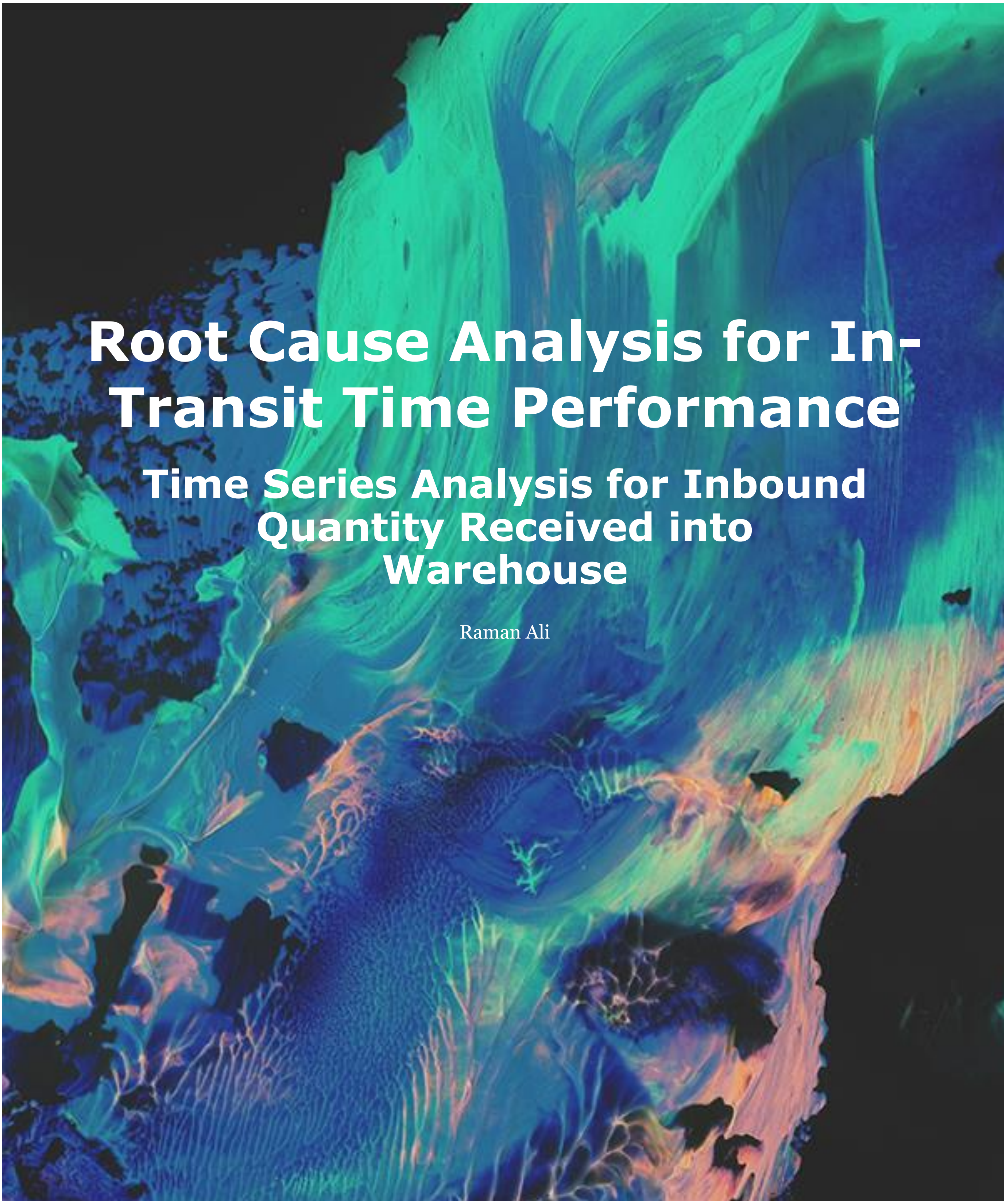
UMEÅ UNIVERSITY



Root Cause Analysis for In-Transit Time Performance

Time Series Analysis for Inbound Quantity Received into Warehouse

Raman Ali

Master's thesis, 30 ECTS

Master of Science in Industrial Engineering and Management, 300 ECTS

Department of Mathematics and Mathematical Statistics

Spring term 2021

ROOT CAUSE ANALYSIS FOR IN-TRANSIT TIME PERFORMANCE
Submitted in fulfilment of the requirements for the degree Master of Science in Industrial Engineering and Management

Department of Mathematics and Mathematical Statistics
Umeå University
SE- 901 87 Umeå, Sweden

Supervisors:
Jun Yu, Umeå University
Bharath Srinivasan, Cytiva

Examiner:
Jonas Westin, Umeå University

Abstract

Cytiva is a global provider of technologies to global pharmaceutical companies and it is critical to ensure that Cytiva's customers receive deliveries of products on-time. Cytiva's products are shipped via road transportation within most parts of Europe and for the rest in the world air freight is used. The company is challenged to deliver products on time between regional distribution points and from manufacturing sites to its regional distribution centers. The time performance for the delivery of goods is today 79% compared to the company's goal 95%

The goal of this work is to find the root causes and recommend improvement opportunities for the logistics organizations inbound in-transit time performance towards their target of 95% success rate of shipping in-transit times.

Data for this work was collected from the company's system to create visibility for the logistics specialists and to create a prediction that can be used for the warehouse in Rosersberg. Visibility was created by implementing various dashboards in the Qlik Sense program that can be used by the logistics group. The prediction models were built on Holt-Winters forecasting technique to be able to predict quantity, weight and volume of products, which arrive daily within five days and are enough to be implemented in the daily work. With the forecasting technique high accurate models were found for both the quantity and weight with accuracies of 96.02% and 92.23%, respectively. For the volume, however, too many outliers were replaced by the mean values and the accuracy of the model was 75.82%.

However, large amounts of discrepancies have been found in the data which today has led to a large ongoing project to solve. This means that the models shown in this thesis cannot be completely reliable for the company to use when a lot of errors in their data have been found. The models may need to be adjusted when the quality of the data has increased. As of today the models can be used by having a glance upon.

Sammanfattning

Cytiva är en global leverantör av teknik till globala läkemedelsföretag och det är viktigt att Cytivas kunder får leveranser av produkter i tid. Eftersom Cytivas produkter skickas via vägtransporter inom de flesta delar av Europa och för resten av världen används flygfrakt, utmanas företaget att leverera produkter i tid mellan regionala distributionsplatser och från tillverkningsställen till sina regionala distributionscentran. Tidsprestandan för leveransen av varor är idag 79% gentemot företagets mål 95%.

Målet med detta arbete är att hitta rotorsakerna och rekommendera förbättringsmöjligheter för logistik organisationen för att förbättra tidsprestandan för leveransen av inkommande varor mot målet på 95%.

Data för detta arbete samlades in från företagets system för att skapa visibilitet för logistik specialisterna samt för att skapa en prediktering som kan komma till användning för lagret i Rosersberg. Visibilitet skapades genom att implementera olika dashboards i programmet Qlik Sense som kan komma användas av logistik gruppen. Modellerna för prediktering är byggda på Holt-Winter's predikterings teknik och predikterera kvantitet, vikt och volymen, som kommer in dagligen inom fem dagars sikt in till varulagret. Med denna predikterings teknik hittades precisa modeller för både kvantitet och volym med precision på 96.02% och 92.23%. För volymen var dock inte modellen lika precis och detta för att fler fel fanns i datat som korrigerades om till medelvärden, detta gav en predikteringsmodel med en precision på 75.82%.

Modellerna är precisa nog för att kunna implementeras i det dagliga arbetet. Dock har stora mängder avvikelser hittats i datat som idag blivit till stora pågående projekt för att lösa. Detta leder till att modellerna som visas i detta examensarbete inte kan litas blint på för företaget då mycket fel i deras data hittats. Modellerna kan komma att behöva korrigeras när kvaliteten på datat ökat och datat är pålitligt. Idag kan modellerna användas som något lagret kan kasta en blick på för att få ett pekfinger på hur mycket som kommer in kommande vecka.

Acknowledgments

I would like to thank supervisor Bharath Srinivasan for all the introduction and guidance through the master's thesis, and for the opportunity to carry out the master's thesis on Cytiva. I would also like to thank other employees, including the Logistic Specialist team and Mattias Thalén for insight, support and help during the work.

I appreciate and thank you very much for the help I received during the thesis work from supervisor Jun Yu, professor at Department of Mathematics and Mathematical Statistics at Umeå university.

Raman Ali
Uppsala, June 8, 2021

Acronyms

UPD - Warehouse in Rosersberg

PDP - Primary distribution point

RDP - Regional distribution point

OEM - Original equipment manufacturer

OTP - On Time Performance

KPI - Key Performance Indicator

SCM - Supply Chain Management

ADF - Augmented Dickey Fuller Test

AR - Auto Regressive model

SARIMA - Seasonal Auto Regressive Integrated Moving average model

Contents

1 Introduction

This section provides an introduction to the thesis. A background of the problem and the company involved is presented following with a problem description together with the aim and the purpose of the thesis. An outline of the thesis is stated in the end of the section as guidance for the reader.

1.1 Background

Transport of goods is critical for our society, without functioning transportation systems it would be impossible to live in the global world we live in today. Functional transport enables economic growth that is critical for us to be able to maintain and improve our standard of living. Further, sustainable transportation is a current topic that affects us all because of the carbon footprint. In order to maintain our current lifestyle, we must all work to make transport more efficient and sustainable. Therefore, one of the important aspects that contributes to sustainability and on-time deliveries is optimized routings. With efficient routings delivery times decrease and the time spent on road, ocean and in air decreases which contributes to sustainability.

Cytiva is a global provider of technologies to global pharmaceutical companies and it is critical to ensure that Cytiva's customers receive deliveries of products on-time. As Cytiva's products are shipped via road transportation within most parts of Europe and for the rest in the world air freight is used. The company is challenged to deliver products on time between regional distribution points and from manufacturing sites to its regional distribution centers. The import volume has been doubled as of January 2020, in comparison to earlier years. This is because as of January 2020, the demand became higher than the supply and the manufacturing sites needed to start manufacture as much as possible. So, in these times, it is even more important to have a functioning transportation system with many transportation options and optimized routings.

Cytiva is a conglomerate with suppliers and customers' from all over the world, in other words there is a global transportation system within the company with a complex supply chain. Hence, air freight is not an bad choice, because of the global reach, fast delivery, its tracking system, reliability and the low risk it includes in terms of minimizing delays and optimizing routings. But since most of its deliveries are transported through air freight, we know as of today this is not a sustainable procedure. Looking through the perspective of the Tripple Bottom Line (TBL), as a response to public concerns about corporate morale, it involves not only reporting on economic results but also on social and environmental aspects [20]. Today, the importance of sustainability is in the spotlight. Since 2016 when the Paris Agreement was signed and adopted by 196 countries to reach the goal of limiting global warming to well below 2, preferably 1.5°C. The importance of sustainability was even more emphasized [6]. For a leading company, one of the main goals is that they have to be cutting-edge in every possible aspect, hence one can argue the importance of efficient routings and sustainability. For an example, by creating variety in the transportation system and decreasing their carbon footprint. The question that arises is if this is possible for such an company to deliver their important products in-time and with the least possible risk of delays? Especially with the circumstances today with the Covid-19 situation that has affected the way of transporting through air works, since flights have been decreased by approximately 50% it is even more important to consider various transportation systems and optimized routings within the supply chain [22].

Further one could argue the expenses of air freight. According to Freightos that are managing global freight comparisons between 10 000 companies, the expenses is five times higher for air shipping than sea transportation in general [7]. Transportation through air is also weather dependent which can impact the delivery time to the costumers. Lastly, the issue for a company that transports through air mostly, is the limitations of both size and weight of the products. To not have the shipping routings efficient is expensive and not efficient in all the three aspects of TBL.

1.2 About Cytiva

Danaher is an American healthcare conglomerate based in Washington, D.C., USA. One of the company's platforms is life science. March 2020 Danaher acquired Cytiva which was the biopharma business unit within GE Life Science [18]. Cytiva makes equipment for both the manufacturing and research of pharmaceuticals. Today, there is two departments in Sweden, one located in Umeå with 550 employees and the other located in Uppsala with approximately 1600 employees. Besides Sweden, the company operates in 39 other countries. Cytiva is a global provider of both services and technologies, that accelerates and advances the development and manufacture of therapeutics. With customers' undertaking life-saving activities, the company's job is to supply them with the right tools. Cytiva brings efficiencies to research and manufacturing workflows where they ensure the development, manufacturing and the delivery of transformative medicines to patients. Where over 90% of top-selling biological manufactured using Cytiva's products [4].

1.3 Problematization

This thesis will explore and identify root causes of the delayed shipping transit times by using the principles of Lean and analyzing data from both the company and its forwarders. From that to come up with new ideas that can improve the routings, shipping transit times and to be able to propose a road map for implementing identified improvements.

As mentioned earlier, in section 1.1, transport of goods is critical for our society due to the globalization as of today. Further one can question the sustainability of today's transportation systems. Cytiva that is not using efficient routings it is even more important to put emphasis on the shipping approach used today. For example, a product delivered from Japan can be shipped to USA, from USA shipped to Sweden and then from Sweden shipped to a customer in South Korea, when it instead could have been delivered from Japan directly to South Korea that is a neighbouring country. Cytiva is a global provider of technologies to global pharmaceutical companies therefore it is critical to ensure Cytiva customers receive deliveries of products on-time. As Cytiva's products are shipped via air freight, the company is challenged to deliver products on time both between the regional distribution centers and from manufacturing sites to its regional distribution centers. The shipping in-transit time performance is as of today 79% versus a target of 95%.

In order to improve the delaying shipping transit times and with that also take sustainability into consideration with an aspect of different transportation system, it has to be mapped and visualised. The specified areas for this thesis will be the shipping from external and internal suppliers to the primary distribution points (PDP) in Sweden. In other words, the inbound into the warehouse in Rosersberg called (UPD).

Figure 1 shows a general value stream map of Cytiva and its supply chain complexity, where the arrows shows the shipping. There are in total 15 manufacturing sites located in different regions. Further, there is three primary distribution points located in USA (PYD), Netherlands (EYD) and Sweden (UPD). Lastly the company has regional distribution points that receives the freight from the PDP and then delivers it to the customers', these are located in Japan (JPD), Singapore (SGD) and China (CNS). As we can see from Figure 1 that the supply chain is complex and shipments are going everywhere, for an example sometimes freight is directly transported from the PDP to customers' and other times not. Lastly, we can quickly identify that the information from the external suppliers is inexplicit. In other words, the transportation from the external suppliers to the PDP in Sweden is as of today not "visible" and the information between the sites is vague.

