ASHOK LEYLAND

INTERNSHIP REPORT

Presented by: Rishiram B

OVERVIEW

01
About Company

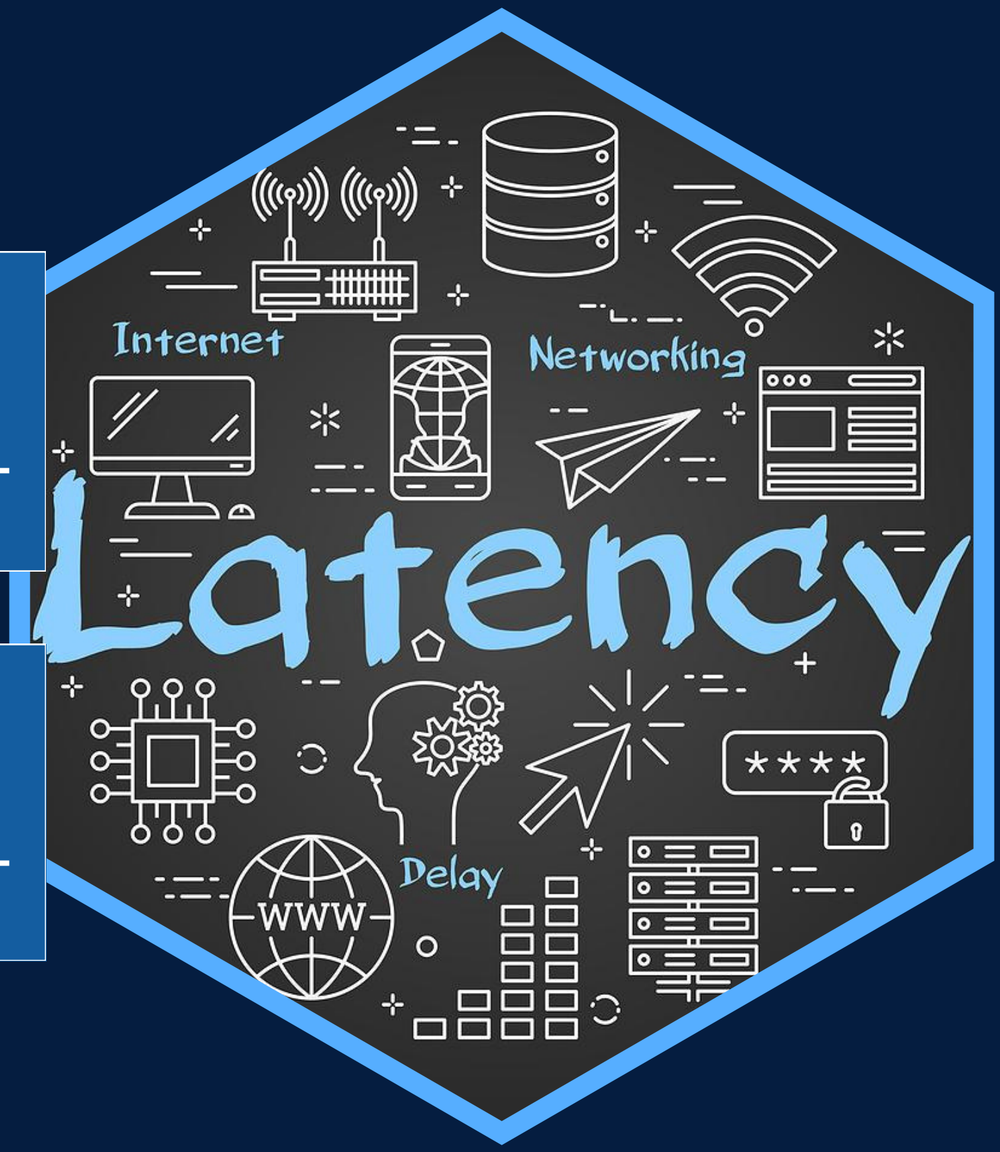
02
About My Project

03
Data Analysis

04
Data Visualization

05
Implementation

06
Conclusion



ABOUT COMPANY

COMPANY OVERVIEW

- Name: Ashok Leyland Limited
- Industry: Automotive Manufacturing and Technology
- Founded: 1948
- Location: Chennai, Tamil Nadu, India

Ashok Leyland, flagship of the Hinduja group, is the 2nd largest manufacturer of commercial vehicles in India, the 4th largest manufacturer of buses in the world, and 19th largest manufacturers of trucks.

CORE BUSINESS AREAS

Commercial Vehicles Manufacturing

Ashok Leyland is a major player in the manufacturing of commercial vehicles, including trucks, buses, and light commercial vehicles. The company produces a wide range of vehicles for various applications such as logistics, construction, mining, and public transportation.

Electric and Hybrid Vehicles

In response to the growing emphasis on sustainable and eco-friendly transportation, Ashok Leyland has been involved in the development and manufacturing of electric and hybrid vehicles. This aligns with global trends towards cleaner and greener mobility solutions.

Defense and Specialized Vehicles

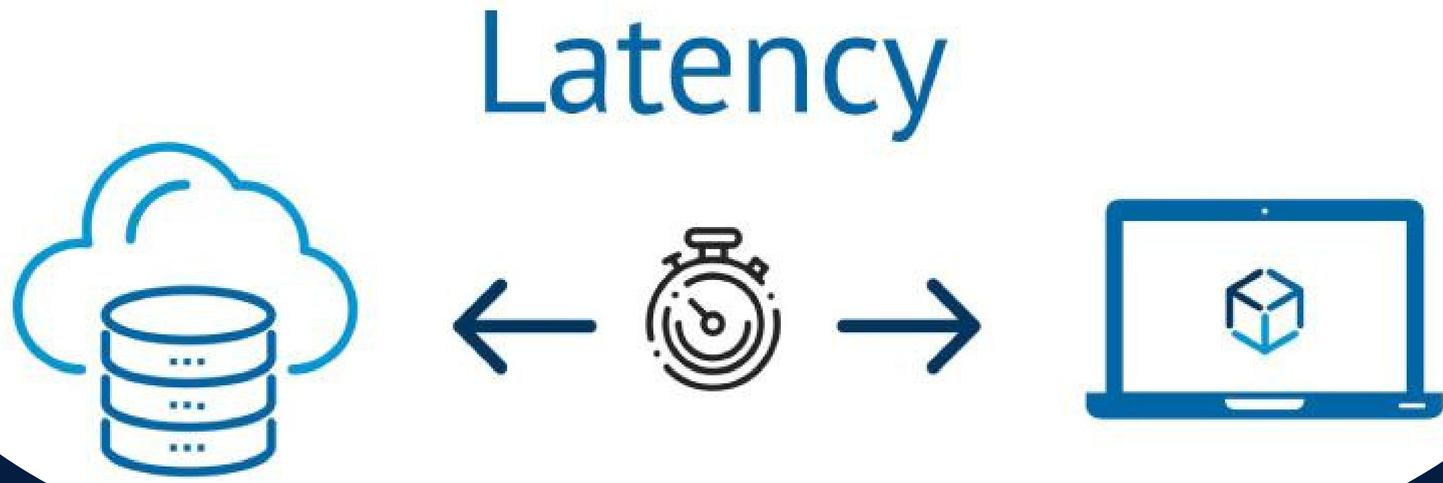
Ashok Leyland is involved in the production of defense and specialized vehicles, including those used by the armed forces. This diversification allows the company to cater to specific requirements in the defense and security sectors.

02

MY PROJECT

BIG DATA LATENCY ANALYSER

WHAT IS LATENCY ?



Latency quantifies the time interval between the initiation of an action and the reception of its response. In computing, it denotes the duration for data to traverse from its origin to its destination. Maintaining low latency is essential for responsive systems, like online gaming or real-time communication, ensuring prompt and effective data processing.

INTRODUCTION

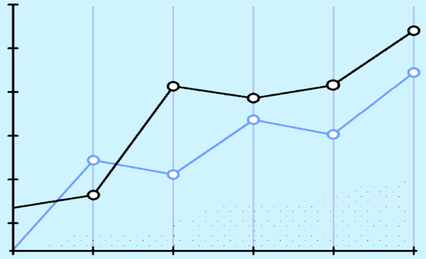
This project provides an in-depth Analysis of the data about the vehicle information stored in the database of the company. The project involved utilizing Python and its libraries like numpy,pandas,sklearn,plotly,matplotlib and Tableau to analyze and visualize the data given. The report covers the project background, objectives, requirements, development process, key features, challenges faced, and future recommendations.

CONCEPT

The concept involves Analyzing and Visualizing the data and should be user-friendly app interface that allows users to conveniently visualize the data . The Analysis provides the details about the cause of the problem and possible solutions that can be made.

OBJECTIVE :

DATA
ANALYSIS

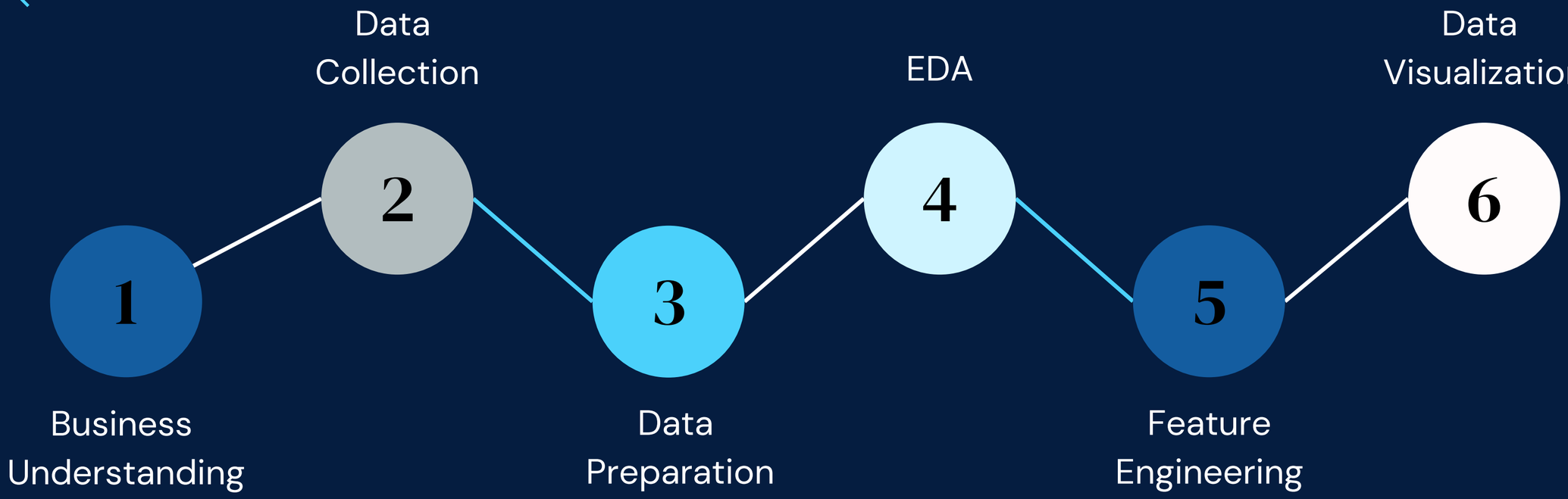


DATA
VISUALIZATION

03

DATA ANALYSIS

PROJECT TIMELINE



BUISNESS UNDERSTANDING

Ashok Leyland limited have been exporting various kind of vehicles, and to track large vehicles with help of the ECU (Electronic Control Unit) from the 3rd party Vendor. Each ECU has its obu id (on-board unit identifier) which has Different kinds of service provider. By the help of the ECU, we can calculate the various parameters like wheel speed, fuel consumption, location, etc. the data collected by the ECU will be published to the vendor database and from the vendor database data is sent to the Company Database.

DATA COLLECTION

Data collection in data science involves gathering relevant information to meet project goals. It includes defining objectives, choosing data sources, and ensuring ethical access. Rigorous quality checks and continuous monitoring are essential for reliable insights.

For this project , we have collected three excels with data of 10 , 100 and their service providers details and with this data we have forwarded for the next step.

