

Restaurant Location Analysis

IBM Applied Data Science Capstone

BY: LUANLUAN XU

JULY, 2019

1. Introduction

- ▶ In the modern society, More and more people choose to eat out or increase the frequency of eating out. Especially in modern metropolises like New York City, restaurants are everywhere, opening and closing every day. But why some restaurants after opening can be filled with customers and make a lot of money, while others have slow bussiness. Location is one of the important factors.

1. Introduction

► 1.1 Bussiness Problem

So how to decide a location-selection?

Based on the data of **2,698** restaurants in five Boroughs of New York City, this project aims to study the influence of household income level, population density, existing competitors and other factors on the location-selection, and find out the patterns.

► 1.2 Targeted Audiences

owners or investors who want to open a restaurant.

2. Data Section

▶ **Tools used for data collection/cleaning/analysis.**

- ▶ Foursquare API
- ▶ Web scraping
- ▶ Python
- ▶ Machine Learning(K-means)
- ▶ Folium

2. Data Section

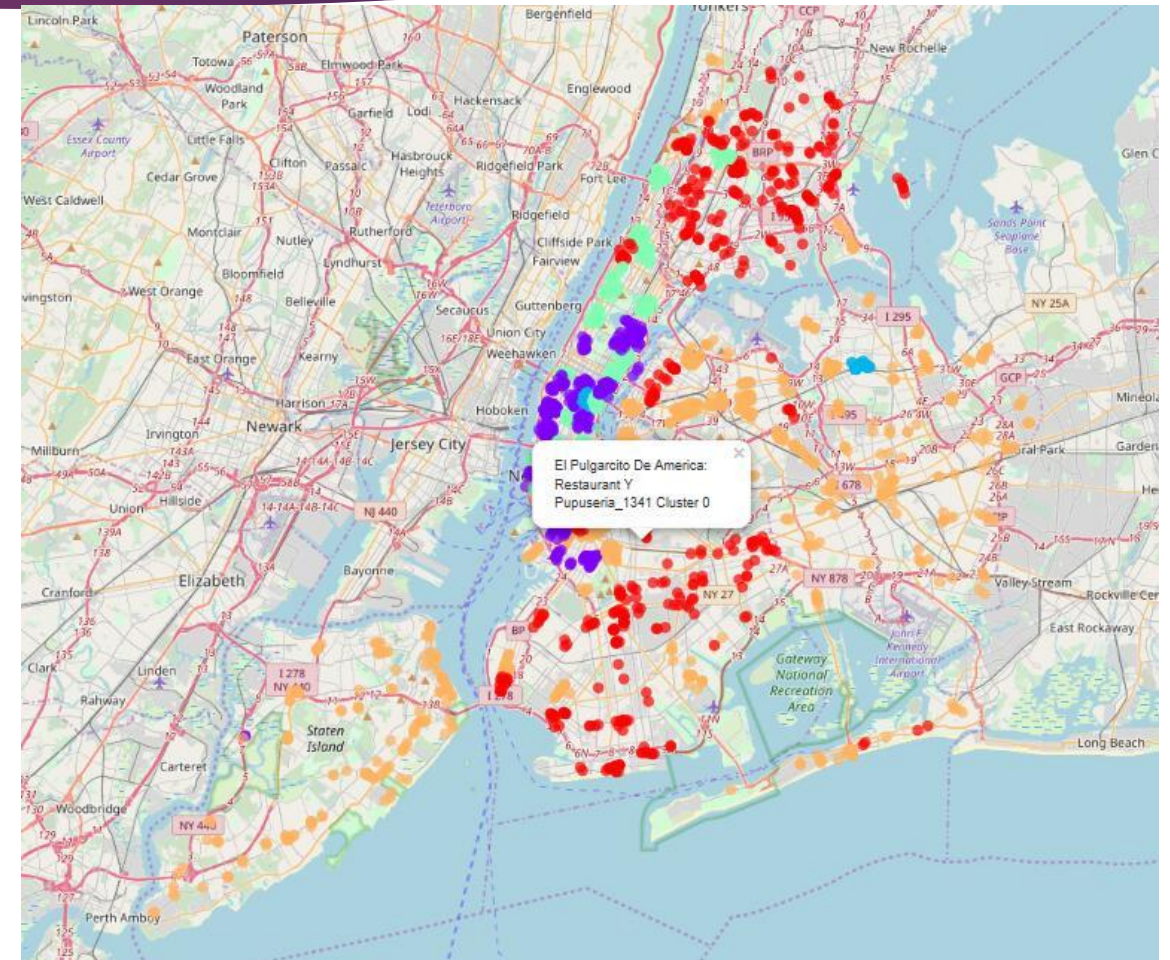
- ▶ Data from all sources are combined into one table.
- ▶ There are 2698 restaurant samples and 11 features.

