Capstone Project - Finding a Place for new Restaurant

Applied Data Science Capstone by IBM/Coursera[1]

Omer Goldbaum

January 6th, 2021

1. Introduction

[2]In this project we will try to find an optimal location for a restaurant. Specifically, this report will be targeted to stakeholders interested in opening an **Argentinian restaurant**, or **BBQ restaurant** in **Argentina**.

Figure 1 - Argentinian original meat restaurant

Argentina is well known with its great meat, and BBQ restaurant. Specifically, there are some main cities in Argentina that became very tourist one, and have a lot of restaurants, including ones that offer local Argentinian food. Since there are lots of restaurants in Argentina, and lots of tourist cities, we will try to detect locations that are not already crowded with restaurants, or with this specific Argentinian meat. We would also prefer locations **as close to city center as possible**, assuming that first two conditions are met.

We will use our data science powers to generate a few most promising neighborhoods based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

[1] This project was created with an help of the example project of IBM "The Battle of the Neighborhoods"
[2] https://pickupthefork.com/2010/05/08/slap-a-steak-on-the-parrilla/

2. Data

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants near the city center (any type of restaurant)
- number of and distance to Argentinian or BBQ restaurants near the city center

Following data sources will be needed to extract/generate the required information:

- CSV data for Argentina's main cities
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **geopy.geocoders**
- number of restaurants and their type and location in every city will be obtained using **Foursquare API**

2.1 Data Cleaning

We will use a CSV file from the web page https://simplemaps.com/. This file contains data for 503 prominent cities in Argentina, including 'name', 'lat', 'lon', 'population' (updated to 2020) for each entry.

In order to reduce the number of interesting cities, we decided to look only at cities with population of at least 50K people. We got 70 cities in Argentina that might be interesting:

Figure 2 Map of the big cities in Argentina

Once that we have our cities, we used Foursquare API to get info on restaurants. We're interested in venues in 'food' category, but only those that are proper restaurants - coffee shops, pizza places, bakeries etc. are not direct competitors so we don't care about those. So, we will include in our list only venues that have 'restaurant' in category name, and we'll make sure to detect and include all the subcategories of specific **Argentinian restaurant** or **BBQ** category, as we need info in the city. Also, we will check those venues within a **radius of 1000 meters** from the cities center.

After we got our relevant restaurants, we calculated the distance from each restaurant to its corresponding city center.

	City	City Latitude	City Longitude	restaurant	restaurant Latitude	restaurant Longitude	restaurant Category	dist2center
0	Buenos Aires	-34.60828	-58.372295	Pertutti	-34.609195	-58.373392	Argentinian Restaurant	142.875942
2	Buenos Aires	-34.60828	-58.372295	Piazzolla Tango	-34.606099	-58.374899	Theme Restaurant	340.024866
3	Buenos Aires	-34.60828	-58.372295	FuraiBo	-34.610105	-58.372549	Japanese Restaurant	203.750072
7	Buenos Aires	-34.60828	-58.372295	Elauge Hermanos	-34.609276	-58.375499	BBQ Joint	313.927511
8	Buenos Aires	-34.60828	-58.372295	Bidou de la Merced	-34.605619	-58.373134	Restaurant	305.112268
...	...	...	...	...	...	...	...	...
1041	Ushuaia	-54.80720	-68.304400	Parrila La Rueda	-54.805330	-68.301049	Steakhouse	299.601331
1043	Ushuaia	-54.80720	-68.304400	Bamboo	-54.805051	-68.299648	Chinese Restaurant	388.071016
1050	Azul	-36.78330	-59.850000	Parrilla El Carrito	-36.789406	-59.858032	BBQ Joint	986.483509
1054	Cruz del Eje	-30.73330	-64.800000	Viejo Munich	-30.736909	-64.791571	German Restaurant	900.942656
1055	Cruz del Eje	-30.73330	-64.800000	El Cardon	-30.737260	-64.791606	BBQ Joint	915.979304

535 rows × 8 columns

Table 1 – Part of the table that contains all the relevant 535 restaurants in 70 cities in Argentina

Figure 3 - An example of the restaurant's span in the city of Bariloche; Argentinian meat restaurants (in red) vs. regular restaurants (in blue). The city center is signed with a black mark

3. Methodology

In this project we will direct our efforts on detecting cities in Argentina that have low restaurant density near the city center, particularly those with low number of Argentinian restaurants. We will limit our analysis to area ~1km around city centers.

