Brummy or Mancunian - Which one are you? [Birmingham|Manchester]

Introduction/Business Problem

When we think of UK, we think of London. Although a great place to visit, London isn't necessarily affordable by everyone as a place to live. Due to this, many prefer living in a city close to London so that not only can they live in a bigger house for a lower rent, and save on many other things while earning a decent income, but also can drop in London within an hour or two. Smart.

A problem with this approach is that if you just want a quick night-out, maybe go to a fancy restaurant, or visit a museum, or a park, you shouldn't have to sit in a train for over an hour, every single time. The city you choose should have things to do that you prefer.

What then, is the next best thing?

A quick Google search revealed people prefer either Birmingham or Manchester. Many say Birmingham's better, even better than London, while many side with Manchester.

But which one's better for you?

Data

I will use Foursquare to explore both the cities. This should help a person select the city they'd like to live in.

I will then use Foursquare to explore the neighborhoods to help the person select a place to call home.

I will get the list of postal codes and neighborhoods from the following:

Birmingham - https://en.wikipedia.org/wiki/B_postcode_area

Manchester - https://en.wikipedia.org/wiki/M_postcode_area

They have details about the 'Postal district', 'Post town', 'Coverage', 'Local Authority Area'.

I will use 'Post town' to filter the data and select only 'BIRMINGHAM' and 'MANCHESTER', respectively.

I will use 'Coverage' (Neighborhoods) to get the latitudes and longitudes needed, using geocoder.

Methodology

1) Creating datasets

- Create the Birmingham and Manchester datasets using the above links.
- Add latitudes and longitudes to the neighborhoods.
- Remove null values and unnecessary columns.

2) Exploration

- Define a function to get nearby venues and their categories – based on the name of the neighborhood and its coordinates.
- Call this function for Birmingham and Manchester.
- Select the top 10 venues with their categories for each neighborhood, in each city.

3) Clustering the neighborhoods

- One hot encode the venue categories and find its 10 most common venues for each neighborhood.
- Group the dataset by neighborhoods for clustering.
- Define the Kmeans model.
- Fit the model using the data frame having the values of the 10 most common venues for each neighborhood. This will give you a label for each cluster.
- Add these labels to your data frame for each neighborhood.

4) Visualize the clusters

- Use folium to visualize the clusters on a map.
- Pass the latitude and longitude values for the respective cities.
  - ➔ Birmingham: lat = 52.4862 | lon = -1.8904
  - ➔ Manchester: lat = 53.4808 | lon = -2.2426
- Add a CircleMarker to the map for each cluster with the name of the neighborhood and cluster on it.

5) Explore the clusters

- Create a separate dataframe for each cluster and see what the common venues are in that cluster.
- Analyze, and make observations. For what kind of people would a particular cluster of neighborhoods be suitable for?

Results

Top 10 Venue Categories:

Similarities

- Pubs
- Indian Restaurants
- Coffee Shops & Cafes
- Bars
- Restaurants
- Supermarkets

Differences

- Birmingham has more Fast Food Restaurants, Italian Restaurants and Soccer Stadiums.
- Manchester has more Parks, Hotels and Grocery Stores.

Birmingham neighborhoods:

- If you are an Indian, check out Edgbaston, Buckland End, Kingshearst and Hamstead.
- If close access to coffee shops is a must for you, check out the City Centre, Vauxhall, Stirchley and West Heath.
- If you like non-veg, many neighborhoods in Birmingham have Fried Chicken Joints.

Manchester neighborhoods:

- If you want close access to a Bar and a Flea Market, check out Deansgate, Tyldesly and Whitefield.
- If you're just visiting Manchester and want to live in a hotel, check out Stretford and Trafford Park.
- If you're a Vegan, almost all neighborhoods will be fine.

Discussion

There are many ways we could improve this project:

- Better Clustering
    - I used K-Means, and I also tried Agglomerative Hierarchical Clustering and Density-Based Clustering. None of them gave good results. In both cases, cluster 1 has the most values. They are outliers, which is not ideal. There must ways to improve the dataset and the clustering algorithms.
- Using Crime Data
    - Getting insights into this would definitely help make a better decision.
- Using Population Data
    - Getting details about the population would be very helpful. Is it a young city/neighborhood or more suitable for the elderly? Are there any neighborhoods where you can find people from your home country?
- Using Weather Data
    - Manchester's in the North of England. How is the weather up there compared to Birmingham?

Conclusion

This project is of course not something you could use alone to make a life decision. For this to reach that level, a lot of improvements need to be moved, some of which have been mentioned above.

I did this project to learn and apply my skills in a clustering problem, and I must say, I learnt a lot. Also learnt a few new things about the two cities.