



Inferencia Estadística

Ejercicio: EDA

DSLlab

noviembre, 2024

Utilizar el dataframe denominado *bank* con el que hemos trabajado en el tema 2 de la asignatura para responder a las siguientes cuestiones:

1. Estructura del DataFrame:

- ¿Cuántos campos y observaciones tiene el dataframe *bank*? Utiliza las funciones *head* y *dim*.

2. Resumen del DataFrame:

- Evalúa el dataframe con la función *summary*.
- ¿Tiene observaciones con elementos faltantes (NA)?
- ¿A qué categorías corresponden las observaciones en la variable *job*?

3. Distribución de la Edad:

- ¿Cuál es la edad máxima y mínima de los clientes en el dataframe?
- ¿Cuál es la media y la mediana de la edad de los clientes?

4. Balance Promedio:

- ¿Cuál es el balance promedio anual (*balance*) de los clientes?
- ¿Cuál es el balance promedio anual de los clientes que han suscrito un depósito a plazo fijo (*y* = "yes")?

5. Frecuencia de Contacto:

- ¿Cuál es el número máximo y mínimo de contactos realizados durante esta campaña (*campaign*)?

6. Análisis de Duración:

- ¿Cuál es la duración media y mediana del último contacto en segundos (*duration*)?
- ¿Cuál es la duración media del último contacto en segundos para los clientes

que suscribieron un depósito a plazo fijo ($y = \text{"yes"}$)?

7. Relación entre Variables:

- ¿Existe alguna relación entre el balance promedio anual y la duración del último contacto? Utiliza una visualización adecuada para responder a esta pregunta.

8. Segmentación por Trabajo:

- ¿Cuál es la media y mediana del balance anual de los clientes agrupados por tipo de trabajo (job)?

9. Análisis de Contactos Anteriores:

- ¿Cuál es el número máximo y mínimo de días que pasaron desde el último contacto de una campaña anterior ($pdays$)?
- ¿Cuál es la media de $pdays$ para los clientes que suscribieron un depósito a plazo fijo ($y = \text{"yes"}$)?

10. Estudio General:

- Haciendo un estudio general de los datos, ¿qué puedes concluir? ¿Existe alguna relación significativa entre las variables `balance`, `duration`, y `campaign`? Se recomienda hacer un análisis visual y estadístico de estas variables.