Project Report

Battle of Neighborhoods: Endgame

-Vinayaka K M

Contents of This Presentation

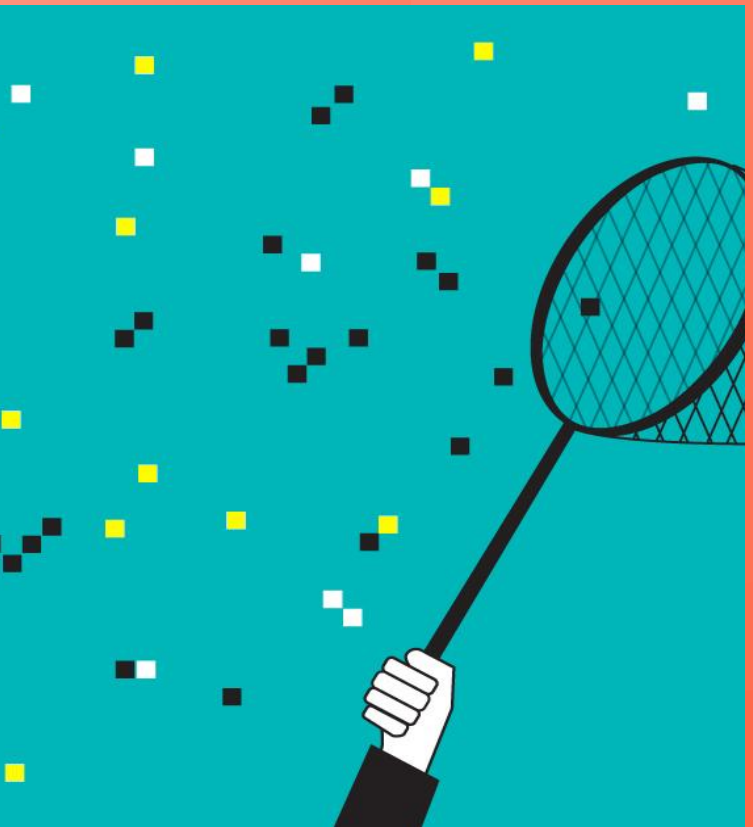
- Problem Description
- Data Presentation
- Methodology
- Results
- **Discussion**
- Conclusion

01. Problem Description

In this project , the problem attempted to solve will be to find the optimum location/neighborhood for a person in Scarborough, Toranto.

Lots of people are migrating to various states of Canada and needed lots of research for good housing prices and reputated schools for their children and other necessities like hospitals,malls etc. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

This Project aim to create an analysis of features for a people migrating to Scarborough to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.



02. Data Presentation

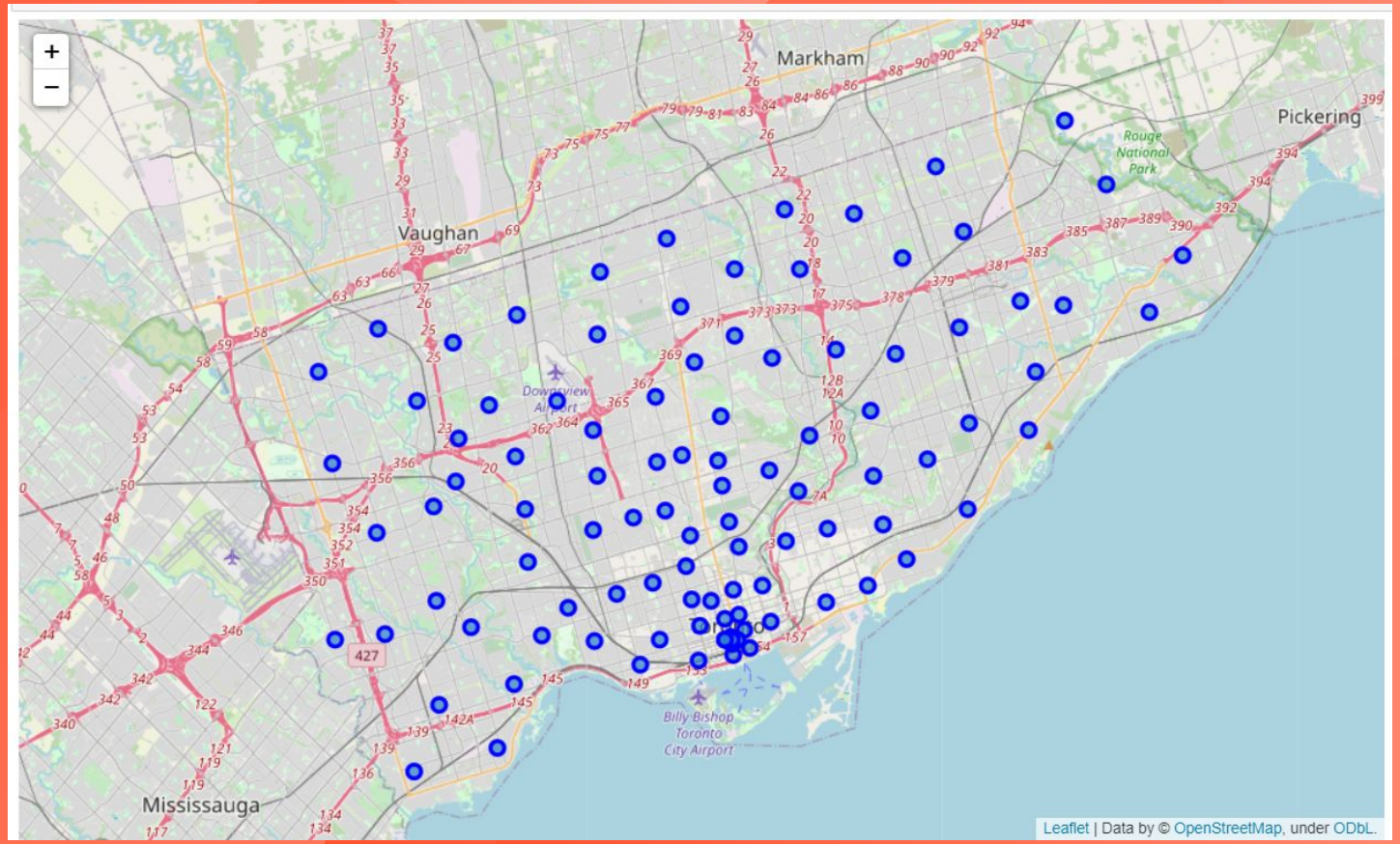
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

Data Extracted

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Neighborhoods of the Scarborough



03.
Methodology

Pipeline

Web Scraping

Data Preprocessing and Analyzing

Modelling and Hyperparameter tuning

Visualization

- The methodology used to approach this problem includes some statistical explorationof the data and some visualizations. The main machine learning technique involved inthe development of this project is clustering, in concrete the K-Means algorithm wasused, implemented with Python

- At a first moment, the main problem was how to obtain the necessary data to build a constructive approach to the problem to be tackled. Usually, to solve these kinds of optimal business location problems, a lot of consumer's data are needed, but for this example and for the sake of simplicity, the focus was put mainly on the population's nationality. A study was carried out over the inhabitants of Scarborough, and it was assumed for this example that the national population from a certain state would prefer restaurants based on their national country and food.Also the person might prefer the facilities similar to their previous neighborhoods

- Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Most common venues of the Neighborhoods

