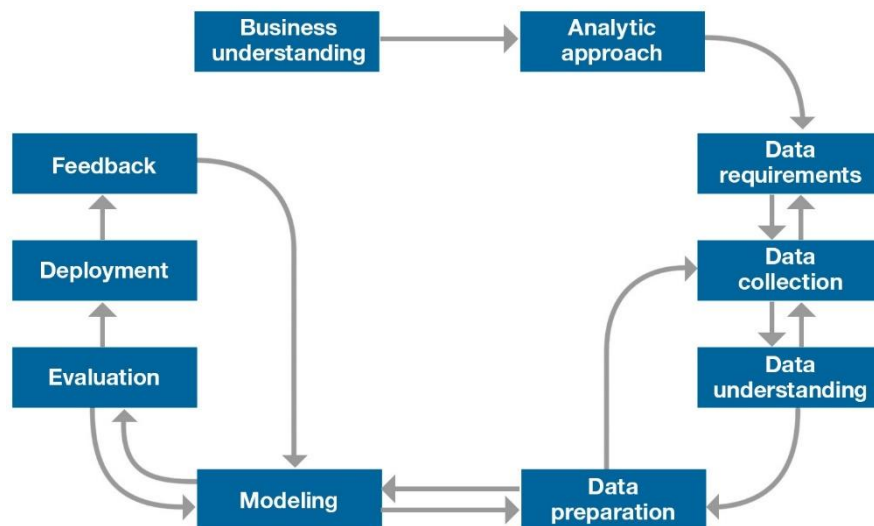# Battle of the Neighbourhoods
Final Report
Jon Tinsley

Introduction
This report will employ IBM's Data Science Methodology, displayed in the graphic above. The structure of the report will mirror that of the sections outlined above.

From problem to approach
Business Understanding

As a business, we help clients find new places to live based on their specific parameters and requirements.

A new client is moving to Scarborough, Toronto, Canada and has requested a report of which neighbourhoods have the least crime, as well as which of these neighbourhoods are seeing a general

decreasing trend in the amount of crime. He has recently had two children and wants to raise them in a safe area while also saving money so he can send them to a good school.

The problem: How can we find the neighbourhoods with the least crime, and where the crime trend is decreasing? How will data science be used to solve this problem?

Analytic Approach

Having gathered that the client would like to know areas with the lowest amount of crime, the analytic approach that will be taken to solve this problem is as follows.

Data Requirements

This project will require Toronto crime data from the last 8 years, in .csv format so it can be read by Python. Additionally, a geojson file of Toronto Neighbourhoods will be required in order to create a choropleth map data visualisation of which neighbourhoods have the least total crime over the last few years.
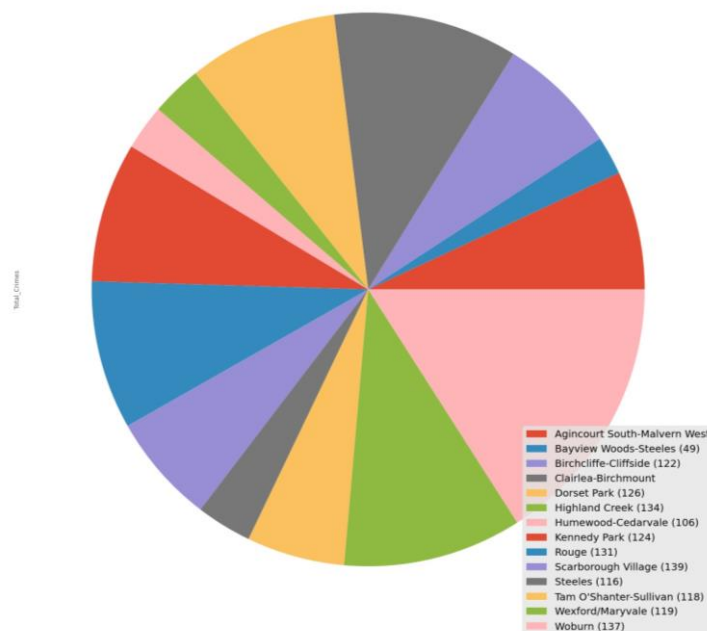
Data Collection

Crime data will be sourced from [Neighbourhood Crime Rates 2020 (torontopolice.on.ca)](torontopolice.on.ca). Geojson data will be sourced from [toronto-geojson/toronto_crs84.geojson at master · jasonicarter/toronto-geojson (github.com)](github.com). Information on postcodes and the neighbourhood of Scarborough in Canada was sourced via Foursquare as well as by web scraping of the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M using the BeautifulSoup module of Python. Latitude and longitude data was gathered from https://cocl.us/Geospatial_data and read using the read_csv function of pandas.

