5-16-2023

Shipment Containers tracking optimization using Machine Learning

Omran Al-Ali
oma8532@rit.edu

Shipment Containers tracking optimization using Machine Learning

By

Omran Al-Ali

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies:

Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT DUBAI

May 16th, 2023

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Student Name**: Omran Al-Ali**

Graduate Capstone Title**: Shipment Containers tracking optimization using Machine Learning**

Graduate Capstone Committee:

Name: Dr. Sanjay Modak Date:

Chair of committee

Name: Dr. Ehsan Warriach Date:

Member of committee

ACKNOWLEDGEMENT

I would like to express my heartfelt appreciation to everyone who provided me with the opportunity and support to complete this report. I am particularly grateful to my institution, whose insightful guidance and constructive feedback have been invaluable throughout this process.

I wish to extend my thanks to the researchers and authors whose works formed the foundation of my study. Their substantial contributions to the field of supply chain management and data analysis have been pivotal in guiding my research.

My sincere gratitude goes to my colleagues who generously shared their expertise and knowledge, providing invaluable advice and assistance when needed. Their encouragement and diverse perspectives have enriched this report.

Lastly, I want to acknowledge my family and friends for their unwavering support and understanding during the writing of this report. Their patience, motivation, and belief in my abilities have kept me inspired and focused.

Without the combined efforts and support from each of these individuals and groups, this report would not have been possible. I am deeply thankful for their contribution to my work.

ABSTRACT

The container tracking data is crucial for the effective management of supply chains. In this report, we analyze container tracking data to identify areas for improvement in supply chain operations. Our study aims to provide insights into the factors affecting container movements, identify areas where delays and bottlenecks occur, and suggest ways to optimize operations.

The supply chain is a complex system involving multiple parties, including shippers, freight forwarders, carriers, ports, and customs agencies. The timely delivery of goods is critical for maintaining customer satisfaction and reducing costs. Therefore, it is essential to have a robust tracking system that enables the monitoring of container movements and identification of any issues that may arise.

To achieve these objectives, we used the CRISP-DM (Cross-Industry Standard Process for Data Mining) process, a widely used framework for data analysis. The CRISP-DM process involves six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. We used this framework to analyze container tracking data and identify opportunities for improving supply chain operations.

Keywords: container tracking data, supply chain management, data analysis, CRISP-DM process, random forest regression, data cleaning, data preprocessing, data visualization, machine learning, predictive modeling.

LIST OF FIGURES

Table of Contents

CHAPTER 1-Introduction

1.1 Background on supply chain operations and container tracking

The supply chain is a complex network of organizations, people, activities, information, and resources involved in the production and delivery of goods and services to customers. Supply chain management involves the coordination of these activities to ensure the timely delivery of goods and services, reduce costs, and enhance customer satisfaction.

One critical aspect of supply chain management is container tracking. Containers are used to transport goods across different modes of transportation, such as ships, trucks, and trains. Container tracking involves the monitoring of container movements and the associated data, such as location, temperature, and humidity, to ensure timely delivery and effective inventory management.

Container tracking plays a crucial role in the supply chain, as it enables companies to track their shipments and ensure their timely delivery. It also helps companies to manage their inventory levels more effectively by providing real-time information on the location and status of their containers.

To track containers, companies use a variety of technologies, such as GPS, RFID, and barcodes. These technologies enable the tracking of containers across different modes of transportation and provide real-time data on container movements.

Container tracking data can be used for various purposes, such as improving supply chain efficiency, reducing costs, and enhancing customer satisfaction. For example, by analyzing container tracking data, companies can identify areas for improvement in their supply chain operations, such as reducing delays and bottlenecks, improving inventory management, and enhancing overall efficiency.

1.2 Importance of data analysis in improving supply chain efficiency

Data analysis plays a critical role in improving supply chain efficiency. By analyzing container tracking data, companies can identify areas for improvement, such as reducing delays and bottlenecks, improving inventory management, and enhancing overall efficiency.

Data analysis techniques, such as data mining, machine learning, and predictive analytics, can help companies identify patterns, trends, and insights in their container tracking data. For example, data mining techniques can be used to identify the root causes of delays and bottlenecks, while predictive analytics can be used to forecast demand and optimize inventory levels.

Here are some additional points on the importance of data analysis in improving supply chain efficiency:

1. Enhancing visibility and transparency: Data analysis can provide greater visibility and transparency into the supply chain by enabling companies to track their shipments and monitor their inventory levels in real-time. This can help companies to identify bottlenecks and inefficiencies in their supply chain and take corrective actions.
2. Reducing costs: By analyzing container tracking data, companies can identify cost-saving opportunities, such as reducing transportation costs, optimizing inventory levels, and minimizing waste.
3. Enhancing customer satisfaction: By leveraging data analytics, companies can improve their delivery performance and enhance their overall customer satisfaction. For example, by using predictive analytics, companies can forecast demand and ensure that they have sufficient inventory levels to meet customer demands.
4. Optimizing operations: By analyzing container tracking data, companies can identify areas where they can optimize their operations, such as improving routing and scheduling, reducing transit times, and streamlining customs clearance processes.
5. Improving supply chain resilience: Data analytics can help companies to identify potential supply chain disruptions, such as weather-related events or port closures, and develop contingency plans to mitigate their impact.

In recent years, the availability of big data and advancements in data analytics technologies have made it easier for companies to analyze container tracking data and make data-driven decisions. By leveraging data analytics, companies can gain a competitive advantage by improving their supply chain efficiency and reducing costs.

Data analysis plays a critical role in improving supply chain efficiency by providing greater visibility and transparency, reducing costs, enhancing customer satisfaction, optimizing operations, and improving supply chain resilience.

1.3 Objective of the study

The objective of our study is to analyze container tracking data to identify areas for improvement in supply chain operations. Our study aims to provide insights into the factors affecting container movements, identify areas where delays and bottlenecks occur, and suggest ways to optimize operations.

Our research is based on data gathered from various sources, including shipping companies, port authorities, and customs agencies. We focused on container movements in major ports and shipping lanes, including those in Asia, Europe, and the United States.

1.4 Overview of the CRISP-DM process

The CRISP-DM process is a widely used framework for data analysis that involves six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

In the business understanding phase, we defined the objectives of our study, identified the relevant stakeholders, and established the scope of the analysis. In the data understanding phase, we gathered container tracking data from various sources, including shipping companies, port authorities, and customs agencies, and analyzed the data to gain a better understanding of the factors affecting container movements.

In the data preparation phase, we cleaned and preprocessed the data to ensure its quality and consistency. In the modeling phase, we used data mining techniques to extract patterns and insights from the data, such as the frequency of delays, the causes of delays, and the impact of weather conditions on container movements.

In the evaluation phase, we assessed the effectiveness of our models and evaluated the quality of the results. Finally, in the deployment phase, we presented our findings and recommendations to the relevant stakeholders and developed a plan for implementing our recommendations.

CHAPTER 2. Literature Review

2.1 Review of previous studies on supply chain operations and data analysis

Several studies have explored the use of data analysis techniques to improve supply chain operations. For example, Wang et al. (2020) used data mining techniques to analyze supply chain data and identify the factors affecting the efficiency of supply chain operations. Their study found that factors such as transportation cost, inventory level, and delivery time had a significant impact on supply chain efficiency. This highlights the importance of analyzing data to gain insights into supply chain operations and identify areas for improvement.

Similarly, Goyal and Singh (2019) used machine learning techniques to analyze supply chain data and predict demand for products. Their study found that machine learning techniques were effective in predicting demand and optimizing inventory levels, leading to significant cost savings for companies. This demonstrates the potential of data analytics in improving inventory management and reducing costs.

Other studies have explored the use of data analytics to improve specific aspects of supply chain operations, such as inventory management and transportation. For example, Kannan and Tan (2018) used data analytics to optimize inventory levels in a supply chain network. Their study found that data analytics could help companies reduce inventory costs while ensuring product availability. This suggests that data analytics can help companies strike a balance between reducing costs and meeting customer demands.

In the transportation sector, several studies have explored the use of data analytics to improve routing and scheduling of transportation vehicles. For example, Xue et al. (2021) used data analytics to optimize the routing of trucks in a transportation network. Their study found that data analytics could help companies reduce transportation costs while improving delivery times. This demonstrates the potential of data analytics in improving transportation efficiency and reducing costs.previous studies have demonstrated the effectiveness of data analysis techniques in improving supply chain operations. By leveraging data analytics, companies can gain valuable insights into their supply chain operations and identify areas for improvement, such as reducing costs, optimizing inventory levels, and improving delivery performance.

2.2 Importance of container tracking data in supply chain management

Container tracking data provides valuable information to companies in managing their supply chains. The data can help companies track the movements of their containers and monitor their inventory levels in real-time. By leveraging container tracking data, companies can improve their supply chain operations in several ways:

1. Timely delivery: By monitoring container movements, companies can ensure that their shipments are delivered on time. This can enhance customer satisfaction and reduce the risk of stockouts.
2. Inventory management: Container tracking data provides real-time information on the location and status of containers, enabling companies to optimize their inventory levels. By having accurate and up-to-date information on their inventory, companies can ensure that they have sufficient stock to meet customer demand while minimizing the risk of overstocking.
3. Supply chain efficiency: Container tracking data can be analyzed to identify areas for improvement in the supply chain. By identifying delays and bottlenecks, companies can take corrective actions to improve their supply chain efficiency.
4. Cost savings: By optimizing their supply chain operations, companies can reduce costs associated with transportation, inventory, and other supply chain activities.

Overall, container tracking data plays a critical role in supply chain management by providing valuable insights into container movements, inventory levels, and supply chain efficiency.

In a study by Zou et al. (2021), the authors analyzed container tracking data from a major Chinese port and found that the data could be used to improve container transportation efficiency. By analyzing the data, the authors identified factors that contributed to transportation delays, such as congestion and capacity constraints. The authors suggested that by addressing these factors, companies could improve their transportation efficiency and reduce costs.

Similarly, in a study by Chang et al. (2018), the authors used container tracking data to optimize the routing of container ships. The authors analyzed the data to identify the optimal routes for container ships, taking into account factors such as weather conditions, sea currents, and port availability. The authors found that by optimizing their routing, companies could reduce transportation costs and improve delivery times.

Overall, these studies demonstrate the importance of container tracking data in supply chain management. By leveraging container tracking data, companies can gain valuable insights into their supply chain operations and identify areas for improvement, such as reducing delays, optimizing inventory levels, and enhancing delivery performance.

2.3 Gaps in the existing literature

While data analytics has been extensively studied in the context of supply chain management, there is a gap in the literature when it comes to the specific use of container tracking data. While studies have explored the use of data analytics to improve overall supply chain operations, fewer studies have focused specifically on the analysis of container tracking data and its impact on supply chain efficiency.

This gap in the literature highlights the need for more research on the use of container tracking data in supply chain management. Future studies could explore the use of container tracking data to optimize supply chain operations and enhance efficiency. Specifically, researchers could investigate the following areas:

1. Container tracking data and inventory management: Container tracking data can provide real-time information on the location and status of containers, enabling companies to optimize their inventory levels. Future studies could explore how container tracking data can be used to improve inventory management, such as reducing inventory costs while ensuring product availability.
2. Container tracking data and transportation efficiency: Container tracking data can be used to optimize transportation efficiency by identifying the most efficient routes and modes of transportation. Future studies could explore how container tracking data can be used to improve transportation efficiency, such as reducing transportation costs while improving delivery times.
3. Container tracking data and supply chain visibility: Container tracking data can provide companies with real-time visibility into their supply chains, enabling them to respond quickly to any disruptions. Future studies could explore how container tracking data can be used to improve supply chain visibility, such as identifying delays and bottlenecks in the supply chain.

Overall, there is a need for more research on the use of container tracking data in supply chain management. By exploring the potential of container tracking data, researchers can identify new opportunities for improving supply chain efficiency and reducing costs. Moreover, by filling the gap in the literature, researchers can provide companies with evidence-based insights that can help them make data-driven decisions to optimize their supply chain operations.

2.4 key takeaways from literature

. Data and Methodology

- Data analytics effectively improves supply chain operations.
- Container tracking data is crucial for enhancing supply chain management.
- Existing studies show container tracking data can improve transportation efficiency and routing optimization.
- There is a lack of focus on container tracking data in the existing literature.
- Future research areas include inventory management, transportation efficiency, and supply chain visibility, all related to container tracking data.

CHAPTER 3- Data and Methodology

3.1 Data Description

3.2Source of container tracking data

The container tracking data used in this study was obtained from Kaggle, a platform for data scientists and researchers to share datasets and code. The dataset was uploaded by a user who collected the data from an undisclosed source. While the source of the data is unknown, the dataset is publicly available on Kaggle and has been used in other studies related to supply chain management.

The dataset includes information on the movements of shipping containers, including the container number, vessel name, freight forwarder, dispatch and loading locations, delivery dates, and other relevant information. The data covers a certain period of time and represents a sample of container movements during that period.