(2698, 11)

	Restaurant name	Neighborhood	Borough	Median Household Income	Neighborhood Latitude	Neighborhood Longitude	Pop./km2	Venue Category	Venue Latitude	Venue Longitude	Competitors
0	White Castle_0	Allerton	Bronx	37816.894737	40.865788	-73.859319	12149.0	Fast Food Restaurant	40.866065	-73.862307	3
1	Chef King_1	Allerton	Bronx	37816.894737	40.865788	-73.859319	12149.0	Chinese Restaurant	40.865561	-73.856752	3
2	Internacional Restaurant & Deli_2	Allerton	Bronx	37816.894737	40.865788	-73.859319	12149.0	Spanish Restaurant	40.863809	-73.856640	3
3	Il Sogno_758	Annadale	Staten Island	66764.200000	40.538114	-74.178549	2593.0	Restaurant	40.541286	-74.178489	3
4	Diesel Bagels_759	Annadale	Staten Island	66764.200000	40.538114	-74.178549	2593.0	American Restaurant	40.540373	-74.177374	3

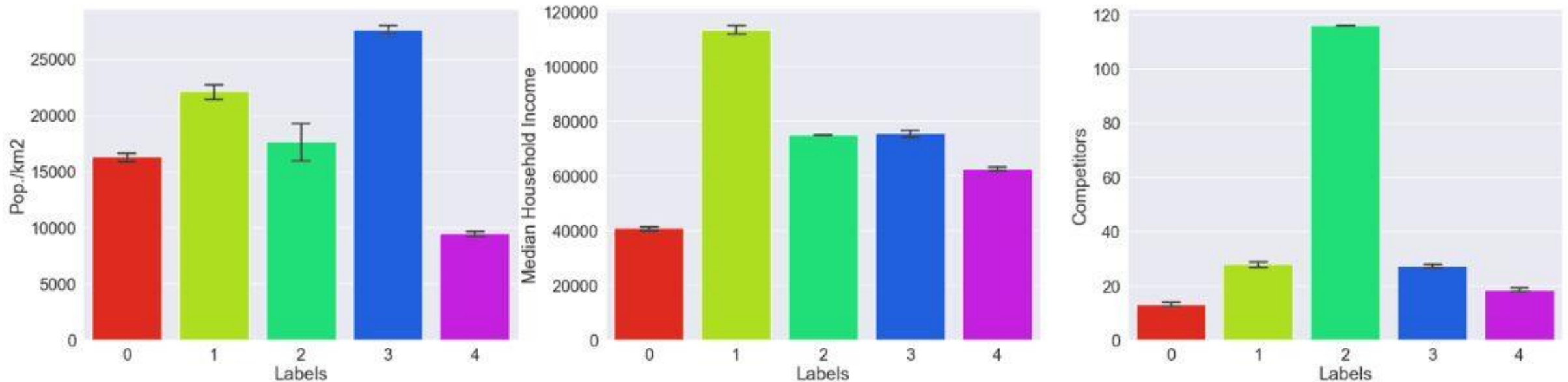
3.1 Using k-means for Restaurants Segmentation

- ▶ I create a new dataframe which only contains 4 features:
 - ▶ 'Restaurant name',
 - ▶ 'Median Household Income',
 - ▶ 'Pop./km2' (Population density)
 - ▶ 'Competitors'.
-
- ▶ I run k-means to cluster the restaurants into 5 clusters, and visualize all restaurants in NYC in a map using Folium.



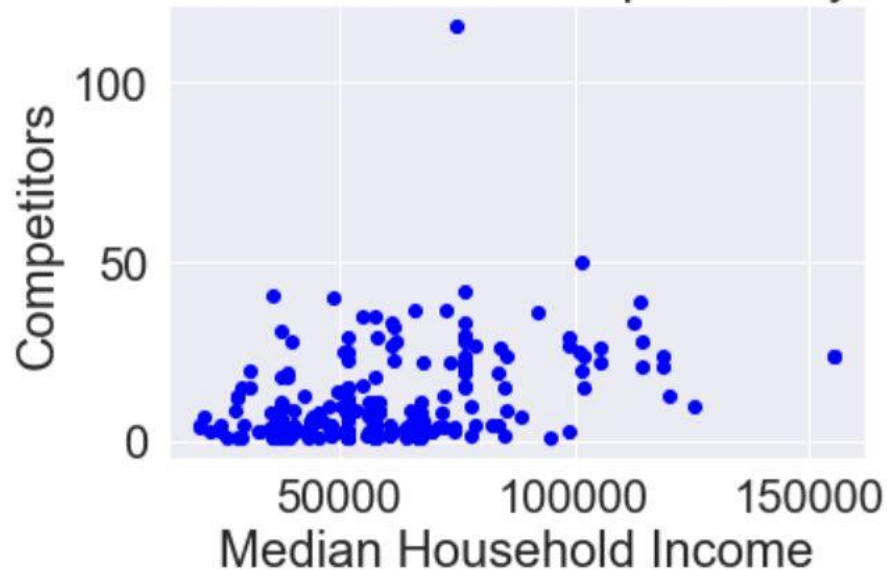
4.2 Analyzing Individual Feature Patterns using Visualization

From the figures, cluster 5 ('Labels'==4) have the lowest population density, and cluster 1 ('Labels'==0) have the lowest Median Household Income, they have less Competitors too. what's the relationships between Median Household Income, Population density and Competitors.



