

Autor(es): Juan Sebastian Pelaez Pardo - Steban Nicolas Tibata Castañeda - Diego Alvaro Morales Medrano
Fecha: 01/12/2024

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. La problemática inicial es identificar cuáles son buenas campañas y qué características tienen, sin embargo, para la solución se necesita efectuar el flujo de datos junto la creación de dashboards, requieren un desarrollo de un API que consuma datos de campañas, junto un proceso claro de ETL que despliegue los datos en Bigquery, esto para el consumo en dashboards como LookerStudio y PowerBI, más la implementación de ciencia de datos que permita una mejor toma de decisiones a la hora de recomendar modelos de inversión al cliente.

Objetivo General

Desarrollar dos productos que optimicen el análisis de las campañas publicitarias: dashboards preliminares para la visualización y modelo de clasificación de Machine Learning que determine el éxito de cada campaña en función de un umbral definido en colaboración con el negocio. Al final se obtiene un producto de datos basado en dos herramientas, la primera como una guía de recomendación acerca de las posibilidades de éxito que tiene las campañas dados unos inputs iniciales, así el cliente puede experimentar con las mejores combinaciones, y además un dashboard que permita ver los insights más significativos de esas campañas exitosas del mes.

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 ver las características más importantes de esas campañas exitosas. 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 costo por click asociado, 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.

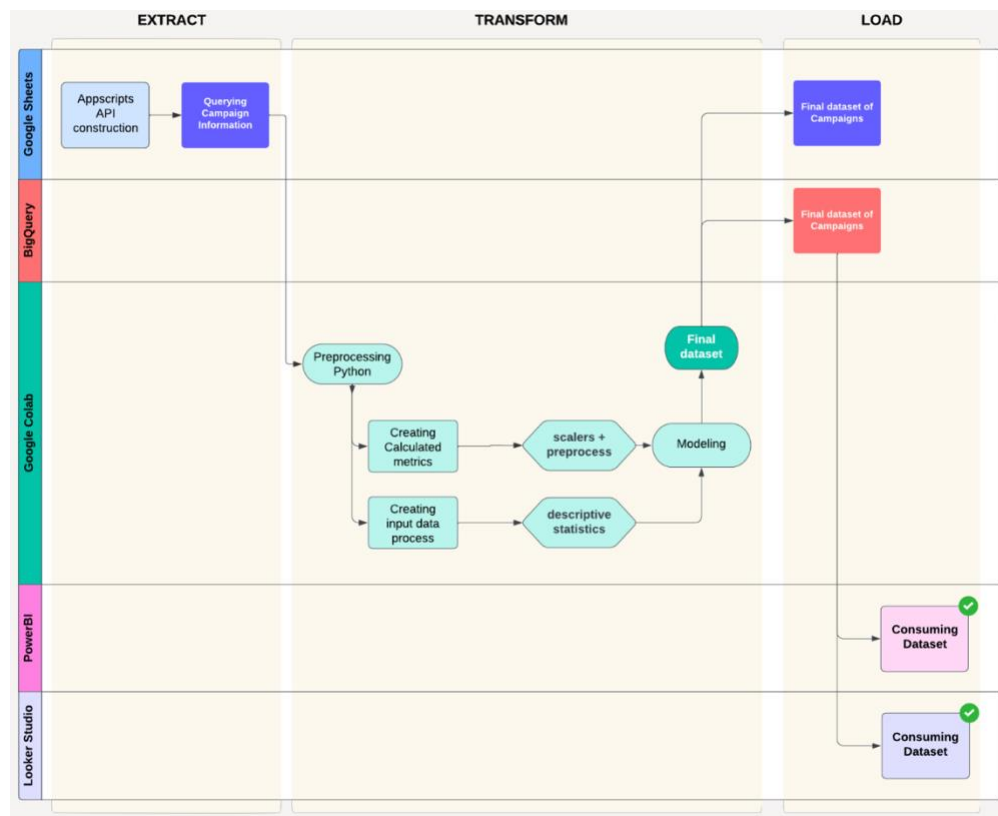


Ilustración 1. Mockup de Solución

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 campañas 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, como desviación estándar y percentiles, para entender el desempeño general de las campañas.

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.

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 del costo por click. 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.

