DATA ANALYTICS WITH COGNOS - PHASE 5

PUBLIC TRANSPORTATION EFFICIENCY ANALYSIS - DOCUMENTATION

Introduction

In Phase 5 of the Public Transport Efficiency Analysis project, we will outline the project's objectives, describe the analysis process, and highlight how the insights generated through data analysis can support transportation improvement initiatives. We will also discuss the development phases and the Design Thinking process that guided our approach.

Project Objectives

The primary objectives of Public Transport Efficiency Analysis project are as follows:

1. Evaluate Public Transportation Efficiency: Assess the efficiency of public transportation services by analysing on time performance, passenger satisfaction, and service efficiency.

2. Collect and Process Data: Dataset is available in Kaggle.

Dataset link: https://www.kaggle.com/datasets/rednivrug/unisys?select=20140711.CSV

Clean and preprocess the data for analysis.

3. Visualize Data Using IBM Cognos: Utilize IBM Cognos to create informative visualizations, dashboards, and reports to present the insights from the analysis.

4. Predict Service Disruptions: Develop models to predict service disruptions and assess the factors contributing to disruptions.

5. Conduct Sentiment Analysis: Analyze passenger feedback to gauge passenger satisfaction and sentiment towards public transportation services.

6. Support Transportation Improvement Initiatives: Provide actionable insights and recommendations to enhance public transportation services, improve resource allocation, and optimize routes.

Design Thinking Process

The Design Thinking process served as the guiding framework for our project. It consists of the following key stages:

1. Empathize: We began by empathizing with public transport user feedbacks to understand their needs and pain points. This involved collecting passenger feedback and understanding their experiences.

2. Define: In the Define stage, we set clear analysis objectives, identified sources of data, and defined the criteria for evaluating public transportation efficiency. This phase established a clear direction for our project.

3. Ideate: During the Ideation phase, we brainstormed various analysis techniques to address the defined objectives. We considered different ways to visualize and analyze the data.

4. Prototype: We developed data collection methods, data cleaning processes, and data visualization techniques as prototypes.

5. Test and Implement: In the Testing and Implementation stage, we collected, processed, and visualized data using IBM Cognos. We implemented predictive models for service disruptions and sentiment analysis. We also tested the effectiveness of these models.

6. Gather Feedback: Continuous feedback from project team members and stakeholders allowed us to make improvements and refinements throughout the project.

Development Phases

The project was divided into several phases:

Phase 1: Problem Definition

- Analysis Objectives: We defined the objectives of analyzing on time performance, passenger satisfaction, and service efficiency.
- Data Collection: We used a dataset from Kaggle which contains fields like TripID, RouteID, StopID, StopName and WeekBeginning.
- Visualization Strategy: We outlined the strategy for creating data models, designing visualizations, and making dashboards interactive.
- Code Integration: We described the code integration process for data collection, cleaning, and analysis.

Phase 2: Data Exploration and Understanding

- We loaded the dataset and explored its structure, data types, and potential relationships between variables.

· Data preprocessing steps were taken to handle missing values, clean data, and prepare it for analysis.

· We applied data validation, missing data handling, data transformation rules, and statistical analysis.

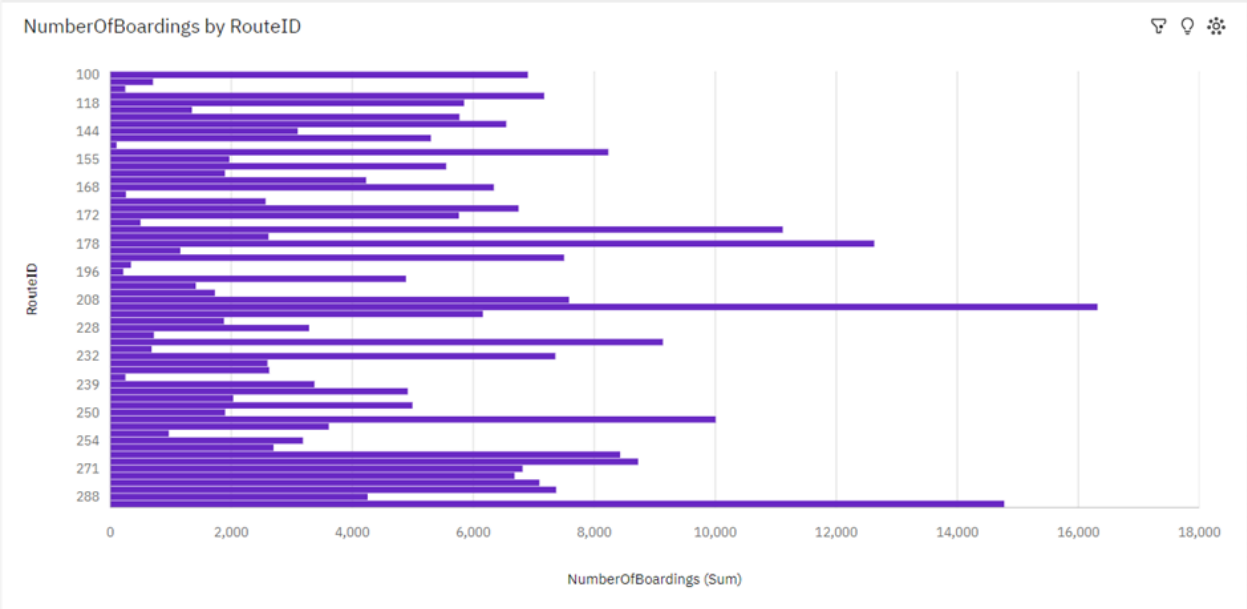
Phase 3: Predicting Service Disruptions

· We aggregated data at the route and week level to calculate service disruptions.

· Decision Tree and Random Forest classifiers were used to predict service disruptions.

· Visualizations were created to display trends in passenger boardings and service disruptions over time.

A bar chart visualizing the noOfBoardings for each route for RouteID ranging from 100 to 288

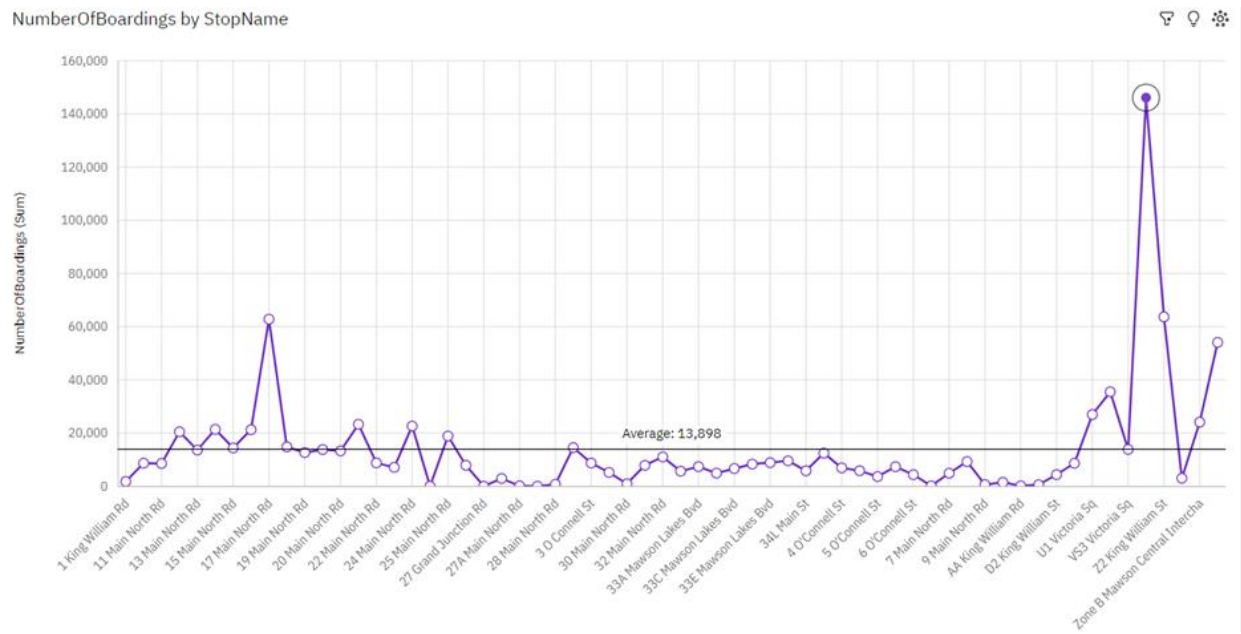


Insights:

RouteID 222.0 has the highest total NumberOfBoardings due to WeekBeginning 2013-07-21.

NumberOfBoardings is unusually high when RouteID is 222 and 300.

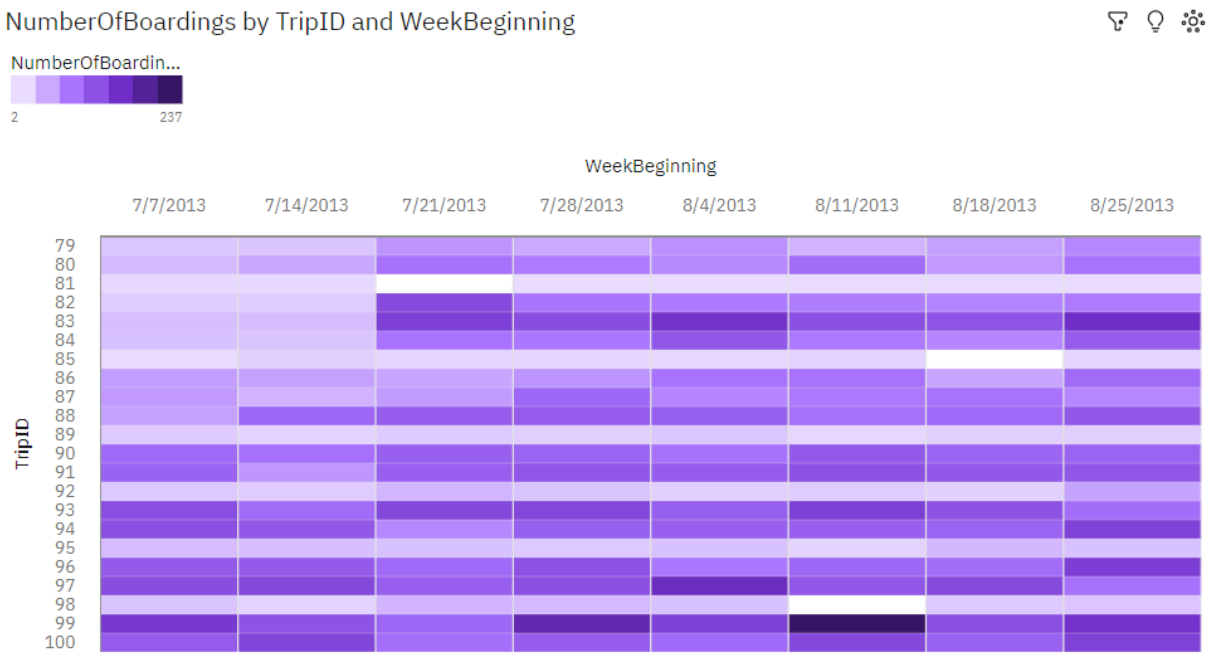
Visualizing NoOfBoardings by StopName



Insight:

NumberOfBoardings is unusually high when StopName is X1 King William St.

A heat map representing NoOfBoardings by TripID for the WeekBeginning from 7/7/2013 to 8/25/2013



Phase 4: Sentiment Analysis for Passenger Feedback

Dataset link: https://www.kaggle.com/datasets/pyrotech/british-airways-reviews-unfiltered

· Data preprocessing and text preprocessing steps were performed to prepare passenger feedback data for sentiment analysis.

· The VADER model was chosen for sentiment analysis due to its suitability for informal text.

· Features were engineered, and a Random Forest model was developed for sentiment analysis.

	Text	OverallSentiment	Positive	Neutral	Negative
0	Not Verified | Our A380 developed a fault taxi...	Negative	0.062	0.812	0.126
1	Not Verified | Horrible airline. Does not care...	Negative	0.032	0.892	0.076
2	✅ Trip Verified | My family and I have flown m...	Positive	0.071	0.895	0.034
3	✅ Trip Verified | This has been by far the wor...	Negative	0.046	0.895	0.059
4	✅ Trip Verified | In Nov 2022 I booked and pai...	Negative	0.024	0.899	0.077
...	...	...	...	...	...
3595	Flew return in CW from LHR to BKK in August 20...	Positive	0.091	0.842	0.067
3596	LHR to HAM. Purser addresses all club passenge...	Positive	0.279	0.721	0.000
3597	My son who had worked for British Airways urge...	Positive	0.071	0.885	0.045
3598	London City-New York JFK via Shannon on A318 b...	Positive	0.282	0.718	0.000
3599	SIN-LHR BA12 B747-436 First Class. Old aircraf...	Positive	0.093	0.809	0.098

[3600 rows x 5 columns]

Insights to Support Transportation Improvement Initiatives

The insights generated from the analysis provide valuable support for transportation improvement initiatives:

1. OnTime Performance Insights: By analyzing ontime performance, transportation authorities can identify specific routes or areas where improvements are needed. This can enhance the overall reliability and timeliness of public transportation.

2. Passenger Satisfaction: Insights from passenger feedback analysis can help address concerns related to cleanliness, staff behavior, and overall comfort. By enhancing passenger satisfaction, transportation services can be improved.

3. Service Efficiency: The analysis of service efficiency can lead to resource optimization, cost reduction, and more sustainable transportation systems. Optimization of routes and resource allocation can enhance efficiency.

4. Service Disruption Prediction: The predictive models for service disruptions provide transportation authorities with the ability to anticipate and mitigate disruptions in realtime. This leads to a more reliable and efficient service.

5. Sentiment Analysis: Sentiment analysis helps in understanding passenger sentiment, identifying common issues, and tracking sentiment trends over time. This information can guide improvements and enhance the passenger experience.

6. Visualizations: The visualizations created using IBM Cognos provide a clear and easily interpretable representation of data trends, enabling decision makers to make informed choices.

Conclusion

In conclusion, the insights gained through this analysis not only enhance public transportation efficiency but also contribute to sustainable urban development and resource optimization. By leveraging data-driven approaches, we aim to provide a more reliable and passenger oriented public transportation experience, ultimately improving the daily lives of commuters and supporting urban mobility initiatives.

Team Members

Yasvinippriyaa S K - 2021506126

Tejasree M S - 2021506113

Sreenithi B - 2021506105

Subithra C - 2021506109

Vaishnavi V - 2021506116