Predicting the Optimal Location and Type of Restaurant in Denver, CO

Jake Mitchell

June 22, 2020

1. Introduction

1.1 Background

Historically, around sixty percent of new restaurants fail within the first year of opening. Often, the most prominent reason is simply location — and management's general lack of awareness of the logistics of that location. Among other metrics, the location of a restaurant is the most critical component for any new restaurant entrepreneur to consider before opening. Additionally, in the last decade, cultural developments in the food and beverage industry have fluctuated greatly, effectively influencing consumer trends and taste. Determining the type of cuisine to be offered at an establishment can make or break the restaurant's success.

1.2 Problem

In Denver, Colorado alone, there are just over seven thousand restaurants, and over twenty different types of restaurant venues within a ten mile radius of the center of the city (according to Foursquare location data based on keyword search query: "restaurant"). However, there are still many available locations in the city for the opening of a small business. Using Python for data analysis, this project's objective is to predict what the optimal type of restaurant to open would be, and the associated best location for the greatest chance of success.

2. Data acquisition and cleaning

2.1 Data sources and origins

The data I acquired was derived exclusively from the Foursquare API, based on the keyword search query: 'restaurant'. Because I needed an origin point for my location data, I researched a central hotel address in downtown Denver. From there, I gathered all of venues within a ten-thousand-foot radius of the hotel's address. This returned around 50 food-based venues.

2.2 Data cleaning and filtering

Once I had the initial data, I began sorting through and cleaning. To do this, I defined functions and loops to keep only the columns from the data set that included relevant information to my objective (name, venue categories, latitude, longitude etc.). This process produced a "cleaner", more legible table that could be acted up for further analysis.

	name	categories	address	lat	lng	labeledLatLngs	distance	postalCode	cc	city	state	country	formattedAddress
0	La Fiesta Mexican Restaurant	Mexican Restaurant	2340 Champa St	39.753410	-104.984636	[{'label': 'display', 'lat': 39.75341022800047...	557	80205	US	Denver	CO	United States	[2340 Champa St, Denver, CO 80205, United States]
1	Rio Grande Mexican Restaurant	Mexican Restaurant	1525 Blake St	39.750223	-105.000036	[{'label': 'display', 'lat': 39.75022256969623...	1922	80202	US	Denver	CO	United States	[1525 Blake St (btwn 17th St & 15th St), Denve...
2	Randolph's Restaurant & Bar	American Restaurant	1776 Grant St	39.744531	-104.983480	[{'label': 'display', 'lat': 39.74453056889666...	1205	80203	US	Denver	CO	United States	[1776 Grant St (at 17th), Denver, CO 80203, Un...
3	The Corner Office Restaurant & Martini Bar	American Restaurant	1401 Curtis St	39.745555	-104.997231	[{'label': 'display', 'lat': 39.74555533418328...	1907	80202	US	Denver	CO	United States	[1401 Curtis St (at 14th St), Denver, CO 80202...

Figure 1. Cleaned and filtered venue data frame

Next, I needed a metric to figure out which of the venues in the data frame have been the most successful. Because I did not have access to any financial from venues, another alternative was used. Foursquare creates an average rating for each venue based on all of their user reviews and allows access through their API.

2c. Find the associated ratings for each venue category:

In [114]:
```
for i, j in zip(dataframe_filtered['id'], dataframe_filtered['categories']):
    venue_id = i
    url = 'https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&v={}'.format(venue_id, CLIENT_ID, CLIEN
    result = requests.get(url).json()
    try:
        print(j,":  ", result['response']['venue']['rating'])
    except:
        print(j,":  ", 'This venue has not been rated yet.')
```

```
Mexican Restaurant :   7.6
Mexican Restaurant :   7.6
American Restaurant :   6.8
American Restaurant :   7.6
Steakhouse :   7.4
Food Service :   This venue has not been rated yet.
Restaurant :   6.1
Indian Restaurant :   6.7
Lounge :   5.9
Japanese Restaurant :   8.6
Seafood Restaurant :   7.3
Mexican Restaurant :   This venue has not been rated yet.
Kitchen Supply Store :   6.3
Pizza Place :   7.1
Food :   This venue has not been rated yet.
African Restaurant :   This venue has not been rated yet.
Mexican Restaurant :   6.5
Mexican Restaurant :   7.5
Brewery :   7.4
Restaurant :   This venue has not been rated yet.
Chinese Restaurant :   This venue has not been rated yet.
Mexican Restaurant :   7.1
Ethiopian Restaurant :   6.6
Gastropub :   This venue has not been rated yet.
Mexican Restaurant :   This venue has not been rated yet.
Mexican Restaurant :   8.7
Ethiopian Restaurant :   7.5
Waste Facility :   This venue has not been rated yet.
Mexican Restaurant :   This venue has not been rated yet.
American Restaurant :   This venue has not been rated yet.
American Restaurant :   This venue has not been rated yet.
Mexican Restaurant :   6.7
Food :   This venue has not been rated yet.
Korean Restaurant :   This venue has not been rated yet.
Chinese Restaurant :   This venue has not been rated yet.
American Restaurant :   This venue has not been rated yet.
Food :   This venue has not been rated yet.
Food :   This venue has not been rated yet.
Office :   This venue has not been rated yet.
Mexican Restaurant :   This venue has not been rated yet.
Food :   This venue has not been rated yet.
Indian Restaurant :   7.5
Furniture / Home Store :   This venue has not been rated yet.
Food :   This venue has not been rated yet.
Breakfast Spot :   8.0
Mexican Restaurant :   This venue has not been rated yet.
Chinese Restaurant :   This venue has not been rated yet.
American Restaurant :   4.7
Food :   This venue has not been rated yet.
American Restaurant :   This venue has not been rated yet.
```