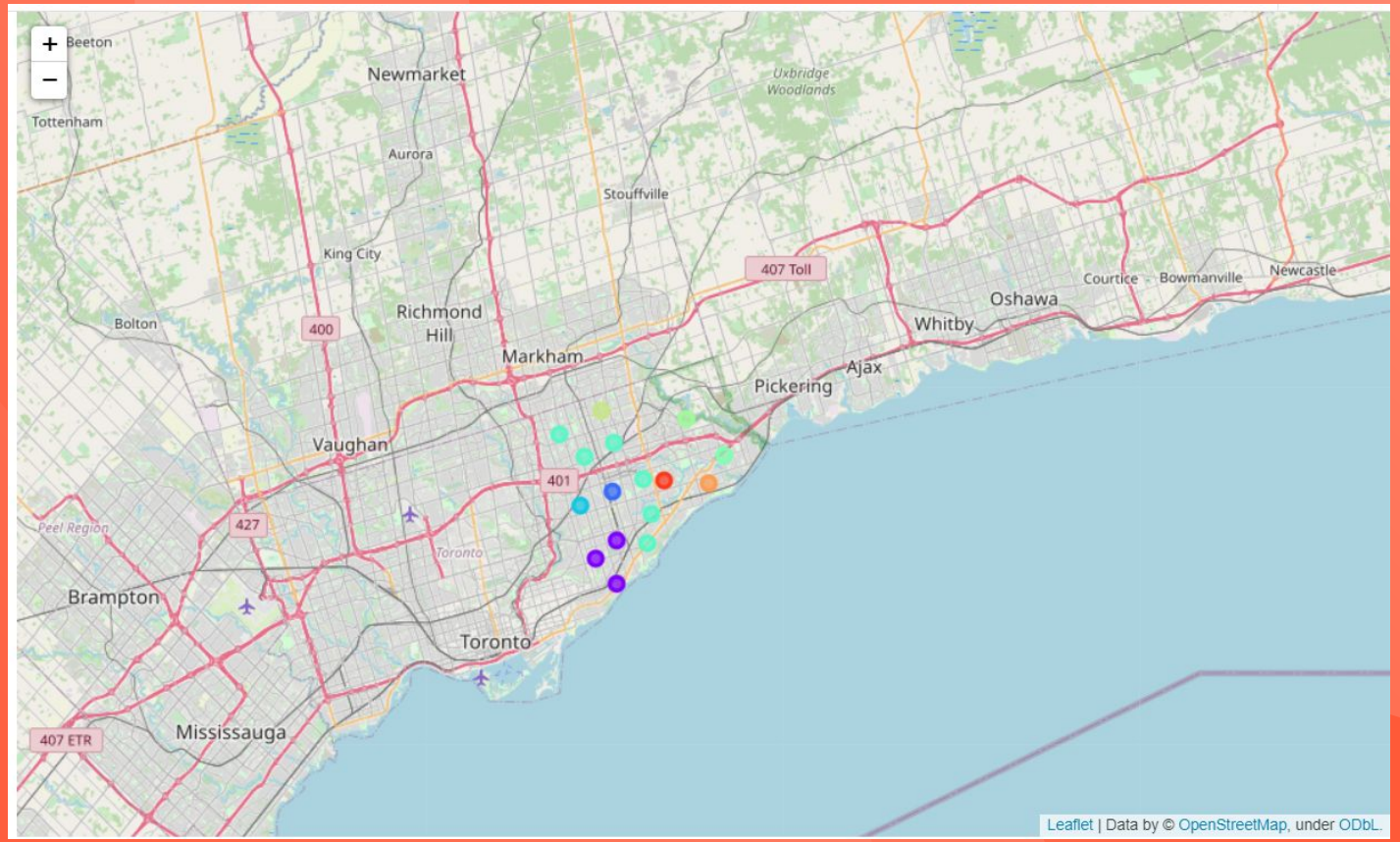
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Shopping Mall	Chinese Restaurant	Breakfast Spot	Coffee Shop	Pizza Place	Supermarket	Sushi Restaurant	Latin American Restaurant	Sandwich Place	Bank
1	Alderwood, Long Branch	Gym	Pharmacy	Print Shop	Pub	Convenience Store	Gas Station	Pizza Place	Coffee Shop	Doner Restaurant	Donut Shop
2	Bathurst Manor, Wilson Heights, Downsview North	Coffee Shop	Park	Pizza Place	Mediterranean Restaurant	Fried Chicken Joint	Sandwich Place	Intersection	Sushi Restaurant	Restaurant	Deli / Bodega
3	Bayview Village	Construction & Landscaping	Flower Shop	Park	Trail	Yoga Studio	Elementary School	Donut Shop	Dry Cleaner	Dumpling Restaurant	Eastern European Restaurant
4	Bedford Park, Lawrence Manor East	Italian Restaurant	Sandwich Place	Coffee Shop	Pizza Place	Indian Restaurant	Butcher	Juice Bar	Sports Club	Thai Restaurant	Restaurant

