New Persian Restaurant in Orange County

Ali Babaie

August 18, 2020

1. Introduction

1.1 Background

New businesses are being opened everyday by different people and different companies. The main goal that everybody has is more revenue.
In the restaurant business, not only should the restaurant provide good quality food and excellent services, but they also need a certain amount of daily orders to stay in business.
Since there are so many Persians in Orange County, California, so it will be a safe choice to open a new Persian restaurant over there. And Americans and other nationalities are also interested in Persian food, which is a plus.

1.2 Problem

Orange County is a county that consisted of many cities. So, the main problem is which city to pick. And then which area of that city?

1.3 Interest

Basically, Restaurateurs are very interested in this project, and this project gives the best possible options.

2. Data

2.1 Data sources

My most important data are provided by Foursquare API, which is consisted of the Name, Location, and type of restaurants.
I also performed web scraping on Wikipedia to obtain some information about the cities in orange county.

2.2 Data cleaning

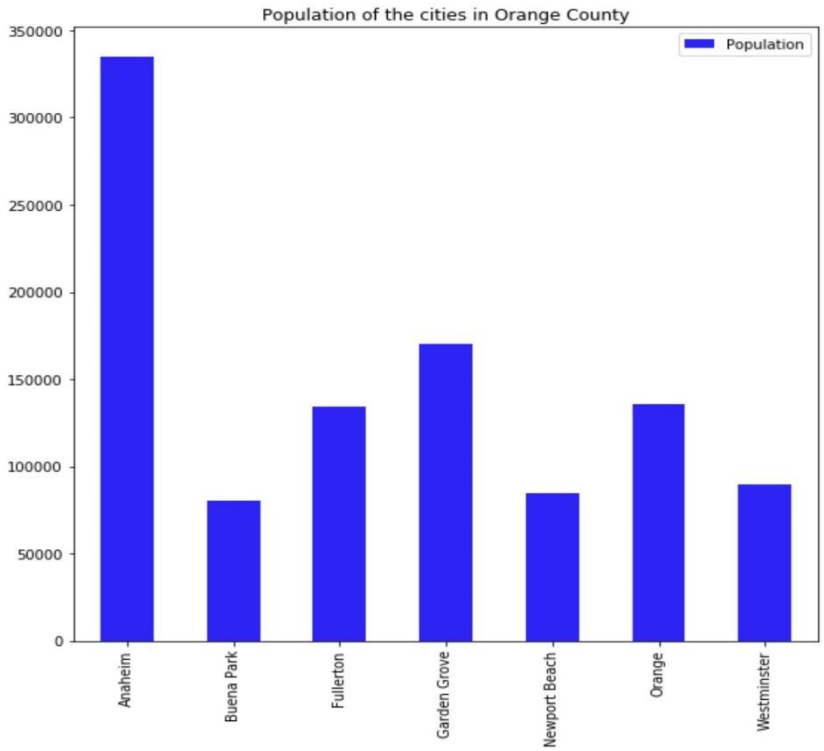
When I got a query from the Foursquare, the results were in the form of a JSON file with so many information. But I was interested in some parts of it, and I cleaned the data based on the venue, category, and name. In the middle of this process, I had to drop some columns to maintain a clean dataset.
I could find the population of the cities in California through web scraping, and I dropped the races and other columns that were not useful for my project. I only kept the population of the cities in orange county because more people in a town means more potential customers.

2.3 Feature selection
After data cleaning, I targeted Persian restaurants in different cities of orange county. I ended up with seven restaurants in six different cities.

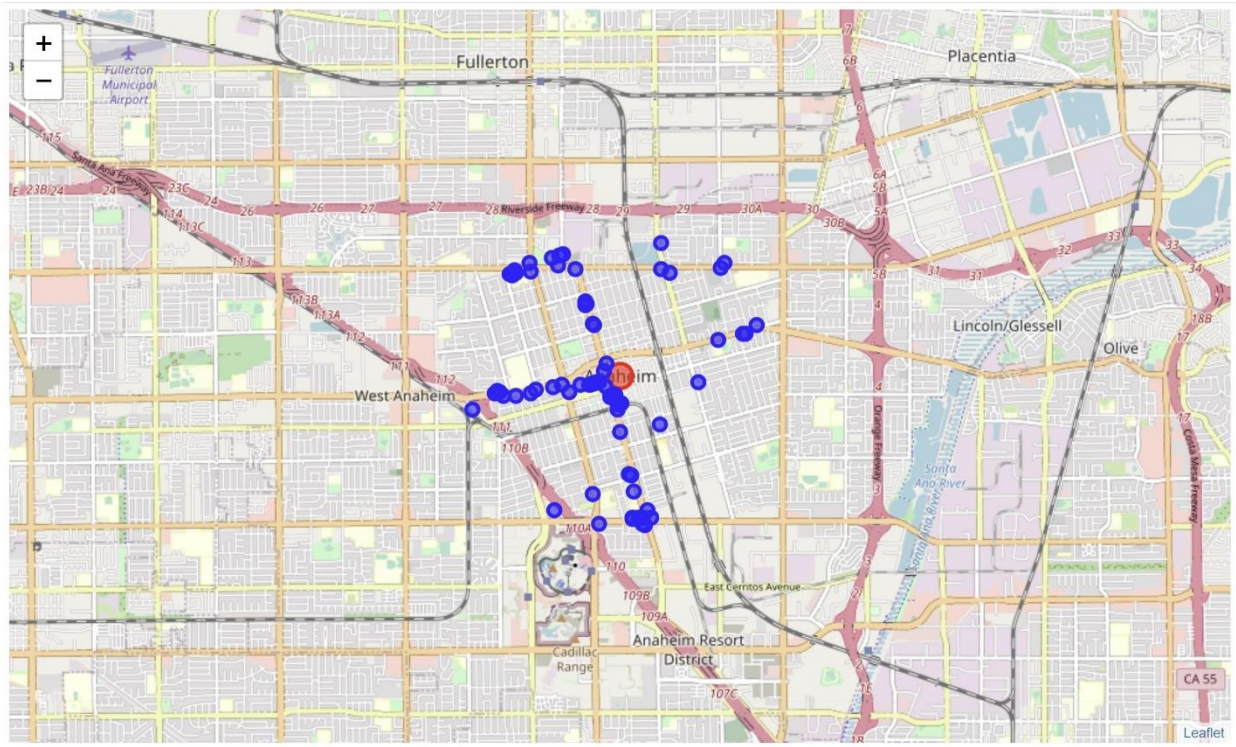
3. Methodology
As I mentioned earlier, cities with more population are better options for opening a restaurant simply because we have more potential customers.

Place	County	Population
Anaheim	Orange	335057
Santa Ana	Orange	325517
Irvine	Orange	205057
Huntington Beach	Orange	189744
Garden Grove	Orange	170148

In the above table, we can see the top five most populated cities. But we already know that we have Persian restaurants in some of them. For example, there is a restaurant in Santa Ana, and there are two in Irvine.
If we take out the cities with Persian restaurants and plot them, we will have this graph based on the population.

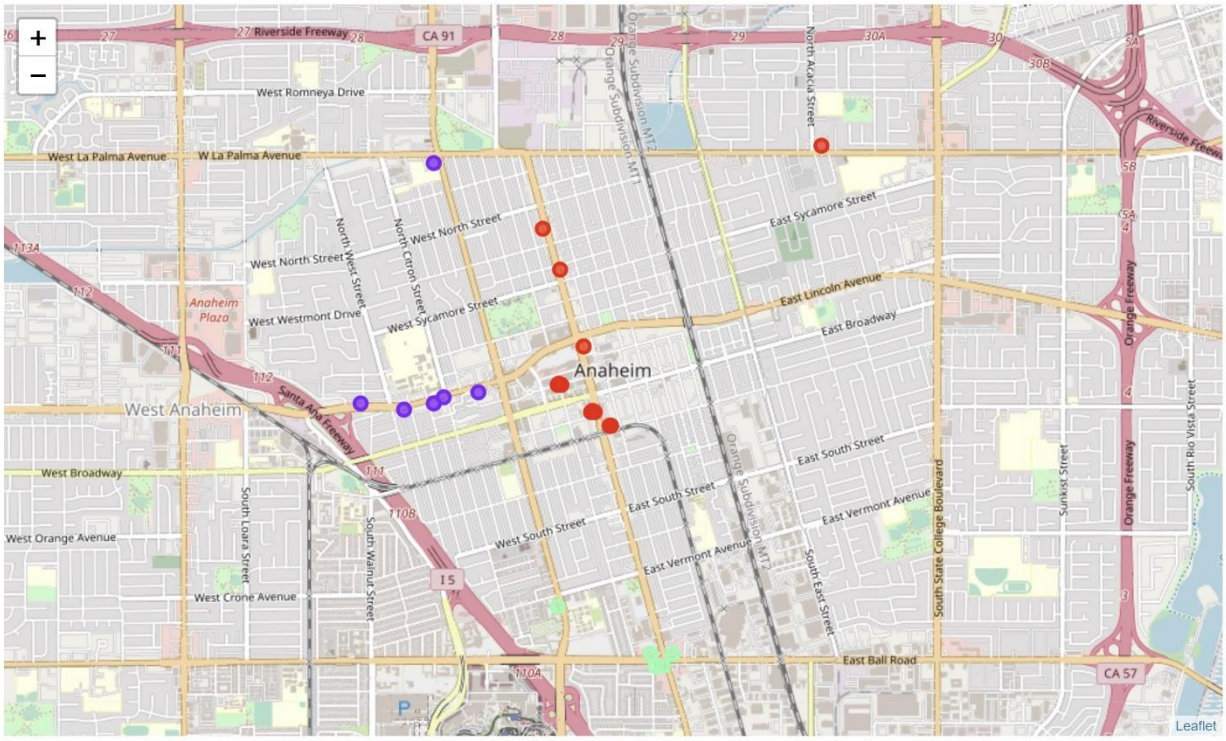


We can see that by far the best city for our new restaurant is Anaheim. It has the most population in the orange county.
By using Foursquare API again, I was able to find different venues in the city of Anaheim, in a radius of two kilometers from the center of the city.
I did some data cleaning again based on the name and category.



We can see the density of the venues.
Then with the help of K-Means Clustering, I was able to put the different restaurants into three distinct clusters. Then it was apparent that which areas are the possible locations for the Persian restaurant.

4. Results and Discussion



We can see the three different clusters on this map, which are from North to West, and South.
So, there is plenty of space for us on the East side of the city through the South.
We can see that we have a couple of blocks available to ourselves.
But in the end, some stakeholders will come up with the final decision on where to start this business. They may prefer any shopping center or plaza over the others in these locations.

5. Conclusion

In this study, I tried to find the best possible locations in Orange County, California, to open a new Persian Restaurant. We can see that with the help of Foursquare API and some extra information, it was not that hard. We need the right data for this project, which is obtained through Wikipedia and Foursquare API. Nevertheless, we need some background information about Python, Data cleaning, Data Analysis, some plotting, Machine learning, and some skills related to Data Science.