



Analytical Backend Dashboard for India Tour Route Planning Using IRCTC Data

Prof Dr. M. G. Panjwani¹, Sonu Tiwari², Adhishree Kayate³, Vaidehi Dambhare⁴, Omkar Sarodey⁵,
Preshit Borkar⁶

¹Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

²Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

³Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

⁴Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

⁵Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

⁶Department of Computer Technology, Priyadarshini College of Engineering, Nagpur, India

Abstract

Indian Railways (IR), often regarded as the country's lifeline, connects every part of the second-most populated nation. The largest logistics carrier's logistical and financial performance has been at a low point for a few years. The progressive decline in IR's health not only grabs people's attention but also makes them curious about the underlying causes of this fragility. The worrying performance of IR during the past few years has attracted a tonne of literature attention, but there has been no satisfactory identification of the illnesses. In light of this, a study of declining performance to identify the underlying causes is necessary. For the aim of analysis, the study uses a variety of metrics, including OR, IR performance, etc. IR is plagued by subpar OR, and IR performance, and is on the edge of going bankrupt. Subventions like subsidies should no longer be used, and outdated pricing strategies should be abandoned in favor of a pricing plan based on how much traffic can be supported to prevent collapse and paralysis. The goal of the analysis-based Python project IRCTC Live Dashboard is to create a dashboard that will provide information about the trains. This dashboard will assist in determining the availability of trains and their specifics from the source station to the destination station. This project is dynamic, takes in many different inputs, and produces results. It is a data analysis-based project where the dashboard is created and analyzed using various Python modules. This project has a huge amount of room for growth in a variety of data science areas, including data analysis.

Keywords: Indian railways; performance of Indian railways; performance of IR; operating ratio (OR); Data Preprocessing, Data cleaning, Data Optimization, Visualization.

I. INTRODUCTION

The interactive dashboard, visualization, and analysis of IRCTC (Indian Railway Catering and Tourism Corporation Limited) data are all described in this paper. The programme is unique in that it enables data blending and in-the-moment cooperation. The dashboard enables the display of numerous visuals in a single glance. Additionally, it is used to display only the most crucial information, which is always changing. The IRCTC encourages determining the availability of trains from the source station to the destination station and their specifics through data visualization and analysis. Data analysis involves looking over, cleaning up, modifying, and modeling data. The goal is to find relevant information, enhance decision-making, and inform conclusions. We forecast useful information or insights, such as graphs, from its data using a variety of datasets. Daily new 3 data are generated in the Real World from several sources, including log files and large volumes of IRCTC data. Data analysis's job is to combine the many data structures into a single format. Data is subjected to a variety of processes, including data collection, processing, cleaning, analysis, modeling, communication analysis, etc.

