

Coursera Capstone Project - Battle of the Neighbourhoods:

Business Analysis on San Salvador (Metropolitan Area), El Salvador

Introduction: Business Problem

This notebook is a study that analyses every type of venue and will recommend the best locations to open a restaurant within the Metropolitan area of San Salvador (MASS). This area of El Salvador, contains 14 municipalities which are the following:

- San Salvador
- Antiguo Cuscatlán
- San Marcos
- Soyapango
- Ilopango
- San Martín
- Tonacatepeque
- Ciudad Delgado
- Cuscatancingo
- Mejicanos
- Apopa
- Nejapa
- Ayutuxtepeque



From the centre of San Salvador, a radius of 15 KMs will set as a limit with the intention of covering the entire metropolitan area. The radius selection is due to lack of venues registered under the Foursquare database in the centre of San Salvador.

Data

The analysis in this notebook will be based on the following requirements of the project to solve the question in hand:

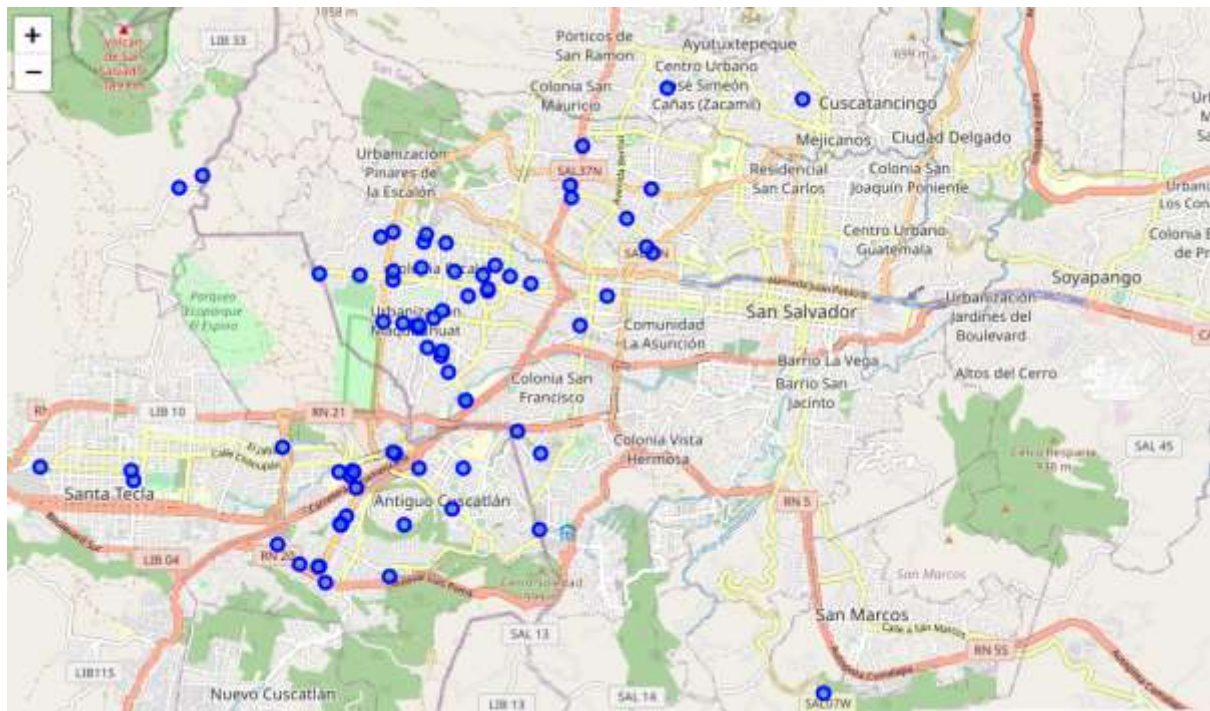
- Types of existing restaurants in the area.
- Top types of restaurants and other venues in the area.

