Coursera IBM Applied Data Science Capstone Project

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

Olumide Olumayegun

May 2020

Introduction

- Manchester is a city and metropolitan borough in Greater Manchester, England, with a population of 547,627 as of 2018
- 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
- Our clients would like us to recommend the best Airbnb listings to book for a short stay in Manchester

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

Data Description

- **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 (http://insideairbnb.com/get-the-data.html)
- **Geographical coordinates of Manchester neighbourhoods:** This was obtained from ArcGIS using geocoder function in Python
- **List and categories of venues in Manchester neighbourhoods:** This dataset contains important information about venues in each Manchester neighbourhood

Methodology

- Load dataset for Airbnb listings in Greater Manchester

	id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availability_365
0	68951	Bolton	Bolton District	53.56214	-2.53305	Entire home/apt	65.0	2	82	196
1	85109	Manchester	City Centre	53.48074	-2.23186	Private room	50.0	2	434	349
2	159189	Rochdale	Rochdale District	53.61537	-2.19885	Shared room	55.0	1	0	349
3	283495	Rochdale	Rochdale District	53.56259	-2.21945	Private room	60.0	3	10	300
4	299194	Stockport	Stockport District	53.37600	-2.04462	Entire home/apt	50.0	2	255	349

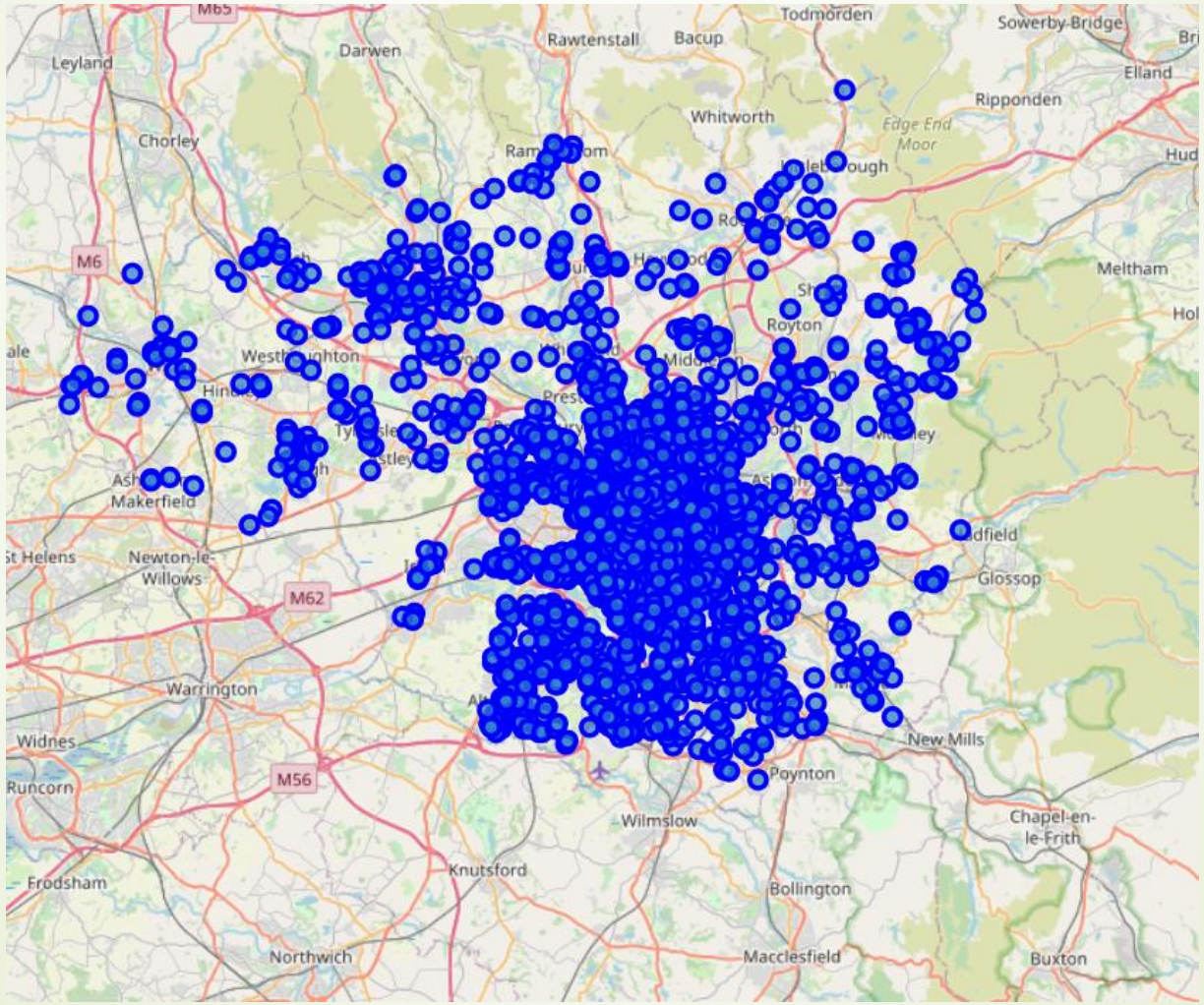
Methodology

- Statistical analysis

	id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availabilit
count	4.924000e+03	4924	4924	4924.000000	4924.000000	4924	4924.000000	4924.000000	4924.000000	4924.0
unique	NaN	10	41	NaN	NaN	4	NaN	NaN	NaN	
top	NaN	Manchester	Salford District	NaN	NaN	Private room	NaN	NaN	NaN	
freq	NaN	2823	787	NaN	NaN	2412	NaN	NaN	NaN	
mean	2.855575e+07	NaN	NaN	53.472945	-2.249493	NaN	88.091998	3.518278	27.871649	153.7
std	1.143300e+07	NaN	NaN	0.045497	0.081715	NaN	173.412694	15.870997	51.629740	138.8
min	6.895100e+04	NaN	NaN	53.345830	-2.708880	NaN	0.000000	1.000000	0.000000	0.0
25%	1.996664e+07	NaN	NaN	53.448817	-2.275375	NaN	30.000000	1.000000	1.000000	0.7
50%	3.044096e+07	NaN	NaN	53.474685	-2.239015	NaN	52.000000	1.000000	7.000000	124.5
75%	3.888886e+07	NaN	NaN	53.488280	-2.218380	NaN	90.000000	2.000000	30.000000	323.0
max	4.345605e+07	NaN	NaN	53.677360	-1.976980	NaN	5539.000000	400.000000	545.000000	363.0

