

Productos digitales de Data y análisis del negocio E-commerce Olist Brasil

Fecha: 13/01/2023

Cliente
olist



Índice

Índice	1
Entendimiento de la situación actual	2
Contexto del mercado	2
Entendimiento del Negocio de Olist	2
Retos del mercado	2
Objetivos del Proyecto	3
Objetivos Generales	3
Objetivos Específicos	3
Alcance	3
Requerimientos dentro del alcance del proyecto	3
Requerimientos fuera del alcance del proyecto	4
Riesgos del proyecto	4
KPIs asociados a los objetivos planteados	4
Explicación de KPIs	5
Desarrollo del Proyecto	5
Roadmap y Diagrama de Gantt	5
Roadmap	5
Diagrama de Gantt	6
Stack Tecnológico	7
Metodología de Trabajo	8
Marco de trabajo: Scrum	8
Gestión de Proyectos: Asana	8
Ceremonias o Eventos: Google Meet Discord	9
Control de Versión del Producto: Github	10
Gestión de archivos cloud: Google Drive	10
Diagramación inteligente y colaborativa: Lucidchart Mural	10
Stakeholders	10
Equipo de Trabajo y Roles	11
Análisis preliminar de calidad de datos	11
Diccionario de Datos	13
Relación preliminar de los datos	16
Presentación	17

Entendimiento de la situación actual

Contexto del mercado

Durante la década de 2010 en adelante, la digitalización en la vida de las personas ha ido en aumento en todos los aspectos y especialmente en América Latina, donde la penetración del acceso a internet ha pasado de un 54% en 2015 hasta un 72% de la población en 2019. Tal transformación digital ha traído consigo una ola de cambios en los hábitos de las personas, y especialmente en los hábitos de compra acudiendo cada vez más al mercado online.

Según estudios de Statista, durante los últimos años, América Latina ha experimentado un crecimiento progresivo de compradores por ecommerce quedando entre las primeras regiones en el mundo con mayor crecimiento en este sector. Solo en 2017 se ha registrado un incremento de 39 puntos respecto al año anterior en compradores eCommerce comparado con cifras de 8 a 9 % de crecimiento en el mundo. Además, dentro de la región, el top 3 de países que lideran el crecimiento en el mercado en línea son Brasil, México y Argentina, en ese orden.

Se espera que dicho sector siga en crecimiento proyectando una cifra de 350 millones de usuarios para 2025.

Entendimiento del Negocio de Olist

En este sentido Olist desde 2015 tiene el objetivo de ofrecer un espacio democrático y basado en un espíritu colaborativo que permite a las pequeñas y medianas empresas adaptarse a la venta online haciendo llegar sus productos a grandes mercados (marketplaces) con la mejor calidad de servicio prestado. Olist presta el servicio de venta online evitando que los minoristas tengan que hacerse cargo de toda la gestión que conlleva el e commerce, acercándolo más con el cliente. Olist está pensado como la “tienda de tiendas”.

Retos del mercado

Luego de conocer los datos estadísticos del sector en Latinoamérica, es notable la rápida evolución que se ha experimentado y que aún queda un largo camino por recorrer con la espera de un gran crecimiento. Esto representa un gran reto y oportunidad por parte de las empresas de adaptarse al cambio y acompañar el crecimiento del ecommerce buscando potenciar su servicio al cliente.

En este sentido, es fundamental que se preste especial atención a aspectos como:

- Almacenamiento de productos
- Control de inventarios
- Disminución de tiempos de entrega

- Alianzas estratégicas
- Satisfacción de los clientes en todo el proceso

Objetivos del Proyecto

Definimos los objetivos generales y específicos. ¿Qué queremos lograr?

Objetivos Generales

- Entender el funcionamiento del mercado ecommerce y el negocio de Olist. Diagnosticar su contexto actual
- Identificar patrones de comportamiento de los usuarios de Olist (Pymes y clientes)
- Descubrir oportunidades de crecimiento para el negocio

Objetivos Específicos

- Analizar los procesos de compra, venta y logística del negocio. Conocer la distribución geográfica de ventas en Brasil.
- Descubrir factores que influyen en el abandono de los procesos de compra y/o que producen insatisfacción de los clientes.
- Analizar la interacción entre los clientes, Olist y las pequeñas y medianas empresas.

