

# Housing Sales Prices & Venues Data Analysis of Mexico City

Jesús Alfonso Juárez

February 2020

## 1.- Introduction

### 1a.-Description and discussion of the background

Greater Mexico City is the second largest metropolitan area of the western hemisphere and the largest spanish-speaking city in the world with **21.3** million of population. Mexico City has by itself **9 million** people gathered in just **1,485 square kilometers** turning into a high-density zone with **6,000 persons** by square kilometer. Mexico has the history embedded in their walls, originally named **Mexico Tenochitlan** by the aztecs has been witness of many stages from the pre-Hispanic to the modern era. Currently, the city is formed by **16 boroughs**.<sup>[1]</sup>

Mexico City is considered a megacity which means that is a high population density zone. Thus, there is a restricted supply of commercial and residential real estate. Moreover, the tendency to the vertical urbanization and the new structures of families demand a new approach in the housing sector. The city residents are seeking zones near to their jobs, with the venues that they attend and where the real estate values are lower, and Furthermore, investors are seeking to establish business in the neighborhoods with lower cost and less competition in the district.

Nowadays does not exist a tool that lead investors and city residents to make a data-based decision of the neighborhood to select. Consequently, we can create a map and information chart where the real estate index is placed on Mexico City and each district is clustered according to the venue density.

### 1b.-Data Description

To solve the problem, we can list the data as below:

- I found the Neighborhoods Coordinates of Mexico City in the Data Repository of the Mexico City government website [2]. The .geojson has the coordinates of all the districts and boroughs ('Delegaciones') of Mexico City
- I used Forsquare API to get the most common venues of given Borough ('Delegación') of Mexico City [3].
- The real estate as other markets has a widespread range of prices in similar housing, thus there is a myriad of information regarding the real estate costs. To overcome this issue, we are going to use the latest square meter Housing Sales Price (HSP)

Average for each Borough ('Delegación') of Mexico City retrieve from the real state retail web page [4].

## **1c.-Data Usage**

The approach that we are going to take with the different sources of information is:

- 1.-The geojson file will be uploaded to the Jupyter notebook file.
- 2.-The information retrieve of Metroscubicos website will be compiled in a csv file then uploaded to the Jupyter notebook file.
- 3.- The gejson file and the csv files are going to pass through a cleansing stage in order to homologue the Boroughs names, and other fields.
- 4.- Afterwards, we are going to set a panda's data frame with the neighborhood name, coordinates, Borough, postcode.
- 5.- Leveraging the foursquare API we are going to retrieve the closest venues to each district in a radio of 700 meters.
- 6.- Finally, we are going to transform this last panda's data frame, establishing the venues categories as columns with the get\_dummies method and grouping by the neighborhood. The resultant data frame will be our input for the k-mean cluster method.

## **2.- References**

[1] Mexico City recover from [Wikipedia](#) February 2020

[2] Coordinates of neighborhoods in Mexico City recover from [CDMX government website](#) February 2020

[3] [Foursquare API](#)

[4] Housing square meter average sales prices of each Borough recover from [Metroscubicos](#) February 2020