

Autor(es): Juan Sebastian Pelaez Pardo - Steban Nicolas Tibata Castañeda - Diego Alvaro Morales Medrano
Fecha: 27/10/2024

DATA SOLUTION: AUTOMATED CAMPAIGN PROCESS

1 DEFINICIÓN DE LA PROBLEMÁTICA Y ENTENDIMIENTO DEL NEGOCIO:

Organización

VML, una de las principales agencias de marketing globales y la segunda más grande del mundo, opera en más de 170 países. Trabaja con clientes importantes como Unilever, Coca-Cola, Microsoft y Colgate, a quienes ofrece soluciones avanzadas de marketing digital para mejorar su posicionamiento y crecimiento en el mercado.

Problemática

La agencia enfrenta la necesidad de mejorar el análisis de campañas publicitarias mediante la automatización y centralización de datos, lo que permitirá obtener insights en tiempo real para ajustar estrategias y maximizar el retorno de inversión. Actualmente, el equipo consulta los datos en Facebook Analytics y los ingresa manualmente en Excel, lo que ralentiza el proceso y limita la agilidad en la toma de decisiones. Para solucionar esto, se desarrollarán dos productos: dashboards preliminares automatizados que eliminen la dependencia del ingreso manual y permitan la visualización y monitoreo de métricas de campaña en tiempo real, y un modelo de clasificación basado en Machine Learning que determine el éxito de una campaña en función de un umbral de éxito definido en conjunto con el negocio.

Objetivo General

Desarrollar dos productos que optimicen el análisis de las campañas publicitarias: dashboards preliminares para la visualización y monitoreo en tiempo real de métricas de campaña y un modelo de clasificación basado en Machine Learning que determine el éxito de cada campaña en función de un umbral definido en colaboración con el negocio.

Objetivos Específicos

1. Construcción de Dashboards Preliminares en Looker Studio y PowerBI:

Configurar dashboards que ofrezcan una visualización clara y en tiempo real de KPIs clave (como clics, impresiones, alcance y costos).

2. Desarrollo del Modelo de Clasificación Basado en Machine Learning:

Definir reglas juntamente con la empresa para poder definir un umbral para clasificar si etiquetar si una campaña es exitosa o no.

Crear un modelo de Machine Learning que clasifique las campañas como exitosas o no exitosas en función de un umbral de éxito definido junto al negocio.

Métricas de negocio (KPI's)

1. Rentabilidad por campaña: Dicho indicador ayudará a medir el impacto de cada campaña dependiendo de su inversión, VML revisará el indicador para poder optimizar el uso de campañas y su impacto respectivo.
2. Alcance generado: Dicho KPI será de suma importancia ya que es la métrica más importante en el mundo del marketing, pues si una campaña llega a más usuarios significa que es efectiva.
3. Visualización efectiva: Dicho indicador será el total de información plasmada dentro del dashboard que sea de importancia para VML. En este caso, es una percepción por más del cliente que nos permita tener un feedback acerca de la construcción del dashboard junto a las métricas más importantes según la necesidad que en este caso es el tema de gestión de las inversiones a lo largo de las campañas.

2 IDEACIÓN:

El producto de datos será enfocado en temas de visualización y ciencia de datos que permita a VML tomar mejores decisiones a la hora de invertir en los diferentes modelos y campañas. Tanto los accionables de flujo de datos, como los descubrimientos estarán plasmados en un dashboard automatizado que permita ver todos los datos en tiempo real y además que permita dar una idea amplia acerca de la toma de decisiones. Debido a que la mayor problemática es que deben tomar decisiones a dedo y sus herramientas de visualización no son automáticas lo cual hace que tomen mucho tiempo en ajustar las visualizaciones y a su vez no tienen el tiempo suficiente de tomar mejores decisiones en las inversiones, además que el equipo de pauta (es decir, los encargados de campañas pagas) está conformado por una sola persona, lo cual hace que la tarea sea muy difícil de cumplir.

