

# Propuesta de Trabajo y análisis preliminar de los datos

E-commerce Olist

Por G9C



## Entendimiento de la situación actual

En Brasil, uno de los mercados más grandes de la economía mundial, el e-commerce ha ganado protagonismo. Su potencial es enorme, esta industria viene creciendo exponencialmente pues mejoran constantemente la rapidez, la seguridad y la eficiencia tanto para comprar como para vender.

Olist es de las principales empresas de e-commerce en Brasil que ayuda a distintos comercios (incluyendo startups) a vender sus productos a nivel local, logrando así captar más terreno en el mercado del país que empresas conocidas como Amazon y Alibaba.

A través de un minucioso trabajo de análisis basado en la ciencia de datos, y junto a un equipo capacitado para este proyecto, cumpliremos los objetivos planteados más adelante, para mejorar su imagen corporativa y que continúen ganando terreno en el mercado e-commerce brasileño.

## Equipo de Trabajo:

- Helio Forero (Data Engineer).
- Martín Eluney Gómez Piñeiro (Data Engineer).
- Ramón Blanc (Data Scientist).
- Blas Anastasio Collias (Data Analyst).
- Facundo Paccioretti (Data Analyst).

## Objetivo General:

Crear planes de acción para los clientes de Olist que les permitan tener mayor impacto en el mercado, a su vez mejorando la imagen corporativa de Olist, creando una mayor exposición de los servicios de Olist para capturar nuevos clientes.

## Objetivos Específicos:

- Comprender detalladamente el estado actual de la operación de Olist.
- Desarrollar KPIs (Key Performance Indicator) que muestren el estado actual de la operación de Olist.
- Crear planes de acción que permitan a los clientes de Olist aumentar sus ventas.
- Crear planes de acción que permitan a los clientes de Olist aumentar el alcance de su oferta.

- Evaluar la efectividad de las campañas de marketing, basados en su conversión de clientes.

## **Alcance:**

Las visualizaciones(Dashboards) entregadas al cliente podrán ser actualizadas por el mismo al incluir más datos recientes. Además los KPIs, visualizaciones y demás se actualizan acorde a esta información, mas no serán modificados a otra información o metas por parte del contratista.

Los planes de acción son resultado del análisis del estado actual de Olist y el contexto actual en el que se encuentra. Estos planes de acción pueden no ser efectivos en otro contexto o en un futuro y tanto su ejecución como los resultados son responsabilidad completa de Olist.

Los análisis entregados reflejan los datos suministrados y el contexto en el que se tienen, estos no se actualizarán con más datos una vez se dé por iniciado el proyecto.

La evaluación de las campañas de marketing se realizará con los datos proporcionados y resume la efectividad de conversión a clientes de las campañas acorde a los clientes actuales. Este análisis de las campañas varía en el tiempo y sus resultados no necesariamente se van a mantener prolongadamente en el tiempo.

Para tener un continuo análisis y generación de planes de acción, evaluación de la implementación y resultados de dichos planes, humildemente sugerimos contratar un plan recurrente con nuestra compañía.

## **KPIs y su relación con el objetivo:**

Los KPIs que se determinaron útiles para alcanzar los objetivos son los siguientes:

1. Variación ventas.
2. Variación Cantidad de productos.
3. Variación Tiempos de Entrega.
4. Variación en promedio reviews.
5. Proporción del costo de envío en el precio total.
6. Porcentaje Conversión de Leads.

Estos KPIs fueron escogidos específicamente para cumplir los objetivos de la siguiente forma:

**Variación de ventas:** Principalmente para dar un vistazo general al crecimiento en el volumen de ventas a lo largo del tiempo. Este KPI es fundamental para entender el panorama actual de cómo opera Olist. (primer y segundo objetivo)

**Variación Cantidad de productos:** Nos enseña la diferencia entre las cantidades de productos solicitadas por los clientes en cada venta. Es útil para tener un mejor control del stock.

**Variación Tiempos de Entrega:** El objetivo de este KPI es medir la variación del tiempo en que se tarda en entregar un producto desde el momento en que se aprueba la venta. Se relaciona con el objetivo de realizar planes de acción para los clientes, mejorando así las reviews y logrando captar más posibles vendedores.

**Variación en promedio reviews:** Este KPI se relaciona con los objetivos de planes de acción para los clientes de Olist. Nos permite identificar productos que están siendo bien recibidos, también se podría considerar bajo el filtro de tiempo de entrega (Otro KPI) para entender el impacto del tiempo de entrega en la satisfacción del cliente.

**Proporción del costo de envío en el precio total:** Este KPI es útil para conocer qué porcentaje del precio de una venta está dispuesto a pagar un cliente de forma adicional para adquirir un producto. Este KPI se relaciona con los objetivos de planes de acción para los clientes de Olist.

