Capstone Project

The Battle of the Neighborhoods - Week 4 Report

Introduction and Problem

As part of life, people move around for many reasons, for career prospects, to be closer to better schools, for culture or to be closer to parents. There could be intercity migration or international migration, but both cases will happen.

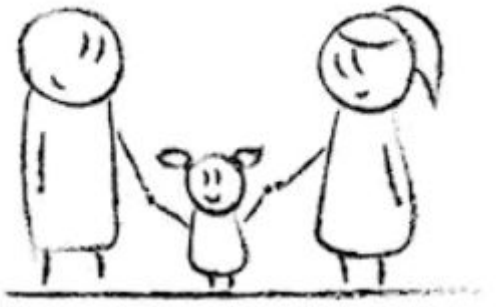
Prior to moving to any city, people will research the potential home and try to make sense of it all. There are multiple features of a potential city or neighbourhood all of which define the character or type of place a neighbourhood is and most importantly whether it is a suitable location to settle in. However, it can be a daunting task to collect all of this information and a very time consuming process especially if the potential location is an unknown location.

Doing a search online for a good neighbourhood will only garner results that are opinionated and possibly be subjective as everyone has their own preferences. A young person may want to be close to places associated with nightlife such as restaurants and bars, whereas a young couple may want to be closer to their place of work and within the reach of all the necessary amenities such as gyms or cafes.

In order to truly know whether a neighbourhood is suitable to settle in, we need hard data. Data such as house prices, schools, crime, weather etc. Being able to get this data will better inform the potential resident of the things they may value more.

This project will focus on Mr Willis, Mr Willis is married to his Wife Mrs Willis for 9 years and they have one Daughter who is currently in Primary School. They are currently overseas

residing in Singapore as expatriates and have been doing so for the past 5 years, however they wish to return to the United Kingdom but at the same time wish to move to one of the major UK cities. One of the major cities that they are targeting is Manchester, however aside from having visited Manchester in the past, they have never spent a great deal of time there and do not know anyone there to assist.



They want to move to a location that is quiet, within budget and has restaurants and schools nearby.

What we want to do is to help Mr Willis and his family determine which of the boroughs in Greater Manchester can be called their future home through the collection and analysis of publicly available data on the web.

Other Target Audiences

- This model can also be used by others who may wish to move to Manchester, and can be reconfigured to analyse other Cities around the world.
- Data Scientists, who want to implement the Exploratory Data Analysis techniques to obtain necessary data and analyze it.

This project can be used by the user at the time of rental apartment or buy house in a locality based on the distribution of various facilities available around the borough. As an example, this project would compare Manchester's boroughs and analyses the venues in each of those boroughs based on the number of visits by people in each of those places.

The project will also use K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc. This would give a better understanding of the similarities and dissimilarities between the various neighborhoods.

City Background

Manchester is a city and metropolitan borough in Greater Manchester, England. It is historically and traditionally a part of the county of Lancashire. It has a population of

547,627 as of 2018 (making it the fifth most populous English district). It lies within the United Kingdom's second-most populous urban area, with a population of 2.5 million and third most populous metropolitan area, with a population of 3.3 million. It is fringed by the Cheshire Plain to the south, the Pennines to the north and east, and an arc of towns with which it forms a continuous conurbation. The local authority for the city is Manchester City Council.

Data

In this project, I will be using the following datasets to help solve Mr Willis' problem.

The main criterion for the Willis Family in selecting a suitable borough will be based on:

- Safety
- Number of Restaurants
- Schools
- House within Budget

Key Libraries being utilised

Python packages and Dependencies:

• Pandas - Library for Data Analysis

• NumPy – Library to handle data in a vectorized manner

• JSON – Library to handle JSON files

• Geopy – To retrieve Location Data

• Requests – Library to handle http requests

• Matplotlib – Python Plotting Module

• Sklearn – Python machine learning Library

• Folium – Map rendering Library

General

Firstly we need to find out how many boroughs are in Greater Manchester and their locations. This information will be obtained by Webscraping Wikipedia and then converted into a dataframe using Pandas.

- List of Manchester Boroughs - https://en.wikipedia.org/wiki/List_of_places_in_Greater_Manchester ,and Manchester PostCodes -

https://simple.wikipedia.org/wiki/M_postcode_area#:~:text=The%20M%20postcode%20area%2C%20also,Manchester%20boroughs%20except%20for%20Stockport.

The Coordinates (Latitude and Longitude) of each Borough will be extracted from Google Maps.

We will also obtain the Demographics of each borough to gauge their size and extract any other information from this that may be of use to the Willis family.

- Demographics of Manchester - https://www.citypopulation.de/en/uk/greatermanchester/ or https://en.wikipedia.org/wiki/List_of_Greater_Manchester_settlements_by_population

Folium- Python visualization library would be used to visualize the neighborhoods cluster distribution of Manchester.

Safety

In order to have an understanding of the safety levels of each borough we shall extract the recorded crime rates in Greater Manchester, this information will be extracted from the UK police database - https://data.police.uk/ and this information will be in the form of a csv file which will be converted into a dataframe using the Pandas library.

We will convert this information into a visualisation of the information either using a Chart or a folum map showing the clusters of crime in their respective areas.

Number of Restaurants

We will use the Foursquare API gather information in relation to venues and places in and around the Greater Manchester boroughs. Foursquare is a social location service that allows users to explore the world around them.

Foursquare has other information such as reviews and photos giving a more layered view of the venues in a certain area. By gathering this information we will be able to get a more rounded view of what amenities and venues are available in a certain borough such as restaurants and cinemas.

HTTP requests would be made to this Foursquare API server using the location of Manchester's boroughs neighborhoods to pull the location information (Latitude and Longitude).

Foursquare API search feature would be enabled to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 50 and the radius parameter would be set to 1000 as the boroughs are quite large.

Extracting this information can give us broad view on whether a particular neighbourhood is primarily residential or more of a mixed use borough with restaurants.

Schools

Finding a good school for the Willis family is also a key requirement. The information source that we will be dealing with is a csv file downloaded from the gov.uk website:

- https://www.compare-school-performance.service.gov.uk/download-data?currentstep=year&regiontype=beforeStep&la=&downloadYear=2018-2019

This information will then be converted into a dataframe and merged with the other dataframes, in order to associate the schools to a borough coordinate.

House within Budget

We would also need to be realistic and consider that the cost of a property must be taken into consideration, as the best neighbourhoods often come at a price. There is a fine balance between a reasonably priced home and a nice neighbourhood to live in.

We shall therefore extract the median household prices paid in Greater Manchester from a csv extracted from https://mappinggm.org.uk/metadata/ which has recorded prices paid in all areas of Greater Manchester across a number of years, however we will use the latest prices paid in 2020 to analyse.

This information will be converted into a dataframe using Pandas and visualised in a graph using Matplotlib.

Analysis

Extensive comparative analysis of the boroughs would be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.

We will then use unsupervised machine learning algorithm K-mean clustering to form the clusters of different categories of places residing in and around the borough.

The primary display formats that will be utilised are folium maps to visualise the collated information on a map of Manchester and the supported using matplotlib charts to display information that could be seen to be better represented as.