DATA PREPERATION

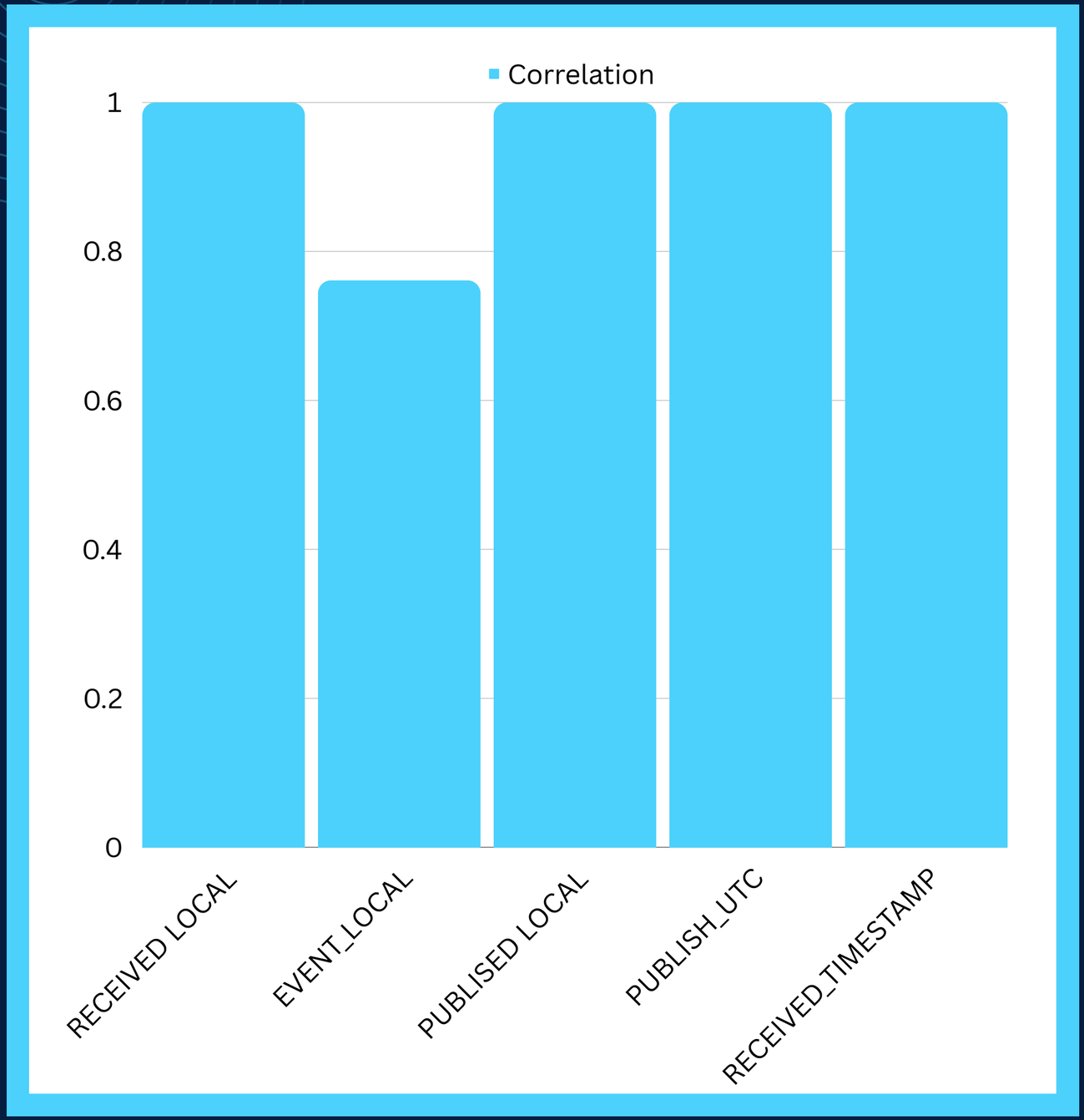
From the given Data, the datasets are cleaned, unnecessary features removed and made more structured, missing values are dealt with, redundancy is eliminated, and preliminary tests are done with the data in order to evaluate the direction of the project. Few important steps are to change the appropriate Dtype and removing the unecessary columns from the dataframe .

EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis (EDA) is a crucial step in data analysis where we explore and understand our dataset. Through visualizations and statistical summaries, EDA helps uncover patterns, trends, and potential outliers in the data. By identifying relationships and distributions, it guides subsequent analyses and informs data-driven decisions. EDA is fundamental for gaining insights, formulating hypotheses, and ensuring a solid foundation for more advanced modeling and interpretation in the data science process.

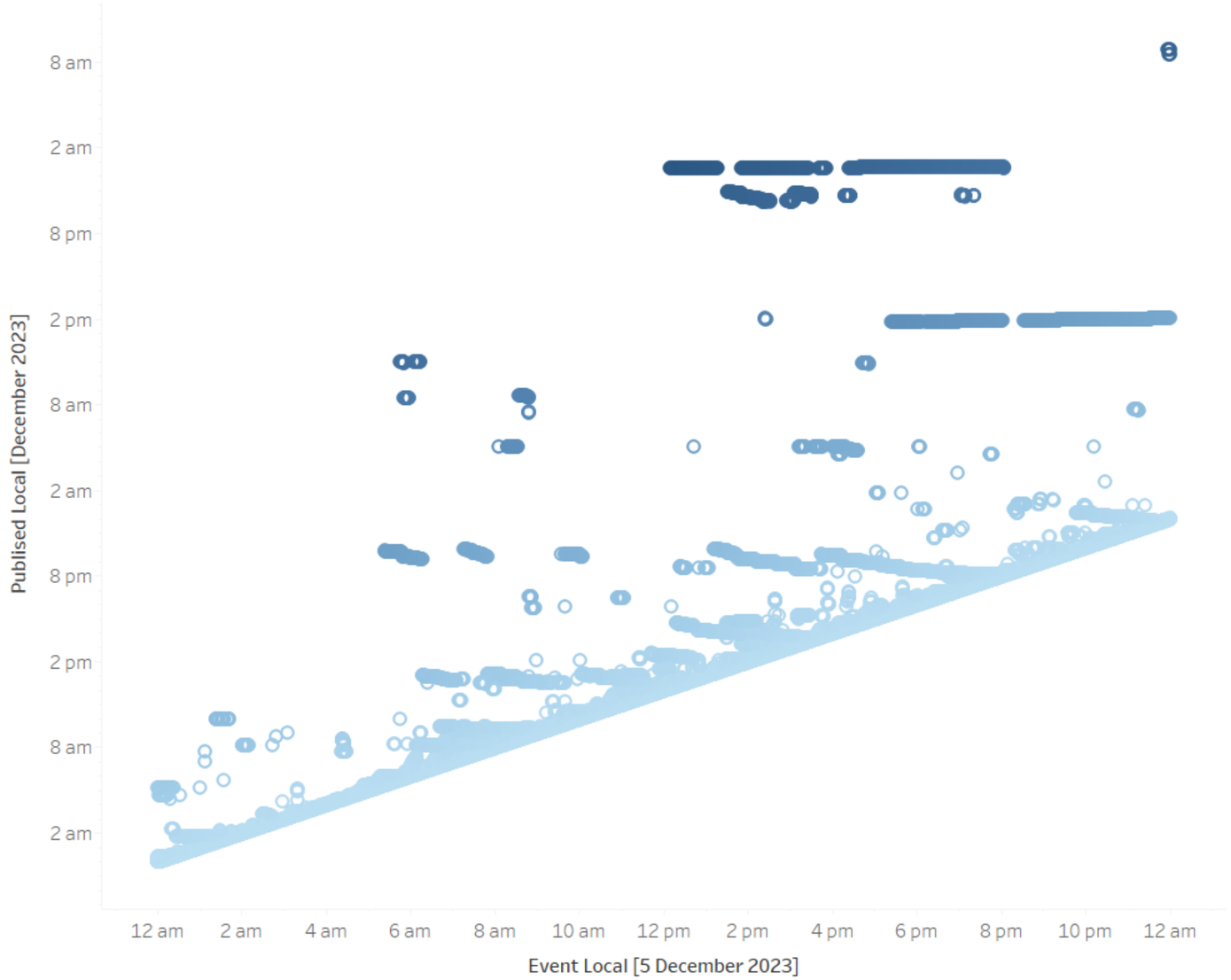
CORRELATION

There is a deviation between event local against the received local showing that there should be some outliers which can be the reason for the latency problem we are dealing with .



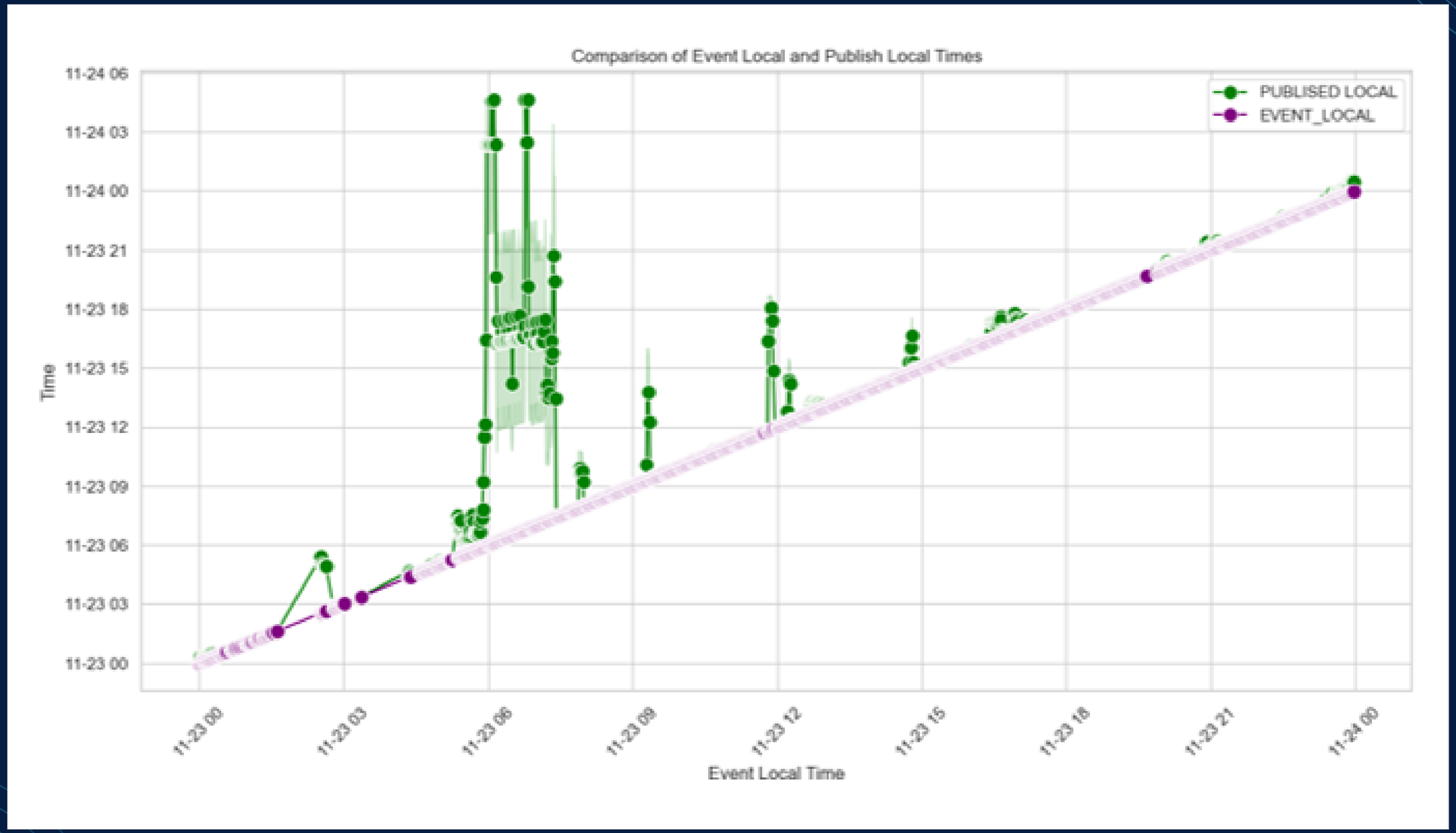
LINEAR REGRESSION

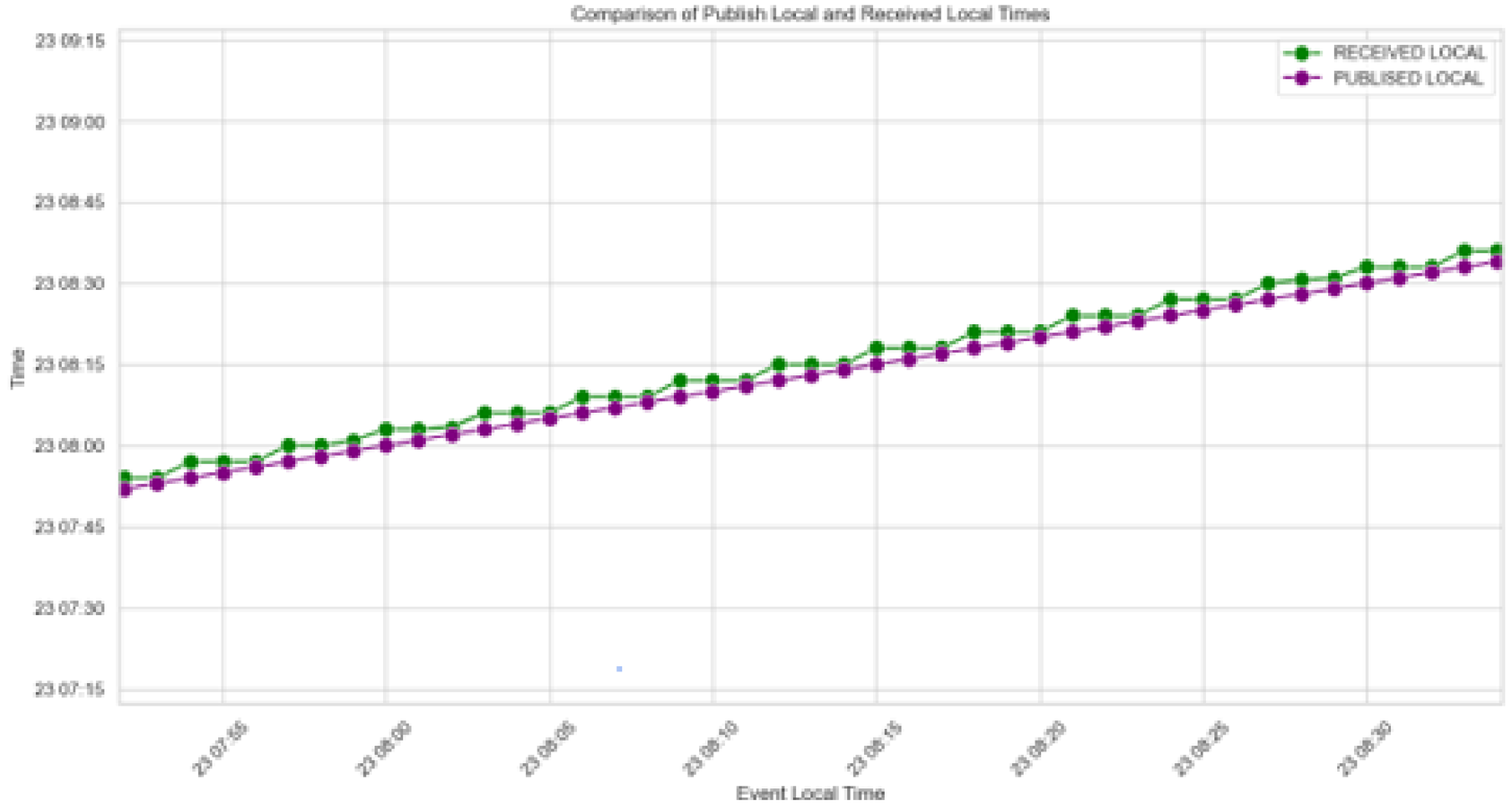
To examine the outliers we have made a graph to plot out the outliers and in this linear regression graph. these outliers are the reason for latency problem caused .



EVENT VS PUBLISH / PUBLISH VS RECEIVED :

Graph to understand the relationship between the local storing time and vendor database receiving time as well as the relationship between the vendor database time and the company database . There is always a 1 - 2 min latency from vendor to company database but the big latency have been occurred due to delay between the local storing and the vendor database.


Comparison of Event Local and Publish Local Times
PUBLISED LOCAL
EVENT_LOCAL
Time
11-24 06
11-24 03
11-24 00
11-23 21
11-23 18
11-23 15
11-23 12
11-23 09
11-23 06
11-23 03
11-23 00
11-23 00
11-23 03
11-23 06
11-23 09
11-23 12
11-23 15
11-23 18
11-23 21
11-24 00
Event Local Time

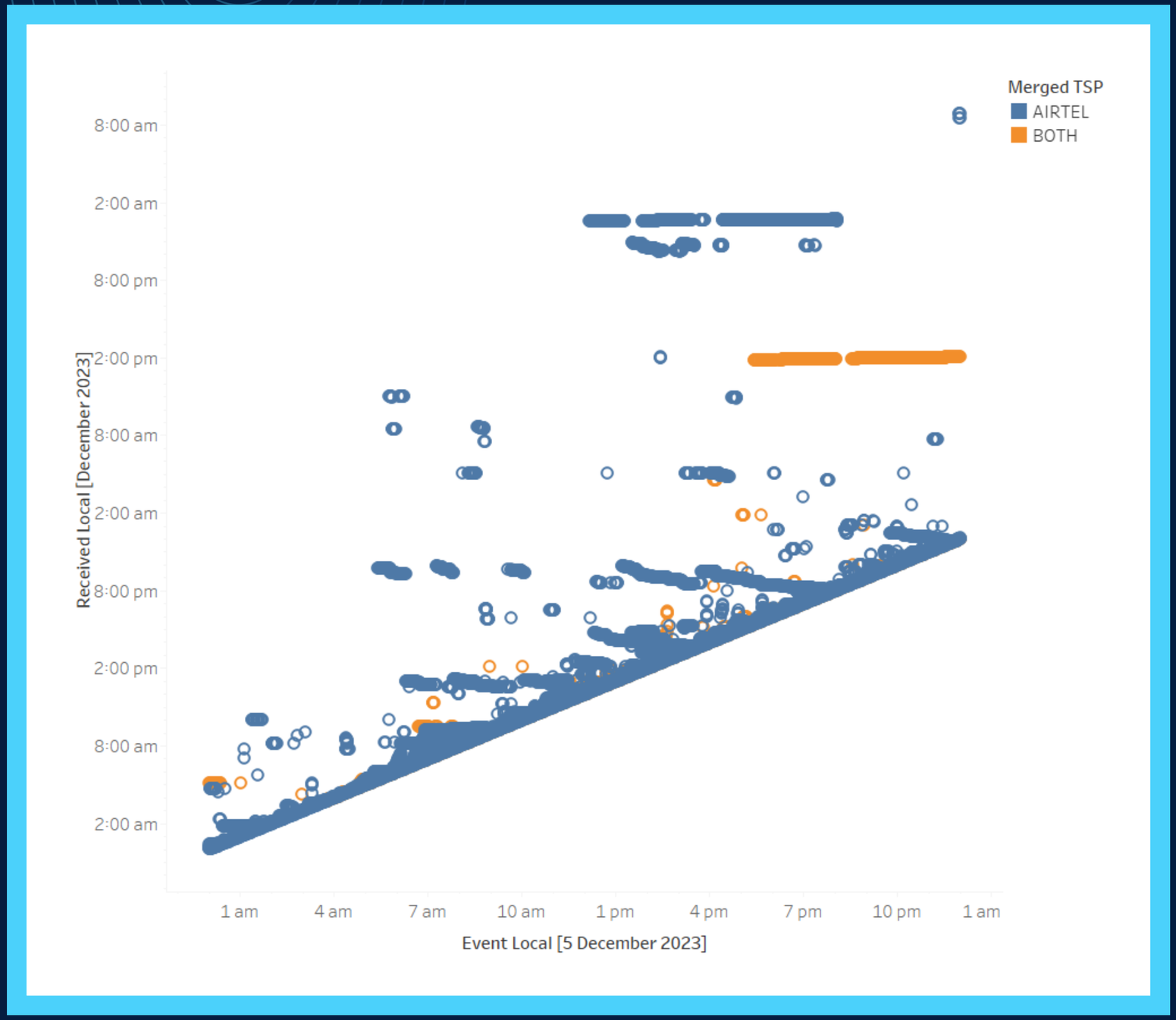

Comparison of Publish Local and Received Local Times
RECEIVED LOCAL
PUBLISED LOCAL
Time
23 09:15
23 09:00
23 08:45
23 08:30
23 08:15
23 08:00
23 07:45
23 07:30
23 07:15
23 07:55
23 08:00
23 08:05
23 08:10
23 08:15
23 08:20
23 08:25
23 08:30
Event Local Time

FEATURE ENGINEERING

Feature engineering enhances raw data in data science for improved insights. Techniques include handling missing values, encoding variables, scaling features, creating interaction terms, and incorporating domain-specific features. The goal is to extract meaningful information and optimize data representation through iterative experimentation with various transformations.

SERVICE PROVIDERS

Company using two service providers one as the primary TSP and a Fallback TSP. some vehicles have both the Service Providers Active and most of them have only one Service provider Active. based on them this graph has been plotted.

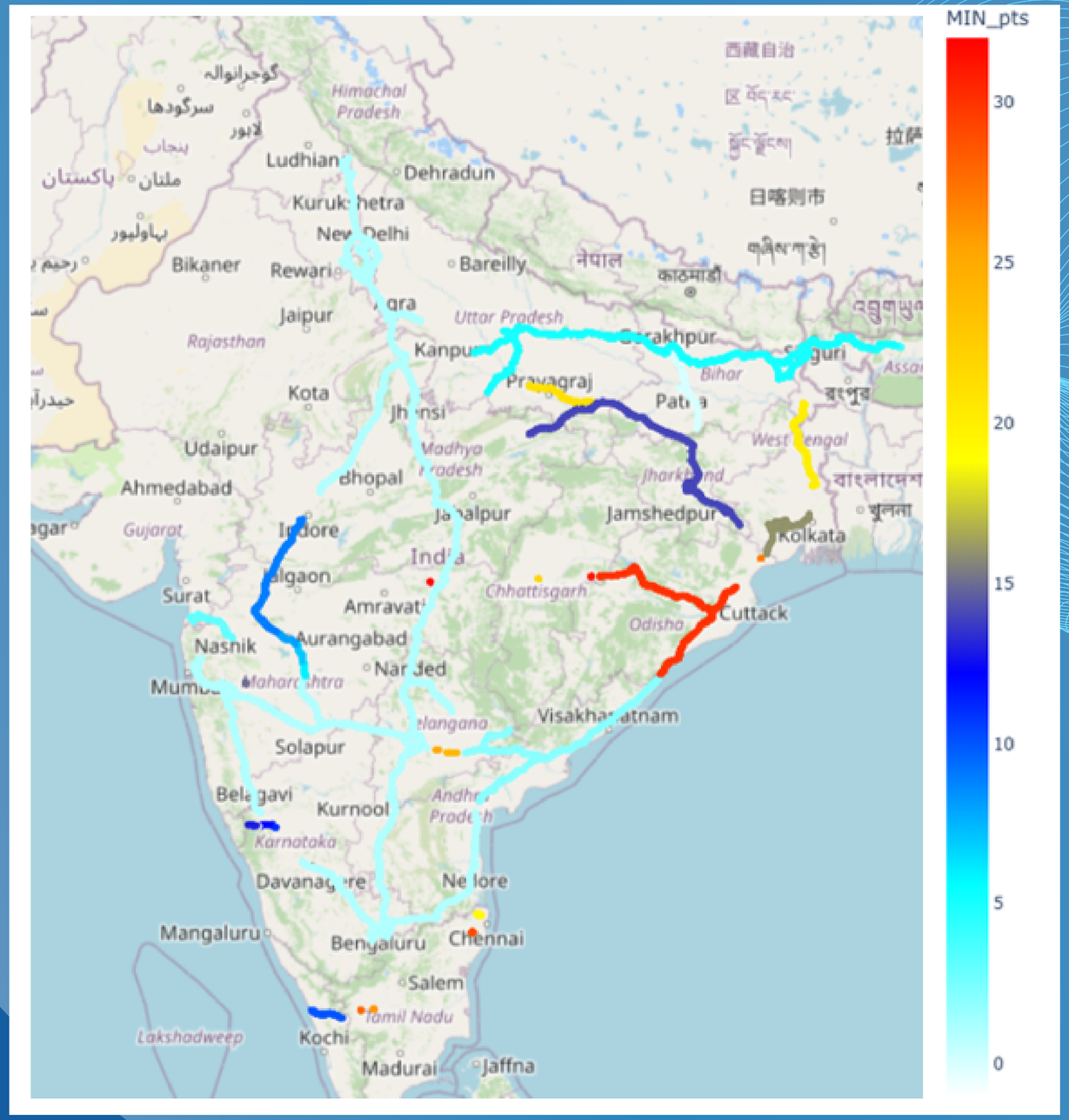


04

DATA VISUALIZATION

CLUSTERING

With help of DBSCAN methodology , clustering nearby signal(nodes) within 0.5 km radius of a sample node , we have clustered them . min_pts here shows the nearby other clusters available nearby to the selected clusters . plotted them in a map using the longitude and latitude from the given data.