Alcance

Requerimientos dentro del alcance del proyecto

Este proyecto es realizado en un plazo de 4 semanas desde el 9 de enero de 2023 por el equipo de desarrollo de profesionales de Data consolidando en la entrega un conjunto de productos y herramientas digitales basado en información histórica del negocio:

- Informe EDA (Exploratory Data Analysis)
- Pipelines de ETL (Extract, Transform and Load)
- Data Warehouse on cloud junto con el Modelo Entidad-Relación
- Automatización de los procesos de ETL
- Validación de datos
- Dashboard con métricas, KPIs y conclusiones del análisis
- Implementación de Modelos de Machine Learning (Motor de recomendación y análisis de sentimiento)
- Documentación
 - Informe de propuesta del proyecto

- Diccionario y estructura de datos (Análisis de datos)
- Diagrama Entidad-Relación detallado
- Workflow del stack tecnológico elegía
- Informe de Análisis de business intelligence
- Documentación de Feature Engineering y modelado

Requerimientos fuera del alcance del proyecto

- Desarrollo de Modelos de Machine Learning no supervisados para la segmentación de clientes.
- Desarrollo de un Data Lake

Riesgos del proyecto

Para el deploy del conjunto de herramientas se utilizarán herramientas cloud en primer lugar gratuitas, que cuentan con ciertas limitaciones relacionadas con la capacidad de almacenamiento y transacciones. El principal riesgo asociado está relacionado con la escalabilidad del proyecto si se superan los límites de procesamiento en la nube. Se analizará dicha situación en la documentación correspondiente y se propondrán alternativas para mitigar los riesgos.

KPIs asociados a los objetivos planteados

- Variación porcentual ventas: comparando con año anterior
- Ticket medio: suma de ingresos totales, dividido por el número de ventas del último año. El ticket medio muestra cuánto gastan sus clientes en sus compras.
- Nivel de Satisfacción: Porcentaje total de reviews 4 y 5 (Muy satisfecho) con respecto al total de reviews.
- Variación porcentual del número total pedidos por trimestre (filtro por producto, pymes, periodos, etc).
- Promedio tiempo de envío: si el cliente vive distinto lugar que la pyme.
- Tasa de entregas dentro del plazo – variación.

Explicación de KPIs

1. Variación porcentual de ventas con respecto al año anterior: permite comparar las ventas actuales con las del año anterior y ver si hay un crecimiento o una disminución en las ventas. Es un indicador clave para medir el desempeño de un

negocio en el tiempo.

2. Ticket medio: muestra el valor promedio de las compras de los clientes. Es útil para diseñar estrategias para aumentar los beneficios ya que si el ticket medio es alto, se pueden aumentar las ventas manteniendo el mismo nivel de clientes.
3. Nivel de satisfacción: permite conocer cómo los clientes perciben la calidad del servicio o producto, ayudando a detectar cuales son las áreas de mejora y aumentar la satisfacción de los clientes.
4. Variación porcentual del número total pedidos por trimestre: permite medir la demanda de productos y servicios, y conocer si hay cambios en la tendencia de la demanda.
5. Promedio tiempo de envío: permite conocer si los tiempos de entrega son adecuados para los clientes y si hay algún retraso en los envíos. Es importante para evitar quejas.
6. Tasa de entregas dentro del plazo: permite medir la eficiencia del proceso de entrega y ayudar a detectar problemas en el proceso de envíos para corregirlos.

Desarrollo del Proyecto

Se establecen los puntos importantes de la solución propuesta respondiendo las preguntas de ¿Qué se va a hacer? y ¿Cómo se lo va a llevar a cabo?.

Roadmap y Diagrama de Gantt

Roadmap

Se realiza un Roadmap sencillo con los entregables pactados para cada Sprint y los hitos principales que aglomeran varias tareas dentro.

[Link para visualización del roadmap en Lucidchart](#)

