**CHOOSING A PLACE TO LIVE DURING RETIREMENT IN SAN FRANCISCO, CA AND SORRENTO, IT BASED ON RESTAURANT CLUSTERS**

**By: Jason Wright**

**June 4, 2020**

**Business Problem**

One of my friends, Andrew, he is a financial planner. A client of his, Greg, is approaching retirement and has had his heart set on splitting time each year of his retirement between San Francisco, CA and Sorrento, Italy. Greg visited Sorrento during a summer trip in college and fell in love with the city from the first moment he arrived. It was then that he had decided this was a place he wanted to spend his years in retirement. Greg had also spent part of his life living in San Francisco, CA and genuinely enjoyed the spring and summer weather. He had decided early on, after vising Sorrento and thinking of the time spent in San Francisco, he wanted to spend his retirement between these two cities. Immediately, he started saving all that he could so that he would be financially secure to realize this dream. Greg fully understands how expensive real estate is in both San Francisco and Sorrento and Andrew has given Greg high assurance that his savings should be enough to cover his living expenses within both cities.

Now that Greg was less than a year away from retirement, it was time to start planning his retirement in San Francisco and Sorrento. One of the main reasons Greg has chosen to spend his time between Sorrento and San Francisco is the memory of the amazing food both cities have to offer. Greg considers himself quite the foodie and he would like to focus his search on finding a place as close to as many great places to enjoy the foods of Sorrento and San Francisco. So, Andrew approached me to see if there was a way to help Greg focus his search on a place to live in both cities. As a data scientist, I knew that Foursquare compiled a vast amount of data on restaurants, bars, and other food related places. I informed Andrew that I would be able to compile a data set of restaurants, etc. within each city and create maps to show how they are clustered throughout San Francisco and Sorrento.

**Data**

It is understood from Andrew that Greg is highly interested in food related establishments and one of his criteria for retirement would be to live within an area that offers the greatest opportunity to partake in these establishments. The data will be compiled by querying Foursquare’s database to download a list of food related establishments (restaurants, bars, etc) within 1000 meters of the city center of Sorrento and 2000 meters for San Francisco. As San Francisco covers a larger area than Sorrento, the radius for San Francisco was expanded.

**Methodology**

Foursquare has categorized every place in their database based on the type of establishment. Furthermore, Foursquare has broken down these categories into subsets and listed each category on their website (<https://developer.foursquare.com/docs/build-with-foursquare/categories/>). Also, Foursquare limits their results per query to 50 venues that fall within the search parameters. So, it was found that using a broad category, i.e. Food, produced 50 results, but this limitation did not afford the opportunity to compile a list that truly encompassed each city’s food related establishments.

The query for San Francisco used categories of Italian Restaurant, Seafood Restaurant, and Steakhouse to compile the data, as state by Greg as his favorite types of venues. Each query returned 50 results. Upon review of the data, some of these venues were outside of the San Francisco city limits. This was a result of the radius encompassing some of the nearby suburbs of San Francisco. These establishments were removed from the data set resulting in 102 unique establishments and summarized in Table 1:



For Sorrento, it was decided to use queries consisting of two subset categories: Italian Restaurants and Bar (note: in Italy, a Bar is not the same type of establishment as it is in the USA. Bars in Italy refer to places where you can enjoy coffee, pastries, etc. Therefore, the term Bar was used, instead of another food related term). Each query returned the maximum result of 50. Upon review of the data, it was seen that each result contained 1 venue that appeared to be mislabeled as an Italian Restaurant or Bar and was therefore removed. This resulted in 98 individual venues within the Sorrento area, summarized in Table 2:



This data set does not encompass all food related venues in either city. However, compiling every establishment in both cities would consist of numerous queries that would be too time consuming. Since both data sets resulted in relatively the same amount of food related venues, I feel confident that this will result in a map for each city that can be utilized in Greg’s search.

**Results**

Now that the data set is complete, we will use it to compile maps using matplotlib in Python to show the distribution within each city. First, we will look at Sorrento, IT. Plotting the list of restaurants gives us the following map:

A picture containing text, map

Description automatically generated

We can take the map one step further and cluster them by location to provide a better idea of the number of restaurants in each area, as seen in the map below:

A picture containing text, map

Description automatically generated

By the numbers, the map shows that the most heavily populated area for restaurants is located closest to the city center of Sorrento and contains 59 venues.

Turning our attention to San Francisco, we can produce similar results in mapping out the restaurants found in our data. Again, looking at the plotting of the restaurants of San Francisco we see this map:

A close up of a map

Description automatically generated

Using the same clustering procedure as Sorrento, we apply this to the map of San Francisco to obtain the following:

A picture containing text, map

Description automatically generated

For San Francisco, we can see that the most heavily populated area for restaurants is 74, located in the northeast section of San Francisco.

**Discussion**

Even though there is a limitation of 50 results per query with Foursquare, we were able to obtain a sufficient set of results for both cities to accurately depict where the highest concentration of restaurants resides in each city. Using both cluster maps, Andrew will be able to relay this information to Greg so that he can work with his real estate agent to find his ideal retirement location within each city.

It can be suggested that there is much more analysis that could be explored using Foursquare’s data. Foursquare’s data also contains other aspects relating to the restaurants used in our data set. If Greg is finding difficulty in locating a place to live in either city, further analyses of Foursquare’s data can be completed. One such criteria to be explored would be the ratings for the restaurants. Analyzing these finding may result in opening his search to new areas of each city. Greg could also explore the reviews of the places contained in the data set. These two factors might come in handy as it is not always the quantity that is the best result, but the quality that may provide better results.

**Conclusion**

By looking at the two maps, it is clear where Greg should be looking for a place to live if he is solely using the criteria of being closest to the most restaurants. For San Francisco, 72.5% (74 out of 102) of the restaurants are in the northeast portion of San Francisco. As for Sorrento, 60.2% (59 out of 98) of the restaurants are located near the city center of Sorrento. These are the two areas where Greg should begin his search for finding his ideal retirement homes.