

# Capstone Project

*Opening a New Pub in Richmond Hill, Ontario*



August 8, 2020

## **Introduction**

The history of pubs can be traced to Roman taverns in Britain in 43 AD (“Historic UK”, n.d., para 3). It means pubs have been existed for almost 2,000 years. What reasons that make pubs last so long? Tribute Cornish Pale Ale conducted a poll about reasons for visiting a pub, and the results showed that catching up with friends, the atmospheres and the opportunity to have a few drinks were the top three reasons (“Beertoday”, 2017). For many peoples, going to a pub is a way to relax and enjoy themselves after work, at weekends or during holidays. On the other hand, the profitability of pubs has strong correlation with sports events (“Financial Time”, n.d.). For example, large number of residents in the neighbourhoods is attracted to nearby pubs on game nights. There are hundreds of pubs in Richmond hill. Therefore, selecting location of a pub is crucial for the pub owner’s investment decision.

## **Business problem**

The main purpose of this project is to find an optimal location for a new pub in Richmond hill, Ontario by using data science methodology, specifically machine learning techniques.

## Data Section

- ***The following data is what we need for this project:***

1. List of neighbourhoods including postal code in Richmond hill, Ontario.
2. Latitude and longitude coordinates of the neighbourhoods.
3. venue data, particularly data related to pubs on the neighbourhoods.

- ***Sources of the data and methods for manipulating***

Richmond hill is a city in south-central York Region, Ontario, Canada. The city contains four major neighbourhoods: Richmond hill (Southeast), Richmond hill (Southwest), Richmond hill (Oak Ridge/Lake Wilcox/Temperanceville) and Richmond hill (Central). The webpage [https://en.wikipedia.org/wiki/Richmond\\_Hill,\\_Ontario#Communities](https://en.wikipedia.org/wiki/Richmond_Hill,_Ontario#Communities) includes a list of neighbourhoods in Richmond hill.

Since the number of neighbourhoods is small, so we can just build a dataframe using pandas of Python data analysis library. Then, we will get the geographical coordinates of the neighbourhoods using Python Geocoder package. Next, we will use Foursquare API to obtain the venue data for the neighbourhoods. Foursquare API will provide a lot of categories of the venue data. However, we are interested in the pub category data which can help us to solve the business problem as mentioned above. Finally, we will clean the dataset, apply machine learning skill K- means clustering and map visualization by using Folium package.