Universiteit Leiden

ICT in Business and the Public Sector

Intermodality and Transport Delays,
A Case Study Integrating a Big Data Approach
to Supply Chain Analytics

Name: Max Doesburg
Student-no: 1518445

Date: 29/06/2022

1st supervisor: Prof.dr. Stefan Pickl

2nd supervisor: Prof.dr. Simcha Jong Kon Chin

MASTER'S THESIS

Abstract

Supply chain networks are a subject where the complexity of the problems encountered is constantly growing as the supply chains of the world become ever more large and complex and with the problems affecting those supply chains, the vulnerability increases.

To compensate for this extra complexity and ever growing problems, special algorithms and adaptive techniques for extracting big data from these networks can be used as well as finding new innovative ways to increase the agility, robustness and resilience of those complex supply chains.
This thesis is a case study on using a certain big data approach from one such supply chain to identify and better understand what the connection is between the transport type in a supply chain network on one site and the amount of delays that are present in that supply chain network on the other site. This is done by taking advantage of some specific methods of data analytics and data science which allows for the extraction and analysis of the relevant data from the data specific set. We compare our approach with existing ones and develop a benchmark comparison for a certain testbed.
First, numerical results and insights will be presented

1 Introduction

In this chapter a brief introduction will be given about the topic, some background information will be given as to why the problem was chosen. After this, the research questions together with the scope of the research will be discussed.

The personal reason for this research was to look at a concrete problem in economics and solve it, or at least look deeper into it, using a real world data set and big data analytics.
Supply chain analytics and supply chain problems have been in the news very prominently the past couple of years, from the supply chain woes of covid and China having their output reduced because of it, to the Ever Given getting stuck in the Suez Canal to the shortage we are experiencing right now in products like gas, oil and grain due to the Ukraine.

Looking at supply chain problems and delays looked, from the outside, like a fairly complex and unexplored puzzle with multiple factors that need to be accounted for. Exploring all of these factors would be out of scope for the timeframe and resources I have, but exploring one of these factors is something that can be done.

The main subject is comparing the efficiency of different transport modes in supply chains, specifically trains, trucks and ships using a big data set.
This type of problem, trying to identify/fix problems in supply chains using a concrete big data set, has not been explored much. The literature and papers that could be found about supply chain analytics focuses more on models and less on the actual usage of big data themselves, with concrete use cases being a rare sight.
This thesis seeks to merge the gap by taking a concrete data set and extracting a useful conclusion about intermodality, the act of swapping transport types for the transport of goods, and the effect of transport modes on delays suffered during freight transport.

1.1 Background

Big data is everywhere and it is growing. Improvements are being made in the field of big data on a daily basis, be it the use of cloud computing, the advancement of the field of big data analytics, through advancements like machine learning, or the exponentially increasing pool of data that can be accessed.
Humans produced 5 exabytes (10^{18} bytes) of data till 2003. This much data is now produced in only two days. The digital universe of data grew to 2.72 zettabytes in 2012 (10^{21} bytes) (Sagiroglu & Sinanc, 2013).

This quantity of data offers many opportunities such as better aimed marketing, more straight business insights, client based segmentation, recognition of sales and market chances, automated decision making, definitions of customer behaviors, greater return on investments, quantification of risks and market trending, comprehension of business alteration, better planning and forecasting, identification consumer behavior from click streams and production yield extension (Russom, 2011).

But it also has many challenges like cannot find or hire big data experts, cost, privation of business sponsorship, hard to design analytical systems, lack of current database software in analytics and fast process times, scalability problems, incapable to make big data usable for end users, slow loading data in current database software, lack of compelling business case (Russom, 2011).

Using big data to get a better picture of supply chains and even using big data to predict supply chain woes (Seyedan & Mafakheri, 2000) is slowly becoming more popular and better understood. Predicting anything can be hard and predicting how supply chains are going to be in the future is an ongoing area of research. From predicting how supply chains are going to react to disruptive events, like the recent 2021 Suez Canal obstruction (Harper, 2021), to things like the bullwhip effect (Lee et al., 1997) and information asymmetry. This on top of figuring out how to ship goods from point A to point B in the most efficient way possible in the future.

The algorithms around the predictive analysis of supply chains are able to handle more and more information with larger accuracy and have come far in recent years. Viellechner, (2022) showcases this with a neural network trained to predict delays with an accuracy of 77% for container vessels, and takes into account a wide variety of factors such as: piracy, time between ports and traffic in maritime chokepoints like the Suez Canal.

The final piece of background information is about the advances in intermodal transport models, which is the focal point of this thesis. For example, Yildiz et al., (2021) tries to figure out how to ship goods from point A to point B in the most efficient way possible with a focus on reducing the cost of money and time of interhub transfers of intermodal transport, which can cause a lot of problems as discussed in the next chapter, Chapter 1.2.

1.2 Problem Statement

One of the areas of application for big data is the supply chain. Transporting goods from location to location has long been a process fraught with problems, just look at the 2021 Suez Canal obstruction (Harper, 2021) as to take an extreme example. Delays have always been present and are caused by certain risk factors, be it due to bad weather or to products sitting in storage waiting for transportation.
These problems can be solved at least in part due to big data (McKinsey, 2016).

One of the major contributors to delays is intermodal and/or modal delays. These delays will be discussed more in depth in Chapter 2, Definition and Organization of Supply Chain Models.

Measuring the effect of the contribution of intermodal delays is, however, not an easy task. One of the ways this can be done is by using big data. A dataset can be used to solve this problem, or at least to predict/detect inefficiencies in the transport of goods along several variables, one of which can be modality (Puil, 1994).
One of the tasks that is also important, is to lay the groundwork for further research into this dataset as well as other datasets similar to this one, as there will be future research into this.

Using big data can be a monumental undertaking. In a world where a prominent problem is overly large and complex datasets (Keim et al., 2008), and time is finite, decisions to limit scope must be made. This comes down to focussing only on the modality of the transport, while trying to minimize the impact of the other variables to get a clear picture. While many other variables can be discussed and examined, like political delays in the form of crossing borders or looking at delays tied to political instability, it simply costs too much time to look at all of them.

When narrowed down, there is one major problem that can be solved by big data in this thesis: delays inherent to intermodal transport.
For intermodal transport the inherent challenge is the inherent challenge of changing transport types in transport hubs (Crainic, 2000) (Engler et al., 2018) which has the potential to cause delays on top of the normal delays already inherent with goods transport. The challenge here lies in trying to find all the underlying causes for delays and trying to untangle all of the interlocking problems that pop up from these delays. This comes from needing an extremely high level of coordination between the different transport types where one small mistake in one of the transport links can butterfly out into massive delays in subsequent transport links unless extensive time buffers are used (Hrušovský et al., 2021).

For this thesis, this problem has been represented in a single anonymized data set. This data set contains a large quantity of instances that depict a transport route across multiple transport modes. With this data set we will be looking at the previously mentioned problems of intermodal transport delays and formulate concrete questions to answer which would shed some light on these problems.

1.3 Research Question

Main question:

- For this specific case study we analyze, is there a correlation between the transport type for goods transportation and the severity of the delays of that transport moment?

When starting from the main question, several sub-questions are asked stemming from the main interaction between the transportation of goods and the delays associated with that as well as any factors that correlate or would be responsible for these delays. These questions are visualized in Figure 1 as how they interact with the delays and transportation of goods.

Sub-questions:

- Which different correlations for delay severity exist in this case study?

- How do these correlations describe intermodality and modality of transport delays?

- How do these correlations influence transport delays?

- How can these delays be reduced with a new approach?

The research method applied will be primarily quantitative research, as such an anonymized data set detailing a large quantity of transport routes is going to be used for this research.

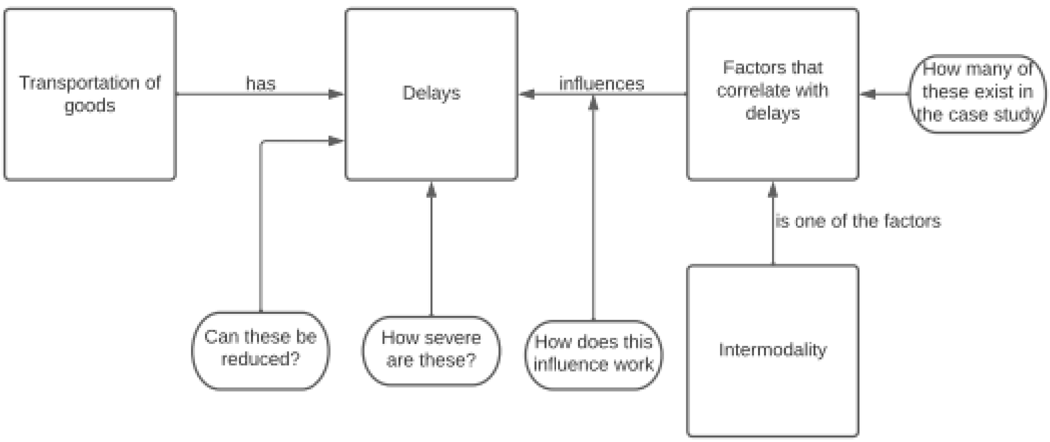

Figure 1: A visualization of the research questions

2 Supply Chains

The data set that is going to be used for this thesis, as was mentioned in the research question, is essentially a link in a supply chain. Because of this a broader overview needs to be given about supply chains to make zooming in on the particular link represented by the research possible.

In this chapter, attention will be given to explaining these supply chains. As this thesis will have a focus on supply chains, it is necessary to provide basic knowledge of these supply chains.

2.1 Definition and Organization of Supply Chain Models

The first thing to do when talking about supply chains is to have a common definition to work with. To use this, the definition used by Beamon (1997) will be used:

A supply chain is an integrated process wherein a number of various business entities (i.e., suppliers, manufacturers, distributors, and retailers) work together in an effort to: (1) acquire raw materials, (2) convert these raw materials into specified final products, and (3) deliver these final products to retailers (Beamon, 1997).

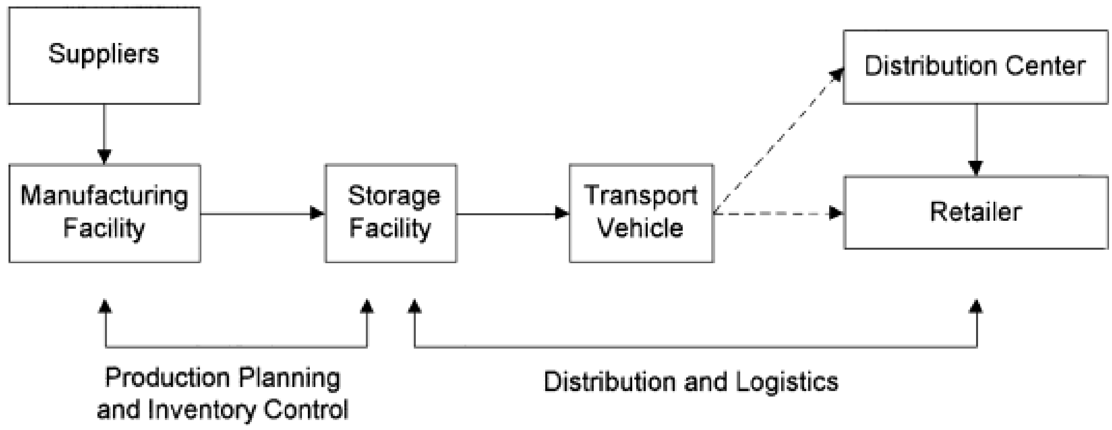

Figure 2: The supply chain process (Beamon, 1997)

An example of a supply chain is Figure 2, which is a five stage supply chain that contains the essential steps of a supply chain and conforms to the earlier given definition. In Figure 2 there are two stages, production planning and inventory control. Production planning describes the manufacture of the goods while inventory control describes how the raw materials, intermediate goods and final goods are stored.

Distribution and logistics governs how the goods are transported to the distribution centers and/or retailers. In this thesis an emphasis will be put on the distribution and logistics. This emphasis is there because the data set itself focuses on the transport vehicle, which means any conclusions and discussion is also going to be focussed on the transport vehicle.

2.2 Supply Chain Disruptions

For this thesis, one of the first questions that was asked was about the already existing correlations between delays in supply chains and supply chain analytics. This was seen in the first research sub-question: Which different correlations for delay severity exist in this case study?.
The way this is to look at disruptions in the supply chain and what their cause is.

Delays have to be caused by a disruption. Which type of disruption causes a delay and how this is handled can be categorized in two parts.
The first part is supply chain resilience, which is the supply chain's ability to be prepared for disruptions.
The second part is the disruption itself, known as supply chain exposures (Prater et al., 2001) or supply chain disruptions (Bode et al., 2013), which represent the different events which can disrupt a supply chain.

Supply chain resilience

The ability for a supply chain to withstand and effectively deal with disruptive events is called supply chain resilience. Supply chain resilience goes hand in hand with supply chain agility.

Supply chain resilience is the ability for a supply chain to deal with a big drop in demand and unexpected risk. Companies do this by, for example, having a high cash reserve or large inventory so that production may continue in case of supply chain disruption.

Supply chain agility is a combination of the speed and flexibility of a supply chain. Speed is the time with which a good is shipped or received and flexibility is measured by the ability to adjust the time it takes to ship or receive a good (Prater et al., 2001). This allows for companies to quickly adapt to shifts in demand.

The recent use of methods like lean and just-in-time has worsened the supply chain resilience significantly (McKinsey Global Institute, 2020). This is because methods like lean and just-in-time rely on an efficient supply chain that constantly produces products they can use and thus, don’t keep a large inventory of intermediate parts.

When the supply chain is disrupted, for example by the COVID-19 pandemic or the Evergreen blocking the Suez Canal, there is not a steady flow of incoming goods. This lack of a buffer in inventory can lead to a big drop in missed sales and/or production when things go wrong, or if there is a sudden spike in sales.

Supply chain risks and vulnerabilities

There are two types of disruptions that cause delays in the supply chain process. Internal and external supply risks (Beamon, 1997) (Prater et al., 2001).
Internal supply risk focuses on the manufacturing base while external supply risk focuses on the transport and logistics of the finished goods from the factory to the store. For the purpose of this text we shall focus on the external supply risks.

Of these external supply risks, there are four main types of supply risk that lead to delays. Intermodal delays, political delays, infrastructure delays and random delays.

Intermodal delays

Intermodal delays are delays that are associated with swapping transport types in a terminal. By swapping types of transport multiple times during the route to the final destination, delays happen. Every additional switch of a good from one type of transport to another causes additional complexity and delays. For goods to swap mode of transportation they need to get transported to a transportation hub, get unloaded off the current mode of transportation, potentially sit in the transportation hub for a time, get loaded on the new mode of transportation and then leave the transportation hub (Engler et al., 2018).

Swapping between transport types compounds the risk of failure, leading to an increasing need to increase reliability. Something can go wrong in any stage of the process of transshipment which reduces the overall reliability of the entire transport process (Puil, 1994).

An example of this is when swapping from truck to airplane. Traffic congestion can cause trucks to be late. This delay can postpone sorting procedures and transferring processes which forces the airplane to either leave without the full load or to be delayed (Lo & Hall, 2008).

Intermodal delays are the main focus for this thesis, figuring out if there is a correlation between the intermodal travel and any delays in this use case and how this correlation describes the intermodality and modality of transport delays as well as how this correlation influences the delays as was discussed in the research questions.

Political delays

Political delays cover a wide variety of bureaucratic delays that come from having to go across country borders and having to conform to the standards of each country that the supply chain travels through.

A country changing their laws can all of a sudden create a situation that needs to be dealt with or laws that require significant change could cause disruptions and delays, countries and the authorities within said countries pose a factor of uncertainty (Bode et al., 2013).
Another problem is going across borders, especially when going across the borders of politically unstable countries. Going across the borders of unstable countries can affect the physical assets, personnel, and operations of foreign firms (Jodice, 1984). Even in countries where political instability is not or less of an issue there is still the problem of border control which costs time.

Infrastructure delays

Infrastructure, both for information and for goods, and its underutilization or underdevelopment can lead to delays (Prater et al., 2001).
Not enough road and rail infrastructure leads to delays in transport. In the Netherlands 10% of the vehicles spent time in congested areas totalling 7% of the total transport cost (Schijndel & Dinwoodie, 2000).

China is an excellent example of this as well, a combination of huge growth in export goods and an infrastructure that is lagging behind its growth. While the ports in China are good and 90% of their exports flow through those ports, the surrounding infrastructure is severely lacking (Zhang & Figliozzi, 2010). The transport network is poorly nationally integrated, both on the road and on rail. There are a lot of internal trade barriers and local protectionism, this makes the national integration of transportation networks harder.
Traffic congestion on the roads is not helped by the lack of a nationally integrated railway network. The railways are responsible for less than 1% of the goods moved through major ports in 2005 (Roth et al., 2008).
China’s warehouse system is also relatively primitive compared to western counterparts (Zhang & Figliozzi, 2010).
All this combined inevitably leads to goods being stuck in traffic or in warehouses which leads to delays.

Random delays

This category of delays deals with “acts of god” like typhoons, earthquakes and tsunamis. These events are beyond our control and the effects are disastrous for the supply chain as production facilities are very susceptible to natural disasters (Prater et al., 2001) (Bode et al., 2013).

2.3 Supply Chain Models

In this chapter there will be a focus on three models that model supply chains and will look at how the models portray delays as well as how the models handle these delays.

There are several studies that try to tackle the problem of delays and create models that try to simulate and reduce the amount of delays that occur in the transport chain. Such models are often limited in scope like the first model from Crainic & Rousseau, (1986) which models a multimodal supply chain network and also formulates the impact of delays and integrates the delays on said model.
The two types of delays it models are intermodal delays and infrastructure delays. The paper tries to optimize it by changing the frequency of use of a transport which in turn changes the intermodal delays. It posits that smaller deliveries lower the cost of intermodal delays but increase the costs incurred by congestion and vice versa. The model itself tries to find the optimal solution between the two factors to reduce the total costs.

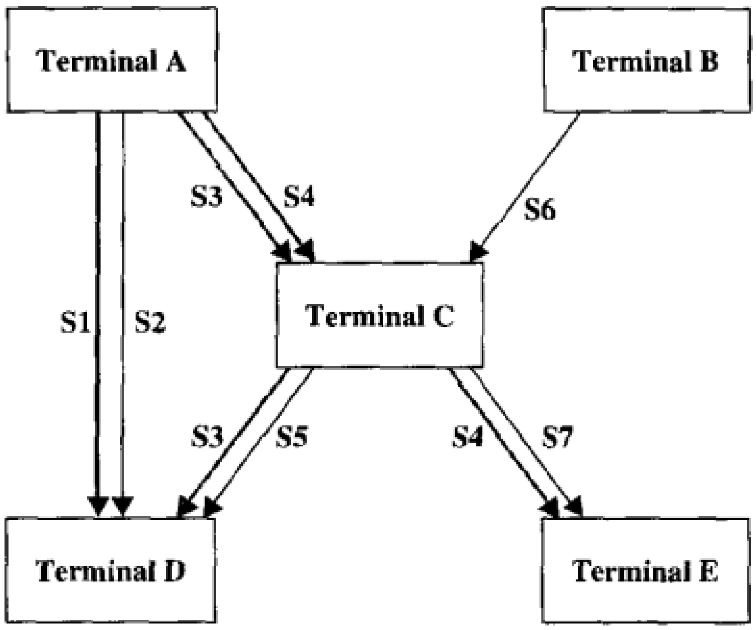

Figure 3: Simplified example of an intermodal service network model (Crainic & Laporte, 1997)

The second of Crainic’s models (Crainic, 2000) designs a service network model. This model has to account for the delays partially mentioned in Chapter 2.2, specifically the ones that are mentioned in this paper are delays, modeling delays as a measure of service quality, mentioning both the tactics used in the intermodal transport hub and the congestion in traffic (Crainic, 2000, 275)
An important factor they also look at with delays is the importance of individual order, some orders are more important than others to get in time and can not afford to be delayed and this needs to be taken into account.

The third and final model (Crainic & Laporte, 1997) has a similar structure to delays as Crainic & Rousseau, (1986) where it creates a service network model as seen in Figure 3. In this model it talks about two main categories of transport, normal direct service (S1, S2, S5, S6 and S7) and services with intermediary stops (S3 which goes from terminal A to terminal D with a stop in terminal C and S4 which goes from terminal A to terminal E with a stop at terminal C).
Within the model the best option for any given transport option, if you want to go from terminal A to terminal D what is your “best option”.
It looks at five issues specifically that are encountered:

- Service network design which looks at the physical route and infrastructure
- Traffic distribution which looks at routing specification for the traffic of each route
- Terminal policies which looks at how well each terminal works and how well traffic is handled there
- Empty balancing which looks at how empty vehicles have to be repositioned within routes
- Crew and motive power scheduling which looks at how well the resources are allocated and which resources are required to make routes run smoothly.

This corroborates the examples from Chapter 2.2, albeit this example has a large focus on infrastructure delays and less so on the other ones (Crainic & Laporte, 1997, 423).

In recent times however, new technologies have forced companies to adapt supply chain models to the new technologies and new knowledge. With AI allowing for robust systems for nonlinear, chaotic environments like supply chains to be more easily predicted or planned for (Fanoodi et al., 2019).

Another advancement since Crainic & Laporte, (1997) is the creation of advanced models with a focus on other factors like sustainability and responsibility like seen in Figure 4 with less emphasis on the factors described in the earlier papers from Crainic (Zimon et al., 2019).

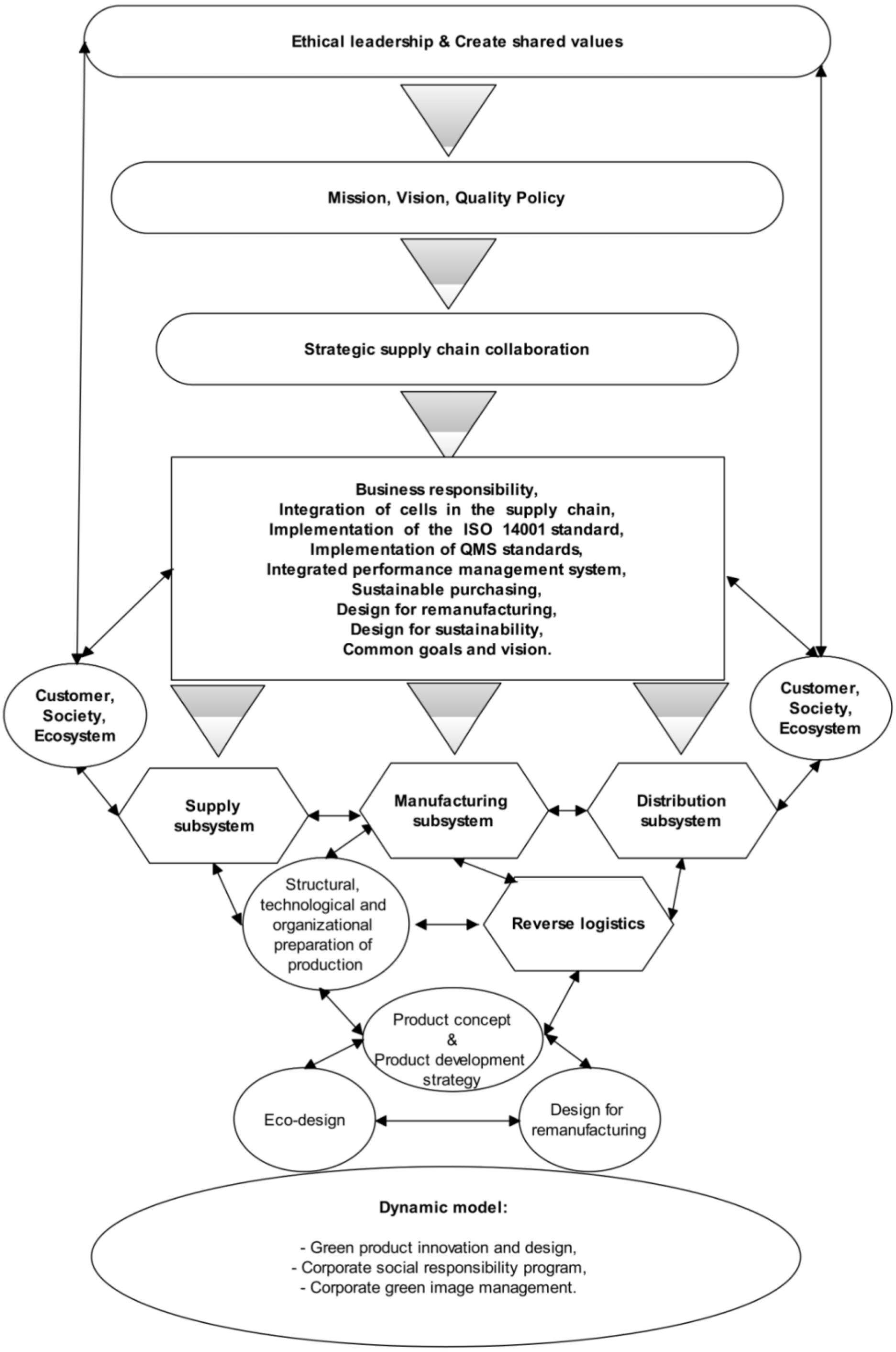

Figure 4: A modern dynamic model focussed on sustainability

2.4 Supply Chain Integration

Supply chain integration is the smooth exchange of information via the use of information and communication systems from multiple stakeholders through one system (MixMove, 2021) (Creative Safety Supply, 2018). This exchange has to both happen within the internal supply chain but also has to happen extensively with the external supply chain, as can be seen in Figure 5.

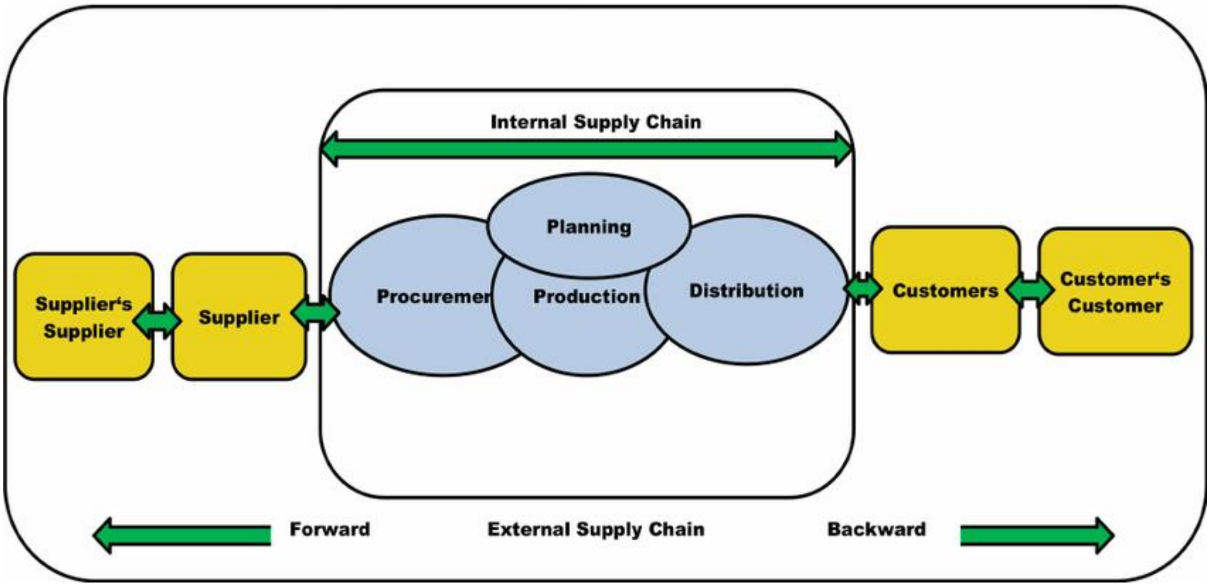

Figure 5: An example of supply chain integration

Supply chain integration is a vital component of making a supply chain run well, but it requires a large amount of data sharing. This type of integration has a strong link with infrastructure delays as it requires a lot of well made data infrastructure to run properly (Gunasekeran & Ngai, 2003).

One of the main components of this type of supply chain integration is the correct use of data. A large amount of data is generated by supply chains, especially when multiple stakeholders are integrated together (IBM, 2021). This can be done by integrating certain parts of information systems close together or even merging entire systems for ease of use. All of the data that is generated by these types of systems need to be handled and analyzed properly using big data analytics (McKinsey, 2016).

How to properly process this data and how to use and present all of the data associated with these supply chains will be discussed in the next chapter.

3 Supply Chain Data Analytics

Using data that is generated from the supply chain and data generated by the transport of goods is a part of the larger subject of data analytics. The larger subject of data analytics will be discussed here.
The definition of big data supply chain analytics will be discussed. After this, there will be a literature review of data modeling, data visualization and visual analytics.

3.1 Definition

There are several definitions for data analytics and big data, (Gartner) has the following definition for data analytics.

Data and analytics is the management of data for all uses (operational and analytical) and the analysis of data to drive business processes and improve business outcomes through more effective decision making and enhanced customer experiences (Gartner).

The National Institute of Standards and Technology has several definitions concerning data science and data analytics that give more insight as well. Specifically their definition for data science and analytics (NIST, 2015)

- Data science is the extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing

- The analytics process is the synthesis of knowledge from information

Another way to define Big Data is the three V's:

- Volume, the amount of data generated

- Velocity**,** the speed at which the data is generated

- Variety, all the structured and unstructured data that has the possibility of getting generated

A set of data needs all three to be present, high-volume, high-velocity and high-variety, to be considered big data.

Like seen in Figure 6, there are varying degrees of being big data.

Big data is made much bigger by its diversity. Big data may come from a number of places and can be classified into three categories: structured, semistructured and unstructured. Unstructured data is random and difficult to examine, whereas structured data enters a data warehouse fully labeled and quickly sorted. Semistructured data contains tags to distinguish data elements rather than fixed fields (Eaton & Zikopoulos, 2011) (Sagiroglu & Sinanc, 2013).

The size of data is increasingly measured in terabytes and petabytes. Traditional data storage and processing approaches are outstripped by the vast quantity and growth of data (Eaton & Zikopoulos, 2011) (Sagiroglu & Sinanc, 2013).

Not only for large data, but for all operations, velocity is necessary. To maximize the value of big data, information should be used as it flows into the organization for time-limited tasks (Eaton & Zikopoulos, 2011) (Sagiroglu & Sinanc, 2013).

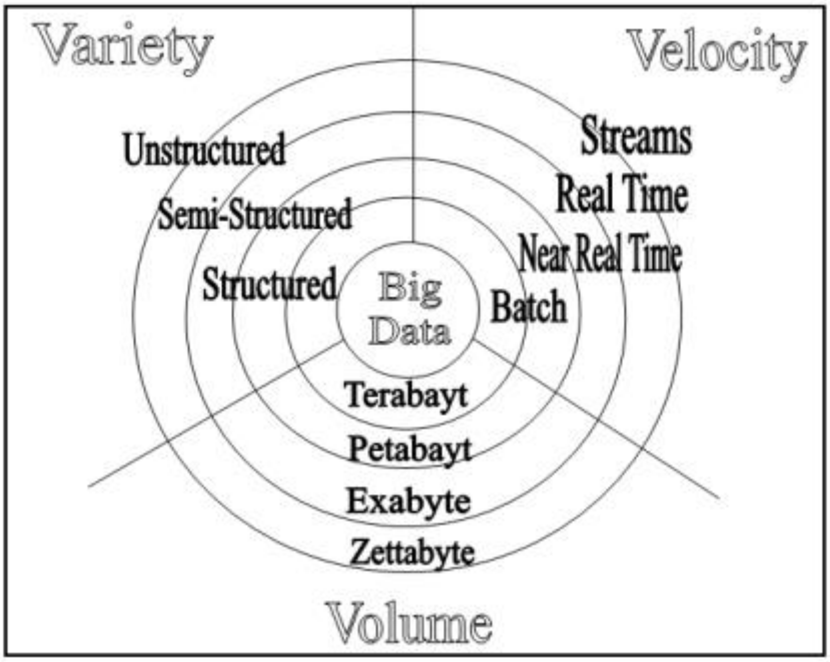

Figure 6: The degrees of Variety, Velocity and Volume (Sagiroglu & Sinanc, 2013)

3.2 Data Modeling with Supply Chains

Data modeling is the process of creating a data model from the gathered information in Chapter 3.1, for an information system by applying certain formal techniques. For this thesis we use data modeling to create a data model that can turn the structured data we found into workable knowledge.

