Title

New York, Paris, and London; do they have similarities?

Is New York City more like London or Paris or vice versa? These cities are in different countries, so obviously we would assume that they should be completely different. These cities are similar in the fact that they are multicultural cities that attract travelers from all over the world, so there is bound to be some similarities.

Purpose

The focus of this project is to use a clustering algorithm - KMeans- to cluster similar Boroughs in these cities together and try to find similarities.

Newyork Data

The newyork data was gotten from this wikipedia page https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City (https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City), found a table containing all the boroughs in Newyork city, neighboourhoods in each borough, the population size and land size. Only the names of boroughs are required for this project, so I did the neccesary data cleaning.

There are 5 Boroughs in Newyork city

Paris Data

The Paris data gotten from this wikipedia page https://en.wikipedia.org/wiki/Arrondissements_of_Paris (https://en.wikipedia.org/wiki/Arrondissements_of_Paris) was a little complicated as column names are in french. Paris has a unique number of 'Arrondisements' and I'll be using them in this project.

Paris has --- Arrondisements

London Data

Got the London data from https://en.wikipedia.org/wiki/List_of_places_in_London#Neighbourhoods (https://en.wikipedia.org/wiki/List_of_places_in_London#Neighbourhoods). The column 'London borough' is the needed column. Here.

London has -- Boroughs

Library Used

1. Pandas
2. Geocoder
3. Foursquare
4. KMeans
5. Unicodedata
6. Json

Steps

After cleaning the cities data, the next step is to get the coordinates of the locations to be analyzed and then pass these coordinates into foursquare API to get the venue information in these cities.

The clustering algorithm will be clustering location/boroughs together based on the type of venues. I'll be working with the top 10 venues in each place, so locations with similar venues will be grouped together.

Geocoder Data

The geocoder library was used to get the coordinates of the boroughs in all the different cities.

I was able to get the longtitude and latitude of each borough using the geocoder library.

Foursquare API Data:

I used the Four-square API for my data collection as it has a database of a lot of places

The information obtained from each borough are as follows:

1. Borough
2. Borough Latitude
3. Borough Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Conclusion

The boroughs in Paris were not clustered with any other borough in London or Newyork, but some boroughs in London and newyork were clustered togther, these can be seen in Clusters 6 and 9.
It is understandable why Paris would not have similar top venues with Londo or Newyork. This could be because of the culture and language difference.

New york and London are most alike among these three cities.

In [ ]: