

Capstone Project – The Battle of Neighborhoods

New Orleans: IndNola Bar and Grill

K. Jhansi Naveena

1. Introduction

A. Introduction:

New Orleans is known for its distinct jazz music, different cuisines and celebrations, and is highly known for its Spanish Creole culture¹. It is one of the states that contains wide variety of Italian, French, Spanish, Cajun, and Chinese flavors infused with one another, and is high of tourism for its festivities, culture and food. It doesn't only indulge in food, but also in outdoor activities, festivals, night life, cultural arts, sports, voodoo, music and shopping². One can only enjoy all of it in one place. Each year the number of tourists increase by at least 5%³ for its attractions. New Orleans being such a diverse and high tourism state, it makes it considerable to open an Indian – Cajun infused restaurant in New Orleans to allow everyone to have a taste of explicit taste of Indian food with light touch of Cajun flavors.

B. Business problem:

The Foursquare API endpoints are utilized to obtain required data to find the most appropriate, possible, location to establish the restaurant that will attract the tourists and natives of New Orleans. The goal is to find which type of restaurants are most visited and liked by people. The results are combined to produce the inclusions of features for the restaurant. Once the features are set, location, or neighborhood to be specific, can be finalized.

2. Data

Though foursquare services are to be utilized to progress towards the mission, to execute the project faster, a New Orleans Districts dataset is used. A CSV file is created that contains the districts' names and the neighborhood in them to help visualize the distinct districts on New Orleans map(s). The dataset contains 74 observations and 5 features, namely, district number (District_N), name of the district (Name), neighborhoods within the district (Neighborhood), and latitude and longitude of the neighborhoods. Sample record is shown below.

	A	B	C	D	E
1	District_N	Name	Neighborhood	Latitude	Longitude
2	District 1a & 1b	French Quarter/CBD	French Quarter	29.9596	-90.063
3	District 1a & 1b	French Quarter/CBD	Central Business District	29.948	-90.0723

Fig 1. Sample CSV file

Since the entire dataset wasn't available online, district number, name and neighborhood were collected from one site, and latitude and longitude were collected from another.

Having such dataset improves efficiency and I believe is easier to segregate and distinguish one neighborhood from another to reach our goal. Associating the chosen dataset and foursquare services (explore, search, venues, tips, etc), it will help us reach both the goals.

3. Methodology

Once the data was read, it was checked for any missing values and are deleted if number of missing values are insignificant. To find the nearby restaurants in New Orleans, the data frame is segregated into 13 data frames as there are 13 districts in New Orleans. Initially, restaurants in French quarter are collected using Foursquare venue search API and stored in a data frame that contains the neighborhood name, restaurant name, restaurant category, address, latitude, longitude, and venue id.

Once the French Quarter data is collected, the neighborhoods in each district are passed to the `getNearbyRestaurants()` function to obtain all the restaurants that are within the neighborhood vicinity. The results are stored in a data frame. The function is called 13 times to get restaurants that are in the respective district. Once all the restaurants are found and stored in their data frames, all the 13 data frames are merged to one data frame. Since no radius was set in the function, there is high possibility that there is presence of duplicate values of venue in the final merged data frame. Hence, the duplicate values are dropped except for the first occurrence. When the duplicate values are dropped, `getRestaurantRateLikes()` function is called to obtain each restaurant's rating and number of likes. The merged data frame is normalized using one-hot encoding and then grouped by neighborhood using mean.

As all the values are numerical and no labeled data, K-Means algorithm is implemented upon the dataset. To find out the best value for 'k', silhouette coefficient is calculated for cluster values of 2 to 10. The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is $(b - a) / \max(a, b)$. The best value is 1 and the worst value is -1. Values near 0 indicate overlapping clusters. The cluster with highest silhouette coefficient is used to fit the data. The cluster values are then inserted in the data frame that contains top 5 restaurants from each neighborhood. Finally, the clustered data frame and the initial final data frame are joined.

