|  |
| --- |
| Finding a New Barbershop location using Data Science |
| Coursera Capstone Project |
| Nicolas Conley  June 2020 |

# Introduction

In this report there will be presented the methodology used to find a location for a new Barbershop in the USA. To do so, a market study will be made to better understand and select the most profitable areas in the country. The main target of this study is people interested in opening their own business, but do not know where to place it. Therefore, it will present information and a list of the top 5 places to open such business, in the author’s opinion.

According to a Forbes article from mid 2017, barbershops are back in trend:

"While soft goods retailers are shuttering stores in record numbers, a cultural resurgence of men's grooming, estimated to reach $26B by 2020, has barbers, brokers and landlords lining up to get their cut."

In addition, an article from 2017 by US Census Bureau claims that there is an opportunity in this kind of business, according to de article



Figure 1 - Interior of the Fellow Barber in San Francisco.

"While more than half of US counties (2,400) have no barber shops, there are only 802 with no beauty salon. Men in these counties probably go to the local beauty salon to get their hair cut".

This shows that even with a growing trend, the country still has a major lack of stores available for their customers. This statement will be tested and analyzed in the following discussions.

# Data

All data will be obtain from the US Census Bureau, DataUSA, FRED and Foursquare. Since there is very little information, many assumptions will be justified later.

The plot below shows the total revenue for barbershops between the year 1999 to 2018 (data extracted from fred.stlouisfed.org).

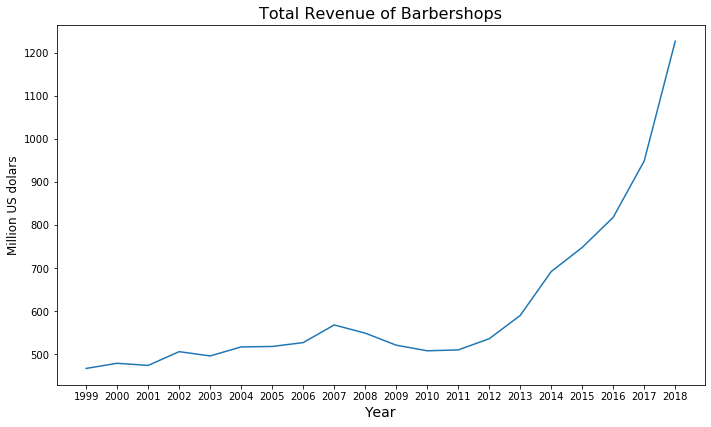
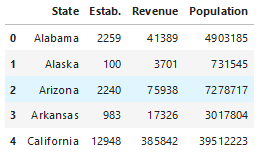


Figure 2 - Total annual revenue for Barbershops in the US

From the Revenue Graph above, a clear revenue increase can be observed starting in the year 2013. This is the main argument for starting a new barbershop.

The population distribution, the number of independents barbers or nonemployer establishments and the annual revenue for those establishments for each state (as shown in the Table 1 below) will be used to find the best location based. All this data was obtained from the US Census Bureau.

Table 1 - First 5 rows of the Census Bureau Data



Foursquare will be used to find the competition near the areas selected, using the API call to retrieve the venues of the same category: Barbershop.

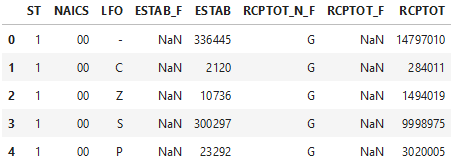
# Methodology

## Statistics

The search begins with exploring the data available, namely the US population and the nonemployer establishments, retrived from the US Census Bureau.

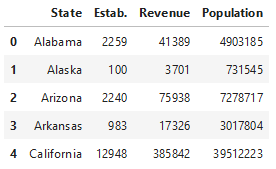
The dataset obtain has a lot of unnecessary and repeated information, like the Table 2 below. That being said, it was cleaned to only show relevant information, like the state number (**ST**), the **NAICS** showing only barbershops (code 812111), nonemployer establishments (**ESTAB**) and their annual receipt in thousands of dollars (**RCPTOT**). Also, the **LFO** column should only show the ‘-‘ lines, because according to the US Census data record layout, it represents “all establishments”, meaning that there will not be any exception regarding the type of organization. This dataframe will be called ***nonemp***.

