IBM Applied Data Science Capstone

Presented by: Sara Rauf

Capstone Project: Opening a traditional dining restaurant in New Orleans by exploring different neighborhoods

1. Introduction and Description of Data:

New Orleans is the most populous city in Louisiana with a population of an estimated 390,144. It is world-renowned for food, music, festivals, and celebrations. Food is one of the main attractions in this historically and culturally rich city. In fact, New Orleans has its own cuisine called Cajun or Creole.

In 2005, New Orleans was majorly struck by Hurricane Katrina which flooded more than 80% of the city also causing a population decline of over 50%. The food industry was also heavily affected but many restaurants bounced back and since then the number of places to eat in the city has grown exponentially.

1.1. Business Problem:

The main purpose of this project is to be able to analyze and choose the best neighborhood in New Orleans Metro area to open a restaurant. By using various Data Science tools and Foursquare location data, different geographical locations will be explored.

1.2. Target Audience:

This data analysis is aimed towards all who are interested to open a successful restaurant in the city of New Orleans which resonates with the traditions as well as cultural values of the residents. This analysis could also be of interest to entrepreneurs willing to invest in restaurant businesses in different neighborhoods. It will specifically help any investor to carefully examine different locations along with their pros and cons and be able to evaluate whether the specific location will be best to open a restaurant.

1.3. Description of Data:

There are various important factors when choosing the best location for a restaurant. All of these factors are taken into consideration and the relevant data will be used to solve the business problem. Following are the data and data collection methods of this project:

1. Total Population by Parish in New Orleans Metro
   Method:
   New Orleans city and Orleans Parish are interchangeable. They have the same

population. However, we will take a good look at all the parishes in the New Orleans Metro area to see which parish is populous. This data will be scraped from https://www.datacenterresearch.org/data-resources/population-by-parish/.

2. Total population by Zipcodes in New Orleans
   Method:
   This data will be scraped from https://www.zipdatamaps.com/zipcodes-new-orleans-la in order to analyze which neighborhood area within a parish is populous.

3. List of Latitudes and Longitudes of Neighborhoods in New Orleans from Wikipedia
   Method:
   The url https://en.wikipedia.org/wiki/Neighborhoods_in_New_Orleans will be used to get latitudes and longitudes of the neighborhoods in New Orleans. The data analysis will be used further to plot the map and get venue data.

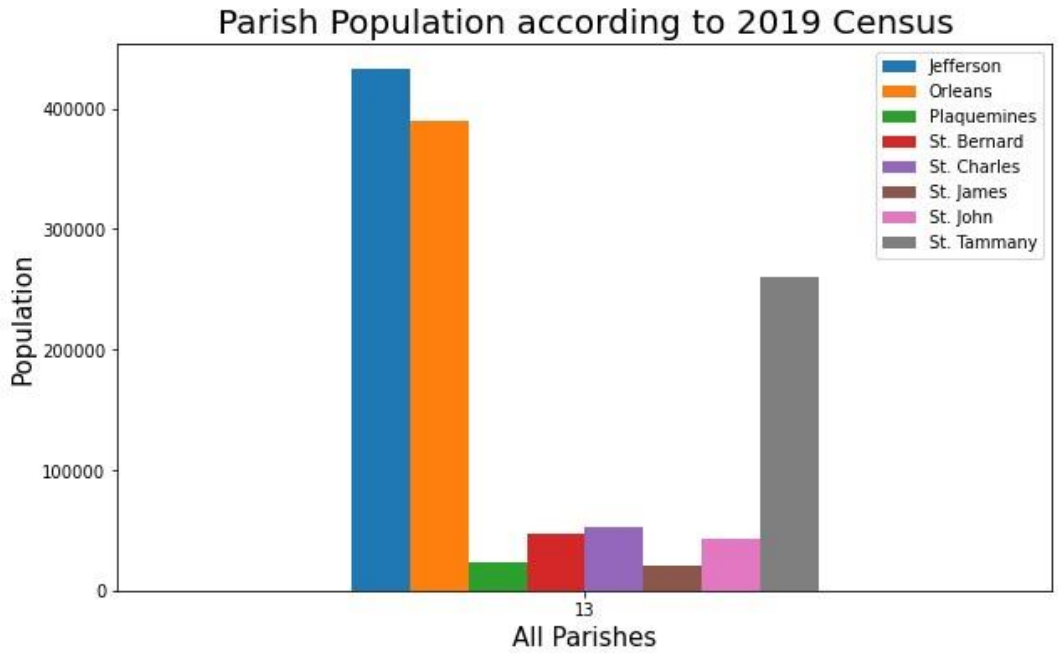
4. Foursquare location data
   Method:
   Foursquare API will be used to download all venues from neighborhoods in New Orleans: https://api.foursquare.com and further perform data analysis on the data.