Roadmap Proyecto OLIST Racont Data Consultant

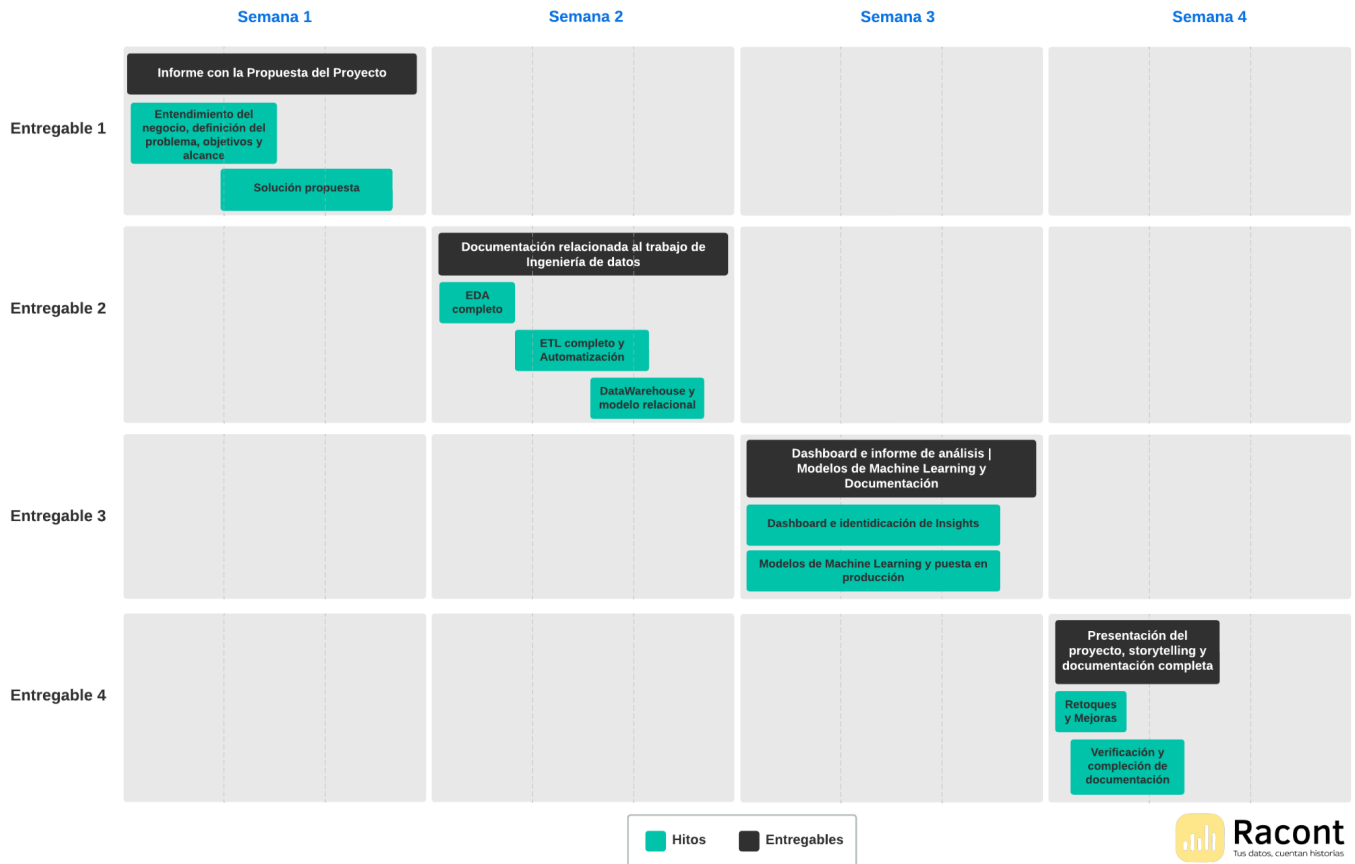
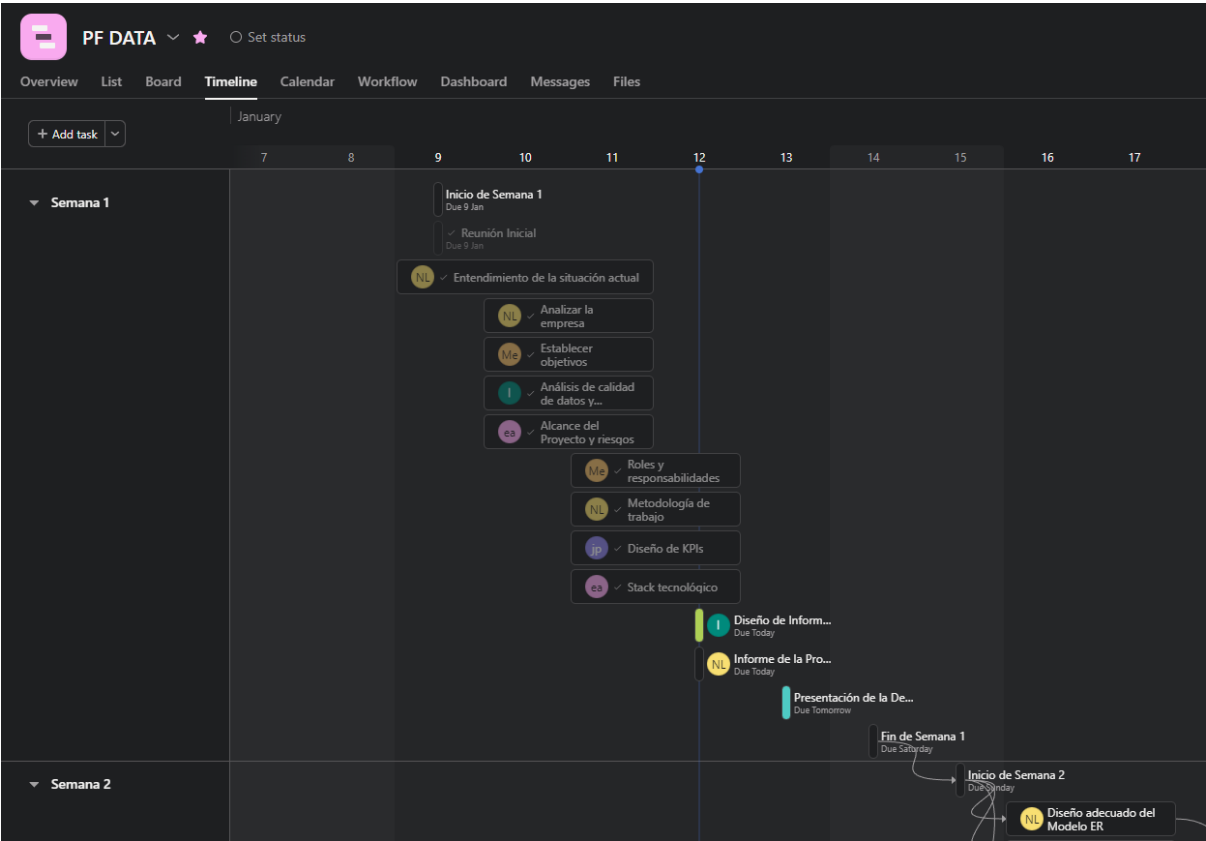


