IBM APPLIED DATA SCIENCE CAPSTONE

Co-working Facility Recommendation based on Attractiveness of Neighbourhood in Toronto

(Toronto_Clusters)

INTRODUCTION

- Coworking and flexi work space - traction in recent years
- Attractive to independent contractors, independent scientists, telecommuting and work-at-home professionals
- Many new companies and institutions which are into providing coworking or flexi workspace facilities
- Analysis of Toronto’s neighbourhoods will be helpful to these companies and also for any company who plans to relocate to Toronto.

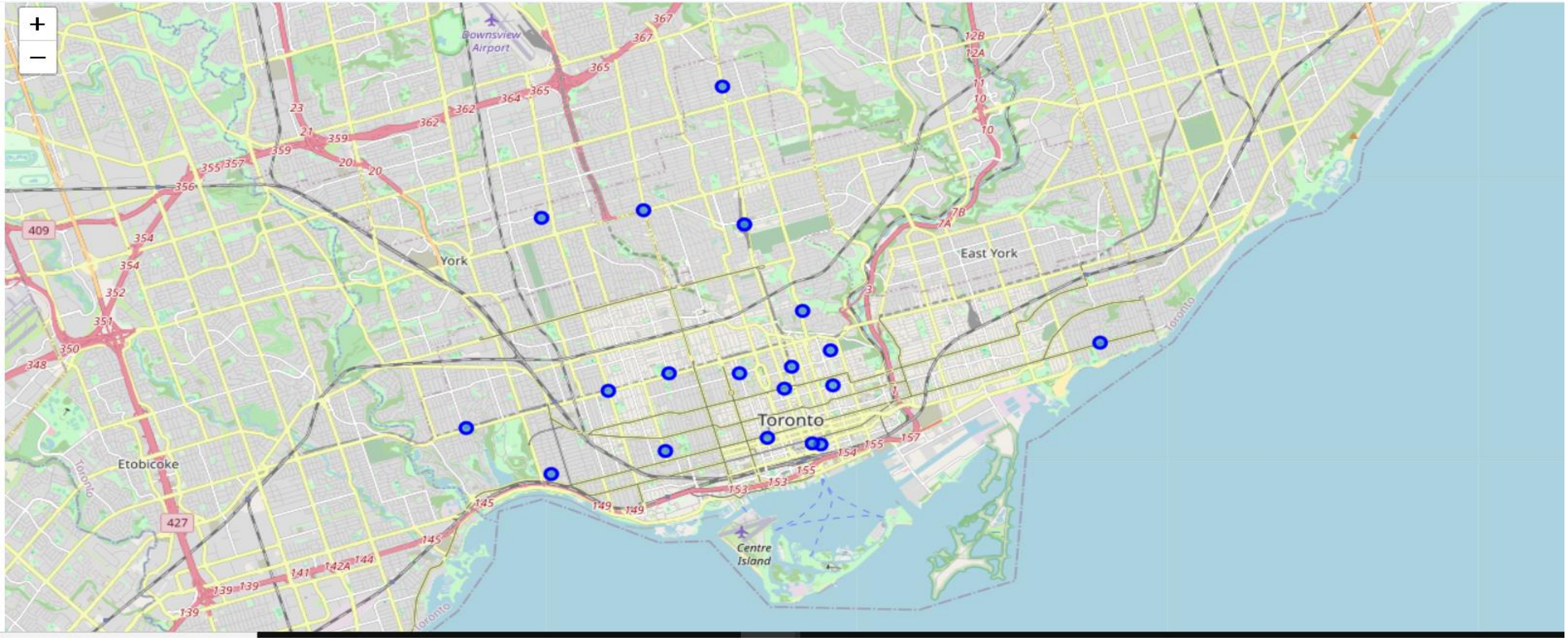
Problem Description

- Issues with Coworking and flexi work space
  - High maintenance costs (rentals, free services such as coffee, gym etc.)
  - Profitability directly depends on the occupancy rates and overhead costs
- **Choosing correct location is a key success factor**
- Most potential neighbourhoods are which have :
  - Surrounding facilities such as cafes, restaurants
  - Average real estate rental is low
  - Cater to larger size of cluster/ population
- Similarity of neighbourhood is very important for success of a coworking facility, so we also need to find the right cluster.

Data

Criteria for selection/ Data Required	Data source
Size of cluster	Number of postal codes is taken as surrogate of largeness of size/ population
Facilities such as cafes, restaurants	Foursquare data (via explore option)
Average real estate rental	Kaggle dataset (pre-cleaned)

Toronto - Neighborhoods



Methodology

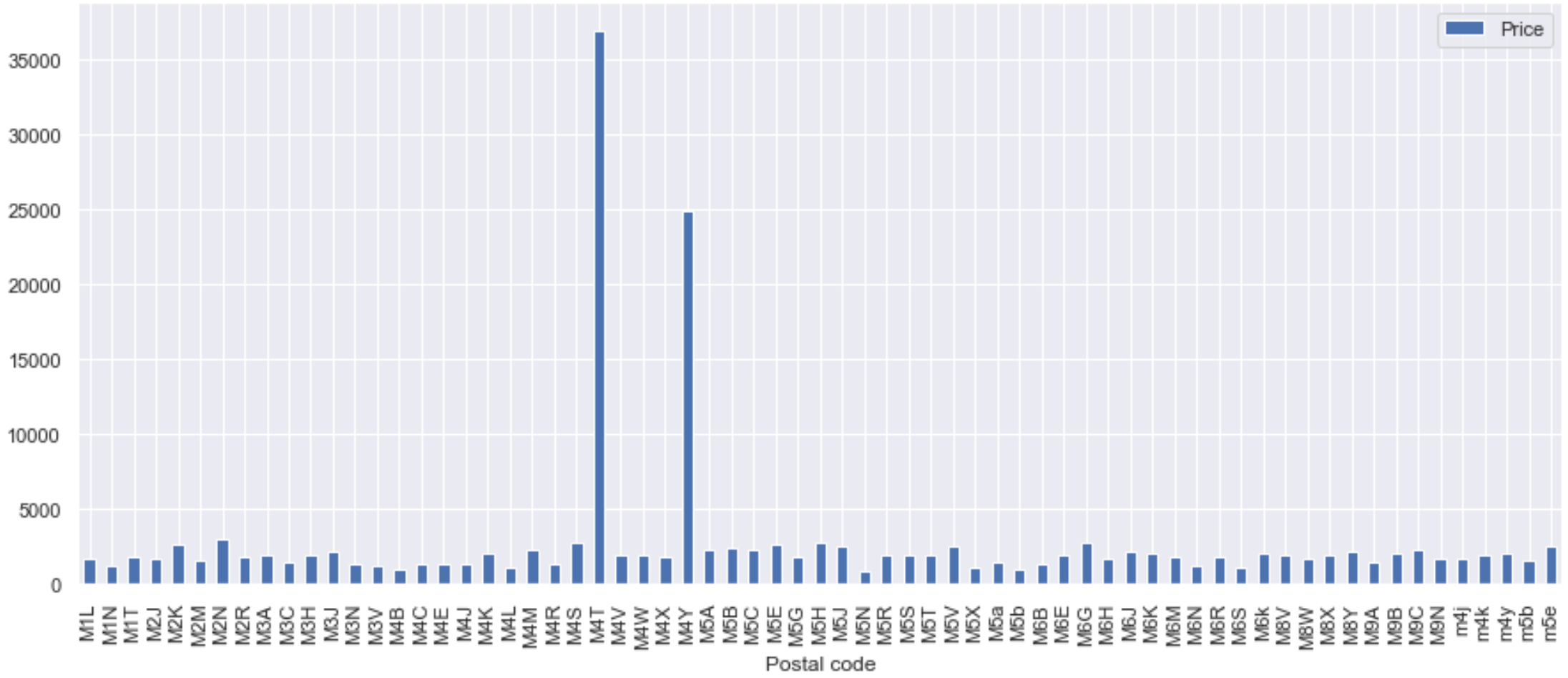
1. Collect & Inspect Data
    1. Drop missing data,
    2. Group Foursquare data based on postal codes
2. Explore data and Preprocessing
    1. Geolocate neighborhood data
    2. Encode data,
    3. Explore rental data to see out-liers
    4. Compute average rental for each neighborhood
3. Data Analysis and Modeling
    1. Identify top 5 common locations in each neighborhood
    2. Using K-Means clustering to identify 4 clusters
    3. Appending average rentals to identify the preferred location

Results

Cluster name	Average Rental	No of postal codes/ Area size	Number of venues	Preferred (Yes/ No)
Cluster 1 (cls_1)	$6163.31	6	High	No
Cluster 2 (cls_2)	$1677.25	8	High	Yes
Cluster 3 (cls_3)	$1300.00	1	Very Low	No
Cluster 3 (cls_3)	$1825.00	1	Very Low	No

As per the analysis above neighbourhood cluster2 (cls_2) is the preferred choice as check all the criteria/ factors i.e. larger areas as highest number of postal codes (8), thereby highest surrounding facilities such as cafes, restaurants in the cluster and lowest average real estate rental ($1677.25).

Appendix


Price
35000
30000
25000
20000
15000
10000
5000
0
M1L M1N M1T M2J M2K M2M M2N M2R M3A M3C M3H M3J M3N M3V M4B M4C M4E M4J M4K M4L M4M M4R M4S M4T M4V M4W M4X M4Y M5A M5B M5C M5E M5G M5H M5J M5N M5R M5S M5T M5V M5X M5a M5b M6B M6E M6G M6H M6J M6K M6M M6N M6R M6S M6k M8V M8W M8X M8Y M9A M9B M9C M9N m4j m4k m4y m5b m5e
Postal code