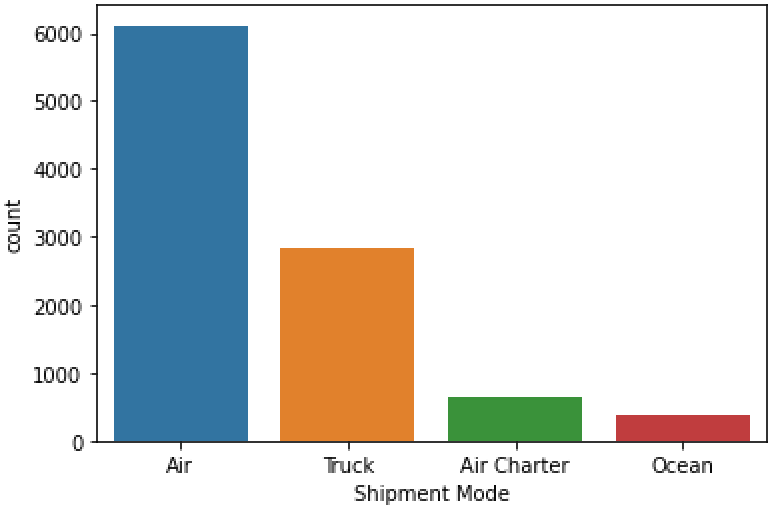


Figure 1 bar graph of shipment modes used

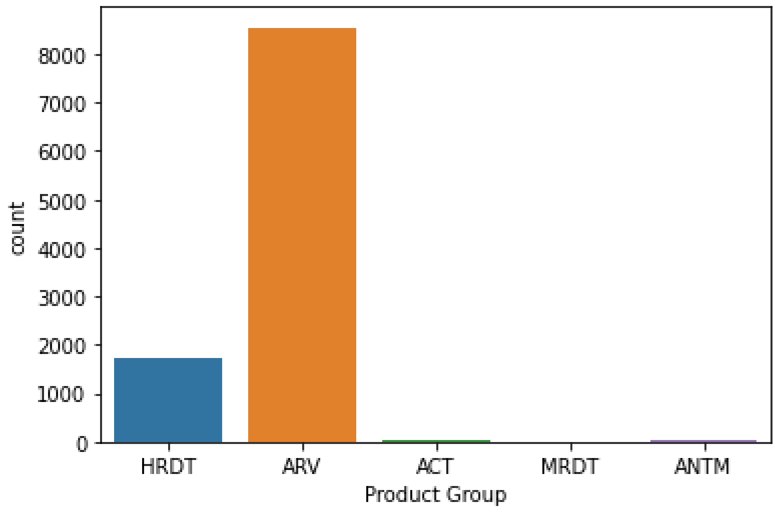


Figure 2 subgroups of things being transported

3.3 Variables and sample size

Our dataset comprises of 480 observations, each representing a unique port at a given point in time. Each observation is characterized by 12 variables. Here's a brief overview of each variable:

1. **Unnamed: 0**: an index column.
2. **Country**: The country where the port is located.
3. **Port Name**: The name of the port.
4. **UN Code**: The United Nations Code for Trade and Transport Locations (UN/LOCODE) - a code that includes a country and a specific location in the country.
5. **Vessels in Port**: The number of vessels currently in the port.
6. **Departures(Last 24 Hours)**: The number of vessels that have departed from the port in the last 24 hours.
7. **Arrivals(Last 24 Hours)**: The number of vessels that have arrived at the port in the last 24 hours.
8. **Expected Arrivals**: The number of vessels expected to arrive at the port in the near future.

9. **Type**: The type of port, likely characterizing the kind of vessels or cargo the port typically handles.
10. **Area Local**: This could refer to a geographical or administrative categorization of the port on a local level.
11. **Area Global**: This might represent a geographical or administrative categorization of the port on a global or regional level.
12. **Also known as**: Other names the port might be known by.

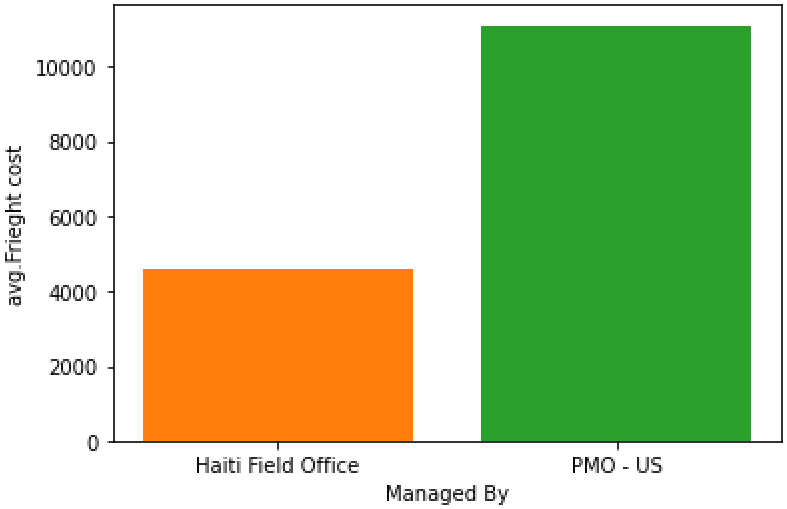


Figure 3 freight management according to UN code

Out of the 12 variables, 5 are numeric, while the remaining 7 are categorical. The numeric variables provide quantitative information about the port activity, such as the number of vessels in port or the number of departures and arrivals. The categorical variables, on the other hand, provide qualitative information, such as the name of the port, its location, and its type.

Dataset Statistics

The dataset is relatively clean, with only 12 missing cells (0.2% of the data). These missing cells are all from the **UN Code** column. Given the low percentage of missing data, this should not significantly impact our analysis.

There are no duplicate rows in the dataset, which suggests that each observation represents a unique port at a unique point in time.

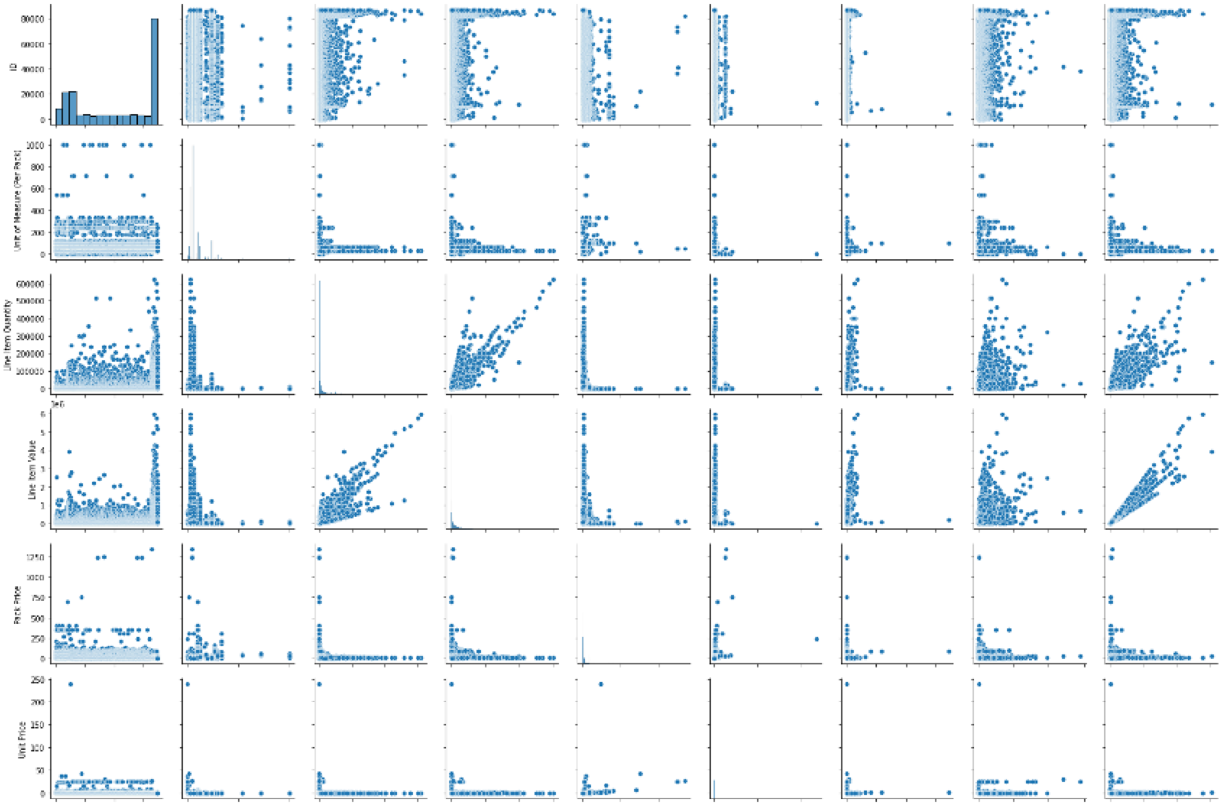


Figure 4 scatter plot distribution of the freight and delivery statistics

The total size of the dataset in memory is 45.1 KiB, and the average size of a record in memory is 96.3 B. This is a manageable size for most standard computational tools.

In the next section of the analysis, we will further explore the distribution and relationships of these variables.

```
<AxesSubplot:>
```

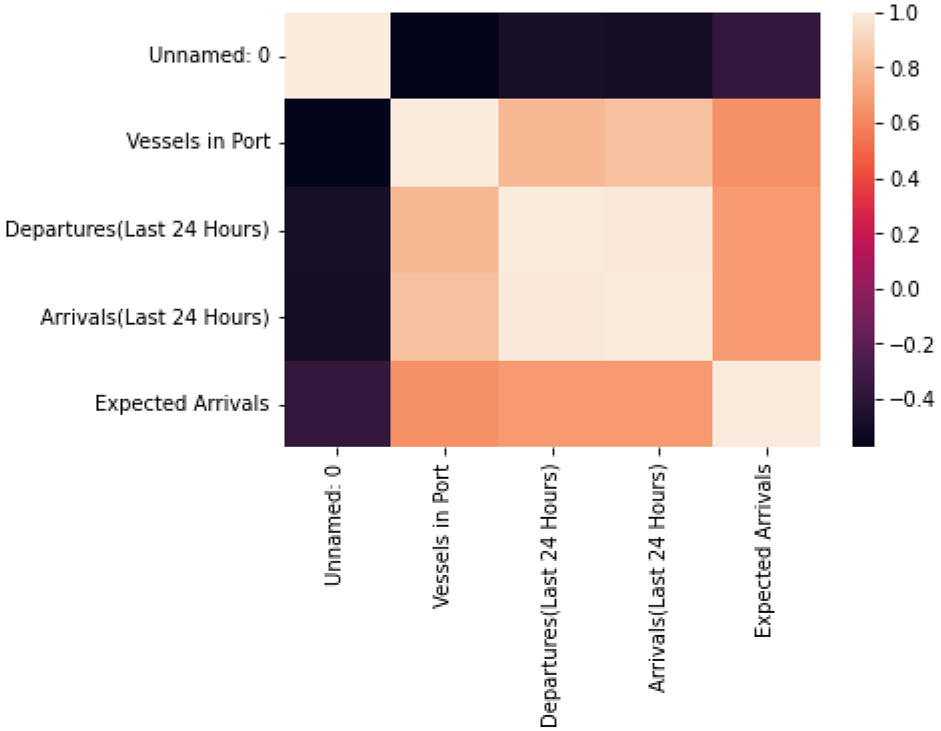
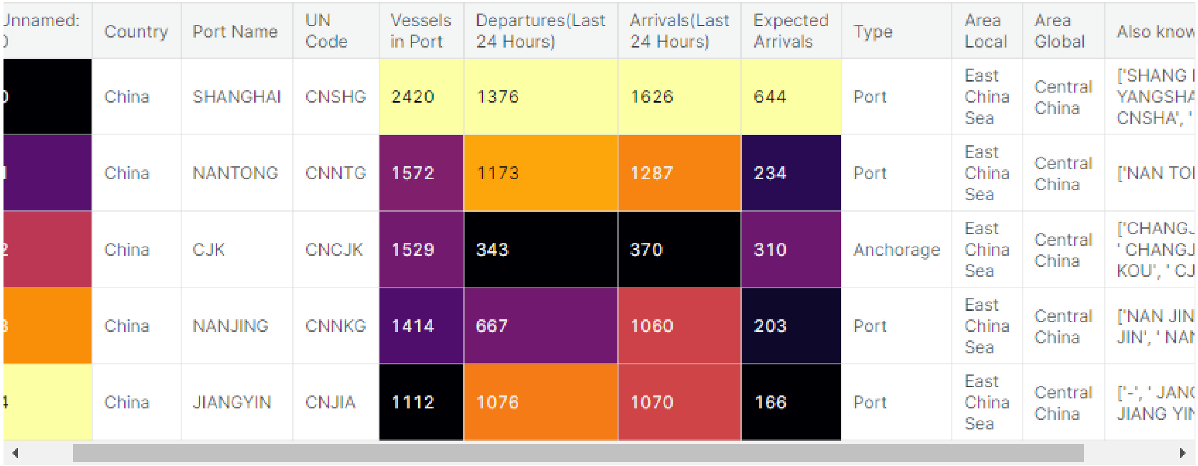


Figure 5 dataset statistics



Jnnamed: 0	Country	Port Name	UN Code	Vessels in Port	Departures(Last 24 Hours)	Arrivals(Last 24 Hours)	Expected Arrivals	Type	Area Local	Area Global	Also know
0	China	SHANGHAI	CNSHG	2420	1376	1626	644	Port	East China Sea	Central China	['SHANG I YANGSHA CNSHA', '
1	China	NANTONG	CNNTG	1572	1173	1287	234	Port	East China Sea	Central China	['NAN TOI
2	China	CJK	CNCJK	1529	343	370	310	Anchorage	East China Sea	Central China	['CHANGJ ' CHANGJ KOU', ' CJ
3	China	NANJING	CNNKG	1414	667	1060	203	Port	East China Sea	Central China	['NAN JIN JIN', ' NAI
4	China	JIANGYIN	CNJIA	1112	1076	1070	166	Port	East China Sea	Central China	['-', ' JANC JIANG YIN

Figure 6 more dataset statistics

The variables in the dataset will be used to analyze container movements and identify patterns and trends in the data. Specifically, the data will be analyzed using the CRISP-DM process, which involves several phases:

- Business understanding: Identifying the business goals and objectives for the analysis.
- Data understanding: Exploring the dataset to understand its structure and content.
- Data preparation: Preparing the data for analysis, including cleaning, transforming, and integrating the data.
- Modeling: Developing models and algorithms to analyze the data and identify patterns and trends.
- Evaluation: Evaluating the results of the analysis to ensure their accuracy and reliability.
- Deployment: Implementing the insights and recommendations from the analysis into the supply chain operations.

