Ficha Técnica: Proyecto de Análisis de Datos

Título del proyecto: Segmentación  
Objetivo: Realizar segmentación de clientes bajo la técnica RfM, preparando la bases de datos, interpretando resultados y extrayendo conclusiones.

Equipo:trabajo individual  
Herramientas y tecnologías: Lista de herramientas utilizadas (por ejemplo, *Google Sheets*, *Looker Studio*)  
Procesamiento y análisis: Breve descripción del preprocesamiento, técnicas aplicadas, códigos utilizados, etc. Resultados y conclusiones: Principales hallazgos, conocimientos extraídos y recomendaciones  
Limitaciones / Próximos pasos: Dificultades encontradas y sugerencias para seguir desarrollando el proyecto  
Enlaces de interés: Link a tu hoja de cálculo, dashboard y otros recursos relevantes

| **Objetivo** | **Punto control** | **Bitácora** |
| --- | --- | --- |
| Importar datos a Google sheets | Importaste los datos automáticamente usando fórmulas o la opción *Archivo > Importar* de Google Sheets? | Inicialmente al acceder a las bases de datos se descargaron los 3 archivos en xls y estos a su vez se guardaron en una carpeta de trabajo del drive y al importarlas a un nuevo sheets no se encontraron disponibles, por lo que tocó abrir un nuevo documento de drive para el proyecto 1, importar los archivos, subiendolos desde la ubicación “descargas” del computador donde estaban guardados, para así poder importarlos al documento PROYECTO1  Luego se hizo el ejercicio de descargar los archivos en extensión .csv en el PC, se guardaron en la carpeta drive y desde ahí se importaron para el sheets PROYECTO1  los libros de google sheets, se unieron en un mismo libro “Proyecto 1” para facilidad en el manejo de los mismos |
| Identificar valores nulos en Google sheets usando fórmulas como countblank o mediante tablas dinámicas | * ¿Identificaste qué variables tienen valores nulos? * ¿Tomaste una decisión sobre cómo tratarlos? | identificando nulos y vacíos con filtros en la variable clave Id\_cliente y en cada una de las otras columnas para determinar si es necesario hacer tratamiento de los mismos  tabla clientes    con formula =FILTER(  FLATTEN(IF(A2:I2241 = "", ADDRESS(ROW(A2:I2241), COLUMN(A2:I2241)), "")),  FLATTEN(ISBLANK(A2:I2241)))  me indica los valores en blanco de toda la tabla donde indican los valores de la columna E perteneciente al ingreso anual dolar , estos valores se proceden a completar con el valor de la mediana de esa misma columna que es sacado mediante tablas dinámicas    con tablas dinamicas puedo ver los blancos y ceros en tabla CLIENTES de las columnas id\_cliente e ingreso anual dolar    tabla resume compra    tabla transacciones  .  tabla dinámica para blancos y ceros de tabla TRANSACCIONES en id\_cliente    hay 7 blancos pero muestra 0, al abrir el reporte de datos se ven los 7 que estan vacios   * TABLA DINÁMICA para blancos y ceros de la tabla RESUMEN\_COMPRAS     las tablas dinámicas me indican que el id\_cliente cero “0” es un cliente que hace transacciones y tiene montos por lo tanto no se borra. los únicos que se proceden a borrar son los 7 blancos que aparecen en la columna ID\_CLIENTES de la tabla transacciones pues no se sabe que cliente es  los espacios blancos de la tabla clientes en columna ingreso\_dolar se pone la mediana de esa misma columna para completar la informacion y que no haga cambios en la distribucuion de la variable , como si los puede generar el promedio  con el uso de countblank se puede saber la cantidad de nulos en una BD sin embargo no dice que celda es la que tiene el nulo por lo que recurrí a chat gpt y me ayudó mediante fórmulas a nombrar las celdas que contienen valores nulos para así identificar la posición del valor nulo e intervenir según la necesidad que se quiera cubrir  =FILTER(  FLATTEN(IF(A2:D22128 = "", ADDRESS(ROW(A2:D22128), COLUMN(A2:D22128)), "")),  FLATTEN(ISBLANK(A2:D22128))  ) |
