**PyBer Analysis**

**Background**

The ride sharing bonanza continues! Seeing the success of notable players like Uber and Lyft, I decided to join PyBer, a fledgling ride sharing company as Chief Data Strategist. In this capacity, I was provided the data on PyBer’s ridesharing for 2018 to analyze and offer data-backed guidance on new opportunities for market differentiation.

To perform my analysis, I was given access to Pyber’s complete record-set of rides for 2016. This data contains information about every active driver and historic ride, including details on: city, driver count, city type, individual fares, date of ride and the ride ID.

**Methods:**

To perform the analysis the following tools were used: Pandas Library for manipulation and analysis of the data; Matplotlib Library for plotting the results of the data analysis; Numpy Library for support of multi-dimensional arrays, matrices and the collection of high-level mathematical functions to operate on these arrays; Jupyter Notebook an integrated development environment (IDE) to write the codes and for data visualizations. Python program was used to perform (run) the overall analysis.

The Pyber rideshare data was contained on two csv files named city\_data and ride\_data. These files were read into Pandas using the Jupyter Notebook as two separate dataframes. These dataframes were then merged on city creating a new dataframe called merged. The merged dataframe was used to build a bubble plot that showcase the relationship between the these four key variables: average fare in dollars per city; total number of rides per city; total number of drivers per city and the city type (Urban, Suburban and Rural). Additionally three pie charts depicting the percentage total by city type of fares, rides and drivers were done to provide more visualization of the analyzed data.

**Results and Observations:**

Based on the analysis of Pyber’s ridesharing data the following results and observations were noted:

* Based on the Bubble Plot, Urban cities have the highest number of rides, the highest number of drivers and the lowest average fare as compared to Suburban and Rural cities.
* The results as shown by the first pie chart based on the analysis of the percentage of total fares by city type showed that 62.7% of total fares were generated in the Urban cities to 30.5% in Suburban cities and 6.8% in Rural cities.
* The results as depicted by the second pie chart showing the percentage of total rides by city types showed that Urban cities accounted for 68.4% of all rides which is more than double the combined percentage of Suburban 26.3% and Rural 5.3% .
* The final pie chart showed the results from analyzing the percentage of total drivers by city type. Urban cities accounted for 80.9% of drivers as compared to 16.5% in Suburban cities and 2.6% in Rural cities.

From the results noted above it appears that Urban cities account for the highest percentages of total fare, total rides and total number of drivers. However, it was also noted that the rural cities accounted for some of the highest average fares per city but the smallest segment of total fares. This could be attributed to the assumptions that Urban cities have larger population, thus more drivers and rides and shorter distances travel per ride but a higher frequency of rides resulting in lower fares.

In reviewing the data and performing the analytics these limitations were noted. The data is only subset of the population of drivers in that it only includes active drivers. The average distance per ride and or the duration of each ride were not included in the data set provided. There was no information on the overall population, migration trends (are more people moving to Urban cities or away from these cities) and the medium household income. The data was only over a limited time period (one year, 2018) and as such no analysis of trends over time was done. No information on the competition or availability of complimentary services such as public transportation and taxis.

From the analysis, these opportunities were noted. Suburban cities appears to offer the best opportunities in that the 16.5% of drivers provide 26.3% of all rides and account for 30.5% of total fare. Also, in general, Suburban cities are in close proximity to Urban cities and such allows for an easier shift or transfer of drivers from Urban cities to Suburban cities. Another opportunity that will also allow for some market differentiation is providing ride services between the respective city types.

**Conclusion**

Based on the set of data provided, methods used, results and limitations observed and opportunities noted, Pyber’s best opportunity is to increase its operation in Suburban cities by shifting drivers from Urban cities. Pyber can also differentiate itself by offering specialized services between city types: Urban and Suburban; Suburban and Rural; Rural and Urban cities.