In first step we have collected the required data: **location and type (category) of every restaurant within 1km from the cities center**. We have also identified local meat restaurants (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of **restaurant density** in every city - we will use **heatmaps** to identify a few promising areas close to center with low number of restaurants in general (and no Argentinian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 250 meters, and we want locations without Argentinian restaurants in radius of 200 meters. We will present map of all such locations but also create clusters using k-means clustering of those locations to identify general zones which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis

We create a map showing heatmap / density of restaurants in every city:

Figure 4 - Example of restaurant density in three cities Córdoba (left), Buenos Aires (middle) and Salta (right)

Córdoba

We decide to focus on the city of Córdoba, the second main city in Argentina. The city of Córdoba located in Argentina is the capital of Córdoba province. It can be found almost in the center of Argentina and is the second most populated city in the whole country. The city is most known for its historical places and culture which can be experienced throughout the whole vacation. Tourists learn about history and the culture of the people without even noticing or making an effort[3].

[3]Córdoba: The City Where Culture Stands Out, Jun 26, 2018, https://skyticket.com/guide/2847

Using Foursquare, there are 9 local restaurants in Córdoba, 22.0% from all the restaurants near the centre (41).

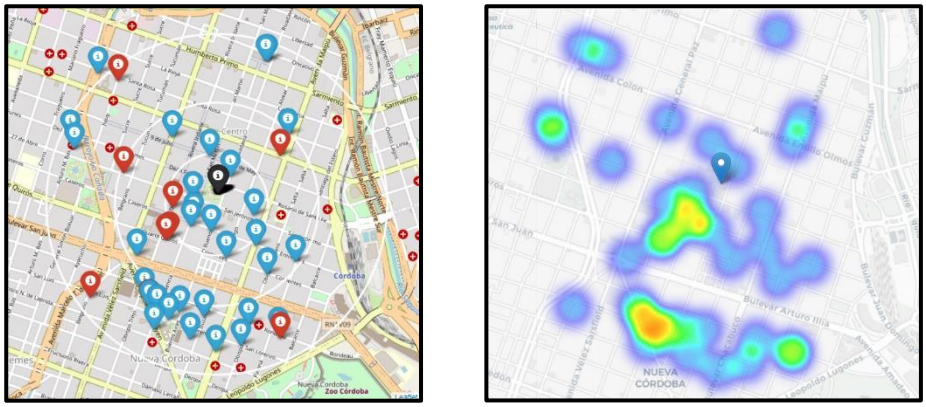

*Figure 5 – **left**: The restaurant's span in the city of Córdoba; Argentinian meat restaurants (in red) vs. regular restaurants (in blue). The city center is signed with a black mark; **right**: Heatmap zoom-in*

This map is not so 'hot' - Argentinian restaurants represent a subset of ~22% of all restaurants in Córdoba (41), but it also indicates higher density of restaurants at the inner radius (300m from the center), and on the southern part within 600-900m from the center. The closest pockets of low restaurant density positioned west and north from city center. Based on this we will now focus our analysis on the western and north areas from Córdoba's center.

We calculated the number of restaurants in vicinity at a radius of **250 meters**, and distance to closest Argentinian restaurant. Then, we filtered those locations, and get only the locations with no more than two restaurants in radius of 250 meters, and no local restaurants in radius of 300 meters.

