**M - Privacy for Medical Data Mart**

*An Internship report submitted in partial fulfilment of the requirements*

*For the award of the Degree of*

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE ENGINEERING

By

K. Mounica

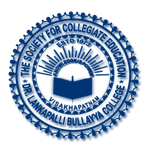
(Regd. No.: 312136410050)

Under the esteemed guidance of

Mrs. T.Arati

Asst.Professor

Department of Computer Science Engineering



**DEPARTMENT OF COMPUTER SCIENCE ENGINEERING**

**Dr. L. BULLAYYA COLLEGE OF ENGINEERING**

(Permanently Affiliated to Andhra University, Visakhapatnam)

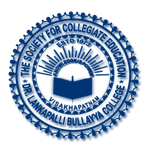
New Resapuvanipalem,Visakhapatnam-530013

Year of submission:2023

**Dr. L. BULLAYYA COLLEGE OF ENGINEERING**

**New Resapuvanipalem, Visakhapatnam-530013**

**Department of Computer Science Engineering**



**Bonafide certificate**

This is to certify that Mr **E. Divij Vignesh** bearing register numbers320136410035student of final yearB.Tech in Computer Science Engineering, has carried out the internship work titled ”**Identifying the optimal shipping liner”** at Sprint Exports Pvt.ltd, Visakhapatnam during the academic year **2022-23.**

**Head of the Department**

Dr.D.Madhavi

Professor

Dept. of Computer Science Engineering



**ACKNOWLEDGEMENT**

I thank Prof.D.Deepak chowdary, Principal Dr.L.Bullayya College of Engineering for permitting us to carry out this Internship.

I would like to thank **Prof. D. Madhavi***,* HOD, Dept. of C.S.E, Dr.L.Bullayya College of Engineering, Visakhapatnam, for her guidance and assistance in producing this work.

I am deeply indebted to my Internship guide **R. Vijay Srinivas**, Finance Manager, Sprint Exports Pvt.ltd, Visakhapatnam, for guiding me throughout the project inspite of her/his busy schedule.

Apart from the efforts of me, the success of this internship depends largely on the encouragement of other Faculty of C.S.E, Dr. L. Bullayya College of Engineering Visakhapatnam. I take this opportunity to express my gratitude to the entire faculty who has been instrumental in the successful completion of this Internship.

Also deserving of thanks are my family and friends, for their support for their confidence in my achievements.

**E. Divij Vignesh**

**Abstract:**

In the dynamic landscape of international trade, the efficient and cost-effective transportation of goods is paramount for the success of export businesses. This project delves into the intricate logistics managed by shipping liners, emphasizing the critical role they play in delivering goods across borders. The focus is on the delicate balance between selecting a shipping liner with a minimal base freight per container and navigating additional charges, such as Terminal Handling Charges. Export businesses often grapple with challenges such as storage costs and unpredictable ship availability, impacting the overall expenditure and profitability.

The learning objectives of this project encompass understanding shipping logistics, conducting a thorough cost analysis, assessing risks associated with export transportation, and mastering the art of vendor evaluation. By systematically evaluating shipping liners based on rates, additional charges, and ship availability, the project aims to empower export businesses to make informed decisions that optimize costs and enhance overall profitability. The ultimate goal is to streamline decision-making processes, reduce uncertainties, and ensure timely deliveries, thereby contributing to the sustainable success of export enterprises in the global market.

**About the organization:**

Delivering excellence for over a decade now, Sprint Exports Pvt Ltd has emerged as one of the leading export companies in India.

Sprint Exports operates in Andhra Pradesh, which is the most favorable state in India for aquaculture business. Established in Visakhapatnam in 2004, the company has adopted world-class farming & production methods to grow synonymous with superlative quality in the shrimp industry. Our integrated operations, across hatchery and farming, span over 500+ acres of water spread area and produce 2000 tonnes of output.Our facility in Visakhapatnam, Andhra Pradesh is approved by the Export Inspection Council for export to all countries excluding Australia and Custom Union. Their facility is also certified with HACCP, BRC & BAP 4 star and Social Accountability Certificates such as Walmart, Kroger, Safeway and SA8000.

Upon harvesting, the shrimps are loaded into specially-designed containers & transported via trucks to the processing unit.Batches of shrimps then pass through several levels of processing & quality control at the unit.

After being thoroughly bubble washed by special equipment, the raw material is made free of extraneous substances before being scientifically tested for quality in the lab.

Upon approval, the shrimps are further processed with the help of state of the art facilities to upgrade its appearance, quality & consumability. A few sophisticated touches then make the shrimp customers ready.

