

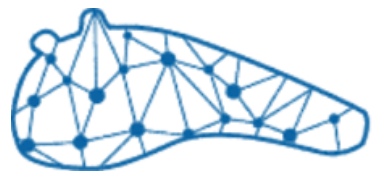
WEATHER OR TIME? UNDERSTANDING WHAT DRIVES CAPITAL BIKESHARE USAGE

DATA 601 FALL 2025

Author: Thanusha Balasubramanian



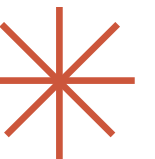
ABOUT THE COMPANY



UC Irvine
Machine Learning
Repository

QUICK FACTS

Launch Year	2010
Initial Fleet	~400 bikes
Current Fleet	6,000+ bikes
Stations	700+
Membership Types	Casual (Single Ride or Day Pass), Registered (Annual)
Coverage Area	Washington, D.C., Maryland, Virginia

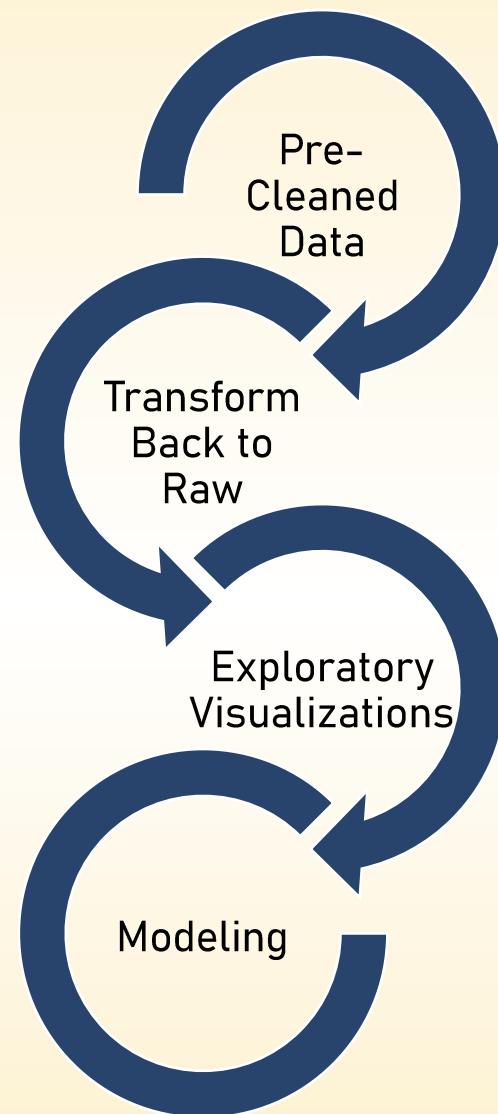


WHAT IS THE
RELATIONSHIP
BETWEEN WEATHER
AND THE NUMBER OF
BIKE SHARE USERS IN
AN HOUR?