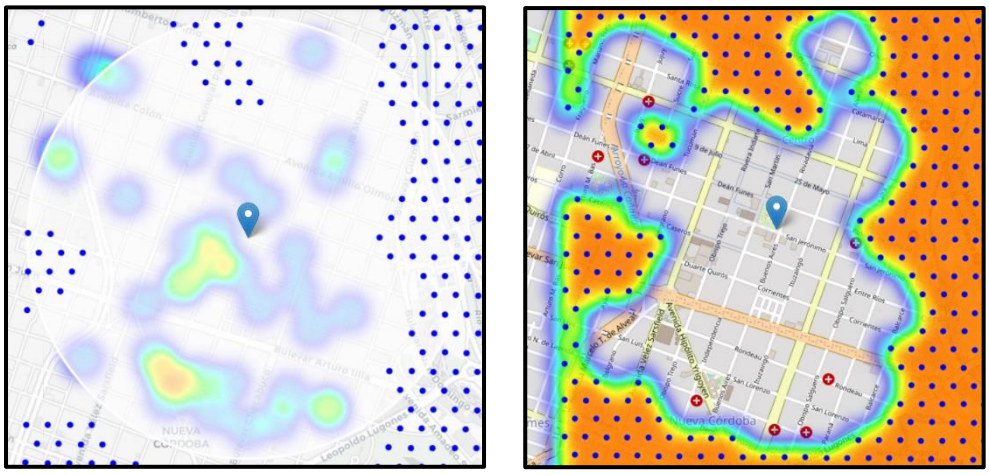

Figure 6 - The blue dots represent the location that satisfied the two conditions

We now have a bunch of locations fairly close to Córdoba's center, and we know that each of those locations has no more than two restaurants in radius of 250m, and no local restaurant

closer than 300m. Any of those locations is a potential candidate for a new Argentinian restaurant, at least based on nearby competition.

Now, we cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis. We do this using **K-mean algorithm** to find those centers. Addresses of those cluster centers will be a starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Figure 7 - The centers of the good locations around the city center

We found 4 locations' centers within 1.5km from the city center:

a. Avenida Pueyrredón, 176 (1.3km)

b. Avenida 24 de Septiembre, 843 (1.2km)

c. Bahía Blanca, 899 (1.3km)

d. Avenida Intendente Mestre Norte, 444 (1.3km)

Figure 8 - Güemes Neighborhood

The neighborhood of Güemes, near Avenida Pueyrredón, will be a good place to be. It is a cool place where visitors spend at least an afternoon. Filled with fancy local bars, trendy cafes, and restaurants, this neighborhood will keep you busy. Güemes is a bohemian style, probably the hippiest and coolest place in Córdoba. Muy Güemes offers many boutiques to see, delicious food joints to try and shopping areas to check out. Strolling down the streets or sitting in a restaurant in Güemes is a refreshing activity in this very alive neighborhood[4].

[4] https://skyticket.com/guide/2847#articleHead_7

5. Results and Discussion

Our analysis shows that there are pockets of low restaurant density fairly close to city center of Córdoba. Highest concentration of restaurants was detected 600-900 meters from the center, so we focused our attention to areas at the outer radius, but still close - about 1 km from the center, corresponding to borough Güemes south-west corner of center. Another borough was identified as potentially interesting (Nueva Córdoba), but our attention was focused on Güemes which offers a combination of popularity among tourists, closeness to city center, and pockets of low restaurant density.

After directing our attention to this narrower area of interest we created a dense grid of location candidates (spaced 100m apart); those locations were then filtered so that those with more than two restaurants in radius of 250m and those with an Argentinian meat restaurant closer than 300m were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 4 zones containing largest number of potential new restaurant locations based on number of and distance to existing venues - both restaurants in general and Argentinian restaurants particularly. This, of course, does not imply that those zones are actually optimal locations for a new restaurant. Purpose of this analysis was to only provide info on areas close to Córdoba center but not crowded with existing restaurants (particularly Argentinian) - it is entirely possible that there is a very good reason for small number of restaurants in any of those areas, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6. Conclusion

The purpose of this project was to identify Cities in Argentina that has good potential to open new Argentinian meat restaurant. Then we focused our analysis to the city of Córdoba, in order to find areas, close to center with low number of restaurants (particularly Argentinian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Argentinian meat restaurant. By calculating restaurant density distribution from

Foursquare data, we have first identified cities that justify further analysis (we continued with Córdoba, but we could choose other city as well), and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.