Recommendation for Business Opportunities in Toronto

1. Introduction

1.1. Background

Toronto is one of the most multicultural cities in the world. Its population consists of many visible minorities. It is easy to recognize the trend, that Toronto is an attractive city for many foreigners around the world, not only for touristic purpose, but also for living, working or education. In 2016, 51,5% of the residents of the city belonged to a visible minority group, with a total population of 2,7 million residents. The population grew by 4,3% from 2011 to 2016 (Source: https://en.wikipedia.org/wiki/Demographics_of_Toronto)

In this capstone project, it will be therefore assumed that a new investment in Toronto is an attractive opportunity due to growing population. Since we are working with Foursquare data, we will discuss the investments in terms of new venues in the city.

Due to new development in the city, we want to especially focus the areas with a high increase of population in the last years. As an example, we will focus on new investment opportunities in the neighborhood "Willowdale East" in the scope of this project, because this neighborhood had one of highest population increase in the last years, due to the demographic statistics in Wikipedia (62,3% change in population since 2001, https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods).

Our approach in this project can be also easily applied to other neighborhoods. But we will restrict our recommendation only to "Willowdale East" for simplicity.

1.2. Problem

As the neighborhood "Willowdale East" is identified as an area which may provide opportunities for the investors or entrepreneurs, the next question is, which venue category will promise the highest potential for them. Therefore, this project aims to provide recommendations in terms of venue categories. Of course, these recommendations must be evaluated by further criteria (e.g. average income of the area, return on investment, etc.), but we aim to provide recommendations as a starting point for further discussions.

1.3. Approach

To answer the question described in the last section, we will use collaborative filtering approach, introduced in the week 5 of the machine learning module in the IBM Data Science Course.

This means, we will search for similar neighborhoods based on the existing venue categories and their frequency (number of venues in each category and neighborhood). Then, we will determine the venue categories which are not existing in Willowsdale East yet, but in the most similar neighborhoods. At the last step, we will also exclude the categories, which already exist in the closest neighborhoods to Willowsdale East.

As result, a list of venue categories will be provided as recommendation, which does not exist in Willowsdale East, and also does not exist in the closest neighborhoods of Willowsdale East.