For practical purposes, the data was extracted from Foursquare Geolocation DB through the API calls. Foursquare is one of the world leaders on geospatial data and keeps its database fresh and constantly updated with the latest venues. **Problem solution** will come from the analysis on the process by which Data was extracted in raw form and processed into data frames, to be later clustered through machine learning algorithm (k-Means), this will provide a clear vision on what types of venues are in a certain area. Understanding what business types develop on certain areas of the city will supply patterns of how the Salvadoran society consumes or accesses a service.

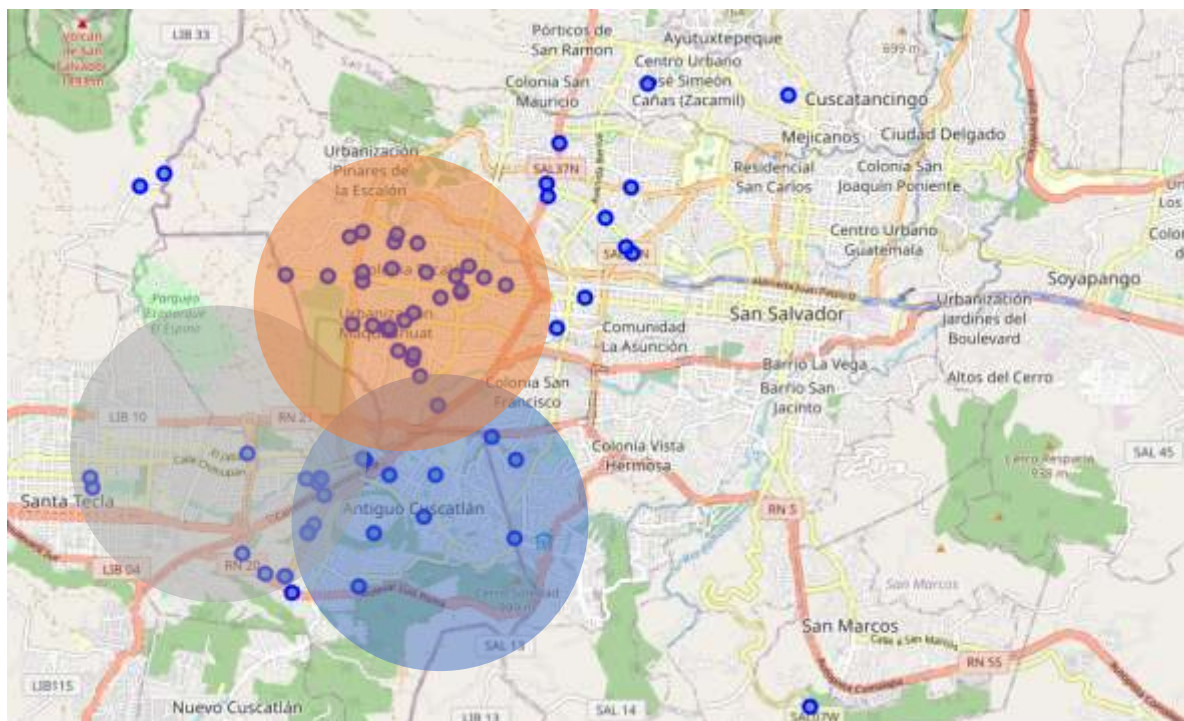
Choosing Municipalities:

We will run several algorithms in the MASS, to later decide where will we focus our analysis. We run our initial study at the city center of MASS, 15 km radius, to be able to grab all municipalities:





After the analysis of venue type distribution through MASS, it becomes clear that three main municipalities excel when thinking about opening a business in the area. Antigua Cuscatlán, Santa Tecla, and San Salvador (North). kNN machine learning algorithm, frequencies and simple counts of the area also reveal which type of venue is most concurrent in the different municipalities:



Methodology

Data extraction:

Data extraction has three main sources:

A. Jupyter Notebooks:

To process the entire study, we will use Jupyter Notebooks.

