Bike Share System in SF Bay Area

Author: Owen Hsu

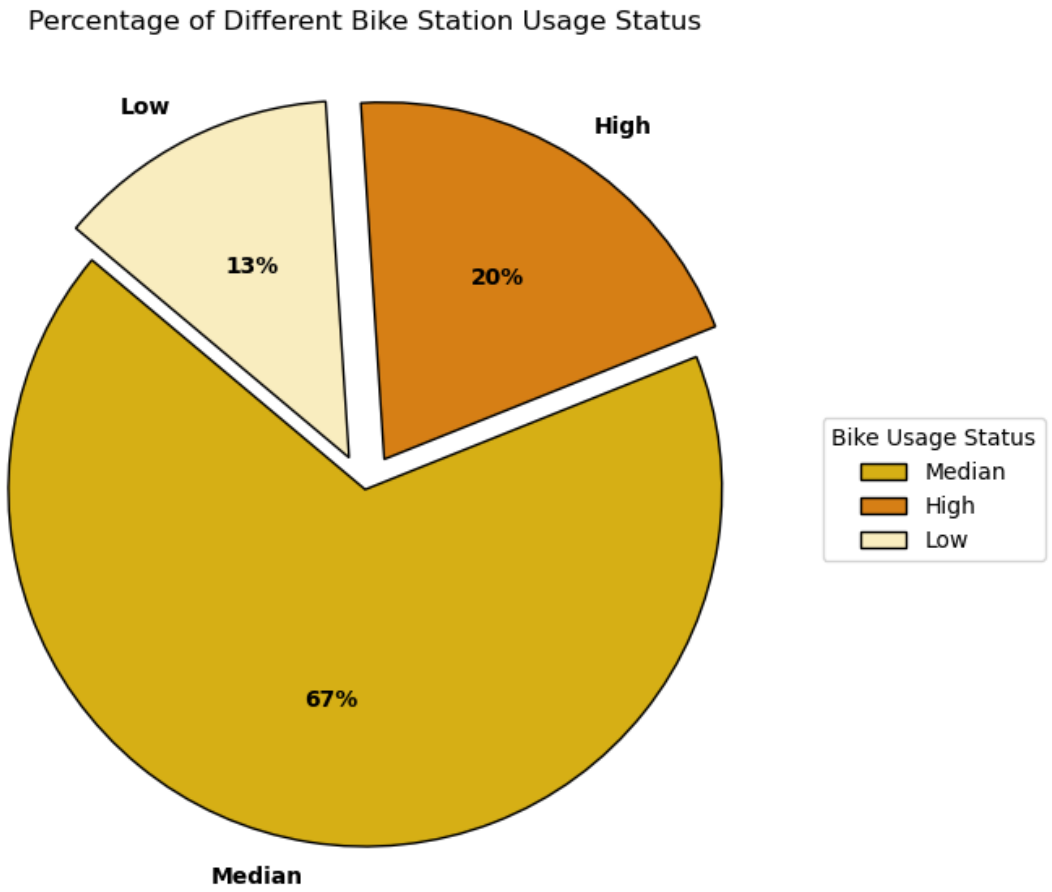
Project Overview

- **The Problem Area:**
  - How might we enhance the efficiency and accessibility of the bike share system in the SF Bay Area?
- **Approach**
  - Building a machine learning model to predict usage patterns, so that we can optimize bike availability through strategic redistribution, ensuring they are accessible where and when they are most needed.
- **Potential Impacts**
  - We aim to enhance user experience and achieve a 3% increase in usage.
  - Cost savings on transportation by $0.3 (Car) – $0.6 (Bus) per mile.
  - Reduction in CO2 emissions by 160 g/mile (Bus).- 400 g/mile (Car).

