TRAFFIC AND CAPACITY ANALYTICS FOR MAJOR PORTS

1. INTRODUCTION

1.1 Project Overview

The transportation of bulk freight and long-distance passengers, traffic and congestion on rail corridors becomes major challenge. Data analytics can be applied to visualize freight transportation and congestion on rail corridors across major railway ports to get better insight of the working of port network and to improve the port connectivity. There are three users, Government sector, Business people and People. By using the dashboard, users can able to track the freight movement patterns across major ports based on rail corridor congestion analysis.

1.2 Purpose

- To develop analytics dashboard that can be used to track the freight movement patterns across major ports.
- To visualize the capacity and to ensure less traffic on the ports.
- To predict delays at the ports due to congestion and makes supply chain efficient.
- To easily track the ports and improve port connectivity.

2. LITERATURE SURVEY

2.1 Existing problem

The Indian Railways has a capital base of about Rs. 100000 crores and is often referred to as the lifeline of the Indian economy because of its predominance in transportation of bulk freight and long distance passenger traffic. The network criss-crosses the nation, binding it together by ferrying freight and passengers across the length and breadth of the country. To solve this problem, develop a data analytics dashboard to track the freight movement patterns across major ports and to track the ports and improve port connectivity

2.2 References

2.2.1 Literature Survey 1

George Stergiopoulos; Evangelos Valvis; Dimitris Mitrodimas; Dimitrios Lekkas; Dimitris Gritzalis, “Analyzing Congestion Interdependencies of Ports and Container Ship Routes in the Maritime Network Infrastructure”, “IEEE Access”, Volume:6,23October 2018.

Work:

They have worked on a risk-based interdependency analysis method capable to detect large-scale traffic congestions between interconnected ports and ship routes in the maritime network and provide solutions to improve flow. The application-oriented, interdisciplinary effort culminated in a prototype tool able to analyze the historical data for container ships in the entire global maritime network and detect congestion dependencies. **Merits**:

The tool can also be used to identify key shipping routes or ports that:

(i) are prone to delays.

(ii)greatly affect the overall maritime network due to position, connections and risk of congestion.

(iii)get affected the most by delays in previous route legs.

Demerits:

The algorithm does not predict future port congestion based on machine learning, but rather extracts congestion patterns and trends through existing, real-world sensor data.

2.2.2 Literature Survey 2

Narinder Singh Punn; Sonali Agarwal, "Modelling railway delay propagation as diffusion-like spreading", Disordered Systems and Neural Networks, arXiv:2105.06111, 13 May 2021.

Work:

They have defined a novel model for studying delays in Railways systems, where they spread across the railway network via a diffusion-like process. They have applied the model to the Belgian railways and studied its performance in simulating the delay propagation in severely disrupted railway situations. They have discussed the role of spatial aggregation by proposing to cluster the Belgian railway system into sets of stations and adapted the model accordingly and founded such aggregation improved the model performance.

Merits:

- Quick computation.
- Ease of applying various statistical tools like spectral methods.

Limitations:

Limitations are related to the directional and discrete nature of delays and the trains carrying them.

2.2.3 Literature Survey 3

Luca Cazzanti, Antonio Davoli, Leonardo M. Millefior ," Automated port traffic statistics: from raw data to visualisation" , "IEEE International Conference on Big Data ", November 2017.

Work:

They worked on big data processing pipeline design and visual analytics to prototype the Maritime Patterns-of-Life Information Service (MPoLIS). They have addressed three main requirements:

a) storing and processing large amounts of data.

b) on-demand availability of statistical summaries of vessel traffic in ports.

c) intuitive and interactive interface for subject matter experts (SMEs) in the maritime domain.

Merits:

1.The modular and scalable architecture supporting MPoLIS makes it easily extensible with a richer set of port statistics to provide a more comprehensive picture of port traffic patterns.

2.The cloud-based Tableau front-end makes MPoLIS immediately deployable to any organization interested in understanding the maritime traffic patterns.

2.2.4 Literature Survey 4

Mingyu Pi; Hanbyul Yeon; Hyesook Son; Yun Jang , "Visual Cause Analytics for Traffic Congestion", "IEEE Transactions on Visualization and Computer Graphics", Volume: 27, 01 March 2021.

