Crescenzo Adriano

Thursday, June 13, 2019

pyber\_starter

An analysis was conducted as the Chief Data Strategist for Pyber. The purpose of this analysis is to provide data supported guidance on new opportunities for the purposes of variation in market exposure.

The source data is constructed from the company wide record set of rides. The extent of detail includes record of every driver, ride history with details relating to the city, the number of drivers, the cost of individual fares and the type of city.

The first analysis conducted reviewed a bubble plot distribution, noting the x axis based on the average cost of each ride fare, the y axis being the total number of rides per city and the bubbles being represented in size by the number of drivers per city. Each individual bubble represents a city, with the gold bubbles being urban cities, the sky blue bubbles being rural cities and the coral bubbles being suburban cities. The findings of the bubble plot revealed a sparse distribution of drivers in rural cities, transitioning more towards a moderate distribution of drivers in suburban cities concentrated between $30 to $35 fares and 15 to 20 total rides per city, and a dense population distribution of drivers in urban cities with $20 to $30 fares with 15 to 35 rides per city, noting two select outliers near 40 fares.

Next a pie chart analysis was performed based on the percentage of total fares by each city type. The distribution found the majority being urban at over 62%, with suburban consisting of over 30% and rural at just under 7%. An additional pie chart analysis was conducted based on the percentage of total rides by each city type, noting the urban population again taking the majority with over 68%, suburban consisting of over 26% and rural cities taking the remainder just over 5% of number of rides per city type. Lastly a pie chart distribution was made for the total number of drivers per each city type, again reflecting urban city with the most at just under 81%, suburban at over 16% and rural at over 2% of the population of drivers.

In conclusion, the findings reveal an expected distribution of the greatest number of drivers, number of rides and fares collected being in the urban cities, with the suburban cities arriving at the middle of the distribution and rural with the least. A more accurate analysis may be conducted by utilizing a calculation for normalizing among the city types by dividing the values found on a per capita for each individual city. Consequently it is recommended to obtain the population size of each city for future analysis.