16 de octubre del 2023

Documentación de proyecto segmentación

Exploración de datos y entendimiento

**Procesado y preparar base de datos**

Meta: Conectar/importar datos a herramientas

**Objetivo 1: Utilizar IMPORTRANGE para importar datos**

1. NOTA: Decidí usar la fórmula de importrange porque fue más fácil y con la técnica e importación no detectaba los archivos que había creado una copia de estos. por lo que para este proyecto usaré un importrange.
   1. #=IMPORTRANGE("https://docs.google.com/spreadsheets/d/1eJxIxHTx9PzQO8f05ZcYhMkDB6yyJhtja8pktGF-Gzg/edit#gid=0";"clientes!A:I")

**Objetivo 2: Identificar y manejar valores nulos**

Identifiqué los siguientes variables donde hay valores nulos o fuera de alcance:

* Use la formula de #countblank para saber en qué celdas había valores nulos, esta es la primera parte de la limpieza
* #=CONTAR.BLANCO(H2:H2241)
* esta fórmula la apliqué para cada columna y saber en donde es que hay datos nulos para que al pasarlos a una pestaña limpia, ya sepa que filtrar con query
  1. CLIENTES LIMPIEZA
     1. Ingreso dólar anual=24
     2. año de nacimiento =3

1. TRANSACCIONES LIMPIEZA
   1. id\_cliente=7

Al final decidí usar la fórmula de query para limpiar mi base de clientes de personas que no tienen registro anual y de personas que no tienen año de nacimiento bien.

Los mandé a llamar de mis datos origen completamente limpios de datos nulos o inconscientes.

#=query(clientes!A1:I2241;"select\* WHERE E>1 and B>1910";1)

**Objetivo 3: Identificar duplicados a través de las herramientas de *Google Sheets***

Revisaré por cada pestaña de datos ya filtrados y limpios si hay datos duplicados y los pasaré en la misma pestaña pero a lado para obtenerlos con datos sin duplicación, al final estos serán los que mandaré a llamar.

Lo primero que haré es identificar dónde hay datos duplicados, guiándome solo de el id del cliente o lista determinada. lo que me interesa es saber que no hay datos duplicados de los clientes.

Encontré los datos duplicados solo en la pestaña de resumen= 9 id repetidos con sus datos iguales.

Con la fórmula #count.si agregué una columna de los datos que están repetidos en la pestaña de resumen\_compras, cuando el dato no se repite se coloca un 1, cuando el dato se repite se pone un 2.

Como yo identifiqué mis datos nulos y duplicados antes de pasarlos a una de limpieza, con la fórmula de query filtre estas dos condicionantes, entonces con una sola fórmula obtuve mis datos limpios sin necesidad de otro procesamiento.

**Objetivo 4: Identificar datos fuera de alcance**

Los datos de transacciones los resumí con una tabla dinámica y así logré obtener el numero total de transacciones por cliente.

Con el filtrado de mis datos a la tabla\_resumen donde se unen las tres pestañas y todos los datos, tomé como datos principal el de \_clean clientes con 2213 datos.

al momento de matchear mis datos de transacciones. Quiere decir que tengo el registro de 2213 clientes con sus datos personales.

**datos clean\_clientes=2213**

**datos clean\_transacciones= 2230**

**datos resumen\_compras=2240**

Al hacer mi tabla resumen tome como principal la de clientes CON 2213 datos unicos, al moment de matchearlos con los datos de transacciones sobraron **17 datos** que no mandé a llamar porque no aparecieron en mi registro de clientes.

Al llamar los datos limpios de la pestaña de resumen, había 2240 datos de clientes con sus respectivas transacciones, sin embargo como no hay datos de todos los clientes al menos sus características ni fecha de registro también se eliminaron. en total solo se lograron matchear/coincidor 2213 datos de id\_clientes y sus datos completos por lo que se eliminaron 27 datos.

filtrado de los 10 datos que no tienen transacciones y que los borre con filtrado y la condicionante de número columna mayor o igual a 1 para que solo me llevara al resumen los que sí tienen transacciones.

Revisar todos aquellos de clientes que se quedaron actuales de 2203 y que coinciden con que si haya en resumen compras y no me de un error.

Borrar inconsistencias de clientes sin registros