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

TARIQ PERVEZ

Outline

- Introduction
- Data
- Methodology
- Results
- Discussion
- Conclusion

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
- But with all these amenities, there lies Toronto's real estate market hype

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
- 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

Data

Data Sources

Source#1

Real estate data web scraped from homefinder.ca website

	PropertyPrice	PropertyDetails	PropertyType	PropertyAddress
0	$849,900	3 + 1 beds | 2 baths	Detached	6 Bolger Pl, Toronto
1	$899,900	3 beds | 2 baths	Detached	534 Rouge Hills Dr, Toronto
2	$3,199,988	2 beds | 1 bath	Detached	400 Hollywood Ave, Toronto
3	$624,900	1 + 1 beds | 1 bath	Condo	5 Marine Parade Dr Unit 211, Toronto

Source#2

Neighbourhood boundaries data extracted from well being Toronto website in csv & shape file format https://open.toronto.ca/dataset/neighbourhoods/.

	_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...
2	5463	25886834	25926664	49885	97	97	Yonge-St.Clair (97)	Yonge-St.Clair (97)	NaN	NaN	-79.397871	43.687859	16491537	2.222464e+06	8130.411276	{u'type': u'Polygon', u'coordinates': (((-79.3...

	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...

Source#3

Neighbourhood venue data extracted using Foursqaure API

	Neighbourhood	NeighbourhoodLatitude	NeighbuorhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Agincourt North	43.805441	-79.266712	Menchie's	43.808338	-79.266288	Frozen Yogurt Shop
1	Agincourt North	43.805441	-79.266712	Saravanaa Bhavan South Indian Restaurant	43.810117	-79.269275	Indian Restaurant
2	Agincourt North	43.805441	-79.266712	Shoppers Drug Mart	43.808894	-79.269854	Pharmacy
3	Agincourt North	43.805441	-79.266712	Booster Juice	43.809915	-79.269382	Juice Bar

Data

Data Gathering

Merger#1

Google map API utilized to find latitude and longitude for real estate data that was web scraped from homefinder.ca website

	PropertyPrice	PropertyType	PropertyAddress	Bedroom	Bathroom	PropertyLatitude	PropertyLongitude
0	849900.0	Detached	6 Bolger Pl Toronto	3 + 1 beds	2 baths	43.7248	-79.5734
1	899900.0	Detached	534 Rouge Hills Dr Toronto	3 beds	2 baths	43.7996	-79.1344
2	624900.0	Condo	5 Marine Parade Dr Unit 211 Toronto	1 + 1 beds	1 bath	43.6299	-79.4756
3	729900.0	Detached	5 Hatfield Cres Toronto	2 + 1 beds	1 bath	43.7137	-79.5463
4	89900.0	Detached	214 Old Mill Campground Timmins	2 beds	1 bath	48.4758	-81.3305

Merger#2

Useful data from shape file & csv file of neighbourhood boundaries data https://open.toronto.ca/dataset/neighbourhoods/.was joined together to a single data frame

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

Merger#3

Merger#1 & 2 data frame was merged together into a single data frame df

	PropertyPrice	PropertyType	PropertyAddress	Bedroom	Bathroom	PropertyLatitude	PropertyLongitude	Neighbourhood	geometry	NeighbourhoodLatitude	NeighbourhoodLongitude
0	849900	Detached	6 Bolger Pl Toronto	3 + 1 beds	2 baths	43.724761	-79.573404	West Humber-Clairville	POLYGON ((-79.5235701287411 43.7094692811074,...	43.716180	-79.596356
1	899900	Detached	534 Rouge Hills Dr Toronto	3 beds	2 baths	43.799639	-79.134369	Rouge	POLYGON ((-79.1970100596851 43.7965219057717,...	43.821201	-79.186343
2	624900	Condo	5 Marine Parade Dr Unit 211 Toronto	1 + 1 beds	1 bath	43.629903	-79.475574	Mimico (includes Humber Bay Shores	POLYGON ((-79.4603951138219 43.6210724571131, ...	43.615924	-79.500137
3	729900	Detached	5 Hatfield Cres Toronto	2 + 1 beds	1 bath	43.713699	-79.546261	Elms-Old Rexdale	POLYGON ((-79.5551162536014 43.7151024596992, ...	43.721519	-79.549863
4	275000	Condo	4645 Jane St Unit 1101 Toronto	2 beds	1 bath	43.769324	-79.519773	Black Creek	POLYGON ((-79.5348614539865 43.7726881362668, ...	43.764890	-79.521979

