IT BROTHERS – Consultora especialista en ciencia de datos

INTEGRANTES DEL STAFF:

JIMÉNEZ, JOSE
ALONSO, LUCILA
ARGUMEDO, HECTOR
TALAVERA, RICARDO
UMBERT, NORBERTO

ÍNDICE

INTRODUCCION

La empresa brasilera Olist nos ha encargado el trabajo de analizar su negocio a través de los datos que fueron recolectando durante los años 2016 a 2018, donde nos solicita que seamos capaces de entregarle una solución innovadora para que sus usuarios puedan vender una mayor cantidad de productos. Asimismo, Olist manifiesta que existe un gran número de quejas de los sellers (vendedores), respecto a ordenes no completadas debido a la diferencia entre el stock real y productos publicados.

Para lo solicitado anteriormente, se elaborará un estudio extensivo del negocio, a fin de poder ofrecer los controles, alertas, predicciones y sugerencias con los datos obtenidos, que permitan a Olist tomar las mejores decisiones y tener una estrategia acorde a su modelo de negocio.

SITUACIÓN ACTUAL

EL MERCADO

En 2021, la venta minorista de productos a través de e-commerce significó un saldo aproximado de 5.2 trillones de dólares en todo el mundo. A raíz de la pandemia, las compras en línea experimentaron un ascenso; tan solo en 2021, el comercio electrónico tuvo un crecimiento del 27% respecto al año anterior y todo esto ha animado a las pymes a incursionar en el comercio digital. Se infiere que la venta minorista de productos a través de e-commerce aumentará un 56% en los próximos años, llegando a los 8.1 trillones en 2026.

Las plataformas de e-commerce actualmente permiten a las PYME llegar a un mayor número de clientes y vender de manera más efectiva. Este tipo de beneficio ha hecho que el 70 % de las pymes que venden online escojan hacerlo por medio de marketplaces y los e-retailers.

LA EMPRESA: OLIST

Olist es una empresa brasileña fundada en el año 2015 con sus oficinas principales en Curitiba. Es una compañía prestadora de servicio e-commerce para PYMES que funciona como un marketplace, es decir, funciona como “tienda de tiendas” donde diferentes vendedores pueden ofrecer sus productos a consumidores finales.

Actualmente la empresa logra generar alrededor de 1MM de visitas (sumadas las visitas en desktop y móviles), las cuales tienen una duración promedio cercano a los 3 minutos. Asimismo, Olist presenta un promedio de 3.2 páginas visitadas por internauta, así como una tasa de rebote (porcentaje medio de visitantes que ven una sola página antes de irse del sitio web) cercano al 61%.

Actualmente la empresa recibe el mayor tráfico de Brasil y a través principalmente de las siguientes redes: WhatsApp Web, Facebook, Youtube y LinkedIn, como se puede apreciar a continuación:

Olist actualmente compite con: americanasmarketplace.com.br (2.1 Millones de visitas al mes) y con melhorenvio.com.br (2.2 Millones de visitas al mes)., Ideris.com.br apenas llega al medio millón de visitas al mes.

Sitio web	Afinidad	Visitas mensuales	Categoría
americanasmarketp...	100%	2.1M	eCommerce y compras > Otros eCommerce y compras
melhorenvio.com.br	82%	2.2M	eCommerce y compras > Otros eCommerce y compras
ideris.com.br	66%	461.0K	eCommerce y compras > Otros eCommerce y compras

OBJETIVOS

PRIMARIO*:*

- ✓ Crear valor a Olist brindando información relevante a través de los datos proporcionados, para la toma de decisiones, en cuanto a la mejora del negocio del e-commerce, evaluando a través de los datos, distintas áreas de la empresa, por ejemplo : ventas, registros de compras, lugares donde se llevan a cabo las compras entre otros.

SECUNDARIOS:

- ✓ Elaborar un informe entendible para el cliente a través del uso de métricas y KPIs.
- ✓ Confeccionar datos que se ajusten a la realidad de la empresa.
- ✓ Confeccionar un Data Lake o Data Warehouse que nos permita manejar los datos de forma transparente y correcta.
- ✓ Determinar mediante modelos predictivos sucesos que beneficien a la empresa.

ALCANCES Y LIMITACIONES

El alcance del proyecto es poder brindar a la empresa una herramienta con la cual se sientan confiados y seguros al momento de tomar decisiones. Brindar un informe donde se puede visualizar y comprender lo ocurrido durante el periodo analizado, para entender el momento presente del negocio y sobre todo poder planificar el futuro de la misma mediante los datos obtenidos. Asimismo, la data está limitada a los años 2016 y 2018.

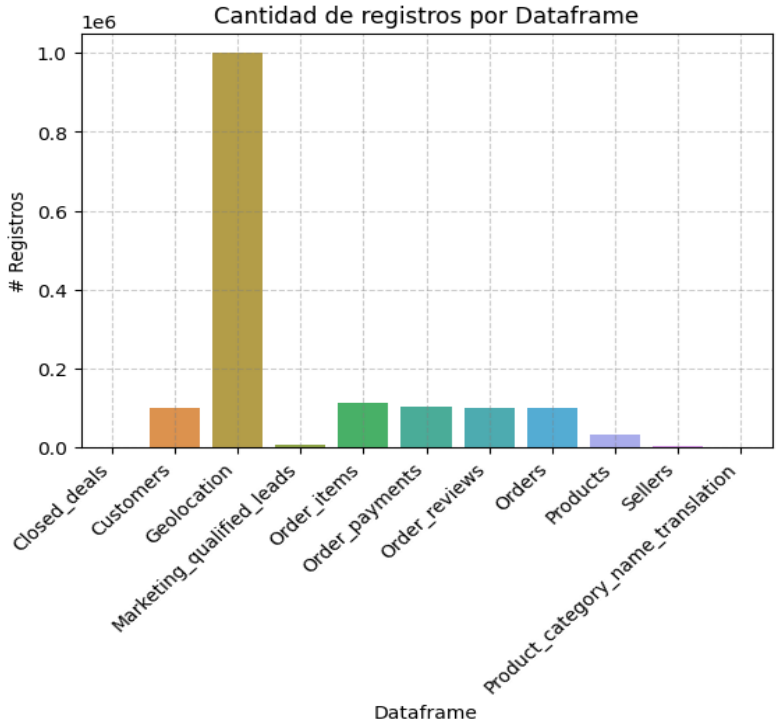
ANALISIS PRELIMINAR DE LOS DATOS (EDA)

Introducción

Para el proyecto se utilizaron como fuente de información 11 tablas provistas por el cliente con información de 2016 a 2018.

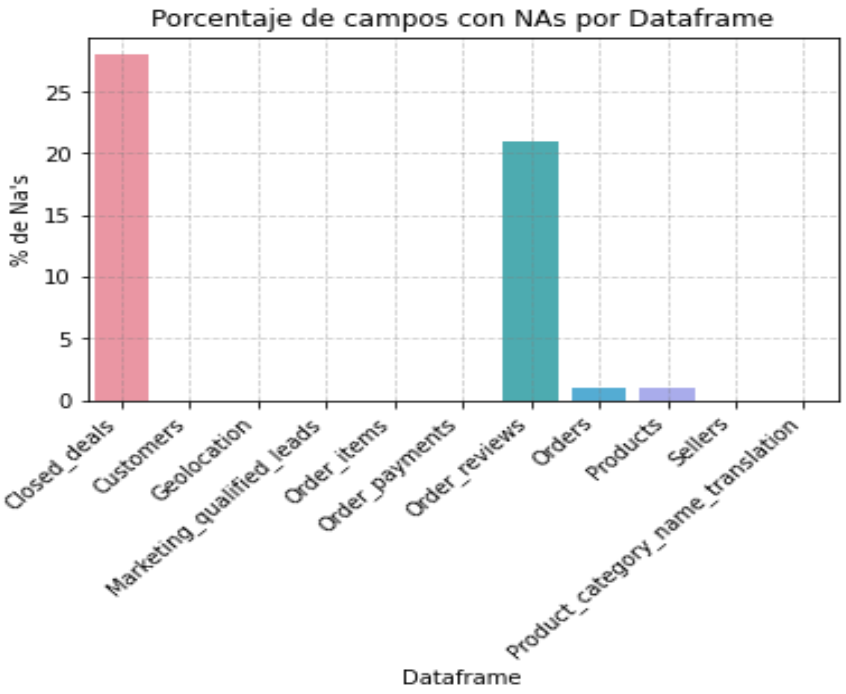
Estas tablas son variadas, conteniendo datos de distintos aspectos del negocio, como datos de ventas, vendedores, clientes, ubicación georeferenciada, etc.

El siguiente gráfico, muestra la cantidad de registros con los que cuenta cada dataset.



Del gráfico, se puede ver de manera fácil y rápida que la tabla con mayor cantidad de registros es “Geolocation” con 1 millón de registros aproximadamente. En segunda instancia, hay varias tablas con un valor cercano a los 100 mil registros. Finalmente, existe otro conjunto con una cantidad de datos menor a los 10 mil registros.

A continuación, se presenta cada tabla con su porcentaje de datos NaN.



La tabla “Closed_Deals” tiene un porcentaje mayor al 25% de valores NaN, y le sigue “Order_reviews” con un porcentaje mayor al 20%.

Esta información es útil para revisar en los próximos pasos la necesidad de estos campos, y cómo puede afectar potencialmente la calidad de un análisis futuro.

Análisis de las tablas

A continuación, analizaremos en detalle cada tabla, con su contenido y tipo de información.

- **Closed Deals (olist_closed_deals_dataset.csv)**

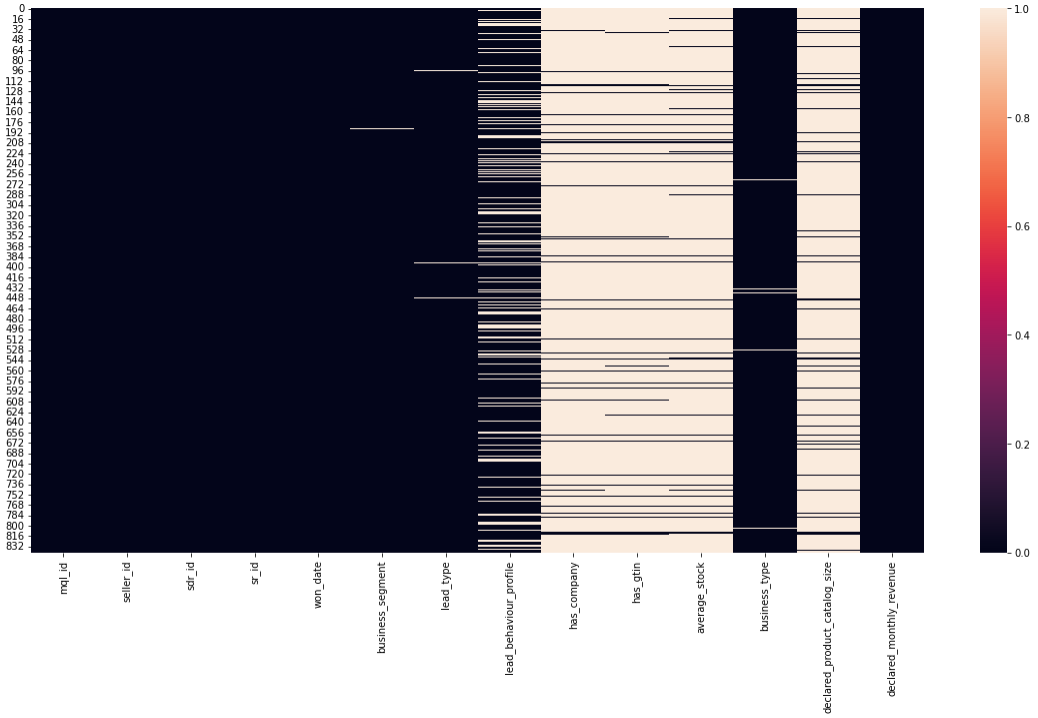
En la tabla “Closed Deals” se puede observar que se tiene información relacionada a los vendedores y datos del negocio en el que se sitúan. Tabla de 842 registros con 14 columnas.

RangeIndex: 842 entries, 0 to 841
Data columns (total 14 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	mql_id	842 non-null	object	Identificador de leads de marketing
1	seller_id	842 non-null	object	Código de identificación del vendedor
2	sdr_id	842 non-null	object	Sin interpretación clara.
3	sr_id	842 non-null	object	Sin interpretación clara.
4	won_date	842 non-null	object	Fecha de incorporación de vendedor
5	business_segment	841 non-null	object	Segmento de negocio
6	lead_type	836 non-null	object	Canal de venta
7	lead_behaviour_profile	665 non-null	object	Perfil de comprador buscado
8	has_company	63 non-null	object	Sin interpretación clara.
9	has_gtin	64 non-null	object	Sin interpretación clara.
10	average_stock	66 non-null	object	Stock promedio
11	business_type	832 non-null	object	Tipo de negocio
12	declared_product_catalog_size	69 non-null	float64	Tamaño de catálogo de producto declarado
13	declared_monthly_revenue	842 non-null	float64	Facturación mensual declarada

dtypes: float64(2), object(12)

Mapa de calor de los datos NaN.



