Best Locations to Open New Coffee Shops in the City of Toronto

Cousera – IBM Capstone Project

BY: RASOUL ARABJAMALOEI

Introduction

- **Objective:**

Use machine learning and foursquare API to optimize the location of new coffee shops in the City of Toronto

- **Audience:**

Investors who are looking to find the best possible investing options in opening new venues

Data scientists who are going to follow this methodology for other cities in the world

Restaurant chain owners who are looking to make a balance in supply-demand in their city of choice.

The data science students who are looking for ideas to work on for their future project.

Data

- Toronto neighborhoods information was extracted from Wikipedia
- Geocoders library was used to find geospatial coordinates
- The final dataframe was built by gathered information:

[65]:

	Name	Population	Density (people/km2)	Average Income	Latitude	Longitude
0	Agincourt	44,577	3580	25,750	43.7854	-79.2785
1	Alderwood	11,656	2360	35,239	43.6017	-79.5452
2	Alexandra Park	4,355	13,609	19,687	43.6508	-79.4043
3	Allenby	2,513	4333	245,592	43.7114	-79.5534
4	Amesbury	17,318	4,934	27,546	43.7062	-79.4835
5	Armour Heights	4,384	1914	116,651	43.7439	-79.4309
6	Banbury	6,641	2442	92,319	43.7428	-79.37
7	Bathurst Manor	14,945	3187	34,169	43.7639	-79.4564
8	Bay Street Corridor	4,787	43,518	40,598	43.6628	-79.3863
9	Bayview Village	12,280	2,966	46,752	43.7692	-79.3767

Methodology

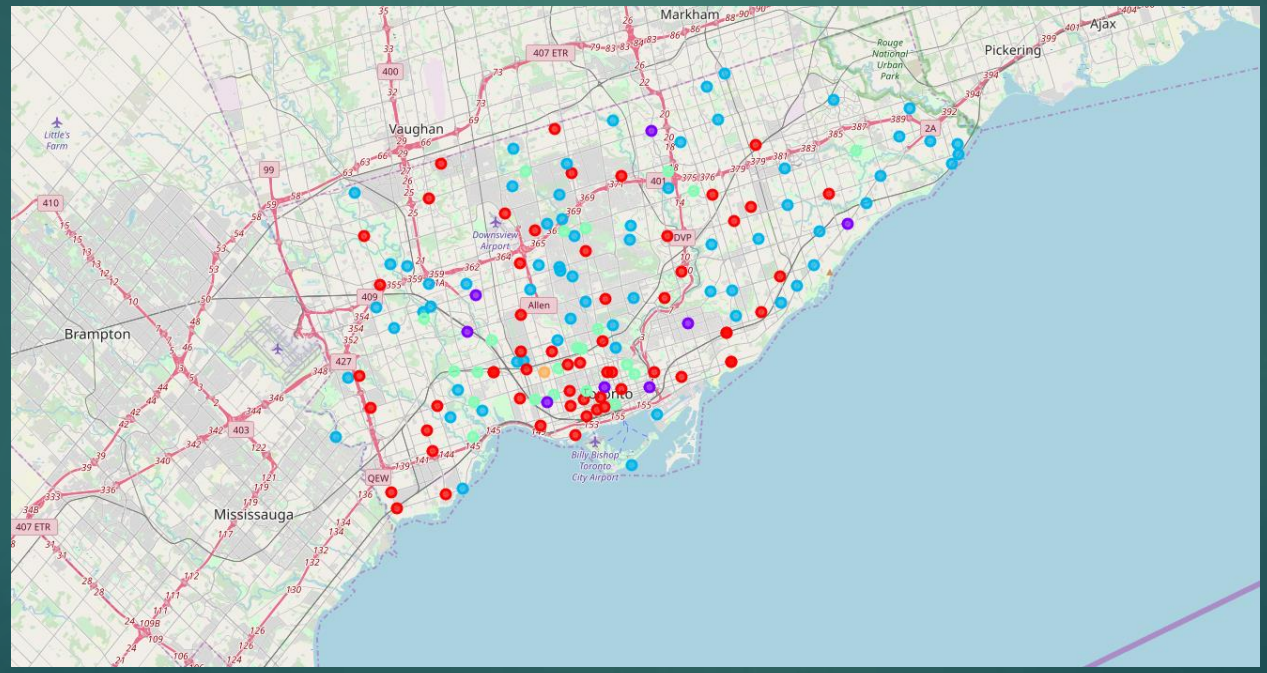
- FourSquare API was used to get the top venues in each neighborhood
- The venues were grouped and mean values of occurrence of each types of venues were calculated
- The final dataframe for clustering was built:

[302]:

	Population	Density (people/km2)	Average Income	venues
0	0.000022	0.000279	0.000039	0.076923
1	0.000086	0.000424	0.000028	0.111111
2	0.000230	0.000073	0.000051	0.060000
3	0.000398	0.000231	0.000004	0.100000
4	0.000058	0.000203	0.000036	0.250000

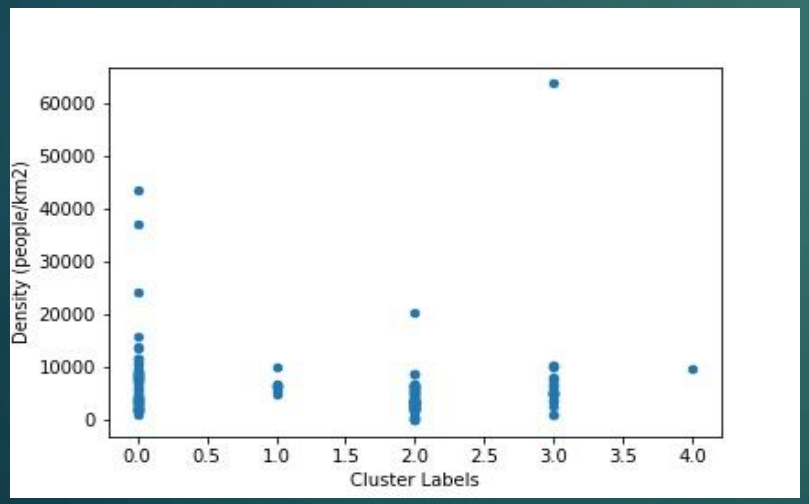
Methodology

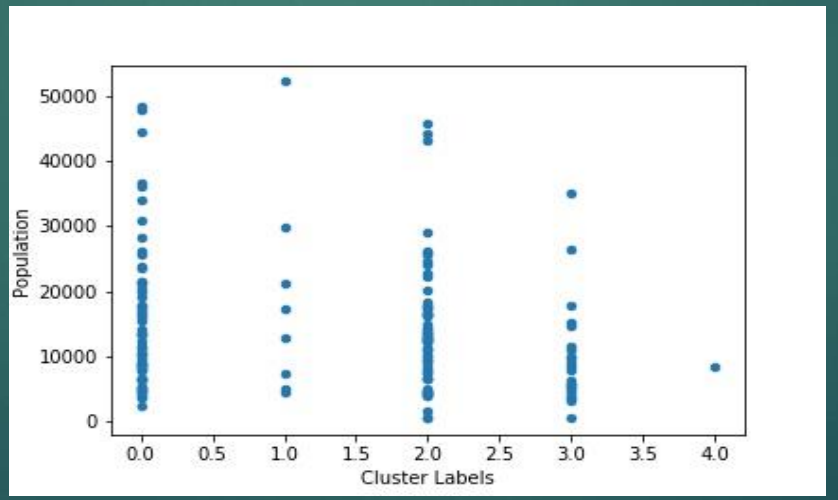
- K-mean clustering was used to divide neighborhoods in 5 clusters.
- Folium was used to locate the neighborhoods and cluster labels:

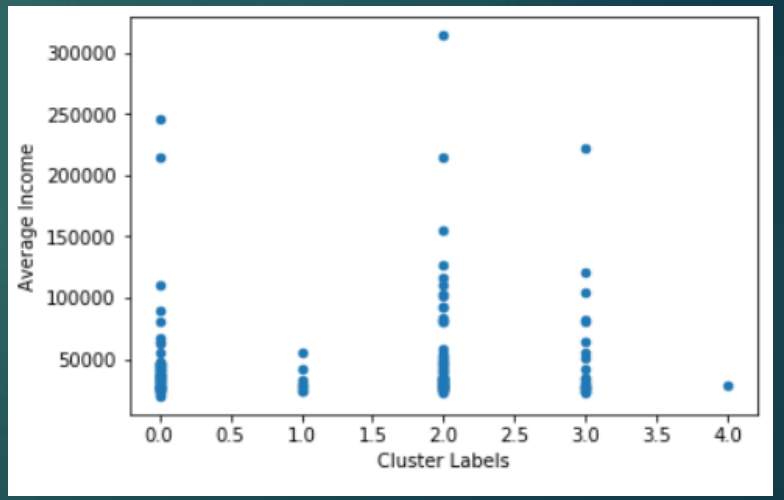


Discussion

- The cluster labels were plotted versus Population density, Population and average income
- No clear relationships between cluster labels and Population density, Population and average income.

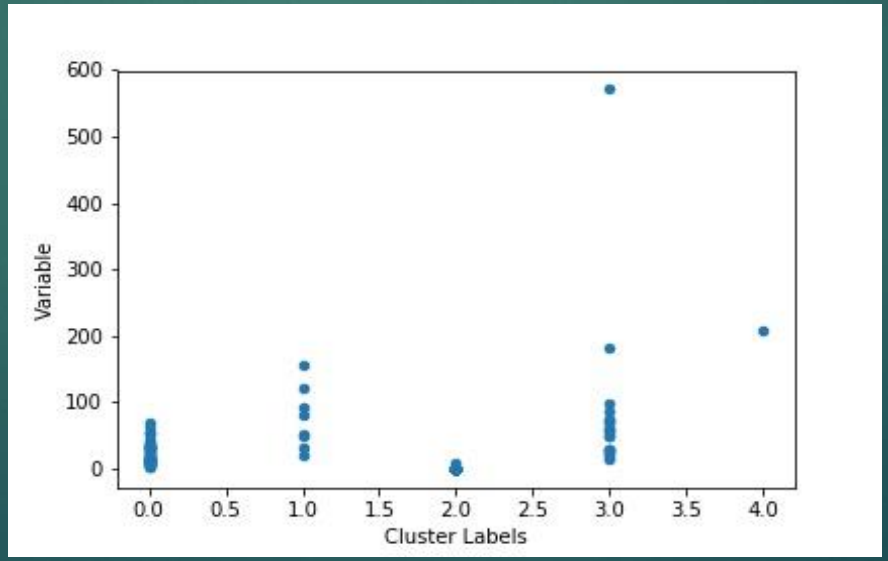






Discussion

- To understand the effect of combination of variables, a new variable was defined by dividing the venues by Population and Average income. The new variable was plotted versus cluster label
- New variable correlates better with the cluster labels



Conclusions

- Foursquare API tool was used to in combination with geocoders, folium, Scikit Learn and matplotlib libraries to find the best locations to open a new coffee shop in the city of Toronto.
- The hypothesis is that the neighborhoods with higher population, higher population density, higher average income and lower existing venues have the best potential for the success rate of a new coffee shop. The results showed that k-mean clustering method could make distinct neighborhood clusters based on the given information.
- This methodology can be used to optimize the process of finding best places for opening other new venues.