Data modeling is divided up into three categories: Conceptual, Logical, and Physical data models. A data model has to go through all three steps for a comprehensive data model to be created as seen in Figure 7.

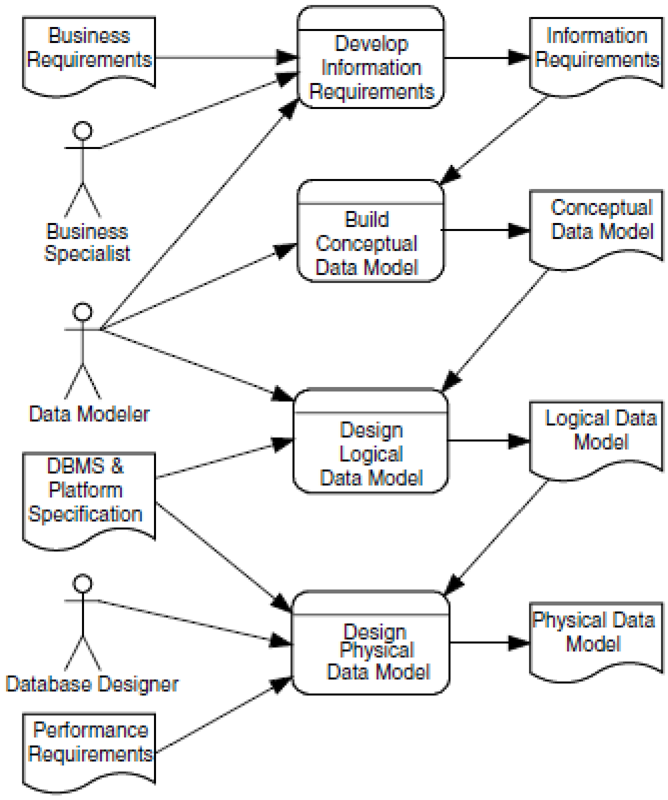

Figure 7: Different steps in data modeling (Simsion & Witt, 2005, 16-17)

Conceptual data model

The conceptual data model is the first stage of the modeling process. The conceptual data model is a representation of the data needed to support business processes. It also keeps up with business events and the metrics that go along with them.
What the system contains is defined by the conceptual model. Instead of focusing on processing flow, this type of data modeling focuses on locating the data used in a business. This data model's primary goal is to organize and establish business rules and concepts. It enables people without a background in data analysis to access any data, such as market data, customer data, and purchase data, for example.

Logical data model

The logical data model is the second stage of the modeling process.
The map of rules and data structures in the logical data model comprises the data required, such as tables, columns, etc. The logical model is created by data architects and business analysts. The logical model can be used to convert it to a database. This data model serves as a foundation for the physical model. There is no secondary or main key defined in this model.

Physical data model

The physical data model is the third stage of the modeling process.
The implementation of a physical data model is discussed using a specific database system. It specifies all of the components and services required to construct a database. The database language and queries are used to create it. Each table, column, and constraint such as primary key, foreign key, not NULL, etc. are represented in the physical data model. The physical data model's principal task is to establish a database. The Database Administrator (DBA) and developers design this model. This type of Data Modelling aids in the creation of the schema by abstracting the databases. This model specifies the data model's specific implementation. We can have database column keys, restrictions, and RDBMS functionality thanks to the physical data model.

3.3 Forms of Data Analytics

There are five separate types of data analytics. Which type is used depends on the type of data being used.

Text analytics

Text analytics (text mining) is a term that describes ways for extracting information from textual data from a variety of sources. Statistical analysis, computational linguistics, and machine learning are all used in text analytics. Large volumes of human-generated text may be converted into relevant summaries using text analytics, allowing organizations to make evidence-based decisions (Gandom & Haider, 2015).

Audio analytics

Unstructured audio data is analyzed and information is extracted using audio analytics. Audio analytics is also known as speech analytics when applied to human spoken language. Because these approaches are mostly utilized with spoken audio, the phrases audio analytics and speech analytics are frequently interchanged. Audio analytics is now used mostly in consumer call centers and healthcare (Gandom & Haider, 2015).

Video analytics

Video analytics, also known as video content analysis (VCA), is a collection of techniques for monitoring, analyzing, and extracting useful data from video streams. This type of analytics is in its infancy compared to the other types of data mining. Various ways for processing real-time and pre-recorded videos have previously been established. Closed-circuit television (CCTV) cameras are becoming more common as well as the increase of popularity of video-sharing websites. These two factors are driving the expansion of automated video analysis (Gandom & Haider, 2015).

Social media analytics

The study of structured and unstructured data from social media platforms is referred to as social media analytics. The phrase "social media" refers to “a multitude of internet platforms that allow users to produce and share information” (Gandom & Haider, 2015).

Predictive analytics

Predictive analytics refers to a set of methodologies that use historical and present data to forecast future results. Predictive analytics may be used in practically any field, from anticipating the failure of jet engines based on data from thousands of sensors to forecasting consumers' future actions based on what they purchase, when they buy, and even what they post on social media.
Predictive analytics is primarily concerned with uncovering patterns and capturing correlations in data. There are two types of predictive analytics methodologies. Moving averages, for example, aim to find past trends in the outcome variable(s) and extrapolate them to the future. Others, like linear regression, attempt to capture and utilize the interdependencies between outcome variables and explanatory factors in order to generate predictions. Methods can also be divided into two areas based on their underlying methodology: regression approaches (e.g., multinomial logit models) and machine learning techniques (e.g., neural networks). Another classification is based on the kind of result variables: linear regression techniques handle continuous outcome variables (e.g., home selling prices), whereas Random Forests approaches address discrete outcome variables (e.g., credit status) (Gandom & Haider, 2015).

Predictive analytics is the one that is primarily going to be used in this thesis, as well as a smaller amount of textual analytics. This is because the primary purpose of this thesis is to identify and predict failure points in the transportation of goods.
The way this thesis uses predictive analytics in uncovering and capturing correlations in data is to try and capture a correlation between transport types used and delays in freight transport.

3.4 Visual Analytics in supply chain networks

Visual analytics is an important process that is used to turn data that is hard to understand without experience from the relevant field into more understandable knowledge. It is going to be used extensively in the results of this thesis to visualize the correlations between delays and intermodality in the use case, as was discussed in the research questions, so it will be briefly explained in this chapter.

Definition

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces, according to Thomas & Cook, (2005).

Visual analytics tools and techniques are used to synthesize information and generate insight from enormous, dynamic, ambiguous, and frequently contradicting data; to identify the anticipated and find the unexpected; and to detect the expected and discover the unexpected. offer evaluations that are timely, defensible, and clear; and effectively convey assessments for action.
Four of the main techniques of visual analytics that were identified are: (Thomas & Cook, 2005).

- Analytical reasoning techniques that let users obtain deep insights that directly support assessment, planning, and decision making;

- Visual representations and interaction techniques that exploit the human eye's broad bandwidth pathway into the mind to let users see, explore, and understand large amounts of information simultaneously;

- Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis;

- Techniques to support production, presentation, and dissemination of analytical results to communicate information in the appropriate context to a variety of audiences.

These techniques make it easier to look at and represent large quantities of data like are found in this thesis, when trying to look for correlations in a data set of thousands efforts must be made to make it easier which these techniques provide.

Goals of visual analytics in supply chain networks

As was discussed in the chapter about data analytics and Big Data, the amount of data that is being generated is increasing quickly and is becoming harder and harder to keep up with. This is known as the information overload problem (Keim et al., 2008). One way to increase our understanding of this Big Data is to use visual analytics, which is a vital step in the process of turning data into knowledge as shown in Figure 8**.**

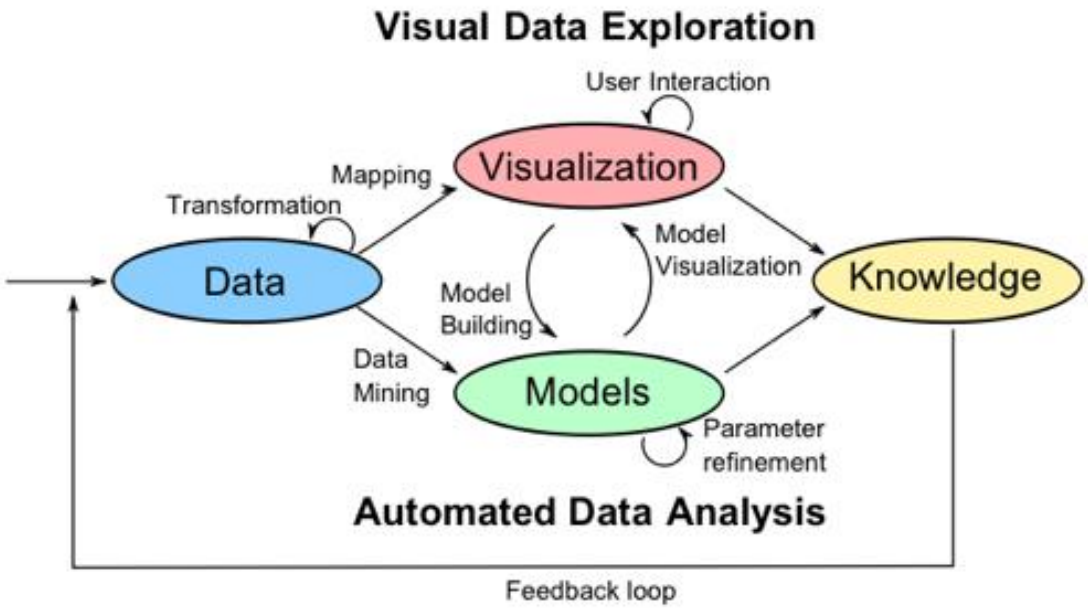

Figure 8: the Visual Data Exploration feedback loop

What Figure 8 shows us, and what Keim et al. (2008) tells us, is that visualization is an important tool for data mining and data analytics. Adding visualization to the data and models allows for a feedback loop to be created where the visualization allows for more people, even those without intimate knowledge of the data set, to interact with the data set without requiring extensive training in the field of computer science and data analytics.

How the ever increasing demand for visual analytics manifests itself specifically for supply chain networks, like described in Park et al. (2016), is in two important ways: the fact that humans struggle with processing large quantities of data and that in recent years the amount of data being used in supply chain networks has exponentially increased as seen in Figure 9. Visual analytics can solve this problem for supply chain analytics by augmenting the human's ability to visually process and examine large-scale data.

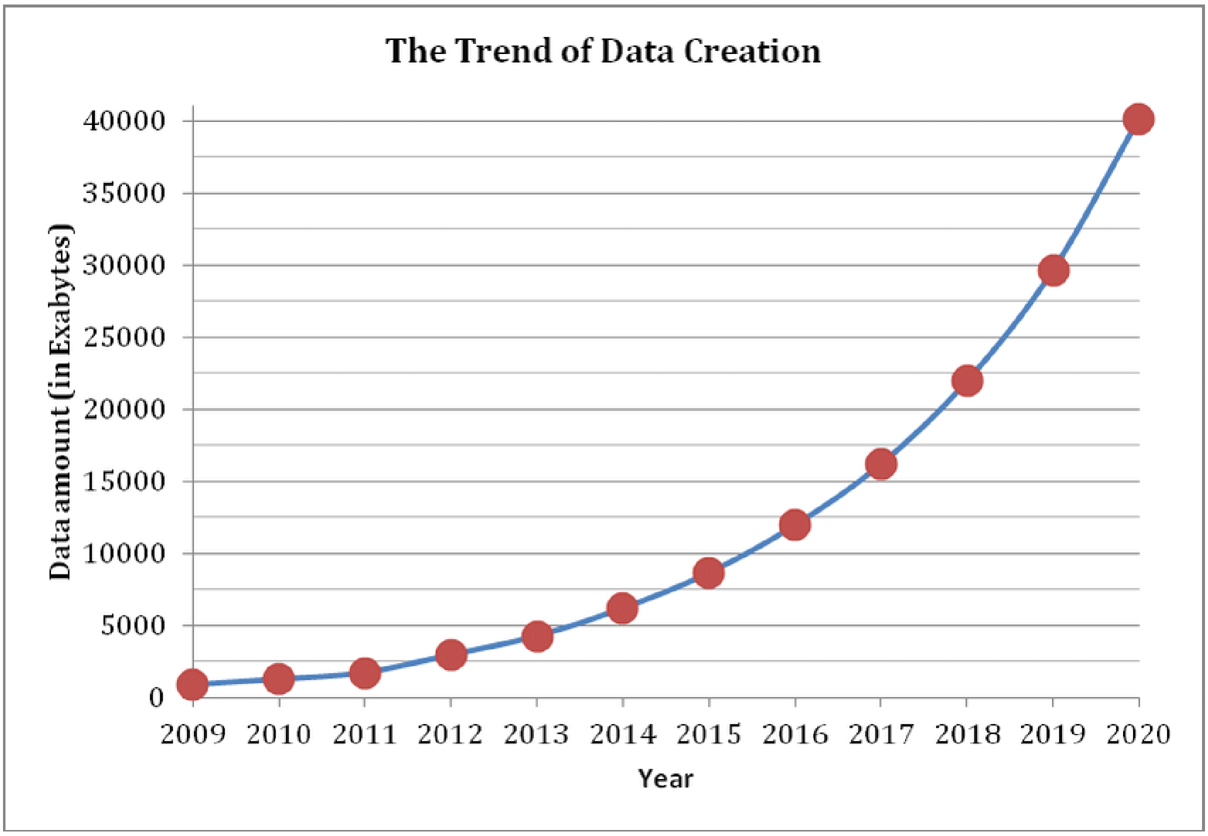

Figure 9: The trend of global data creation

Data visualization will be used extensively in Chapter 4 and Chapter 5. Where it will be used to take a closer and more detailed look at the big data set and results and to make the large quantity of data more understandable.

4 The Case Study

The problem that needs to be looked at is, as stated in the problem statement and the research questions, the impact of transport modes on the time it takes to transport goods and any correlations between said transport of goods and other factors like intermodality. For this an anonymized data set has been used.

In this chapter we will be characterizing the data set, what is in the data set, which parts of the data set that will be used and how these parts are going to be used.

4.1 Data Set Description

The data set describes a large quantity of transit routes of goods. There are 8 separate data points per data entry. They are all depicted together in Table 1 which depicts a singular data entry.

- ID: A unique ID to identify the route
- Route: contains all the transport moments belonging to the route
- Departure: Location where the transport moment departs from
- Destination: Location where the transport moment arrives at
- Container type: What container type is used
- Contents: the anonymized contents, empty due to anonymization
- Sequence: what transport moment this is in the route
- Market: Country/market of the final destination
- Start Time: the time and day the good leaves it’s departure point
- Arrival Time: the time and day the good arrives at its destination

ID	Departure	Container type	Sequence	Start Time
3VW5Q80G8A2208	RTRDM	20” standard	2	17-05-2019 5:12:00 PM
Route	**Destination**	**Contents**	**Market**	**Arrival time**
MNCH-RTRDM-RTRDM-TKYO	TKY	-	JPN	19-05-2019 11:28:00 AM

Table 1: Data entry example

Data entry

Each data entry in the data set represents a transport moment, each transport moment represents a singular movement of a good or a set of goods from the departure point to the destination point. These transport moments make up routes, each route consists of at least one transport moment and the route entry. For example the data entry in Table 1 would have another data entry that is extremely similar, with an identical ID, an identical route but would represent the transport moment of MNCH-RTRDM, as can be seen in Table 2.

This transport moment has a sequence field with a value of 1, instead of the value of 2 for the one in Table 1. The sequence indicates what part of the route the transport moment is, if it’s the first transport moment of the route like Table 1 the sequence will be 1, if it’s the second transport moment of the route it will be a 2, if it is the third it will be three etc.

The remaining parts of the data entry start time and arrival time, which represent the time it leaves its departure point and arrives at its destination respectively for this particular transport moment.

ID	Departure	Container type	Sequence	Start Time
3VW5Q80G8A2208	MNCH	20” standard	1	15-05-2019 2:00:00 PM
Route	**Destination**	**Contents**	**Market**	**Arrival time**
MNCH-RTRDM-RTRDM-TKYO	RTRDM	-	JPN	16-05-2019 6:35:30 PM

Table 2: Data entry example 2

This type of dataset is comparable to other data sets, with the ID, starting time and location, and arrival time and location all being found in other data sets (Brunel University, 2019). This type of data set allows for easy extraction of data about the routes that the supply takes and allows for detailed questions to be answered about the nature of the routes.

4.2 Data Set Methodology

Now that what was in the data set was discussed, the first question that needs to be answered is: What is going to be characterized from this data set?
There are two main categories that are going to be measured:

- Travel time per route
- Waiting time per route

The travel time per route is measured by:
- Taking all the transport moments in a route
- Calculating the travel time for each transport moment in a route (Arrival time - Start time)
- Adding all the transport times together

When all the times are gathered, the average travel time of a route can be calculated and this average can then be used to determine the amount of travel delays that are present in the route.

The waiting time per route is measured by:
- Taking all the transport moments in a route
- Calculating the waiting time between the transport moments, this is done by subtracting the arrival time from the next transport moment in the sequence from the start time
- Adding all the waiting time together

Waiting time is a little bit more complicated than the travel time: partly because there are several more edge cases we need to worry about, like routes existing out of only one transport moment, as well as having to calculate the waiting time using multiple transport moments. After we have gathered the waiting times, the average waiting time can be calculated and the amount of delays pertaining to waiting times that are present in a route can be calculated.
There are several reasons that the travel and waiting time per route were used, and not the travel and waiting time per transport moment. The first reason is that there is a potential for a butterfly effect in transport networks (Jing & Liao, 2017). A delay in a

transport moment with a lower sequence might have an effect on later transport moments. If we only look at individual transport moments, this effect would not be able to be captured, as certain transport moments occur in multiple routes.

This effect also applies to waiting times, a long waiting time/travel time can lead to even longer waiting times. An example of this is that if, on certain routes, the goods are too late for a ship, the ship will not wait for the goods. It will leave at the time it's supposed to leave. This can lead to ever more increasing waiting times as one transport time that takes a bit too long can snowball into longer waiting times. This is why we have to look at the entirety of the route instead of looking at individual transport moments.

The second part is to determine what part of the dataset is going to be used. There are several routes that don't occur as much while some routes occur in the data set hundreds or even thousands of times as seen in Figure 10. In the end it was decided on to look at each route that occurred over 100 times.

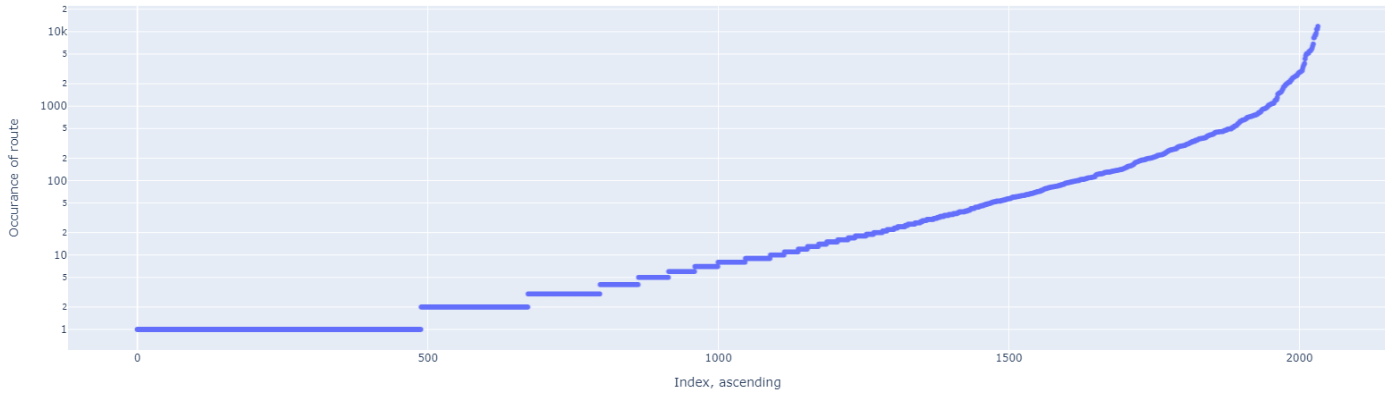

Figure 10: a representation of how often a route occurs in the data, each dot represents a route. The X axis represents the amount of routes, ordered by how often it occurs ascendingly and the Y axis is how often each route occurs.

Measurements development

Now that we have the part of the dataset that is going to be discussed, we need to discuss what part of that dataset is relevant to this examination.

It was decided early on that, due to the nature of the problem, the first focus was going to be on routes that had overly long delays, which was a rather vague concept. To sharpen this concept both “overly long” needs to be defined and “delay” needs to be defined.

First, for this study we narrowed our definition of delay, in this particular case a route is delayed when the waiting time or travel time is longer than average. This is because the focus of the research is on goods arriving late, as discussed in the chapter on scope. We do not care for small delays, the only delays we care about are overly long delays which will be measured primarily in hours.

Secondly, in this particular case for “overly long” it was decided that anything that had a delay larger than 1 standard deviation over the average. This industry allows for enough routes to be selected to give us a clear picture of what is wrong, but also provides us with a statistically significant result.

These definitions together form the part of the set that is classified as “problematic routes'', namely any transport moment where the average delay is one standard deviation above average. Using this measurement and the routes that fall within this measurement allow for conclusions to be drawn about the interaction between transport type and delays. But first this measurement needs to be extracted from the raw data set which will be done in the next chapter.

5 The Model

With the part of the data set that is going to be worked on properly defined, we need to turn the big data of the use case into usable knowledge. To do this we need to create a model.

Using Figure 11, there are 4 products that need to be delivered: the information requirements, the conceptual data model, the logical data model and the physical data model.

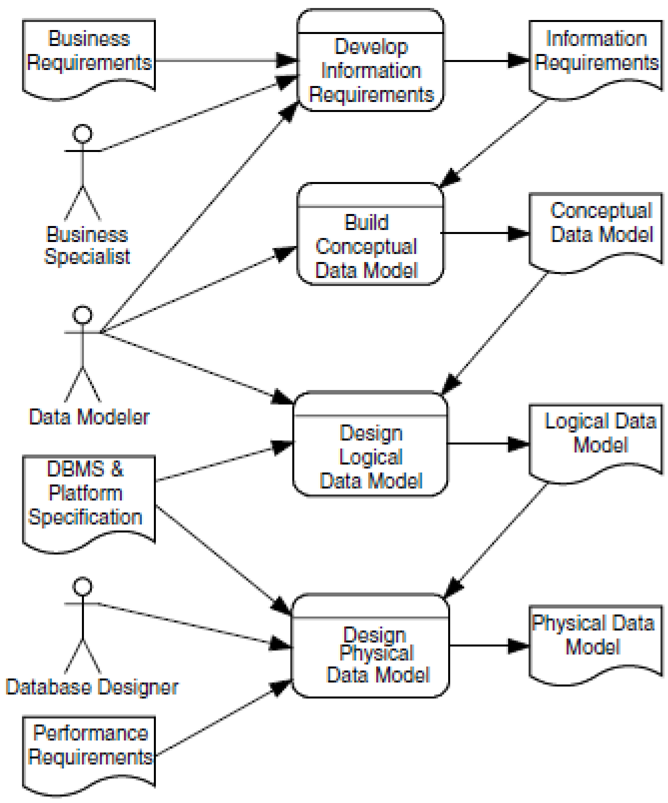

Figure 11: Different steps in data modeling (Simsion & Witt, 2005, 16-17)

The requirements for the thesis have already been discussed in earlier chapters, specifically in the chapter about the Data Set as well as in the Introduction. In this chapter the conceptual data model, the logical data model and the physical data model will be discussed, starting with a recap of what the model actually represents.

Conceptual data model

The conceptual data model is the first stage of the modeling process. The conceptual data model is a representation of the data needed to support business processes.
What the system contains is defined by the conceptual model. Instead of focusing on processing flow, this type of data modeling focuses on locating the data used in a business. This data model's primary goal is to organize and establish business rules and concepts. It enables people without a background in data analysis to access any data.

Logical data model

The logical data model is the second stage of the modeling process.
The map of rules and data structures in the logical data model comprises the data required, such as tables, columns, etc. The logical model is created by data architects and business analysts. The logical model can be used to convert a data set to a database. This data model serves as a foundation for the physical model. There is no secondary or main key defined in this model.

Physical data model

The physical data model is the third and final stage of the modeling process.
The implementation of a physical data model is discussed using a specific database system. It specifies all of the components and services required to construct a database. The database language and queries are used to create it. Each table, column, and constraint such as primary key, foreign key, not NULL, etc. are represented in the physical data model. The physical data model's principal task is to establish a database.

The three models for the conceptual, logical and physical data model are based on eachother, only the final model, the physical data model, will be used. The others are there to shape the design process as well as show the process used to attain the final model. the physical data model is there to model the process used to turn the raw data set into the final refined set of problematic routes from which we can extract conclusions about the correlation between transport type and delays
These three models will be made using UML.

5.1 Conceptual Data Model

A conceptual data model consists of a rough model of the process that was discussed at the start of the thesis. This will be a high level model that was built with the information that was available at the start.

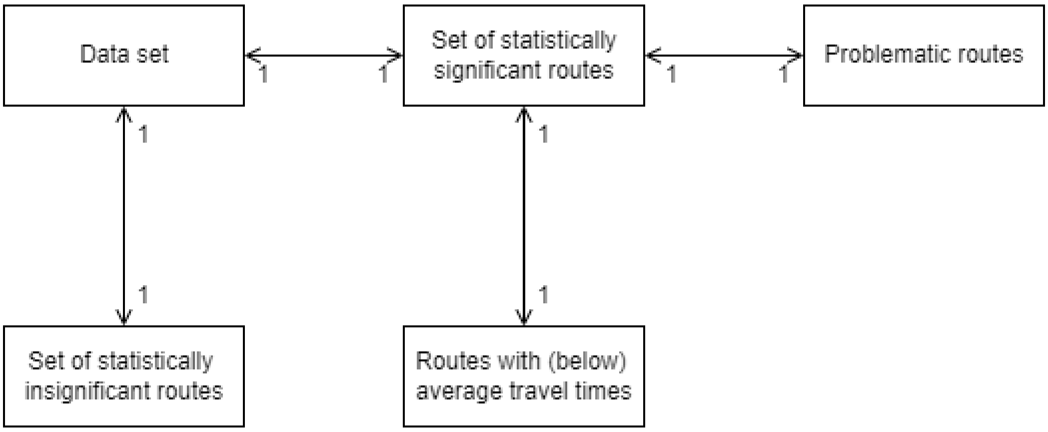

Figure 12: Conceptual data model as developed in Chapter 3.3

The original model in Figure 12 represents the conceptual data model. This model has two major steps: calculate the set of statistically significant routes and determine the problematic routes, both of these were defined in chapter. These steps have two byproducts that will not be used, the set of statistically insignificant routes and the unproblematic routes.

List of variables

Here is a legend of the variables from Figure 13 and Figure 14 with a short description of each variable used.

Data Set:

- ID: Unique identifier of the item in the route
- Departure: Place of departure for the transport moment
- Destination: Destination for the transport route
- Route: Full route the transport moment is a part off
- Market: country and or market of the destination
- Sequence: Depicts which part of the route this transport moment is
- Start time: Time and date of when the transport moment started
- Arrival time: Time and date of when the transport moment ended
- Transport type/mot_name: Type of transport the transport moment uses

All times of routes:

- Route: Full route the transport moment is a part off
- ID: Unique identifier of the item in the route
- Route: Full route the transport moment is a part off
- Start time: Time and date of when the transport moment started
- Transport time: The amount of time it takes for the transport moment to complete

Average waiting/travel times per route:
One file contains the average waiting time for every route and another file contains the average travel time for every route

Average deviation for travel time per route:
One file contains the deviation from the average waiting time for every route and another file contains the deviation from the average travel time for every route

Waiting/travel time deviations per route:

- ID: Unique identifier of the item in the route
- Travel/waiting time: Contains how long a travel/waiting moment took in hours
- Start time: Time and date of when the transport moment started
- Deviation from average travel time in percent: The time a travel moment deviated from the average in percent
- Deviation from median travel time in percent: The time a travel moment deviated from the median in percent
- Deviation from average waiting time in hours: The time a waiting moment deviated from the average in percent
- Deviation from median waiting time in hours: The time a travel moment deviated from the median in percent

Problematic route with waiting/travel times:

- Route: Full route the transport moment is a part off
- Sequence: Depicts which part of the route this transport moment is
- Departure: Place of departure for the transport moment
- Destination: Destination for the transport route
- Market: Country and or market of the destination
- Transport type & mot_name: Type of transport the transport moment uses

5.2 Logical Data Model

In this case, the logical data model is there to represent the actual way the process is done, how the data is transformed into actual insight. This model shows a simple version of the way the code is structured, but is significantly more detailed compared to the conceptual data model.

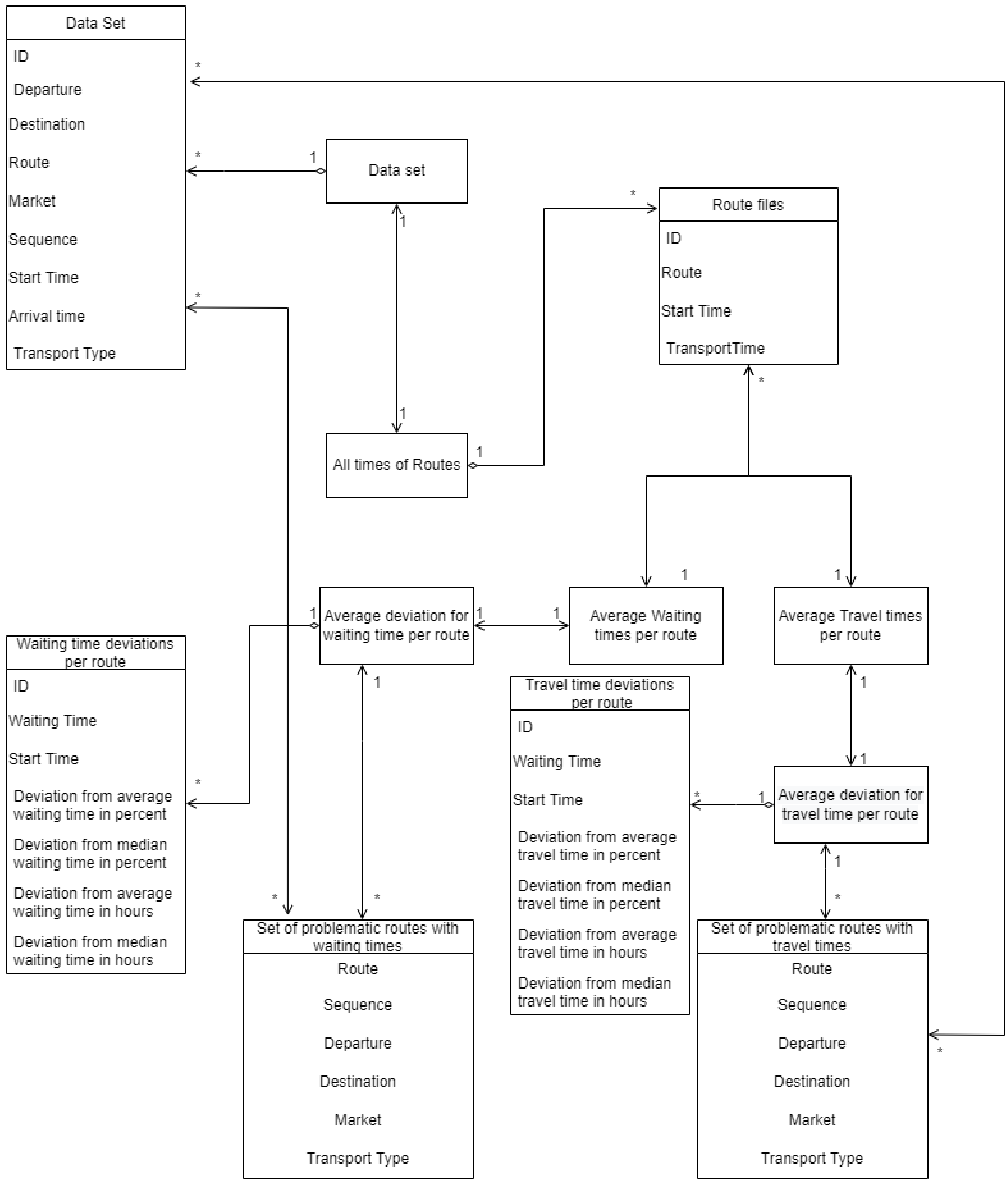

Figure 13: The Logical data model as developed in Chapter 3.3

The model above in Figure 13 is a representation of the states of data implemented in the code. As discussed in the chapter on the conceptual data model, there are broadly 3 parts. These parts are the data set, the set of statistically significant routes and the problematic routes. In each part it is explained to which part each box in Figure 13 belongs to.