Figure 2. Associated Foursquare user ratings for each instance of a venue category

3. Methodology

3.1 Exploratory data analysis

As seen above, there were many instances of a certain venue not having an associated rating. To supplement, I eliminated all of these rows and constructed a new data frame with only venues that had at least one instance of a rating. From there, I sorted each venue category by rating, to see which types of restaurants in downtown Denver area were being rated the highest amount Foursquare users.

2e. Sort the venue categories by highest rating:

In [100]:
```
ratings_df.sort_values(by=['Avg_Rating'], ascending = False)
```

Out[100]:

	Venue Category	Avg_Rating
6	Japanese Restaurant	8.600000
11	Breakfast Spot	8.000000
2	Steakhouse	7.400000
9	Brewery	7.400000
7	Seafood Restaurant	7.300000
8	Pizza Place	7.100000
10	Ethiopian Restaurant	7.050000
0	Mexican Restaurant	6.462500
1	American Restaurant	6.366667
3	Restaurant	6.100000
5	Lounge	5.900000
4	Indian Restaurant	4.733333

Figure 3. Average rating per venue sorted descending

With the most popular restaurant type in downtown Denver determined ("Japanese Restaurant"), I began exploring the optimal location for the new venture. For preliminary analysis, I created a Folium map of all of the venue locations in the original data frame for the downtown Denver area.

3. Map of all local restaurants in the dowtown Denver area:

In [119]:

```
venues_map = folium.Map(location=[latitude, longitude], zoom_start=13) # generate map centred around the Hotel


# add the Italian restaurants as blue circle markers
for lat, lng, label in zip(dataframe_filtered.lat, dataframe_filtered.lng, dataframe_filtered.categories):
    folium.features.CircleMarker(
        [lat, lng],
        radius=5,
        color='blue',
        popup=label,
        fill = True,
        fill_color='blue',
        fill_opacity=0.6
    ).add_to(venues_map)

# display map
venues_map
```

Out[119]:

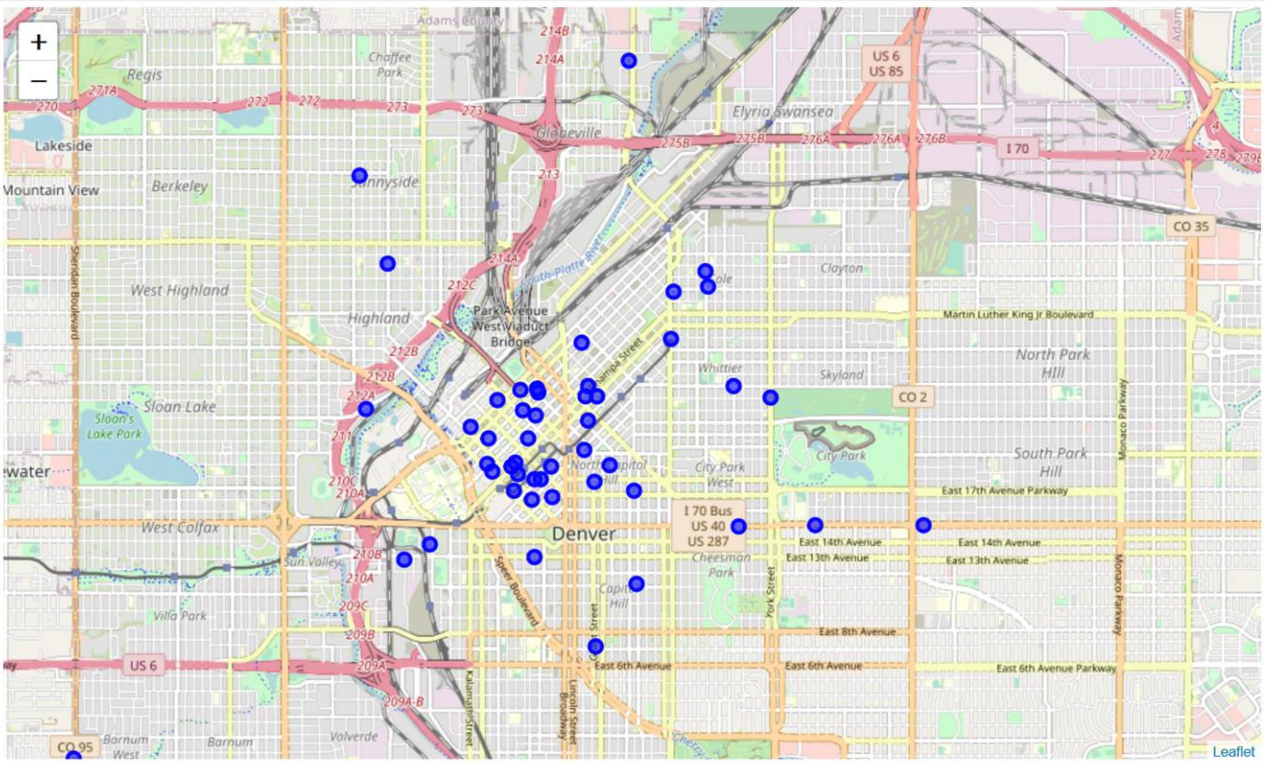

Figure 4. Average rating per venue sorted descending

It is also worth noting that I wrote algorithms to determine the highest “trending” venues in the area based on foot traffic, however at the time of this report (based on time of day), there was no available trending venue data in the Foursquare API.

The next step was to find out the total number of each type of venue in Denver. To perform this, I defined the function “getNearbyVenues” that would return all of the venues (restaurant or not) within a five-thousand-foot radius of the central hotel address. These nearby venues were then loaded into a separate data frame, where I was then able to count each instance of the associated venue category for each venue.

```
In [122]: denver_venues['Venue Category'].value_counts().head(15)

Out[122]: Mexican Restaurant      11
          American Restaurant     10
          Bar                      6
          Coffee Shop              5
          Sandwich Place           4
          Hotel                    4
          Marijuana Dispensary     4
          Thai Restaurant          3
          Sushi Restaurant         2
          Burger Joint             2
          Japanese Restaurant      2
          Noodle House             2
          Brewery                  2
          Seafood Restaurant       2
          Italian Restaurant       2
          Name: Venue Category, dtype: int64
```

Figure 5. Count of unique venue categories within five-thousand-foot radius of the central hotel

The depiction of the data frame category reveals some informative insights. The most common type of restaurant in the Denver area are “Mexican Restaurants” (eleven) followed by “American Restaurants” (ten). It is worth noting that there are only two “Japanese Restaurants” in the near surrounding area of the central hotel’s address, regardless of its highest rating among Foursquare users. This suggests there may be a shortage of Japanese restaurants when they are in fact popular and in-demand among the Foursquare users.

3.2 Inferential statistical analysis

Finally, to determine the optimal location for a new restaurant, I began clustering each of the venues in downtown Denver. To do this, I had to convert “denver_venues” populated with various strings into a binary, “one-hot encoded” structure. This would allow for appropriate clustering of the venues on a folium map. To cluster, I used the “kmeans” approach, clustering the data into seven different clusters in the Denver area.

From there, I created a Folium map with predicted optimal location based on the cluster locations of different venue categories and their associated Foursquare ratings