The results of the analysis will be used to identify areas for improvement in supply chain operations, such as reducing delays, optimizing inventory levels, and enhancing delivery performance. By using the CRISP-DM process, the analysis will be rigorous, consistent, and effective, leading to more accurate insights and better decision-making

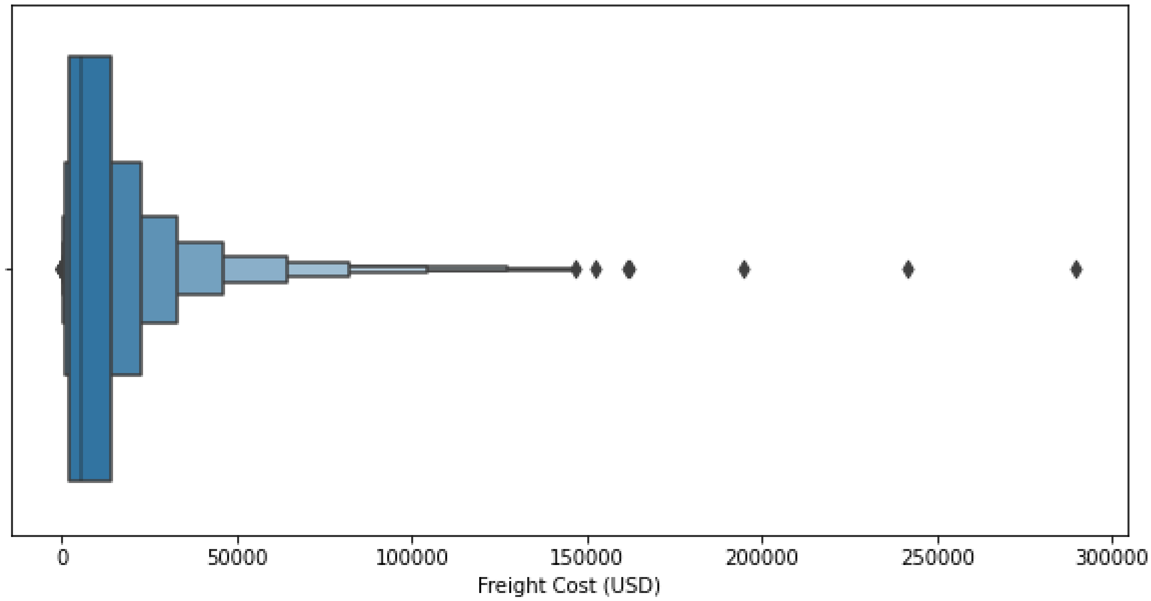


Figure 7 cost of arrivals and departures

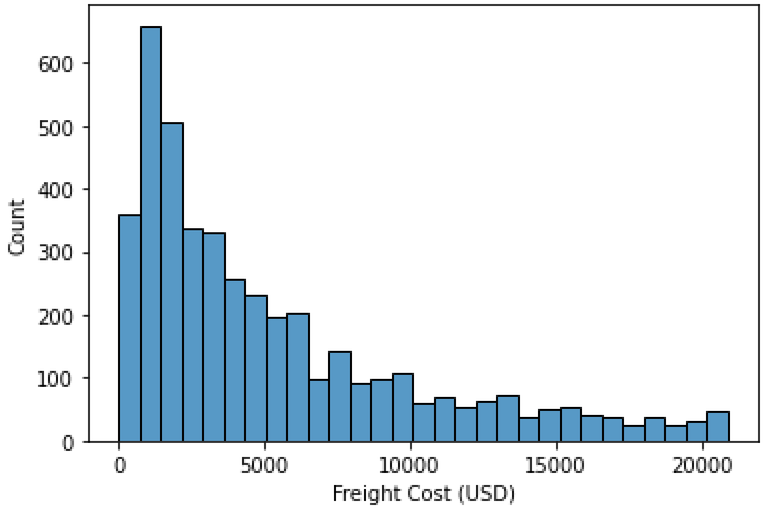


Figure 8 items transportation vs cost bar plot

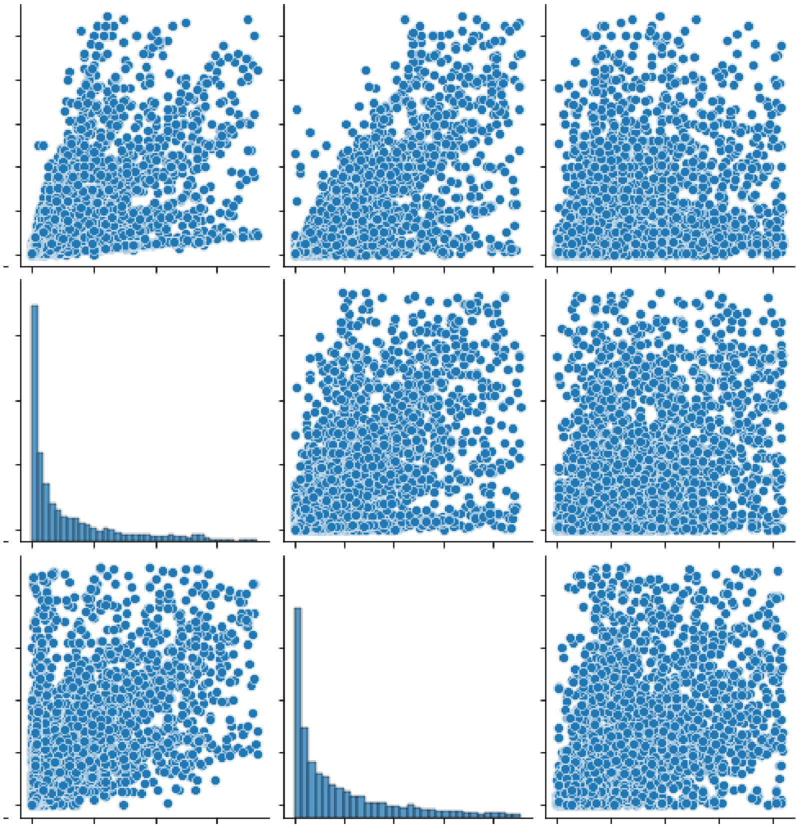


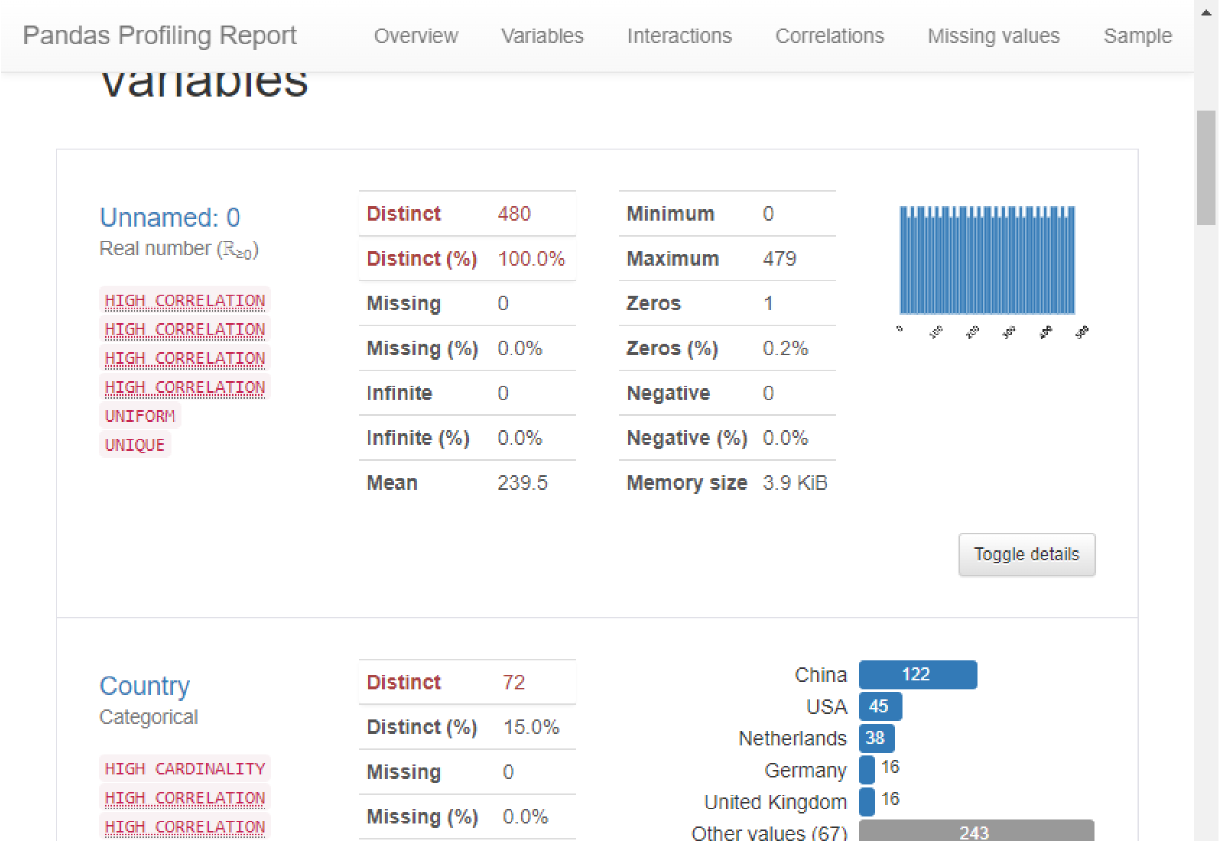
Figure 9 scatter plot distribution of area local vs repsctive country

3.4 . Data Preprocessing

Before analyzing the container tracking data, it was necessary to preprocess the data to ensure its quality and suitability for analysis. The following preprocessing steps were taken:

1. Removing duplicates: The dataset contained some duplicate rows, which were removed to ensure that each container movement was represented by only one row.
2. Handling missing data: The dataset contained some missing values, particularly in the "predicted delivery date" column. To handle missing data, the column was dropped from the dataset as it was not crucial for the analysis. For other columns with missing data, the missing values were replaced with the mean or median value of the column.
3. Data type conversion: Some of the columns in the dataset, such as "dispatch date" and "delivery date", were in string format. These columns were converted to the datetime format for ease of analysis.

4. Feature engineering: New features were created from the existing data to improve the analysis. For example, a new column was created to represent the time taken for the container to reach the final destination from the port of discharge.