Dataset Overview

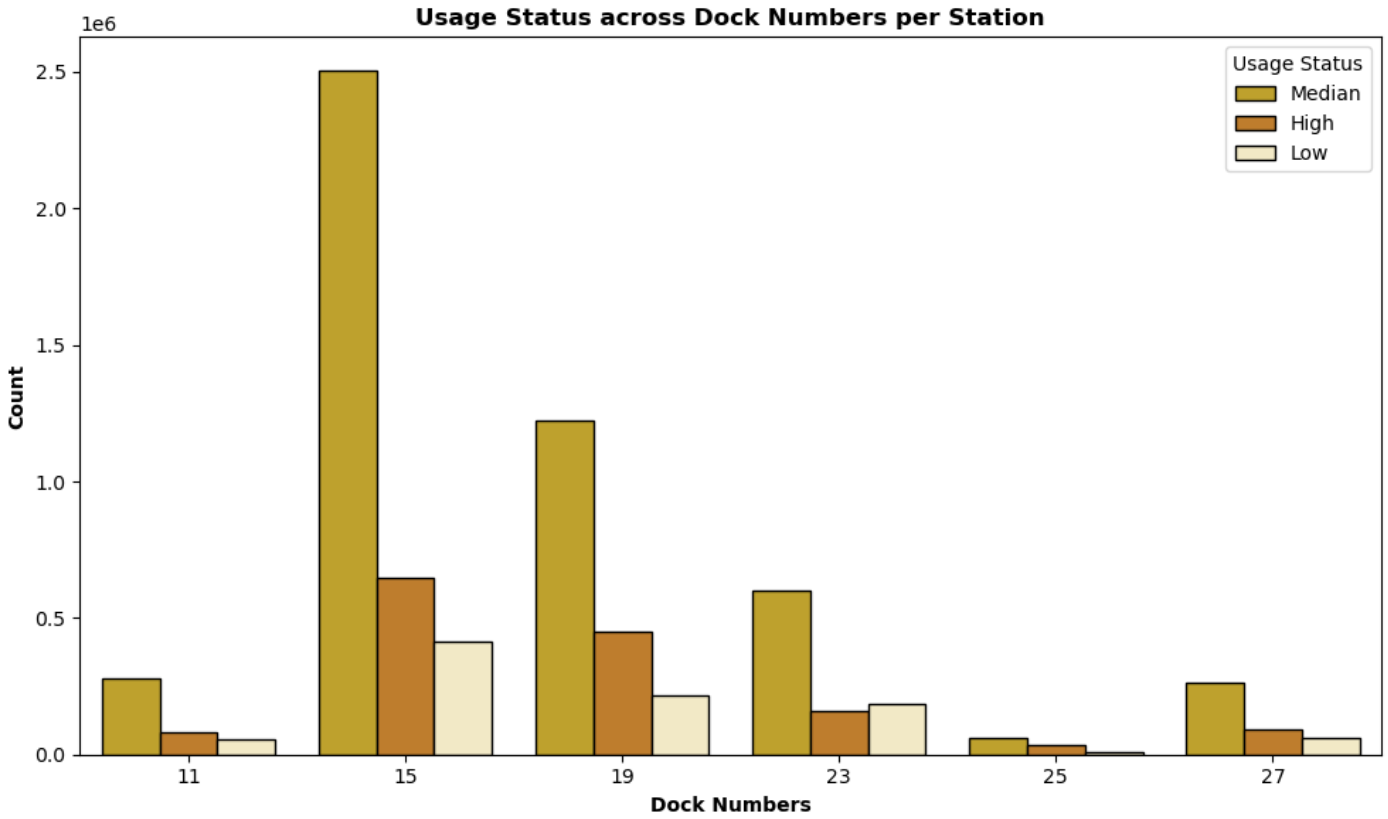
- **Original Datasets:**
  - Station, Status, and Weather datasets
  - 71,984,434 rows; 7 columns, 4 columns, and 24 columns respectively
- **Merged, Filtered, and Cleaned Dataset:**
  - 7,337,194 rows; 15 columns (13 numerical and 2 datetime)
  - Station ID, Date/Time, Available Bikes, Available Docks, Total Docks, and Weather Info

EDA



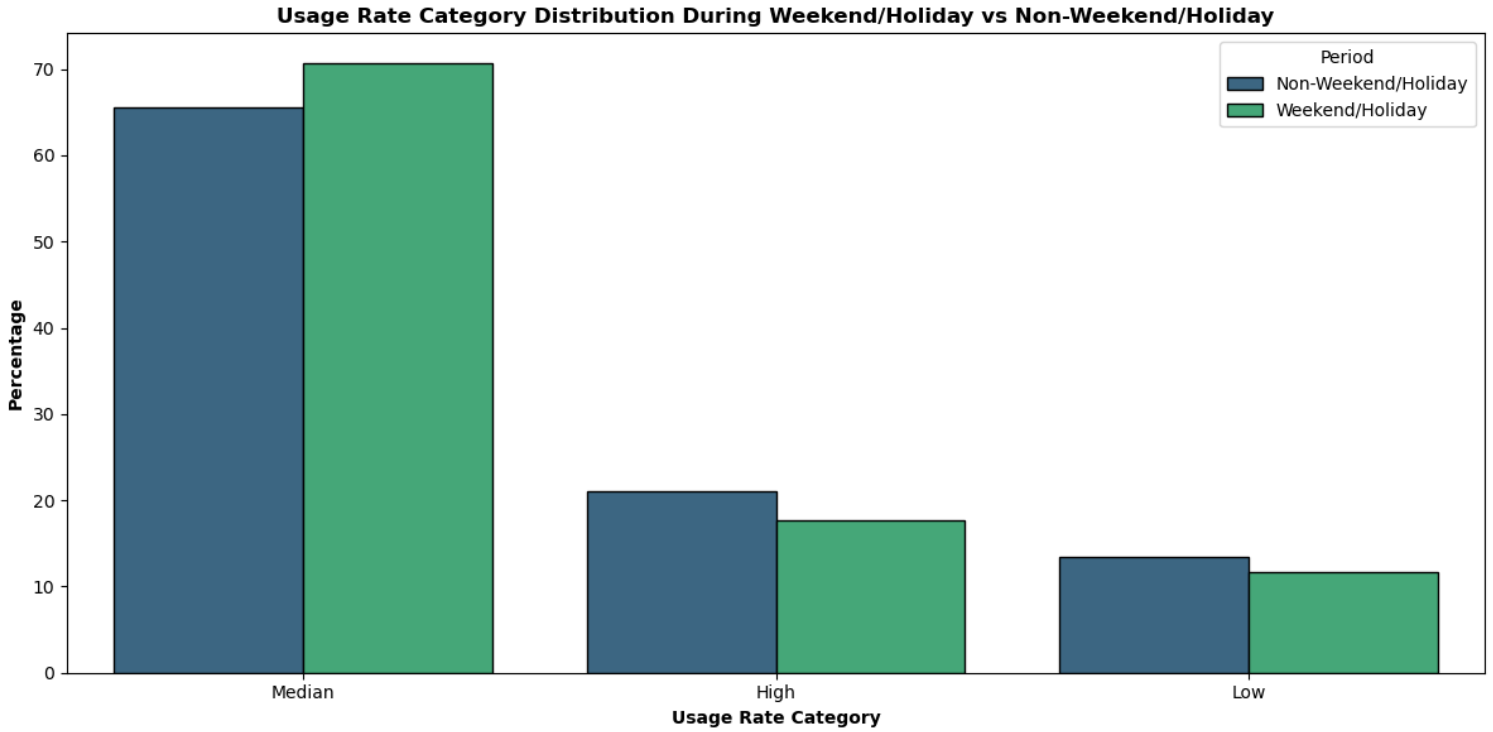
- **Target Variable - Bike Usage Rate**
  - Median: 67%
  - High: 20%
  - Low: 13%