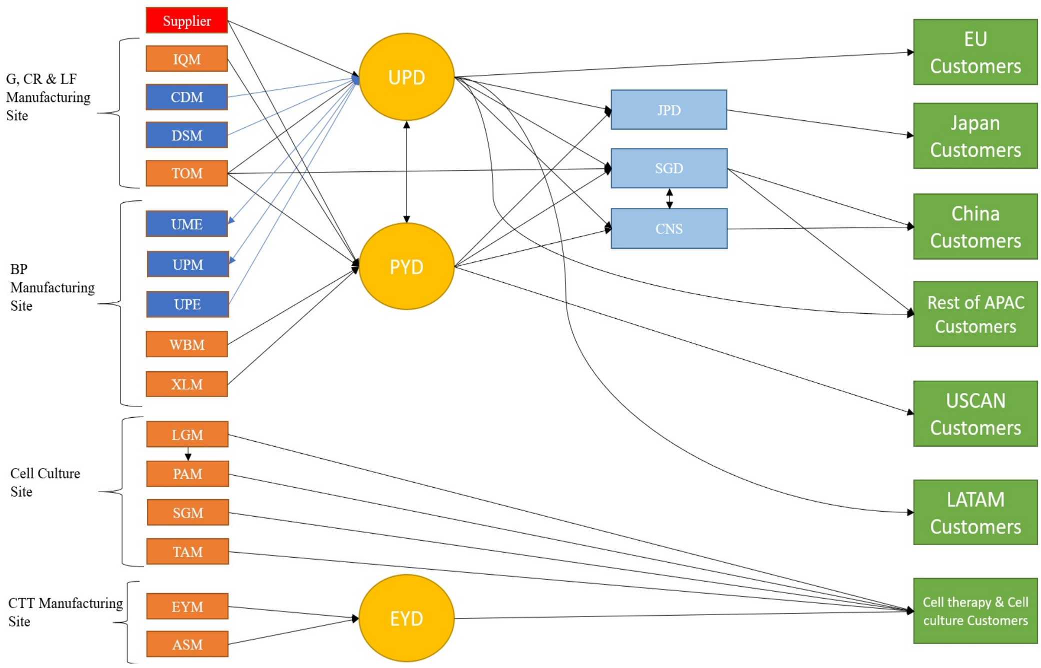


Figure 1: *A general value stream map of the company's transportation's from start to end. The arrows represent shipping.*

By creating a high level value stream map one can visualise and identify waste. Further, one can by processing and analysing the data identify the root causes of the waste and the reasons for the delaying shipping transit times. One of the challenges is to identify the reasons for the low performance of the in-transit time and find possible solutions. From this problematization the following three specific questions have been raised for this thesis work:

- *How can statistical and logistical methodologies contribute to identify the root cause of delayed shipping transit times within Cytiva and suggest improvements in the supply chain routings and other identified initiatives to reach the company's goal of 95%?*
- *Are there any key performance indicators highlighted from logistical tools that could be included in the Daily Management system?*
- *How can the identified initiatives be prioritized and implemented within the logistics organization?*

1.4 Project Goal

The goal of this work is to find the root causes of the late in-transit time performances and recommend improvement opportunities for the logistics organizations in-transit time performance towards their target of 95% success rate of shipping in-transit times. In other words, find the root causes and create visibility for the logistic specialist team. Further, possible solutions to the root causes will be investigated and presented to the logistic specialist team as well. Lastly, the purpose is to propose key performance indicators that could be included in the Daily Management System for the logistics operations team that could improve the in-transit time performances.

1.5 Limitations

The objective is to find the reasons for the delayed shipping transit times, in other words identify root causes within the supply chain of the company. To do so, one main obstacle will be the complex supply chain the company has. Therefore, limitations are essential, if the scope is too big, time will not suffice and the project will most likely lead nowhere.

On Time Performance (OTP) is the logistic specialist teams key performance and one of Cytiva's eight Key Performance Indicators (KPI) that they are measuring. In-Transit times is a part of creating good OTPs and therefore the scope will lay within the inbound from both the external and the internal suppliers. The original equipment manufacturers (OEM) which are the external suppliers, the manufacturing sites and the other RDPs are the internal suppliers.

In this thesis, the scope will lay within the inbound in-transit time performances for the warehouse located in Rosersberg or so called UPD. To visualise the scope one can observe Figure 2 and think of the arrows only pointing towards UPD that is the inbound for the warehouse.

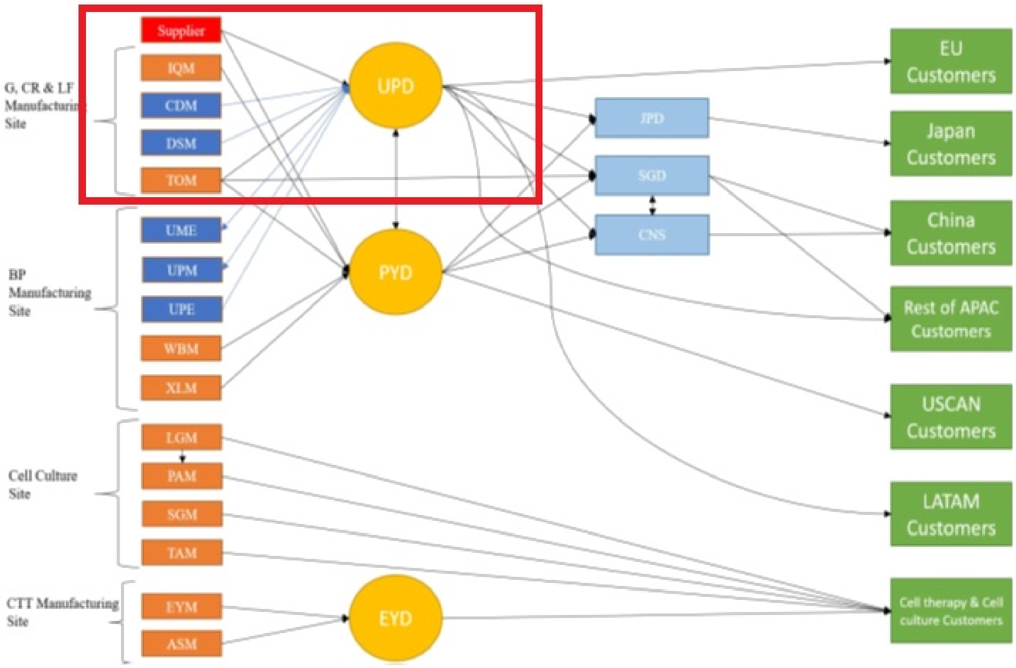


Figure 2: *A general value stream map of the company's transportation's from start to end. The arrows represent shipping. The marked line is the scope for this thesis.*

1.6 Project Go Through

In this thesis a root cause analysis will first be done to identify reasons of the late in-transit time performance introduced above. Further, the root cause analysis will be a standpoint for creating visibility for the logistic organization at Cytiva. Lastly, possible solution to a found root cause will be made. In this case, a time series analysis will be done to forecast weight, volume and quantity coming into the warehouse in Rosersberg due to warehouse capacity and planning problems.

1.7 Outline

The overall structure of this master's thesis will follow a chronological order of the project work, with the purpose of making the report easily understood regardless of what kind of background the reader has. This thesis is arranged in the following way:

Section 1, gives a background and introduction to the goal and objective of this thesis, but also a background to give some context to the problem.

Section 2, a theoretical background about both logistical and statistical methodologies will be presented

Section 3, the proceeded method will be presented and the usage of the data will be presented; how the data was collected, reprocessed and analyzed.

Section 4, will provide the results given by the applied methodologies.

Section 5, presents discussions and conclusions from the obtained results and what these results imply if the methods and models were to be implemented in a practical use.

Section 6, the final section will contain my final thoughts regarding the thesis and my personal thoughts that I wanted to share with the reader. It also involves my recommendations for the logistical organization on how they should interpret the outcome of the project work.

2 Theory

This section will provide the theory behind the methodologies which are used in the thesis. Presenting the necessary logistical methodologies and statistical methodologies.

2.1 Logistical methodology

2.1.1 Supply Chain Management

Supply Chain Management (SCM) is a way of describing the way in which supply processes are managed and structured [27]. SCM covers all movement and storage of raw materials, work-in-process, inventory and finished products from origin to point of consumption [14]. The concept is linked to logistics, where SCM is a concrete mindset to see the logistical functions from an organization's point of view. The chain also includes organizations and processes needed to create and deliver products and services to the customer.

Logistics is, according to Bowersox et al. [3], part of the overall supply chain and refers to the work required to place, synchronize and move layers within it. Further Bowersox et al [3] emphasized that logistics, by coordinating and positioning warehouses, is one value-creating process. This means that products or services have no value if they are not available at the right time and in the right place when a customer wants to consume them [1]. The structure and strategic results of the supply chain are determined by how well the company operates and adapts to customers as well as the supporting distributor and supplier networks in order to gain competitive advantage [3].

SCM is about how external and internal material processes are handled by the company, thus one can divide it between outgoing and incoming material flow. The outgoing material flow deals with how finished products are distributed to customers while the incoming material flow covers all activities required to optimize freight flows from suppliers to the point of consumption within the own company [27]. Through targeted improvements of the whole or parts of the supply chain, for example, routine costs can be reduced, the quality of the flow of information is improved and inventories are reduced and in this way the company can create sustainable competitive advantages [1]. The supply chain can be seen as a long value chain built up of a number of companies in a row where the material goes from raw material to end product. Every single company's value chain builds up and contributes to how well the entire supply chain performs [1].

2.1.2 Porter's Value Chain

A chain is never stronger than its weakest link is a well-known saying that can be applied when activities are analyzed. One of the important links in the business chain is the purchasing and delivery function, especially considering the amount of money involved in purchasing decisions [27]. Porter's value chain breaks the company down into nine activities with the intention of trying to understand the costs of a specific department and of finding potential sources of competitive advantage [11]. The activities are divided into primary and supportive activities (See Figure 3). Primary activities are performed in the company's physical flow, including logistics for inflow, manufacturing, logistics for outflow, marketing and service in a manufacturing company. Supporting activities are activities that must be carried out in order for the primary activities to be carried out, for example, product development, investments, administration and business management.

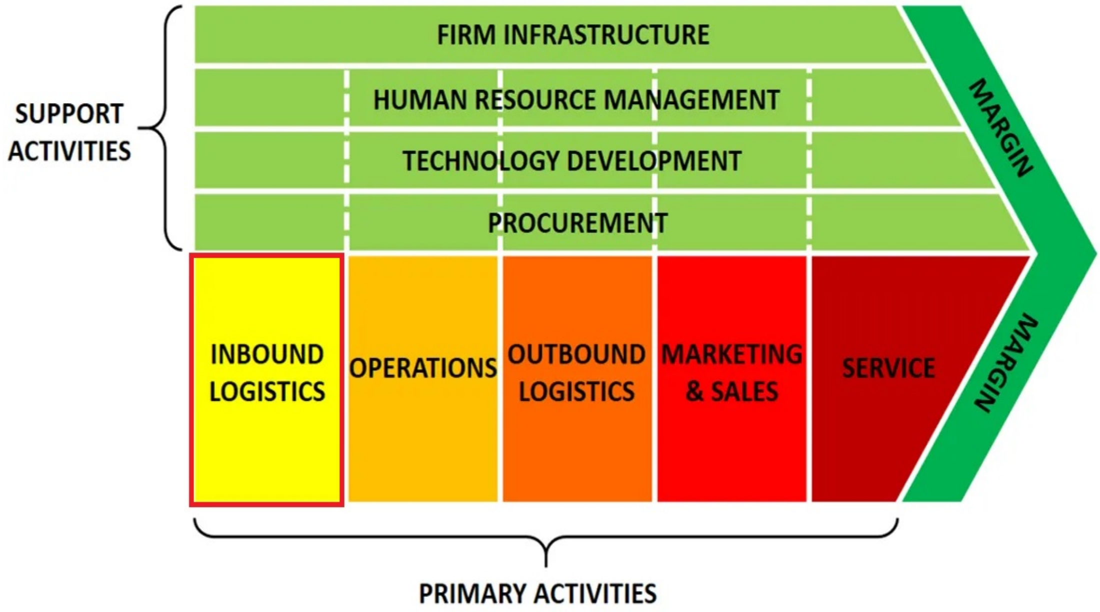


Figure 3: *Porter's Value Chain*

On the whole, Porter believes that the difference between the total value of the product and the total cost of all activities within the value chain constitutes the margin, and through good performance through the chain, a company can create sustainable competitive advantages [17]. In this study, the focus is on the area of inbound logistics, which involves relationships with suppliers and includes all the activities required to receive and store externally purchased material and internally shipped material. This can also be seen in Figure 3, where the red marking illustrates the focus area in this study.

2.1.3 Inbound Logistics

Inbound logistics according to Bowersox et al [3] include purchasing and organizing incoming materials or finished goods from supplier to warehouse, manufacturing or directly to store. It is one of the primary processes and disruptions that can lead to a company not being able to meet the requirements and obligations they have towards the end customer. This is because a large part of the disruptions in outbound logistics stem from disruptions in incoming logistics [23]. Furthermore, the author believes that the disturbances are dependent on the company's inventory levels. For a company with, "lean thinking", therefore with reduced inventories, this problem gets one greater impact. It is therefore a crucial task for the company to achieve an appropriate balance between inventory and the occurrence of disturbances. It is through this balance that a company can achieve the best result in a competitive business environment [25]. During the time it takes to manufacture a product, the product is in transport for a much longer time than it is processed purely physical. This means that incoming transports are an important part of the incoming logistics, at the same time as the transports link supplier and customer [13].

2.1.4 Incoming transport

Procurement of materials has long been an important function and will probably continue to be so in the future, as it is directly linked to the inflow into the company [21]. Production is dependent on inflows, which makes transport one of the most important activities in incoming logistics [10]. According to Stock and Lambert [21], companies should separate between incoming and outgoing transport, as there are three major differences. First, it is considered market demand, which generates outbound shipments, generally to be uncertain and fluctuating. The production flow, on the other hand, which generates the incoming transports, is more stable and easier to forecast. This is because companies often plan their production to be able to meet demand in the best way, which means that they can thus create a relatively good idea of the production flow. Because of this, decisions regarding incoming transports are not same as those that arise for the outgoing transports, which means that they should be distinguished. Secondly, there are probably differences between the movement of raw materials and finished goods in form of handling, size, weight, etc. which affects the mode of transport [21]. Finally, companies generally have less cost control over incoming shipments as they tend to look only at the total delivery cost. In the case of Cytiva the company has no visibility between their manufacturing sites and their PDPs, rather the focus from the supply chain team is laid on the outbounds only. Hence, this is the reason of the visibility not been investigated within the company. Something that also can be identified is that the inbound delayed performance is one of the big reasons for why the in-transit performance is as of today 79% versus the 95% of the company's target.

Incoming transports are not analyzed as often and deeply as outgoing transports, despite the fact that there are opportunities to save costs and streamline, by creating synergy effects between these. It has become almost impossible to continue with the traditional division of incoming and outgoing transports, from a financial perspective. A merger can in addition of cost savings also have positive effects from an environmental perspective. It is however, important to note that new governance and management may require organizational restructuring and consolidation of divisions [24].

2.1.5 Carrier

The logistical role of transport has changed dramatically over the last three decades, and is no longer seen only as a movement of products from one place to another, but as slightly larger. Through deregulation and technological development, conditions have been created for transport companies to offer a large number of services. It can for example, be real-time positioning of freight transport in the supply chain or advanced delivery information or tailor-made transport solutions. Transport is the operational field of logistics, which moves and positions warehouses geographically [3]. A company's need for transport can be managed on three basic levels; arranging transport with its own equipment, outsourcing the transport or using several companies that provide different transport services depending on what is required for the current shipment. However, taking care of the transports with own equipment is in in many cases an excluded option as companies from a financial point of view cannot compete with companies that only deal with this [2]. When using several carriers that provide different transport services, it is set producing company prior to the choice of carrier. This choice has a major impact on a company's transport costs as well as the efficiency of the supply chain, as these are linked for example customers, suppliers and warehouses in the event of a physical flow of goods. This choice becomes especially important for companies in Sweden due to its geographical location with long distances to the global markets.

2.1.6 Delivery terms for incoming transports

Delivery terms are an important instrument for the company's business, both legally and financially. For example, the company's transport finances are affected depending on the choice of conditions. Properly used delivery terms can be valuable instruments with the help of which competitive advantages can be won. Because of this, and that, it is important to make sure that the delivery terms fit into the company's holistic view of warehousing, transport agreements, insurance contracts, internal transport, etc., it is a prerequisite that the relevant staff at the company master the delivery terms [26].

Incoterms is a collection of internationally recognized rules that are used as delivery terms international trade and are established by the International Chamber of Commerce (ICC). These delivery terms are revised regularly to always be adapted to the development of international trade, the latest edition came into force on 1st January 2020 and is a revision of Incoterms 2020. In the current edition there are 11 different delivery terms. These delivery terms are available with different levels of commitments regarding which costs and risks fall on sellers respectively buyer. By referring to these rules, both buyers and sellers can avoid uncertainty arising from interpretations and also reduce the risk of misunderstandings. Incoterms aims to indicate which of the seller and buyer pays for the shipping, takes care of the documentation, takes care of it customs clearance, possibly insuring the goods, risk distribution during transport, etc [8].

2.1.7 Root Cause Analysis

Creating solutions with appropriate methodology is at the core of engineering. To develop a solution to a problem in a careful, analytical and methodological way is desirable to find correct and sustainable solutions.