2. Methodology

Our main purpose is to find which neighborhoods are the most populated and which neighborhoods have common clusters of restaurants in New Orleans

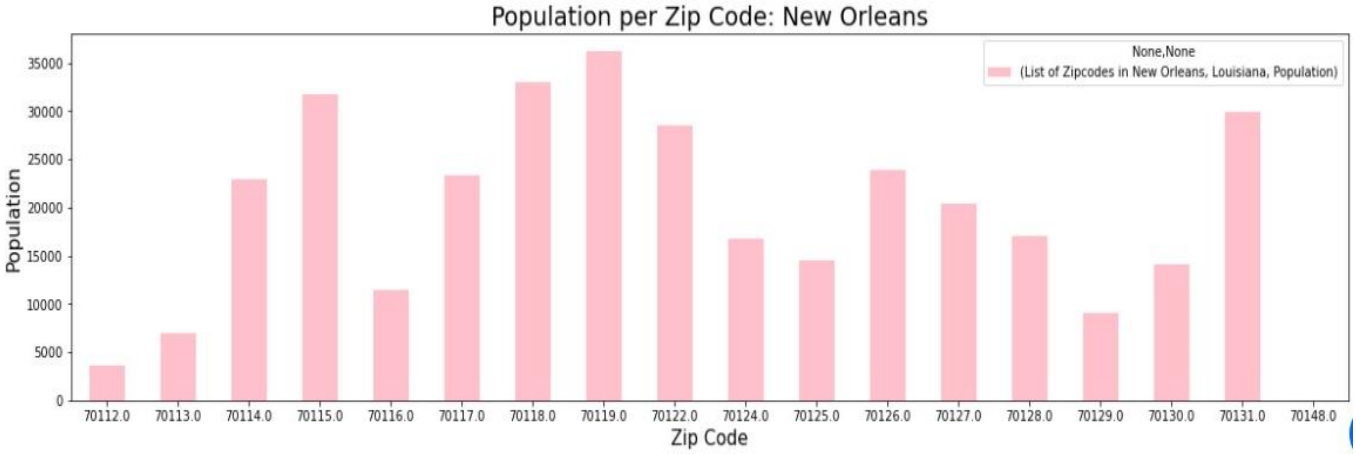
2.1. Data Wrangling and Cleaning- After loading all the data into the project and converting it into pandas dataframes, the data was cleaned.

2.1.1.1. A bar chart was plotted to see the highest population per parish.



According to the chart, it is evident that in the New Orleans Metro Area, Jefferson and Orleans Parishes are the most populated. We choose Orleans Parish.

2.1.1.2. Next, a bar chart was plotted to see the highest population per zip code area in the Orleans Parish.



According to our results, among the most populated zip codes are 70115, 70118, 70119, 70122 and 70131.

2.2. The Foursquare API is used to explore neighborhoods in New Orleans:

-An API request URL is created
- A GET request is made.
- We retrieve relevant information regarding different venues
- We get all the venues of New Orleans
-We check how many venues each neighborhood has
-We retrieve unique categories for all the returned venues and pick out the restaurant category only.
- We use one hot encoding and analyze each neighborhood and category
-We take the mean of the frequency of occurrence of all the categories
-The top 5 restaurant venues for each neighborhood are printed
-We retrieve 5 most common restaurant categories for each neighborhood. This feature is used to group neighborhoods into clusters.

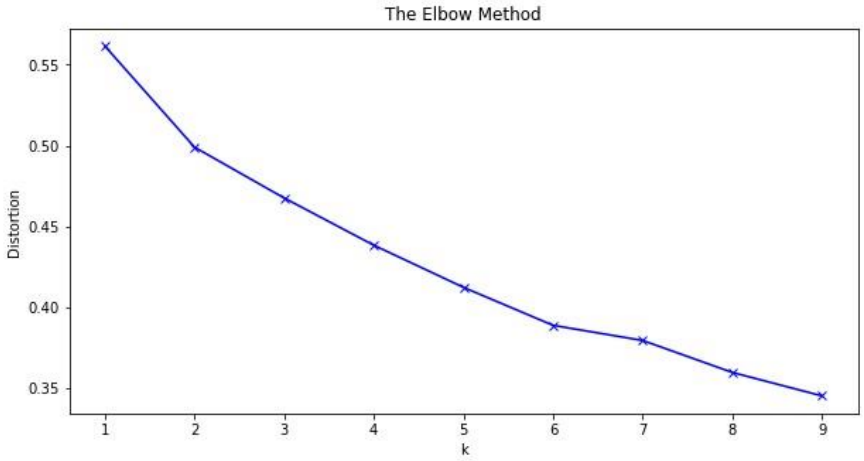
2.3. The clustering model used in this project is k-means. This model was selected because it is useful to discover quick insights from unlabeled data.

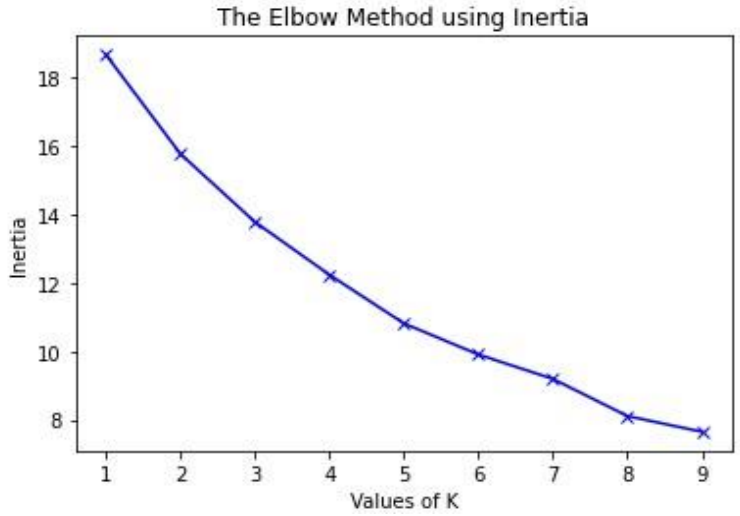
2.3.1. Choosing the optimal for k in kmeans: The Elbow Method

The **Elbow Method** is one of the most popular methods to determine the optimal value of k. In the elbow method, the sum of squares at each number of clusters is calculated and graphed and the user looks for a change of slope from steep to shallow (an elbow) to determine the optimal number of clusters. The clustering model is built and values of Distortion and Inertia are calculated. Distortion and Inertia can be defined as follows:

Distortion: It is calculated as the average of the squared distances from the cluster centers of the respective clusters. Typically, the Euclidean distance metric is used.

Inertia: It is the sum of squared distances of samples to their closest cluster center.