Directas de Facebook Ads
Métricas de input del cliente

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 campana
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

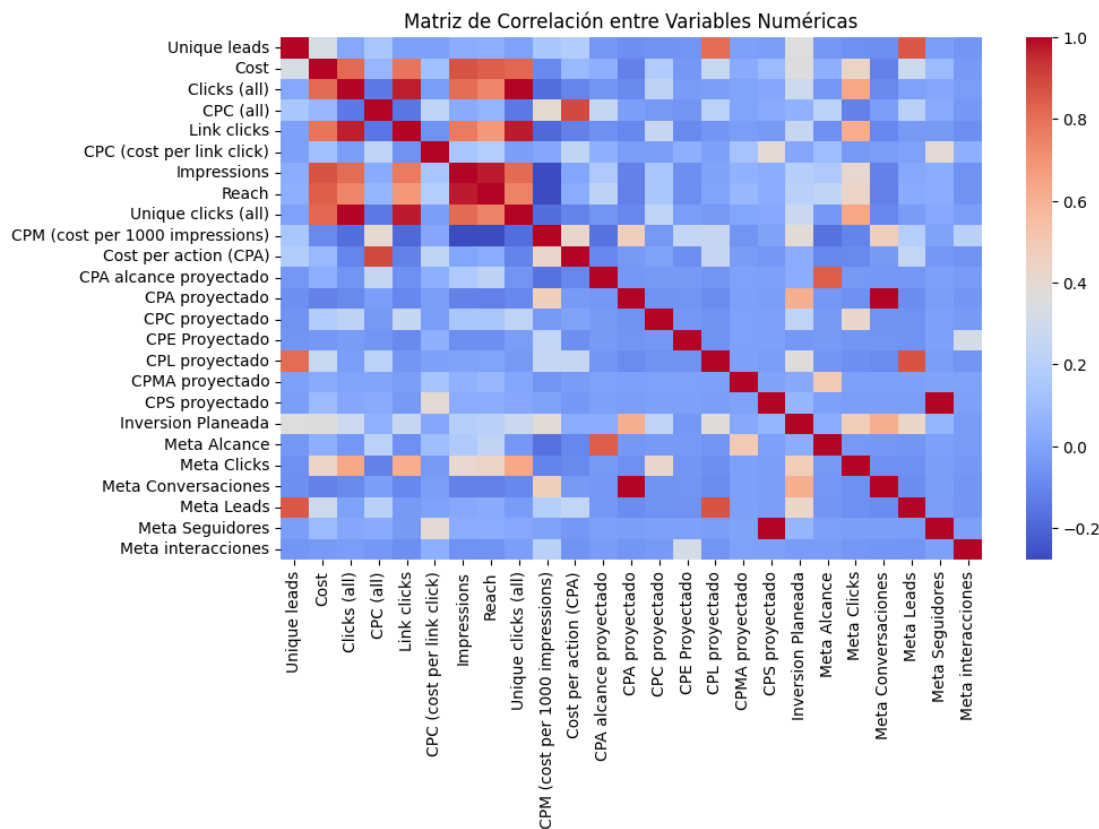


Ilustración 2. Matriz de Correlación

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

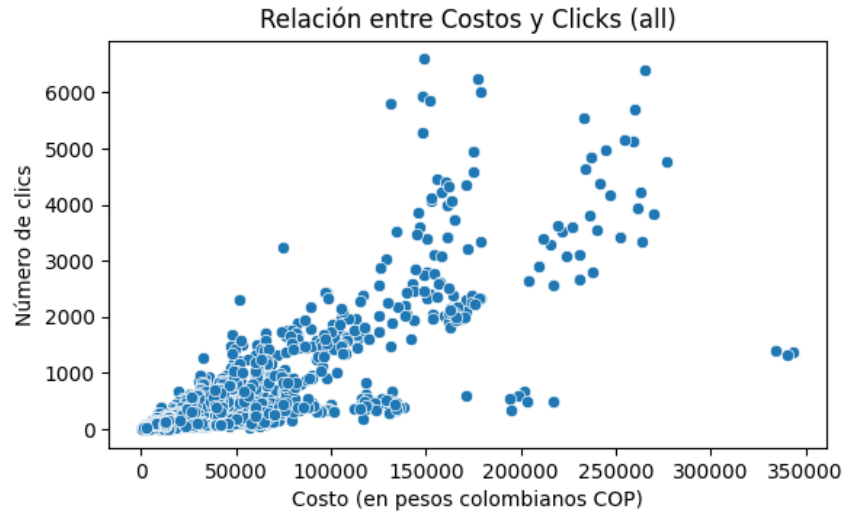


Ilustración 3. Gráfico de Dispersión: Relación entre Costos y Clicks (all)