Table 2 – Nonemployer Establishment Dataset by the US Census Bureau



After the cleaning the table will still be missing the actual States names, they were added by merging the nonemp dataframe with the states dataframe, that shows all states names and their FIPS code, then it looks like Table 3.

**Table 3 - Final result of the 'nonemp' dataframe**



To get a better view of the population data, a Choropleth map was needed, using the US geotypical location and the Folium library.

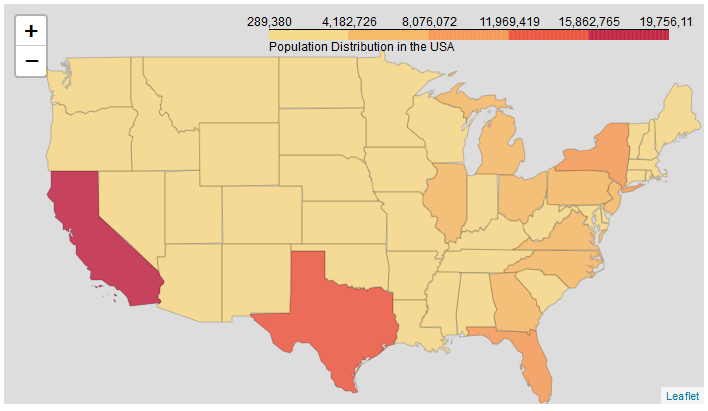


Figure 3 - Choropleth map showing the Population Distribution in the USA

Let’s check if there is an actual relation between the number of establishments and the revenue.

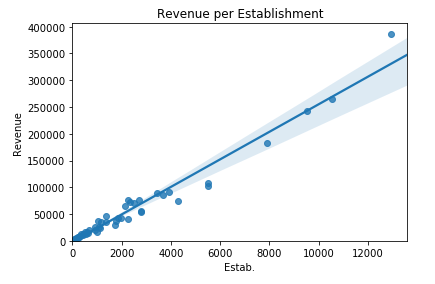


Figure 4 - Revenue per Establishments Regression Plot

There is a clear relationship between those two variables, meaning that even with a great number of barbers, there is also a good number of clients. Also the revenue distribution map is important, to see where the population get more haircuts.

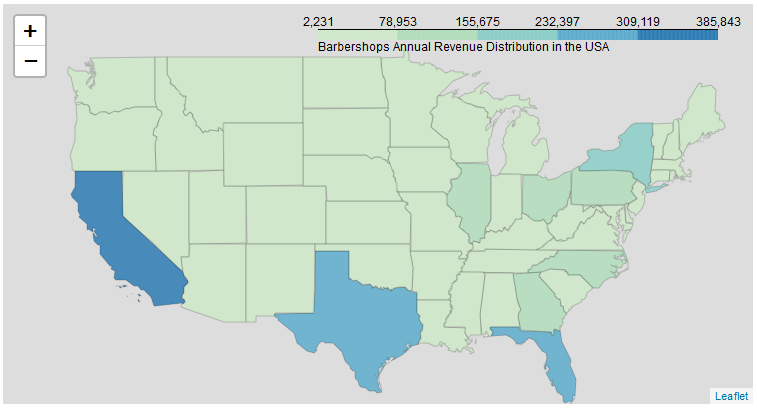
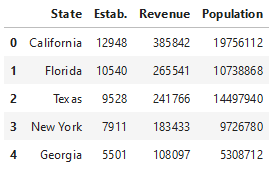


Figure 5 - Choropleth map of the annual Revenue Distribution in the USA in thousands of dollars

There are 13 states that pop out more, since the main objective of this report is to find the best 5 location, only the Top 5 will be selected. This new dataframe will be called Top5. In addition, since the main focus is the male costumer, the population column should represent the male population. According to the US Census Bureau, nearly 50% of the US population is male so, in order to be conservative, the population will be cut in half.

Table 4 – ‘Top5’ dataframe with male population



To simplify the analysis, only one city was selected from each state, being them:

* Los Angeles, CA.
* Miami, FL.
* New York City, NY.
* Dallas, TX.
* Atlanta, GA.

So there are the cities that will be studied, and the ones that will be referring to when their state in named later.

## Foursquare

After analyzing the states and defining the city for each one, a Foursquare search was made.

For every search, the max number of venues retrieved was set to 20 and the radius to 5000 meters. After receiving the results and formatting them, each location get a different dataframe with the names of the venues and their geographical coordinates, as the one below.

Table 5 – First 5 venues in Los Angeles



One problem found is that Foursquare does not distinguishes Barbershops from Beauty Salons, and since there is no way to know which is which, a manual selection may be needed. For the California venues dataframe, the venues where selected by the name that most likely be for a barbershop and pointed in red.

Table 6 – Barbershops in Los Angeles, CA



# Results

As shown in the Table 6, there are 8 barbershops for every 12 beauty salons in Los Angeles. The new dataframe was name BarberCA for the California Barbershops, as the table 7 below. This method was done for each location resulting in the Tables 7 to 11.

Table 7 - 'BarberCA' containing only barbershops and their coordinates

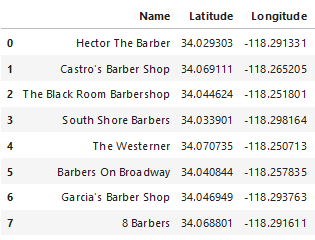


Table 8 - 'BarberFL' containing only barbershops and their coordinates



Table 9 - 'BarberNY' containing only barbershops and their coordinates

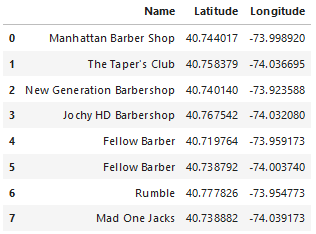
****

Table 10 - 'BarberTX' containing only barbershops and their coordinates

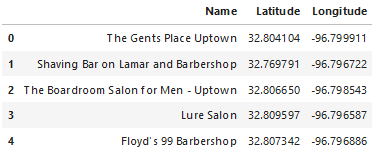
****

Table 11 - 'BarberGA' containing only barbershops and their coordinates

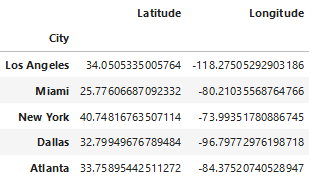
****

## Cluster

To find the best location, a cluster method to find the center of the barbershops group in each city was programmed using the K-Mean method with 5 iterations.

The coordinates of the new barbershops, obtained from the code, are in the Table 12 below.

Table 12 - New Barebershop location for each City



# Discussion

Next step is to show those venues in the map to see the proper location of the new Barbershop. The location of the competition Barbershops are in blue and selected location for the new Barbershop is shown in red.

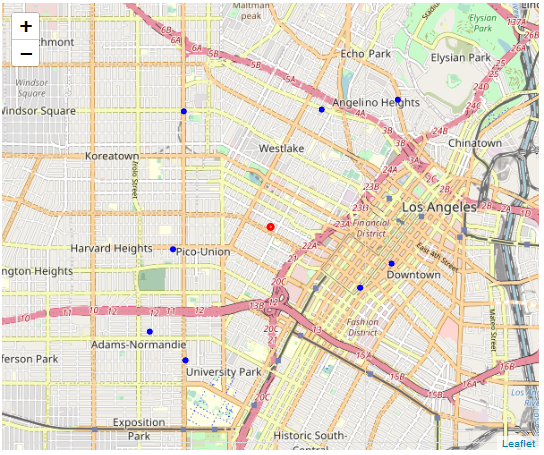


Figure 6 - Los Angeles venues with the selected location

The Barbershop in Los Angeles has a really good location, fairly distanced from the other barbershops, yet in the center, so can get customers from all directions.

