Capstone Project Report

1- ***Description of the data and how it will be used to solve the problem***

Description of the data

The following data sets will be used:

A- List of national capitals from the Wikipedia page:

Sample snapshot of the data:

City/Town	Country/Territory	Notes
Abidjan (former capital; still has many government offices)	Ivory Coast	
Yamoussoukro (official)		
Abu Dhabi	United Arab Emirates	
Abuja	Nigeria	Lagos was the capital from 1914 to 1991.
Accra	Ghana	
Adamstown	Pitcairn Islands	British Overseas Territory.
Addis Ababa	Ethiopia	
Aden (de facto, temporary)	Yemen	Sana'a has been occupied by Houthis rebels since February 2015. Aden is Yemen's acting capital. See also: Yemeni Civil War (2015–present).
Sana'a (de jure)		
Algiers	Algeria	
Alofi	Niue	Self-governing in free association with New Zealand.
Amman	Jordan	
Amsterdam (official)		The Dutch constitution refers to Amsterdam as the "capital".

For the complete dataset, data please refer to:

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

B- Geo-Location data of each national capital from geocoding web services.

	City	Country	lat	lng
0	Abidjan	Ivory Coast	5.32036	-4.01611
1	Yamoussoukro	Ivory Coast	6.80911	-5.27326
2	Abu Dhabi	United Arab Emirates	24.4748	54.3706
3	Abuja	Nigeria	9.06433	7.4893
4	Accra	Ghana	52.4934	4.80368
5	Adamstown	Pitcairn Islands	-25.0667	-130.1
6	Addis Ababa	Ethiopia	9.01079	38.7613
7	Aden	Yemen	12.8333	44.9167
8	Sana'a	Yemen	15.3539	44.2059
9	Algiers	Algeria	36.7754	3.06019
10	Alofi	Niue	-19.0534	-169.919

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

C- National capitals important venues from Foursquare API.

Example:

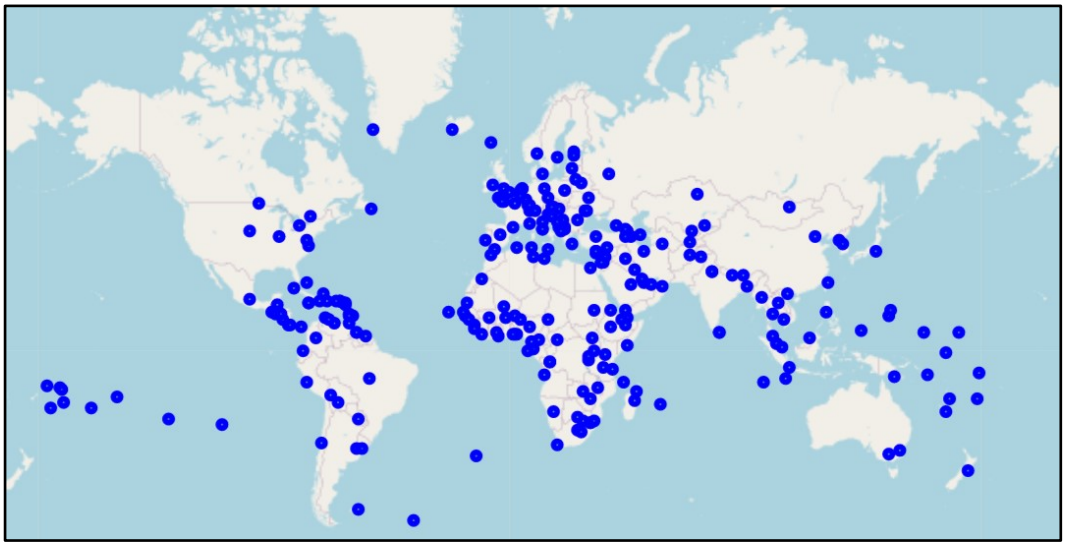
	World Capital	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Abidjan	5.320357	-4.016107	Sofitel Abidjan Hôtel Ivoire	5.327097	-4.004801	Hotel
1	Abidjan	5.320357	-4.016107	Norima	5.363668	-3.992067	American Restaurant
2	Abidjan	5.320357	-4.016107	Cap Sud	5.298763	-3.987246	Shopping Mall
3	Abidjan	5.320357	-4.016107	Bao Café	5.348778	-3.996881	Coffee Shop
4	Abidjan	5.320357	-4.016107	Pink Club	5.305360	-3.988696	Nightclub
5	Abidjan	5.320357	-4.016107	Nice Cream	5.291398	-3.982492	Ice Cream Shop
6	Abidjan	5.320357	-4.016107	Lifestar	5.324086	-4.015354	Nightclub
7	Abidjan	5.320357	-4.016107	Des Gateaux & Du Pain	5.360270	-3.989671	Bakery
8	Abidjan	5.320357	-4.016107	Di Sorrento	5.288542	-3.987629	Italian Restaurant

For technical details, please refer to

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

D- The world map GIS data from Folium

Example



2-How the data will be used to solve the problem?

A- Data of national capitals will be read from the Wikipedia page through python libraries.

B- 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

C- 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

D- 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

E- Cluster national capital based on venue categories (Coffee Shop Hotel Restaurant)

F- Visualize the resulting clusters in the world map with colour coding

G- 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

H- 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)