Methdology

Data Understanding and Exploratory Analysis

In order to understand the data, a pie chart and a bar graph were used to see the distributions of total crimes (**fig.1**). While pie charts are not usually the best representations of data, it helped to identify the areas with the most and least crime, which then allowed for further exploratory data analysis.
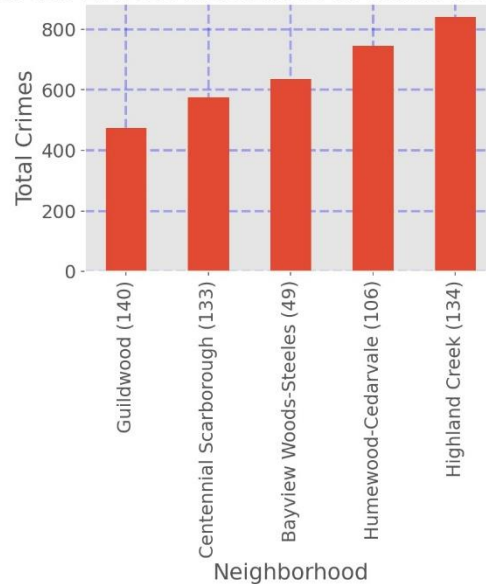

Fig.1: *Pie chart of total number of crimes committed per neighbourhood*

The next step was to produce a bar graph of the 5 areas with the least amount of crime. This was achieved using the Python package 'Pandas'. This showed that Guildwood, Centennial Scarborough,

Bayview Woods-Steeles, Humewood-Cedarvale and Highland Creek were the neighbourhoods with the lowest amounts of crime (**fig.2**).

Fig.2: *Top 5 areas with the lowest amount of crime in Scarborough*

Data Preparation

The data was pre-processed and cleaned using the 'Pandas' package in order to make the data suitable for comparison and analysis. Neighbourhood postcodes and coordinates were identified for the Scarborough neighbourhood names using BeautifulSoup, which allowed for Scarborough. Each dataframe was joined together after which a new dataframe was made that contained all of the necessary data for the main analysis. Neighbourhoods of Scarborough were plotted to a map to visualise their location and proximity to each other (**fig.3**).
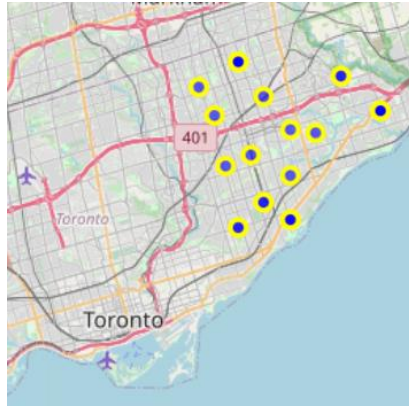
Fig.3: *Neighbourhoods of Scarborough plotted onto a map*

## Results: K-Means Model with Choropleth Map of Total Crime
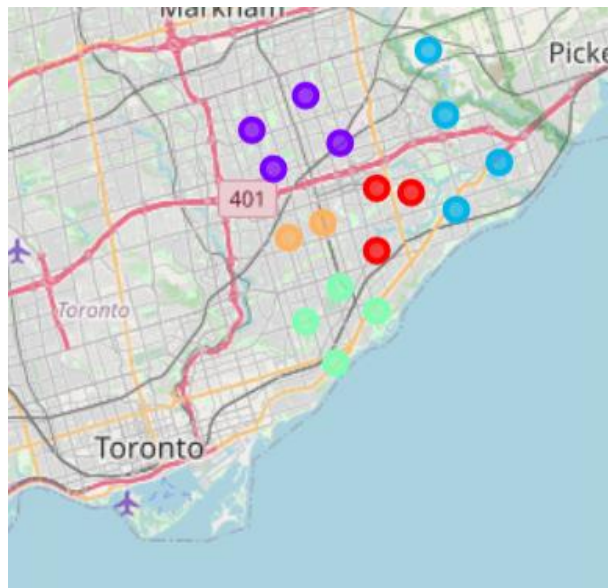
Fig.4: *k-Means map of the neighbourhoods in Scarborough*

