

Obtención de estadísticas descriptivas

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

1. Carga los datos usando tu lector de csv o con pandas. Es recomendable hacerlo con pandas.
2. Verifica la cantidad de datos que tienes, las variables que contiene cada vector de datos e identifica el tipo de variables.
3. Analiza las variables para saber qué representa cada una y en qué rangos se encuentran. Si la descripción del problema no te lo indica, utiliza el máximo y el mínimo para encontrarlo.

Link

del

Github:

<https://github.com/Aram32mm/A01657142/tree/main/D%C3%ADA%202>

4. Basándose en la media, mediana y desviación estándar de cada variable, ¿Qué conclusiones puedes entregar de los datos?

Basándose puramente en esas 3 herramientas estadísticas. Primero, tomando únicamente en cuenta la media, podemos decir que los usuarios en promedio tienen más seguidores que amigos, es decir, siguen a más personas que las que los siguen y tienen en promedio una mayor cantidad de usuarios marcados como “favoritos.” Por otro lado, la mediana aunque pudiera resultar interesante, no aporta ningún dato de valor en la analítica de los datos, por lo que es mejor analizar la desviación estándar. Dicha desviación nos muestra la dispersión del conjunto de datos numéricos y podemos notar claramente que en donde más dispersión existe es en los seguidores, mientras que le siguen los favoritos y por último, los amigos. Tomando todos estos datos y analizando todo lo obtenido en el código, como diagramas de cajas y bigotes, histograma, correlación, mapa de calor, etc. Podemos concluir que la mayoría de los tweets publicados fueron por las mismas personas, e incluso podríamos concluir que fue gracias a bots(máquinas), esto por la gran discordancia entre amigos y seguidores, tiempo de publicación, localización, etc.