**CONTENTS OF THE REPORT**

* Title Page
* College Certificate
* Internship Certificate
* Acknowledgement Page
* Abstract
* About the Organization
* Table of contents
* Chapters

1. Learning Objectives (Goal)
2. Prerequisite Basic Skills
3. **Learning Process**
4. **Learning Outcomes**
5. Executive Summary
6. Future scope (Optional)
7. References as per APA format(Mendeley Tool)

* Appendices (Generated Reports/Output Screens)

Times New Roman 12pt Body, Justified

Side Headings 12pt Bold,

Title TNM 24 First letter Cap of each word

Minimum 20 pages spiral

Use diagrams appropriately with captions

Prepare Softcopy and get approval for printing

**1. Learning Objectives (Goal)**

* Understanding Shipping Logistics: Gain a comprehensive understanding of the logistics involved in international trade, specifically focusing on the role of shipping liners in transporting goods between nations.
* Cost Analysis: Learn how to analyze and calculate the total cost of shipping, including base freight, Terminal Handling Charges, and other additional fees. Understand the impact of these costs on the overall expenses of an export business.
* Risk Assessment: Develop the ability to identify and assess potential risks in the export business related to shipping, such as storage costs, delays due to ship availability, and other unforeseen circumstances that could affect the transportation process.
* Vendor Evaluation: Learn the criteria for evaluating and selecting the most cost-effective shipping liner. Understand the importance of considering not only the base freight per container but also additional charges, ship availability, and reliability in scheduling.
* Profitability Analysis: Gain skills in analyzing how the choice of a shipping liner impacts the overall profitability of an export business. Understand that a lower base freight per container may not necessarily result in overall cost savings if additional charges are high or if there are delays.

**Goals:**

* Optimizing Shipping Costs: The primary goal is to optimize the costs associated with shipping by selecting the most suitable shipping liner. This involves minimizing both the base freight and additional charges while ensuring timely and reliable transportation.
* Enhancing Profit Margins: Improve the overall profitability of the export business by making informed decisions regarding shipping liners. The project aims to demonstrate that a comprehensive evaluation of shipping options can lead to increased profits.
* Streamlining Decision-Making: Develop a systematic approach for export businesses to evaluate shipping options. The goal is to streamline the decision-making process by providing a framework for considering various factors, including costs, schedules, and reliability.
* Reducing Uncertainties: Mitigate the uncertainties associated with shipping logistics by equipping export businesses with the knowledge and tools to anticipate and address potential challenges. This contributes to more stable and predictable operations.
* Ensuring Timely Deliveries: Emphasize the importance of selecting a shipping liner that not only offers cost advantages but also ensures timely deliveries. Timeliness is crucial in maintaining a competitive edge and satisfying customer expectations.

**2. Prerequisite Skills Required**

* **Data Analysis:** Statistical Analysis: Proficiency in statistical techniques to analyze and interpret business data trends and patterns.
* **SQL**: Proficiency in SQL for querying and manipulating data in relational databases.
* **Data Visualization**: Familiarity with tools like Matplotlib, Seaborn, or any other data visualization platforms to effectively communicate complex data through charts, graphs, and other visual elements.
* **Programming Skills:** Proficiency in scripting languages like Python for data manipulation, analysis, and automation.
* **Domain Knowledge:** A good understanding of the specific industry or business domain to interpret data in a meaningful business context.
* **Data Storytelling:** Ability to convey insights and findings in a clear and compelling manner to non-technical stakeholders.
* **Problem-Solving:** Critical Thinking: Strong critical thinking skills to approach complex data challenges and solve problems effectively.

**3. Learning Process:**

* Clustering:

This method helps in the process of finding similar data points. This will help in the understanding of different liners and their charge per container to a destination. Even though the sample size is not that large (around 600 container data) we can find some similarity among liners.

* Regression:

Creating a regression model which will predict the price per container to a destination can be done. The main importance of this model is that we can estimate the expenses spent by the company on transportation. This cost estimate will help the company to estimate their profitability upon a purchase order. Predicting the price per container depends on various factors like

* + Time taken for loading the cargo inside the container in the factory.
  + The time of receiving the purchase order.
  + Transit time.
* Availability of ships (Depends on the shipping liner)

Considering all of the above factors we can estimate the cost per container accurately. With the given data by the company, we can find out up to what level these factors affect the price of the container.

* Monte- Carlo Simulations:

This can be used to simulate various scenarios in which different shipping liners can be chosen to estimate who is giving the best price for transporting a container. In some situations the buyer might require the cargo as soon as possible, in this situation reducing the transit time is the main goal. To achieve less transit time we can run these simulations to understand which liner is taking less amount of time for delivering the cargo. This probability based simulation will help the company in understanding the best possible scenario choosing the shipping liner.

* Matplotlib:

This python library is used to visualize variations in data in a graphical mode. Understanding many of the findings

**4. Learning Outcomes:**

The learning outcomes of Business data management encompass a multifaceted skill set that empowers individuals to extract meaningful insights from complex datasets. Through this project of business data management, I developed proficiency in statistical analysis and data manipulation using tools like Python, gaining the ability to uncover patterns and trends within information. They become adept at utilizing machine learning algorithms to build predictive models and make data-driven decisions. Furthermore, I acquired expertise in data visualization, effectively communicating findings to diverse audiences through charts and graphs.

**5. Executive Summary :**

This project focuses on Sprint Exports Pvt. Ltd. a Private Limited Company founded by Mr.G.S. Siva Kumar, incorporated on 19.3.2003 (Vide Certificate of Incorporation No.01-40688) with Registrar of Companies, A.P. Hyderabad. This business is a B2C type and deals with manufacturing and exporting of Shrimp.

Delivering the goods to the planned location and at the appropriate time is crucial in any export or import business. Shipping liners handle the logistics of moving these goods from one nation to another or from one city to another. The arrangement of containers and their delivery to the stores are entirely the responsibility of these businesses. It is important for the company's economy to select the best Shipping liner that charges a minimum charge and lesser transit time per container of goods. Shipping liner companies try to charge a base freight per container, but there are other costs like Terminal Handling Charges and other charges. Export businesses are required to pay these additional fees because of the following circumstances: The business will be required to pay additional storage costs if the goods are prepared for export but there are no ships available from the nearby port. Even though the shipping liner offers a lower base freight per container, the export expenditure as a whole is increased by additional costs. Frequently, export companies face difficulties in selecting the most cost-effective shipping method. Poor ship scheduling, underestimating waiting times, and other additional costs could be to blame for this issue, which will raise transportation costs and directly affect the company's profitability.

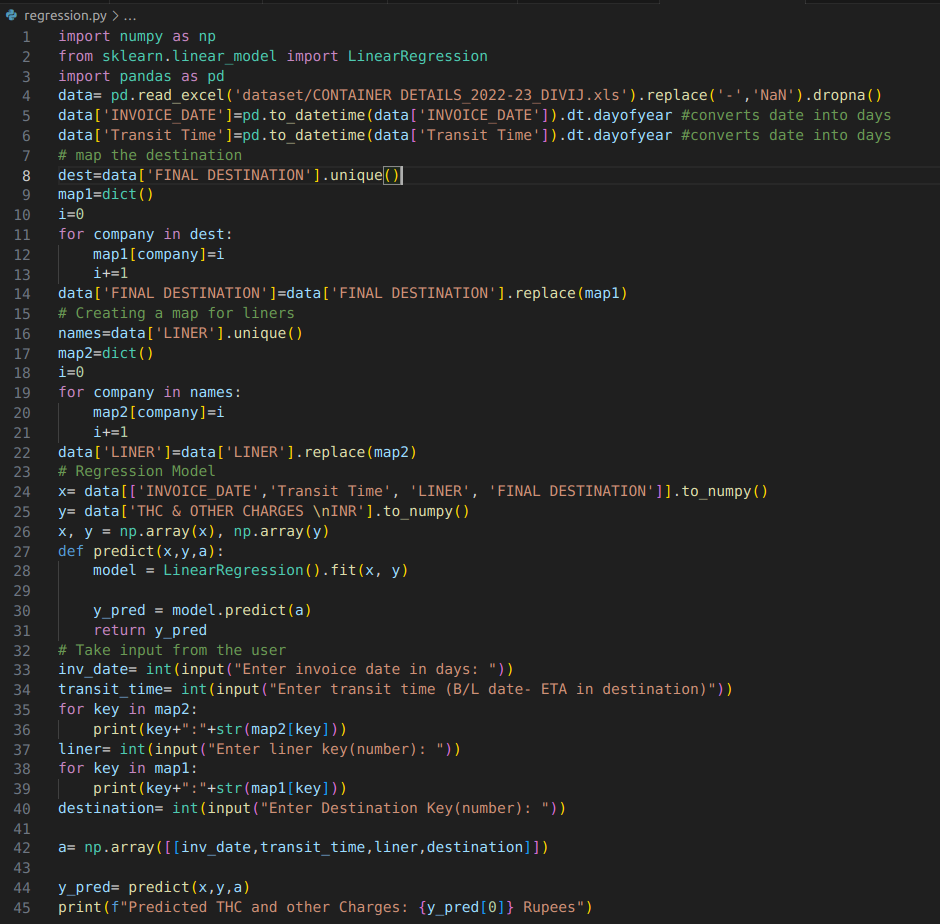
By evaluating the rates of all shipping liners, including any additional charges, and the availability of ships, the company will be able to select the most suitable shipping liner. Profits will rise as a result.

**6. Implementation Details:**

**Predicting THC and other charges using Regression:**

THC and other charges have a significant impact on the profitability per container for the company. Every liner has fixed Freight charges to a particular destination which does not change and are agreed upon by the company and the liners with an agreement i.e each liner will change the freight charges per month or per year based on the agreement. So there isn't much which we can do to reduce the Freight charges.

Therefore the overall determining profitability factor depends on the THC and other charges. Different liners charge different THC and other charges based on circumstances. These charges should be considered as variable costs as it depends on various factors (like cargo loading time into the container, plugin charges etc to name a few). This importance of THC and other charges makes it a much important feature which is to be estimated so that the company can have a better understanding regarding the profitability they can have per container.

The below python code demonstrates the implementation of Regression using Python and Scikit-learn library:

* **Input fields**: Invoice Date, Transit Time, Liner and Final Destination

Output Field: Predicted THC charges in Rupees

* **Preparing the data**:

Here, we used Pandas dataframe to read the excel sheet and get the data. I’ve converted INVOICE\_DATE and Transit Time date fields into integers(days of the year) for the ease of feeding the integer data into the regression model( 21st Jan is outputted as number 21 as it's the 21st day of the year).

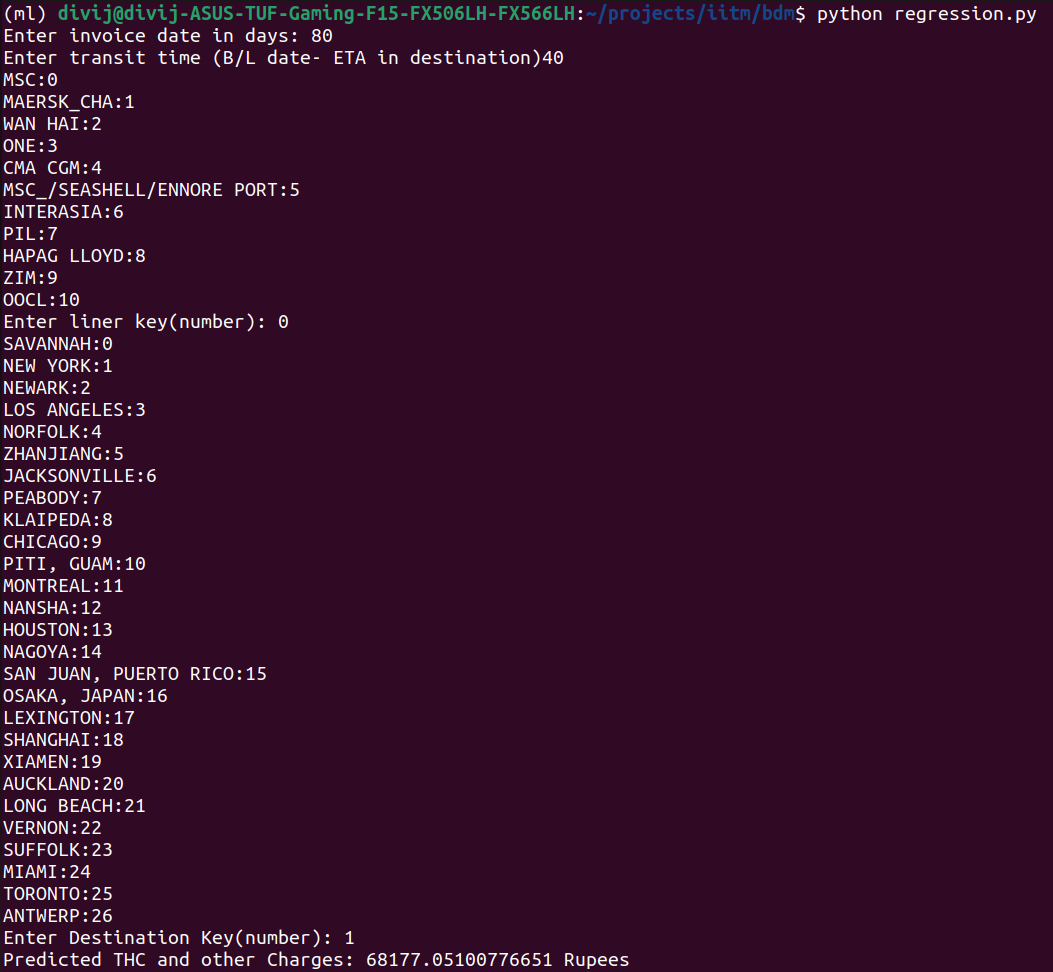
We know that Liner details and Final Destination is not Numeric kind of data, rather it is a Categorical kind of data which we cannot directly feed into the regression model. So I've created a dictionary data type map which maps each liner and destination to a unique key integer value which can now be used in the regression model. During the data inputting the user should enter the key values of the destination and liners as input.

* **Output**:

In the below example after running the regression.py file( contains the above image code), we can see a prompt asking for the input fields. I’ve entered the Invoice date as 80 i.e it is the 80th day of the year which is March 21st. Transit time is also mentioned in the form of days.

After entering these two fields, Liner map which which was created will be displayed in the output for the user to enter the proper key value of the Liner, followed by entering the destination key( in the output the destination key map is not displayed as it contains more elements which cannot be displayed in a single screenshot).

