Comparing Dining in Major Cities

Coursera Capstone

6/20/2020

Marrs

INTRODUCTION/BUSINESS PROBLEM

When traveling, good food can make all the difference. For tourists, it is important to know whether a certain travel destination offers the type of food or dining they might enjoy. This code will answer the question, "How can we categorize major cities by the types of dining available?" by comparing dining in major cities. This will allow people to gauge how similar a new destination is to a known city in terms of food.

DATA

Two sets of data are used. One data frame is created with the latitude and longitude of 22 major cities around the world (Figure 1). The second data set comes from a search query for ‘restaurants’ to foursquare. After the data is cleaned it contains the names and types of all dining within 10,000 m of the center of each major city (Figure 2).

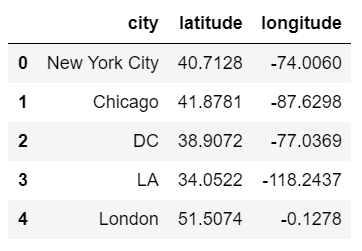


Figure Created Data Frame Sample



Figure Foursquare Data Frame Sample

METHODOLOGY

The final output of this code is a table which contains the cluster number, and the top 10 most common restaurant types for each city(Figure 3). In addition, the code returns a map of the clustered cities.

To get these outputs a series of steps were taken. First libraries were imported and foursquare credentials were defined. Then a new data frame was manually created containing the coordinates of 22 major cities around the world (Figure 1). Then a map with markers for each city was displayed.

The next step was to get the most common restaurant types from every city. This required to parts. First, to create a function that would return a table containing the name and category of every restaurant withing 10 km of each city (Figure 2). The second step was to transform the data into a new frame which contained the cluster number, and the top 10 most common restaurant types for each city (Figure 3).

The final step was clustering each city and creating a map. To cluster, KMeans function was used, and cluster labels were merged with the table in Figure 3 to create Figure 5. Finally, a map was created using folium. Each color represents a unique cluster of cities with similar types of restaurants (Figure 4).



Figure common restaurant types by city sample

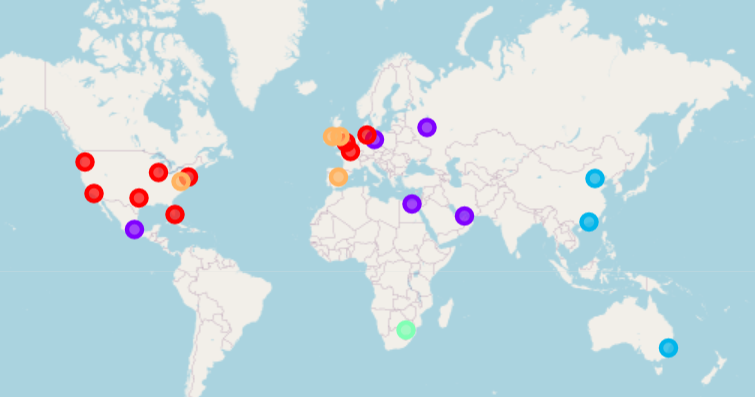


Figure Final clustered map



Figure Final results table sample

RESULTS

Based on this data, restaurants in Mexico city compare to cities near the Middle East. Food venues in the United states and western Europe seem to be similar, even though together they make up two clusters. Johannesburg in South Africa did not compare to any other cities and was clustered by itself. Finally, cities in the far east are shown to have similar dining venues. Overall, the data seems to suggest that dining venues compare well regionally. In other words, cities within a certain region are likely to offer similar dining venues.

DISCUSSION

It would be interesting to see how the results change with more cities included. For example, Johannesburg did not compare well with other cities, but it was also the only city representing its region. In addition, restaurant types from metropolitan cities sometimes included ‘Office’ if the office was associated with a restaurant chain. Small errors like these may have affected the final result. Finally, the scope of restaurants provided by foursquare is somewhat limited. For example, the number of restaurants in NYC measures near 26,000, however foursquare provided less than 100.

CONCLUSION

This Data Science project fulfilled the goal to compare cities around the world by restaurant type. This data will be useful to tourists, travel agencies, and perhaps even airline industries. Overall, cities within a certain region are likely to offer similar dining venues.