2020

# Safest Neighborhood in Toronto for opening a commercial establishment

CAPSTONE PROJECT

FER VÁZQUEZ

## 1. INTRODUCCION:

Toronto is a great place to live, the shopping is great, thousands of restaurants and cafes to get a fantastic meal, and there are lots of things you can do at any hour from strolling through parks, catching a movie or concert, or watching some live sports. But opening a business in Toronto isn't always so good, especially if you ask about crime. Fortunately, if you want to open your business in Toronto, with this project we will be looking to understand the crime, and which will be the best neighborhood to open your own business.

## 2. BUSINESS PROBLEM:

The purpose of this project is to understand which neighborhood will be the best to open a commercial business in Toronto and which type of commercial business. The first task will be to understand which neighborhood is the safest by analyzing the crime data and the second task will be to analyze the 10 most common venue in this neighborhood. We will use our knowledge of Data Science to do this analysis.

## 3. DATA SOURCE

The data of crimes I will use the real data that it is published in Kaggle dataset for this page: https://www.kaggle.com/kapastor/toronto-police-data-crime-rates-by-neighbourhood

In the next table I describe the columns and the transformation that I will apply for each column:

| Column | Description | Transformation |
|---|---|---|
| **X** | Latitude | Remove |
| **Y** | Longitude | Remove |
| **Index_** | Unique ID | I will use as unique id |
| **event_unique_id** | Event ID | Remove |
| **occurrencedate** | Date of crime occurred | Remove |
| **reporteddate** | Date of crime reported | Remove |
| **premisetype** | Location of crime occurred (commercial, house, apartment, outside, other) | I will use to filter the premise only with commercial and outside types. |
| **ucr_code** | Code | Remove |
| **ucr_ext** | Ext | Remove |
| **offence** | Crime description | Remove |
| **reportedyear** | Year of the report | Remove |
| **reportedmonth** | Month of the report | Remove |
| **reportedday** | Day of the report | Remove |

| | | |
|---|---|---|
| **reporteddayofyear** | Year day of the report | Remove |
| **reporteddayofweek** | Week day of the report | Remove |
| **reportedhour** | Hour of the report | Remove |
| **occurrenceyear** | Year of the crime occurred | Remove |
| **occurrencemonth** | Month of crime occurred | I will use to known which month has more crimes |
| **occurrenceday** | Day of crime occurred | Remove |
| **occurrencedayofyear** | Year Day of crime occurred | Remove |
| **occurrencedayofweek** | Day of week of crime occurred | I will use to known which day of week has more crimes |
| **occurrencehour** | Hour of crime occurred | Remove |
| **MCI** | Type of crime | I will use to know the type of crime |
| **Division** | Division | Remove |
| **Hood_ID** | Neighborhood Id | Remove |
| **Neighbourhood** | Neighborhood | I will use to know the name of the Neighborhood |
| **Long** | Longitude | I will use to create the map |
| **Lat** | Latitude | I will use to create the map |
| **ObjectId** | Object ID | Remove |

For data of Toronto Neighborhoods, I will use the WikIpedia source:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This list, I will clean it to have the next dataframe:

- PostCode
- Borough
- Neighborhood

Then, I will use another dataset to get the Latitude and Longitude of each neighborhoods, the final dataframe will be:

- PostCode
- Borough
- Neighborhood
- Latitude
- Longitude

And Finally, I will use the Foursquare location data to know the 10 most common venue in the safest neighborhood.