For a business to be successful in its constant improvement work requires that waste is detected, analyzed and measures taken so that the wastes do not recur [16]. This applies for deviations and problems that do not add value to the end customer, because he is not willing to pay for it. The work of analyzing found discrepancies is about finding the real cause of the deviations, the so-called root cause. When this cause is found, measures can be taken to prevent these (root) cause-related deviations from recurring. It is with a background of this reasoning that the understanding becomes elementary about, why the interest is great of finding the root cause of each deviation.

However, a root cause analysis can be time consuming, and many people may have to work to find these causes. There are many methods and approaches for this work. It is therefore important to be able to prioritize the deviations after, among other things, effort and distribution of the finding, analysis and measures taken. It can therefore be advantageous to mix large and small (more or less time-consuming) deviations for efficient improvement work. The smaller deviations can often be solved by the people closest to the process and operations themselves, while the larger ones may require help from support functions, specialists or improvement groups. Since the time required can be extensive to find the root cause for each specific deviation, the working methodology in root cause analysis is often largely about collecting data about what happened, in order to be able to analyze it in a later phase. Root cause analysis is a central part of the Lean philosophy. [16].

2.1.8 Lean Production

The car manufacturer Toyota's production system, Toyota Production System (TPS), which within Toyota's organization called "The Toyota Way", is what underlies what we today call Lean Production. The Lean concept contains everything from methods and values to tools and philosophy and is about creating added value for the customer by having a way of working and approach that suits the business. An example of a tool is the "5S method" which is one way to have in order within the workplace and "Just In Time" (JIT) which means production and delivery of materials in exactly the quantity and time needed [12].

Lean thinking does not mean resource minimization but resource efficiency and applies to every resources that the company uses, such as people, machines, materials and time. The focus is on creating value for the customer and where everything that provides no value is one potential waste. In order to identify all activities that can be streamlined, it is important that the business is analyzed. Today, the term Lean is used as one collective name for improvement work. Lean was created for the highly competitive automotive industry but is used today in most industries and also more and more in the construction industry. Toyota worked with Lean for decades without documenting the theory behind the system though when the working methods eventually matured and became more complicated, that was all harder to teach. To simplify the teaching, the "Lean Temple" was illustrated. The temple shows the tools, methods and principles used, see Figure 4 [12].

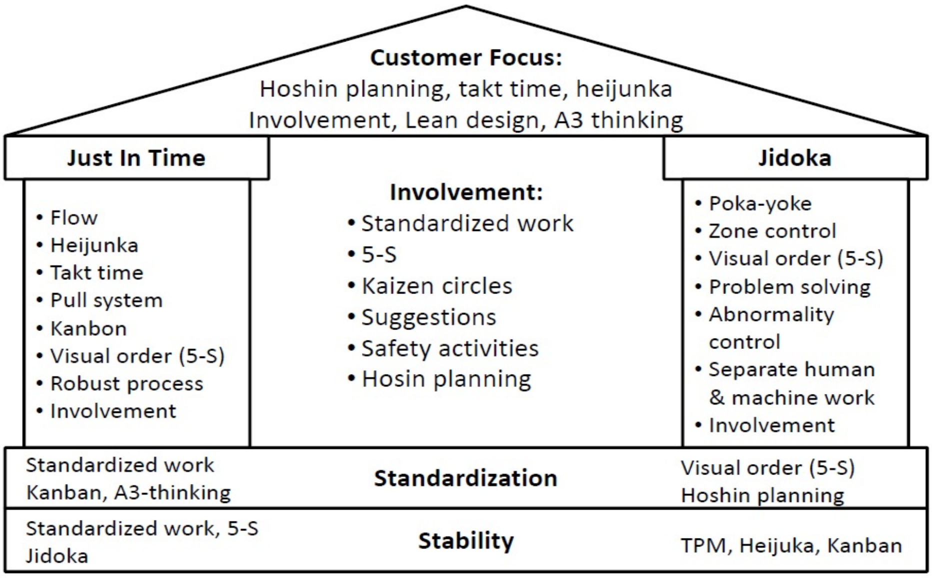


Figure 4: *The various basic principles are illustrated as the supporting elements in the house.*

2.2 Statistical methodology

A forecast is a prediction of some future event or events. As suggested by Neils Bohr, "it's difficult to Make Predictions, Especially About the Future" [19]. Forecasting is an important problem that spans many fields such as economics, industry, medicine etc. Often forecasting problems are classified as long-term, medium-term and short-term. These forecasting problems involve then predicting events only a few time periods such as days, weeks and months into the future. Long-term forecasts can extend beyond months while medium-term forecasts usually extends from 1 to 2 week and short-term forecasts extends for less than a week.

Both medium and short term forecasting are typically based on first identifying then modeling and lastly extrapolating patterns found in the historical data. Historical data sometimes exhibit inertia and do not change dramatically and so for medium and short term forecasting statistical methods are useful. It is usual that forecasting problems involve the use of time series data. The reason that forecasting is so important is that prediction of future events is a critical input into many types of planning and decision-making processes, with application to areas such as industrial process control, finance and risk management or like in our case for planning the warehouse capacity. [5]. But, before building a model and start predicting, one important principle in time series is to see if the data is stationary or non-stationary. The definitions given below in this section are cited from Brockwell & Davis [15].

2.2.1 Time Series Analysis

A time series is a series of data points $\{Y_t, t = 1,2\ldots\}$ that are observed over a given period. The most common type of time series, is a sequence taken at successive points in time with equal distance between the measurements. Examples of time series are weather records, economic indicators, and patient health evolution metrics. Time series analysis is about analyzing time series with the aim of extracting statistics and other characteristic features of the data. Time series prediction is about using a statistical model to try to predict future values based on either extrapolation of historical data or with the help of other time series that are believed to have an impact. Examples of models that can be used for a time series analysis include Auto Regressive (AR) model, Seasonal Auto Regressive Integrated Moving Average (SARIMA) model, and Holt-Winters exponential smoothing model [15].

A prerequisite for being able to use the above mentioned methods is, that the time series is stationary. By stationary, is meant that we have a constant mean, variance and a autocovariance function that is not dependent on time. If the original series would not be stationary one can remedy this, through various transformations. The most common methods to achieve this is to differentiate or logarithmize the data set. To differentiate the series means that differences between observations at time interval k are formed. Differencing the original series Y_t leads to the new series $Z_t = Y_t - Y_{t-k}$. When differentiation is necessary, it is common to differentiate once at time interval 1 and or once at current seasonal distance, as data show seasonal effect. Then, to differentiate as far as necessary to get stationary data but still with caution, so that the differentiations do not lead to data losing connection to reality. New autocorrelation can occur if too many differentiations are made [15].

2.2.2 Stationarity

One of the key roles in time series analysis is played by processes whose properties or some of them do not vary with time. So, if we want to make predictions, clearly we have to assume that something does not vary with time. It is common practice, to assume that either the function itself or one of its derivatives is constant when trying to extrapolate the deterministic functions. So by the assumption of a constant first derivative, it leads to linear extrapolation as a means of prediction. The goal in time series analysis is, to predict a series, that typically is not deterministic but contains a random component. If this component is stationary, then we can use powerful techniques to forecast its future values [15]. For stationary time series there is weak and strict stationarity which will be defined below.

A time series $\{X_t, t = 0, \pm 1, \ldots\}$, is stationary if it has statistical properties, similar to those of the time-shifted series $\{X_{t+h}, t = 0, \pm 1, \ldots\}$ for each integer h. Restricting attention to those properties that depend only on the first and second order moments of $\{X_t\}$, one can make this precise with the following definitions

Let $\{X_t\}$ be a time series with $E\left(X_t^2\right) < \infty$. The mean function of $\{X_t\}$ is

$$\mu_X(t) = E(X_t),$$

and the covariance function of $\{X_t\}$ is

$$\gamma_X(r,s) = \mathrm{Cov}(X_r, X_s) = E\left[\left(X_r - \mu_X(r)\right)\left(X_s - \mu_X(s)\right)\right],$$

for all integers r and s.

So $\{X_t\}$ is **weakly stationary** if

$$\text{(i) } \mu_X(t) \text{ is independent of } t\,,$$

and

$$\text{(ii) } \gamma_X(t+h, t) \text{ is independent of } t \text{ for each } h\,.$$

Strict stationarity of a time series $\{X_t, t = 0, \pm 1, \ldots\}$ is defined by the conditon that, $(X_1, \ldots, X_n)$ and $(X_{1+h}, \ldots, X_{n+h})$ have the same joint distributions, for all integers h and $n > 0$. It is simple to check that, if $\{X_t\}$ is strictly stationary and $X_t^2 < \infty$ for all t. Then $\{X_t\}$ is also weakly stationary. To test if a given time series is stationary or not, one may use the Augmented Dickey-Fuller (ADF) test [15]. Throughout the thesis weak stationarity is used.

2.2.3 Auto Regressive Model

An autoregressive model should explain an observation at time t with a linear composition of previous observations. The general model AR(p) is written as

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t \tag{1}$$

where c is a constant term, ϕ_j is the j:th running average value parameter and Z_t is the random term at time t.

2.2.4 Augmented Dickey-Fuller test

The ADF-test is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The test is a systematic approach for testing the presence of a unit root of an auto-regressive polynomial, to decide if a time series should be differenced or not. The null hypothesis is set that a unit root is present in a time series sample, while the alternative hypothesis is that the time series is stationary. The ADF-test is defined by Brockwell & Davis [15] as follows

Let $X_1, \dots, X_n$ be the observations from an AR(1) model

$$X_t - \mu = \phi_1 \left(X_{t-1} - \mu\right) + Z_t, \quad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right) t = 1, 2, \dots, n \tag{2}$$

where $\left|\phi_1\right| < 1$ and $\mu = EX_t$. The maximum likelihood estimator $\hat{\phi}_1$ of ϕ_1, is approximately $\mathrm{N}\left(\phi_1, \left(1-\phi_1^2\right)/n\right)$ for large n. Normal approximation is not applicable for the unit root case, neither is it applicable asymptotically. This precludes its use the hypothesis of the unit root which is as follows

$$H_0 : \phi_1 = 1, \tag{3}$$

$$H_1 : \phi_1 < 1. \tag{4}$$

To construct a test of H_0 one can write model (2) as

$$\nabla X_t = X_t - X_{t-1} = \phi_0^* + \phi_1^* X_{t-1} + Z_t, \quad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right) \tag{5}$$

where $\phi_0^* = \mu\left(1-\phi_1\right)$ and $\phi_1^* = \phi_1 - 1$. Let now $\hat{\phi}_1^*$, be the ordinary least squares estimator of ϕ_1^* found by regressing ∇X_t on both X_{t-1} and 1. The estimated standard error (SE) of $\hat{\phi}_1^*$ will then be

$$\widehat{\mathrm{SE}}\left(\hat{\phi}_1^*\right) = S\left(\sum_{t=2}^{n}\left(X_{t-1} - \bar{X}\right)^2\right)^{-1/2} \tag{6}$$

where $\bar{X}$ is the sample mean of $X_1, \dots, X_{n-1}$, and $S^2 = \sum_{t=2}^{n}\left(\nabla X_t - \hat{\phi}_0^* - \hat{\phi}_1^* X_{t-1}\right)^2 / (n - 3)$. Then, derive the limit distribution as $n \to \infty$ of the t-ratio.

$$\hat{\tau}_\mu := \hat{\phi}_1^* / \widehat{\mathrm{SE}}\left(\hat{\phi}_1^*\right) \tag{7}$$

with the unit root assumption, that $\phi_1^* = 0$ from where a test of the null hypothesis H_0 can be constructed. For the limit distribution of $\hat{\tau}_\mu$, the quantiles 0.01, 0.05 and 0.10 are -3.43, -2.86 and -2.57. This means that the ADF-test rejects the null hypothesis of a unit root. For an example, for the level 0.05 if $\hat{\tau}_\mu < -2.86$. One can quickly observe that this cutoff value for this specific statistic test is smaller than the standard cutoff, which is -1.645, that is retrieved from the normal approximation to the t-distribution. This is so that the unit root hypothesis is, less likely to be rejected using the correct limit distribution.

The procedure explained above could be extended, to the case where $\{X_t\}$ follows the AR(p)-model with a mean μ as follows

$$X_t - \mu = \phi_1\left(X_{t-1} - \mu\right) + \cdots + \phi_p\left(X_{t-p} - \mu\right) + Z_t, \quad \{Z_t\} \sim \mathrm{WN}\left(0, \sigma^2\right) \tag{8}$$

this can be rewritten as

$$\nabla X_t = \phi_0^* + \phi_1^* X_{t-1} + \phi_2^* \nabla X_{t-1} + \cdots + \phi_p^* \nabla X_{t-p+1} + Z_t \tag{9}$$

where $\phi_0 = \mu\left(1-\phi_1-\cdots-\phi_p\right)$, $\phi_1^* = \sum_{i=1}^{p}\phi_i - 1$, and $\phi_j^* = -\sum_{i=j}^{p}\phi_i$ for $j = 2,\ldots,p$. If the AR polynomial have a unit root at 1 then, $\phi(1) = -\phi_1^* = 0$ and the differenced series $\{\nabla X_t\}$ is an $AR(1-p)$ process. Testing the hypothesis of a unit root at 1 of the AR polynomial is equivalent to testing $\phi_1^* = 0$. As such was done for the AR(1) model described above, ϕ_1^* can be estimated as the coefficient of X_{t-1} in the ordinary least square regression of ∇X_t onto $1, X_{t-1}, \nabla X_{t-1}, \ldots, \nabla X_{t-p+1}$ and so for large n the t-ratio will be

$$\hat{\tau}_\mu := \hat{\phi}_1^* / \widehat{\mathrm{SE}}\left(\hat{\phi}_1^*\right) \tag{10}$$

where $\widehat{\mathrm{SE}}\left(\hat{\phi}_1^*\right)$, is the estimated SE of $\hat{\phi}_1^*$, that has the same limit distribution as the test statistic in (6). So the ADF-test in this case, is applied in the same way as for the AR(1)-model case using the test statistic in (9) and the cutoff values which are mentioned above.

Another approach is to examine information criteria such as Bayesian information criterion (BIC) or the Akaika information criterion (AIC). The ADF-statistic used in the test is a negative number. So, The more negative it is the stronger the rejection of the hypothesis that there is a unit root at some level of confidence [15].

2.2.5 Auto Regressive Moving Average Model

The Auto Regressive Moving Average Model (ARMA) model is defined by Brockwell & Davis. [15] as

Let $\{X_t\}$ be an $ARMA(p,q)$ process, and for every t and if $\{X_t\}$ is stationary then

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{11}$$

where the polynomials $\left(1-\phi_1 z - \ldots - \phi_p z^p\right)$ and $\left(1+\theta_1 z + \ldots + \theta_q z^q\right)$ and $\{Z_t\} \sim \mathrm{WN}\left(0,\sigma^2\right)$

2.2.6 Seasonal Auto Regressive Integrated Moving Average Model

The Seasonal Auto Regressive Integrated Moving Average (SARIMA) model is defined as follows

if d and D are nonnegative integers, then $\{X_t\}$ is a **Seasonal ARIMA(p,d,q)** $\times$ $\boldsymbol{(P,D,Q)_s}$ **process with period** s, if the differenced series $Y_t = (1-B)^d\left(1-B^s\right)^D X_t$ is a causal ARMA process defined by

$$\phi(B)\Phi\left(B^s\right)Y_t = \theta(B)\Theta\left(B^s\right)Z_t, \quad \{Z_t\} \sim \mathrm{WN}\left(0,\sigma^2\right) \tag{12}$$

where $\phi(z) = 1-\phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1-\Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1+\theta_1 z + \cdots + \theta_q z^q$, and $\Theta(z) = 1+\Theta_1 z + \cdots + \Theta_Q z^Q$ [15].

