

Introduction / Business Problem

It has been a dream of mine to open a pizzeria. According to CNBC, "60 percent of new restaurants fail within the first year. And nearly 80 percent shutter before their fifth anniversary." They attribute the number 1 reason of failure to restaurant location. This sets up the problem for the case: how to determine the best location to open a successful pizzeria? Unfortunately it is not always enough to make a quality product to make a business successful. You have to be unique and you have to have the good foresight to pick a location that will attract customers. As a business owner, any advantage that you can use must be taken advantage of. For the data scientist with a passion for cooking pizzas, you can tilt the odds in your favor by analyzing geographic locations.

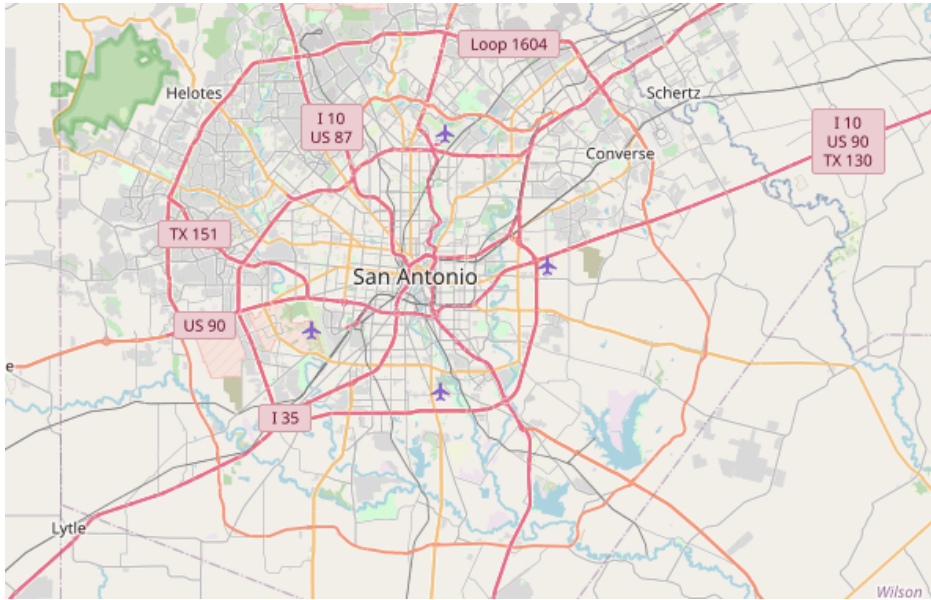
This case will utilize the Foursquare location to analyze the best location to open a pizzeria in San Antonio, TX.

Data

This analysis will utilize Foursquare location data. This will be completed using API calls to Foursquare within an IBM Jupyter Notebook. Further, I will obtain a list of San Antonio zip codes, via a webscrape from WorldPostalCode.com, to be assessed. Having a list of San Antonio zip codes in conjunction with Foursquare's venue capabilities, I will determine where popular dining areas are located. Further, this data will allow me to see if direct competitors are located nearby a prospective location. K-Means Clustering can also be applied to segment the data.

