Contenido

Objective

The main objective of this project is to apply the different skills learned in the course IBM Analytics by Coursera and does not pretend to be a Business Case.

We will use Python for this project together with some libraries like pandas, Seaborn, Matplotlib, Choropleth, Folium, Foursquare and others that will help us to do the analysis.

The target audience could be people who are starting their journey on Business Analytics with Python.

Introduction.

For the final project of IBM Analytics course, it was requested to perform an analysis using the different skills learned in the course and Foursquare to explore location data for one or more cities.

We have decided to explore México and other country with similar population, and analyze how the foreign cuisine is positioned in the other country. We will not analyze all type of cuisines, but only those local cuisines for each country, in other words, we will analyze Mexican Cuisine in the other country and Cuisine X (because at this point we don't know which country we are going to use) for Mexico.

With this information, we pretend to see how developed are those foreign cuisines for similar cities of each countries. This could be the starting point to understand the distribution of the foreign cuisine on each city, and maybe at the end we conclude if it's convenient to open the local cuisine on the other country or not.

Data

There are several steps that we will go thru this project, but the most important one is getting the resources of information and for that, this process will include but not be limited to:

Finding the data

Search a country with a similar population to México. There are a lot of resources to get this information, but we have decided to explore them thru the World Bank, an organization that provide high-quality international statistics and maintains a number of macro, financial and sector databases. We will explore https://datos.bancomundial.org to find out which country has a similar population to Mexico.

Understanding and Analizing data

Understanding the structure of the country in terms of administrative divisions (states, municipalities, etc.) will allow us to pick up the cities that we are going to explore in each country. Information about administrative divisions can be obtained from Wikipedia, and in other cases can be obtained from government resources of each country. For both countries we will show the population information segregated by administrative divisions.

Exploring data

We will use Foursquare to explore venues (in our case cousins types) around the areas selected and K-means to find out if there are similarities among the two cities selected

Methodology.

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
WorldBank.ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/WorldBank.ipynb	Extract country with similar population to Mexico	Mexico_General_Population Mexico_StatePopulation

Understanding administrative divisions for each country

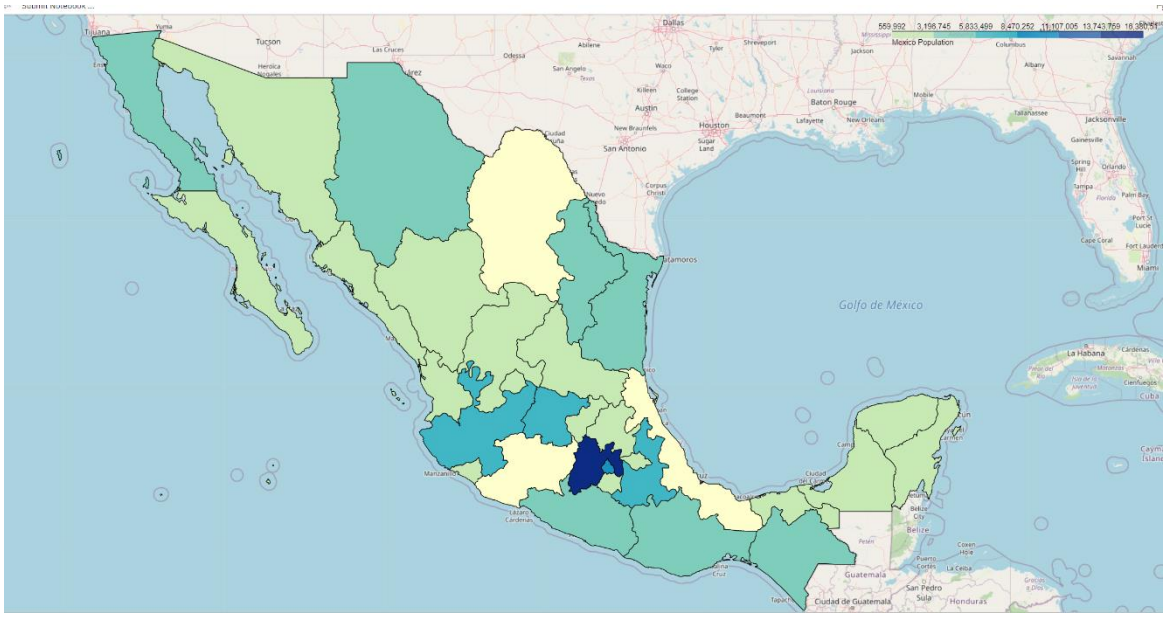
Mexico Analysis.

