
## IBM DATA SCIENCE CAPSTONE PROJECT

# CITIES OF THE WORLD

# INTRODUCTION

You might be a seasoned traveller that's been many places. You're a foodie and love theatres. You've travelled to many cities before and you don't really know where you would like to go next? Wouldn't it be nice to get new travel ideas? To find new cities based on how similar they are to the ones you liked in your previous travels?

Or, you might be a business man that travels around a world to sign deals. Everywhere you go, you need to take your futur clients out and entertain them to talk business in a relaxed setting. Wouldn't it be nice if you could know how similar the next city you're going to go to is to the ones you know already?

Usually, all these travels start with landing in a major city. Often, these are either country capitals or financial centers. The problem we're going to solve is to answer these questions: How is one city similar to the others? What are the most common venues that set these cities appart?

To answer these questions, we're going to compare and cluster capitals and financial centers based on the Foursquare API explore venues feature.

# DATA

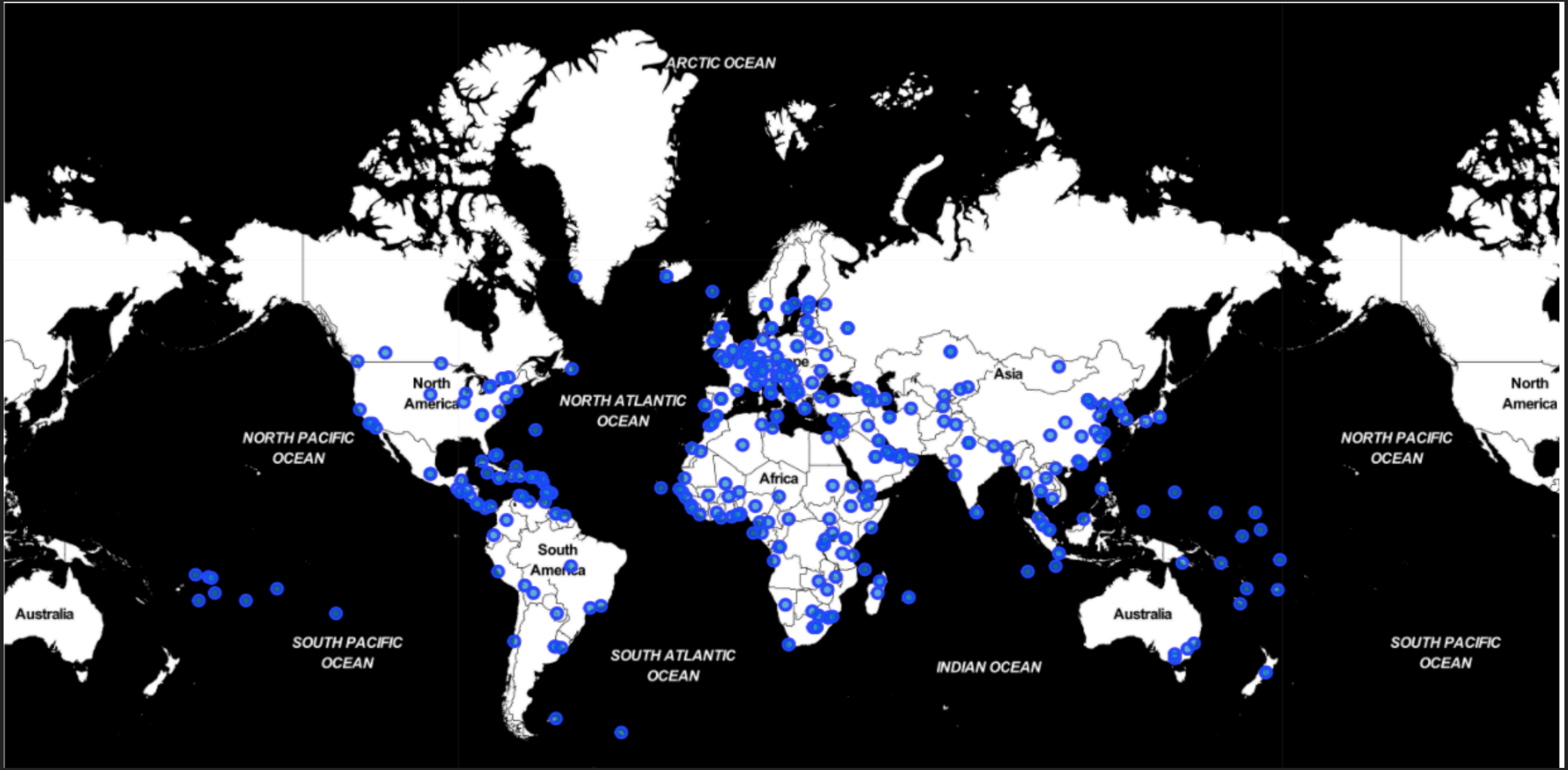
- https://en.wikipedia.org/wiki/List_of_national_capitals
- https://en.wikipedia.org/wiki/Global_Financial_Centres_Index
- geopy.geocoders Nominatim
- Foursquare API