04. Results

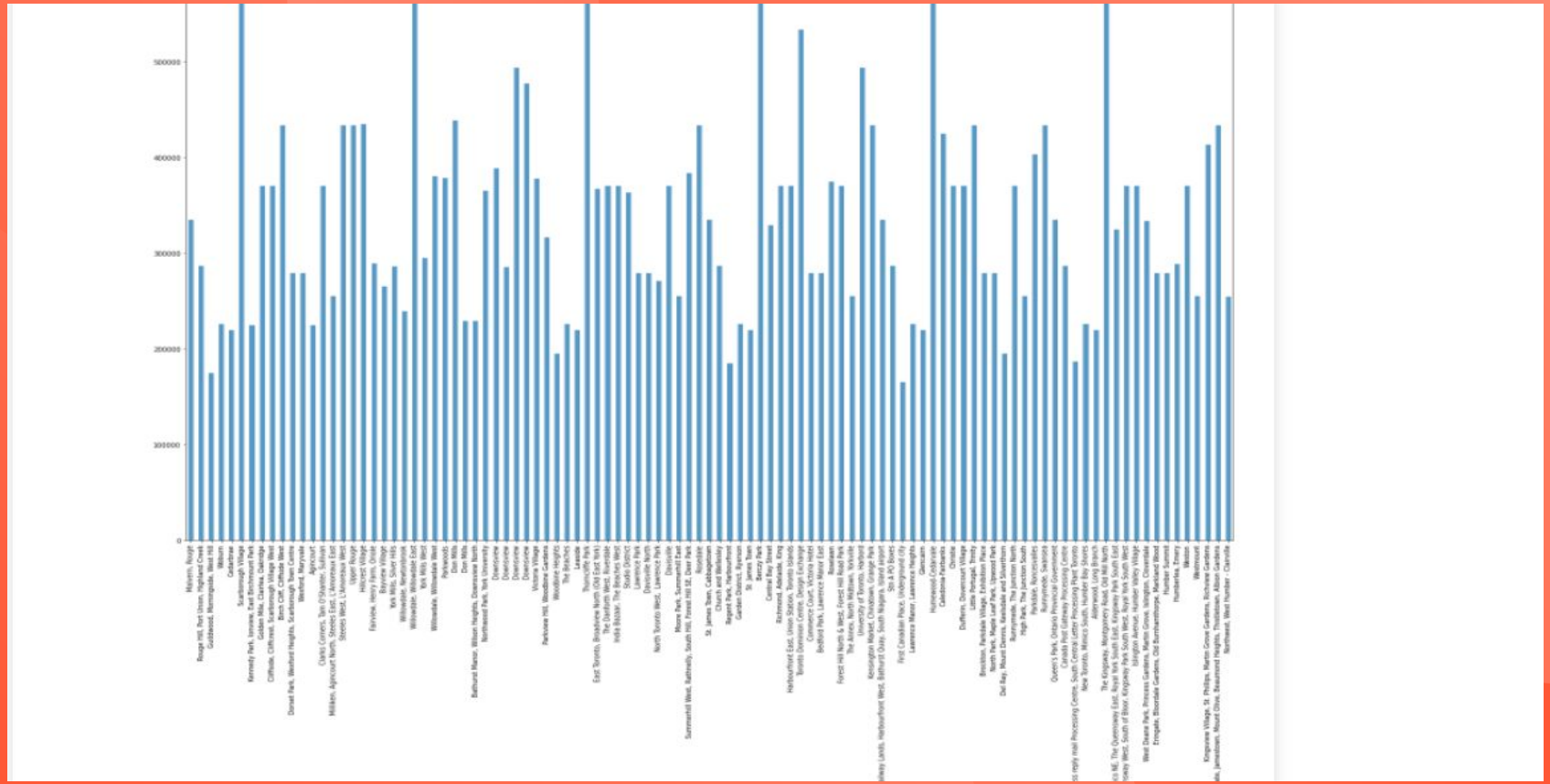
The Location:

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Canada has been one of on the rise.