Diagrama de Gantt

Se utiliza la herramienta de gestión de proyectos Asana para la creación y seguimiento de las tareas en un Timeline interactivo entre los miembros del proyecto. Se muestra a continuación una imagen del Diagrama generado que contiene las tareas con su duración y miembro del equipo asignado.

[Link de visualización del Diagrama de Gantt en Asana](#)



Stack Tecnológico

Entregable	Stack
EDA	<ul style="list-style-type: none">• Python• Pandas• Matplotlib• Seaborn
ETL	<ul style="list-style-type: none">• Python- PySpark• Apache Spark• Databricks
Cloud	<ul style="list-style-type: none">• Azure
Data Warehouse	<ul style="list-style-type: none">• PostgreSQL• Delta Lake• Databricks
Dashboard	<ul style="list-style-type: none">• Power BI o Tableau

Metodología de Trabajo

Se define como se lleva a cabo el trabajo en base a una organización ágil y eficiente que aseguren la calidad del proyecto y la satisfacción del cliente

Marco de trabajo: Scrum

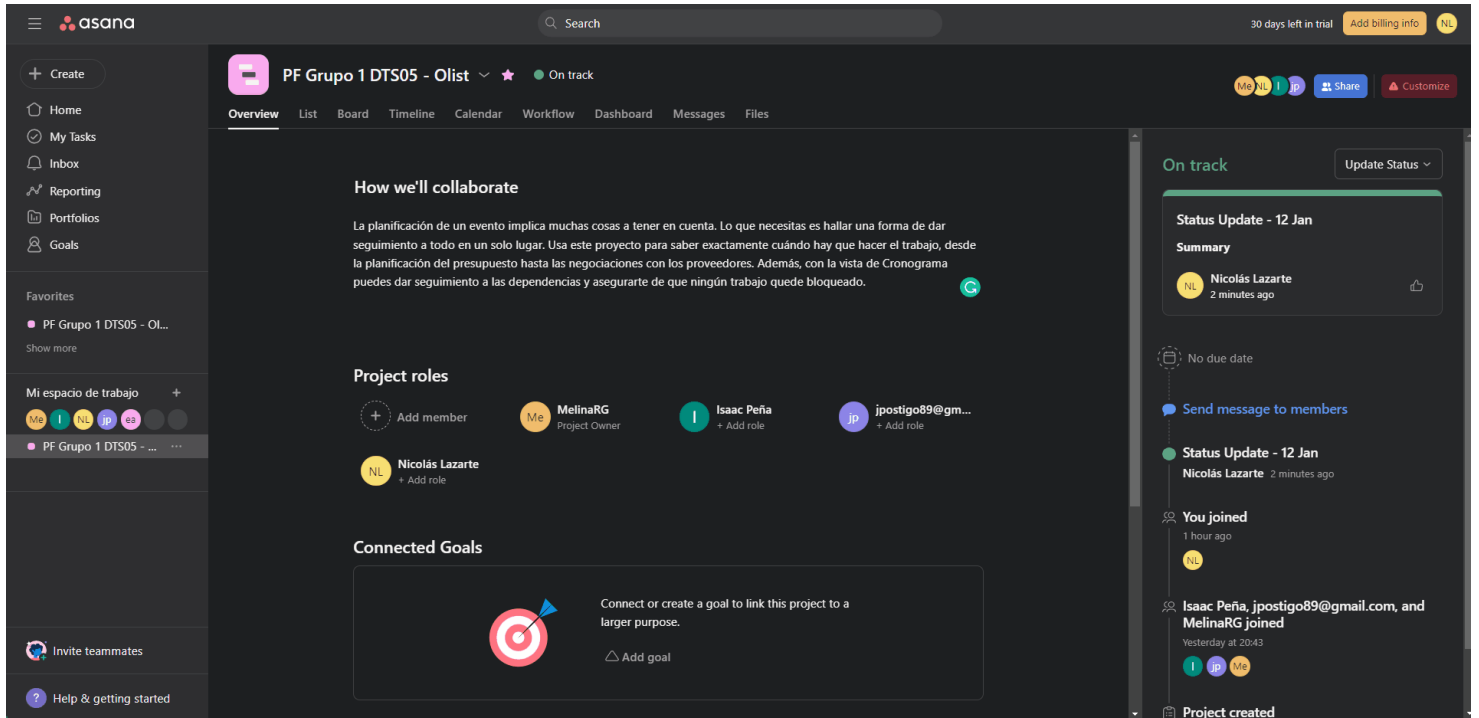
Para el desarrollo del proyecto se utilizará Scrum como metodología ágil, organizado de la siguiente manera:

- El Proyecto se diseña con cuatro Sprints de una semana de duración cada uno
- La estructura de cada Sprint está dada por:
 - Sprint planning (día lunes de cada semana)
 - Dailys a una hora pactada durante la mañana con nuestro equipo de trabajo para revisión del backlog, tareas en curso y finalizadas, planteo de problemas y posibles soluciones
 - Dailys con Henry Mentor después del mediodía para revisión del avance de los entregables y resolución de dudas
 - Sprint Review: Demo a fin de semana (viernes) con el Product Owner para presentar entregable correspondiente al sprint y evaluar el cumplimiento de los objetivos
 - Sprint Retrospective: Reunión con el equipo de trabajo luego del sprint review para reflexionar sobre el trabajo realizado y detectar acciones de mejora.

Gestión de Proyectos: Asana



Como herramienta para la gestión de proyectos utilizamos Asana. Se gestiona el backlog dividiendo las tareas por Sprints semanales, asignando responsables y fechas de inicio y finalización estimadas. Las tareas se muestran en forma de lista, tarjetas, en un Timeline (diagrama de Gantt) con gestión de dependencias y en un Workflow. Se evalúa el progreso a través del análisis del Dashboard que ofrece la herramienta.

El siguiente Gif muestra un overview del espacio de trabajo y las herramientas de gestión.



Ceremonias o Eventos: Google Meet | Discord

Para el desarrollo de los eventos se utilizan las herramientas de Google Meet y Discord con horarios pactados que contribuyen a la organización de trabajo.

Evento	Propósito	Herramienta
Daily con el equipo de desarrollo (Previo a la reunión con Henry Mentor)	Ver avanzar, identificar problemas, proponer soluciones	
Daily con Henry Mentor (Pactado a través de Google Calendar)	Evacuar y resolver dudas en el avance del proyecto	
Demo con PO (Último día de cada Sprint)	Evaluar los entregables semanales y el cumplimiento de objetivos propuestos	

Control de Versión del Producto: Github

En el repositorio podrán encontrar todos los archivos del proyecto y el historial de revisiones de cada uno de ellos. Cada integrante del grupo va a trabajar en una rama diferente y al finalizar el desarrollo del mismo se harán las uniones (merge) correspondientes. Por último el archivo README.md contará con un resumen detallado del proyecto.

