Capstone Project - The Battle of Neighborhoods

COMPARISON BETWEEN LONDON, MOSCOW AND SAN FRANCISCO

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.

- However travelling abroad is always very expensive and time consuming for most people. The flights and travel expenses are themselves huge amount of money.

- In this project, the city of San Francisco is considered as a city Jamie (or a group of friends) has traveled.

Data and Cleansing

- The raw data of the project is scraped using beatifulsoup from the Wikipedia.
- The formatted dataframe looks like this

	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

Data and Cleansing(cont)

- 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.

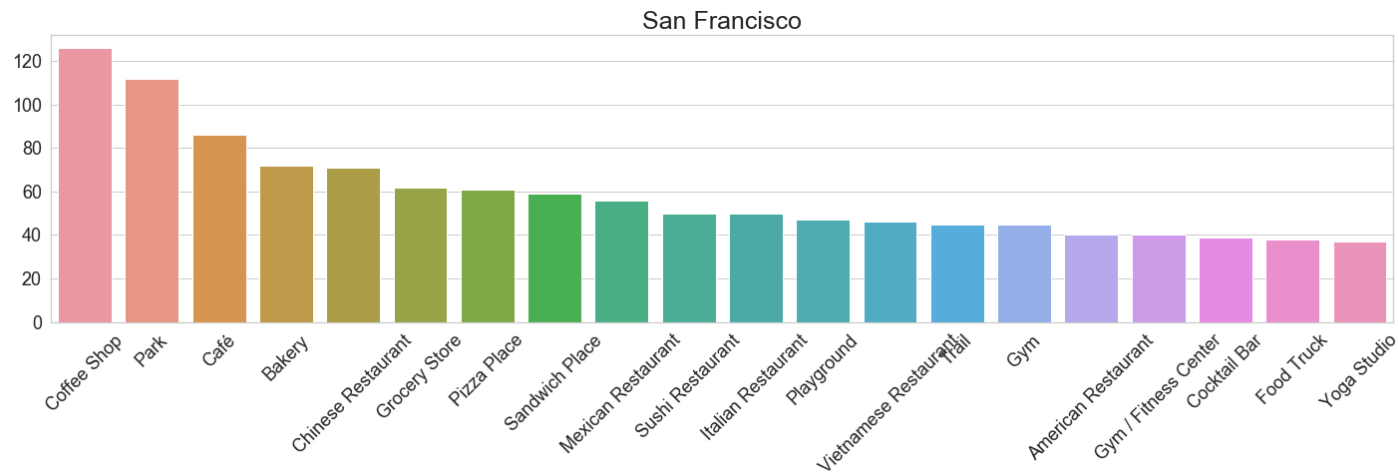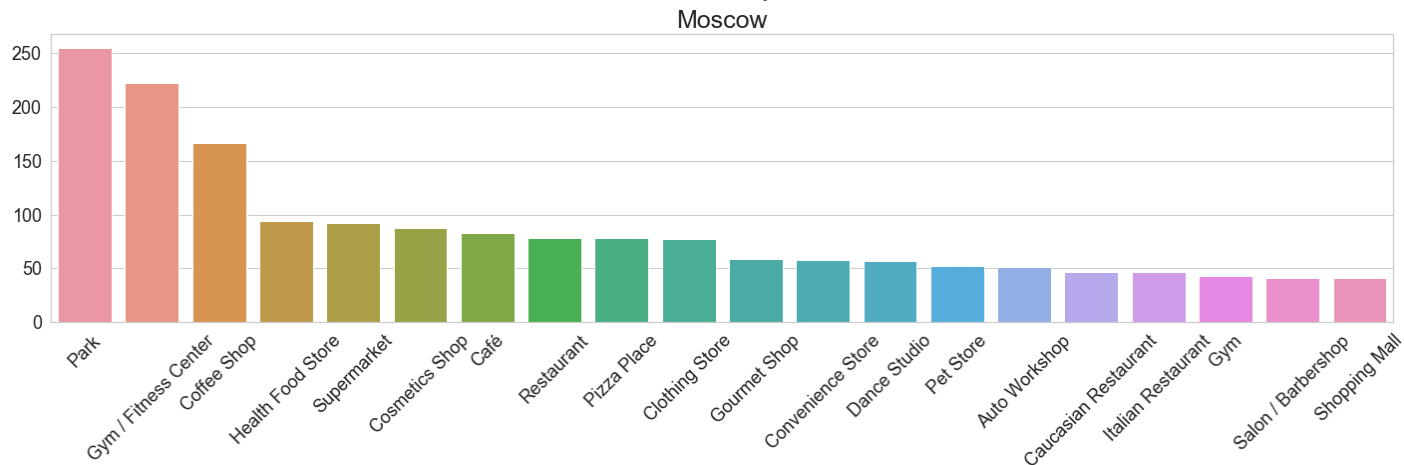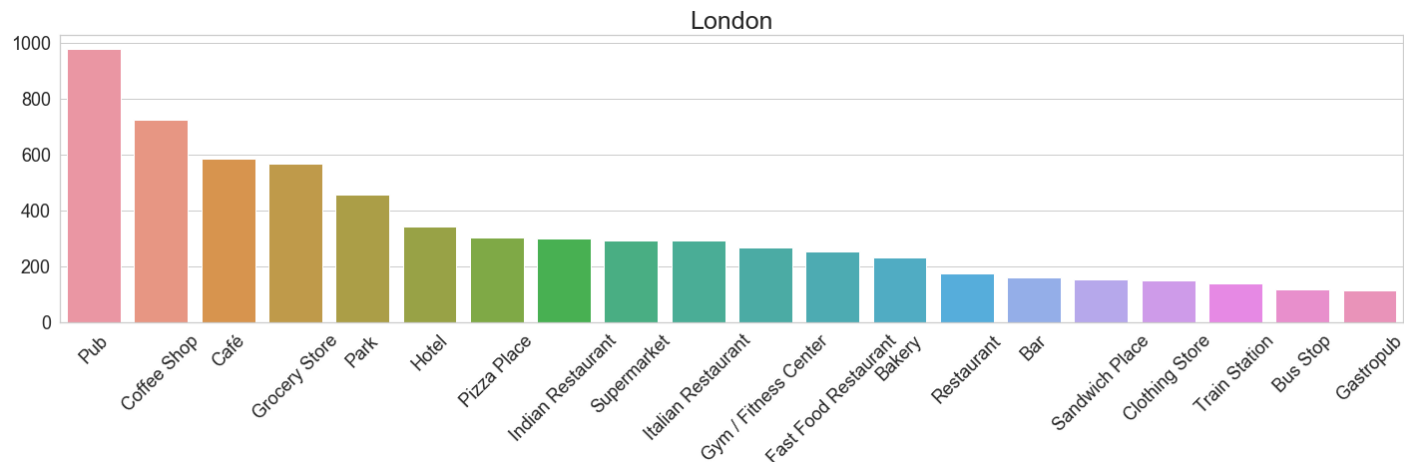|	name	categories	lat	lng
39	Brill	Coffee Shop	51.525767	-0.109477
18	PizzaExpress	Pizza Place	51.406065	0.016451
73	Tariro Fairtrade Coffee House	Coffee Shop	51.401712	-0.195487
38	Jun Ming Xuan	Chinese Restaurant	51.595409	-0.242935
24	The London Borough of Barking & Dagenham Stadium	Soccer Stadium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

Methodology

- In this project we will use be comparing the similarities between three cities, mainly between San Francisco and London, San Francisco and Moscow.

- For each city the venues will be divided into several categories.

- Then for the venues in London, a clustering analysis will be performed to see how they are distributed in the city.

- DBSCAN and Birch clustering algorithms are used because they don't need the number of clusters to begin with.