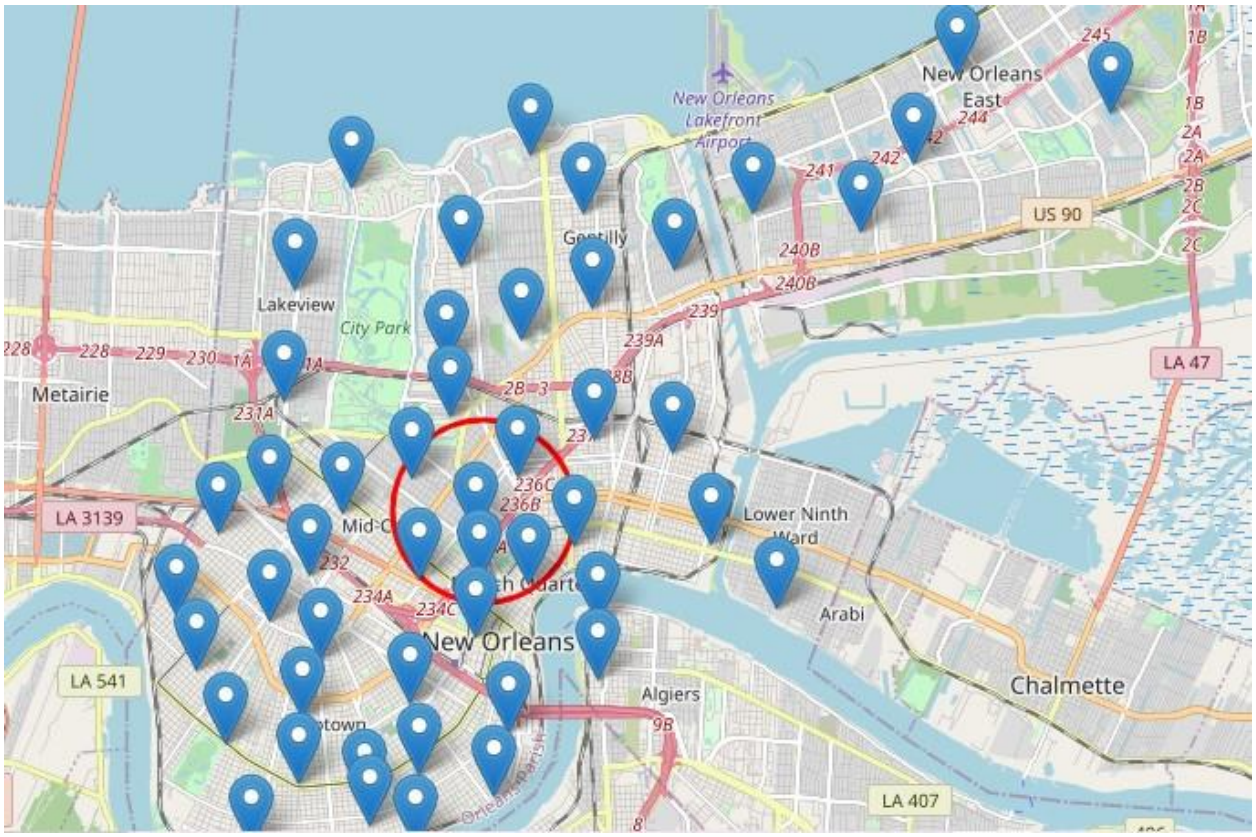
According to this method, we can choose k from 6 to 9. We choose 7.

2.3.2. Run K means and segment data into 7 clusters and generate labels.
2.3.3. Merge the NOLA data with neighborhood coordinates data
2.3.4. Add clustering labels
2.3.5. Create new data frame with both cluster labels and common venues
2.3.6. Calculate the most representative point within the cluster group - the centroid. Usually, this is the mean of the values of the points of data in the cluster. This will also be our ideal location to open a restaurant.

2.4. Geocoding API - Opencage will be used to convert latitude and longitude values to addresses to see which neighborhood area is the best to open a restaurant.

2.5. The Folium library will be used to visualize the neighborhoods in New Orleans and the resulting clusters.

3. Results



The best neighborhood to open a restaurant in New Orleans is Treme/Lafitte and the address is 1228 North Johnson Street, New Orleans, LA 70116, United States of America

4. Discussion

According to our results, Treme/Lafitte is the best neighborhood to open a traditional restaurant. It should be noted that Treme/Lafitte is not highly populated, but it has a potential for successful emerging and booming startups as it is adjacent to some of the most accessible areas in New Orleans, like the French Quarter which is the heart of New Orleans.

Upon observing the top venues per neighborhood, it can be concluded that mostly, Treme/Lafitte has Southern / Soul Food Restaurants. It also has some wing joints, bars, Cajun/creole restaurants as well as coffee shops.

4.1. Future Recommendations:

4.1.1.1. Four square venue location accuracy can be improved to give better results. More extensive software development tools can be built, and the data can be tested up to date.

4.1.1.2. The performance of kmeans is not fully accurate compared to other clustering techniques and it is not enough for classification. If there are any changes in data, it could lead to more inaccuracy. Clusters are also assumed to be spherical and evenly sized, which further reduces the accuracy of kmeans clustering.

4.1.1.3. It is possible that the Geocoding API can return inaccurate locations as many places have the same name. One way to improve this would be to build larger databases and keep the data up to date.

5. Conclusion

Our main goal was to open a traditional restaurant that speaks to New Orleans culture as well as caters to tourists who come solely to enjoy New Orleans food. Our goal is accomplished through the results obtained in this project. It is best to open a restaurant with a Southern theme that offers creole/Cajun food - which is what New Orleans is known for.