Ficha Técnica: Proyecto de Segmentación de Clientes

**1. Introducción**

Este proyecto tiene como objetivo analizar la base de clientes de la tienda de productos alimenticios importados EL MERCADO, utilizando técnicas de segmentación y análisis de patrones de compra. A través de este análisis, se busca comprender mejor a los clientes, identificar grupos prioritarios y desarrollar estrategias efectivas para la retención y el aumento de ingresos. Este trabajo fue desarrollado de manera individual, con una duración de 4 sprints.

## **2. Objetivos**

* Identificar características y comportamientos clave.
* Segmentar a los clientes para focalizar estrategias de marketing y retención.
* Desarrollar estrategias para incrementar las ventas y la lealtad del cliente.

## **3. Metodología**

### *Procesar y Preparar la Base de Datos:*

1. Conectar/Importar datos a herramientas
2. Identificar y manejar valores nulos
3. Identificar y manejar valores duplicados
4. Identificar y manejar datos fuera del alcance del análisis
5. Unir tablas para crear una base de datos consolidada que incluyera toda la información necesaria para el análisis.
6. Crear nuevas variables a partir de los datos existentes para enriquecer el análisis.
7. Construir tablas auxiliares para facilitar el análisis y la manipulación de datos.

### *Análisis Exploratorio:*

1. Agrupar datos según variables categóricas para identificar patrones y tendencias.
2. Visualizar las variables categóricas para identificar las más relevantes.
3. Aplicar medidas de tendencia central para comprender mejor la distribución de los datos.
4. Visualizar distribución.
5. Analizar el comportamiento de los datos a lo largo del tiempo.
6. Calcular cuartiles, deciles o percentiles para segmentar los datos de manera más precisa.

### *Aplicar Técnica de Análisis:*

1. Aplicar segmentación utilizando la técnica de RFM (Recencia, Frecuencia, Valor Monetario) para segmentar a los clientes.
2. Aplicar análisis por cohorte para evaluar la retención de clientes a lo largo del tiempo.

### *Resumir Información en un Dashboard o Reporte:*

1. Representar datos a través de tabla resumen o scorecards para ver los datos de manera clara y concisa.
2. Representar datos a través de gráficos simples: Se utilizaron gráficos simples para ilustrar los hallazgos clave.
3. Representar datos a través de gráficos o visuales avanzados: Se emplearon visualizaciones avanzadas para resaltar patrones y tendencias complejas.
4. Aplicar opciones de filtros para manejo e interacción: Se añadieron filtros interactivos para permitir una exploración más profunda de los datos.

### *Presentar Resultados:*

1. Seleccionar gráficos e información relevante para incluir en la presentación final.
2. Crear una presentación utilizando Google Slides para comunicar los resultados del análisis.
3. Presentar resultados con conclusiones y recomendaciones basadas en los hallazgos.

## **4. Procesamiento y Análisis**

### **Importación y Preparación de Datos:**

### Se importaron las tablas "clientes", "resumen\_compras" y "transacciones" utilizando la opción de importación de Google Sheets.

### Se aplicaron filtros en cada tabla para ordenar los "id\_clientes" de A-Z, identificando y eliminando duplicados y espacios en blanco. En "resumen\_compras", se encontraron y eliminaron 9 duplicados. En "transacciones", se identificaron 7 clientes sin ID, para los cuales se utilizó la función UNIQUE y FILTER para crear una nueva columna con IDs únicos, eliminando los valores nulos.

### **Creación de Variables:**

### En la tabla "clientes", se extrajeron variables como edad, rango de edad y total de hijos. Utilizando la función YEAR, se creó la columna "fecha de nacimiento", y con DATEIF se calculó la variable "edad".

### Para los rangos de edad, se definieron categorías consistentes y relevantes, divididas en intervalos de 10 años: 24-34, 35-44, 45-54, 55-64, 65-74, y 75-81. Estos se crearon usando la función IF.

### En la tabla "resumen\_compras", se calcularon las variables "monto\_total" y "monto\_promedio". En "transacciones", se extrajeron variables como "id\_clientes", "transacciones\_enlinea", "transacciones\_tienda", "transacciones\_2020 (2021, 2022)", "primera\_transacción", "última\_transacción", "total\_días" y "promedio\_días".

### **Unión de Tablas y Creación del Dataset:**

### Se utilizaron las funciones IMPORTRANGE y VLOOKUP para consolidar las tablas en una nueva pestaña llamada "dataset". Primero, se limpió la tabla "transacciones" creando una nueva tabla "transacciones\_limpia", y luego se unieron a la "dataset" con VLOOKUP, asegurando que cada celda corresponde a cada "id\_cliente".

### Se revisó y validó que la tabla estuviera ordenada, sin valores nulos, y que todos los datos fueran consistentes.

### **Análisis Exploratorio:**

### Se utilizó la tabla "dataset" para crear la tabla dinámica llamada "Pivot\_categorias", respondiendo preguntas clave sobre la base de clientes, como niveles de educación, rangos de edad, estado civil, ingreso anual promedio, número promedio de hijos, respuesta a campañas de marketing, frecuencia promedio de compra y valor monetario promedio por categoría de producto.

### Se utilizaron gráficos de barras para visualizar las variables categóricas más relevantes.

### **Segmentación RFM:**

### En la tabla "transacciones\_limpias", se crearon nuevas variables como "fecha\_max" (última compra), "total\_transacciones", "total\_montos" y "dias\_entre\_compras". Luego, se trasladaron estos datos a la "dataset" y se creó la pivot table "pivot\_cuartiles" para calcular cuartiles de recencia, frecuencia y valor monetario.

| **cuartiles** | **recencia** | **frecuencia** | **monto** |
| --- | --- | --- | --- |
| 0 | 0 | 1 | 8 |
| 1Q | 54 | 5 | 69 |
| 2Q | 133 | 10 | 396.5 |
| 3Q | 243 | 14 | 1044.5 |
| 4 | 879 | 27 | 2525 |

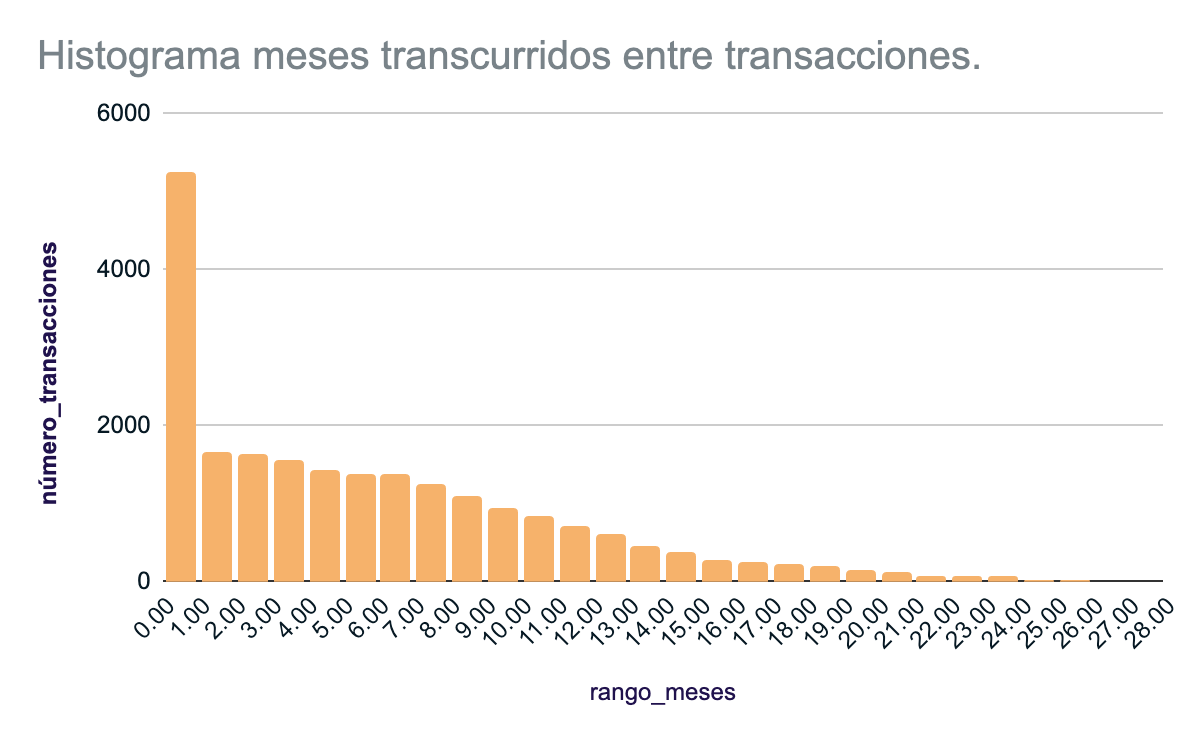
### Se utilizaron fórmulas IF para asignar los cuartiles a cada cliente en la pestaña de la tabla "RFM\_cuartiles":

