A Tale of Three Cities: A Comparison between London, Moscow and San Francisco

Hongfang Lu

May 14 2019

1. Introduction

The city of San Francisco is one of the most famous people in the world. It attracts millions of tourists each year by its numerous landmarks, restaurants, cultural sites and so on. It might just be the starting point of a traveler on his/her journey to travel around the world.

However travelling abroad is always very expensive and time consuming for most people. The flights and travel expenses are themselves huge amount of money, let alone the planning phase of which landmark or restaurant to visit. Additionally there are so many cities around the world, which can serve as the destinations of journey. Which city should a traveler choose to be his/her next wonderland? The problem can be solved by comparing the city (cities) he/she has explored with the potential options.

In this project, the city of San Francisco is considered as a city Jamie (or a group of friends) has traveled. From that point, Jamie is deciding which city to explore next. London and Moscow the capitals of the United Kingdom and Russia respectively, and they are the options in Jamie's mind. By comparing how similar or dissimilar London and Moscow are to San Francisco, we can recommend Jamie to travel to either one of them or both. The same process can theoretically apply to any other different cities in the world.

2. Data and Cleansing

The raw data of the project is scraped using beatifulsoup from the following three Wikipedia pages.

https://en.wikipedia.org/wiki/List_of_areas_of_London

https://en.wikipedia.org/wiki/Administrative_divisions_of_Moscow

https://en.wikipedia.org/wiki/Category:Neighborhoods_in_San_Francisco

The neighbourhoods, areas and boroughs (in Moscow they are named differently) are scraped and formatted. The areas include their latitude and longitude information in their respective pages. The format of geographical coordinates are in different styles, such as '1'37°45'16"N 122°27'33" or "37.76424°N 122.42366°W". A cleansing of the formats are done so that the latitudes and longitudes are all in decimal formats with positive signs (N, E) and negative signs (S, W).

The formatted dataframe looks like the following:

	Link	Location	London_borough	Post_town	Postcode_district	Dial_code	OS_grid_ref	Latitude	Longitude
0	/wiki/Abbey_Wood	Abbey Wood	Bexley, Greenwich [1]	LONDON	SE2	020	TQ465785	51.486400	0.110900
1	/wiki/Acton,_London	Acton	Ealing, Hammersmith and Fulham[2]	LONDON	W3, W4	020	TQ205805	51.513519	-0.270661
2	/wiki/Addington,_London	Addington	Croydon[2]	CROYDON	CR0	020	TQ375645	51.358300	-0.030500
3	/wiki/Addiscombe	Addiscombe	Croydon[2]	CROYDON	CR0	020	TQ345665	51.381000	-0.066300
4	/wiki/Albany_Park,_Bexley	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728	51.426400	0.102600

The formatted geometrical coordinates are fed into the foursquare api to generate recommended venues. All the areas/neighbourhoods/locations form a web and a Delaunay analysis is performed to find out the average distance between the vortices. This distance is used as the radius parameter in the foursquare api, so that as many as venues can be generated/captured. The venues information is similar to what we did in the lab, as following

	name	categories	lat	lng
39	Brill	Coffee Shop	51.525767	-0.109477
18	PizzaExpress	Pizza Place	51.406065	0.016451
73	Tariro Fairtrade Coffee House	Coffee Shop	51.401712	-0.195487
38	Jun Ming Xuan	Chinese Restaurant	51.595409	-0.242935
24	The London Borough of Barking & Dagenham Stadium	Soccer Stadium	51.547615	0.160125

Duplicates will appear during the exploration process because the explorative circles of neighborhoods overlap in a lot of cases. The duplicates are removed. Altogether there are ~12k, ~4k, ~3k unique venues generated respectively in London, Moscow and San Francisco. Then the venues are used to analyze the similarities between the cities.

3. Methodology

In this project we will use be comparing the similarities between three cities, mainly between San Francisco and London, San Francisco and Moscow. The types of venues are extracted and will be the criteria based on which a recommendation of London or Moscow will be made.

For each city the venues will be divided into several categories. The percentage of each category is calculated and compared across different cities. The top venue types in San Francisco will be the basis of comparison.

Then for the venues in London, a clustering analysis will be performed to see how they are distributed in the city. London has ~12k nodes and is a good example to perform clustering analysis. There are no classification or regression models used in this project. DBSCAN and Birch clustering algorithms are used because they don’t need the number of clusters to begin with. However the geometrical locations are not evenly distributed, which lead to meaningless clustering.

4. Analysis and Results

4.1 Exploratory Data Analysis

4.1.1 General Categories

Each of these three cities has their corresponding popular venue types. As can be seen from Figure 1, top venues in London, Moscow, San Francisco are (Pub, Coffee Shop, Cafe, Grocery Store, Park), (Park, Gym / Fitness center, Coffee Shop, Healthy Food Store, Supermarket), (Coffee Shop, Park, Cafe, Bakery, Chinese Restaurant).

As can be seen the drinking places are definitely among the most popular venues in all cites. In London there are ~1000 pubs and ~700 coffes shops. For San Francisco coffee shop and cafe rank 1st and 3rd in all the venues.

Park also comes in the top five venues in all three cities, which is a sign of how relaxation and entertainment matter in people's lives.