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

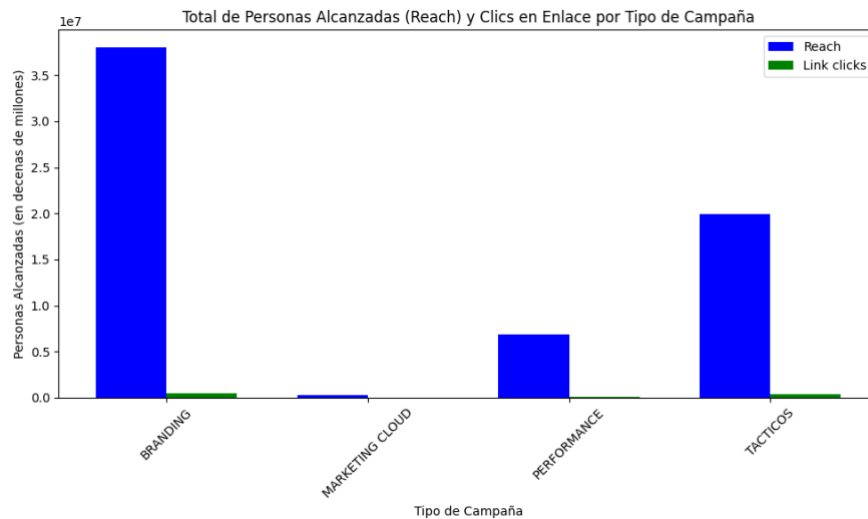


Ilustración 4. Gráfico de Barras: Total de Personas Alcanzadas (Reach) y Clics en Enlace por Tipo de Campaña

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

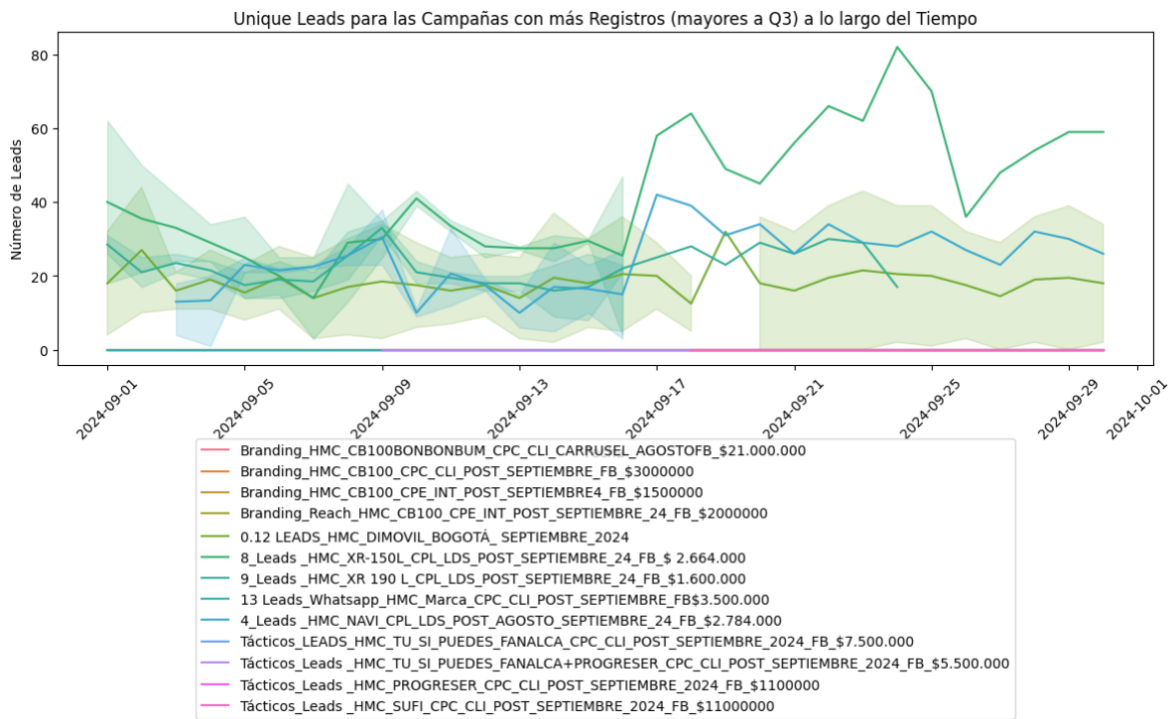


Ilustración 5. Gráfico de Líneas: Unique Leads para las Campañas con Más Registros (mayores a Q3)

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.

8 PREPARACIÓN DE DATOS

En primer lugar, se decidió eliminar las filas de la columna 'Inversión Planeada' que contienen valores iguales a 0, ya que estos no son adecuados para el modelo y podrían afectar negativamente su desempeño, estos representan el 20% del total de los datos.

Se estableció un umbral de éxito para el modelo en colaboración con los expertos del negocio, definiendo como exitosas aquellas campañas cuyo costo por clic sea inferior a 80, ya que este valor representa una inversión eficiente. Un costo por clic por debajo de 80 se considera rentable, permitiendo optimizar los recursos y maximizar el retorno de la inversión.

Para este análisis, las campañas exitosas fueron etiquetadas con un 1, mientras que las no exitosas recibieron un 0 en la columna 'Success', que se utilizó como variable objetivo para predecir el éxito de futuras campañas.

Anticipar si una campaña será exitosa antes de ejecutarla puede proporcionar un gran ahorro al cliente, ya que permite ajustar el tipo de campaña o recalibrar las metas, logrando un punto óptimo donde los recursos invertidos generen una campaña eficiente y exitosa.

En colaboración con los expertos del negocio, se identificaron las principales variables que influyen en el éxito de una campaña. Estas variables son: 'Inversión Planeada', 'Meta Leads', 'Meta Conversaciones', 'Meta Clicks', 'Meta Seguidores', 'Meta Interacciones', 'Meta Alcance' y 'Campaign Type'. Estas características fueron seleccionadas como variables de entrada para el modelo. Todas las variables son numéricas, excepto 'Campaign Type', que es categórica. Para esta última, se aplicó un proceso de codificación mediante one-hot encoding, lo que generó nuevas columnas que representan los tipos de campaña: Performance, Branding y Tácticos. Finalmente, luego del procesamiento de los datos, los valores de entrenamiento presentan las siguientes características:

```

Index: 2173 entries, 5 to 2770
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   Inversion Planeada     2173 non-null   int64  
1   Meta Leads             2173 non-null   int64  
2   Meta Conversaciones    2173 non-null   int64  
3   Meta Clicks            2173 non-null   int64  
4   Meta Seguidores        2173 non-null   int64  
5   Meta interacciones     2173 non-null   int64  
6   Meta Alcance           2173 non-null   int64  
7   Campaign type          2173 non-null   object  
8   Success                2173 non-null   int64  
dtypes: int64(8), object(1)
memory usage: 169.8+ KB

```

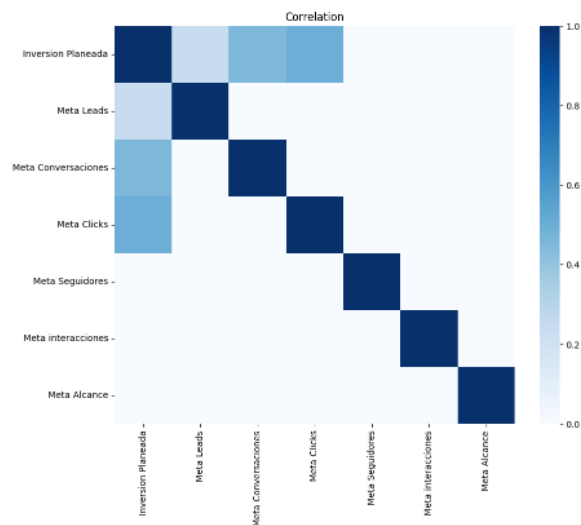


Ilustración 6. información de las variables de input y matriz de correlación entre estas.

