**Data Science Capstone Project**

**Housing Prices and Social Venues of Buenos Aires City**

***By Matias Rapagna***

***June 22 2020***

# Index

[1. Index 2](#_Toc43736921)

[2. Introduction 3](#_Toc43736922)

[2.1 Description & Disscusion of the Background 3](#_Toc43736923)

[2.2 Data Description 3](#_Toc43736924)

[3. Methodology 4](#_Toc43736925)

[4. Discussion 11](#_Toc43736926)

[5. Conclusion 12](#_Toc43736927)

[6. References: 13](#_Toc43736928)

# 

# Introduction

## Description & Disscusion of the Background

Greater Buenos Aires, also known as the Buenos Aires Metropolitan Area, refers to the urban agglomeration comprising the autonomous city of Buenos Aires and the adjacent 24 partidos (districts) in the Province of Buenos Aires. Thus, it does not constitute a single administrative unit. The conurbation spreads south, west and north of Buenos Aires city. To the east, the River Plate serves as a natural boundary.

Urban sprawl, especially between 1945 and 1980, created a vast conurbation of 9,910,282 inhabitants in the 24 conurbated partidos, as of 2018, and a total of 14,967,000 including the City of Buenos Aires, a third of the total population of Argentina and generating more than half of the country's GDP. This makes Buenos Aires, the 13th largest city in the world.

To try to simplify the exercise, I will focus only in the City of Buenos Aires, the capital of Argentina. It is divided in 48 districts, has a 203 square kilometers of area, and 3,075,646 inhabitants, with a population density of 15,151 hab/m2.

As you can see from the figures, Buenos Aires is a city with a high population and population density. When making a decision to purchase a house or apartment in the city, investors would want to choose a district with low property value, but with hi social venues density. Today, there is no consolidated information regarding this topic.

When we consider these problems, we can create a map and information chart where the real estate index is placed on Buenos Aires and each district is clustered according to the venue density.

## Data Description

To consider the problem we can list the datas as below:

* I have found Geo Referential Data regarding districts in Buenos Aires, from a National Government site <https://datosgobar.github.io/georef-ar-api/download/> where you can obtain json files to use in your code. I obtained also the information regarding each district to use as a base data frame. I will obtain 2 files, one is a GeoJson to be able to draw Choropleth map of the districts in Buenos Aires based on the prices of Sq M2. The other file contains latitude and longitude data for each of the districts.
* I had to create an Excel data with the average price per square meter in each of the districts. This was based on information obtained by <https://data.buenosaires.gob.ar/>, from the City of Buenos Aires government. This file will be merged with the Location information of each of the districts and the top 10 venues.
* I used **Forsquare API** to get the most common venues of given each of the districts in Buenos Aires.
* I also used Wikipedia information regarding general data of Buenos Aires City, Greater Buenos Aires, and other useful data.