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
- 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

- Load the cities table from Wikipedia
- Resample the data to get only 50 of them. (due to the quota limitation by the Foursquare free account)
- Enrich the data by querying for geolocation via the OSM API
- Enrich the data by querying the Foursquare API for popular venue in each city
- List the top venues in each city and their corresponding frequency

Example

----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----

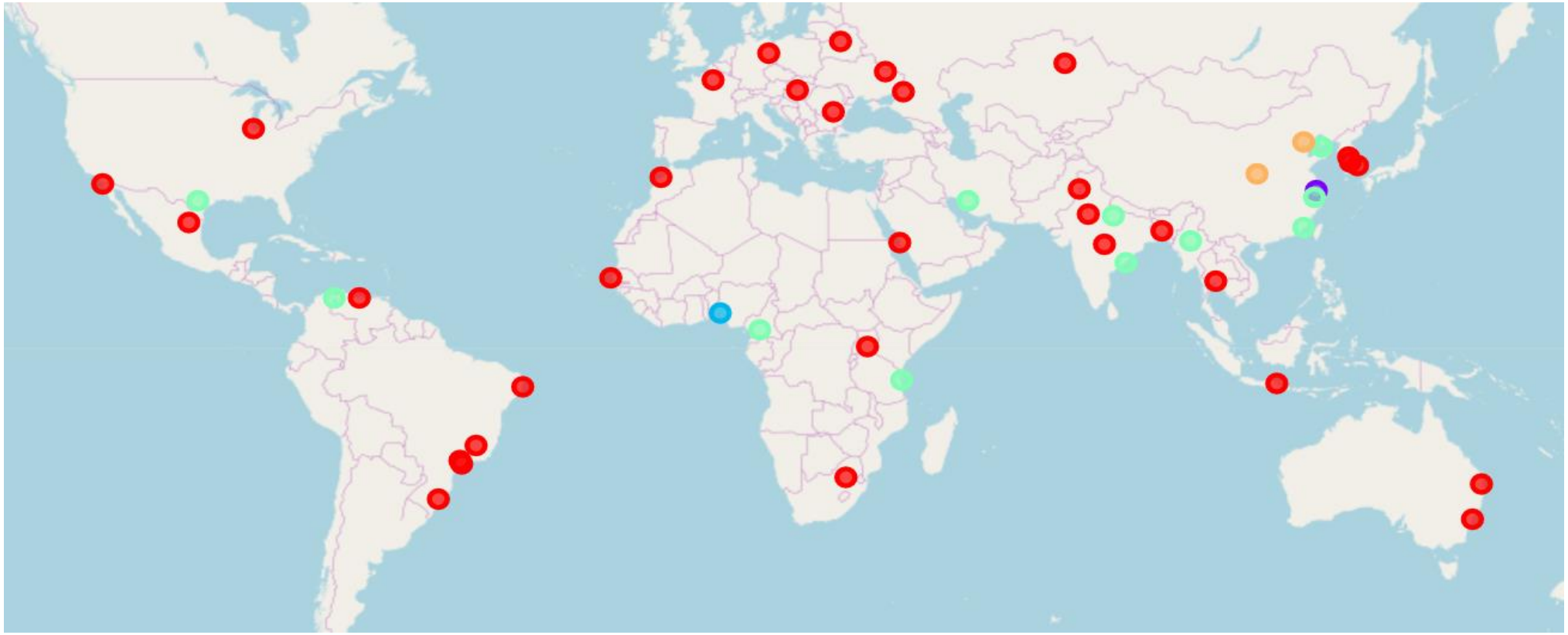
	venue	freq
0	Bar	0.11
1	Restaurant	0.06
2	Brazilian Restaurant	0.06
3	Gym / Fitness Center	0.05
4	Café	0.04

----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

Clustering

- We applied K-Means with k=5



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