Graham Clinning

Capstone assignment

Part 1

1. Introduction
   1. Background

Traveling to places around the world is something which many look forward to doing which provides an opportunity to see the world. It allows one to explore different cultures, food, people, venues and so much more. Traveling can be daunting if you have never been to that country, borough or neighbourhood and do not know which the top venues are to visit on your trip to make it worthwhile.

People who travel overseas usually do research beforehand to see where they want to go, the neighbourhoods they want to explore and the top venues they want to visit.

Afterall, you want to make the most of your trip and visit the top venues in each neighbourhood.

* 1. Problem/idea

People visiting The Bronx in New York and York in Canada, who have never been to these boroughs before, do not know the neighbourhoods or top venues to visit. Even people who have visited these neighbourhoods before may not be aware of the top venues so they may have been missing out and want to return and visit the top venues.

The idea is to identify the neighbourhoods in the Bronx and York, to identify the top 5 venues per neighbourhood and to show them on a map so that people can see where they are located.

* 1. Stakeholders

The stakeholders are anyone who is visiting the Bronx or York for the first time, people who want to know what the top venues are in each neighbourhood and people who want to know what the neighbourhoods are in both boroughs.

* 1. Interest

As mentioned in the above sections, traveling can be daunting when you are not familiar with the destination, the neighbourhoods or venues you want to visit.

People who require additional information regarding all of these factors will be interested in the insights generated from this project which will better help them understand and get to know each neighbourhood.

* 1. Conclusion

This project will help shed insight on the Bronx and York in terms of their neighbourhoods, top venues and map locations for anyone who require this information, whether visiting them for the first time, wanting to visit the top venues, wanting to see the where the locations are on a map or, just have a general interest in these locations and want to know more about them

1. Data

2.1 Data sources

2.1.1 The Bronx

The data source for the Bronx will come from a data set of New York which include the Borough Bronx and the relevant neighbourhoods. This data source also includes the longitude and latitude co-ordinates for boroughs and neighbourhoods to be able to plot their locations on a map. <https://geo.nyu.edu/catalog/nyu_2451_34572>

2.1.2 York

The data source for York will come from a Wikipedia page on Canada which includes York as one of the boroughs and the relevant neighbourhoods as well. A second data source (Excel sheet) with the longitude and a latitude co-ordinates for the borough and neighbourhoods will also be used so that the locations can be plotted on the map. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

2.2 Data cleaning

2.2.1 The Bronx

The data for New York will be converted into a Pandas data frame and then cleaned so that it can be easily worked with and used for the relevant data analysis. The cleaning will remove all unnecessary data and ensure that only the relevant data is used in the data frame to be able to perform the required analysis of said data.

2.2.2 York

The data from the Wikipedia source needs to be scrapped so that it can be imported into the notebook and then converted into the data frame. The data frame will then be cleaned to remove any unnecessary data and where rows are not assigned, convert the neighbourhood to borough.

The second data source for the longitude and latitude co-ordinates must be converted into a data frame and then combined with the first data frame so have one data frame with all the relevant data do allow for the data analysis to take place.

Data cleaning must be done to get to a data frame which includes all relevant data and can be easily used for data analysis.

2.3 Feature selection

2.3.1 The Bronx

A new data frame will be created which will only include the borough of the Bronx, the neighbourhoods and the longitude and latitude co-ordinates. This new data frame will be used for the analysis.

2.3.2 York

A new data frame will be created which will only include the borough of York, the neighbourhoods and the longitude and latitude co-ordinates. This new data frame will be used for the analysis.

2.4 Foursquare

Foursquare location data will be used to get the top 5 venues for each neighbourhood in the Bronx and York. The data will be limited to 100 venues which will then be cute down accordingly.

This data will be used to create a new data frame for each borough which will allow me to get the frequency per venue and then ultimately convert that into the final data frame to show the top 5 venues in order from first to fifth.

2.5 Folium map

A folium map of the Bronx and York will also be created to visually show the neighbourhoods location on a map so that people can see where they are located.

2.6 Conclusion

The data will be collected, cleaned and converted into the relevant data frames for New York and Canada.

New data frames will then be created specific to Bronx and York so that analysis may take place.

Foursquare location data will be used to get the top venues for the Bronx and York which will then allow new data frames to be created which list each neighbourhood and their top venues for first to fifth.

Finally, a folium map for the Bronx and York will be create to visualize the data and show where they are located.