Toronto Transit Commission (TTC) Bus Delay

Group 4

Agam Sanghera, Ashita Diwan, Cheng Zhang, Yichun Liu

Introduction

Source: Toronto Star

Introduction

Home › Opinion › TTC delays cause frustration among riders

TTC delays cause frustration among riders

OPINION

BY KATHLEEN WRIGHT

PUBLISHED ON MARCH 2, 2012

Frustration is a feeling associated with Toronto's transit. There are several issues that need change and the most common issue is the amount of delays and lack of on-time service.

The Toronto Transit Commission (TTC) reports that on a regular business day, approximately 504,000 people depend on its service – streetcars, buses and the subway. In a city with a population of 2.5 million people, driving can be difficult.

Adam Kurey, 25, doesn't like driving in the city and sees the TTC as an alternate method of transit.

"The streets of Toronto tend to be very congested and there seem to be a lot of aggressive drivers," Kurey said. "Public transit is well structured way to get around the congestion that Toronto faces," he said.

CENTENNIAL COLLEGE

School of Communications, Media, Arts and Design

Source: Toronto Observer

Introduction

- Importance of the topic
  - Impacts on commuters
  - Effects on transportation network

Introduction

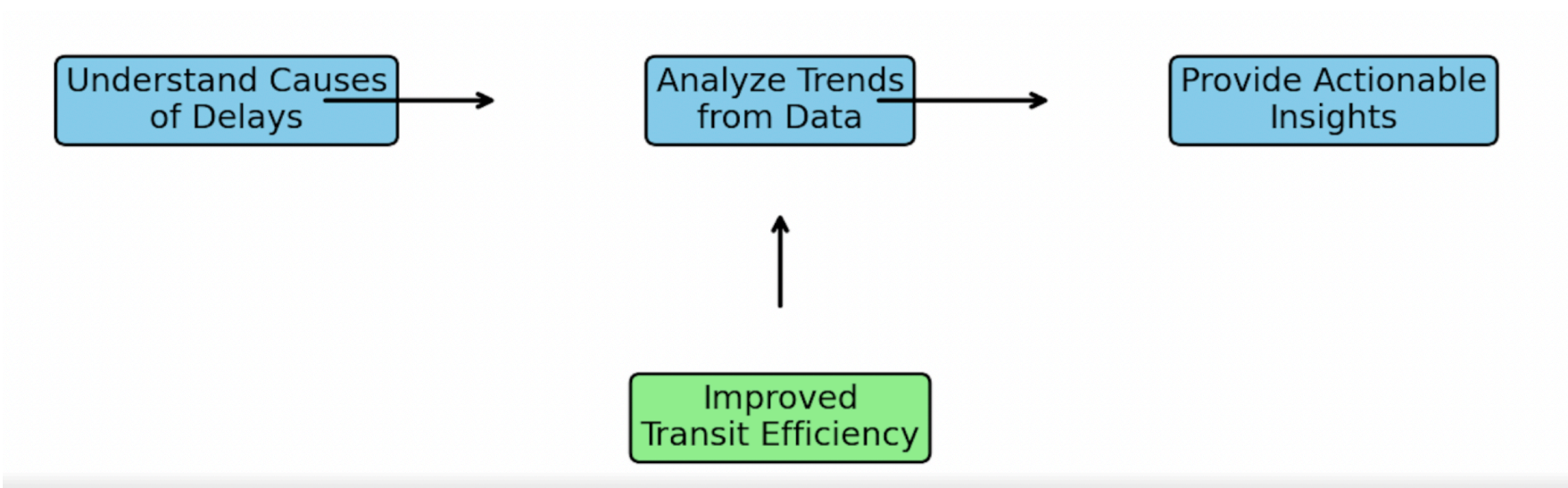
- Purpose of the report
  - To analyze bus delays in the TTC system
  - To identify key patterns, causes, and potential solutions for these delays

Introduction

- End product: build a logistic regression model to predict future bus delay duration
  - Better allocate resources
  - Enhance bus service precision

Objectives

- Understand the primary causes of bus delays.
- Analyze trends based on data (e.g., time of day, location, type of delay).
- Provide actionable insights to improve service efficiency.



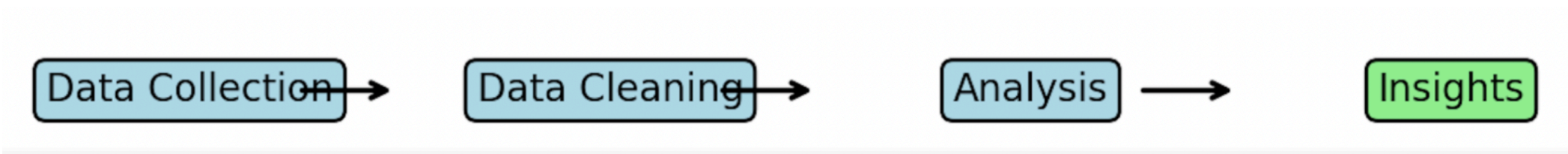
Methodology

Data Source: TTC bus delay logs.