2.2.7 Forecasting Technique for Additive Holt-Winter Triple Exponential Smoothing model

Exponential smoothing is a procedure, for continually revising a forecast in the light of more recent experience. As the observation gets older, exponential smoothing assigns exponentially decreasing weights to that observation. In other words, older observations are given less weight when forecasting, than the recent observations. If the series $\{Y_1, Y_2, \ldots, Y_n\}$, contains both a trend and a seasonality with a period d then a generalization of the forecast function is as

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h}, \quad h = 1, 2, \ldots \tag{13}$$

where $\hat{a}_n$, $\hat{b}_n$ and $\hat{c}_n$ can be thought of as estimates, of the trend level for $\hat{a}_n$, as a trend slope for the $\hat{b}_n$ and as a seasonal component for $\hat{c}_n$ at time n. If k is the smallest integer such that

$$n + h - kd \leq n \tag{14}$$

then we set

$$\hat{c}_{n+h} = \hat{c}_{n+h-kd}, \quad h = 1, 2, \ldots \tag{15}$$

The values $\hat{a}_i$, $\hat{b}_i$ and $\hat{c}_i$ can be found by recursing analogously as

$$\hat{a}_{n+1} = \alpha \left(Y_{n+1} - \hat{c}_{n+1-d}\right) + (1-\alpha)\left(\hat{a}_n + \hat{b}_n\right), \tag{16}$$

$$\hat{b}_{n+1} = \beta \left(\hat{a}_{n+1} - \hat{a}_n\right) + (1-\beta)\hat{b}_n \tag{17}$$

$$\hat{c}_{n+1} = \gamma \left(Y_{n+1} - \hat{a}_{n+1}\right) + (1-\gamma)\hat{c}_{n+1-d} \tag{18}$$

having the initial conditions

$$\hat{a}_{d+1} = Y_{d+1} \tag{19}$$

$$\hat{b}_{d+1} = \left(Y_{d+1} - Y_1\right)/d \tag{20}$$

$$\hat{c}_i = Y_i - \left(Y_1 + \hat{b}_{d+1}(i-1)\right), \quad i = 1, \ldots, d+1 \tag{21}$$

from this, (16) – (18) can be solved successively, for $\hat{a}_i, \hat{b}_i$, and $\hat{c}_i, i = d+1, \ldots, n$ and the predictors $P_n Y_{n+h}$ from (13). The forecast depends on the parameters α, β and γ that can, either be prescribed arbitrarily between 0 and 1, or chosen in a systematic way, to minimize the sum of squares of the one-step errors as follows

$$\sum_{i=d+2}^{n} \left(Y_i - P_{i-1} Y_i\right)^2 \tag{22}$$

obtained when the algorithm is applied to the already observed data [15].

3 Method

This section provides the practical and methodological work used throughout the thesis project. It includes the description of how the root cause analysis were done and how relevant data have been collected and how the goals of this project have been reached. The purpose of this section is to give the reader a more in-depth view on how the practical work have been carried out and what kind of challenges have appeared during the project work.

First, trying to understand the complex supply chain and its inbound was a challenge and the most important thing to do before starting to dive deep into anything at all. This was done by having meetings with my supervisor, a logistics specialist team, sourcing team in different countries such as US and India, meetings with the sustainability team were also done because they had good understanding of the transportation system. Cytiva also has its own “Supply Chain University” where I could find interesting documents to dig deep into and get a greater understanding of the SC within the company. Further, two training sessions were attended to about the main pillars of supply chain within Cytiva. Lastly, warehouse specialists were also involved in retrieving information and creating a good understanding of how everything is done within the company and its inbound. With all this information gathered a high look up figure was made to visualize the complex supply chain. Small changes were then made with the help of the network design & optimization leader and the supervisor within the company and approved, see Figure 1.

3.1 Data Collecting and Preprocessing

3.1.1 Internal & External suppliers

When the processes were understood the next step was to collect relevant data for both internal and external inbounds. In other words gather data from the OEMs, RDPs and the other manufacturing sites. For the RDPs and manufacturing sites the company has its own system Oracle and a data analytics platform Qlik Sense containing data about every internal shipment made and received including its timestamps, quantity, what type of product, product number, part origin etc. All the data collected within the company was imported to Excel and it was about 370 thousand number of lines. This data was collected from when they started collecting data since January 2, 2020 until March 22, 2021. The data retrieved from the analytics platform was not raw but still needed to be preprocessed to make it more understandable in the purpose for this thesis. Further, data was collected from the forwarders with the same time interval as well containing shipments from different carriers that were not included within the dataset retrieved from the system within the company. When all the data required were gathered from both the company’s system and the forwarders, a cleaning process was done to remove missing values and smooth out noise while identifying outliers. The outliers were specific type of products not shipped to the warehouse and this could be identified with the help of the logistic specialist team. These outliers were as well removed. The number of outliers in this case were very small and could be removed without impacting the quality of the data. Next step became to integrate the two data sets from the company and the forwarders and transforming the data into more appropriate categories. Lastly, the data was reduced to only necessary information so that performances of different suppliers shipping from warehouses and manufacturing sites could be calculated and measured against the in-transit time target. The reduction was done in the form of removing unnecessary columns.

For the external inbounds, so called OEMs, oddly enough the company did not have any visibility, so the first challenge was to collect which the carriers was and what incoterms was agreed upon. This was much more difficult than thought, and discussions with sections from everywhere in the company were made to be able to get all the carriers from the original equipment manufacturers. All the carriers were found but time did not suffice to gather all the information as the company's system had for their internal shipments, the idea was though to do the same procedure for the OEMS as describe above for the RDPs and Manufacturing sites.

3.1.2 Warehouse Prediction

For the prediction of quantity, weight and volume received into the warehouse in Rosersberg, more data was required and this could be retrieved by having further access to the system Oracle and using the program language SQL. With help of SQL and the company's own system Oracle, data from 2015 until June 2021 could be retrieved from the company's server. The data was though reduced and only used from January 1, 2018 until March 02, 2021 since the receiving of goods had changed since then.

The data received from the system contained delivery timestamp, item id, amount of quantity received, item unit volume and item unit weight. To be able to get the volume and weight received into UPD, the volume and weight for each product had to be calculated and then multiplied with the amount of quantity received. This was done with help of existing standard weight and volume data for each item id. When this was done the standard volume and weight were set to cubic decimeter and kg, these were recalculated to cubic meter and tonnes. Lastly, the data were resampled from delivery timestamp to days. This means that the amount of quantity, volume and weight received into the warehouse was summed for every day. The amount of observations went from 658 thousand to 1165. The dataset was then divided into three datasets, one containing the quantity, one containing the weight and one containing the volume

Approximately 30% of the data here was found to be incorrect by removing the outliers and comparing it with the original data. The outliers could be identified by visualising the data and identifying the maximum and minimum values of the datasets, and so the data used is questionable. The data collected for the time series analysis were preprocessed by removing all holidays and weekends that the warehouse do not receive any goods. From this there still were a lot of questionable outliers. These outliers seemed to be errors within the company's system, these errors were too many to be just removed from the data set and so discussions with a warehouse specialist were made to get a suitable mean value to replace the outliers with. The errors were also presented to the supervisor and are now an ongoing project for a team to find the reasons for these errors within the system. There were no correlation between the quantity, volume and weight and this most probably due to the number of different products received into the warehouse.

3.2 Root Cause Analysis

To identify the root cause one has to understand the process and its underlying problem. This was done after the data had been collected and preprocessed by calculations that was made in the way of measuring from the time that the shipment of the product started to the time the goods had been ship confirmed and received in the warehouse as follows:

$$\text{Receiving date} - \text{Shipped date} = \text{Actual in-transit time} \tag{23}$$

From this the actual in-transit times was calculated and could now be compared with the standard in-transit times. Standard in-transit times are predetermined by the company based on were the product is coming from. To identify late deliveries as following:

$$\text{Actual in-transit time} - \text{Standard in-transit time} = \text{Delivery Status} \tag{24}$$

This was done for all lines shipped from every supplier from all the internal warehouses and manufacturing sites into the warehouse in Rosersberg. Further, the different supplier's performance were calculated as following:

$$\frac{\text{Total lines on time}}{\text{Total lines shipped}} = \text{Performance} \tag{25}$$

Another thing calculated was the suppliers' impact of their transportations, this was calculated based on carrier shipping from specific manufacturing site or regional distribution point and then divided by the total amount of lines shipped from the specific manufacturing site or RDP. In other words, it was calculated as follows:

$$\frac{\text{Number of lines shipped from specific carrier}}{\text{Total lines shipped from specific manufacturing site or RDP}} = \text{Performance} \tag{26}$$

Lastly, their performances and impact were compiled. The performances were not living up to the standards of the company, this could be identified by comparing the results given from the calculation above with the standards of the company based on every carrier, RDP and manufacturing site. The calculated results were dramatically different from month to month and could go from 80% to 0% within in a month which can be observed from Figure 5 and 6 and so further investigations had to be made. High variation can be seen by the figures mentioned, therefore no statistical tools were used to identify the exact number of variation for each RDP and manufacturing site.

The timestamp from when the shipping starts from each carrier to when the warehouse has received the goods were split into two parts. The first part was from when the shipping started from the specific carrier to when it arrived to UPD. The other part was when the goods had arrived from the supplier to when the goods got ship confirmed and put away into the warehouse. For this, standard times are also set within the company for how long it should take to ship confirm and place the goods into the warehouse. The reason for this split up was due to incoterms and the different contracts between different suppliers' and their responsibilities. The focus became what the reasons were for the late timestamps between when the goods were arrived until the goods were available in the warehouse. The motive for this, is because it is the company's responsibility.

From this a dashboard containing the *Actual in-transit time*, *Delivery status*, *Performances* and the *Impact* of every inbound supplier shipping internal from the manufacturing sites and the other regional distribution points into the Roserberg warehouse were made. The dashboard was made with the help of calculations mentioned above and tools used within Qlik Sense to create the different barcharts. The dashboard where then set with the help of Qlik Sense into updating every 15th minute, so the dashboard becomes a dynamic live report containing the different barcharts. This dashboard was implemented in the company's system Qlik Sense with the approval of the supervisor from the company. The idea of this was so that everyone with access to the dashboard and especially the logistics specialist team in Cytiva can quickly identify the different suppliers' performances, impact and see if improvements are made in the future.

Further, another dashboard was implemented containing the in-transit lines that are due today & not due, in-transit lines overdue and commented in-transit lines that are overdue. The dashboard was made with the help of calculations mentioned above and with tools used within Qlik Sense the different barcharts could be made by using the different categories such as In-Transit lines due today & Not due, commented overdues and lines overdue. For example, the tools within Qlik Sense could search for comments within the database and show that there were no comments for in-transit lines. This dashboard is also implemented as a dynamic live report updating every 15th minute. This was implemented so the logistic specialist team can identify one of their weaknesses which is commenting the reasons for the delays, and so by visualising this for them it will improve their ability to comment. The idea of the comments was that they have to be standardized in a way that reasons for delayed shipments have to be categorized. For example, if weather is the biggest reason for delayed shipments then it is nothing you can impact in comparison to if the largest reason for delayed shipments is that the products cannot be received into the warehouse because of warehouse capacity.

One of the main causes due to the low in-transit performance within UPD was found to be that the in-transit time is measured from when the goods are shipped until it is available in the warehouse and so the shipments can be arrived in time but the warehouse might not be able to receive it. What was identified as one of the causes of delay was the "warehouse capacity" this due to that the warehouse knows how many goods will arrive only within 24 hours and most likely cannot plan in-time to receive the goods arrived. The reason behind why the warehouse do only know when the goods will arrive within 24 hours is because, that is when the carrier informs the warehouse that the product will arrive. The only information existing in the order book is when the products were shipped. Now when one of the causes was identified a possible solution was proceeded.

3.3 Predicting Quantity, Weight & Volume

The solution is thought of, creating a forecasting model over a week for the inbound quantity, volume and weight. In other words, forecast how much will arrive into the warehouse and what is the expected time for the goods to be available in the warehouse. With this tool then the warehouse can plan and receive the goods when they arrive. It will affect the in-transit time performance positively since the in-transit time performance key takes the receiving of the goods in consideration when calculated. This was done with time series analysis prediction and comparing the three models which are explained in Section 2.2.3, 2.2.6 and 2.2.7. The reason of the time series analysis is because the information given within the order books only tell when the product is shipped and due to many late deliveries the standard times that are set by the company are not fulfilled. Therefore, the company also only knows within 24 hours the amount of quantity that will be received. The chosen timeperiod of the dataset were due to three reasons. One because of that is the data i had, two to find seasonality, three due to the datasets being resampled into daily timestamps.

The time series analysis was done in the programming language Python with the help of the libraries *statsmodels* and *pmdarima*. To use proper time series model for each prediction of quantity, weight and volume, first thing to observe was if there is some sort of trend or seasonality for each of the time series data. This was done with the help of an Autocorrelation function (ACF) and an Partial Autocorrelation function (PACF) which is the correlation of a signal with a delayed copy of itself as a function of delay. Informally, it is the similarity between observations as a function of the time lag between them. Fundamentally the pattern of the ACF's and the pattern of the PACF's are used to identify which model might be a good starting model. By plotting both the ACF and PACF one can identify visually if there is any seasonality or trends within the time series. This was done for all three of the time series and gave a good identification for predicting for all three of the time series quantity, weight and volume. When observed this visually, the AR(p) model was interesting to use on all the three time series data, this model is described in Section 2.2.3.

For the quantity data an AR(3) model was set as a good start, for the weight an AR(1) was set as a good start and lastly for the volume AR(4) was set as a good start. Before using the model though, for an AR-model, stationarity is assumed and so a stationarity test needs to be done, in this case an ADF-test was done. The ADF-test explained in the Section 2.2.4 could be done with the help of the function *adfuller* from the *statsmodels* library. The test was done with a constant only regression and an autolag to minimize the AIC. In this case the ADF-test rejects the null hypothesis for the level 0.05 which means that if $\hat{\tau}_\mu < -2.86$. When all three time series had been tested for stationarity the next step became to use a function called *auto.arima* from the *pmdarima* package to fit the best AR(p) model to the univariate time series. This function is based on the commonly-used R function *"forecast:auto.arima"*. In this case the *auto.arima* function gave the lowest AIC for the quantity, weight and volume such as which we earlier had observed from the PACF plots.

For the quantity time series the AR(3)-model was set as following:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + Z_t \tag{27}$$

where c is a constant term, ϕ_1, ϕ_2, ϕ_3 are running average value parameters and Z_t is the random term at time t, with $Z_t \sim WN(0, \sigma^2)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.1 Table A1.

For the weight time series the AR(1)-model was set as following:

$$X_t = c + \phi_1 X_{t-1} + Z_t \tag{28}$$

where c is a constant term, ϕ_1 is the running average value parameter and Z_t is the random term at time t, with $Z_t \sim WN(0, \sigma^2)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.1 Table A2.

For the volume time series the AR(4)-model was set as following:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + Z_t \tag{29}$$

where c is a constant term, ϕ_1, ϕ_2, ϕ_3, ϕ_4 are running average value parameters and Z_t is the random term at time t, with $Z_t \sim WN(0, \sigma^2)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.1 Table A3.

The results from these models were though very bad and so a Seasonal Auto Regressive Integrated Moving Average (SARIMA) model was tested out. The reason for the SARIMA-model is to include seasonality which is reasonable that the datasets could have. The parameters $(p, d, q)(P, D, Q)s$ for the SARIMA-model were chosen by inspecting the ACF and PACF plot of the residuals of the data which was after the trend and seasonality were removed from the data set of the quantity, volume and weight. Further the *auto.arima* function was also used to find a suitable starting point. From this a grid-search was done to find the most suitable $(p, d, q)(P, D, Q)s$ for each time series.

The final model for the quantity time series was a

$$SARIMA(1,0,2)(2,0,3)_{91} \tag{30}$$

Generally a model with period 91 can be written with the operator polynomials,

$$\phi(B)\Phi\left(B^{91}\right) Y_t = \theta(B)\Theta\left(B^{91}\right) X_t \tag{31}$$

and if we consider our model we get,

$$\left(1 - \phi_1 B\right)\left(1 - \Phi_1 B^{91} - \Phi_2 B^{182}\right) Y_t = \left(1 - \theta_1 B - \theta_2 B^2\right)\left(1 - \Theta_1 B^{91} - \Theta_2 B^{182} - \Theta_3 B^{273}\right) X_t \tag{32}$$

where $X_t \sim WN(0, \sigma^2)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.2 Table A4.

