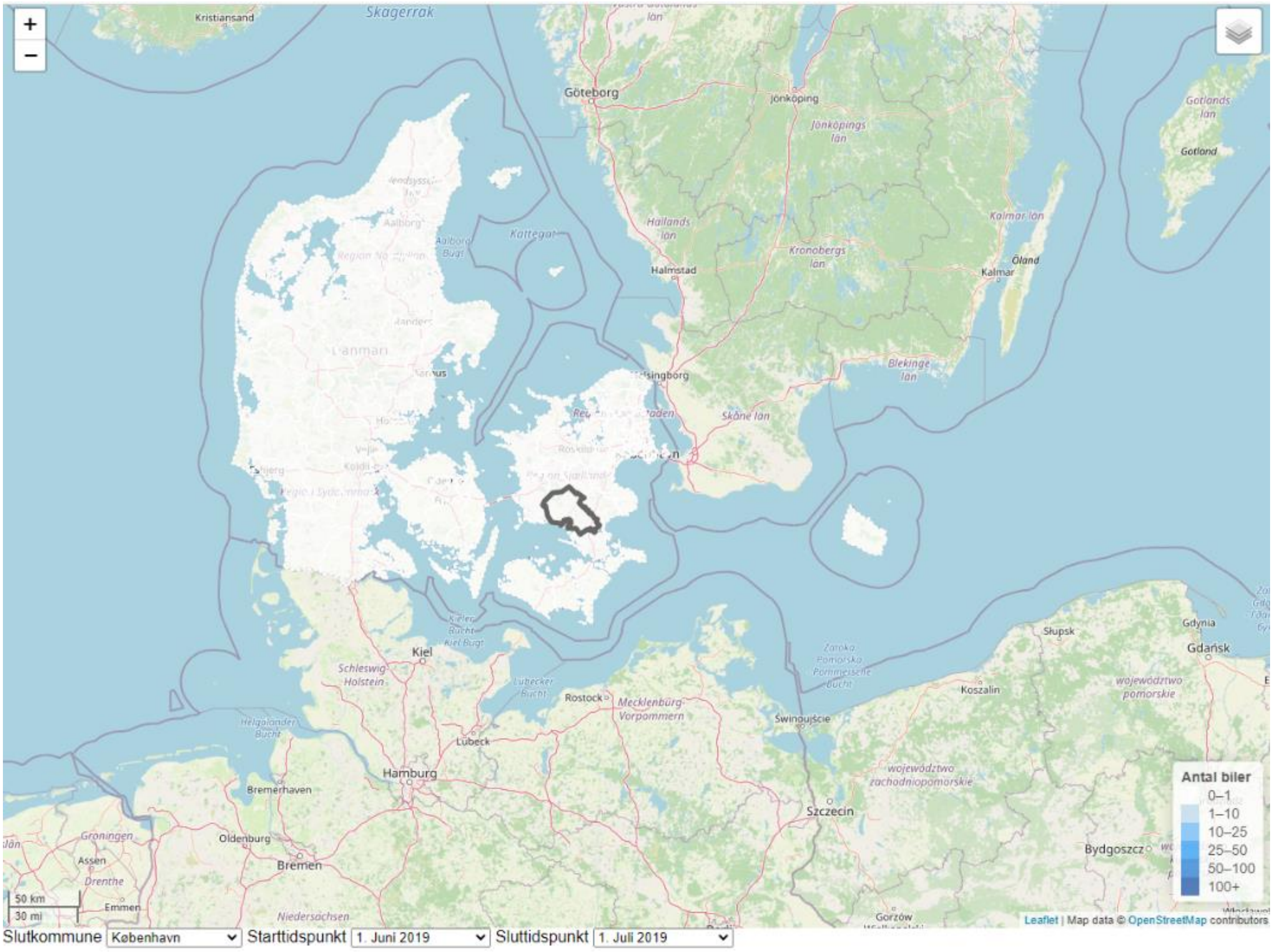


Figure 14: Mouseover function added to the municipalities layer. This and a click function could be implemented in the tool to make it more user-friendly.

Even though the data set contains data from many cars, it is only a small percentage of the total flow of cars in Denmark. It is estimated to make up around 4% of the total number of cars. In addition to this, the GPS data has only been collected from a group of cars, the Semler group. This might give a result that does not represent the reality, because the distribution of cars looks different. There might even be differences in travel behaviour depending on which type of car one chooses to drive, in which case the used data might not be that representable. Though, it has been evaluated that this is not relevant for the purpose of this dissertation, as is just an assumption. For this study, it has been assumed that the data is representative enough, based on the distribution of data both geographically and by type of car, as presented in section 5.1 *Data description*.

In Table 3 the number of cars going into Copenhagen municipality from respective municipalities during June 2019 are listed. When breaking this data down into smaller intervals such as days or even hours, the data volume is not quite as high, if even existent, as one could have wished for. Accordingly, the data in the analyses above have been grouped by morning and afternoon peak hours (weekdays only) to get a somewhat trustworthy result of the flow of cars from one are to another. This is a general trend, so as to get a more accurate representation of reality, a larger time scale should be considered.