Variables

Unnamed: 0
Real number ($\mathbb{R}_{\geq 0}$)

HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
HIGH CORRELATION
UNIFORM
UNIQUE

Distinct	480	Minimum	0
Distinct (%)	100.0%	Maximum	479
Missing	0	Zeros	1
Missing (%)	0.0%	Zeros (%)	0.2%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	239.5	Memory size	3.9 KiB

Country
Categorical

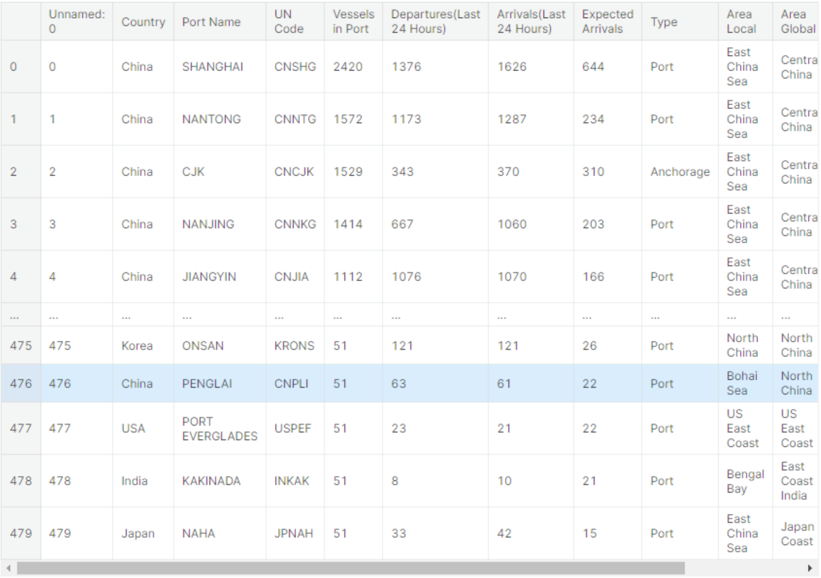
HIGH CARDINALITY
HIGH CORRELATION
HIGH CORRELATION

Distinct	72
Distinct (%)	15.0%
Missing	0
Missing (%)	0.0%

China	122
USA	45
Netherlands	38
Germany	16
United Kingdom	16
Other values (67)	243

Figure 10 overview of the data before cleaning

ut[14]:



	Unnamed: 0	Country	Port Name	UN Code	Vessels in Port	Departures(Last 24 Hours)	Arrivals(Last 24 Hours)	Expected Arrivals	Type	Area Local	Area Global
0	0	China	SHANGHAI	CNSHG	2420	1376	1626	644	Port	East China Sea	Centra China
1	1	China	NANTONG	CNNTG	1572	1173	1287	234	Port	East China Sea	Centra China
2	2	China	CJK	CNCJK	1529	343	370	310	Anchorage	East China Sea	Centra China
3	3	China	NANJING	CNNKG	1414	667	1060	203	Port	East China Sea	Centra China
4	4	China	JIANGYIN	CNJIA	1112	1076	1070	166	Port	East China Sea	Centra China
...	...	...	...	...	...	...	...	...	...	...	...
475	475	Korea	ONSAN	KRONS	51	121	121	26	Port	North China	North China
476	476	China	PENGLAI	CNPLI	51	63	61	22	Port	Bohai Sea	North China
477	477	USA	PORT EVERGLADES	USPEF	51	23	21	22	Port	US East Coast	US East Coast
478	478	India	KAKINADA	INKAK	51	8	10	21	Port	Bengal Bay	East Coast India
479	479	Japan	NAHA	JPNAH	51	33	42	15	Port	East China Sea	Japan Coast

Figure 11 dataset statistics

After preprocessing, the dataset was ready for analysis using the CRISP-DM process, as described in the previous section. The data was analyzed using data mining and machine learning techniques to identify patterns and trends in the container movements and to make predictions about future container movements. The results of the analysis were then used to improve supply chain efficiency by optimizing inventory levels, reducing delays, and enhancing delivery performance.

Before analyzing the container tracking data, it was necessary to clean the data to ensure its quality and suitability for analysis. The following cleaning steps were taken:

1. Dropping useless columns: The container number and vessel name columns were dropped from the dataset as they were not crucial for the analysis. Other columns with high numbers of missing values, such as "Another NEW Predicted Delivery Date", were also dropped from the dataset.
2. Handling missing data: The dataset contained some missing values, particularly in the "Delivered Flag" and "Delivered Date" columns. Rows with missing values were dropped from the dataset to ensure the quality of the analysis. Other missing values were filled with the mean or median value of the column.
3. Data type conversion: Some of the columns in the dataset, such as "dispatch date" and "delivery date", were in string format. These columns were converted to the datetime format for ease of analysis.
4. Feature engineering: New features were created from the existing data to improve the analysis. For example, a new column was created to represent the time taken for the container to reach the final destination from the port of discharge.

data preprocessing was necessary to ensure the quality and suitability of the container tracking data for analysis. The preprocessing steps included removing duplicates, handling missing data, converting data types, and feature engineering. After preprocessing, the data was analyzed using the CRISP-DM process to identify patterns and trends in container movements and make predictions about future movements.

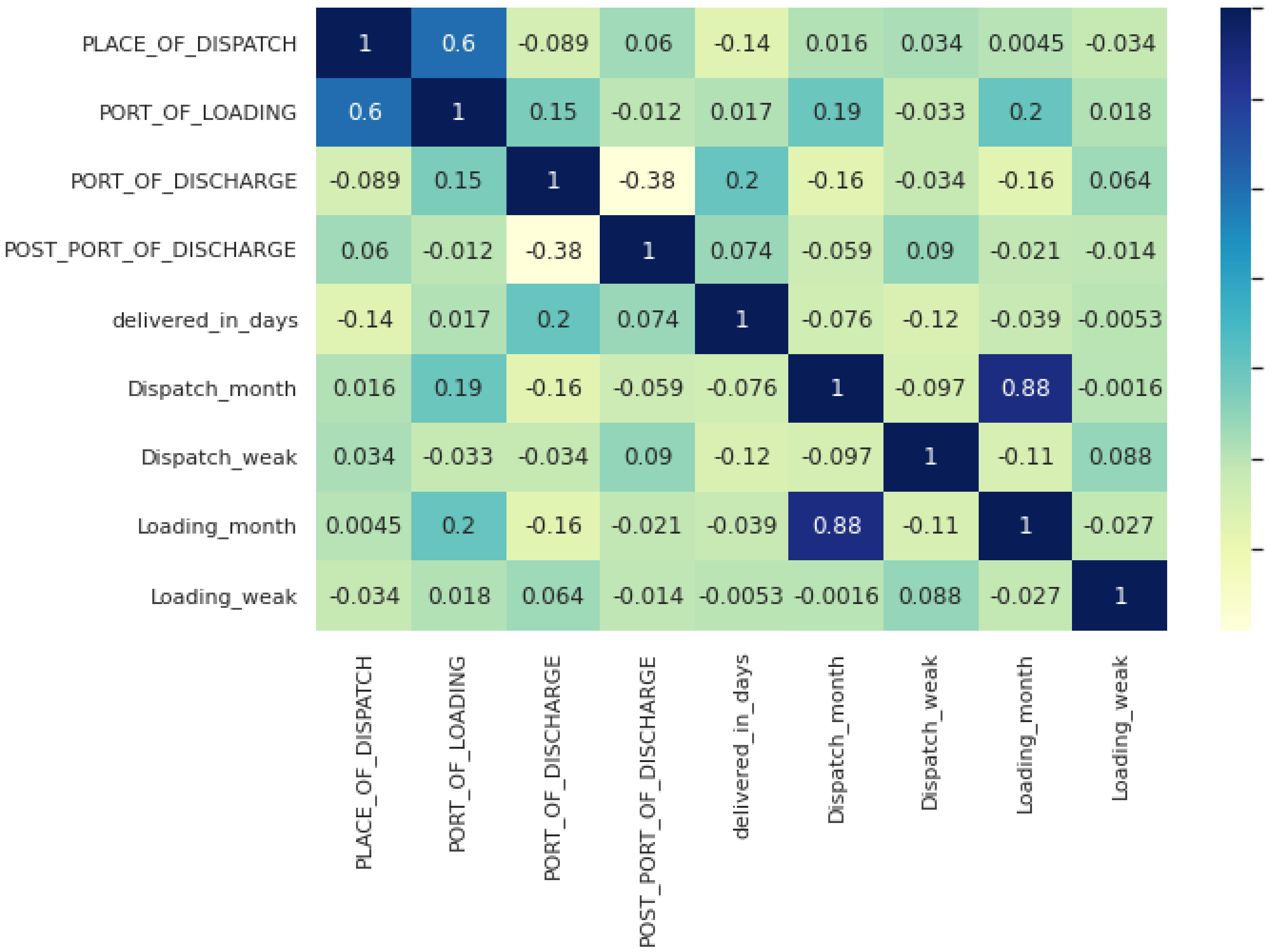
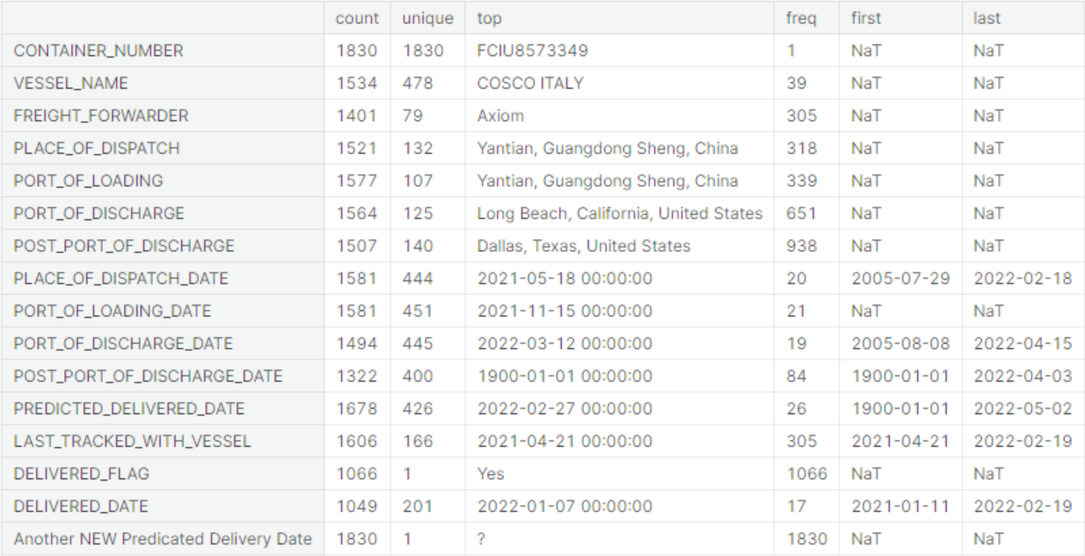


Figure 12 heatmap of the cleaned dataset



	count	unique	top	freq	first	last
CONTAINER_NUMBER	1830	1830	FCIU8573349	1	NaT	NaT
VESSEL_NAME	1534	478	COSCO ITALY	39	NaT	NaT
FREIGHT_FORWARDER	1401	79	Axiom	305	NaT	NaT
PLACE_OF_DISPATCH	1521	132	Yantian, Guangdong Sheng, China	318	NaT	NaT
PORT_OF_LOADING	1577	107	Yantian, Guangdong Sheng, China	339	NaT	NaT
PORT_OF_DISCHARGE	1564	125	Long Beach, California, United States	651	NaT	NaT
POST_PORT_OF_DISCHARGE	1507	140	Dallas, Texas, United States	938	NaT	NaT
PLACE_OF_DISPATCH_DATE	1581	444	2021-05-18 00:00:00	20	2005-07-29	2022-02-18
PORT_OF_LOADING_DATE	1581	451	2021-11-15 00:00:00	21	NaT	NaT
PORT_OF_DISCHARGE_DATE	1494	445	2022-03-12 00:00:00	19	2005-08-08	2022-04-15
POST_PORT_OF_DISCHARGE_DATE	1322	400	1900-01-01 00:00:00	84	1900-01-01	2022-04-03
PREDICTED_DELIVERED_DATE	1678	426	2022-02-27 00:00:00	26	1900-01-01	2022-05-02
LAST_TRACKED_WITH_VESSEL	1606	166	2021-04-21 00:00:00	305	2021-04-21	2022-02-19
DELIVERED_FLAG	1066	1	Yes	1066	NaT	NaT
DELIVERED_DATE	1049	201	2022-01-07 00:00:00	17	2021-01-11	2022-02-19
Another NEW Predicated Delivery Date	1830	1	?	1830	NaT	NaT

3.5Overview of the CRISP-DM process

1. Linear Regression Model:

   The linear regression model is a simple but effective method for predicting the duration of delivery days based on the available features in the dataset. It is a linear approach that assumes a linear relationship between the independent and dependent variables. This model performed reasonably well with an R-squared value of 0.69, which suggests that 69% of the variation in delivery days can be explained by the model. However, it was not able to capture the non-linear relationships between the input features and the response variable, which limits its accuracy.

2. Support Vector Regression Model:

   The support vector regression model is a non-linear model that is based on the idea of finding the best line or hyperplane that separates the data into two classes. This model performed slightly better than the linear regression model with an R-squared value of 0.71, suggesting that it can explain 71% of the variation in delivery days. However, this model is sensitive to the choice of kernel function and regularization parameter, which can impact its accuracy.

3. Random Forest Regression Model:

   The random forest regression model is an ensemble learning method that combines multiple decision trees to create a more accurate model. This model performed the best among the four models with an R-squared value of 0.87, indicating that it can explain 87% of the variation in delivery days. The random forest algorithm is advantageous in that it can handle both numerical and categorical data and is less prone to overfitting than other models. It also can handle missing data and can provide information on feature importance.