# METHODOLOGY

- Build the dataset
- 

| | City | Latitude | Longitude |
|---|---|---|---|
| **0** | New York City | 40.712728 | -74.006015 |
| **1** | London | 51.507322 | -0.127647 |
| **2** | Shanghai | 31.232276 | 121.469207 |
| **3** | Tokyo | 35.682839 | 139.759455 |
| **4** | Hong Kong | 22.350627 | 114.184916 |

# VISUALIZING THE CITIES



# GETTING THE VENUES FROM FOURSQUARE


```
'venue': {'id': '4c6a79b5d0bdc9b6ba80a80b',
 'name': 'The Ritz-Carlton, Hong Kong (香港麗思卡爾頓酒店)',
 'location': {'address': '1 Austin Road West, International Commerce Centre, Tsim Sha Tsui',
  'lat': 22.3034608,
  'lng': 114.1602184,
  'labeledLatLngs': [{'label': 'display',
    'lat': 22.3034608,
    'lng': 114.1602184}],
  'distance': 5833,
  'cc': 'HK',
  'city': 'West Kowloon',
  'country': '香港',
  'formattedAddress': ['1 Austin Road West, International Commerce Centre, Tsim Sha Tsui',
   '香港']},
 'categories': [{'id': '4bf58dd8d48988d1fa931735',
   'name': 'Hotel',
   'pluralName': 'Hotels',
   'shortName': 'Hotel',
   'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_',
    'suffix': '.png'},
```

