

1. Data acquisition and cleaning

1.1 Data sources

The average house rental prices are found in the Canada Mortgage and Housing Corporation dataset [here](#). The dataset found from the Canada Mortgage and Housing Corporation provided average house leasing price for every neighbourhoods in Canada, however, the dataset does not provide the coordinates for each neighbourhoods and the neighbourhoods are named with different complexity, which it caused the python library Geocoder unable to accurately search the neighbourhood's coordinates based on neighbourhood's name. To fix this issue, I manually searched the coordinates for each Greater Toronto Area neighbourhoods by using Google Map and created a new dataset based on the one founded in the Canada Mortgage and Housing Corporation website. The new dataset can be found [here](#) where I uploaded to my Github repository. The final venue category data are acquired by using Foursquare API to explore the first 100 venues around neighbourhoods within a range of 1000 meters.

1.2 Data Cleaning

The Data downloaded from the Canada Mortgage and Housing Corporation were included with rows of data descriptions and corporate information, and thus these rows were removed from the dataset. Furthermore, the whole data were filled with both English and French, but only the English data were needed, therefore removed the French parts from the dataset.

After that, I changed the name of columns such as "Total" to "Average Price", the names of province "Ont." to "Ontario", the Zones like "Etobicoke (South)" to "Etobicoke" that are inside the Greater Toronto Area and split the Neighbourhoods with a slash "/" that has multiple neighbourhoods in the same row and shares the same average rental price, then created new rows for each split neighbourhoods in order to support accurate venue search and consistent data structure.

1.3 Feature Selection

From examining each feature, there were some redundancy in the features. The redundant features are "Dwelling Type" and "Bedroom Sizes", there are no specific rules on selecting

building types for office as people may select any type of buildings. Since this project is only aiming at the desirable office locations but not building types, if trying to specify every single dwelling types, bedroom or room sizes will cause the result harder to acquire and create unnecessary confusions, therefore I only focused on “Total” for all the dwelling types and “Total” for all the bedroom sizes to get an overall average house rental price for the neighbourhoods in Greater Toronto Area, then dropped the rest of the dwelling types, bedroom sizes and neighbourhoods that are not inside Greater Toronto Area. Eventually there are 5 features and 155 Neighbourhoods were selected.

Kept features	Dropped features	Reason for dropping features
Dwelling Type: Total	Dwelling Type: Row, Apartment & Other	Focusing on overall average house rental price for all the building types to better interpret the result.
Average Rental Price: Total	Average Rental Price: Bachelor Studios, 1 Bedroom, 2 Bedroom, 3 Bedroom	Focusing on overall average house rental price for all the bedroom or room sizes to better interpret the result.
Province, Zone: Ontario, Greater Toronto Area	Province, Zone: Other Provinces in Canada	This project only interested in searching the desirable office setup locations in Greater Toronto Area.