Popular Cities Battle

# Introduction

Tourists always like to visit new cities and explore different cultures. Finding a new city to visit is not an easy task since a wrong decision will cost tourists money and time. In this project, we will study how similar/dissimilar are the largest cities in the world. The results will help tourist choose a relevant city according to their preferences.

# Data

The list of cities is collected from the following Wikipedia page: https://en.wikipedia.org/wiki/List\_of\_largest\_cities.

City locations were queried via the geocoder open street maps API in Python.

# Methodology

The first step was to load the cities table from Wikipedia. Using Pandas, I was able to quickly scrap a list of city names for the project.

The next step was to resample the data to get only 50 of them. This is due to the quota limitation by the Foursquare free account.

Since only city names alone are not enough, I decided to enrich the data by querying for geolocation via the OSM API.

The resulting data set now is a list of city names with their corresponding location. After having the location, I was able to enrich the data by querying the Foursquare API for popular venue in each city.

From this data, I was able to list the top venues in each city and their corresponding frequency as shown below:

----Bangkok----

venue freq

0 River 0.25

1 Buffet 0.25

2 Food Truck 0.25

3 Boat or Ferry 0.25

4 Pastry Shop 0.00

----Belo Horizonte----

venue freq

0 Bar 0.11

1 Restaurant 0.06

2 Brazilian Restaurant 0.06

3 Gym / Fitness Center 0.05

4 Café 0.04

----Berlin----

venue freq

0 Hotel 0.10

1 German Restaurant 0.05

2 History Museum 0.04

3 Theater 0.04

4 Café 0.03

----Brisbane----

venue freq

0 Café 0.15

1 Coffee Shop 0.07

2 Hotel 0.06

3 Bar 0.05

4 Burger Joint 0.04

----Bucharest----

venue freq

0 Coffee Shop 0.10

1 Hotel 0.06

2 Café 0.04

3 Gastropub 0.04

4 Bookstore 0.04

----Budapest----

venue freq

0 Hotel 0.10

1 Coffee Shop 0.07

2 Restaurant 0.07

3 Plaza 0.06

4 Hungarian Restaurant 0.05

----Campinas----

venue freq

0 Bar 0.09

1 Brazilian Restaurant 0.07

2 Pizza Place 0.06

3 Vegetarian / Vegan Restaurant 0.04

4 Hotel 0.04

----Caracas----

venue freq

0 Pharmacy 0.13

1 Plaza 0.09

2 Theater 0.07

3 Café 0.07

4 Bakery 0.05

----Casablanca----

venue freq

0 Café 0.14

1 Hotel 0.13

2 Fast Food Restaurant 0.06

3 French Restaurant 0.06

4 Hotel Bar 0.04

----Chicago----

venue freq

0 Coffee Shop 0.05

1 Hotel 0.04

2 Italian Restaurant 0.03

3 Garden 0.03

4 Pizza Place 0.03

----Daegu----

venue freq

0 Coffee Shop 0.21

1 Korean Restaurant 0.08

2 Café 0.05

3 Chinese Restaurant 0.05

4 Multiplex 0.05

----Daejeon----

venue freq

0 Coffee Shop 0.24

1 Korean Restaurant 0.09

2 Fast Food Restaurant 0.06

3 Bakery 0.05

4 Multiplex 0.05

----Dakar----

venue freq

0 African Restaurant 0.25

1 Plaza 0.25

2 Park 0.25

3 Market 0.25

4 Pedestrian Plaza 0.00

----Dalian----

venue freq

0 Hotel 0.21

1 Coffee Shop 0.12

2 Department Store 0.09

3 Shopping Mall 0.07

4 Plaza 0.05

----Dar es Salaam----

venue freq

0 Hotel 0.31

1 BBQ Joint 0.08

2 American Restaurant 0.08

3 Indian Restaurant 0.08

4 Restaurant 0.04

----Dhaka----

venue freq

0 Asian Restaurant 0.14

1 Fried Chicken Joint 0.09

2 Shopping Mall 0.09

3 Indian Restaurant 0.09

4 Park 0.09

----Ekurhuleni----

venue freq

0 Burger Joint 0.11

1 Fast Food Restaurant 0.11

2 Grocery Store 0.11

3 Department Store 0.11

4 Gym / Fitness Center 0.11

----Hangzhou----

venue freq

0 Hotel 0.13

1 Shopping Mall 0.13

2 Bookstore 0.07

3 Coffee Shop 0.07

4 Convenience Store 0.03

----Ibadan----

venue freq

0 Market 0.33

1 Historic Site 0.33

2 Business Service 0.33

3 Pastry Shop 0.00

4 Pet Store 0.00

----Jaipur----

venue freq

0 Indian Restaurant 0.20

1 Historic Site 0.15

2 Café 0.10

3 Market 0.10

4 Jewelry Store 0.05

----Jeddah----

venue freq

0 Middle Eastern Restaurant 0.12

1 Hookah Bar 0.05

2 Breakfast Spot 0.05

3 Seafood Restaurant 0.05

4 Café 0.05

----Kampala----

venue freq