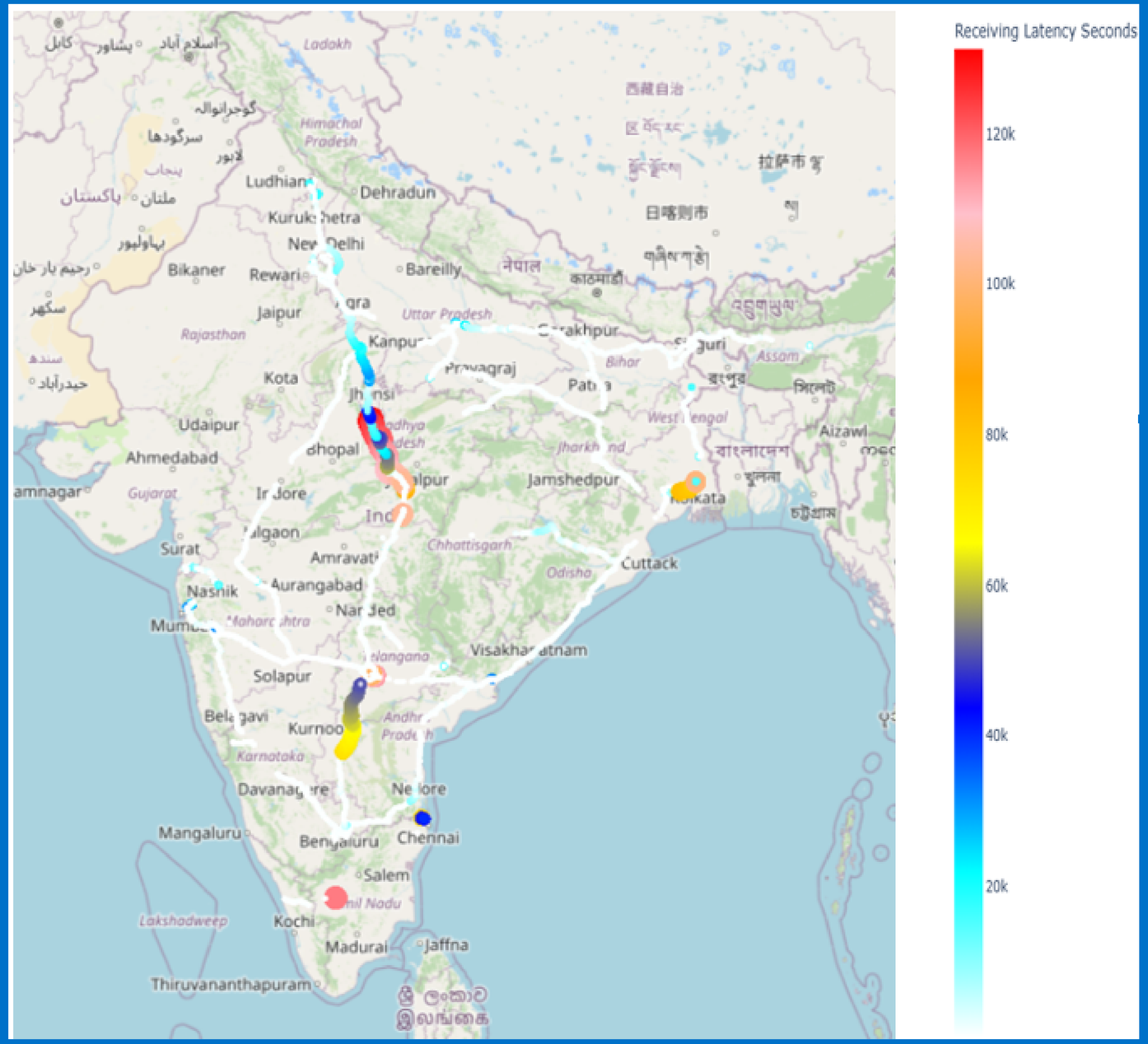


RED ZONE

For some specific range irrespective of the service providers or ECU damage or any hardware issues there is always a high latency and sometimes the signals gets interrupted and has no signals for some part. The reason behind this is due to the Forest Area where we don't have much access for network to store in the database.

