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. For example, a family traveling to Europe would need to make sure they could find a city that offers dining venues that meet the needs of every member of their family. 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 and what types of dining venues they offer. Not only will this be helpful information to travelers, but also to travel agencies who can make recommendations to tourists.

DATA

Two sets of data are needed. One data frame is created within the code containing the latitude and longitude of 22 major cities around the world (Figure 1). This will allow us to map the clusters of cities for a helpful visual representation using folium. 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). This data is used to find relevant information about the types of dining venues of each city which will be used for clustering.

	city	latitude	longitude
0	New York City	40.7128	-74.0060
1	Chicago	41.8781	-87.6298
2	DC	38.9072	-77.0369
3	LA	34.0522	-118.2437
4	London	51.5074	-0.1278

Figure 1 Created Data Frame Sample

	city	name	category
0	New York City	The Shops & Restaurants at Hudson Yards	Shopping Mall
1	New York City	WFC Shops & Restaurants	Food Court
2	New York City	Restaurants Open 24	Falafel Restaurant
3	New York City	Zhou Restaurants	Food
4	New York City	Barilla Restaurants	Italian Restaurant

Figure 2 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).

	city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bejing	Chinese Restaurant	French Restaurant	Indian Restaurant	New American Restaurant	Diner	Comfort Food Restaurant	Corporate Cafeteria	Coworking Space	Cruise	Cupcake Shop
1	Berlin	German Restaurant	Restaurant	Eastern European Restaurant	Sandwich Place	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Sri Lankan Restaurant	Boarding House	French Restaurant
2	Cairo	Restaurant	Hotel	Yemeni Restaurant	Food Court	Cruise	Diner	College Cafeteria	Comfort Food Restaurant	Corporate Cafeteria	Coworking Space
3	Chicago	Sandwich Place	Food	Office	Pizza Place	Mexican Restaurant	Fish Market	Event Space	Indian Restaurant	Eastern European Restaurant	Middle Eastern Restaurant
4	DC	Sandwich Place	Food	Hotel	Bank	Diner	Office	Middle Eastern Restaurant	Mexican Restaurant	BBQ Joint	American Restaurant

Figure 3 common restaurant types by city sample

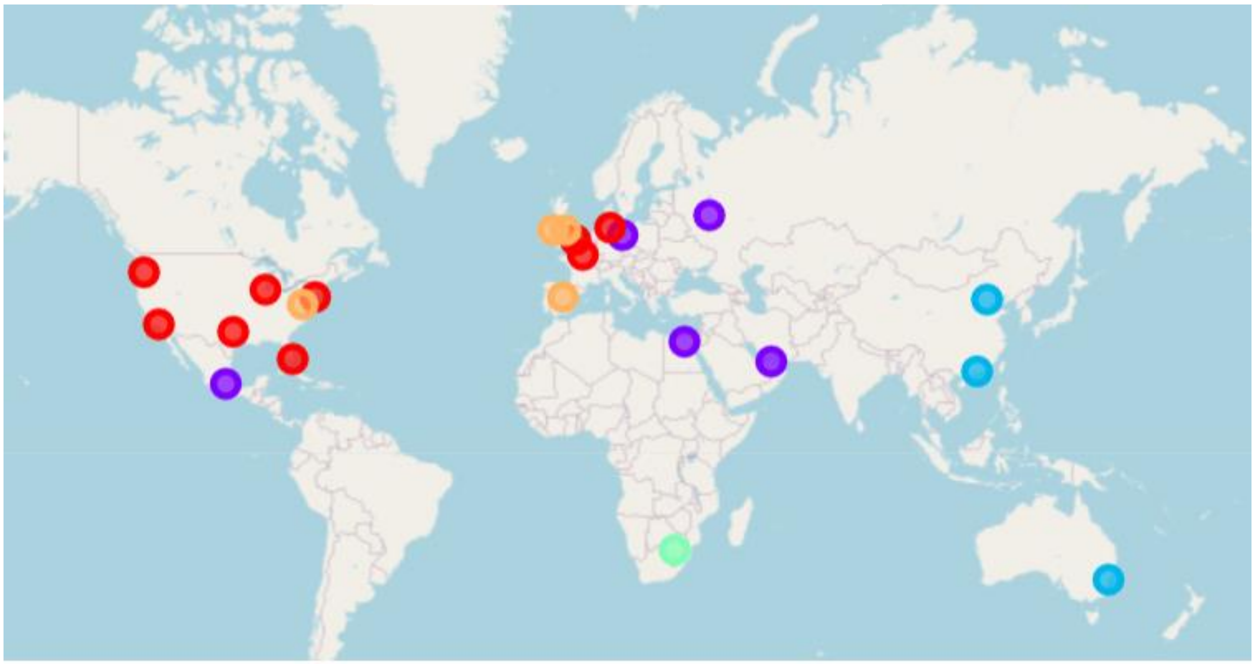

Figure 4 Final clustered map

	city	latitude	longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	New York City	40.7128	-74.0060	0.0	Food	Office	Sandwich Place	Miscellaneous Shop	Furniture / Home Store	Italian Restaurant	Food Court	American Restaurant	Falafel Restaurant	Coworking Space
1	Chicago	41.8781	-87.6298	0.0	Sandwich Place	Food	Office	Pizza Place	Mexican Restaurant	Fish Market	Event Space	Indian Restaurant	Eastern European Restaurant	Middle Eastern Restaurant
2	DC	38.9072	-77.0369	4.0	Sandwich Place	Food	Hotel	Bank	Diner	Office	Middle Eastern Restaurant	Mexican Restaurant	BBQ Joint	American Restaurant
3	LA	34.0522	-118.2437	0.0	Food	Sandwich Place	Mexican Restaurant	Japanese Restaurant	Restaurant	Spanish Restaurant	American Restaurant	Seafood Restaurant	Department Store	Peruvian Restaurant
4	London	51.5074	-0.1278	0.0	Sandwich Place	Restaurant	Office	African Restaurant	Turkish Restaurant	Italian Restaurant	Miscellaneous Shop	Pizza Place	Seafood Restaurant	Food Court

Figure 5 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.