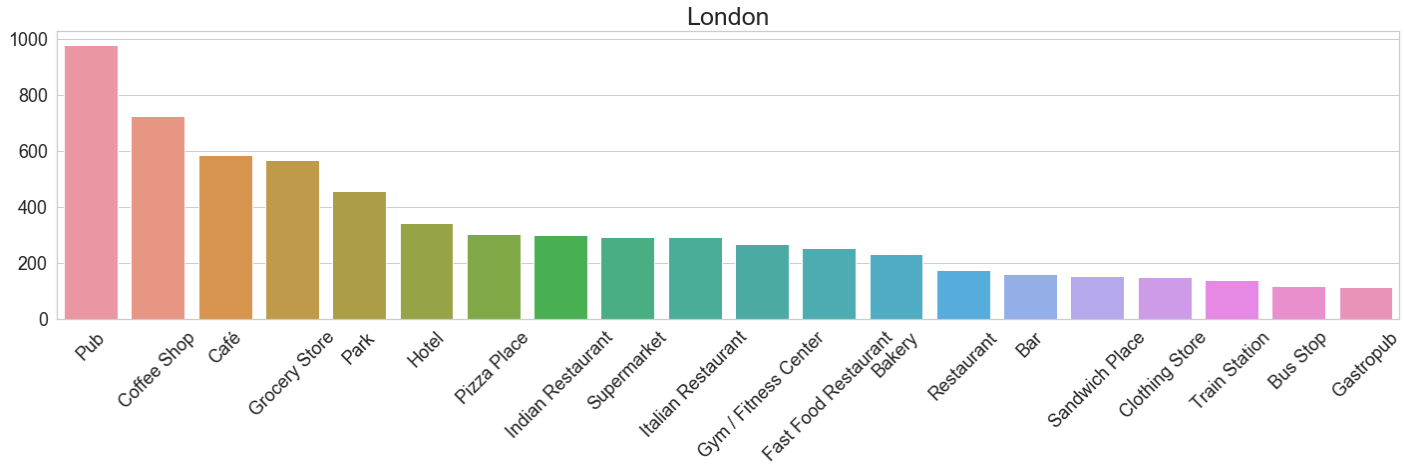


Figure 1(a)

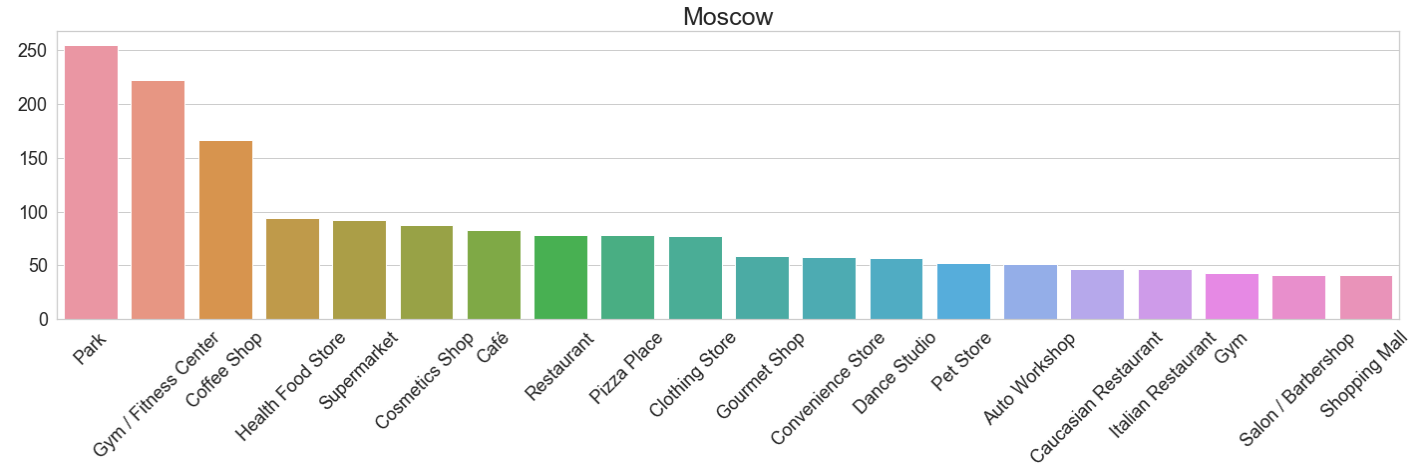


Figure 1(b)

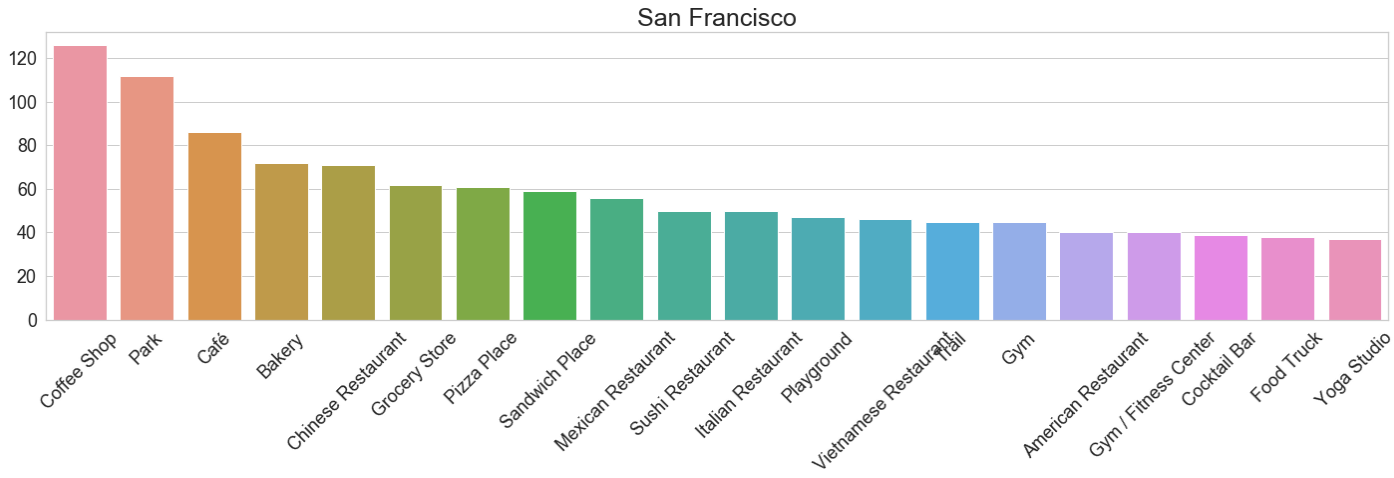


Figure 1(c)

There are all together mainly 6 types,

- Dining places: food trucks / restaurants / diners / …
- Drinking places: bars / pub / coffee / …
- Gyms: fitness / gym / yoga / swimming / rugby / …
- Groceries: supermarket / grocery store / market / …
- Entertainments: cultural sites / museums / arts / nature / park / cinema / …
- Education: University / colleges / school / library / …

The distribution of these six types is plotted in Figure 2. It can be seen that this major six categories take up about 60% of all venues in all cities. And the top three categories are dining, drinking and entertainments in London and San Francisco, however in Moscow it is dining, entertainments and gym.

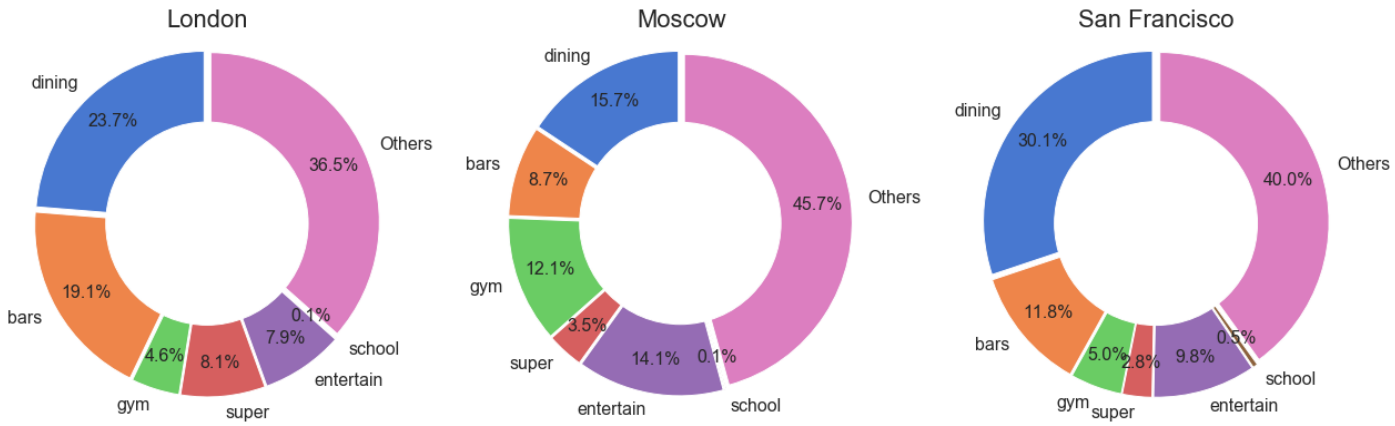


Figure 2

4.1.2 Dining places

Dining places are definitely the most important aspect of a city, of which the ranking of top 20 restaurants are shown in Figure 3 for each city. The top 5 restaurants in the cities are,

- **London** - Indian, Italian, Fast Food, Turkish, Chinese
- **Moscow** - Health Food, Italian, Caucasian, Fast Food
- **San Francisco** - Chinese, Mexican, Italian, Sushi, Vietnamese

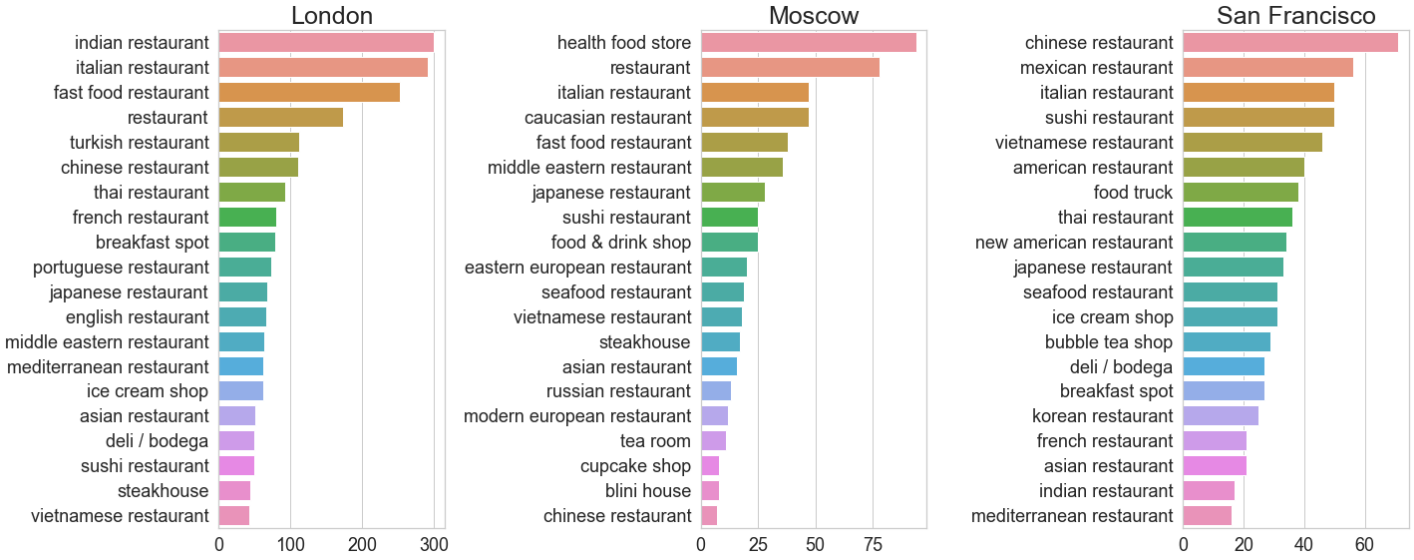


Figure 3

According to Demographics of San Francisco (https://en.wikipedia.org/wiki/Demographics_of_San_Francisco), in 2010, there are 41.9% Non-Hispanic White, 33.3% Asian and 15.1% Hispanic. These top three ethnical groups correspond to the top three restaurant types in San Francisco. One interesting fact is that 3 out of the top 5 types are Asian, namely Chinese, Sushi and Vietnamese. It originates from the food culture.

The distribution of top five restaurants in San Francisco is shown in Figure 4. Chinese restaurants distribute in the western