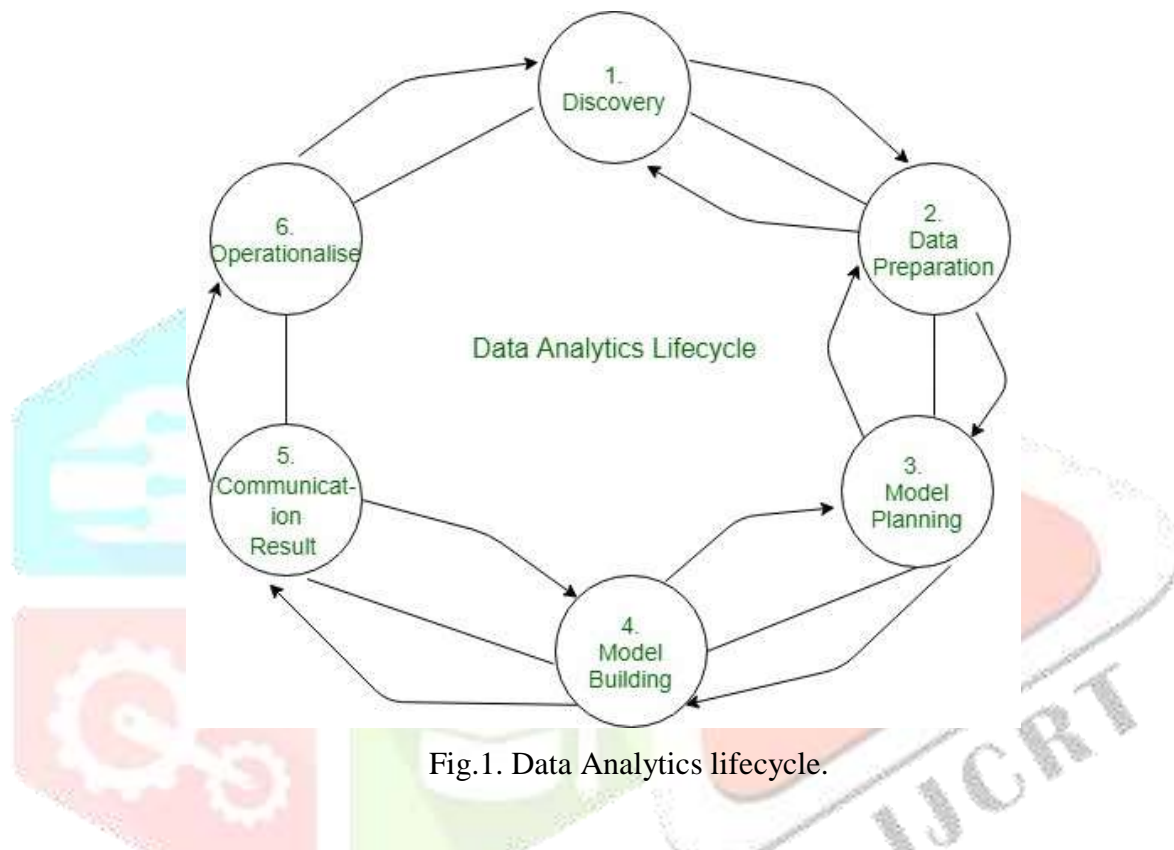


Fig.1. Data Analytics lifecycle.

The train timetable is the basis of a train operation organization which can also effectively present a comprehensive plan of transportation production activities. It is generally acknowledged that the effectiveness of the train timetable can directly impact operational safety, passenger satisfaction, and financial advantages of the urban rail transit (URT) system. It is crucial to assess the effectiveness of the train timetable since it can offer a wealth of helpful and insightful data for enhancing the level of service

II. LITERATURE SURVEY

| S. No. | Title of the Paper | Key Points | Conclusion | Reference |
|--------|--|--|---|-----------|
| 1 | Data analytics approach for train timetable performance measures using automatic train supervision data. | Data analytic framework, data preparation, data analysis. | Evaluation of the execution of the schedule is essential for raising the standard of the railway service. This study uses ATS data to propose a methodical data analytics methodology for TTPM. | [1] |
| 2 | An analysis of the performance of Indian railways. | Performance of Indian railway, financial measures of IR, safety measures of IR, operating ratio, capital-output ratio. | One of the important contributions of this study is the thorough examination of the effectiveness of IR from its two primary financial and logistical perspectives. | [2] |
| 3 | Data analyzing in Indian railway: A survey to analyze applications of datamining. | RAC (Reservation against cancellation), KNN (K-Nearest Neighbor), ARP – Advanced reservation period, clustering, TTE. | The review of other works using data analysis methodologies to identify TTE findings for ticket allocation, identify train delay causes, and forecast delay times was conducted in this study. | [3] |
| 4 | A study on passenger satisfaction towards inline booking in IRCTC application. | Passenger satisfaction, service quality, technology changes, flexibility. | Despite being created for a unique situation, the model may be applied to other services of a similar nature and help raise overall satisfaction while also enhancing the quality of life for the public. | [4] |

III. METHODOLOGIES

Planning & Scheduling

The planning of the project was divided into 6 phases. The phases are as follows:

- Problem Specification
- Research phase
- Understanding the Data set
- Coding
- Testing
- Dispatch

During the Seminar, the company gave us instructions on what they wanted from us and asked us to work accordingly. Requirements for the project were identified, and we started to work on the data set by understanding the data given to us later, we performed analysis, and code was written for the dashboard.