Work:

They worked on a technique to analyze the cause of traffic congestion based on the traffic flow theory.They have extracted vehicle flows from traffic data, such as GPS trajectory and Vehicle Detector data. They have also detect vehicle flow changes utilizing the entropy from the information theory. Finally they built cumulative vehicle count curves (N-curve) that can quantify the flow of the vehicles in the traffic congestion area.

Merits:

The system classifies the causes of traffic congestion and can be used efficiently in road planning.

They have also found the possibility of classifying abnormal traffic congestion patterns, such as traffic accidents and bad weather, by combining external data.

Limitations:

If the distance between the intersections is small, the glyphs can overlap.

It can take a long time for an analyst to understand traffic conditions at the intersection since a single glyph contains much information.

2.2.5 Literature Survey 6

Francesco, Rotoli; Gabriele, Malavasi; Stefano, Ricci, "Complex railway systems: capacity and utilisation of interconnected networks", European Transport Research Review volume 8, Article number: 29 (2016).

Work:

They have worked in evaluating capacity and utilisation rates of rail systems at different geographical scales. The method has provided a first identification of 'weak' links or nodes for which, then, specific and detailed analyses should be carried out, taking into account more in depth their actual configuration, the technical characteristics and the real composition of their traffic.

Merits:

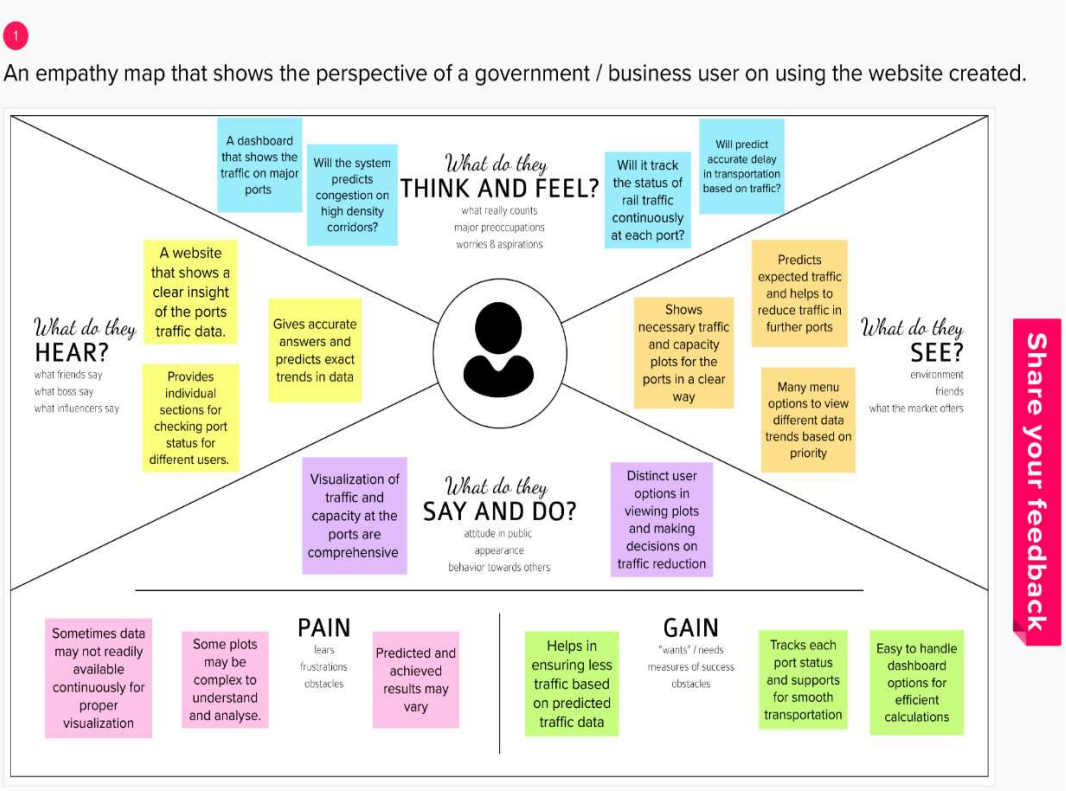
The approach proposes a schematization of typical components of a rail network be applied in case of lack of more detailed data.

