Coursera IBM Applied Data Science Capstone Project

(The Battle of Neighbourhoods)

Airbnb Recommendation System for Manchester Neighbourhoods Using K-Means Clustering and Foursquare API in Python

By

Olumide Olumayegun

May 2020

Contents

1. Introduction and Business Problem 3

1.1 Introduction 3

1.2 Business Problem 3

2. Data Description 3

1. Introduction and Business Problem

1.1 Introduction

Manchester is a city and metropolitan borough in Greater Manchester, England, with a population of 547,627 as of 2018 (making it the fifth most populous English district) [1]. Manchester is the third most visited city in the UK, after London and Edinburgh. In 2019, it surpassed Edinburgh to become the second most visited city in the UK after London [2]. It is notable for its architecture, culture, musical exports, media links, scientific and engineering output, social impact, sports clubs and transport connections. There are three universities in the City of Manchester: The University of Manchester, Manchester Metropolitan University and Royal Northern College of Music. The University of Manchester is the largest full-time non-collegiate university in the United Kingdom. Manchester is well known as a city of sport with two decorated Premier League football clubs bear the city name – Manchester United and Manchester City. Therefore, Manchester is very popular with visitors seeking short-term accommodations. These include sport fans, business visits, school open day's visits, conferences, e.t.c.

Our client, who is planning a visit to Manchester, would like us to recommend the best Airbnb listings to book for a short stay in Manchester. However, this is not a straight forward choice as the clients have preferences for Airbnb as well as the kind of neighbourhood the Airbnb is located.

1.2 Business Problem

To build segregation and clustering models that will recommend to an individual suitable Airbnb accommodations in Manchester City area based on available data of Airbnb listings and venues categories in Manchester neighbourhoods.

2. Data Description

The development and analysis of this system will require a variety of data sources. Below is a brief description of each dataset as well as links to the original sources where the data was downloaded from.

- **Airbnb listing for Greater Manchester, England:** This dataset contains information for Airbnb properties for the area of Greater Manchester. It is provided by Inside AirBnb. It was downloaded from the website at http://insideairbnb.com/get-the-data.html. It is available as listings.csv file under Greater Manchester, England, United Kingdom on the website. This file was downloaded and saved in the working directory of this project. The dataset contains information and metrics for Airbnb listings in Greater Manchester. Some useful observations in the data include: Listing id, neighbourhood group, neighbourhood, latitude and latitude of the listing, room type, price, number of reviews for the listing, and availability. K-Means algorithm will be applied on this dataset to cluster the Airbnb listings into different groups.
- **Geographical coordinates of Manchester neighbourhoods:** This is the latitude and longitude of the 32 neighbourhoods in Manchester area. This was obtained from ArcGIS using geocoder function in Python. A table of Manchester neighbourhoods and their latitude and longitude was created by pulling the neighbourhood geographical

coordinates with Python calls. The neighbourhood geographical coordinates will be used to generate maps of Manchester with Folium. It will also form part of API request for venues in the neighbourhoods.

- **List and categories of venues in Manchester neighbourhoods:** This dataset contains important information about venues in each Manchester neighbourhood. This was obtained through API requests to Foursquare server. K-means algorithm will be applied on this dataset to cluster the Manchester neighbourhoods based on the categories of venues in the neighbourhood.

References

[1] "Population estimates for the UK, England and Wales, Scotland and Northern Ireland mid-2018". www.ons.gov.uk. Office for National Statistics. 26 June 2019. Retrieved 27 May 2020.

[2] "London visited by 50% of UK's tourists". BBC News. Archived from the original on 8 June 2013. Retrieved 21 May 2020.