De las 14 columnas se aprecia que 5 columnas cuentan con bastantes valores NaN, en donde 4 de ellas tiene menos del 1% utilizable, y la columna restante un 79,45%. A estas columnas, se debe incorporar la columna “declared_monthly_revenue” que tiene el 95,33% con valor cero (0).

Hay 2 columnas que no se entiende bien su finalidad, ya que refieren a un código particular que no se repite en otras columnas: “sdr_id”,y “sd_id”.

En conclusión, de las 14 columnas se tiene 4 columnas que no se pueden utilizar por gran cantidad de valores NaN, 2 columnas no se entiende su uso, y una columna se puede usar de forma parcial ya que cuenta con un 79,45% de datos cargados de forma correcta.

- **Customers ('olist_customers_dataset.csv')**

La tabla Customers hace referencia a información de los clientes y su ubicación geográfica.

Tabla de 5 columnas y 99.441 registros.

RangeIndex: 99441 entries, 0 to 99440
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	customer_id	99441 non-null	object	Código de identificación de cliente
1	customer_unique_id	99441 non-null	object	Código único de identificación de cliente
2	customer_zip_code_prefix	99441 non-null	int64	Código postal del cliente
3	customer_city	99441 non-null	object	Ciudad del cliente
4	customer_state	99441 non-null	object	Estado del cliente

dtypes: int64(1), object(4)

Mapa de calor de los valores NaN

No se observan valores NaN, ni particularidades de esta tabla. Es una tabla para ser utilizada en todo su espectro de ser necesaria.

- **Geolocation ('olist_geolocation_dataset.csv')**

En esta base se tiene la geo-posición de gran cantidad de barrios, con su código postal incluido.

Tabla de 5 columnas y 1.000.163 de registros.

RangeIndex: 1000163 entries, 0 to 1000162
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	geolocation_zip_code_prefix	1000163 non-null	int64	Código postal de la ubicación
1	geolocation_lat	1000163 non-null	float64	Latitud (coordenadas)
2	geolocation_lng	1000163 non-null	float64	Longitud (coordenadas)
3	geolocation_city	1000163 non-null	object	Ciudad
4	geolocation_state	1000163 non-null	object	Estado

dtypes: float64(2), int64(1), object(2)

Mapa de calor de los valores NaN:

No se observan datos faltantes, aunque si tiene una gran cantidad de valores duplicados. Estos alcanzan un total de 261.831 registros duplicados. De ser removidos, la tabla quedaría con un total de 738.332 registros, 73,82% del total actual.

- **Marketing Qualified Leads ('olist_marketing_qualified_leads_dataset.csv')**

Cuenta con información relacionada al origen de cada lead generado, su origen, y la fecha de primer contacto.

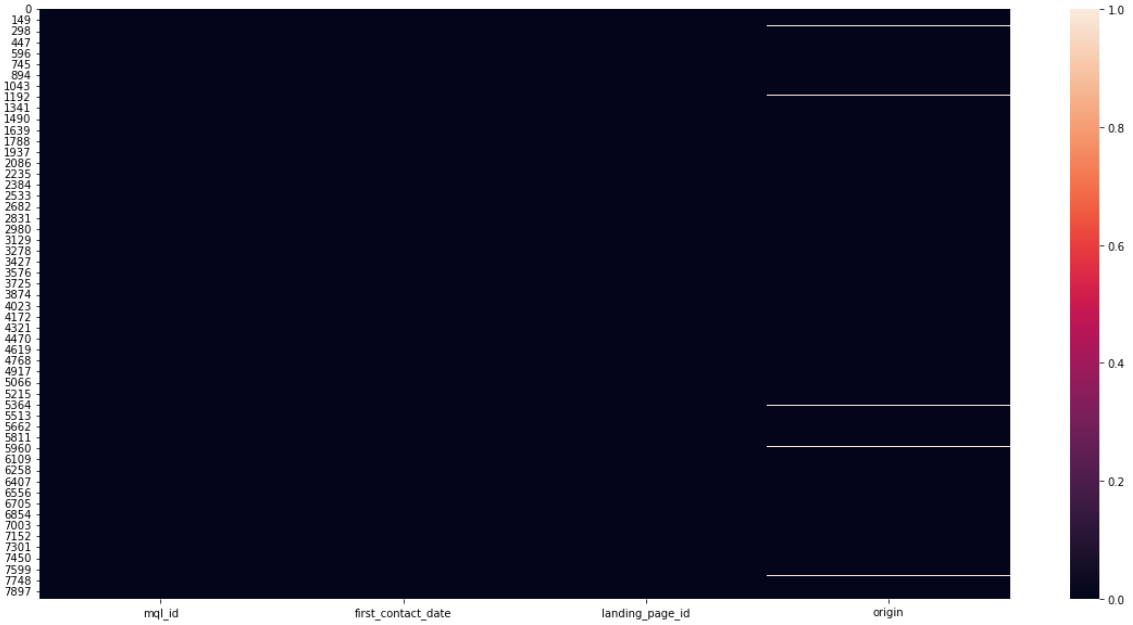
Tabla de 4 columnas y 7.999 registros.

RangeIndex: 8000 entries, 0 to 7999
Data columns (total 4 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	mql_id	8000 non-null	object	Código identificador de marketing leads
1	first_contact_date	8000 non-null	datetime64[ns]	Primer día de contacto
2	landing_page_id	8000 non-null	object	Identificador de página de landing de mkt
3	origin	7940 non-null	object	Origen del lead de marketing

dtypes: datetime64[ns](1), object(3)

Mapa de calor de los valores NaN



Tiene un buen nivel de información. Sin embargo, vale aclarar que la columna “origin” tiene 60 valores NaN y 1.099 valores calificados como “unknown”. Esto implica que esta columna tiene el 86,26% de valores utilizables.

- **Order Items ('olist_order_items_dataset.csv')**

Este dataset cuenta con información relacionada a las órdenes de compra, los productos involucrados, datos del vendedor, precios, y costos de envío.

Tabla de 7 columnas y 112.650 registros

RangeIndex: 112650 entries, 0 to 112649
Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	order_id	112650 non-null	object	Código de identificación de orden
1	order_item_id	112650 non-null	int64	Número de item por orden
2	product_id	112650 non-null	object	Código de identificación de producto
3	seller_id	112650 non-null	object	Código de identificación del vendedor
4	shipping_limit_date	112650 non-null	datetime64[ns]	Día límite de envío
5	price	112650 non-null	float64	Precio
6	freight_value	112650 non-null	float64	Costo de flete

dtypes: datetime64[ns](1), float64(2), int64(1), object(3)

Mapa de calor de valores NaN:

Buen nivel de información. No se observan particularidades para destacar.

