BATTLE OF NEIGHBORHOODS: BUENO AIRES

DATA SCIENCE CAPSTONE

Abstract

Finding the optimal location for a new Restaurant in Buenos Aires by analyzing Neighborhood, Venues, Population and socioeconomic data.

Coursera IBM Capstone Project

Table of Contents

Introduction ***2***

Data ***3***

Methodology ***4***

Finding data **4**

Data Cleaning **4**

Data Analysis **5**

Results ***6***

Conclusion ***8***

Introduction

People in Argentina like good gastronomic quality. Going to restaurants and sharing a meal among friends and family, spending a whole afternoon chatting while enjoying a post-meal coffee is one the most Argentina traditions of all time. For this reason, restaurants are a big deal.

With this in mind, a Chef who lives in Buenos Aires, Argentina, has won a cooking competition and received some price money to set up his Restaurant. However, he is not sure where he can set it up.

The goal of the project is to select an optimal location for his new restaurant, and to find this answer we will try to answer the following questions

- How many venues in each neighborhood?
- How many categories in neighborhood?
- How many venues in each neighborhood?
- What are the most popular categories in each neighborhood?

Also, it is important to mention that the price is to build the "Restaurant of your dreams Prize" and the chef has an unlimited budget.

Data

To find the location for the new restaurant we will be using public available data to answer question that can lead us for best possible location for the restaurant.

1. List of Neighborhoods of Buenos Aires

2. Lists of Food Venues in Buenos Aires

3. Socio-Economic data of Buenos Aires

Below a data description is included

List of Neighborhoods of Buenos Aires

- A list of all neighborhoods will allow us to understand the different areas of the city where the restaurant can be set

Food Venues

- It is important to understand what other types of restaurants are out there so as not to create a price war against incumbent venues

Socio-Economic Data

- This data will allow us to learn in what areas of the city has higher income and have an idea where would it be more suitable to put the restaurant

Methodology

Finding data

We obtained a List of Neighborhoods of Buenos Aires from the Wikipedia site: https://en.wikipedia.org/wiki/Neighbourhoods_of_Buenos_Aires

The Food Venues data was obtained by requesting it to the Buenos Aires Chamber of Commerce

Socio-Economic data: this kind of data can be found in the INDEC Socio-Economic indicators site. INDEC is the Institute of Statistics and Census from Argentina.

Data Cleaning

Data cleaning was performed in the IBM Watson studio using python. After cleaning the data, data frames for each were created. The following snapshots illustrate the data cleansed that resulted.

3]:

	Name	Area in km2	Population	Avg Income
0	Almagro	4.1	128206	46150
1	Barracas	7.6	73377	41200
2	Belgrano	6.8	126816	87200
3	Boedo	2.6	45563	62450
4	Caballito	6.8	170309	65300
5	Flores	7.8	142695	34750
6	La Boca	3.1	43413	35000
7	La Paternal	2.2	19058	32750
8	Liniers	4.3	42083	36600
9	Mataderos	7.3	62206	35000

	Restaurants	Number
0	Chinese	27
1	Italian	242
2	Vietmamese	7
3	Grill-BBQ	252
4	Home food	97
5	Mexican	12
6	Indian	3
7	Sushi	27

Data Analysis

Once the data was cleaned, we proceed to the data analysis. In this part of the project we focus in following a logic to finding the best location for the new restaurant. The approach taken was to first understand what factors were most important to the population and then how the population was formed in different neighborhoods of the city.

