

CODING BOOTCAMPES ESPOL: Data-Driven Decisions Specialist

Data Visualization

Grupo 4

Proyecto: Campañas y Respuestas

Entregable 1

Integrantes:

Forero Villota Katherine Sheila

Gonzales Ochoa Jean Pierre

León Santos Henrry Josue

Mora Verdezoto Luis Enrique

Morales Lucas José Daniel

Estado Inicial del Dataset.....	3
Estado Actual del Dataset con Mejoras Aplicadas.....	3
Enfoque Analítico, Preguntas de Negocio y KPIs.....	4
Preguntas de Negocio.....	4
Perfil del Cliente.....	4
Definición de KPI's.....	4

Estado Inicial del Dataset

El dataset presentaba varios problemas de calidad que afectaban directamente su uso para análisis y modelado. Las columnas monetarias estaban almacenadas como texto debido a la presencia del símbolo “\$”, lo que impedía realizar cálculos numéricos correctamente. Además, existían variables categóricas representadas mediante múltiples columnas one-hot, lo que aumentaba innecesariamente la dimensionalidad del dataset y dificultaba su interpretación. Aunque el número de registros era consistente y no había valores nulos, se detectaron valores inválidos como montos negativos en columnas monetarias y una variable binaria (AcceptedCmpOverall) con valores fuera de su dominio esperado. También se encontraron registros duplicados y columnas que no aportaban valor analítico, lo que indicaba que el dataset contenía ruido, redundancia y errores lógicos que podían distorsionar cualquier análisis posterior.

Estado Actual del Dataset con Mejoras Aplicadas

Tras el preprocesamiento, el dataset fue transformado en una versión más limpia, coherente y lista para análisis. Las columnas monetarias fueron convertidas a formato numérico, eliminando el símbolo “\$” para permitir cálculos y agregaciones correctas. Se reconstruyeron las variables categóricas Marital_Status y Education a partir de sus columnas one-hot, concentrando la información en una sola columna por variable y reduciendo la dimensionalidad y redundancia. Se identificaron y trajeron valores inválidos como montos negativos (también se dio un ejemplo de como tratar una edad inválida dadas las instrucciones, pero realmente en el dataset no existían valores fuera de lo normal), proponiendo imputaciones basadas en estadísticas robustas como la mediana. Asimismo, se eliminaron duplicados y columnas irrelevantes que no aportaban valor analítico. En conjunto, estos cambios mejoran significativamente la calidad del dataset al aumentar su consistencia, reducir el ruido, corregir errores lógicos y hacer que los datos sean más interpretables y confiables para análisis estadísticos y modelos de datos.

Enfoque Analítico, Preguntas de Negocio y KPIs

Preguntas de Negocio

Perfil del Cliente

Se proponen 5 preguntas estratégicas diseñadas para entender cómo las características demográficas influyen en el comportamiento de compra.

1. **¿Qué nivel educativo presenta mayor gasto total promedio?**
 - *Justificación:* Determinar si existe una correlación directa entre la formación académica (como proxy de estatus socioeconómico) y el volumen de compra.
2. **¿Los clientes con carga familiar (hijos/adolescentes) tienen un ticket promedio menor?**
 - *Justificación:* Validar la hipótesis de que los clientes con dependientes destinan menos presupuesto a productos de consumo premium o discrecional.
3. **¿La antigüedad del cliente impacta positivamente en el gasto acumulado?**
 - *Justificación:* Analizar si la fidelidad (medida en días desde el registro) se traduce en un mayor *Customer Lifetime Value*.
4. **¿Existe una brecha significativa de ingresos anuales según el estado civil?**
 - *Justificación:* Entender la capacidad adquisitiva de los diferentes segmentos (Solteros vs. Casados vs. Divorciados) para personalizar la oferta de productos según su "bolsillo".
5. **¿Qué grupo generacional (Rango de Edad) muestra mayor disposición a aceptar campañas de marketing?**
 - *Justificación:* Identificar qué perfil demográfico es más reactivo a los estímulos publicitarios para optimizar la inversión en marketing (ROAS).

Definición de KPI's

Hemos definido 5 indicadores clave alineados a las preguntas de negocio planteadas anteriormente. Cada KPI ha sido diseñado para medir una dimensión específica del perfil demográfico del cliente.

Matriz de Definición de KPIs - Perfil de Cliente

#	Pregunta de Negocio	Nombre del KPI	Definición Técnica	Propósito del Negocio
1	¿Qué nivel educativo presenta mayor gasto?	Gasto Promedio por Nivel Educativo	Suma(MntTotal) / Total Clientes (por Nivel)	Identificar segmento académico de mayor valor (High-Value).
2	¿Los clientes con hijos gastan menos?	Gap de Ticket Promedio (Padres vs. No Padres)	Promedio(Sin Hijos) - Promedio(Con Hijos)	Cuantificar contracción del consumo por carga familiar.
3	¿La antigüedad impacta el gasto?	Valor del Cliente por Antigüedad	Promedio(MntTotal) agrupado por años de registro	Validar si la lealtad se traduce en rentabilidad.
4	¿Existe brecha salarial por estado civil?	Ingreso Medio Anual por Estado Civil	Promedio(Income) por cat. Marital_Status	Segmentar ofertas según capacidad de pago real.
5	¿Quién acepta más campañas?	Tasa de Conversión por Edad	(Aceptaciones / Total Clientes en Rango Edad) * 100	Definir target generacional para optimizar publicidad.