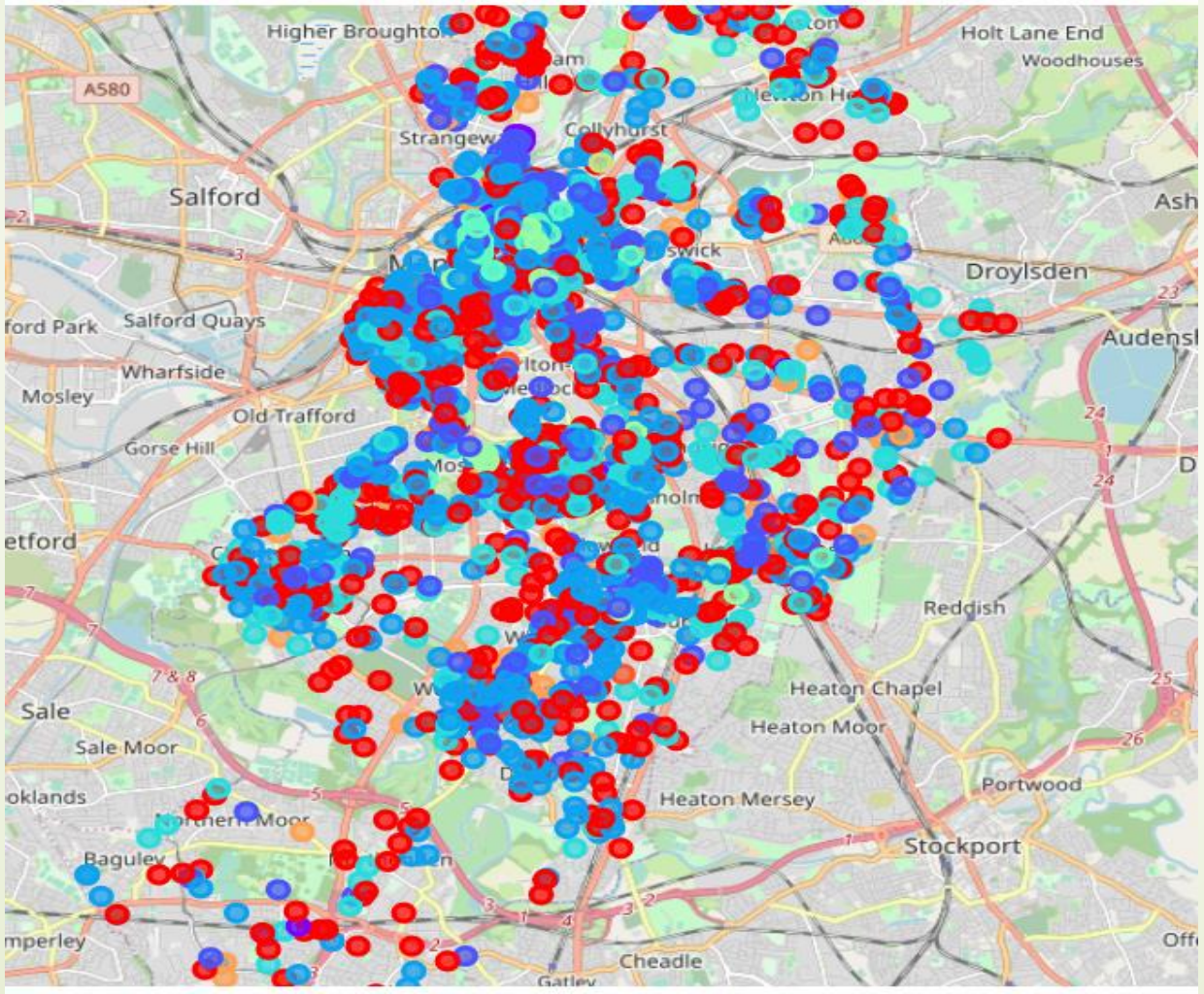
Methodology

- Map of Greater Manchester showing Airbnb listings



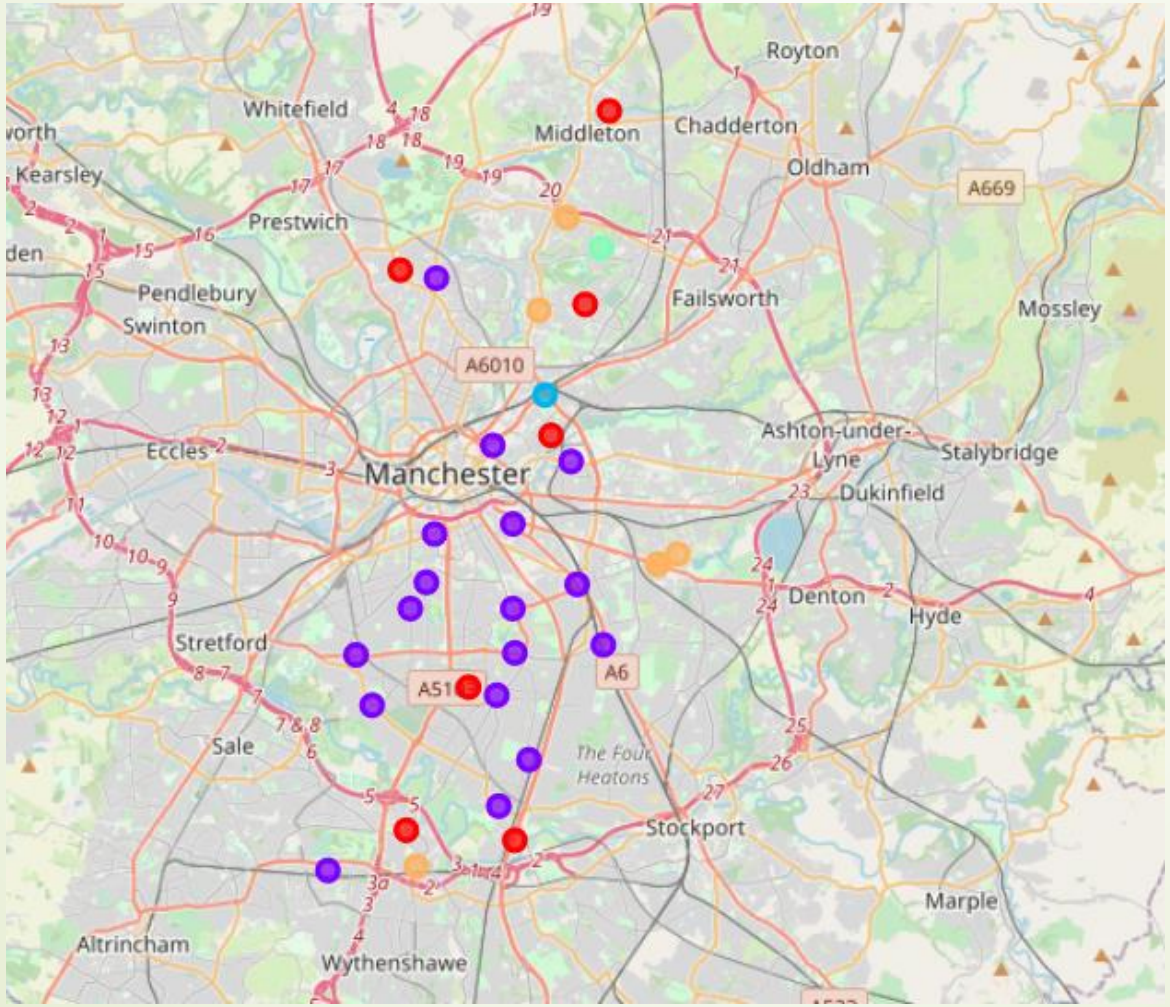
Methodology

- K-Means clustering of Airbnb listings



Methodology

- K-Means clustering of Manchester neighbourhoods

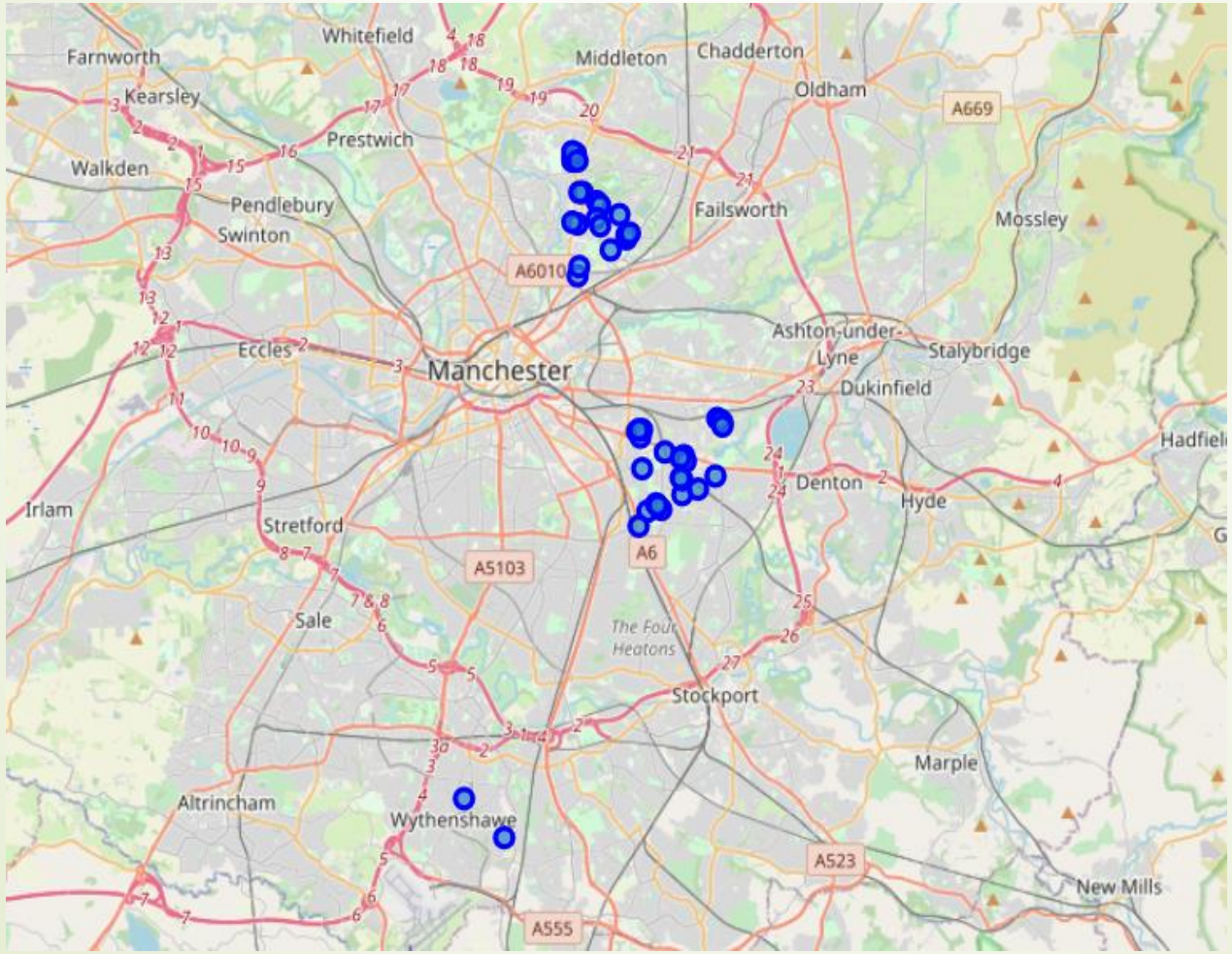


Results and Discussion

- Use case scenario
  - The customer preferences for Airbnb accommodation is as follows:
    - Room type: private room
    - Price: £55
    - Minimum night: 2
    - Number of reviews: 100
    - Availability: 340 days
  - In addition, the customer prefers accommodations in a neighbourhood with a lot of Chinese and fast food restaurants and supermakets/markets

Results and Discussion

- Recommended Airbnb listings for the client



Conclusions

- In this project, we have been able to deploy data science powers of dataset wrangling, machine learning models (K-Means clustering) and geolocation APIs to develop a system to recommend Airbnb listings to our clients looking to book accommodation in Manchester for short stays.
- The capability of the models were demonstrated through a use case scenario.
- Out of the over 3000 Airbnb listings, our system was able to pick about 50 listings suitable for our client.
- It should be emphasized that this 50 listings can be further narrowed down for the client by a stricter application of the criteria.