4. Results

Matplotlib and seaborn packages are utilized for visualization. A bar plot is drawn to visualize the distribution of all restaurant ratings. The restaurants with highest ratings are listed to check how many restaurants there are in that neighborhood. Then, a list of the top 10 most liked restaurants is produced as a data frame. Another bar chart is drawn to see the top 10 unique categories of restaurants. The visualizations and the lists produced are used to conclude the possible restaurant features. To select the neighborhood for the restaurant establishment, a bar chart is visualized with top 10 neighborhoods that have the highest number of likes and ratings.

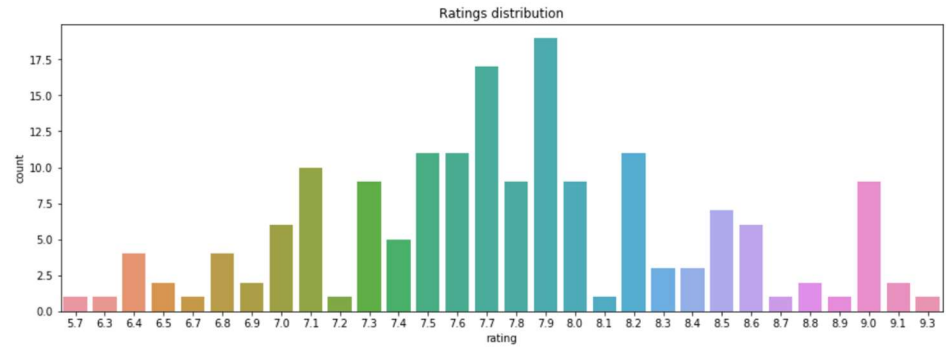


Fig 2. Ratings distribution

	name	categories	number of likes	neighborhood	rating
19	House of Blues Restaurant & Bar	Music Venue	1407	French Quarter	9.3
2	Mother's Restaurant	Cajun / Creole Restaurant	1077	French Quarter	8.0
33	Cochon Restaurant	Southern / Soul Food Restaurant	638	Central Business District	9.1
3	Felix's Restaurant & Oyster Bar	Seafood Restaurant	478	French Quarter	8.4
8	Deanie's Seafood Restaurant in the French Quarter	American Restaurant	394	French Quarter	7.9
43	Atchafalaya Restaurant	Cajun / Creole Restaurant	248	East Riverside	9.0
14	Mulate's Cajun Restaurant	Cajun / Creole Restaurant	247	French Quarter	7.4
38	Katie's Restaurant & Bar	American Restaurant	227	Central City	9.0
17	Restaurant R'evolution	Cajun / Creole Restaurant	227	French Quarter	9.1
5	Antoine's Restaurant	Cajun / Creole Restaurant	219	French Quarter	8.0

Fig 3. Top 10 liked restaurants

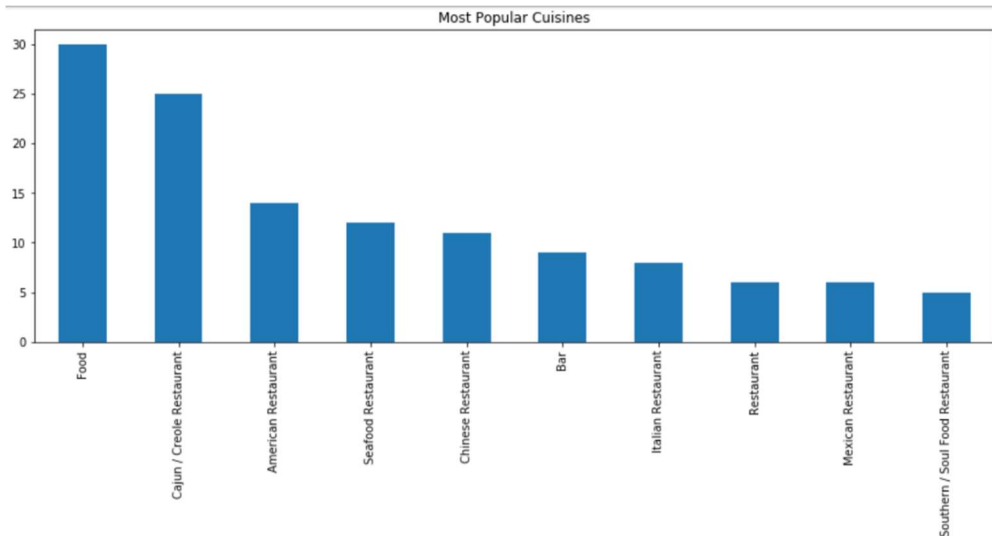


Fig 4. Most popular cuisines

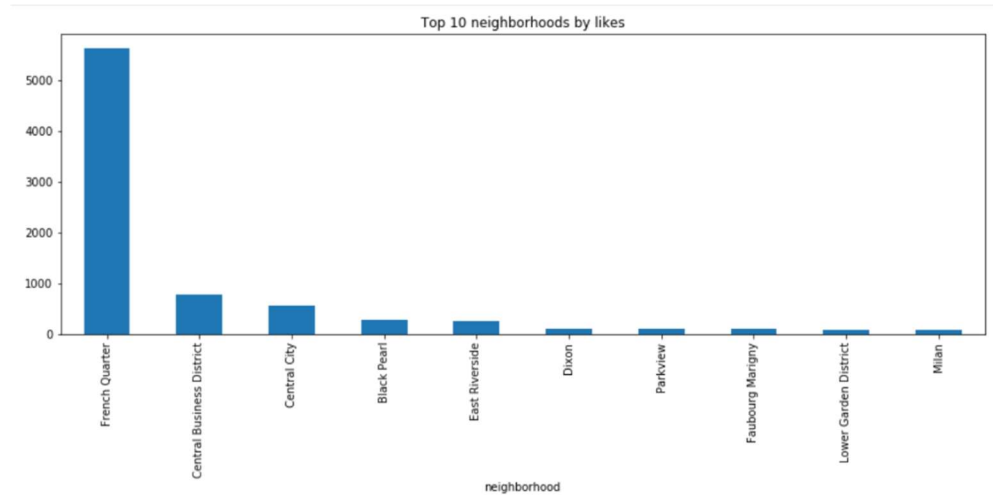


Fig 5. Top 10 liked neighborhoods

5. Discussion

The results produced by the dataset and Foursquare clearly states where tourists and locals most likely tend to go, and which type of cuisine is preferred by them. However, apart from the data collected, there are so many other factors for opening a restaurant. Some of the basic factors are⁴:

- Choosing a solid restaurant concept
- Checking local competition
- Selecting a good location
- Writing a business plan (executive summary, company description and market analysis)
- Meeting with investors
- Hiring staff
- Writing the menu
- Ordering equipment
- Holding a grand opening
- Marketing

All the required data can be leveraged using Foursquare data. For instance, premium API endpoints are utilized for checking local competition by comparing each restaurant with another using values such as reviews, likes, check-ins, and tips. Even best investing companies can be found and compared using Foursquare API endpoints.

The current outcomes are forthright, as it didn't require in-depth analysis to show conclusions. It is highly probable due to the demographics and geographical data. For instance, if New Orleans was to be compared with New York City, it is clearly seen that land area of New Orleans is half of New York's and population is immense in New York City than New Orleans.

6. Conclusion

The first goal is to find which type of cuisines are liked by people and to include their distinctive features in restaurant. The features are based on the visualizations and exploratory data analysis performed after cleaning and extracting the data. In conclusion, the possible features for restaurant are:

- Bar: liquor, cocktails and Indian street food with various liquor assortments
- Music venue: live music at night
- Indian/Cajun cuisine: Amalgamate Indian cuisine with Cajun cuisine

The final goal is to find best suited location for opening the restaurant. Once K-Means algorithm is implemented, the cluster with high frequency count is further analyzed by grouping neighborhood. The neighborhood that has the highest count is chosen as the opening location. Hence, French Quarter is chosen to open the restaurant.

7. References

1. “New Orleans”, https://en.wikipedia.org/wiki/New_Orleans
2. “Things to Do”, <https://www.neworleans.com/things-to-do/>
3. “New Orleans Breaks Tourism Records for Visitation and Visitor Spending in 2016”, FQBA – Brittany, In News, March 23, 2017
4. “Basic Checklist for Opening a Restaurant”, <https://www.thebalancesmb.com/basic-checklist-for-opening-a-new-restaurant-2888369>, Lorri Mealey, June 25, 2019