Capstone Project - The Battle of the Neighborhoods

Table of Contents

Introduction

The objective of this project is to *find a suitable place for a restaurant in Buenos Aires, Argentina*.

The outcome of this work is oriented to *help stakeholders interested in opening a Peruvian Restaurant in Ciudad de Buenos Aires, Buenos Aires, Argentina*.

Buenos Aires is a top tourist destination in Latin America, is known for its preserved eclectic European architecture and rich cultural life. Since it is a multicultural city, being home to multiple ethnic and religious groups, there are many restaurants, for that reason, looking for an adequate place to open a new one is an important challenge.

Some of the *main drivers for decision making* are related with the following factors:

- places not currently crowded with restaurants
- places in areas with group of ethnic restaurants (with no Peruvian restaurants)
- places that are close to city downtown

Data Science methods, processes and tools will be applied in order to help identifying city areas candidates based on the aforesaid criteria.

Some of questions that this work may help to answer:

- Is this a good place to open a restaurant?
- Is it a promising place that will allow the Business to have good financial performance?
- Will it be a place that will help the Business to be expanded to other city areas in the future?

Data

Considering the Business Challenge, the 3 main drivers to be considered for collecting, wrangling, enrichment, and modeling the Data are as follows:

- places not currently crowded with restaurants
- places in areas with group of ethnic restaurants but with no Peruvian restaurants)
- places that are as close to city downtown

The 3 selected data sources for this project are:

1-***Wikipedia***: Communes of Buenos Aires: information of the 15 communes Buenos Aires:

English: https://en.wikipedia.org/wiki/Communes_of_Buenos_Aires

Spanish: https://es.wikipedia.org/wiki/Comunas_de_la_ciudad_de_Buenos_Aires

JSON file for Geolocation:
http://cdn.buenosaires.gob.ar/datosabiertos/datasets/comunas/CABA_comunas.geojson

2-***Geolocator API***: coordinates of Ciudad de Buenos Aires will be obtained using services for geocoding

3-***Foursquare API***: information on restaurants and type of restaurants and location in the city https://foursquare.com/

Methodology

The steps followed in order to fulfill the objective of the project were oriented to identify areas in Ciudad de Buenos Aires that are suitable for opening a new restaurant.

The Steps followed are:

a) Understanding of the Business context

- Basic knowledge of Ciudad de Buenos Aires Geography and neighborhoods: to proceed with these steps, GEO information gathered from Government is used. JSON files from government site were used to have a map of the city and its 15 neighborhoods or "Comunas"

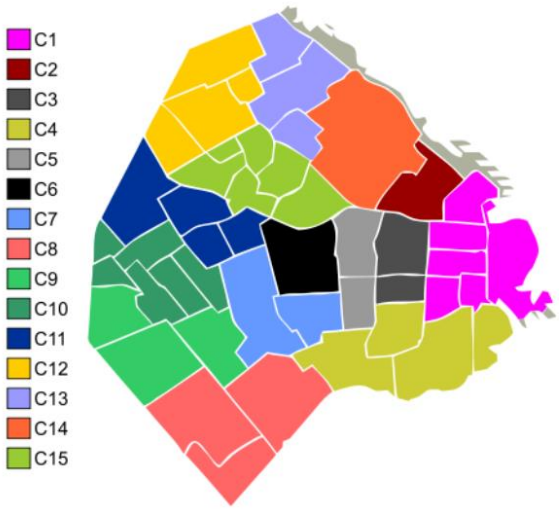
- Data visualization of neighborhoods: in order to understand the "big picture" folium graphs were created

- Project objective reconfirmation was done with stakeholders. As an outcome, 1 of the neighborhoods were set as the targeted area (CABALLITO) due to its central position in Ciudad de Buenos Aires, and the particular interest in developing the business there

Data Source 1: Wikipedia Communes of Buenos Aires: information of the 15 communes Buenos Aires

The city of Buenos Aires is administratively divided into fifteen comunas. Each comuna (commune) of the city encompasses one or more neighbourhoods (barrios), which are represented in the respective community centres for administrative purposes.

This source of information indicates a high level view on how the city area (10 x 10 km) is divided.