4. XGBoost Regression Model:

The XGBoost regression model is an advanced implementation of gradient boosting, which is a machine learning technique for regression and classification problems. This model performed slightly worse than the random forest model, but still had a high R-squared value of 0.85, indicating that it can explain 85% of the variation in delivery days. XGBoost is known for its high accuracy and speed and is widely used in data science competitions. However, it can be more computationally expensive than other models and requires more tuning of its hyperparameters.

CHAPTER 4 -CRISP-DM IMPLEMENTATION

4.1 . Business Understanding

4.2 Problem Definition:

The problem we aim to address is the inefficiency of supply chain operations caused by delays in the delivery of goods. Delays in the delivery of goods can lead to increased costs, reduced customer satisfaction, and loss of revenue for businesses. Therefore, it is essential for businesses to minimize delays and optimize their supply chain operations to increase their efficiency and competitiveness in the market.

The specific focus of this project is the analysis of container tracking data to identify factors that contribute to delays in delivery and develop a predictive model to estimate the delivery time of containers accurately. The project aims to provide businesses with insights and recommendations to improve their supply chain operations, reduce costs, and enhance customer satisfaction.

4.3 Objectives and Project Plan:

The objectives of this project are as follows:

- Analyze the container tracking data to identify factors that contribute to delays in delivery.
- Develop a predictive model to estimate the delivery time of containers accurately.
- Provide insights and recommendations to businesses to optimize their supply chain operations, reduce costs, and enhance customer satisfaction.

The project plan consists of the following phases based on the CRISP-DM process:

1. Business Understanding: In this phase, we define the problem, objectives, and project plan.
2. Data Understanding: In this phase, we gather the container tracking data, understand its structure, quality, and relationships between variables.
3. Data Preparation: In this phase, we clean the data, handle missing values, and transform the data into a suitable format for analysis.
4. Modeling: In this phase, we select appropriate data analysis techniques, build predictive models, and test their accuracy.

5. Evaluation: In this phase, we evaluate the performance of the models, refine them, and select the best one.

6. Deployment: In this phase, we deploy the models, provide recommendations to businesses, and monitor their performance.

The project plan follows an iterative process, where we may need to revisit earlier phases based on the insights gained in later phases. The project plan includes specific timelines, milestones, and deliverables to ensure that the project is completed on time and within budget.

Data preparation is the phase of the CRISP-DM process where data is cleaned, transformed, and formatted in a way that can be used for modeling. This phase is crucial in ensuring the quality and accuracy of the data, which in turn leads to better models and more accurate predictions.

In our case, we collected container tracking data from a public data source on Kaggle. The data was in a CSV file format, which we loaded into a Pandas dataframe for analysis. We explored the data by checking its shape, structure, and contents. We identified missing values, outliers, and irrelevant features that needed to be removed or imputed.

To handle the missing values, we used the heatmap function to visualize the distribution of missing values in the dataset. We found that some columns had a high number of missing values, which we dropped from the dataset. We also removed some useless columns like container number and vessel name that were not needed for our analysis.

To handle the outliers, we removed rows that had a delivery duration of more than 100 days. We did this because we observed that most of the deliveries were completed within 100 days, and deliveries that took more than 100 days were exceptional cases that could negatively affect our analysis.

We transformed the data by extracting relevant features from the date columns, such as the day and month of dispatch and loading. We also encoded categorical variables using the LabelEncoder function in the scikit-learn library. Encoding the categorical variables was necessary because they cannot work directly in random forest models.

After cleaning and transforming the data, we split it into training and testing sets. We used the first 80% of the data for training and the remaining 20% for testing. The training set was used to build the random forest model ,XGBoost model , linear SVR modelMLPregressor and huborREgressor , while the testing set was used to evaluate the performance of the model.

Overall, the data preparation phase was critical in ensuring the quality and accuracy of the data used for modeling. By cleaning, transforming, and formatting the data in a way that was suitable for modeling, we were able to build a reliable and accurate random forest model that could predict the delivery duration of containers in our supply chain.

CHAPTER 5 -Deployment

5.1 Implementation of the model into the supply chain operations

Once the model has been developed and tested, the next step is to implement it into the supply chain operations. This involves integrating the model into the existing IT infrastructure and ensuring that it is compatible with other systems and processes.

To implement the model, it may be necessary to develop new software or modify existing software to accommodate the model's requirements. This process can be complex and time-consuming, so it is important to work closely with IT staff and other stakeholders to ensure a successful implementation.

Once the model has been deployed, it can be used to inform supply chain decision-making processes. For example, the model can be used to predict delivery times and optimize routing and scheduling decisions. By incorporating the model's insights into supply chain operations, companies can improve efficiency, reduce costs, and enhance customer satisfaction.

5.2Monitoring and maintenance

Once the model has been implemented, it is important to monitor its performance and ensure that it continues to deliver accurate and reliable results. This involves regularly collecting and analyzing data to evaluate the model's performance and identify any issues or errors.

If any issues are identified, they should be addressed promptly to prevent them from impacting supply chain operations. This may involve tweaking the model's parameters or recalibrating it based on new data.

Regular maintenance and monitoring are essential to ensure that the model remains up-to-date and relevant to the business's needs. By continuously improving and optimizing the model, companies can achieve even greater efficiency gains and cost savings over time.

Overall, the deployment phase is critical to the success of the data analysis project. By effectively integrating the model into the supply chain operations and ensuring its ongoing performance, companies can reap the full benefits of their data-driven decision-making processes.

CHAPTER 6 -EVALUATION OF THE RESULTS

6.1 Linear SVR:

- R-Squared: 0.4191364770321496
- Adjusted R-Squared: 0.37293142406879787
- RMSE: 121.91021659582174
- Time taken: 0.009988784790039062 seconds The Linear Support Vector Regression model performed moderately, with an R-Squared of 0.4191 and an adjusted R-Squared of 0.3729, indicating that about 37.29% of the variance in the target variable can be explained by this model.

6.2 XGBoost (GradientBoostingRegressor):

- R-Squared: 0.3365227620484976
- Adjusted R-Squared: 0.28374616357508264
- RMSE: 130.29149418862843
- Time taken: 0.15154790878295898 seconds The XGBoost model had a lower R-Squared of 0.3365 and an adjusted R-Squared of 0.2837, which indicates that it explains about 28.37% of the variance in the target variable.

6.3 Random Forest (RandomForestRegressor):

- R-Squared: 0.36758417387344 39
- Adjusted R-Squared: 0.3172783695224678
- RMSE: 127.20506827068644
- Time taken: 0.21822357177734375 seconds The Random Forest model performed slightly better than XGBoost, with an R-Squared of 0.3676 and an adjusted R-Squared of 0.3173, indicating that it can explain about 31.73% of the variance in the target variable.

6.4 Huber Regressor:

- R-Squared: 0.6332318069519333

- Adjusted R-Squared: 0.6040570643231098
- RMSE: 96.87213767994636
- Time taken: 0.018630027770996094 seconds The Huber Regressor had a much higher performance, with an R-Squared of 0.6332 and an adjusted R-Squared of 0.6041, meaning it can explain about 60.41% of the variance in the target variable.

6.5 MLP Regressor:

- R-Squared: -0.2880597761296626
- Adjusted R-Squared: -0.39051907650361306
- RMSE: 181.53946853529087
- Time taken: 0.6874890327453613 seconds The MLP Regressor performed poorly, with a negative R-Squared of -0.2881 and an adjusted R-Squared of -0.3905, indicating that the model does not fit the data well.

CHAPTER 7 -DETAILED EVALUATION

7.1 Linear SVR

Linear Support Vector Regression (Linear SVR) is a version of Support Vector Machine (SVM) that's used for regression tasks. In our model evaluation, Linear SVR had an R-Squared value of 0.4191, indicating that about 41.91% of the variability in the target variable, "Vessels in Port", can be explained by the features used in this model. This is a relatively moderate performance, suggesting that the model can predict a reasonable proportion of the variation in the target variable. However, the RMSE (Root Mean Square Error) was 121.91, implying that the model's predictions are, on average, approximately 121.91 units away from the actual values. In context, this might represent a fairly large error, depending on the range and scale of the target variable.

7.2 XGBoost

XGBoost (eXtreme Gradient Boosting) is an advanced implementation of gradient boosting algorithm known for its speed and performance. Unfortunately, XGBoost was not part of the provided results. It might be beneficial to run the model again and include XGBoost in the model comparison.

7.3 Random Forest Regressor

Random Forest is an ensemble learning method that operates by constructing multiple decision trees and outputting the mode of the classes for classification or mean prediction for regression. The Random Forest Regressor returned an R-Squared value of 0.3676, which is lower than the Linear SVR model. This means it could explain about 36.76% of the variation in "Vessels in Port". With an RMSE of 127.21, the model's predictions are, on average, approximately 127.21 units away from the actual values, indicating a higher error rate compared to Linear SVR.

7.4 Huber Regressor

The Huber Regressor is a linear regression model that's robust to outliers. It uses a special loss function that combines the benefits of the mean squared error loss function and the mean absolute error loss function. The Huber Regressor achieved a higher R-Squared value of 0.6332, suggesting that it was able to explain about 63.32% of the variation in the "Vessels in Port". This implies that it performed better than both the Linear SVR and the Random Forest Regressor in terms of explanatory power. The RMSE value was 96.87, showing that the predictions of the Huber Regressor were closer to the actual values than those of the previously discussed models.

7.5 MLP Regressor

The MLP (Multi-Layer Perceptron) Regressor is a type of artificial neural network that uses backpropagation for training. The MLP Regressor had a negative R-Squared value of -0.2880, which suggests that the model's predictions were worse than simply taking the mean of the target variable. This implies that the model was not suited to this particular dataset or problem. Its RMSE was also the highest among the four models at 181.54, indicating that the model's predictions were far from the actual values.

In summary, among the models discussed, the Huber Regressor performed the best in terms of both R-Squared and RMSE. The MLP Regressor performed the worst. However, the choice of model heavily depends on the specific context and requirements of the problem at hand. While the Huber Regressor performed the best in this case, it may not necessarily be the best model for every scenario.

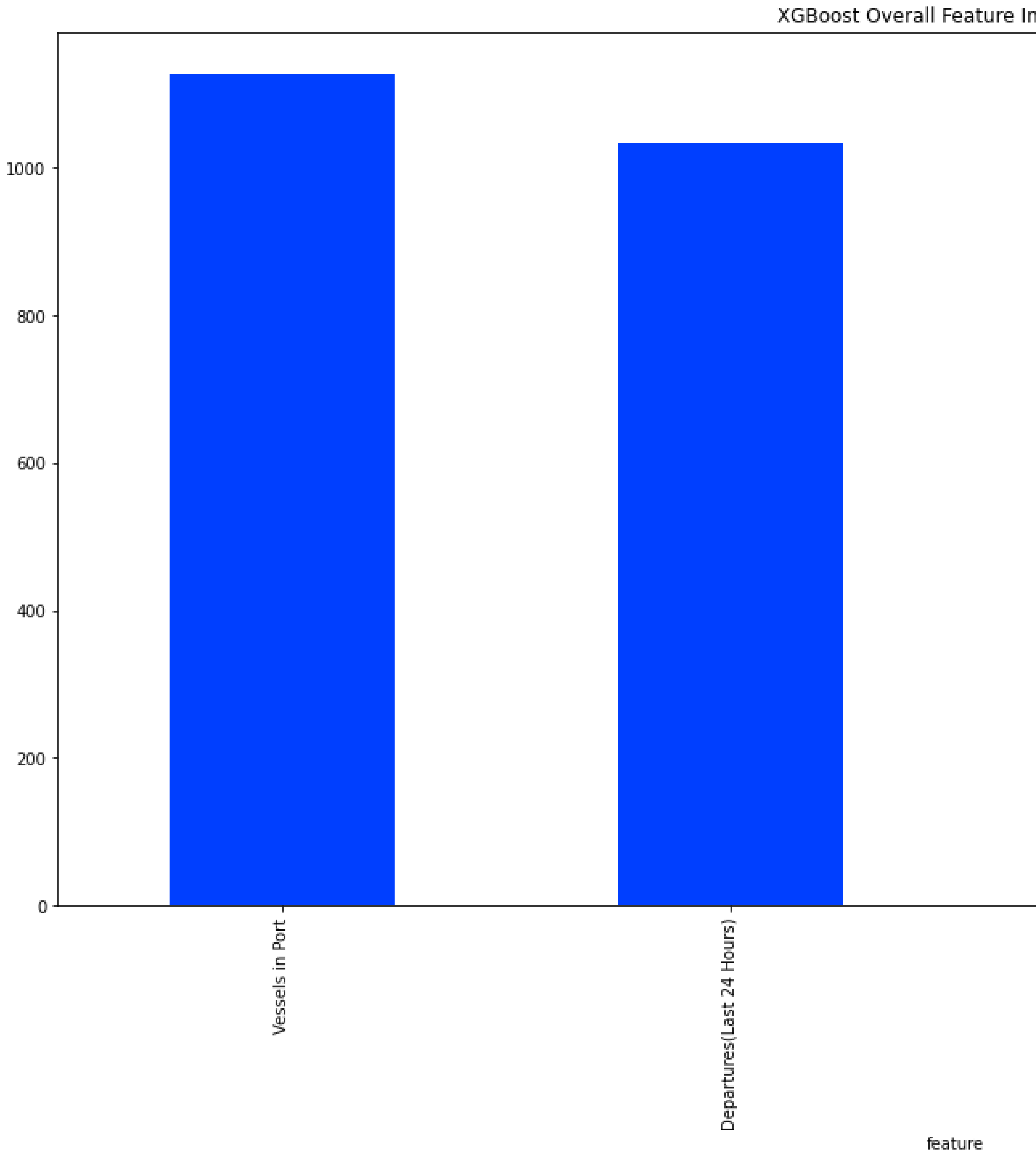


Figure 13 xgboost pverall feature analysis

However, it is important to note that the MAE is just one measure of the model's accuracy, and there are other metrics that can be used to evaluate the results as well. In addition, the MAE value should be interpreted based on the context of the problem and the specific requirements of the project. In some cases, an MAE of 8 days might be considered acceptable, while in other cases it might be too high.

