Capstone Project – The Battle of the Neighborhoods

Business Problem and Background

Applied Data Science Capstone by IBM/Coursera

Author: Maia Ludmila Budziñski

Date: July 01, 2020

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. Toronto is the fastest growing city in North America. and is the anchor of an urban agglomeration, known as the Golden Horseshoe in Southern Ontario, located on the northwestern shore of Lake Ontario.

Toronto encompasses a geographical area formerly administered by many separate municipalities. These municipalities have each developed a distinct history and identity over the years, and their names remain in common use among Torontonians. Former municipalities include East York, Etobicoke, Forest Hill, Mimico, North York, Parkdale, Scarborough, Swansea, Weston and York. Throughout the city there exist hundreds of small neighborhoods and some larger neighborhoods covering a few square kilometers.

Having such vast population and big geographical area, there also exists a big competition between different types of businesses. Therefore it became very challenging for stakeholder or new business to decide in which area should their start a business to get higher revenue with lowest possible competition.

In this project we will try to find an optimal location for a restaurant and identify the ideal type of food it should serve keeping in mind the competitors. Specifically, this report will be targeted to stakeholders interested in opening a restaurant in Toronto, Canada.

Since there are lots of restaurants in Toronto, we will try to detect locations that are not already too crowded with restaurants but also show a prosper territory for this kind of business.

By using data science methods and tools along with machine learning algorithms such as clustering we will identify the most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.