B. Foursquare API:

The Foursquare API will allow us to analyse neighbourhoods and venues in each area of the city. A 15 km radius was chosen for data extraction

C. GDB data of the country, states, and municipalities:

To understand how the municipalities are divided we require data extraction from GDB files. The GDB files were downloaded from the GeoSpatial.

Data Processing:

Standardization:

A. Foursquare API:

The raw data received from the API call a cleansing process is done through Python with its Pandas library for data manipulation functions and conversion of this raw information into data frame.

B. GDB:

The GDB files are already standardized, therefore no changes will occur to this type of file.

C. Variable ID:

The use of tags to keep track of which variable, and class belong to which municipality.

- AC = Antiguo Cuscatlan.
- ST = Santa Tecla
- SS or Ss = San Salvador.

Information Generation:

After main data frames have been created, and GDB files have been loaded successfully, produced from the original data, new data frames and graphs will be generated, to examine different views which will allow us to analyse the problem and provide a solution.

Preliminary Data Analysis:

Individually counting the type of venue will reveal how common a venue could be, and although it bids the idea of commonness, the analysis would not be complete. Precision is required to understand the economic activity of the different municipalities in MASS. By now, it is evident where most of the economic activity in MASS occurs, three main municipalities are shown:

- Antiguo Cuscatlán
- Santa Tecla
- San Salvador (North)

Foursquare works by the amount of times a person checks in a venue. Someone can subscribe venues, and if this venue becomes popular, then it will show up on top searches. If this study is focused from the city center of San Salvador (even by using 4KM radius), the amount of venues in the city center, is too low (57+- venues), the study could reflect faulty results. Therefore, this investigation will now redirect its focus to the three municipalities mentioned above; to determine, where in MASS could the Stakeholder open a restaurant. The process of analysis continues by extracting data of all nearby venues, of the one hundred sites originally summoned from Foursquare. This data will be processed to new data frames with values that are to be transformed from alphabetic to numeric, not only for standardization purposes, but also to generate new frames of data, with a frequency measuring the amount of times a type of venue concurs in a neighbourhood. This data set of frequencies can now tell us commonness of venue types. The last step of the process is to merge (inner join by neighbourhood) the original purified data frame, with the commonness data frame. This last information table shall be processed by kNN Machine Learning Algorithm who should provide a clear answer on what type of restaurant and best location for type. New data frames comparing the concurrence per cluster, will help the Stakeholder understand the economic activity in the area, and how to invest. Although there are other factors to consider when choosing a location, like price and security; this study will not take this data into account since more data is required from local authorities, who at the moment do not possess the technology infrastructure to extract this data from.

Visualization:

All stages of the data analysis are properly graphed with matrices representing the data frames, scatterplot maps representing geolocation of venues, other types of maps helping represent an idea of the area, some of the maps were made in Tableau, as well many charts where done by using PowerBI, charts that are part of this report, and presentation.

Analysis

A radius of 2Kms, from each municipality center in Antigua Cuscatlán, Santa Tecla; for San Salvador, a Neighborhood known as Escalon Norte was chosen for its center, since the study reflects more economic movement on this area. The behavior of the areas has been set to extract all nearby venues:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Calle Principal Antigua Cuscatlan	13.67225	-89.240188	4 Monkeys Café	13.673681	-89.238330	Breakfast Spot
1	Calle Principal Antigua Cuscatlan	13.67225	-89.240188	Biscuit Factory Café Bistro	13.672250	-89.240188	Café
2	Calle Principal Antigua Cuscatlan	13.67225	-89.240188	Sushi King	13.677708	-89.238621	Sushi Restaurant
3	Calle Principal Antigua Cuscatlan	13.67225	-89.240188	The Donut Place	13.675482	-89.244515	Donut Shop
4	Calle Principal Antigua Cuscatlan	13.67225	-89.240188	Oasis Jugos y Licuados	13.677677	-89.239972	Juice Bar

For us to compare, and run Algorithms that will show us we need to turn certain data into numbers, the technique used during this study was One Hot Encoding:

Neighborhood	American Restaurant	Arcade	Argentinian Restaurant	Art Gallery	Asian Restaurant	Bakery	Bar	Beach	Beer Bar	—	Sushi Restaurant	Taco Place	Ten Room	Theater	Theme Park	Ty / Game Store	Veget. /V. Resto
0 Calle Principal Antigua Cuscatlan	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	0
1 Calle Principal Antigua Cuscatlan	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	0
2 Calle Principal Antigua Cuscatlan	0	0	0	0	0	0	0	0	0	—	1	0	0	0	0	0	0
3 Calle Principal Antigua Cuscatlan	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	0
4 Calle Principal Antigua Cuscatlan	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	0

With these files converted, we are now able to measure mean and therefore frequency:

```

----1a Calle Pte block P #21----
      venue  freq
0      Coffee Shop  0.07
1          Café  0.05
2      Pizza Place  0.04
3      Shopping Mall  0.04
4  Italian Restaurant  0.04

----1ra Calle Oriente----
      venue  freq
0      Coffee Shop  0.07
1      Restaurant  0.04
2      Shopping Mall  0.04
3      Italian Restaurant  0.04
4  Latin American Restaurant  0.04

```

With the frequency, we can now determine concurrency of venues in neighborhood, thus opening the possibility of another perspective.

Knowing the type of venues in the area, gives us an idea of how the economy moves in certain locations.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	1a Calle Pte block P #21	Coffee Shop	Café	Restaurant	Pizza Place	Shopping Mall	Italian Restaurant	Sushi Restaurant	Latin American Restaurant	Bakery	Salad Place
1	1ra Calle Oriente	Coffee Shop	Latin American Restaurant	Italian Restaurant	Hotel	Pizza Place	Restaurant	Shopping Mall	Café	Bakery	Sushi Restaurant
2	1ª Calle Oriente	Coffee Shop	Latin American Restaurant	Italian Restaurant	Hotel	Pizza Place	Restaurant	Café	Shopping Mall	Bakery	Salad Place
3	Antigua Via Ferrocarril	Café	Coffee Shop	Hotel	Restaurant	Shopping Mall	Italian Restaurant	Bakery	Gastropub	Sushi Restaurant	Salad Place
4	Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall	Hotel	Pizza Place	Restaurant	Bakery	Italian Restaurant	Salad Place	Latin American Restaurant

Results and Discusion

After the analysis of venue type distribution through MASS, it becomes clear that three main municipalities excel when thinking about opening a business in the area. Antiguo Cuscatlan, Santa Tecla, and San Salvador (North). kNN machine learning algorithm, frequencies and simple counts of the area also reveal which type of venue is most concurrent in the different municipalities:

- Antiguo Cuscatlan: Coffee Shops or Café
- Santa Tecla: Bars
- San Salvador: Pizza Place

Most concurred areas are also shown by the kNN algorithm:

- Antiguo Cuscatlan: Neighborhoods in Cluster 0
- Santa Tecla: Neighborhoods in Cluster 2
- San Salvador: Neighborhoods in Cluster 0 and 4

Antiguo Cuscatlán - Cluster 0

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Calle Principal Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall
Ave. Rio Amazonas	Coffee Shop	Café	Hotel
Ave. Antiguo Cuscatlan	Coffee Shop	Café	Hotel
hacienda los miranda	Café	Coffee Shop	Hotel
Antiguo cuscatlan	Coffee Shop	Café	Hotel
Clle. del Mediterráneo	Café	Coffee Shop	Hotel
1a Calle Pte block P #21	Coffee Shop	Café	Restaurant
Calle Cuscatlán Poniente	Coffee Shop	Restaurant	Pizza Place
Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall
1ra Calle Oriente	Coffee Shop	Latin American Restaurant	Italian Restaurant
Calle Circunvalación pasaje #2 Condominio Vict...	Coffee Shop	Pizza Place	Latin American Restaurant
Peatonal de la UCA	Coffee Shop	Café	Pizza Place
Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall
Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall
Antiguo Cuscatlan	Coffee Shop	Café	Shopping Mall
Calle las Rosas	Coffee Shop	Café	Hotel
Antiguo cuscatlan	Coffee Shop	Café	Hotel
1° Calle Oriente	Coffee Shop	Latin American Restaurant	Italian Restaurant
Boulevard Los Próceres	Café	Hotel	Coffee Shop
Calle Del Mediterraneo	Coffee Shop	Café	Hotel
Calle del Mediterráneo #17	Coffee Shop	Pizza Place	Café
C.C. La Sultana	Coffee Shop	Café	Hotel
Avenida Albert Einstein,Block "J", #3	Coffee Shop	Latin American Restaurant	Hotel
Calle Antigua Ferrocarril	Coffee Shop	Café	Restaurant
Calle y Colonia Cumbres de Cuscatlán	Coffee Shop	Café	Latin American Restaurant
Av. Rio Amazonas #3	Coffee Shop	Café	Hotel
Antigua Vía Ferrocarril	Café	Coffee Shop	Hotel
La Sultana	Coffee Shop	Pizza Place	Café
Boulevard Los Próceres	Café	Hotel	Coffee Shop
Blvd. La Sultana	Coffee Shop	Pizza Place	Café

Santa Tecla - Cluster 2

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
5a. Ave norte	Bar	Pizza Place	Latin American Restaurant
5ta Calle Oriente 3-2, Santa Tecla	Bar	Latin American Restaurant	Pizza Place
1° Calle Pte.	Bar	Latin American Restaurant	Pizza Place

7 avenida norte	Bar	Pizza Place	Latin American Restaurant
Paseo El Carmen	Bar	Latin American Restaurant	Pizza Place
7a. Avenida Norte	Bar	Pizza Place	Latin American Restaurant
1era. Calle Oriente 3-7	Bar	Pizza Place	Latin American Restaurant
Calle José Ciriaco López	Bar	Latin American Restaurant	Pizza Place
1 Calle Poniente, local 2-A	Bar	Latin American Restaurant	Pizza Place
Paseo El Carmen, entre 3 y 5 Av. Norte	Bar	Pizza Place	Latin American Restaurant
Calle Don Bosco #3-2	Bar	Pizza Place	Latin American Restaurant
Av. Manuel Gallardo	Bar	Pizza Place	Shopping Mall
Paseo Del Carmen	Bar	Pizza Place	Latin American Restaurant
Paseo El Carmen	Bar	Latin American Restaurant	Pizza Place
Paseo El Carmen	Bar	Latin American Restaurant	Pizza Place
1 Calle Poniente	Bar	Latin American Restaurant	Pizza Place
Primera Calle Oriente No 4-4	Bar	Pizza Place	Latin American Restaurant
1ra Calle Orients N° 2-6	Bar	Latin American Restaurant	Pizza Place
Centro Histórico de Santa Tecla	Bar	Latin American Restaurant	Pizza Place
Paseo del Carmen	Bar	Pizza Place	Latin American Restaurant
Avenida Dr. Manuel Gallardo Santa Tecla La Li...	Bar	Latin American Restaurant	Pizza Place

San Salvador - Cluster 0			
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Centro Comercial Masferrrr, frente a Burger K...	Chinese Restaurant	Pizza Place	Hotel
Plaza Masferrer	Pizza Place	Chinese Restaurant	Hotel
Calle masferrer	Pizza Place	Chinese Restaurant	Hotel
Paseo General Escalon	Pizza Place	Chinese Restaurant	Hotel
Colonia Escalon	Pizza Place	Chinese Restaurant	Café
Calle Maximo Jerez	Pizza Place	Chinese Restaurant	Italian Restaurant
Colonia Escalon	Pizza Place	Chinese Restaurant	Café
Paseo General Escalon	Pizza Place	Chinese Restaurant	Hotel
Plaza Barcelona, Colonia Escalón	Pizza Place	Chinese Restaurant	Café
3a Calle Poniente #5254 Lomas Verdes, Colonia ...	Pizza Place	Chinese Restaurant	Café
Plaza Masferrer	Pizza Place	Chinese Restaurant	Hotel

