

Continúa aprendiendo sobre EDA

Este es el fin de curso de Exploración de Datos, antes de hablar sobre como continuar aprendiendo, hablaremos sobre las conclusiones:

Conclusiones

- **Las preguntas son la fuente de toda exploración.** Asegúrate de definir qué quieres encontrar y quién necesita consultar los resultados desde un comienzo.
 - **Es fundamental identificar el tipo de análisis de datos y variables que se requieren.** Explora las dimensiones de tu conjunto de datos y qué tipos de variables contiene.
-
- Las preguntas son la fuente, así todo tendrá un inicio y un fin, así nosotros podremos saber que explicar y que resultados nos pueden arrojar esas preguntas. Es importante para tener acotado la obtención de la información.
 - Hay que identificar que tipo de variables tenemos, si tenemos valores nulos y como se comportan las variables.

Conclusiones

- **Siempre visualiza los estadísticos.** Todos los conjuntos de datos son diferentes, conócelos más allá de sus números de resumen.
 - **Visualiza una o varias variables de distintas maneras.** La diversidad de gráficas te permitirá conocer a detalles los matices de los datos.
-
- Siempre visualiza los estadísticos, todos los conjuntos de datos son diferentes así que lo principal es conocerlos e ir más allá de un número o resumen que me pueda entregar un parámetro estadístico. Recuerda el caso de la **paradoja de Simpson**, los números de manera aislada no tienen un significado y hasta nos podrían estar ocultando cierta información, sin embargo con el apoyo visual los números empiezan a tener sentido. Por eso siempre visualiza los estadísticos.
 - Visualiza de distintas formas las variables: La diversidad de gráficas nos va a permitir entender y comprender los matices que existen en los datos. Darnos cuenta que si una relación parecería negativa, al incluir una nueva variable cambia completamente su comportamiento de la relación y cobra sentido, este tipo de situaciones suelen pasar a menudo, simplemente por no explorar el conjunto de datos de la manera correcta. Juega con las gráficas y preséntalas de diferentes formas hasta que sea clara la información que transmiten y que estamos encontrando, para asegurarnos que la exploración está avanzando

Continúa explorando:

- **Nuevos datos.**
- **Estadísticos.**
- **Gráficas.**
- **Comunicación de resultados.**

- Busca nuevos conjuntos de datos. Y replica el análisis o hazlo a tu propio modo que entiendas.
- Aprende nuevos temas estadísticos, entiende sus casos de uso para saber cuando aplicarlos en los datos y cuando no.
- Aprende sobre nuevas gráficas, nuevas formas de representar los resultados. Esto nos permitirá expandir ideas y no quedarnos cortos, además de ver los matices que pueden tener los conjuntos de datos.
- Continúa aprendiendo como comunicar los resultados, ya sea a través de un **notebook, dashboard**, como escribir en prosa lo que significa cada resultado hallado en los estadísticos. ¿Cómo podrías presentarlo a una persona que nunca ha visto tus datos?

Extra:

Aporte Carlos Mazzaroli

Bueno gente, les tiro un trucazo, en un código corran este código

```
import palmerpenguins
import pandas as pd
import pandas_profiling as pp
```

```
df = palmerpenguins.load_penguins()
```

```
pp.ProfileReport(df)
```

Tira un reporte general de los datos bastante bueno para el principio del EDA

- [Librerías de Python para EDA](#)