

# Análisis de Campañas de Marketing

📅 Start	@1 de julio de 2024
📅 End	@5 de julio de 2024
🕒 Last edited time	@6 de septiembre de 2024 18:49

## Task

## Proyecto



Detalles del Proyecto

## Objetivos para desarrollar el proyecto

### Objetivo:

Desarrollar un proyecto de Ingeniería de Datos enfocado en el análisis detallado de campañas de marketing digital. Este proyecto permitirá mejorar las habilidades en la metodología ELT, utilizando herramientas como DBT, BigQuery, y APIs de marketing, para extraer, transformar y analizar datos de múltiples campañas digitales. El objetivo es proporcionar insights sobre el rendimiento de las campañas, ayudando a optimizar las estrategias de marketing.

## Diseño de la arquitectura

- Bigquery
- Python para consumo de API
- SQL
- DBT
- Looker

### ▼ Selección de las herramientas y tecnologías adecuadas

La elección de las herramientas y tecnologías adecuadas es crucial para el éxito de cualquier proyecto de ingeniería de datos. Las herramientas y

tecnologías deben ser:

- **Capaces de cumplir con los requisitos del proyecto.**
- **Fáciles de usar y mantener.**
- **Escalables para adaptarse al crecimiento de los datos.**
- **Compatibles con los sistemas existentes.**
- **Rentables y que se ajusten al presupuesto del proyecto.**

#### **Factores a considerar al seleccionar herramientas y tecnologías:**

- **Tipo de datos:** ¿Qué tipo de datos se manejarán en el proyecto?
- **Volumen de datos:** ¿Qué volumen de datos se espera manejar?
- **Casos de uso:** ¿Cuáles son los casos de uso específicos del proyecto?
- **Habilidades del equipo:** ¿Qué habilidades y experiencia tiene el equipo de ingeniería de datos?
- **Presupuesto:** ¿Cuál es el presupuesto disponible para herramientas y tecnologías?

#### **▼ Diseño de la arquitectura de datos**

La arquitectura de datos define la estructura general de cómo se almacenarán, procesarán y analizarán los datos en el proyecto. Una arquitectura de datos bien diseñada debe ser:

- **Modular:** La arquitectura debe dividirse en módulos independientes que puedan ser desarrollados y mantenidos por separado.
- **Escalable:** La arquitectura debe ser capaz de adaptarse al crecimiento de los datos sin afectar el rendimiento.
- **Segura:** La arquitectura debe proteger los datos contra accesos no autorizados y ataques cibernéticos.
- **Eficiente:** La arquitectura debe utilizar los recursos de manera eficiente para optimizar el rendimiento y los costos.

#### **Componentes clave de una arquitectura de datos:**

- **Capa de ingesta:** Esta capa se encarga de recopilar y cargar los datos en el sistema.

- **Capa de transformación:** Esta capa se encarga de limpiar, transformar y preparar los datos para su análisis.
- **Capa de almacenamiento:** Esta capa se encarga de almacenar los datos de manera segura y eficiente.
- **Capa de análisis:** Esta capa se encarga de ejecutar análisis y modelos de aprendizaje automático sobre los datos.
- **Capa de visualización:** Esta capa se encarga de comunicar los resultados del análisis a las partes interesadas.

#### **Ejemplos de arquitecturas de datos comunes:**

- **Lago de datos:** Un lago de datos es un repositorio central que almacena todos los datos en su formato bruto.
- **Almacén de datos:** Un almacén de datos es un repositorio estructurado que almacena datos limpios y preparados para su análisis.
- **Data Mesh:** Data Mesh es una arquitectura distribuida que desglosa la responsabilidad de los datos en dominios específicos.





