

# **Comparison of High-rated Restaurants in Raleigh and Charlotte**

**Priyansha Choukse**

**March 28, 2020**

## **1. INTRODUCTION**

The city of Raleigh, NC and Charlotte, NC are the major cities of North Carolina. They have amazing restaurants of different types. Being a foodie, I am interested in exploring dining in both cities. Therefore, I am comparing restaurants between these two cities to see best places to go to in both cities and what type of restaurant are popular in both cities. This analysis can give many local reader lists of restaurant based on their interested category and rating.

## **2. BUSINESS PROBLEM:**

Analysis is done based on Foursquare likes, restaurant category and geographic location data for restaurants in Charlotte and Raleigh. Then with the help of visualization and cluster analysis will discover which category is popular in both cities and their specialties based on Foursquare user feedback.

## **3. ABOUT DATA**

The Foursquare API is used to pull the following location data on restaurants in Raleigh, NC and Charlotte, NC:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- No. of Likes

## **4. DATA ACQUISITION APPROACH**

For Data Acquisition, Geolocator latitude and longitude coordinates for Raleigh, NC and Charlotte, NC is used. Then with the help of Foursquare API I extracted list of all venues in Raleigh, NC and Charlotte, NC and filtered for venues of category Restaurant.

### **Algorithm Used:**

Used gathered data to create a k-means clustering that groups restaurants into different clusters so that comparison can be made to see which type popular and which restaurants are the best to eat at in Raleigh and Charlotte.

## 5. DATA EXTRACTION AND METHODOLOGY

With the help of geolocator, extracted the geographical coordinates of Raleigh and Charlotte. Then created URL for all the venues in Charlotte and Raleigh. This URL was used to request JSON file using beautiful soup library. The data extracted by parsing json file was stored in form of dataframe. The venues pulled out were filtered using their category.

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>
<b>0</b>	Blumenthal Performing Arts Center	4bad5829f964a52071483be3	Performing Arts Venue	35.227930	-80.841951
<b>1</b>	Belk Theater	4b058640f964a520495a22e3	Concert Hall	35.227711	-80.841663
<b>2</b>	The Capital Grille	4b05863ef964a520f45922e3	American Restaurant	35.228216	-80.841974
<b>3</b>	Vapiano	4e2ffb921838f1c552ca5f70	Italian Restaurant	35.226234	-80.844124
<b>4</b>	The Punch Room	549af5b3498e3a660685323b	Hotel Bar	35.226354	-80.841747

*Fig 1. Screenshot of Data for Charlotte*

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>
<b>0</b>	North Carolina Museum of Natural Sciences	4ad4c009f964a520eeec20e3	Science Museum	35.782217	-78.639375
<b>1</b>	North Carolina Museum of History	4ad4c009f964a520edec20e3	History Museum	35.781516	-78.638747
<b>2</b>	Gallo Pelón	53c2e7a1498e06db62294821	Cocktail Bar	35.779320	-78.638517
<b>3</b>	Centro	4ad90386f964a5201d1721e3	Mexican Restaurant	35.779236	-78.638264
<b>4</b>	Raleigh Raw	56d20a97498e40011a57fbba	Juice Bar	35.778308	-78.639610

*Fig 2. Screenshot of Data for Raleigh*

Venues having restaurant category was filtered and rest all categories were dropped. Total venues of restaurant type were 25 for Charlotte and 27 for Raleigh. Foursquare API was again called to pull out total number of likes for each restaurant type venue with the help of venue ID. These number of likes were stored as a new variable in the two datasets.

To classify venues based on their user rating, histogram was built for both datasets to see the distribution of Total likes variable. And function was defined having four categories based on bins value obtained for both datasets. Categories were - high rated, above average rated, below average rated and poor rated. These labels were added to the dataset in the Rating column.

