

K-Means Clustering Analysis

- Neighbourhoods in Scarborough, Canada

Average of Housing Price & Top School Rates

Author Name: Jabin Choi

Course: IBM Data Science (Applied Data Science Capstone Project)

INTRO

Problem Description

- ▶ Scarborough in Canada is one of the cities that recently attracted a lot of attention among immigrants
- ▶ This project will result in data related to an average of housing price and top school rates for neighbourhoods in Scarborough
- ▶ This project is useful for people planning to move to Scarborough
- ▶ In this project, the app called Foursquare is used for venues data
- ▶ For more details on Foursquare, please visit the following url:
* URL: [https:// foursquare.com](https://foursquare.com)

WHAT IS FOURSQUARE API?

- ▶ It is a social location service
- ▶ It provides users to explore the location around the world
- ▶ It enables users to download it through iPhone, Blackberry or Android and then connects their account to other social media accounts
- ▶ It is a RESTful set of addresses to which developers can send requests
- ▶ Users has nothing to download onto their server
- ▶ It allows developers to interact with the Foursquare platform
- ▶ Users can gather data source as it has a dataset of millions of places
- ▶ API has the ability to perform location search, sharing, and details on a business

DATA APPROACH

Methodology

- Data table from wiki url
(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

Out[4]:

	Postalcode	Borough	Neighbourhood
0	0	0	0
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village

- ▶ Data cleaning:
 - ▶ Drop meaningless contents in the table - meaning handle 'Not assigned' and None values
 - ▶ Group and re-index the table
 - ▶ Add two columns called 'Latitude' and 'Longitude'
- ▶ Once completed, the result is as follows:

Out[10]:

	Postalcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.811525	-79.195517
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765815	-79.175193
3	M1G	Scarborough	Woburn	43.768369	-79.217590
4	M1H	Scarborough	Cedarbrae	43.769688	-79.239440

- ▶ Need Foursquare locational information as follows:
 - ▶ Neighbourhoods
 - ▶ Neighbourhoods Latitude / Longitude
 - ▶ Venues
 - ▶ Venue names such as restaurants or stores
 - ▶ Venue Latitude / Longitude
 - ▶ Venue Category
- ▶ Extract information - venue names, categories and latitude / longitude

Out[25]:

	name	categories	lat	lng
0	SEPHORA	Cosmetics Shop	43.775017	-79.258109
1	Disney Store	Toy / Game Store	43.775537	-79.256833
2	American Eagle Outfitters	Clothing Store	43.776012	-79.258334
3	DAVIDsTEA	Tea Room	43.776613	-79.258516
4	Hot Topic	Clothing Store	43.775450	-79.257929

K-MEANS CLUSTERING

- ▶ What is K-MEANS CLUSTERING?
 - ▶ A form of unsupervised machine learning is used for data analysis
- ▶ What will we discover by k-means clustering?
 - ▶ The average of housing price in Scarborough
 - ▶ Top school rates in Scarborough

