**The Battle of Neighbourhoods (Week 1)**

**Capstone Coursework Submission – Innes McLaren**

**2. Sources of Data & how data sources will be used to solve the problem.**

**2.1 Wikipedia Page for Manchester**

The Wikipedia Manchester web page was used to source the Greater Manchester area. Wikipedia Page is available unrestricted on the internet and is found at address

<https://en.wikipedia.org/wiki/M_postcode_area>

Example screenshot of the data is below

Graphical user interface, text, application, email

Description automatically generated

The Wikipedia data will be used to solve the problem of obtaining a comprehensive list of the Manchester Postal Codes which need to be considered for this report.

The Postal Code information will be scraped from the Wikipedia page using Beautiful Soup Python package. Beautiful Soup can extract data from XML and HTML data from web pages and convert into common Python data types for analysis. Beautiful Soup package is available on IBM Watson Studio Python utility.

**2.2 Geopy Nominatim Python webservice**

For this work, Geopy will be used to solve problem of how to get latitude and longitude co-ordinates for all of the Manchester Postal Code areas. Geopy Nominatim Python webservice is available on IBM Watson Studio Python utility

The webservice will be called using Python code for all of the Manchester Postal Codes and will return a set of latitude and longitude co-ordinates. See example below

Text

Description automatically generated with medium confidence

Latitude and longitude values are required to interact with Foursquare (see below). The Geopy

**2.3 Google maps**

Google maps is the online web mapping service developed by Google. Google Maps is available unrestricted on the internet.

For this work, Google maps was used as a backup source of latitude and longitude values for Manchester area Postal Codes. Google maps was used to provide a central latitude and longitude point for each of the Manchester area Postal Codes.

Map

Description automatically generated

The latitude and longitude values were stored in a CSV file which was loaded into Python and merged with the Postal Codes address information.

Below is a sample of the resulting CSV file is given below

A picture containing text, receipt

Description automatically generated

Latitude and longitude values are required to interact with Foursquare (see below)

**2.4 Foursquare Developer Console**

Foursquare company has an extensive database of locations frequently visited by general population. For this project, access to ‘Places by Foursquare’ webservice for Manchester will be used. A user account needed to access Foursquare Developer Console.

The Foursquare Developer Console will be used to solve the problem of identifying

* Postal Code Areas which have a large number of restaurants
* Types of restaurant per Postal Code Area

The picture below shows example of the venue data returned by Foursquare when the Webservice is provided with a latitude, longitude co-ordinate pairing

Table

Description automatically generated with low confidence