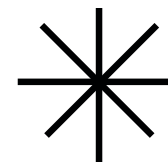
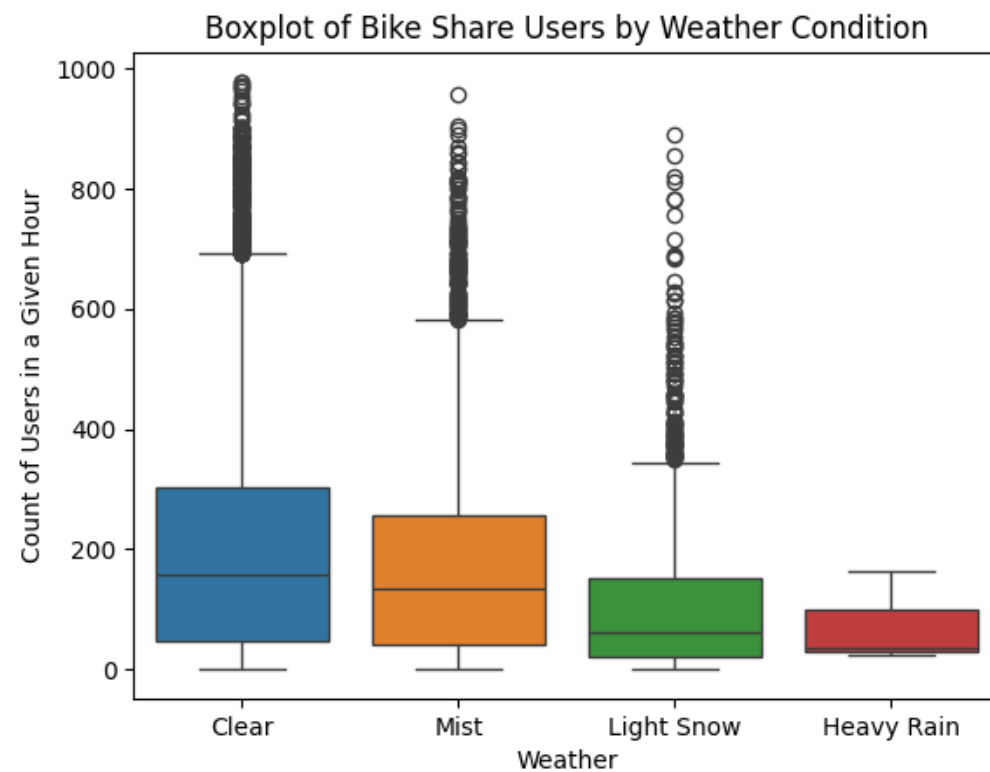
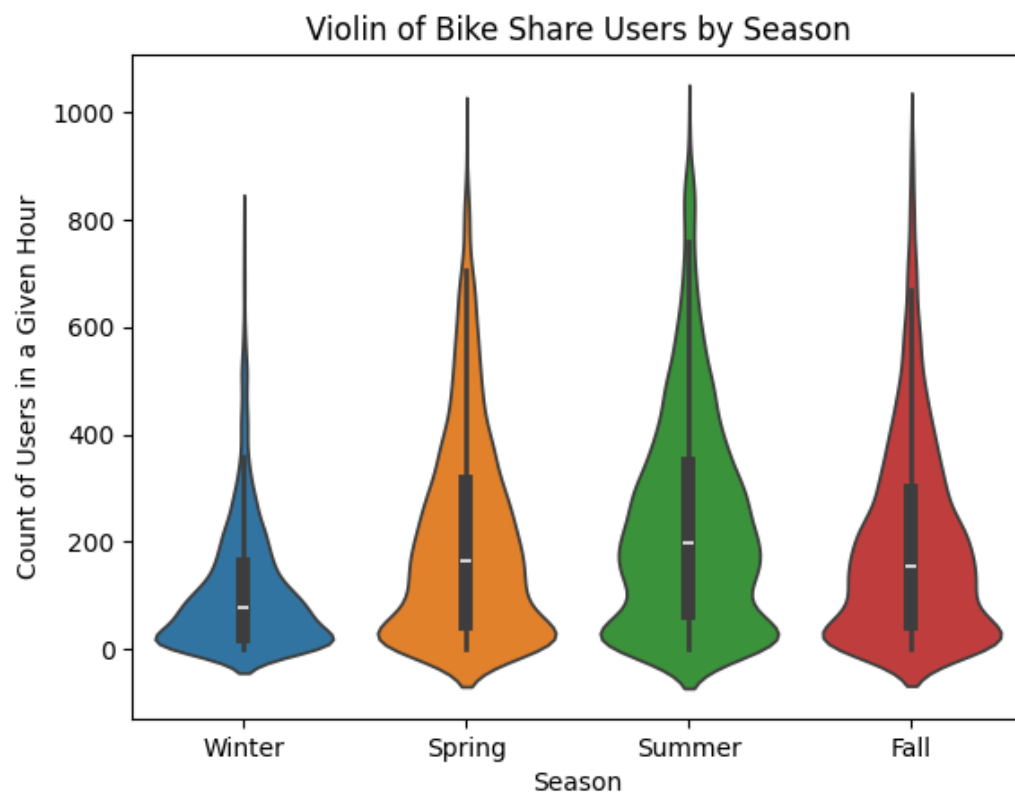


DATASET & WORKFLOW

Variable Name	Role	Type	Description	Units	Missing Values
temp	Feature	Continuous	Normalized temperature in Celsius. The values are derived via $(t-t_{min})/(t_{max}-t_{min})$, $t_{min}=-8$, $t_{max}=+39$ (only in hourly scale)	C	no
atemp	Feature	Continuous	Normalized feeling temperature in Celsius. The values are derived via $(t-t_{min})/(t_{max}-t_{min})$, $t_{min}=-16$, $t_{max}=+50$ (only in hourly scale)	C	no
hum	Feature	Continuous	Normalized humidity. The values are divided to 100 (max)		no
windspeed	Feature	Continuous	Normalized wind speed. The values are divided to 67 (max)		no
casual	Other	Integer	count of casual users		no
registered	Other	Integer	count of registered users		no
cnt	Target	Integer	count of total rental bikes including both casual and registered		no

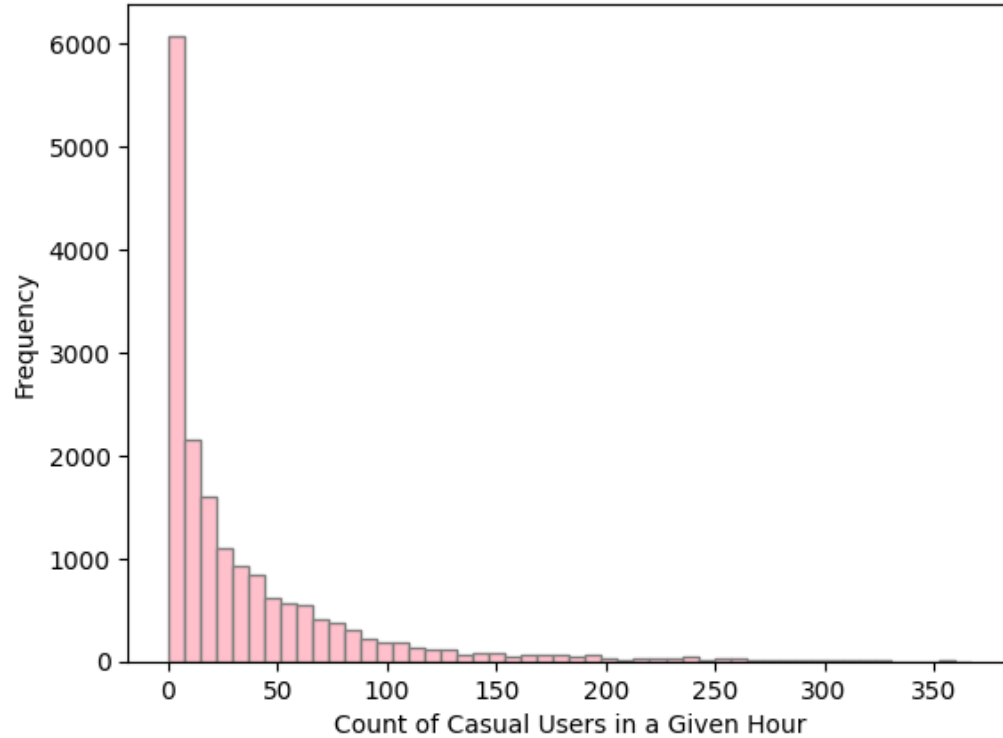


WEATHER

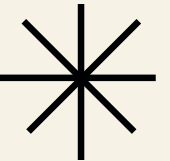
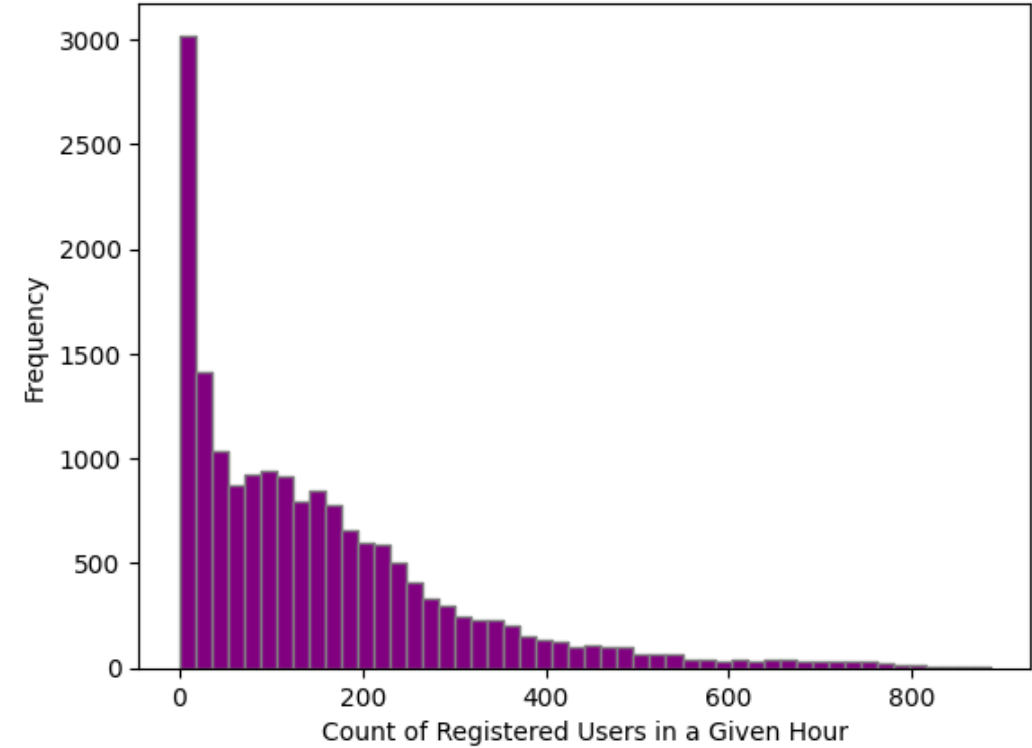


USER TYPE

Histogram of Casual Users in a Given Hour



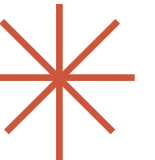
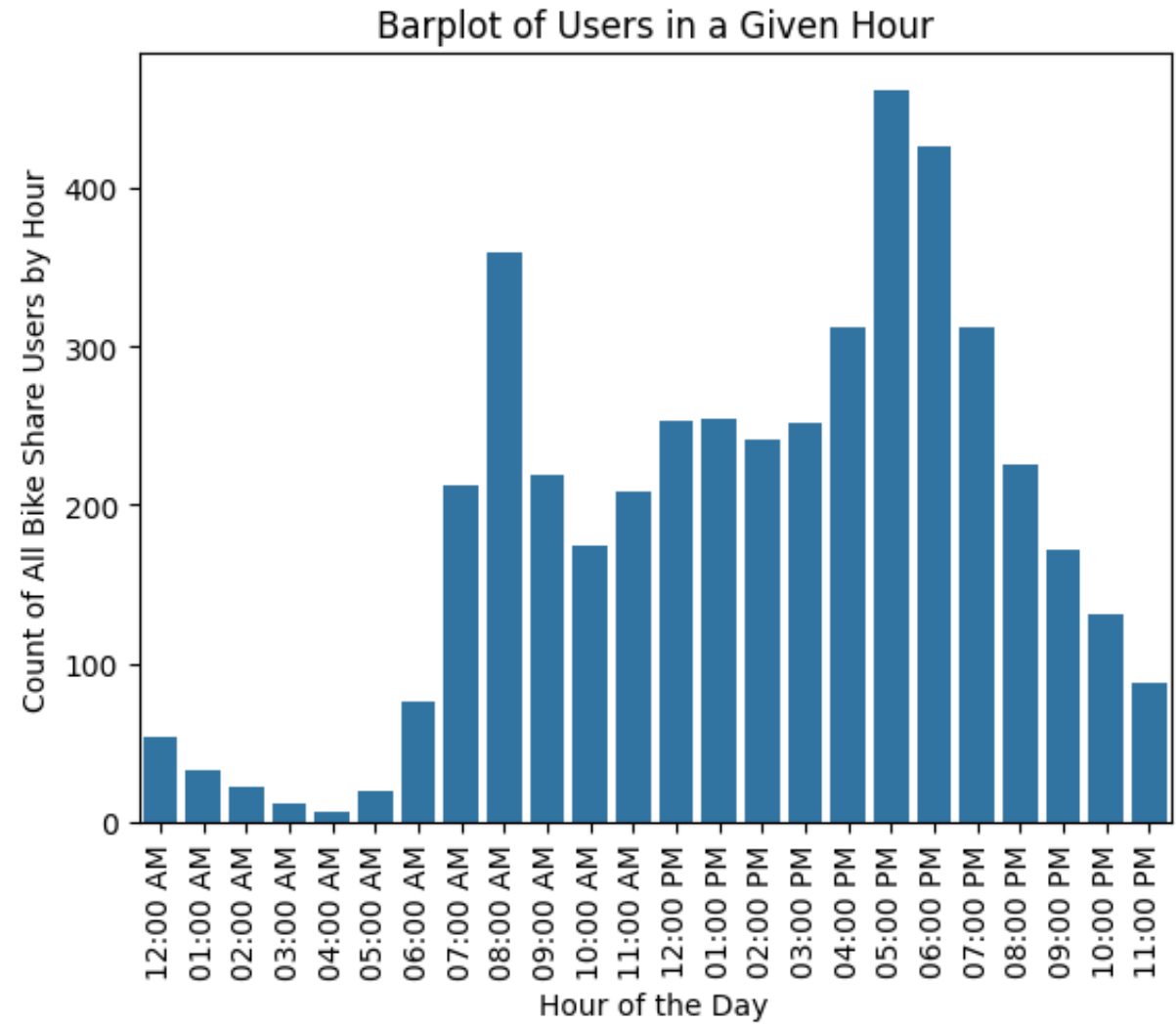
Histogram of Registered Users in a Given Hour



