Capstone Project: City Eating

Introduction

A key aspect when deciding where to travel are the cuisines and cultures of the destination. The aim of this project is to quantify the range of cuisines available in the most popular cities around the world, thus allowing travelers to make an informed decision about their next destination. Using machine learning techniques cities will be clustered based on the similarity of their cuisines.

Data

Data for the location of each city centre will be sourced using the geopy package.

	City	Lat	Long
14	Barcelona	41.382894	2.177432
7	Berlin	52.517037	13.388860
2	Cairo	30.048819	31.243666
1	Delhi	28.651718	77.221939
13	Lisbon	38.707751	-9.136592

Data for each city's most popular restaurant venue and range of cuisines will be sourced using the Foursquare API within python. 16 of the most popular travelling destinations were chosen for the cities. For each city the top 50 restaurant venues were located within 10km of the city centre. This data will then be used to group cities into clusters based on the similarity of their cuisines.

	City	City Latitude	City Longitude	Venue	Cuisine
0	Tokyo	35.682839	139.759455	Restaurant Rosette (レストラン ロゼット)	French
1	Tokyo	35.682839	139.759455	The Restaurant by AMAN (ザ・レストラン by アマン)	Mediterranean
2	Tokyo	35.682839	139.759455	MOTIF RESTAURANT & BAR	Restaurant
3	Tokyo	35.682839	139.759455	Restaurant ベラージュ	French
4	Tokyo	35.682839	139.759455	Tokyo Athlete Restaurant (東京アスリート食堂)	Japanese
...	...	...	...	...	...
795	Moscow	55.750446	37.617494	Restaurant Russian	B & B
796	Moscow	55.750446	37.617494	Novikov	Asian
797	Moscow	55.750446	37.617494	The Waiters	Restaurant
798	Moscow	55.750446	37.617494	Yoko	Japanese
799	Moscow	55.750446	37.617494	Гамбит Lounge	Nightclub

800 rows × 5 columns

Methodology/Results

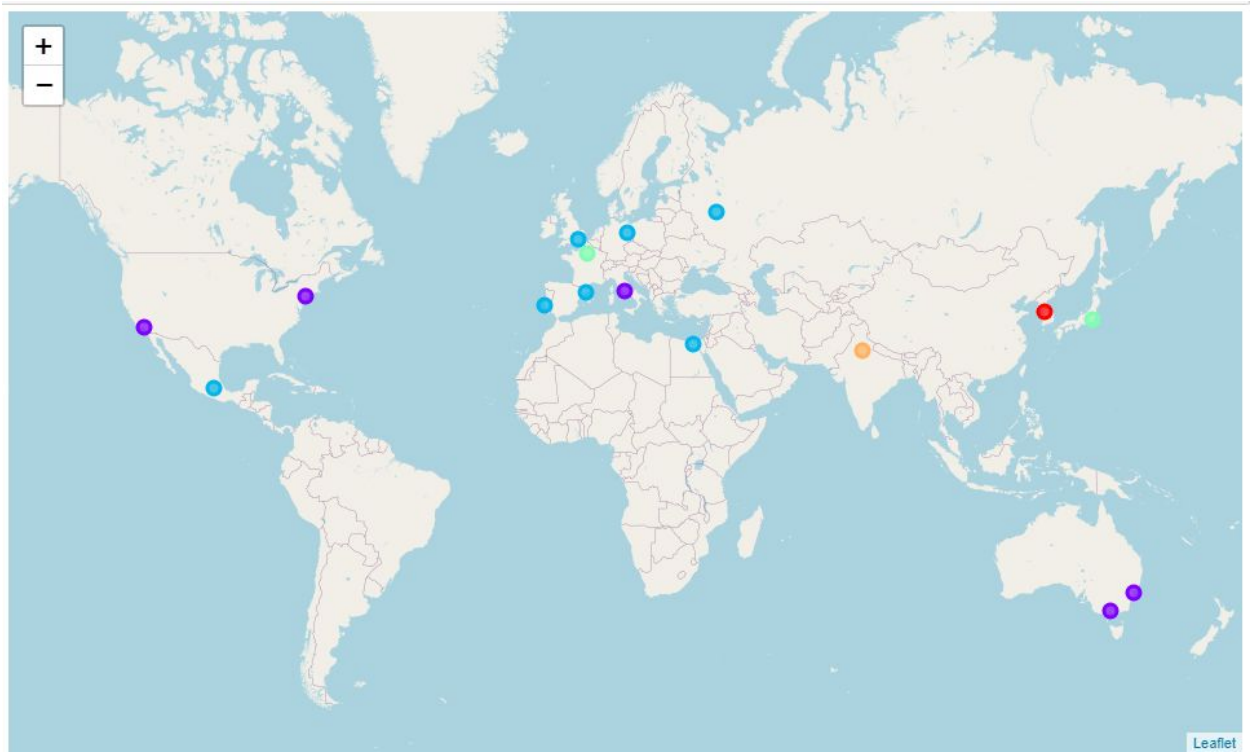
Firstly I performed one-hot-encoding on the city venue data and then grouped the data by city to obtain the mean of each cuisine within that city. I then ranked the top 5 cuisines within each city.