Analysis and Results : Exploratory Data Analysis

- ## General Categories
  - Each of these three cities has their corresponding popular venue types. As can be seen from Figure, 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).
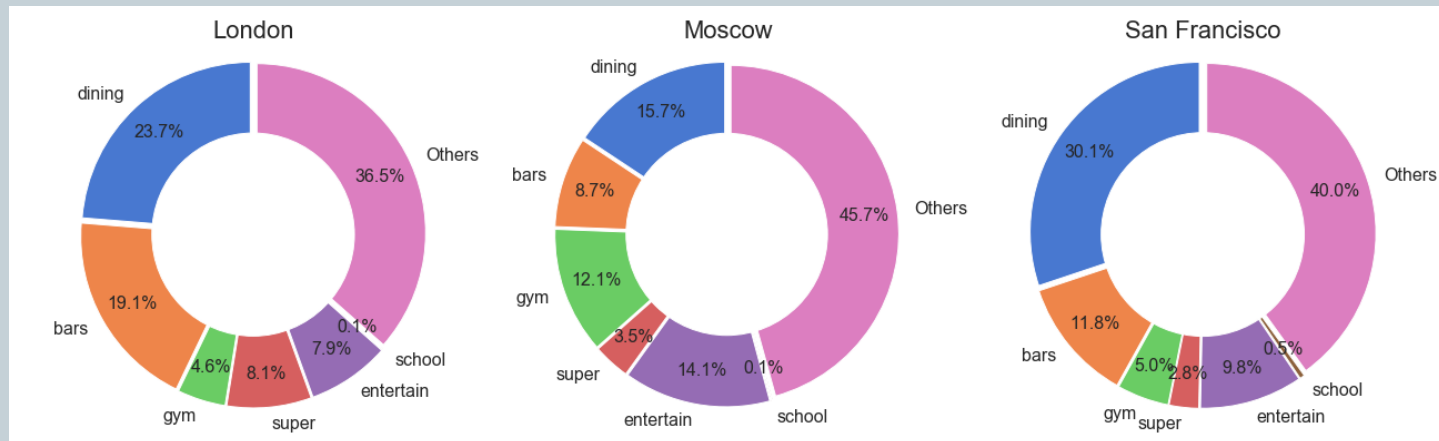
London

Pub, Coffee Shop, Café, Grocery Store, Park, Hotel, Pizza Place, Indian Restaurant, Supermarket, Italian Restaurant, Gym / Fitness Center, Fast Food Restaurant, Bakery, Restaurant, Bar, Sandwich Place, Clothing Store, Train Station, Bus Stop, Gastropub

Moscow

Park, Gym / Fitness Center, Coffee Shop, Health Food Store, Supermarket, Cosmetics Shop, Café, Restaurant, Pizza Place, Clothing Store, Gourmet Shop, Convenience Store, Dance Studio, Pet Store, Auto Workshop, Caucasian Restaurant, Italian Restaurant, Gym, Salon / Barbershop, Shopping Mall

San Francisco

Coffee Shop, Park, Café, Bakery, Chinese Restaurant, Grocery Store, Pizza Place, Sandwich Place, Mexican Restaurant, Sushi Restaurant, Italian Restaurant, Playground, Vietnamese Restaurant, Trail, Gym, American Restaurant, Gym / Fitness Center, Cocktail Bar, Food Truck, Yoga Studio

Analysis and Results : Exploratory Data Analysis

- ## General Categories
  - The distribution of these six types is plotted in Figure. 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.
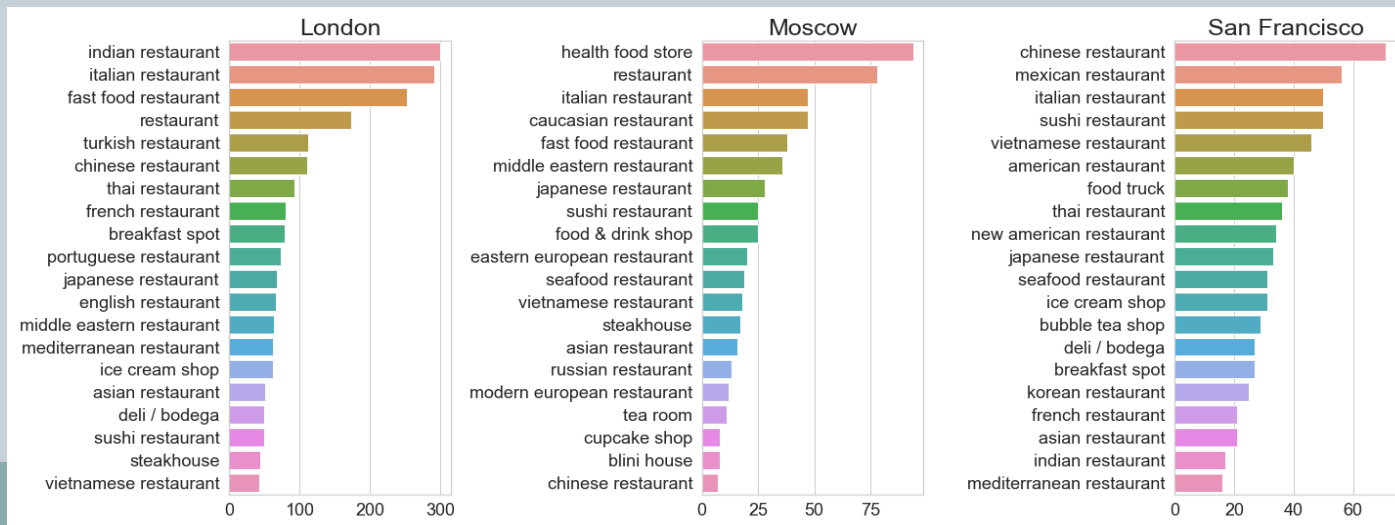
- **Dining places**
  - Dining places are definitely the most important aspect of a city, of which the ranking of top 20 restaurants are shown in Figure for each city. The top 5 restaurants in the cities are
    - **London** - Indian, Italian, Fast Food, Turkish, Chinese
    - **Moscow** - Health Food, Italian, Caucasian, Fast Food
    - **San Francisco** - Chinese, Mexican, Italian, Sushi, Vietnamese
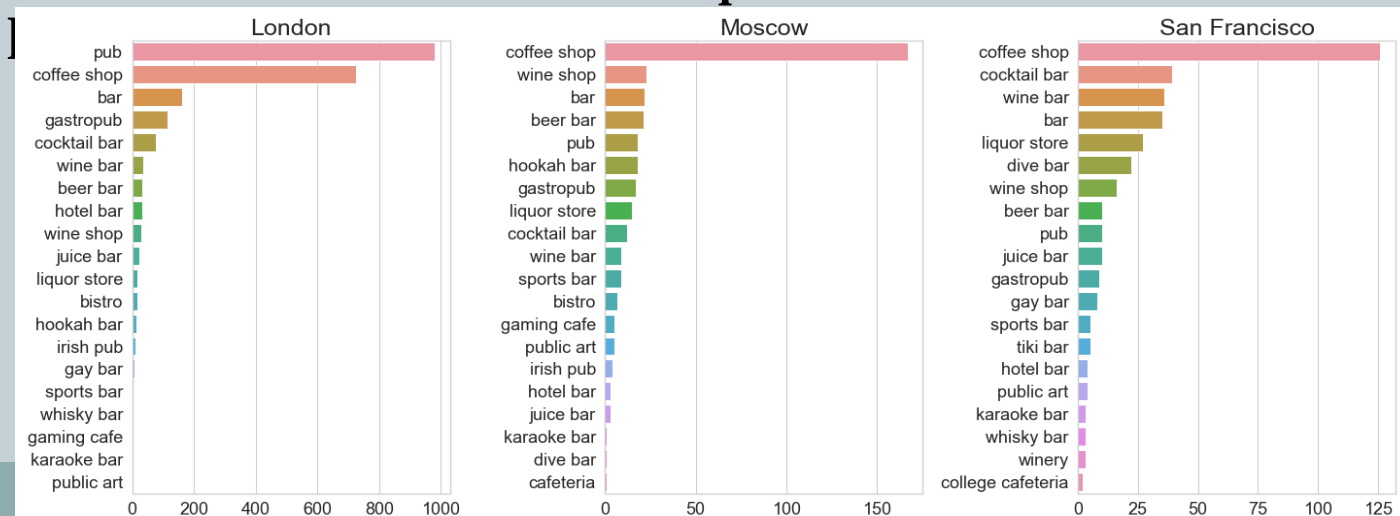
Analysis and Results : Exploratory Data Analysis