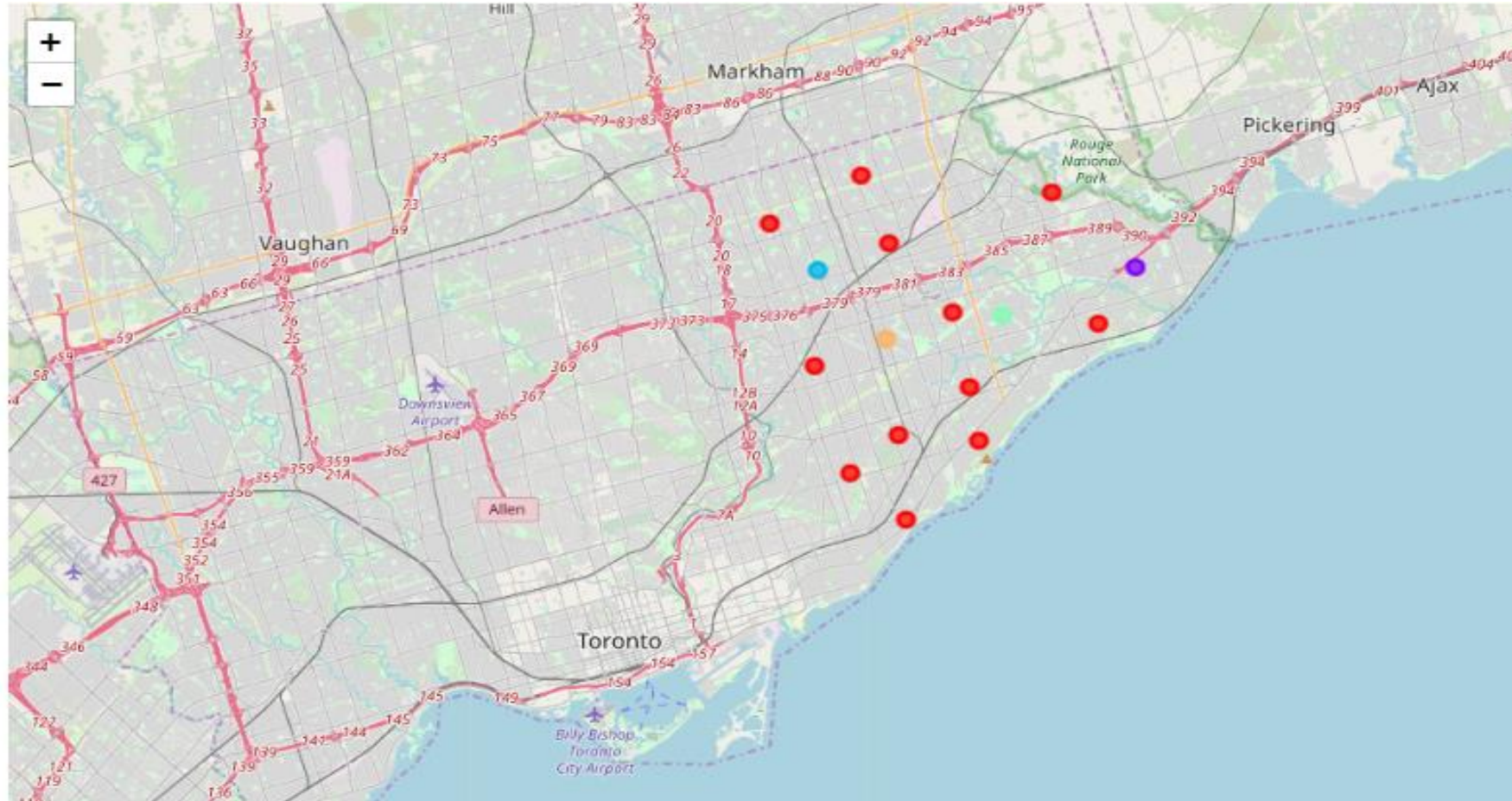
► Sorted data table (k=5)

Out[38]:

	Neighbourhood	1st Most common venue	2nd Most common venue	3rd Most common venue	4th Most common venue	5th Most common venue	6th Most common venue	7th Most common venue	8th Most common venue	9th Most common venue	10th Most common venue
0	Agincourt	Shopping Mall	Chinese Restaurant	Discount Store	Vietnamese Restaurant	Bakery	Grocery Store	Sushi Restaurant	Supermarket	Hong Kong Restaurant	Bubble Tea Shop
1	Agincourt North, L'Amoreaux East, Milliken, St...	Pharmacy	Zoo Exhibit	Gym	Golf Course	General Entertainment	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Electronics Store	Discount Store
2	Birch Cliff, Cliffside West	Gym	College Stadium	General Entertainment	Gym Pool	Skating Rink	Park	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Electronics Store
3	Cedarbrae	Playground	Zoo Exhibit	College Stadium	Golf Course	General Entertainment	Fried Chicken Joint	Flower Shop	Fast Food Restaurant	Electronics Store	Discount Store
4	Clairlea, Golden Mile, Oakridge	Bakery	Bus Line	Intersection	Coffee Shop	Soccer Field	Bus Station	Metro Station	Convenience Store	General Entertainment	Fried Chicken Joint

- Result of k-means clustering ($k = 5$)