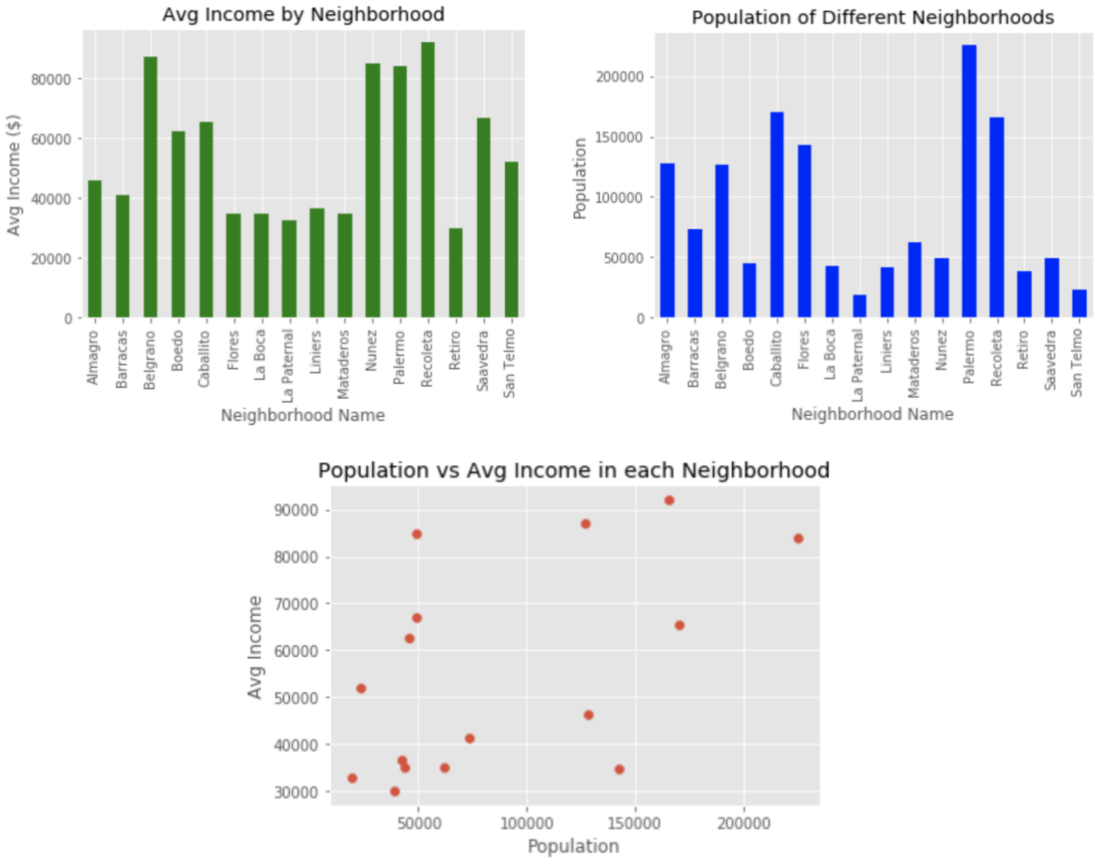
By finding the distribution of different types of cuisine, the amount of people living and that frequent each part of the city, as well as the avg income per household we can take an educated decision and selection of a good place to set up the restaurant.

To achieve this, histograms were plotted and the mean avg income for the population, segmented by neighborhood were obtained.

Then after finding the best type of venue, and the larger population and the largest incomes, we selected an area where the intersection of the two latter groups was represented.

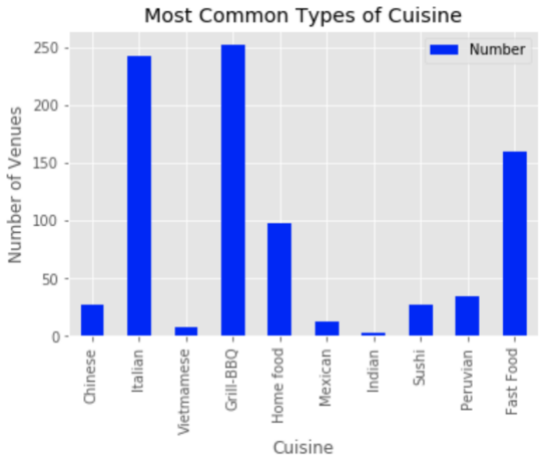
Furthermore, we also focused the area so that it would have major foot traffic and frequent visitors from other neighborhoods during a given week.