9 ESTRATEGIA DE VALIDACIÓN Y SELECCIÓN DE MODELO

Para la selección del modelo, se implementará una estrategia basada en cuatro modelos de clasificación binaria, los cuales serán evaluados exhaustivamente para determinar su rendimiento. A través de pruebas de A/B Testing, se compararán los resultados entre un modelo base y los modelos candidatos, permitiendo identificar la solución más efectiva y robusta para incluirla en el producto final.

En la primera etapa, los datos se dividirán en dos conjuntos principales: un conjunto de entrenamiento que representará el 80% de los datos disponibles, y un conjunto de prueba que abarcará el 20%. Esta división asegura que el modelo pueda ser evaluado en datos no vistos durante el entrenamiento, proporcionando una estimación precisa y confiable de su desempeño en situaciones reales.

Adicionalmente, se verificará que las distribuciones de las variables se mantengan consistentes en los subconjuntos de entrenamiento y prueba en comparación con el conjunto original. Esto garantizará una representación adecuada de las características de los datos y evitará sesgos que puedan comprometer los resultados del modelo.

Una vez entrenados y evaluados, los modelos serán analizados mediante métricas clave como precisión, recall, F1-score y el área bajo la curva ROC (AUC-ROC). Estas métricas no solo permitirán evaluar la capacidad predictiva de los modelos, sino también comparar su eficacia en diferentes contextos. Al finalizar, se generará un reporte detallado que incluirá el rendimiento técnico de cada modelo, junto con los resultados de las pruebas estadísticas realizadas durante el A/B Testing, facilitando la selección del modelo más adecuado para cumplir los objetivos del proyecto.

Por último, podemos observar que la división entre test y validación representan correctamente la distribución del set de datos original:

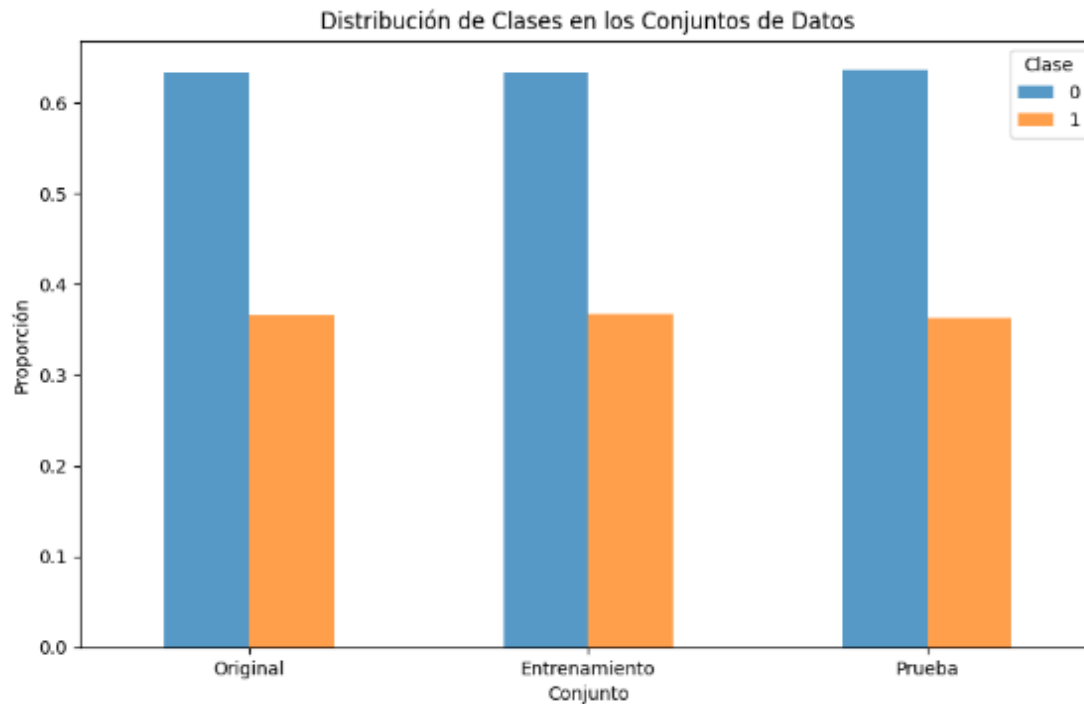


Ilustración 7. distribución de los subconjuntos de datos.