Data set

The data set itself, which is described by the two data set fields as depicted in Figure 13. The data set was described in Chapter 4 about the data set, as seen in Table 1. The only change in the actual model is that the transport type was provided. This transport type is either Ship, Train, Truck or Plane. Out of these three, planes have an extremely low occurrence at 0.267% so will be ignored because it’s not statistically relevant.

Set of statistically significant routes

The set of statistically significant routes, any route which occurs over 100 times, is represented by the route files and the files of average waiting/travel times per route as depicted in Figure 13.
The route files are a set of intermediary files. Each file directly correlates with a route and contains every time a route occurs in the data set together with the ID, Route, start time and travel/waiting time of each occurrence of that route.
This is done for two reasons:

- Information about individual routes is more easily available for closer study

- The information about individual routes is easier to use when condensed in singular files.

The files of average waiting/travel times are a file that contains the average travel/waiting times in hours for each and every route. This allows us to calculate the average delay for each transport moment.

Problematic routes

The problematic routes are represented in the model by the average deviation for waiting/travel time per route and the problematic routes with waiting/travel times as seen in Figure 13.

From the waiting and travel time averages, the deviations (or delays) are calculated. These delays form the basis of the results and from these delays we can calculate the problematic routes as they were defined in Chapter 3, a route that has an average delay greater than one standard deviation.
The exact way that the problematic routes are calculated will be discussed in the next chapter.

5.3 Physical Data Model

The physical data model as seen in Figure 14 is very similar to the logical data model but shows a more indepth version of the logical model. Data types are specified and variable names are used.

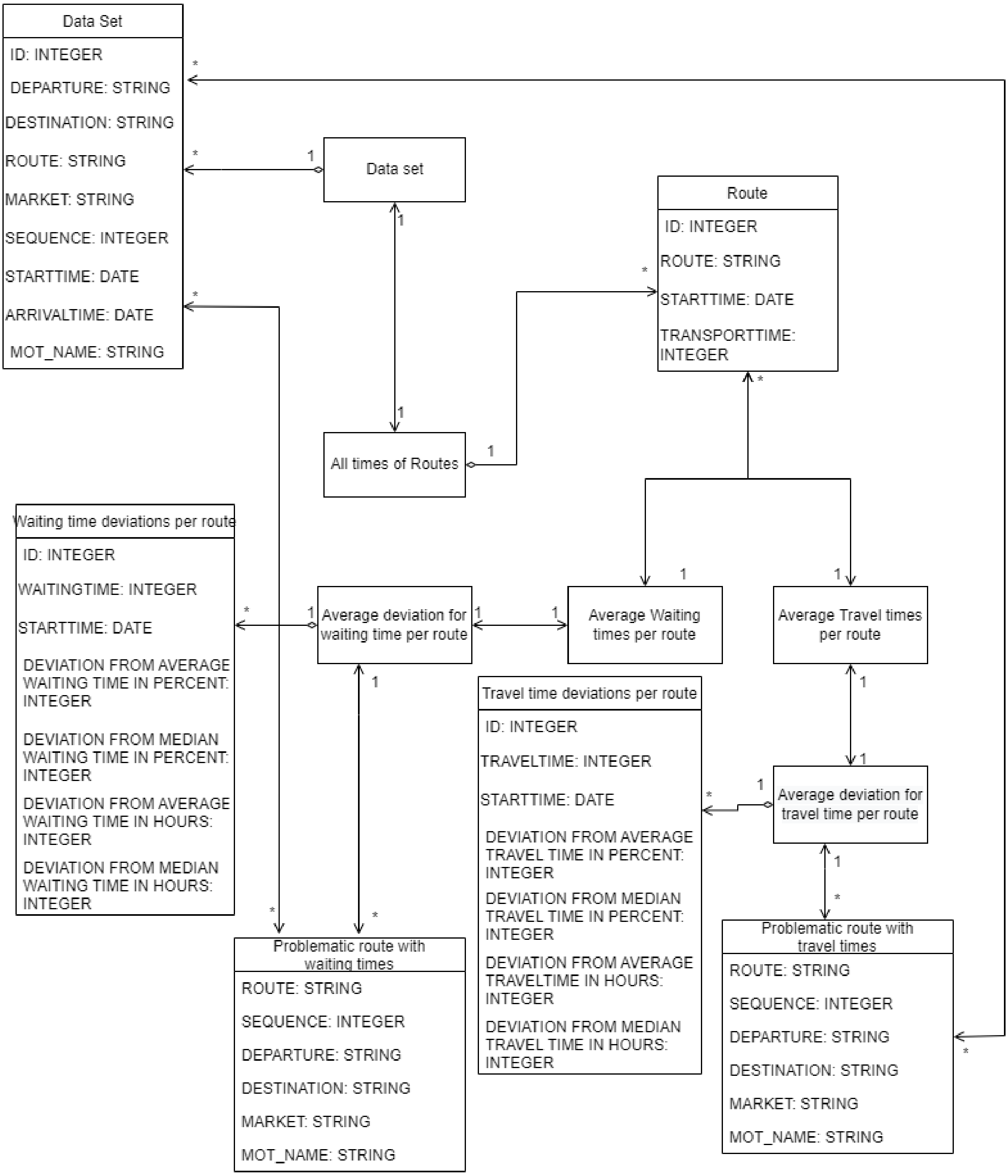

Figure 14: Physical data model as developed in Chapter 3.3

The implementation of the model

A better way to describe the physical model is by describing how the model was implemented using code and which code was written and which libraries were used.

The physical data model was created using Python, using Jupyter Notebook.
The libraries that were used are:

- Pandas was used to allow Python to read in and store the data in a dataframe that is easy to interact with

- Plotly was used for the visualization of the results

- Numpy was used for fast and efficient computations with arrays

- Datetime and Time was used to simplify calculation with date and timestamps

The first step in implementing the code is to divide the data up into a set of files. Each file in this set contains all the entries of the data belonging to a certain route. This step makes it easier to look at and work with individual routes.

With those routes now defined better and the data in workable form to work with, the data set is now able to be split into two parts: The waiting times, which is the time spent in between transport moments not traveling and travel times, the time spent on a transport going from point A to point B. Waiting times represent the time spent between transport moments while travel time represents the time spent in those transport moments. From this point onward, the files representing the waiting time and the files representing the travel times are going to be looked at separately.

Calculating travel time and calculating waiting time is a matter of using Datetime to subtract the time of arrival from the time of departure of the transport moment for travel time and for waiting time subtract the departure time from the arrival time of the next route.

For both the travel time and the waiting time, the average time for each route was then calculated. Using the average for each route, the deviation of travel time and waiting time for each instance of that route was calculated.

This was originally done using four different metrics across three categories. These categories were used to determine which of the metrics produce the best results. In the end only one of the metrics and one of the categories will be chosen.

The first category is simply adding all the deviations together, both negative deviation and positive deviation as visualized in Figure 15. This means that instances of a route that are faster compensate for the instances of routes that are slower. This would be a good way to measure if the preferred outcome values faster routes as well and values inconsistency less.

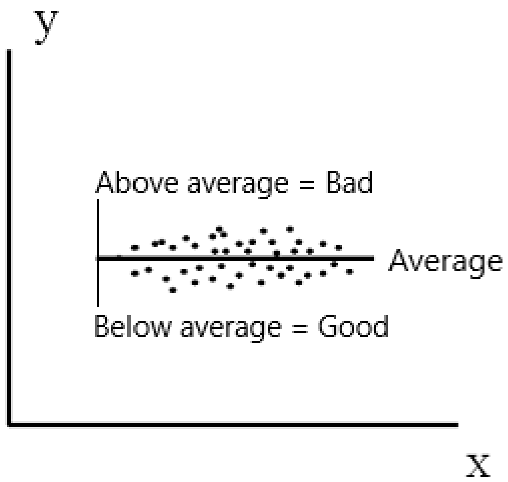

Figure 15: A visual representation of the first category of deviations

The second category was calculated by taking all the absolute deviations and adding them together. This was done under the assumption that every deviation from the average is bad as visualized in Figure 16, it can be more beneficial to have a smaller overall deviation from the average over a lower average time.

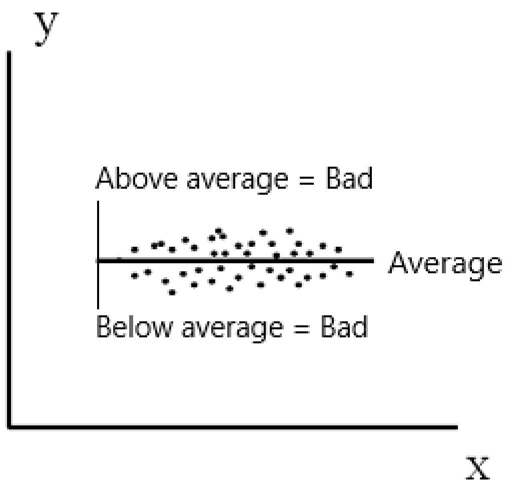

Figure 16: A visual representation of the second category of deviations

The third category, which is the one we will use, is looking at only routes that are slower than average, not at the ones faster than average as visualized in Figure 17. This is because one of the goals of this paper was to reduce the amount of delays and this goal is more easily achievable when looking at only the delays. This represents the set of problematic routes as depicted in Figure 13 and Figure 14 as well as was discussed in Chapter 5.2.

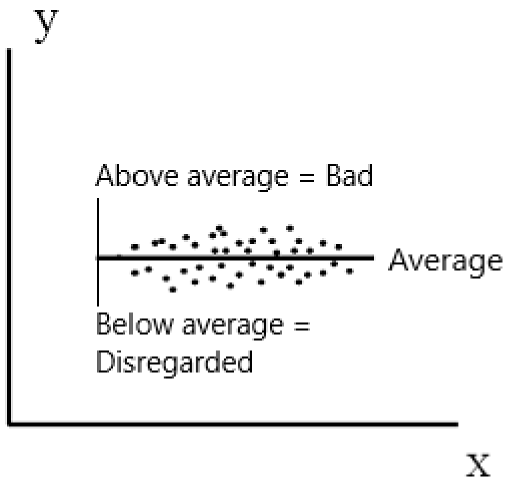

Figure 17: A visual representation of the third category of deviations

The four metrics are as follows:

- Standard deviation from travel/waiting time in hours
- Standard deviation from travel/waiting time in percent
- Average deviation from median travel/waiting time in hours
- Average deviation from median travel/waiting time in percent

In the end “Standard deviation from travel/waiting time in hours” was chosen as the metric that was going to be used for determining if a route was a problematic route. Standard deviation from travel/waiting time was chosen over average deviation from median travel/waiting time because of the chosen method of determining the problematic routes. Standard deviation selection like this is common in statistical analysis (Lee et al., 2015), and as such was chosen for this one.

Hours were chosen over percentage because when percent was used the routes that were deemed problematic all had extremely low times total travel/waiting, with most routes having under an hour of delay for those routes. This meant that problematic routes had, with percentages used, a relatively minor impact when looking at the whole set.
When using hours instead of percent, the routes that were selected as problematic routes were longer which is why they were chosen.

The final product is also depicted in Figure 14 as “problematic route with travel/waiting times”. The only difference between waiting and travel times is that, of course, waiting times don’t have a transport type. The two variables that are used instead are the transport type of the incoming transport and the transport type of the outgoing transport type.

ID	Departure	Container type	Sequence	Start Time
3VW5Q80G8A2208	RTRDM	20” standard	2	17-05-2019 5:12:00 PM
Route	**Destination**	**Contents**	**Market**	**Arrival time**
MNCH-RTRDM-RTRDM-TKYO	TKY	-	JPN	19-05-2019 11:28:00 AM

Table 3: Data entry example 1

ID	Departure	Container type	Sequence	Start Time
3VW5Q80G8A2208	MNCH	20” standard	1	15-05-2019 2:00:00 PM
Route	**Destination**	**Contents**	**Market**	**Arrival time**
MNCH-RTRDM-RTRDM-TKYO	RTRDM	-	JPN	16-05-2019 6:35:30 PM

Table 4: Data entry example 2

Using Table 3 and 4 as examples for two travel instances in a route, with Table 3 and 4 being sequence 1 and 2 respectively. The waiting time in this particular case would be roughly 2 days and 17 hours. The incoming transport type of the waiting time would be Train and the outgoing transport type would be Boat.
Using this you get your two sets of outbound waiting time and inbound waiting time.

With this final data set, we have the ability to generate a set of results. With this we can look more closely at the new data, see what the new data set represents. This will be done in the next chapter.

6 Results

With the structure for how the results are going to be measured and calculated, these results can now be measured, refined and classified. With all these anomalies, what is going to be measured that might give some insight and what is needed to answer the research questions. The specific questions that are of interest to this part of the thesis are the two questions: “How do these correlations describe intermodality and modality of transport delays?” and “How do these correlations influence delays?”

6.1 Problematic Routes Specifics

After calculating the problematic routes and anomalies, the part of the data set that is left to look at is 32 problematic transport moments for waiting times and 62 transport moments for the travel times as achieved by the methods in Chapter 5.

There are significantly less waiting moments because there are less waiting moments compared to travel moments for two reasons:

- There is always one more transport moment in a route than there are waiting moments. This is because for a waiting moment to exist it needs to be between two transport moments.
- The waiting moments that make up the routes that are part of the set of problematic routes contain a large amount of duplicates.

In the next couple of subchapters the results and measurements will be analyzed, these results will then be discussed. What is intended to be extracted from this data set is a causality or a correlation between the delays observed in the data set and the transport types which will be expanded on at the end of this chapter and in the conclusion.

The measurements will be structured into two separate categories with each containing four separate groups. The two categories are the occurrence group, which contains how often certain transport types appear in the data set and the weighted set, which contains how often certain transport types appear and are weighted by the severity of the delay.

Each of these two categories contains identical four groups, each of which consists of four transport types: Trucks, Trains, Ships and Planes.

Control group: Contains all of the transport moments from the data set counted up together and then divided up based on transport type and how often this transport type occurs in the data set.
This group is here as a baseline.

Travel times: Contains all the transport types of the problematic transport moments from the travel group as was discussed in Chapter 5.

Waiting times inbound: Contains all the transport types of the problematic transport moments from the waiting group as was discussed in Chapter 5. This particular group contains the transport types of the transport that brings the goods to the waiting terminal as was discussed in Chapter 5 on page 40 and 41.

Waiting time outbound: Contains all the transport types of the problematic transport moments from the waiting group as was discussed in Chapter 5. This particular group contains the transport types of the transport that takes the goods from the waiting terminal as was discussed in Chapter 5 on page 40 and 41.

This type of feature selection is important and aligns with papers like Brintrup et al., (2019) where the important categories are identified and selected carefully.

When these four groups are looked at across two categories, a causation or correlation could be established but first the data needs to be analyzed, which will be done in the upcoming chapters.