Techniques Used:

- Statistical methods to identify trends and correlations.
- Data visualization for pattern recognition.

Tools/Software: Python (Pandas, Matplotlib, Seaborn).



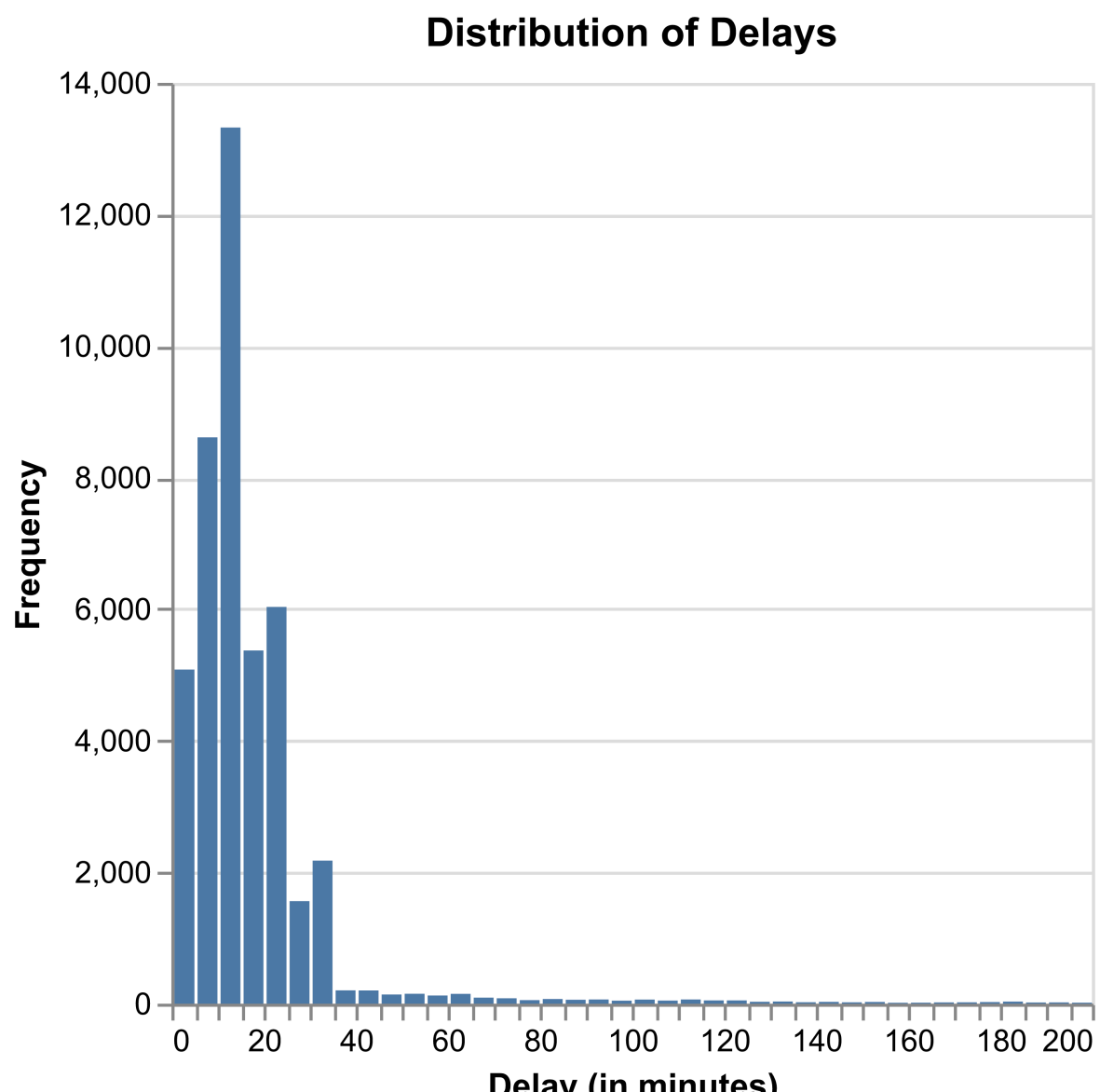
Analysis

Analysis

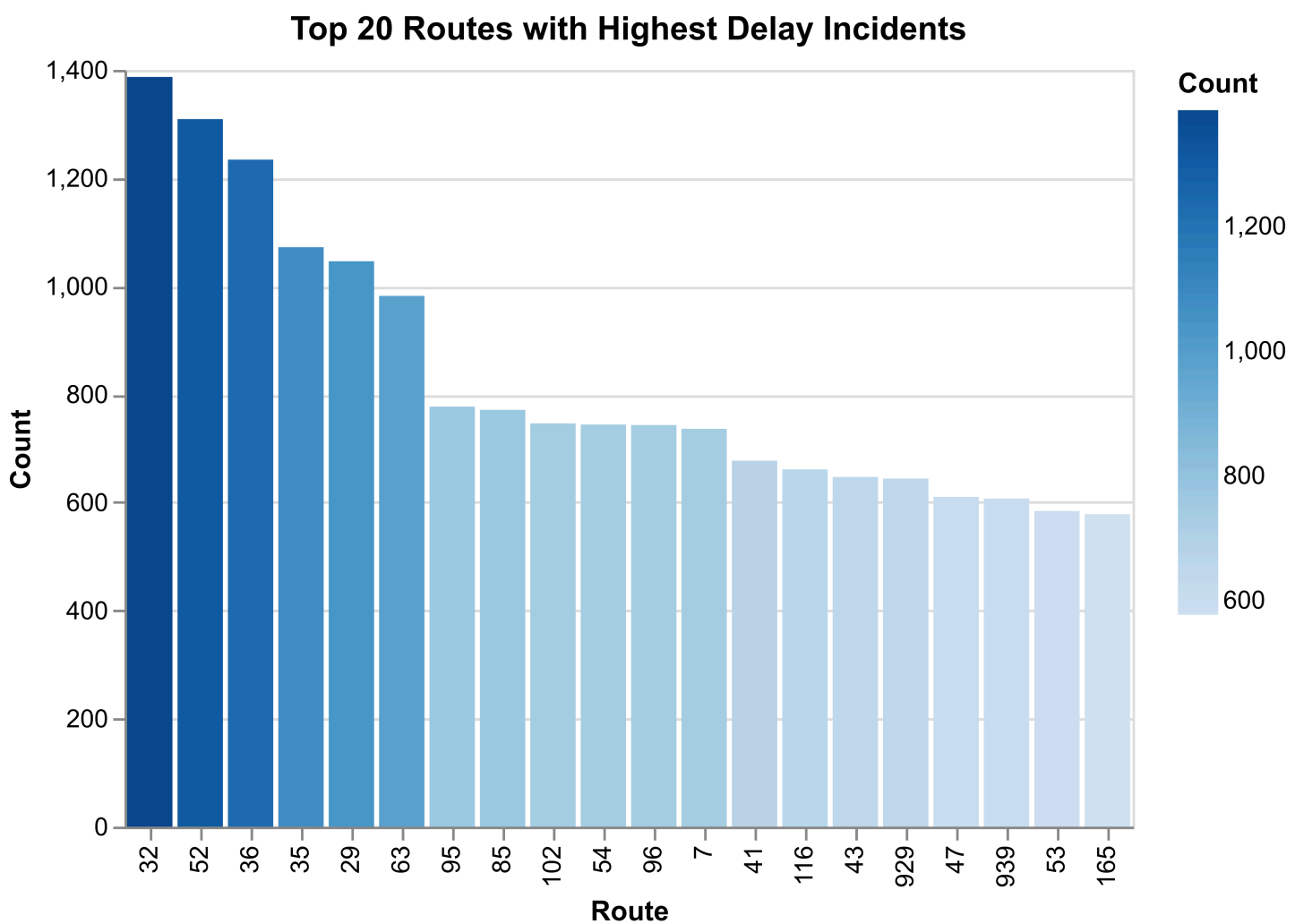
1. **Loading and Preprocessing Data**
2. **Visualization**
3. **Modelling**

Analysis: Loading and Preprocessing Data

Analyze distribution of delays



Identify top routes and locations with frequent delay incidents



Visualize delays based on day and incident type

Analysis: Modelling

- Logistic Regression model to predict
  - “Short”, “Medium” or “Long” duration
- Cross-validation and randomized grid for hyperparameter tuning

Results and Conclusions

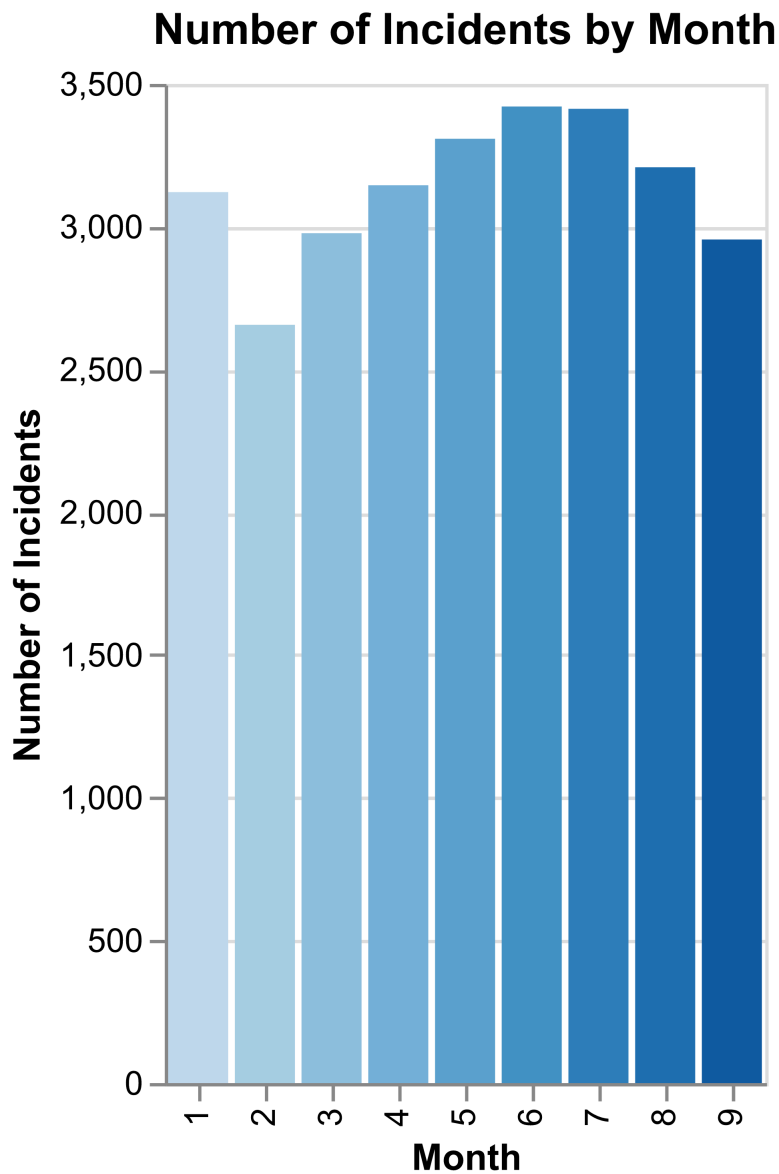
Results

- EDA
- Model

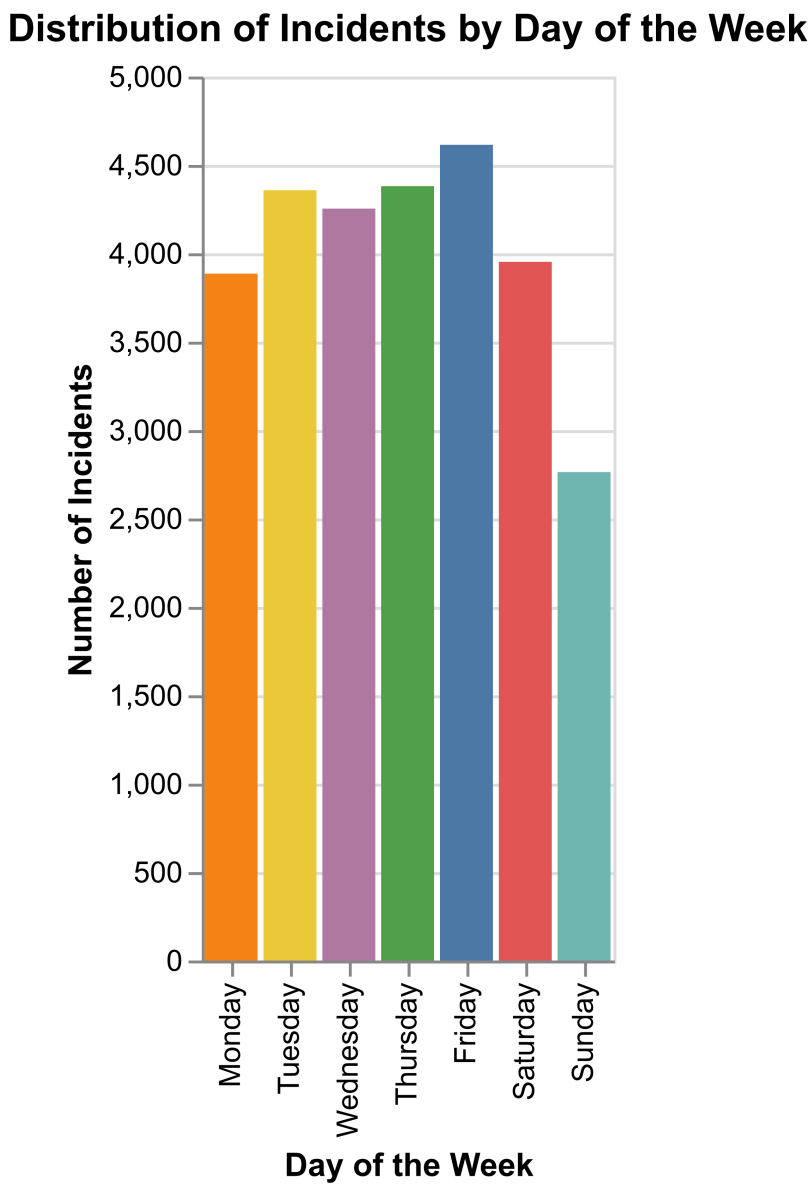
EDA

- **Distributions**