De manera general, los requerimientos están divididos en 3 puntos: El primero es acerca del consumo de datos ya que debe ser de manera automatizada que permita ahorrar tiempos de consulta, de manera tecnológica será el uso de un API a través de Google Sheets para recoger la información. El segundo es el tema de visualizaciones en el cual debe mostrar toda la información necesaria a la hora de tomar las mejores decisiones para invertir el dinero, en términos tecnológicos deben ser visualizaciones en LookerStudio ya que todas sus herramientas están desarrolladas en Google Cloud, y el tercero hace referencia a un modelamiento donde VML pueda revisar cuál campaña está dando buenos resultados en términos de inversión y rendimiento por campaña, ya que así pueden revisar cuales modelos son los más rentables y a cuáles se debe enfocar una estrategia de marketing distinta, en términos tecnológicos hace referencia a un modelamiento de clasificación que permita mostrar esas campañas sobresalientes, todo desarrollado en Python para poder conectarlo en el futuro a BigQuery.

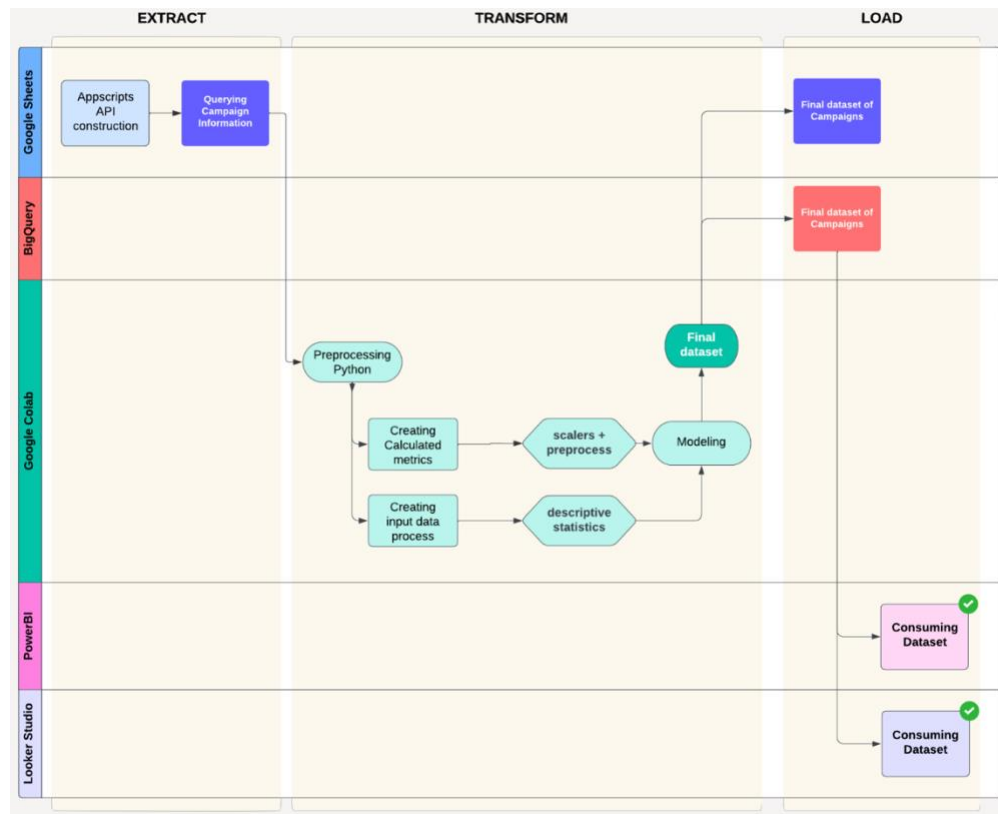


Ilustración 1. Mockup de Solución

3 RESPONSABILIDAD ÉTICA Y LEGAL:

Para garantizar un análisis ético en el uso de datos y Machine Learning en marketing, es fundamental priorizar la privacidad y transparencia, asegurando que los usuarios comprendan cómo se recolecta y utiliza su información y usar datos privados y sensibles de forma anonimizada. Además, prevenir el sesgo en los modelos es crucial, lo cual implica trabajar con datos representativos y realizar auditorías periódicas para mantener la equidad en las predicciones. Finalmente, los resultados generados por los modelos deben actuar como un apoyo en la toma de decisiones, dejando la interpretación final al equipo de marketing para usar los insights de manera ética y responsable, promoviendo decisiones bien fundamentadas y alineadas con los valores de la organización.¹

4 ENFOQUE ANALÍTICO:

Las preguntas que se quieren responder son ¿Qué campañas de marketing son exitosas? ¿Qué características tienen las principales campañas de marketing? ¿Cuáles campañas de marketing son exitosas?

Para responder a estas preguntas, se implementarán diversas técnicas estadísticas y de Machine Learning. En primer lugar, se llevará a cabo un análisis descriptivo para resumir y analizar los datos de las campañas, utilizando medidas de tendencia central, como promedios y medianas, y dispersión,

¹ **Ethical Data Use in Marketing:** Malthouse, E. C., & Li, H. (2017). "Opportunities for and Pitfalls of Using Big Data in Marketing." *Journal of Interactive Marketing*, 39, 44-54. Discute la importancia de usar los insights de datos de manera ética, alineando estrategias de marketing con prácticas responsables y transparentes.

como desviación estándar y percentiles, para entender el desempeño general de las campañas. Además, se crearán visualizaciones de datos, incluyendo gráficos de barras y diagramas de dispersión, que permitirán visualizar la relación entre diferentes variables, como clics, impresiones, alcance y ROI. Esto facilitará la identificación de características comunes en las campañas exitosas. También se desarrollará un modelo de clasificación utilizando técnicas de Machine Learning, como regresión logística o árboles de decisión, para predecir el éxito de las campañas en función de sus características y con un umbral que se establecerá dependiendo de las variables descriptivas que definan el éxito de una campaña. El uso de métodos de validación cruzada asegurará la robustez del modelo y su capacidad de generalización.

Finalmente, se establecerán métricas específicas para evaluar la calidad del modelo. La precisión se medirá como la proporción de verdaderos positivos sobre el total de predicciones, lo que indica la efectividad del modelo al identificar campañas exitosas. El recall, que mide la capacidad del modelo para detectar todos los casos reales positivos, será otra métrica clave. La F1-Score, que representa la media armónica entre precisión y recall, proporcionará un balance útil entre ambas métricas, especialmente en escenarios con clases desbalanceadas. También se evaluará el área bajo la curva ROC (AUC-ROC) para medir la capacidad del modelo para distinguir entre clases. Por último, se analizarán métricas de retorno de inversión (ROI) para determinar la efectividad de las estrategias de marketing. Estas hipótesis, técnicas y métricas en conjunto proporcionarán un marco sólido para el proceso de experimentación y asegurarán que los resultados sean relevantes y útiles para la toma de decisiones dentro de la agencia.

5 RECOLECCIÓN DE DATOS:

Los datos de cada campaña, con una granularidad diaria, se extraen a través de la conexión de Apps Script con la API de Facebook. Posteriormente, estos datos se procesan en Python y se cargan en BigQuery y Google Sheets, permitiendo su consumo en los dashboards de Looker Studio y Power BI para un análisis visual en tiempo real.

Diccionario de Datos

A continuación, se presenta un diccionario de datos que detalla todas las columnas disponibles en el dataset de campañas para facilitar su interpretación y análisis.

Columna	Descripción
Date	Fecha de periodicidad diaria, captura el registro de las métricas de cada una de las campañas.
Campaign type	Tipo de campaña (Branding, Performance, Tácticos, Marketing Cloud)
Campaign name	Nombre de la campaña
Ad set name	Junto de anuncios dentro de cada campaña, es decir un Campaign Name puede tener varios ad set name
Model	Modelo de moto respectiva

Unique leads	Leads únicos se refiere a cada persona que ha completado el formulario de generación de leads a través de los anuncios de la campaña.
Cost	Costo de cada anuncio (o también conocido como inversión ejecutada)
Clicks (all)	Total de clicks
CPC (all)	Costo por click, costo por cada click a un anuncio
Link clicks	Total de link clicks
CPC (cost per link click)	Costo por cada link click
Impressions	Cada vez que el anuncio sale en la pantalla de un usuario
Reach	Cuántas personas han visto el anuncio
Unique clicks (all)	Clicks únicos que el usuario a ingresado al anuncio. Es decir, si un mismo usuario hace click en el mismo anuncio 3 veces, solo se cuenta una vez
CPM (cost per 1000 impressions)	Costo asociado por cada 1000 veces que el anuncio sale en la pantalla de un usuario.
Cost per action (CPA)	Costo asociado a una adquisición de un usuario nuevo
Inversion Planeada	(Periodicidad mensual) Inversión planeada por parte del cliente
CPL proyectado	(Periodicidad mensual) Costo por lead generado planeado
Meta Leads	(Periodicidad mensual) Meta de leads planeada
Meta Conversaciones	(Periodicidad mensual) Meta de conversaciones, específicamente para campañas de whatsapp
Meta Clicks	(Periodicidad mensual) Meta de clicks planeados
CPC proyectado	(Periodicidad mensual) Costo por click planeado
Meta Seguidores	(Periodicidad mensual) Meta de seguidores planeado
CPS proyectado	(Periodicidad mensual) Costo por seguidor
Meta interacciones	(Periodicidad mensual) Meta de interacciones planeado
CPE Proyectado	(Periodicidad mensual) Costo por engagement generado
Meta Alcance	(Periodicidad mensual) Meta de alcance por campaña planeado
CPMA proyectado	(Periodicidad mensual) Costo por cada mil alcances proyectado
CPA proyectado	(Periodicidad mensual) Costo por Alcance proyectado

6 ENTENDIMIENTO DE LOS DATOS:

Dimensiones y Variables:

- Total de registros: 4,735
- Variables: 31 (25 cuantitativas y 6 cualitativas)
- Duplicados: 0 registros duplicados detectados

Matriz de Correlación

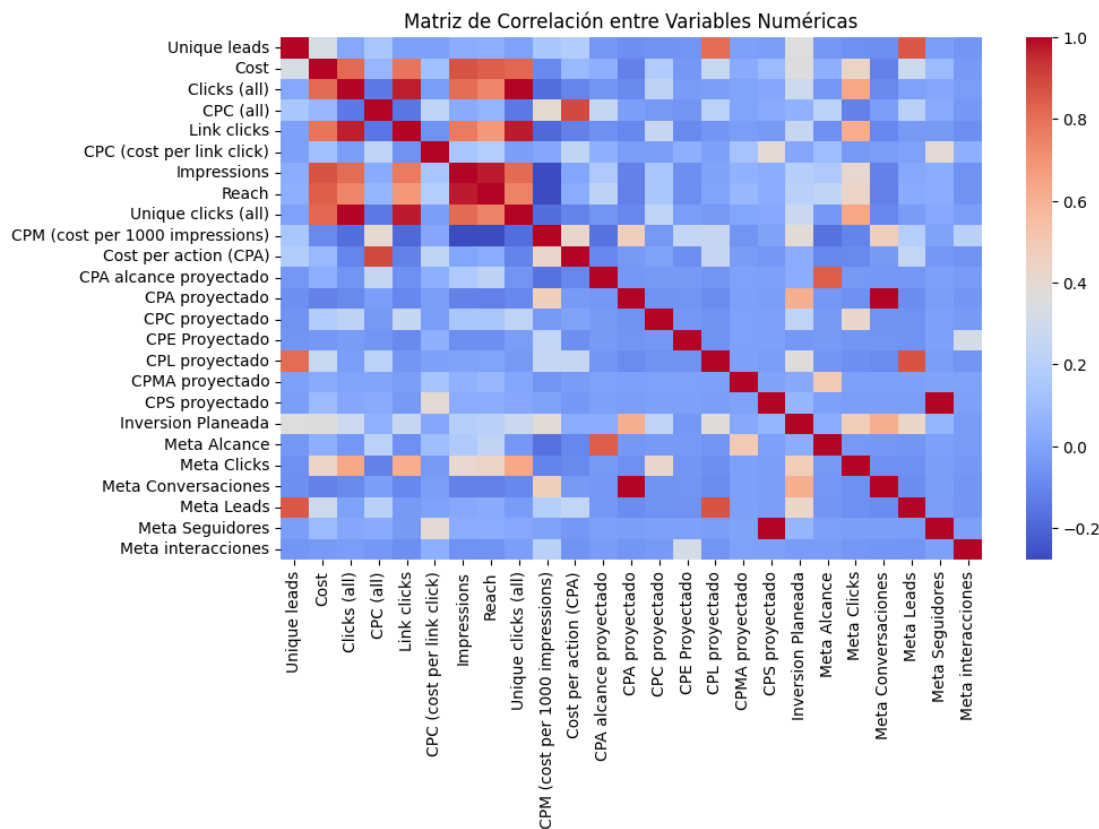


Ilustración 2. Matriz de Correlación

La matriz de correlación muestra la relación entre diversas métricas de las campañas. Se observa una correlación positiva entre el costo de las campañas y el número de clics obtenidos, lo cual indica que un mayor presupuesto suele asociarse con un aumento en la cantidad de clics. Sin embargo, al analizar el **Costo por Clic (CPC)** y los **Link Clicks**, no se encuentra una correlación tan alta como se esperaría, sugiriendo ineficiencias en la relación entre el gasto y los clics en enlace.

Relación entre costos y clics

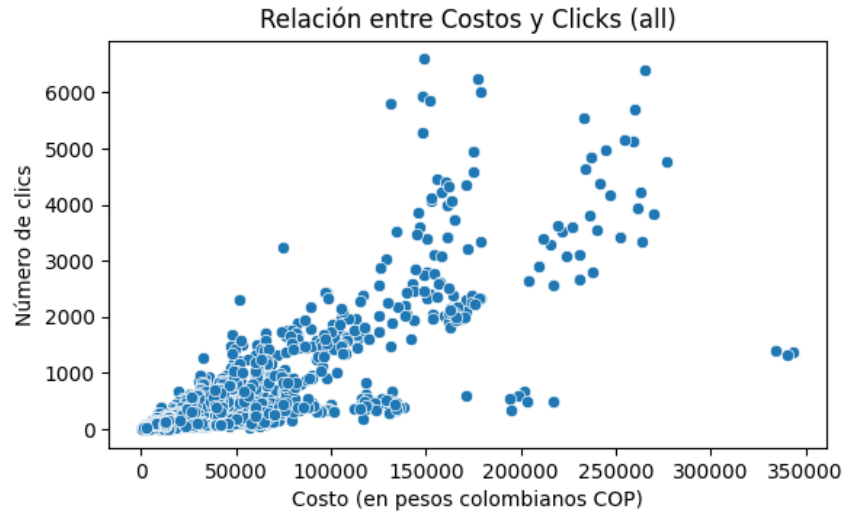


Ilustración 3. Gráfico de Dispersión: Relación entre Costos y Clicks (all)

Este gráfico de dispersión muestra la relación entre los costos de las campañas (expresado en pesos colombianos) y el número de clics obtenidos. Se observa que, en general, a mayor inversión en la campaña, mayor es el número de clics. Sin embargo, se identifican algunos puntos dispersos que representan campañas con altos costos, pero un bajo rendimiento en términos de clics, lo cual sugiere ineficiencias que podrían ser optimizadas. Este análisis indica una correlación positiva entre el costo y los clics, aunque no siempre proporcional, sugiriendo que algunas campañas no están maximizando su impacto.

