



Tecnológico de Monterrey

Actividad Evaluable 3 - Actividad evaluable: Mapas de calor y boxplots

Gael González Arbesú - A01611800

Brandon Kevin Saavedra Cortes- A01748300

En esta actividad trabajarás con el conjunto de datos asignado para el reto. Tienes que replicar los pasos vistos durante las horas de salón en un diferente conjunto de datos.

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

```

Edit Selection View Go Run Terminal Help
CargaDatosPython.ipynb kmeans.ipynb Entregable_Patrones con K-means.ipynb calorBoxplot.ipynb analisisavocado.ipynb
C:\Users\pinkl\Desktop\Programas\Laboratorio\Laboratorio_A01611800\Actividad 04\calorBoxplot.ipynb > M*Actividad evaluable: Mapas de calor y boxplots
+ Code + Markdown ▶ Run All ↺ Restart ⚙ Clear All Outputs 📄 Variables 📖 Outline ...

^ Actividad evaluable: Mapas de calor y boxplots

En esta actividad trabajarás con el conjunto de datos asignado para el reto. Tienes que replicar los pasos vistos durante las horas de salón en un diferente conjunto de datos.

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

▶ ~
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb

datos = pd.read_csv("covid19_tweets.csv")
[27] ✓ 0.5s
```

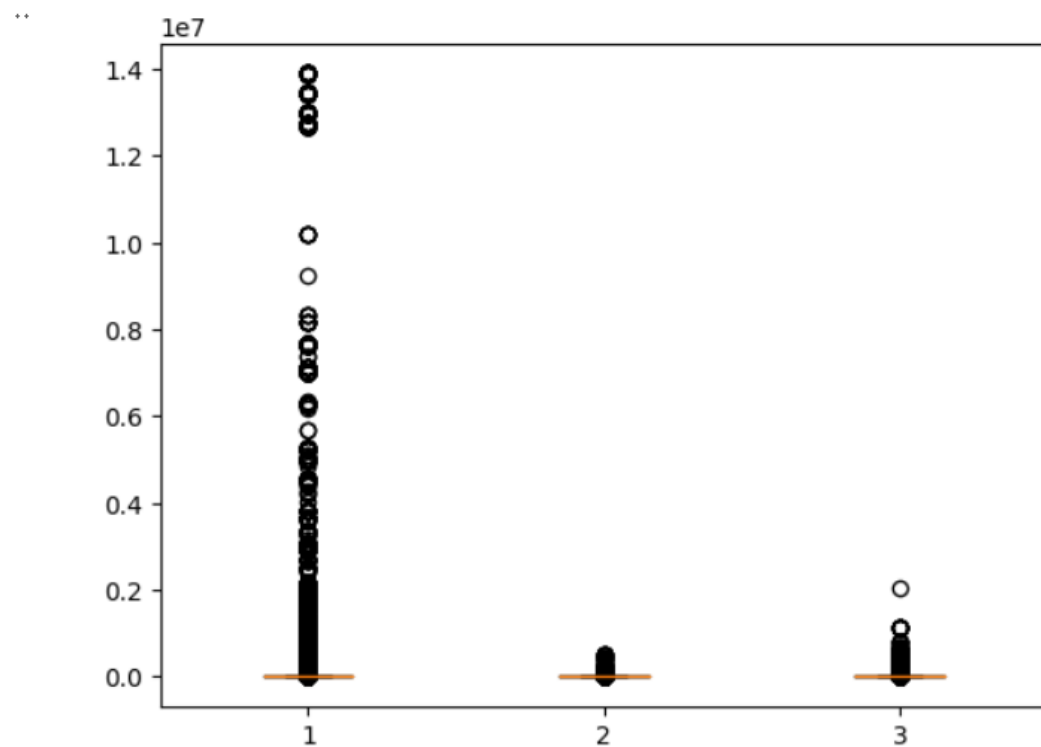
Realiza el análisis de las variables usando diagramas de cajas y bigotes, histogramas y mapas de calor.



Tecnológico de Monterrey

Diagrama de cajas y bigotes:

```
> ~  
plt.boxplot(datos[["user_followers","user_friends","user_favourites"]])  
  
plt.show()  
38] ✓ 0.1s
```