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>	<b>total likes</b>	<b>Rating</b>
<b>2</b>	The Capital Grille	4b05863ef964a520f45922e3	American Restaurant	35.228216	-80.841974	96	high rated
<b>3</b>	Vapiano	4e2ffb921838f1c552ca5f70	Italian Restaurant	35.226234	-80.844124	208	high rated
<b>5</b>	Dandelion Market	4b466381f964a5206f1f26e3	Tapas Restaurant	35.228656	-80.842479	236	high rated
<b>9</b>	Aria Tuscan Grill	4bb14a46f964a520fa893ce3	Italian Restaurant	35.226615	-80.842317	85	high rated
<b>11</b>	Stoke	57a56d1f498e5952a14f86f4	Restaurant	35.228057	-80.843475	33	below avg rated

Fig 3. Screenshot of Final Dataset for Charlotte

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>	<b>total likes</b>	<b>Rating</b>
<b>3</b>	Centro	4ad90386f964a5201d1721e3	Mexican Restaurant	35.779236	-78.638264	125	high rated
<b>6</b>	Sitti	4ad93789f964a520261921e3	Mediterranean Restaurant	35.778237	-78.638266	231	high rated
<b>8</b>	Gravy	4a592915f964a520d6b81fe3	Italian Restaurant	35.778482	-78.638143	139	high rated
<b>10</b>	Death & Taxes	54a1adbe498e54b4407ce331	American Restaurant	35.778245	-78.640375	64	above avg rated
<b>15</b>	Trophy Tap & Table	588663ddd60c8b2e0e1b1527	American Restaurant	35.777417	-78.638184	46	below avg rated

Fig 4. Screenshot of Final Dataset for Raleigh

## 6. ANALYSIS & RESULTS

Cluster analysis was done using k-means clustering algorithm that grouped venues into 4 clusters, these clusters would help people get a better insight when they are looking to have nice eating experience in Raleigh and Charlotte.

	<b>lat</b>	<b>lng</b>	<b>total likes</b>
<b>count</b>	27.000000	27.000000	27.000000
<b>mean</b>	35.777776	-78.639461	104.074074
<b>std</b>	0.002151	0.003450	95.887725
<b>min</b>	35.774581	-78.646165	14.000000
<b>25%</b>	35.776105	-78.642841	43.000000
<b>50%</b>	35.777227	-78.638264	64.000000
<b>75%</b>	35.778859	-78.636781	132.500000
<b>max</b>	35.782864	-78.634479	453.000000

Fig 5. Descriptive Analysis of Raleigh's Dataset

	lat	lng	total likes
count	25.000000	25.000000	25.000000
mean	35.227405	-80.842793	79.480000
std	0.001253	0.001957	69.044140
min	35.224981	-80.847295	7.000000
25%	35.226428	-80.843548	19.000000
50%	35.227504	-80.842806	44.000000
75%	35.228216	-80.841523	112.000000
max	35.229240	-80.839710	236.000000

