

Actividad Evaluable: Mapas de calor y boxplots

Valeria Aispuro Gomez A01642923

- 1. Carga los datos usando tu lector de csv o con pandas. Es recomendable hacerlo con pandas.**
- 2. Realiza el análisis de las variables usando diagramas de cajas y bigotes, histogramas y mapas de calor.**

Para esta actividad use la base de datos que vamos a usar en nuestro proyecto final, los datos hablan de los sectores de economía en México. El documento es este [d.csv](#).

Responde las siguientes preguntas:

-¿Hay alguna variable que no aporta información?

Usamos todas las variables proporcionadas ya que todas son importantes para poder hacer el análisis completo. Para asegurarnos de eso sacamos la varianza de todas las variables, si alguna salía muy baja significa que no era importante tomarla en cuenta porque no cambiaba mucho a través de los años, sin embargo todas tienen una varianza alta.

-Si tuvieras que eliminar variables, ¿cuáles quitarías y por qué?

Excluimos las variables relacionadas con Comercio, Restaurantes, Servicios de Alojamiento, Transportes, Comunicaciones, Correo, Almacenamiento, Servicios Profesionales, Financieros, Corporativos, Servicios Sociales, Servicios Diversos, Gobierno y Organismos Internacionales, ya que estos sectores no son las principales fuentes de contribución a la economía de México. En su lugar, nos enfocamos en variables asociadas a Agricultura, Ganadería, Silvicultura, Caza, Pesca, Industria Extractiva, Electricidad, Industria Manufacturera y Construcción, que son los sectores fundamentales de la economía mexicana.

-¿Existen variables que tengan datos extraños?

En el segundo trimestre de 2020, todas las variables tenían valores de NA (No Aplicable). Esto podría deberse a demoras en la obtención, procesamiento o divulgación de datos en ese período. por eso decidimos también quitar los datos del segundo trimestre de este año.

-Si comparas las variables, ¿todas están en rangos similares? ¿Crees que esto afecte?

Las variables relacionadas con Agricultura, Ganadería, Silvicultura, Caza, Pesca, y las de Industria Extractiva y Electricidad tienen rangos de valores más limitados en comparación con las variables de Industria Manufacturera y Construcción. Diferentes rangos de valores dificultan la comparación directa entre las variables, ya que una variable con valores más grandes podría tener un impacto desproporcionado en el análisis.

-¿Puedes encontrar grupos que se parezcan? ¿Qué grupos son estos?

Se puede utilizar el algoritmo K-Means para identificar grupos de variables similares en el conjunto de datos. K-Means es una herramienta efectiva para encontrar conjuntos de datos con características parecidas. Los grupos resultantes pueden interpretarse como categorías o segmentos de datos que comparten características similares.