**Report – Manchester Restaurant Recommendation**

**Final Coursera Capstone Assignment**

**Innes McLaren**

**27/02/2021**

**1. Introduction - Description of Business Problem & Background**

**1. 1 Business Problem**

Our client is a Venture Capitalist who wants to invest in the catering industry in the United Kingdom (UK). The client wants to open a restaurant in the Manchester area of England, UK.

The client is looking to open restaurant in a popular “foodie” area of Manchester where there is already a significant presence of restaurants.

The client has made the following requests

* Use the Greater Manchester area of Britain will be target area for analysis
* Provide a recommendation on Type of restaurant (i.e., cuisine served)
* Provide a recommendation on the Area of Manchester where restaurant should be opened

Justifications for any recommendations must be made using Data Science techniques with evidence presented.

The client expects that, if they were to open a new restaurant , the chances of profitability would be maximised based on the information our report provides on the popularity of cuisine type and location.

**1.2 Discussion of Background**

Manchester is a major city in the UK which has a busy commercial centre. The expectation prior to running any analysis is that there will be a large number of restaurants in the greater Manchester area and a likelihood that many will be near or in the city centre. This assumption needs to be validated by data analysis.

The Manchester restaurant analysis will reveal which types of cuisine are most popular in the area as a whole. There is also a need to understand what cuisine types are most popular per Postal code area. This information will be a primary consideration when making recommendations to the client on which area to open their new restaurant.

It is thought that Clustering will be a good approach to identify common areas in terms of restaurant popularity.

**2. Sources of Data and 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 Postal Codes. 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

**Section 3. Methodology - Data Analysis & Recommendation Approach**

* The segmentation Manchester into component areas was done using Postal codes. Postal codes are the UK equivalent of US ZIP codes which break
* Investigation of reliable data sources for Manchester Postal Codes on the Internet was done. Wikipedia was judged to be the most suitable data source with a structured table available. Data Scraping will be used to extract the Postal Codes data into our Python Notebook using the Beautiful Soup package. Data cleansing was done to ensure final list of Postal Codes is suitable to be used in Analysis process.
* (NOTE: M3 listed twice in the list of Manchester Postal codes in Wikipedia due to the Postal Code being spread over 2 administrative regions. One value was removed to avoid duplications in analysis.)
* Once final list of Manchester Postal Codes was obtained these were converted into Latitude / Longitude co-ordinates for use with FourSquare (see later).
* Initially Geopy Python webservice was used to provide Latitude / Longitude co-ordinates but it was found that it was unreliable when returning Latitude and Longitude values for Manchester Postal Code areas. Multiple areas returned the same latitude and longitude values when, clearly, they should have been different. (The webservice did not return any errors to indicate there was defaulting happening but it was noticed when analysing the API data returns.
* As a result, Google Maps was used as a backup and provided accurate longitude and latitude measurements for the Manchester Postal Code areas.
* Folium was used to visualizing geospatial data of the Postal Codes to ensure that there was well segmented regional coverage for Manchester area
* Foursquare Venue API was called (using Latitude, Longitude co-ordinated) to return all venues within the radius of the Postal Codes area. The resulting list was then filtered to keep only the venues which were Restaurants
* The Restaurant Data was grouped by Postal Code area to find most popular “foodie” areas of Manchester. Python Dataframe visualisation was used to make the determination of which Postal Code area had the most restaurants.
* Data was grouped by Restaurant type to identify most popular forms of cuisine consumed in the Manchester area. Python Dataframe visualisation was used to make the determination of which Restaurant types were most common.
* It was necessary to understand which areas had common characteristics in terms restaurant type popularity. K means clustering was judged an appropriate approach to provide this picture. The K means clustering analysis provided a good view of restaurant popularity per Postal Code area.
* Folium was again used after K means clustering to get a visual picture of Postal Code clusters which had common restraint type prevalence.
* The clustering AI revealed clusters of areas where particular restaurant types are common and critically areas where particular restaurant types are not the most common. This provided the necessary data input to the decision making process for restaurant recommendation.
* The recommendation approach was then to recommend the opening of popular restaurant types in areas where the AI clustering has shown they were potentially underrepresented today.

**Section 4 Results**

Postal Code List for Manchester

**Table

Description automatically generated**

Postal Code List for Manchester – enhanced with Latitude & Longitude co-ordinates

Table

Description automatically generated

Postal Code coverage of Manchester – visualisation

Map

Description automatically generated

The Most popular areas in terms of restaurants per Postal Code area are shown in the graph below

Chart

Description automatically generated

Most common restaurant types in Manchester are shown in the graph below

Chart

Description automatically generated

K means Clustering of Postal code areas by common restaurant types.

Chart

Description automatically generated

The clustering colour coding is as below:

* **Indian / Vegan**
* **Eclectic area - Indian / Italian / Asian / Middle Eastern**
* **Chinese and Asian**
* **Fast Food & Vegan**
* **Fast Food / Mixed**
* **Italian and Ethiopian**

Postal code regions in the “eclectic” cluster which contains the busiest food areas

**Table

Description automatically generated**

**Section 5: Observations and client recommendations**

Postal code areas M2, M3 and M14 had the highest number of restaurant venues. M4, M13, M20, M21 and M50 are in the second tier of popularity.

Indian, Italian and Fast Food restaurants are the most common the in the Manchester area. The second tier of popularity includes Asian, Chinese, Middle Eastern and Vegan restaurants.

The K means clustering of restaurant data showed that the most popular areas for eating had the widest range of restaurant types. On the clustering diagram above M2, M3 and M14 all reside on the “eclectic” cluster.

Focusing on M2, M3 and M14, Indian restaurants were the most common overall, they were the most common in the busy M2, M3 areas but **not** in the M14 Postal Code regions where Middle Eastern Restaurants were most common.

From a coverage perspective, the busiest postal code areas, I conclude that Indian restaurants while Indian restaurants are well represented in the M2, M3 and M14 areas already. If the client is to open another Indian restaurant in one of these areas, then M14 is the least saturated in terms of current restaurant numbers.

**Recommendation:** Open an Indian restaurant the of the M14 region of Manchester.

**Notes on results**

The Postal Code areas of Manchester greatly vary in size and shape. I used a fixed 1000m radius parameter when using Foursquare Webservice. Since postal code areas vary in size, the1000m radius was a “best can do” value which was slightly large for some areas, a good fit for most areas and too small for others.

Some duplication was evident in the venues due to the 1000m radius pulling in the same venue for different Postal codes. This was not a problem as the restaurants were accessible from the area even if not strictly speaking inside the PC area.

Some of the Restaurants had no cuisine type classifications. In a real-world exercise, it would be reasonable to ask Foursquare to add ethnic classification to these restaurants in their database.

**Section 6: Conclusion**

My report has systematically analysed the Greater Manchester region, by Postal Code area, to determine the following:

* Most Popular Restaurant types in the Greater Manchester region - Indian
* Most Popular Eating areas in Greater Manchester – M2, M3 and M14

The client wants to open a popular restaurant type in a popular eating area. Based on these results and the client specifications I recommend the following:

Open an Indian Restaurant in the M14 Postal Code area of Manchester.