After entering the inputs, the Predicted THC and other charges field will be displayed in Rupees. Users can vary the liner key by keeping the invoice date, transit time and destination constant to get different values of THC and other charges. These values can be used to understand how THC varies from liner to liner to a fixed destination.

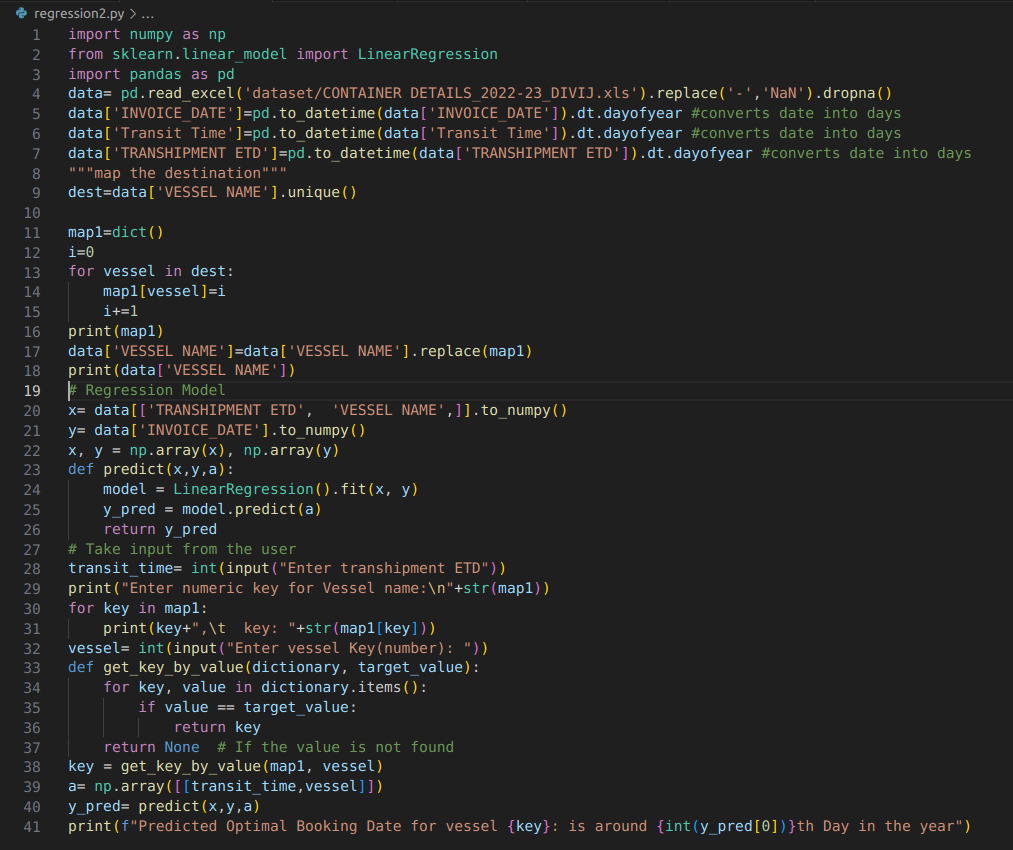


**Predicting Optimal Date at which a ship is available:**

Understanding the frequency at which motherships are moving from are also important as we can have an estimate on which date we should load the container. If the mothership is departing from the port on the 20th day of the month and we load the container at the start of the week, it itself will increase the charges imposed by the Liner. Plugin charges, trailer charges etc, will increase when containers are loaded and sent to the port way ahead of the departure time. To prevent this the Company should utilize the cooling storage available in the factory and contact the liner to send the trailer to the factory for loading the container 2 or 3 days before the departure. This will decrease the additional charges imposed by the liner to a significant amount.

As we need to decrease these THC and other charges, the importance of understanding the availability and departure of the ships are crucial. Thus predicting the optimal date for each ship will allow us to understand the availability of ships and choose the optimal liner based on the available ships.

The below code will help us in understanding the optimal booking day of each ship.



* **Input fields**:

Vessel Name and Estimated Time of Departure from Transhipment port

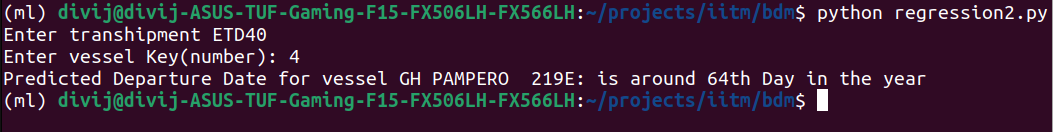
* **Preparing the data**:

Since we want to convert all of the data into numeric type for feeding the data into the regression model, ETD of Transhipment field is converted into Days of the year and a map is created for Vessel Name as it is a categorical type of data and the created map will assign a unique integer value to convert the data into numerical type.

* **Output**:

In the below code output , we can observe that after entering the Transhipment ETD(in days) and vessel key( map is not printed as it contains more than 60 ships which cannot fit in the window) the predicted departure day for the Vessel is displayed as the output.

This prediction helps the company to understand the frequency of each ship and thus provide a decent amount of understanding on when the departure of the ship happens from the Transhipment Port. After obtaining the date, the company should book the appropriate ship owning liner , at least 10 days before the predicted date. As it will roughly take a week for the container to reach the transhipment port from Viskhapatnam.



**High Freight due to COVID:**

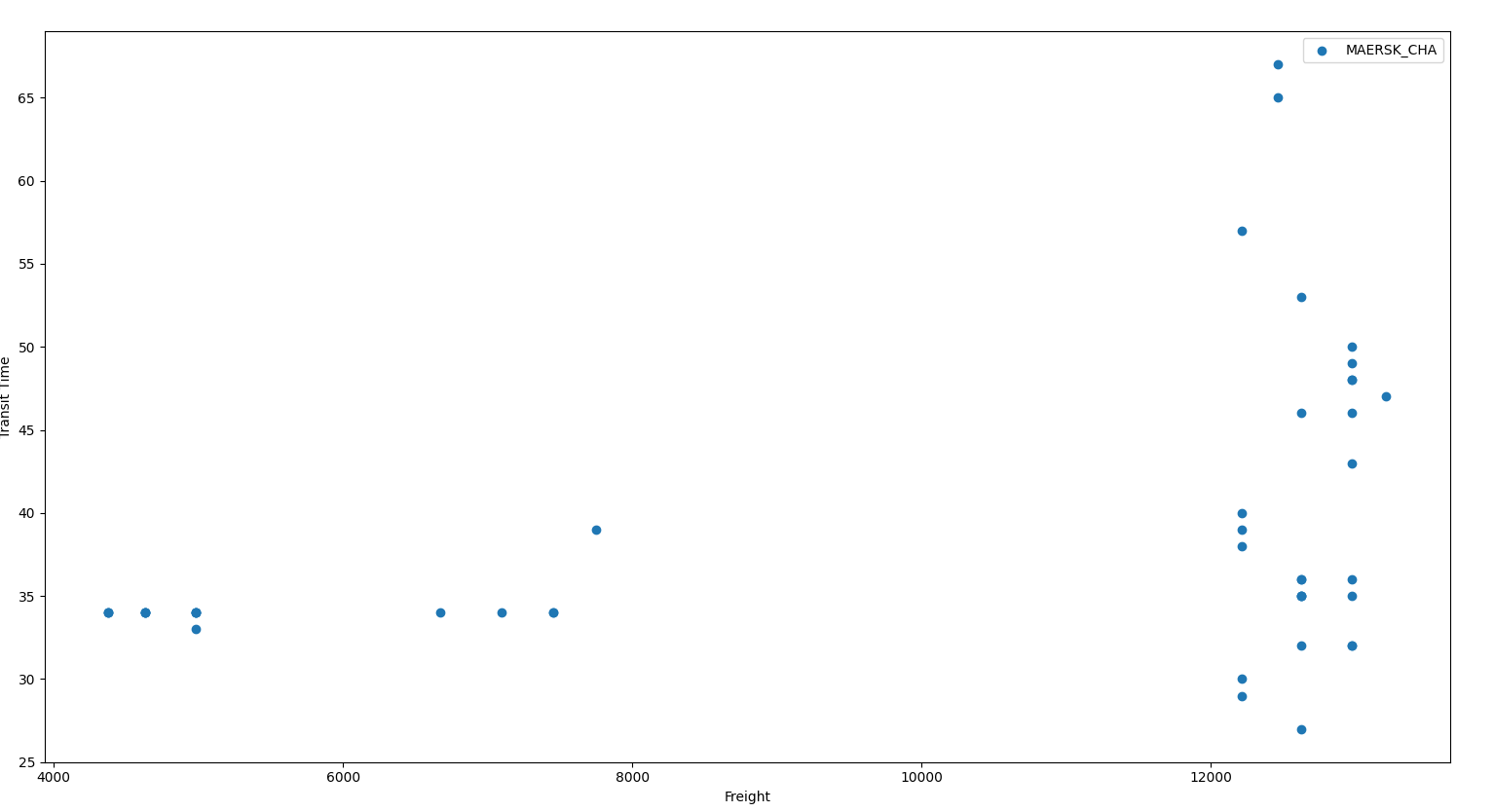
The below graph plot is Freight against Transit time of the containers for Location Chicago.

We came to a conclusion before that the freight of a container is fixed and is generally agreed upon by both parties( company and the liner). The agreement could be monthly, quarterly or yearly also depending upon the liner. This means that to a particular destination in a year we should observe similar Freight charges varying for a few thousand dollars but in the below graph we can observe a spike in the Freight charges which take more than 12,000 USD as freight charges. This is unusual because based on the latest data freight charges are generally around 4,000 USD mark and now the majority of the data points are having 3 times more freight charges and increased transit time also . This kind of behavior was also observed for different Locations like Newark etc.

