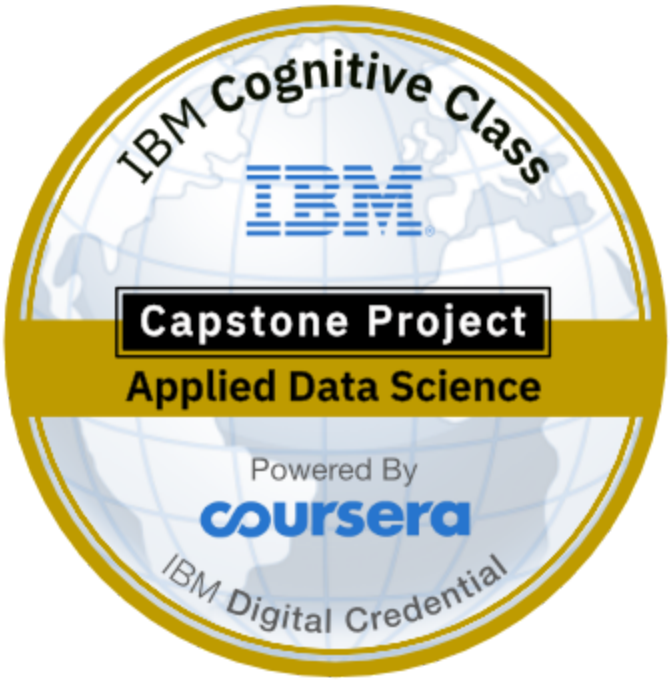
Data Science Project
REPORT

Segmenting Neighbourhoods of Toronto on the basis of Livability



ANANTH S G

29/Feb/2020

1. INTRODUCTION

- ## Background

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of about 3 Million as of July 2018. Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. Toronto is the New York of Canada.

Toronto ranks in the top 5 of the most livable cities in the world according to the Economist Intelligence Unit. Thanks to the booming economy, the inflow of immigrants the real estate market of Toronto is heating up.

For prospective home buyers and businessmen, it will be useful to know which neighborhoods of Toronto are more attractive for investment.

- ## Problem

We know that Toronto is among the top livable cities of the world. But no comprehensive study done to rank the livability of the neighborhoods of Toronto. Livability Score of a neighborhood depends on proximity to Restaurants, Nightlife options, Shops, Schools, Hospitals, Public Transport and Offices. For simplicity we only take into consideration the Restaurants, Bars, Schools and Shops to access the Livability Score.

- ## Interest

Grouping neighborhoods by Livability Scores will be extremely useful for the following:

1. **Real Estate agents would want to know how the commercial & housing prices vary with Livability Score**
2. **Online listing companies such as Realtor.com, Zillow can provide customers the options to browse through livability scores of neighbourhoods**
3. **Most importantly the customers would like compare the livability score of neighbourhoods with the Housing Prices before making the purchase**

2. Data

- Data Sources

  The list of Toronto Neighborhoods and the postal codes can be found in wikipedia(web-scraping). The location data of each neighbourhood can be obtained from geocoder api. Foursquare API provides the list of Restaurants, Bars, Schools and Shops in a neighborhood.

- Data Cleaning

  The neighborhood data was scraped from the website and merged with the location data.

  The neighbourhood data had some unassigned boroughs which were removed. If the neighborhood name was unassigned but it had a corresponding borough value, the borough name was used the neighborhood name.

  The location data that we got was based on the postal codes. That is why the neighborhood data had to be grouped. The grouping was done on the basis of postal code. If multiple neighbourhood names shared a single postal code, then the names were concatenated with 'comma' separator.

  The final merged data set containing the neighborhood & location data had data points for Toronto & its suburbs. For this analysis we only consider the data points for Toronto city.

  This is what the merged data set looks like.

	Postal-Code	Borough	Neighbourhood	Latitude	Longitude
0	M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568
1	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
2	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191
3	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558
4	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420

- **Feature Selection**

  The livability score of a Toronto neighbourhood depends on how many venues are present within 500m (walking distance) of its geographical location.

  For simplicity, we assume that the livability score depends on criteria such as lifestyle and the convenience. The criteria Lifestyle is the number of Bars & Restaurants nearby. The criteria convenience is the number of schools and shops nearby. High lifestyle score is preferable for young people and a high convenience score is preferred by everybody.

  So, Livability Score = Restaurants,Bars,Schools & Shops

  We can extract the count of all these venue categories using the Foursquare API.

  Once we fetch the number of Restaurants,Bars,Schools & Shops for every neighborhood, we merge this data with the original data set.

  This is a snapshot of Toronto neighbourhood data along with the feature set needed to group neighbourhoods by Livability Score.

	Postal-Code	Borough	Neighbourhood	Latitude	Longitude	Restaurants	Bars	Shops	Schools
0	M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568	42.0	50.0	50.0	12.0
1	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306	13.0	45.0	24.0	7.0
2	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	4.0	9.0	5.0	3.0
3	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	0.0	3.0	2.0	0.0
4	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0.0	0.0	0.0	1.0

3. Exploratory Data Analysis

- Calculation of Target Variable
- Relationship
-

4. Predictive Modelling

- Algorithms
- Solution to the problems
- Performance to different models
-

5. Conclusions

6. Future Directions

7. References