Improving Bzzt Podtaxi Operations in Sweden: How Machine Learning Can Anticipate Rider Demand.

Alexis Casas
Supervisor: Shivani Shukla
University of San Francisco
Business Analytics Department
aicasas@dons.usfca.edu

Abstract

This analysis uses K-Means, unsupervised learning algorithm to identify different boroughs within Stockholm, Sweden.

Introduction

Bzzt PodTaxi is a clean energy transportation service located in Stockholm, Sweden. Their low emission vehicles and podtaxis half the size of normal cars aim to reduce carbon emissions and congestion in cities. This report analyzes the pick up and drop off locations provided by Bzzt in order to optimize Podtaxi pick ups. Anticipating rider demand is an important aspect of gaining user popularity. To do this, K-Means cluster analysis is applied to spatial coordinate locations. This unsupervised machine learning algorithm separates these coordinates into boroughs based on different 'hot spots' throughout Stockholm.

Data Cleansing

The data set used comes directly form Bzzt Podtaxi's records. In order to apply K-Means, the data set of 27,000 rows had to be cleansed. First, the column variables needed to be separated and named. Date and times of each Podtaxi pick up was prepared for analysis. The particularly important columns for identifying pick up and drop off boroughs were the columns regarding origin and destination points. These address columns, neighborhoods, and zip codes were then geocoded into pick up and drop off longitudes and latitudes using a Google Maps API. Communication to the Google Maps server was time consuming. In order to overcome this challenge, only the first three months of data was geocoded. Using three months created a large enough sample size to generate significant results (17,000 rows). Finally, the two separate longitude and latitude coordinates for the pick up and drop off locations were ready to undergo the K-Mean cluster analysis.

Analysis

The frequency of rides per weekday is plotted as a histogram in Figure 1. It is clear that Friday and Saturday tend to be Bzzt Podtaxi's busiest hours. Other city transportation data suggests individuals tend to use transportation services during peak commuter hours to and from work. To determine whether this is true for Bzzt a histogram was generated based on hour of the day. Figure 2 The maximum points from the density result suggests that Podtaxi users are in fact taking advantage of the service at peak commuter times.

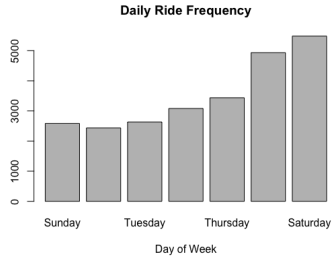

Figure 1: Frequency of Rides by Day

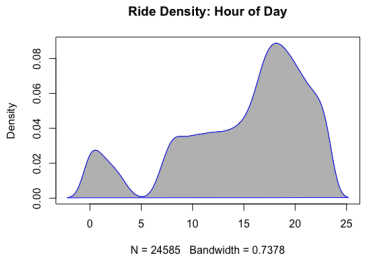

Figure 2: Rides per Hour of Day

Next, the spatial coordinates for pick up and drop off locations were analyzed using the Elbow Method to determine the optimal amount of clusters to be used. Both pick up and drop off coordinates were decided to have an optimal number of 2 clusters Figures 34. The number of clusters were then used to for the K-Means, nearest neighbor method to separate Stockhom into boroughs based on the number of

rides taken to or from that coordinate. Figures 56. illustrates the resulting separation of clusters.

The cluster analysis showed a clear separation between pick up and drop off spatial coordinates. Figure 7. illustrates the results plotted against a street view on google maps for boundary clarity. The boroughs both had about the same number of coordinates for both pick up and drop off locations; However, borough 2 ranged from 80-400 more points than borough 1.

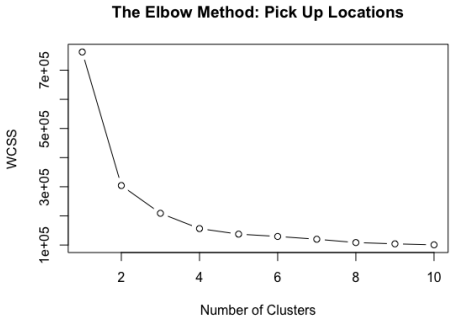

Figure 3: Optimal Pick Up Clusters

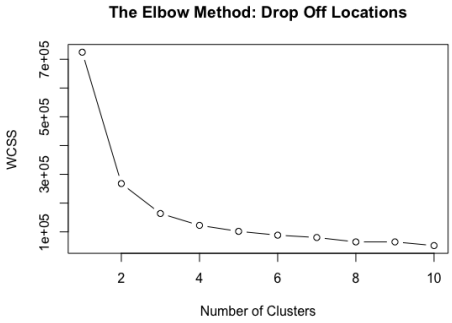

Figure 4: Optimal Drop Off Clusters

Conclusion

Over the next few years, Bzzt Podtaxi aims to globally expand their services. In order to overcome competitors like Uber and Lyft, Bzzt needs to anticipate user pick up and drop off locations to create a quick and efficient interface with available Podtaxis. Uber and other transpiration services frequently use K-Means to optimize the use of their vehicles. Setting up the vehicles in the two boroughs shown in figure 3 would help optimize and adjust their estimated cost, arrival time, and other important variables.

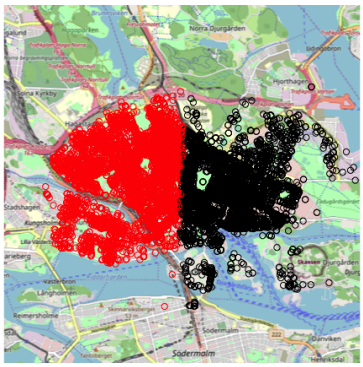

Figure 5: K-Means: Pick Up

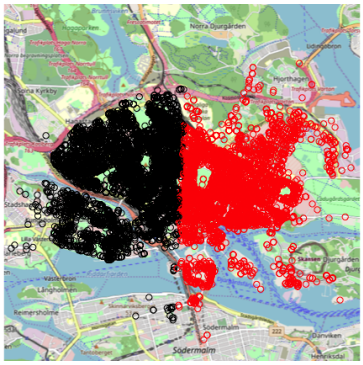

Figure 6: K-Means: Drop Off

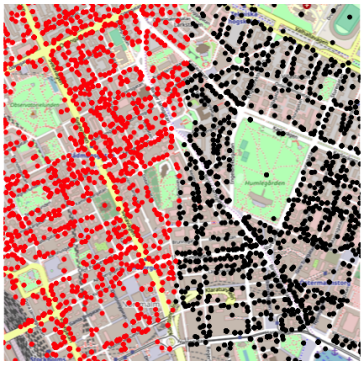

Figure 7: Borough Separation