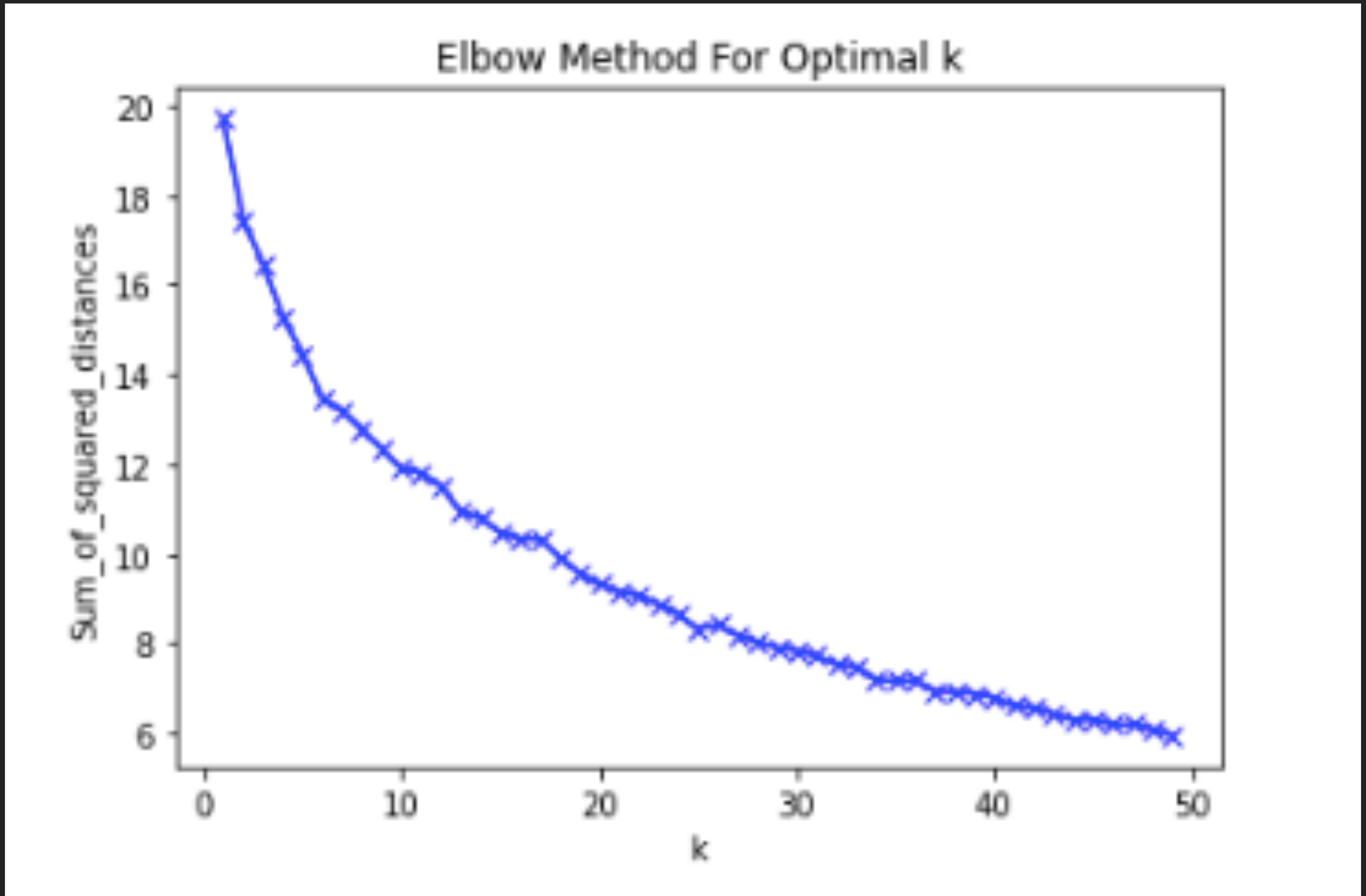
# BUILDING THE VENUES DATA FRAME

| | City | Latitude | Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | New York City | 40.712728 | -74.006015 | Aire Ancient Baths | 40.718141 | -74.004941 | Spa |
| 1 | New York City | 40.712728 | -74.006015 | 9/11 Memorial North Pool | 40.712077 | -74.013187 | Memorial Site |
| 2 | New York City | 40.712728 | -74.006015 | Crown Shy | 40.706187 | -74.007490 | Restaurant |
| 3 | New York City | 40.712728 | -74.006015 | Los Tacos No. 1 | 40.714267 | -74.008756 | Taco Place |
| 4 | New York City | 40.712728 | -74.006015 | The Rooftop @ Pier 17 | 40.705463 | -74.001598 | Music Venue |

# SORTING VENUES FROM MOST TO LEAST COMMON

| | City | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | Abidjan | African Restaurant | Nightclub | Shopping Mall | Hotel | Ice Cream Shop |
| 1 | Abu Dhabi | Hotel | Resort | Beach | Shopping Mall | Food Truck |
| 2 | Abuja | Fast Food Restaurant | Shopping Mall | Hotel | Bed & Breakfast | Restaurant |
| 3 | Accra | Hotel | Shopping Mall | Cocktail Bar | Pizza Place | African Restaurant |
| 4 | Adamstown | Nature Preserve | Zoo | Food Stand | Food Court | Food & Drink Shop |