Table 3: Distribution of cars going in to Copenhagen municipality in June. SQL: select count(*), start_kommune from test where end_kommune='København' and start_time between '2019-06-03 00:00:00' AND '2019-07-01 00:00:00' group by start_kommune

8411	København	104	Furesø	38	Kalundborg	3	Esbjerg
520	Tårnby	93	Hillerød	37	Faxe	2	Aabenraa
484	Frederiksberg	76	Fredensborg	35	Sorø	2	Samsø
382	Gentofte	74	Næstved	29	Ringsted	2	Haderslev
317	Hvidovre	71	Halsnæs	24	Lolland	2	Skanderborg
275	Gladsaxe	70	Køge	16	Odense	2	Hedensted
230	Rødovre	70	Slagelse	16	Stevns	2	Varde
195	Lyngby-Taarbæk	69	Allerød	14	Vejle	2	Faaborg-Midtfyn
141	Rudersdal	68	Egedal	11	Assens	2	Tønder
138	Odsherred	66	Albertslund	9	Middelfart	1	Vejen
135	Ballerup	65	Frederikssund	6	Århus	1	Skive
132	Herlev	59	Ishøj	5	Svendborg	1	Aalborg
126	Greve	59	Holbæk	4	Herning	1	Ringkøbing-Skjern
123	Glostrup	55	Vallensbæk	4	Horsens	1	Billund
119	Dragør	53	Lejre	4	Kolding	1	Silkeborg
115	Brøndby	52	Solrød	4	Bornholm	1	Randers
114	Gribskov	49	Hørsholm	4	Fredericia	1	Rebild
113	Roskilde	45	Vordingborg	4	Kerteminde	1	Langeland
107	Høje-Taastrup	41	Nyborg	4	Holstebro		
107	Helsingør	38	Guldborgsund	3	Nordfyns		

It should be mentioned that each set of O-D pair might not be the actual origin or destination of a trip; a trip starts from ignition and lasts until it is turned back off. A new trip ID is provided when turning on the ignition. Accordingly, longer trips with a pit stop, or a trip home from work with a stop at the grocery store, is not seen as one, but two separate trips. This might give the idea that more cars are driving in an area within a period of time, than what is actually the case. It is especially a problem for the islands that are connected by ferry instead of a bridge because it seems as if there are no cars going in/out of the islands. This problem could have been corrected for, as the raw data from Connected Cars contains a car id, but for the sake of privacy issues, the car id has been excluded from the aggregated data used in this study.

Working with O-D data has its advantages and disadvantages. For many purposes, O-D data is sufficient to analyse and determine patterns in time and space, and for these purposes it is the perfect choice of data. By not including the points of the entire trip, the sample size is decreased drastically, which decreases the processing time as well. Though, there might be cases where a greater detail is needed or where the points between the origin and the destination is vital for the analysis, to which the O-D pairs are not enough. In the Copenhagen OTM zones example in Figure 9, it can be seen how many cars are destinating in each zone, from which the hotspots of destination can be analysed to find out where to change the infrastructure or other, in order to limit the number of cars going there. However, this type of data does not show the number of cars on the roads at a given time, which might also be relevant for planners to know, as some areas could be congested and containing numerous cars on the road at the same time, but not being the destination zone of the trips. Hence, the O-D pairs might result in a wrong distribution of cars. However, it was expected that the zones that has most destinations would naturally also have many cars within them, and hence this method was acceptable for the purpose of this study.

7.3.1 Visualization of O-D patterns

When choosing a way of visualizing the flow of traffic from one zone to another, different variants were considered. The style of choice ended, as can be seen on Figure 6-9 and Figure 15 on a colour scale with breakpoints. This was chosen because it is easily interpreted, as the colours will make the different zones and O-D patterns rather distinct. Another way of visualisation that was considered was showing the O-D pairs by trajectories, either as individual trajectories for each trip or aggregated trajectories visualized by various sizes. The downside to these is that it can be quite difficult to see what is below the trajectories on the map and it can be a bit confusing with trajectories in different directories, especially for the single ones (Graser, 2016). Accordingly, the colour ramp was chosen as the visualisation method.

The results of the aggregated data could either be shown as percentage of total flow of cars with destination in the selected area or based on a total number. The option with percentage was first selected, combined with an interpolating style. This included more gradients on the map, and it is easier for a static map to compare the colours from different zones. However, when comparing two different zones or two different time periods, it looked like some areas included a total number of originating cars that were greater than for other periods/areas, which was not the case. Instead, a static colour scale, consisting of five breakpoints, was selected.

Based on the Figure 3 showing the distribution of cars by car type, and Figure 2 the distribution of cars in Denmark, it is assumed that the percentage is to some extent close to the percentage of all cars. Another way of visualizing the results could be to upscale the flow of cars from 4 % to 100 %, again based on the assumption that the total sum of cars is distributed as the cars from the data set, and that their patterns are the same. Furthermore, for the sake of visualization, it is easier to create a static legend, instead of a dynamic one; if the map showed the results of total number of cars, it would have to be dynamic to fit the scale of the time period in order to show details. Moreover, a static legend will make comparisons of different periods easier, because of the scale being the same. The downside of working with a static scale instead of percentages is that it is not as dynamic for multiple datasets, and the breakpoints might have to be adjusted to fit the scale.

