2020

IBM Applied Data Science Capstone Project Report

EVALUATING HOUSING PRICES FOR TORONTO NEIGHBOURHOODS UTILIZING REAL ESTATE AND NEAR BY VENUES DATA

TARIQ PERVEZ

Contents

Introduction

Background

Toronto is the capital hub of Ontario province with enormous employment and business opportunities. These recompenses have gained attention of several local Canadians and international immigrants, aspiring them to settle and lead a good lifestyle in the city of Toronto. But with all these amenities, there lies Toronto's real estate market hype. For the last few years, significant influx of locals from other provinces of Canada and immigrants from worldwide have outpaced housing demand in comparison to supply. This has lead prices inch steadily upward developing a bubble in Toronto real estate market.

Business Problem

Immigrants and locals moving into Toronto city founds selection of house challenging for them and their families due to hot housing market. Affordability, accommodation features and nearby facilities always remain qualifying parameters for selection of an appropriate place to live for them.

Price variation with respect to housing attributes for different Toronto neighbourhoods in conjunction with nearby venues needs to be addressed to resolve the encounters faced by any individual during selection of a place. The project seeks to explore real estate data to get an insight of property price variation in combination with its traits and near services available based on location data along all the neighbourhoods to establish relationship between them, which can be used as a recommendation while selecting a place to live.

Audience & Stakeholders

The target audience for this report are:

- Locals planning to settle from other provinces of Canada to Toronto city
- Immigrants across the globe planning to settle in the city of Toronto
- Potential house sellers and buyers residing in Toronto city who can optimize their advertisements
- Potential business investors setting up their businesses in Toronto city and aims to accommodate their workforce
- City planning authorities to set up more amenities in neighbourhoods with less venues

Data

This section describes the data sourced for this project.

Data Sources

This project integrates data sources such as Home finder Canada, Google Map, Wellbeing Toronto as well as Foursquare data. This section describes each of these data sources and provides examples of the data.

Home finder Toronto Data

Home finder Canada is a leading real estate website covering details of commercial and residential properties available for sale. The website consists of around more than 4000 registered properties. In view to the scope of the project, family living options such as Detached, Condo & Town Houses in Toronto city are considered. This

data aims to provide details such as property type (Detached/Condo/Town House), price, details such as number of bedrooms & bathrooms with their respective addresses. Since the data is available on homefinder.ca website, the data have been web scraped from the following link;

https://www.homefinder.ca/homes-for-sale/47618-toronto-real-estate?utf8=%E2%9C%93&sort_order=suggested&bedrooms=Any&advanced_filters%5Blifestyle%5D%5Badult_living%5D=Adult+Living&advanced_filters%5Blifestyle%5D%5Bfamily%5D=Family&advanced_filters%5Blifestyle%5D%5Byoung_professional%5D=Young+Professional&advanced_filters%5Blifestyle%5D%5Bspacious_living%5D=Spacious+Living&commit=Filter+Search&property_types%5Bcondos%5D=Condos&property_types%5Bdetached%5D=Detached&property_types%5Btowns_semi%5D=Towns+%26+Semi

Home Finder Toronto Data Sample

Following is a sample of the scraped data extracted from homefinder.ca website showing the property price, details, type and address:

	Property Price	Property Details	Property Type	Property Address
0	$624,900	1 + 1 beds | 1 bath	Condo	5 Marine Parade Dr Unit 211, Toronto
1	$729,900	2 + 1 beds | 1 bath	Detached	5 Hatfield Cres, Toronto
2	$899,900	3 beds | 2 baths	Detached	534 Rouge Hills Dr, Toronto
3	$849,900	3 + 1 beds | 2 baths	Detached	6 Bolger Pl, Toronto

Google Map API Data

Since Home Finder data sample doesn't have latitude and longitude data related to each property, which will be a pre-requisite for visualizing data on choropleth map using folium. Moreover, it will also help in finding out nearby venues using Foursquare API. In this regard, google map API is used to extract latitude and longitude of every property

Home Finder Toronto with Google Map API Latitude & Longitude Data sample

Following is a sample of the Home Finder data with latitudes and longitudes extracted utilizing google map API:

	Property Price	Property Details	Property Type	Property Address	Property Latitude	Property Longitude
0	624900	1 + 1 beds | 1 bath	Condo	5 Marine Parade Dr Unit 211 Toronto	43.6299	-79.4756
1	729900	2 + 1 beds | 1 bath	Detached	5 Hatfield Cres Toronto	43.7137	-79.5463
2	899900	3 beds | 2 baths	Detached	534 Rouge Hills Dr Toronto	43.7996	-79.1344
3	849900	3 + 1 beds | 2 baths	Detached	6 Bolger Pl Toronto	43.7248	-79.5734

Wellbeing Toronto Neighbourhood Boundaries Data

Wellbeing Toronto (WbTo) provides neighbourhood-level datasets about Toronto services, facilities, and well-being. The datasets are segmented by community indicators grouped under 11 categories, including

demographics, civics, and health as examples. While primarily a mapping application, and underlying datasets are downloadable.

