

Análisis de nuevas ubicaciones de CAP's sugeridas utilizando el Modelo de K-Means

Javier Horacio Pérez Ricárdez

Marzo del 2025

Abstract

En este documento se presenta un análisis para sugerir ubicaciones óptimas en los 32 estados de la República Mexicana, utilizando el modelo de **K-Means**. Se consideran datos de la Población Económicamente Activa (PEA) por municipio, junto con las coordenadas geográficas (latitud y longitud en formato decimal) de los municipios, así como la ubicación de diversos Centros de Atención al Público (CAP) de las AFOREs.

Introducción

El objetivo de este análisis es determinar la ubicación óptima para centros de atención al público utilizando el modelo de **K-Means**, un algoritmo de aprendizaje no supervisado que permite agrupar datos basándose en su similitud. Se trabaja con la Población Económicamente Activa (PEA) por municipio y las coordenadas geográficas de los mismos. Los datos de entrada incluyen ubicaciones de Centros de Atención al Público (CAP) de diversas Administradoras de Fondos para el Retiro (AFORE), tales como:

- AZTECA
- CITIBANAMEX
- COPPEL
- INBURSA
- INVERCAP
- PENSIONISSSTE
- PRINCIPAL
- PROFUTURO
- SURA
- XXI-BANORTE

El modelo de K-Means se emplea para identificar la ubicación óptima basada en la concentración de la PEA y la distribución geográfica de los municipios.

Modelo de K-Means

El modelo de **K-Means** es un algoritmo de clustering que busca agrupar un conjunto de datos en k clusters, basándose en la similitud de los puntos. En este caso, se utiliza $k = 1$ para encontrar el centroide que representa la ubicación óptima. El algoritmo sigue los siguientes pasos:

1. **Inicialización:** Seleccionar k centroides iniciales (en este caso, $k = 1$).
2. **Asignación:** Asignar cada punto de datos al centroide más cercano.
3. **Actualización:** Recalcular los centroides como el promedio de todos los puntos asignados a cada cluster.
4. **Iteración:** Repetir los pasos 2 y 3 hasta que los centroides converjan.

Formulación Matemática

1. ****Distancia Euclidiana**:** La distancia entre un punto x_i y un centroide c_j se calcula como:

$$d(x_i, c_j) = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2}$$

Donde x_{i1} y x_{i2} son las coordenadas (latitud y longitud) del punto x_i , y c_{j1} y c_{j2} son las coordenadas del centroide c_j .

2. ****Función de Coste**:** El objetivo del algoritmo es minimizar la suma de las distancias cuadradas entre cada punto y su centroide asignado:

$$J = \sum_{i=1}^n \sum_{j=1}^k d(x_i, c_j)^2$$

3. ****Centroide Óptimo**:** El centroide óptimo c_j se calcula como el promedio de todos los puntos asignados al cluster j :

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Donde C_j es el conjunto de puntos asignados al cluster j .

Aplicación del Modelo

En este análisis, el modelo de K-Means se aplica de la siguiente manera:

- **Datos de Entrada:** Latitud, longitud y PEA de cada municipio.
- **PEA Normalizada:** Para ponderar la influencia de cada municipio, se normaliza la PEA:

$$\text{PEA}_{\text{normalizada}} = \frac{\text{PEA}_i}{\sum_{j=1}^n \text{PEA}_j}$$

- **Coordenadas Ponderadas:** Las coordenadas de cada municipio se ponderan por su PEA normalizada:

$$\text{Latitud}_{\text{ponderada}} = \text{Latitud}_i \times \text{PEA}_{\text{normalizada}}$$

$$\text{Longitud}_{\text{ponderada}} = \text{Longitud}_i \times \text{PEA}_{\text{normalizada}}$$

- **Ubicación Sugerida:** La ubicación sugerida es el promedio ponderado de las coordenadas de los municipios:

$$\text{Latitud}_{\text{sugerida}} = \sum_{i=1}^n \text{Latitud}_{\text{ponderada}}$$

$$\text{Longitud}_{\text{sugerida}} = \sum_{i=1}^n \text{Longitud}_{\text{ponderada}}$$

Ejemplo Numérico

Supongamos que tenemos los siguientes datos para tres municipios en un estado:

Municipio	Latitud	Longitud	PEA
Guadalajara	20.6597	-103.349	5000
Zapopan	20.7230	-103.385	3000
Tlaquepaque	20.6400	-103.310	2000

Calcular la PEA Normalizada

$$\text{PEA}_{\text{total}} = 5000 + 3000 + 2000 = 10000$$

$$\text{PEA}_{\text{normalizada}}(\text{Guadalajara}) = \frac{5000}{10000} = 0.5$$

$$\text{PEA}_{\text{normalizada}}(\text{Zapopan}) = \frac{3000}{10000} = 0.3$$

$$\text{PEA}_{\text{normalizada}}(\text{Tlaquepaque}) = \frac{2000}{10000} = 0.2$$

Calcular las Coordenadas Ponderadas

$$\text{Latitud}_{\text{ponderada}}(\text{Guadalajara}) = 20.6597 \times 0.5 = 10.3298$$

$$\text{Longitud}_{\text{ponderada}}(\text{Guadalajara}) = -103.349 \times 0.5 = -51.6745$$

$$\text{Latitud}_{\text{ponderada}}(\text{Zapopan}) = 20.7230 \times 0.3 = 6.2169$$

$$\text{Longitud}_{\text{ponderada}}(\text{Zapopan}) = -103.385 \times 0.3 = -31.0155$$

$$\text{Latitud}_{\text{ponderada}}(\text{Tlaquepaque}) = 20.6400 \times 0.2 = 4.1280$$

$$\text{Longitud}_{\text{ponderada}}(\text{Tlaquepaque}) = -103.310 \times 0.2 = -20.6620$$

Calcular la Ubicación Sugerida

$$\text{Latitud}_{\text{sugerida}} = 10.3298 + 6.2169 + 4.1280 = 20.6747$$

$$\text{Longitud}_{\text{sugerida}} = -51.6745 + (-31.0155) + (-20.6620) = -103.3520$$

Por lo tanto, la ubicación sugerida por el modelo es aproximadamente **20.6747, -103.3520**.