Limitations:

The presented methodology might be valuable only in case of feasible studies (when time, cost and complexity of more comprehensive approaches would be less appealing).

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

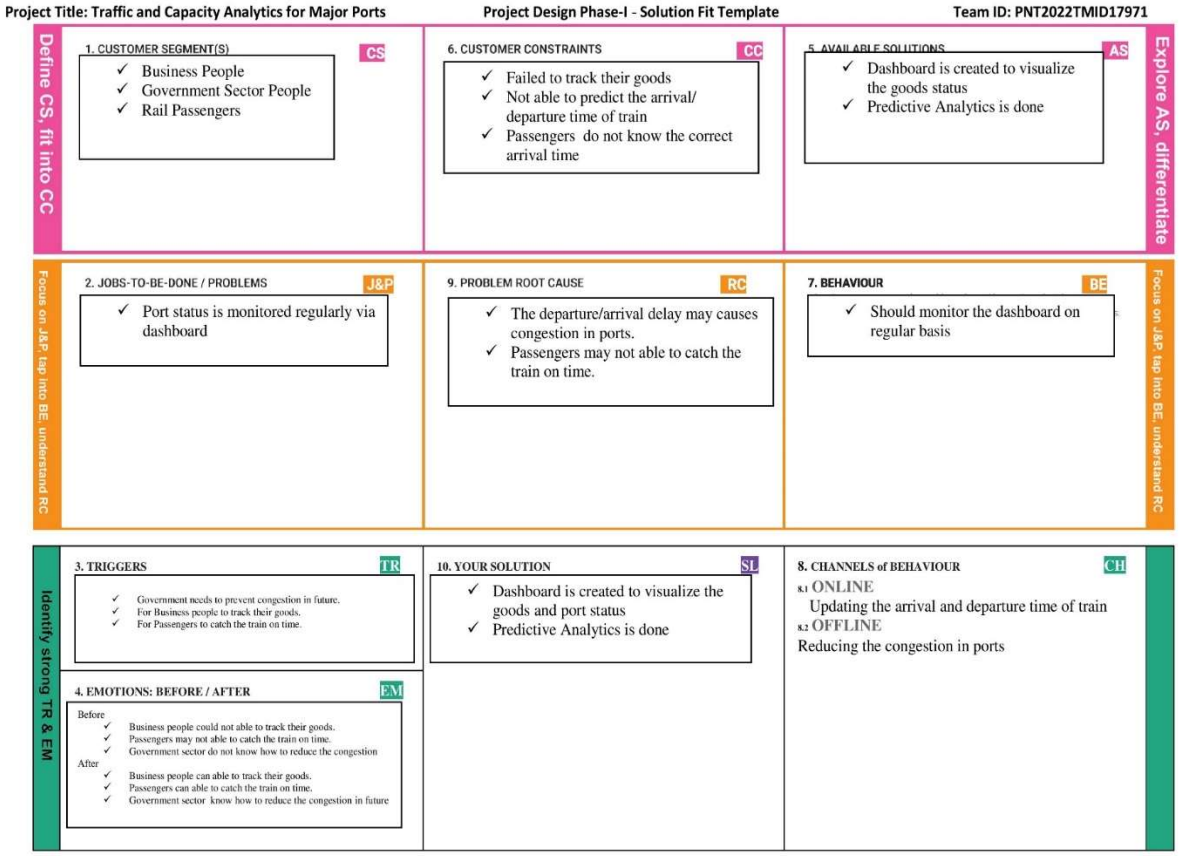


3.2 Ideation & Brain Storming

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The Indian Railways has a capital base of about 1 lakhs crores and is often referred to as the lifeline of the Indian economy. As it includes transportation of bulk freight and long-distance passengers, traffic and congestion on rail corridors becomes a major challenge.
2.	Idea / Solution description	Data analytics can be applied to visualize freight transportation and congestion on rail corridors across major railway ports to get better insight of the working of port network and to improve the port connectivity.
3.	Novelty / Uniqueness	Can also predict the time at which the particular train will arrive and depart.
4.	Social Impact / Customer Satisfaction	Adequate resources will be provided for the customers regarding the arrival, departure and delay of the trains.
5.	Business Model (Revenue Model)	Businesses using railway ports can easily track the trains. Government can use data analytics dashboard to ensure less traffic on the ports.
6.	Scalability of the Solution	The solution can be used almost for all modes of transportation including the ships and so on. Thus it is scalable for almost all modes of transportation.

3.4 Problem Solution Fit



4. Requirement Analytics

4.1 Functional Requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email
FR-3	User Input Acceptance	The dashboard accepts user input by means of selecting the location of the ports.
FR-4	Options for User to filter location of ports	The user can use filter options to view ports by countries.
FR-5	Visualization of ports.	The dashboard provides various visualization techniques to understand the flow.
FR-6	Providing Delay Information of trains.	The dashboard is able to provide the user the information like delay of a particular train to the ports.

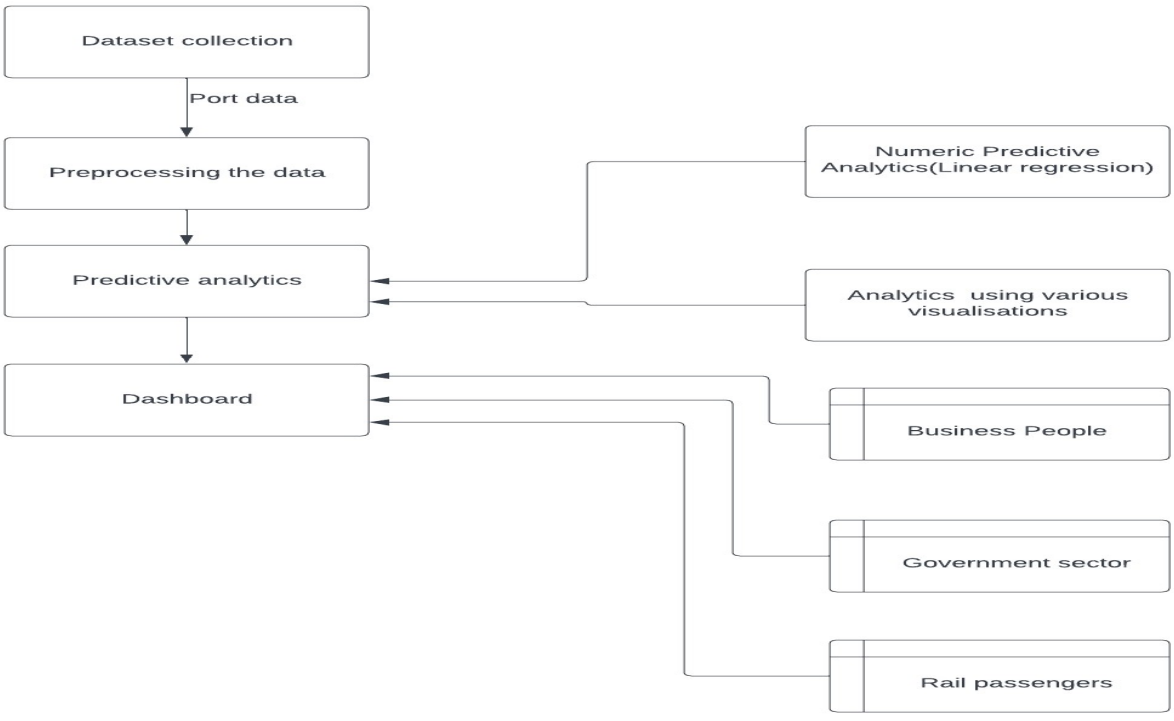
4.2 Non Functional Requirement

Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The dashboard is able to provide the users the consistency and the aesthetic they expect. The user can constantly use the dashboard without any flaw in the visual quality.
NFR-2	**Security**	The dashboard is much secured that the data of the users are kept confidential and also it is not prone to any kind of attacks.
NFR-3	**Reliability**	The failure rate is minimal and the failure can easily be rectified using the measures. Thus this makes the dashboard much reliable.
NFR-4	**Performance**	The dashboard gives better performance. It provides the user a convenient and flexible User Interface.
NFR-5	**Availability**	The dashboard is always available to serve the users. The availability is ensured in such a way that the user can access the dashboard any time anywhere.
NFR-6	**Scalability**	The dashboard is highly scalable. It can withstand any increase or decrease of loads.