- **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
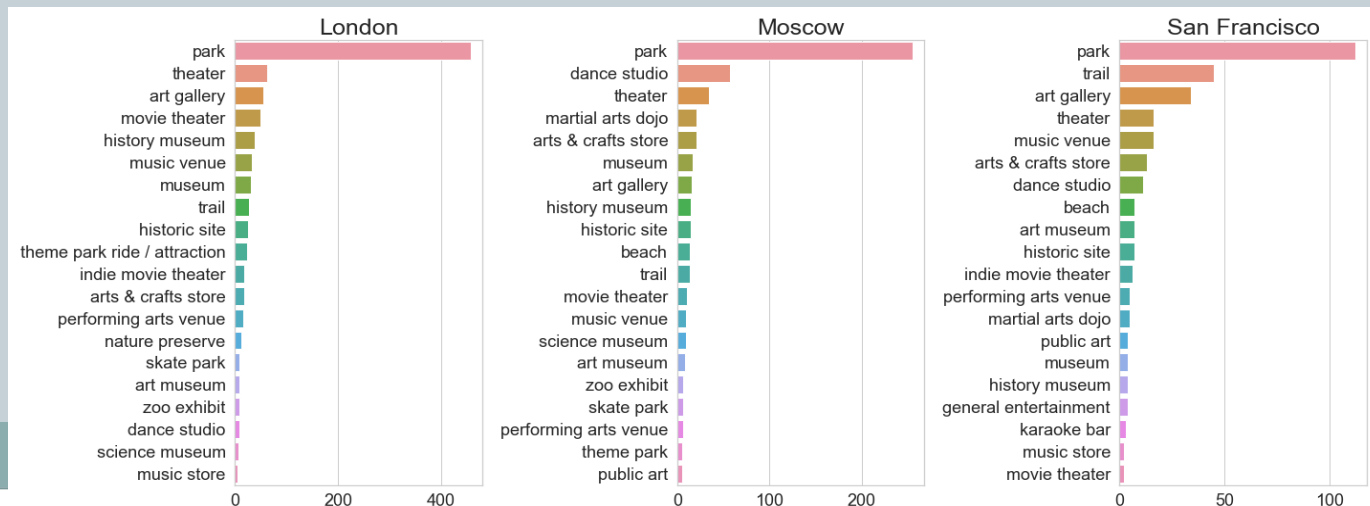   - **San Francisco** - Coffee shop, cocktail bar, wine bar, bar,

