Recommending Your Ideal Restaurant

with synopsis per Neighborhood

Cleveland, OH

Jeff Kortis

March 14, 2019

1. Introduction

1.1 Background

We'll start in the city of Cleveland, OH. A town rich with history that went through some hard times leading up to the turn of the century. Since 2009, the city has been rejuvenated with an ample amount of new construction projects bringing new and exciting venues to old buildings once barren.

In this notebook we will work through an individual looking to get recommended a restaurant within the city and its surrounding suburbs. This individual will be looking to have a recommendation built upon their current ratings within their Foursquare user profile.

We will use the Foursquare API to pull location data on all existing restaurants in the area and group them by category to perform a count on the number of options. From there we will plot the existing locations into neighborhoods where they are located to detect the best neighborhood choice for this new restaurant. The optimal choice will be a restaurant that is most closely aligned with the user's previous category choices and ratings.

In this notebook we will be using Foursquare location data to pull existing restaurant locations within the city of Cleveland, OH and surrounding suburbs. From there we will create a dataframe of the items and group them by type and neighborhood to create a visual map layout of the locations. Lastly, we will recommend a few restaurants that are closely related in rating to the user's current choices and neighborhood preferences.

1.2 Problem

As the city of Cleveland continues its resurgence upon the world and fine dining culture, there has been an abundance of new restaurants popping up throughout the neighborhoods. The varying rates of success of these locations can be attributed to several different factors such as neighborhood choice/location, saturation of the neighborhood with other similar establishments, or a perception of quality based on surrounding restaurants with a long-standing footprint within that neighborhood. With all these items being a factor to an individual looking to choose a dinner location, we aim to identify which neighborhoods provide the best options and which restaurants are most closely aligned with the user's previous choices and ratings.

1.3 Interest

Cleveland is becoming a notorious hotspot for fine dining with celebrity chefs the likes of Michael Symon and Dante opening numerous restaurants in the city. Anyone looking to explore the local eateries will be able to utilize this paper as a baseline for their evaluation. While I focus on venues with 'restaurant' in the target category, there are findings represented here that cover all category types available within the current Cleveland market. Any category type not covered within this paper could also be identified as a potential new footprint for someone looking to potentially open a restaurant.

2. Data Acquisition and Cleansing

2.1 Data Sources

We have three main data sources used for analysis. The first comes from a Wikipedia page listing Cleveland, OH neighborhoods here:
https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Cleveland.

The next is a CSV file listing geographic information for the scraped neighborhoods from the Wikipedia page that was populated through GeoHack search information and uploaded to Kaggle.

The final data source comes from the Foursquare API to pull venue information from the Cleveland neighborhoods.

2.2 Data Cleansing

One of the aspects is first to cleanse the data of neighborhoods based on the information at hand. GeoHack did not have latitude and longitude coordinates for two of the neighborhoods outlined within our data frame. The choice was made to drop the two neighborhoods of Nottingham, Ohio and St. Clair-Superior. The initial data frame is displayed in Figure 1.1.

Figure 1.1

	Neighborhood	Latitude	Longitude
0	Asiatown, Cleveland	41.508833	-81.680417
1	Bellaire-Puritas, Cleveland	41.433682	-81.800140
2	Broadway-Slavic Village	41.458056	-81.644722
3	Brooklyn Centre	41.453446	-81.699402
4	Buckeye-Shaker	41.483889	-81.590556
5	Campus Disctict	41.497778	-81.670000
6	Central, Cleveland	41.500000	-81.666667
7	Clark-Fulton	41.666670	-81.716667
8	Collinwood	41.558000	-81.569000
9	Detroit-Shoreway	41.479062	-81.737795
10	Downtown Cleveland	41.498889	-81.689722
11	East 4th Street District (Cleveland)	41.498889	-81.690000
12	Edgewater, Cleveland	41.431559	-81.702332
13	Fairfax, Cleveland	41.483889	-81.590556
14	The Flats	41.492000	-81.696000
15	Glenville, Cleveland	41.533347	-81.616588
16	Goodrich-Kirtland Park	41.512500	-81.663056
17	Hough, Cleveland	41.512334	-81.635213
18	Industrial Valley	41.483333	-81.666667
19	Kamm's Corners	41.444286	-81.818492
20	Kinsman, Cleveland	41.558000	-81.569000
21	Lee-Miles	41.440140	-81.564786
22	Nine-Twelve District	41.499444	-81.685833
23	Nottingham, Ohio	NaN	NaN
24	Ohio City, Cleveland	41.483611	-81.710278
25	Old Brooklyn	41.431559	-81.702332
26	St. Clair-Suprerior	NaN	NaN
27	Stockyards, Cleveland	41.483889	-81.590556
28	Tremont, Cleveland	41.473611	-81.688611
29	Union-Mills Park	41.454889	-81.614389
30	University Circle	41.508611	-81.605278

