# **Capstone Project - The Battle of the Neighborhoods**

### **Applied Data Science Capstone by IBM / Coursera for Pedro G Jimenez Gutz**

1. **Introduction: Business Problem**

Mexico City abbreviated as CDMX, is the capital city of Mexico and the most populous city in North America. The city has 16 subdivisions, formerly known as boroughs. Only 1 in 10 Mexicans has insurance for private medical expenses. Insurance companies in the health sector have the opportunity to increase the percentage of the insured population for the private sector. The public health sector is applying the strategy of improving existing capacity, or temporarily reconverting internal areas to deal with the covit-19 pandemic global.

The purpose of this project is to identify boroughs in CDMX, with a low number of public hospitals. Health is a priority issue, and it is essential to have effective universal medical coverage. The challenge is to expand the health system. Where should there be another hospital?

**2. Data**

**2.1 Data Acquisition**

Data requirements to solve the problem:  
List of CDMX boroughs with population density and coordinates.  
List of public sector hospitals.  
Search for health interest venues in each borough.

Hospitals and Health Centers in Mexico City. <https://datos.cdmx.gob.mx/explore/dataset/hospitales-y-centros-de-salud/>

Wikipedia.

<https://es.wikipedia.org/wiki/Ciudad_de_M%C3%A9xico>

<https://es.wikipedia.org/wiki/Anexo:Delegaciones_de_la_Ciudad_de_M%C3%A9xico_por_IDH>

Google Geocoding API.

Foursquare API.

## 3. Methodology

The methodology in this project consists of two parts:

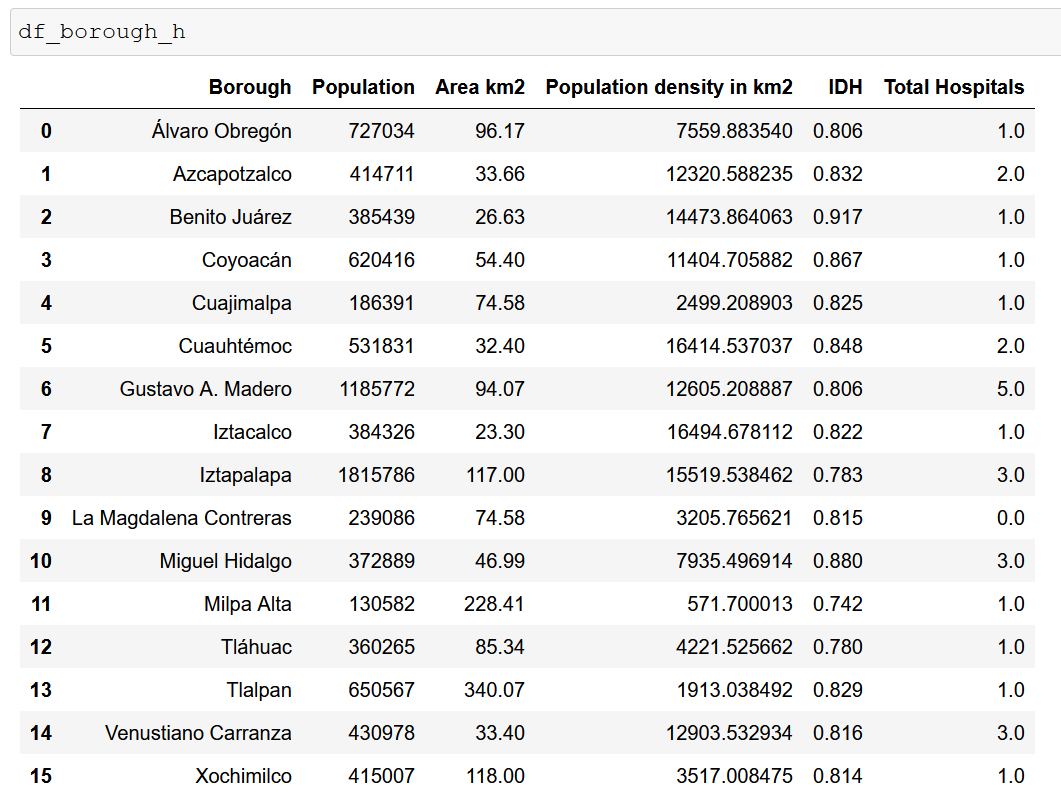
[Exploratory Data Analysis](http://localhost:8888/notebooks/Downloads/IBM%20Practicas%20Python%20Notebooks%20-%20cert1/Capstone%20Project/project_final.ipynb#part1) View the total of Hospitals by boroughs in CDMX, and extract neighborhoods of the five selected boroughs.

[Modelling](http://localhost:8888/notebooks/Downloads/IBM%20Practicas%20Python%20Notebooks%20-%20cert1/Capstone%20Project/project_final.ipynb#part2) To help stakeholders interested in knowing, using or improving the hospital health system in CDMX. We will use K-means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 5 for this project.

### **Exploratory Data Analysis**

Captura de pantalla de un celular

Descripción generada automáticamente



Una captura de pantalla de una red social

Descripción generada automáticamente

Imagen que contiene texto, mapa

Descripción generada automáticamente

### **Modelling**

**Finding all the venues within a 3000 meter radius of each neighborhood. Perform one hot ecoding on the venues data. Grouping the venues by the neighborhood and calculating their mean. Performing a K-means clustering (Defining K = 5**

Imagen que contiene captura de pantalla

Descripción generada automáticamente

Captura de pantalla de un celular con letras

Descripción generada automáticamente

Captura de pantalla de computadora

Descripción generada automáticamente

**Clusters**

Una captura de pantalla de una red social

Descripción generada automáticamente

Una captura de pantalla de una red social

Descripción generada automáticamente

Una captura de pantalla de una red social

Descripción generada automáticamente