Link: [Github](#) 

Gestión de archivos cloud: Google Drive

Se utiliza Google Drive para la organización de archivos del proyecto en forma simultánea entre todos los integrantes del equipo de trabajo para facilitar el trabajo colaborativo. Allí se alojan:

- Assets (Imágenes, logos, plantillas, etc.)
- Datasets y fuentes de datos
- Bibliografía
- Drafts
- Informes
- Presentaciones

Diagramación inteligente y colaborativa: Lucidchart | Mural

El trabajo colaborativo combinado con herramientas visuales de diagramación es de suma importancia para que el equipo pueda proponer ideas, aunar criterios y mantener una documentación ordenada y visual del trabajo. Las herramientas usadas son sitios web donde se crean espacios de trabajo colaborativo para el equipo (brainstorming, roadmaps, diagramas Entidad Relación, etc.)

- Lucid – <https://www.lucidchart.com/>
- Mural – <https://www.mural.co/>

Stakeholders

Los interesados identificados del proyecto son:

- Client: Olist
- Product Owner (PO): Juliana Aragón
- Scrum Master | Henry Mentor (HM): Leandro Franchina

Equipo de Trabajo y Roles



Melina Griffo | **Data Engineer**

Experta en los procesos de ETL y desarrollo del backend en conjunción con las herramientas cloud



Isaac Peña | **Data Engineer**

Experto en análisis de datos, procesos ETL y desarrollo del backend



Julio Postigo | **Data Analyst**

Experiencia en análisis e interpretación de datos. Diseño y presentación de reportes.



Emmanuel Corral | **Analista Funcional**

Experiencia en procesos EDA y ETL. Especialista en interpretación y exposición de datos.



Nicolás Lazarte | **Data Scientist**

Científico de datos con experiencia en el desarrollo de modelos supervisados de Machine Learning

Análisis preliminar de calidad de datos

Para la realización de este proyecto Olist nos proporcionó 2 links de Kaggle como fuente de datos en donde están alojados sus datasets y nos indicó la versión a utilizar..

Primera fuente: Brazilian E-Commerce Public Dataset by Olist

Descripción

Datos públicos de comercio electrónico brasileño de pedidos realizados en la tienda Olist. Se tiene información de cien mil pedidos de 2016 a 2018 realizados en múltiples mercados en Brasil.

Lista de Datasets

- o 2.- olist_customers_dataset
 - o 3.- olist_geolocation_dataset
 - o 5.- olist_order_items_dataset
 - o 6.- olist_order_payments_dataset
 - o 7.- olist_order_reviews_dataset
 - o 8.- olist_orders_dataset
 - o 9.- olist_products_dataset
 - o 10.- olist_sellers_dataset
 - o 11.- product_category_name_translation
- Fuente: [Link](#)
 - Fecha de descarga: 09/01/2023
 - Fecha de última actualización: 01/10/2021

Segunda Fuente: Marketing Funnel by Olist

Descripción

Datos de embudo de marketing de vendedores que completaron solicitudes de contacto para vender sus productos en Olist Store. Se tiene información de ocho mil clientes potenciales calificados de marketing (MQL) que solicitaron contacto entre el 1 de junio de 2017 y el 1 de junio de 2018 que fueron seleccionados aleatoriamente del total de MQL.

Lista de Datasets

- o 1.- olist_closed_deals_dataset
 - o 4.- olist_marketing_qualified_leads_dataset
- Fuente: [Link](#)
 - Fecha de descarga: 09/01/2023

- Fecha de última actualización: 16/11/2018.

Diccionario de Datos

En las siguientes tablas se explica el significado de cada columna y el tipo de dato que contienen por cada dataset para su correcto entendimiento y aplicación en el presente proyecto.

1.- olist_closed_deals_dataset		
Recopilación y verificación de información de un qualified lead:"cliente potencial calificado" después de completar un formulario que es comunicado a un representante de desarrollo de ventas, los qualified lead después de cerrar el trato se convierten en vendedores de Olist.		
Columna	Descripción	Tipo de dato
mql_id	Id del lead	Alfanumérico
seller_id	Id del lead convertido en vendedor de Olist	Alfanumérico
sdr_id	Id del representante de desarrollo de ventas	Alfanumérico
sr_id	Id del representante de ventas	Alfanumérico
won_date	Fecha en que se cerró el trato de Olist con el lead	Fecha y hora
business_segment	Sector de negocio del lead	Alfanumérico
lead_type	El tipo al que pertenece el lead	Alfanumérico
lead_behaviour_profile	Perfil de comportamiento del lead	Alfanumérico
has_company	El lead ¿tiene empresa?	Booleana
has_gtin	El lead ¿Tiene número gtin (Número Global de Artículo Comercial) para sus productos?	Booleana
average_stock	Promedio de existencias declarado por el lead	Alfanumérico
business_type	Tipo de negocio del lead	Alfanumérico
declared_product_catalog_size	Tamaño de catálogo de productos declarado por el lead	Decimal
declared_monthly_revenue	Ingresos mensuales declarados por el lead	Decimal

