IBM Data Science Professional Certificate

Capstone Project – The Battle of Neighborhoods

Karol Kozłowski

Introduction

Establishing a restaurant involves setting many criteria and having a vision of what you want to deliver to your potential customers. I mean here such aspects as the type of cuisine served, the definition of the type of customers for which it should be a place, decor, location, etc..

Recommendation systems are now very popular in many aspects such as e-commerce services, movie rating (Netflix), social networking and a lot of services helping users finding preferred items, in this work I will focus on (LBSN) location-based social networks.

Business Problem

This analysis is for those who know what kind of cuisine (e.g. Italian) they want to serve in their restaurant. I will focus on identifying good/not good locations for this restaurant. The choice must be based on an analysis of many factors such as: the population of the district, the popularity of the area among people, the proximity of popular public transport, the proximity of other urban infrastructure and the most important is the number of units in the area offering the same type of cuisine.

Data Sources

- List of Toronto postal codes from WIKIPEDIA.

- Geospatial Coordinates CSV file for Toronto postal codes from http://cocl.us/Geospatial_data

- Foursquare API - to explore venues in Toronto neighbourhoods

Data Sources

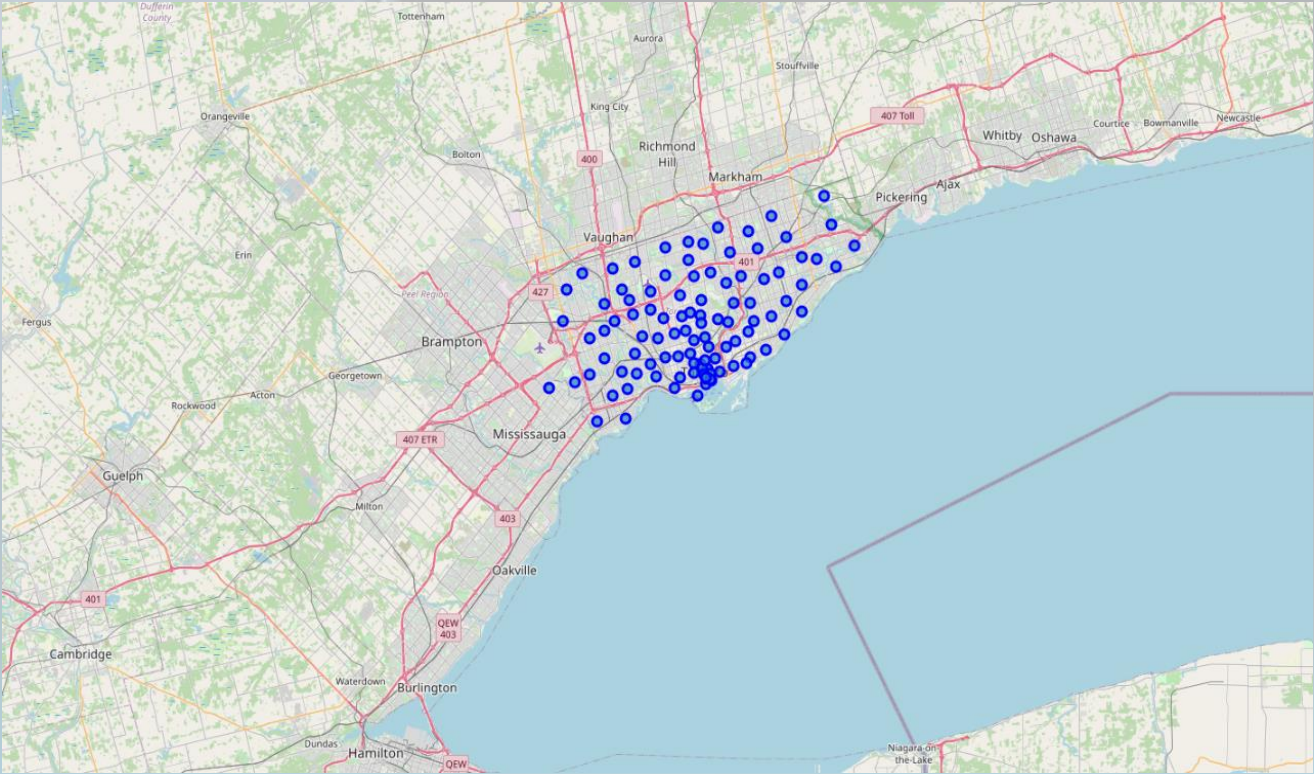List of Toronto postal codes from [WIKIPEDIA.](#)