EDA



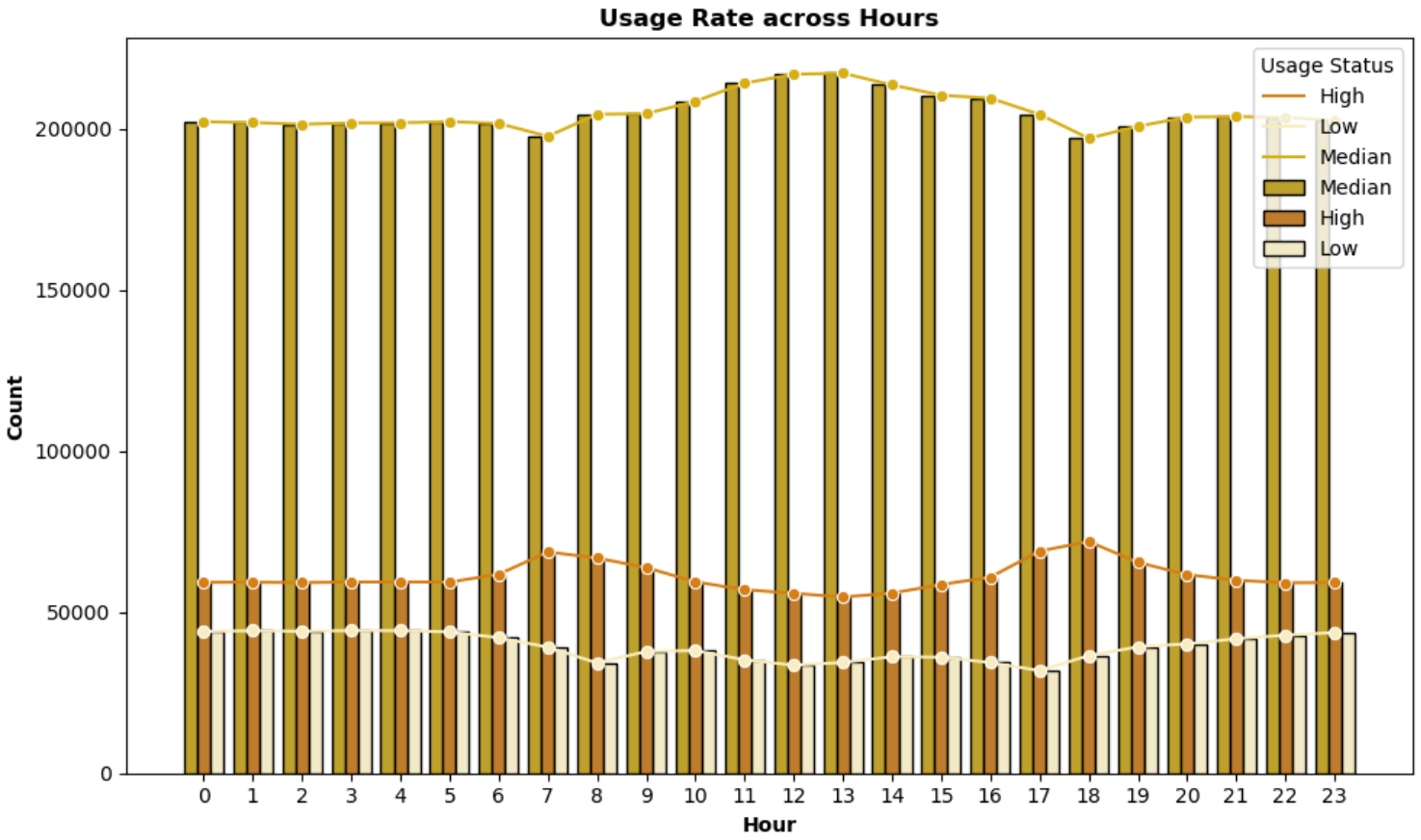
- The lower the dock number of the station, the more likely it is to have either a higher or lower usage rate.
- This is expected, as smaller bike stations' usage rates are more sensitive to the available bike and dock numbers.

EDA



- High usage rates:
  - 21% to 17%
- Low usage rates:
  - 13% to 11%
- Median usage rates:
  - 65% to 70%