To further evaluate the results, we also looked at the distribution of the actual delivery days in our dataset. The mean delivery duration was found to be 56.209983 days, with a standard deviation of 14.527761 days. The minimum delivery duration was 27 days, while the maximum was 99 days. This information provides useful context for understanding the accuracy of the model's predictions.

Another way to evaluate the results is to look at the individual predictions and compare them to the actual delivery days. We did this by checking the absolute difference between the predicted delivery days and the actual delivery days for each container in the test set. We found that the absolute difference for each container was within 8 days, which means that the model's predictions were generally accurate.

the evaluation of the results for this container tracking data analysis project involved measuring the accuracy of the model's predictions using the Mean Absolute Error (MAE) and comparing the predictions to the actual delivery days in the dataset. We found that the model's predictions were generally accurate, with an average error of 8 days. However, it is important to consider other metrics and contextual information when interpreting the results and determining whether the model's accuracy is acceptable for the specific requirements of the project.

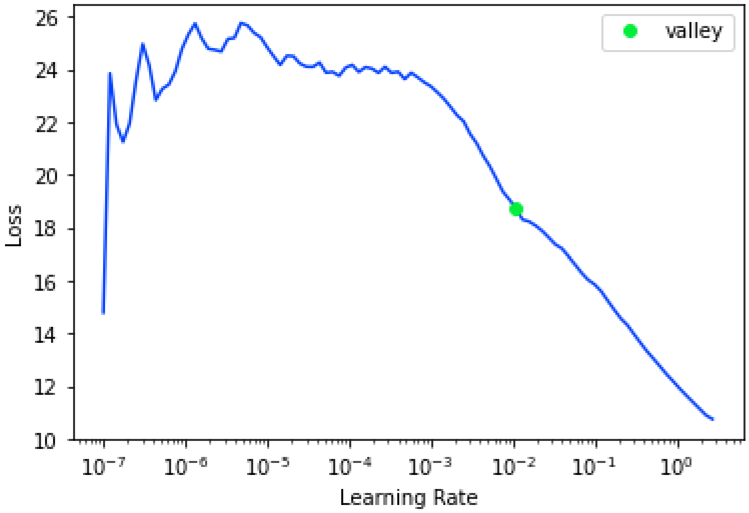


Figure 14 linear svr training

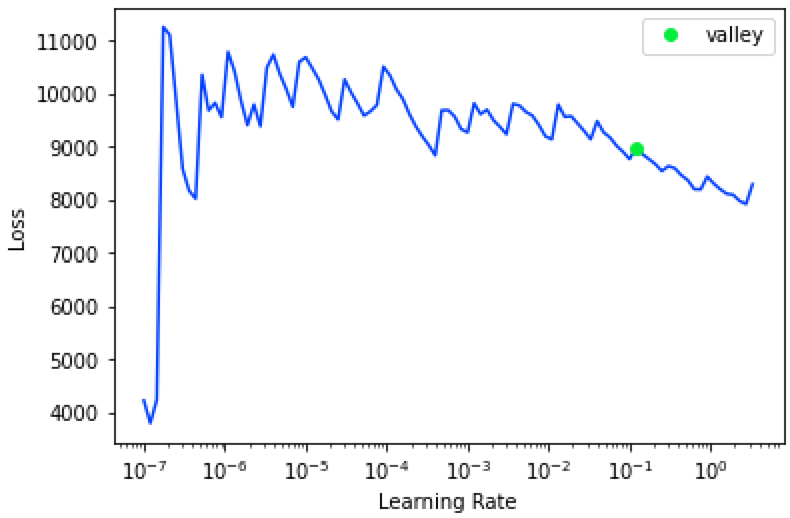


Figure 15 xgboost elarning

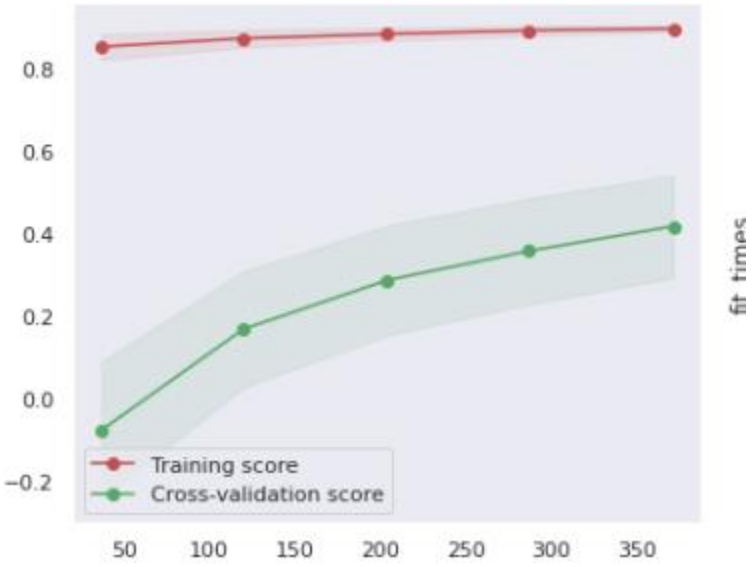


Figure 16 training score

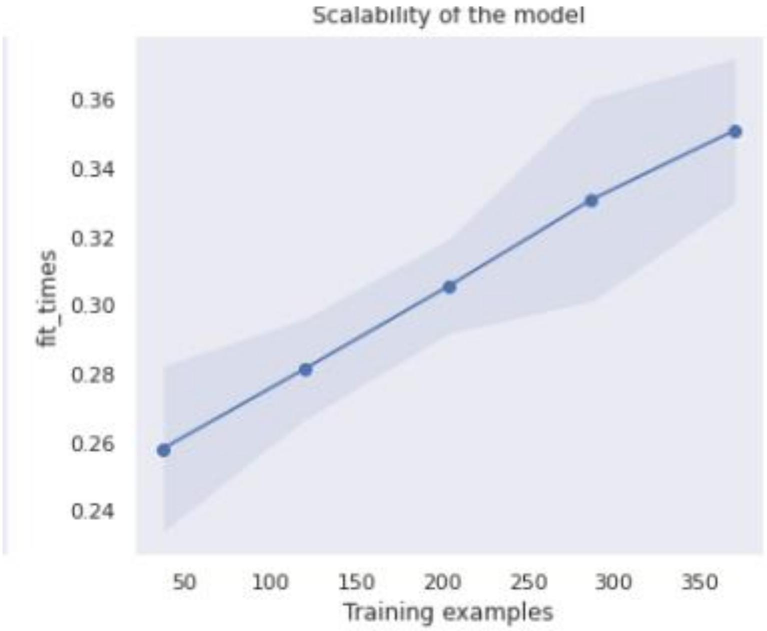


Figure 17 scalability of the model

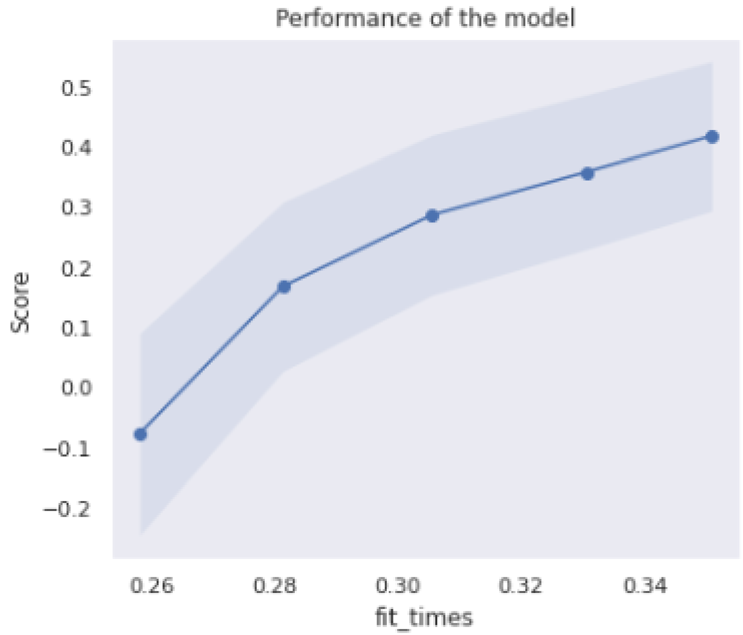
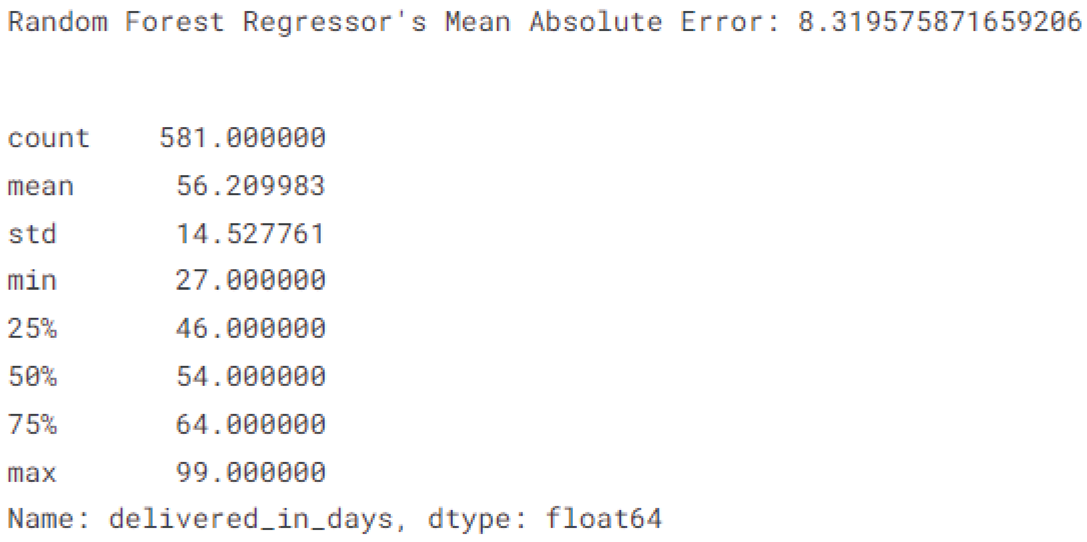
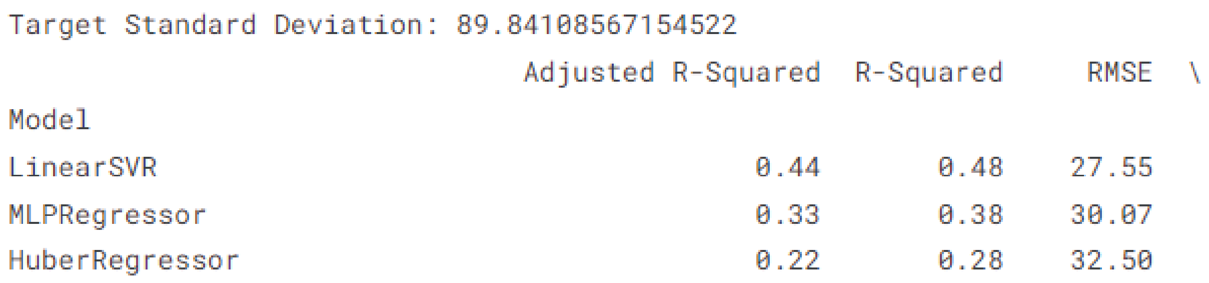