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 (based on the 2015 Census):

```
# 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 Mexico City 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.

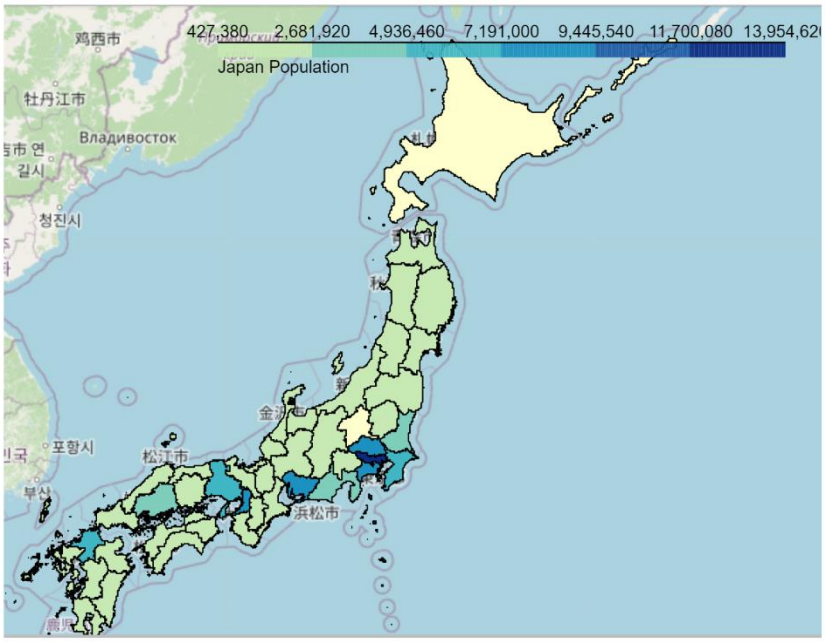
The external resources that we used were:

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 (based on 2018 information):

```
# 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

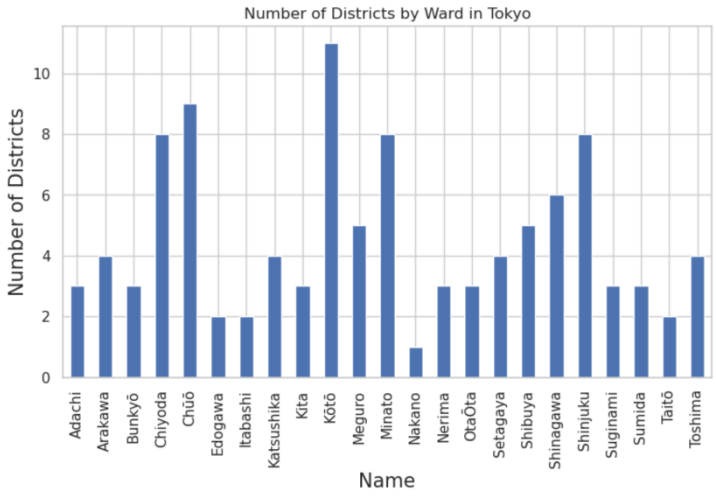
Now that we have the two countries to work with, the next step would be to decide which cities we are going to use for the exploration.

