



DATA ANALYTICS IN PYTHON FINAL PROJECT



TIP PREDICTOR FOR DRIVERS

PRESENTED BY
Charlotte Lin, Kunjia Shi,
Ruijie Shi, Shanrong Zhou, Xinyi Ma

HELP FHV DRIVERS EARN MORE

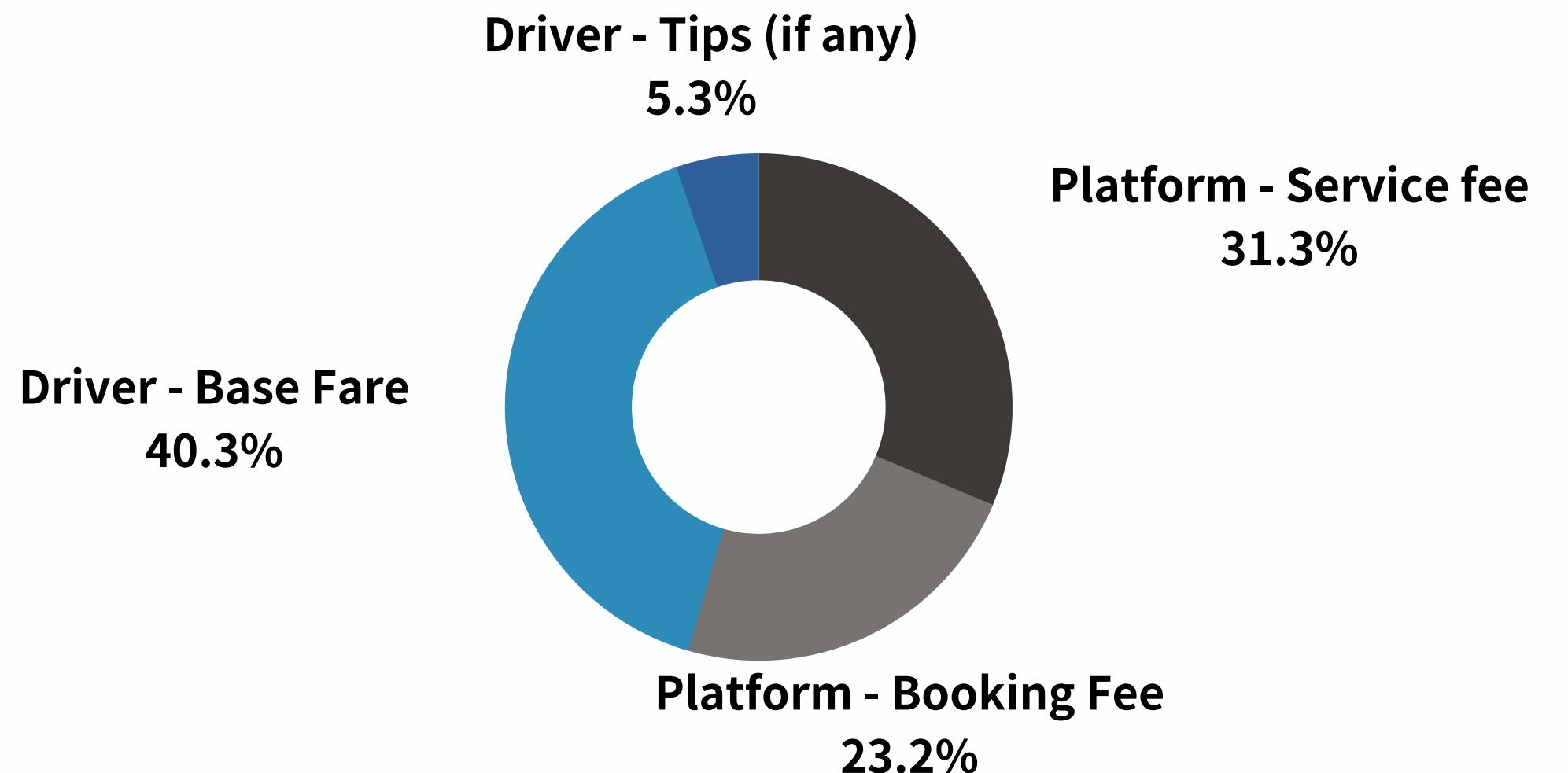
FHV (For-Hire Vehicles) Platform commission fee takes at least 25% for each ride and is still increasing