- **Order payments ('olist_order_payments_dataset.csv')**

Esta tabla cuenta con datos relacionados a las órdenes de compra, y distintos aspectos de las formas de pago.

Tabla de 5 columnas y 103.886 registros.

RangeIndex: 103886 entries, 0 to 103885
Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	order_id	103886 non-null	object	Identificador de orden
1	payment_sequential	103886 non-null	int64	Número de secuencia de pago
2	payment_type	103886 non-null	object	Forma de pago
3	payment_installments	103886 non-null	int64	Cantidad de cuotas
4	payment_value	103886 non-null	float64	Valor de pago

dtypes: float64(1), int64(2), object(2)

Mapa de calor de los valores NaN:

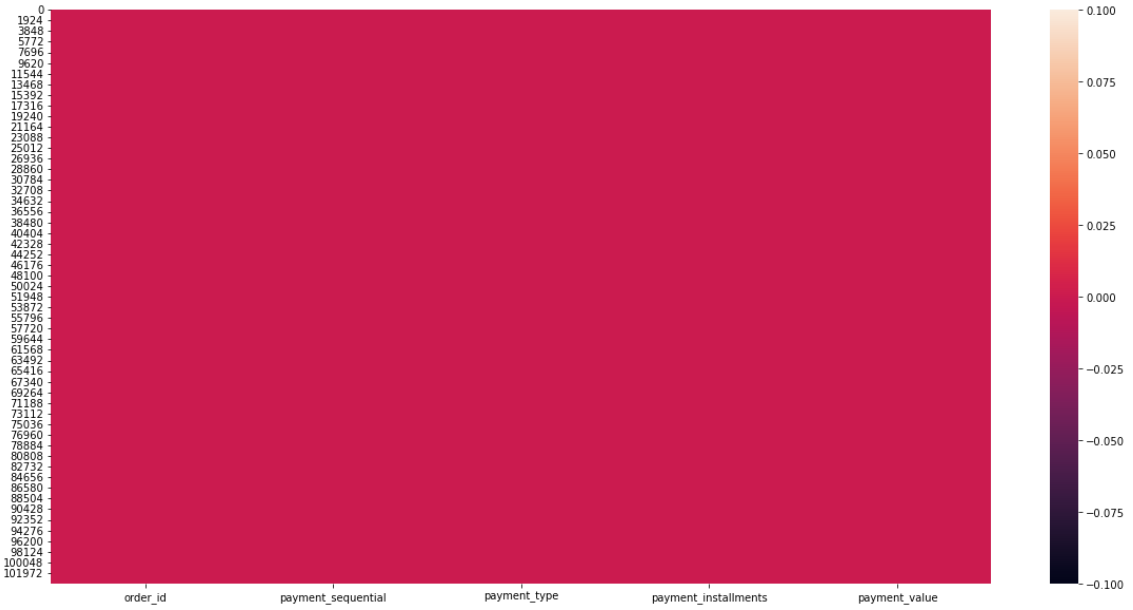


Tabla con datos completos y listos para su uso. No se observan puntos negativos a resaltar.

- **Order Reviews ('olist_order_reviews_dataset.csv')**

La siguiente base de datos tiene información relacionada con la devolución de los clientes y su satisfacción con la experiencia de compra.

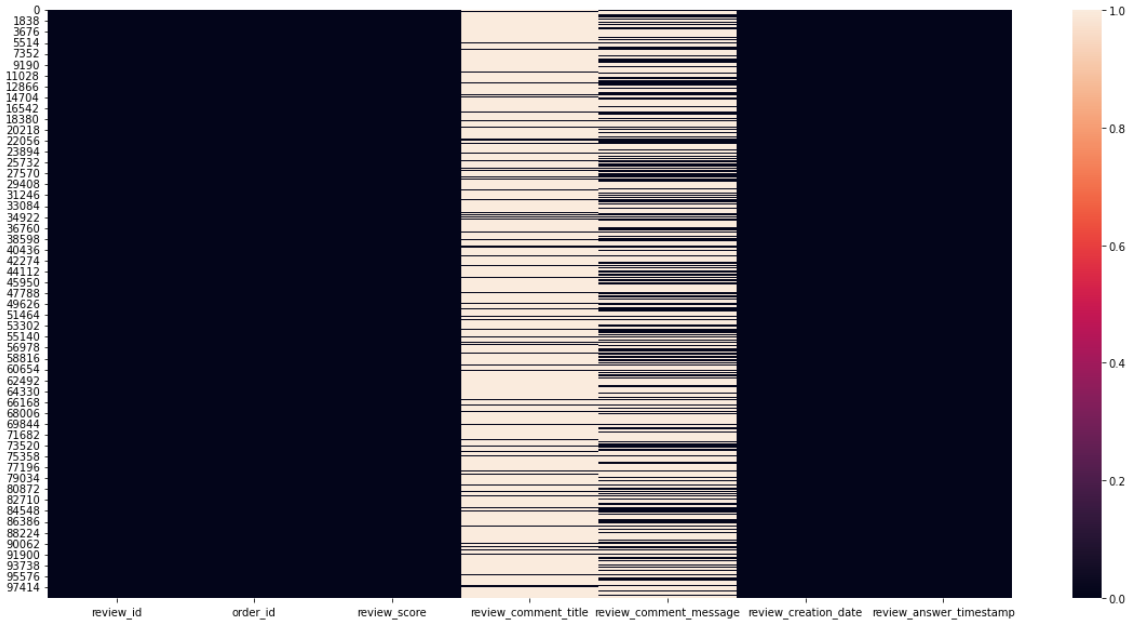
Tabla de 7 columnas y 99.224 registros.

RangeIndex: 99224 entries, 0 to 99223
Data columns (total 7 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	review_id	99224 non-null	object	Código de indentificación del review
1	order_id	99224 non-null	object	Código de indentificación de orden
2	review_score	99224 non-null	int64	Valorización del review
3	review_comment_title	11568 non-null	object	Título del comentario del review
4	review_comment_message	40977 non-null	object	Comentario del review
5	review_creation_date	99224 non-null	datetime64[ns]	Día de generación del review
6	review_answer_timestamp	99224 non-null	datetime64[ns]	Respuesta al review

dtypes: int64(1), object(4), datetime64[ns](2)

Mapa de calor de los valores NaN:



De los registros de las devoluciones, se tiene 2 columnas con mayor cantidad de datos faltantes: “review_comment_title” y “review_comment_message” con un 88,34% y 58,70% de valores faltantes respectivamente. Como positivo el campo “review_score” tiene los datos completos y una buena distribución de sus opciones, haciendo una primera suposición que es un campo que puede resultar de gran utilidad.

