

Mapas de calor y boxplots

Jorge Iván Sánchez González A01761414

Ana Camila Jiménez Mendoza A01174422

Gustavo José Ortiz Zepeda A01637220

**Carga los datos usando tu lector de csv o con pandas. Es recomendable hacerlo con pandas.*

**Realiza el análisis de las variables usando diagramas de cajas y bigotes, histogramas y mapas de calor.*

Responde las siguientes preguntas:

¿Hay alguna variable que no aporte información?

Sí, hay algunas variables que no aportan información tan relevante o que afecten directamente a otras variables, además muchas de ellas presentan ausencia de información:

```
[ ] #Cantidad de informacion ausente  
print(df.isna().sum()/len(df)*100)
```

Estación SIMAJ	0.000000
Fecha	0.000000
Hora	0.000000
O3	36.438356
NO	63.287671
NO2	63.287671
NOX	63.287671
SO2	74.520548
CO	50.958904
PM10	76.164384
PM2.5	93.595890
TMPI	11.027397
TMP	15.342466
RH	15.068493
WS	15.068493
WD	15.342466
PP	9.863014
RS	100.000000
PBA	100.000000
UV	100.000000
UVI	100.000000

dtype: float64

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

Las variables que quitaremos son todas a excepción de:

- Hora
- O3
- CO
- TMP

Esto debido a que son de las variables que menos ausencia de información presentan, además de que se relacionan de buena manera en los mapas de calor exceptuando el CO.

¿Existen variables que tengan datos extraños?

No hay datos extraños, pero las variables de estación y fecha son datos de tipo string que realmente son irrelevantes.

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

Si comparamos variables similares, y por similares me refiero a que se relacionan o son del mismo “tipo”, podemos observar que la mayoría se encuentra en rangos similares, claro que esto afecta porque nos indica cierta correlación que va a afectar unas variables con otras arrojando diferentes resultados al ser modificadas. Cabe aclarar que no todas las variables están en rangos similares porque miden diferentes cosas y unas no se correlacionan tanto o nada con otras.

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

Sí, dos grupos que se parecen son el ozono y el dióxido de nitrógeno, que son partículas esencialmente en el aire por lo que son similares como anteriormente mencionamos. Sin embargo podemos clasificar la información en varios grupos como lo son:

- Contaminantes atmosféricos: Este grupo incluiría variables como O3 (ozono), CO (monóxido de carbono), NO2 (dióxido de nitrógeno) y otros contaminantes atmosféricos. Estas variables están relacionadas con la calidad del aire y su impacto en la salud y el medio ambiente.
- Variables climáticas: Este grupo podría incluir TMP (temperatura), HR (humedad relativa), WS (velocidad del viento) y otras variables relacionadas con las condiciones meteorológicas. Estas variables son importantes para comprender el clima y su influencia en la calidad del aire.
- Variables de tiempo: Este grupo comprendería las variables de fecha y hora, que son importantes para el análisis temporal y la identificación de patrones estacionales en los datos.
- Variables irrelevantes: Este grupo incluiría las variables de estación y cualquier otra que se considere no relevante para el análisis de calidad del aire o clima.