EDA

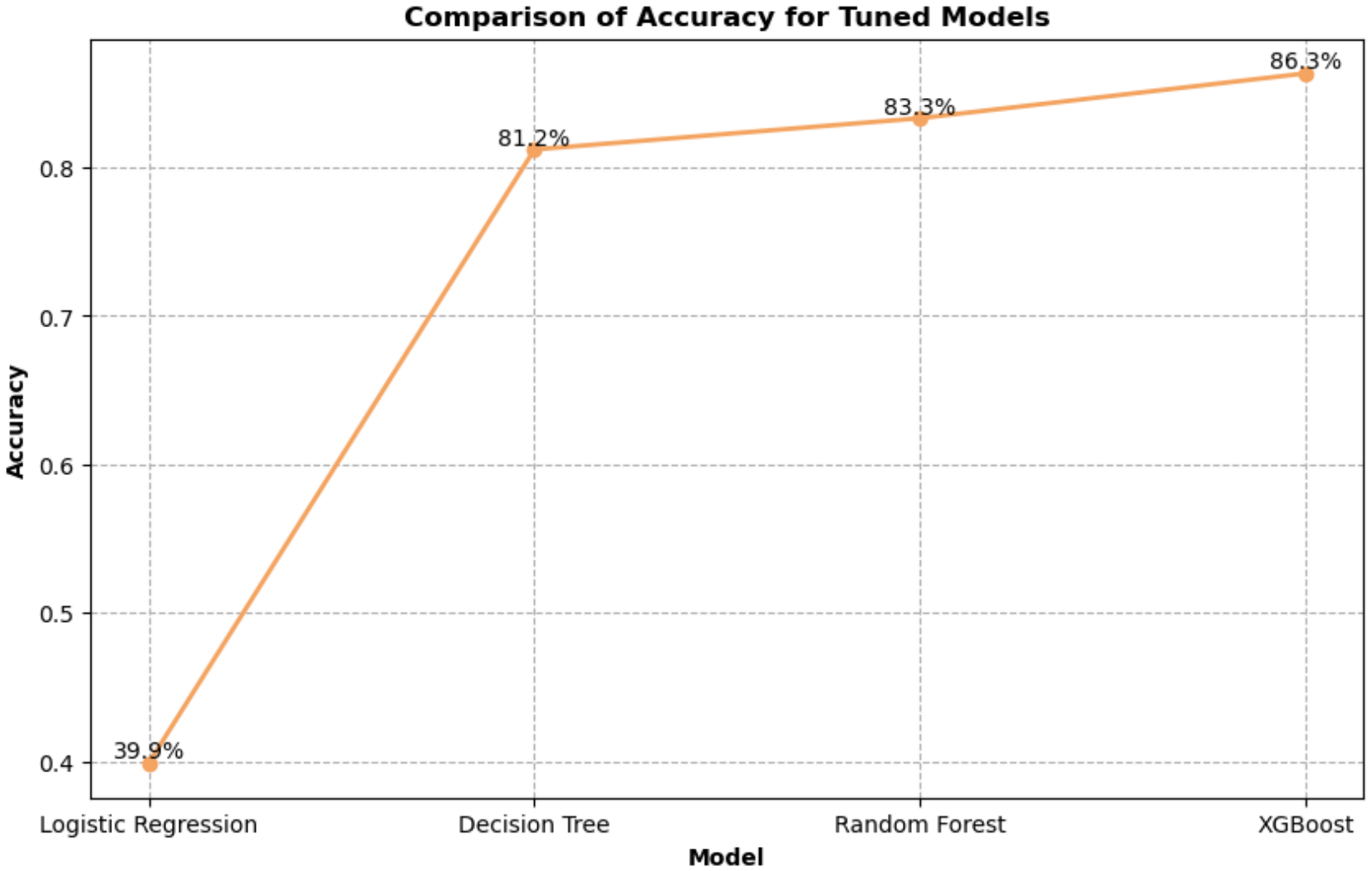


- 7 - 9 AM and 5 - 7 PM have the highest high bike usage rates during the day.
- 8 AM and 5PM have the lowest low bike usage rates during the day.
- This could be commuter influence.