6.2 Control Group

To start off with measuring the impact of the transport types, ship, truck and plane, a standard to measure against must be established. This is done by measuring the occurrence of each transport type in the entire data set. As seen in Figure 18, this is displayed as a percentage.

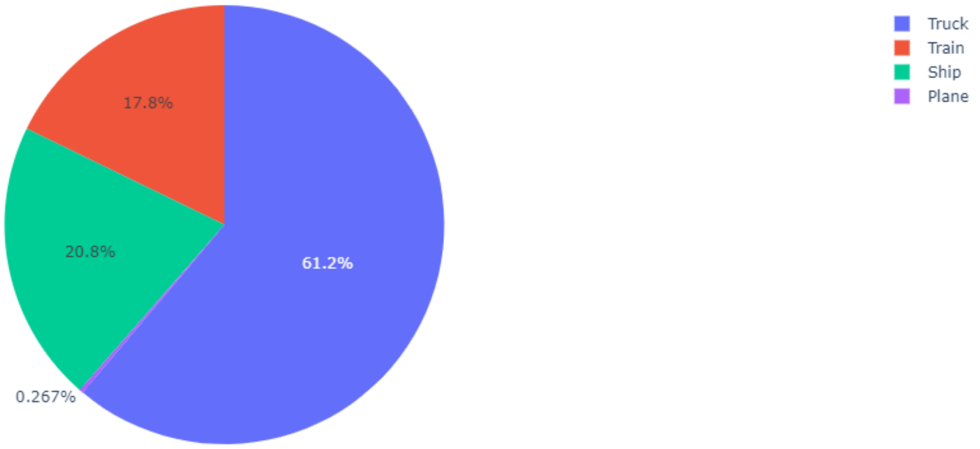

Figure 18: The amount of occurrences of each transport type

This chart counts every occurrence of all four transport types. As can be seen, planes are almost non-existent and will not be taken into serious discussion as they make up less than half a percent of the total transport moments.
Of the other three types, trucks make up 61.2 percent of the chart, trains 17.8 percent and ships about 20.8 percent. This means that trucks are by far the most prevalent form of transportation.
These percentages show a relatively low prevalence of ships, an example from the EU shows that 45.8% was done via truck, 36.9% via sea, 10.2% via rails and 3.8% via inland waterways (European Commission, 2012).
We don’t have any information about the contents of the goods, so it might well be possible that if looking at the total amount transported, trucks might not be the most prevalent transport type when it comes to amounts of goods transported.

If we compare this with similar analysis we can identify that depending on how far the goods are transported, that this could be representative. Depending on the distance, Eurostat, (2022) and U.S. Department of Transportation, (2017) show similar numbers as will be discussed in Chapter 6.5.

6.3 Occurrence Comparison

The first comparison that can be made is to travel times and to waiting times. There are two separate waiting times categories, checking the inbound and the outbound transport types as was described in Chapter 4.2, data set methodology.

Travel times

The first category that is going to be discussed and compared to the original numbers are the problematic routes for travel times. These times are depicted in Figure 19.

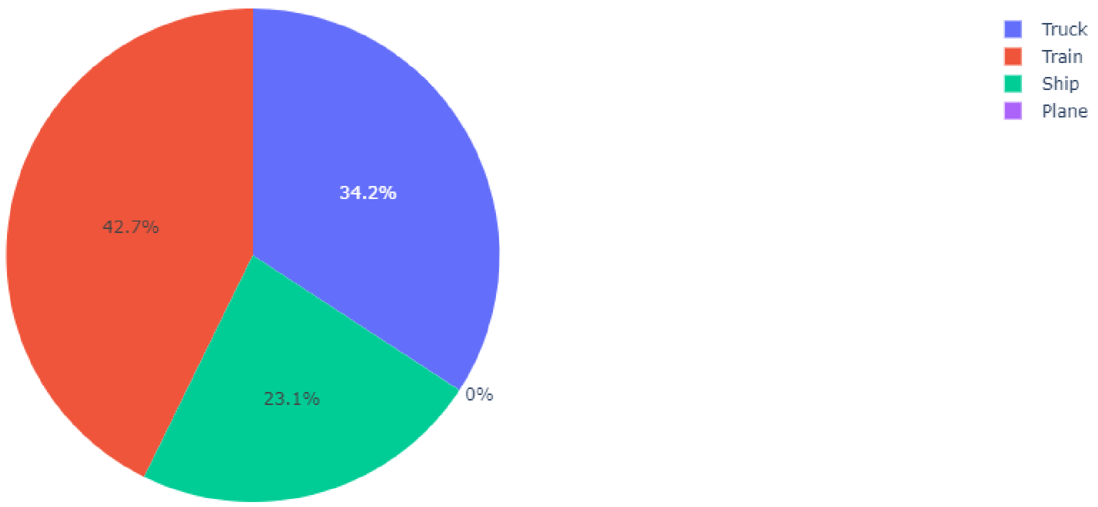

Figure 19: The amount of times each transport type occurs for the set of travel times

In this case, the occurrence of ships in problematic routes is similar to the control group, with only a difference of 2.3 percentage points. The big difference is in the ship and truck percentage points. Trucks have a reduction of 27 percentage points while trains have an increase of 24.9 percentage points.

Trains are significantly more common in this set while trucks are significantly less common. Ships are mostly unchanged when compared to the control group.
For conclusions to be drawn from this data, a look must be given to the surrounding results. When compared with, for example, the control group we can surmise that trucks are more reliable than average, with ships being roughly average and trains being responsible for a lot of delays.

To get more out of this data we also need to look at the weighted comparison which will be discussed later in this chapter.

Waiting times (inbound)

The set of problematic routes about inbound waiting times is about the transport type of the transport route before the waiting time and is depicted in Figure 20. This is described more in depth in Chapter 4.2, data set methodology.

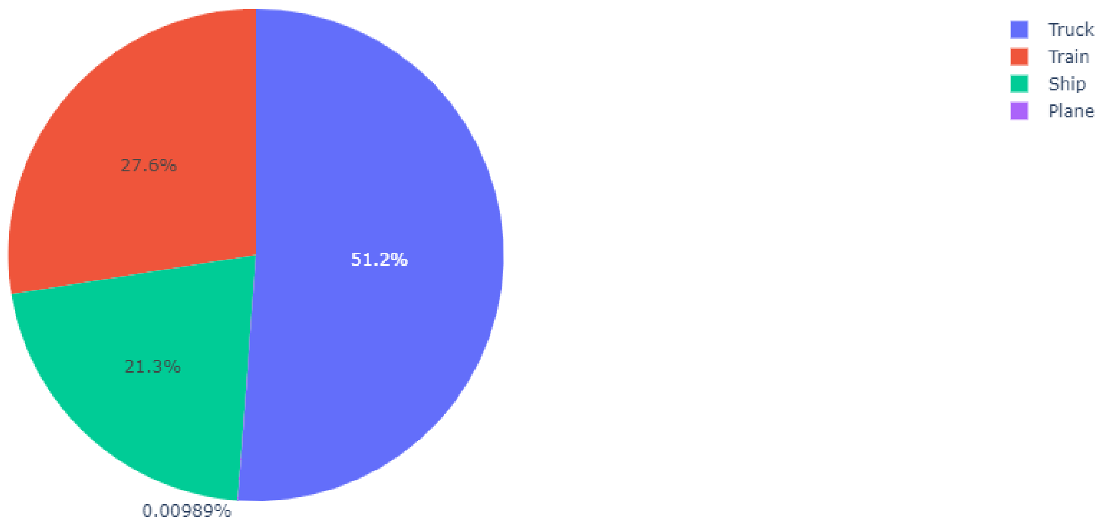

Figure 20: The amount of times each transport type occurs for the set of inbound waiting times

The ship-based travel is, as was with the travel time, essentially the same as the control group with only half a percentage point difference. The difference comes in the form of truck and train numbers, with trucks losing 10 percentage points to the trains.

This reason this measurement was taken into account was due to the concerns that if a transport arrived too late at its destination, a deadline could potentially be missed which would lead to higher delays.
Sometimes transports won't wait for the delayed goods to arrive which will require rescheduling which can be both expensive and time consuming.

One of the important factors to talk about in intermodal or serial transport is that if there is a delay in one transport moment of the route, that this can balloon out in more delays as deadlines are missed. An example of this would be ships, ships tend to be a fairly inflexible but very cheap way to transport goods, a ship isn't going to wait for a couple of missing containers that arrive late. This leads to fairly long delays as goods have to reschedule.

Waiting times (outbound)

This is the second result that is majorly different from the control group, with the first one being travel times. The set of problematic routes that consist of the outbound waiting times, consists of the transport types of the transport route after the waiting time as seen in Figure 21. This is described more in depth in Chapter 4.2, data set methodology. The main goal of this particular set is to look at what our packages are waiting for when they are in storage, waiting to be transported.

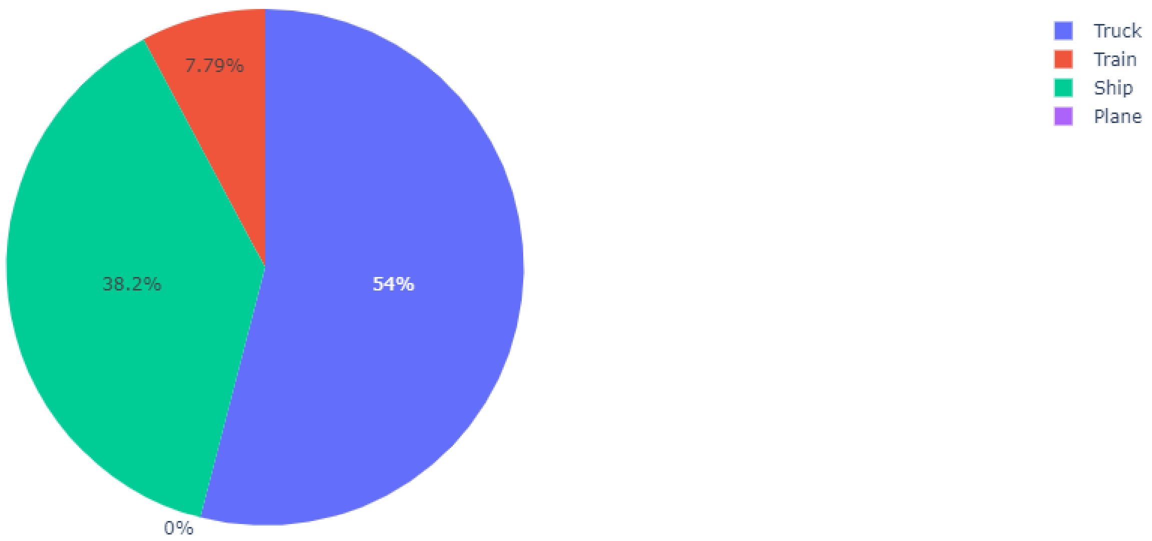

Figure 21: The amount of times each transport type occurs for the set of outbound waiting times

Trucks lose 7.2 percentage points and trains lose 10 percentage points. Ships however, gain 17.5 percentage points.

Outbound waiting time has a relatively low difference compared to the control group, with only two statistically significant differences. In this chart trains are significantly less common and ships are significantly more common.

The outbound waiting times indicate what type of transport the goods are waiting for. This can be indicative of two things:

The first one is the same as with inbound waiting times. Due to delays, goods could have arrived late and missed the transport if the transport didn’t wait for the goods to arrive.

The second one is that the transport type itself leaves later than expected. This could be for a large number of reasons but that would require further research and would be route specific.
One of the major factors that would explain this is that ports tend to be major bottlenecks for transport (Stergiopoulos et al., 2018). This leads to overutilization of infrastructure and long waiting times in the port itself as the port is overbooked.

Another factor that could explain this is that trains tend to be more punctual because of the highly regulated environment (Grechi & Maggi, 2018). This regulation allows for more control over the entire process that the train goes through and would allow for more flexibility in planning for goods as there are less outside forces that influence the departure time when compared to cars and ships.

The final major factor that could explain these types of delays is that shipping routes are more sensitive to errors that happen earlier in the transport route.
When goods are transported by ship, you don’t rent or use the entire ship, you rent individual container space. Multiple other goods are transported on that ship and as such, ships will generally not wait for one or two containers if they arrive late, they are inflexible when it comes to their departure date. This can lead to large waiting times for goods that miss that departure date.

6.4 Weighted Comparison

The weighted comparison is a secondary set of measurements that take into account the severity of the delay; this is simply done by adding all the hours of delays together for the problematic routes instead of adding all the occurrences.

Weighted comparison gives us more information about the problematic routes and more information about the transport types and what the impact is of those transport types.

Travel times

Travel times have a different set of results when the delays associated with travel time are being taken into account as seen in Figure 22.

Figure 22: The amount of times each transport type occurs for the set of travel times when weighted with the amount of hours per transport moment

The part of the pie that represents trains has doubled in size and ships have increased 27.2 percentage points at the cost of trucks which has shrunk by 44.8 percentage points to 16.4 percent.

The difference between this part of the data set and the control group is that trucks are much less common, trains are more common, and ships are much more common.

This indicates that while trucks are the most common delay overal, they aren't the largest source of delays. This is explainable by two reasons, the first reason is that transport moments that use trucks are shorter than the ones that use ships and trains time-wise. In the 20 transport moments with the largest delays there are only around 100 occurrences of truck transport moments while there are around 3000 ship transport moments in the 20 problematic routes with the highest delays.

For ships to have such a large part there are several factors at play, two of which are discussed in Stergiopoulos et al. (2018).

The first reason is that ports are often major bottlenecks for transport and are extremely congested, a large amount of traffic is happening in ports and the infrastructure just can't keep up.
A second reason is that routes that are used by ships tend to be very very long when compared to either trains or trucks.

When comparing the data found here to the control group and to the data of the occurrence travel times to data from the weighted travel time, several things can be noted in addition to this.

- Trucks have the least severe delays and occur less often on average as trucks are overrepresented in the control group.

- Trains have delays of roughly average severity but the delays occur often as trains are overrepresented in the occurrence travel times but become slightly less common when you weigh in the amount of time the delays cost.

- Ships have delays that have a high severity and average occurrence. Because the occurrence of ships barely changes between the control group and the occurrence travel times. When the amount of time every delay costs is taken into account however, the representation of ships becomes way bigger.

The exacted data can be seen in Table 5 in Chapter 6.5.

Waiting times

As seen below the differences for both inbound and outbound travel time are extremely similar to the control group as can be seen in Figure 23 and 24. The only significant difference is in the outbound waiting times where trains make up a significantly smaller part of the pie with 7.1 percentage points less.