3.3 Neighborhoods analysis

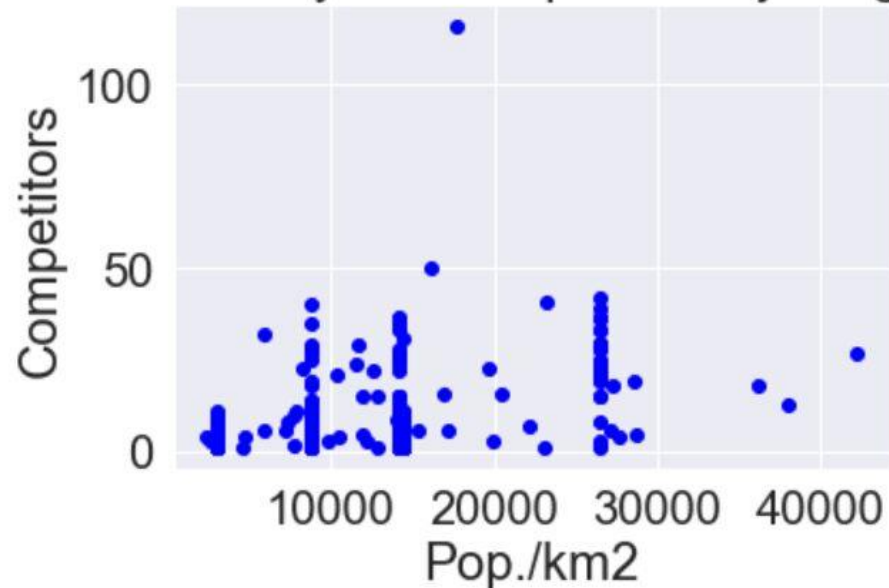
Median Household Income VS Competitors by Neighborhoods



- ▶ The Pearson Correlation Coefficient is $r^* = 0.353$ with a P-value of $P = 3.116e-09$.
- ▶ Since the p-value is $\ll 0.001$, the correlation between Median Household Income and Competitors is statistically significant, although the linear relationship isn't extremely strong (~ 0.35).
- ▶ From the 2D Scatter-plot, we can see that when the neighborhood tends to have more competitors with a higher Median Household Income.

3.3 Neighborhoods analysis

Population density VS Competitors by Neighborhoods



- ▶ The Pearson Correlation Coefficient is $r = 0.406$ with a P-value of $P = 5.953e-12$.
- ▶ Since the p-value is $\ll 0.001$, the correlation between Population density and Competitors is statistically significant, although the linear relationship isn't extremely strong (~ 0.41).
- ▶ From the plot, less obviously with noise which is caused by the missing information, competitors tend to rise as the population density increases.

4. Results

Based on the patterns, Let's find the neighborhoods :

- ▶ **Median Household Income > 66764(75%)**
- ▶ **Population density > 14353(75%)**
- ▶ **Competitor < 11(mean=10.18).**

	Neighborhood	MHI>66764	Com<11	Pop>14353
0	Edgewater Park	67549.0	8	14353.379074
1	Forest Hills	67881.0	6	15279.000000
2	North Riverdale	78895.0	5	14353.379074
3	Roosevelt Island	98797.0	3	26482.390996
4	Schuylerville	67549.0	4	14353.379074
5	Spuyten Duyvil	67534.0	1	14353.379074
6	Stuyvesant Town	95022.0	1	26482.390996
7	Throgs Neck	67549.0	3	14353.379074

5. Discussion

Numbers of restaurants

Neighborhood	Venue Category	
Edgewater Park	American Restaurant	1
	Asian Restaurant	1
	Chinese Restaurant	1
	Fast Food Restaurant	1
	Italian Restaurant	4

- ▶ There are many other factors to be considered while deciding a location selection. For example, if the owner wants to open a Italian restaurant. Edgewater Park is not a good choice.
- ▶ Because there have been 4 Italian Restaurants in this area. If he does so, the new restaurant may spend too much time on marketing because of the abundance of competitors around neighborhood. It's better to choose a location from the rest.

6. Conclusion

- ▶ Overall, location selection is a very complicated process. There are many considerations that should be made in the selection.
- ▶ According to this report, 8 out of 306 neighborhoods are recommended to owners or investors who want to open a restaurant. Although it can't give the exact location, it helps to narrow it down!

Data Sources & Reference

- ▶ 1) Neighborhoods in New York City Coursera-Course 9
- ▶ 2) Restaurants in New York City Foursquare API
- ▶ 3) Population density by Neighborhoods
Wikipedia page
(["https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City"](https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City))
- ▶ 4) Median Household Income by Neighborhoods
web page
(<https://ny.curbed.com/2017/8/4/16099252/new-york-neighborhood-affordability>)



Thank you!