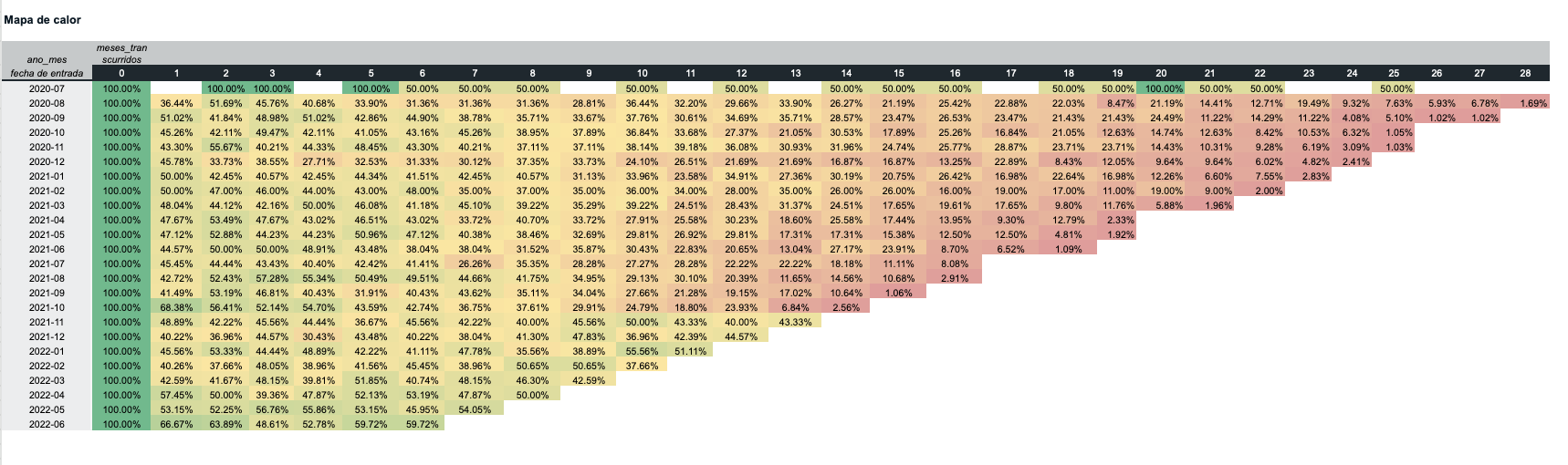
* + Fórmula para recencia: =IF(B2<=$F$10, "c 1",IF(B2<=$F$11, "c 2", IF(B2<=$F$12, "c 3", "c 4")))
  + Fórmula para frecuencia: =IF(L2<=$P$9, "c 1", IF(L2<=$P$10, "c 2", IF(L2<=$P$11, "c 3", "c 4")))
  + Fórmula para valor monetario: =IF(V3<=$Z$9, "c 1", IF(V3<=$Z$10, "c 2",F(V3<=$Z$11, "c 3", "c 4")))

| **cuartil\_dias\_entre\_compras** | | | | |
| --- | --- | --- | --- | --- |
| **0** | **Q1** | **Q2** | **Q3** | **4** |
| 0 | 54 | 133 | 243 | 879 |
| cuartil 1 | |  |  |  |
|  | cuartil 2 | |  |  |
|  |  | cuartil 3 | |  |
|  |  |  | cuartil 4 | |

* 1. Se creó la tabla "Segmentacion\_RFM" utilizando VLOOKUP para traer los valores de"RFM\_cuartiles". Se asignaron puntajes en una escala de 1 a 4 para cada dimensión (recencia, frecuencia, valor monetario) y se concatenaron estos resultados en "RFM\_score" para obtener un puntaje combinado.
  2. Los clientes en el cuartil superior(Q3) para "frecuencia" y "valor\_monetario" recibieron una calificación más alta(4). Para "recencia", los clientes en el cuartil inferior(Q1) recibirán una calificación más alta(4) ya que se considera mejor haber comprado más recientemente.
  3. Se utilizaron fórmulas IFS para clasificar a los clientes en categorías como "mejores clientes", "clientes leales", "clientes nuevos", "clientes en riesgo" y "clientes perdidos" basándose en su puntaje RFM.
     1. Ejemplo: =IF(B2<=54, 4, IF(B2<=133, 3, IF(B2<=243, 2, IF(B2<=879, 1, ""))))
  4. También se creó la variable “RFM\_score\_sum” donde se sumaron los resultados de las variables para tener la opción de medirlos por puntaje numérico y no solo por texto. El resultado de la clasificación de los clientes fue:

| **Clasificación de clientes** | **Puntaje (mínimo 3 - máximo 12)** | **Definición** |
| --- | --- | --- |
| Mejores clientes | >=10 | ha comprado recientemente, compra frecuentemente, gasta bastante |
| Clientes leales | 8-9 | compran a menudo, pero la recencia y el monto pueden variar |
| Clientes nuevos | 6 - 7 | han comprado recientemente,pueden ser su primera o segunda compra, el monto puede variar |
| Clientes en riesgo | 4 - 5 | han comprado frecuentemente en el pasado, pero hace tiempo que no compran |
| Clientes perdidos | <=3 | han dejado de comprar |

* 1. Se aplicó la fórmula de segmentación usando la función IFS:
     1. Ejemplo: =IFS(I2>=10,"mejores clientes",I2>=8,"clientes leales",I2>=6, "clientes nuevos",I2>=4, "clientes en riesgo",I2<=3, "clientes perdidos",TRUE,"otros")
  2. Un formato condicional en colores se aplicó a cada variable resultante de la columna “clasificacion\_clientes”
  3. Finalmente se creó una gráfica de barras para visualizar el resultado de la clasificación de clientes.