Tecnológico de Monterrey





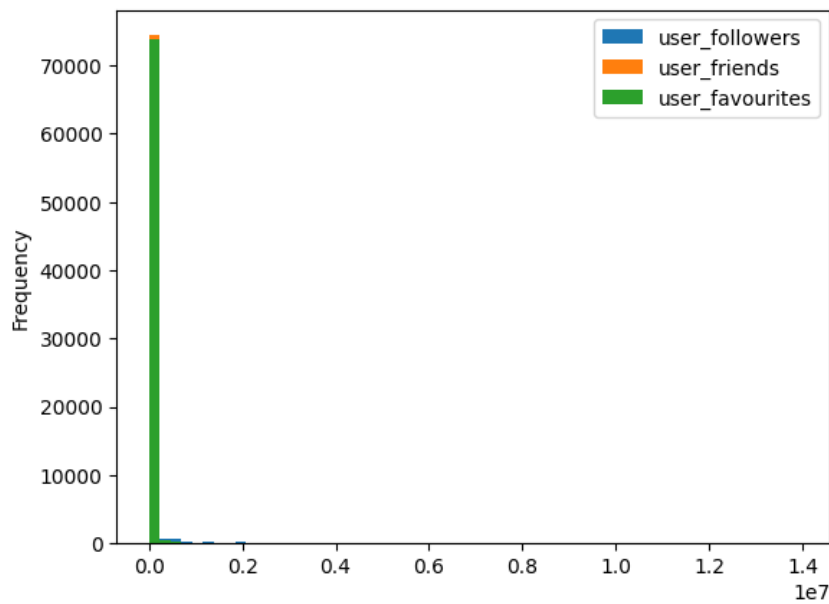
Tecnológico de Monterrey

Histograma:

```
ax = datos[["user_followers", "user_friends", "user_favourites"]].plot.hist(bins=60, alpha=1)
```

[40] ✓ 0.3s

...

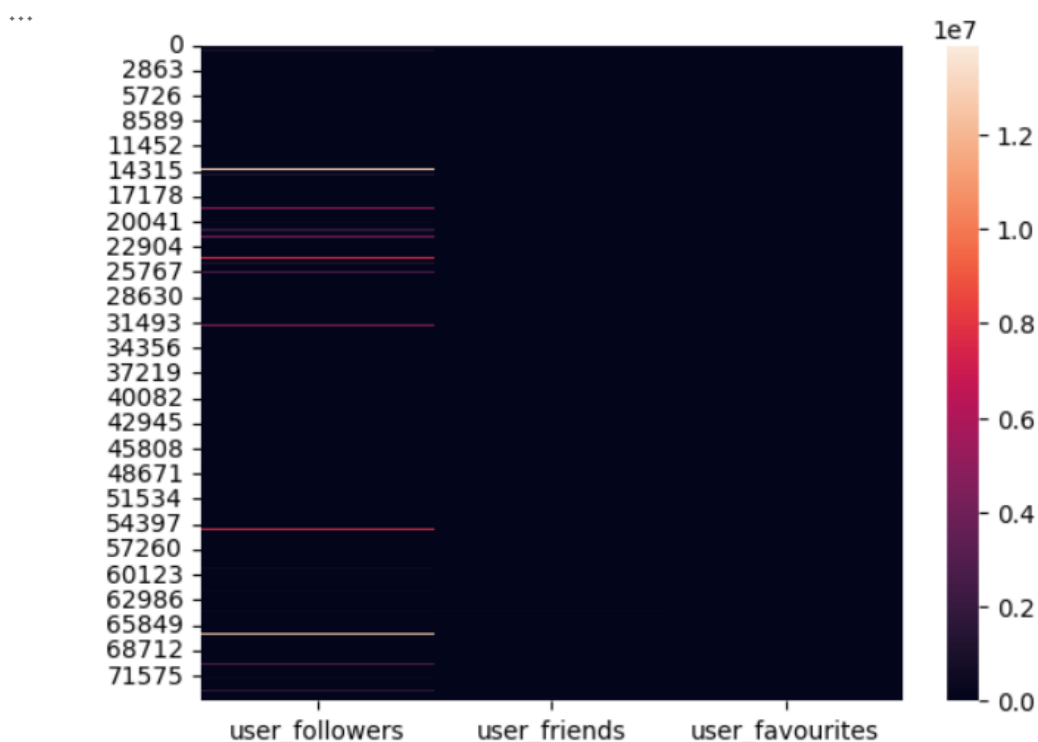




Tecnológico de Monterrey

Mapa de calor:

```
ax = sb.heatmap(datos[["user_followers","user_friends","user_favourites"]])  
[41] ✓ 0.5s
```



Responde las siguientes preguntas:

¿Hay alguna variable que no aporte información?

- Dentro de las utilizadas para el análisis todas funcionan y por mismas razones fueron seleccionadas y esto lo podemos ver al graficar pues todas brindan información relevante

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

- Si se refiere a los datos en general quitamos source y user_description, esto pues source no le hayamos utilidad y si bien user_description podría servir



Tecnológico de Monterrey

como referencia a ver el perfil de quienes publican, es mucha información que puede llegar a ser más innecesaria que necesaria.

¿Existen variables que tengan datos extraños?

- Si, algunas de las columnas como lo son user_description, text y hashtags presentan símbolos extraños que se podrían entender como el uso de emojis o diferentes lenguajes que no son reconocidos y son mostrados como datos extraños.

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

- No realmente, si bien hay datos que se encuentran en rangos similares hay muchos más que se encuentran muy dispersos lo cual hace que en los análisis se vean datos que no cuadren como el promedio, lo cual hace que se pierda efectividad y precisión al hacer análisis.

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

- Si, pues como se ha mencionado anteriormente hay grupos que comparten rasgos similares lo cual hace que se parezcan, estos pueden ser seguidores, retweets, hashtags y datos como favoritos o amigos.