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0, however, the postal code M0R 8T0 is assigned to an Amazon warehouse in Mississauga, suggesting

Postal Code ⬥	Borough ⬥	Neighborhood ⬥
M1A	Not assigned	
M2A	Not assigned	
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
M8A	Not assigned	
M9A	Etobicoke	Islington Avenue
M1B	Scarborough	Malvern, Rouge

Data Sources

Geospatial Coordinates CSV file for Toronto postal codes from http://cocl.us/Geospatial_data

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

Data Sources

Foursquare API - to explore venues in Toronto neighbourhoods

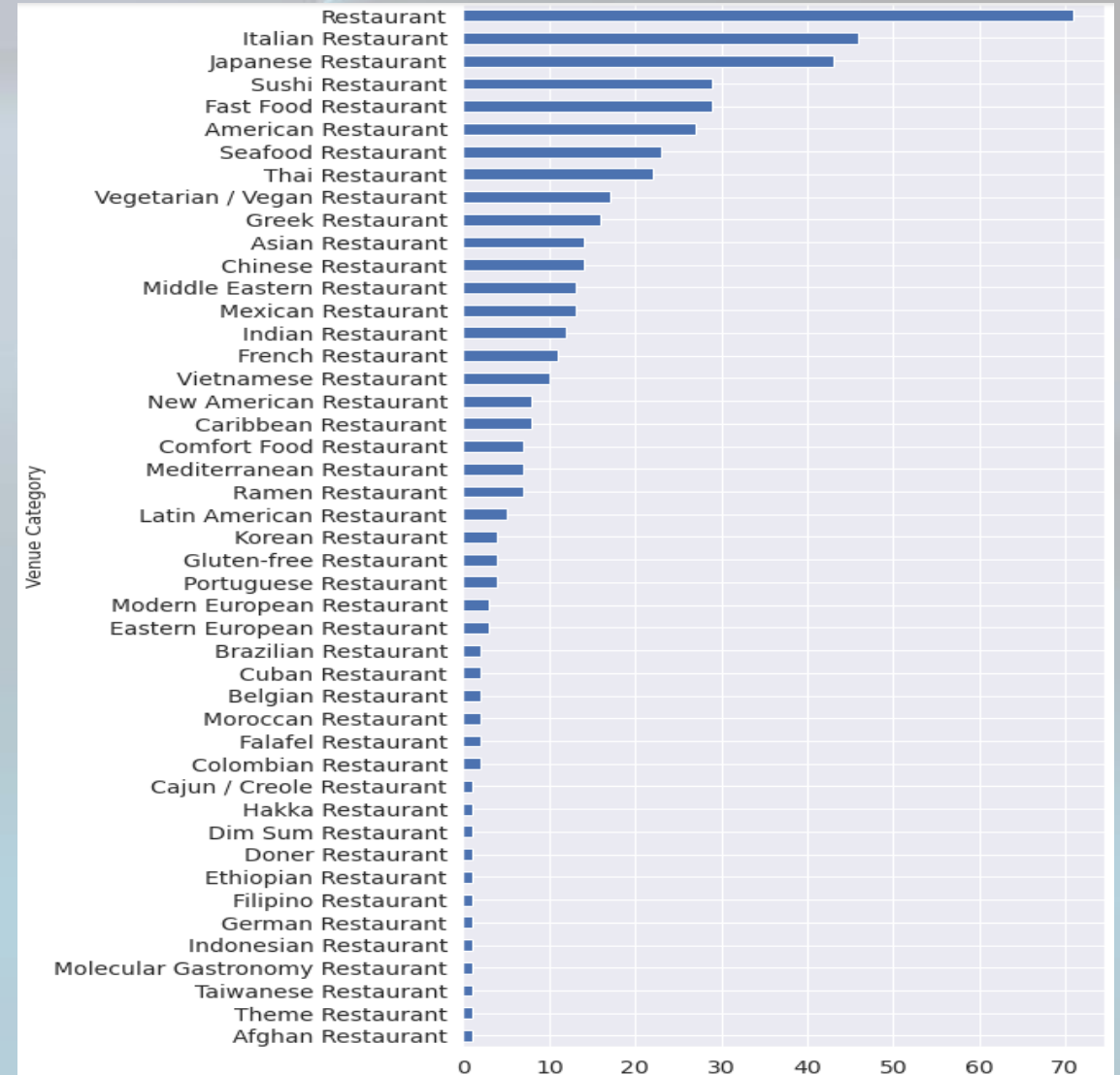|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Id	Venue Name	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	4e6696b6d16433b9ffff47c3	KFC	43.754387	-79.333021	Fast Food Restauran
1	Victoria Village	43.725882	-79.315572	4f3ecce6e4b0587016b6f30d	Portugril	43.725819	-79.312785	Portugues Restauran
2	Victoria Village	43.725882	-79.315572	4d689350b6f46dcb77ee15b2	The Frig	43.727051	-79.317418	French Restauran
3	Regent Park, Harbourfront	43.654260	-79.360636	5612b1cc498e3dd742af0dc8	Impact Kitchen	43.656369	-79.356980	Restauran
4	Regent Park, Harbourfront	43.654260	-79.360636	53a22c92498ec91fda7ce133	Cluny Bistro & Boulangerie	43.650565	-79.357843	French Restauran