Ave Masferrer	Pizza Place	Chinese Restaurant	Hotel
Redondel Masferrer	Pizza Place	Chinese Restaurant	Café
Colonia Escalon	Pizza Place	Chinese Restaurant	Café
Paseo General Escalón #5338	Pizza Place	Seafood Restaurant	Chinese Restaurant
Avenida Masferrer Norte	Pizza Place	Chinese Restaurant	Café
Paseo General Escalón y 99 Avenida Norte Plaza...	Chinese Restaurant	Pizza Place	Hotel
calle cuscatlan oriente	Pizza Place	Chinese Restaurant	Hotel
Avenida Alberto Masferrer	Pizza Place	Chinese Restaurant	Café
Av. Masferrer Norte, Col. Escalón	Pizza Place	Chinese Restaurant	Café
Avenida Masferrer, Texaco Masferrer Local 3	Pizza Place	Chinese Restaurant	Café
San Salvador - Cluster 4			
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
3a. calle poniente # 4817	Café	Hotel	Pizza Place
Torre Futura	Café	Pizza Place	Italian Restaurant
89 Av Norte Escalon	Café	Italian Restaurant	Hotel
Col. Escalón 85 Av. Norte Pie. Dordely # 4407	Café	Italian Restaurant	Pizza Place
91 Avenida Norte #532	Café	Pizza Place	Seafood Restaurant
Plaza Futura	Café	Pizza Place	Seafood Restaurant
79 Av Norte Y 3 Calle Poniente,colonia Escalon	Hotel	Italian Restaurant	Café
81 Ave. Sur Col. Escalon	Italian Restaurant	Café	Hotel
Plaza Futura	Café	Pizza Place	Seafood Restaurant
Plaza Futura	Café	Pizza Place	Seafood Restaurant
Calle Del Mirador	Café	Hotel	Seafood Restaurant
C.C. El Paseo	Hotel	Italian Restaurant	Café
89 Av. Nte. Y 9a. Calle Poniente #4612	Pizza Place	Seafood Restaurant	Café
Plaza Futura	Café	Pizza Place	Seafood Restaurant
Plaza Futura locales 12B-13	Café	Pizza Place	Italian Restaurant
77 Av Norte #346 Col. Escalon	Hotel	Café	Pizza Place
Calle Cuscatlan #130	Hotel	Italian Restaurant	Restaurant
Centro Comercial Orión (frente ex-Kismet)	Hotel	Italian Restaurant	Café
3a. Calle Poniente	Hotel	Italian Restaurant	Café
World Trade Center	Pizza Place	Seafood Restaurant	Café
Paseo General Escalón	Hotel	Italian Restaurant	Mexican Restaurant
Paseo General Escalón	Hotel	Italian Restaurant	Mexican Restaurant
Calle del mirador #4134 Colonia Escalón, San S...	Hotel	Burger Joint	Café

The recommendation would be to open a Cafe, a Bar or a Pizza place since it was shown by kNN that Pizza, Bars and Cafes are still concurrent type of venues in other areas. It might be an idea worth considering but opening a Bar that specializes in cooking pizza and has advanced coffee cooking and serving techniques would be a safe bet. It is also recommended to expand the study of the competition in the areas and do a SWOT analysis on them, to know how to compete against other venues.

Conclusion:

The study reflects clearly where and how to invest money in MASS. El Salvador does have FTA with the United States, UE, Japan and many other countries, these treaties help investors in many ways; El Salvador does count with a government entity that assists for all foreign investment (PROESA). Now then, this study can be taken by anyone interested who wants to invest in the MASS area, not only foreigners. In the past, El Salvador has had problems with security, but in these last year, crime rates have hit a new low, stimulating for investment.

This study does not correlate crime data, and real state pricing, the country does not have free access to this type of information. Therefore, other studies in terms of security and real state should be considered to build a complete business plan.