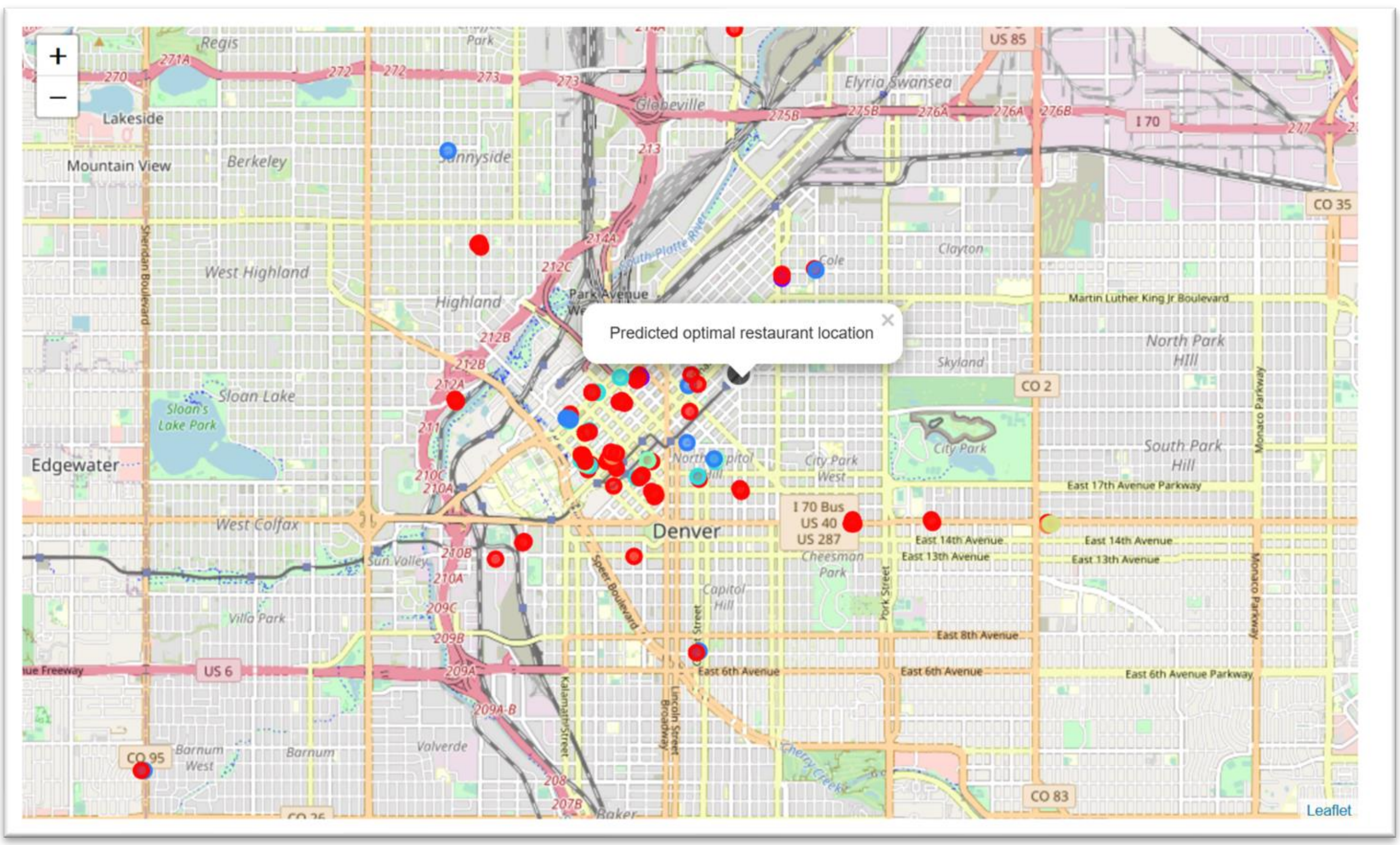

Figure 6. Predicted optimal restaurant location in for Denver area

4. Results

Based on the Foursquare user ratings and kmeans clustering, it appears the optimal type of restaurant and location are a Japanese oriented cuisine and between the confines of 27th Street, 26th Street and Welton Street respectively.

5. Discussion

Because the point associated on the cluster map is most likely associated with a building that is already occupied, opening a Japanese restaurant within one or two blocks would be adequate as well. It is worth noting that in some of exploratory analysis (Figure 5), there are some venues listed that are not of the food or restaurant variety, I took this into account when creating my folium maps by only including food oriented establishments. Additionally, as seen in Figure 2, there were many instances of venues without ratings that could have contributed to more accurate predictions had they had appropriate ratings. For a similar analysis, a more consistent data set with more accurate ratings could be used to produce better overall results.

5. Conclusion

The low quantity of Japanese restaurants in the Denver area paired with their high ratings among Foursquare users indicates that it is the specific type of restaurant venture to pursue in the area. In tandem with the Kmeans clustering of all of the local venues, the optimal location in downtown Denver is predicted, however more generally within a block or two of the predicted data point. Finally, it is worth reiterating a similar analysis can be done for a different location with more consistent and integrous data. Although powerful, the Foursquare API contained too many missing values to more accurately depict the objectives.

Figure 2. Associated ratings for instance of a venue category