Opening a restaurant or bar in New Orleans

Tom VandenBoom

1 Introduction

1.1 Background

New Orleans is a city (population 390,845 in 2019) located at the mouth of the Mississippi river on the central gulf coast. One of the city's foremost industries is tourism, with nearly 20 million visitors spending over $10 billion USD in the city in 2019. Prior to the COVID-19 pandemic, the tourism industry had experienced year-over-year growth for over a decade. Since March 2020, however, many New Orleans restaurants and bars have been forced to close due in part to federal and local government restrictions on travel and indoor gatherings. As more of the world becomes vaccinated against COVID-19, these recent closures present a unique opportunity for prospective restaurateurs and bar owners to establish a foothold in one of the premier tourism cities in the United States.

1.2 Problem

In this project, I explore the neighborhoods in New Orleans in which to open a new restaurant or bar by identifying those neighborhoods which are similar to the very popular French Quarter in downtown New Orleans.

1.3 Interest

The target audience for this report is any potential business owner who is unfamiliar with the city who may have interest in opening a restaurant or bar in New Orleans.

2 Data acquisition and cleaning

2.1 Data sources

To address the problem in question I utilized the following data:

- A list of neighborhoods of New Orleans, along with their latitudes and longitudes, from Wikipedia: `https://en.wikipedia.org/wiki/Neighborhoods_in_New_Orleans`
- Calls to the Foursquare API to retrieve a list of venues close to the center of a given neighborhood.
- A call to the OpenStreetMap Nominatim geocoding service to find the central coordinates of the city of New Orleans.

2.2 Data cleaning

I first scraped the Wikipedia page above to find the latitudes and longitudes of the centers of each neighborhood in New Orleans. Given these latitudes and longitudes, I made calls to the Foursquare API to find the data for the (at most) 100 nearest venues to the center of a given neighborhood; an example five rows of such data are shown below:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	U.S. NAVAL BASE	29.946085	-90.026093	The Mighty Missisippi	29.949695	-90.023710	Boat or Ferry
1	ALGIERS POINT	29.952462	-90.051606	Tout de Suite Café	29.952121	-90.051090	Café
2	ALGIERS POINT	29.952462	-90.051606	Congregation Coffee Roasters	29.951918	-90.053395	Coffee Shop
3	ALGIERS POINT	29.952462	-90.051606	The Crown & Anchor	29.951416	-90.054220	Bar
4	ALGIERS POINT	29.952462	-90.051606	Old Point Bar	29.954940	-90.050226	Bar

Figure 1: A snapshot of the data, including the neighborhood, coordinates, venue, and venue categories of five sample venues.

2.3 Feature selection

I will examine the frequencies of "Venue Categories" in neighborhoods as pulled from the Foursquare API.

3 Methodology

Each venue was categorized numerically by converting the Venue Category feature using the OneHot encoding method `pandas.get_dummies()`. This data was then grouped by neighborhood. I took the mean value of each category and performed the `sklearn` k-means unsupervised clustering algorithm for various values of k ($1 \leq k \leq 30$). I observed the graph of distortions of the models for each k and applied the "elbow method" to determine the optimal value of k to select a particular clustering model. My fundamental hypothesis is that the cluster containing the French Quarter contains ideal neighborhoods in which to open a new restaurant in New Orleans.

3.1 Exploratory data analysis

After scraping the Wikipedia URL above, I plotted the neighborhoods using the `folium` library in Python:

Some of these neighborhoods (Viavant - Venetian Isles, Village de Lest, Lake Catherine, and New Aurora - English Turn) were well outside the city limits and not reasonable places to consider opening a restaurant; I removed them from consideration for this reason.

3.2 Clustering Analysis

After grouping the OneHot encoded venues by neighborhood and running the k-means clustering algorithm for $1 \leq k \leq 30$, we plot the distortions of the models against k for each such number of clusters:

Ideally, this plot would have an "elbow", where the amount the distortion decreases by increasing k becomes significantly less than lower values of k. This plot does not have an apparent elbow, other than perhaps at $k = 8$ and $k = 13$. I defaulted to the smaller value of k to avoid overfitting.