---

# Discovery de Fuentes

## | Investigación de fuentes

### Ejemplo de contenido de base de datos:

- **Campañas Google Ads:**
  - ID de campaña, Nombre de campaña, Impresiones, Clics, Conversiones, Costo por clic (CPC), Costo total, Fecha de inicio, Fecha de fin.
- **Campañas Facebook Ads:**
  - ID de campaña, Nombre de campaña, Alcance, Interacciones, Clics, Costo por clic (CPC), Costo total, ROI, Fecha de inicio, Fecha de fin.
- **Campañas Instagram Ads:**
  - ID de campaña, Nombre de campaña, Impresiones, Clics, Me gusta, Comentarios, Conversiones, Costo total, Fecha de inicio, Fecha de fin.
- **Métricas agregadas:**
  - ID de campaña, Canal (Google, Facebook, Instagram, Twitter), ROI total, Tasa de conversión, CPA (Costo por adquisición), Costo por canal, Resultados por canal.

### ▼ Búsqueda de fuentes de datos

El primer paso en el discovery de fuentes es identificar todas las posibles fuentes de datos que podrían ser relevantes para el proyecto. Algunas de las fuentes de datos más comunes incluyen:

- **Datos internos:** Estos datos se generan dentro de la organización, como registros de ventas, datos de clientes o datos de producción.
- **Datos externos:** Estos datos se generan fuera de la organización, como datos gubernamentales, datos de investigación o datos de redes sociales.
- **Datos abiertos:** Estos datos son gratuitos y de libre acceso para cualquier persona, como conjuntos de datos de Kaggle o datos del Banco Mundial.

### Herramientas para la búsqueda de fuentes de datos:

- **Motores de búsqueda:** Google, Bing y DuckDuckGo pueden ser útiles para encontrar fuentes de datos en línea.
- **Repositorios de datos:** Existen repositorios en línea que albergan datasets de diversas fuentes, como Kaggle y Data.gov.
- **Sitios web gubernamentales:** Muchos gobiernos publican datos abiertos sobre diversos temas.
- **Organizaciones sin fines de lucro:** Algunas organizaciones sin fines de lucro recopilan y comparten datos sobre temas de interés público.
- **Empresas privadas:** Algunas empresas ofrecen acceso a sus datos a través de API o suscripciones.

#### ▼ **Identificación de fuentes de datos relevantes**

Una vez que hayan identificado una lista de posibles fuentes de datos, es necesario evaluar cada una de ellas para determinar si son relevantes para su proyecto. Algunos de los factores a considerar incluyen:

- **Temática:** ¿Los datos se ajustan al tema del proyecto?
- **Calidad:** ¿Los datos son confiables, completos y precisos?
- **Disponibilidad:** ¿Los datos están disponibles en un formato accesible?
- **Frecuencia de actualización:** ¿Con qué frecuencia se actualizan los datos?
- **Costo:** ¿Existe algún costo asociado con la obtención de los datos?

#### ▼ **Selección de fuentes de datos**

Después de evaluar las fuentes de datos relevantes, es necesario seleccionar las que mejor se ajusten a las necesidades del proyecto. Algunos de los criterios de selección incluyen:

- **Relevancia:** ¿Los datos son esenciales para el proyecto?
- **Calidad:** ¿Los datos son de alta calidad y confiables?
- **Disponibilidad:** ¿Los datos están fácilmente disponibles y accesibles?
- **Frecuencia de actualización:** ¿Los datos se actualizan con la frecuencia suficiente?
- **Costo:** ¿Los datos son asequibles dentro del presupuesto del proyecto?

### Find Open Datasets and Machine Learning Projects | Kaggle

Download Open Datasets on 1000s of Projects + Share Projects on One Platform. Explore Popular Topics Like Government, Sports, Medicine, Fintech, Food, More. Flexible Data Ingestion.

 <https://www.kaggle.com/datasets>

### Mockaroo - Random Data Generator and API Mocking Tool | JSON / CSV / SQL / Excel

A free test data generator and API mocking tool - Mockaroo lets you create custom CSV, JSON, SQL, and Excel datasets to test and demo your software.

