Ruslan Ibragimov
Data Science Applied Capstone - Coursera
IBM Professional Certificate

Research on effects of different types on venues on the price of property in the neighborhoods of Toronto

10.09.2019

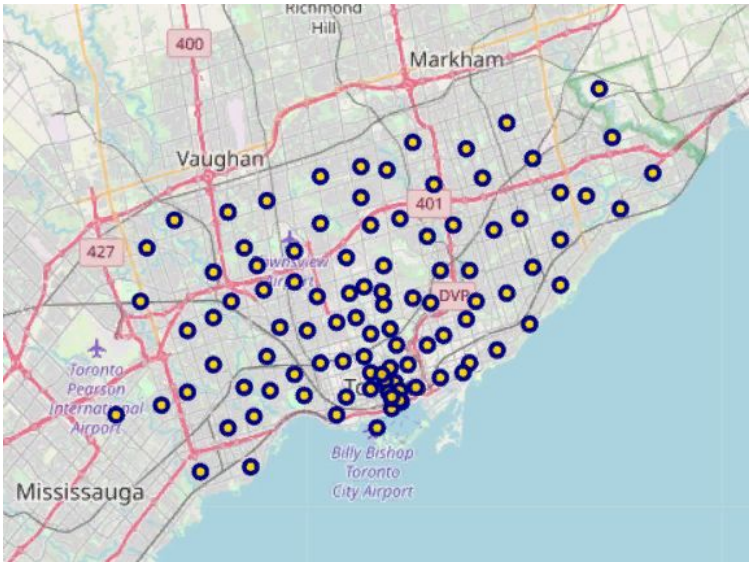

fig-1. Neighborhoods of Toronto.

I. Introduction and Business Understanding

In this paper we will cover the analysis of impact of different venues in area on the price of private property in different neighborhoods of Toronto (Canada). As it is commonly known, the price of housing in specific area is strongly dependent on the number of different venues, namely shops, restaurants and et criteria. The variation and the accessibility of them can be one of the first things to consider for customer during real estate purchase.

For better management and market research, it would be useful to know what kinds of venues influence the price more, than the others, therefore it explains the importance of this research. To do it we will use open datasets (**Data**) and different methods (**Methodology**). The main target audience would be city service workers, businessmen and people interested in buying or sell their property.

II. Data

We are going to need the next two datasets:

1. Dataset provided by Coursera, that includes the longitude and latitude of neighborhoods, the names of neighborhoods and their postal codes. (http://cocl.us/Geospatial_data)
2. Dataset provided by city of Toronto administration on house prices in neighborhoods. (https://open.toronto.ca/dataset/wellbeing-toronto-economics)

At the end we need to get the dataset that has the next format:

Name of neighborhood	Venues encoded by one hot encoding	Price

Features for evaluation will be venues encoded by one hot encoding, while labels will be present as price.

Features will give us understanding about what kind of venues is present in area of neighborhood, so we can connect it with the price.

The dataframe above is going to be created using above mentioned two sets:

1. We will use **Foursquare API** in order to get information of venues in neighborhoods, using longitude and latitude from the table.
2. We will export housing price from the second data set and merge it with the frame from the first step.
3. Then we will drop unused columns (‘Borough’ or ‘Post code’) and finish wrangling of dataset.