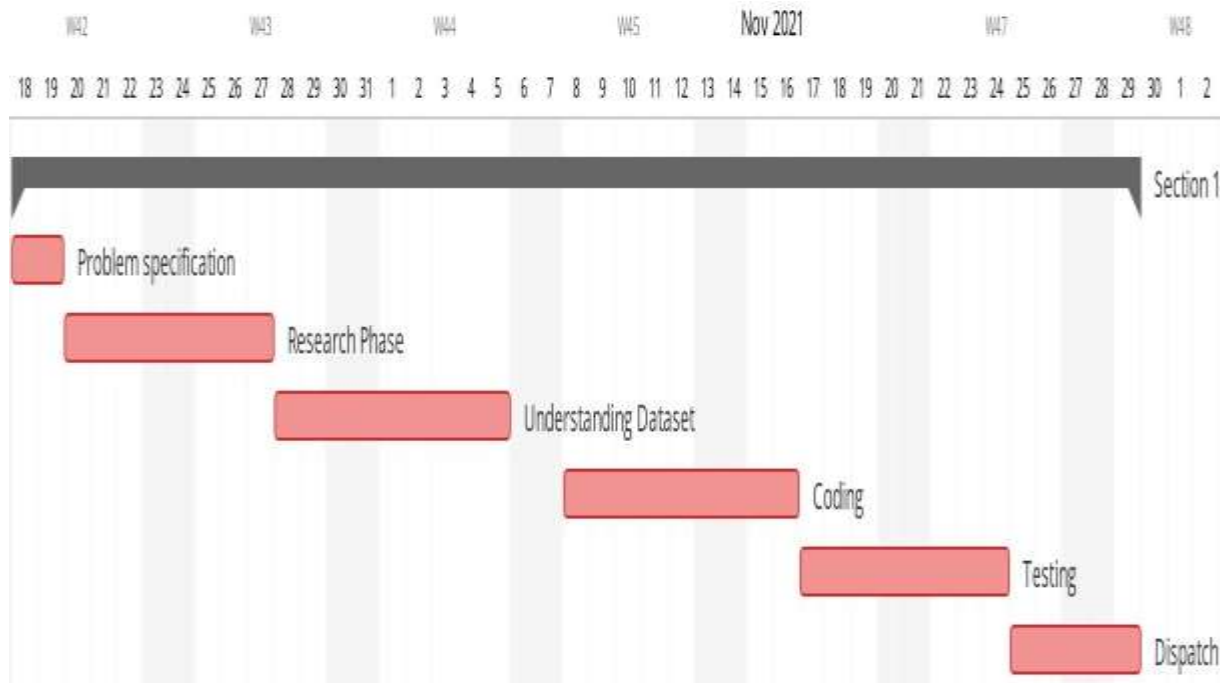


Fig.2. Planning and Scheduling.

Requirement Analysis

Demand Analysis is the first and utmost pivotal phase of software development in effectively developing any design. I started to list out all the functionalities that my operation should give.

- **FUNCTIONAL REQUIREMENTS**

1. Help the customer to see the output faster.
2. Access the dashboard anytime.
3. Correct data is shown.
4. Python-Supported Operating System.

- **NON- FUNCTIONAL REQUIREMENTS**

1. Non-functional requirements are not directly related to the functional behavior of the system.
2. Software must be simple, interactive & user friendly.
3. The user interface is designed in such a way that novice users with little knowledge of the software, should be able to access this application.
4. The system should be giving desired information as output in varied formats.
5. The system should perform analysis, importing and displaying the desired result.

Proposed Techniques or Methods to be Implemented

In programming, a module is a piece of software with specific functionality.

Modules in Python are simply Python lines with a .py extension. The name of the module will be the name of the train. A Python module can have a set of functions, classes, or variables defined and enforced. The colorful Python module used in

The various Python module used are:

NumPy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with an extensive collection of high-level mathematical functions to operate on these arrays.

Pandas

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

Matplotlib

Matplotlib is a conniving library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-acquainted APL for bedding plots into operations using general-purpose GUI Toolkits

Seaborn

Seaborn is a library for making statistical plates in Python. It builds on top of matplotlib and integers with the panda's data structure. It provides a high-position interface for drawing seductive and constructive visualization.

Folium

Folium is a important Python library that helps you produce several types of Leaflet maps. By dereliction, Folium creates a chart in a separate HTML train. Since Folium results are interactive, this library is veritably useful for dashboard structure. You can also produce inline Jupyter maps in Folium.

Data Preprocessing

Preprocessing is a data mining technique used to transform raw data into a useful andefficient format.

Steps involved in Data Preprocessing.

Data Cleaning –

Missing data This situation arises when some data is missing in data. The data can have numerous in applicable data missing corridor. To handle this part, data cleaning is done. Ignore the tuples.

Fill in the missing values.

Data Optimization

Data optimization refers to collecting company data and managing it efficiently to maximize the speed and effectiveness of extracting, analyzing, and utilizing criticalinformation.

An optimization problem consists of maximizing or minimizing an objective function by systematically choosing input values from an allowed set and computing the value of the function. It helps find the best solution to a problem.

Tentative Project Flowchart

The data flow diagram shows a system's information transfers and process steps. A general concept is an approach to depicting how input occurs in a system, different processes, and what runs out. DFD aims to accomplish understanding between developers and users. Data flow diagrams are maintained with other methods of structured system analysis. A DFD illustrates how a system processes data in terms of input and output.