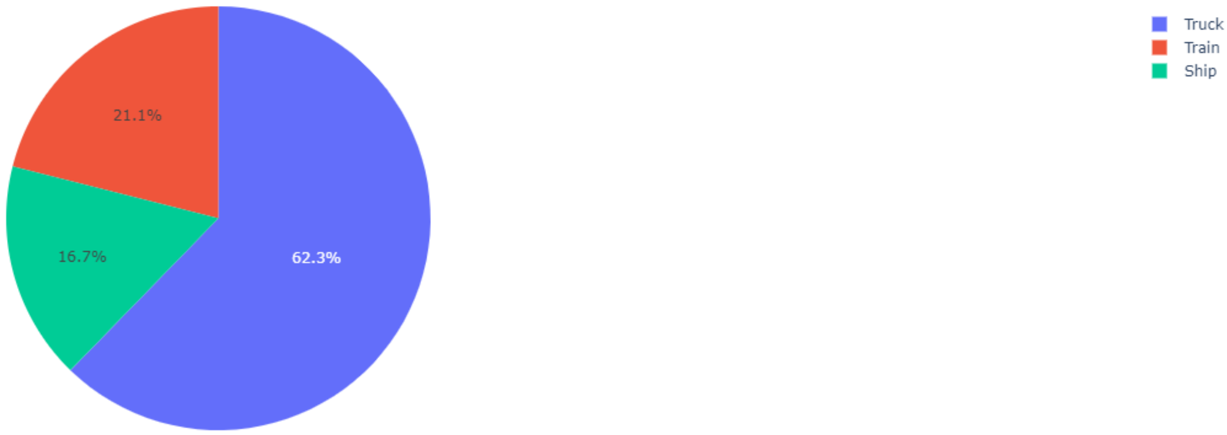

Figure 23: The amount of times each transport type occurs for the set of inbound waiting times when weighted with the amount of hours per transport moment

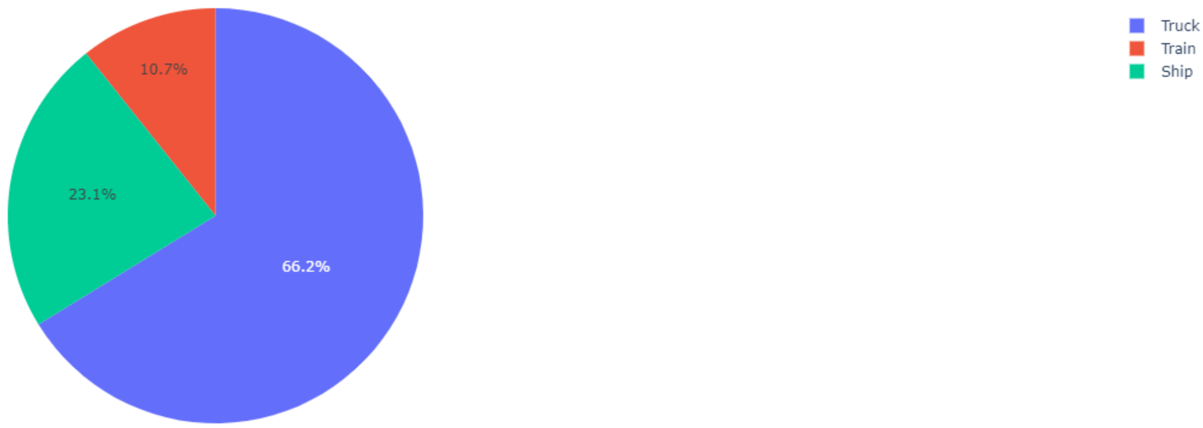

Figure 24: The amount of times each transport type occurs for the set of outbound waiting times when weighted with the amount of hours per transport moment

6.5 Discussion

	Occurrence comparison	**Weighted comparison**
Control group	Truck: 61.2% Train: 17.8% Ship: 20.8%	
Travel time	Truck: 32.4% Train: 42.7% Ship: 23.1%	Truck: 16.4% Train: 35.6% Ship: 48%
Waiting time (outbound)	Truck: 54% Train: 7.8% Ship: 38.2%	Truck: 66.2% Train: 10.7% Ship: 23.1%
Waiting time (inbound)	Truck: 51.2% Train: 27.6% Ship: 21.3%	Truck: 62.3% Train: 21.1% Ship: 16.7%

Table 5: Combination of all the percentage of all the results.

There are three categories that are going to be looked at, are the travel time from the weighted comparison, travel time and outbound waiting time from the occurrence comparison. As can be seen in Table 3 and as was discussed in the previous chapter, these three categories are the only ones with major differences to the control group. In the next set of subchapters these changes will be discussed in depth.

Travel time (weighted)

The difference between this part of the data set and the control group is that trucks are much less common, trains are more common, and ships are much more common.
This indicates that while trucks are the most common delay overall, they aren't the largest source of delay. This is explainable by two reasons, the first reason is that transport moments that use trucks are shorter than the ones that use ships and trains time-wise. In the 20 transport moments with the largest delays there are only around 100 occurrences of truck transport moments while there are around 3000 transport moments that represent ships in the 20 problematic routes with the highest delays.

As seen in Figure 25 generally trucks tend to be more prevalent in shorter distance travel compared to trains and ships. Trucks aren’t significantly slower or faster than trains and so generally the amount of hours traveled is less. This leads to the amount of hours of delay also being lower.

The results of the problematic routes are roughly in line with the data presented in Figure 25, Trucks are most prevalent in the shorter distances in dark blue while trains travel in the middle distances in light blue with ships taking over the long distance bands in green.

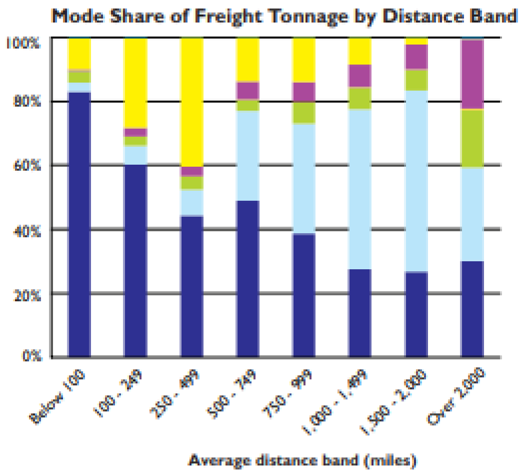

Figure 25: The percentage of each mode of transport
of freight by tonnage in 2015 (U.S. Department of Transportation, 2017, Figure 2-1)

Another agency that paints a similar picture is the Eurostat, which measures both shorter range internal trade as well as longer range international trade. Figure 26 depicts trade between European trade between states, generally shorter ranger while Figure 27 depicts longer range trade between Europe and the rest of the world, reinforcing the causality between delays and distance traveled.

When comparing this data with our own analysis and our own calculations and statistics, a similar story plays out with the split between transport modes as well as being consistent with the analysis of delay length.
The shortest delays belonging to trucks is consistent with the trucks traveling short routes
The middling delays belonging to trains is consistent with trains traveling the routes with middle distance
The longest delays belonging to ships is consistent with ships traveling the routes with the longest distance

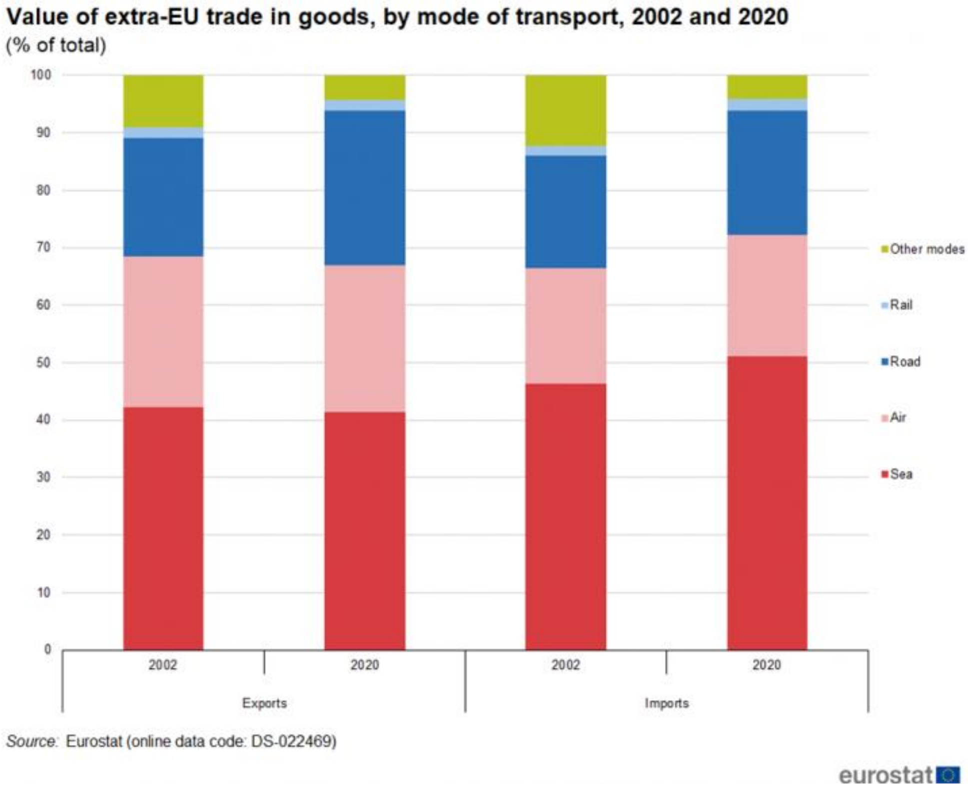

Figure 26: The amount of trade between EU states and states outside of the EU depicted in percentages

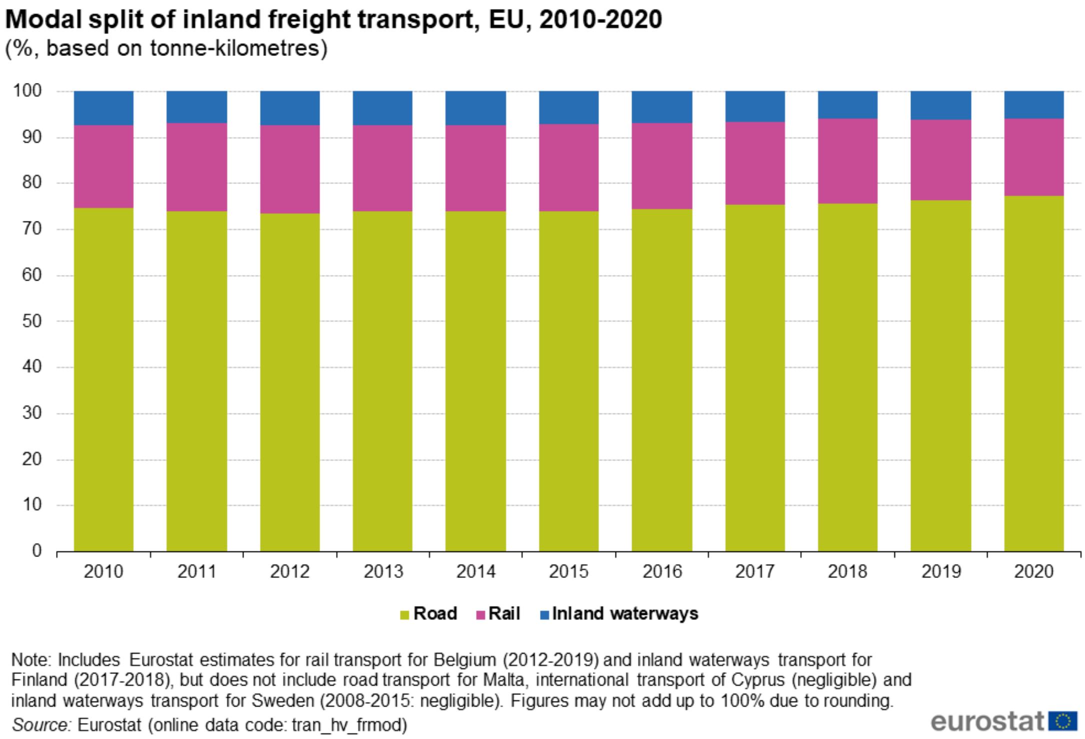

Figure 27: The amount of internal trade between EU states

For ships to have such a large part there are several factors at play, two of which are discussed in Stergiopoulos et al. (2018). The first reason is that ports often serve as bottlenecks for transport and are extremely congested. The second reason is that shipping routes are more sensitive to errors that happen earlier in the transport route.

More reasons for delays exist beyond the distance of travel that make determining the impact of the transport type more difficult. Three of the main reasons are listed below:

Border checks:
Countries that have extensive border checks are a common sight throughout the world, every country has its own rules and regulations on import and export after all. Borders and the steps taken to get through them may significantly lengthen the time it takes for goods to reach their destination and make trading more difficult (Kim et al., 2021).

Infrastructure:
The quality of ports for ships, the quality of railways and roads for trains and trucks, and the condition of airports for aircraft, as well as how well these facilities are used, are all examples of this. Underutilization and underdevelopment are two factors that might cause delays (Prater et al., 2001).
Several examples of delays have previously been shown (Zhang & Figliozzi, 2010) (Roth et al., 2008), as well as how this form of delay is very site dependent or local, with most research being conducted on a national scale (Zhang & Figliozzi, 2010) (Dappe & Lebrand, 2021).

Political Stability:
Another issue is passing through the countries themselves, particularly when passing through politically unstable ones. Crossing the borders of unstable countries can have an impact on international companies' physical assets, staff, and operations (Jodice, 1984).

Travel time (occurrence)

When looking at how often routes occur in the problematic route set for travel times, there is one major change. Trains are significantly more common in this set at the expense of trucks. Ships are mostly unchanged when compared to the control group. For conclusions to be drawn from this data, a look must be given to the surrounding results, specifically the control group, the weighted travel time and the waiting time as seen in Table 5.

When comparing this group to data from the weighted travel time, as can be seen in the red bars in Figure 28, it should be noted that train travel is slightly less represented in the weighted travel time as compared to the occurrence travel time, but when compared to the control group, as can be seen in the blue bars in Figure 28, the occurrence travel time has a significantly bigger representation.

The data indicates that the train delays are relatively more common when compared to the control case, but the severity of the delay is below average.

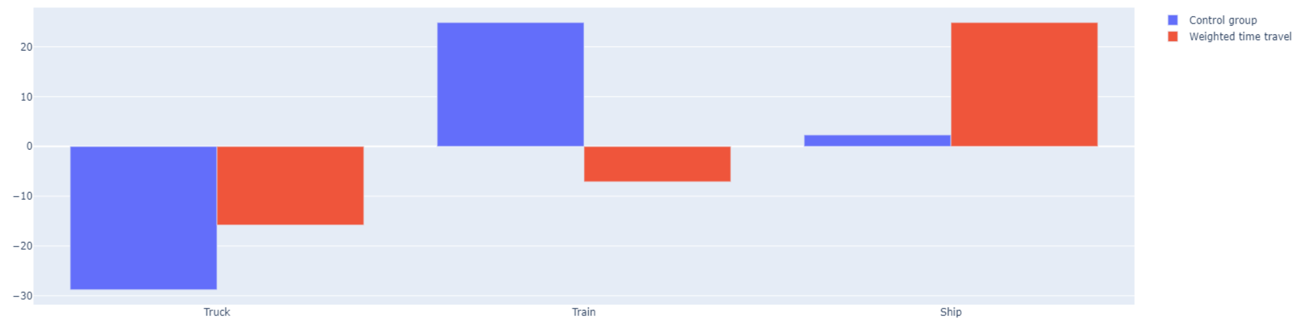

Figure 28:The difference between prevalence transport types between the control group and travel time (occurrence) in blue and between travel time (occurrence) and travel time (weighted) in red

When looking more at the data in Figure 28, a pattern can be seen. Specifically, when looking at the severity of the delays of each type can be determined: Trucks have the least severe delays, as trucks occur way less in the data that represents the occurrence of delays and even less so when that data set is weighed by the severity of the delays as is seen in the red bar in Figure 28.

Trains have delays of average severity but are very common, as can be seen by the blue bar in Figure 28, they occur significantly more often in the data that represents the occurrence of delays. But when looking at the severity of the delays, trains lose 7.1 percentage points indicating that the severity of the delays is average/below average

Ships have high severity delays that are averagely common. In Figure 28 and Table 5 there is only a 2.3 percentage point difference between the control group and the data that represents the occurrence of delays. Where ship traffic takes a hit is in the severity of the delays where the weighted travel time gains 24.9 percentage points compared to the occurrence travel time.

This result reinforces the correlation between distance and delays from the Chapter Travel time (weighted). In that chapter

Outbound waiting time

Outbound waiting time has a relatively low difference compared to the control group, with only two statistically significant differences. In this chart trains are significantly less common and ships are significantly more common.
What this chart essentially describes is what transport the goods are waiting on.
For ships this is partially because of the earlier described phenomenon of ports being a major bottleneck (Stergiopoulos et al., 2018). This leads to larger delays in waiting time as the amount of work that piles up in ports causes a cascading effect of more and more significant delays.
For trains to be as effective as they are in this set is because of the relatively highly regulated and controlled environment that train traffic represents as well as the emphasis that is put on punctuality in a train network (Grechi & Maggi, 2018). In Grechi & Maggi, (2008) there is also further explanation about reasons for why this phenomenon could occur, including a well understood maintenance cycles and failure points which reduces the amount of delays caused by train failures and the relatively deep understanding of what factors cause failure/delays for trains allow for a consistent departure time.

6.6 Combined Results

When looking at the results together instead of looking at individual charts with different results there is more information to extract as was briefly discussed. The list of individual results top summarize is:

In terms of severity of delay, ships have longer delays than trains which have longer delays than trucks. Due to the anonymous nature of the goods it’s hard to fully determine the cause of this and would require further, more indepth research.