| identificar y gestionar valores duplicados | * ¿Identificaste los valores duplicados? * ¿Los eliminaste correctamente? | para los valores duplicados en la tabla CLIENTES de la columna id\_clientes use dos tipos de fórmula que me recomedó gpt. la primera me dice que valores son duplicados:  =FILTER(A2:A2250, COUNTIF(A2:A2250, A2:A2250) > 1)  COUNTIF(A2:A100, A2:A100) devuelve un array con el número de veces que aparece cada valor de la columna A.  FILTER(A2:D100, ...) selecciona todas las filas donde la cuenta es mayor que 1 (es decir, hay duplicados).  la segunda manera de saber que valores duplicados hay en id de clientes y resumen compra, es generar una columna nueva que me muestre por fila y las etiqueto en : duplicado o único según la columna que es la llave clave, en este caso la columna A  =IF(COUNTIF($A$2:$A$2250, A2) > 1, "Duplicado", "Único"  de esta manera procedo a eliminar los duplicados en resumen\_compra pues corresponden a registros de un mismo cliente |
| Unir tablas en *Google Sheets* usando fórmulas como BUSCARV, ÍNDICE + COINCIDIR. | * ¿Identificaste la clave de unión entre las tablas? * ¿Generaste una única tabla consolidada usando fórmulas (no con copiar/pegar)? | determine que la clave de unión para las tablas es la columna Id\_clientes  En una hoja nueva llamada CONSOLIDADO, Inicie la unión de la tabla de clientes y transacciones pues esta última me parecía más compleja de relacionar y unir por el tema de repeticiones en el id del cliente que indican la cantidad de transacciones hechas por ese cliente. entonces para ello hice conteo del número de transacciones por cliente con la fórmula  =IFERROR(COUNTIF(transaccionesSINvaciosNIceros!B2:B22109, A2), 0)  Seguidamente para la unión de la columna fecha de transacción de la tabla transacciones a la hoja CONSOLIDADO, genere una nueva columna de nombre **recurrencia\_compra** donde saco el **promedio de días** entre cada compra de un mismo cliente para relacionar numéricamente la columna id\_clientes (con valores repetidos) de la tabla transacciones a la tabla CONSOLIDADO la fórmula  =AVERAGE(ARRAYFORMULA(  LAMBDA(i,CHOOSEROWS(SORT(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2)),i))(SEQUENCE(COUNTA(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2))-1,1,2))-  LAMBDA(i,CHOOSEROWS(SORT(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2)),i))(SEQUENCE(COUNTA(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2))-1))))  para unir la columna lugar\_de\_compra a CONSOLIDADO, hice dos nuevas columnas donde se muestran los valores del conteo de las compras si es en línea o en tienda  =COUNTIFS(  transaccionesSINvaciosNIceros!B:B, A2,  transaccionesSINvaciosNIceros!D:D, "en línea"  )  =COUNTIFS(  transaccionesSINvaciosNIceros!B:B, A2,  transaccionesSINvaciosNIceros!D:D, "tienda"  )  para la unión de las demás columnas donde solo se relaciona un solo id\_cliente, use Vlook up para unirlas todas a la hoja CONSOLIDADO  una vez terminada la unión de tablas, hago ajustes en la columna recurrencia\_compra de la hoja CONSOLIDADO, pues aparece #VALUE!:  la unión de tablas en la nueva hoja CONSOLIDADO genera valores #VALUE! en la columna recurrencia\_de\_compra cuando la columna total\_transa\_por\_cliente es no hay valor o es 1 , por lo que les coloco un 0 en recurrencia, pues ese cliente no tiene recurrencia en compra      para el calculo dela nueva columna ultima fecha de compra, a los id caliente ocn transacción cero, se le puso total transaccion cliente en 1 y lcon la fecha de entrada se lleno en ultima fecha compra |
| análisis exploratorio **Agrupación de datos según variables categóricas** Utilizar tablas dinámicas para resumir información de variables categóricas. | * ¿Cuál es el nivel educativo más común? * ¿Cuál es el rango de edad? * ¿Cuántos clientes están casados? * ¿Cuál es el ingreso promedio?   👍 ¿Has creado tablas dinámicas para resumir educación, estado civil, ingresos, hijos, respuesta a campañas, frecuencia y gasto promedio? | Para la generación de tablas dinámicas agregamos nuevas columnas a la base de datos CONSOLIDADO. entre estas estan: recurrencia compra, edad, rango de edad, ingresos anuales por rangos y antiguedad cliente en meses  grado superior con 50.3%  el rango de edad más relevante so los mayores de 50, seguido de los de entre 40-50 y los de 30-40. la menor poblacion son los menores a 30  según los datos, 864 clientes están casados. de estos, 394 tienen niños hasta 10 años y 442 tienen niños con mas de 10 años-  52.237.98  si |
|  |  |  |