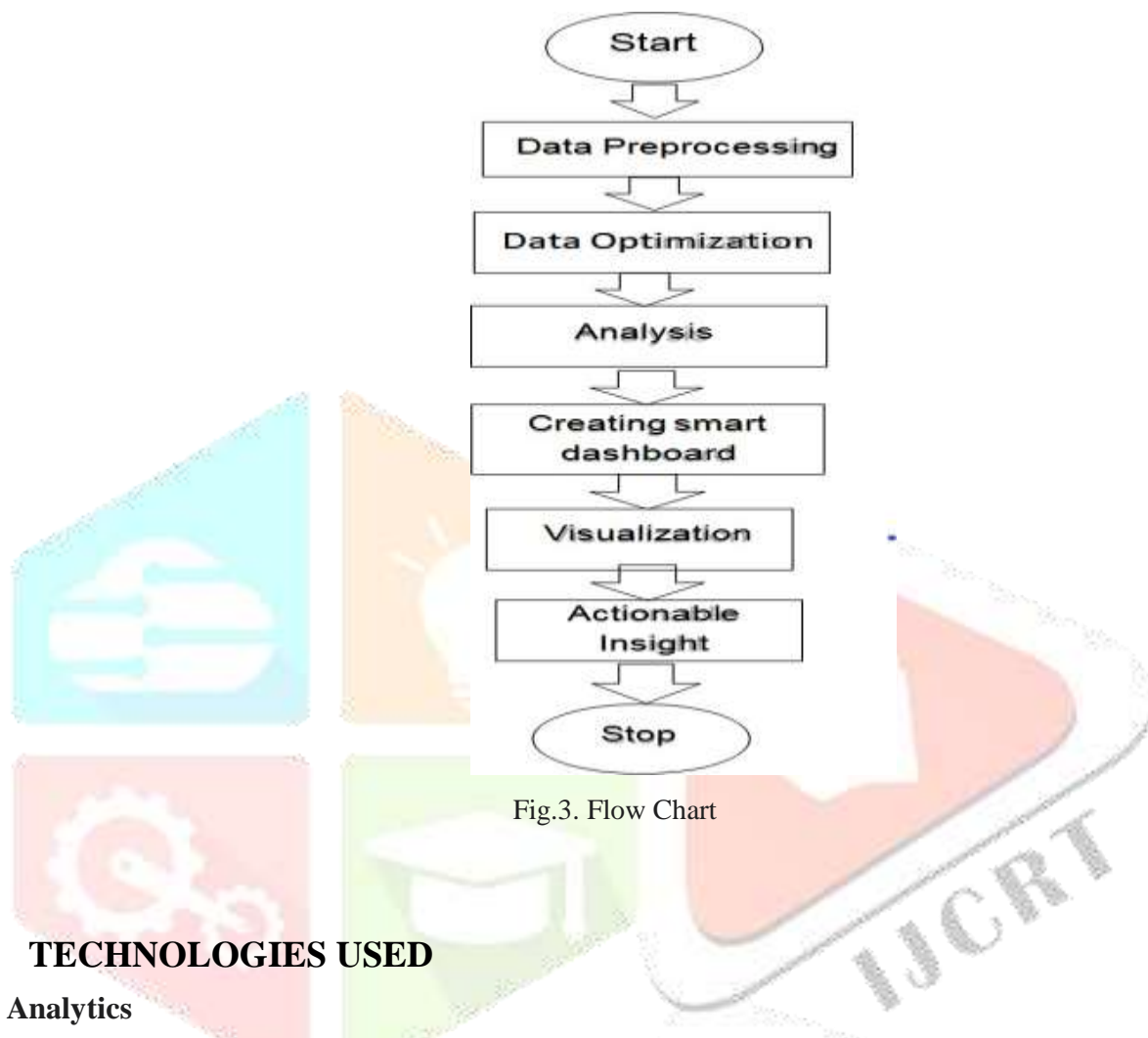


Fig.3. Flow Chart

IV. TECHNOLOGIES USED

Data Analytics

Data analytics converts raw data into practicable. It includes a range of tools, technologies, and processes used to find trends and break problems by using data. Data analytics can shape business processes, ameliorate decision- timber, and foster business growth.

Data analytics helps companies gain further visibility and a deeper understanding of their processes and services. It gives them detailed perceptivity into the client experience and client problems. By shifting the paradigm beyond data to connect perceptivity with action, companies can produce substantiated client gests, make affiliated digital products, optimize operations, and increase hand productivity.

Python

Python has come one of the most popular programming languages in the world in recent times. It's used in everything from machine literacy to erecting websites and software testing. It can be used by inventors andnon-developers likewise Python, one of the most popular programming languages in the world, has created everything from Netflix's recommendation algorithm to the software that controls tone- driving buses.

Python is a general- purpose language, which means it's designed to be used in a range of operations, including data wisdom, software and web development, robotization, and generally getting stuff done.

Jupyter Notebook

The IPython Tablet is now known as the Jupyter Tablet. It's an interactive computational terrain where you can combine law prosecution, rich textbook, mathematics, plots, and rich media. For further details on the Jupyter Tablet, please see the Jupyter website.