7.3.2 Replication of analysis

Traffic patterns differ from city to city, and hence one solution that is suitable for one place, is likely not suitable for other places, even though they have similar infrastructure and size (Feldthaus & Odgaard, 2011). Though, the methods used could be the same, and as already stated above, the tool is created to the processing and extraction of GPS data for urban and traffic planners. The tool is created under the requirement that it should be possible to redo the analysis for various areas. In terms of recreating the visual, geographic analysis of another area and a different period of time, it is possible to do so, without any change of the tool as it is now. Though, in order to examine areas of other scale than the two included or other data, a bit more processing is required, but it is applicable to the tool as is.

One of the advantages of having a tool like this, is that is can be used for future analysis in other contexts. Moreover, the tool could be used by planners, who do not have any knowledge or competencies within database management using SQL, as soon as the basics has been set up including the selection options, as for this study. In that sense, the tool itself is applicable in other projects, but with the aim of getting full flexibility, it requires a bit more development or a bit of programming competencies from the user.

The analytical tool was developed solely by using Open-Source services that are combined to fulfil the proposition initiated by COWI. The process of creating the exploratory analytics is not black-boxed, since it has been described clearly which technologies has been included, as well as how they are combined. The basics of it is a basemap, to which a WMS layer is added, which is dependent on user input. The technologies are all Open-Source and free to use, so the exploratory analytics in this study could be used for further development by other planners to fit the purpose of other analyses and cases. The methods and technologies are transparent and available for everyone to use, which allows for others to use this as inspiration, and as a template for further development, where other layers, functionalities and time periods could be included. The work done in this study could be used as it is now, functionalities could be added or removed, or it could be used simply as inspiration of how to carry out similar analyses.

7.3.3 Other use cases

As the situation of the world changes, the demand for analytics and a greater understanding of these changes and their impacts increases. During the last months, the Corona Virus Disease (COVID-19) has been all over the media and it has undoubtedly had an implication on every activity during the that period. The analytical tool could be used to examine the flow of cars, ships, airplanes, or other modes of transport before, after and during the pandemic. An analysis like this could be based on GPS data like the one used in this study, but it could also be other kinds of data, such as itineraries or schedules of planes, stating when they are taking off and landing, as well as where. By extracting data about the arrivals from e.g. March-June 2020 with data for the same period last year, a clear drop would likely be seen. Analysing data for various modes of transport on a greater scale than for one nationality, it might be possible to examine the order in which the different countries were affected, based on their travelling patterns and find possible underlying patterns that could have been a side effect to the pandemic. As nations turn back to a somewhat normal daily life, more cars will naturally flow through the streets, because people are going back to work. In relation to that it could be interesting to inspect the patterns in car flow before and after the pandemic, to see if there has been a change in how many people commute or in the way they do it. Experts has recently been speculating in whether there might be a change in how we work and that more people will work from home (VIA Ritzau, 2020; Pröschold, 2020). If this is to happen, it will naturally lead to fewer cars on the streets. At least for a period, because after a while, the less congestion might lead to more people choosing the car instead of other modes of transport, as mentioned in section 6.2 *Usage of GPS based O-D tool in traffic planning*.

Putting COVID-19 aside, UN's Sustainable Development Goals (SDG) are highly prioritized in many enterprises, and many projects work as a step towards improving the sustainability in relation to one or more of these goals (verdensmaal.org, n.d.). As stated above in section 7.3.2 *Replication of analysis*, the exploratory analytical tool could be used as inspiration or a template for future analyses, where some aspects were changed. Accordingly, the analytical tool could also be the base of analyses for examining and detection improvements within reaching the goals. The Danish government wishes to keep close track of how Denmark is progressing towards meeting each of the 17 SDGs and proposes that this will happen in the municipalities and regions as well, in order to ensure anchorage and fulfilment of the goals (Udenrigsministeriet, 2017). Using an analytical tool like the one proposed in this study can be helpful in this case.

Denmark is already rather close towards reaching the goals, but there still needs to be great focus on improving. In this regard, the Danish government has prioritized some of the development goals, which they think are most important to focus on and grouped them into four categories; growth and prosperity, people, environment and climate, and safe and sound communities. Some are more suitable for spatio-temporal tracing of improvements, especially some of the goals that related to the environment and climate. One of these is to protect the aquatic and marine environment, which involves protecting the fish stocks and the water conditions, which is part of SDG number 6 and number 14 (Udenrigsministeriet, 2017). Having repeatedly collected data about the number of fishes for individual lakes, creeks, inlets and more, the data could be implemented in a similar exploratory analytical tool, where the period of time and area of interest could be selected. The amount of data would not be as large as the one used in this study, as the GPS data is collected automatically, while data about fish would have to be collected in a more manual way. Accordingly, the selection of time during the day would not make much sense, but the tool can be suited for this purpose as well, by setting up the basics of it in the HTML code, where the selection options are created.

The same could be done for the goal to stop the decline of biological diversity (goals number 12, 14, 15) (Udenrigsministeriet, 2017), which in the same way could include time and space as user inputs for the detection of changes in condition of the diversity, where the seasonal changes will have to be of higher interest than the hourly changes.

8 Conclusion

This chapter will sum up the results carried out through this study, related to the problem statement – how an exploratory analytical tool can be created to indicate spatio-temporal patterns, such as congested areas by the use of GPS data; how such a tool can be applied to a unique dataset of GPS points and; and how such a tool can benefit urban and traffic planning.

Through this study, an exploratory analytical tool has been created with focus on having a tool that is suitable for multiple cases and multiple datasets. The tool is suitable for carrying out spatio-temporal analysis within a variety of fields of research.

