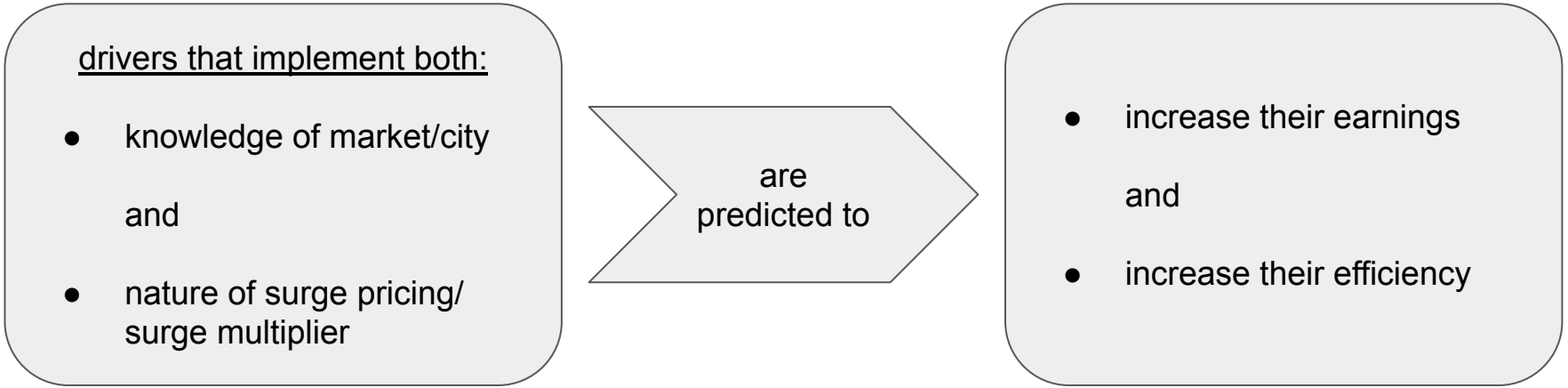
Capstone: Sprint 1

Garrett Ard

A non-technical overview of the subject area and the problem statement / opportunity you identified

Uber and Lyft drivers can gain from insights to increase earnings and efficiency.

An overview of your proposed vision for tackling the problem using Data Science



An estimate of the potential impact of such a solution

surge multipliers typically range from:

Uber

- 1.2x - 3x during busy times
- 3x - 5x during big events
- even > 5x in congested areas

Lyft

- 1.5x - 2.5x during busy times
- 3x - 4x during big events
- even > 4x in congested areas

Surge multipliers can vary significantly depending on demand and location.

These figures and driver earnings can fluctuate based on factors like events, weather, and time of day.

An introduction to the dataset, including data quality concerns and findings from preliminary EDA.

cab_rides.csv dataset

- 693071 rows and 11 columns
- measured for ~ 1 week in Nov.
- initially used to predict demand

weather.csv dataset

- 6276 rows and 8 columns
- support to measure weather's impact on demand

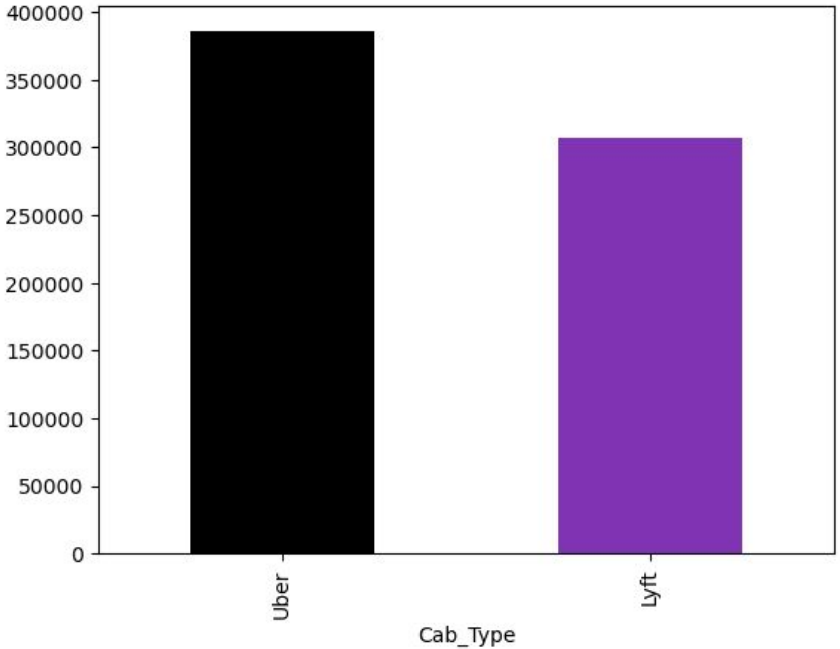
Data quality concerns

- missing values in `precipitation` and `price` columns
- unknown values in `product_id` column

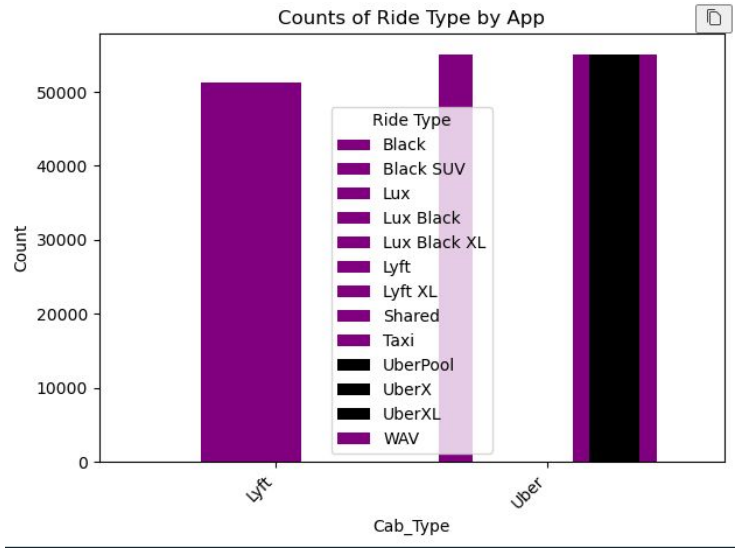
Findings from preliminary eda

- 56% Uber : 44% Lyft in this dataset
- UberXL ordered the most - 55,096

Cab_Type Count between Apps



Counts of Ride Type by App



Next steps in terms of data processing, feature engineering and baseline modeling.

Find optimal blend of variables between weather and cab_rides datasets that are likely to activate surge pricing to appear on each app's gps.

starting the analytics process with:

- time_of_day (weather.csv | cab_rides.csv)
- location (weather.csv | cab_rides.csv)
- impactful weather variables such as temp., rain, wind, humidity (weather.csv)