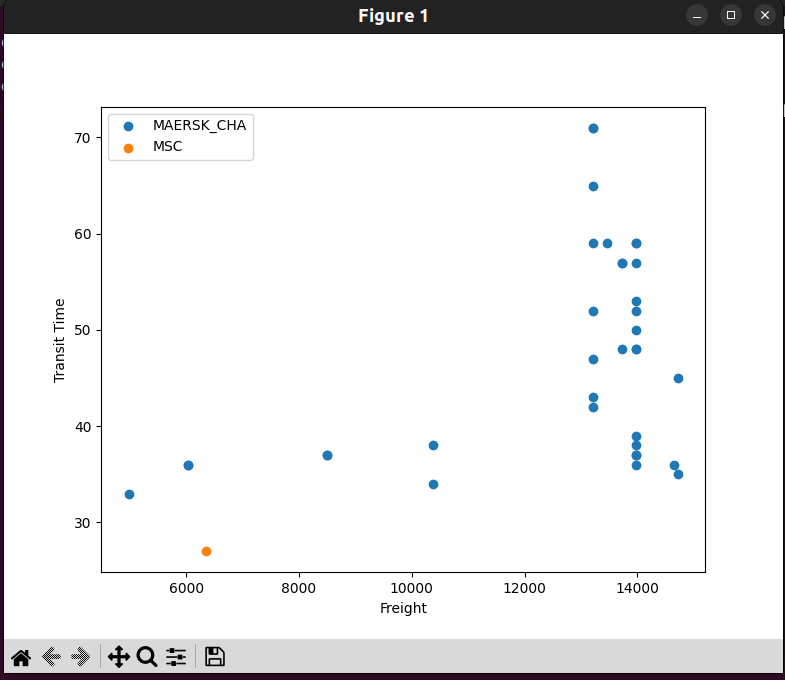
Upon searching these data points and upon discussing with the company personnel , the conclusion which I came to is that there were shortages in containers during the early 2022 year due to COVID-19 and these shortages had made the freight charges to become more than twice the initial price. These abnormal situations have led to bad performance of the company during these post COVID years. Increase in Freight had decreased the profit margin to a significant level.

As an addition to shortages of containers, the Transit time of the containers has also increased 1.5 times the initial transit days which used to be less than 40 days. This led to an overall increase in Freight and decrease in performance of the company.

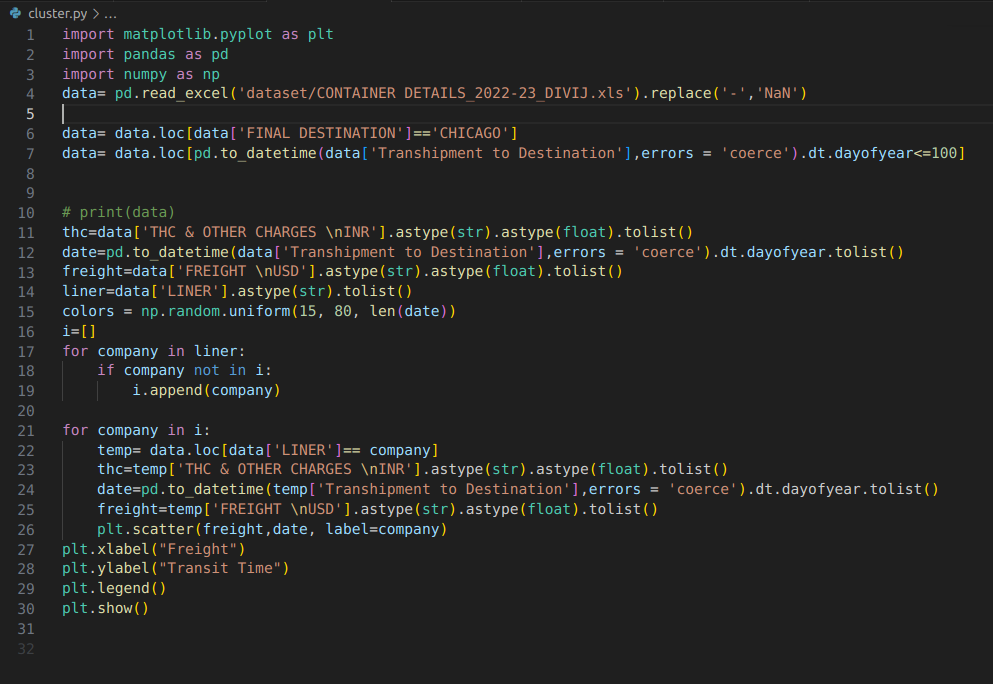
The newer data points are showing significant improvement over others as container shortages are restored and the freight also decreased as a result.