By the use of GPS data collected in Denmark through 2019, a case study was conducted, using the analytical tool. This was done in order to show the functionality of the analytical tool, and examine a large dataset based on the results that can be drawn from the tool.

The spatio-temporal pattern of cars in Denmark can be analyzed in various scales; for the purpose of this case study the scales where municipality and OTM layers. When examining the spatio-temporal patterns on municipality layer, it was concluded that seasonal changes could be found in the data set, by selecting various times of the year in the tool. Additionally, adjacent municipalities more often were the origin of a trip, than municipalities from further away. Focusing on Copenhagen and Frederiksberg municipalities, it seems that a lot of cars were entering these zones both in the morning and the afternoon, though more people came from further away in the morning, while in the afternoon, the traffic going into Copenhagen and Frederiksberg were more concentrated in some areas. It was clear in both periods of time that many trips were conducted internally in the two municipalities. Using the exploratory tool as a complementary method, additional analyses of the OTM layer were carried out in QGIS, to get a better understanding of the internal trips. This shows in more detail where internal trips end on a smaller zone level, which gives a better idea of where the most problematic areas are during the peak hours. Additional analysis of the OTM layers were carried out by comparing the zones' incoming traffic with the infrastructure of the city. From an overlay analysis in QGIS, there seemed to be no clear insufficiencies in the infrastructure, so a more qualitative follow-up analysis could be helpful in understanding these patterns and choice of transport mode, which could be useful in some cases, in order to plan the most optimal solution for the congested areas.

The analytical tool can be useful in urban and traffic planning as an exploratory tool for various purposes. When working on a project, it could be used as the main tool for examining a given area to find patterns in time and space. Though, it can also be used as a complementary tool to point out which areas to dive more into, using other methods. The tool is created to be rather user-friendly, as soon as the data is loaded and aggregated into the wished zonal boundaries. Once this is done, the data can be used for analyses for various areas and periods of time. Regarding the case study and the used GPS data, planners from COWI's Urban Planning and Transport department can use the analytical tool for several future analyses. As the process of aggregating the data for smaller areas would be rather repetitive, it could save the planners time using this tool, as the data is already pre-processed and aggregated into different zone layers of various scale.

The tool provides interactivity, which makes it easy to use for all planners, as only a few clicks are needed to extract aggregated data for the area and time interval of interest. However, additional functionalities and refinements of the existing functionalities could be included to improve the user interface of the tool and make it even easier to use.

The analytical tool is created using only Open-Source technologies. By creating an analytical tool of combined Open-Source technologies, the analytical tool can benefit planners working on other projects as well, either by exchanging the data for the type of data needed and keeping the rest, or as a template for inspiration. The case study conducted in this study is just one example of use cases for the analytical tool, but it is suitable for other purposes and datasets as well. It could e.g. be very useful in detecting the changes made towards meeting the Sustainable Development Goals, or the change in traffic flow caused by the Covid-19 pandemic.

9 Future work

Due to the limited time period for the thesis, there were some limitation to what was possible to do. By having ongoing status meetings with COWI, different ideas came to the table; though not all was possible within the time span, and the analyses and possible features to the tool had to be prioritized. Some of these will be presented in this final chapter, alongside with some other possible future work that could be carried out in relation to this topic.

The tool functions as it is now but could be made much faster and more flexible. Two of the possible ways of decreasing the processing time would be to either pre-process data in PostGIS, where tables could be created for aggregated data in hours or time periods such as peak hours, etc. This way the query does not have to count the total number of cars within the specified time period in the SQL code in GeoServer. Another option, which would make the loading of data faster and keep the flexibility of choosing the time down to minutes, is switching to client-side rendering instead of server-side rendering, as it is now. Moreover, implementing caching would be useful in order to only let the map update when an update button is clicked, whereas now it updates whenever the zoom level changes or the map is dragged to a new area (Breux, n.d.). Switching to server-side rendering would include using WFS layers instead of WMS, and this would allow more interactivity with the layers, such as clicking the area of interest on the map instead of choosing from a drop-down menu bar.

For the purpose of further development of the tool, more flexibility would be great to include; being able to add on any vector layer and getting the O-D results for that specific layer would be one option that would be great to include. Though, this was not prioritized for this first step of development, because it was decided that if multi-scaled layers were already a solution, then there might not be that big a need for customized layers.

Regarding the analyses that could be carried out from the results of the O-D analyses, there are a long, long list. Some of those that were considered including were comparison of traffic models such as Ørestadstrafikmodellen (OTM) or Landstrafikmodellen (LTM). By comparing the two, it could be used to validate the results of estimations based in the models with real traffic data. But as already stated in section 7.2 *Benefits of the exploratory tool*, it requires training to use LTM, and OTM as well, accordingly it was not possible to carry out an analysis to compare with for this thesis.

Another analysis could be to examine the infrastructure of public transport to see if there are any hot spots of cars starting or ending their trips, but not any good connections with public transport. It could also be used to analyse the basis for carpooling; if many cars go from a closely related area to another closely related area in the same period of time most days, there might be a possibility for focussing on getting more people to share rides – e.g. by creating an app to help people find drivers or passengers to share the ride with.

For super markets or shopping malls it might be relevant to see where their customers come from, which can also be drawn from this data set. In this way they can target their marketing the right places to attract more customers.