part, Italian restaurants mainly distribute in the northeastern part and Mexican restaurants mainly distribute in the southeastern part. According Map of racial distribution in San Francisco Bay Area, there are more Asian people on the west, more white people on the northeast, more Hispanic people in the southeast, which relates to the observation of the distribution of restaurants.

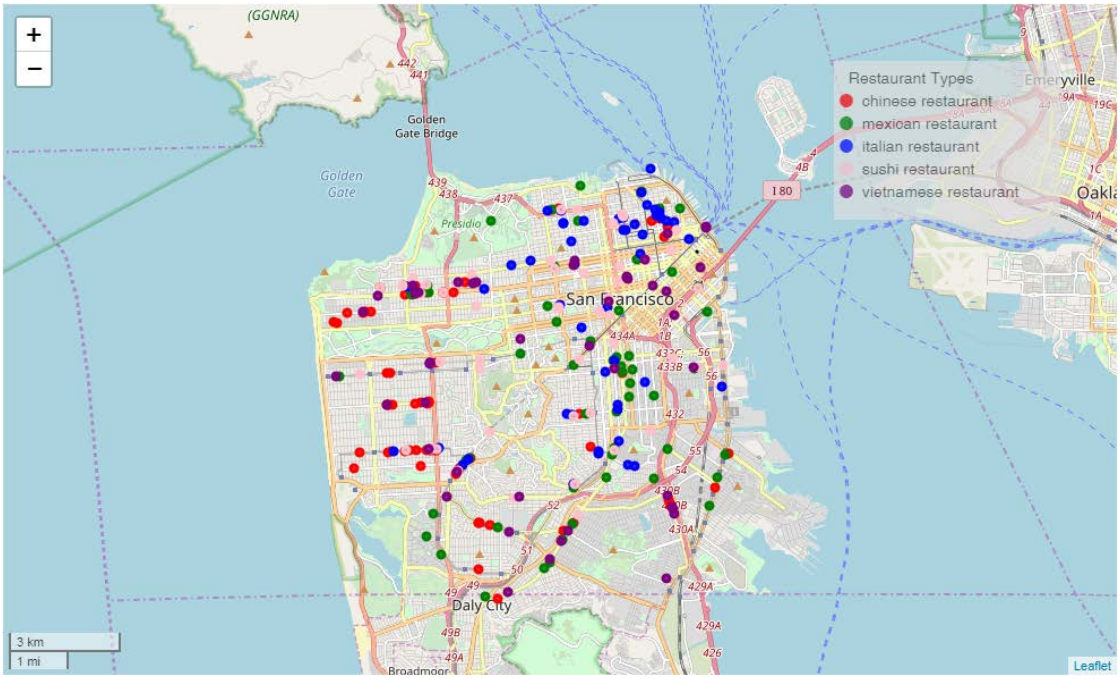
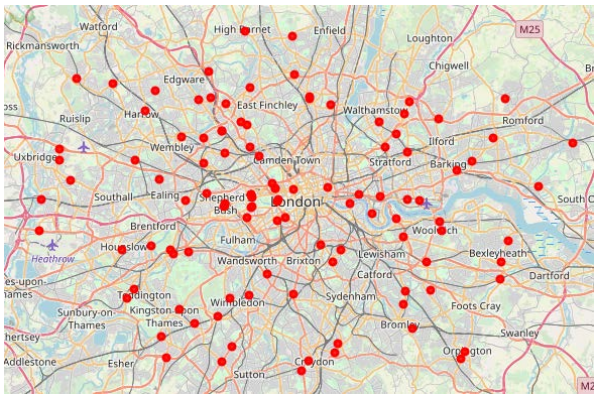
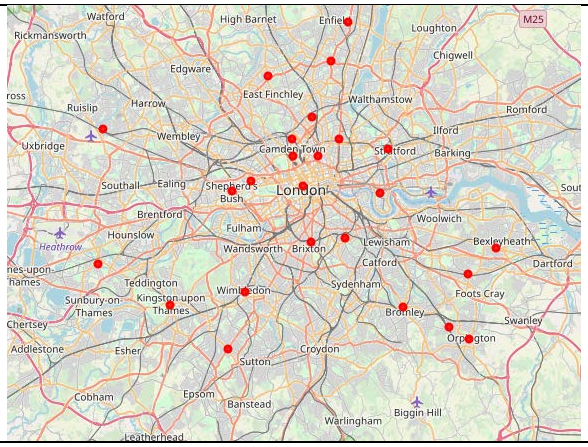
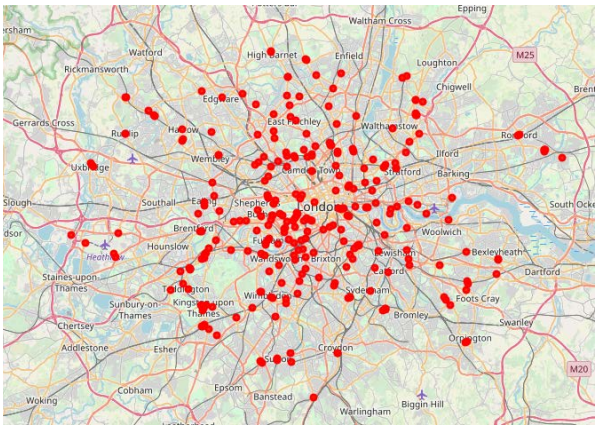
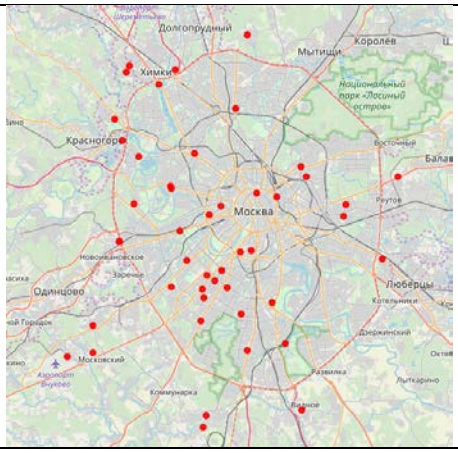


Figure 4

As a comparison, the Chinese, Mexican and Italian restaurants are shown in the following figures.

Table 1

	London	Moscow
Chinese		

Mexican		There is no Mexican restaurants in Moscow according to the data in this project.
Italian		

The apparent lack of Chinese restaurants and absence of Mexican restaurants in Moscow has its historical reasons. Moscow has only 0.03% / 3222 Chinese people as of 2010, and almost no Mexican people. So there are no Mexican restaurants. In consideration of the inaccuracy of this project there are actually a few mexcian restaurants according to google maps. However indeed there is not a lot (only 19). Italian restaurants are popular in every country, so the result is not surprising.

4.1.3 Drinking places

Drinking places are also important but less diverse compared to the case of dining places. The top five drinking places in each city is

- **London** - Pub, coffee shop, bar, gastropub, cocktail bar
- **Moscow** - Coffee shop, wine shop, bar, beer bar, hookah bar
- **San Francisco** - Coffee shop, cocktail bar, wine bar, bar, liquor store

Coffee is import in western world, from this aspect all cities has really high ratio of coffee shops. However Moscow has relatively fewer coffee shops because out of its ~4k venues only ~150 are coffee shops. Different bars and pubs in all cities are similar, but London and San Francisco has more in common.

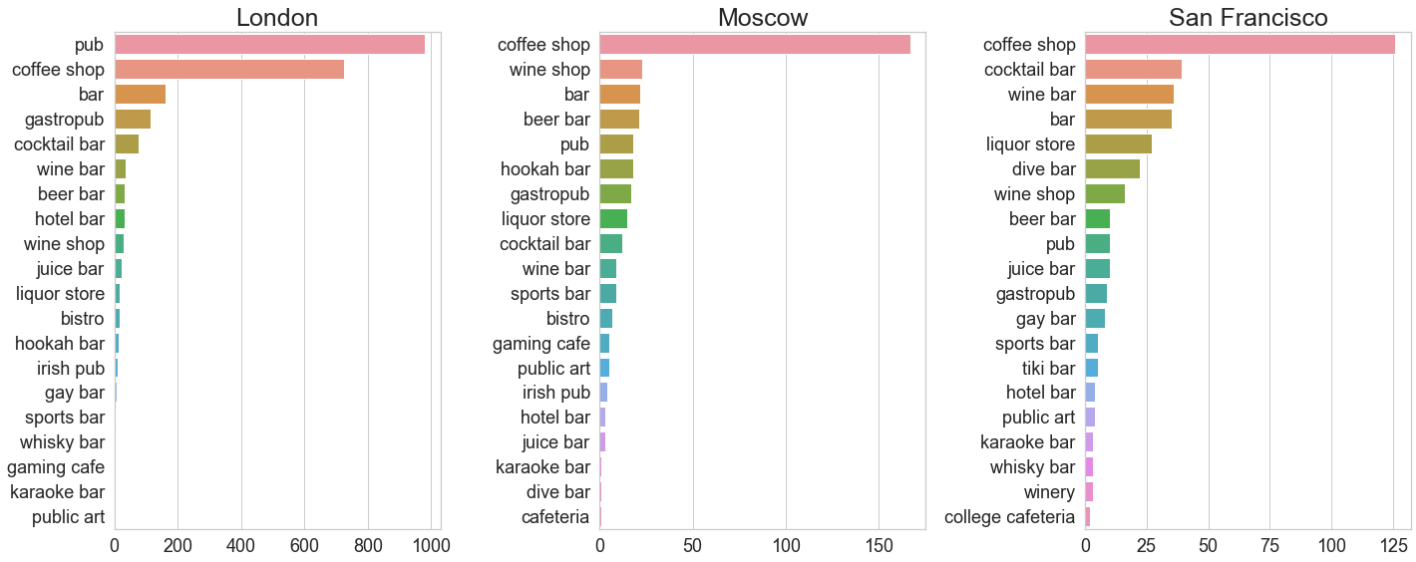


Figure 5

4.1.4 Entertainments

Entertainments in all cities are also very popular, as can be imagined. Parks are the most welcome entertainments, however different cities has its preference. The top five entertainments in the cities are,

- **London** - Park, theater, art gallery, movie theater, history museum
- **Moscow** - Park, dance studio, theater, atrial arts dojo, arts and crafts store, museum
- **San Francisco** - park, trail, art gallery, theater, music venue

It can be seen that cultural sites are very popular in London, with art gallery and history museum ranking 3rd and 5th place. One observation in San Francisco is that it has trail ranking 2nd place, because the state of California does have a lot of national parks and trails. In Moscow however, dance studio ranks 2nd, which correlates to the fact that they have higher ratio of gyms. Other than that, the three cities have similar ratio of entertainments and distribution of entertainments. Parks, theaters, art galleries, and museums are among some of the most popular ones.

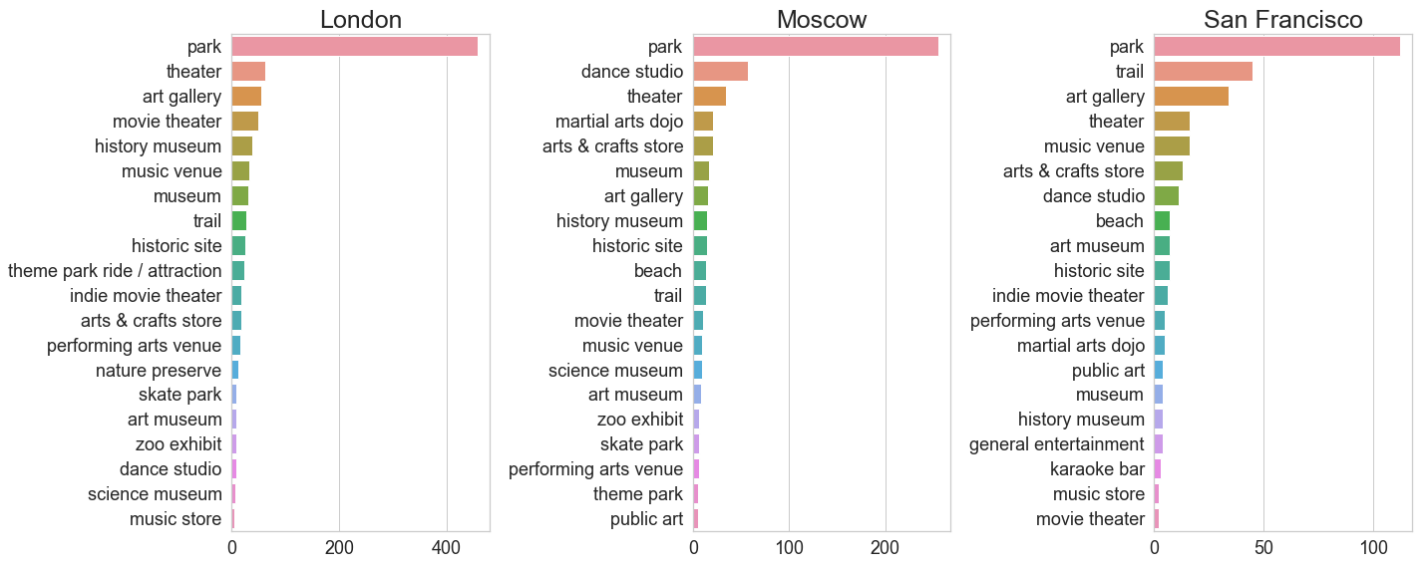


Figure 6

4.1.5 Fitness

Fitness is extremely import in Moscow, so the distribution of gym / fitness related venues are also explored in the three cities, the top five types of workout places are.

- **London** - Gym / fitness center, gym, sporting goods shop, soccer stadium, athletics and sports
- **Moscow** - Gym / fitness center, gym, sporting goods shop, pool, skating rink
- **San Francisco** - Gym, gym / fitness center, baseball field, sporting goods shop, sports bar

It can be seen that even though Moscow only has ~4k total venues, compared to London's ~12k total venues, it has relatively the same amount of fitness related venues. London has its characteristic fitness place, soccer stadium, and Moscow has skating rink ranking 5th in the distribution, which is due to the cultural and climate difference. San Francisco has baseball field in the 3rd place apparently due to the popularity of baseball in the United States.