 <https://www.mockaroo.com/>

## Modelado(Opcional)





---

## Desarrollo

*Desglose del proyecto en tareas a realizar con documentación sobre funcionalidad*

### ▼ Implementación del código y las herramientas

La etapa de desarrollo implica la implementación del código y las herramientas necesarias para transformar, analizar y visualizar los datos. Esto puede incluir:

- **Limpieza y preparación de datos:** Eliminar valores atípicos, manejar valores faltantes y normalizar los datos.
  - **Transformación de datos:** Agregar nuevas características, combinar conjuntos de datos y crear variables derivadas.
  - **Modelado de datos:** Desarrollar modelos de bases de datos
  - **Visualización de datos:** Crear gráficos, cuadros y paneles para comunicar los resultados del análisis.
- Python
  - Pandas
  - Request
  - GCP

### ▼ Herramientas comunes para el desarrollo de ingeniería de datos:

- **Lenguajes de programación:** Python, R, Scala, Java
- **Frameworks de big data:** Apache Spark, Hadoop, Flink
- **Bibliotecas de aprendizaje automático:** scikit-learn, TensorFlow, PyTorch
- **Herramientas de visualización de datos:** Tableau, Looker Studio, Metabase

### ▼ Buenas prácticas para el desarrollo de ingeniería de datos:

- **Utilizar un control de versiones:** Git
- **Escribir código limpio y documentado:** PEP8, Google Java Style Guide

- **Implementar orquestacion:** Github Actions, Apache Airflow

Aa Name	👤 Assign	⚙️ Status
<u>Facebook Ads Extract &amp; Load</u>		Done
<u>Google Ads Extract &amp; Load</u>		Done
<u>Twitter Ads Extract &amp; Load</u>		Done

## Evaluación de resultados

### Evaluación del cumplimiento de los objetivos iniciales

Es importante evaluar si el proyecto ha cumplido con los objetivos iniciales que se definieron en la etapa de diseño. Esto implica:

- **Revisar los objetivos iniciales:** Recordar los objetivos específicos, medibles, alcanzables, relevantes y con un plazo definido (SMART) que se establecieron al inicio del proyecto.
- **Comparar los objetivos con los resultados:** Comparar los resultados obtenidos con los objetivos iniciales y determinar si se han alcanzado.
- **Identificar las áreas de éxito y fracaso:** Identificar las áreas del proyecto que tuvieron éxito y las que no cumplieron con las expectativas.
- **Aprender de las lecciones aprendidas:** Reflexionar sobre las lecciones aprendidas durante el desarrollo del proyecto para mejorar en futuros proyectos.

### Documentación del diseño de la solución

Es importante documentar el diseño de la solución de ingeniería de datos, incluyendo:

- **Arquitectura de datos:** Describir la arquitectura de datos del proyecto, incluyendo los componentes, las relaciones entre ellos y el flujo de datos.
- **Tecnologías utilizadas:** Documentar las tecnologías y herramientas específicas que se utilizaron en el proyecto.

- **Justificación de decisiones:** Explicar las decisiones de diseño que se tomaron en el proyecto y la justificación detrás de ellas.

La documentación del diseño de la solución es importante para:

- **Facilitar la comprensión del proyecto:** Permitir que otras personas comprendan cómo funciona la solución y cómo se implementó.
- **Facilitar el mantenimiento del proyecto:** Facilitar la actualización y el mantenimiento de la solución en el futuro.
- **Compartir el conocimiento:** Compartir el conocimiento y las mejores prácticas con otros equipos de ingeniería de datos.

---

## Task

- ✓ ~~Análisis de los resultados obtenidos en el proyecto.~~
- ✓ ~~Evaluación del cumplimiento de los objetivos iniciales del proyecto.~~
- ✓ ~~Documentación del diseño de la solución, incluyendo la arquitectura, tecnologías utilizadas y justificación de decisiones.~~
- ✓ ~~Creación de un prototipo o demo que muestre la funcionalidad del proyecto y el flujo de datos.~~
- ✓ ~~Preparación de una presentación clara y concisa para comunicar los resultados del proyecto a audiencias técnicas y no técnicas.~~

## Documentacion