Methodology

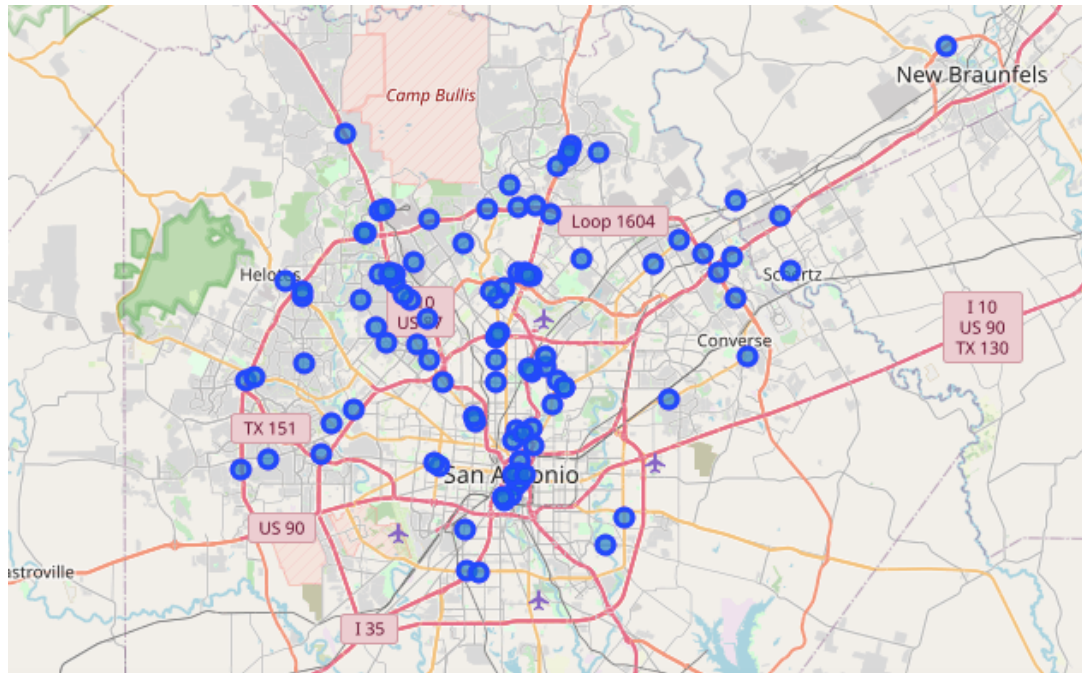
To begin, I imported all relevant python libraries needed to perform the analysis. Next, a CSV file of San Antonio zipcodes was uploaded and written into a pandas dataframe. I was then obtained the longitude and latitude coordinates of downtown San Antonio and created the following map as a starting point for the analysis:



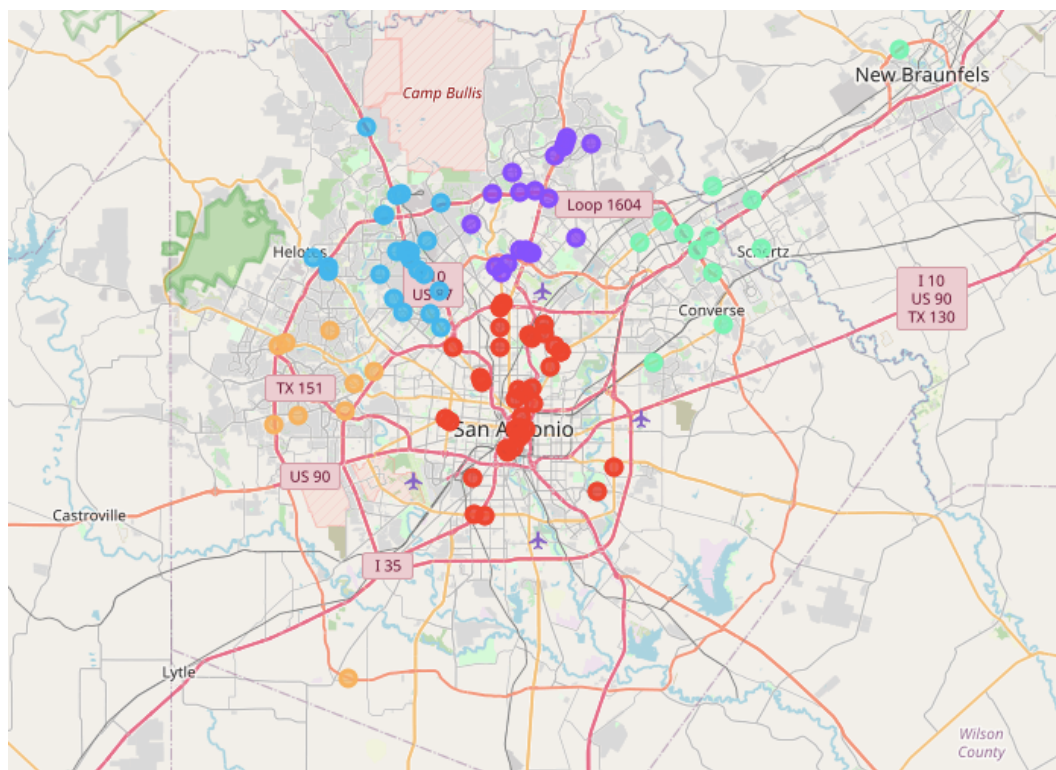
After the San Antonio map was created, I connected to the Foursquare API using my Client ID and Client Secret. My parameters to make the API call was limited to venue types as “Pizza Places” within a 50,000 meter (~31 mile) radius of downtown San Antonio and limited to 1,000 results. The retrieved data was written to a dataframe with the following result structure:

	name	categories	lat	lng
0	Barbaro	Pizza Place	29.455675	-98.491693
1	DOUGH Pizzeria Napoletana	Pizza Place	29.519182	-98.507337
2	Florio's Pizza	Pizza Place	29.499785	-98.467363
3	Main St. Pizza	Pizza Place	29.447439	-98.494090
4	The Station Cafe	Pizza Place	29.417457	-98.490465
5	Pizza Classics	Pizza Place	29.456757	-98.479276
6	Pizza Italia	Pizza Place	29.575156	-98.439420
7	Grimaldi's Pizzeria	Pizza Place	29.495976	-98.478831
8	Stella Public House	Pizza Place	29.409961	-98.495550
9	Alamo Pizza	Pizza Place	29.385755	-98.532876
10	California Pizza Kitchen	Pizza Place	29.497705	-98.481490

The pizza place coordinates were then plotted on the San Antonio map:

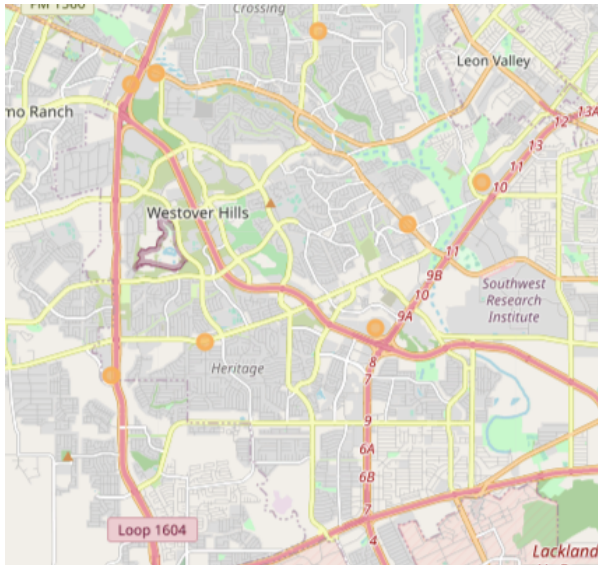


Next, I performed a K-Means cluster analysis and segmented to pizza places into 5 clusters. The results of the K-Means analysis (0-4 segments) were then written back to the listing of pizza places. With this data, I could then create my final map to show pizza place clusters in San Antonio:



Results

The results of the K-Means Cluster Analysis yielded interesting insights on where to open a pizzeria in San Antonio. It was important to perform this type of analysis because of how crucial it is to choose the right location. The location must be in a desirable part of the city and have few adjacent direct competitors. Further, you must also consider rent prices for the location to ensure profitability and long-term success. Based on the K-Means Cluster Analysis, an opportunity presents itself to open a pizzeria on the Northwest Side of San Antonio. This corresponds to the lower-density section of Cluster 4:



Discussion

In creating this analysis, I was tempted to retrieve Foursquare data for each zip code from the pandas dataframe created in the beginning. This would require coding a loop and I did not see an advantage over using a single coordinate and expanding the radius. In the end, that is exactly what I did and have higher confidence in my results as this just necessitated one API call. In analyzing the clusters, it was important to not automatically choose a low-density section of San Antonio. In reviewing the map, the East side of San Antonio stood out to me; however, I chose to forgo this area as it did not meet my other criteria for a desirable restaurant location.

Conclusion

This analysis was very useful in determining where to open a new pizzeria. The Northwest side of San Antonio is desirable due to the low-density cluster of competing pizzerias, is in a popular area, and an owner could expect moderate rent prices.