	Borough	Neighborhood	Latitude	Longitude
0	2.0	RECOLETA	-34.570024	-58.380004
1	5.0	ALMAGRO - BOEDO	-34.614116	-58.412870
2	6.0	CABALLITO	-34.607047	-58.430606
3	7.0	FLORES - PARQUE CHACABUCO	-34.629745	-58.451995
4	9.0	LINIERS - MATADEROS - PARQUE AVELLANEDA	-34.633010	-58.519254
5	10.0	FLORESTA - MONTE CASTRO - VELEZ SARSFIELD - VE...	-34.620160	-58.488337
6	11.0	VILLA DEL PARQUE - VILLA DEVOTO - VILLA GRAL....	-34.596368	-58.498385
7	12.0	COGHLAN - SAAVEDRA - VILLA PUEYRREDON - VILLA ...	-34.593856	-58.503311
8	14.0	PALERMO	-34.552023	-58.426759
9	3.0	BALVANERA - SAN CRISTOBAL	-34.598003	-58.411919
10	15.0	AGRONOMIA - CHACARITA - PARQUE CHAS - PATERN...	-34.596614	-58.426024
11	8.0	VILLA LUGANO - VILLA RIACHUELO - VILLA SOLDATI	-34.651166	-58.449132
12	4.0	BARRACAS - BOCA - NUEVA POMPEYA - PARQUE PATRI...	-34.660917	-58.399077
13	1.0	CONSTITUCION - MONSERRAT - PUERTO MADERO - RE...	-34.573650	-58.368541
14	13.0	BELGRANO - COLEGIALES - NUÑEZ	-34.540224	-58.440331

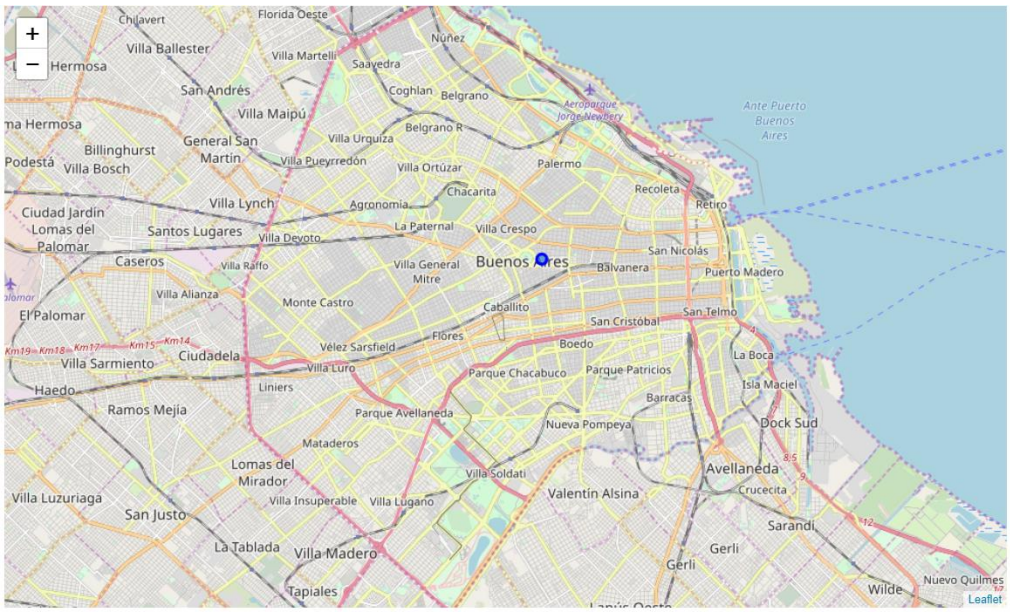
b) Collecting all the necessary data for the analysis

- Location and type (category) of every restaurant in the nearby area of interest (Ciudad de Buenos Aires): Tools used FourSquare API, and Geo libraries

- Reconfirming central point of interest for the analysis: Tools used: Wikipedia sites, and GEO JSON file

CABALLITO, Central point of Ciudad de Buenos Aires Neighborhoods

Location using Folium tool



c) Data exploration, data wrangling/enrichment and data modeling

- Identification of venues in the nearby of the area of analysis (CABALLITO): use of data modeling techniques to breakdown CABALLITO area into sub-areas (matrix area view with centroids equally placed)

- Isolation of the correct venues: filtering data in order to have only restaurant sites, and discard references to other places such as coffee, spa, and so on

- Data modeling and data enrichment in order to simplify multiple restaurant types (19 types) into 3 new created types: Local (Argentinean restaurants), Fast foods, and Ethnical restaurants (where the only one Peruvian restaurant is included)