Methodology

Exploratory Data Analysis

- Histogram
- Box Plot
- Bar Plot

Data Visualization

- Chloropleth Map

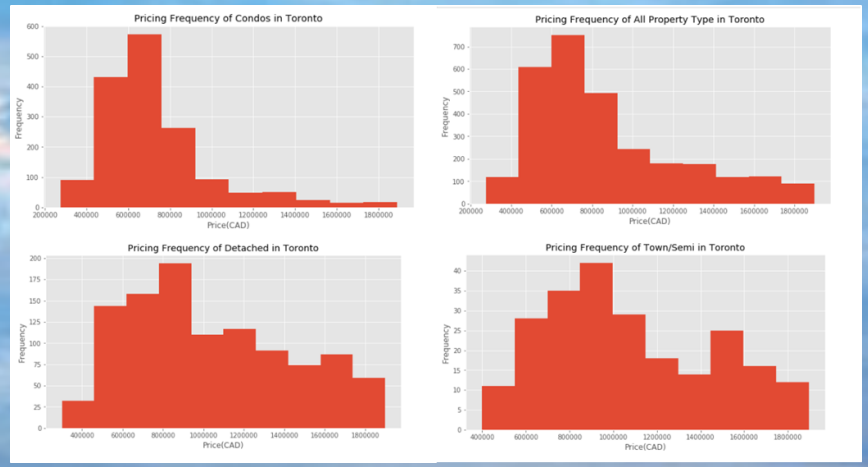
Model Development & Evaluation: Machine Learning Approach

- Linear Regression
- Polynomial Regression
- K Nearest Neighbour Regression
- Random Forest Regression

Exploratory Data Analysis

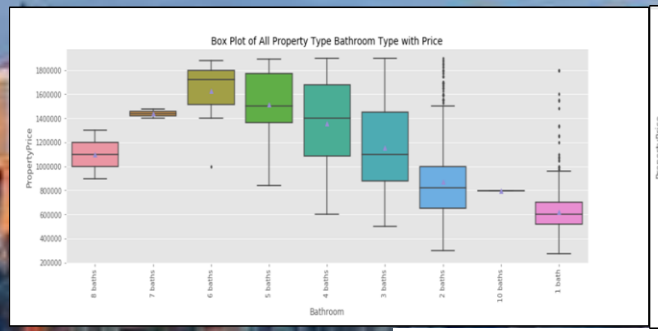
Histogram

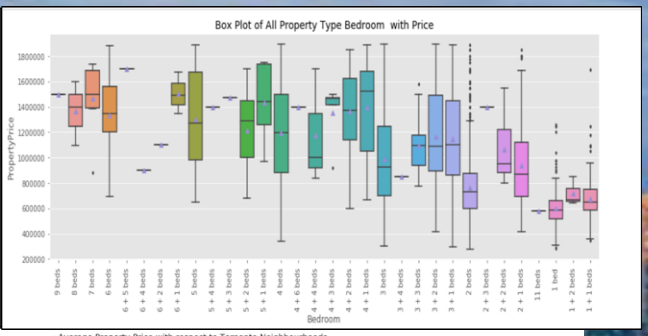
- This exploration of data present four histograms for comparison of property prices for all property type, condos, detached & town/semi houses



Box Plot

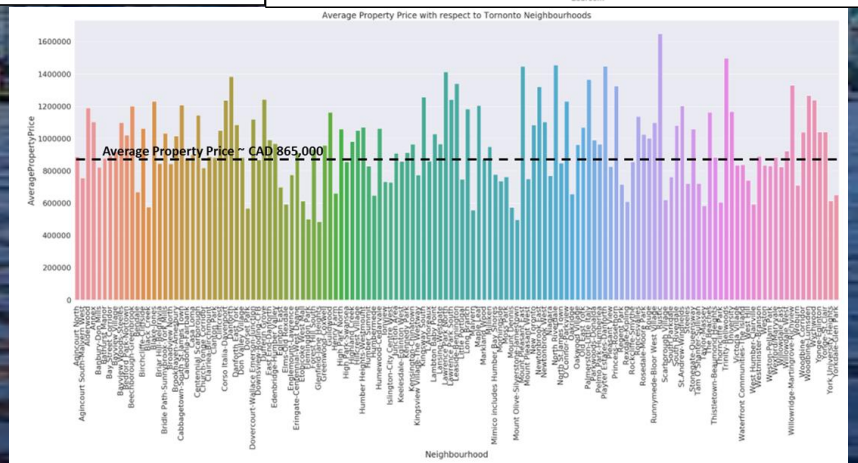
This exploration of data presents boxplots in order to investigate the distribution of property price with respect to any property number of bathrooms & bedrooms for all the cases as mentioned above in histogram section





Bar Plot

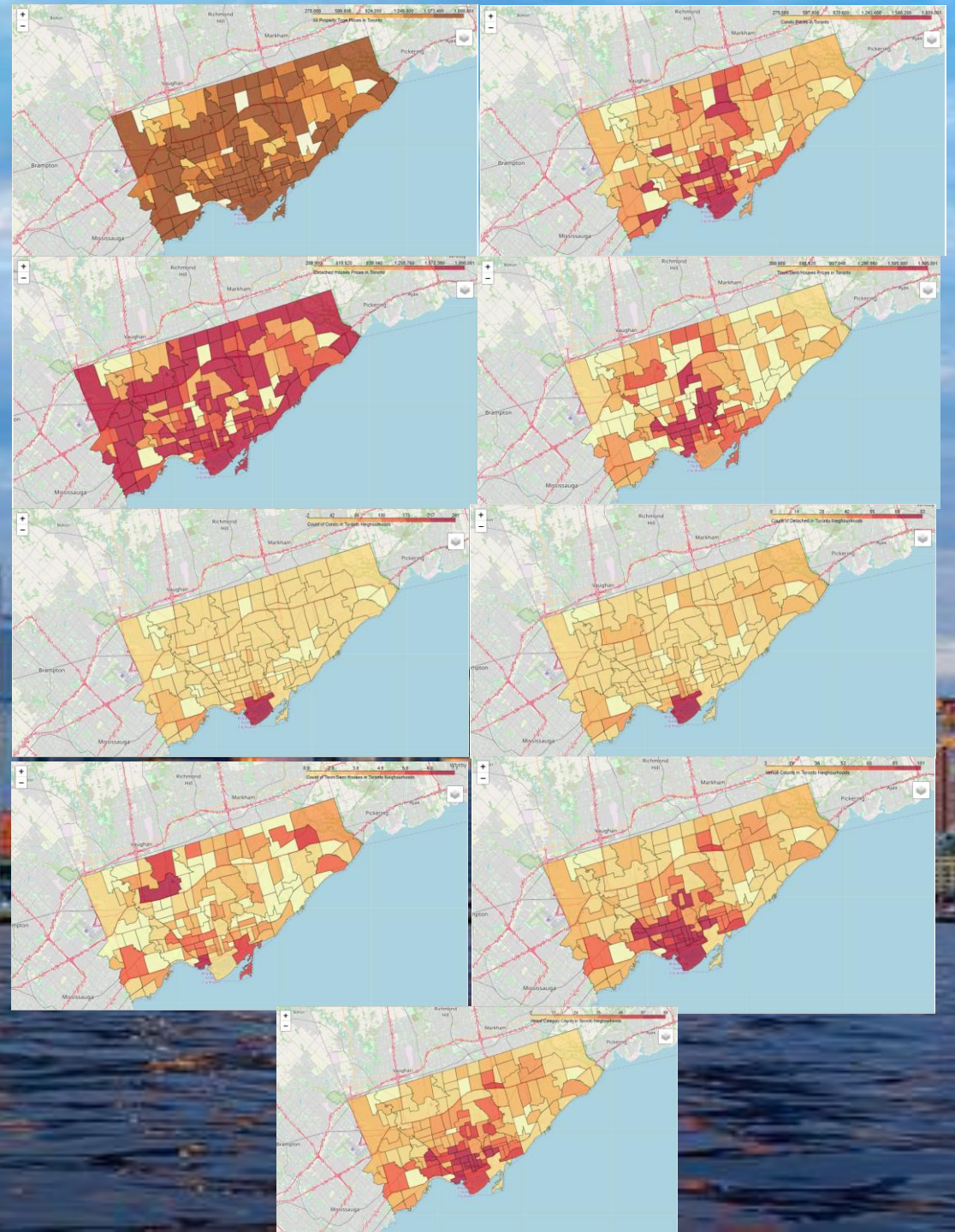
This exploration of data presents the average property price of Toronto's neighbourhoods



Data Visualization

This section will cover the visual exploration of the data utilizing Folium Chloropleth map to get an insight of the following attributes;

- Property price for all property type
- Property price for Condos
- Property price for Detached Houses
- Property price for Town/Semi Houses
- Count of Condos
- Count of Detached Houses
- Count of Town/Semi Houses
- Venue Count
- Venue Category Count



Model Development & Evaluation: Machine Learning Approach

Machine learning approach regression was chosen because of its simplicity and with the aid of Sklearn library implementation of model is quick and easy which is perfect to start the analyzing process. Regression approach was used to develop model for following dependent & independent variable;

- Number of bedrooms versus price for all property types
- Number of bathrooms versus price for all property types
- Number of bedrooms versus price for Condos
- Number of bathrooms versus price for Condos
- Number of bedrooms versus price for Detached Houses
- Number of bathrooms versus price for Detached Houses
- Number of bedrooms versus price for Town/Semi Houses
- Number of bathrooms versus price for Town/Semi Houses
- Neighbourhood venue count versus average property price
- Neighbourhood venue category count versus average property price
- Neighbourhood venue category versus average property price

Linear Regression

Below is the tabulated summary of Linear Regression results for all cases;

	Linear Regression Results											
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count		Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom				
Regression Coefficient (R^2)	-1.41E+24	0.41	0.15	0.26	-1.23E+26	0.45	0.09	0.25	-0.02	-0.02		-2.29
Mean Squared Error (MSE)	1.81E+35	7.59E+10	5.89E+10	5.15E+10	1.98E+37	8.96E+10	1.61E+11	1.34E+11	6.04.E+10	6.04E+10		1.96E+11

In addition to this distribution plots were generated to visualize actual versus fitted values based on linear regression model

Polynomial Regression

Below is the tabulated summary of Polynomial Regression results for all cases;

	Polynomial Regression Results											
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count		Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom				
Regression Coefficient (R^2)	-1.10E+23	0.41	0.15	0.26	-1.14E+23	0.44	-1.45E+26	0.25	-0.01	-0.01		-1.31
Mean Squared Error (MSE)	1.42E+34	7.625E+10	5.89E+10	5.12E+10	1.85E+34	8.98E+10	2.58E+37	1.33E+11	5.99E+10	5.98E+10		1.37E+11