The address to this location is: 953 Beacon Avenue, Westlake, Los Angeles

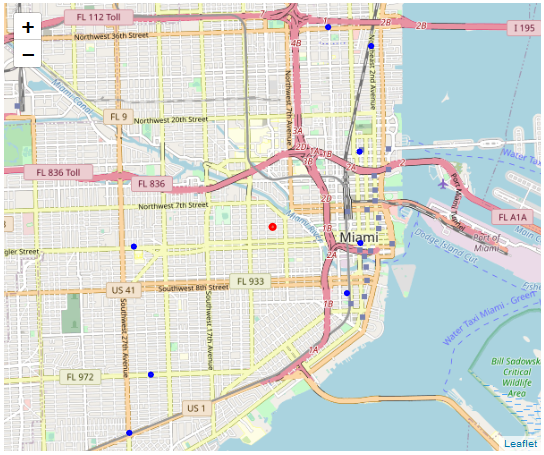


Figure 7 - Miami venues with the selected location

As well as the location in Los Angeles, this has a very good position. The address of this location is: 971 Northwest 2nd Street, Little Havana, Latin Quarter, Miami

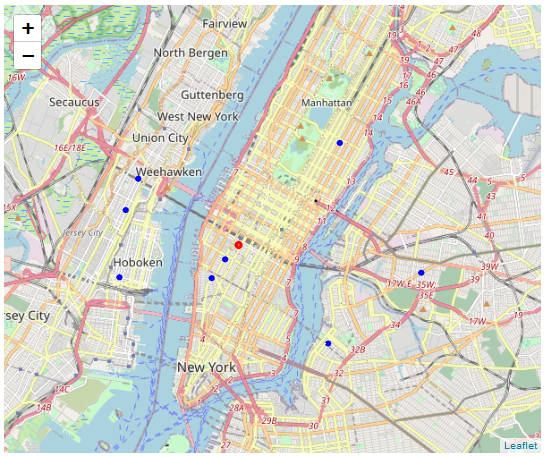


Figure 8 – New York venues with the selected location

This location is not as good as the two before. It can be improved by locating it to the south, but the Manhattan area is good place to start.

The address of this location is: 215 West 29th Street, Manhattan, New York.

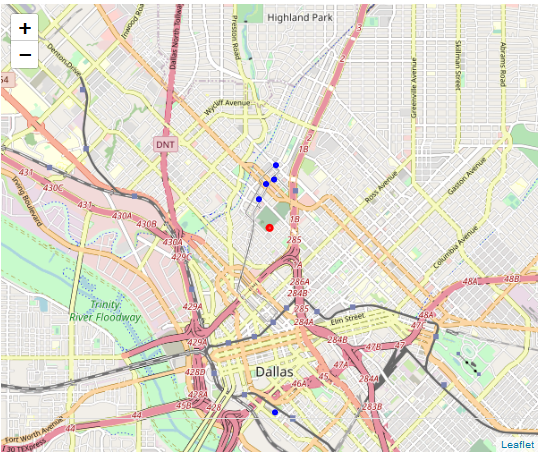


Figure 9 - Dallas venues with the selected location

Just like the New York location, this one it is not the best place, because its proximity to the rest of the barbershops. To improve it, it would be better to place it more to the south.

The address to this location is: 6322 Royal Crest Drive, Meaders, Dallas, Texas

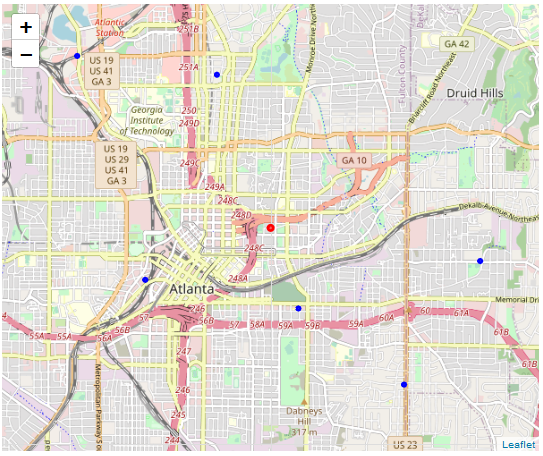


Figure 10 - Atlanta venues with the selected location

This one is very good located, again, fairly distanced from the other barbershops, but not isolated from the town.

The address to this location is: I 20 Stratford, Carey Park, Atlanta

# Conclusion

This method to place 5 Barbershops around the country, was successful. The Shops were places in the Top 5 most populated states of the US, to catch the larger number of customers as possible.

After getting the Top 5 states (California, Florida, New York, Texas and Georgia), and getting the venues around the very center of the selected cities, a K-means method was used to find the center of the group of shops, this center is the author’s suggestion for the location for the new Barbershop. All addresses were mentioned above in the best location given the data obtained in the beginning of the study.