- **Reclassified Labels**

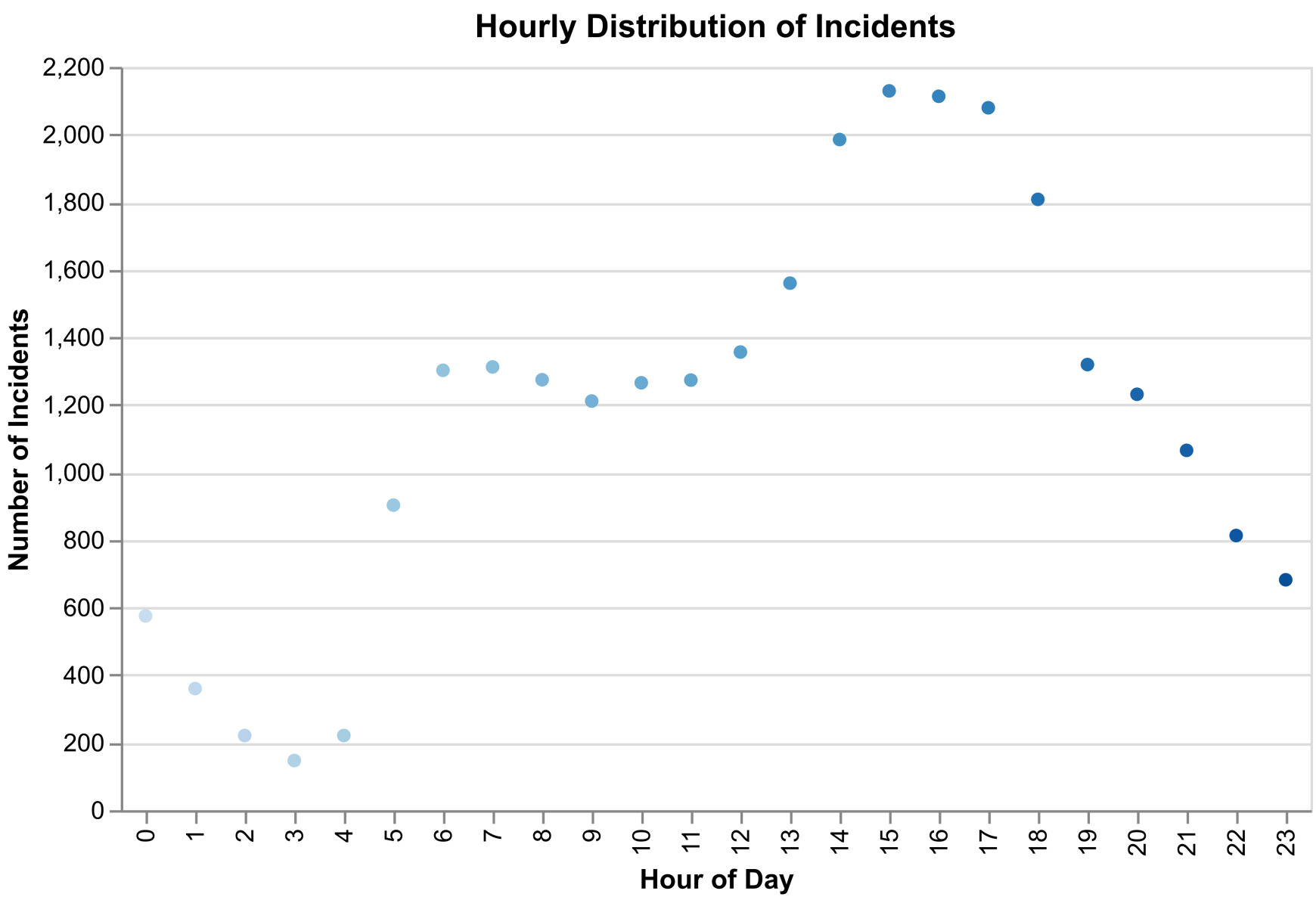
Distribution: Monthly Incidents



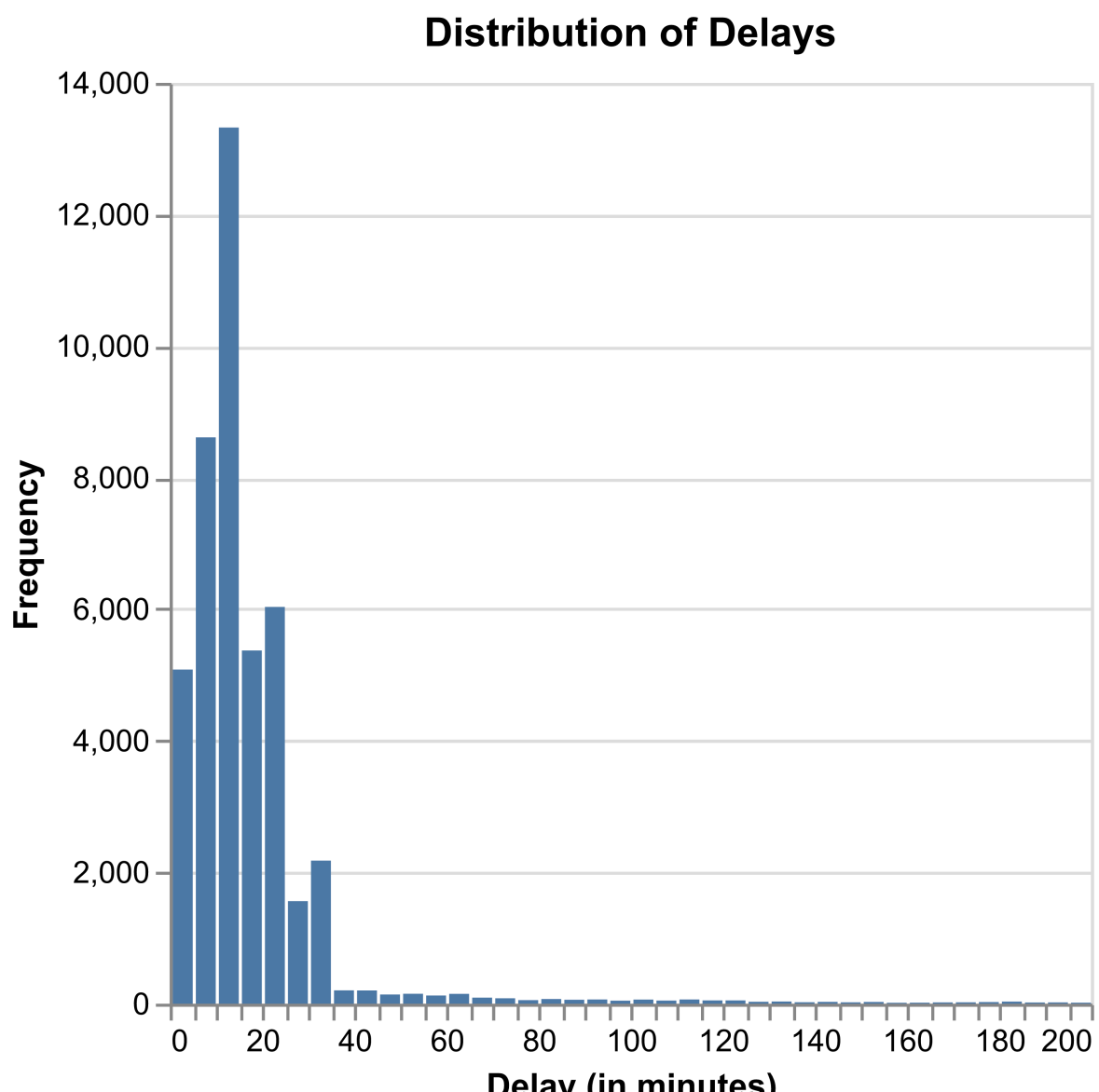
Distribution: Daily Incidents



