

Project3-Ridesharing Analysis(WriteUp)

Xufei Li

Abstract:

The goal for this project is to help Lyft increase market share - Uber has 68% market share while Lyft has 32% nationwide. Scope and design the project, brainstorm the possible reasons which cause the difference. My focus is on **demand** and **price**, and my analysis is mainly based on total revenue of “**pick up location**” and “**time of the day**”. I analyzed & visualized the price and rides count relationships of Uber and Lyft to get some insights by using Excel. Then, I use Tableau for further analysis by building an interactive dashboard to communicate my results.

Design:

- **Business Impact:** Help Lyft increase market share
- **Possible reasons:** Service & UI, Demand & Supply, Price
- **Data Science Solution Path:**

I am assuming price is the key factor, I want to build 2 regression models on Uber & Lyft dataset separately to predict price. Then identifying which locations and time of the day Lyft is overpriced by comparing the price with Uber. (Set a benchmark, eg: if Lyft's price is 10% over than Uber, it will be identified as overprice)

- **Impact Hypothesis:**
By finding out the difference in predicted price between Lyft & Uber, Lyft will be able to adjust “surge-multiplier” to provide a more competitive rate for each specific time slot & locations which result in a greater market share
- **Risk** : lyft drivers will get less profit each ride, they may leave for another platform eventually.
- **Cost:** we are reducing profit for each ride, if this doesn't bring in more rides, then we will lose this part of the revenue.

Data:

[Boston's dataset](#) cab_price data on Kaggle include 600,000+ data points.

I only selected one day's data(11/26/2018 Monday) which includes 31,587 rows due to the capacity of Excel.

Assumptions:

Assumption 1:

Lyft UI & Service are good, almost no user complains about it. Then demand and price may be the key factors of the difference.

Assumption 2:

For data cleaning, I am assuming the Uber's missing values is "missing at random", the price for Uber is right skewed, so I filled null with median of Uber's price(I didn't drop null is because I wanted to keep the ratio of the ride counts between lyft and Uber, and all missing values are

Uber's data)

Assumption 3:

I am assuming there's no difference in average distance of the rides between Uber & Lyft users for each location, then we can just compare the average final price.

Tools:

- Python Pandas for data cleaning(fill out null value, adding zip code)
- Excel for preliminary data analysis and visualization(found out the locations with a potential that Lyft can increase revenue)
- Tableau for further analysis with an interactive dashboard

Communication:

“Beacon Hill”(16:00 to midnight), “Haymarket Square”(14:00, 17:00, 19:00, 21:00), “South Station”(14:00, 18:00, 22:00), “West End”(9:00-12:00, 22:00) have potential to gain more revenue. However, “average final price” doesn't have a big difference between Uber & Lyft, the “rides counts” have a big impact. So, the best strategy may be to encourage drivers to work more at those areas and time with more demand, and then attract more drivers to join in Lyft in those locations.

Interactive dashboard screen recording -

<https://drive.google.com/file/d/1KkeB188PaN3Yzwp5b3LPjBmOV3p2mFyk/view>