Figure 6. Descriptive Analysis of Charlotte's Dataset

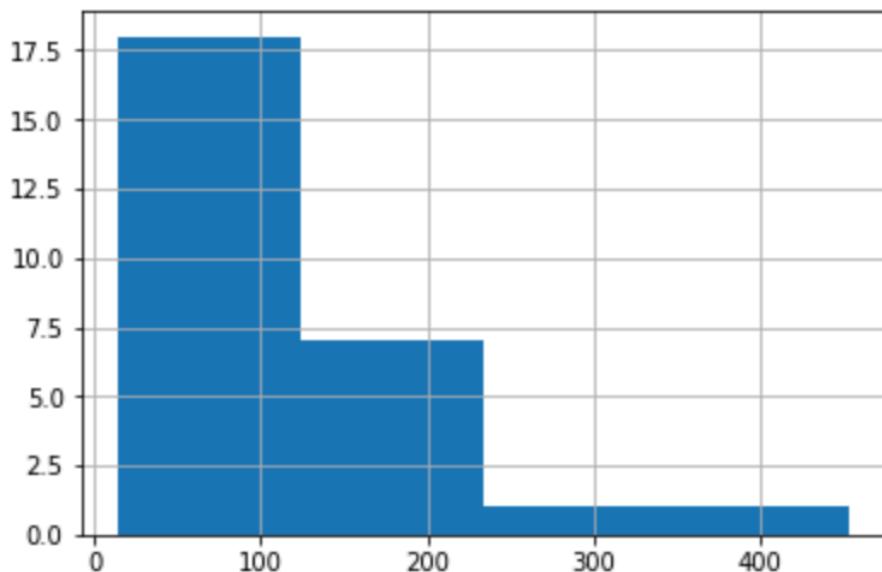


Fig 7. Distribution of total likes of restaurants in Raleigh

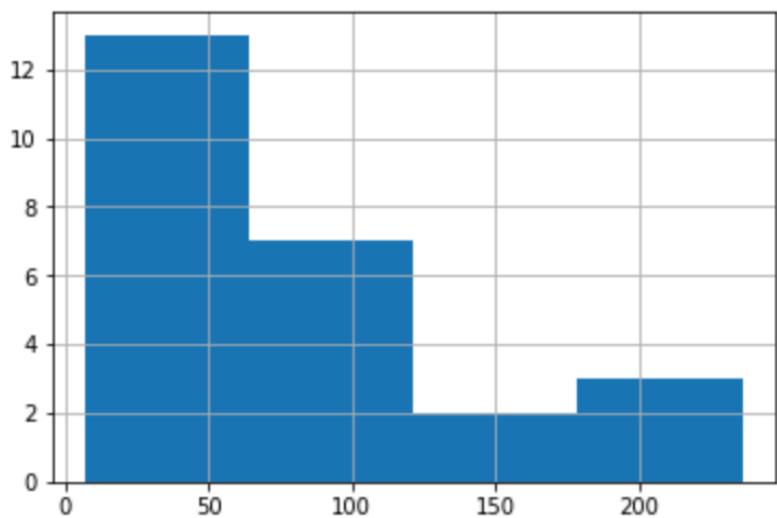


Fig 8. Distribution of total likes of restaurants in Charlotte

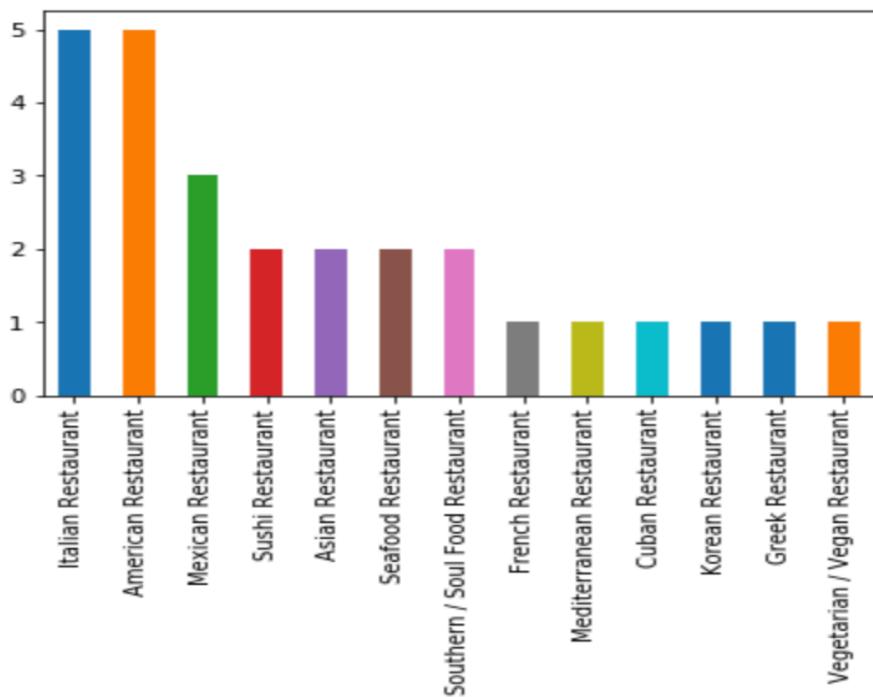
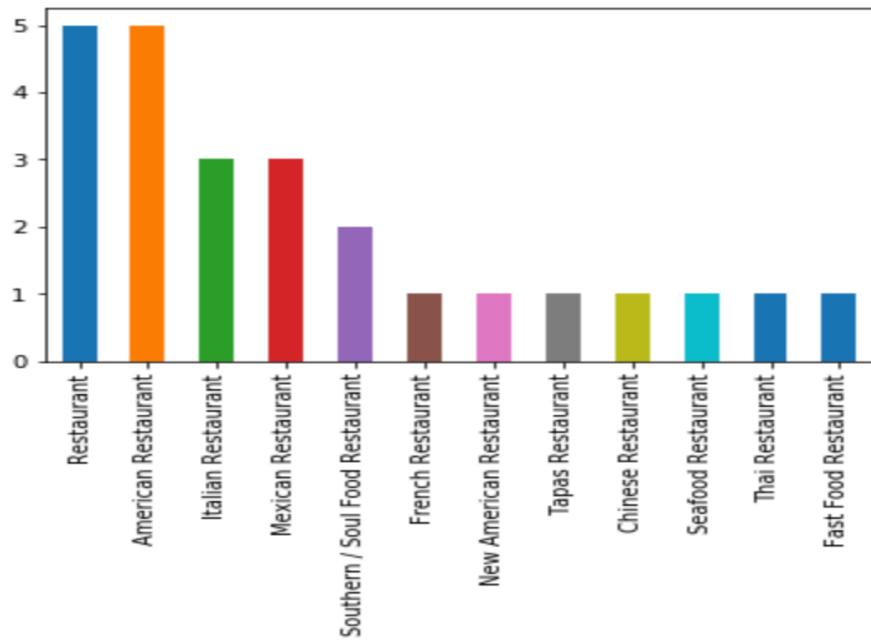


