Project Development

Finding, understanding and analizing data

Thru this processes we will download, clean and analyze data for both countries.

Country with similar population to Mexico.

The source of information was extracted https://datos.bancomundial.org/.

First, we obtained the Mexico total population:

```
TotMex.head()
```

	Population, total
date	
2019-01-01	127575529.0

```
TotMex.head()
print('Total population in Mexico is ', TotMex['Population, total'][0])
```

```
Total population in Mexico is  127575529.0
```

and next, we search for a country with population less or equal to Mexico population:

```
df = wbdata.get_dataframe(indicators, data_date=fecha,convert_date=True,country='all')
```

```
# select the first country that has a population <= Mexico population
dfsort = df[df['Population, total'] <  TotMex['Population, total'][0]]
dfsort = dfsort.sort_values(by='Population, total',ascending=False)
dfsort.head(1)
```

	Population, total
country	
Japan	126264931.0

Summary: the country with similar population to Mexico is Japan.

References:

NoteBook	Description	Output data
MexicoJapanData.ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/MexicoJapanData.ipynb	Extract and clean data	Mexico_General_Population Mexico_StatePopulation

Understanding administrative divisions for each country

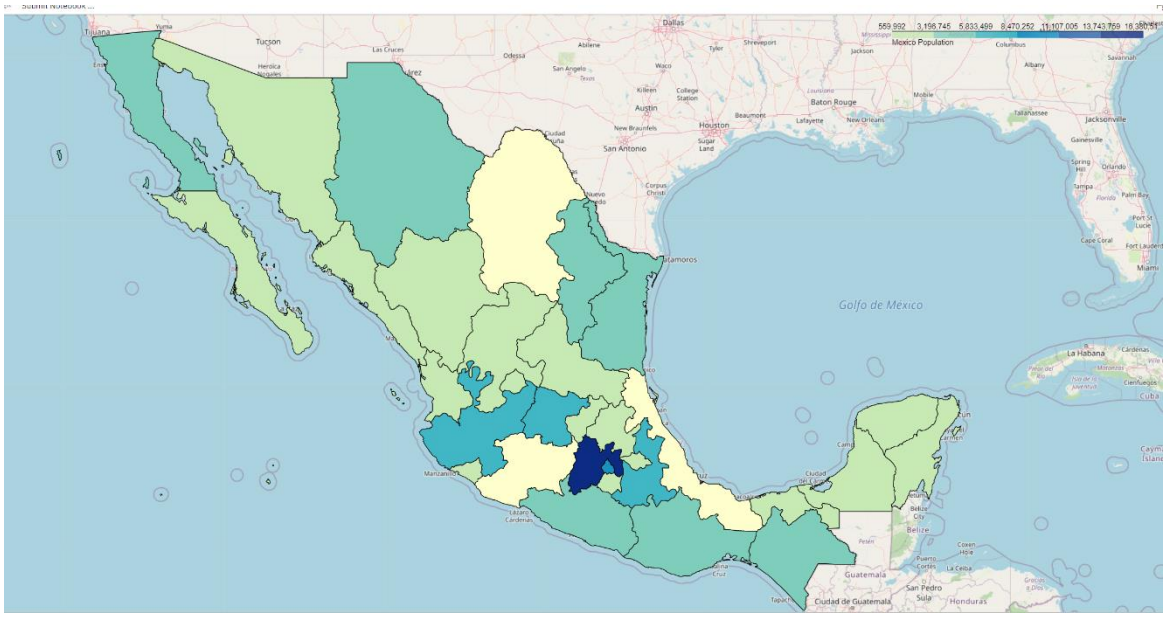
Mexico Analysis.

