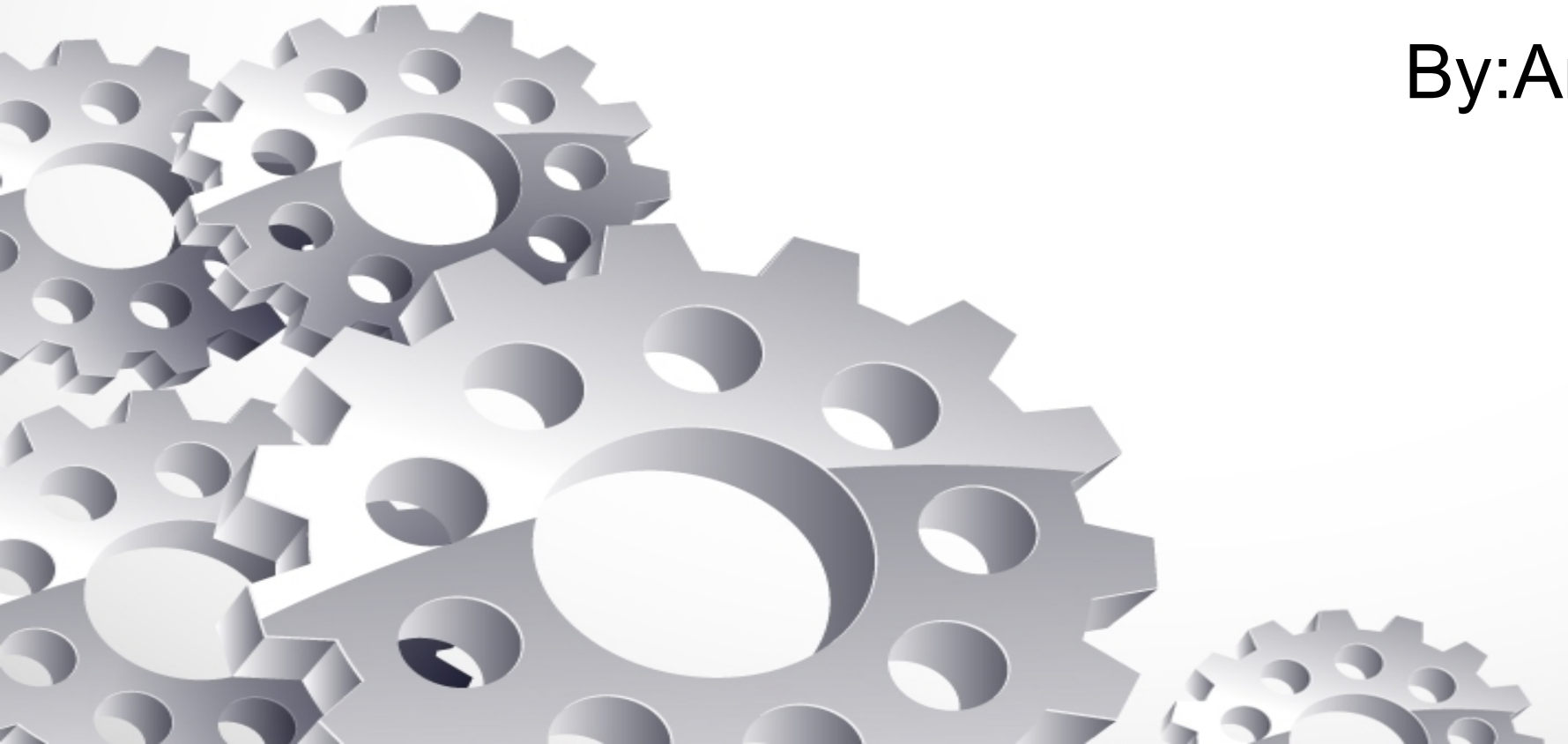
Opening a New Shopping Mall in Toronto, Canada

By:Arundhati Varma



Business Problem

- The objective of this capstone project is to analyse and select the best locations in the city of Toronto, Canada to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Toronto, Canada, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience

- Target Audience of this project This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Canada i.e. Toronto.

Data To solve the problem

- we will need the following data:
- • List of neighborhoods in Toronto. This defines the scope of this project which is confined to the city of Toronto, the capital city of the country of Canada in Northern America
- • Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- • Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Sources of data

- This Wikipedia page This Wikipedia page ('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M') contains a list of neighborhoods in Toronto, with a total of 70 neighborhoods.

Methods to extract Data

- We will use **web scraping techniques** to extract the data from the Wikipedia page, with the help of Python requests and **beautifulsoup packages**. Then we will get the geographical coordinates of the neighborhoods using Python **Geocoder package** which will give us the latitude and longitude coordinates of the neighborhoods. After that, we will use **Foursquare API** to get the venue data for those neighborhoods.

Results

The results from the k-means clustering show that we can categorize the neighborhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:

- Cluster 0: neighborhoods with moderate number of shopping malls
- Cluster 1: neighborhoods with low number to no existence of shopping malls
- Cluster 2: neighborhoods with high concentration of shopping malls

Conclusion

- This project recommends property developers to capitalize on these findings to open new shopping malls in neighborhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 2 which already have high concentration of shopping malls and suffering from intense competition.

Thank You