

# **Prediciendo la mejor colonia para abrir una pizzería en Villahermosa, Tabasco**

**Emmanuel Rodriguez Zamora**

**21 /Mayo/2020**

## **Introducción**

### **Antecedentes**

Actualmente en el mundo existe una enorme tendencia al emprendimiento, lo cual conlleva la creación de pequeñas empresas. Dentro del grupo de emprendedores siempre existen aquellos que desean abrir restaurantes con sus habilidades de cocina o recetas secretas de sus propias familias. Las pizzerías son sitios de comida en los que se pueden obtener muchas ganancias, esto debido a el alimento en sí mismo (considerado como el más adictivo del mundo), por lo que es una gran idea si se piensa en poner un restaurante. Sin embargo, existe una gran cantidad de cadenas multinacionales que es común que absorban la mayoría los clientes, por lo que es ideal estar alejado de estos lugares.

### **Problema**

A través de los datos podemos elegir los lugares ideales para poner una pizzería en la ciudad, esto teniendo en cuenta la cercanía de las otras pizzerías y la densidad de población en el lugar. El fin del proyecto es identificar los mejores sitios para poner una pizzería en la ciudad de Villahermosa, Tabasco.

### **Interesados**

Todas las personas interesadas en realizar un emprendimiento con un restaurante que su platillo principal sea la pizza.

## **Adquisición y limpieza de datos**

### **Fuente de datos**

Los datos de los nombres y códigos postales de las colonias de Villahermosa se obtuvieron de la página oficial del estado de Tabasco donde se enlista toda la información [aquí](#). A partir de los códigos postales se obtuvieron la población y el área de cada uno de ellos en página de códigos postales nacional [aquí](#). Después de tener la información básica de cada colonia se obtuvo la ubicación dada en latitud y longitud de cada una de las colonias a través de la librería geopy. Finalmente, la información concerniente a las pizzerías se obtuvo a través de la página Foursquare.

### **Limpieza de datos**

Se realizaron una gran cantidad de acciones para lograr una limpieza de datos satisfactoria. Inicialmente los datos obtenidos a través de las APIs se encuentran en formato json, por lo que se tiene que filtrar la información deseada de cada petición, para crear las tablas deseadas.

Una vez obtenidos todos los datasets, se procedió a eliminar las colonias de las que no se logró obtener información útil como la ubicación o la cantidad de población. Además de ello, después de realizar diversas pruebas con la API de geopy se llegó a la conclusión de que el artículo “de” impedía a la librería diversas ciudades, por lo que se eliminó de todas las colonias encontradas en el dataframe.

En las mediciones de área se eliminaron las letras representativas de la dimensión de los datos y se estandarizaron todos los datos ya que algunos se encontraban en metro mientras que otros estaban en kilómetros. Finalmente, los dos dataframes referentes a las colonias se unieron a partir de los códigos postales existentes.

## **Selección de características**

A partir de los datos en el dataframe, debido a las variantes de área que posee cada código postal, se decidió unir al conjunto de datos una variable que relacionara el área y la población, por lo que se obtuvo la densidad de población de cada colonia. Para la realización del análisis es necesario seleccionar únicamente las características numéricas, debido a que no había variables categóricas en este dataframe. Además de ello, se realizó el cálculo de la cantidad de pizzerías cercanas a cada colonia en un radio de 700 metros para ayudar a escoger los lugares adecuados. Por lo tanto, se escogieron las siguientes características para pasar la siguiente fase:

- Posición en x de la colonia
- Posición en y de la colonia
- Población por código postal
- Área por código postal
- Densidad por código postal
- Cantidad de pizzerías 700 metros a las redondas