The external resources that we used were:

a) The population at both levels (national and state administrative divisions) was extracted from INEGI https://www.inegi.org.mx/, an autonomous public body responsible for regulating and coordinating the National System of Statistical and Geographical Information of the country.

b) Also because we wanted to use Choropleth maps, we use a geojson file (with geospatial information) that we find in https://raw.githubusercontent.com/angelnmara/geojson/master/mexicoHigh.json

Below is the choropleth map with information of population by state:



On the above map we can see that Mexico (well known as Estado de México) is the most populated state in México.

The information of the top 3 states with more population are as follow:

```
# Verifying which states have de maximun population
df.sort_values(by=['Population'],ascending=False).head(3)
```

	cve_entidad	desc_entidad	cve_municipio	Population
14	15	México	0	16225409.0
8	9	Ciudad de México	0	8985339.0
29	30	Veracruz de Ignacio de la Llave	0	8127832.0

The top 3 most populated Municipalities in Ciudad de Mexico are:

	cve_entidad	desc_entidad	cve_municipio	desc_municipio	Population
5	9	Ciudad de México	7	Iztapalapa	1815786.0
3	9	Ciudad de México	5	Gustavo A. Madero	1185772.0
8	9	Ciudad de México	10	Álvaro Obregón	727034.0

Summary: Mexico has 31 states and México City (32 states), all of them are divided into municipalities . Mexico City is divided into 16 *delegaciones* or municipalities.

References:

NoteBook	Description	Output data
MexicoJapanData.ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/MexicoJapanData.ipynb	Extract and clean data	Mexico_General_Population Mexico_StatePopulation
MexicoVisualization ipynb	Generate maps and graphs for México	

Japan Analysis.

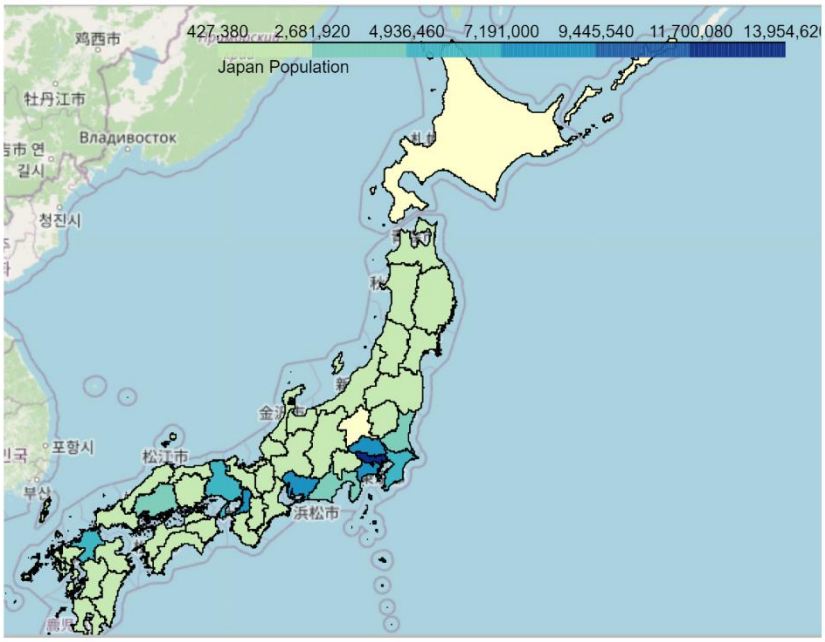
The external resources that we used were:

a) Population at national level were gathered from https://www.e-stat.go.jp/, a portal Site of Official Statistics of Japan.

b) Tokyo Wards, were obtained the information from https://en.wikipedia.org/wiki/Special_wards_of_Tokyo that we scraped to obtain the population of each ward in Tokyo,

c) Also because we wanted to use Choropleth graphics, we use a geojson file (with geospatial information) that we find in https://raw.githubusercontent.com/dataofjapan/land/master/japan.geojson

Below is the map that shows the population by prefecture:



The top 3 most populated perfectures in Japan are:

```
# perfectures order by Population
dfJapan.sort_values(by=['Population'],ascending=False)
```