V. Result

We provided the user with two options in the first option user will enter the source station from the given options and the destination from the given options. Then from the source station to the destination station all the available trains will show with all details, then the user will select one train and enter the train number in the next field after that user will see all the details of that train. By seeing all this information user can plan their tour by themselves.

In the second option user will provide four options silver, gold, platinum, and luxury packages user have to select any one package after selecting of package user will show their route on the map and also the details information about that state and city. And in gold, platinum, the user can eliminate the city he doesn't want to visit.

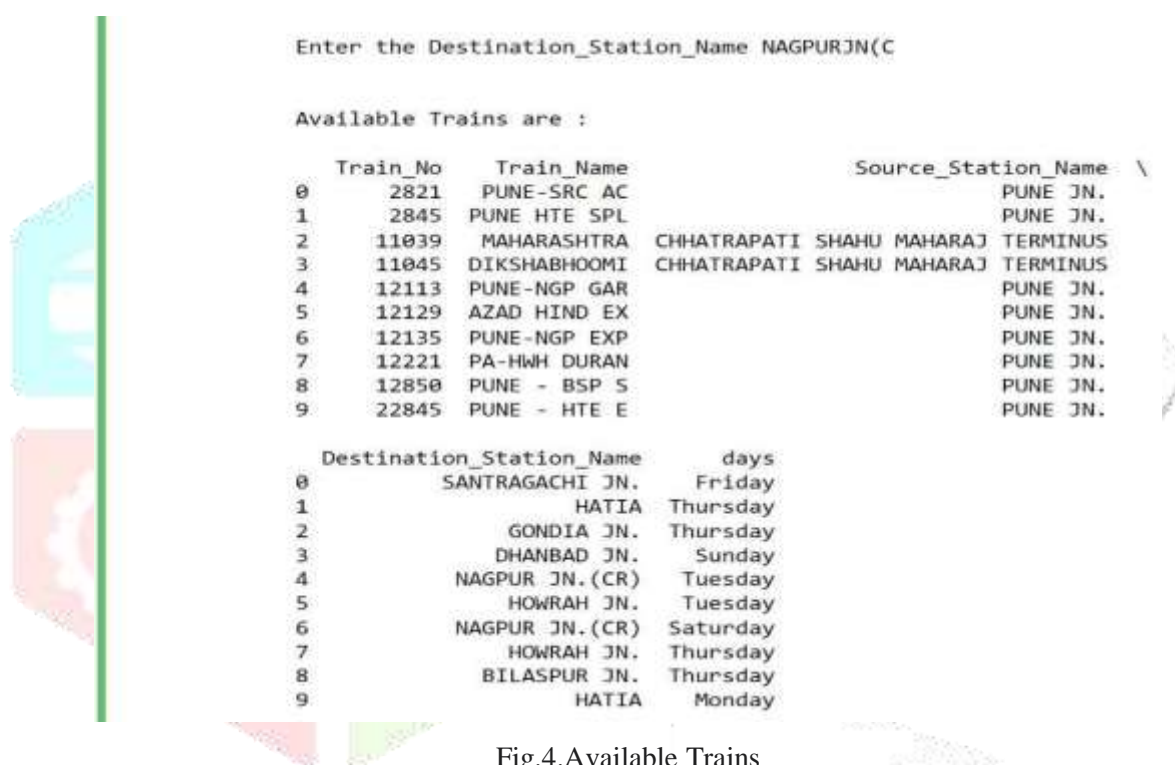


Fig.4.Available Trains

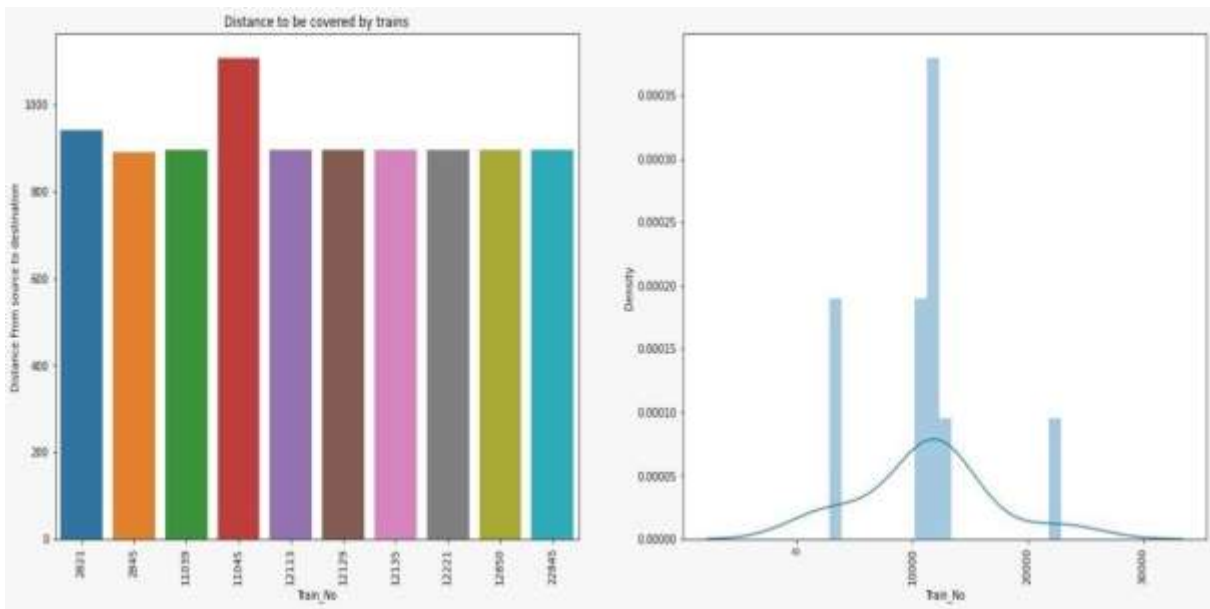


Fig.5.Available Trains in Graphical Format



Fig.6.Number of Trains Available

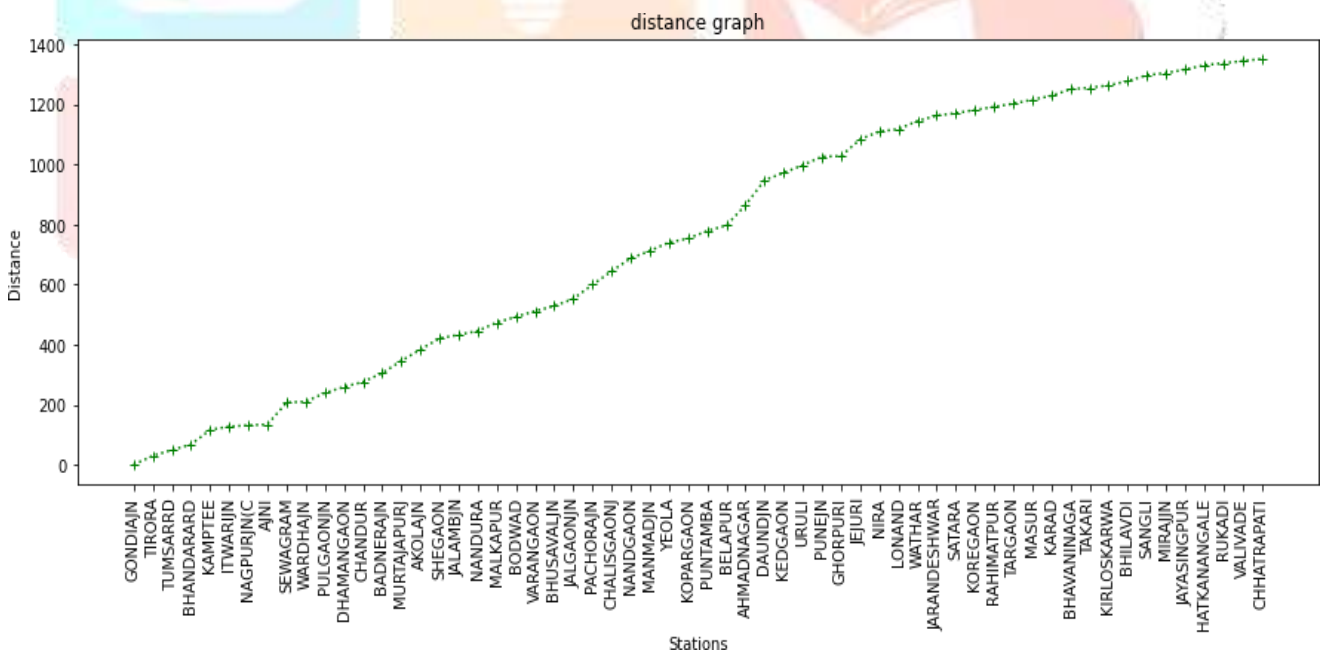
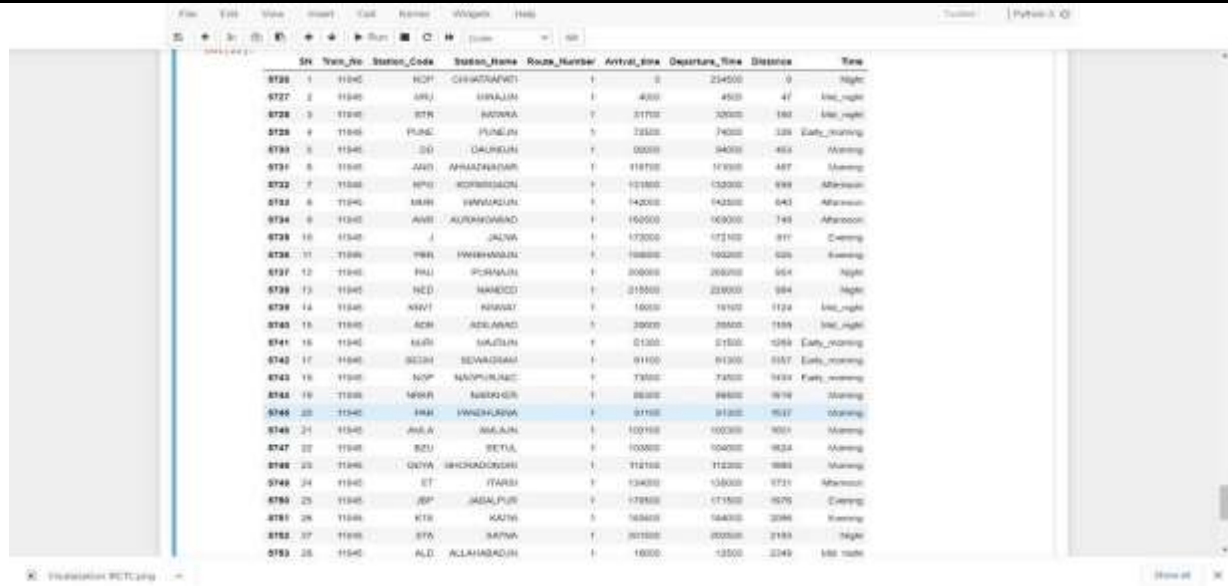