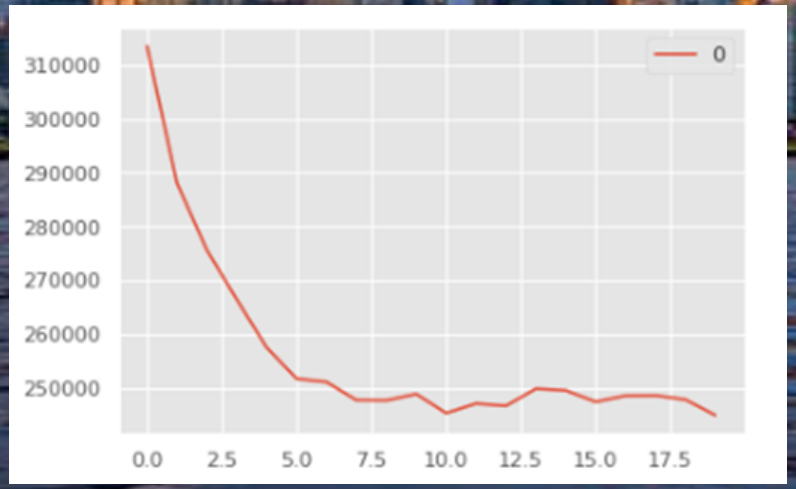
In addition to this for every case polynomial degree sensitivities were run to optimize regression coefficient (R^2) value

K Nearest Neighbour Regression

Below is the tabulated summary of K Nearest Neighbour Regression results for all cases;

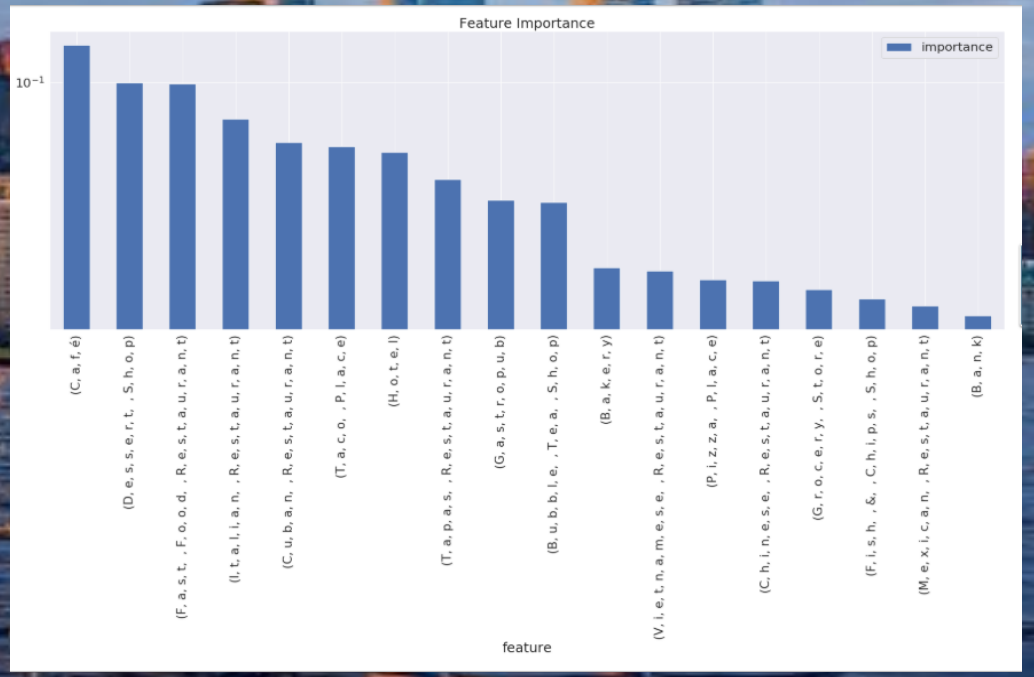
	K Nearest Neighbour Regression Results										
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count	Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom			
Regression Coefficient (R^2)	0.34	0.37	0.10	0.26	0.15	0.44	0.03	0.20	-0.07	-0.06	-0.01
Mean Squared Error (MSE)	8.49E+10	8.159E+10	6.26E+10	5.14E+10	1.37E+11	9.04E+10	1.72E+11	1.41957E+11	6.35E+10	6.33E+10	5.99E+10
Root Mean Squared Error(RMSE)	2.91E+05	2.86E+05	2.50E+05	2.27E+05	3.71E+05	3.01E+05	4.15E+05	3.77E+05	2.52E+05	2.52E+05	2.45E+05

