IBM Data Science Capstone Project

(IBM Data Science Professional Certificate, Coursera)

**Rent Prices and Social Environment Analysis
of Vancouver Neighbourhoods**

Prepared by: Beibyt Azymbek

February 2020

Table of Contents

1. Introduction

1.1 Background

Vancouver is the largest city in the province of British Columbia, and the eighth largest municipality in Canada. The Greater Vancouver metropolitan area (which includes neighbouring cities such as Burnaby, Richmond, and Surrey) is the third largest in Canada. The City is famous for its beautiful surrounding nature and vibrant lifestyle it provides for the habitants [1]. With its scenic views, mild climate, and friendly people, Vancouver is known around the world as both a popular tourist attraction and one of the best places to live. Vancouver has now also become one of the fastest growing cities in the world by attracting large populations of internal and international migrants. According to the World Population Review, population of Greater Vancouver metropolitan area has risen from 1.96 million in 1990 to 2.58 million in 2020 [2]. Due to its geography, being surrounded by the Pacific Ocean from West and Rocky Mountains on East, Vancouver has limited space for its expansion and being able to accommodate its growing population, which reasonably leads to high housing and living expenses. With the size of only 114 square kilometers, City of Vancouver has now become one of most expensive to live cities in Canada and in the world, especially in terms of accommodation and real estate. According to the 15th Annual Demographia International Housing Affordability Survey, Vancouver has been listed as 2nd least affordable city in the world as of 2019 [3].

1.2 Problem

Above stated issue has now become a major concern for not only current habitants of the area but prospective migrants as well, who seek affordable rental living when considering relocation. Usually a lot of time and effort is spent on deciding what neighbourhood in the City to settle in. Work conducted in this paper is a data-driven analytical approach of comparing different neighbourhoods of Vancouver in terms of their rental prices and social environments. As part of the analysis, this work aims to identify what factors, specifically what types of venues and their presence in a neighbourhood affect rental prices.

1.3 Interest

The project will be interesting for people who plan renting accommodation in Vancouver or generally consider moving to the City. By providing analytics and comparison on the neighbourhoods, results of this work would assist in finding optimal neighbourhood in accordance with the needs and budget of those who plan to live there.

2. Data

2.1 Data Sources and Description

The following data sources and methods on obtaining them were used within this work:
- Vancouver Rental Market Statistics Summary provided by Zones from Canada Mortgage and Housing Corporation [4]. This is a database on housing and mortgage information

managed by a state corporation, which provides information in interactive as well as in textual format (csv). Within this work, I have downloaded csv file for Vancouver housing market data specifically for 1-bedroom residences, which is one of the most common residence sizes rented in highly priced cities such as Vancouver. There is obviously data available for larger (2+ bedroom) residences which can easily be obtained and similar analysis could be conducted on. This source provided both list of neighbourhoods of Vancouver as well as average rental price for 1-bedroom residences within those neighbourhoods.

- Geocoder Python package () was used to obtain geographical coordinates of the neighbourhoods. This information would allow us to pass coordinates as parameters when creating neighbourhood maps and collecting information on venues within those neighbourhoods.

- Foursquare Application Programming Interface (API) was used to gather information on venues, categories of venues, their location, quantity and quality (based on user reviews) within each of the neighbourhoods of Vancouver [5]. This information was used to obtain and classify social environmental background of the neighbourhoods and identify potential factors which might directly impact on the rent prices in that neighbourhood.

2.2 Data Cleansing and Preparation

Data on rental market (csv file) had too many unnecessary details which had to be cleansed before proceeding with analysis. For example, vacancy and availability rates, information on number of available units was not relevant within our subject of work and had to be removed from for further processing. Also, I have simplified the data at hand to cover only those related to 1-bedroom residences as usually they are enough to provide indicators on the associated neighbourhood prices as any other larger residence types, i.e. if renting 1-bedroom apartment in neighbourhood A would cost higher than renting same 1-bedroom apartment in neighbourhood B, high chances are the same will be true for 2 and 3-bedroom apartments as well.

Moreover, it was necessary to keep relevant formatting of data to be able to cluster the neighbourhoods. Some column names or names of the Vancouver neighbourhoods had unnecessary characters, such as commas or brackets, which would cause false results of classification algorithms. Further, as long as new data was gathered associated with venues, I had to merge all information into a single dataset in order to proceed with running analysis on it. More detailed

3. References:

[1] https://vancouver.ca/about-vancouver.aspx
[2] http://worldpopulationreview.com/world-cities/vancouver-population/
[3] http://demographia.com/media_rls_2019.pdf
[4] https://www.cmhc-schl.gc.ca/en/data-and-research/data-tables
[5] https://developer.foursquare.com/
[6] https://geocoder.readthedocs.io/index.html