



# FIND ME A HOUSE

An Analytics approach to segment Toronto's  
neighbourhoods based on common venues

## Abstract

When a person migrates to a new place, it becomes really essential for them to find a house. However, their decision to rent/buy a house is also based on the nearby venues to the house. Using Data Science and Machine Learning, we will group different neighbourhoods of Toronto into different segments and provide suggestions on which segment could be better for users.

Nitish Bhardwaj

## Contents

1. Business Problem .....	2
2. Data .....	2
References .....	3

## 1. Business Problem

---

Whenever a person migrates to a new place a person instantly starts the search for an ideal house. It is generally noticed that a decision to buy/rent a house is highly supported by the location [1][2] or nearby venues of the house. It becomes difficult for a user to find nearby venues to the houses. They end up doing their research either by visiting the location or by visiting various websites.

Toronto is a highly populated city in the Ontario province of Canada. It is also considered as the Silicon Valley of Canada [3]. Every month a lot of people do migrate to Toronto in search of better opportunities and a world-class lifestyle. This project addresses the business problem of the users who are looking for a house in Toronto. The solution will suggest the common venues near the neighbourhoods in the Toronto area.

## 2. Data

---

The Toronto postal code data for this project is fetched from a Wikipedia page [4]. The Wikipedia page [4] is having the names of the Boroughs and Neighbourhoods based on each postal code of the Toronto region.

The neighbourhood data for each postal code fetched from Wikipedia is then joined with a local copy of the Toronto postal code data file having the geolocation details for each postal code.

Once we have a single dataset having the geolocation details of each neighbourhood in Toronto, the details of nearby venues are fetched for each location using Foursquare API [5].

The data is then cleaned and processed to have the final dataset ready to feed for the clustering algorithm.

## References

---

- [1] "10 Important Features to Consider When Buying a House | HOMEiA.com", *HOMEiA*, 2020. [Online]. Available: <https://homeia.com/10-important-features-to-consider-when-buying-a-house/>. [Accessed: 30-Apr- 2020].
- [2] "The 5 Factors of a 'Good' Location", *Investopedia*, 2020. [Online]. Available: <https://www.investopedia.com/financial-edge/0410/the-5-factors-of-a-good-location.aspx>. [Accessed: 30-Apr- 2020].
- [3] V. Monga, "Silicon Valley Looks North as Tech Giants Expand in Toronto", *WSJ*, 2020. [Online]. Available: <https://www.wsj.com/articles/silicon-valley-looks-north-as-tech-giants-expand-in-toronto-11566054001>. [Accessed: 30- Apr- 2020].
- [4] "List of postal codes of Canada: M", *En.wikipedia.org*, 2020. [Online]. Available: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). [Accessed: 30- Apr- 2020].
- [5] "Endpoints | Places API", *Developer.foursquare.com*, 2020. [Online]. Available: <https://developer.foursquare.com/docs/places-api/endpoints/>. [Accessed: 30- Apr- 2020].