	City	1	2	3	4	5
0	Barcelona	Restaurant	Spanish	Mediterranean	Chinese	Seafood
1	Berlin	Restaurant	Breakfast	German	French	Turkish
2	Cairo	Restaurant	Middle Eastern	Falafel	Yemeni Restaurant	Kebab
3	Delhi	Indian	Restaurant	Chinese	North Indian	Diner
4	Lisbon	Portuguese	Restaurant	Indian	Asian	Himalayan
5	London	Restaurant	Chinese	English	Indian	Italian
6	Los Angeles	Mexican	Chinese	Food	Japanese	American
7	Melbourne	Chinese	Korean	Indian	Asian	Restaurant
8	Mexico City	Restaurant	Mexican	Bar	Buffet	Chinese
9	Moscow	Restaurant	Russian	Seafood	Nightclub	Italian
10	New York	Chinese	Italian	Dim Sum	Shop	Seafood
11	Paris	French	Restaurant	Turkish	Middle Eastern	Cafeteria
12	Rome	Italian	Restaurant	Sushi	Chinese	Pizza
13	Seoul	BBQ	Korean	Indian	Asian	Italian
14	Sydney	Japanese	Chinese	Australian	Italian	Seafood
15	Tokyo	French	Restaurant	Café	Italian	Yoshoku

I then performed k means clustering on the encoded data with 5 clusters.

	Cluster Labels	City	1	2	3	4	5
0	2	Barcelona	Restaurant	Spanish	Mediterranean	Chinese	Seafood
1	2	Berlin	Restaurant	Breakfast	German	French	Turkish
2	2	Cairo	Restaurant	Middle Eastern	Falafel	Yemeni Restaurant	Kebab
3	4	Delhi	Indian	Restaurant	Chinese	North Indian	Diner
4	2	Lisbon	Portuguese	Restaurant	Indian	Asian	Himalayan
5	2	London	Restaurant	Chinese	English	Indian	Italian
6	1	Los Angeles	Mexican	Chinese	Food	Japanese	American
7	1	Melbourne	Chinese	Korean	Indian	Asian	Restaurant
8	2	Mexico City	Restaurant	Mexican	Bar	Buffet	Chinese
9	2	Moscow	Restaurant	Russian	Seafood	Nightclub	Italian
10	1	New York	Chinese	Italian	Dim Sum	Shop	Seafood
11	3	Paris	French	Restaurant	Turkish	Middle Eastern	Cafeteria
12	1	Rome	Italian	Restaurant	Sushi	Chinese	Pizza
13	0	Seoul	BBQ	Korean	Indian	Asian	Italian
14	1	Sydney	Japanese	Chinese	Australian	Italian	Seafood
15	3	Tokyo	French	Restaurant	Café	Italian	Yoshoku

Using folium I then mapped the cities, with each represented by the colour of their associated cluster.



Discussion

We can see that Delhi and Hong Kong have their own clusters, reflecting their unique cuisine landscape. We can also see that Sydney, Melbourne, NY and LA have a similar range of cuisines, Rome interestingly is also clustered with these cities. This may be due to the heavy influence of Italian in the other four cities. Lastly Tokyo and Paris are strangely clustered together, however when I researched each city it is apparent that the restaurants around Tokyo city centre are heavily influenced by French cuisine.

Conclusion

This project was able to successfully highlight some commonalities between the major restaurant cuisines in cities around the world. It should be noted that better results may be obtained by extending the radius of the Foursquare search, increasing the number of cities and also increasing the number of restaurants used in the data. However to keep computational times low and to avoid exceeding the maximum number of requests to the Foursquare API these numbers were kept low.