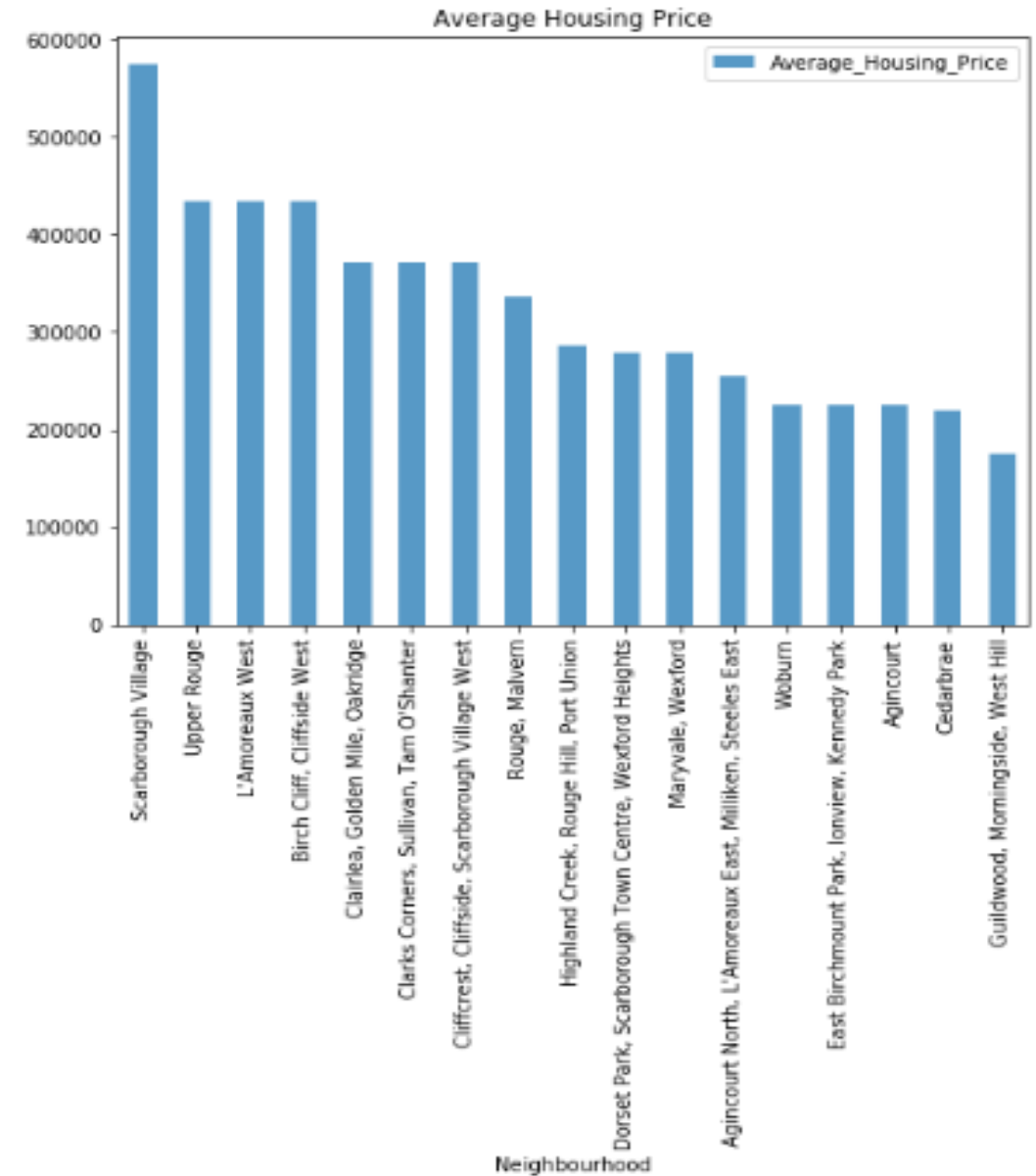
Out[47]:



DISCOVERY THROUGH DATA ANALYSIS

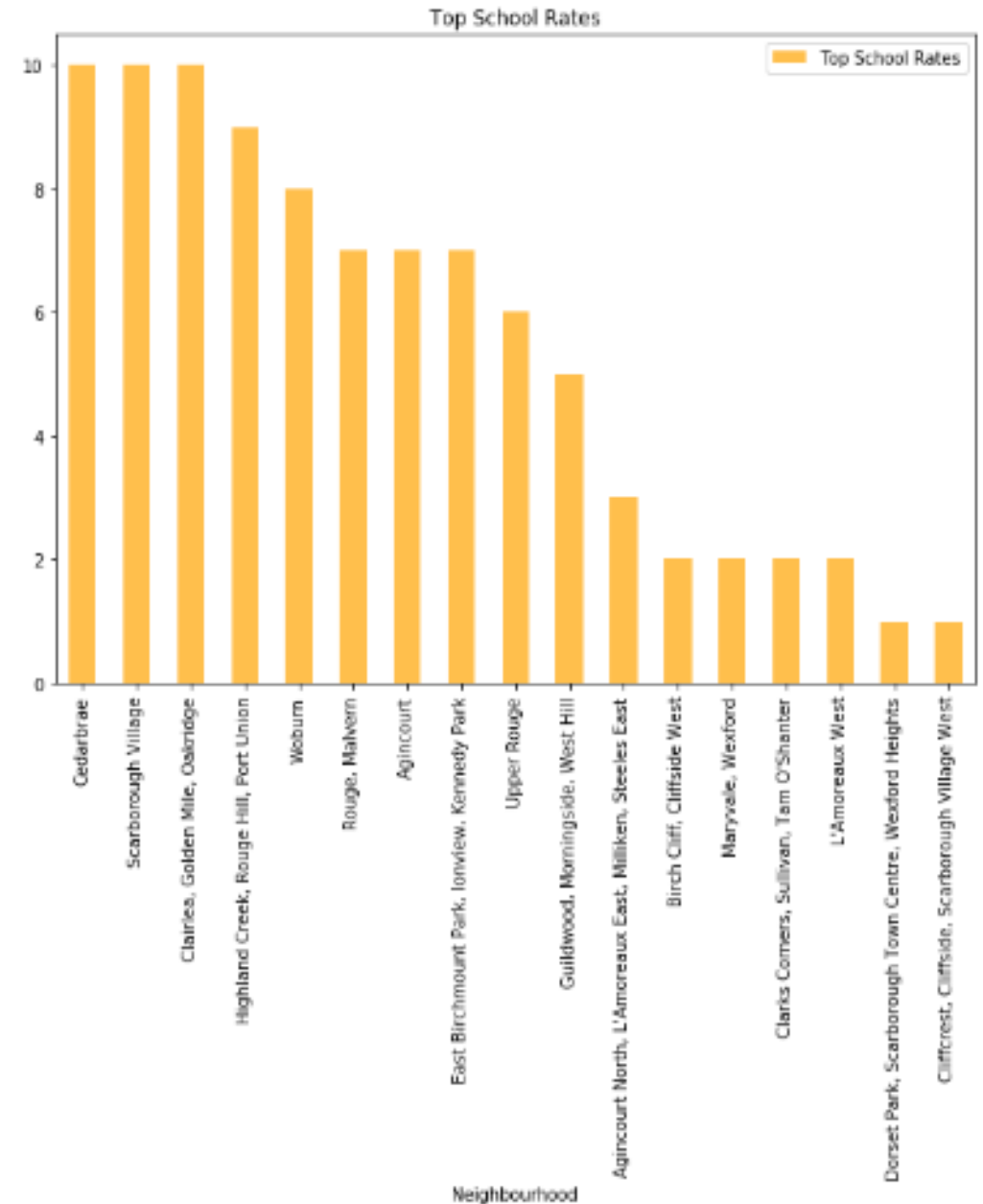
- The result of bar plot after sorting values by the average of housing price

Out[34]: <matplotlib.axes._subplots.AxesSubplot at 0x1539c09a308>



- The result of bar plot after sorting values by top school rates

Out[31]: <matplotlib.axes_subplots.AxesSubplot at 0x1539be17a88>



CONCLUSION

► Result of Barplot 1

- From a bar plot named 'Average Housing Price':
'Scarborough Village' is the neighbourhood that has the highest average of housing price

► Result of Barplot 2

- From a bar plot named 'Top School Rates':
'Cedarbrae' is the neighbourhood that got the highest school rates

THANK YOU!