Analysis and Results : Exploratory Data Analysis

- **Entertainments**
  - 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
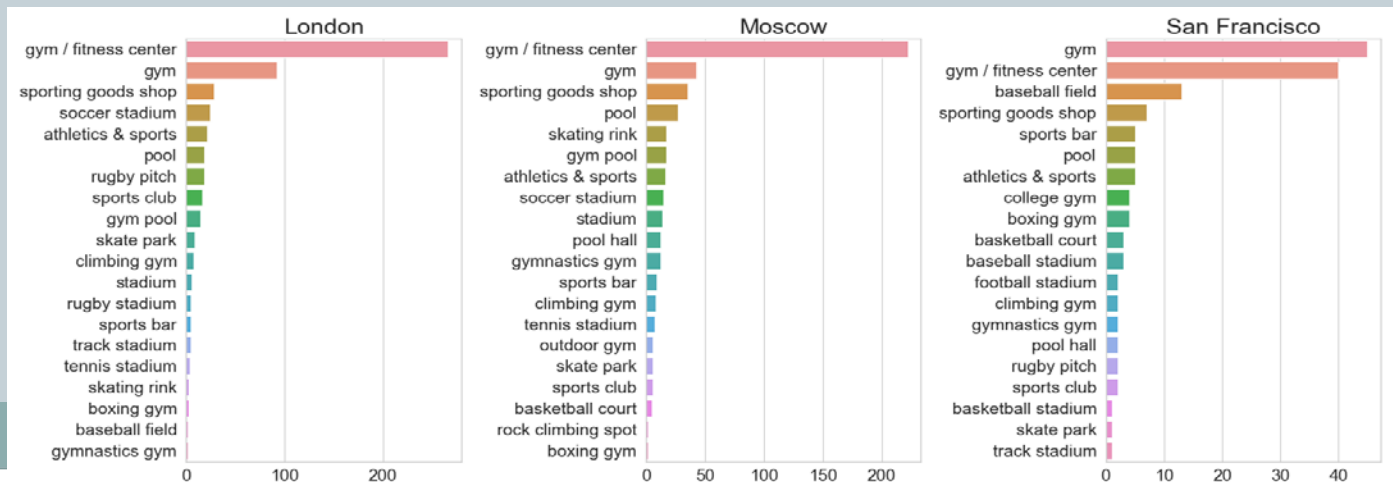   - **San Francisco** - park, trail, art gallery, theater, music venue