Methodology

- Addresses are converted into their equivalent latitude and longitude values.

- Foursquare API is used to explore neighborhoods in Toronto, Ontario.

- After that, explore function to get restaurant categories in each neighborhood.

Methodology

The most popularity types of restaurants in Toronto

Methodology

The most popularity types of restaurants in each Neighborhood

Methodology

Then using this feature to group the neighborhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Manhattan and its emerging clusters.

	Neighborhood	1st Most Common cuisine	2nd Most Common cuisine	3rd Most Common cuisine	4th Most Common cuisine	5th Most Commo cuisine
0	Agincourt	Venue Category_Latin American Restaurant	Venue Category_Vietnamese Restaurant	Venue Category_Dim Sum Restaurant	Venue Category_German Restaurant	Venue Category_French Restaurant
1	Bathurst Manor, Wilson Heights, Downsview North	Venue Category_Sushi Restaurant	Venue Category_Restaurant	Venue Category_Middle Eastern Restaurant	Venue Category_Vietnamese Restaurant	Venue Category_Cuban Restaurant
2	Bayview Village	Venue Category_Japanese Restaurant	Venue Category_Chinese Restaurant	Venue Category_Vietnamese Restaurant	Venue Category_Dim Sum Restaurant	Venue Category_German Restaurant
3	Bedford Park, Lawrence Manor East	Venue Category_Restaurant	Venue Category_Italian Restaurant	Venue Category_Comfort Food Restaurant	Venue Category_American Restaurant	Venue Category_Thai Restaurant
4	Berczy Park	Venue Category_Seafood Restaurant	Venue Category_Restaurant	Venue Category_Greek Restaurant	Venue Category_Thai Restaurant	Venue Category_Vegetari / Vegan Restauran