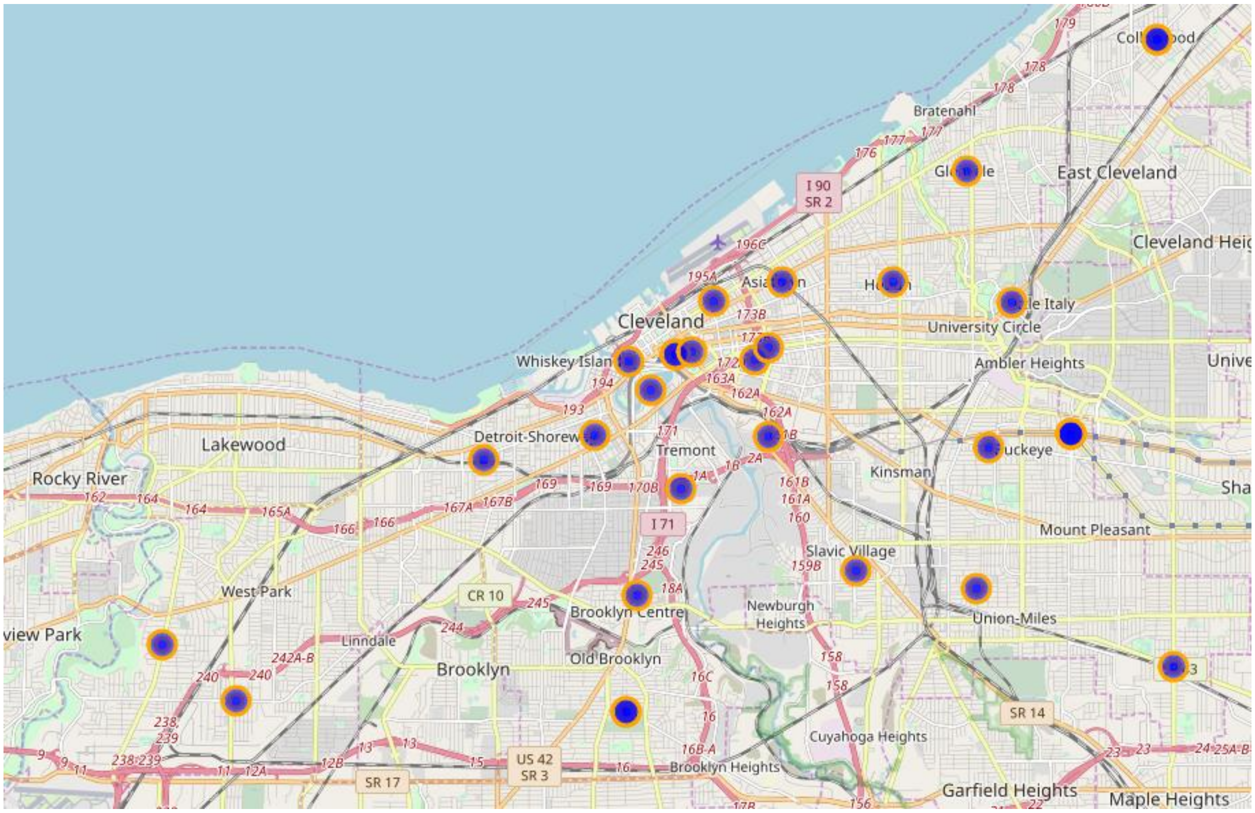
The cleansed data set it now represented below in Figure 1.2.