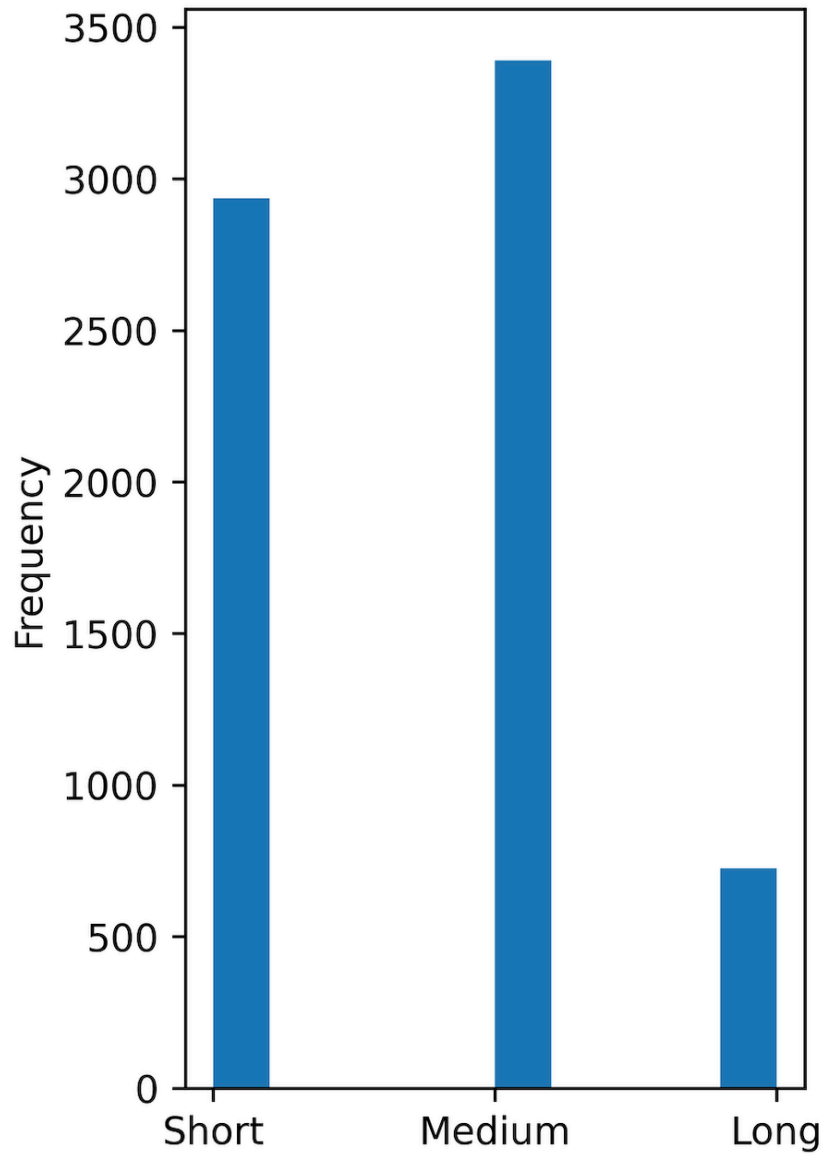
Distribution: Hourly Incidents



Distribution of Delays



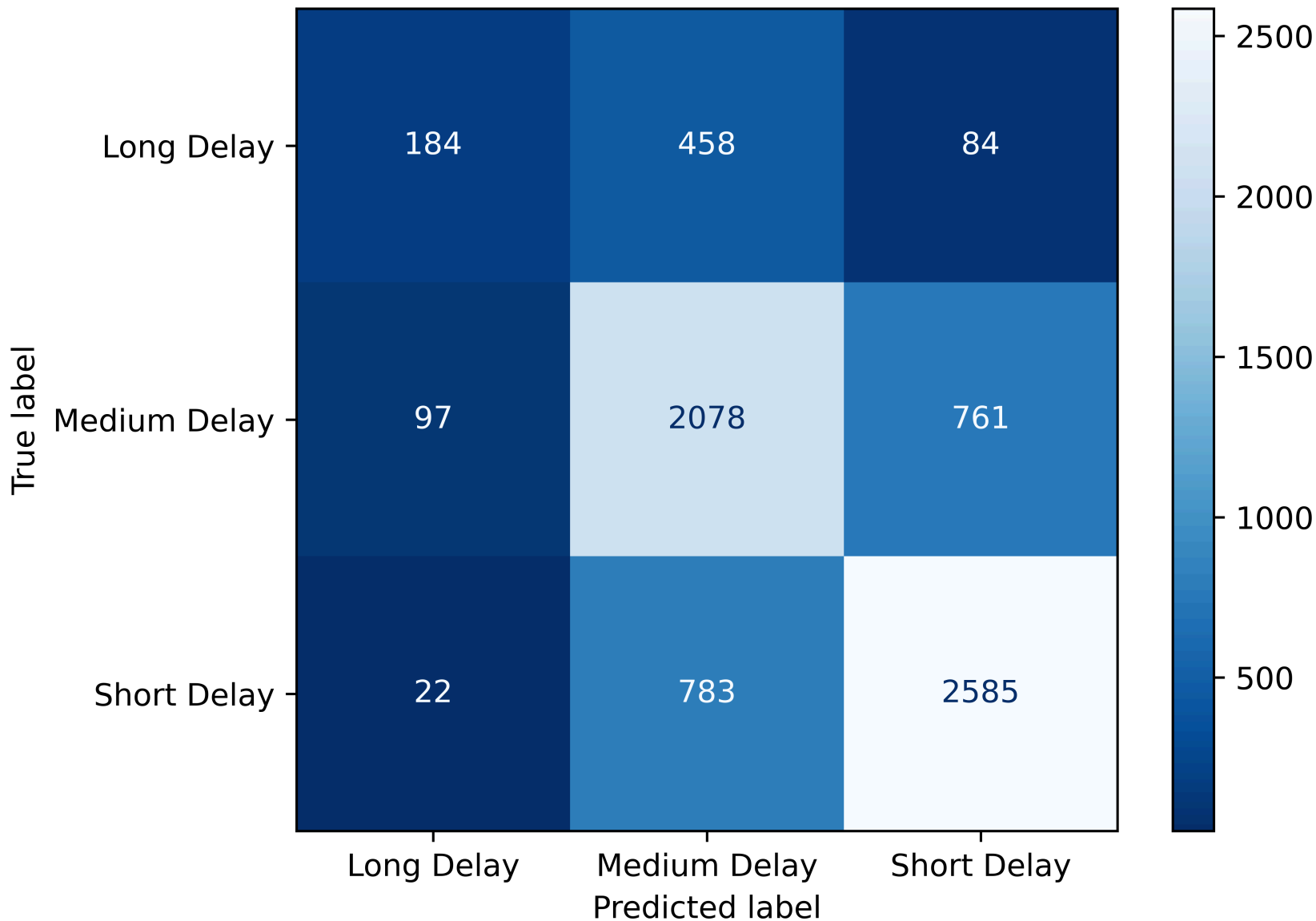
Reclassification of Delays



Model

- Hyperparameter Tuning
  - C : 0.1
- Model Scores
  - Test Accuracy : 68.32%
  - Training Accuracy: 71.63%
  - Cross Validation Accuracy: 69.53%

