Capstone Project Report

* 1. ***escription of the data and how it will be used to solve the problem***

***Description of the data***

The following data sets will be used:

* + 1. List of national capitals from the Wikipedia page:

Sample snapshot of the data:



For the complete dataset, data please refer to:

<https://en.wikipedia.org/wiki/List_of_national_capitals>

* + 1. Geo-Location data of each national capital from geocoding web services.



for technical information, please refer to <https://geopy.readthedocs.io/en/stable/>.

* + 1. National capitals important venues from Foursquare API.

Example:

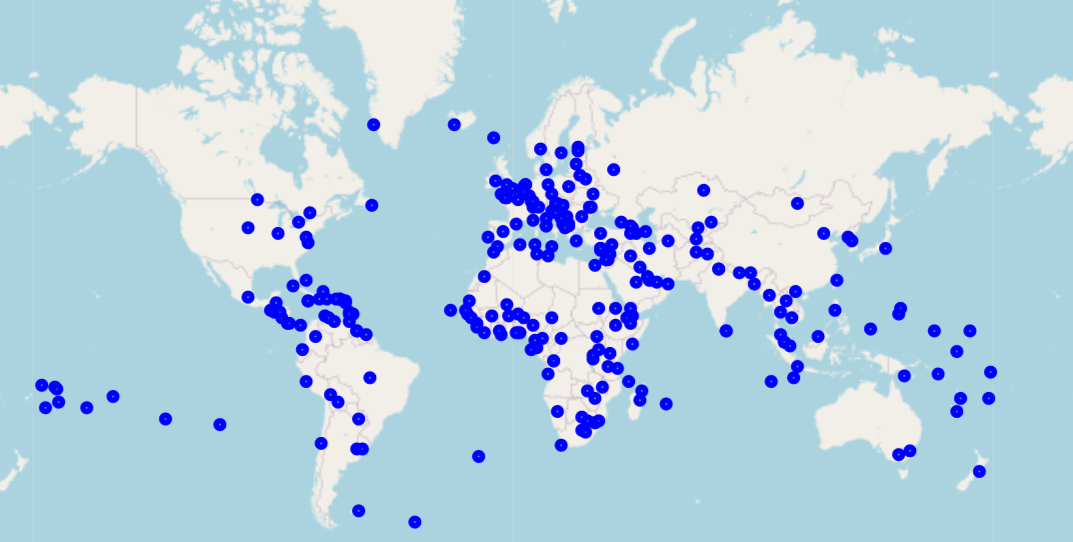


For technical details, please refer to

<https://developer.foursquare.com/docs/places-api>

* + 1. The world map GIS data from Folium

Example



***How the data will be used to solve the problem?***

* + 1. Data of national capitals will be read from the Wikipedia page through python libraries.
    2. Data will be prepared as following:
       - Renaming columns of city and country according to standard naming conventions
       - Removing parentheses and all data within using Pandas
       - Adding the lat. and long coordinate columns structure to the data frame structure
       - Obtaining the coordinates data of the national capitals. If an exception occurs, insert nan values in the coordinates columns and print the word "nan inserted".
       - Drop rows with nan values in latitude or longitude fields (if any)
       - Reset df index in case nan rows were dropped
       - Creating a World Map with all Countries' Capital Cities superimposed on top
    3. Exploring the World Capitals
       - Reading the venue data through Four -quare API
       - Clean the json and structure it into a pandas data frame.
       - Fnd out how many unique categories can be curated from all the returned venues
       - Calculate the total number of venues in each category
       - Exclude all venues except Hotels, Coffee Shops and Restaurants
    4. Analyse each national capital
       - Do one hot encoding
       - group rows by World Capital Cities and by taking the mean of the frequency of occurrence of each category
       - print each capital city along with the venues of interest
    5. Cluster national capital based on venue categories (Coffee Shop Hotel Restaurant)
    6. Visualize the resulting clusters in the world map with colour coding
    7. Print the required statistics
       - National capitals assigned to each section (i.e. each cluster)
       - Potential customers in the section’s business line in each national capital
       - Potential customers in the other two business line in each city
    8. Visualize the results in the form of 4 pie chart (one pie chart for each of the three sections and a fourth pie chart for the total number of potential customers world-wide)