Figure 1.2

	Neighborhood	Latitude	Longitude
0	Asiatown, Cleveland	41.508833	-81.680417
1	Bellaire-Puritas, Cleveland	41.433682	-81.800140
2	Broadway-Slavic Village	41.458056	-81.644722
3	Brooklyn Centre	41.453446	-81.699402
4	Buckeye-Shaker	41.483889	-81.590556
5	Campus Disctict	41.497778	-81.670000
6	Central, Cleveland	41.500000	-81.666667
7	Clark-Fulton	41.666670	-81.716667
8	Collinwood	41.558000	-81.569000
9	Detroit-Shoreway	41.479062	-81.737795
10	Downtown Cleveland	41.498889	-81.689722
11	East 4th Street District (Cleveland)	41.498889	-81.690000
12	Edgewater, Cleveland	41.431559	-81.702332
13	Fairfax, Cleveland	41.483889	-81.590556
14	The Flats	41.492000	-81.696000
15	Glenville, Cleveland	41.533347	-81.616588
16	Goodrich-Kirtland Park	41.512500	-81.663056
17	Hough, Cleveland	41.512334	-81.635213
18	Industrial Valley	41.483333	-81.666667
19	Kamm's Corners	41.444286	-81.818492
20	Kinsman, Cleveland	41.558000	-81.569000
21	Lee-Miles	41.440140	-81.564786
22	Nine-Twelve District	41.499444	-81.685833
24	Ohio City, Cleveland	41.483611	-81.710278
25	Old Brooklyn	41.431559	-81.702332
27	Stockyards, Cleveland	41.483889	-81.590556
28	Tremont, Cleveland	41.473611	-81.688611
29	Union-Mills Park	41.454889	-81.614389
30	University Circle	41.508611	-81.605278
31	Warehouse District, Cleveland	41.497500	-81.701667
32	Woodland Hills, Cleveland	41.481389	-81.611389

Figure 1.3 shows a map of the Cleveland neighborhoods.

Figure 1.3



2.3 Feature Selection

Our features we've chosen are neighborhood, restaurant name, restaurant category, and restaurant rating.

I then decided to only focus on neighborhoods that had two features with certain parameters to best represent our target result. We chose to eliminate neighborhoods with less than 10 total restaurants, had zero venues with 'restaurant' within the category, or had an aggregate restaurant rating (that is the mean of all restaurant ratings within the neighborhood) of less than 4.5 via Foursquare. The basis for this is that neighborhoods with zero venues that had 'restaurant' in their categories are not an area we would consider to begin with. The aggregate or mean rating is related to the quality of the restaurants within the neighborhood. Based on the aggregate rating, we can make a general assumption that neighborhoods with less than adequate ratings are of lower quality and not worthy of our recommendation. We also chose to eliminate neighborhoods with less than 10 total restaurants as that was a good benchmark to lend credibility to the neighborhood as a location that values quality cuisine.

Using the Foursquare API, we retrieved a list of restaurants within the Cleveland neighborhoods. After associating a radius to the neighborhood list, I then looped through our list and appended a neighborhood/s to the restaurant for association. Our final list is the data frame image below.

3. Exploratory Data Analysis

We first wanted to explore the dataset by generating a few plots that represent the total number of restaurants within each neighborhood and by category of restaurant.

3.1 Relationship between Neighborhood and Restaurant Type

What we found is that the **East 4th Street District** had the highest concentration of restaurants available via Foursquare at **44%.** While that neighborhood consisted of nearly half of all the available restaurants the greater downtown metro which encompasses the top 4 neighborhoods was the most represented neighborhood within our set.

