**Uber Vs Lyft, A Comparison.**

Adam Flory, Rory Berry, and Spencer Hurley.

**Dataset:**

We used Kaggle.com to find a large set of data collected on Uber and Lyft rides provided in Boston Massachusetts. The Data set we used is titled “Data Mining Project – Boston” uploaded by use SophiaLiu. This data set consisted of over 600,000 rides provided by the rideshare services over the course of 18 days in late November and early December 2018. The data set include the drop off and pick up points for each ride as well as the date, time, price, company and service used(ex. Uber Black), and various weather conditions(ex. Temperature, precipitation). Luckily the data set was already very clean so we only really had to narrow things down to the specific variables we wanted to use. The largest problem with the data set was that it was collected over such a short period of time. This made trying to track usage taking into account things like weather somewhat unfruitful, as it would be challenging to establish a verifiable pattern with such little data. This led us to look into other items most specifically the overall usage of each company and there services when taking things like time of day, type of service and pick-up/drop-off point.

**Adam’s Section:**

**A screenshot of a cell phone

Description automatically generatedA picture containing writing implement

Description automatically generated**

**Are rideshare services used more often to some areas of the city than others?**

Short answer yes, while the graph below does not appear to show a much difference between different destinations, running a Chi-Squared test rejects that each destination receives the same amount of traffic. To run the chi squared test I found how many rides ended in each destination, think the burrows of New York, to find the observed values. For the expected values I divided the total number of rides by the number of possible destinations, so every destination had the same number of rides. After running the Chi-Squared test which showed a p-value of .00129. This means that we can reasonably say some areas of Boston see more traffic from ride share services than other, which is not to surprising.

A screenshot of a cell phone

Description automatically generated

**Is the service used affected by the destination?**

Since the answer to the last question seemed pretty obvious we also decided to look at whether any areas of Boston see more than expected traffic from one ride share service or the other. in order to took into this we decided to again run a Chi-Squared test, however this time the observed values were calculated using the difference between the number of uber rides dropping off in a particular area and the Lyft rides dropping off in that area. The expected values where the difference between all Uber and Lyft rides in the data set divided by number of drop off points. This time the Chi-Squared test produced a value of .4004. which means that we can reasonably say that there is no difference in service preferences between different destinations.

A screenshot of a cell phone

Description automatically generated

**Rory’s Section:**

**A close up of a logo

Description automatically generatedA screenshot of a cell phone

Description automatically generated**