Fig 9. Popularity of Restaurant Categories in Raleigh



*Fig 10. Popularity of Restaurant Categories in Charlotte*

In case of Charlotte,

Cluster 3 and Cluster 0 - High rated restaurants

Cluster 2- Poor rated restaurants

Cluster 1- Below Average rated restaurants

There were no restaurants found in above average rating category.

In case of Raleigh,

Cluster 0 - High rated restaurants

Cluster 1- Above Average rated restaurants

Cluster 2- Below Average rated restaurants

Cluster 3 - Poor rated restaurants

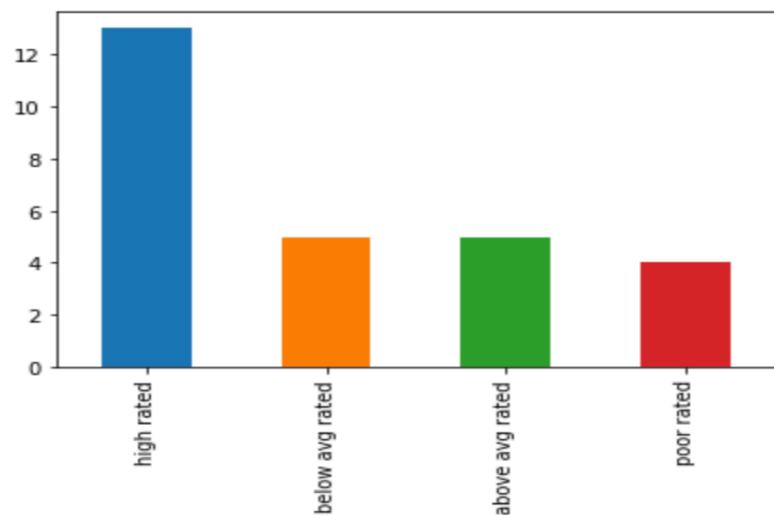


Fig 11. Frequency Distribution of Restaurant based on ratings in Raleigh

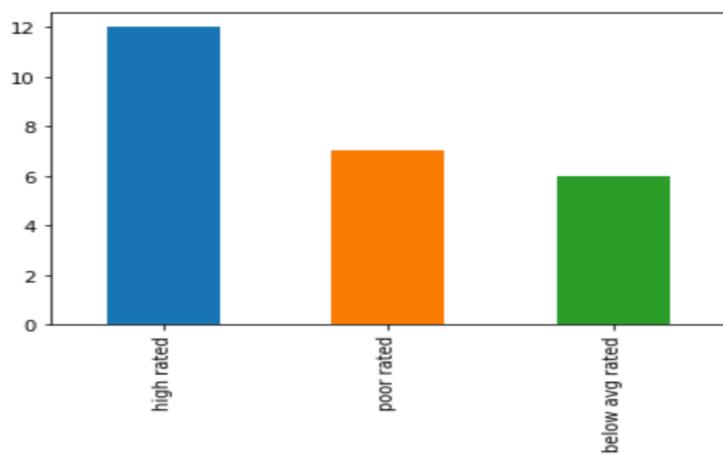


Fig 12. Frequency Distribution of Restaurant based on ratings in Charlotte

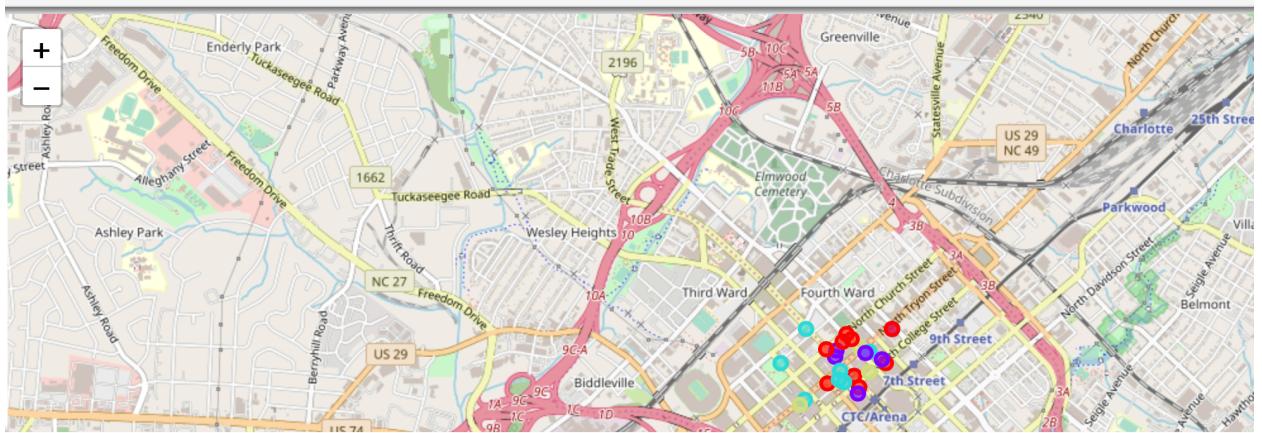


Fig 13. Folium map obtained for Charlotte city.

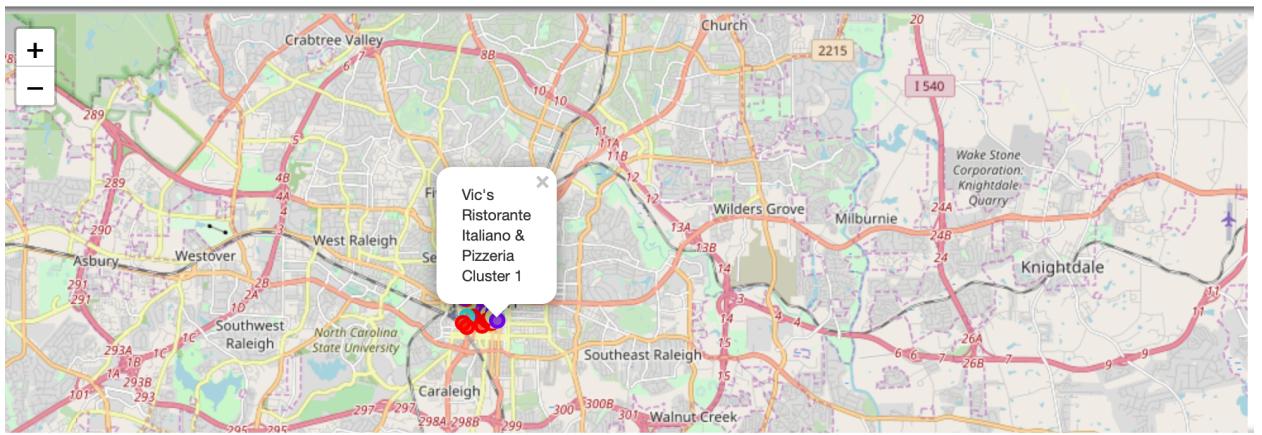


Fig 14. Folium map obtained for Raleigh city.

These clusters were plotted on Map as well using folium library.

## 7. CONCLUSION

It was observed that Italian type of restaurants are common in Raleigh NC. While in charlotte, general type restaurant and American type restaurants are common. Also, in Raleigh there are more high rated restaurants as compared to Charlotte based on data pulled from Foursquare API.