The final model for the weight time series was a

$$SARIMA(2,0,1)(2,1,1)_{315} \tag{33}$$

Generally a model with period 315 can be written with the operator polynomials,

$$\phi(B)\Phi\left(B^{315}\right)\left(1 - B^{315}\right) Y_t = \theta(B)\Theta\left(B^{315}\right) X_t \tag{34}$$

and if we consider our model we get,

$$\left(1-\phi_1 B-\phi_2 B^2\right)\left(1-\Phi_1 B^{315}-\Phi_2 B^{630}\right)\left(1-B^{315}\right) Y_t=(1-\theta_1 B)\left(1-\Theta_1 B^{315}\right) X_t \tag{35}$$

where $X_t \sim WN\left(0,\sigma^2\right)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.2 Table A5.

The final model for the volume time series was a

$$SARIMA(1,1,2)(1,0,1)_{154} \tag{36}$$

Generally a model with period 154 can be written with the operator polynomials,

$$\phi(B)\Phi\left(B^{154}\right)(1-B)\, Y_t=\theta(B)\Theta\left(B^{154}\right) X_t \tag{37}$$

and if we consider our model we get,

$$\left(1-\phi_1 B\right)\left(1-\Phi_1 B^{154}\right)(1-B)\, Y_t=\left(1-\theta_1 B-\theta_2 B^2\right)\left(1-\Theta_1 B^{154}\right) X_t \tag{38}$$

where $X_t \sim WN\left(0,\sigma^2\right)$. The estimates of the coefficients, their standard errors and confidence intervals can be found in Appendix A.2 Table A6.

Unfortunately these models did not give the desired outcome this can be identified by inspecting in the Appendices A.1 and A.2 where most of the points are not within the 95%-confidence interval. One can though observe that the SARIMA-model outperforms the AR-model, and this might be due to seasonality included. When these two models did not give the desired outcome further investigations had to be made. Since the data for all three time series had no clear patterns and short-term predictions will be made, Holt-Winters Triple Exponential Smoothing forecasting technique was found to be useful. Another reason for Holt-Winters technique is due to its suitability in predicting time series with repeated seasonal patterns. It was set as a model with an additive trend and an additive seasonality. To get the most reasonable seasonal period for the data a systematic approach was made by starting from 2 which is the minimum seasonal period and looped up to 365 which is a seasonal period of a year. The reason for the additive seasonality is because it is preferred to have it when the seasonal variations are roughly constant through the series. This technique was used with help of the function *ExponentialSmoothing*. The optimal parameters for the smoothing-level α, the smoothing trend β, the smoothing seasonality γ and the damping trend ϕ were found by systematically minimizing the sum of squares of the one-step errors for each time series data of the quantity, weight and volume. The seasonal periods could not be identified with the help of ACF and PACF-plots they were therefore experimented by trying out with weekly, monthly, quarterly, half years and yearly periods.

For the quantity time series the prediction model was set as

$$P_n Y_{n+h}=\hat{a}_n+\hat{b}_n h+\hat{c}_{n+h}, \quad h=1,2,\ldots \tag{39}$$

the values of $\hat{a}_n$, $\hat{b}_n$ and $\hat{c}_n$ were found from recursions analogous as described in Section 2.2.7 with the estimated parameters $seasonalperiod=91$, $\alpha=0.00$, $\beta=0.23$, $\gamma=0.99$ and $\phi=1.00$. The smoothing parameters and initial estimates for the components have also been estimated by minimising RMSE, in this case $RMSE=746.27$. In

Table 1 below one can observe the actual values y_t, level $\hat{a}_n$, trend $\hat{b}_n$, seasonal $\hat{c}_n$, and the predicted values $\hat{y}_t$ for the fitted model.

Table 1: Applying Holt-Winters method with additive seasonality for forecasting quantity into warehouse in Rosersberg. The smoothing parameters and initial estimates for the components have been estimated by minimizing RMSE ($\alpha = 0.00$, $\beta = 0.23$, $\gamma = 0.99$, $RMSE = 746.27$)

DateTime	y_t	$\hat{a}_n$	$\hat{b}_n$	$\hat{c}_n$	$\hat{y}_t$
2018-01-02	16334.00	17618.04	1.00	0.92	16348.22
2018-01-03	14981.00	17633.08	1.00	0.85	15006.90
2018-01-04	4684.00	17647.91	1.00	0.26	4696.06
2018-01-05	2268.00	17662.52	1.00	0.12	2275.73
2018-01-06	16632.78	17676.92	1.00	0.94	16703.18
...	...	...	...	...	...
2021-03-02	15236.00	20750.65	1.00	0.76	16190.59
2021-03-03	9633.00	20711.59	1.00	0.67	15461.05
2021-03-04	9503.00	20668.63	1.00	0.70	16254.40
2021-03-05	14119.00	20661.81	1.00	0.71	15123.83
2021-03-06	648.00	20562.61	1.00	0.60	16469.59

For the weight time series the prediction model was set as

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h}, \quad h = 1, 2, \ldots \tag{40}$$

the values of $\hat{a}_n$, $\hat{b}_n$ and $\hat{c}_n$ were found in the same way as described for the quantity model, with the estimated parameters $seasonalperiod = 315$, $\alpha = 0.10$, $\beta = 1.00$, $\gamma = 0.60$ and $\phi = 0.83$. The smoothing parameters and initial estimates for the components have also been estimated by minimising RMSE, in this case $RMSE = 2.85$. In Table 2 below one can observe the actual values y_t, level $\hat{a}_n$, trend $\hat{b}_n$, seasonal $\hat{c}_n$, and the predicted values $\hat{y}_t$ for the fitted model.

Table 2: Applying Holt-Winters method with additive seasonality for forecasting weight into warehouse in Rosersberg. The smoothing parameters and initial estimates for the components have been estimated by minimizing RMSE ($\alpha = 0.10$, $\beta = 1.00$, $\gamma = 0.60$, $RMSE = 2.85$)

DateTime	y_t	$\hat{a}_n$	$\hat{b}_n$	$\hat{c}_n$	$\hat{y}_t$
2018-01-02	11.08	13.49	1.00	0.82	11.08
2018-01-03	8.67	13.49	1.00	0.64	8.67
2018-01-04	6.79	13.49	1.00	0.50	6.80
2018-01-05	1.98	13.49	1.00	0.14	1.98
2018-01-06	14.30	13.49	1.00	1.05	14.30
...	...	...	...	...	...
2021-03-02	13.54	14.07	1.00	0.96	18.79
2021-03-03	24.03	14.07	1.00	1.70	22.31
2021-03-04	39.92	14.07	1.00	2.81	14.47
2021-03-05	14.37	14.07	1.00	1.03	29.48
2021-03-06	1.01	14.07	1.00	0.08	14.46

For the volume time series the prediction model was set as

$$P_n Y_{n+h} = \hat{a}_n + \hat{b}_n h + \hat{c}_{n+h}, \quad h = 1, 2, \ldots \tag{41}$$

the values of $\hat{a}_n$, $\hat{b}_n$ and $\hat{c}_n$ were found from recursions analogous as described in section 2.2.7 with the estimated parameters $seasonalperiod = 154$, $\alpha = 0.00$, $\beta = 0.23$, $\gamma = 0.99$ and $\phi = 1.00$. The smoothing parameters and initial estimates for the components have also been estimated by minimising RMSE, in this case $RMSE = 30.00$, which in both cases gave approximately the same α, β and γ values. In Table 3 below one can observe the actual values y_t, level $\hat{a}_n$, trend $\hat{b}_n$, seasonal $\hat{c}_n$, and the predicted values $\hat{y}_t$ for the fitted model.

Table 3: Applying Holt-Winters method with additive seasonality for forecasting volume into warehouse in Rosersberg. The smoothing parameters and initial estimates for the components have been estimated by minimizing RMSE ($\alpha = 0.00$, $\beta = 0.23$, $\gamma = 0.99$, $RMSE = 30.00$)

DateTime	y_t	$\hat{a}_n$	$\hat{b}_n$	$\hat{c}_n$	$\hat{y}_t$
2018-01-02	116.75	114.85	0.99	1.01	116.73
2018-01-03	61.35	114.82	0.99	0.53	61.32
2018-01-04	47.39	114.80	0.99	0.41	47.36
2018-01-05	11.68	114.77	0.99	0.10	11.67
2018-01-06	130.00	114.74	0.99	1.13	129.84
...	...	...	...	...	...
2021-03-02	100.09	87.53	0.99	1.68	177.22
2021-03-03	122.35	87.51	0.99	1.27	104.31
2021-03-04	130.40	87.49	0.99	1.33	108.77
2021-03-05	97.40	87.47	0.99	1.29	123.65
2021-03-06	4.05	87.44	0.99	0.84	118.98

When an appropriate model was found last step became to validate the model. To see how robust the model is different mean values was tried out with a maximum deviation of 10%. This did not impact the results of the models. To validate a prediction model one saves a sample of a test set to compare with what the model predicts. For these models, since a short-term prediction will be made the test set saved was only one week, in this case the recent week. In other words, 5 observerations were saved as a test set. In this case, for the Holt-Winters model, mean percentage average error (MAPE) and Root mean squared error (RMSE) have been used to validate the model. To make the predictions more reliable a 95% confidence interval was made from the predicted values. Lastly, the accuracy of the models are based on the MAPE which is a useful measurement tool for forecasting.

4 Results

This section will provide the results from the root cause analysis and the possible solutions found for the causes.

As stated in the method section first thing to see was the in-transit time performances from the different manufacturing sites into the warehouse located in Rosersberg, which were calculated from the data within Cytiva's database. This could be visualised in Figure 5.

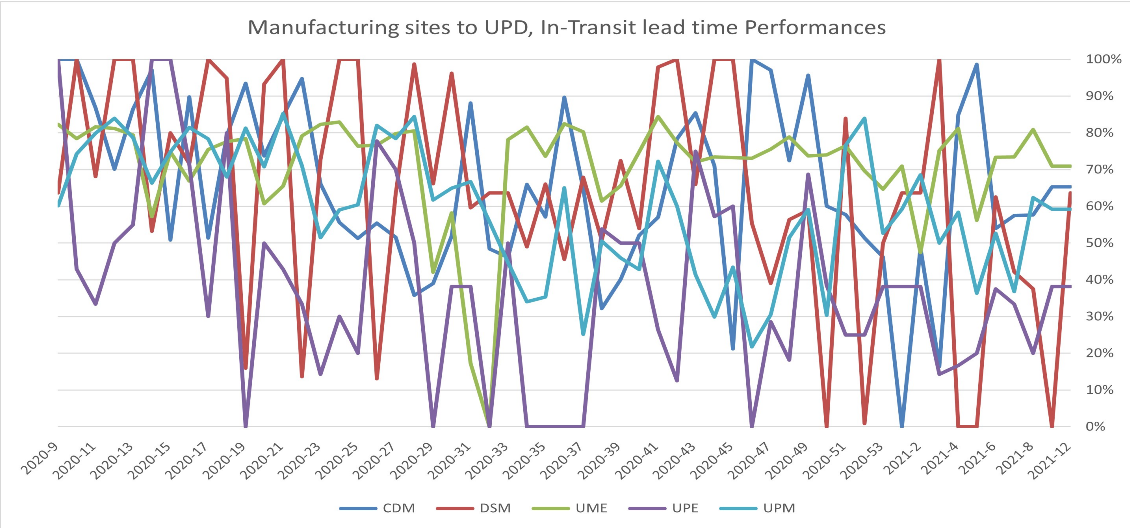


Figure 5: *In-transit lead time performances from manufacturing sites to UPD since 2020 to 2021.*

The same procedure as mentioned above could be done for the other distribution points in-transit time performances into the Rosersberg warehouse, and this could be visualised as in Figure 6.

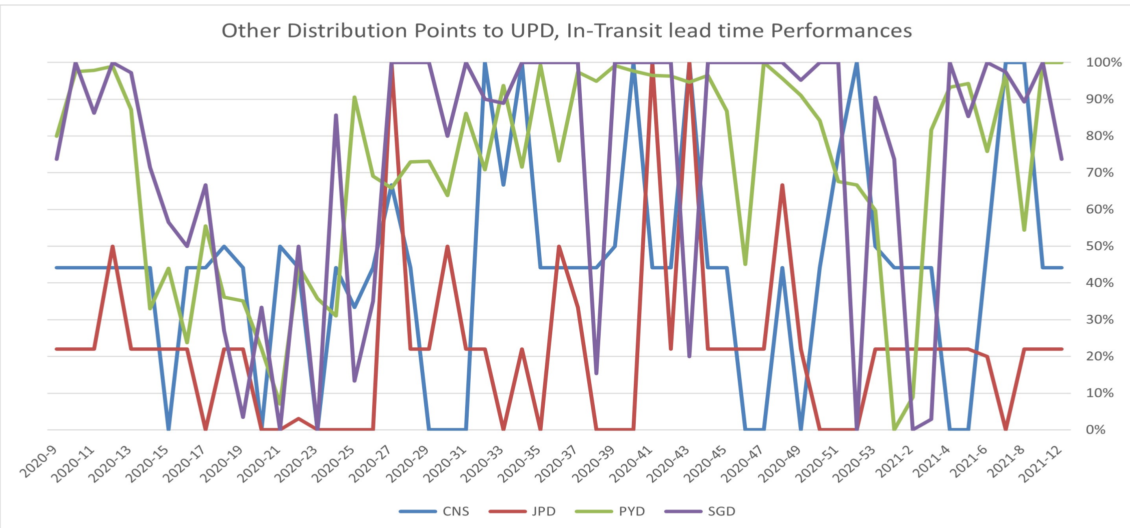


Figure 6: *In-transit lead time performances from other distribution points to UPD since 2020 to 2021.*

These two figures were presented for the logistic specialist team, in the purpose of creating visibility and to show the inconsistency of the inbound logistics of UPD.

In Figure 7 one can observe a snapshot of the dashboard i have created in Qlik Sense and implemented in the system that could be of use for the Logistic Specialist team. From this we can also observe the comment section being empty today. Additional thing to point out here is the importance of the bar chart titled "In-Transit lines Overdue" which gives an hint of which carrier represents the most overdued lines within the inbound.

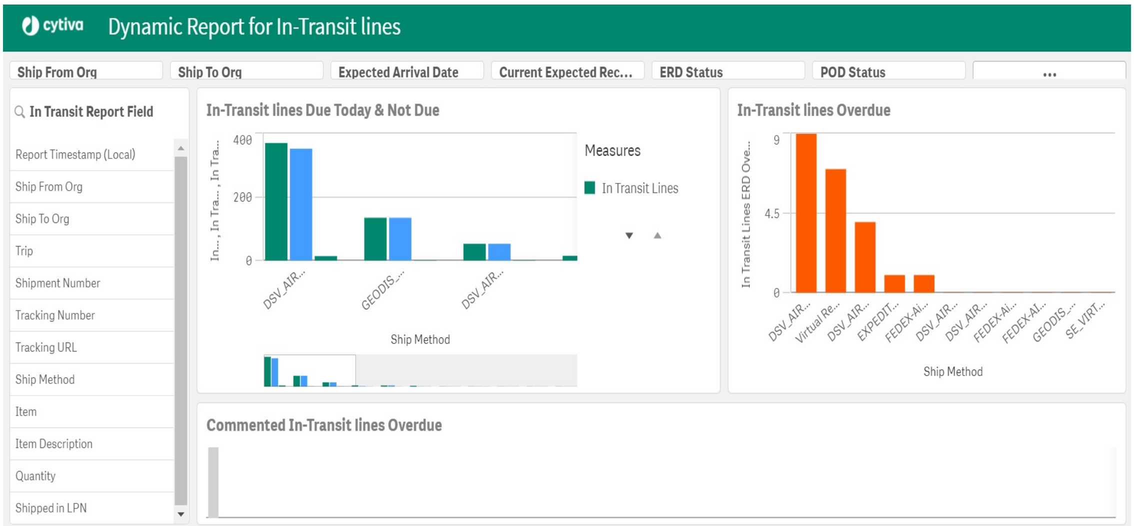


Figure 7: *Snapshot of a dashboard that gives visibility for the In-transit lines that are overdue, not due for the day and lines that are due for the day. In the bar chart to the bottom, the comments for the in-transit lines can be observed.*

In Figure 8 we can observe another dashboard i have created in Qlik Sense and implemented it in the company's system which gives visibility for the current in-transit lines and what their performances are. Further, the pie-chart gives the visibility of which shipping method that have the greatest impact on the inbound in-transit lines. The two barcharts in the bottom show the different shipping methods in-transit time performances and the expected receipt date for the in-transit lines.