In addition to this for every case K value was determined through Grid Search CV & elbow plot was also generated to find the optimum K value



Random Forest Regression

- Based on previous regression techniques we came to a conclusion that bathroom count seems to show a predictability with respect to price, but in terms of nearby venues above developed models show poor relationship. Therefore, based on this concern random forest regression with feature importance is generated to determine which venues play a vital role in inching any property price upward.
- Random forest regression estimates regression coefficient of ~ -0.016 with mean squared error (mse) of ~ 6.04E+10
- Based on feature importance cafes, dessert shops & fast food restaurants have major impact on property price



Results

This results section provides an overview of the outcomes of the methodology and their relevance to the original business problem

Histogram

- Pricing frequency in Toronto lies between CAD 450,000 - 920,000 for all property types
- Condos possess less pricing frequency in comparison to other property types making them more affordable for an individual

Box Plot

- For all the cases strong to moderate association was observed with the number of bathrooms & property price, whereas weak to moderate association was observed with the number of bedroom & property in the case of all property types & detached houses
- In view to this, spearman's rank correlation was applied to the dataset which also showed similar results. Strength of correlation (Rs) was observed more in case of bathrooms in comparison to bedroom.

Bar Plot

- Average property price across Toronto city is CAD 865,000 and most of neighbourhood lies above it
- Rustic & Trinity Bellwoods are considered as the most expensive neighbourhoods of Toronto city

Results

Data Visualization

Following insights were gathered as a result of generating Folium Chloropleth map for all the cases;

- Majority of neighbourhoods fall under the highest bracket of price for all property types leaving few that lies near to the boundary of Toronto city in North York Borough and in the central part of Scarborough District
- Condos located in the downtown and midtown Toronto falls in the highest price bracket
- Detached houses in most of the neighbourhoods in Toronto falls under the highest price bracket, except few neighbourhoods which are located near to boundary of Toronto city in North York Borough and the ones which are in the central part of Scarborough district
- Most of town/semi houses falls under lower to mid-price bracket except few ones which are in midtown & west end district
- Maximum number of condos & detached houses are concentrated in downtown Toronto in Waterfront Communities-The island
- Neighbourhoods including Downsview Roding-CFB & Niagara have maximum number of town/semi houses
- Maximum number of venues & unique venues category are concentrated in Midtown & Downtown Toronto

Results

Model Development & Evaluation: Machine Learning Approach

	Linear Regression Results											
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count		Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom				
Regression Coefficient (R^2)	-1.41E+24	0.41	0.15	0.26	-1.23E+26	0.45	0.09	0.25	-0.02	-0.02		-2.29
Mean Squared Error (MSE)	1.81E+35	7.59E+10	5.89E+10	5.15E+10	1.98E+37	8.96E+10	1.61E+11	1.34E+11	6.04.E+10	6.04E+10		1.96E+11

	Polynomial Regression Results											
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count		Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom				
Regression Coefficient (R^2)	-1.10E+23	0.41	0.15	0.26	-1.14E+23	0.44	-1.45E+26	0.25	-0.01	-0.01		-1.31
Mean Squared Error (MSE)	1.42E+34	7.625E+10	5.89E+10	5.12E+10	1.85E+34	8.98E+10	2.58E+37	1.33E+11	5.99E+10	5.98E+10		1.37E+11

	K Nearest Neighbour Regression Results											
	All Property Types		Condos		Detached Houses		Town/Semi Houses		Venue Count	Venue Category Count		Venue Category
	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom	Bedroom	Bathroom				
Regression Coefficient (R^2)	0.34	0.37	0.10	0.26	0.15	0.44	0.03	0.20	-0.07	-0.06		-0.01
Mean Squared Error (MSE)	8.49E+10	8.159E+10	6.26E+10	5.14E+10	1.37E+11	9.04E+10	1.72E+11	1.41957E+11	6.35E+10	6.33E+10		5.99E+10
Root Mean Squared Error(RMSE)	2.91E+05	2.86E+05	2.50E+05	2.27E+05	3.71E+05	3.01E+05	4.15E+05	3.77E+05	2.52E+05	2.52E+05		2.45E+05

Results

Model Development & Evaluation: Machine Learning Approach

Based on regression model results;

- Number of bathrooms shows moderate to strong association with property price whereas number of bedrooms shows poor association with property price
- One key observation is that the KNN regression tends to improve R^2 value for bedrooms in the case of all property type and detached houses
- Venue data shows poor association with property price for all the three regression techniques. In this regard, random forest regression with feature importance was performed which suggested cafes, dessert shops & fast food restaurants as the top three venues that play a vital role in inching any property price upward

Discussion & Recommendations

As a result of subject analysis, following trail can be followed in order to find suitable place to live;

- For any newcomer, the main district of interest for finding accommodation should be the neighbourhood located at the border of Toronto city in North York and in the central part of Scarborough
- Secondly, property type condo should be the focus due to its affordability in comparison to other property types
- Thirdly, number of bathrooms requirement to be optimized as per family need as it has a moderate to strong association with property price
- Lastly, consider those neighbourhoods which are without or at least have minimum number of venues such as cafes, dessert shops & fast food restaurants as they have a major impact on property price

One more key fact which has always kept Toronto's real estate market in a bubble is due to less supply of accommodations in comparison to demand. This is due to continuous influx of locals from other provinces of country and immigrants from all around the globe which has led increase in Toronto population with a smaller number of places to live. Since Canada aims to increase invitations for more immigrants in coming years from all around the globe which will keep rising Toronto population leaving real estate market hype.

Conclusion

Since the objective of the project was to investigate the relationship between property attributes and nearby venues with respect to property price. As a result of subject study and analysis done with integration of real estate and location data, following are the outcomes and observations;

- Exploratory data analysis in conjunction with modelling suggests bathroom count as key attribute for property price prediction
- Rustic and Trinity Bellwoods seems to be the most expensive neighbourhoods in Toronto
- Based on data visualization, neighbourhoods closer to the boundary of Toronto city in NorthYork and in the central part of Scarborough district are the least expensive ones
- Based on data visualization, Neighbourhoods in Mid Town & Down Town Toronto including Waterfront Communities-The Island, Niagara, Kensington-ChinaTown, Trinity Bellwoods, Palmerston-Little Italy, University, Baystreet Corridor, Church-Young Corridor, Moss Park, Annex, Dovercourt-Wallace Emerson-Juncti leads with maximum property prices, venue counts & unique category venue with respect to all property types
- With respect to modelling including Linear, Polynomial and KNN regression shows similar results in terms of regression coefficient and mean squared error, with exceptions of bedroom case where the all property types and detached case shows improved regression coefficient and mean squared error when KNN regression is applied on the data
- Since regression techniques including linear, polynomial and knn regression didn't show any association of venue category data with neighbourhood average property price. In this regard, random forest regression with feature importance was performed which improved regression coefficient and showed feature importance with respect to average property price of neighbourhoods
- Based on feature importance results; cafes, dessert shops & fast food restaurants are considered as top three venues that have major impact on average property price of neighbourhoods

So, in a nutshell, locals moving to Toronto from other provinces of Canada and international immigrants planning to settle in Toronto should consider above mentioned factors as qualifiers for selection of an appropriate place for living