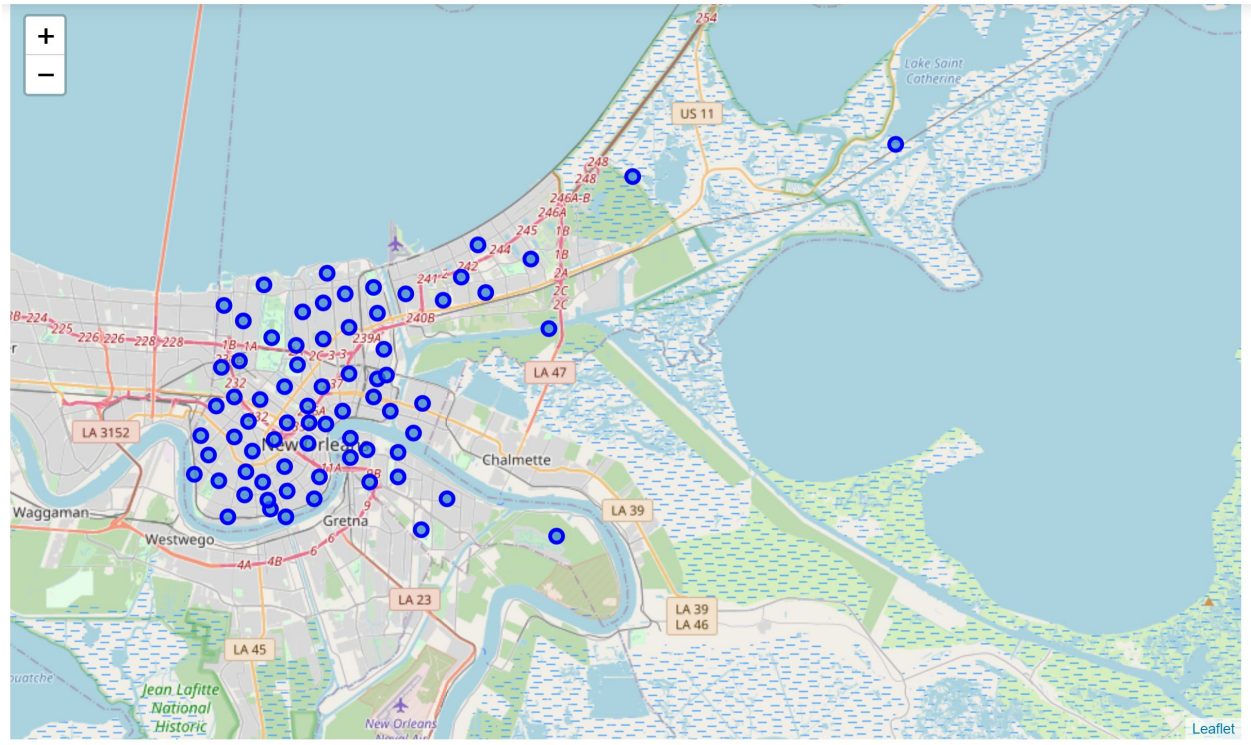

Figure 2: A map with the neighborhoods of New Orleans from the above Wikipedia URL.

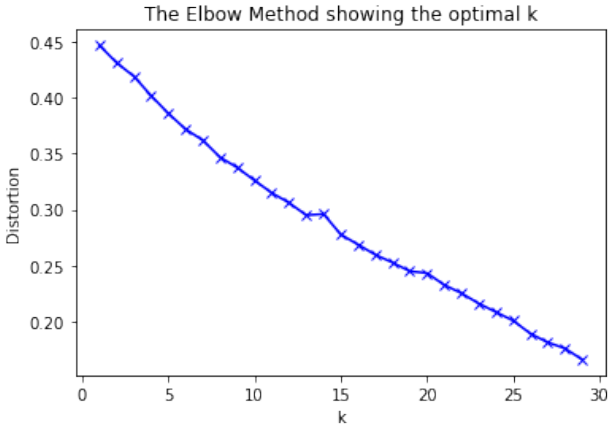

Figure 3: The distortions of the k-means clustering models for $1 \leq k \leq 30$.

For the k-means clustering model with $k = 8$, we fit the data to the model and plotted the neighborhoods of New Orleans using `folium`, colored by cluster, in Figure 4.

The neighborhoods which are in the same cluster as the French Quarter, as well as their three most common venue categories, are displayed in Figure 5 below.

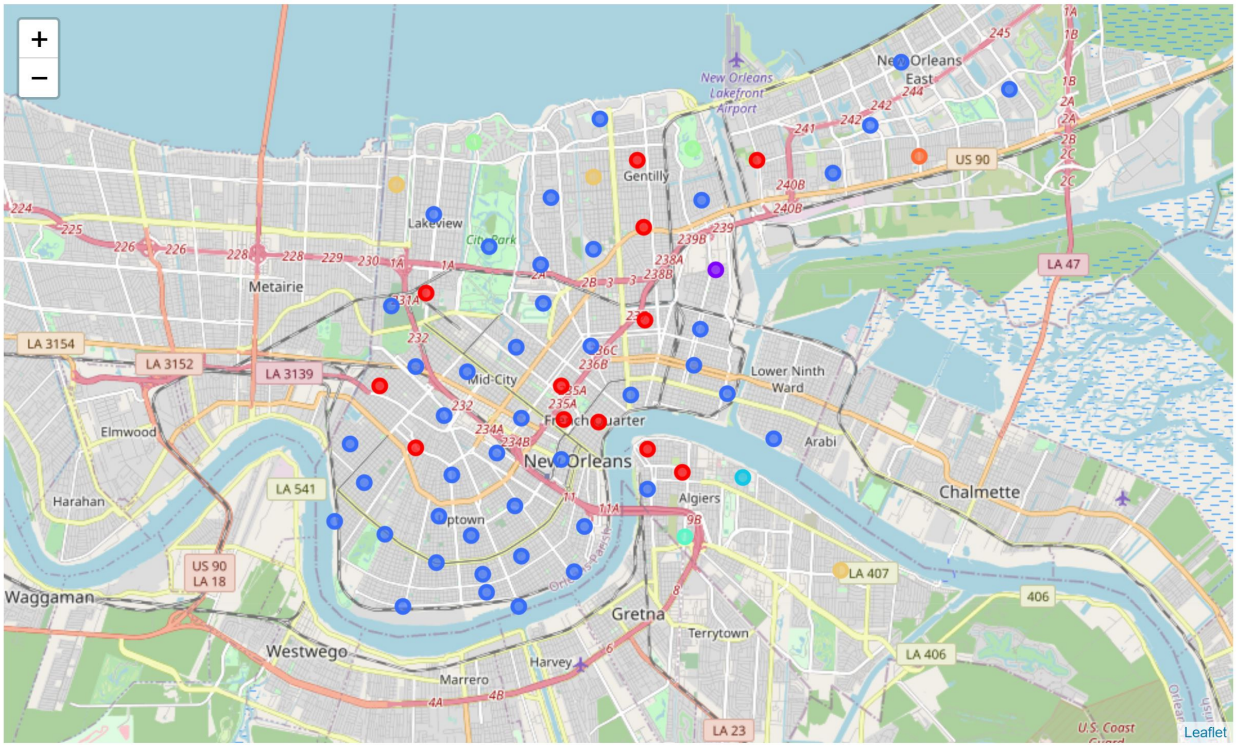

Figure 4: A map of the neighborhoods of New Orleans, organized into $k = 8$ clusters using the k-means unsupervised clustering algorithm. The red dots belong to the same cluster as the French Quarter.

	Neighborhood	Longitude	Latitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	ALGIERS POINT	-90.051606	29.952462	0	Boat or Ferry	Bar	Historic Site
2	WHITNEY	-90.042357	29.947200	0	Nightclub	Clothing Store	Yoga Studio
10	MARLYVILLE - FONTAINEBLEAU	-90.113468	29.953001	0	Bar	Campground	Nightclub
15	FRENCH QUARTER	-90.064373	29.958839	0	Cajun / Creole Restaurant	Bar	Gay Bar
36	HOLLYGROVE	-90.122995	29.967278	0	Plaza	Bar	Paper / Office Supplies Store
38	TREME - LAFITTE	-90.074415	29.967278	0	Southern / Soul Food Restaurant	Trail	Music Venue
41	ST. ROCH	-90.052099	29.982149	0	Hardware Store	Bar	Sandwich Place
44	NAVARRE	-90.110378	29.988468	0	Café	New American Restaurant	Coffee Shop
54	MILNEBURG	-90.054245	30.019165	0	Cosmetics Shop	Pool Hall	Clothing Store
57	GENTILLY TERRACE	-90.052357	30.003706	0	Food Truck	Pool	Bar
64	PINES VILLAGE	-90.022144	30.019165	0	Seafood Restaurant	Nightlife Spot	Discount Store
71	IBERVILLE	-90.073686	29.959499	0	Bar	Hotel	Performing Arts Venue

Figure 5: The neighborhoods clustered with the French Quarter alongside their most common venue categories.

4 Discussion

A few interesting neighborhoods far from downtown end up in the same cluster as the French Quarter. The Gentilly neighborhood (Gentilly Terrace, Milneburg, and Pines Village) apparently

cluster with the French Quarter, as do the more uptown neighborhoods Hollygrove and Marlyville - Fontainebleau.

As can be seen in Figure 5, some of the neighborhoods which cluster with the French Quarter do not have a restaurant of any kind in their top 3 most common venues; namely, Algiers Point, Whitney, Hollygrove, Milneburg, and Iberville have market room for expansion in the restaurant industry. Similarly, the Whitney, Treme - Lafitte, Navarre, Milneburg, and Pines Village neighborhoods do not have Bars among their common venues (although some have bar-adjacent industries, like Nightclubs and Pool Halls). These neighborhoods would make natural candidates to explore opening a restaurant or bar, respectively.

Several of the clusters consist of single neighborhoods, particularly those on the outskirts of the city. It's not so surprising, for example, that the neighborhood "U.S. Naval Base" belongs to a unique cluster. This tendency for neighborhoods to cluster individually is reflected in the Elbow Method chart Figure 3; adding additional clusters continually reduces the distortion of the data relative to the model in a uniform fashion. This might be suggestive of extreme heterogeneity of the neighborhoods of New Orleans, as extremely diverse neighborhoods would be difficult to cluster effectively with any small number of clusters.

5 Conclusions

The analysis above suggests that, if one were to consider opening a new restaurant in a neighborhood similar to the French Quarter, one should consider the Algiers Point, Whitney, Hollygrove, Milneburg, and Iberville neighborhoods. It likewise suggests to consider opening a new bar in the Whitney, Treme - Lafitte, Navarre, Milneburg, and Pines Village neighborhoods. Each of these neighborhoods was clustered by the model with the French quarter while simultaneously demonstrating a lack of the corresponding venue type in the Foursquare data.

That being said, I would hesitate to act on this analysis without a more thorough investigation of the problem, perhaps incorporating additional data. Indeed, the elbow graph indicates that our particular choice of k-means clustering was not especially optimal, suggesting significant heterogeneity of the venue categories in a given neighborhood. In some sense this is a failure of the Foursquare API, which classifies Venue Categories in seemingly arbitrary ways (e.g., distinguishing "Cajun/Creole Restaurants" from "Seafood Restaurants", and "Bars" from "Gay Bars"). While the analysis above is a good start, one could perhaps develop more predictive models (and draw more meaningful conclusions) by categorizing these venues in such a way so as to increase similarities between neighborhoods. Furthermore, incorporating additional data, such as the cost of leasing or purchasing retail space in a neighborhood, could further inform a potential restaurateur's decision in an impactful way.