Uber/Lyft drivers' trouble: shrinking share of their pie

QUICKEST WAY TO INCREASE DRIVERS' PROFIT: TIPS

Predictable tips facilitate a better decision making process for driver



Negotiation with platforms takes time but tips are instant \$ at hand

OUR SOLUTION: TIPTOK

An app plug-in to help driver decide which ride worths more



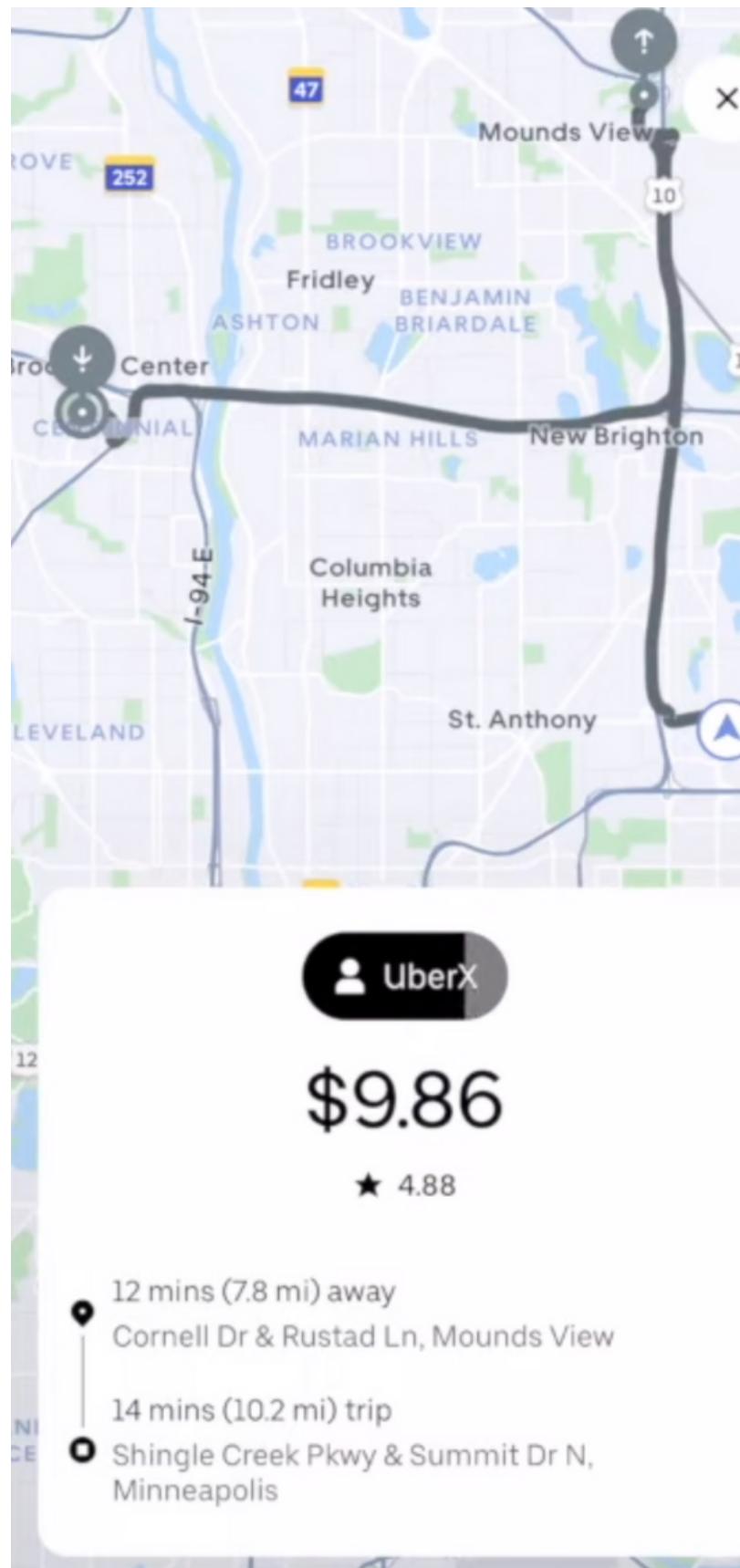
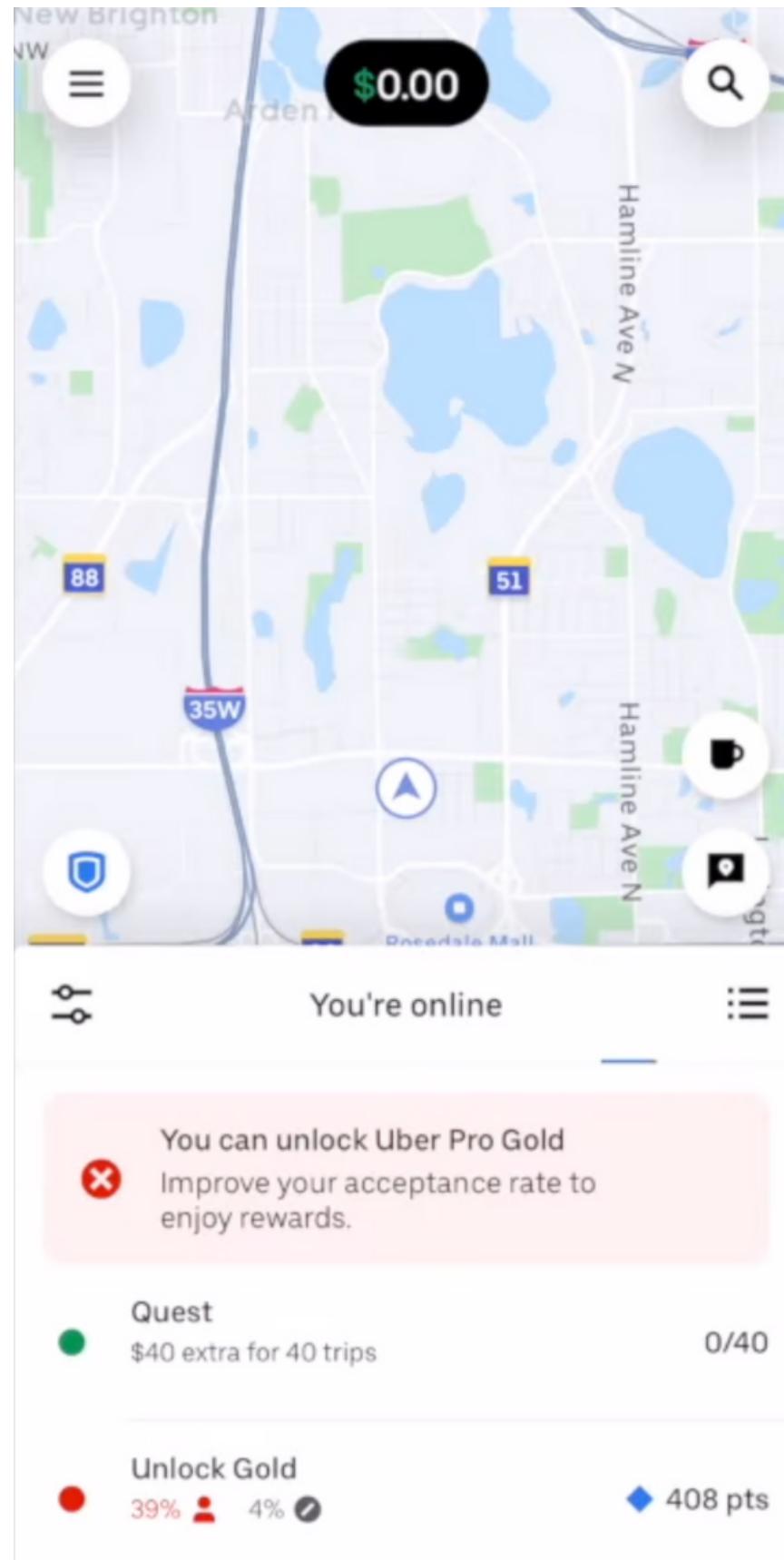
Your trip, your tip