WHAT IF WE LOOK
AT THE HOUR?



COMMUTES?



ANALYSIS: MODELING APPROACH AND RESULTS



MODEL 1 – LINEAR REGRESSION

RMSE = 139.78

Mean of Users = 189.46

Median of Users = 142

FEATURE	COEFFICIENT
Hour	53.50
Feel Temperature	44.53
Year	40.71
Humidity	37.60
Season	21.89
Temperature	11.76

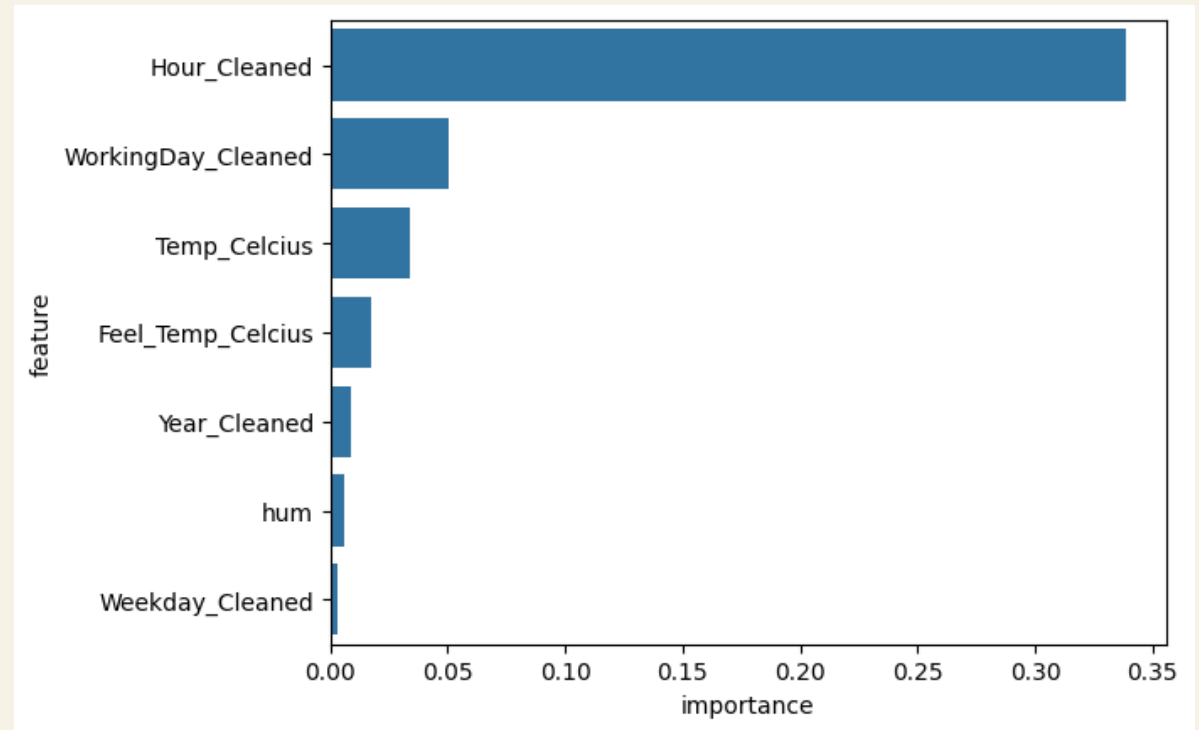
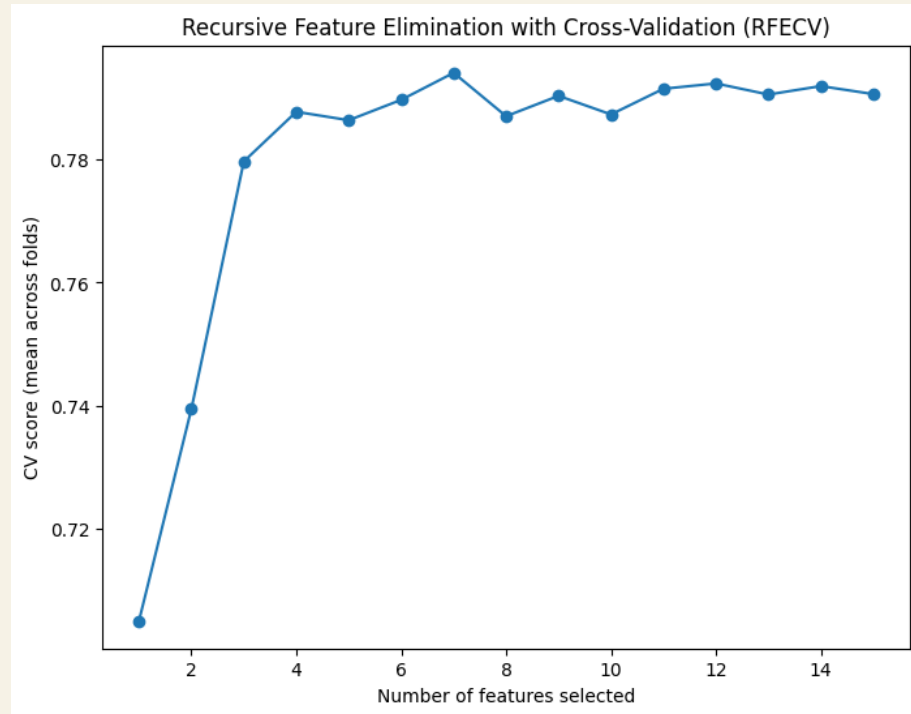


MODEL 2 – RANDOM FOREST CLASSIFICATION: BUCKETING

USER COUNT BUCKET	NO. OF USERS
Low Usage	<40 users ~ Bottom 25%
Medium Usage	40 – 281 Users ~ Middle 50%
High Usage	>281 Users ~ Top 25%