10 CONSTRUCCIÓN Y EVALUACIÓN DEL MODELO

Se establecieron cuatro modelos de clasificación binaria, los cuales son:

- **Regresión logística:** El modelo de Regresión Logística alcanzó un 87.36% de accuracy, destacándose por su capacidad para identificar campañas exitosas con un recall del 99%, lo que significa que detecta casi todas las campañas exitosas correctamente. Sin embargo, su precisión en esta clase es del 74%, lo que indica una mayor propensión a falsos positivos. En contraste, la clase 0 (campañas no exitosas) obtuvo una precisión perfecta del 100%, aunque con un recall del 81%, lo que sugiere que algunas campañas no exitosas son clasificadas erróneamente como exitosas. En general, el modelo muestra un equilibrio razonable entre precisión y recall, con un F1-score ponderado de 88%

Accuracy: 0.8735632183988846

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.81	0.89	277
1	0.74	0.99	0.85	158
accuracy			0.87	435
macro avg	0.87	0.90	0.87	435
weighted avg	0.90	0.87	0.88	435

- **Árbol de Decisión:** Alcanzó un 89.65% de accuracy, mostrando un rendimiento sólido en la clasificación binaria de campañas exitosas y no exitosas. La clase 0 (campañas no exitosas)

obtuvo un desempeño destacado con un F1-score de 92%, mientras que la clase 1 (campañas exitosas) alcanzó un F1-score de 85%.

Accuracy: 0.896551724137931

Classification Report:				
	precision	recall	f1-score	support
0	0.91	0.94	0.92	277
1	0.88	0.83	0.85	158
accuracy			0.90	435
macro avg	0.89	0.88	0.89	435
weighted avg	0.90	0.90	0.90	435

- Random Forest: Alcanzó un 89.65% de accuracy, mostrando un rendimiento sólido en la clasificación binaria de campañas exitosas y no exitosas. La clase 0 (campañas no exitosas) obtuvo un desempeño destacado con un F1-score de 92%, mientras que la clase 1 (campañas exitosas) alcanzó un F1-score de 85%.

Accuracy: 0.896551724137931

Classification Report:				
	precision	recall	f1-score	support
0	0.91	0.94	0.92	277
1	0.88	0.83	0.85	158
accuracy			0.90	435
macro avg	0.89	0.88	0.89	435
weighted avg	0.90	0.90	0.90	435

- Support vector machine: El modelo basado en SVM (Máquinas de Soporte Vectorial) destacó entre las opciones evaluadas al lograr un 88.51% de accuracy y un F1-score ponderado de 89%, demostrando un excelente equilibrio entre precisión y recall. Su alto rendimiento en ambas clases lo convierte en la mejor opción para la tarea de clasificación binaria. La clase 0 (campañas no exitosas) obtuvo una precisión perfecta del 100% y un buen recall del 82%, mientras que la clase 1 (campañas exitosas) sobresalió con un recall del 99%, asegurando que casi todas las campañas exitosas fueran identificadas correctamente. Aunque existe una ligera propensión a falsos positivos, el balance global del modelo, junto con su capacidad para generalizar, lo posiciona como la alternativa más robusta y adecuada para el sistema de clasificación.

Accuracy: 0.8850574712643678

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.82	0.90	277
1	0.76	0.99	0.86	158
accuracy			0.89	435
macro avg	0.88	0.91	0.88	435
weighted avg	0.91	0.89	0.89	435

Todos los modelos obtuvieron resultados satisfactorios en las métricas evaluadas. Para determinar el modelo más adecuado, se llevó a cabo un A/B Testing comparando cada modelo candidato con el modelo SVM, que se estableció como modelo de control. Los resultados del análisis no proporcionaron evidencia suficiente para rechazar la hipótesis nula, indicando que las diferencias entre SVM y los demás modelos no fueron estadísticamente significativas. En consecuencia, se seleccionó el modelo SVM como la mejor opción, dado que demostró ser adecuado para resolver el problema planteado, manteniendo un rendimiento robusto y balanceado en todas las métricas clave.

11 CONSTRUCCIÓN DEL PRODUCTO DE DATOS

El desarrollo del producto de datos está compuesto por dos componentes principales: un **Panel de Control (Dashboard)** y un **Modelo de Predicción**, cada uno diseñado para atender necesidades específicas de análisis y predicción para la empresa VML.

