Peer-graded Assignment: Capstone Project

The Battle of Neighborhoods

Problem Description

Opening a business can be tricky and investors often make the wrong choice before they even start, by choosing the wrong location. In this project, I will attempt to explore the effects that population size, crime and the existence of other business types have on the success or failure and distribution of businesses in three cities.

I believe it may be possible to see a relationship between these three points which can be used to patch 'holes' in cities, where a business can succeed but there is none there now. This information can be useful to investors and entrepreneurs.

Data

For this project, I will have gathered three data sets for the cities of New York, Chicago, and Toronto for the year 2019. I will also use foursquare to acquire venue data for each city.

1. Population data by district/borough for each city

I extracted this data from different census websites, as well as Wikipedia. I did the best I could to ensure the consistency of the estimates provided by these sites.

2. Crime information for each city

Most major cities have police departments that compile all of the major crime cases that occur in their city on a yearly basis. It was not difficult to acquire crime statistics from these sites. It was surprisingly more difficult to find the population data.

3. The distribution and number of venues within these cities.

For this data, I will use the foursquare API to collect location data on venues in each city. This should provide a good estimate of the geographic distribution and frequency of venues within each district.

Approach

I will use the population and crime data to plot a map that shows the relative occurrence of crime within each city. I will then spread the various venue types over the map and attempt to discover any correlations or causation. I will also use regression to investigate any possible relationships that might exist between the three features.

Questions

1. Does the rate of crime in an area affect the success or availability of various business types, between New York, California, and Toronto?

2. Does the existence of particular business types affect the existence of others?

3. Can any patterns or groupings be found among the three features?

4. Is it possible to determine areas that should have a business operating there and the only reason it doesn't is an oversight?

Conclusion

From this exploration, I expect we should have a good idea of the features surrounding why some businesses are located where they are, both the city of choice and their location within that city. Of course, the accuracy would be significantly higher with a more detailed and varied data set.

Thanks for reading!