Based on the results of the previous section, the two candidates cities (based on population) should be México with a population of 16,225,409 and Tokyo with 13,822,000, however the density/km2 is significantly different (655,9 hab/km² and 6,158 respectively) and we think this is an important factor in the analysis.

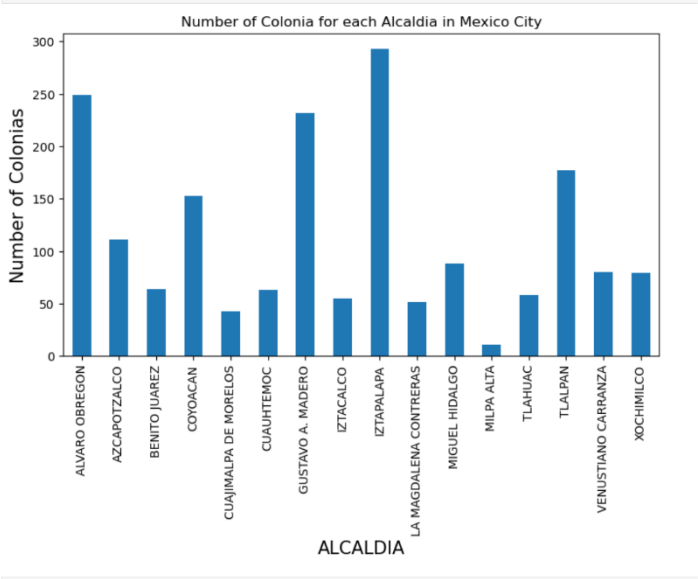
Having said that, we select Mexico City with a population of 8,985,339 (census 2015) with 5,936.8 density hab/km2 and Tokyo with a density of 6158 (very similar).

Foursquare and K-Mean

Before exploring Foursquare, we generate some graphs of the number of districs per ward, and colonias by Alcaldia

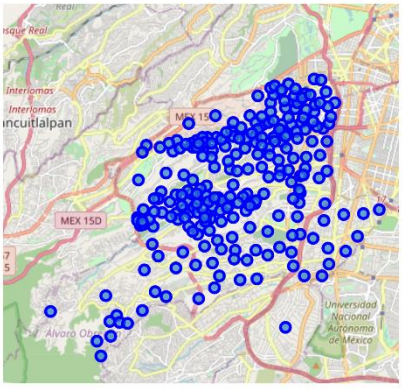


Tokyo number of Districts by Ward



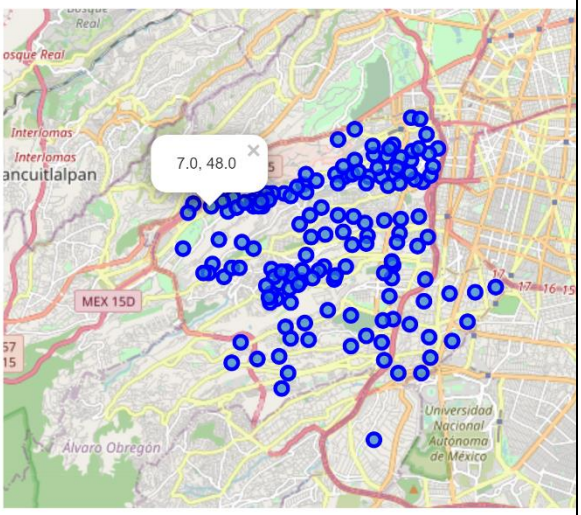
Mexico City number of Colonias by Alcaldia

Due to the vast number of Colonias in México City, we decided to pick ALVARO OBREGON Alcaldia only (which is the second Alcaldia with more Colonias) as shown in the above graph, and we will select all wards from Tokyo.