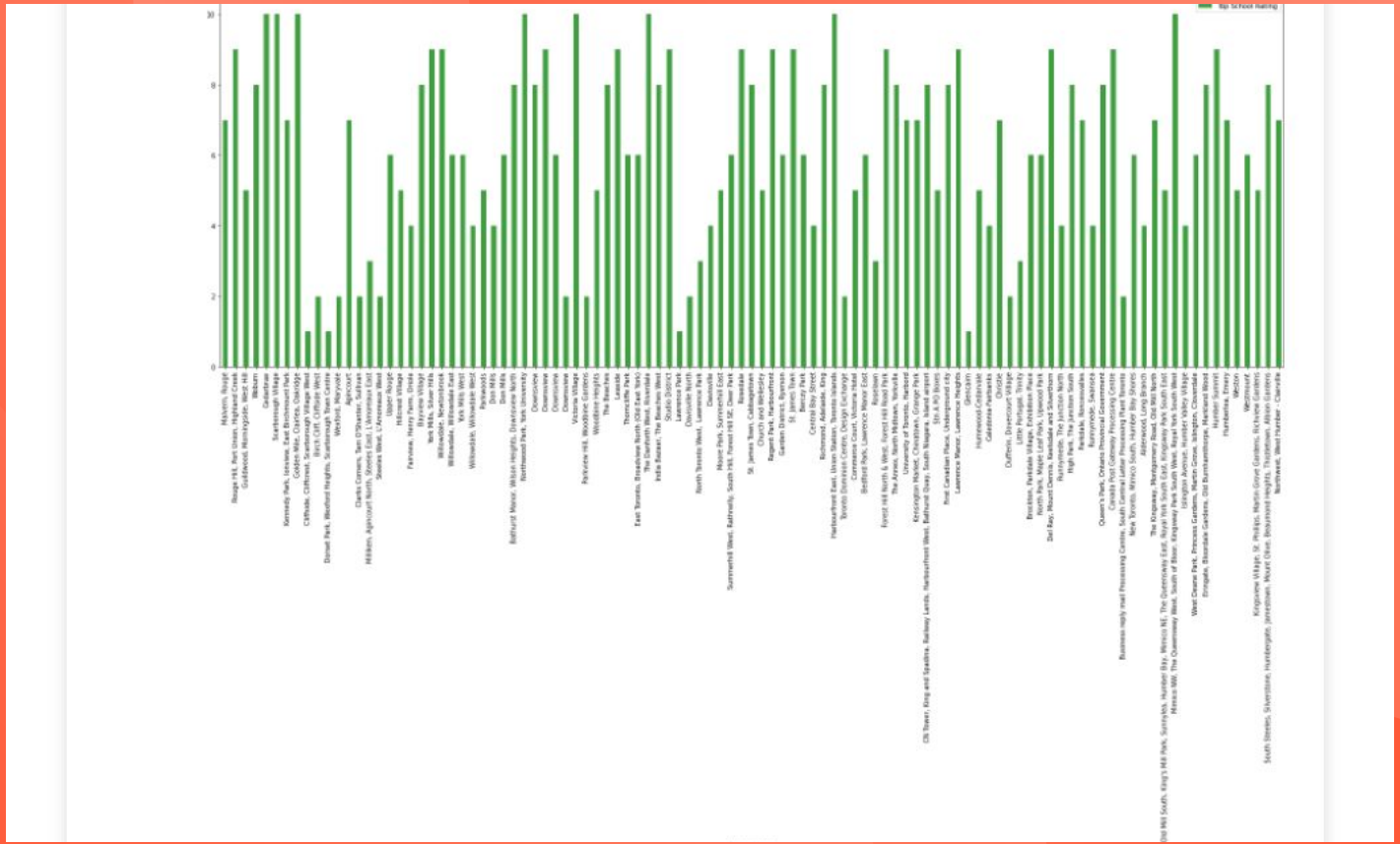
Clusters of Optimal Neighborhoods



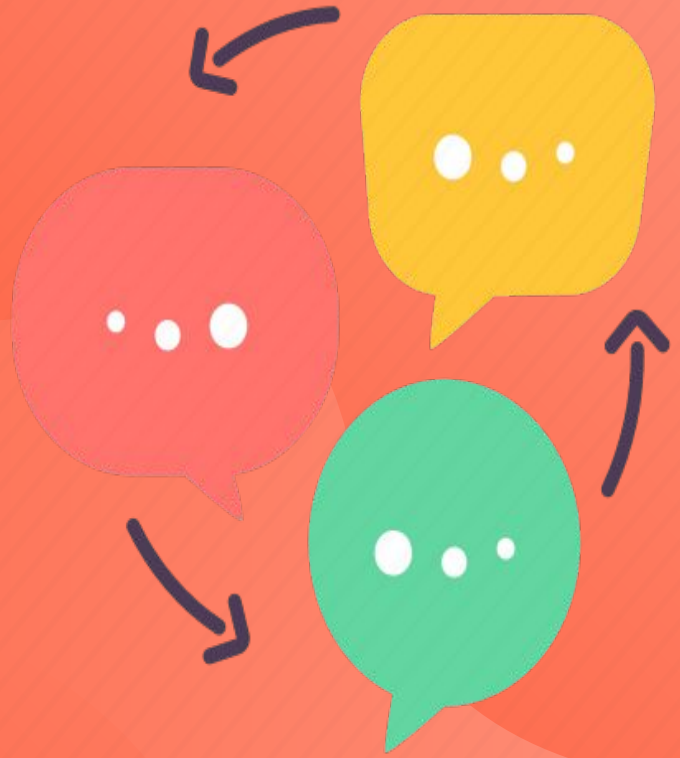
Average Housing price by Clusters



Schools ratings by clusters



05. Discussion Section



Problem Which Tried to Solve:

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

1. Sorted list of house in terms of housing prices in a ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews

6.Conclusion

In this project, using k-means cluster algorithm I separated the neighborhood into 15 different clusters and for 103 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

By this approach a local optimum solution can be suggested to a person i.e a optimum neighborhood can be found out by clustering that has reduced distance between facilities/venues, good ratings and least possible housing price.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Thanks!


slidesgo