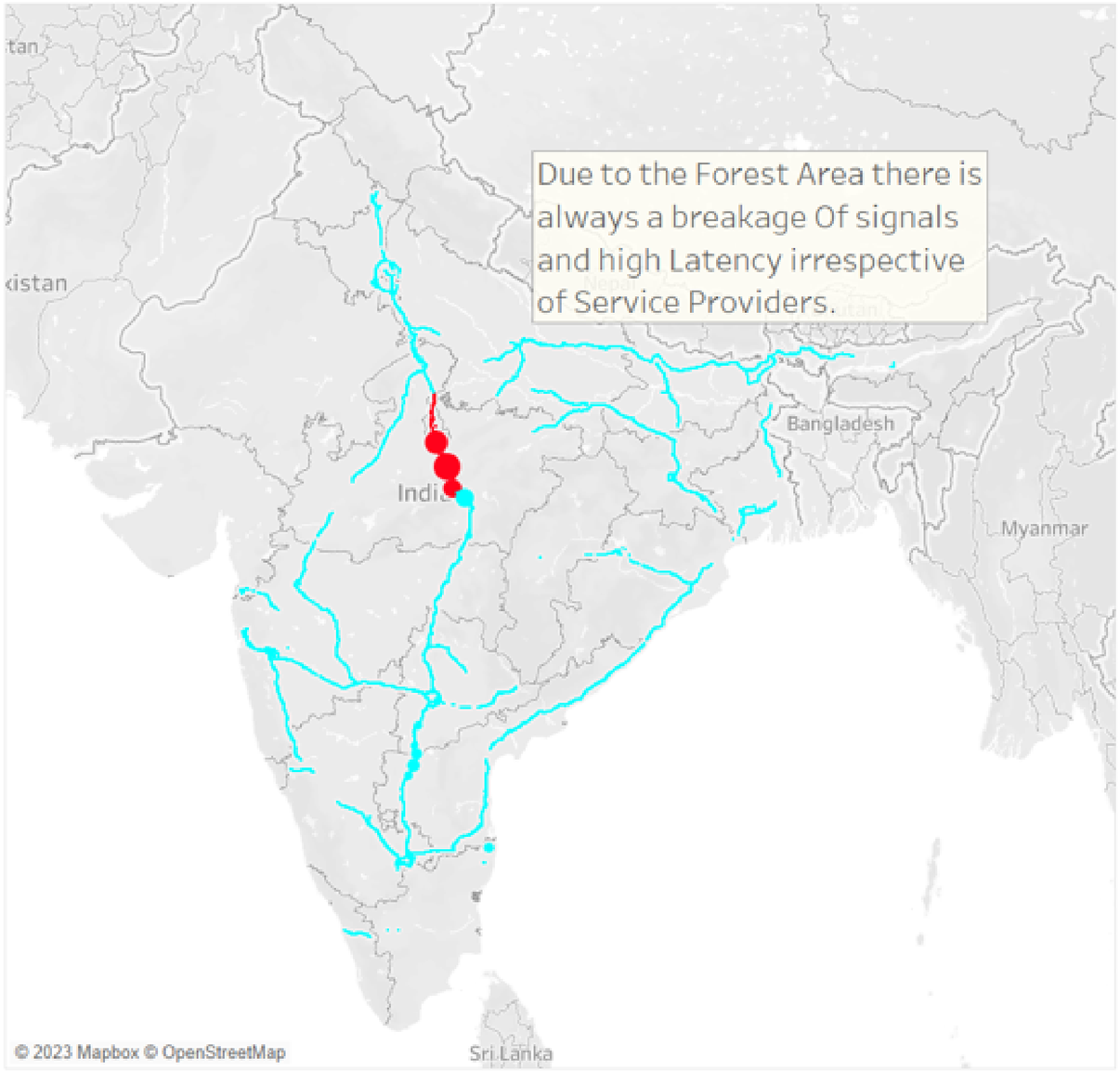
05

IMPLEMENTATION



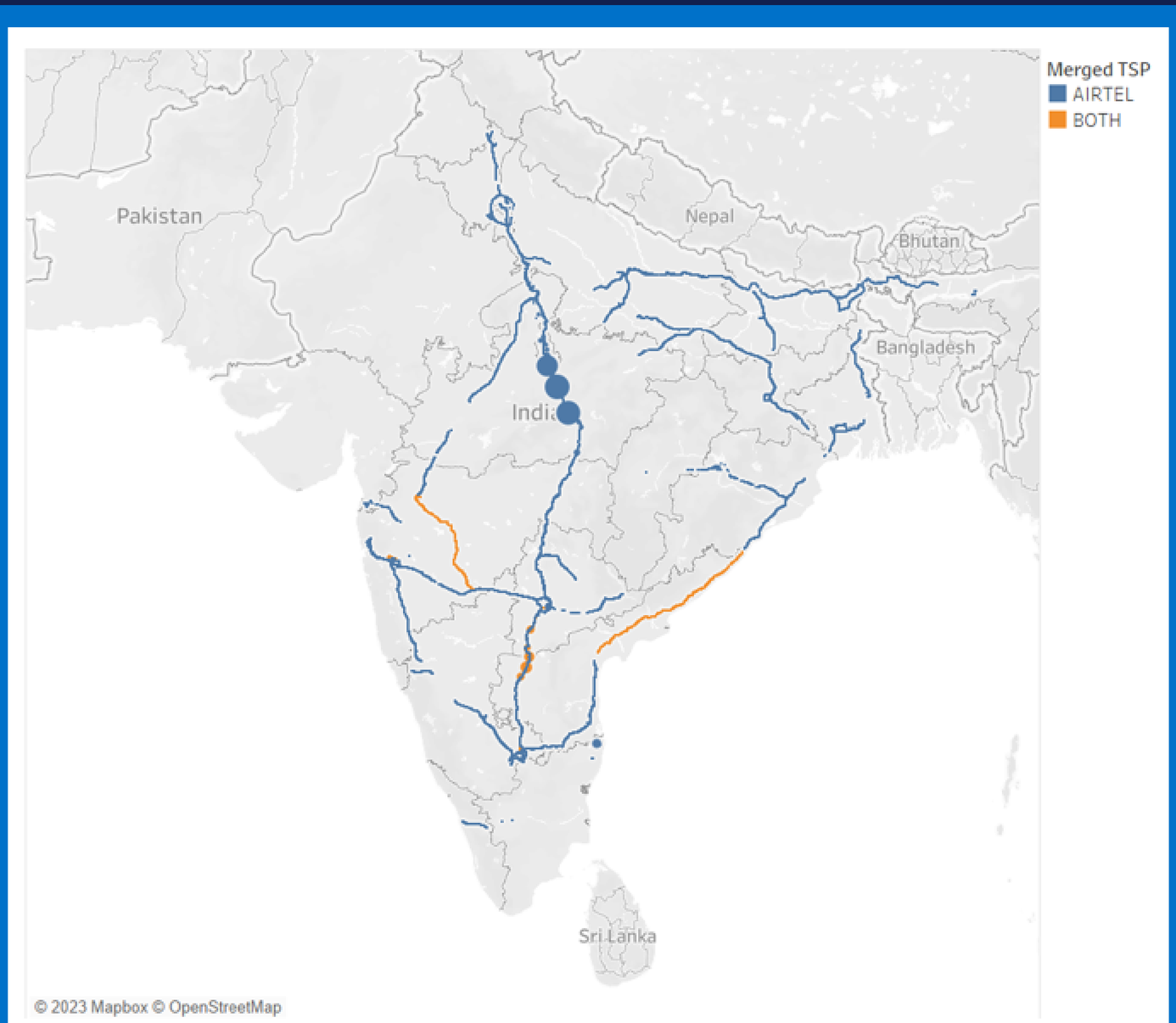
BASED ON LATENCY

After clustering , the size of the nodes are changed based on the latency received in seconds . the higher the delay , the size increases. we can notice at some particular areas there are multiple gaps between the signals received.to understand this we should find the possible redzone areas.



LATENCY BASED ON REDZONE

After clustering , the size of the nodes are changed based on the latency received in seconds . the higher the delay , the size increases. we can notice at some particular areas there are multiple gaps between the signals received.to understand this we should find the possible redzone areas.



LATENCY BASED ON SERVICE PROVIDERS

This map is to differentiate service providers in the map with the latency so that we can observe the latency of each service provider. From our given sample data, most of the vehicle have 1 active profile and majority of the service provider for them is AIRTEL. Ignoring the redzone area, AIRTEL doesn't really have any high latency problem but vehicles with 2 Service providers have high latency and while analyzing the data, the primary TSP is BSNL which means we can assume that compared to BSNL, AIRTEL have reliable network and for further clarification we need more different kinds of sample data.

06

CONCLUSION

high latency problem in the specific range is caused irrespective of the service providers and many other problems, when the Vehicle is in the REDZONE there will be always high latency and in worst case sometimes we may not able to get signals. Other reason can be because of the service provider as we saw most of the vehicle uses AIRTEL and some vehicles have two active profiles in which we can Assume that the BSNL having the high latency comparatively to AIRTEL. some other reasons can be due to ECU hardware Damage, Vendor database Connectivity, Distance between server and the vehicles' current location and can be because of poor configuration of ECU. In conclusion, there is no problem on company database and there are some REDZONES where the latency is always high with some possible difference in latency between service providers in the particular area.

Thank you

Contact

rishiram20757@gmail.com