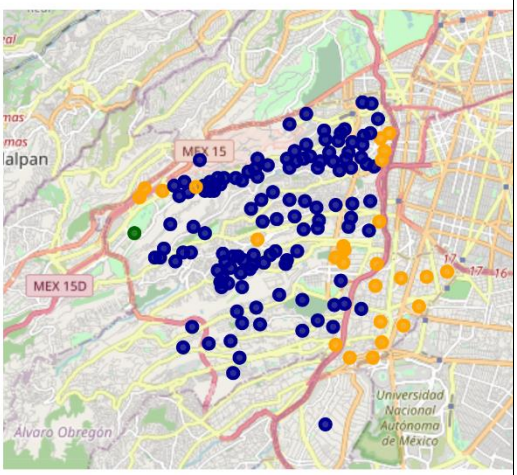
Mexico City Colonias for Alcaldia = ALVARO OBREGON		Japan Wards
		

And now is time to use Foursquare. Foursquare is a social location service that allows users to explore the world around them and allow companies to use it for marketing purposes. We are going to use the API in Python to interact with this platform and gain access to a wide range of venues using geolocalization. Mainly, we will gather the number of venues near to each colonia in ALVARO OBREGON Alcaldía for Mexico City and to all Wards in Tokyo based on two categories: Mexican Food and Japanese Food.

We will explore venues in a radius of 500 meters and the search will be limited to 100 results. Foursquare platform returned the following results:

Japanese Restaurant in ALVARO OBREGON (by Colonia) Label : Japanese Restaurants, Mexican Restaurants		Mexican Restaurants in Tokyo (by Ward) Label : Ward, Mexican Restaurnt, Japanese Restaurant
		

Within this information we proceed to create 3 Clusters for both cities:

Clusters for ALVARO OBREGON		Clusters for Tokyo
		

From the above maps we can resume:

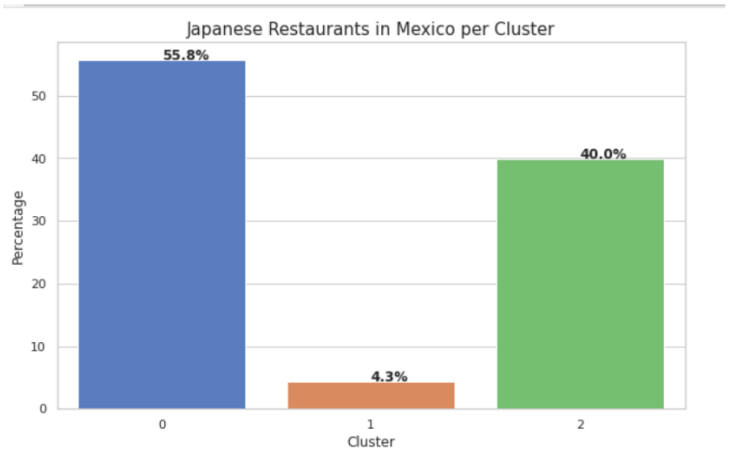
México City:

Cluster 0 and Cluster2 represents almost 99% of the total of Japanese restaurants and it's a mix of residential zones and popular (with tradition) zones.

Cluster 0 (darkblue) : Majority of the Colonias belong to this Cluster, but the range of Japanese restaurant are between 1-3 per Colonia, this represents 55.8% of the Japanese Restaurant total found in Alvaro Obregón.

Cluster 1: (darkyellow) : It include 26 colonias, and has between 4 and 9 Japanese restaurants, this represents 40.0 % of the Japanese Restaurant total found in Alvaro Obregón.

Cluster 2: (dargreen) One Colonia has the total of the 16 Japanese restaurants in this Cluster, this represents 4.3% of the Japanese Restaurants total found in Alvaro Obregón. This is a cluster with a lot of new residentials developments and a lot of variety of choices of cuisines. Urban development is hugh and lot of investment in infrastructure has been done to communicate this Colonia with the rest. I think we will see more services on the following years.



