TTC Bus Delay

Agam Sanghera, Ashita Diwan

Summary

Introduction

Analysis

Analysis

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

Analysis: Loading and Preprocessing Data

- Handle missing values,
- Convert timestamp data to day parts, and
- Clean data fields irrelevant to the analysis

Analysis: Visualization

- Analyze distribution of delays,
- Identify top routes and locations with frequent delay incidents, and
- 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**

Model

- **Model Scores**
- **Hyperparameter Tuning**
- **Cross Validation**

Conclusion

- **Interpretation**
- **Future Scope**

Interpretation

- **Comparison of Actual vs Prediction**
- **Exlpanation of Results**

Future Scope

- **Improve Prediction of Long Delays**
- **Experiment with other techniques**

Thank you for your attention