- **Orders ('olist_orders_dataset.csv')**

La tabla “Orders” provee información de las fechas involucradas en el proceso de compra como la fecha de compra, aprobación, entrega al despachante, entrega al cliente, y fecha estimada de entrega.

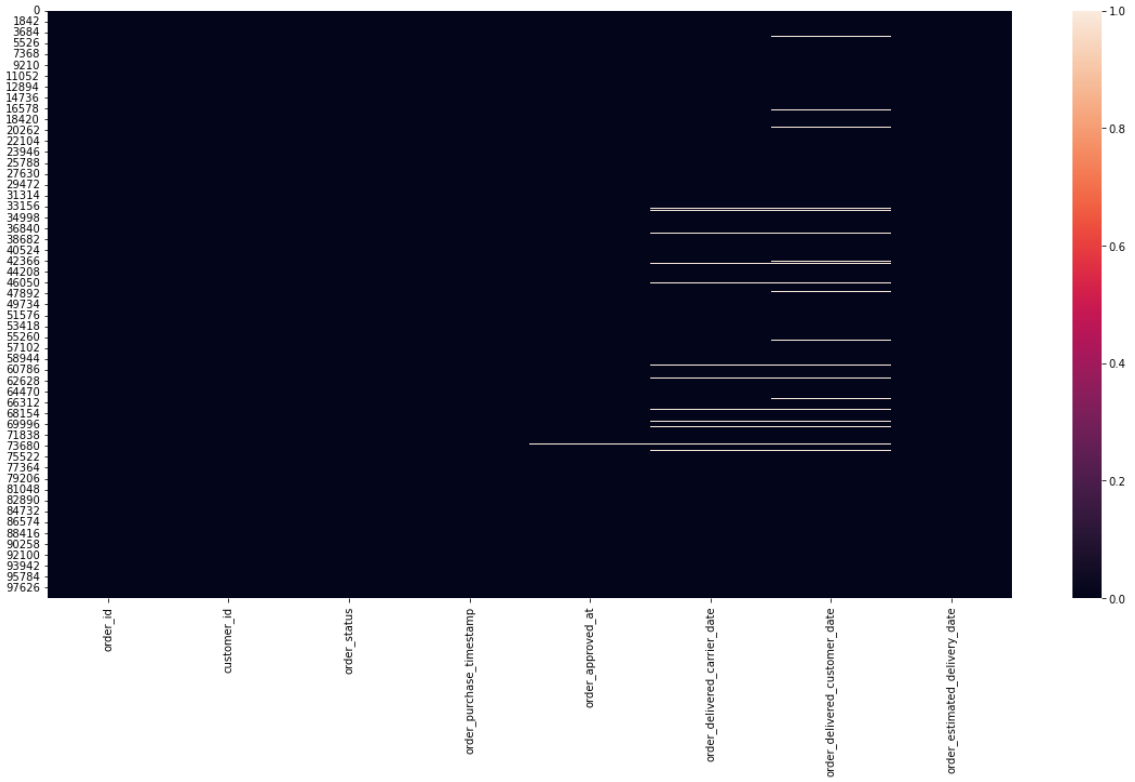
Tabla de 8 columnas y 99.441 registros

RangeIndex: 99441 entries, 0 to 99440
Data columns (total 8 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	order_id	99441 non-null	object	Código de identificación de orden
1	customer_id	99441 non-null	object	Código de identificación de cliente
2	order_status	99441 non-null	object	Estado de la orden
3	order_purchase_timestamp	99441 non-null	datetime64[ns]	Fecha de compra
4	order_approved_at	99281 non-null	datetime64[ns]	Fecha de aprobación de transacción
5	order_delivered_carrier_date	97658 non-null	datetime64[ns]	Fecha de entrega al carrier de transporte
6	order_delivered_customer_date	96476 non-null	datetime64[ns]	Fecha de entrega al cliente
7	order_estimated_delivery_date	99441 non-null	datetime64[ns]	Fecha estimada de entrega

dtypes: datetime64[ns](5), object(3)

Mapa de calor de los valores NaN:



Si bien se observan algunos valores NaN en 3 columnas, el porcentaje de datos completos es de 97% para “order_approved_at”, 98,2% para “order_delivered_carrier_date” y 99,84% para “order_delivered_customer_date”. Es un alto grado de datos completos para poder hacer un análisis adecuado, en caso de ser necesario.

- **Products ('olist_products_dataset.csv')**

En este dataset, se tiene información acerca de las características del producto, de su publicación, y características físicas como peso y tamaño.

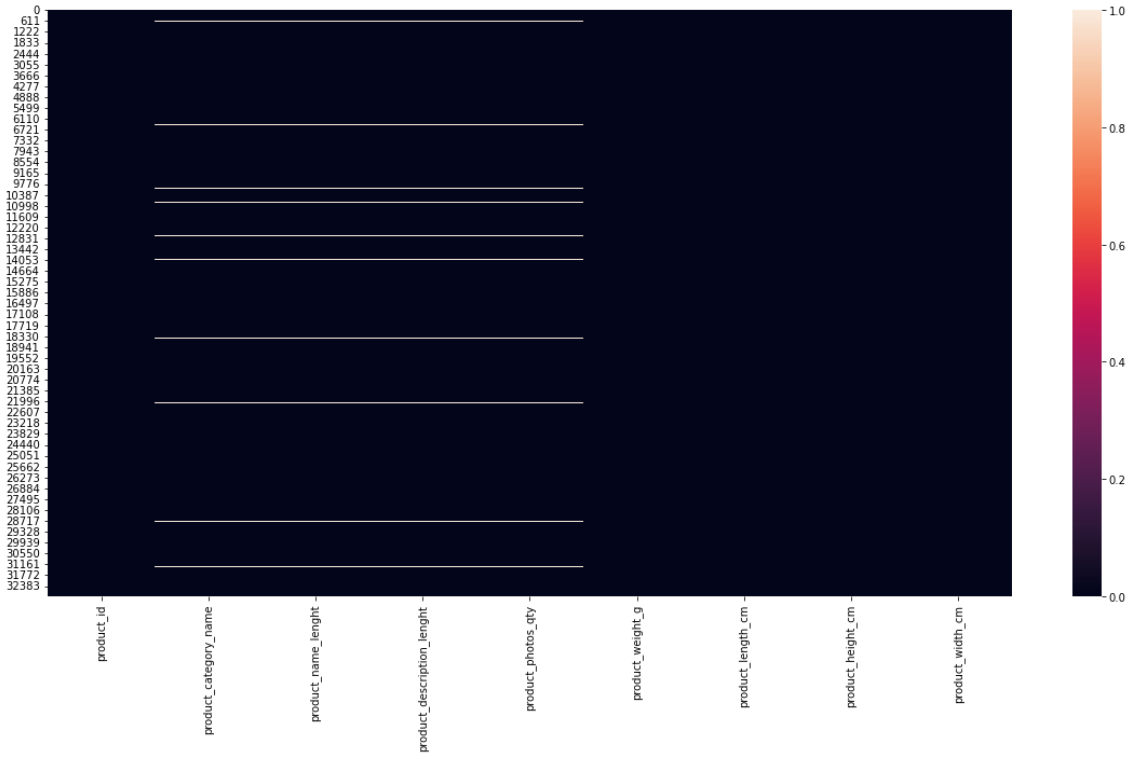
Tabla de 9 columnas y 32.951 registros.