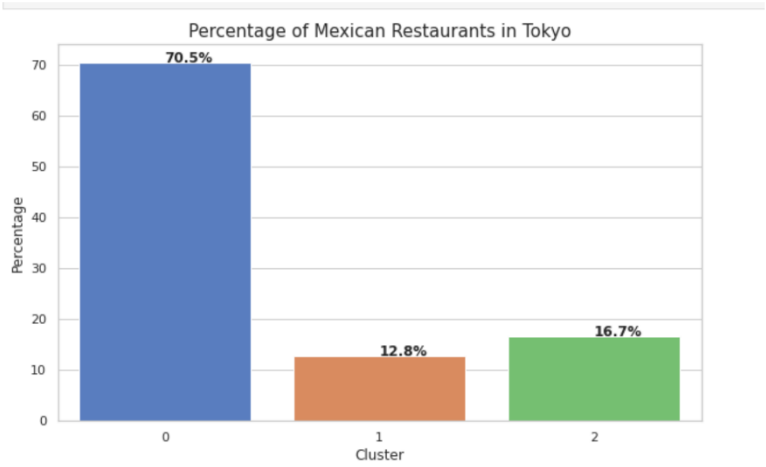
Tokyo:

Only 10 out of 23 Wards had Mexican restaurants.

Cluster 0 (darkblue) : a range from 11 to 16 mexican restaurant belong to this cluster and represents 14.8% of the total Mexican restaurants in Tokyo. This cluster has districts with very popular spots, shopping malls, corporate offices (Amex, Microsoft, Iwatsu Electric,etc.), business centers, residential and public areas, and could be the reason why this cluster has the majority of the mexican restaurants (due to the diversity of services).

Cluster 1 (darkgreen) : a range from 1 to 5 mexican restaurants belong to this cluster and represents 12.8% of the total Mexican restaurants in Tokyo. This cluster include Shinjuku which is the most important Neighorhood on West Tokyo with lots to shopping, restaurants and night live, but it seems to be that Mexican cuisine is not well established here.

Cluster 2 (red) : : a range from 6 to 7 mexican restaurants belong to this cluster and represents 16.7% of the total Mexican restaurants in Tokyo. Some spot here include the Skeytree and the oldest temple in Tokyo and as with Cluster 1, Mexican cuisine is not well established here.



NoteBook	Description	Output data
FourSquareMEXJPN.ipynb https://github.com/ASanchez3793/Coursera_Capstone/blob/main/Notebooks/FourSquaareMEXJPN.ipynb	Explore venues in Foursquare and Cluster	

Conclusion and Discussions

Conclusion

Regarding México City (Alcaldía Alvaro Obregón) I would say that due to the strong social contrast on this Alcaldía, where on one hand there are exclusive residential zones focus on high and medium-high classes, on the other hand it also includes medium-low and low classes. I would suggest to do a segregation based on income in order to reach the right target.

Regarding Tokyo, If I should have to do a recommendation for opening a new Mexican restaurant, I will certainly look for more information on districts in Cluster 0.

I see similarity on Cluster0 among the two cities : lot of services besides restaurants (residentials developments, corporate offices) and a diverse variety of other services.

Discussions.

To improve the outcome of this project, I would definitively :

a) I would not explore all Wards in Tokyo, but only a district;

b) I would include variables like income, Land price, security index and other that could help to have a stronger result;

c) explore Foursquare on a deeper level to get for instead : quality, price, type of restaurant, average cost of meal or other characteristics that will help to take a more accurate decision.

Related with the tools, I can say there are other platforms that provide similar information to Foursquare, like Yelp, Tripadvisor, Wikitravel, Swarm, Loopt, whrrl, etc., ¿which is the best? Definitely it depends on the company budget, target to achieve, type of analysis to be done, etc..

And last but not less important, like any other project, 75% of the project is planning – in this case, searching for accurate data – and the 35% left is execution. So, I suggest to take the time to search for the right sources and make sure to understand the stakeholder(s) needs.