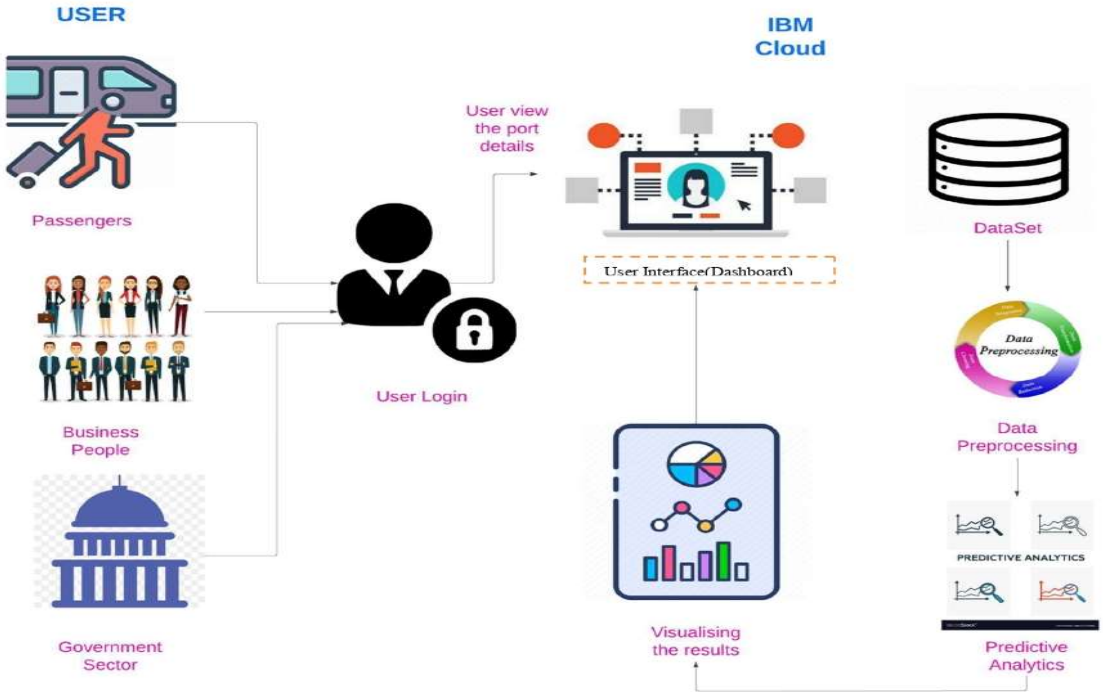
5. Project Design

5.1 Data Flow Diagram



5.2 Solution & Technical Architecture

Technical Architecture:



5.3 User Stories

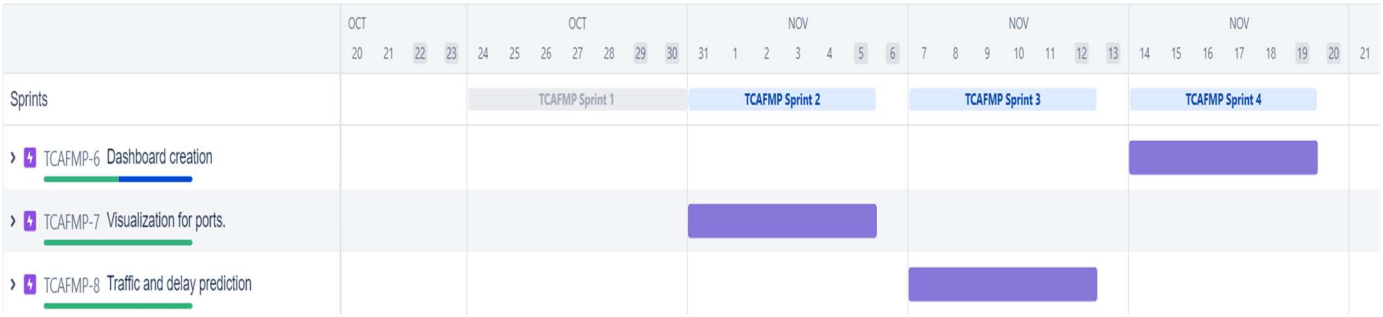
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Business People	Monitoring	USN-1	As a user, I can view the dashboard to see the port status .	I can visualize the port status in dashboard.	High	Sprint-1
	Tracking	USN-2	As a user,I can track the goods.	I can track the goods by it's arrival/departure time	High	Sprint-1
Government Sector People	Viewing	USN-1	As a user,I can view the port status regularly	I can able to know the port status	Low	Sprint-2
	Predicting	USN-2	As a user,I will reduce the congestion in ports by predicting the port congestion through dashboard.	I can able to predict the congestion in future	High	Sprint-2
Passengers	Tracing	USN-1	As a user, I can trace the arrival/departure time of rail in ports.	I can able to track the correct time of rail.	High	Sprint-2

6. Project Planning and Scheduling

6.1 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.2 Reports from JIRA



7. Coding and Solutioning

7.1 Port Traffic & Capacity Data Statistics

Traffic and capacity data for different ports are processed and its statistics provides insight on measures on capacity and traffic throughput for major ports and shows the port-wise summary of the traffic and capacity.

Code:

```
import numpy as np
import pandas as pd
df = pd.read_csv('D:/ibm/datafile_02.csv')
print(df.columns)

Index(['Port', 'Traffic in Eleventh Plan (MT) (2011-12)Proj.',
       'Traffic in Eleventh Plan (MT) (2011-12) Ach.',
       'Traffic in Eleventh Plan (MT) (2011-12) %',
       'Total Capacity  in Eleventh Plan (MT) (2011-12) Proj.',
```

```
       'Total Capacity  in Eleventh Plan (MT) (2011-12) Ach.',
       'Total Capacity  in Eleventh Plan (MT) (2011-12) %'],
      dtype='object')
```

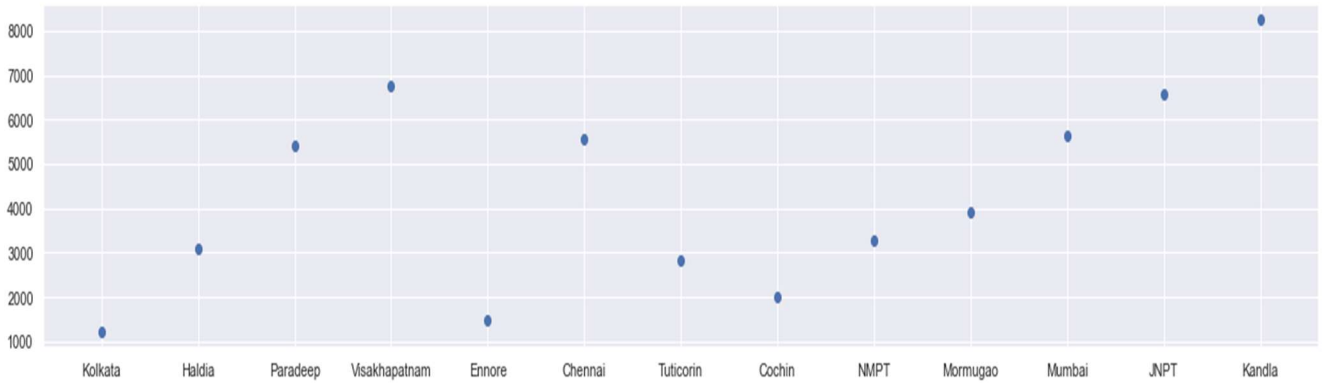
```
df.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-
12)Proj.':'Traffic_Projected','Traffic in Eleventh Plan (MT) (2011-12)
Ach.':'Traffic_Achieved', 'Total Capacity  in Eleventh Plan (MT) (2011-12)
Proj.':'Total_Capacity_Projected', 'Total Capacity  in Eleventh Plan (MT) (2011-12)
Ach.':'Total_Capacity_Achieved'}, inplace = True)


Traffic_Percent = round((df.Traffic_Achieved/df.Traffic_Projected)*100,2)

Total_Percent = round( (df.Total_Capacity_Achieved/df.Total_Capacity_Projected)*100,2)

# Replacing the existing columns with newly created columns
df.rename(columns = {'Traffic in Eleventh Plan (MT) (2011-12) %':'Traffic_Percent%','Total
Capacity  in Eleventh Plan (MT) (2011-12) %':'Total_Percent%'}, inplace = True)
df.iloc[:,3:4] = Traffic_Percent
df.iloc[:,6:] = Total_Percent
df
```