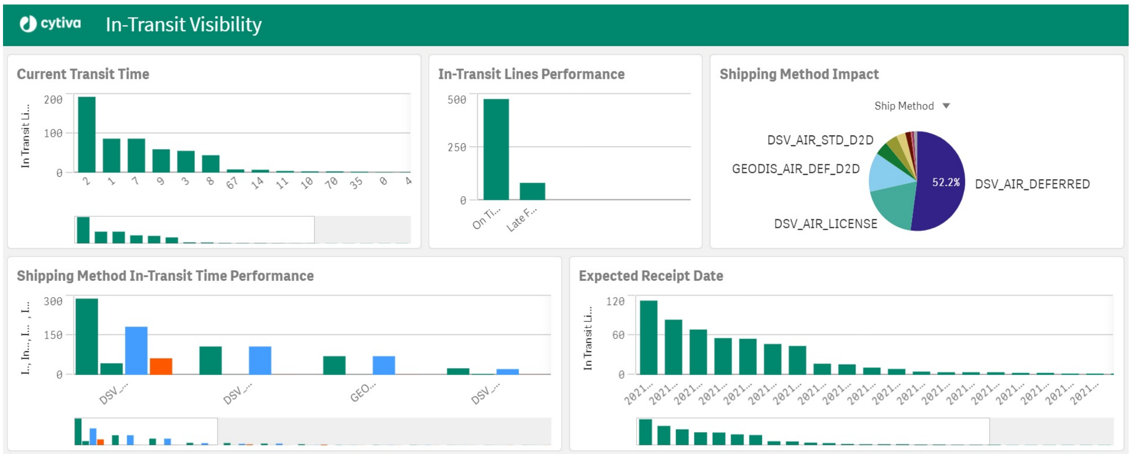


Figure 8: *Snapshot of a dashboard that gives visibility for the In-transit lines performances, the different shipping methods impact and what the shipping methods performances are. In the bottom barcharts one can observe the expected receipt date for the in-transit lines and the in-transit time performances for the different shipping carriers and methods.*

From Porter's Value Chain we know that inbound logistics is one of the five primary activities. By breaking the inbound logistics activity as well into parts as done in this case, and visualised in the figures above, we know this will contribute to finding potential sources of competitive advantages. As mentioned in subsection 2.1.2 Porter believes that the difference between the total value of the product and the total cost of all activities within the value chain constitutes the margin, and through good performance through the chain, a company can create sustainable competitive advantages in this case the focus layed within the inbound logistics.

4.1 Predicting Quantity

In Figure 9 one can observe the time series data of the quantity. From this it is hard to draw any conclusion about trend, seasonality or stationarity at all. Further one can quickly identify that some days very dramatic changes occur or one may question if they were not systematic errors.

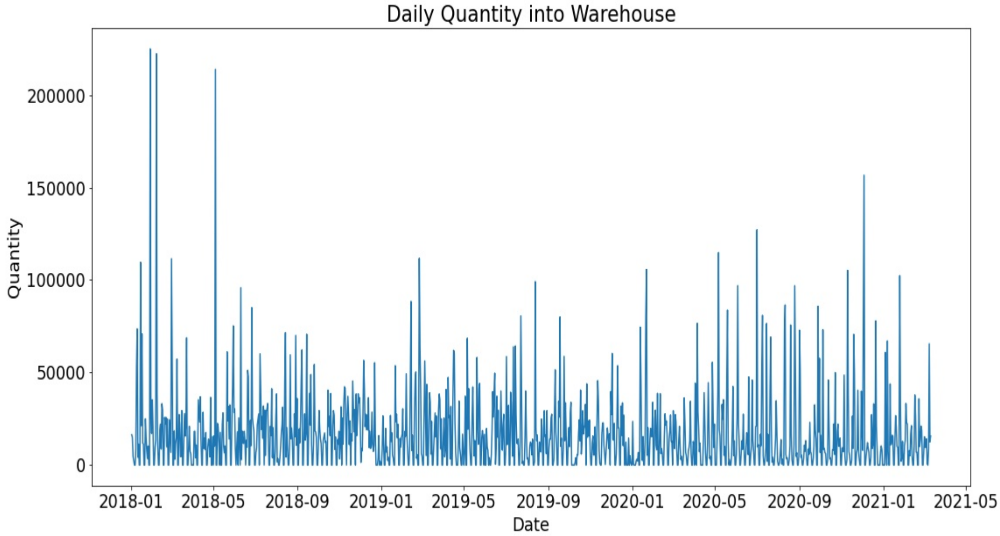


Figure 9: *Time series of the daily quantity received into the warehouse in Rosersberg*

By further investigations of this data and presenting it for the supervisor within the Company. It has now led to an ongoing project to identify why when scanning the goods to identify the quantity it is not uploaded correctly in the system.

After discussing this with the warehouse specialists in Rosersberg a reasonable limit of 50 000 quantity received daily was set and so everything above this amount of quantity was replaced with the mean of the data. Figure 10 presents this data that will be the starting point for the time series analysis of the quantity.

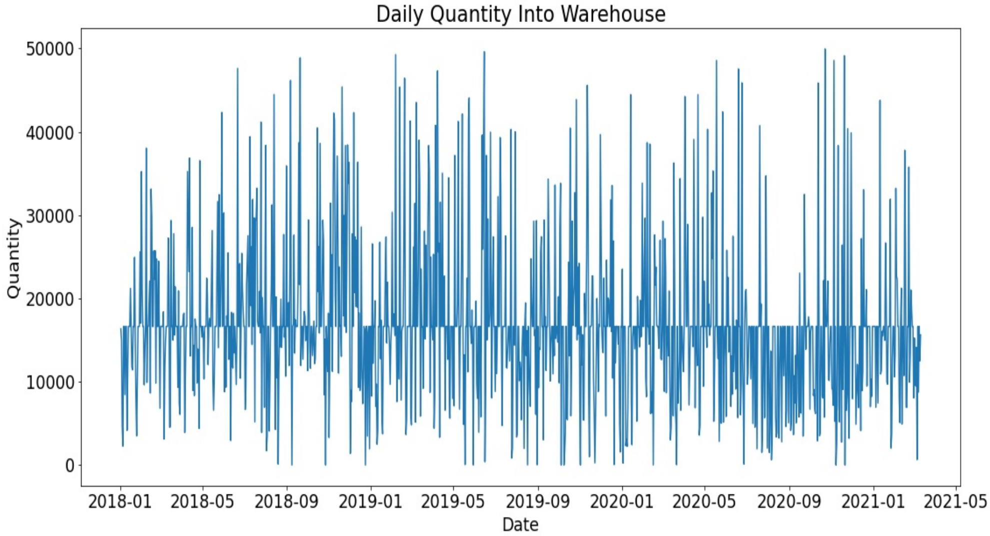


Figure 10: *Preprocessed time series of the daily quantity received into the warehouse in Rosersberg*

From Table 4 one can observe the results from the ADF-test. By inspecting the p-value and from Section 3.1.2 mentioned that the null hypothesis level is set to 0.05, the null hypothesis is rejected in this case since **p-value** < 0.05. In other words, this proves that the time series data for the quantity is stationary.

Table 4: Augmented Dickey Fuller test for the quantity time series data.

Test Statistic	-31.59
p-value	0.00
#Lags Used	0.00
Number of Observations Used	1164

In Figure 11 one can observe 5-day prediction for the quantity into the warehouse, being made with the Holt-Winters forecasting technique mentioned in Section 3.3.

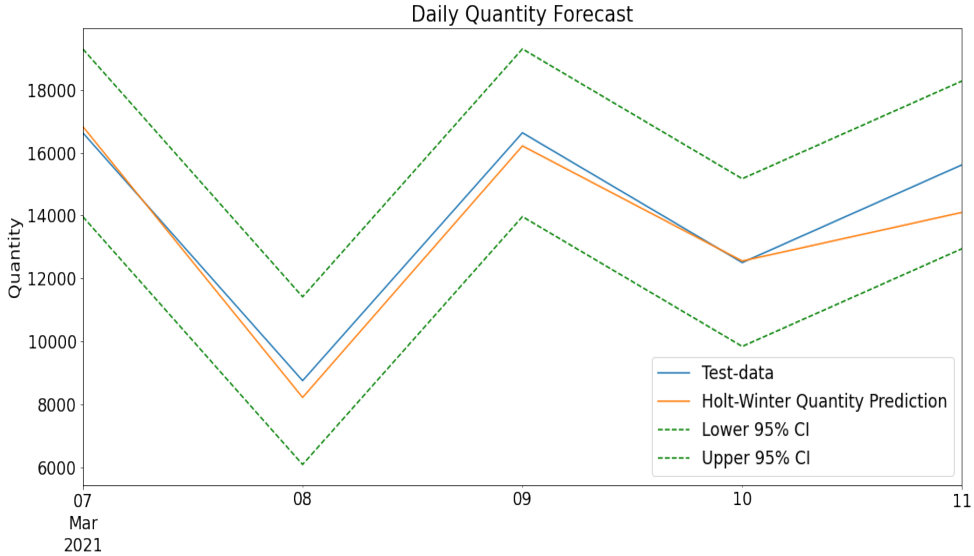


Figure 11: *Holt-Winters model with a 5-day prediction of incoming quantity into the warehouse in Rosersberg.*

The obtained Holt-Winters prediction may give a decent hint about coming weeks quantity received into the warehouse. The validation of this prediction gave $MAPE =$ 3.98%, $RMSE$ = 746 and the 95% confidence interval can be observed in Figure 11 for the upcoming week. This indicates that the prediction model have an accuracy of 96.02% based on the MAPE.

4.2 Predicting Weight

From Figure 12 one can quickly identify the questionable errors from the data here as well. It indicates that the warehouse some days receives weight over 100-200 tonnes and once even receiving over 1500 tonnes. These outliers were discussed with the warehouse specialists about how much a reasonable maximum of received weight during a day could be. This also indicates some problems from the scanning system that the company have, when scanning the weight of the received goods.

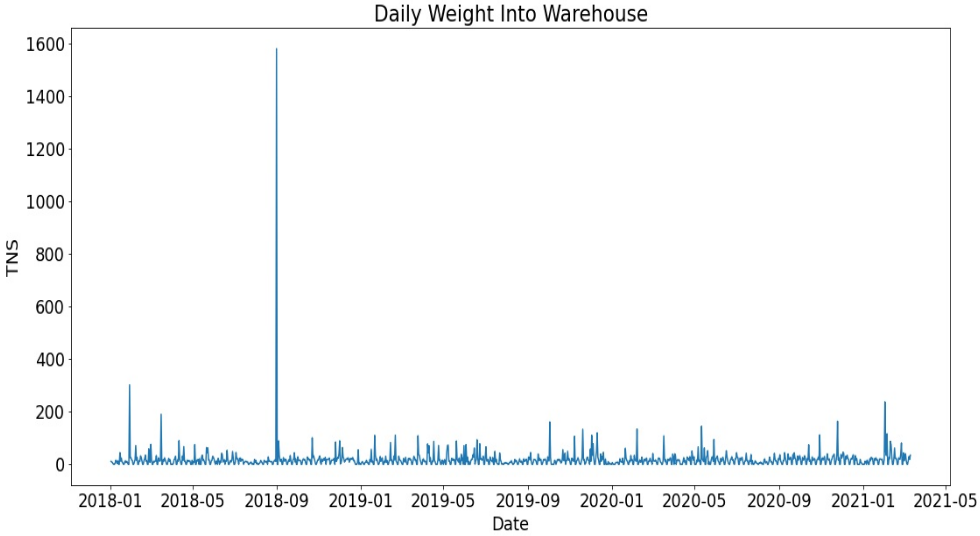


Figure 12: *Time series of the daily weight received into the warehouse in Rosersberg*

Figure 12 and the identified errors have been presented for the supervisor as well and is now also an ongoing project about finding the reason about these dramatically system problems when scanning the products weight today.

For all the errors in this case, everything presenting a daily receiving over 50 tonnes were replaced with a suitable mean of 14 tonnes. One can now see more reasonable data presented in Figure 13 that could be used for a time series analysis for the weight received into the warehouse in Rosersberg.

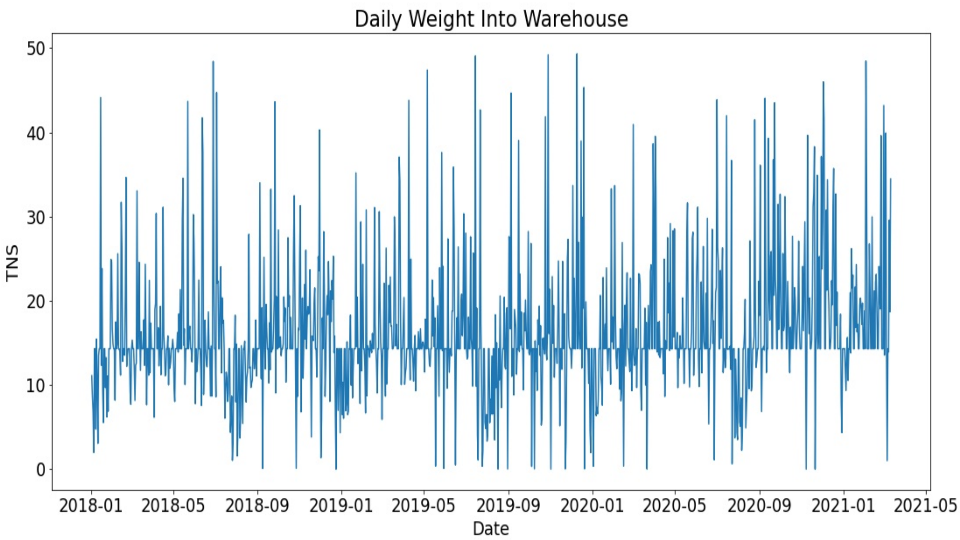


Figure 13: *Preprocessed time series of the daily weight received into the warehouse in Rosersberg*

To know if the time series data of the weight is stationary, an ADF-test was proceeded and the results presented in Table 5 show that the null hypothesis can be rejected since **p-value** < 0.05. Just as mentioned earlier, informally, this means that the time series data for the weight is also stationary.

Table 5: Augmented Dickey Fuller test for the weight time series data.

Test Statistic	-6.24
p-value	4.75e-08
#Lags Used	13.00
Number of Observations Used	1151

In Figure 14 one can observe 5-day prediction for the weight into the warehouse, being made with the Holt-Winters forecasting technique mentioned in Section 3.3.

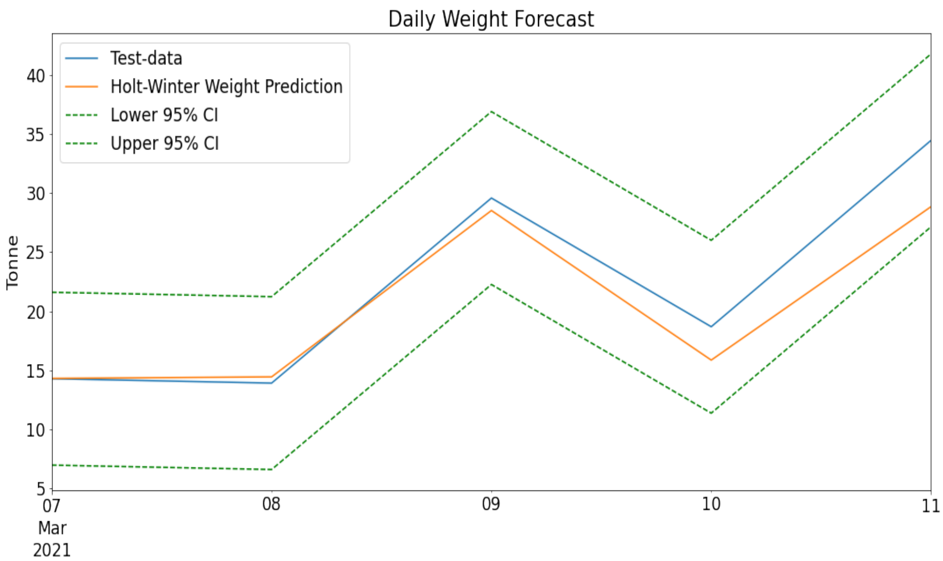


Figure 14: *Holt-Winters model with a 5-day prediction of incoming weight into the warehouse in Rosersberg.*

From Figure 14 one can observe that the model may give a decent hint about coming weeks weight received into the warehouse. The validation of this prediction gave $MAPE = 7.77\%$, $RMSE = 2.85$ and the 95% confidence interval can be observed in Figure 14 for the upcoming week. This indicates that the prediction model have an accuracy of 92.23% based on the MAPE.