Neighborhoods — A String

East 4th Street District (Clevelan...	44%
Ohio City, Cleveland	20%
Nine-Twelve District	16%
Downtown Cleveland	8%
Other (3)	12%

Valid	25	100%
Mismatched	0	0%
Missing	0	0%
Unique	7	
Most Common	East 4th S	44%

Based on categories, we found that Cleveland is a very diverse area when it comes to restaurant categories with the largest category being **Italian Restaurants** which consisted of **12%** of our data.

Category — A String

Italian Restaurant	12%
Mexican Restaurant	8%
New American Restaurant	8%
Food	8%
Other (16)	64%

Valid	25	100%
Mismatched	0	0%
Missing	0	0%
Unique	20	
Most Common	Italian Re:	12%

After looking into the options, we found that **East 4th Street District** encompassed 1 restaurant from every category on our list making it the most diverse but also providing the most options of restaurants to a user.

3.2 Relationship between Neighborhood and Mean Rating

After looking at each neighborhood and their corresponding categories, we next examined the ratings of each of those categories.

Below you'll see that the average overall rating of a restaurant in Cleveland came in at **7.95** with a standard deviation of **0.75**, you'll also see that the lowest rating was a 6 and the highest a **9.3** for **Lola Bistro**.

Rating

\# Decimal

Valid	19	76%
Mismatched	0	0%
Missing	6	24%
Mean	7.95	
Std. Deviation	0.75	
Quantiles	6	Min
	7.6	25%
	8	50%
	8.4	75%
	9.3	Max

With that information we also looked at specific sects of neighborhoods and found that **East 4th Street District** was the top-rated neighborhood with an average rating on its restaurants of **8.35**.

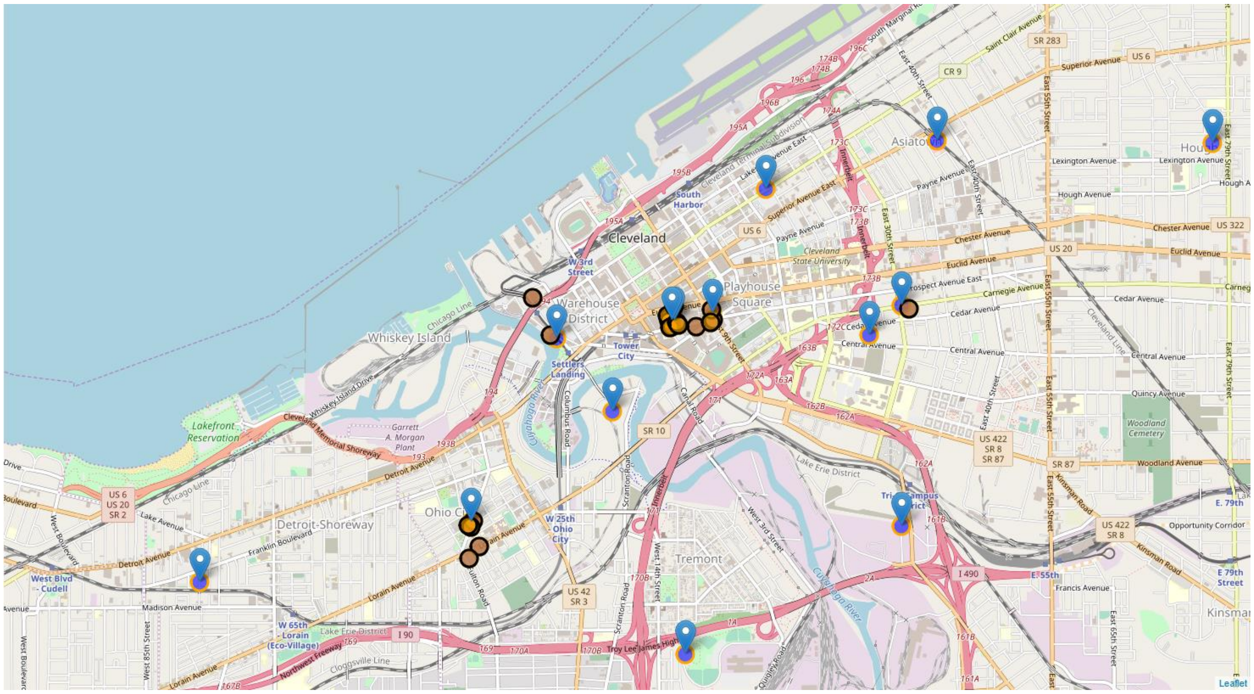
This gives us a few ideas around which neighborhoods have the most established venues, which categories are currently represented the strongest based on volume, and which neighborhoods have the highest aggregate rating for all their restaurants.