Analysis and Results : Exploratory Data Analysis

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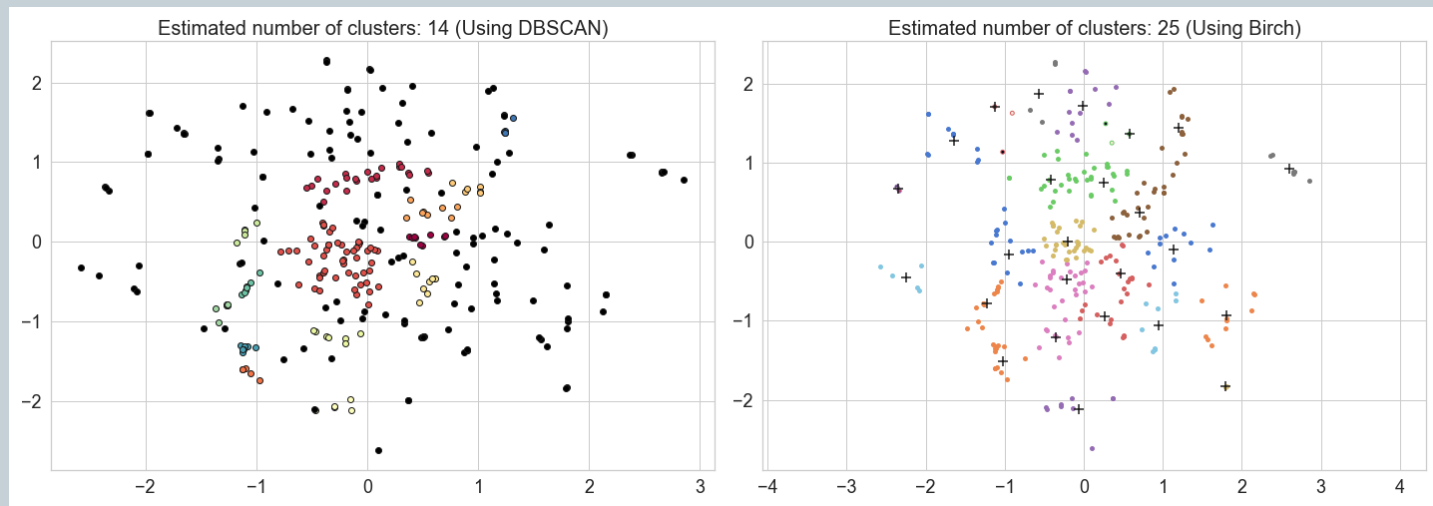   - **London** - Gym / fitness center, gym, sporting goods shop, soccer stadium, athletics and sports
    - **Moscow** - Gym / fitness center, gym, sporting goods shop, pool, skating rink
    - **San Francisco** - Gym, gym / fitness center, baseball field, sporting goods shop, sports bar
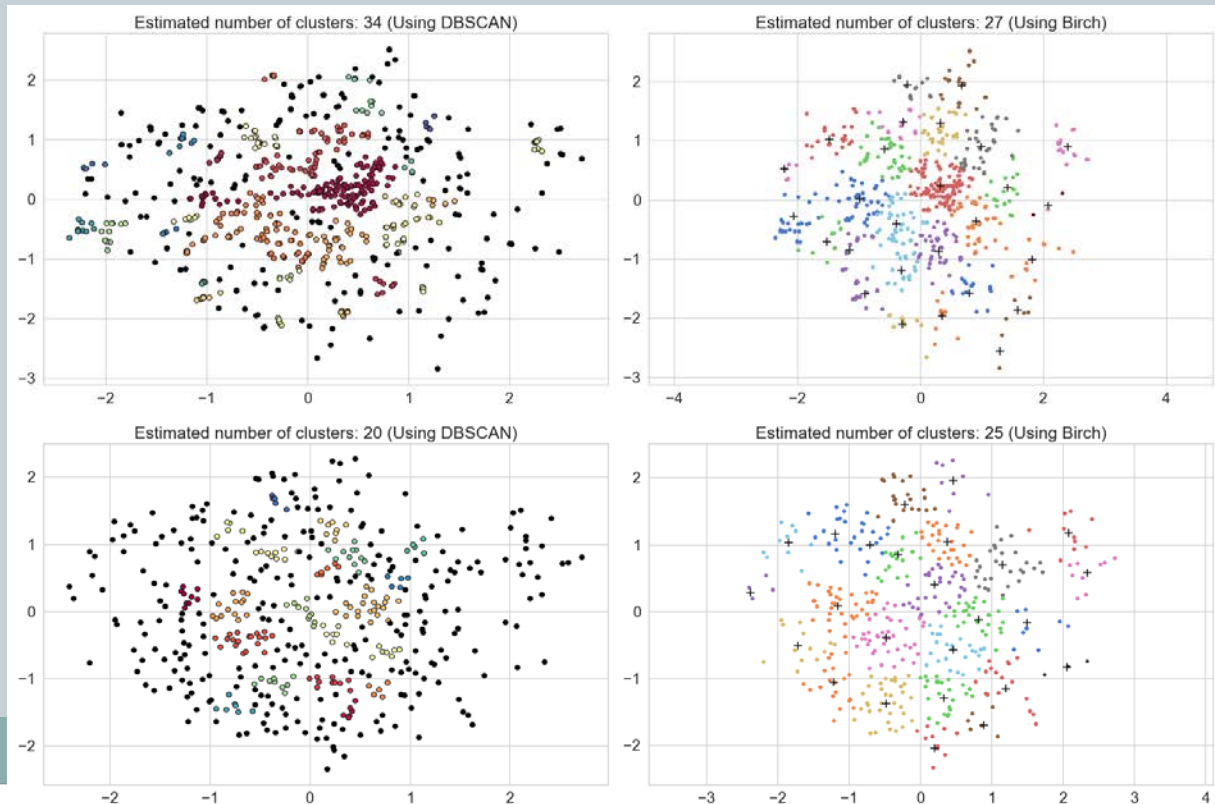
Analysis and Results : Clustering Analysis

- It can be seen that the Italian restaurants are mainly distributed in the city centers.

Estimated number of clusters: 14 (Using DBSCAN)   Estimated number of clusters: 25 (Using Birch)

Analysis and Results : Clustering Analysis

- 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.

Discussion

- The distribution of restaurants in London is more similar to San Francisco compared to Moscow.

- 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.

Discussion(cont)

- London people showed their love for coffee and drinks, while Moscow people have a higher preference on fitness and gyms.

- Drinking places and entertainments has similar distribution in these three cities.

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.