Confusion Matrix



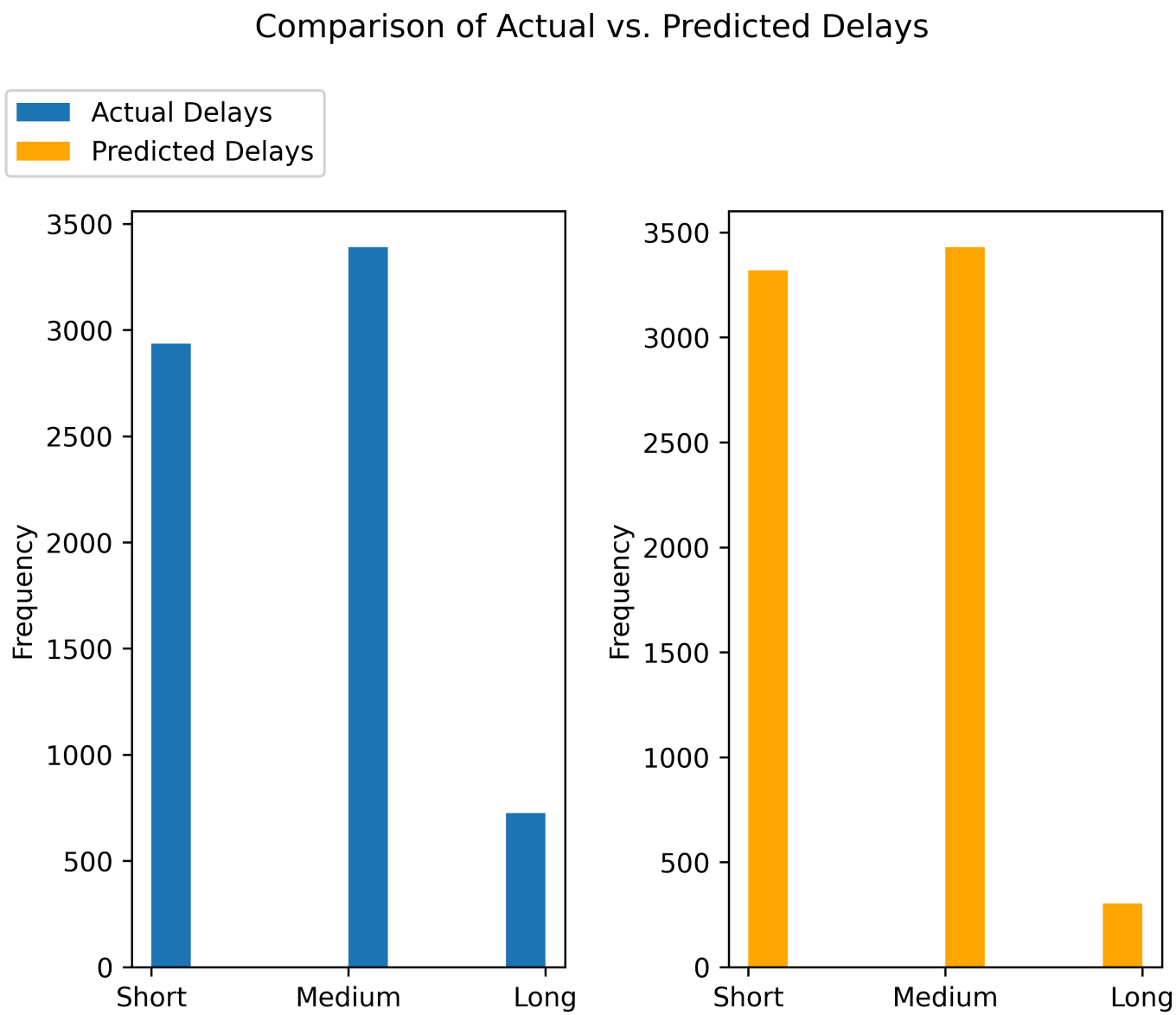
Conclusion

- Interpretation
- Future Scope

Interpretation

- Comparison of Actual vs Prediction
- Exlpanation of Results

Comparison of Actual vs Predictions



Future Scope

- Improve Prediction of Long Delays
- Experiment with other techniques

Thank you for your attention