Restaurants in Glendale, CA

By Amir Tayyebian

10/2019

IBM Data Science Specialization

Capstone Project

TABLE OF CONTENTS

Introduction 1

Data 1

Methodology 2

Results 3

Discussion 4

Conclusion 5

**Introduction**

1. **Background**

Glendale California is a city in Los Angeles county with a population of over 200,000, due to the cities’ large Armenian population, Middle Eastern are also widely available. Believing that too much competition exists to open a Middle Eastern restaurant, our client is looking to open either a Thai, Mexican or Italian restaurant in Glendale. Our client is primarily concerned about the number of competitors for each type of restaurant. Additionally, our client wants to know which Glendale zip code (other than heavily residential zip codes 91208 and 91207) is best suited for this new restaurant

This report will first discuss the data requirements and methodology used to make this recommendation.

1. **Problem**

The purpose of this project is to answer the following questions:

1. Can Foursquare location data be used to provide a recommendation for our client regarding which of the three cuisine types should they choose.
2. Can Foursquare location data be used to provide a recommendation for which zip code is best suited for this new restaurant
3. **Interest**

This project will be of interest particularly to anyone who is trying to open a restaurant in Glendale CA. Additionally, anyone who is looking to open a restaurant anywhere in the United States can use the methodology in this project for similar insights.

**Data**

1. **Requirements**

There are several data entities that is required for this project. First, a list of all the zip codes in Glendale, CA is needed, along with the longitude and latitude values for the city of Glendale and each of these zip codes. This data will be used alongside the Foursquare API to gather data on the venues in the city of Glendale in order to decide which restaurant type should be chosen. After the restaurant type is decided, additional data is required to find the best zip code to open the restaurant in. This data will be included population data to find the number of potential customers. Population data will be combined with additional Foursquare API data on the different venues in each zip code, and the ratings for these venues.

1. **Sources**

United States Zip code data is publicly available in csv format, and can be downloaded by clicking the following link:

<https://simplemaps.com/data/us-zips>.

1. **Cleaning**

Relevant data entities in this file includes Zip Codes, City, Population per zip code, and the longitude and Latitude values for each zip code. These columns will be stripped from the csv file in Python to create the primary database for this project. The Foursquare API will also be cleaned to include only venues that fall under “Thai” and “Italian” and “Mexican” restaurants located inside Glendale. Additionally, these venues will be grouped by zip code for more detailed analysis.

Foursquare venue data was gathered by separately searching the queries Mexican, Italian, and Thai, with the three resulting data frames cleaned and ultimately joined to create the 37 resulting venues. Additionally, the data was cleaned in order to find non-restaurant venues that may have been included in the data.

**Methodology**

1. **Explorotory Data Analysis**

Both the Foursquare venue data and the Glendale zip code data needed be analyzed in order to get a better understanding of the data.

* Zip Code Data

Figure 1 shows the distribution of Glendale’s population by each zip code. The extremely low totals for the 91210-zip code led to the removal of the zip code from the database, and the removal of any venues located within the zip code, as such a low population is not desirable for our client. Keep in mind that 91208 and 91207 are already eliminated due to a lack of office space.

FIGURE 1

A screenshot of a cell phone

Description automatically generated

* Venue Data

The distribution of the 36 venues that fall under the three specified categories is shown in figure 2.

A close up of a logo

Description automatically generated

Figure 2

1. Methodology

The objective for this analysis was to first find the number of competitors for each restaurant type, and if there was no clear outlier regarding the number of competitors, we moved on to the quality of these competitors in the form of Foursquare ratings. If there is no obvious recommendation at this point, we planned to move on to analyzing the eligible zipcodes, and then repeating the process if necessary for only the resturants within the target zipcodes.

**Results**

1. **Number of Competitors**

Based on the previous figures, we can see that there isn’t noticeable difference between the number of potential competitors when customers are looking at Thai, Mexican, and Italian Restaurants. As a result, considering our client’s preferences for having fewer competitors, we can’t eliminate a specific category.

1. **Quality of Competitors**

To provide a recommendation regarding which of the three cuisine types provides the most opportunity, our next step was to analyze the ratings of the restaurants in each category. Our client’s second factor of importance in making this decision was to see which type of restaurant is more likely to stand out from its competitors. Our goal of this analysis was to figure out which type of restaurant is on average rated lower and has weaker competition in terms of quality. As a result, a python script was created to extract the reviews for each venue on Foursquare, and the average rating and the distribution of ratings for each restaurant type was calculated and added to the database. Figure 3 demonstrates these findings.

A screenshot of a social media post

Description automatically generated

Figure 3

As evident by figure 3, Mexican restaurants are higher rated than Thai or Italian restaurants, meaning that a new Mexican Restaurant would face better competition. As a result, we can eliminate Mexican restaurants for our clients. However, the average ratings between Thai and Italian restaurants are far too close to eliminate either category.

1. **Zip Code Analysis**

As there is no distinct edge regarding competitor quantity or ratings looking at the remaining two types of cuisine, we looked at each of the 6 remaining zip codes in order to find the best location for our client’s restaurant. Using this data, combined with the population data for each zip code, we calculated the ratio of people to restaurants in each zip code. Figure 4 displays those findings A screenshot of a cell phone

Description automatically generated

It is evident that the 91202-zip code has by far the highest number of populations per restaurant, meaning that it is the best zip code for our client to open his restaurant.

*It is additionally important to note that there were no trending restaurants while conducting this analysis*

**Best restaurant type for the 91202-zip code:**

In the above sections, we learned that neither remaining restaurant type, Italian and Thai, has an advantage over the others regarding number of competitors and the quality of the competitors. As a result, we chose to prioritize zip code location first before choosing the ideal cuisine type. After learning that the 91202-zip code is best suited for this new restaurant, the final step was to analyze the types of restaurants in the 91202-zip code in order to decide which restaurant is best suited for this zip code.

To make this analysis, we looked at all the restaurants of the three types in the 91202 and the surrounding eligible zip codes, 91201,91203,91207. The total number of restaurants by type in these zip codes is displayed on figure 5 below.

A close up of a logo

Description automatically generated

As a result, we found that focusing on and around the 91202-zip code, we can eliminate Thai food for our client’s restaurant as there is more competition compared to Italian restaurant, with the only eligible restaurant currently in 91202 also being a Thai restaurant. As a result, an Italian restaurant is most ideal for our client.

**Discussion**

As stated, limiting we to a city of Glendale with three cuisine types resulted in a relatively small sample size. Our client’s first concern for choosing a restaurant type was the number of competitors, and we learned there was no major distinction regarding the number of competitors for each restaurant type. As a result, we moved on to their second concern, the quality of the competitors, which led to us eliminating Mexican food.

Still unable to make a clear recommendation of the two remaining restaurant types, we turned our attention to first finding what zip code is best suited for a new restaurant. Our population and restaurant per population analysis led to us choosing the 91202-zip code. The final step was to limit the database to venues located in the 91202-zip code, or in a neighboring zip code, and two repeat the analysis for the remaining two restaurant types. After doing this we eliminated Thai food as a restaurant type, with our final recommendation being an Italian Restaurant in the 91202-zip code.

**Conclusion**

In this study we analyzed the zip codes in the city of Glendale, and the venues in Glendale that can be qualified as Mexican, Thai, or Italian food. Using a combination of available data and the Foursquare API, we narrowed down our recommendation to a specific type of restaurant in a zip code, based on the client’s primary concerns: the number of competitors & the quality of these competitors.

**Future Direction**

This study merely focused on the number of competitors and the quality of the competitors based on data available on Foursquare.

It is important to note that there are several factors that should be explored in the future regarding this study. It is important to note that demographic information such as age, race and income can potentially change the quality of the recommendations made. Additionally, commercial real estate prices and availability is another important factor that can be analyzed in the future. Further, surveys can be conducted to gather more data to understand the demand of each restaurant, and other methods such as Yelp ratings can be considered to analyze the competition.