RangeIndex: 32951 entries, 0 to 32950
Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	product_id	32951 non-null	object	Código de identificación de producto
1	product_category_name	32341 non-null	object	Categoría de producto
2	product_name_lenght	32341 non-null	float64	Extensión de nombre de producto
3	product_description_lenght	32341 non-null	float64	Extensión de descripción de producto
4	product_photos_qty	32341 non-null	float64	Cantidad de fotos de productos
5	product_weight_g	32949 non-null	float64	Peso del producto
6	product_length_cm	32949 non-null	float64	Longitud del producto
7	product_height_cm	32949 non-null	float64	Altura del producto
8	product_width_cm	32949 non-null	float64	Ancho del producto

dtypes: float64(7), object(2)

Mapa de calor de los valores NaN:



Excepto la columna “product_id” todas las demás cuentan con valores NaN. Tiene un alto nivel de valores completos de forma correcta, superando el 98%. Es una tabla que puede resultar de utilidad en caso de ser necesaria.

- **Sellers ('olist_sellers_dataset.csv')**

En esta base, se encuentran los datos del vendedor y su ubicación geográfica.

Tabla de 4 columnas y 3094 registros.

RangeIndex: 3095 entries, 0 to 3094 Data columns (total 4 columns):				
#	Column	Non-Null Count	Dtype	Descripción
0	seller_id	3095 non-null	object	Código de identificación del vendedor
1	seller_zip_code_prefix	3095 non-null	int64	Código postal del vendedor
2	seller_city	3095 non-null	object	Ciudad del vendedor
3	seller_state	3095 non-null	object	Estado del vendedor
dtypes: int64(1), object(3)				

Mapa de calor de los valores NaN:

Se observa una buena calidad de la información.

- **Product Category Name Translation ('product_category_name_translation.csv')**

Es una tabla de conversión de nombres de categorías de productos del portugués al inglés.

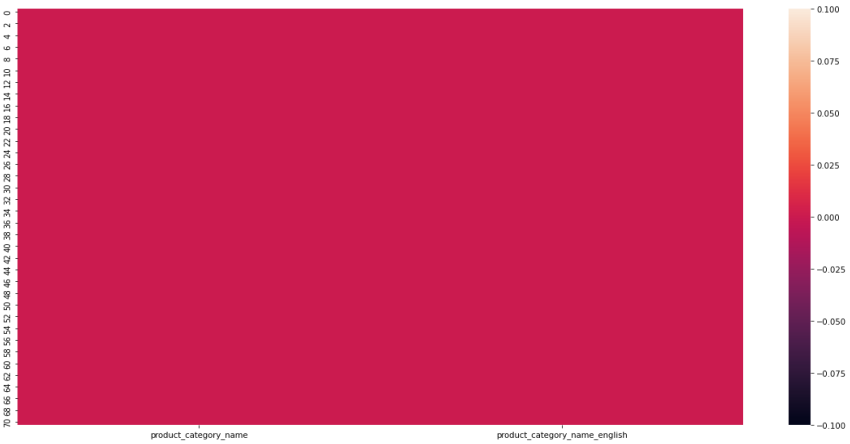
Tabla de 2 columnas y 70 registros.

RangeIndex: 71 entries, 0 to 70
Data columns (total 2 columns):

#	Column	Non-Null Count	Dtype	Descripción
0	product_category_name	71 non-null	object	Nombre de la categoría del producto
1	product_category_name_english	71 non-null	object	Traducción al inglés de la categoría

dtypes: object(2)

Mapa de calor de los valores NaN:



Se presenta como una tabla con buena calidad de información.

Conclusión

Al iniciar este análisis se observaron una gran cantidad de valores faltantes en dos tablas en particular. Sin embargo, analizando su utilidad y relacionamiento con otras bases, se infiere que no son críticas para el análisis a desarrollar y, por ende, no afectan el estudio.

Realizando algunas transformaciones a las tablas, se puede avanzar con el estudio de interés.

De cambiarse el foco del análisis de estudio, se debe analizar nuevamente estos campos y su criticidad para desarrollar un nuevo proyecto.

KPIS Y MÉTRICAS ASOCIADOS (PLANTEO)

Durante la comprensión de los datos, observamos la oportunidad de poder plantear ciertas métricas para poder entender de qué manera se está desarrollando el negocio en ciertas áreas, y así poder elevar algún informe al respecto para su posible mejora.

Las métricas y KPIs como candidatos iniciales a definir son:

- Aumento y disminución de ventas por los drivers de: región, tipo de cliente, canal, proporcionalidad del envío respecto al precio. Asimismo, se plantean las frecuencias tanto anual como trimestral.

- El comportamiento de las ventas anualmente o trimestral, por región abre interrogantes : ¿Tiene fotos?, ¿A qué categoría pertenecen los productos?

- Observar los ingresos y su evolución por año, ciudad, región, etc.

- Porcentaje de comentarios positivos y/o negativos, por vendedor y/o producto.

- Valor promedio de las ventas totales y por driver.

- Tiempo de respuesta del vendedor hacia el cliente.

- Tipo de producto que más se vende, con respecto a sus tamaños, tipo de negocio, etc.

- Con respecto a los vendedores. ¿Dónde se vende más? ¿Dónde se vende menos? ¿Por qué? ¿Qué podemos hacer con respecto a eso?

- Qué se observa respecto al tipo de pago y qué se puede plantear con respecto a eso. Por ejemplo, financiamientos, descuentos en efectivo, etc.

- Observación de la tardanza en las entregas y ver si es necesario una mejora.

- Promedio de un número de ventas de un mismo comprador a un mismo vendedor.