Results

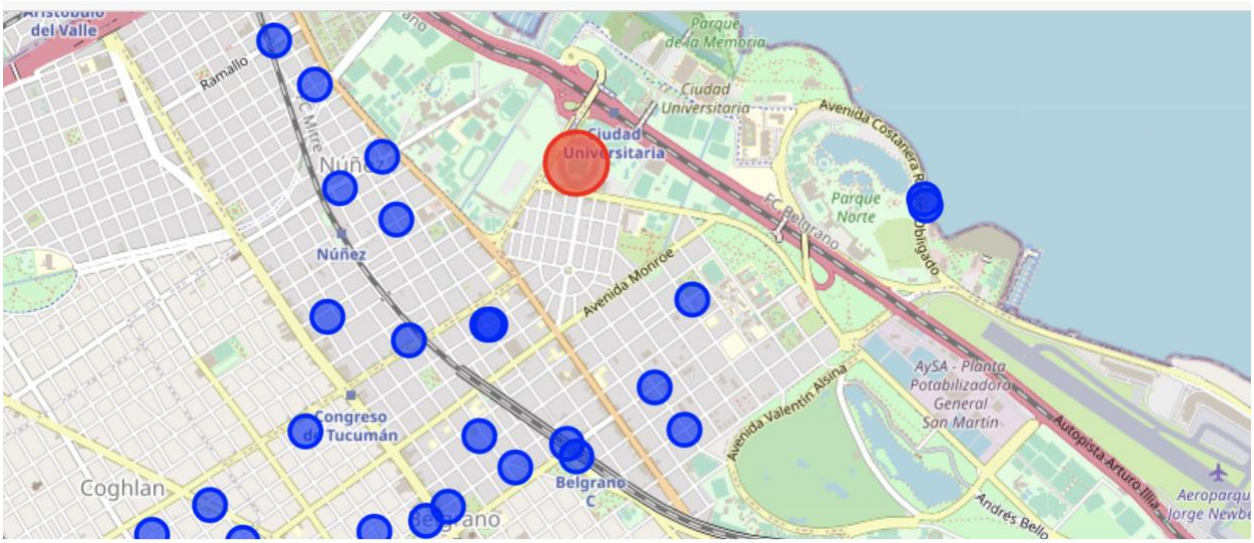


From the generated graphs we can see that Belgrano, Palermo and Recoleta are High-Income and Highly Populated areas. From this insight we can infer that these neighborhoods would be candidates to set up a restaurant there.

We can see that the Italian and Grill/BBQ cuisines are very predominant against other cuisines, like Indian, Sushi or Vietnamese. Fast food has a relevant share, but still cannot compete against the top 2.



Moreover, Football/Soccer is a very popular sport and people go to the stadiums every weekend to see their Team. Our friend would like his restaurant to be located close the Biggest football stadium, this being River Plate's Stadium with a capacity of over 70k spectators. Around this area there is an enormous amount of foot traffic and potential customers during the weekend. This happens to be the Belgrano Neighborhood, with 127K people and $87,200 of avg yearly income.



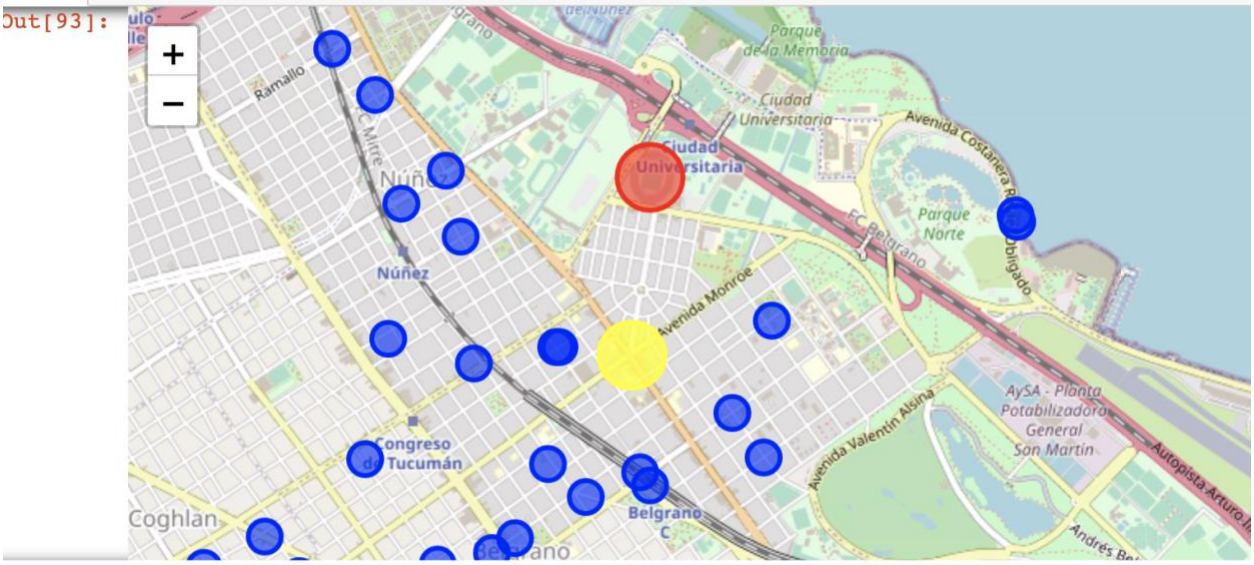
This map picture shows the River Plate Stadium and the different grill venues given by the Four-Square database.

Conclusion

The Best Location for the restaurant was found to be the intersection of Av. Monroe and Av. Libertador.

```
[93]: # display map
      venues_map
```

Out[93]:



Now, we can see in the map where the new restaurant shoud be located

Based on the map and the distribution of other grills around the River Plate Stadium, we chose the best new location for the new restaurant that would be closest to the stadium and closest to main foot traffic areas.

The intersection of Av. Monroe, and Av. Libertador seems like an excellent place. Also, this location has stops for all major buses and is close to the D subway line, that connects downtown Buenos Aires to most northern residential areas.