MODEL 2 – RANDOM FOREST CLASSIFICATION: FEATURES

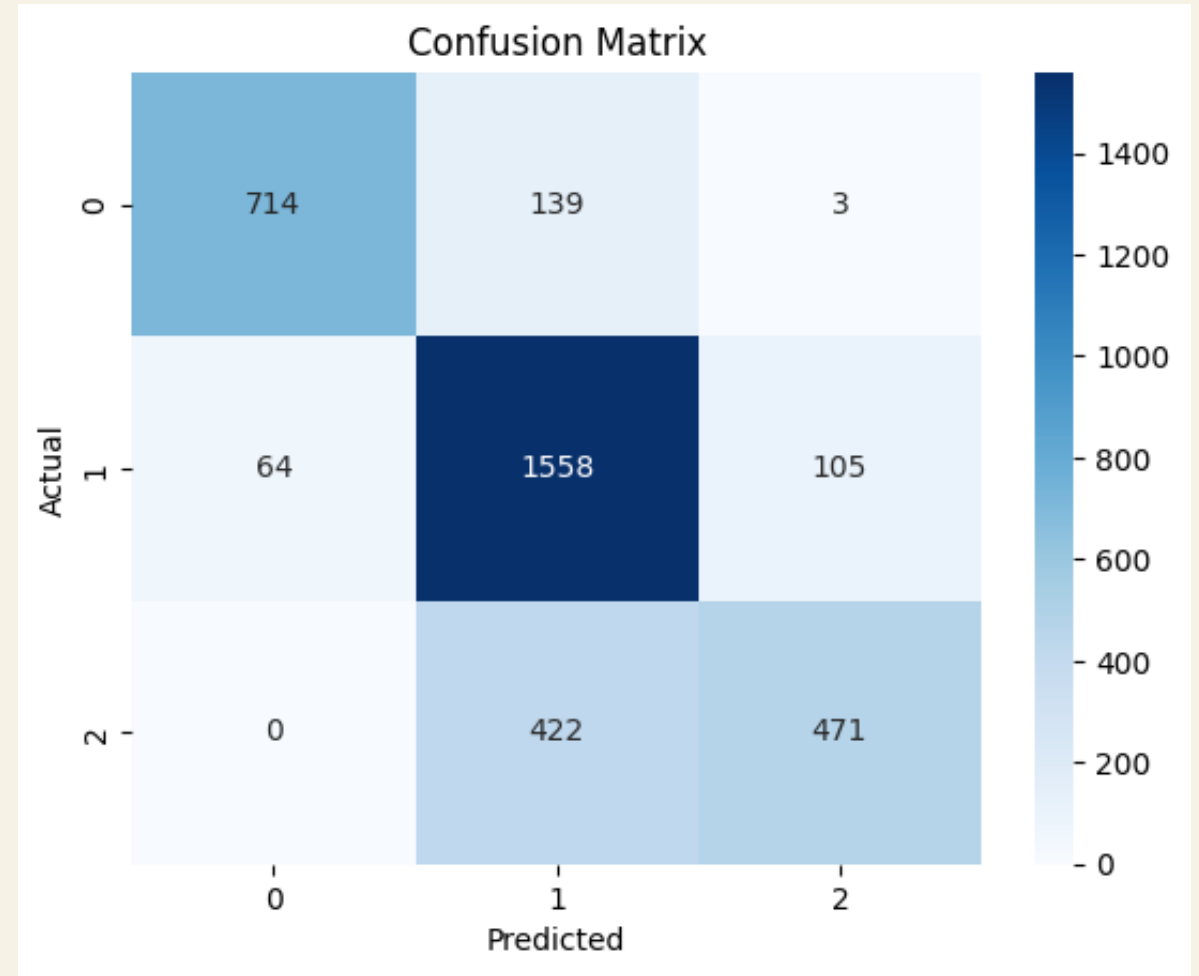


MODEL 2 – RANDOM FOREST CLASSIFICATION: MODEL ACCURACY

Accuracy = 0.7891

Precision = 0.8003

Recall = 0.7891



APPLICATION: BUSINESS RECOMMENDATIONS



BUSINESS INSIGHTS & APPLICATIONS



APPLICATION	RECOMMENDATION
Dynamic Pricing	Apply surcharge during peak hours
Operational Efficiency	Charge and redistribute bikes during low-use hours
Marketing Opportunities	Incentivize off-peak rides with discounts or loyalty offers



SOMETIMES
YOU FIND THE
UNEXPECTED



FUTURE OF PROJECT

Improvement on Regression Model

The regression model was built before our class on regression, so I may proceed to fine tune the model. (ie. Adjust features included, maybe transform the data). Check the conditions of the model (Data 602) to see if the residuals are normally distributed and check for homoscedasticity.

Additional Fine Tuning of Classification Model

I will change the hyperparameters, beyond number of features (ex: max depth) to improve model. Try a different classification for usage.