4.3 Predicting Volume

Same as mentioned for both quantity and weight received daily into the warehouse, one can question the errors of over 10 000 CO/day. By further investigations of this data and presenting this data as well for the supervisor within the company, additional work were put into the ongoing project of identifying the reasons of why the system interprets the scanned quantity, weight and now volume included so wrong.

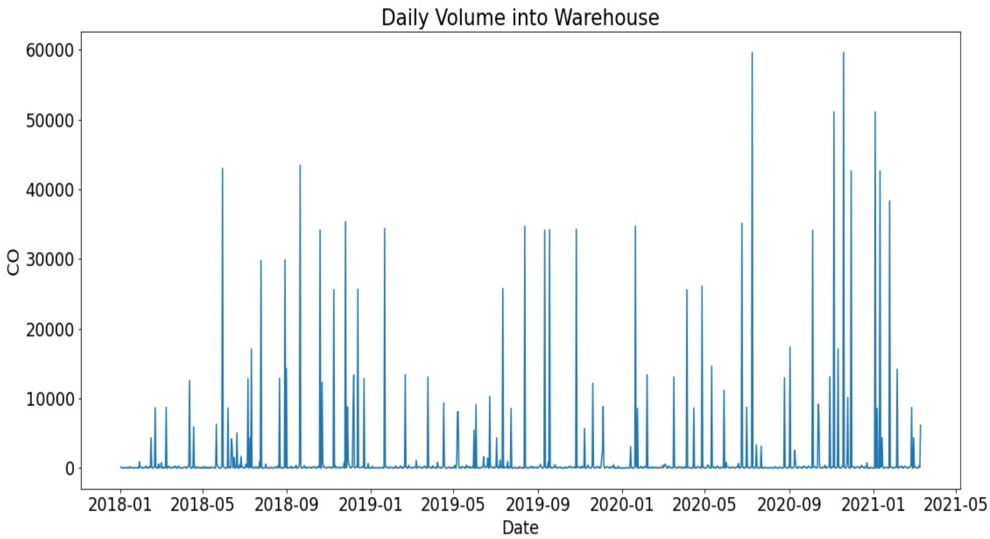


Figure 15: *Time series of the daily volume received into the warehouse in Rosersberg*

Same procedure with the warehouse was done including discussions about what a reasonable limit of the amount of CO in one day could be. When this was done all the outliers over 500 CO were replaced with a suitable mean of 130 CO. Figure 16 presents this data that will be the starting point for the time series analysis of the volume.

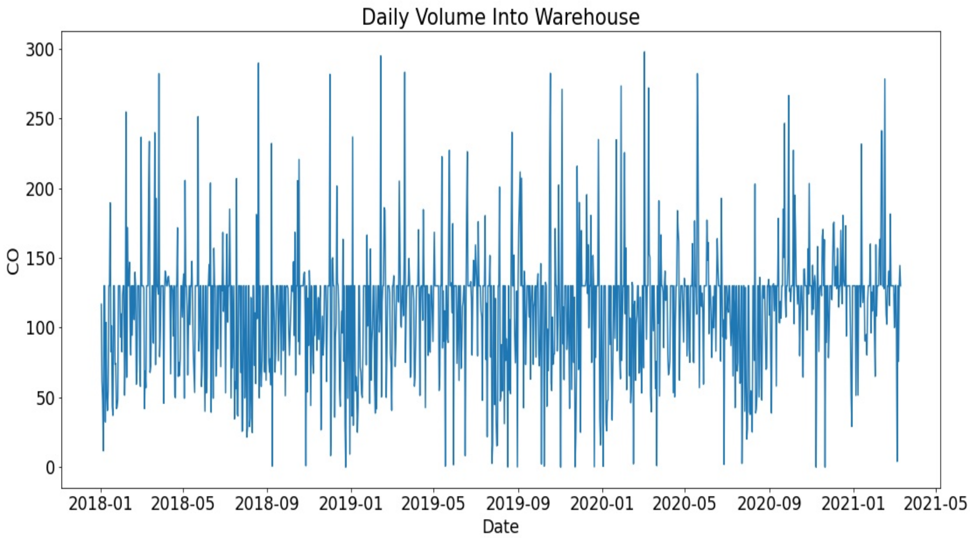


Figure 16: *Preprocessed time series of the daily volume received into the warehouse in Rosersberg*

To know if the time series data of the volume is stationary, an ADF-test was proceeded and the results presented in Table 6 show that the null hypothesis can be rejected since **p-value** < 0.05.

Table 6: Augmented Dickey Fuller test for the volume time series data.

Test Statistic	-8.83
p-value	0.00
#Lags Used	7.00
Number of Observations Used	824

In Figure 17 one can observe 5-day prediction for the volume into the warehouse, being made with the Holt-Winters forecasting technique mentioned in Section 3.3.

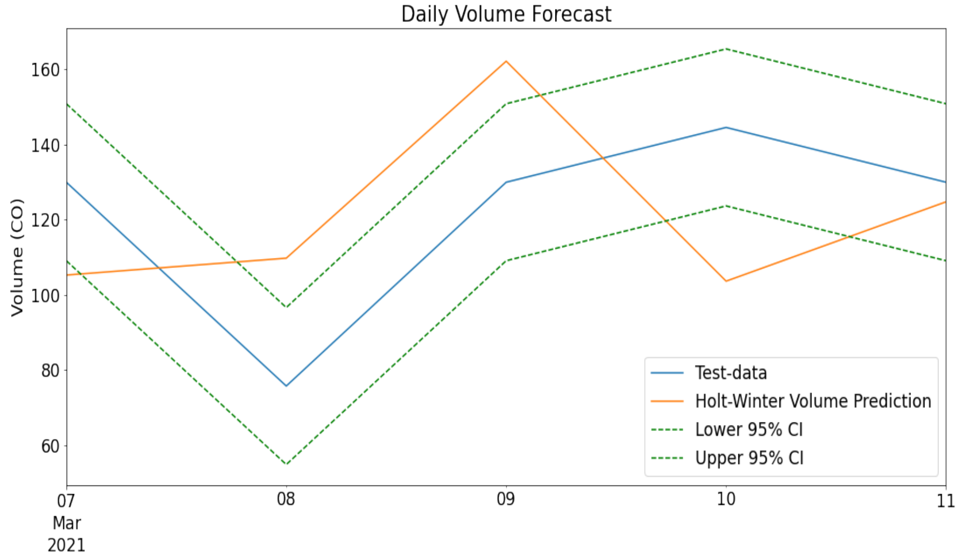


Figure 17: *Holt-Winters model with a 5-day prediction of incoming volume into the warehouse in Rosersberg.*

Figure 17 shows that the model may give nothing more than a hint about coming weeks volume received into the warehouse. The validation of this prediction gave $MAPE =$ 24.18%, $RMSE = 30.01$ and the 95% confidence interval can be observed in Figure 17 for the upcoming week. This indicates that the prediction model have an accuracy of 75.82% based on the MAPE.

5 Discussion and Conclusion

This section will include the conclusions from the obtained results and discussions regarding how Cytiva could implement the outcome from the project in the future.

To understand the complex supply chain in the company is something everyone on site should do. This does not have to be a detailed understanding but at least a higher level of understanding of the whole SCM. One has to understand the importance of the inbound and its impact on the outbound, as of today mentioned in the method, most of the focus is laid on the outbound from the logistic specialist team. As mentioned in Section 2.1.4, many easy points can be collected by directing some focus on the incoming transports and not only the outgoing transports. This indicates that the logistic specialist team has to be trained about the importance of the inbound and its impact on the outbound and direct some focus on the inbound to be able to create a higher level understanding of the supply chain.

By looking at Figure 5 and 6 one can quickly identify that there is no consistency within the in-transit lead time performances from both the PDPs and manufacturing sites. Further, one can identify that there is no visibility at all with the OEMs, this is something that those in charge for that have to take responsibility for. The ability to not even measure the in-transit time performances for the OEMs is something that has to change dramatically. Mentioned in section 3.1.1, the logistic specialist team as of today has no compiled data of who the carries for the OEMs are and what the incoterms are agreed upon. This was something that was investigated throughout the whole project and yet not a fully compiled document was found or could be made from all the collected information from the responsible for the OEMs. To conclude from this as mentioned earlier, responsibilities have to be taken for the OEMs.

Inspecting Figure 7 one can observe the empty comment section of the reasons of the delayed inbound shipments. By creating this snapshot, the logistic specialist team can visualize what they have to improve, in this case for an example commenting the reasons for late shipments. Commenting on deviations must be done in a similar way for everyone, the comments must be standardized, which we see falls back within the framework of logistics in the concept of Lean. Through standardized comments, the system understands better, for example, it will be able to sum up the number of different deviations in comparison with everyone writing in their own way, the system will not be able to read it. Furthermore, standardized comments will also make it easier for employees by letting everyone know what the different comments stand for, so that there is no misunderstanding when someone reads another colleague's comment. It should be in the principle that when you read someone else's comment, it should be as if you wrote it yourself. From this then meetings once every month can be set to discuss the top three commented reasons of delayed incoming shipments and investigate it. Looking at Figure 8 it facilitates for the logistic team to quickly view the actual in-transit time performances and delivery status from both the manufacturing sites and the PDPs. One of the basics within Lean as one can observe in Figure 4 is standardization which is something that has to be worked upon and implemented.

For the data of the quantity, weight and volume for the inbound into the warehouse in Rosersberg it is questionable and as mentioned earlier it is as of today projects going on investigating why the data is having this big errors. By looking at Figure 11 and 14 one can though identify that it is possible to predict how much will come in within the five coming days in an accurate way with the help of Holt-Winters forecasting technique. The results of the validation can also be seen as good for the quantity having $MAPE = 3.98\%$ and $RMSE = 746$ and for the weight having $MAPE = 7.77\%$ and $RMSE = 2.85$. For the volume on the other hand, the prediction was not as good with $MAPE = 24.18\%$ and $RMSE = 30.01$. One of the reasons for this is because the volume data had a lot more errors within it than the quantity and weight data. This also answers a question one might have wondered, why I created three models instead of one, because of the datasets having different outliers in different timestamps. If one model would have been created for this the replaced outliers would have been approximately up to 80% of the total dataset which is not acceptable for a model. Instead of as of today that the warehouse only knows how much will come in within the 24 hours they will know approximately how much will come in within the five coming days. One can quickly identify this by inspecting Figure 16, the same goes for the quantity and weight but not to the same extent as for the volume data and therefore those models predict more accurate. Lastly, one important thing to note for the models is that even though there is no visbility with the OEMs as of today, they are baked in with the prediction model which eases for the warehouse even further.

As of today these models would be recommended as of something the warehouse only can have a glance upon. Though when the investigation project of the data is done and the data collection becomes correct this could be more reliable and of usage.

All in all to conclude from the discussion above and to answer the first specific questions raised in Section 1.3 is that by creating value stream maps, higher level understanding and looking if there is standardized processes implemented one can identify if those working with the supply chain and the in-transit times knows the reasons of the delayed shipments. By this methodology one can identify the root causes throughout how the shipments are handled and identify the visibility which in this case was found to not be that high. Further, with the help of the logistical methodologies one can visualize the in-transit time performances and the reasons of the delayed shipments which in this case was found out to be the receiving of goods from the warehouse and a solution for this was made to facilitate for the warehouse to be able to improve the status of the receiving of the goods. Lastly, the two other specific questions are answered above but to conclude them, one can mention that it might not be any KPI to highlight but more of having tools to create visibility for the inbound that can be used within the daily management system. Further, the implementation of the identified initiatives can be observed in Section 4.

6 Final Thoughts and Recommendations

This section suggests how further studies can be done within the area and in what other context these results can be used and applied. Finally, this part describes the recommendations given to Cytiva based on the work that has been done.

6.1 Further studies

For the transportation systems, further studies could be upon the environmental aspect of only using air freight and the cost savings that could be made from changing the transportation system. Another aspect to investigate would be to follow products and see their transportation routine to identify unnecessary transportations. This thesis has made an attempt of using time series analysis to predict the quantity, weight and volume of incoming goods into a warehouse. Further studies for the Holt-Winters models could include spatio temporal statistics, in order to take into account the spatial information in the data such as different warehouses and delivery sites. Additional studies could also be to investigate the interplay between the quantity, weight and volume to make the models even more precise.

6.2 Recommendation

By creating visibility for the in-transit lines of the inbound my recommendation to the Logistic Specialist team in Cytiva would be to start commenting the reasons of the late in-transit time performances. By doing so, one could then summarize this comments and rank them after how often they occur and for the top three commented reasons start to have meetings about why this reasons occur and what could be done to prevent them from occurring. By doing so, the team has to also direct focus to inbound and not only focus on the outbound. Looking at the snapshot in Figure 7 one can observe an example of how this could look like.

Another recommendation would be to create a dashboard in Qlik Sense to show the logistic specialist team the actual in-transit times versus the in-transit time targets. Further, one could show how every carrier shipping from both the manufacturing sites and other regional distributional points into the warehouse at Rosersberg how their performances are and what kind of impact they have on the total lines that have been shipped from the specific place. Looking at the snapshot in Figure 8 one can observe an example of how this could look like.

Last recommendation would be for the warehouse to be able to plan, as one would like to know approximately what amount of quantity will come in tomorrow, the day after tomorrow or maybe approximately what amount of quantity will come in next week. To do so, my recommendation is to compare the predictions that I have presented with the actual outcomes for the nearest future, to see if it gives some sort of sure instinct. If it so does, one may use it to give the warehouse a heads up that next weekend might be tough. This would improve the in-transit time performance because the loading of the goods and put it on the shelf is included in the in-transit time. This will give the warehouse enough information to be able to plan the warehouse and in that way improve their time to load and transfer the goods. If this does not affect the in-transit time performance. Then the Logistic Specialist team will know for sure the reason of the late in-transit lines, and that would be the carriers. Because that would be the only thing not investigated for the in-transit time performances. These models have to though be used by caution due to the instability of the data and huge amount of errors, but one can observe that it is possible to create accurate predictions which absolutely should be investigated when the data is more reliable.

References

[1] Ballou, R. H. (2004). *Business Logistics / Supply Chain Management*. Upper Saddle River: Pearson Prentice Hall.

[2] Bardi, E. J., Novack, R. A., & Coyle, J. J. (2006). *Management of Transportation*. Mason, Ohio: Thomson/South Western.

[3] Bowersox, D. J., Closs, D. J., & Cooper, M. B. (2010). *Supply Chain Logistics Management*. New York: McGraw-Hill.

[4] Cytivalifesciences, Cytiva, About us. https://www.cytivalifesciences.com/en/us/about-us. [Acessed 2021-02-03].

[5] Montgomery, D.C., Jennings, C.L., & Kulhci, M. (2015), *Introduction to Time Series Analysis and Forecasting*, Second Edition, Wiley.

[6] Europa Commission, Climate Action, Paris Agreement. https://ec.europa.eu/clima/policies/international/negotiations/paris_en. [Acessed 2021-02-04].

[7] Freightos, Shipping Air or Ocean Freight. https://www.freightos.com/freight-resources/air-freight-vs-ocean-freight-making-the-decision/. [Accessed 2021-02-11].

[8] ICC, Incoterms 2020 https://iccwbo.org/resources-for-business/incoterms-rules/incoterms-2020/. [Retrieved - 2021-02-24].

[9] Jonsson, P., & Mattsson, S.-A. (2005). Logistik - Läran om effektiva materialflöden. Lund: Studentlitteratur.

[10] Jonsson, P. (2008). *Logistics and Supply Chain Management*. Maidenhead: McGraw-Hill Education.

[11] Kotler, P., & Armstrong, G. (2008). *Principles of Marketing*. Upper Saddle River: Prentice Hall.