0 Café 0.13

1 Hotel 0.13

2 Hotel Bar 0.05

3 Casino 0.05

4 Coffee Shop 0.05

----Kharkiv----

venue freq

0 Coffee Shop 0.08

1 Restaurant 0.06

2 Café 0.05

3 Yoga Studio 0.05

4 Park 0.05

----Lahore----

venue freq

0 Pakistani Restaurant 0.15

1 Restaurant 0.05

2 Bakery 0.05

3 Café 0.05

4 Camera Store 0.05

----Lucknow----

venue freq

0 Shopping Mall 0.25

1 Hotel 0.25

2 American Restaurant 0.25

3 Indian Restaurant 0.25

4 Pastry Shop 0.00

----Mandalay----

venue freq

0 Hotel 0.42

1 Bed & Breakfast 0.12

2 Bistro 0.08

3 Burmese Restaurant 0.08

4 Italian Restaurant 0.04

----Maracaibo----

venue freq

0 Hotel 0.29

1 Bakery 0.29

2 Big Box Store 0.14

3 Café 0.14

4 Liquor Store 0.14

----Minsk----

venue freq

0 Cocktail Bar 0.09

1 Bar 0.07

2 Restaurant 0.05

3 Coffee Shop 0.04

4 Park 0.04

----Monterrey----

venue freq

0 Convenience Store 0.07

1 Taco Place 0.07

2 Fast Food Restaurant 0.04

3 Café 0.04

4 Plaza 0.04

----Nagpur----

venue freq

0 Indian Restaurant 0.33

1 Dhaba 0.17

2 Convenience Store 0.17

3 Toll Plaza 0.17

4 Moving Target 0.17

----Nur-Sultan----

venue freq

0 Asian Restaurant 0.14

1 Coffee Shop 0.09

2 Gym 0.05

3 Fast Food Restaurant 0.05

4 Park 0.05

----Paris----

venue freq

0 French Restaurant 0.10

1 Plaza 0.06

2 Ice Cream Shop 0.06

3 Art Gallery 0.04

4 Restaurant 0.03

----Porto Alegre----

venue freq

0 Buffet 0.08

1 Café 0.07

2 Coffee Shop 0.06

3 Brazilian Restaurant 0.04

4 Theater 0.03

----Recife----

venue freq

0 Brazilian Restaurant 0.08

1 Historic Site 0.06

2 Ice Cream Shop 0.04

3 Bookstore 0.04

4 Sandwich Place 0.04

----Rostov-on-Don----

venue freq

0 Coffee Shop 0.08

1 Café 0.07

2 Cocktail Bar 0.05

3 Hotel 0.04

4 Gym 0.03

----San Antonio----

venue freq

0 Hotel 0.16

1 Mexican Restaurant 0.08

2 Seafood Restaurant 0.05

3 Cocktail Bar 0.04

4 Steakhouse 0.04

----Shiraz----

venue freq

0 Historic Site 0.21

1 Persian Restaurant 0.14

2 Hotel 0.14

3 Jewelry Store 0.07

4 Farmers Market 0.07

----Surabaya----

venue freq

0 Food Truck 0.11

1 Indonesian Restaurant 0.09

2 Chinese Restaurant 0.07

3 Convenience Store 0.05

4 Shopping Mall 0.05

----Suwon----

venue freq

0 Korean Restaurant 0.18

1 Coffee Shop 0.14

2 Hotel 0.09

3 Seafood Restaurant 0.07

4 Italian Restaurant 0.05

----Suzhou----

venue freq

0 Jiangsu Restaurant 1.0

1 Photography Lab 0.0

2 Pet Store 0.0

3 Persian Restaurant 0.0

4 Perfume Shop 0.0

----Sydney----

venue freq

0 Café 0.09

1 Australian Restaurant 0.08

2 Scenic Lookout 0.06

3 Hotel 0.04

4 Hotel Bar 0.03

----São Paulo----

venue freq

0 Japanese Restaurant 0.06

1 Brazilian Restaurant 0.06

2 Theater 0.04

3 Gift Shop 0.04

4 Grocery Store 0.04

----Tangshan----

venue freq

0 Shopping Mall 0.33

1 Coffee Shop 0.33

2 Department Store 0.17

3 Hotel 0.17

4 Accessories Store 0.00

----Tijuana----

venue freq

0 Fast Food Restaurant 0.09

1 Grocery Store 0.07

2 Café 0.05

3 Convenience Store 0.05

4 Shopping Mall 0.05

----Visakhapatnam----

venue freq

0 Indian Restaurant 0.21

1 Hotel 0.14

2 Bakery 0.07

3 Mobile Phone Shop 0.07

4 Bookstore 0.07

----Xi'an----

venue freq

0 Shopping Mall 0.50

1 Coffee Shop 0.25

2 Park 0.25

3 Pedestrian Plaza 0.00

4 Pharmacy 0.00

----Xiamen----

venue freq

0 Hotel 0.15

1 Pizza Place 0.15

2 Coffee Shop 0.12

3 Fast Food Restaurant 0.06

4 Café 0.06

----Yaoundé----

venue freq

0 Bakery 0.24

1 Hotel 0.19

2 Nightclub 0.10

3 Bar 0.10

4 Platform 0.05

Now it is time for applying machine learning to cluster the different cities. We applied K-Means algorithm with K=5. The results are shown in the Map below. We notice that some poor areas are in different cluster.



# Conclusion

This study focused on showing the difference between venue types in different cities around the world and grouping them in different cluster. This may help tourists find similar/dissimilar locations for their next visit according to the venue types in each city. We found that famous cities in rich areas are more likely to be found in the same cluster.