**Porcentaje Conversión de Leads:** Este KPI se relaciona con el objetivo de evaluar la efectividad de las campañas de marketing. Permite entender cuáles han sido más exitosas y puede verse tanto en cantidad de clientes como en volumen de ventas por cliente.

Estos mismos KPIs pueden ser filtrados por región, escala temporal, proveedor o segmento del producto para analizar soluciones a problemáticas específicas y determinar el objetivo para cada KPI o para la segmentación específica del KPI (Ejemplo: Objetivo de ventas totales vs Objetivo de ventas por región)

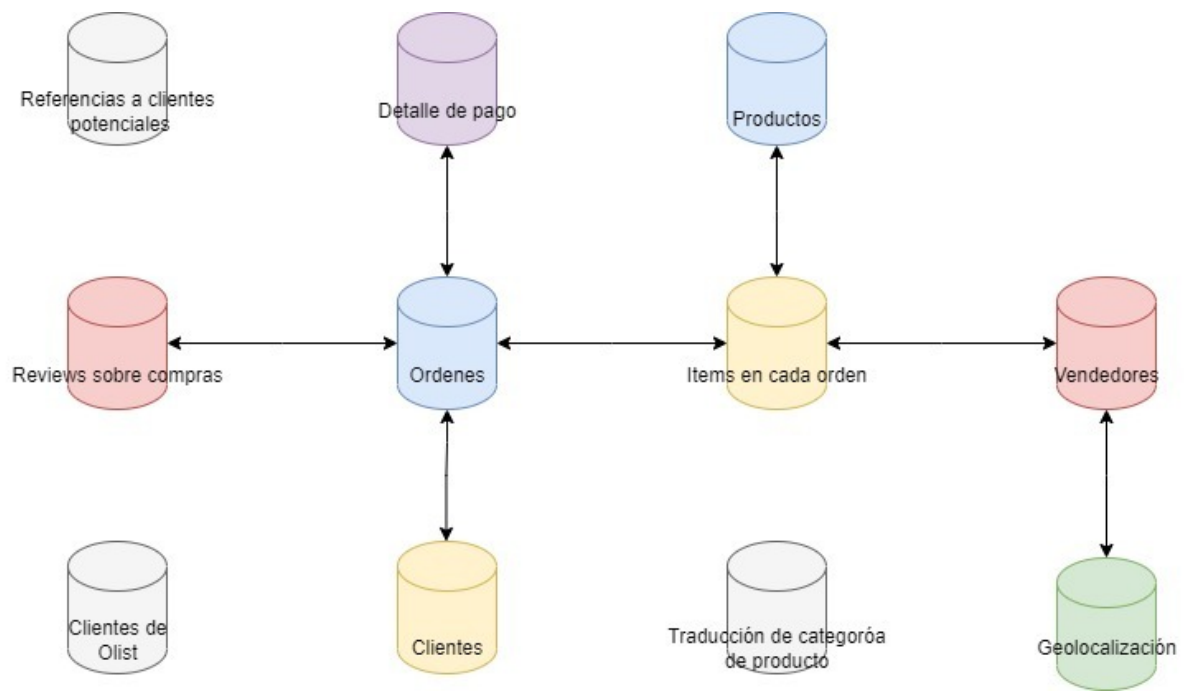
## Entregables:

A la conclusión del proyecto, estarán disponibles para Olist los siguientes entregables:

- Acceso y uso del Data Warehouse, el modelo de machine learning y las visualizaciones(dashboards).
- Análisis de la situación actual de Olist.
- Planes de acción para los clientes de Olist.
- Evaluación de las campañas de marketing.

## Análisis preliminar de calidad de datos:

Para este análisis utilizaremos el dataset suministrado por la empresa brasileira de e-commerce 'Olist', el cual contiene 11 tablas (archivos .csv), con información de pedidos (+100k registros), clientes, datos de geolocalización y otros, de la empresa entre los años 2016 y 2018.



Todo el análisis preliminar se encuentra en detalle documentado en el archivo nombrado 'Anexo: Análisis preliminar de calidad de datos'.

### Resumen por tabla:

- **olist\_closed\_deals\_dataset:** esta tabla cuenta con 14 columnas y 842 registros, está dedicada a llevar a cabo una lista de clientes de Olist

#### observaciones importantes:

- En la columna 'lead\_behaviour\_profile' hay algunos datos combinados como por ejemplo: eagle, wolf
- en las columnas 'has\_company', 'has\_gtin', 'average\_stock' y 'business\_type', se observa aproximadamente un 90% de nulos
- en la columna 'declared\_monthly\_revenue' cuenta con 797='0.0', lo cual no aporta mucha información
- **olist\_customers\_dataset:** Esta tabla cuenta con 5 columnas y 99441 registros, corresponde a un listado de clientes.

#### Observaciones importantes:

- En la tabla customer\_zip\_code\_prefix, un código postal en Brasil tiene el siguiente formato 12345 - 678, donde los primeros 5 números corresponden al prefijo y los últimos 3 al sufijo. Dentro de la tabla el sufijo no se encuentra presente
- **olist\_geolocation\_dataset:** Esta tabla contiene 5 columnas y 1000163 registros, contiene los datos de geolocalización, por ciudad, prefijo de código postal y estado.

**Observaciones importantes:**

- La cantidad de registros es muy superior a la del resto de las tablas.
  - en la columna `geolocation_city` : se observa el doble de valores únicos con respecto a la columna `'customer_city'` de la tabla anterior.
- **olist\_marketing\_qualified\_leads\_dataset:** Esta tabla cuenta con 4 columnas y 8000 registros, hace referencia a los potenciales clientes/sellers.

**Observaciones importantes:**

- En la tabla `'origin'`: se desconoce el detalle de algunas denominaciones
- **olist\_order\_items\_dataset:** Esta tabla cuenta con 7 columnas y 112650 registros, representa un listado de los Ítems en cada orden.

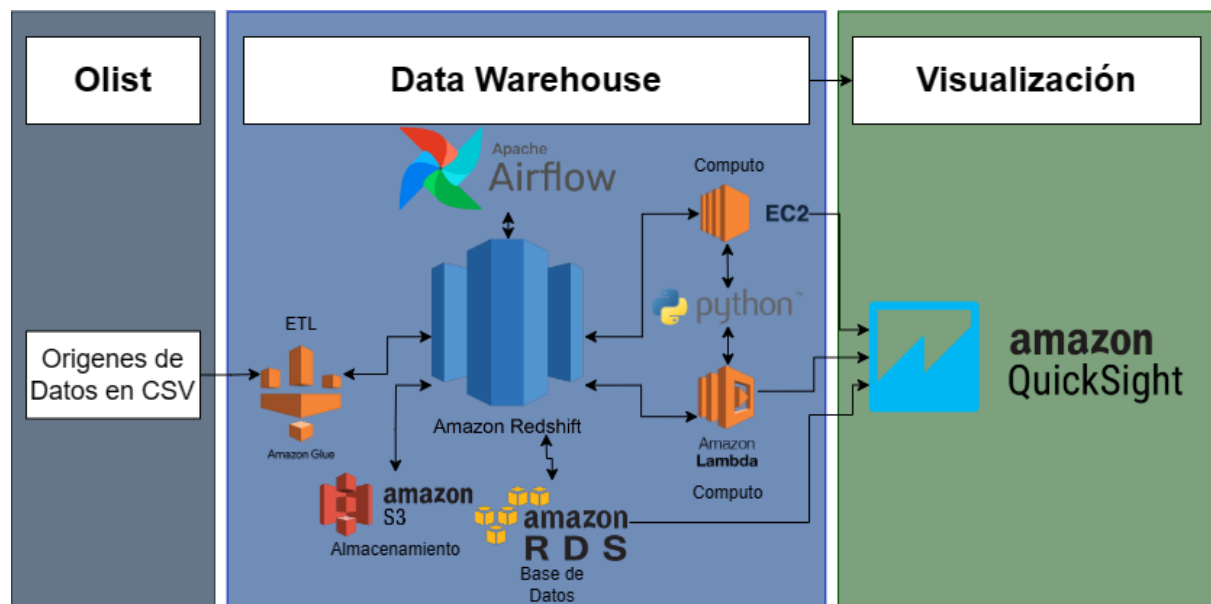
**Observaciones importantes:**

- en la tabla `'shipping limit date'`: La fecha más reciente está fuera de rango con respecto a la fecha máxima de otras tablas.
  - en la tabla `'price'`: el valor mínimo observado es de 0.85 Reales, deberemos ver a qué producto se refiere este precio para ver si es congruente.
- **olist\_order\_payments\_dataset:** Esta tabla cuenta con 5 columnas y 103886 registros, representa los detalles sobre los pagos.
  - **olist\_order\_reviews\_dataset:** el tamaño es de 99224 registros y 7 columnas. Trata de las opiniones/Reviews sobre compras.

