

DOCUMENTACION DEL PROCESO / PROCESS DOCUMENTATION

LIMPIEZA Y PROCESAMIENTO DE DATOS / DATA CLEANING AND PROCESSING

- 1) Revisamos si existían valores duplicados y procedimos a eliminar estos (En total existían 26 valores duplicados los cuales fueron eliminados).
We checked for duplicate values and proceeded to remove them. In total, there were 26 duplicate values, which were deleted.
- 2) Revisamos cada una de las columnas y verificamos si existían campos blancos, en este caso todos los campos estaban correctamente diligenciados.
We reviewed each of the columns and checked for any blank fields; in this case, all fields were properly filled out.
- 3) La columna “Marital Status” y “Gender” tenían la “M” como valor en común por lo que organizamos estos valores y reemplazamos los valores “M” por “MARRIED” y “S” por “SINGLE” en la columna “Marital Status” y en la columna “Gender” reemplazamos “M” por “MALE” y “F” por “FEMALE”
"The 'Marital Status' and 'Gender' columns had 'M' as a common value, so we organized these values and replaced 'M' with 'MARRIED' and 'S' with 'SINGLE' in the 'Marital Status' column, and in the 'Gender' column, we replaced 'M' with 'MALE' and 'F' with 'FEMALE'."
- 4) Cambiamos el formato de la columna “income” para que tuviera formato de moneda y quitamos 2 decimales para que fuera más agradable a la vista.
We changed the format of the 'income' column to currency and removed two decimal places to make it more visually appealing.
- 5) Luego utilizando la función =SI() creamos una nueva columna “group_income” para agrupar los datos de esta columna en 3 (\$10.000-\$60.000) (\$60.000- \$120.000) (More than 120.000)
Then, using the IF() function, we created a new column called 'group_income' to group the data from this column into three categories: (\$10,000-\$60,000), (\$60,000-\$120,000), and (More than \$120,000).

- 6) Al igual que el paso anterior agrupamos el grupo de edades de la siguiente manera (menores a 31 “adolescent” , entre 31 y 55 “adult”, mayores a 55 “old”

Similar to the previous step, we grouped the age ranges as follows: under 31 as 'adolescent', between 31 and 55 as 'adult', and over 55 as 'old'.

* Luego de implementar los pasos anteriores ya teniamos nuestro archivo organizado y listo para visualizacion en Power BI

After implementing the previous steps, we had our file organized and ready for visualization in Power BI.

- 7) Mientras estaba creando el dashboard con la informacion reelevante hice un calculo utilizando DAX para calcular el porcentaje de respuestas que “si” tenian bicicleta y el porcentaje que “no”

While creating the dashboard with the relevant information, I used a DAX calculation to determine the percentage of responses that 'yes' had a bicycle and the percentage that 'no' did not.

VISUALIZACION CON POWER BI

Luego continúe con la creación del dashboard interactivo, teniendo en cuenta que entre mas limpio y claro se vea el dashboard con la informacion mas importante en el inicio se creo el titulo, se incluyeron 2 filtros interactivos de (“Marital Status” y “Purchased Bike”) luego se pusieron el porcentaje de respuestas positivas y negativas que calculamos con codigo DAX en Power Bi

Grafico 1: Muestra el numero de respuestas positivas y negativas por grupo de personas

Grafico 2: Muestra el numero de bicicletas en base a la respuesta de distancia brindada por las personas que hicieron parte del dataset

Grafico 3: Muestra el porcentaje por genero sexual

Grafico 4: Muestra el comportamiento por region.

"Next, I continued with the creation of the interactive dashboard, keeping in mind that the cleaner and clearer the dashboard looks, with the most important information at the forefront, the better. I created the title and included two interactive filters ('Marital

Status' and 'Purchased Bike'). Then, I displayed the percentages of positive and negative responses that we calculated using DAX in Power BI.

Chart 1: Shows the number of positive and negative responses by group of people.

Chart 2: Shows the number of bicycles based on the distance responses provided by the people in the dataset.

Chart 3: Shows the percentage by gender.

Chart 4: Shows the behavior by region."