SOLUCION PROPUESTA

Teniendo en cuenta que la empresa es la encargada de tomar la última decisión en administrar sus recursos, se proponen los siguientes pasos para optimizar el negocio según lo solicitado por Olist:

PASO 1:

Elaborar un dashboard de inteligencia de negocios, conteniendo los KPIs y métricas que informen de las tendencias más relevantes para el negocio, de manera de tomar decisiones de forma rápida y acertada.

PASO 2:

Elaborar un modelo de análisis de sentimiento que permita tener conocimiento del posicionamiento de la marca en las redes.

PASO 3:

Definir los drivers donde se puede mejorar venta y buscar asociaciones que expliquen los resultados.

PASO 4:

Generar un plan de acción sugerido de acuerdo a las interpretaciones del estudio, con medidas estratégicas y/o operativas, según sea necesario.

METODOLOGIA PROPUESTA

La metodología de trabajo a adoptar será Scrum, la cual consiste en llevar a cabo un conjunto de tareas de forma regular con el objetivo primordial de trabajar de forma conjunta.

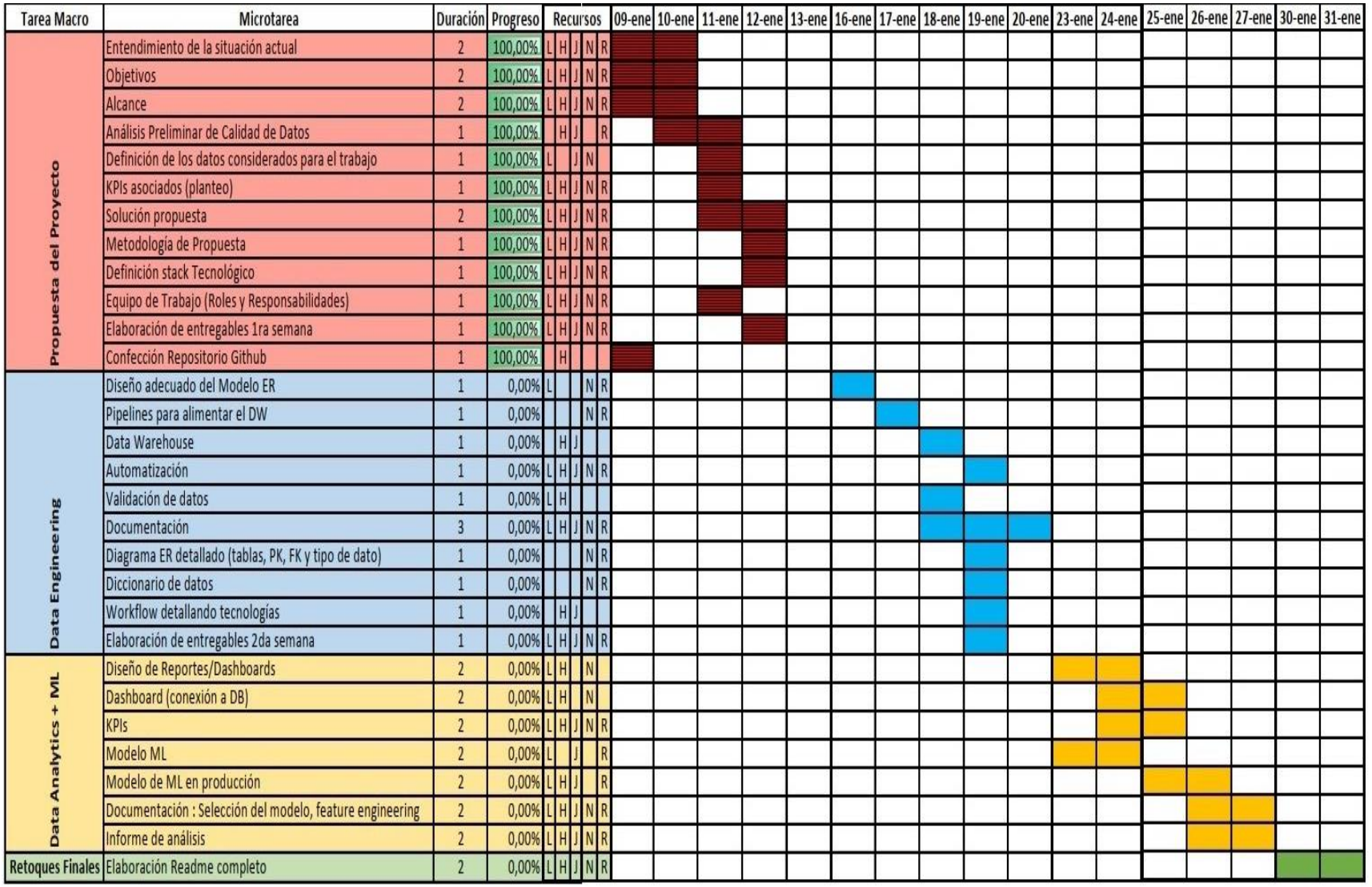
Se tendrán dos reuniones en equipo durante el día, la primera durante el horario de la mañana y la segunda durante la noche. En la primera reunión se evaluarán las tareas a desarrollar en el proyecto de forma diaria, y en la segunda se evaluará el progreso obtenido durante la jornada. Cada reunión tiene la primicia de 15 minutos de duración, la cual podrá extenderse a no más de 30 minutos cada una. Se respetará lo pautado, sin embargo, habrá lugar para reuniones de emergencia que solucionen problemas puntuales surgidos durante evolución del proyecto.

También se coordinará diariamente con el Scrum Máster, para poder tener coordinaciones de avance y novedades posibles de incorporar al proyecto.

De igual forma se tendrá una reunión cada viernes con el Product Owner, con el fin de presentar el trabajo realizado hasta el momento, y así tener un feedback del mismo.

Finalmente, como guía diaria se adjunta la carta Gantt para organizar, planificar y controlar las tareas del proyecto, el cual permite hacer un correcto uso del tiempo disponible en cada tarea.

