A description of the problem and a discussion of the background.

Main idea came form a question:

‘I liked this city. Where should I go to be satisfied as much as then?’   
This question regards all the places on the world. Having a tool that helps to answer this question one would know where to go for another vacations. Such tool would be also very welcome by every Travel Agency.

This project however, is focused on capital cities similarity. This is just due the number of data that should be collected and computed if it focused on every travel destination.

Main assumption is that appropriate clustering method together with well suited data-set shall combine similar cities in the same cluster.

A description of the data and how it will be used to solve the problem.

First of all a data with all (or almost all) capitals names and its geographical location shall be collected.

I will use data taken from this particular web page:  
'<http://techslides.com/list-of-countries-and-capitals>'  
Data set contains information like:

* Country Name
* Capital Name
* Capital Latitude
* Capital Longitude
* Country Code
* Continent Name

Another set of data will be the Foursquare responses for each geographical location exploration query:

url = '<https://api.foursquare.com/v2/venues/explore?&client_id>={}&client\_secret={}&v={}&ll={}, {}&radius={}&limit={}'.format(CLIENT\_ID, CLIENT\_SECRET, VERSION, lat, lng, radius, LIMIT)

Having this, data will be extended by results of each venue exploration:

url = 'https://api.foursquare.com/v2/venues/{}?&client\_id={}&client\_secret={}&v={}'.format(venue\_id, CLIENT\_ID, CLIENT\_SECRET, VERSION)

Another data-set I would like to incorporate into my project is List of countries by GDP (PPP) per capita taken from: <https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita>

I will focus on World Bank (2019) data set.

And the final one is List of countries by life expectancy taken from: <https://en.wikipedia.org/wiki/List_of_countries_by_life_expectancy>

I will focus on Countries and regions by life expectancy at birth in 2018 (2019 report) and in particular the Overall column.

The final data-set would contain all the below:  
Capital Name

Capital Latitude

Capital Longitude

Venue Name

Venue Category

Venue Rating

Venue Number of Tips

With such prepared data-set I will try to cluster the capitals in a way to get clusters of the cities with similar attractiveness.