Bibliography

Anon., n.d. *Clustering: An Introduction.* [Online]
Available at: http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/index.html
[Accessed 28 02 2020].

Arbex, R., da Cuncha, C. B. & Speicys, R., 2019. Before-and-after evaluation of a bus network improvement using performance indicators from historical smart card data. *Public Transport.*

Bahuleyan, H. & Vanajakshi, L. D., 2017. Arterial Path-Level Travel-Time Estimation Using Machine-Learning Techniques. *Journal of Computing in Civil Engineering,* 31(3).

Bandeira, J., Guarnaccia, C., Fernandes, P. & Coelho, M., 2018. Advanced Impact Integration Platform for Cooperative Road Use. *International Journal of Intelligent Transportation Systems Research,* 16(1), pp. 1-15.

Bellemans, T., De Schutter, B. & De Moor, B., 2002. Models for traffic control. *Journal A,* 43(3-4), pp. 13-22.

Breux, G., n.d. *Client-side vs. Server-side vs. Pre-rendering for Web Apps.* [Online]
Available at: https://www.toptal.com/front-end/client-side-vs-server-side-pre-rendering
[Accessed 06 01 2020].

Cambridge Dictionary, n.d. *grid.* [Online]
Available at: https://dictionary.cambridge.org/dictionary/english/grid
[Accessed 27 02 2020].

Cerqueira, V., Moreira-Matias, L., Khiari, J. & van Lint, H., 2018. On Evaluating Floating Car Data Quality for Knowledge Discovery. *IEEE Transactions on Intelligent Transportation System,* 18(11), pp. 3749-3760.

Chen, C. et al., 2017. ScenicPlanner: planning scenic travel routes leveraging heterogeneous user-generated digital footprints. *Frontiers of Computer Science,* 11(1), pp. 61-74.

Connected Cars, n.d. *Customer-centric workshop management.* [Online]
Available at: https://connectedcars.dk/workshop/
[Accessed 20 03 2020].

COWI, 2019. *Salgspræsentation_MASTER.* Lyngby: Unpublished.

Danmarks Statistik, 2020 a. *FOLK1A: Folketal den 1. i kvartalet efter område, køn, alder og civiltilsand.* [Online]
Available at: https://www.statistikbanken.dk/10021
[Accessed 18 05 2020].

Danmarks Statistik, 2020 b. *BIL5: Nyregistrerede motorkøretøjer efter køretøjstype.* [Online]
Available at: (https://www.statistikbanken.dk/10091
[Accessed 18 05 2020].

Danmarks Statistik, 2020 c. *ERHV5: Arbejdssteder efter område og sektor.* [Online]
Available at: https://www.statistikbanken.dk/10096
[Accessed 18 05 2020].

Dirmeyer, P. A. et al., 2016. Confronting weather and climate models with observational data from soil moisture networks over the United States. *Journal of hydrometeorolog,* 17(4), pp. 1049-1067.

DTU, 2017. *Om modellen.* [Online]
Available at: http://www.landstrafikmodellen.dk/om_modellen
[Accessed 15 04 2020].

DTU, 2019. *Adgang til modellen.* [Online]
Available at: www.landstrafikmodellen.dk/adgang-til-modellen
[Accessed 20 05 2020].

esri, n.d. a. *Why hexagons?.* [Online]
Available at: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-whyhexagons.htm
[Accessed 02 03 2020].

esri, n.d. b. *Cell size of raster data.* [Online]
Available at: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/cell-size-of-raster-data.htm
[Accessed 01 04 2020].

European Global Navigation Satellite Systems Agency, 2018. *Programme.* [Online]
Available at: https://www.gsa.europa.eu/galileo/programme
[Accessed 05 03 2020].

European Global Navigation Satellite Sytems Agency, 2020. *FAQ.* [Online]
Available at: https://www.gsa.europa.eu/european-gnss/galileo/faq#FOC
[Accessed 01 06 2020].

Feldthaus, P. & Odgaard, G., 2011. *Kronik: Bæredygtig bevægelse i byen.* [Online]
Available at: https://politiken.dk/debat/kroniken/art5445569/Bæredygtig-bevægelse-i-byen
[Accessed 29 05 2020].

Frederiksberg Kommune, n.d.. *Parkeringspladser.* [Online]
Available at: https://www.opendata.dk/city-of-copenhagen/parkeringspladser
[Accessed 20 05 2020].

Gan, M. et al., 2018. The identification of truck-related greenhouse gas emissions and critical impact factors in an urban logistics network. *Journal of Cleaner Production,* Volume 178, pp. 561-571.

García-Bartual, R. & Andrés-Doménech, I., 2016. A two-parameter design storm for Mediterranean convective rainfall. *Hydrology and Earth System Sciences Discussions,* pp. 1-19.

Geofrabrik, n.d. *Download OpenStreeMap data for this region: Denmark.* [Online]
Available at: https://download.geofabrik.de/europe/denmark.html
[Accessed 20 05 2020].

GeoServer.org, n.d. a. *What is Geoserver?.* [Online]
Available at: http://geoserver.org/about/
[Accessed 20 05 2020].

GeoServer, n.d. b. *Introduction to SLD.* [Online]
Available at: https://docs.geoserver.org/stable/en/user/styling/sld/introduction.html
[Accessed 29 05 2020].

Google, 2020. *Google Cloud Platform Overview.* [Online]
Available at: https://cloud.google.com/docs/overview
[Accessed 25 04 2020].

Google, n.d. *Compute Engine.* [Online]
Available at: https://cloud.google.com/compute
[Accessed 25 04 2020].

Graser, A., 2016. *Free and Open Source GIS Ramblings.* [Online]
Available at: https://anitagraser.com/2016/11/07/movement-data-in-gis-3-visualizing-massive-trajectory-datasets/
[Accessed 10 05 2020].

Guo, D. et al., 2012. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Transaction in GIS,* 16(3), pp. 411-429.

Hoffman, C., 2017. *Beginner Geek: How to Create and Use Virtual Machines.* [Online]
Available at: https://www.howtogeek.com/196060/beginner-geek-how-to-create-and-use-virtual-machines/
[Accessed 25 04 2020].

HTML.dk, n.d. *Den ultimative guide til HTML.* [Online]
Available at: https://www.html.dk/tutorials/html/#Hvad_er_HTML
[Accessed 28 04 2020].

Hu, Z. et al., 2017. Computing a hierarchy favored optimal route in a Voronoi-based road network with multiple levels of detail. *International Journal of Geographial Information Science,* 31(11), pp. 2216-2233.

Hyun, K. & Ritchie, S., 2017. Sensor location decision model for truck flow measurement. *Transportation Research Record,* 2644(1), pp. 1-10.

Hägerstraand, T., 1970. What about people in regional science?. *Papers in Regional Science,* 24(1), pp. 7-24.

Jiang, X., Zheng, C., Tian, Y. & Liang, R., 2015. Large-scale taxi O/D visual analytics for understanding metropolitan human movement patterns. *Journal of Visualization,* 18(2), pp. 185-200.

Jiang, Y. & Li, X., 2013. Travel time prediction based on historical trajectory data. *Annals of GIS,* 19(1), pp. 27-35.

Knapen, L. et al., 2016. Determining structural route components from GPS traces. *Transportation Research Part B,* Volume 90, pp. 156-171.

Kyes, J., 2020. *What is GPS?.* [Online]
Available at: https://www.geotab.com/blog/what-is-gps/
[Accessed 28 05 2020].

Københavns Kommune, n.d. a. *Cykeldata.* [Online]
Available at: https://www.opendata.dk/city-of-copenhagen/cykeldata
[Accessed 30 05 2020].

Københavns Kommune, n.d. b. *P-pladser.* [Online]
Available at: https://www.opendata.dk/city-of-copenhagen/parkeringspladser
[Accessed 20 05 2020].

Lee, T., Yeo, G.-T. & Thai, V. V., 2014. Changing concentration ratios and geographical patterns of bulk ports: The case of the Korean West Coast. *The Asian Journal of Shipping and Logistics,* 30(2), pp. 155-173.

Linux.com, n.d. *What Is Linux?.* [Online]
Available at: https://www.linux.com/what-is-linux/
[Accessed 27 04 2020].

Li, S. et al., 2017. A study on bus passenger travel origin and destination based on Spatio-Temporal data fusión analysis. *Advances in Transportation Studies,* 1(Special), pp. 31-38.

Li, Y.-S.et al., 2014. A near-real-time automatic orbit determination system for cosmic and its follow-on satellite mission: Analysis of orbit and clock errors on radio occultation. *IEEE Transactions on Geoscience and Remote Sensing,* 52(6), pp. 3192-3203.

Lovelace, R. & Leigh, E., n.d. *Origin-destination data with stplanr.* [Online]
Available at: https://cran.r-project.org/web/packages/stplanr/vignettes/stplanr-od.html
[Accessed 10 03 2020].

Luo, S. & Miller, J. I., 2014. On the spatial correlation of international conflict initiation and other binary and dyadic dependent variables. *Regional Science and Urban Economics,* Volume 44, pp. 107-118.

Malinga, J. et al., 2019. Investigating the drivers of the spatio-temporal patterns of genetic differences between Plasmodium falciparum malaria infections in Kilifi County, Kenya. *Scientific Reports,* Volume 9.

Matejicek, L., 2010. Spatial Modelling of Air Pollution Based on Traffic Emissions in Urban Areas. In: *Environmental modeling with GIS.* New York: Nova Science Publishers, Inc., pp. 85-99.

Matejicek, L., 2011. Estimating the Environmental Effects on Residential Property Value with GIS. *Environmental Modelling with GIS,* pp. 113-127.

Matejicek, L., 2013. Estimating the environmental effects on residential property value with GIS. *Geographic Information Systems,* pp. 119-132.

Ma, X., Mccormack, E., Wang, Y. & Ma, X., 2011. Processing Commercial Global Positioning System Data to Develop a Web-Based Truck Performance Measures Program. *Transportation Research Record,* Issue 2246, pp. 92-100.

Ma, X. & Wang, Y., 2014. Development of a data-driven platform for transit performance measures using smart card and GPS data. *Journal of Transportation Eningeering,* 140(12).

MDN contributors, 2020 a. *JavaScripts basics.* [Online]
Available at: https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/JavaScript_basics
[Accessed 01 06 2020].

MDN contributors, 2020 b. *About JavaScript.* [Online]
Available at: https://developer.mozilla.org/en-US/docs/Web/JavaScript/About_JavaScript
[Accessed 15 04 2020].