7.2 Port Data Visualization

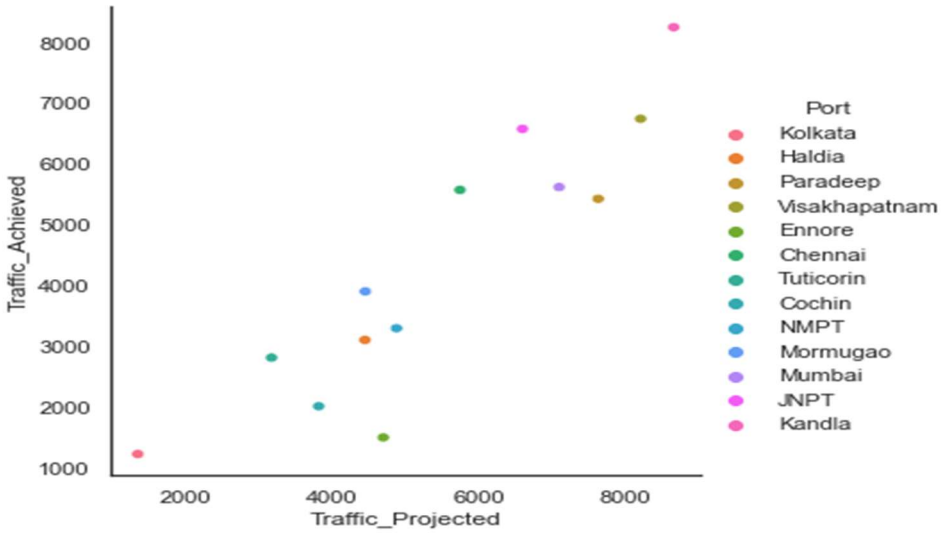
Different visualization charts in python are plotted for better understanding of the traffic and capacity data of major ports.

Code:

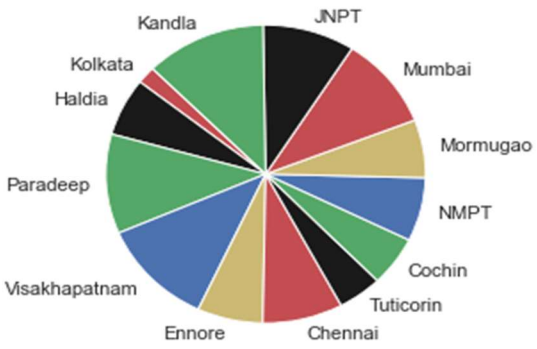
```
import matplotlib.pyplot as plt
import seaborn as sns
plt.scatter(df.Port,df.Traffic_Achieved)
sns.set_style('white')
sns.set_context('notebook')
```



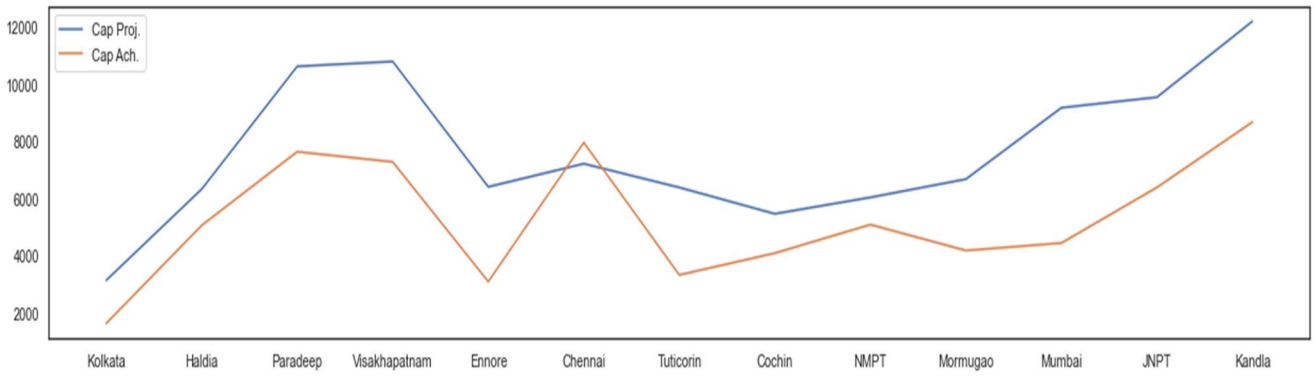

```
sns.relplot(data=df,x="Traffic_Projected",y='Traffic_Achieved',hue='Port')
```



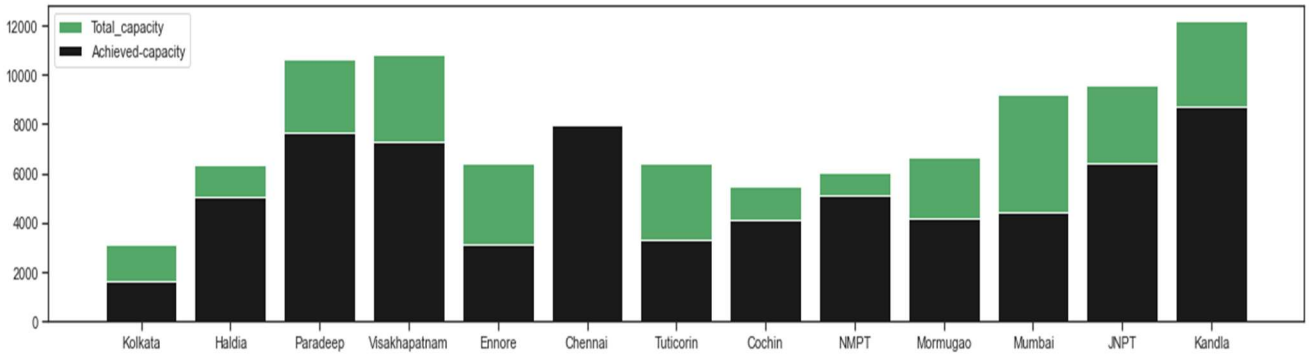

```
colors=['r','k','g','b','y']
plt.pie(df.Traffic_Projected,labels=df.Port,colors=colors,startangle=135)
```




```
plt.plot(df.Port,df.Total_Capacity_Projected,label='Cap Proj.')
plt.plot(df.Port,df.Total_Capacity_Achieved,label='Cap Ach.')
plt.legend()
```




```
plt.bar(df.Port,df.Total_Capacity_Projected,label='Total_capacity',color='g')
plt.bar(df.Port,df.Total_Capacity_Achieved,label='Achieved-capacity',color='k')
plt.legend()
```



7.3 Traffic and delay prediction for ports

Traffic and congestion prediction for the rail ports are achieved using the Machine learning technique called Linear Regression.

Code:

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)

from sklearn.linear_model import LinearRegression
mlr=LinearRegression()
mlr.fit(x_train,y_train)

#Prediction using the model
mlr.predict(x_test[0:5])

#Performance evaluation metrics
from sklearn.metrics import r2_score
r2_score(mlr.predict(x_test),y_test)

from sklearn.metrics import mean_squared_error
a = mlr.predict(x_test)
mean_squared_error(a,y_test)
```

8. Conclusion

Thus the proposed Traffic and Capacity Analytics for Major Ports can able to track the movement of freight movement patterns across ports and visualized the capacity, ensured the less capacity traffic on the major ports. The proposed system predicted the delays at the ports due to congestion and make the supply chain efficient.