3.3 Creation of Restaurants/Neighborhoods/Rating data frame

We had 2 main data frames to start with, the Neighborhoods of Cleveland, OH and the Restaurants of those neighborhoods with their rating from Foursquare.

After some exploratory analysis that included mapping and viewing the data frames graphically, I decided to combine the 2 data frames into 1 master data frame that would be our final data frame for which our recommender system will be built. This data frame consisted of the restaurant name, rating, neighborhood/s with its geographical coordinates for mapping and is the result show below.

	Name	Category	Latitude	Longitude	Neighborhoods	id	Rating
0	Saigon Restaurant & Bar	Vietnamese Restaurant	41.498871	-81.690149	East 4th Street District (Cleveland)	4ae1cb35f964a520738721e3	8.3
1	Zocalo Mexican Grill & Tequileria	Mexican Restaurant	41.498656	-81.690084	East 4th Street District (Cleveland)	4ad4bff2f964a520b8e920e3	6.5
2	Society Lounge	Cocktail Bar	41.498800	-81.690069	East 4th Street District (Cleveland)	51410105e4b00f288c5fbd54	7.6
3	Butcher and the Brewer	New American Restaurant	41.498938	-81.690169	East 4th Street District (Cleveland)	53f66de8498ea900655dd238	8.4
4	Mabel's BBQ	BBQ Joint	41.498938	-81.690169	East 4th Street District (Cleveland)	56c48226cd104835eaebc79f	8.5
5	The Greenhouse Tavern	Gastropub	41.499078	-81.690295	East 4th Street District (Cleveland)	4ace8227f964a520f5d020e3	8.7



4. Recommender System

4.1 Collaborative Filtering

Now, time to start our work on recommendation systems.

The first technique we're going to look at is called **Collaborative Filtering**, which is also known as **User-User Filtering**. As hinted by its alternate name, this technique uses other users to recommend items to the input user. It attempts to find users that have similar preferences and opinions as the input and then recommends items that they have liked to the input. There are several methods of finding similar users (Even some making use of Machine Learning).

4.1.1 Pearson Correlation Function or User-User

Our recommender system, not surprisingly, has select the top 5 restaurants this user would go to based on their input to all be in the **East 4th Street District**. Our systems top choice for a restaurant this user should visit is the **New American** restaurant of **Lola Bistro** which was in the highest rated neighborhood and was the highest rated restaurant.

Out[39]:

```
(Restaurant ID                                   7
 Name                                   Lola Bistro
 Category                   New American Restaurant
 Latitude                                   41.4988
 Longitude                                 -81.6902
 Neighborhoods    East 4th Street District (Cleveland)
 id                        4ad4bff1f964a5208be920e3
 Rating                                         9.3
 Name: 6, dtype: object, [3], [8], [5], [1])
```

4.1.2 Content-Based or Item-Item

While there was insufficient data to produce a content-based recommender system, should we have had the data we could have compared the two systems to determine the best approach.

5. Conclusions

In reviewing the data I've found that not a lot of venues are tagged within the Cleveland metro area with the category of 'restaurant'. Our list was a cursory portion in comparison to my familiarity with the location of choice. That being said, this recommender system would not, in my opinion, be a good choice for restaurant selection. We would need to explore new/other API connections to other sites potentially like GrubHub, Opentable, etc. that have a fuller list of options and more up to date details.

6. Future Directions

From here we could also delve deeper into each restaurant category by dividing the types into sub-categories offered through the Foursquare API. A dataframe that consists of the sub-types of Italian restaurants or neighborhoods could help an individual looking to target a specific sect of Italian or neighborhood. The future synopsis could continue to do into deeper detail should the individual choose.