MDN contributors, 2020. *HTML: Hypertext Markup Language.* [Online]
Available at: https://developer.mozilla.org/en-US/docs/Web/HTML
[Accessed 30 05 2020].

Mehndiratta, S. & Alvim, B., 2014. *Big Data comes to transport planning: how your mobile phone helps plan that rail line.* [Online]
Available at: https://blogs.worldbank.org/transport/big-data-comes-transport-planning-how-your-mobile-phone-helps-plan-rail-line
[Accessed 10 03 2020].

Miller, H. J., 2017. Time Geography. *Encyclopedia of GIS,* pp. 2235-2242.

Millward, H., Spinney, J. & Scott, D., 2013. Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography,* Volume 28, pp. 101-110.

Naghizade, E., Chan, J. & Tomko, M., 2020. From small sets of GPS trajectories to detailed movement profiles: quantifying personalized trip-dependent movement diversity. *International Journal of Geographical Information Science,* pp. 1-26.

Packt, n.d. *Importing shape files using shp2pgsql.* [Online]
Available at:
https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781784391645/1/ch01lvl1sec2/importing-shape-files-using-shp2pgsql
[Accessed 06 03 2020].

pgAdmin, n.d. *pgAdmin 4.* [Online]
Available at: https://www.pgadmin.org/docs/pgadmin4/latest/index.html
[Accessed 28 04 2020].

PostGIS.net, n.d. a. *About PostGIS.* [Online]
Available at: https://postgis.net/
[Accessed 28 04 2020].

PostGIS.net, n.d. b. *Feature List.* [Online]
Available at: https://postgis.net/features/
[Accessed 28 04 2020].

Postgresql, n.d. a. *What is PostgreSQL?.* [Online]
Available at: https://www.postgresql.org/about/
[Accessed 28 04 2020].

Postgresql, n.d. b. *1. What Is PostgreSQL?.* [Online]
Available at: https://www.postgresql.org/docs/12/intro-whatis.html
[Accessed 28 04 2020].

Pröschold, J., 2020. *Ny undersøgelse: Hver femte chef vil have ansatte til at udføre mere hjemmearbejde efter coronakrisen.* [Online]
Available at: https://politiken.dk/indland/art7743336/Hver-femte-chef-vil-have-ansatte-til-at-udf%C3%B8re-mere-hjemmearbejde-efter-coronakrisen
[Accessed 30 05 2020].

qgis.org, n.d. a. *Introducing GIS.* [Online]
Available at: https://docs.qgis.org/3.10/en/docs/gentle_gis_introduction/introducing_gis.html#figure-gis-application
[Accessed 28 04 2020].

qgis.org, n.d. b. *DB Manager Plugin.* [Online]
Available at: https://docs.qgis.org/2.14/en/docs/user_manual/plugins/plugins_db_manager.html
[Accessed 03 06 2020].

Rabbel, W., Koulakov, I., Dinc, A. N. & Jakovlev, A., 2011. Arc-parallel shear deformation and escape flow in the mantle wedge of the Central America subduction zone: Evidence from P wave anisotropy. *Geochemistry, Geophysics, Geosystems,* 12(5).

Rao, W. et al., 2018. Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data. *Transportation Research Part C,* Volume 95, pp. 29-46.

Rasmussen, T. K., Nielsen, O. A., Watling, D. P. & Prato, C. G., 2017. The Restricted Stochastic User Equilibrium with Threshold model: Large-scale application and parameter testing. *European Journal of Transport and Infrastructure Research,* 17(1), pp. 1-24.

Redondi, R., Malighetti, P. & Paleari, S., 2011. Hub competition and travel times in the world-wide airport network. *Journal of Transport Geography,* 19(6), pp. 1260-1271.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016. *Regulations on the protection of natural persons with regard to the processing of personal data and on the free.* [Online]
Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
[Accessed 20 04 2020].

Rich, J. & Nielsen, O. A., 2018. *Den danske landstrafikmodel Version 2.0.* [Online]
Available at: http://asp.vejtid.dk/Artikler/2018/11/9148.pdf
[Accessed 15 05 2020].

Rouse, M., 2017. *Google Cloud Platform.* [Online]
Available at: https://searchcloudcomputing.techtarget.com/definition/Google-Cloud-Platform
[Accessed 25 04 2020].

Rouse, M., 2019. *Public cloud.* [Online]
Available at: https://searchcloudcomputing.techtarget.com/definition/public-cloud
[Accessed 23 03 2020].

Schneider, B., 2018. *CityLab University: Induced Demand.* [Online]
Available at: https://www.citylab.com/transportation/2018/09/citylab-university-induced-demand/569455/
[Accessed 23 05 2020].

Schoier, G. & Burroso, G., 2012. Spatial Data Mining for Highlighting Hotspots in Personal Navigation Routes. *International Journal of Data Warehousing and Mining,* 8(3), pp. 45-61.

Scopus, n.d. *About.* [Online]
Available at: blog.scopus.com/about
[Accessed 01 03 2020].

Skovgaard, H.-O., 1996. *Beskrivelse af Ørestadstrafikmodellen.* s.l., s.n.

SMATS, n.d. *Origin Destination.* [Online]
Available at: https://www.smatstraffic.com/solutions/traffic-management/origin-destination/
[Accessed 10 03 2020].

Song, Y. & Miller, H. J., 2015. Beyond the Boundary: New Insights from Inside the Space-Time Prism. *Space-Time Integration in Geography and GIScience,* pp. 189-209.

