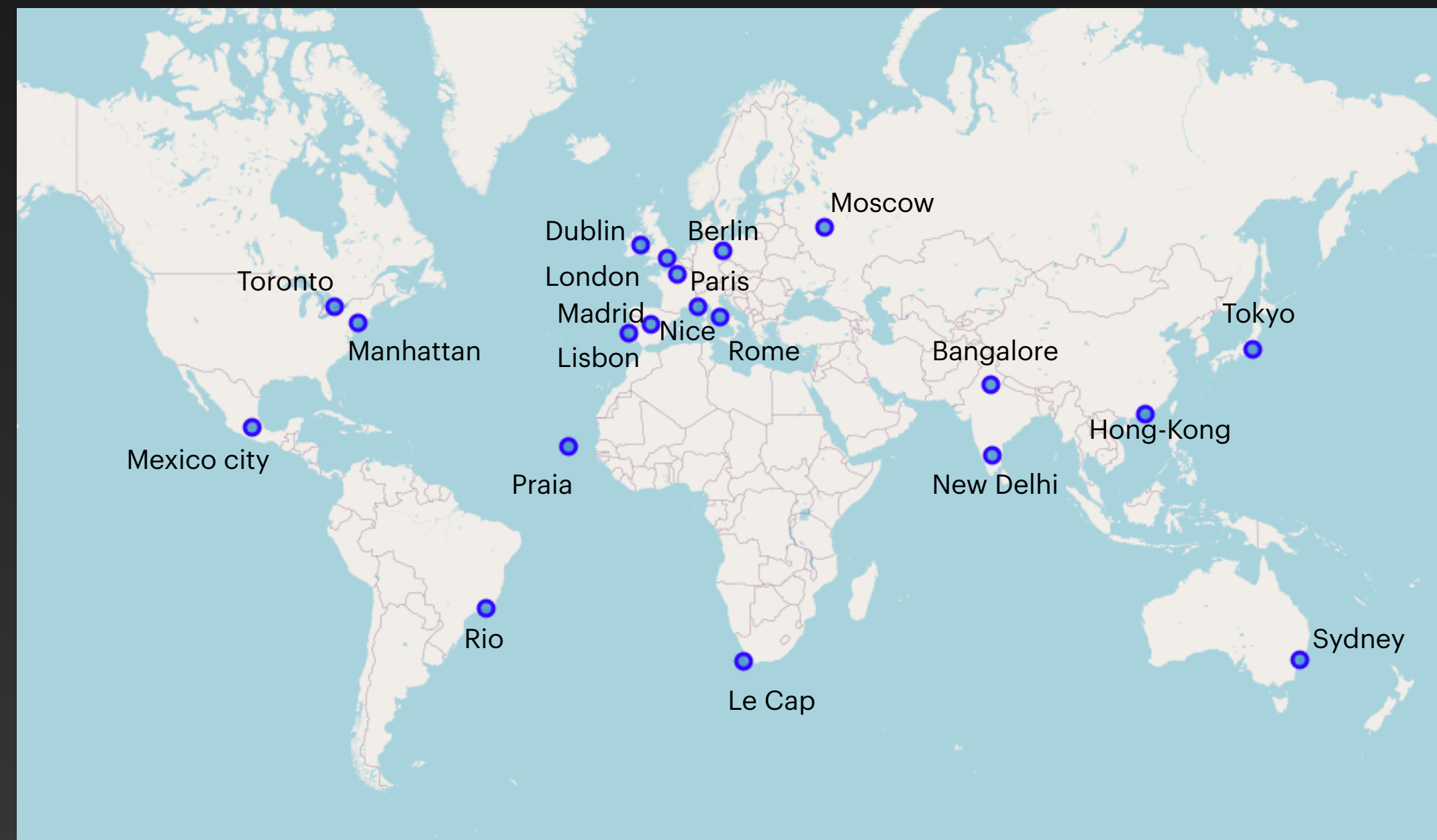# Exploration around the world

Junior Fernandes - 16 Oct 2020

Business problem and interest

- The goal is to compare several cities by knowing their most common or frequent places nearby.

- This project can be applied to several interests

  - A person leaving his city for another for professional reasons

  - A person going in vacations and don't have informations about other cities

  - A person being curious about similarity between several well-known cities

Data used

- The Foursquare API will be used to know the different categories of venues for each city in the map below
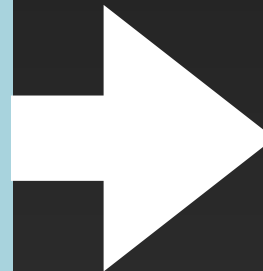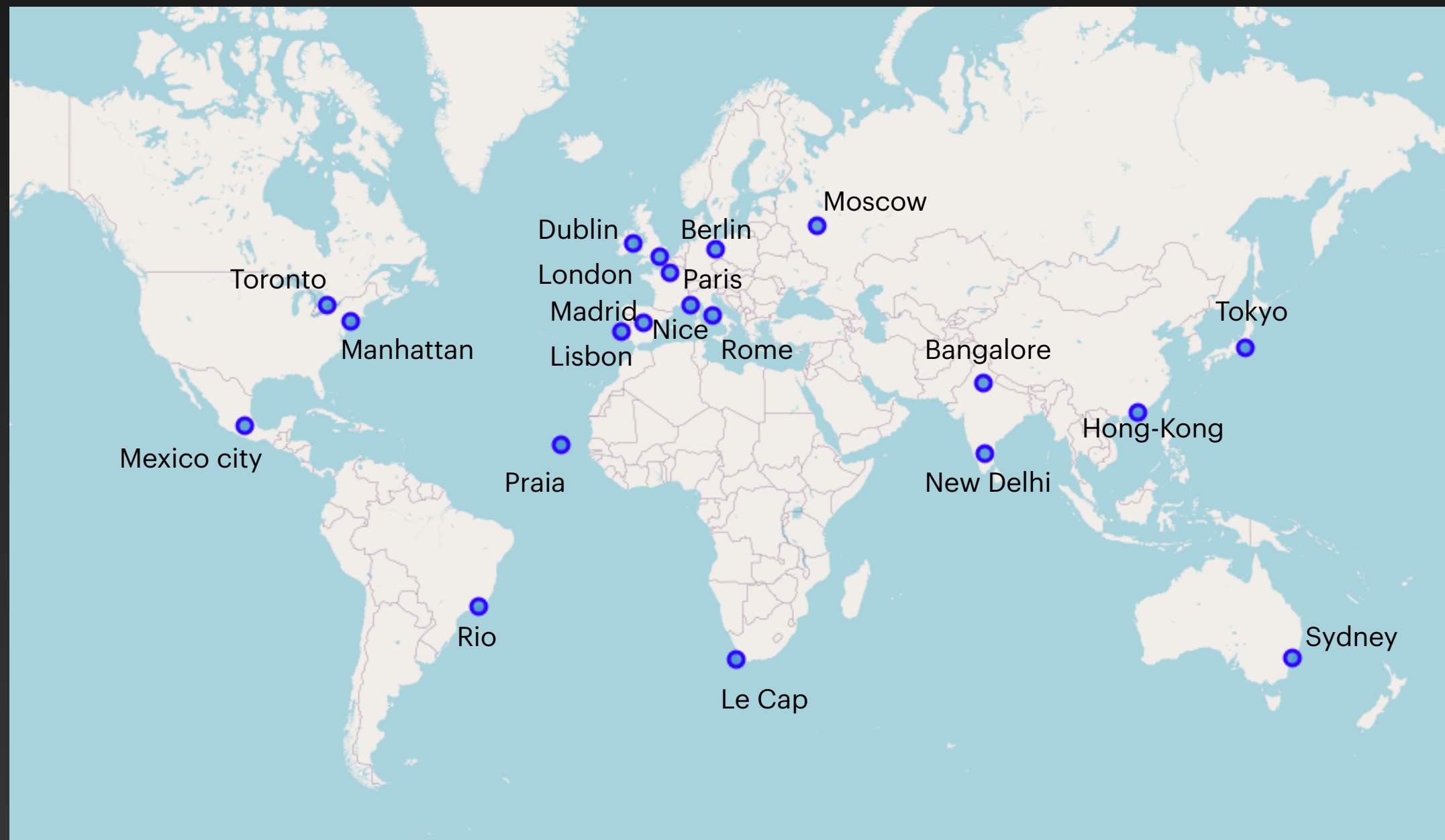
Methodology

- Find for each city the top 20 and most present categories of venues

- Example below with the top 5 categories of venues for Bangalore, Le Cap, Berlin, Lisbon, Dublin and London

- The 'freq' represents the frequency of apparition of each category

```
----Bangalore, India----              ----Le Cap, South Africa----
               venue  freq                           venue  freq
0    Capitol Building  0.25            0               Café  0.08
1              Hotel  0.25            1   Italian Restaurant  0.08
2            Vineyard  0.25            2                Pub  0.06
3               Park  0.25            3             Museum  0.06
4             Palace  0.00            4         Restaurant  0.04


----Berlin, Germany----               ----Lisbon, Portugal----
               venue  freq                           venue  freq
0              Hotel  0.12            0   Portuguese Restaurant  0.18
1            Wine Bar  0.06            1                Bar  0.10
2          Restaurant  0.04            2              Plaza  0.08
3       Clothing Store  0.04            3             Hostel  0.06
4               Plaza  0.04            4       Ice Cream Shop  0.04


----Dublin, Ireland----               ----London, England----
               venue  freq                           venue  freq
0         Coffee Shop  0.12            0              Hotel  0.12
1                Pub  0.12            1            Theater  0.08
2       Clothing Store  0.08            2        Art Gallery  0.06
3           Bookstore  0.06            3             Garden  0.06
4               Café  0.06            4              Plaza  0.06
```

Results

- Cluster 1 (in purple) seem to have common recurrent category of venues which make them similar and thus on the same cluster.

- Other clusters with only one city = these cities have uncommon places and different cultures which make them isolated on another cluster.

Improvements

- Use Neighbourhoods's of each city to find the most common categories of venues

- Use meteorological data to find out cities with similar/different climates