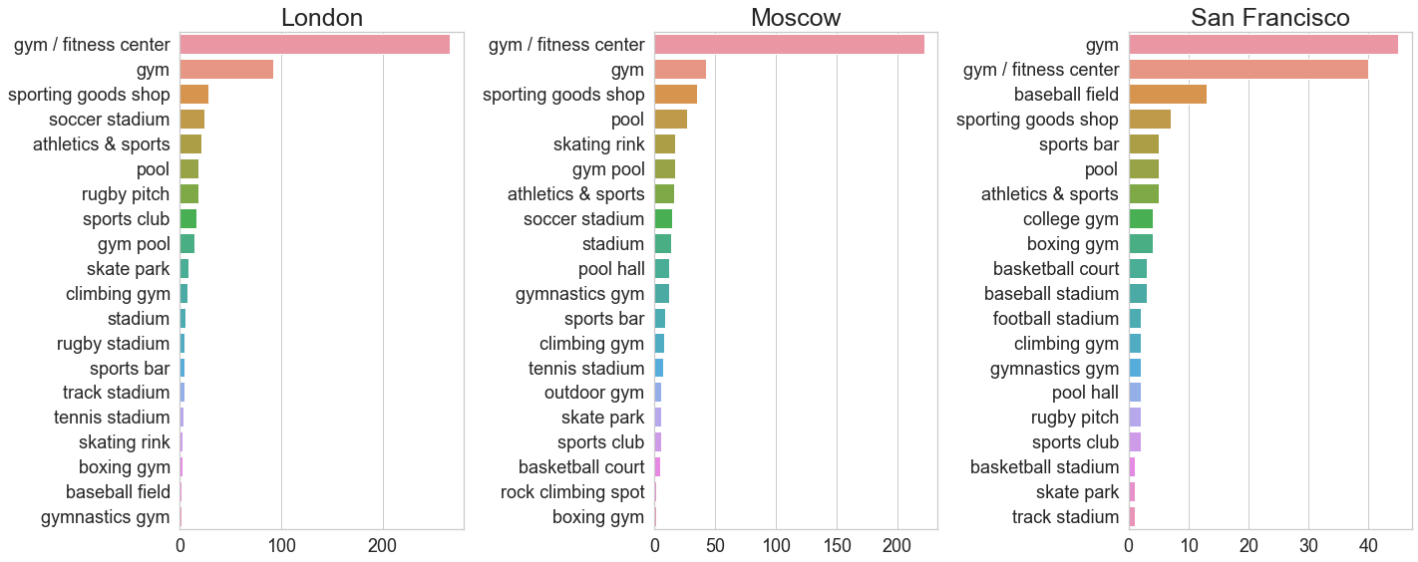


Figure 7

4.2 Clustering Analysis

The clusters of Italian restaurants, coffee shops and parks in London are analyzed, in Figure 8, 9, 10. It can be seen that the Italian restaurants are mainly distributed in the city centers, however the coffee shops and parks are distributed everywhere. Italian restaurants often have higher price compared to a coffee shop, so the distribution is reasonable.

Also coffee shops have really high density in the downtown area, but parks have a relatively uniform distributed density. Parks also cover a larger area compared to Italian restaurants or coffee shops.