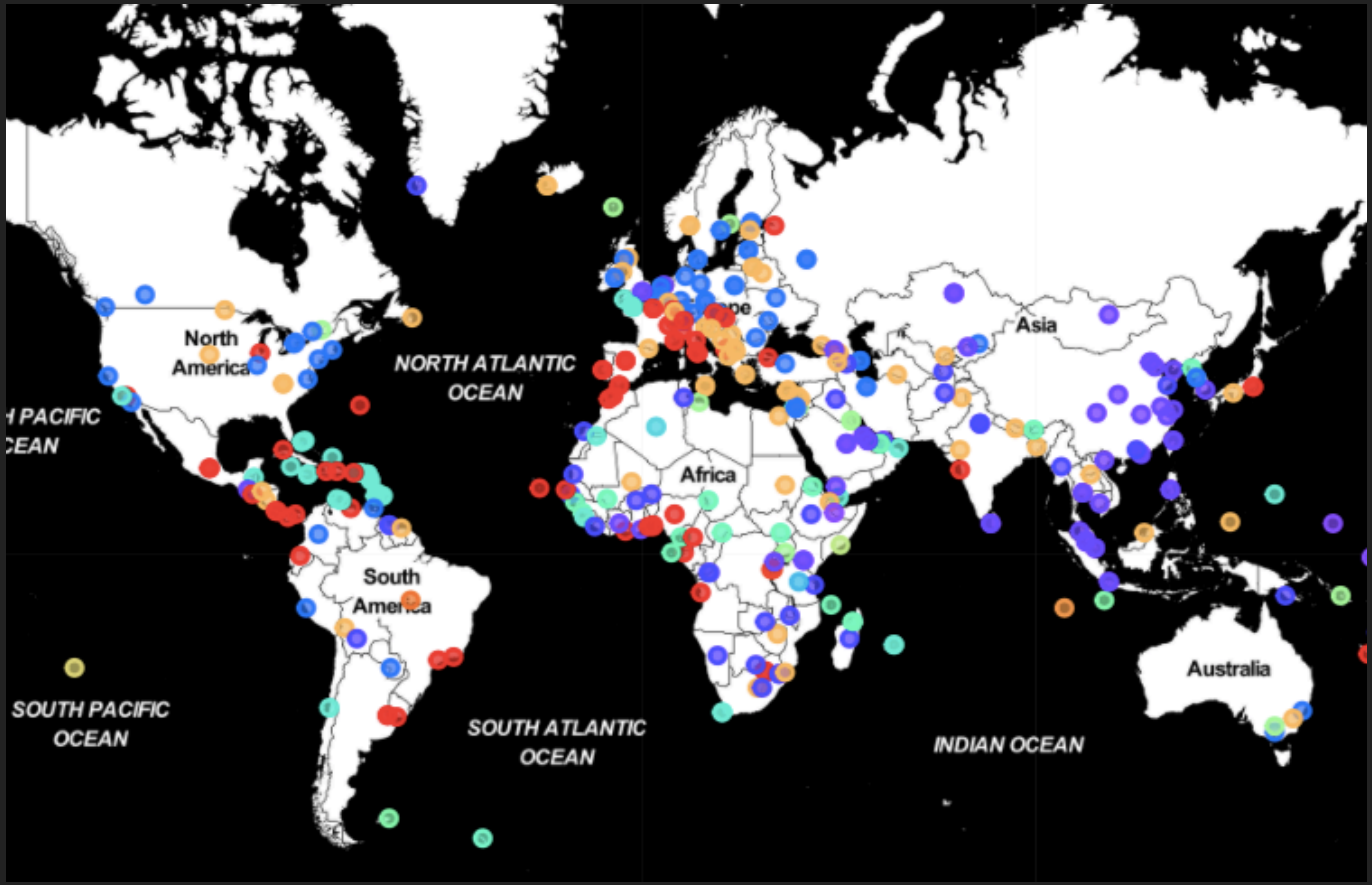
# USING K MEANS – FINDING THE BEST K WITH THE ELBOW METHOD



# CLUSTERING THE CITIES WITH K MEANS

| | City | Latitude | Longitude | Cluster Labels |
|---|---|---|---|---|
| **0** | New York City | 40.712728 | -74.006015 | 4 |
| **1** | London | 51.507322 | -0.127647 | 2 |
| **2** | Shanghai | 31.232276 | 121.469207 | 2 |
| **3** | Tokyo | 35.682839 | 139.759455 | 19 |
| **4** | Hong Kong | 22.350627 | 114.184916 | 2 |

# RESULTS



North America
NORTH ATLANTIC OCEAN
Asia
Africa
South America
SOUTH PACIFIC OCEAN
SOUTH ATLANTIC OCEAN
INDIAN OCEAN
Australia

# CLUSTER DETAILS (CLUSTER 8)

| | City | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6 |
|---|---|---|---|---|---|---|---|
| 44 | Tel Aviv | Beach | Café | Park | Ice Cream Shop | Coffee Shop | |
| 54 | Jersey | Beach | Hotel | Historic Site | Restaurant | French Restaurant | |
| 62 | Mauritius | Resort | Beach | Hotel | Café | Park | |
| 66 | Cape Town | Beach | Hotel | Coffee Shop | Café | Trail | |
| 68 | Bahamas | Beach | Resort | Hotel | Coffee Shop | Ice Cream Shop | |
| 71 | British Virgin Islands | Beach | American Restaurant | Bar | Caribbean Restaurant | Hotel | |
| 77 | Cayman Islands | Caribbean Restaurant | Hotel | Resort | Italian Restaurant | Beach | |
| 78 | Guernsey | Beach | Hotel | Restaurant | French Restaurant | Café | S |
| 85 | Barbados | Caribbean Restaurant | Beach | Hotel | Café | Shopping Mall | |
| 118 | Aden | Hotel | Surf Spot | Resort | Airport | Airport Food Court | |

# DISCUSSION

The results show that cities can be grouped together by looking at their venues. At first glance, we can see that very large cities or cities from countries with similar cultural backgrounds seem to have a tendency to group together.

If we look closer into each cluster, we can see that some cities are outliers and belong to their own small clusters. This seems to happen for smaller cities with a lower number of venues. Also, some cities have coordinates the put them in wrong places. Further work would be needed to correct these. For now, looking at both the map and the cluster details is sufficient to work around this problem.

By analyzing the clusters, we can find useful information about cities that might be similar in experience. For example, cluster 4 have a group of cities where parks are the most common venue and cluster 12 has cities with a lot of cafés. Cluster 8 consists of a group of cities where beaches are the most common.

# CONCLUSION

The questions we wanted to answer were the following. How is one city similar to the others? What are the most common venues that set these cities appart? In the discussion section, we set forward that the current city clustering analysis was able to answer these questions.

If you are a seasoned globetrotter, by looking at the data, you might decide you want to explore the idea of a trip to Tel Aviv if you like beaches. Maybe you would'nt have thought of going to a city that large and expect to find this kind of venue.

If on the other hand, you are a traveling business man, you might plan your meetings better by knowing that Singapore, where youhave never been, is similar to London that you know very well.