The causes that were found were distance of travel, transport via ship generally takes longer than travel via train which takes longer than transport via truck. When there are longer distances, that means that delays also become longer as was discussed in Chapter 6.5.

The other piece of information that could be extracted was that train based transport has a tendency to leave on time more so than average while the opposite is true for ships. This can be correlated to their environment. With ships having to deal with highly congested ports and inflexible travel times, and trains having significantly less factors to deal with which can disrupt departure times.

7 Conclusions

In this chapter final conclusions will be drawn about the results and how this thesis contributes to the larger problem of supply chain analytics as well as discussing the impact of the model made in this thesis.

7.1 Research Questions

This thesis started with the objective of merging the gap between a concrete use case and the more analytical, scientific models like Crainic, (2000) and Engler et al., (2018). To this end the original research question that was asked was:

- For this specific case study we analyze, is there a correlation between the transport type of a transport moment and the severity of the delays of that transport moment?

The goal was to figure out if there was a correlation between the transport type and any delays that happen in the transport moment. This was to be done by taking an anonymized data set and using big data analytics to calculate where delays are in the data set and which of those delays matter and are relevant to look at.

The sub-questions of this thesis are:

- Which different correlations for delay severity exist in this case study?
- How do these correlations describe the intermodality and modality of transport delays
- How do these correlations influence transport delays?
- How can these delays be reduced with a new approach?

The immediate question that springs up after the research question is stated, is which correlations there are between transport type and delays. This was discussed in Chapter 2 together with the question of how these correlations describe the intermodality of transport delays. These questions were confirmed in Chapter 5. When these correlations were achieved, how the correlations interact with modality and how these correlations influence delays and in what way they do.
How can these correlations be reduced using the new approach developed in this thesis was discussed in the latter half of Chapter 6 where the detracting reasons for such delays were pointed out.

As described by Beamon, (1997) and Prater et al., (2001) in Chapter 2.2, there are a lot of factors that go into a route and the delays of a route. Some of these factors were explored in our results, like distance and transport type and some factors were not explored in that amount of depth. This choice of certain topics was mainly for reasons of time and information. The factors that have not been discussed are still an area of research that can be explored and can bring further insight into the problem.
Examples of the factors that weren't able to be researched in depth are divided up into several categories:

Political factors

In Chapter 2.2, political delays are covered. Political delays concern a large amount of bureaucratic delays. These delays are often found in the form of border crossings and checkpoints or when goods are to be transferred between political entities with differing sets of rules and regulations (Bode et al., 2013) (Jodice, 1984).

Impact of countries on the travel time:
One of the factors that rears its head every so often is overly byzantine bureaucracies. Border checks can severely increase the amount of time goods take to reach their destination and make trade harder (Kim et al., 2021). An example of how long border delays can take is sub-Saharan Africa where border delays, in some cases, take up to 6 times as long as in countries like Germany and Ireland (Marius, 2022).

Impact of political stability factors on travel time:
Another problem is going across borders, especially when those borders belong to politically unstable countries. Going across the borders of unstable countries can affect the physical assets, personnel, and operations of foreign firms (Jodice, 1984). Even in countries where political instability is not or less of an issue there is still the problem of border control which costs time.

Having these factors which cost time, incorporated into the delay calculations would grant a clearer picture about what is happening in the routes of the datasets and would allow for an easier assessment about the impact of other variables.

Infrastructure factors

Impact of the infrastructure of the transport modes:
Examples of this are the quality of ports for ships, the quality of railway and roads for trains and trucks and the quality of the airports for planes as well as how well these facilities are utilized. The underutilization and underdevelopment can both lead to delays (Prater et al., 2001).
Several examples of delays have already been given (Zhang & Figliozzi, 2010) (Roth et al., 2008), as well as how this type of delay is very location dependent or very local, with most studies being on a national level. Like looking at the impact of rail on the Chinese transport network (Zhang & Figliozzi, 2010) or the impact of infrastructure in the Horn of Africa (Dappe & Lebrand, 2021).

Intermodal delays
Intermodal delays are delays that are associated with swapping transport types in a terminal. By swapping types of transport multiple times during the route to the final destination, delays happen. Every additional switch of a good from one type of transport to another causes additional complexity and delays. For goods to swap mode of transportation they need to get transported to a transportation hub, get unloaded off the current mode of transportation, potentially sit in the transportation hub for a time, get loaded on the new mode of transportation and then leave the transportation hub (Engler et al., 2018).

Swapping between transport types compounds the risk of failure, leading to an increasing need to increase reliability. Something can go wrong in any stage of the process of transshipment which reduces the overall reliability of the entire transport process (Puil, 1994) (Lo & Hall, 2008).

Out of all of the mentioned factors that influence the use case, a correlation was observed between the modality of transport and the delays that happen during the transport, as was seen in Chapter 5. To this end we introduced a model in Chapter 5. Via this model we managed to gather relevant data from the data set to prove this correlation in Chapter 6.

In this thesis a specific correlation was found between transport types and the length of the delay. Ships have greater delays than trains, which have greater delays than trucks.
The anonymous nature of the items of the data set in this thesis makes it difficult to pinpoint the exact reasons for these delays. An additional in-depth investigation would be needed to pinpoint those exact reasons and what share of the delay can be attributed to them.
The distance of travel was discovered to be one of the factors to have an impact on the delays attributed to travel modes: Traveling by ship takes longer than traveling by rail, which takes longer than traveling by truck. When there are larger distances between points, there are also lengthier delays.

Another piece of information that was gathered was that train-based transportation had a higher tendency to leave on time than average, but ships have the opposite tendency. This can be linked to their surroundings. Ships must deal with congested ports, whereas trains have far fewer elements to contend with that might cause delays in departure timings (Grechi & Maggi, 2018).

To figure out the remaining factors as well as researching them thoroughly is, however, out of the scope of this thesis and will be discussed in future works.

7.2 Advances and Potential in this Case Study

This case study has focused on the effect of transportation modes and intermodality in the supply chain on delays, as shown in the Chapter 7.2, specifically by using big data and big data analytics. The advances made by this model are there to contribute to merging the gap between the literature of supply chain analytics and the real world use cases this literature would apply to.

The model for processing the information in this thesis is there to extract an image of the impact of transport models on the travel time. This is unlike other papers like Zokaee et al.,(2014) and Schijndel & Dinwoodie,(2000) which focus mostly on cost or focus on multiple factors trying to get a comprehensive picture.
This use case tries to narrow the focus and scope of the research to get a more narrow and focused answer. The studies that focus on more wider models like Crainic (2000), do however get a clearer and larger picture of the whole situation.

This model, as described in Chapter 5 and Chapter 6, allows for the transformation of big data into specific knowledge about the transport delays. The use case that was set up using that model has shown us that there is a correlation between transport delays and transport types.
This model and the data can be further refined into giving a clearer picture of the exact causes but the current model gives a clear focussed picture of this conclusion.

This thesis and the results therein serve as a platform to launch deeper research into this subject and as a way to merge the gap between theoretical models for Big Data and supply chains and the practical application of said data in a concrete scenario. This case study found a correlation between the modality of transport types and the delays incurred by the transport of goods.
For this case study, an impact on delays is shown by the different transport types, trains, trucks and ships.

References

Beamon, B. M. (1997, April 24). Supply chain design and analysis: Models and methods. (Supply chain design and analysis). https://doi.org/10.1016/S0925-5273(98)00079-6

Bode, C., Kemmerling, R., & Wagner, S. M. (2013). Internal versus External Supply Chain Risks: A Risk Disclosure Analysis. http://dx.doi.org/10.1007/978-3-642-32021-7_6

Brintrup, A., Pak, J., Ratiney, D., Pearce, T., Wichmann, P., Woodall, P., & McFarlane, D. (2019, Februari 23). Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. https://doi.org/10.1080/00207543.2019.1685705

Brunel University. (2019, December 13). *Supply Chain Logistics Problem Dataset*. Brunel University London. Retrieved June 20, 2022, from https://brunel.figshare.com/articles/dataset/Supply_Chain_Logistics_Problem_Dataset/7558679

Crainic, T. G. (2000). Service network design in freight transportation. *European Journal of Operational Research*, *122*. https://doi.org/10.1016/S0377-2217(99)00233-7

Crainic, T. G., & Laporte, G. (1997). Planning models for freight transportation. *European Journal of Operational Research*, *97*, 409-438. https://doi.org/10.1016/S0377-2217(96)00298-6

Crainic, T. G., & Rousseau, J.-M. (1986). Multicommodity, multimode freight transportation: a general modeling and algorithmic framework for the service network design problem. https://doi.org/10.1016/0191-2615(86)90019-6

Creative Safety Supply. (2018, February 22). *Information Sharing in Supply Chain Integration*. Creative Safety Supply. Retrieved March 15, 2022, from https://www.creativesafetysupply.com/articles/supply-chain-integration/

Dappe, M. H., & /Lebrand, M. S. M. (2021, November). Infrastructure and Structural Change in the Horn of Africa. http://hdl.handle.net/10986/36646

Eaton, C., & Zikopoulos, P. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. Mcgraw-hill.

Engler, E., Gewies, S., Banys, P., & Grunewald, E. (2018, 3 12). TRAJECTORY-BASED MULTIMODAL TRANSPORT MANAGEMENT FOR RESILIENT TRANSPORTATION. http://doi.org/10.21307/tp.2018.13.1.8

European Commission. (2012). *EU Transport in Figures 2012* (Union européenne. Commission européenne, Ed.). Publications office of the European Union.

Eurostat. (2022, March). Freight transport statistics - modal split. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics_-_modal_split#Modal_split_in_the_EU

Fanoodi, B., Malmir, B., & Jahantigh, F. F. (2019, October). Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. https://doi.org/10.1016/j.compbiomed.2019.103415

Gandom, A., & Haider, M. (2015, April). Beyond the hype: Big data concepts, methods, and analytics. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gartner. (n.d.). *Gartner Glossary*. https://www.gartner.com/en/information-technology/glossary/data-and-analytics

Grechi, D., & Maggi, E. (2018, December). The importance of punctuality in rail transport investigation on the delay determinants. https://www.researchgate.net/publication/330322491_The_importance_of_punctuality_in_rail_transport_investigation_on_the_delay_determinants

Gunasekeran, A., & Ngai, E. W. T. (2003, November 6). Information systems in supply chain integration and management. doi:10.1016/j.ejor.2003.08.016

Harper, J. (2021, March 26). Suez blockage is holding up $9.6bn of goods a day. *BBC*. https://www.bbc.com/news/business-56533250

Hrušovský, M., Emrah Demir, Jammernegg, W., & Woensel, T. V. (2021, January 20). Real-time disruption management approach for intermodal freight transportation. https://doi.org/10.1016/j.jclepro.2020.124826

IBM. (2021). *What is supply chain analytics?* IBM. Retrieved March 17, 2022, from https://www.ibm.com/topics/supply-chain-analytics

Jing, K., & Liao, W.-C. (2017, 12 17). The Butterfly Effect of Transport Network Connectivity and Welfare Implication. http://dx.doi.org/10.2139/ssrn.3089360

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008, April 11). Visual Analytics: Definition, Process, and Challenges. 23. https://hal-lirmm.ccsd.cnrs.fr/lirmm-00272779/file/VAChapter_final.pdf

Kim, K., Mariano, P., & Abesamis, J. (2021). Trade Impact of Reducing Time and Costs at Borders in the Central Asia Regional Economic Cooperation Region, Emerging Markets Finance and Trade. https://doi.org/10.1080/1540496X.2021.2007877

Lee, D. K., In, J., & Lee, S. (2015, May 28). Standard deviation and standard error of the mean. https://doi.org/10.4097/kjae.2015.68.3.220

Lee, L. H., Padmanabhan, P., & Whang, S. (1997, April). Information distortion in a supply chain: The bullwhip effect. (Management Science), 546-558. https://doi.org/10.1287/mnsc.43.4.546

Lo, S., & Hall, R. W. (2008). The design of real-time logistics information system for trucking industry. https://doi.org/10.1016/j.cor.2007.01.023

Marius, I. (2022, January). Time delays at the border: macroeconomic consequences for sub-saharan african economies. https://steg.cepr.org/sites/default/files/2022-01/Full%203%20Adom.pdf

McKinsey. (2016, February 16). *Big data and the supply chain: The big-supply-chain analytics landscape (Part 1)*. McKinsey. Retrieved March 17, 2022, from https://www.mckinsey.com/business-functions/operations/our-insights/big-data-and-the-supply-chain-the-big-supply-chain-analytics-landscape-part-1

McKinsey Global Institute. (2020). Risk, resilience, and rebalancing in global value chains. https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/risk%20resilience%20and%20rebalancing%20in%20global%20value%20chains/risk-resilience-and-rebalancing-in-global-value-chains-full-report-vh.pdf?shouldIndex=false

MixMove. (2021, November 9). *What is supply chain integration?* MIXMOVE. Retrieved March 15, 2022, from https://www.mixmove.io/post/supply-chain-integration

NIST. (2015, September). NIST Big Data Interoperability Framework: Volume 1, Definitions. http://dx.doi.org/10.6028/NIST.SP.1500-1

Park, H., Bellamy, M. A., & Basole, R. C. (2016, November). Visual analytics for supply network management: System design and evaluation. https://doi.org/10.1016/j.dss.2016.08.003

Prater, E., Biehl, M., & Smith, M. A. (2001). International supply chain agility, Tradeoffs between flexibility and uncertainty. 823-839.

Puil, J. (1994). Knopen uit de keten, NBT.

Roth, A. V., Tsay, A. A., Pullman, M. E., & Gray, J. V. (2008). Unraveling the food supply chain: Strategic insights from China and the 2007 recalls. *Journal of Supply Chain Management*, *44*, 22-39. https://doi.org/10.1111/j.1745-493X.2008.00043.x

Russom, P. (2011). TDWI Best Practices Report | Big Data Analytics. https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf

Sagiroglu, S., & Sinanc, D. (2013). Big Data: A Review. doi:10.1109/cts.2013.6567202

Schijndel, W.-J. v., & Dinwoodie, J. (2000). Congestion and multimodal transport: a survey of cargo transport operators in the Netherlands. https://doi.org/10.1016/S0967-070X(00)00023-8

Seyedan, M., & Mafakheri, F. (2000, July 25). Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. *53*. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00329-2

Simsion, G. C., & Witt, G. C. (2005). *Data modelling Essentials*.

Stergiopoulos, G., Valvis, E., Mitrodimas, D., Lekkas, D., & Gritzalis, D. (2018). Analyzing Congestion Interdependencies of Ports and Container Ship Routes in the Maritime Network Infrastructure. *IEEE Access*, *Volume 6*, 63823-63832. https://doi.org/10.1109/ACCESS.2018.2877659

Thomas, J.J., & Cook, K.A. (2005). *: Illuminating the Path. IEEE Computer Society Press*.

U.S. Department of Transportation. (2017). https://www.bts.dot.gov/sites/bts.dot.gov/files/docs/FFF_2017_Full_June2018revision.pdf

Viellechner, A. M. (2022, February 5). The New Era of Predictive Analytics in Container Shipping and Air Cargo. https://opus4.kobv.de/opus4-whu/files/899/Viellechner_Adrian_WHU_Diss_2022.pdf

Yildiz, B., Yaman, H., & Karaşan, O. e. (2021, October 5). https://doi.org/10.1287/trsc.2021.1070

Zhang, Z., & Figliozzi, M. A. (2010, March). A Survey of China's Logistics Industry and the Impacts of Transport Delays on Importers and Exporters. *Transport Reviews*, *30*, 179-194. https://doi.org/10.1080/01441640902843232

Zimon, D., Tyan, J., & Sroufe, R. (2019, December 10). Implementing Sustainable Supply Chain Management: Reactive, Cooperative, and Dynamic Models. https://doi.org/10.3390/su11247227

Zokaee, S., Jabbarzadeh, A., Fahimnia, B., & Sadjadi, S. J. (2014, November 30). Robust supply chain network design: an optimization model with real world application. https://doi.org/10.1007/s10479-014-1756-6