Popular high rated restaurants found in both cities are following:

For Charlotte,

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>	<b>total likes</b>	<b>Rating</b>	<b>label</b>
<b>2</b>	The Capital Grille	4b05863ef964a520f45922e3	American Restaurant	35.228216	-80.841974	96	high rated	3
<b>3</b>	Vapiano	4e2ffb921838f1c552ca5f70	Italian Restaurant	35.226234	-80.844124	208	high rated	0
<b>5</b>	Dandelion Market	4b466381f964a5206ff1f26e3	Tapas Restaurant	35.228656	-80.842479	236	high rated	0
<b>9</b>	Aria Tuscan Grill	4bb14a46f964a520fa893ce3	Italian Restaurant	35.226615	-80.842317	85	high rated	0
<b>13</b>	Basil Thai Cuisine	4b09c3fef964a520311d23e3	Thai Restaurant	35.228941	-80.842806	112	high rated	0
<b>22</b>	Rooster's Wood Fired Kitchen II	4f516ba8e4b063296bf04765	American Restaurant	35.226856	-80.841229	90	high rated	3
<b>24</b>	5 Church	4f8b5454e4b00a6263a38774	New American Restaurant	35.228548	-80.843054	212	high rated	0
<b>25</b>	The King's Kitchen	4bd07af1caff952185dfcef0	Southern / Soul Food Restaurant	35.228093	-80.844206	98	high rated	0
<b>39</b>	Mert's Heart & Soul Restaurant	4b05863ff964a520345a22e3	Southern / Soul Food Restaurant	35.227318	-80.840130	168	high rated	0
<b>44</b>	Vida Cantina	4c23a655f3859c7420c56488	Mexican Restaurant	35.226011	-80.841920	137	high rated	0
<b>64</b>	Duckworth's Grill & Taphouse	54ecae87498ec256aee9450d	Restaurant	35.229240	-80.839710	121	high rated	0
<b>76</b>	Mimosa Grill	4b05863bf964a520775922e3	American Restaurant	35.224981	-80.845958	101	high rated	3

For Raleigh,

	<b>name</b>	<b>id</b>	<b>categories</b>	<b>lat</b>	<b>lng</b>	<b>total likes</b>	<b>Rating</b>	<b>label</b>
<b>3</b>	Centro	4ad90386f964a5201d1721e3	Mexican Restaurant	35.779236	-78.638264	125	high rated	0
<b>6</b>	Sitti	4ad93789f964a520261921e3	Mediterranean Restaurant	35.778237	-78.638266	231	high rated	0
<b>8</b>	Gravy	4a592915f964a520d6b81fe3	Italian Restaurant	35.778482	-78.638143	139	high rated	0
<b>16</b>	Beasley's Chicken + Honey	4e091acc1f6e21103396e069	Southern / Soul Food Restaurant	35.776968	-78.638175	453	high rated	0
<b>17</b>	Bida Manda Laotian Restaurant and Bar	4ee404309911944900b07ad5	Asian Restaurant	35.777227	-78.636812	262	high rated	0
<b>30</b>	Sono	4ad90968f964a520651721e3	Sushi Restaurant	35.775919	-78.638562	146	high rated	0
<b>32</b>	Taverna Agora	5577a05f498e75efda31c6a9	Greek Restaurant	35.780776	-78.644406	93	high rated	0
<b>40</b>	Whiskey Kitchen	57a7bf38498ea30d9431bde1	Southern / Soul Food Restaurant	35.776861	-78.642111	89	high rated	0
<b>53</b>	Oak City Meatball Shoppe	516798e0e4b0cda68217b5ed	Italian Restaurant	35.775520	-78.637234	126	high rated	0
<b>63</b>	42nd St Oyster Bar	4ae6edcaf964a5208ea721e3	Seafood Restaurant	35.782864	-78.646165	202	high rated	0
<b>81</b>	José and Sons	52251fd611d21c8d037f3127	Mexican Restaurant	35.775410	-78.644782	149	high rated	0
<b>91</b>	Fiction Kitchen	50a5647be4b079eb079361fa	Vegetarian / Vegan Restaurant	35.774581	-78.643745	123	high rated	0
<b>94</b>	Jimmy V's Osteria + Bar	51eeeca15498e3ffb88aaa1fe	Italian Restaurant	35.774977	-78.639655	102	high rated	0

## 8. LIMITATIONS & FUTURE SCOPE

For more accurate analysis, more data about restaurants is needed. Many restaurants are not in the list. In order to overcome this problem, it is recommended to pull data from other APIs like yelp reviews and google reviews. Other classification machine learning models like SVM can be applied on the extracted data.