All Neighborhoods in Ciudad de Buenos Aires

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
AGRONOMIA - CHACARITA - PARQUE CHAS - PATERNAL - VILLA CRESPO - VILLA ORTUZAR	39	39	39	39	39	39
ALMAGRO - BOEDO	15	15	15	15	15	15
BALVANERA - SAN CRISTOBAL	38	38	38	38	38	38
BARRACAS - BOCA - NUEVA POMPEYA - PARQUE PATRICIOS	1	1	1	1	1	1
BELGRANO - COLEGIALES - NUÑEZ	5	5	5	5	5	5
CABALLITO	20	20	20	20	20	20
COGHLAN - SAAVEDRA - VILLA PUEYRREDON - VILLA URQUIZA	17	17	17	17	17	17
CONSTITUCION - MONSERRAT - PUERTO MADERO - RETIRO - SAN NICOLAS - SAN TELMO	1	1	1	1	1	1
FLORES - PARQUE CHACABUCO	29	29	29	29	29	29
FLORESTA - MONTE CASTRO - VELEZ SARSFIELD - VERSALLES - VILLA LURO - VILLA REAL	4	4	4	4	4	4
LINIERS - MATADEROS - PARQUE AVELLANEDA	9	9	9	9	9	9
PALERMO	15	15	15	15	15	15
VILLA DEL PARQUE - VILLA DEVOTO - VILLA GRAL. MITRE - VILLA SANTA RITA	23	23	23	23	23	23
VILLA LUGANO - VILLA RIACHUELO - VILLA SOLDATI	4	4	4	4	4	4

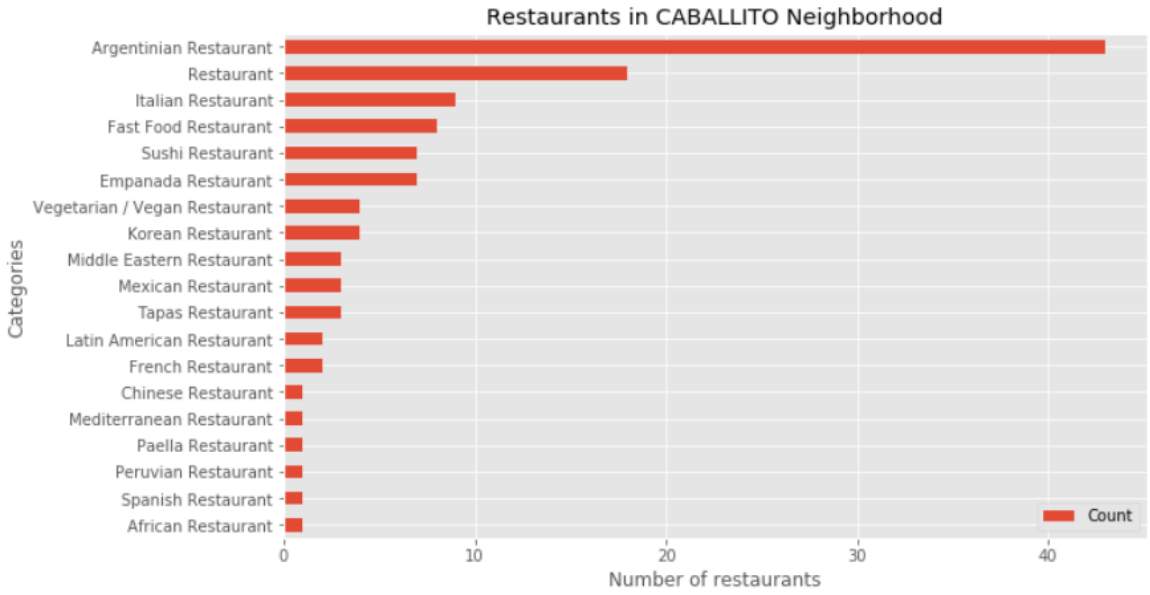
Splitting CABALLITO into 25 sub-areas to allow drill down analysis:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
CABALLITO 1	73	73	73	73	73	73
CABALLITO 10	8	8	8	8	8	8
CABALLITO 11	24	24	24	24	24	24
CABALLITO 12	38	38	38	38	38	38
CABALLITO 13	63	63	63	63	63	63
CABALLITO 14	22	22	22	22	22	22
CABALLITO 15	14	14	14	14	14	14
CABALLITO 16	16	16	16	16	16	16
CABALLITO 17	18	18	18	18	18	18
CABALLITO 18	30	30	30	30	30	30
CABALLITO 19	24	24	24	24	24	24
CABALLITO 2	39	39	39	39	39	39
CABALLITO 20	36	36	36	36	36	36
CABALLITO 21	12	12	12	12	12	12
CABALLITO 22	6	6	6	6	6	6
CABALLITO 23	5	5	5	5	5	5
CABALLITO 24	4	4	4	4	4	4