1. **Análisis de Cohorte:**
   1. Se creó una tabla auxiliar utilizando la función QUERY para extraer variables necesarias de "transacciones". Se generaron nuevas variables como "año\_mes\_entrada" y "año\_mes\_transaccion" usando DATEDIF, ROUND, TEXT y CONCATENATE.
   2. Seguido, se creó una tabla para sacar las medidas de tendencia central: moda, mediana, media, deviation tipica, minimo, máximo y se genero un histograma para comprender el comportamiento a traves del tiempo.
   3. Se construyó una tabla dinámica "Pt\_cohorte\_clientes" para definir cohortes y visualizar los porcentajes de retención utilizando formato condicional y un mapa de calor, donde el color verde se refiere a una alta retención de clientes y del color amarillo a rojo, una disminución en la retención, con rojo indicando las tasas más bajas.
   4. En una pestaña llamada "tabla\_retencion", se calcularon los porcentajes de retención y abandono, creando gráficos de línea para facilitar la visualización de tendencias a lo largo del tiempo.

## 

## 

## **5. Conclusiones**

#### *Perfil de los Clientes:*

* Datos demográficos de Clientes: Los clientes tienen entre 24 y 81 años, con el 50% en el rango de 45 a 64 años. Su ingreso anual promedio es de $52,247 y poseen estudios de secundaria, grado y postgrado. En promedio, tienen un hijo y la mayoría están casados, solteros o en unión de hecho.
* Respuesta a Campañas de Marketing: Solo el 14.8% de los clientes respondió a la campaña de marketing, lo que indica una oportunidad para mejorar la efectividad de las campañas de marketing.

#### *Preferencias de Compra:*

#### Productos Populares: Los vinos son los productos más populares entre los clientes, seguidos de cerca por las carnes. Esto sugiere una fuerte demanda en estas categorías.

#### Productos Menos Populares: Las frutas son el producto menos consumido, lo que indica una posible área de mejora en términos de promoción y oferta.

* Canales de Compra: Los clientes están divididos casi por igual entre compras en línea y en tienda. Sin embargo, en 2022, se observa una ligera inclinación general hacia las compras en tienda física, lo que podría deberse a la preferencia por experiencias de compra físicas.

*Segmentación de Clientes:*

* Clientes Nuevos: El 19.4% de los clientes son nuevos, lo que es una señal positiva de que se están atrayendo nuevas personas a la tienda.
* Clientes en Riesgo: El 26.9% de los clientes se encuentran en riesgo de dejar de comprar, esto indica la necesidad de estrategias efectivas de retención para este segmento.

*Análisis de cohortes, retención y abandono:*

* Alta retención inicial: Todos los clientes están activos en el mes 0 con un 100%.
* Disminución progresiva: Los porcentajes de retención disminuyen conforme pasan los meses desde la primera compra.
* Patrones Cíclicos: Hay ciertos meses en los que la retención cae de manera más pronunciada.
* En los años 2020-2021, se observa una caída significativa en la retención de clientes, especialmente después de los primeros 3 a 6 meses.
* En el año 2022, la retención en los primeros meses parece mejorar en comparación con los años anteriores.
* En general, entre 6 y 7 meses y entre 12 y 13 meses, hay una notable disminución en la retención.

## **6. Recomendaciones**

* Segmentación de Mejores Clientes y Clientes Leales: Implementar programas exclusivos y descuentos.
* Segmentación de Clientes Nuevos y en Riesgo: desarrollar estrategias de retención temprana.
* Promociones y Eventos: Atraer más clientes mediante promociones y eventos especiales.
* Campañas de Marketing: Aumentar el interés en productos menos vendidos.
* Estrategia de Producto: Mantener o aumentar la variedad y calidad en las categorías de vinos y carnes, y mejorar la oferta de productos menos populares.
* Diversificación: Explorar la introducción de nuevos productos complementarios y programas de fidelización para mantener el interés de los clientes.
* Diseñar programas de fidelización y recompensas: para mantener a los clientes activos durante más tiempo, puede ser campañas enfocadas en estaciones o épocas del año para maximizar la retención.

## **7. Referencias**

* Herramientas utilizadas: Google Sheets, Google Slides.

Elaborado por: Natalia Alejandro González.  
Fecha: miércoles 12 de junio de 2024.