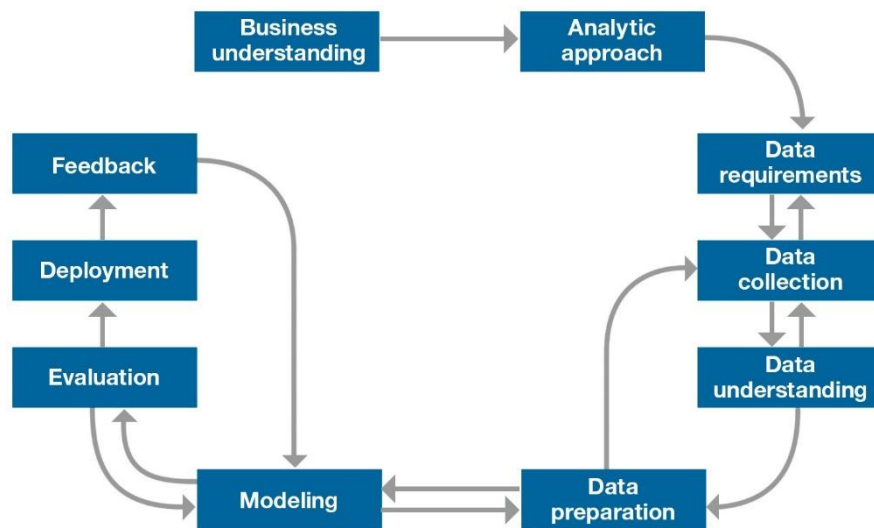# Battle of the Neighbourhoods
Initial Report
Jon Tinsley

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)](#).
- Geojson data will be sourced from [toronto-geojson/toronto_crs84.geojson at master · jasonicarter/toronto-geojson (github.com)](#).
- Information on the neighbourhoods will be gathered from FourSquare.
- Postcodes in Canada was sourced from the following Wikipedia page via webscraping using the BeautifulSoup module of Python: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Latitude and longitude data will be gathered from https://cocl.us/Geospatial_data and read using the read_csv function of pandas.

Data Description

In order to solve this problem, **crime data** will be needed. This will be sourced as described above. This data contains neighbourhood information and the different types of crimes and the frequency they are committed.

However, this information will need to be complemented **by data on the neighbourhood.** FourSquare will provide the neighbourhood data, while Wikipedia will provide the postcodes.

In order to plot the neighbourhoods and the crimes to the map, **coordinates of the Scarborough neighbourhoods** are required. This will be gathered as described above.

To generate a map of Scarborough and the neighbourhoods within it, a GEOJSON file is needed containing geospatial data (i.e. the polygons of the individual neighbourhoods and the coordinates of these polygons). This will be gathered as described above.

In concert, all of these data harnesses what is required to plot, model and visualise the areas of Scarborough with the least and most amounts of crime, so as to inform families which area they should move to.