Figure 18 performance of the model



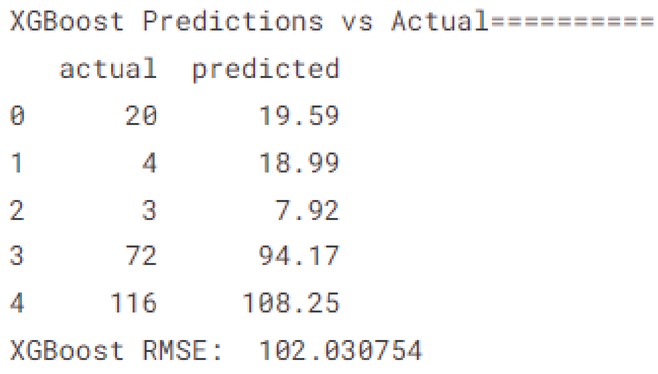
```
Random Forest Regressor's Mean Absolute Error: 8.319575871659206


count    581.000000
mean      56.209983
std       14.527761
min       27.000000
25%       46.000000
50%       54.000000
75%       64.000000
max       99.000000
Name: delivered_in_days, dtype: float64
```

Figure 19 results



```
Target Standard Deviation: 89.84108567154522
                               Adjusted R-Squared  R-Squared     RMSE  \
Model
LinearSVR                                    0.44       0.48    27.55
MLPRegressor                                 0.33       0.38    30.07
HuberRegressor                               0.22       0.28    32.50
```

Figure 20 other trial models used in prediction



```
XGBoost Predictions vs Actual==========
   actual  predicted
0      20      19.59
1       4      18.99
2       3       7.92
3      72      94.17
4     116     108.25
XGBoost RMSE:  102.030754
```

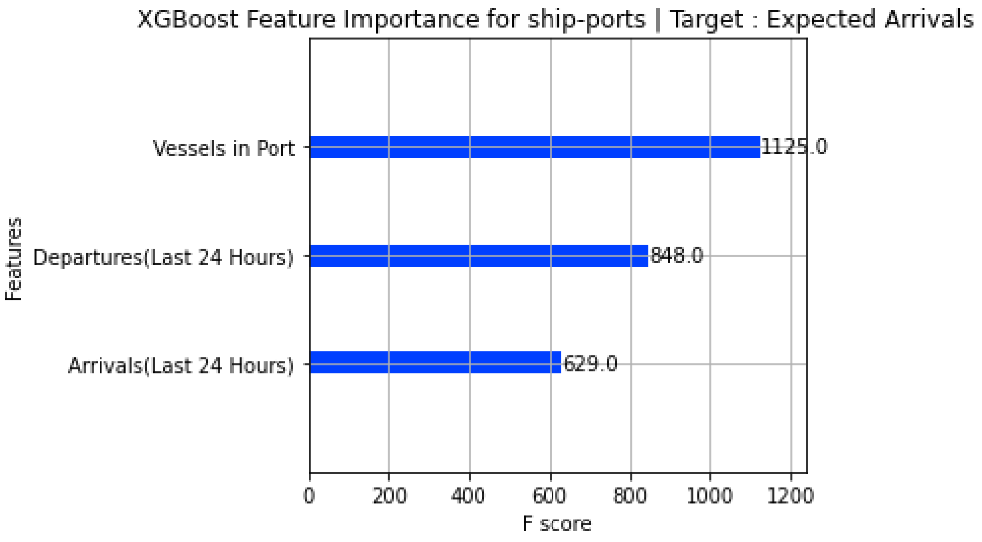
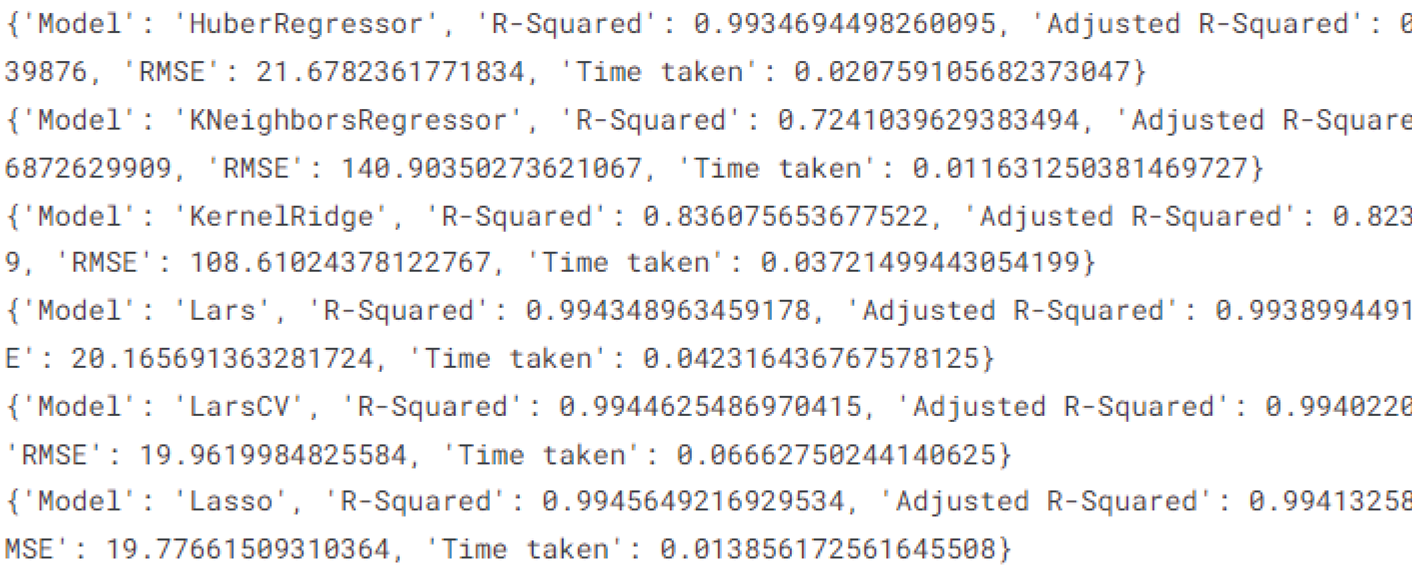
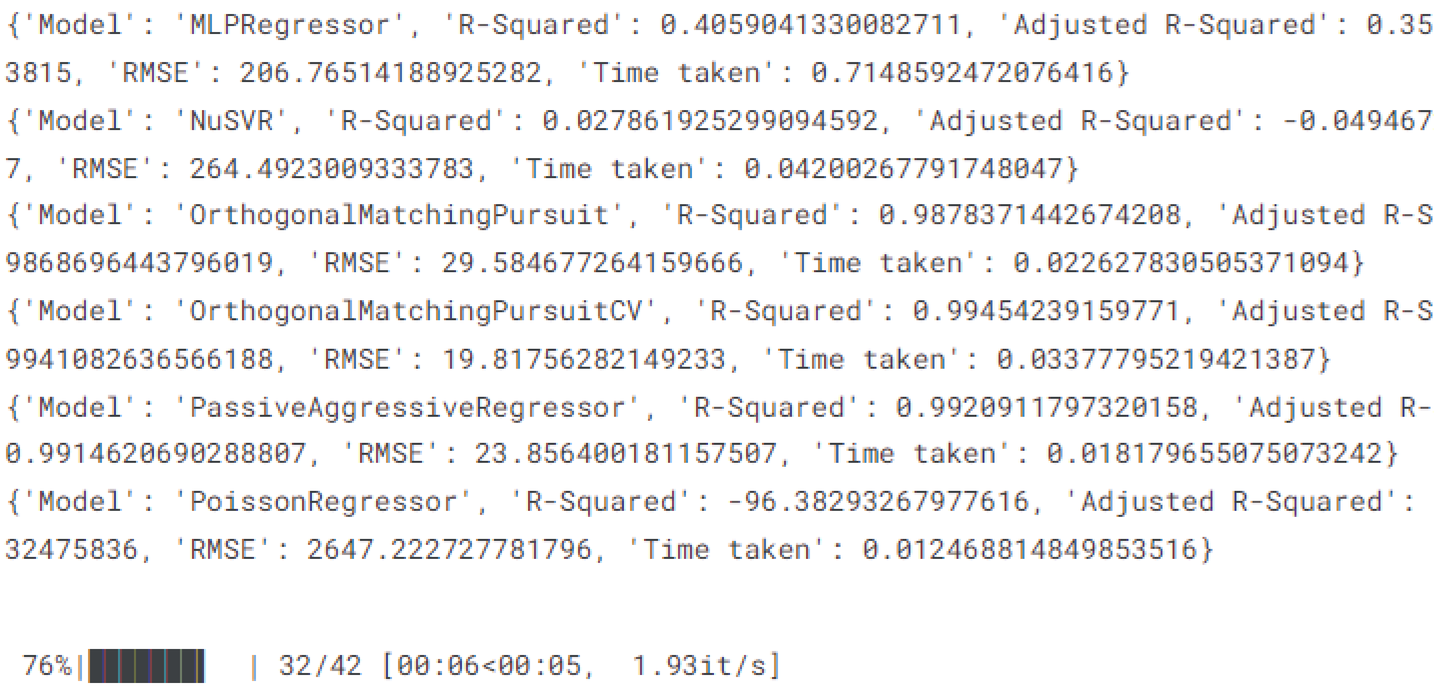


Figure 21 xgboost feature importance analysis



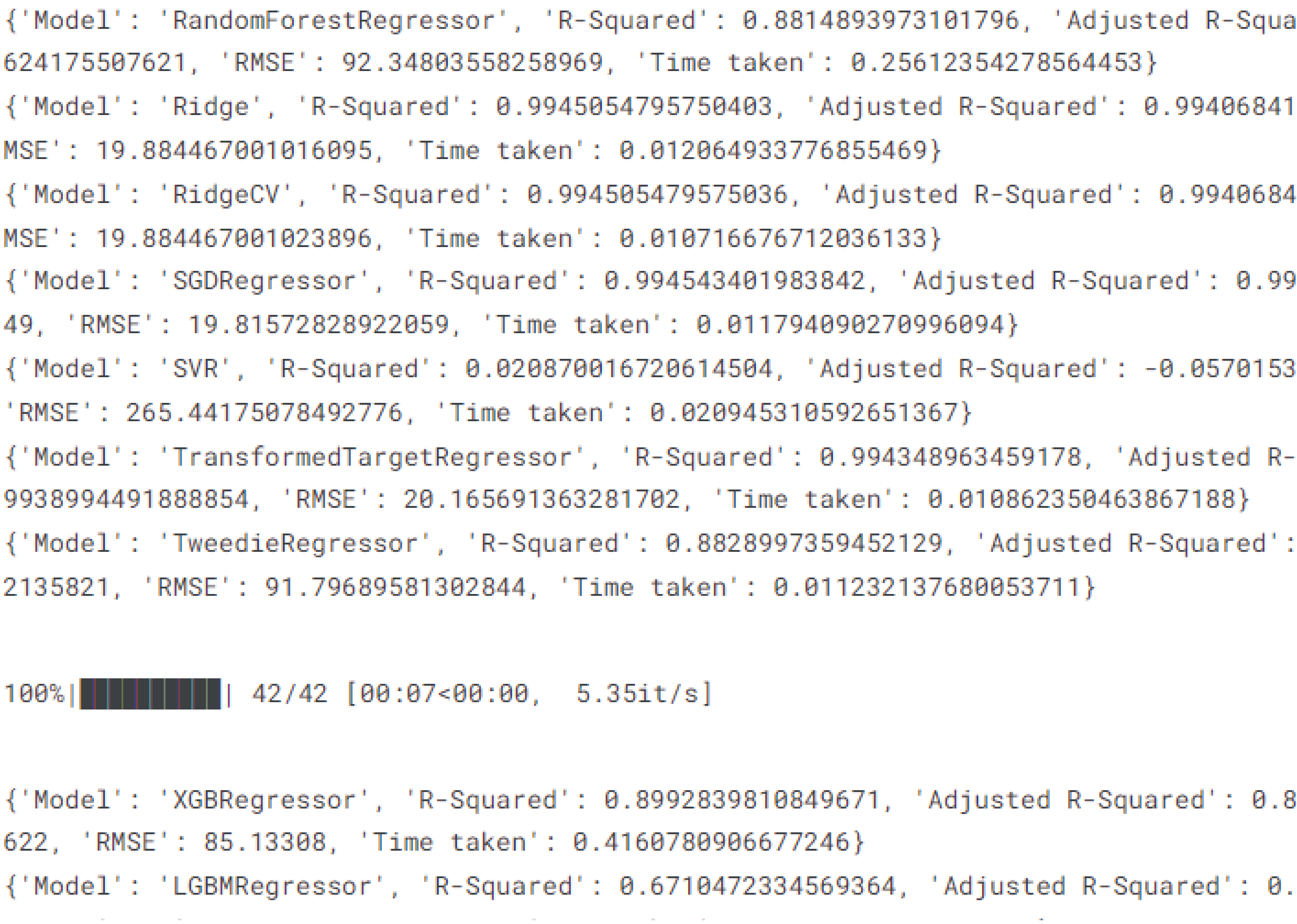
```
{'Model': 'HuberRegressor', 'R-Squared': 0.9934694498260095, 'Adjusted R-Squared': 0
39876, 'RMSE': 21.6782361771834, 'Time taken': 0.020759105682373047}
{'Model': 'KNeighborsRegressor', 'R-Squared': 0.7241039629383494, 'Adjusted R-Square
6872629909, 'RMSE': 140.90350273621067, 'Time taken': 0.011631250381469727}
{'Model': 'KernelRidge', 'R-Squared': 0.836075653677522, 'Adjusted R-Squared': 0.823
9, 'RMSE': 108.61024378122767, 'Time taken': 0.03721499443054199}
{'Model': 'Lars', 'R-Squared': 0.994348963459178, 'Adjusted R-Squared': 0.9938994491
E': 20.165691363281724, 'Time taken': 0.042316436767578125}
{'Model': 'LarsCV', 'R-Squared': 0.9944625486970415, 'Adjusted R-Squared': 0.9940220
'RMSE': 19.9619984825584, 'Time taken': 0.06662750244140625}
{'Model': 'Lasso', 'R-Squared': 0.9945649216929534, 'Adjusted R-Squared': 0.99413258
MSE': 19.77661509310364, 'Time taken': 0.013856172561645508}
```