**Observaciones importantes:**

- En las tablas `'review_comment_title'` y `'review_comment_message'` : se observa una gran cantidad de valores nulos
- **olist\_order\_dataset:** Esta tabla contiene 99441 registros y 8 columnas, y trata sobre las órdenes de compra. .
  - **olist\_products\_datasets:** . La tabla tiene 32951 registros y 9 columnas, y contiene un listado de los productos..
  - **olist\_sellers\_dataset:** se compone de 3095 registros y 4 columnas, que tratan sobre un listado de vendedores y no contienen nulos.
  - **product\_category\_name\_translation:** Traduce del portugués al inglés las categorías de los productos. Tabla simple, de 71 registros y 2 columnas

**Stack Tecnológico:**

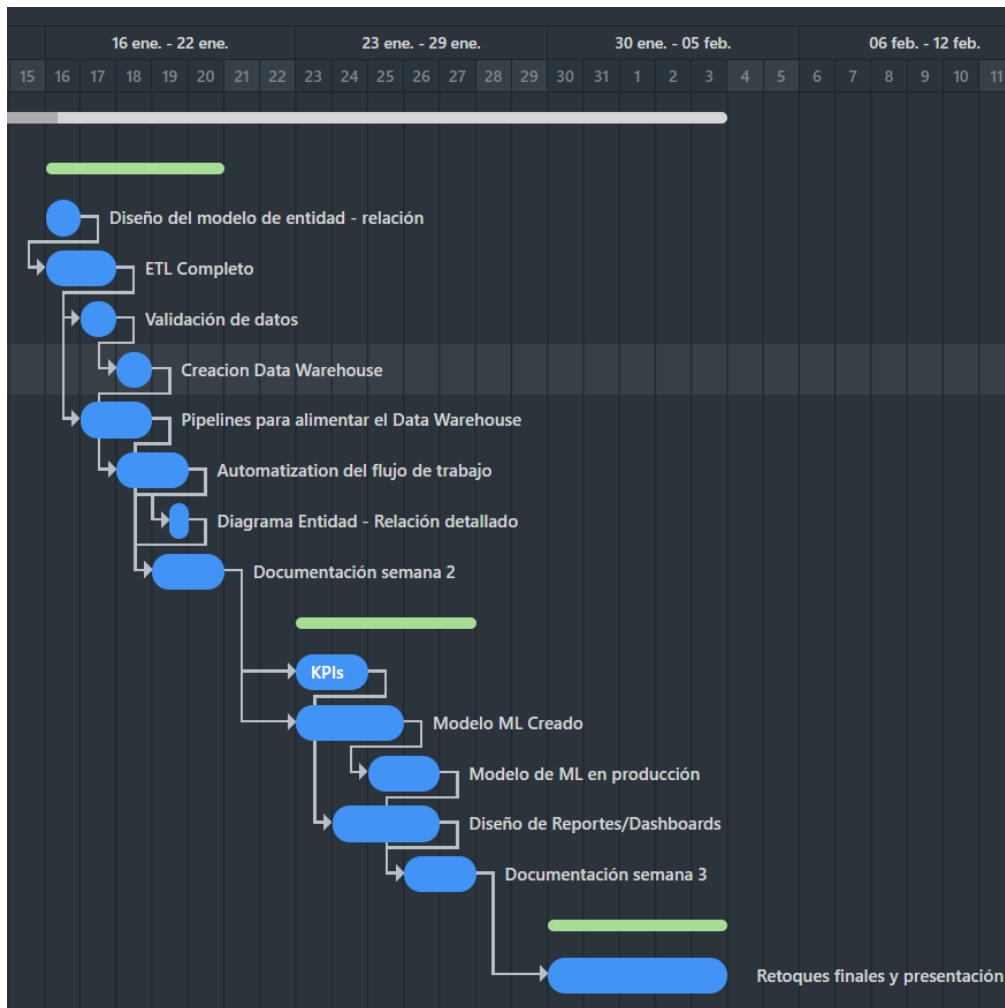


Para tener una disponibilización web de los resultados, escogimos Amazon Web Services como nube de desarrollo para el proyecto y específicamente se utilizarán los siguientes componentes y plugins, cada uno con su respectivo rol.

- Amazon Glue: ETL (Extraction, Transformation, and Load)
- Amazon S3: Almacenamiento General
- Amazon RDS: Bases de datos (RDS: Relational Database Service)
- Amazon Lambda: Resuelve código y su uso es por ejecución.
- Amazon EC2: Crea una instancia de cómputo y su uso es por hora.
- Python: Se utilizará para computó especialmente en Machine Learning
- Amazon Redshift: Utilidad para crear y administrar Data Warehouse
  
- Airflow: Automatización del flujo de trabajo. Permite que automáticamente resultados de una parte del proceso sean usados por otra parte cuando son creados nuevamente.
- Amazon QuickSight: Herramienta de visualización y Business Intelligence.

## Duración y programación del proyecto

A continuación se visualiza un diagrama de Gantt generado en Click Up, la aplicación que escogimos para el seguimiento del proyecto. Se indican las tareas, sus dependencias y el tiempo estimado para completarlas.



El proyecto inicia ejecución el lunes 16 de enero y se completarán los entregables el 27 de enero. Estos entregables serán socializados y entregados al cliente a más tardar el 3 de febrero.