**Help driver earn more
through machine learning**

UBER / LYFT APP

Driver Interface - Main Page

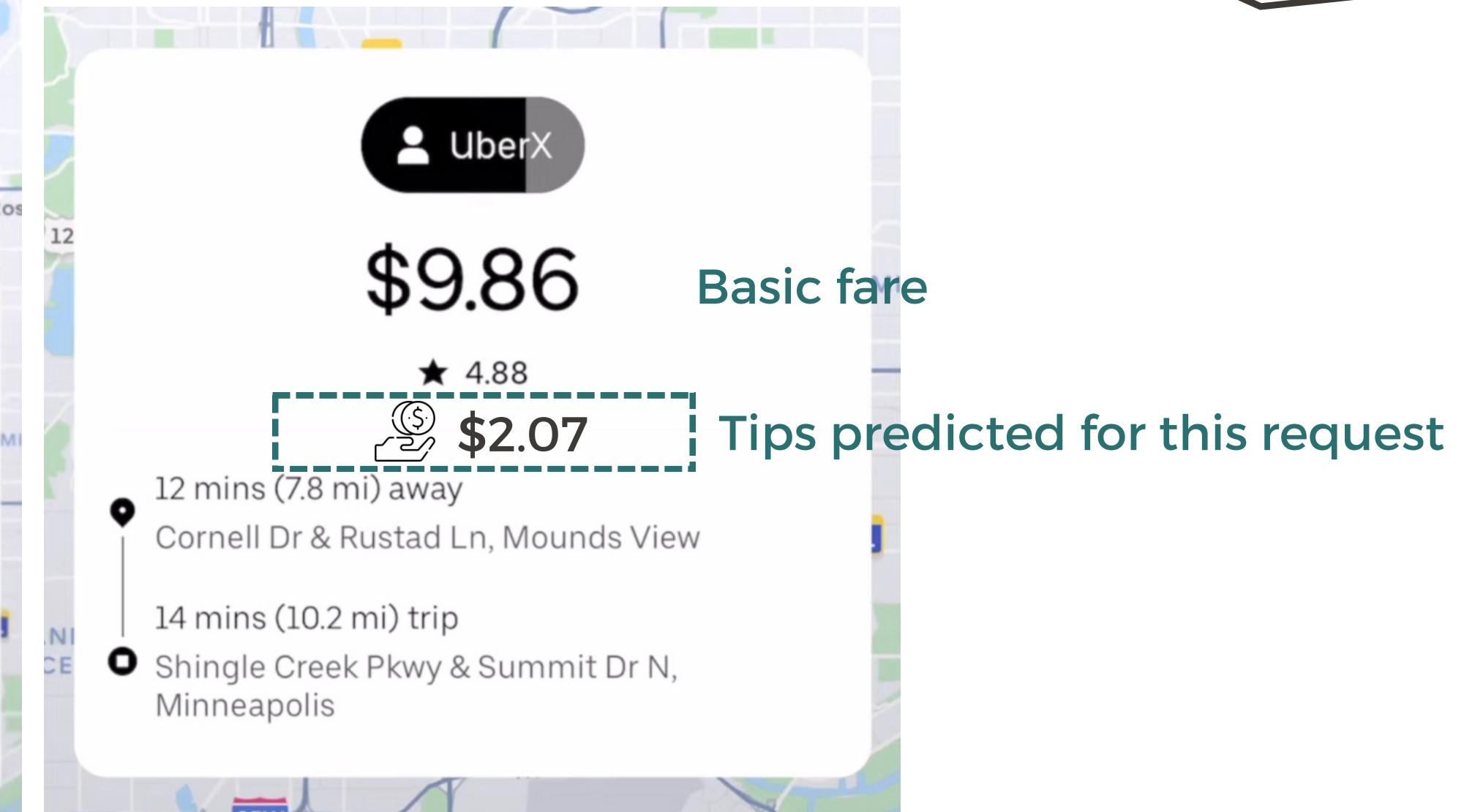
[Before] Getting a Request



PROTOTYPE

UI/UX DESIGN Before & After

[After] Getting a Request



PREDICTORS



We collected the features on FHV app driver's interface for NYC trip requests. The transactional data does not contain customer information.

- 1) Vehicle License
- 2) Trip Datetime
- 3) Route
- 4) Government Charges
- 5) Special Requests

Trip Datetime

- pickup_datetime
- request_datetime

Route

- PULocationID (pickup)
- DOLocationID (dropoff)
- trip_miles
- base_passenger_fare

Government Charges

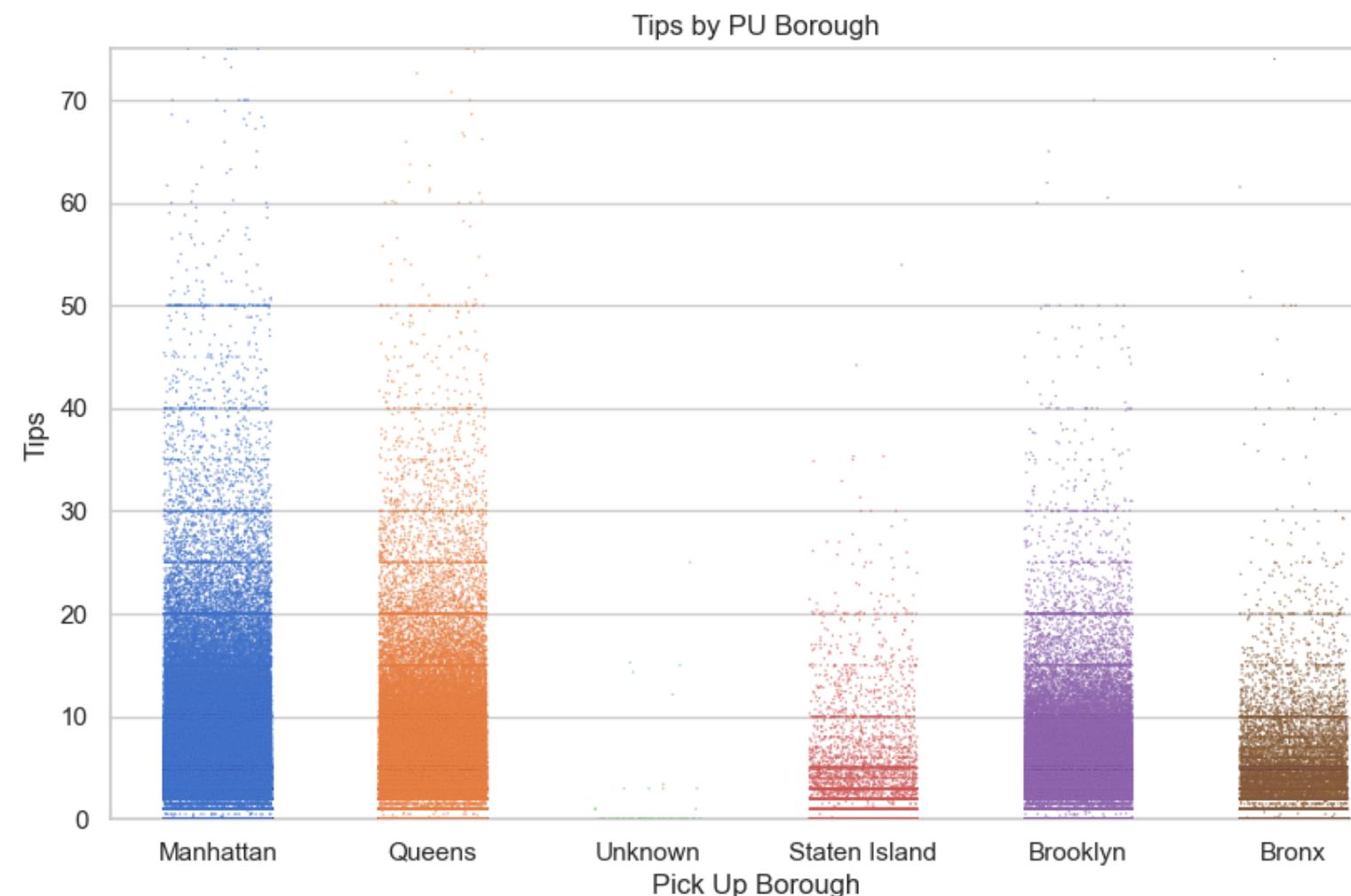
- tolls
- congestion_surcharge
- airport_fee

Special Requests

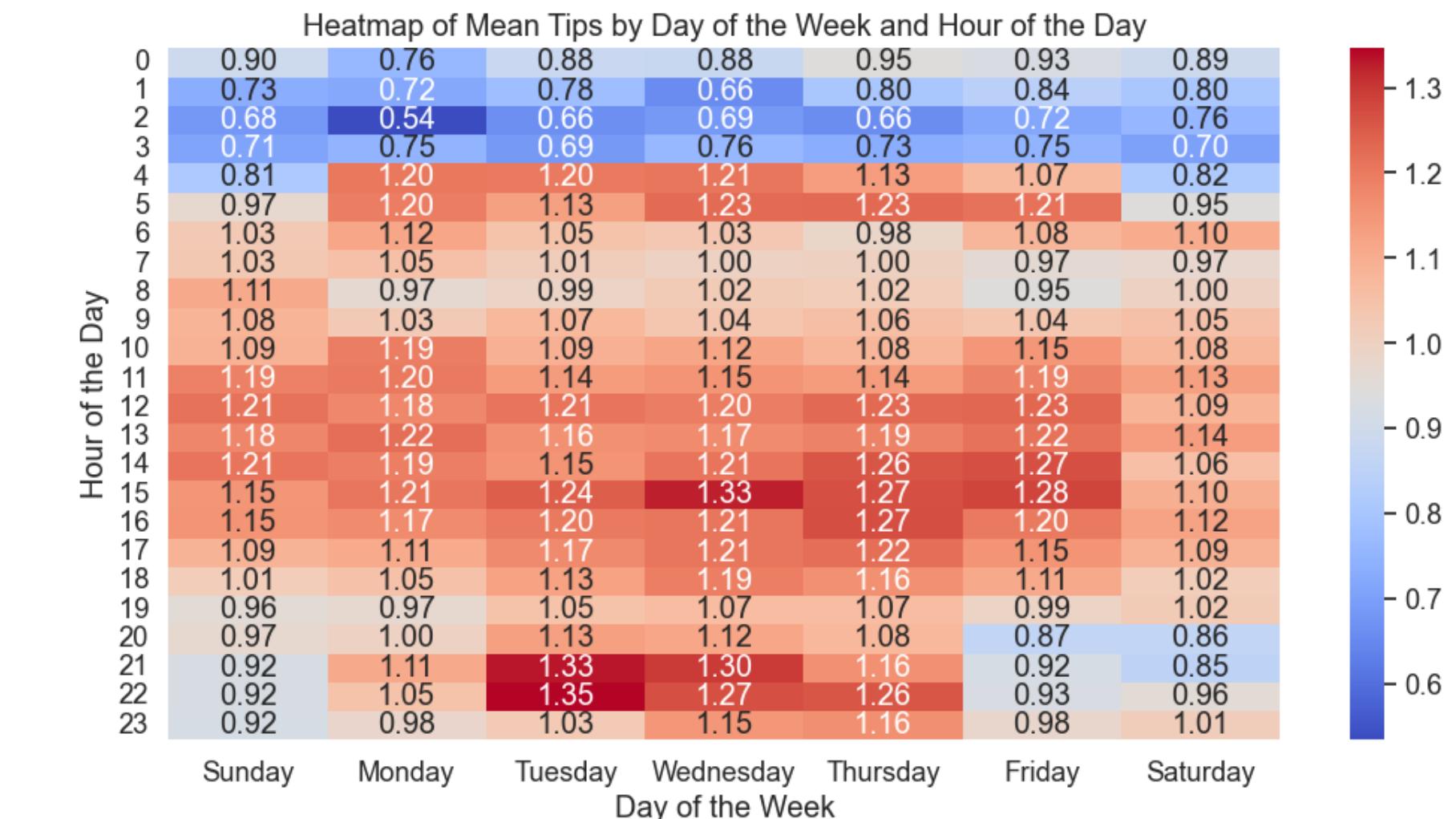
- shared_request_flag
- wav_request_flag
(WAV: wheelchair-accessible vehicle)

VARIABLE VISUALIZATION

Pickup Location vs. Tips



Pickup DateTime vs. Tips



Order by the average tips among boroughs,
Manhattanites tip higher

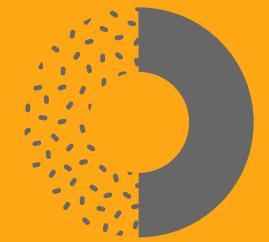
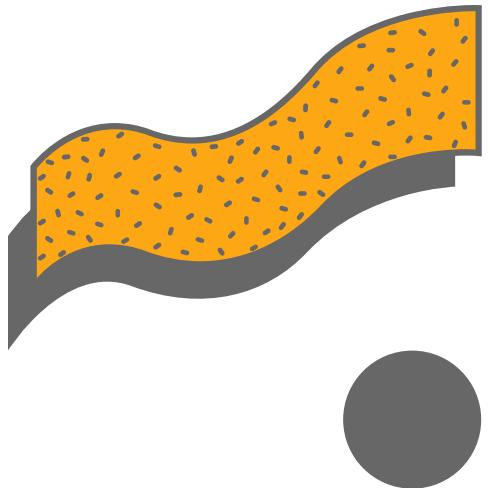
Highest tips on weekdays
evening rush hours

QUANTITATIVE MODEL SELECTION

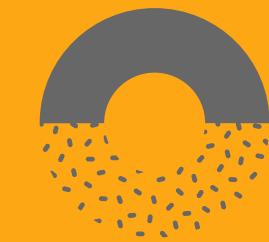
Model	RMSE	Model Rationale
Regression (Ridge)	2.805808	<ul style="list-style-type: none">- most common approach for predicting continuous variables- Ridge helps to shrink the coefficients towards zero because of multi-collinearity
Random Forest	2.799820	<ul style="list-style-type: none">- easy to explain non-linear relationships between predictors and the target variable- handle a mixture of continuous and categorical variables
XGBoost	2.783945	<ul style="list-style-type: none">- similar to RF, and- handle higher volume of data and deliver more accurate results than RF
GBM	2.789692	<ul style="list-style-type: none">- similar to XGBoost, but focusing on the errors made by the previous trees, whereas XGBoost builds the trees in parallel
Neural Network	2.788817	<ul style="list-style-type: none">- handle non-linear interactions between the predictors

Note: We trained our models on 2 million trips * 57 features

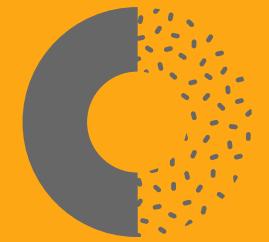
IMPACT TO DRIVERS



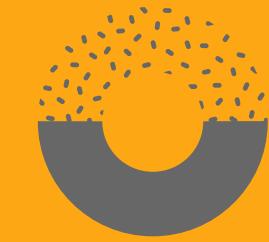
Prioritize high-earning requests



Motivate excellent customer service



Reduce income uncertainty



Increase job satisfaction



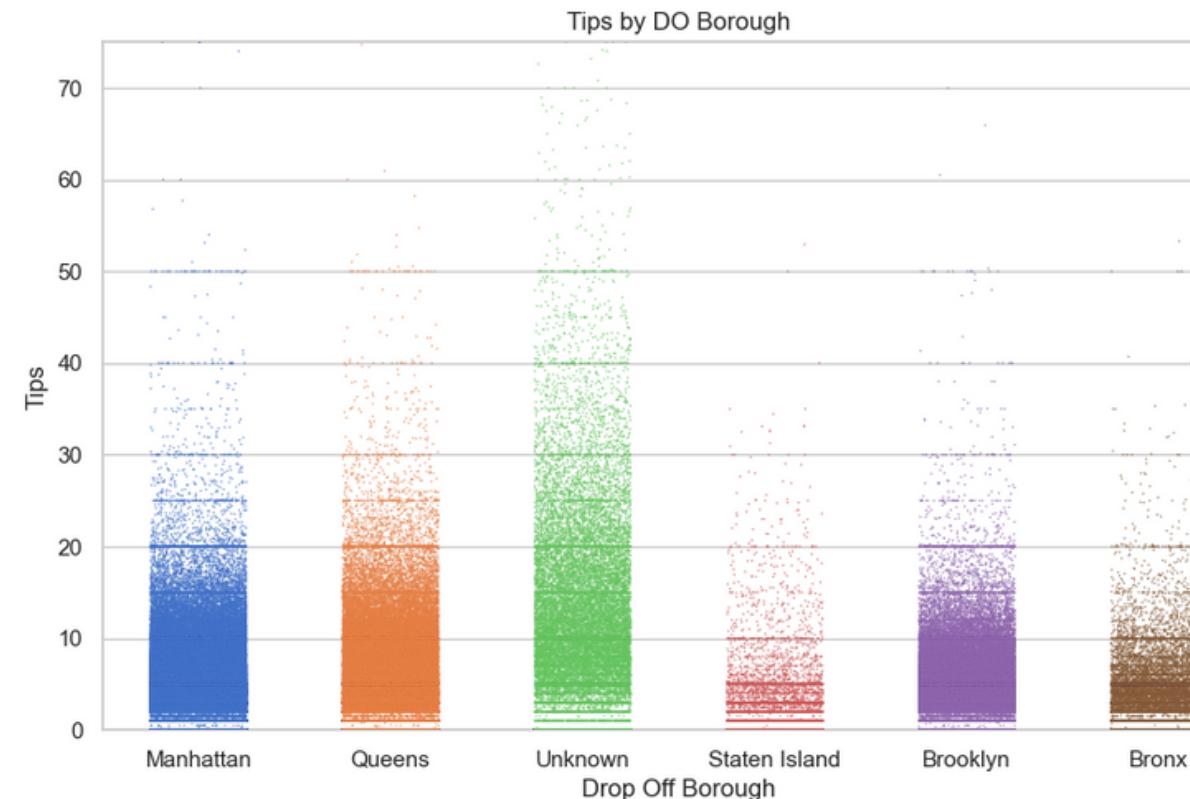
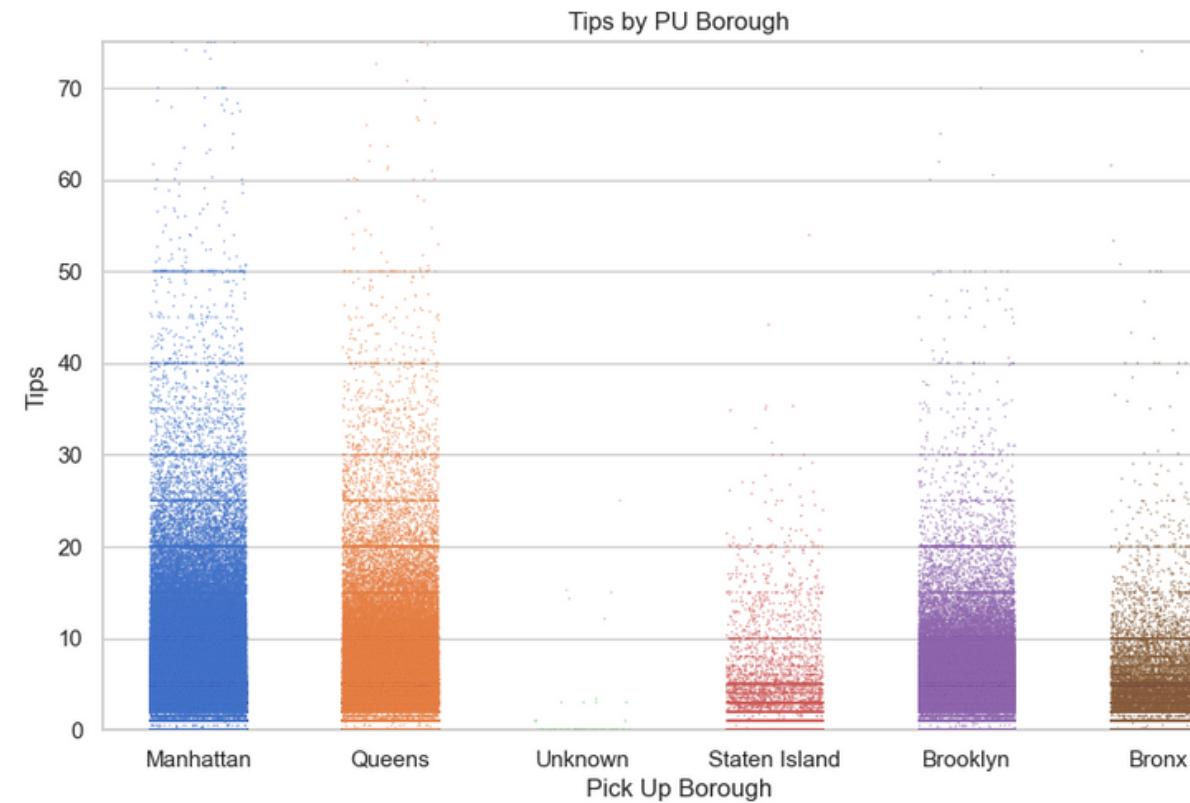
TIPTOK

THANK YOU!

WELCOME FOR FURTHER DISCUSSION

APPENDIX - VARIABLE VISUALIZATION II

Tips vs Locations



Tips vs Trip Volume