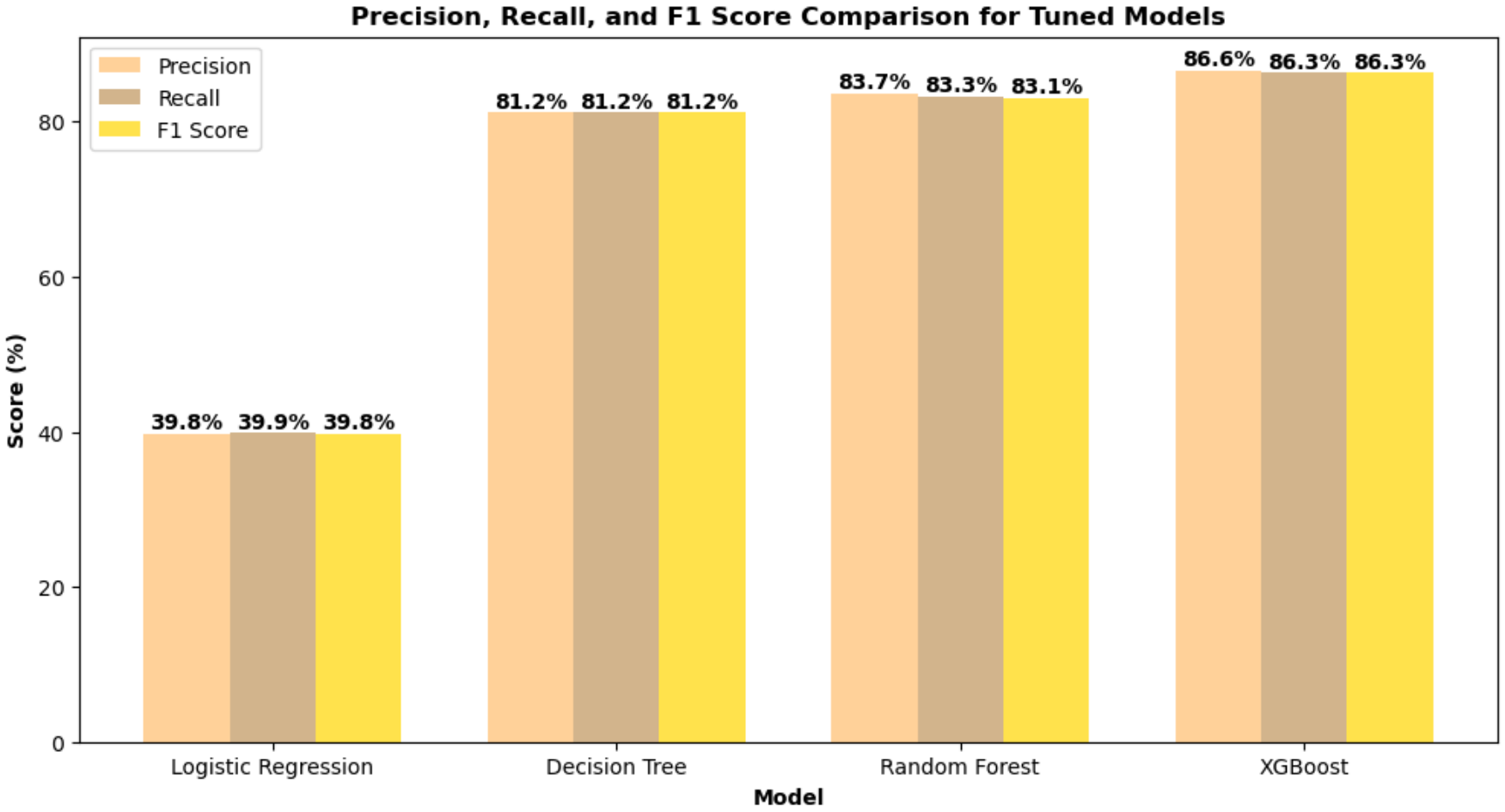
Summary of Tuned Models

Model	Train Score (%)	Test Score (%)	Elapsed Time (seconds)
Logistic Regression	40.03%	39.87%	11.38
Decision Tree	82.72%	81.20%	21.60
Random Forest	84.50%	83.31%	753.97
XGBoost	87.31%	86.33%	65.50