In order to decide which neighbourhoods of Scarborough should be grouped together as an area, a k-Means clustering model was produced in Python, as shown in **Fig.4**.

Choropleth Map of Total Crime per Area in Scarborough

The final step of the analysis was to show how the identified areas overlapped with crime levels in Scarborough. A choropleth map was created of total crimes committed since 2014, and the k-Means clusters were overlain to visualize the distribution of crime per area.
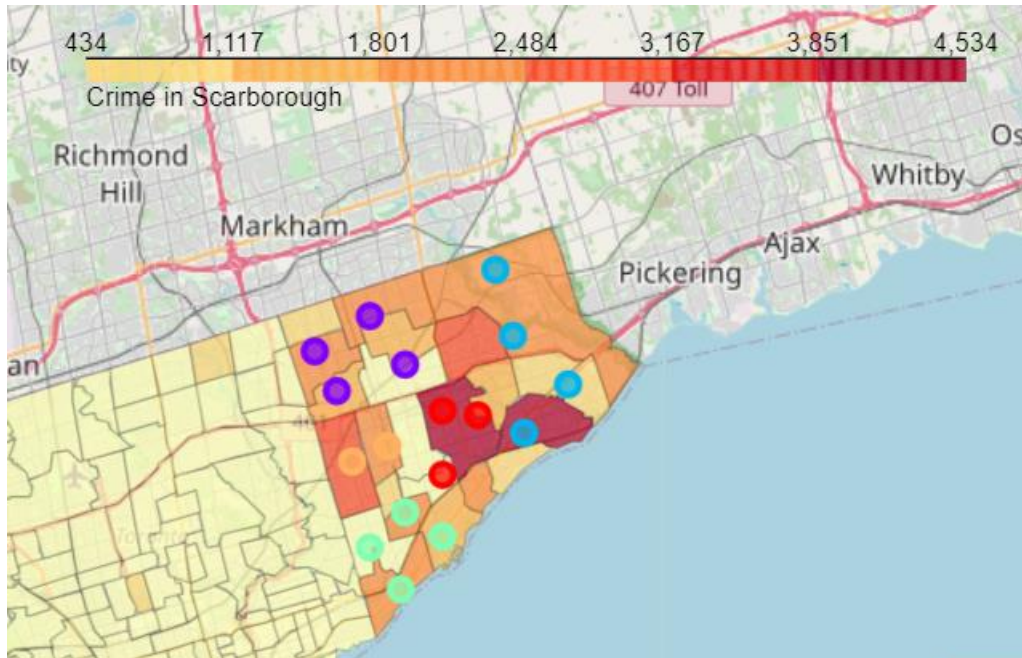
Fig.5: Choropleth map with k-Means clustering algorithm overlain, showing how neighbourhoods are distributed across the crime levels of Scarborough

Discussion

The results shown above in **Fig.5** and **Fig.2** show that Guildwood, Centennial Scarborough, Bayview Woods-Steeles, Humewood-Cedarvale and Highland Creek were the neighbourhoods with the lowest amounts of crime. It is therefore recommended that the family move to one of these 5 areas.

Model Deployment and Feedback

As the model is yet to be deployed to the client, obtaining feedback on its deployment and functionality is beyond the current remit of this report.

Conclusion

This report shows how Python was used to identify areas of Scarborough that had the lowest areas of crime, thus recommending the family to move into one of Guildwood, Centennial Scarborough, Bayview Woods-Steeles, Humewood-Cedarvale or Highland Creek.