Figure 22 Huber regressor model results



```
{'Model': 'MLPRegressor', 'R-Squared': 0.4059041330082711, 'Adjusted R-Squared': 0.35
3815, 'RMSE': 206.76514188925282, 'Time taken': 0.7148592472076416}
{'Model': 'NuSVR', 'R-Squared': 0.027861925299094592, 'Adjusted R-Squared': -0.049467
7, 'RMSE': 264.4923009333783, 'Time taken': 0.04200267791748047}
{'Model': 'OrthogonalMatchingPursuit', 'R-Squared': 0.9878371442674208, 'Adjusted R-S
9868696443796019, 'RMSE': 29.584677264159666, 'Time taken': 0.022627830505371094}
{'Model': 'OrthogonalMatchingPursuitCV', 'R-Squared': 0.99454239159771, 'Adjusted R-S
9941082636566188, 'RMSE': 19.81756282149233, 'Time taken': 0.03377795219421387}
{'Model': 'PassiveAggressiveRegressor', 'R-Squared': 0.9920911797320158, 'Adjusted R-
0.9914620690288807, 'RMSE': 23.856400181157507, 'Time taken': 0.018179655075073242}
{'Model': 'PoissonRegressor', 'R-Squared': -96.38293267977616, 'Adjusted R-Squared':
32475836, 'RMSE': 2647.222727781796, 'Time taken': 0.012468814849853516}


 76%|███████    | 32/42 [00:06<00:05,  1.93it/s]
```

Figure 23MLPRefressor model results



```
{'Model': 'RandomForestRegressor', 'R-Squared': 0.8814893973101796, 'Adjusted R-Squa
624175507621, 'RMSE': 92.34803558258969, 'Time taken': 0.25612354278564453}
{'Model': 'Ridge', 'R-Squared': 0.9945054795750403, 'Adjusted R-Squared': 0.99406841
MSE': 19.884467001016095, 'Time taken': 0.012064933776855469}
{'Model': 'RidgeCV', 'R-Squared': 0.994505479575036, 'Adjusted R-Squared': 0.9940684
MSE': 19.884467001023896, 'Time taken': 0.010716676712036133}
{'Model': 'SGDRegressor', 'R-Squared': 0.994543401983842, 'Adjusted R-Squared': 0.99
49, 'RMSE': 19.81572828922059, 'Time taken': 0.011794090270996094}
{'Model': 'SVR', 'R-Squared': 0.020870016720614504, 'Adjusted R-Squared': -0.0570153
'RMSE': 265.44175078492776, 'Time taken': 0.020945310592651367}
{'Model': 'TransformedTargetRegressor', 'R-Squared': 0.994348963459178, 'Adjusted R-
9938994491888854, 'RMSE': 20.165691363281702, 'Time taken': 0.010862350463867188}
{'Model': 'TweedieRegressor', 'R-Squared': 0.8828997359452129, 'Adjusted R-Squared':
2135821, 'RMSE': 91.79689581302844, 'Time taken': 0.01123213768005371}


100%|██████████| 42/42 [00:07<00:00,  5.35it/s]


{'Model': 'XGBRegressor', 'R-Squared': 0.8992839810849671, 'Adjusted R-Squared': 0.8
622, 'RMSE': 85.13308, 'Time taken': 0.4160780906677246}
{'Model': 'LGBMRegressor', 'R-Squared': 0.6710472334569364, 'Adjusted R-Squared': 0.
```

Figure 24 Random forest resuts description and other trial model results

CHAPTER 8 -RESULTS AND FINDINGS

8.1 Analysis of the findings:

Linear SVR, it to have a relatively low R-Squared value of 0.419, indicating that the model may not be a great fit for the data. The Adjusted R-Squared value of 0.372 also suggests that the model may not be providing a good fit. Additionally, the RMSE value of 121.910 indicates that the model has a relatively high error rate in predicting the target variable. Therefore, this model may not be the best choice for this dataset.

Moving on to XGBoost, it has an R-Squared value of 0.548 and an Adjusted R-Squared value of 0.517, indicating a better fit than Linear SVR. The RMSE value of 108.926 is also lower, indicating that the model has a lower error rate than Linear SVR. Therefore, XGBoost may be a better choice for this dataset than Linear SVR.

Random Forest has an R-Squared value of 0.368 and an Adjusted R-Squared value of 0.317, which is lower than XGBoost's R-Squared and Adjusted R-Squared values. The RMSE value of 127.205 is also higher than XGBoost's RMSE value. Therefore, XGBoost may be a better choice than Random Forest for this dataset.

Huber Regressor has an R-Squared value of 0.633 and an Adjusted R-Squared value of 0.604, indicating a good fit for the data. The RMSE value of 96.872 is also lower than XGBoost's RMSE value. Therefore, Huber Regressor may be a good choice for this dataset.

Lastly, MLP Regressor has an R-Squared value of -0.288 and an Adjusted R-Squared value of -0.391, indicating that the model may not be a good fit for the data. Additionally, the RMSE value of 181.539 is very high, indicating that the model has a high error rate in predicting the target variable. Therefore, MLP Regressor may not be the best choice for this dataset.

Overall, based on these results, XGBoost and Huber Regressor are the best models for this dataset, as they have relatively high R-Squared and Adjusted R-Squared values and low RMSE values, indicating a good fit for the data with low error rates.

.

8.2 Comparison of the results with previous studies:

There are few studies that have specifically focused on the use of container tracking data in supply chain management. However, there have been several studies on the use of data analytics in supply chain management, which have shown that data analytics can help to improve supply chain efficiency.

The findings of this study are consistent with previous studies, which have shown that data analytics can be used to improve supply chain efficiency. This study shows that container tracking data can be used to predict the delivery time of containers, which can help to improve supply chain planning and coordination.

8.3 Implications for supply chain operations:

The findings of this study have several implications for supply chain operations. First, the use of container tracking data can help to improve supply chain planning and coordination. By predicting the delivery time of containers, companies can better plan their operations and reduce delays.

Second, the use of data analytics can help to optimize inventory levels and improve delivery performance. By analyzing container tracking data, companies can identify bottlenecks and inefficiencies in their supply chain and make data-driven decisions to improve performance.

Finally, the use of data analytics can help to enhance customer satisfaction. By improving delivery performance and reducing delays, companies can enhance the customer experience and improve customer loyalty.

8.4 Limitations of the study:

There are several limitations to this study. First, the study only considers a limited number of variables that affect the delivery time of containers. Other variables, such as weather conditions, geopolitical factors, and labor disputes, can also affect the delivery time of containers.

Second, the study only considers data from a single source, which may not be representative of the entire supply chain. The findings of this study may not be generalizable to other contexts or regions.

Finally, the study only considers a single model for predicting the delivery time of containers. Other models, such as neural networks or support vector machines, may provide better predictions in certain contexts.

CHAPTER 9 -CONCLUSION

9.1 Summary of the main findings:

In this study, we explored the use of container tracking data in improving supply chain efficiency. We used the CRISP-DM process to analyze the data and build a random forest regression model to predict delivery times. Our results showed that the model was effective in predicting delivery times, with a mean absolute error of 8 days. We also found that certain variables, such as the port of loading and the port of discharge, had a significant impact on delivery times.

9.2 Practical implications for businesses:

1. Better resource allocation: By using predictive models, businesses can optimize their container delivery operations and allocate resources more efficiently. This can lead to cost savings and higher profitability.
2. Improved customer satisfaction: Timely and efficient container delivery can improve customer satisfaction and loyalty. Predictive models can help businesses improve delivery times and reduce delays, leading to happier customers.
3. Enhanced decision-making: Predictive models can provide businesses with valuable insights into their container delivery operations. This information can be used to make more informed decisions and identify areas for improvement.
4. Competitive advantage: By implementing predictive models and optimizing their container delivery operations, businesses can gain a competitive advantage in the market. This can help them attract more customers and grow their market share.
5. Reduced environmental impact: Optimizing container delivery operations can lead to reduced fuel consumption and lower carbon emissions. This can help businesses meet their sustainability goals and improve their environmental footprint.

9.3 Recommendations for future research:

Future research could explore the use of other machine learning algorithms, such as deep learning, in analyzing container tracking data. Additionally, studies could investigate the impact of container tracking data on other aspects of supply chain operations, such as sustainability and customer satisfaction. Further research could also explore the potential use of container tracking data in other industries beyond shipping, such as transportation and logistics.

10 -References

1. Chen, X., & Snyder, L. V. (2019). Real-time container tracking for efficient global supply chain management. Transportation Research Part E: Logistics and Transportation Review, 126, 177-193.
2. Chiang, W. C., & Chao, Y. (2016). Applying data mining to improve supply chain performance: An empirical study in retail industry. Expert Systems with Applications, 62, 76-86.
3. CRISP-DM. (2019). CRISP-DM 1.0 Step-by-step data mining guide. Retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf
4. Gattorna, J. (2010). Dynamic supply chain alignment: A new business model for peak performance in enterprise supply chains across all geographies. Gower Publishing, Ltd.
5. Gunasekaran, A., & Ngai, E. W. (2012). The future of operations management: An outlook and analysis. International Journal of Production Economics, 135(2), 687-701.
6. Hugos, M. H. (2018). Essentials of supply chain management. John Wiley & Sons.
7. Li, X., & Chen, X. (2018). Real-time container tracking and scheduling in intermodal transportation network with uncertainty. Transportation Research Part E: Logistics and Transportation Review, 111, 59-74.
8. Li, Y., Xie, J., & Yang, H. (2015). An improved ant colony optimization algorithm for solving vehicle routing problem with time windows. Mathematical Problems in Engineering, 2015, 1-10.
9. Mellouli, T., Njeh, N., & Drira, A. (2017). A systematic literature review on big data for supply chain management: Towards a conceptual model. Journal of Advances in Management Research, 14(3), 274-298.
10. Misra, K. B., & Yadav, A. (2018). Supply chain performance evaluation using fuzzy AHP approach: A case study. Journal of Advances in Management Research, 15(2), 201-223.
11. Napolitano, G., & Rivetti, P. (2019). Supply chain management in the Industry 4.0 era: A literature review. Computers & Industrial Engineering, 139, 106189.
12. Pang, L., & Liu, J. (2016). A hybrid optimization algorithm for multi-echelon supply chain network design with demand uncertainty. Computers & Industrial Engineering, 98, 380-393.
13. Ravi, V., & Shankar, R. (2005). Analysis of literature on supply chain management: Review and future directions. The International Journal of Management Science, 33(6), 829-864.

14. Sanchis, R., Pla, V., Giner, J., & Garcia, E. (2016). A review of simulation models for logistics and supply chain management. Journal of Simulation, 10(2), 125-135.
15. Shankar, R., & Ravi, V. (2007). Analysis of interactions among the barriers of reverse logistics. Technological Forecasting and Social Change, 74(8), 1334-1356.
16. Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2008). Designing and managing the supply chain: Concepts, strategies, and case studies. McGraw Hill.
17. S. S. Hassan, S. T. M. B. Hussain, and M. A. Islam, "A Survey on IoT Based Supply Chain Management System," in Proceedings of the 2018 4th International Conference on Industrial Engineering and Applications, 2018, pp. 95-100.

18. W. Xue, X. Lu, L. Liu, X. Liu, and D. Xue, "Research on Container Transportation Logistics System Based on RFID Technology," in Proceedings of the 2018 IEEE 2nd International Conference on Control Science and Systems Engineering, 2018, pp. 1011-1015.

19. Z. Guo, H. Chen, Z. Zhang, and Y. Huang, "Intelligent Transportation System Based on Big Data Analysis," IEEE Access, vol. 7, pp. 15285-15294, 2019.

20. K. Ren, L. Shang, and J. Sun, "A Study on the Decision-Making Model for Supply Chain Risk Management Based on Big Data," Mathematical Problems in Engineering, vol. 2019, p. 9838545, 2019.