Fig.7.Distance Graph



| SN | Train_No | Station_Code | Station_Name | Route_Number | Arrival_Time | Departure_Time | Distance | Time | |
|------|----------|--------------|--------------|--------------|--------------|----------------|----------|------|---------------|
| 8720 | 1 | 1194C | NCP | CHANDRANWADI | 1 | 0 | 234500 | 9 | Night |
| 8727 | 2 | 1194C | SRJ | SIRAJULI | 1 | 400 | 450 | 47 | Mid_night |
| 8728 | 3 | 1194C | STR | SATARA | 1 | 3170 | 3050 | 180 | Mid_night |
| 8729 | 4 | 1194C | PUNE | PUNEJN | 5 | 2250 | 7400 | 5150 | Early_morning |
| 8730 | 5 | 1194C | DD | DAUND | 7 | 0500 | 9400 | 450 | Morning |
| 8731 | 6 | 1194C | JAD | AKHADKAR | 1 | 11870 | 17300 | 487 | Morning |
| 8732 | 7 | 1194C | SPN | SOMNATH | 1 | 12300 | 13200 | 900 | Midnoon |
| 8733 | 8 | 1194C | SRH | SARHAD | 1 | 14200 | 14200 | 640 | Afternoon |
| 8734 | 9 | 1194C | ANR | AURHAWAD | 1 | 16000 | 16000 | 740 | Afternoon |
| 8735 | 10 | 1194C | J | JALNA | 1 | 17300 | 17310 | 811 | Evening |
| 8736 | 11 | 1194C | PHS | PHULE | 1 | 18800 | 18200 | 620 | Evening |
| 8737 | 12 | 1194C | PHI | PHATA | 1 | 20000 | 20000 | 664 | Night |
| 8738 | 13 | 1194C | NED | NANDED | 1 | 21000 | 22000 | 984 | Night |
| 8739 | 14 | 1194C | NAV | NAVAR | 1 | 1600 | 1910 | 1124 | Mid_night |
| 8740 | 15 | 1194C | ADR | ADAR | 5 | 2000 | 2000 | 1168 | Mid_night |
| 8741 | 16 | 1194C | NDR | NANDED | 1 | 0130 | 0130 | 1268 | Early_morning |
| 8742 | 17 | 1194C | SDR | SARHAD | 1 | 9110 | 9130 | 1317 | Early_morning |
| 8743 | 18 | 1194C | NCP | NANDED | 1 | 7300 | 7200 | 1634 | Early_morning |
| 8744 | 19 | 1194C | NDR | NANDED | 1 | 8800 | 8800 | 1678 | Morning |
| 8745 | 20 | 1194C | PHR | PHULE | 1 | 9110 | 9100 | 1617 | Morning |
| 8746 | 21 | 1194C | AKA | AKHADKAR | 1 | 10010 | 10030 | 1601 | Morning |
| 8747 | 22 | 1194C | BDU | BETUL | 1 | 10000 | 10000 | 1624 | Morning |
| 8748 | 23 | 1194C | GUW | GURUDWAR | 1 | 11210 | 11220 | 1681 | Morning |
| 8749 | 24 | 1194C | ET | ETAH | 1 | 13400 | 13400 | 1721 | Midnoon |
| 8750 | 25 | 1194C | JBP | JALAPUR | 1 | 17810 | 17150 | 1676 | Evening |
| 8751 | 26 | 1194C | KTR | KATRA | 5 | 16400 | 16400 | 2046 | Morning |
| 8752 | 27 | 1194C | STN | SATNA | 7 | 20300 | 20300 | 2188 | Night |
| 8753 | 28 | 1194C | ALD | ALLAHABAD | 1 | 1800 | 1200 | 2249 | Mid_night |