Result

Grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category

	Neighborhood	Venue Category_Afghan Restaurant	Venue Category_American Restaurant	Venue Category_Asian Restaurant	Venue Category_Belgian Restaurant	Venue Category_Brazilian Restaurant	Venue Categor / Creole Restaur
0	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0
1	Bathurst Manor, Wilson Heights, Downsview North	0.0	0.0	0.0	0.0	0.0	0.0
2	Bayview Village	0.0	0.0	0.0	0.0	0.0	0.0
3	Bedford Park, Lawrence Manor East	0.0	0.1	0.0	0.0	0.0	0.0
4	Berczy Park	0.0	0.0	0.0	0.0	0.0	0.0

Result

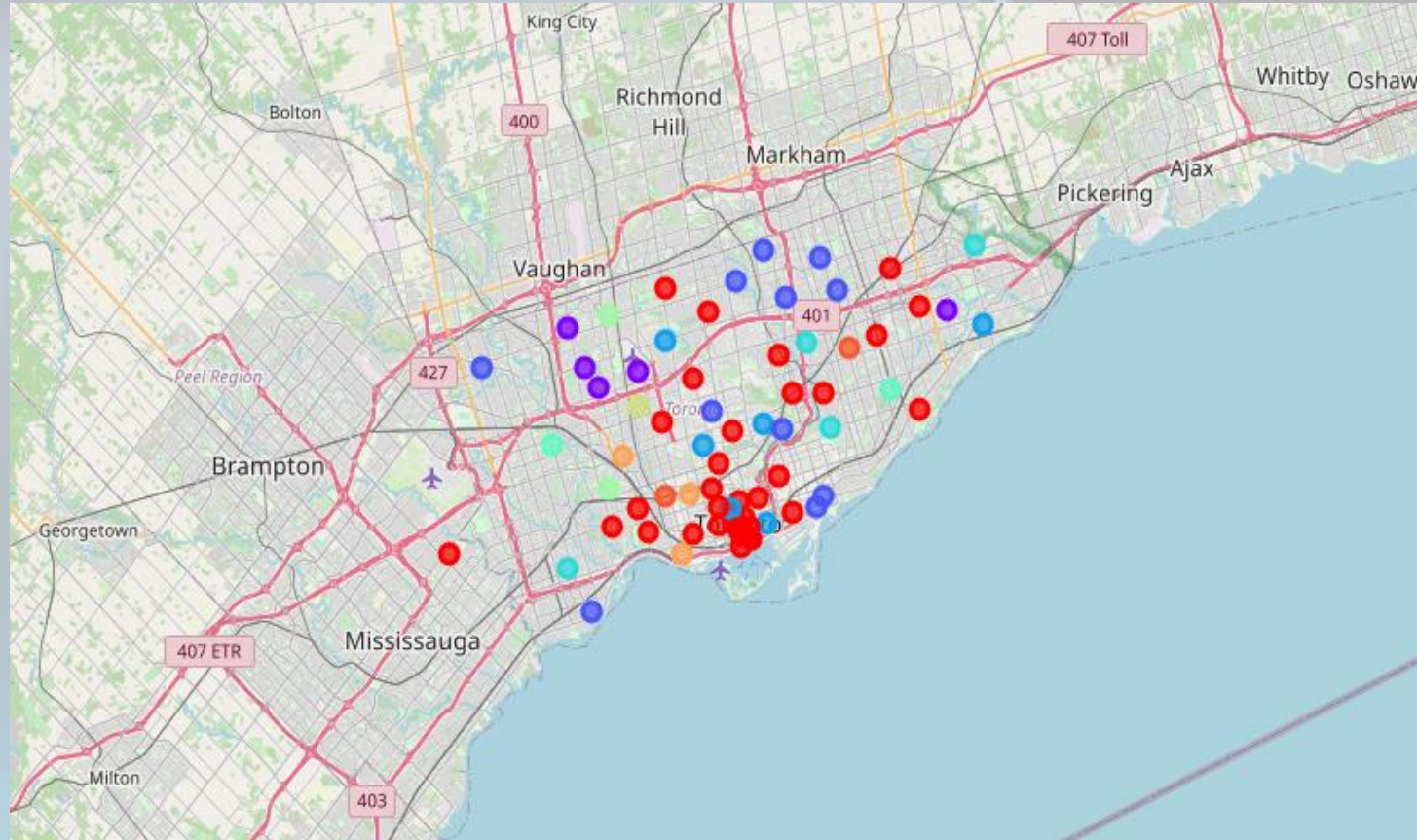Number of positions in each cluster

```
Cluster Labels
0       35
1        5
2       11
3        6
4        4
5        2
6        2
7        1
8        3
9        2
```
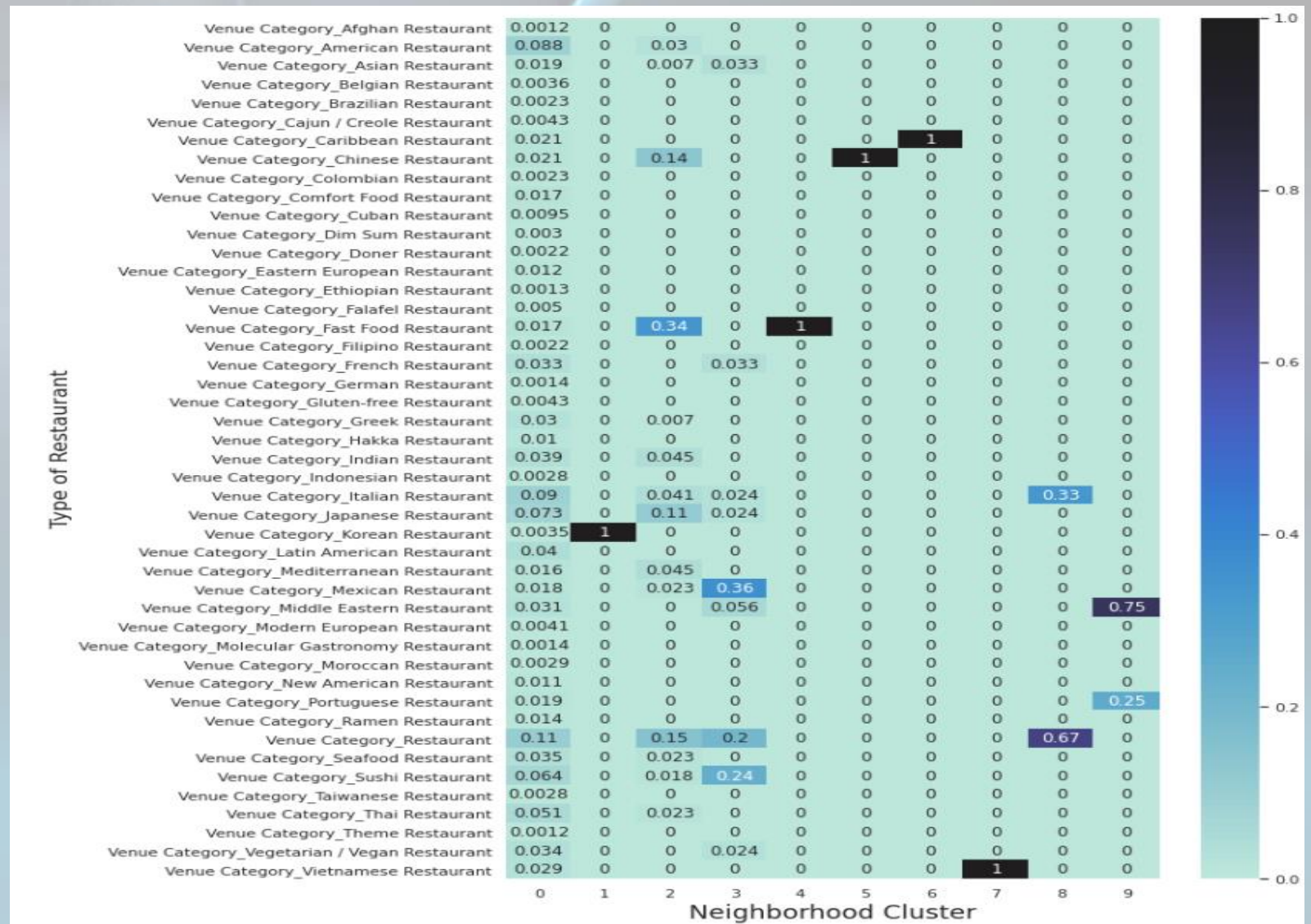
Result