A continuación, se presenta un diagrama ilustrativo que describe cómo estos componentes interactúan dentro de la solución tecnológica:

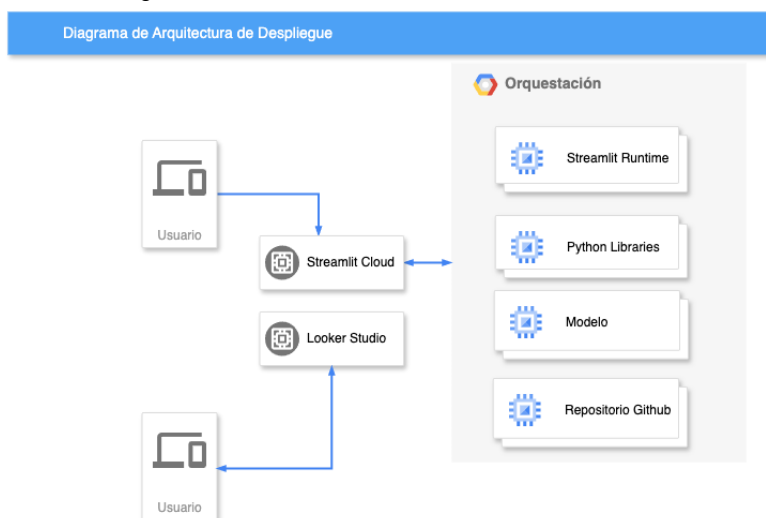


Ilustración 8. Diagrama de Arquitectura de Despliegue

11.1 Panel de Control (Dashboard)

Como se muestra en el diagrama, el **Panel de Control** ha sido desarrollado utilizando **Looker Studio**, y su principal función es consolidar y visualizar los datos provenientes de plataformas publicitarias como **Facebook Ads** y **Google Ads**. Este componente utiliza:

- **Google Sheets** como fuente de datos, alimentada por llamadas **API REST** que consolidan la información publicitaria.
- **Looker Studio** para desplegar gráficos y reportes interactivos, proporcionando una herramienta visual efectiva para analizar métricas clave de las campañas.



Ilustración 9. Panel de Control (Dashboard)



El dashboard permite a los usuarios realizar análisis en tiempo real y tomar decisiones fundamentadas basadas en datos confiables. El dashboard está disponible en el siguiente enlace: https://lookerstudio.google.com/u/0/reporting/9e091957-6c52-4ad3-a98d-b1e4b99c9ca7/page/p_kmxd5gj8ld. Sin embargo, por políticas de la empresa no puede estar público, pero estaremos pendientes de la necesidad de accesos personales.

11.2 Modelo de Predicción

El **Modelo de Predicción**, también representado en el diagrama, ha sido implementado mediante **Streamlit** y **Python**. Su desarrollo e integración se basa en los siguientes pasos:

- **Exportación del mejor modelo de Machine Learning**, optimizado para las necesidades de predicción de campañas publicitarias.
- **Gestión de dependencias**: A través de un repositorio en **GitHub**, se asegura la correcta instalación de librerías necesarias para la ejecución del modelo.
- **Ejecución del modelo**: Utilizando **Streamlit Runtime**, se habilita una interfaz web que permite a los usuarios interactuar fácilmente con el modelo y obtener predicciones en tiempo real.

Predicción de Éxito en Campañas Publicitarias

Por favor, ingresa los valores de las siguientes variables:

Inversión Planeada (en pesos colombianos):

50000,00 - +

Meta Conversaciones:

20 - +

Indica la inversión planeada en pesos colombianos (valor decimal).

Número de conversaciones objetivo generadas por la campaña.

Meta Leads:

50 - +

Meta Seguidores:

200 - +

Número objetivo de leads que esperas alcanzar con la campaña.

Número objetivo de seguidores nuevos obtenidos.

Meta Clicks:

100 - +

Meta Alcance:

1000 - +

Número total de clics que esperas lograr.

Cantidad de usuarios que esperas que la campaña alcance.

Meta interacciones:

20 - +

Selecciona el tipo de campaña:

BRANDING

Número esperado de interacciones en redes sociales.

Selecciona el tipo de campaña que estás planificando.