2.- olist_customers_dataset		
Tiene información sobre los clientes que compran productos en Olist y su ubicación		
Columna	Descripción	Tipo de dato
customer_id	Id del cliente	Alfanumérico
customer_unique_id	Id único de cliente, único por persona	Alfanumérico
customer_zip_code_prefix	Código postal de cliente	Entero

customer_city	Ciudad de cliente	Alfanumérico
customer_state	Estado del cliente	Alfanumérico

3.- olist_geolocation_dataset		
Información sobre los códigos postales de Brasil y sus coordenadas de latitud y longitud		
Columna	Descripción	Tipo de dato
geolocation_zip_code_prefix	Código postal	Entero
geolocation_lat	Latitud	Decimal
geolocation_lng	Longitud	Decimal
geolocation_city	Ciudad	Alfanumérico
geolocation_state	Estado	Alfanumérico

4.- olist_marketing_qualified_leads_dataset		
Información acerca de los leads:"clientes potenciales" que completaron un formulario en una página de aterrizaje después de haber sido filtrados para vender sus productos en Olist, siendo ahora qualified lead : "Clientes potenciales calificados"		
Columna	Descripción	Tipo de dato
mql_id	Id del lead	Alfanumérico
first_contact_date	Fecha de solicitud del primer contacto	Fecha
landing_page_id	Id de la página de aterrizaje donde se adquirió el cliente potencial	Alfanumérico
origin	Tipo de medio en donde se adquirió el lead	Alfanumérico

5.- olist_order_items_dataset		
Información acerca de productos comprados por clientes en Olist en cada orden		
Columna	Descripción	Tipo de dato
order_id	Id de la orden	Alfanumérico
order_item_id	Número secuencial que identifica el número de ítem para una misma orden	Entero
product_id	Id del producto	Alfanumérico
seller_id	Id del vendedor	Alfanumérico
shipping_limit_date	Fecha límite de envío hacia logística	Fecha y Hora
price	Precio del producto	Decimal
freight_value	Costo de transportar el producto(si una orden tiene más de un ítem, el costo se divide entre ítems	Decimal

6.- olist_order_payments_dataset		
Información sobre opciones de pago de las órdenes de compra		

Columna	Descripción	Tipo de dato
order_id	Id de la orden	Alfanumérico
payment_sequential	Número de pagos parciales para una misma orden	Entero
payment_type	Método de pago	Alfanumérico
payment_installments	Número de cuotas	Entero
payment_value	Valor de pago	Decimal

7.- olist_order_reviews_dataset		
Datos sobre las reseñas hechas por los clientes después de recibir el producto o cuando caduque la fecha estimada		
Columna	Descripción	Tipo de dato
review_id	Id de la reseña	Alfanumérico
order_id	Id de la orden	Alfanumérico
review_score	Nota dada por el cliente en su reseña	Entero
review_comment_title	Título de la reseña dejada por el cliente	Alfanumérico
review_comment_message	Mensaje en la reseña dejada por el cliente	Alfanumérico
review_creation_date	Tiempo en que se envió la reseña al cliente	Fecha y hora
review_answer_timestamp	Tiempo en que la reseña del cliente fue completada	Fecha y hora

8.- olist_orders_dataset		
Información de cada una de las órdenes de compra de Olist		
Columna	Descripción	Tipo de dato
order_id	Id de la orden	Alfanumérico
customer_id	Id del cliente	Alfanumérico
order_status	Estado de la orden	Alfanumérico
order_purchase_timestamp	Tiempo de compra	Fecha y hora
order_approved_at	Tiempo de aprobación de pago de la compra	Fecha y hora
order_delivered_carrier_date	Tiempo de entrega al transportista	Fecha y hora
order_delivered_customer_date	Tiempo real de llegada del pedido al cliente	Fecha y hora
order_estimated_delivery_date	Tiempo estimado de llegada del pedido al cliente	Fecha y hora