| **5.2.2 🟣 Visualizar variables categóricas** Objetivo: Crear gráficos de barras para representar variables categóricas. | 👍 ¿Creaste gráficos de columnas para representar grupos por variable categórica? | he creado graficos de columnas y circulares para mirar proporciones en las diferentes variables categóricas |
| **5.2.3 🟣 Calcular cuartil, decil o percentil** Objetivo: Usar fórmulas para dividir clientes en grupos según métricas RFM. | 👍 ¿Calculaste cuartiles o percentiles para cada atributo RFM? | he aplicado cuartiles y percentiles a variables numéricas como: total\_transa\_cliente, ticket\_promedio, ingreso\_anual, montos\_categoria, recurrencia, antigüedad (calculada con la columna fecha ingreso)  y también para los atributos RFM;  par el calculo de la recencia se crean dos columnas, una para encontrar la última fecha de compra de la tabla transacciones con la fórmula  =MAX(FILTER(transacciones!C:C, transacciones!B:B = A2)) y la columna de recencia se calcula a fecha 31 de diciembre de 2022 con la siguente formula =IFERROR(vDATE(2022,12,31) - MAX(FILTER(transacciones!C:C, transacciones!B:B = A2)),  " " )  de esta manera, aparecen clientes con recencia en blanco, po rlo que procedo a ponerles cero pues son clientes que solo compraron una vez |
| **🟥 5.3 Aplicar la técnica de análisis** En este paso aplicarás la técnica de segmentación de clientes mediante RFM (*Recencia, Frecuencia, Monetario*), muy usada en marketing para clasificar perfiles de clientes y mejorar estrategias comerciales. **5.3.1 🔴 Aplicar segmentación** Objetivo: Crear categorías basadas en cuartiles RFM usando SI o IFS. | 👍 ¿Creaste fórmulas automatizadas para clasificar clientes?  ¿Tus categorías están balanceadas y la mayoría de los clientes fueron clasificados? | Cálculo de recency  =IFERROR(  DATE(2022,12,31) - MAX(FILTER(transacciones!C:C, transacciones!B:B = A2)),  "")  Calculo recurrencia, promedio entre días de compra  =AVERAGE(ARRAYFORMULA(  LAMBDA(i,CHOOSEROWS(SORT(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2)),i))(SEQUENCE(COUNTA(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2))-1,1,2))-  LAMBDA(i,CHOOSEROWS(SORT(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2)),i))(SEQUENCE(COUNTA(FILTER(transacciones!$C$2:$C, transacciones!$B$2:$B = A2))-1))))  Monetary  Puntaje Quartil Recency  =IF(M2<=QUARTILE($M$2:$M$2241,1), 4,  IF(M2<=QUARTILE($M$2:$M$2241,2), 3,  IF(M2<=QUARTILE($M$2:$M$2241,3),2,1)))  Puntaje Quartil Frequency  =IF(O2<=QUARTILE($O$2:$O$2241,1), 1,  IF(O2<=QUARTILE($O$2:$O$2241,2), 2,  IF(O2<=QUARTILE($O$2:$O$2241,3),3,4)))  Puntaje Quartil MOnetary  =IF(Q2<=QUARTILE($O$2:$O$2241,1), 1,  IF(Q2<=QUARTILE($O$2:$O$2241,2), 2,  IF(Q2<=QUARTILE($O$2:$O$2241,3),3,4)))  Bloque RFM  concateno los puntajes de los quartiles para formar bloques y luego estos bloques los etiqueto  =N2&P2&R2  en la nueva columna segmento clasifico los grupos  =IFS(  AG2="444","VIP",  LEFT(AG2,1)="4","Recientes",  MID(AG2,2,1)="4","Frecuentes",  RIGHT(AG2,1)="4","High Spenders",  AG2="111","Perdidos",  TRUE,"Otros"  ) |
| **5.4 Resumir información en un tablero o informe** Un tablero o *dashboard* es una herramienta visual que permite entender los datos rápidamente, mostrando métricas clave de forma resumida. Puedes construirlo en una nueva pestaña de tu *Google Sheets* utilizando tablas dinámicas, tarjetas de resultados y gráficos. **5.4.1 🟠 Crear tabla resumen o tarjeta de resultados** Objetivo: Usar tablas dinámicas o fórmulas para generar un resumen visual del negocio. | ¿Creaste tarjetas de resultados o tablas que resumen métricas clave del negocio? |  |
| **5.4.2 🟠 Incluir gráficos simples** Objetivo: Agregar gráficos circulares o de barras para representar los resultados. | ¿Has creado al menos un gráfico para mostrar resultados del análisis o la segmentación? |  |
|  |  |  |
|  |  |  |
|  |  |  |

5.1.1 Conectar/importar datos a herramientas

#### **1. Toma de decisiones en la etapa "Procesar y preparar la base de datos"**

* ¿Qué hacer si hay valores nulos?
* ¿Cómo manejar valores duplicados?
* ¿Cómo identificar valores fuera de rango?

#### **2. Segmentación en la etapa "Aplicar la técnica de análisis"**

* ¿Cuántos perfiles de cliente debo crear?
* ¿Cómo defino las reglas para segmentarlos?

#### **3. Presentación en la etapa "Presentar resultados"**

* ¿Qué información es verdaderamente relevante?
* ¿Cómo priorizo el tiempo limitado en mi presentación?