Newark Graph



Chicago Graph



The above code contains all of the data of CHICAGO containers and a graph is plotted against the Freight data and Transhipment port to Destination port time difference( in days).

**Observing nature of clusters formed:**

Understanding the clusters formed upon plotting a graph will tell us a lot about understanding the nature of the data points. Using this particular cluster we can understand the kind of common features which we have among the data points.

Hypothesis: There is a direct correlation among Destination , Transit Time and Freight as we know that depending upon the destination the Transit time( time taken to reach the destination port) and freight vary. This kind of correlation between different features can be observed by the below graph which is plotted against Transit Time and freight.

After plotting the graph, we can immediately observe 2 major clusters forming in the graph and there are significant outliers in the data mainly due to the sharp rise in Freight during COVID-19 which we observed in the previous finding. Without worrying much about these outliers we can try to have some understanding of the nature of the clusters formed.

We know that there is a direct correlation of Transit Time and freight on Destination and Transhipment ports as it is obvious if the destination is far away from the starting point, more time is spent traveling on sea which in turn impacts freight charges. This justifies the correlation among these features.

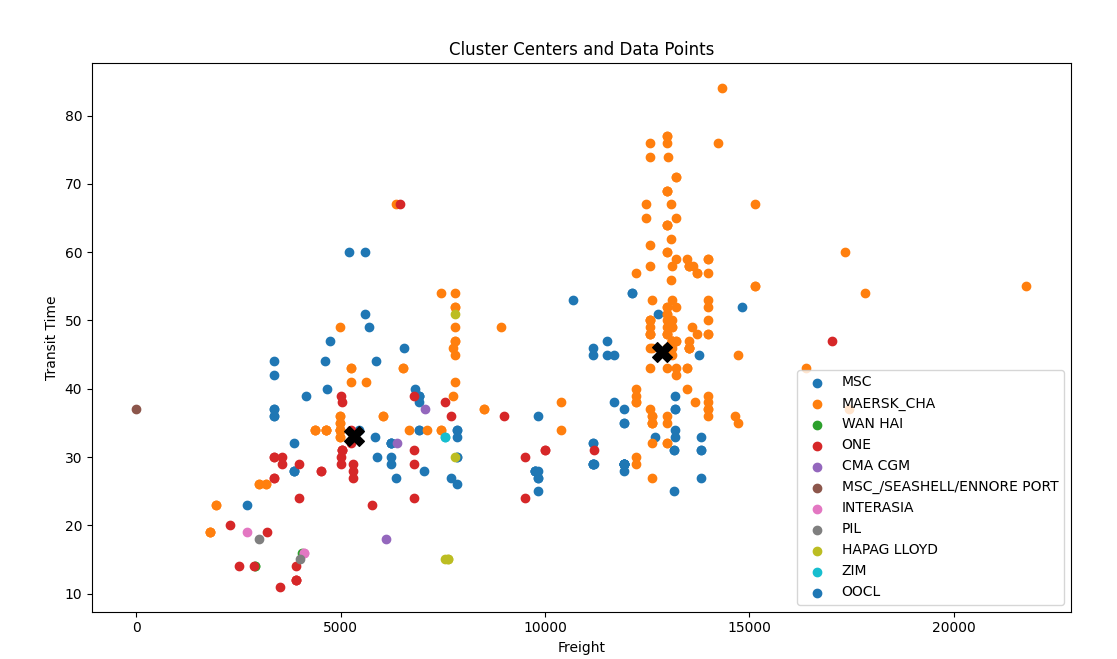
So upon plotting the data, finding the cluster centers (k - number of clusters) is important and we know in clustering the value of k could be anything. We should find that k which will give the least Squares Error.

**Least Squares Error** = Sum of all ( Data - Cluster Mean)^2 + Penalty

Upon varying the k Value we will have different error values, choose that K which produces reasonable error. Here the goal is to have least Squares error by varying the value of k and the penalty will ensure that the correct K value is chosen.

I’ve found out the k value using the Elbow method which will tell us the correct k value. Upon running the elbow method on the data the k value was given to be 2.

Here in the below diagram,



We can observe major clusters formed(by using KMeans Algorithm) and cluster centers represented by an X symbol. These centers tell us about the common features in the surrounding data points. To observe the common features I wrote a python code which will give those data points which are in the proximity of the cluster center.I took a freight condition of 4000 USD to 10,000 USD and Transit time condition between 20 to 40 days. Similarly for cluster 2 freight range from 12,000 USD to 15,000 USD and transit time between 30 to 60 days.

* **Observations**:

In cluster 1, the common feature which they have is that majority of the containers in these data points go through same transhipment port (i.e Singapore and Tanjung pelepas).

In cluster 2, the common feature which was observed is that the majority of the containers in these points go through Colombo (Transhipment Hub).

* **Confirming the hypothesis after observations**:

These observations in the cluster features support the initial hypothesis which we had that these Freight and Transit Time values are directly correlated on the Transhipment port and destination.