Song, Y., Miller, H. J., Stempihar, J. & Zhou, X., 2017. Green accessibility: Estimating the environmental costs of network-time prisms for sustainable transportation planning. *Journal of Transport Geography,* Volume 64, pp. 109-119.

Soycan, M. & Ocalan, T., 2011. A regression study on relative GPS accuracy for different variables. *Survey Review,* 43(320), pp. 137-149.

SSH.com, n.d. a. *What Is an SSH Server?.* [Online]
Available at: https://www.ssh.com/ssh/server
[Accessed 26 04 2020].

SSH.com, n.d. b. [Online]
Available at: (https://www.ssh.com/ssh/best_ssh_clients
[Accessed 26 04 2020].

Strimas-Mackey, M., 2020. *Fishnets and Honeycomb: Square vs. Hexagonal Spatial Grids.* [Online]
Available at: https://strimas.com/post/hexagonal-grids/
[Accessed 01 04 2020].

Styrelsen for Dataforsyning og Effektivisering, n.d. b. *Danmark 1:1.000.000, vektor.* [Online]
Available at: https://download.kortforsyningen.dk/content/danmark-11000000-vektor
[Accessed 20 05 2020].

Styrelsen for Dataforsyning og Effektivisering, n.d. *Danmarks Administrative Geografiske Inddeling 1:10.000.* [Online]
Available at: https://download.kortforsyningen.dk/content/danmarks-administrative-geografiske-inddeling-110000
[Accessed 15 03 2020].

Talavera, E. et al., 2018. Performance Comparison of Geobroadcast Strategies for Winding Roads. *Electronics,* 7(3), p. 32.

Tang, W. & Cheng, L., 2016. Analyzing multiday route choice behavior of commuters using GPS data. *Advances in Mechanical Engineering,* 8(2), pp. 1-11.

Tarasenko, A., 2019. *Leaflet vs OpenLayers.* [Online]
Available at: https://www.geoapify.com/leaflet-vs-openlayers/
[Accessed 27 04 2020].

The Python Software Foundation, n.d. *The Python Tutorial.* [Online]
Available at: https://docs.python.org/3/tutorial/index.html
[Accessed 29 04 2020].

Ton, D., Duives, D., Cats, O. & Hoogendoorn, S., 2018. Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society,* Volume 13, pp. 105-117.

Trafikselskabet Movia, 2017. *Hvorfor transporterer vi os, som vi gør?.* [Online]
Available at: https://www.moviatrafik.dk/media/6760/hvorfor-transporterer-vi-os-som-vi-goer_pixi.pdf
[Accessed 21 05 2020].

tutorialspoint, n.d. *PostgreSQL - INDEXES.* [Online]
Available at: https://www.tutorialspoint.com/postgresql/postgresql_indexes.htm
[Accessed 02 05 2020].

Udenrigsministeriet, 2017. *Handlingsplan for FN's verdensmål - Danmarks opfølgning på FN's verdensmål for bæredygtig udvikling,* København K: Regeringen.

United Nations, 2018. *The World's Cities in 2018,* s.l.: Department of Economic and Social Affairs.

verdensmaal.org, n.d. *Virksomheder.* [Online]
Available at: https://www.verdensmaal.org/virksomheder
[Accessed 28 05 2020].

VIA Ritzau, 2020. *Næsten halvdelen vil arbejde mere hjemmefra efter coronatiden.* [Online]
Available at: https://via.ritzau.dk/pressemeddelelse/naesten-halvdelen-vil-arbejde-mere-hjemmefra-efter-coronatiden?publisherId=11341474&releaseId=13592814
[Accessed 29 05 2020].

visualstudiocode.com, n.d. *Code editing. Redefinted..* [Online]
Available at: https://code.visualstudio.com/
[Accessed 25 04 2020].

w3schhols, n.d. *CSS Tutorial.* [Online]
Available at: https://www.w3schools.com/css/
[Accessed 01 05 2020].

Wu, L. et al., 2014. Intra-urban human mobility and activity transition: evidence from social media check-in data.. *PloS one,* 9(5).

Ye, P., Wu, B. & Fan, W., 2016. Modified betweenness-based measure for prediction of traffic flow on urban roads. *Transportation Research Record,* 2563(1), pp. 144-150.

Zeng, W., Miwa, T., Wakita, Y. & Morikawa, T., 2015. Application of Lagrangian relaxation approach to α-reliable path finding in stochastic networks with correlated link travel times. *Transportation Research Part C,* Volume 56, pp. 309-334.

Zhai, X., Zhu, X., Xiao, Z. & Weng, J., 2009. Research on the comparison of extension mechanism of cellular automaton based on hexagon grid and rectangular grid. *Proceedings of SPIE,* 7492(1), pp. 74924K-74924K-10.

Zhan, X., Hasan, S., Ukkusuri, S. V. & Kamga, C., 2013. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C,* Volume 33, pp. 37-49.

Zhao, W. & Goodchild, A. V., 2011. Truck travel time reliability and prediction in a port drayage network. *Maritime Economics & Logistics,* 13(4), p. 387.

Zhua, R. et al., 2020. Understanding spatio-temporal heterogeneity of bike-sharing and scooter-sharing mobility. *Computers, Environment and Urban Systems,* Issue 81.