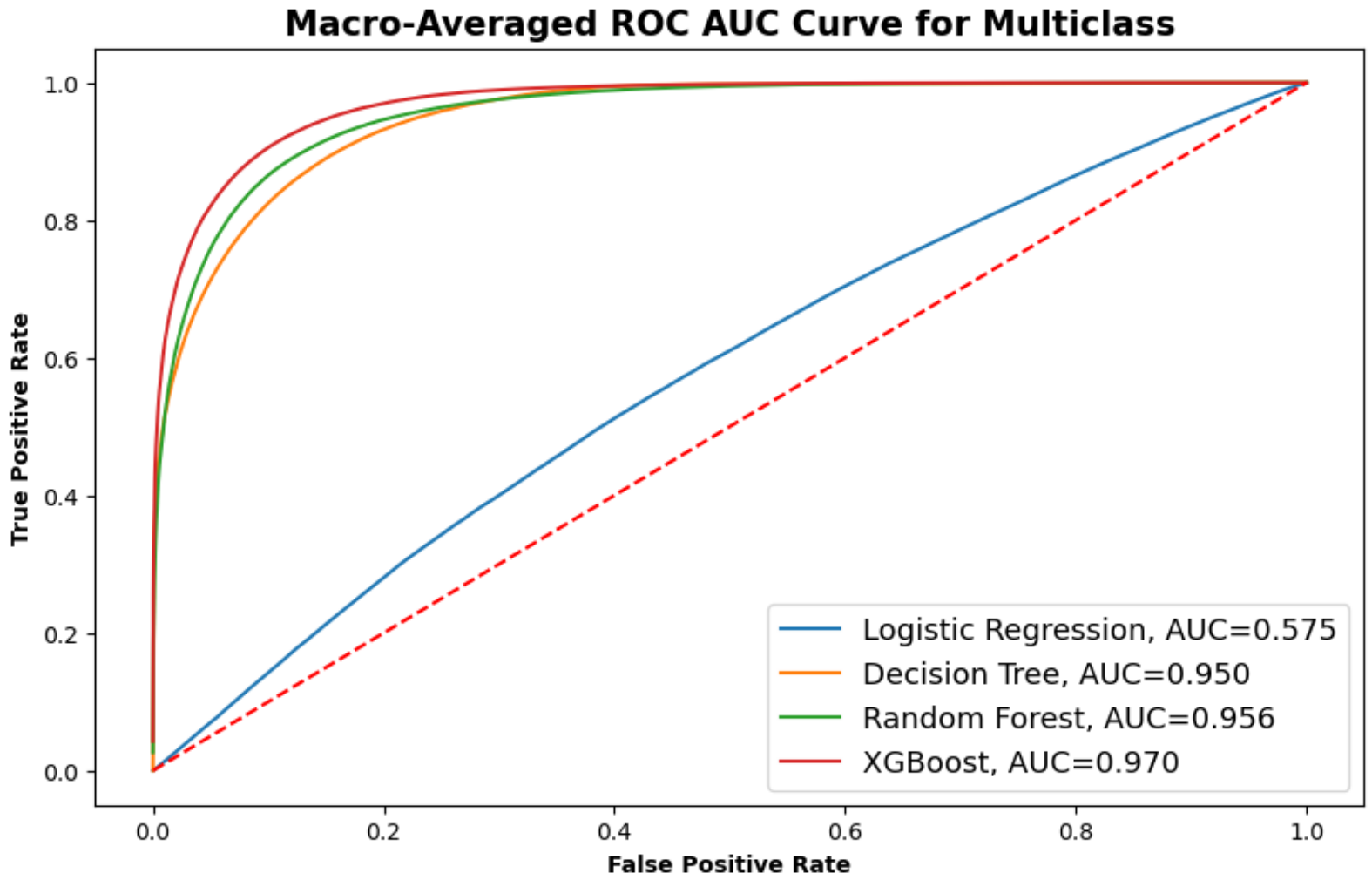
Summary of Tuned Models



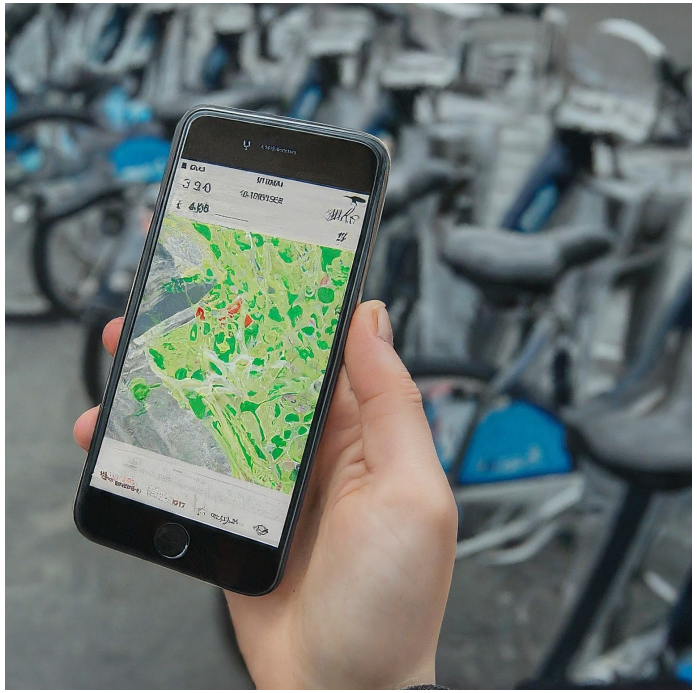
Summary of Tuned Models



Summary of Tuned Models



Product - Web App Concept



- **Interactive Map**
  - A map that shows live bike availability and station status across the city.
  - Users can click on stations to see real-time data such as the number of bikes available, and the station's popularity.
- **Predictive Insights**
  - Display predictions on the best time to ride based on past usage patterns and weather data.
  - Offer personalized recommendations for routes that optimize for scenery, speed, or minimal traffic.

Thank you