For this project, the dataset of interest includes latitude and longitude data of each neighbourhood and polygon co-ordinates showing limit for each of Toronto's 140 neighbourhood. These data are downloadable in csv/shape/geojson/geopackage format file. Wellbeing Toronto is publicly accessible at https://open.toronto.ca/dataset/neighbourhoods/.

Wellbeing Toronto Neighbourhood Boundaries Data Sample

Following is a sample of the neighbourhoods boundaries shape file which includes geometry:

	geometry
0	POLYGON ((-79.43592 43.68015, -79.43492 43.680...
1	POLYGON ((-79.41096 43.70408, -79.40962 43.704...
2	POLYGON ((-79.39119 43.68108, -79.39141 43.680...
3	POLYGON ((-79.50529 43.75987, -79.50488 43.759...
4	POLYGON ((-79.43969 43.70561, -79.44011 43.705...

Following is a sample of the neighbourhoods boundaries csv file which includes neighbourhood names with their latitude, longitude and geometry;

	_id	AREA_ID	AREA_ATTR_ID	PARENT_AREA_ID	AREA_SHORT_CODE	AREA_LONG_CODE	AREA_NAME	AREA_DESC	X	Y	LONGITUDE	LATITUDE	OBJECTID	Shape__Area	Shape__Length	geometry
0	5461	25886861	25926662	49885	94	94	Wychwood (94)	Wychwood (94)	NaN	NaN	-79.425515	43.676919	16491505	3.217960e+06	7515.779658	{u'type': u'Polygon', u'coordinates': (((-79.4...
1	5462	25886820	25926663	49885	100	100	Yonge-Eglinton (100)	Yonge-Eglinton (100)	NaN	NaN	-79.403590	43.704689	16491521	3.160334e+06	7872.021074	{u'type': u'Polygon', u'coordinates': (((-79.4...

Above mentioned both datasets are joined using AREA_NAME which is actually neighbourhood name, LONGITUDE, LATITUDE with the geometry in shape file. In this way, now we have a new dataset of all the neighbourhoods with their defined latitude, longitude and boundaries in terms of polygon which is as follow;

	AREA_NAME	LONGITUDE	LATITUDE	geometry
0	Wychwood (94)	-79.425515	43.676919	POLYGON ((-79.43592 43.68015, -79.43492 43.680...
1	Yonge-Eglinton (100)	-79.403590	43.704689	POLYGON ((-79.41096 43.70408, -79.40962 43.704...
2	Yonge-St.Clair (97)	-79.397871	43.687859	POLYGON ((-79.39119 43.68108, -79.39141 43.680...
3	York University Heights (27)	-79.488883	43.765736	POLYGON ((-79.50529 43.75987, -79.50488 43.759...
4	Yorkdale-Glen Park (31)	-79.457108	43.714672	POLYGON ((-79.43969 43.70561, -79.44011 43.705...

This data set aims to supplement Home Finder data set by assigning each property their respective neighbourhood name, latitude and longitude by checking property latitude and longitude from Home Finder data set that whether it lies within the specified neighbourhood polygon co-ordinates provided in above attached table. Moreover, neighbourhood boundaries Geojson file is utilized for visualizing neighbourhood extents.

Foursquare Data

Foursquare provides a mobile app that allows users to search for near-by venues and see information and reviews. Users also feed information back to Foursquare both passively, as the app tracks users' locations, and actively as users enter venue names, locations, and reviews.

Since 2009, users have provided Foursquare with location data on over 105 million venues, with over 75 million tips from local experts. As one of the largest sources of location-based venue data, the company describes itself as a technology company that uses location intelligence to build meaningful consumer experiences and business solutions.

This project will access Foursquare venue data for all neighbourhoods. The Foursquare venue data will particularly seek to identify venues that have significant impact on property prices. These data will then be used for subsequent comparison and categorization to provide insight to the business problem.

The Foursquare venue data are accessible via application programming interface (API). A free developer account is used to access the data from https://developer.foursquare.com/places-api.

Foursquare Data Sample

Following is a sample of the imported data showing particularly the venues (by name) and the respective venue categories for each neighbourhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop