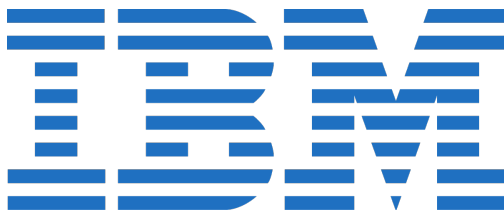# Real Estate Investment Opportunities in the GTA

Capstone Report

Jonathan Densil

IBM Data Science Professional Certificate

Coursera Inc.

August 2, 2020

Contents

1 Introduction

The Greater Toronto Area (GTA) is the most populous metropolitan area housing around 20% of the population of Canada. Serving as the anchor of the GTA, the second-largest financial sector in North America. Toronto is the 4th largest city in North America and Canada's business and financial capital. The GTA also has nearly 100,000 new immigrants annually and 130 million people within a 500 mile radius; such a large flux of people would provide ample real estate investment opportunities (https://torontoglobal.ca/Discover-Toronto-region/Toronto-region-quick-facts).

Figure 1: Municipalities in the GTA

2 Business Problem

As the old adage goes: location, location, location. Using machine learning and artificial intelligence to determine the best locations to buy a home or commercial real estate will be a key tool for realtors in advising clients of where to invest. Developing this kind of model would also be of great value to immigrants and internal migrants seeking to buy a home in an new and unknown location, giving them a sense for the community before committing to a life-changing decision.

3 Data

In order to create the dataset needed to train the machine learning algorithm, we will need the following data:

- Real estate sales in Ontario to determine a range of affordability options

- Popular venues and attractions that will potentially raise the house prices and attract traffic

- Geolocation data to visualise the potential areas of interest

Housing sales data was taken from a kaggle dataset (https://www.kaggle.com/mnabaee/ontarioproperties) since the Canadian Real Estate Association (CREA) prohibits the publication of any analyses, graphs, charts, results, conclusions or forecasts of the MLS® House Price Index (HPI) data without the prior written consent of CREA. Geolocation location was taken from the Atlas of Canada and popular venue data was queried using the Foursquare API.

4 Methodology

5 Results

6 Discussion

7 Conclusion

8 References

9 Appendix