Alcance y clics en enlace por tipo de campaña

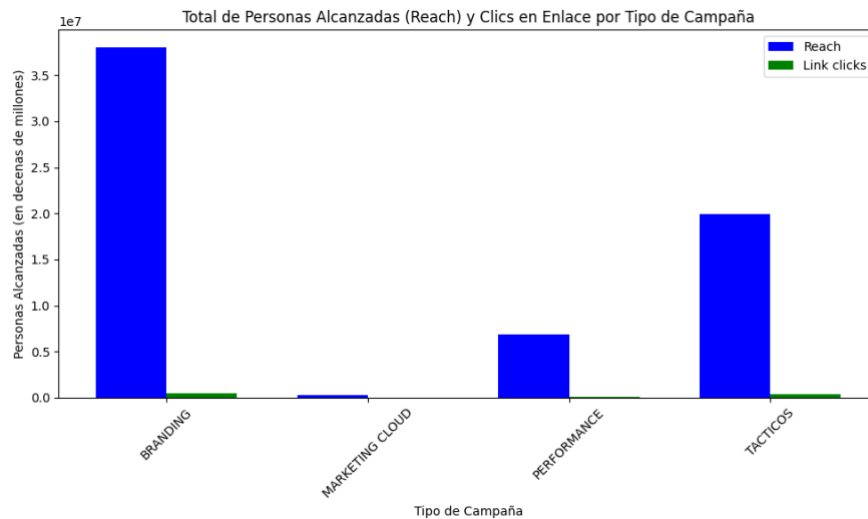


Ilustración 4. Gráfico de Barras: Total de Personas Alcanzadas (Reach) y Clics en Enlace por Tipo de Campaña

Este gráfico compara el alcance total y los clics en enlace de cada tipo de campaña (Branding, Marketing Cloud, Performance y Tácticos) en un mes. Las campañas de Branding destacan significativamente en alcance, alcanzando a más de 35 millones de personas, aunque generan relativamente pocos clics en enlace, lo que podría indicar un alto nivel de exposición, pero una baja interacción en comparación con su alcance. Las campañas Tácticas también muestran un buen

alcance, mientras que las campañas de Marketing Cloud tienen menor impacto. Este análisis sugiere que, aunque las campañas de Branding logran una amplia visibilidad, es necesario ajustar la estrategia para aumentar el engagement y mejorar la conversión.