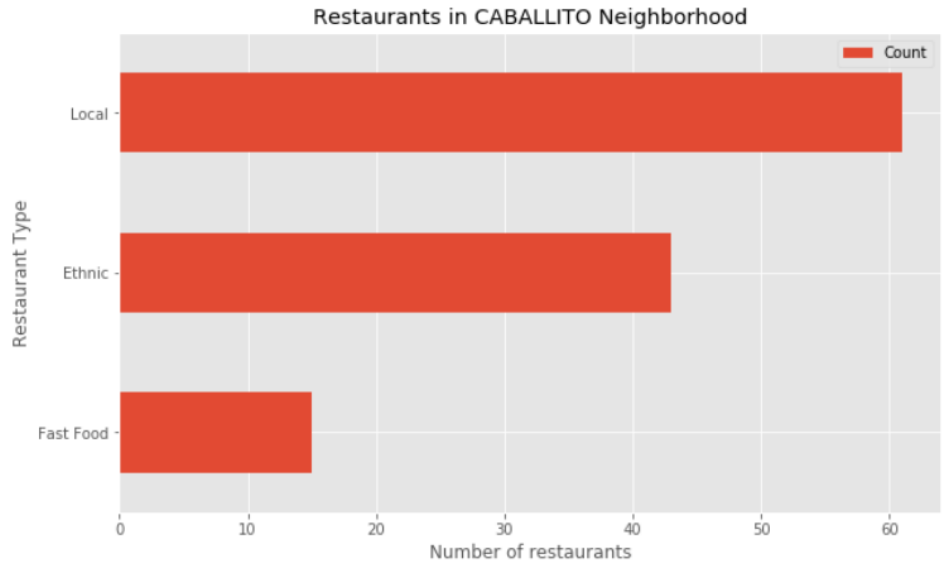
d) Inferential and statistical testing

- Review statistics on restaurants in the target area to do drill down; top down / bottom up analysis comparing local Argentinean restaurants and non-Argentinean restaurants (Mediterranean, Korean, Japanese, Italian, Peruvian, and so on)