[12] Liker, J (2009). The Toyota Way - Lean för världsklass, 6th edition. Malmö: Liber AB.

[13] Lumsden, K. (2006). Logistikens grunder. Lund: Studentlitteratur.

[14] Paulsson, U., & Nilsson, C.-H. T. (2000). Flödesekonomi: supply chain management. Lund: Studentlitteratur.

[15] Brockwell, P.J. & Davis, R.A. (2016). *Introduction to Time Series and Forecasting*, Third Edition, Springer.

[16] Petersson, P., and Ahlsén, S. (2009). Lean: gör avvikelser till framgång. Part Development.

[17] Porter, M. (1998). *The Competitive Advantage of Nations*: with a new introduction. Basingstoke: Macmillian.

[18] PR Newswire. March 31, 2020. "Danaher Completes Acquisition Of The Biopharma Business Of General Electric Life Sciences; Business Will Be Called Cytiva" (Press release). https://www.prnewswire.com/news-releases/danaher-completes-acquisition-of-the-biopharma-business-of-general-electric-life-sciences-business-will-be-called-cytiva-301032965.html. [Accessed 2021-02-03].

[19] Quote Investigator, It's Difficult to Make Predictions, Especially About the Future, 2013. https://quoteinvestigator.com/tag/niels-bohr/. [Retrieved - 2021-04-15].

[20] Slaper, T.E. & Hall, T.J. The triple bottom line: What is it and how does it work. *Indiana Business Review*, 86.1: 4-8. http://www.ibrc.indiana.edu/ibr/2011/spring/article2.html.

[21] Stock, J. R., & Lambert, D. M. (2001). *Strategic Logistics Management.* Boston: McGrawHill.

[22] Sun, X.Q., Wandelt, S., & Zhang, A.M. (2020). How did COVID-19 impact air transportation? A first peek through the lens of complex networks. *Journal of Air Transport Management*, 89: 101928. https://www.sciencedirect.com/science/article/abs/pii/S0969699720305111?via%3Dihub.

[23] Svensson, G. (2001). The impact of outsourcing on inbound logistics flows. *The International Journal of Logistics Management*, 12(1): 21-35.

[24] Svensson, G. (2002). A conceptual framework of vulnerability in firms' inbound and outbound logistics flows. *International Journal of Physical Distribution & Logistics Management*, 32(2): 110-134..

[25] Svensson, G. (2003). The principle of balance between companies inventories and disturbances in logistics flows. *International Journal of Physical Distribution & Logistics Management*, 33(9): 765-784.

[26] Sveriges exportråd. (1999). Exporthandboken: internationell affärsutveckling för små och stora företag. Stockholm: Sveriges exportråd: Industrilitteratur.

[27] Weele, A. J. (2009). *Purchasing and Supply Chain Management*: analysis, planning and practice. Hampshire: Cengage Learning.

[28] Zacharia, Z.G., Sanders, N.R., & Nix, N.W. (2011). The emerging role of the third-party logistics provider (3PL) as an orchestrator. *Journal of Business Logistics*, 32(1): 40-54.

A Appendices

A.1 Auto Regressive model estimates

Table A1: The estimates of the coefficient and standard error (SE) for the quantity time series.

	coef	SE	p-value
c	1.70e+04	310.089	0.000
ϕ_1	0.078	0.029	0.007
ϕ_2	-0.023	0.031	0.474
ϕ_3	0.040	0.031	0.192
σ_2	8.942e+07	0.018	0.000

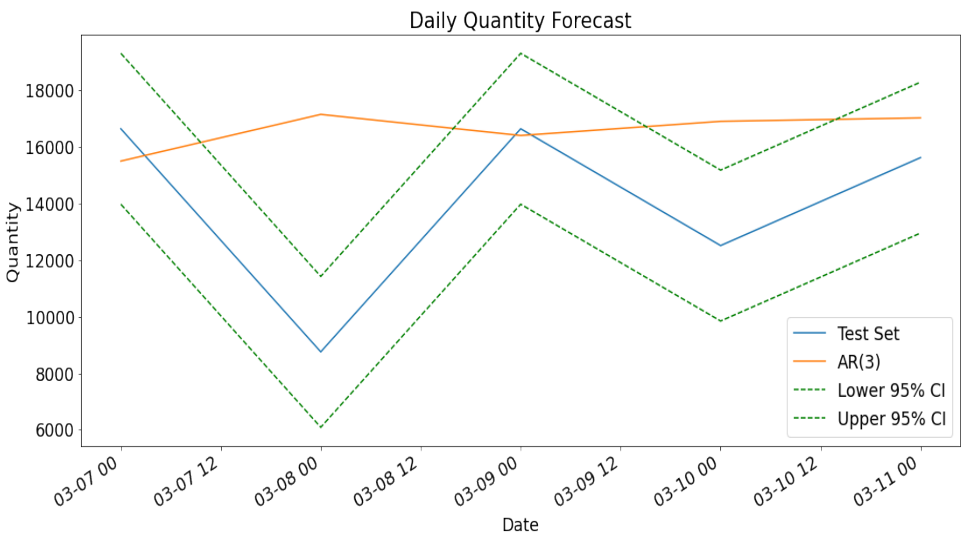


Figure 18: *Prediction AR(3) model for quantity received into UPD.*

Table A2: The estimates of the coefficient and standard error (SE) for the weight time series.

	coef	SE	p-value
c	16.367	0.344	0.000
ϕ_1	0.159	0.028	0.000
σ_2	61.012	1.964	0.000

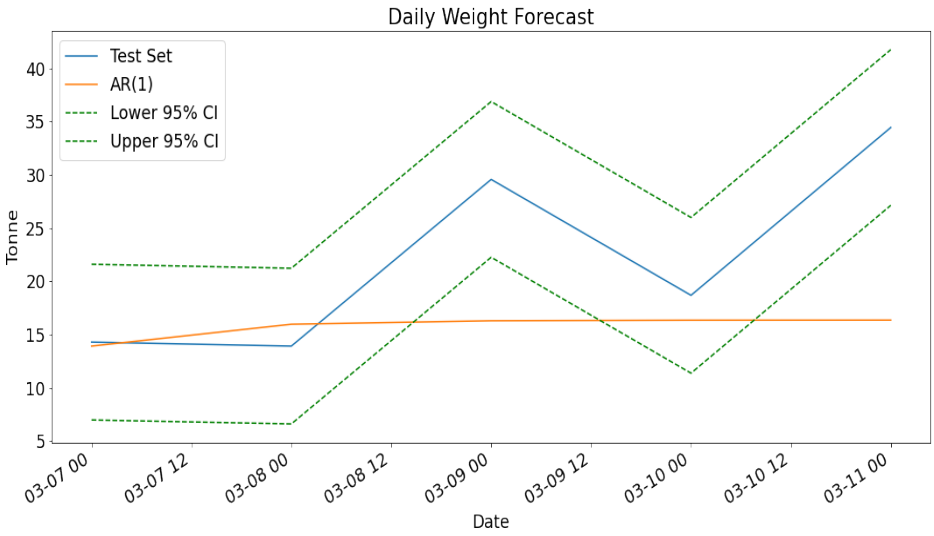


Figure 19: *Prediction AR(1) model for weight received into UPD.*

Table A3: The estimates of the coefficient and standard error (SE) for the volume time series.

	coef	SE	p-value
c	113.281	1.650	0.000
ϕ_1	0.091	0.026	0.001
ϕ_2	-0.009	0.033	0.767
ϕ_3	0.042	0.031	0.187
ϕ_4	0.049	0.030	0.101
σ_2	2101.507	64.625	0.000

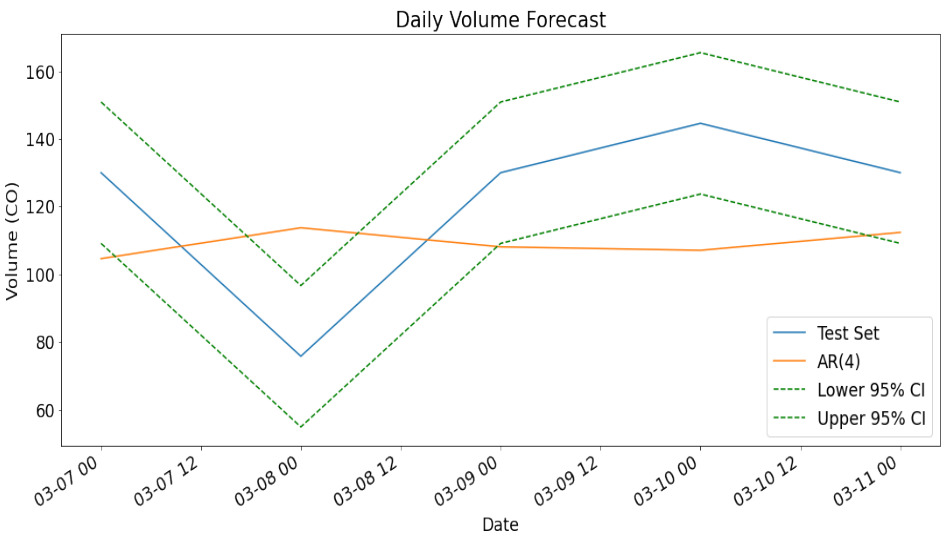


Figure 20: *Prediction AR(4) model for volume received into UPD.*

A.2 Seasonal Auto Regressive Integrated Moving Average model estimates

Table A4: The estimates of the coefficient and standard error (SE) for the quantity time series.

	coef	SE	p-value
ϕ_1	0.997	0.005	0.000
θ_1	-0.895	0.035	0.000
θ_2	-0.086	0.035	0.014
Φ_1	0.012	0.042	0.782
Φ_2	0.988	0.082	0.000
Θ_1	0.059	0.081	0.469
Θ_2	-0.987	0.128	0.000
Θ_3	-0.070	0.039	0.070
σ_2	1.027e+08	6.1e-09	0.000

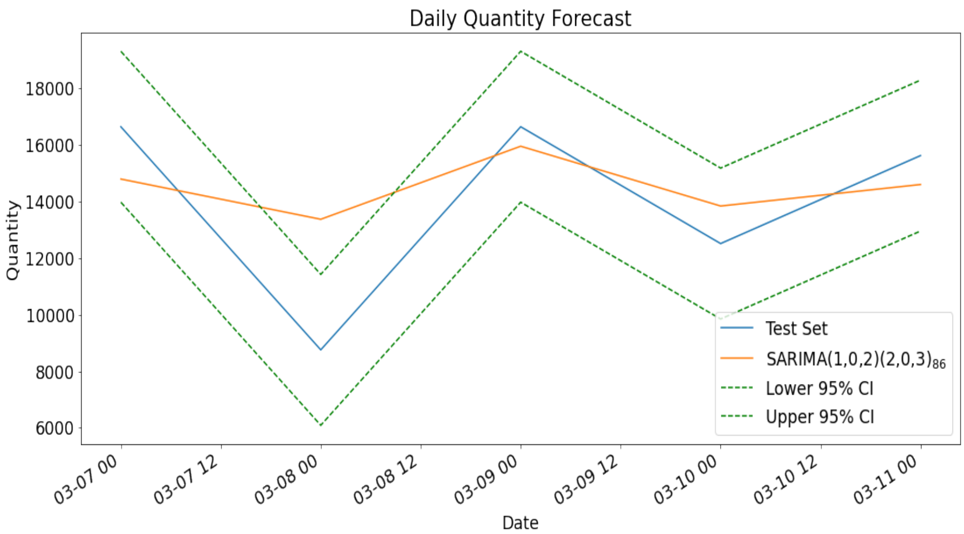


Figure 21: *Prediction $SARIMA(1,0,2)(2,0,3)_{86}$ model for quantity received into UPD.*

Table A5: The estimates of the coefficient and standard error (SE) for the weight time series.

	coef	SE	p-value
ϕ_1	-0.690	0.531	0.194
ϕ_2	0.144	0.075	0.056
θ_1	0.848	0.533	0.112
Φ_1	-0.010	0.037	0.787
Φ_2	0.017	0.036	0.632
Θ_1	-0.981	0.052	0.000
σ_2	62.637	2.783	0.000

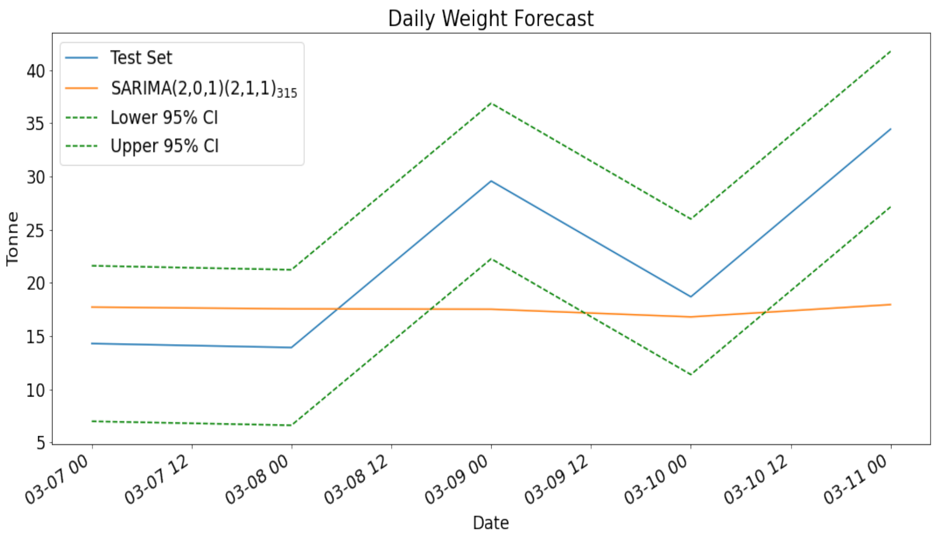


Figure 22: *Prediction $SARIMA(2,0,1)(2,1,1)_{315}$ model for weight received into UPD.*

Table A6: The estimates of the coefficient and standard error (SE) for the volume time series.

	coef	SE	p-value
ϕ_1	-0.381	0.416	0.360
θ_1	-0.526	0.409	0.199
θ_2	-0.427	0.393	0.278
Φ_1	0.906	0.252	0.000
Θ_1	-0.917	0.241	0.000
σ_2	2106.589	61.769	0.000

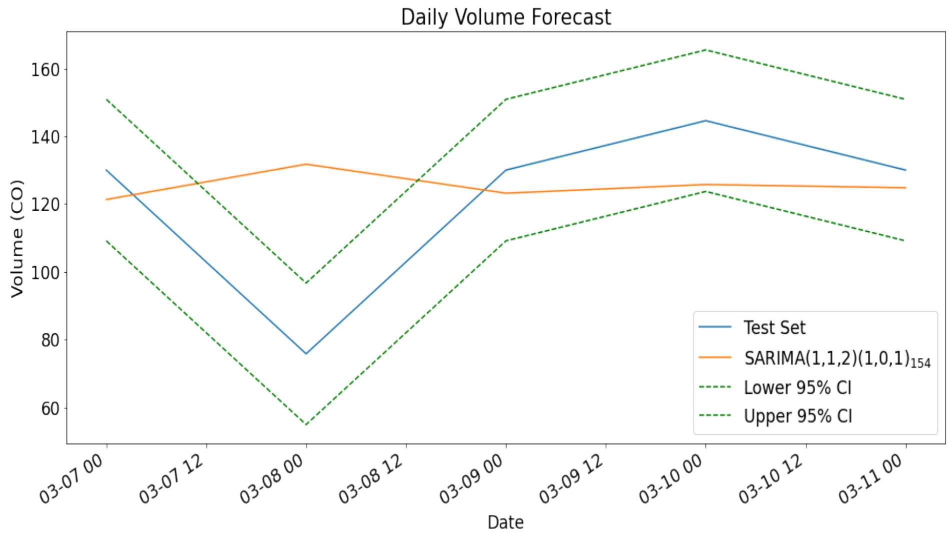


Figure 23: *Prediction* $SARIMA(1,1,2)(1,0,1)_{154}$ *model for volume received into UPD.*

A.3 Source code Data Retrieving, SQL

The code used for retrieving the data from the system will not be made available upon request due to confidential material.

A.4 Source code Forecasting, Python

The code used for time series forecasting for the inbound within the company will not be made available upon request due to confidential material.