Leads únicos para las campañas con más registros a lo Largo del tiempo

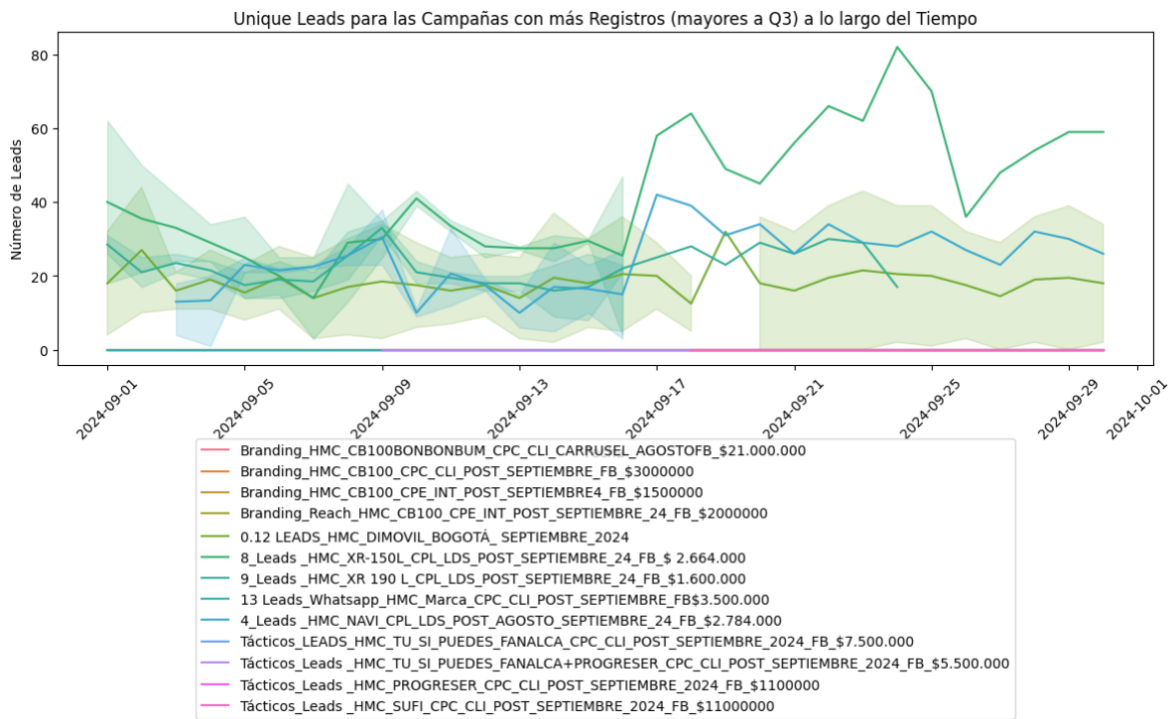


Ilustración 5. Gráfico de Líneas: Unique Leads para las Campañas con Más Registros (mayores a Q3)

Este gráfico muestra la evolución de los leads únicos generados a lo largo del tiempo para las campañas con el mayor número de registros, destacando aquellas que están en el cuartil superior (mayores a Q3). Las campañas de Branding y Tácticas tienden a generar un número más elevado de leads de manera consistente. Además, se observan picos significativos en algunas fechas específicas, lo cual podría estar relacionado con promociones o eventos especiales. Este comportamiento indica que ciertos periodos son más efectivos para captar leads, y estas tendencias pueden ser aprovechadas para futuros lanzamientos estratégicos.

7 CONCLUSIONES INICIALES:

A partir del análisis exploratorio de datos (EDA) se obtuvo los siguientes insights:

Matriz de Correlación:

- La matriz de correlación muestra que, aunque existe una relación positiva entre muchas variables lo que se esperaba, la baja correlación entre el **Costo por Clic (CPC)** y los **Link Clicks** indica posibles ineficiencias que deben ser analizadas para optimizar el gasto publicitario.

Alto alcance, pero baja Interacción en campañas de Branding:

- Las campañas de Branding logran un amplio alcance, pero presentan una baja tasa de clics en enlace, lo que sugiere que la visibilidad no se traduce en interacción significativa.

Identificación de picos temporales en captación de Leads:

- Se observan picos de generación de leads en fechas específicas, lo cual sugiere que ciertos periodos son más efectivos para captar leads.

Rendimiento bajo de campañas de marketing Cloud:

- Las campañas de Marketing Cloud tienen un rendimiento relativamente bajo en términos de alcance y generación de leads.

Ineficiencias en costos de Campañas con bajo rendimiento en clics:

- Existen campañas con altos costos que no logran un número proporcional de clics, lo que indica ineficiencias en el gasto.

Próximas acciones

Las siguientes acciones consisten en compartir los hallazgos con VML para definir conjuntamente las reglas que establecerán el umbral de éxito de las campañas. Paralelamente, se iniciará el desarrollo de los dashboards que permitirán a la empresa visualizar sus datos de manera efectiva y facilitarán la toma de decisiones.

Referencias:

1. Ethical Data Use in Marketing: Malthouse, E. C., & Li, H. (2017). "Opportunities for and Pitfalls of Using Big Data in Marketing." *Journal of Interactive Marketing*, 39, 44-54.