119 Restaurants classified in 19 categories according to type of “cousine”

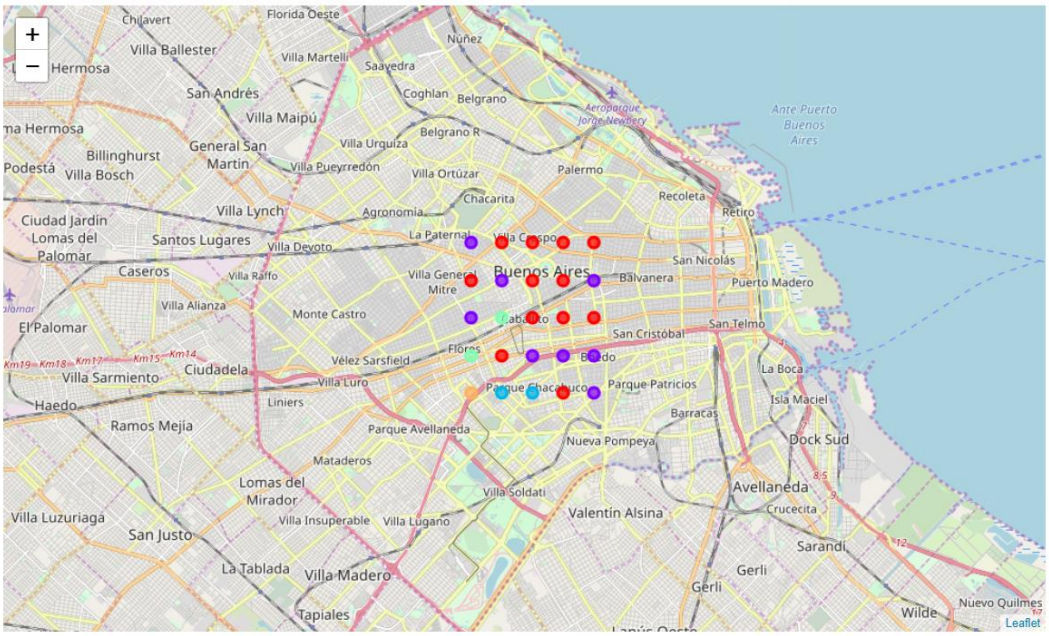


119 Restaurants classified in 3 group types



e) Knowledge Discovery for making recommendation using Machine learning techniques

- Use of k-means clustering in order to identify adequate areas within CABALLITO Comuna for putting a new Peruvian Restaurant. This is to plot at geographical level neighborhood / street the suitable place for starting the new business



Please refer to the project Notebook for more details on how the methodology steps highlighted here have been developed.

Part 1 of 2: https://jp-tok.dataplatform.cloud.ibm.com/analytics/notebooks/v2/98622bb6-170d-4690-83ee-a97dd016cf75/view?access_token=187729180c46e88e38f7bdd12fb3a65dcbbd1f9ca45f2dcde5c28b0e1faf2410

Part 2 of 2: https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/645af8c9-69ed-4be6-a2fe-6b73a83586de/view?access_token=d30464576c0b349116e545e6fde5e3e9888c65462ffdf6d5e0b302694af445cf

Results

- 119 restaurants have been identified in the target area: CABALLITO neighborhood, Ciudad de Buenos Aires, Argentina.
- All of them can be grouped in 19 type of restaurants: Argentinean, Fast Foods, and an important sub-group of Ethnic restaurants (Korean, Italian, Mediterranean, and so on).
- Among the Ethnic restaurants, there is only one **Peruvian restaurant**! This represents an **opportunity** since the target area is a broad one, 2.5 km radius from a CABALLITO Neighborhood central point. There is space to open new business there with a minimal number of competitors.
- Considering the Clustering done in order to determine different groups, the results were as follows:
    - 5 clusters created:
        - Cluster 1: dominated mainly by Argentinean restaurants, with some participation of Ethnic restaurants. There is only one Peruvian restaurant and it is included in this cluster
        - Cluster 2: dominated mainly by Argentinean restaurants, and the second type is related to vegetarian / vegan cousine
        - Cluster 3: dominated by Korean cousine
        - Cluster 4: dominated by Vegetarian and Latin American cousine
        - Cluster 5: dominated by Lain American cousine
- Recommendation for setting up the new restaurant is avoiding Cluster 1

Discussion

Based on the aforesaid results, there are good opportunities for setting up a Peruvian Restaurant in CABALLITO Neighborhood, Ciudad de Buenos Aires Argentina.

Decision on the place should consider not using neighborhood of cluster 1, that cluster is mainly dominated by Argentinean restaurants and some Ethnic restaurants, this is to avoid a competitive scenario with the already existing Peruvian Restaurant in that neighborhood. Other clusters can be considered as potential destination place of the new restaurant.

Further analysis can be done considering other inputs from the stakeholders.

Conclusion

The objective of this project was to find a suitable place for a restaurant in Buenos Aires, Argentina. The project activities were oriented to help stakeholders interested in opening a Peruvian Restaurant in Ciudad de Buenos Aires, Buenos Aires, Argentina.

The main drivers for decision making were:

- places not currently crowded with restaurants
- places in areas with group of ethnic restaurants (but with no Peruvian restaurants)
- places that are as close to city downtown

Applying Data Science Methods, Processes and Techniques, and supported by tools such as Foursquare data, some areas have been identified as adequate places for opening a Peruvian restaurant in Ciudad de Buenos Aires, Argentina, in the nearby of CABALLITO Comuna.

Clustering techniques have helped to indicate potential areas for business development in the restaurant arena. The ultimate word on the most adequate location for opening the new business is in charge of the stakeholders, that can use this work as input in their decision process.

There are a set of opportunities that indicates there are good prospect in the CABALLITO area to set up a new business opening a new Peruvian Restaurant: one 1 competitor, when for other Ethnic restaurants (such as Korean restaurants) there are many.

Further analysis can be performed using more elements that can be added from the stakeholder’s point of view.