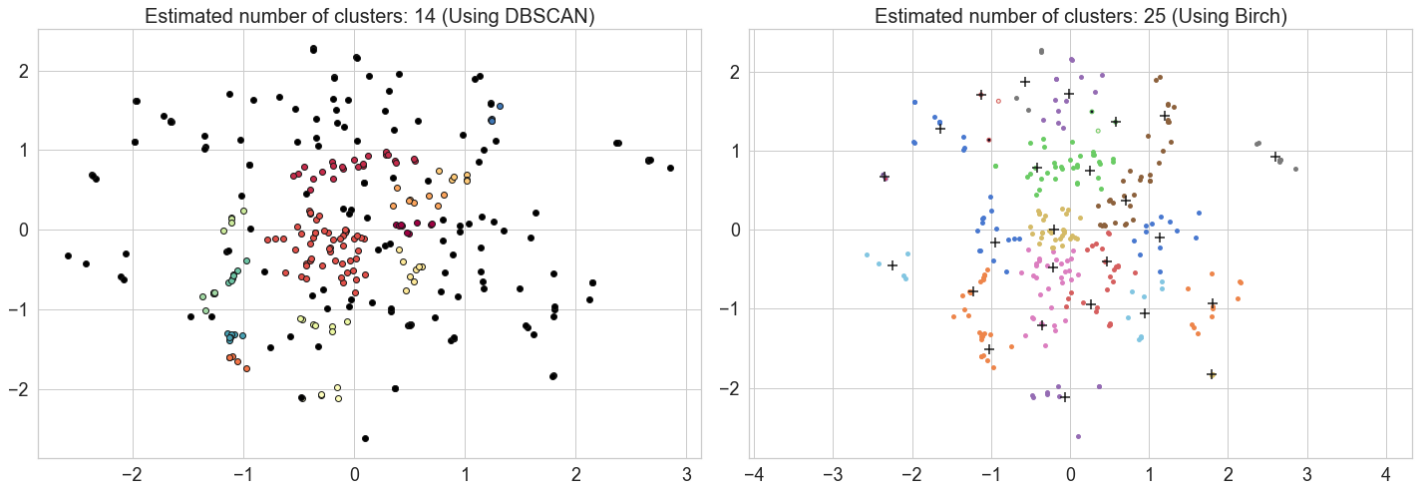


Figure 8: Italian Restaurants in London

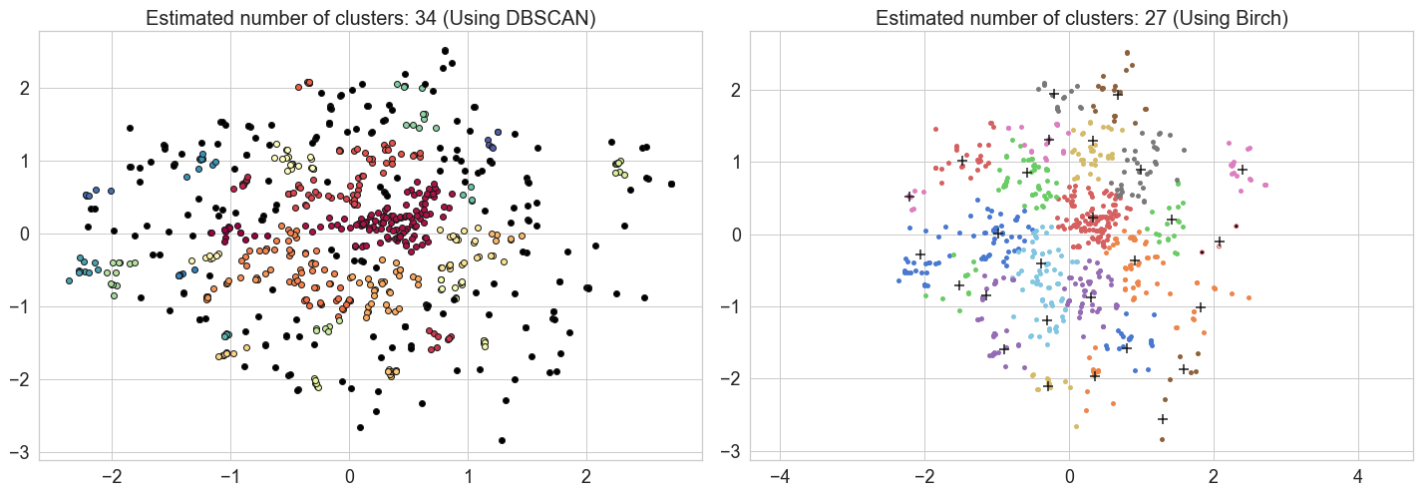


Figure 9: Coffee Shops in London

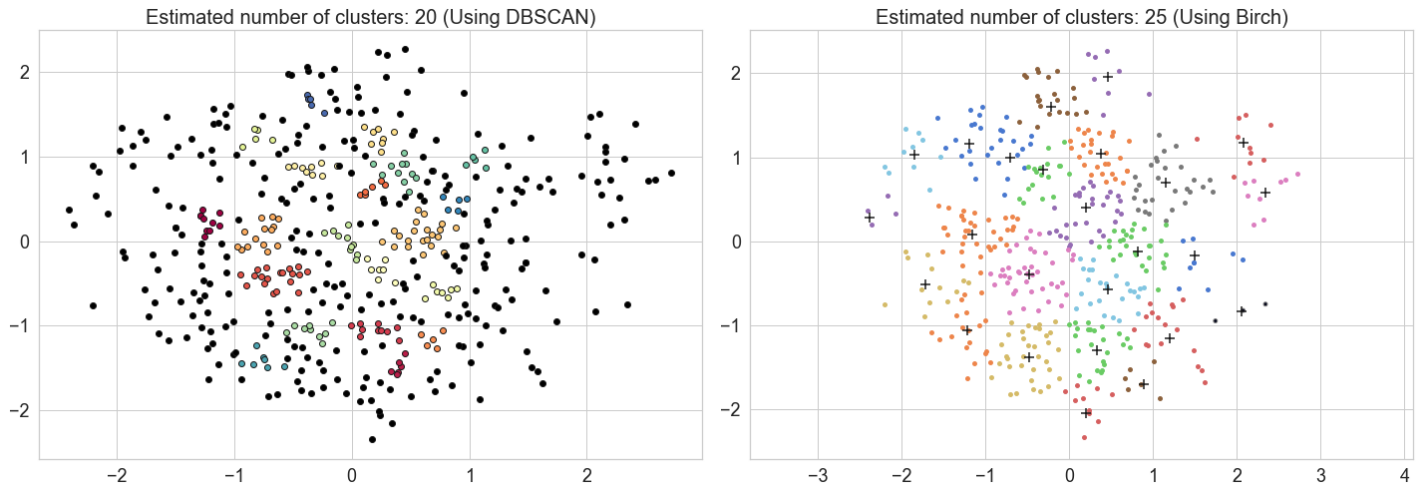


Figure 10: Parks in London

5. Discussion

From above results we can see, the distribution of restaurants in London is more similar to San Francisco compared to Moscow. There are historical and cultural reasons. Also London has a more diverse racial distribution compared to Moscow. For travelling in a city, restaurants are definitely important. From this aspect London should be a better choice compared to Moscow if the tourist has travelled in San Francisco already.

Drinking places and entertainments has similar distribution in these three cities, with a slightly different emphasis. London people showed their love for coffee and drinks, while Moscow people have a higher preference on fitness and gyms. Gyms are not an important aspect when it comes to travelling, however if one decides to live in Moscow, it might be a good and healthy choice.

From the above analysis, we can see the distribution of dining places, drinking places and entertainments are actually related to the culture, climate and racial distribution.

6. Conclusion

Through a thorough analysis of venues in all three cities, I recommend London for people who have travelled in San Francisco before. The similar analysis can be done to other cities. Further quantity related analysis can be done, for example number of restaurants around metro stations and hotels.