# Determining Possible Locations for a Premium Barbershops in San Antonio, TX

by

## Olin Kennedy

**1. Introduction**

**1.1 Background**

Premium Men’s Barbershops are a recent trend in the United States and their popularity is exploding. These barbershops are characterized by tasteful and themed décor, a lengthier list of barber services such as straight razor shaves, massages, beard grooming, and often these places serve complimentary beverages, alcoholic or otherwise. From a business perspective, they have taken a low-margin business and increased margins and profitability by creating and serving a premium niche in the barbershop marketplace.

San Antonio, Texas is the 2nd largest city in Texas. The city is economically vibrant and growing, pulling in migrants from around the United States. Texas is one of the fastest growing states, both economically and population wise, in the United States and San Antonio is well placed to take advantage of both trends.

Additionally, San Antonio is typical in spatial arrangement for a US city, in that the city and its residents rely on personally owned vehicles to get from home to work, places of play, etc. The historic city center is ‘conveniently’ walkable, but for everywhere else, residents must drive or take a bus to get to where they need to go. A vast majority or residents live in the suburbs of San Antonio.

**1.2 Problem**

Premium Barbershops are no longer a secret phenomenon and successful first-movers in this arena have been able to expand into multiple locations or franchise out their business model and brand. However, there is reason to believe that the market for premium barbershops is still underserved and, that long-term demand for these services is still growing.

So, given the presence of competition, where should an entrepreneur locate his new premium barbershop? Or where should an existing premium barbershop owner choose to open a new location? Since this space in the marketplace is no longer in its infancy, all the obvious places to locate a premium barbershop are likely already served. Can we identify favorable locations using data?

**1.3 Interest**

The obvious business interest in this problem would come from businessmen or entrepreneurs looking to open a new premium barbershop in San Antonio, Texas.

**2. Data Acquisition and Cleaning**

**2.1 Data Sources**

In order to identify the best places to locate a premium business, we are looking for economic and social data tied to a location. Additionally, we are looking for this high-quality data to be packaged in the smallest and most discrete form available. Luckily, the US Census collects detailed data down to the ZIP code level. ZIP codes are the smallest unit of area data for which there is availability. For example, ZIP codes are smaller than counties and cities, in the context of US political organization.

Zip-codes.com is a data packaging service which is easier to use than the US Census Bureau website and pulls all their data from the US Census Bureau. They are also free, so I used them to pull demographic data for all the zip codes of San Antonio, Texas.

To determine which zip codes fall within the political boundaries of San Antonio, I did a Google search and then saved the table from the website ‘zip-codes.com’.

Additionally, for mapping, I needed to find geographical boundaries of each ZIP code. I found a geoJSON of all the zip codes and their shapes/boundaries on GitHub, posted by user ‘enactdev’ and this was the basis for mapping much of our data.

Lastly, I used the Foursquare API to pull all the Salons and Barbershops for San Antonio, Texas.

**2.2 Data Cleaning**

The first task was to clean the data that I pulled open-source for the zip codes of San Antonio. I dropped data not relevant to the task at hand such as the Area codes (phone numbers) associated with that zip code and the county within which that ZIP code resided. I also decided to drop the ‘population’ from this column, opting to pull it later from the zip-codes.com API so I would be working with the most current population data. The most relevant data cleaning I did on this list was eliminating the ‘P.O. Box’ only zip codes, which are not associated with a geographic area and do not have associated demographic data attached to them. I also changed the way that ZIP codes were listed from strings to integers to make using this data as a callable reference easier when working with other datasets.

Next, I took the list of validated San Antonio zip codes and used it to slice the geoJSON I had for Texas zip codes and their boundaries. I used the mapshaper.org service to do this. I used my list of San Antonio Zip Codes to procedurally generate the JavaScript command to slice the Texas geoJSON into a smaller and more relevant San Antonio geoJSON. With this file, I can later map our data and our model.

**2.3 Feature Selection: ZIP Code Demographic/Economic Data**

The next step in solving the business problem was to gather the demographic data by zip codes. The zip-codes.com API returns 103 data points for each zip code, so then the task was to select the relevant features about each zip code that will help us predict where a good location to open a premium barbershop is.

Broadly speaking, I selected features that could indicate or answer 3 things:

* What is the wealth of the residents and employees in each zip code? This indicates potential customers who live or work nearby with the prerequisite wealth to afford premium haircuts.
* What social features might affect the proportion of a population who is willing or able to pay for a premium barbershop experience?
* What is the population density of a given area? This would indicate more potential customers.

Based on the above, I selected the following features:

|  |  |
| --- | --- |
| **Feature** | **Reason** |
| ZIP Code | This is the item being described by the other features |
| Latitude | Location Data |
| Longitude | Location Data |
| Land Area | Used to determine Population Density |
| Population | Used to determine Population Density |
| Average House Value | Indicator of Wealth of Residents |
| Income per Household | Indicator of Wealth of Residents |
| Business Employment | Used to determine average wealth of workers in each zip code |
| Business Payroll | Used to determine average wealth of workers in each zip code |
| Median Age Male | Age may indicate a likelihood to use premium barbershop services |
| Average Family Size | Family Size effects disposable income and thus likelihood to use a premium barbershop service. |

I also created a population density feature for each zip code by dividing the population of a zip code by its land area. This also helps control for the fact that ZIP codes are not uniform in size. Lastly, I created an average payroll feature by dividing the number of jobs in a zip code by the payroll of that zip code.

**2.4 Feature Selection: FourSquare Data**

For the Foursquare data, I queried the API based up the venue category ‘Salon / Barbershops’ and by zip code. From the data returned on this query, I selected venue name, latitude, and longitude. This query returned 2215 items and not all of them were relevant. To make this data useful for our purposes, I conducted the following:

* Eliminated duplicates in the results by comparing the latitude and longitude of the results
* I eliminated salons that were not oriented towards cutting hair (i.e. tanning, makeup, nails, etc.)
* I also eliminated salons that were oriented towards woman and kids.

I did this by looking at the name of the salon. When I was unsure, I conducted a google search of the venue. My method for filtering the results was that I built a list of keywords and names by which to filter the results by Venue name.

When this was complete, I had a list of barbershops where men could be expected to get their haircuts. Then, I needed to split this list into premium venues and regular venues. I used the same methodology as above, creating a list of keywords and venue names to assign a premium label to the premium establishments. Once again, if I was unsure about a venue, I google searched it so I could classify it correctly.

The result of the foursquare analysis is two lists: a ‘premium list’ and an ‘economy’ list of barbershops. With these, we can plot the locations of the venues which serves two purposes in this study: validate the clustering and classification of neighborhoods to determine suitability of a zip code for hosting a premium barbershop venue and to show visually to business stakeholders where their competition is located on the map.

**3. Methodology**

**3.1 Hypothesized model for good places to locate a business**

**3.2 K-means Clustering based upon the above variables**

**3.3 Adding location data to validate the model and check for competition**

**4. Results**

**4.1 Generalized Observations**

**4.2 Unexpected observations or Anomalies**

**5. Discussion**

**5.1 Insights Gained**

**5.2 Recommendations for Clients**

**6. Conclusion**