Fig.8.Train Detail

VI. CONCLUSION

The "Analytical Backend Dashboard for India Tour Route Planning Using IRCTC Data" project was completed successfully in a clear-cut and orderly fashion. This dashboard makes it simple for users to choose the ideal train based on their preferred time and distance. Additionally, graphical depictions are crucial for quickly comparing trains and choosing the one we want. Thanks to this added feature, our dashboard is now fairly clever and user-friendly for every passenger.

VII. REFERENCES

- [1] Apathy Abhinay, T. Y. (Volume 13, Number 12 (2018) pp). Passenger Flow Analysis in Metro Network Using Machine Learning fashion. International Journal of Applied Engineering Research, 10604-10606
- [2] Rasika Ingle, M. K. (Volume 3, Issue 3, March, 2013). An Approach for Effective Use of Pattern Discovery for Detection of Fraudulent Patterns in Railway Reservation Dataset. International Journal of Computational Engineering Research, 26-29.
- [3] Prof.Ashish Saxena, M. N. (Volume 5, Issue 03, March - 2018). Indian Railway System Monitoring Regards Passenger Allocation and Reservation. International Journal of Advance Engineering and Research, 608-611.
- [4] Prof.Ashish Saxena, M. N. (Volume 5, Issue 03, March - 2018). Indian Railway System Monitoring Regards Passenger Allocation and Reservation. International Journal of Advance Engineering and Research, 608-611.
- [5] Khan, J. A. (Volume 8, No. 5, May June 2017). Waiting for Ticket Optimization using Reservation Chart Cluster for Indian Railway. International Journal of Advanced Research in Computer Science, 913-916.
- [6] Palak Baid, Apoorva Gupta, Neelam Chaplot. Sentiment Analysis of Movie Reviews using Machine Learning Techniques.
- [7] Satyakrishna, J., Sagar, R. K., & Tech, M. (2018). Train Delay Prediction Systems Using Big Data Analytics. 6(3), 7.

- [8] Hansen, I.A.: 'Station capacity and stability of train operations'. Computers in Railways VII, 2000, pp. 809–816
- [9] Goverde, R.M.P., Odijk, M.A.: 'Performance evaluation of network timetables using PETER'. Computers in Railways VIII, 2002, pp. 731–740
- [10] Goverde, R.M.P.: 'Railway timetable stability analysis using max-plus system theory', Transp. Res. B, Methodol., 2007, 41, (2), pp. 179–201
- [11] Delorme, X., Gandibleux, X., Rodriguez, J.: 'Stability evaluation of a railway timetable at station level', Eur. J. Oper. Res., 2009, 195, (3), pp. 780–790
- [12] Cacchiani, V., Toth, P.: 'Nominal and Robust train timetabling problems', Eur. J. Oper. Res., 2012, 219, (3), pp. 727–737
- [13] Meng, L., Goverde, R.M.P.: 'A method for constructing train delay propagation process by mining train record data', J. Beijing Jiaotong Univ. Nat. Sci., 2012, 36, (6), pp. 15–20
- [14] Li, D., Daamen, W., Goverde, R.M.P.: 'Estimation of train dwell time at short stops based on track occupation event data: a study at a Dutch railway station', J. Adv. Transp., 2016, 50, (3), pp. 1–20
- [15] Muller, T.H.J., Furth, P.G.: 'Trip time analyzers: key to transit service quality'. Transp. Res. Rec., 2001, 1760, pp. 10–19
- [16] Kimpel, T.J., Strathman, J.G., Bertini, R.L., et al.: 'Analysis of transit signal priority using archived TriMet bus dispatch system data', Transp. Res. Rec., 2005, 1925, pp. 156–166
- [17] Furth, P.G., Hemily, B.J., Muller, T.H.J., et al.: 'Using archived AVL-APC data to improve transit performance and management'. Transp. Res. Rec. TCRP Report 113, 2006
- [18] Elgeneidy, A., Horning, J., Krizek, K.J.: 'Using archived ITS data to improve transit performance and management'. Minnesota Department of Transportation Research Services Section, Report 2007-44, 2007