	YEAR	AREA Code	AREA	Population
12	2018	13000	Tokyo-to	13822000.0
13	2018	14000	Kanagawa-ken	9177000.0
26	2018	27000	Osaka-fu	8813000.0

And as we mention earlier, in order to obtain the Tokyo wards, we did some web scrapping and from the saved file, we obtain the 23 municipalities:

Tokyo wards

```
df = pd.read_csv('Japan_WardTokyo.csv')
df.head()
```

	No.	Name	Kanji	Poblacion	Density(/km2)	Area(km2)	MajorDistricts
0	1	Chiyoda	千代田区	59441.0	5100.0	11.66	Nagatachō, Kasumigaseki, Ōtemachi, Marunouchi,...
1	2	Chūō	中央区	147620.0	14460.0	10.21	Nihonbashi, Kayabachō, Ginza, Tsukiji, Hatchōb...
2	3	Minato	港区	248071.0	12180.0	20.37	Odaiba, Shinbashi, Hamamatsuchō, Mita, Roppong...
3	4	Shinjuku	新宿区	339211.0	18620.0	18.22	Shinjuku, Takadanobaba, Ōkubo, Kagurazaka, Ich...
4	5	Bunkyō	文京区	223389.0	19790.0	11.29	Hongō, Yayoi, Hakusan

```
# verify the number of wards
df.shape
```

```
(23, 7)
```

The top 3 most populated wards in Tokyo are:

```
df.sort_values(by=['Poblacion'],ascending=False).head(3)
```

	No.	Name	Kanji	Poblacion	Density(/km2)	Area(km2)	MajorDistricts
11	12	Setagaya	世田谷区	910868.0	15690.0	58.05	Shimokitazawa, Kinuta, Karasuyama, Tamagawa
19	20	Nerima	練馬区	726748.0	15120.0	48.08	Nerima, Ōizumi, Hikarigaoka
10	11	OtaŌta	大田区	722608.0	11910.0	60.66	Ōmori, Kamata, Haneda, Den-en-chōfu

Summary: Japan has 47 prefectural entities and Tokyo has 23 municipalities or special wards (together make up the core and the most populous part of Tokyo Metropolis, Japan).

NoteBook	Description	Output data
MexicoJapanData.ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/MexicoJapanData.ipynb	Extract and clean data from Japan	Japan_General_Population

BeautifulSoap ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/BeautifulSoap.ipynb	Wrap Wikipedia page for Tokyo Wards	Japan_WardTokyo.csv
Japan Visualization ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/Japan Visualization.ipynb	Maps and graph for Japan and Tokyo	

Selecting data

Now that we have the two countries to work with, the next step would be to select at least one of the wards (for Japan) and one municipality (for México) to find out if there are similarities on the type of cousins between those two cities.

Foursquare and K-Mean

Under development and will be delivered on Week5 of Module 9.

However, just for learning purposes we repeat the lab done with New York, but with México City data (see Notebook: FourSquare.ipynb).

Five clusters were generated and the majority of the Clusters has as 1st Most Common venue, food and drink places, except for Cluster zero where we can see more variety (Gyms, Electronics)

In most of the Clusters, gyms are in 3rd and 4rd most common venues, except for Cluster 4 that is in 5th position.

Cluster zero has more “Colonias” included (Neighorhoods).

Cluster zero is the only cluster that shows a museum.

Pharmacies appear in all Clusters, but is more common in Cluster zero.

My conclusion based on the above finding is that Cluster zero has much more variety than other clusters. I live in México and I can understand why : nice neighorhoods like Lomas de Guadalupe, San Angel Inn, Guadalupe Inn, Paseo de las Lomas, San Angel, Florida, Las Aguilas are in this cluster. It’s a reality that developed neighorhoods have much more services available and less developed neighborhoods have less services available (like Cluster 4).

Conclusions

Under development and will be delivered on Week5 of Module 9.