Project3-MVP (ridesharing analysis)

Q: Define a realistic business problem faced by your client.
A: Lyft's current market share is 32% while Uber is 68%.
Opportunity: Help Lyft increase market share. How can Lyft gain a competitive advantage to catch up with Uber?

Q: Develop a fully-scoped proposal for how data science methods can be used to drive impact towards the defined problem,

- **Impact hypothesis**

  A: Lyft is more expensive than Uber in general(hypothesis 1 - need to explore data), and lower the price per ride per hour will attract more customers in order to increase annual revenue. (hypothesis 2)

- **Solution paths**

  A:**Step 1: List and Assess Alternatives**

  Alternative 1: In order to increase annual revenue, instead of lowering the price, we can also provide a better UI or have promotion to gain more riders. (gain more customers)

  Alternative 2: Analysis the demand & supply of the Boston Market, to see if increasing driver numbers can increase more rides. If so, we can encourage Lyft current drivers to work longer hours. (stimulate current drivers)

  Alternative 3: For the drivers who're using both Uber & Lyft, we can stimulate drivers using Lyft more than Uber. Also, we can attracts more new drivers to join in Lyft by providing some benefit (get more new drivers)

  Step 2: Clarify the Output

  Goal: Bring in 10% of the increase in rides in order to increase the annual revenue by

  lower 5% of current price.

  Step 3: Consider the Solution Path Critically

  Risk: driver gets less profit per ride, they may be less motivated to work as a Lyft driver.

  Costs: We are reducing the price for each ride, if it can't bring in more rides/customers, we will lose this part of our revenue.

  Potential benefits and losses: increase market share; losing drivers and not make good profit per trip.

  Data availability: Uber & Lyft cab_rides.csv on Kaggle https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices The data is approx. for a week of Nov '18, for a few hot locations in Boston.

- **Measures of success, both technical and non-technical**

A: **Non-technical:**

Lyft has more riders than before and annual income does increase by decreasing unit price (or gain more drivers to gain more rides).

Technical:

(1)We can build a regression model to see the price difference between Uber and Lyft on the same distance.

Feature: ‘distance’, ‘time_of_the_day’, ‘price’, ‘cap_type’(Uber/Lyft), ‘destination’, ‘surge_multiplier’, ‘name_of_cap_type’

Target: ‘Price’

(2)We can start a survey for Lyft users as well as drivers, then clustering(Unsupervised ML) them into 4 groups based on loyalty and satisfaction, then find the specific strategy for each group to move them to a high loyalty high satisfaction group.

(3)After knowing price is a factor, take market pairs based on similarity as control and test, then adjust price for a couple locations in the test group, and then compare with their control group to see if we are indeed getting more income/rides. (Test by market)

Risks and assumptions

A:Trade off: getting more rides to increase revenue: we want customers to pay less money, at the same time want the driver to be happy with their income.

Q: Tools:

Python Pandas - clean data, separate data to 2 sheets, generate random 100,000 for each sheet.
Spreadsheet - EDA
Tableau - Visualization

Q: MVP: Preliminary Analysis - Be able to find the difference on price between Uber & Lyft.