A continuación el diagrama de Gantt:

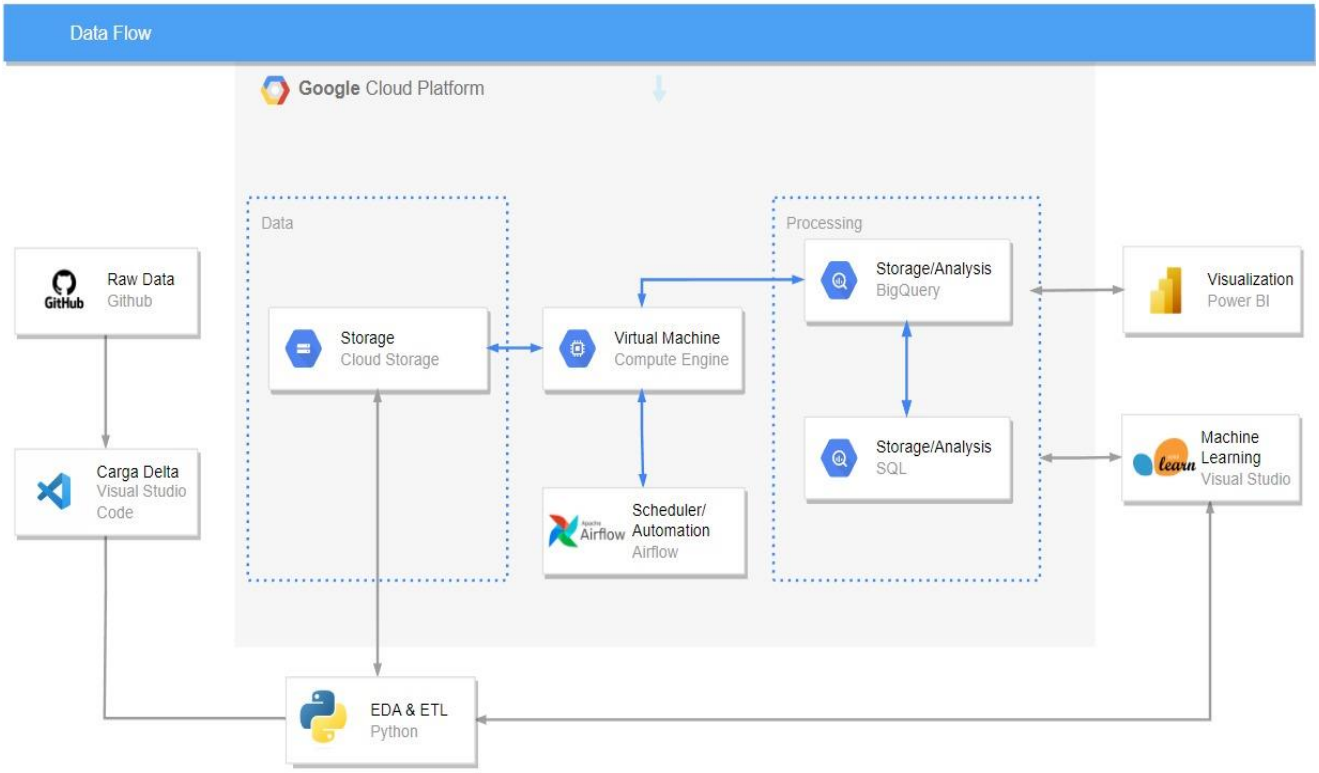


Tarea Macro	Microtarea	Duración	Progreso	Recursos	09-ene	10-ene	11-ene	12-ene	13-ene	16-ene	17-ene	18-ene	19-ene	20-ene	23-ene	24-ene	25-ene	26-ene	27-ene	30-ene	31-ene
Propuesta del Proyecto	Entendimiento de la situación actual	2	100,00%	L H J N R																	
	Objetivos	2	100,00%	L H J N R																	
	Alcance	2	100,00%	L H J N R																	
	Análisis Preliminar de Calidad de Datos	1	100,00%	H J R																	
	Definición de los datos considerados para el trabajo	1	100,00%	L J N																	
	KPIs asociados (planteo)	1	100,00%	L H J N R																	
	Solución propuesta	2	100,00%	L H J N R																	
	Metodología de Propuesta	1	100,00%	L H J N R																	
	Definición stack Tecnológico	1	100,00%	L H J N R																	
	Equipo de Trabajo (Roles y Responsabilidades)	1	100,00%	L H J N R																	
	Elaboración de entregables 1ra semana	1	100,00%	L H J N R																	
	Confección Repositorio Github	1	100,00%	H																	
Data Engineering	Diseño adecuado del Modelo ER	1	0,00%	L N R																	
	Pipelines para alimentar el DW	1	0,00%	N R																	
	Data Warehouse	1	0,00%	H J																	
	Automatización	1	0,00%	L H J N R																	
	Validación de datos	1	0,00%	L H																	
	Documentación	3	0,00%	L H J N R																	
	Diagrama ER detallado (tablas, PK, FK y tipo de dato)	1	0,00%	N R																	
	Diccionario de datos	1	0,00%	N R																	
	Workflow detallando tecnologías	1	0,00%	H J																	
	Elaboración de entregables 2da semana	1	0,00%	L H J N R																	
Data Analytics + ML	Diseño de Reportes/Dashboards	2	0,00%	L H N																	
	Dashboard (conexión a DB)	2	0,00%	L H N																	
	KPIs	2	0,00%	L H J N R																	
	Modelo ML	2	0,00%	L J R																	
	Modelo de ML en producción	2	0,00%	L H J R																	
	Documentación : Selección del modelo, feature engineering	2	0,00%	L H J N R																	
	Informe de análisis	2	0,00%	L H J N R																	
Retoques Finales	Elaboración Readme completo	2	0,00%	L H J N R																	

DEFINICIÓN STACK TECNOLÓGICO

Para la elaboración del proyecto debatimos y analizamos el uso de las siguientes herramientas en función de las tareas a desarrollar, estas son:

- TRABAJO DIARIO: Github, Google Meet, Python
- INGENIERIA DE DATOS: Python, Google Cloud, Big Query
- ANALISIS Y VISUALIZACION DE DATOS: Python, Power Bi
- MODELOS DE MACHINE LEARNING: Python



EQUIPO DE TRABAJO (ROLES Y RESPONSABILIDADES)

De acuerdo a la experiencia y perfil de los integrantes del equipo de trabajo, las tareas, líderes y ayudantes para cada etapa del proyecto quedaron definidas de la siguiente manera:

	LUCILA ALONSO	JOSE JIMENEZ	RICARDO TALAVERA	HECTOR ARGUMEDO	NORBERTO UMBERT
ARQUITECTURA DE DATOS	AYUDANTE	AYUDANTE	LIDER	AYUDANTE	LIDER
ANALISIS DE DATOS	LIDER	AYUDANTE	AYUDANTE	LIDER	LIDER
MACHINE LEARNING	LIDER	LIDER	LIDER	AYUDANTE	AYUDANTE
INGENIERIA DE DATOS	LIDER	LIDER	AYUDANTE	LIDER	AYUDANTE

CONFECCIÓN DEL GITHUB

Para poder visualizar los trabajos de forma remota y conciliar la evolución de versiones, se adoptó la tecnología GIT, por lo tanto se creó un repositorio colaborativo en la plataforma GitHub, donde se podrá visualizar, modificar y crear los contenidos que se irán requiriendo a lo largo del proyecto.