9.- olist_products_dataset		
Información sobre productos que se venden en Olist		
Columna	Descripción	Tipo de dato
product_id	Id del producto	Alfanumérico



product_category_name	Nombre de categoría del producto	Alfanumérico
product_name_lenght	Longitud del nombre del producto	Decimal
product_description_lenght	Longitud de la descripción del producto	Decimal
product_photos_qty	Número de fotos publicadas del producto	Decimal
product_weight_g	Peso del producto en gramos	Decimal
product_length_cm	Longitud del producto en centímetros	Decimal
product_height_cm	Altura del producto en centímetros	Decimal
product_width_cm	Ancho del producto en centímetros	Decimal

10.- olist_sellers_dataset		
Información sobre los vendedores de Olist que cumplieron ordenes		
Columna	Descripción	Tipo de dato
seller_id	Id del vendedor	Alfanumérico
seller_zip_code_prefix	Código postal del vendedor	Entero
seller_city	Ciudad del vendedor	Alfanumérico
seller_state	Estado del vendedor	Alfanumérico

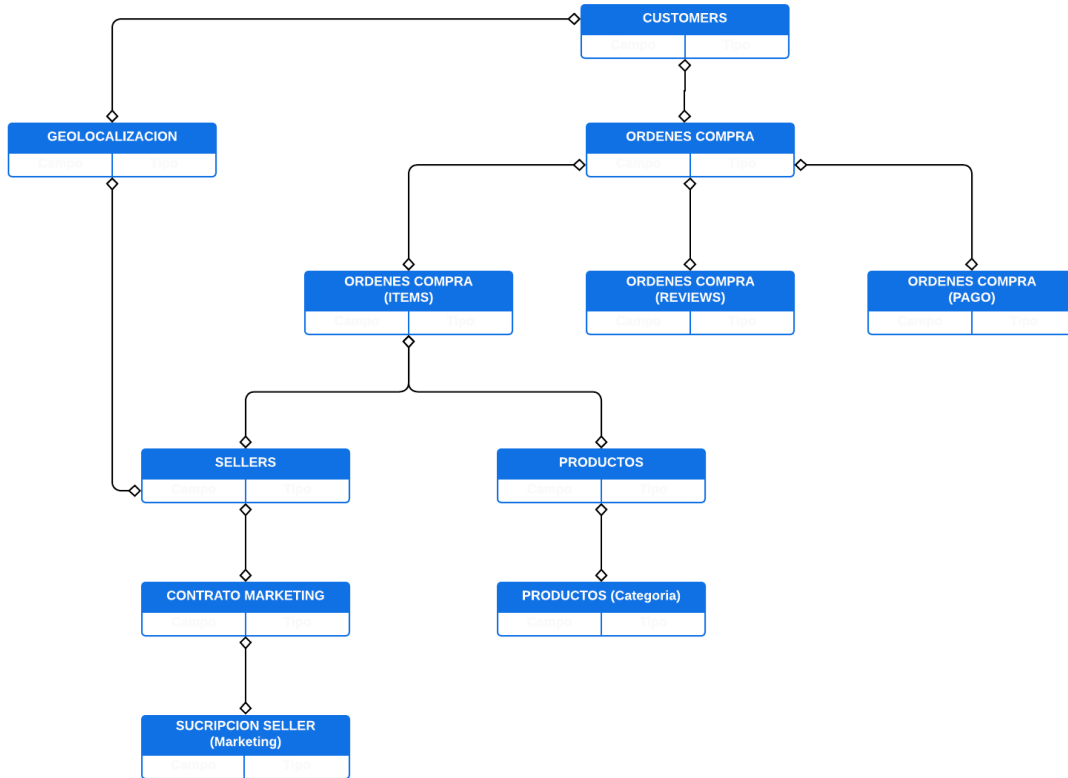
11.- product_category_name_translation		
Traducción de nombre de categoría de producto de portugués a inglés		
Columna	Descripción	Tipo de dato
product_category_name	Nombre de categoría de producto en Portugués	Alfanumérico
product_category_name_english	Nombre de categoría de producto en Inglés	Alfanumérico

Relación preliminar de los datos

En la siguiente imagen se muestra la relación preliminar existente entre los 11 datasets para su correcto entendimiento, uso y posterior generación de base de datos.

OLIST DATASETS 2016-2018

Racont | January 13, 2023



Presentación

Se presenta al cliente la propuesta. Se adjunta el link a dicha presentación.

[Link Presentación DEMO](#)