Predecir

¡La campaña será exitosa! (88.69% de confianza)

Ilustración 10. Modelo de Predicción

Este componente brinda a VML una solución accesible y eficaz para predecir el éxito de las campañas, integrándose de manera armoniosa con el resto de la infraestructura tecnológica. El modelo de predicción está disponible en el siguiente enlace: <https://cdadeployment-97gn69kmzjdhwzuwcbzhg6.streamlit.app/>.

12 RETROALIMENTACIÓN POR PARTE DE LA ORGANIZACIÓN

Al presentar el producto de datos a VML quedaron satisfechos frente a los resultados de visualización de las campañas exitosas, ya que reflejan cómo se comportan de manera diaria en términos de impresiones o alcance (métricas demasiado importantes para medir performance en marketing). Frente al sistema de recomendación basado en el modelamiento de clasificación les llamó mucho la atención en cómo dicha herramienta es útil a la hora de definir las mejores características.

Sin embargo, dentro del producto de datos basado en visualizaciones no hubieron observaciones a mejorar, pero en el sistema de recomendación además de poder experimentar con las características de las campañas para definir si una campaña será exitosa o no, les gustaría poder saber qué se debe mejorar dentro de los inputs del modelo para que dicha campaña aumente la probabilidad de confianza o permita que la campaña tenga un mejor performance, de esta manera la herramienta no solo ayudaría a definir cuáles son las mejores características sino permitiría dar cierto feedback al usuario sobre como realizar unos mejores ajustes a las campañas pautadas. Además, se abrió la discusión acerca de agregar más features o características más finas de las campañas para entender mejor su comportamiento, dicha retroalimentación se tomará en cuenta para fases posteriores.

13 CONCLUSIONES

- ¿Se cumplieron los objetivos del proyecto?

Sí, los objetivos del proyecto se cumplieron exitosamente, ya que se desarrollaron tanto el dashboard en Looker Studio, permitiendo la visualización en tiempo real de KPIs clave como clics, impresiones y costos, y se implementó un modelo de clasificación basado en Machine Learning que determina el éxito de las campañas en función de un umbral definido en conjunto con el negocio, logrando optimizar el análisis y la toma de decisiones. Además de brindar una herramienta de recomendación hacia el cliente que permite la experimentación con las características de las campañas.

- b) ¿Cuáles fueron las mayores dificultades que se obtuvieron durante su desarrollo?

Las principales dificultades encontradas estuvieron relacionadas con la selección de características para entrenar el modelo, debido a la abundancia de variables con información relevante y similar, lo que complicó la identificación de las más significativas. Además, la definición del umbral y las reglas para clasificar una campaña como exitosa o no representó un desafío importante, ya que requería una colaboración activa con el negocio, siendo este un elemento crítico para garantizar la efectividad y precisión del modelo.

- c) ¿Qué estimación se puede dar respecto a cómo se impactarían las métricas de negocio (KPIs) definidas?

Gracias al producto de datos los KPI mencionados al inicio del informe serán afectados de manera positiva ya que mejora el rendimiento de las campañas futuras entendiendo el comportamiento de estas y además entendiendo mejor las características necesarias para que una campaña sea exitosa, para el tercer KPI también tendrá una mejora ya que la tarea manual de visualización se convertiría en un proceso automático y escalable, reduciendo tiempos y costos de mantenimiento, además de plasmar todo el contenido necesario para mejores tomas de decisiones, tal cual como lo confirmó el cliente en la retroalimentación.

- d) ¿Qué condiciones considera que deberían tener los datos para obtener mejores resultados? Más datos, nuevas características, menor sesgo, etc.

Para mejorar los resultados del modelo, se requiere más cantidad de datos representativos. Además, incorporar otras características relevantes y garantizar que los datos estén libres de sesgos permitirá al modelo capturar mejor la complejidad del problema, mejorando su capacidad predictiva y robustez.

- e) ¿El mejor modelo obtenido es suficiente para dar solución al problema?

Sí, el modelo obtenido logró realizar predicciones con métricas destacadas, lo cual es muy positivo. Tras consultar con el negocio, se verificó que las predicciones generadas están alineadas con la realidad operativa y estratégica, ofreciendo un alto valor agregado. Esto permite que el modelo no solo sea técnicamente sólido, sino también relevante para generar impacto tangible para el cliente.

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.