Venue positions on the Toronto map, in colors depending on the membership of a given cluster.
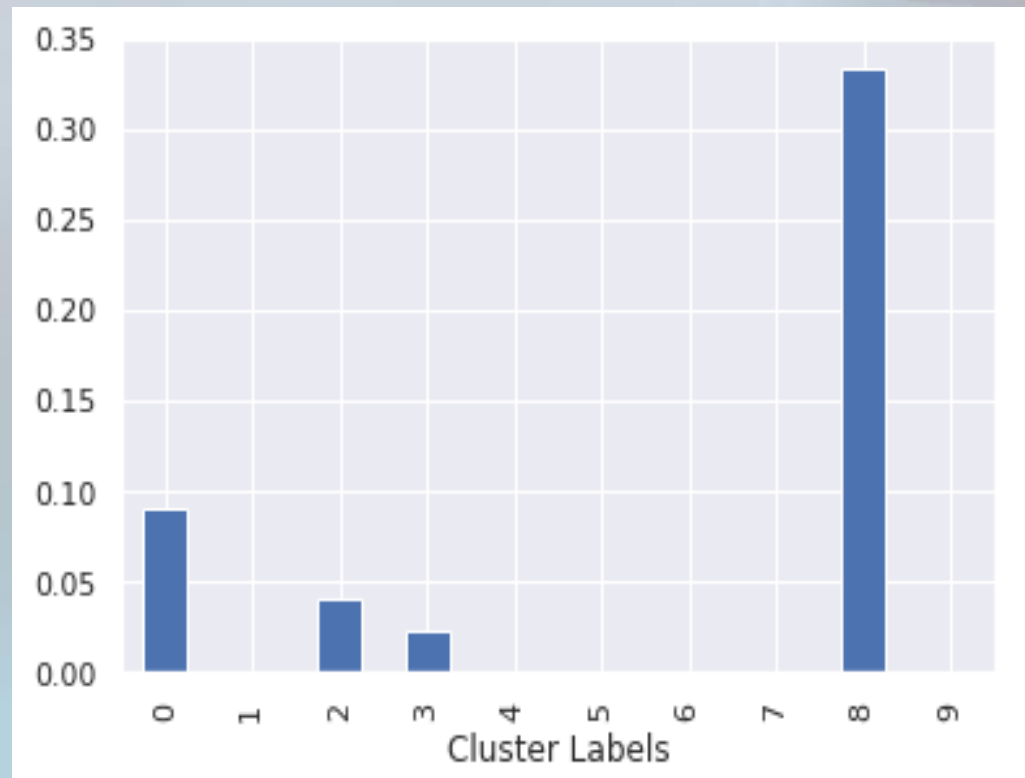
Result

Heat map of restaurant categories and cluster membership

Result

The occurrence of restaurants serving Italian cuisine is frequent in cluster no. 8. Therefore, it is unreasonable to open restaurants with such cuisine in a district qualified for this cluster.

Discussion

- This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

- There is high competition in neighborhoods belonging to cluster 8 there is a lot of competition among Italian restaurants, so it is very risky to open business in these areas.

- It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Conclusion

Although all of the goals of this project were met there is definitely space for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be developed into a fully pledged application that could support the opening a business idea in an unknown location.