



DeCeLeNe Consultora

ALCANCE DEL PROYECTO

FECHA: 28/07/2022

1- EQUIPO DE TRABAJO - ROLES Y RESPONSABILIDADES

-Nelson Rodriguez (Project Manager), Data Analyst, Ingeniero/Analista de datos, reportes de KPI's.

-Delfina Ubiedo, Ingeniera/Analista de datos, Limpieza y estructuración de datos, Documentación, visualización de datos

-Leandro Paez, Ingeniero/Analista de datos, Creación y carga de DW.

-Cecilia Romero, Software Engineer, Ingeniera/Analista de datos, Investigación e implementación de tecnologías, Automatización y creación de Pipelines, visualización de datos.

2- ENTENDIMIENTO DE LA SITUACIÓN ACTUAL

Teniendo en cuenta la propuesta de expansión de su organización hacia el mercado brasilero, nuestro equipo efectuó un análisis del mercado del comercio electrónico para dicho país. Para esto, nos basamos en los datos públicos de la plataforma Olist, reconocida empresa de e-commerce que opera en el mercado brasilero. El conjunto de datos a utilizar contiene información sobre órdenes de ventas, productos, vendedores, clientes e incluso reviews de las ventas escritos por diversos clientes, desde el 2016 hasta el 2018.

3-OBJETIVO

Realizar el estudio y comprender el comportamiento para comenzar la expansión en el mercado del e-commerce en Brasil.

4- ALCANCE DEL PROYECTO

Conocer el mercado, la distribución geográfica y el comportamiento de los clientes (vendedores/compradores), productos más elegidos y mejor valorados, reconocimiento de los medios de pago y planes de pago, demora y costos de entrega de pedidos.

A partir de este análisis efectuaremos una serie de recomendaciones para llevar a cabo posibles planes de acción en el mercado brasilero.

5- FUERA DE ALCANCE

Queda fuera del alcance del proyecto:

- Análisis de comisiones cobradas a los vendedores.
- Estudio de alternativas en el área de logística.
- Temáticas referentes a cuestiones financieras del modelo de negocio.

6- SOLUCIÓN PROPUESTA

Como **objetivo principal**, proponemos obtener el **15% del mercado que abarca Olist en el primer año**.

Para verificar el avance de este objetivo, vamos a centrarnos en las siguientes KPI's:

- Ventas en el primer año.
- Cantidad de clientes.
- Cantidad de vendedores.
- Cantidad de productos.

La creación de un *dashboard* para la visualización del análisis de las métricas.

La implementación de dos modelos de *predicción de Machine Learning*:

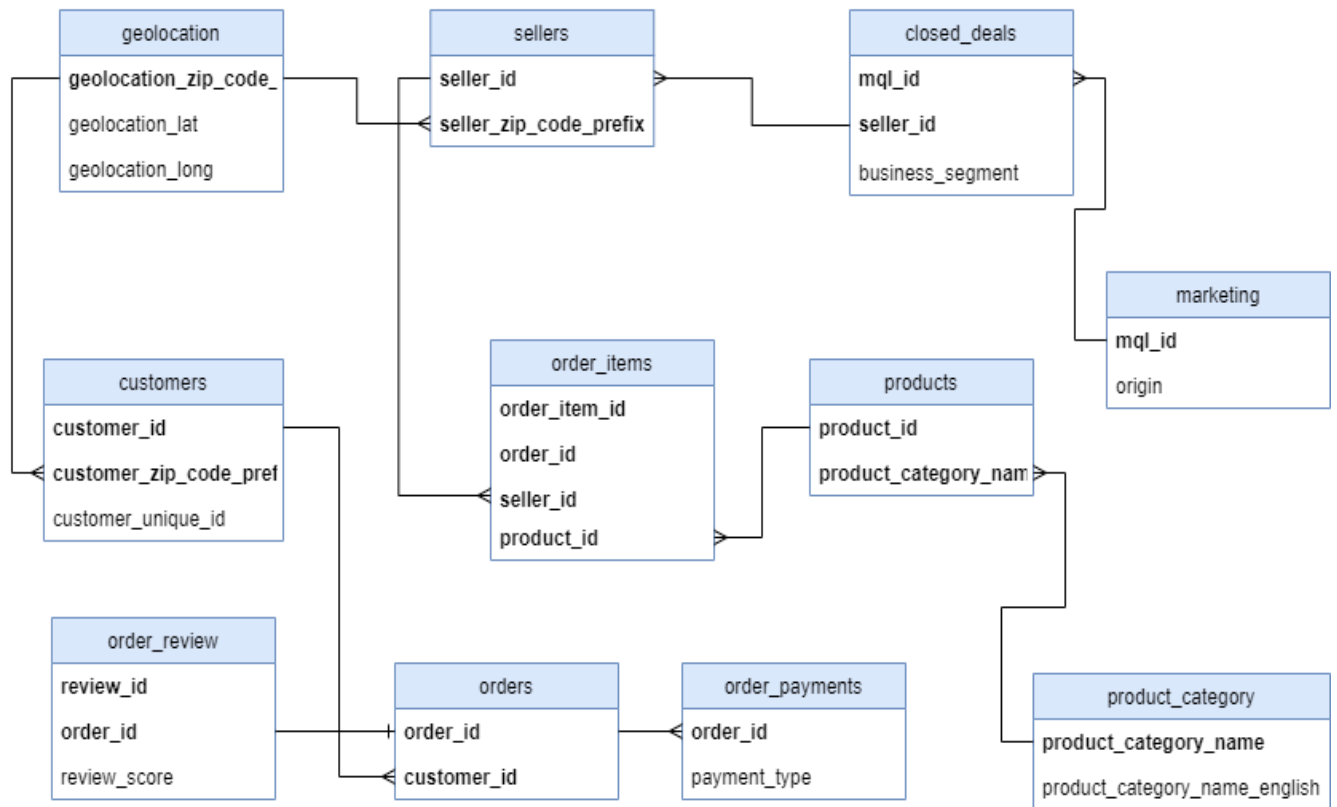
- Modelo de regresión para predecir el costo de envío.
- Modelo de clasificación para determinar si el producto enviado puede tener delay.

7 - HERRAMIENTAS UTILIZADAS

- **AWS IAM:** Administración de usuarios, perfiles y roles.
- **AWS LAKE FORMATION:** Administración de seguridad del DataLake.
- **AWS GLUE:** Administración de tablas dentro del DataLake.
- **AWS S3:** Storage del DataLake.
- **AWS LAMBDA:** Automatizar la carga incremental de datos.
- **AWS ATHENA:** Aplicar consultas al DataLake.

- **BOTO3:** Conexión de Python con los servicios de AWS S3 y AWS Athena.

8 - MODELO DE DATOS



9- TRATAMIENTO DE DATOS

Se exploraron los datos para realizar los siguientes cambios:

Se reemplazó datos nulos en las tablas *order_review*, *closed_deals* y *marketing* con el texto “Sin Dato”.

Se reemplazó datos nulos en las tablas *product* con el valor 0.0.

Se eliminaron columnas con gran cantidad de nulos en la tabla *closed_deals*.

Se realizó la detección y limpieza de outliers en la tabla *order_payments*.

Se realizó la detección y limpieza de duplicados en la tabla *order_review*.

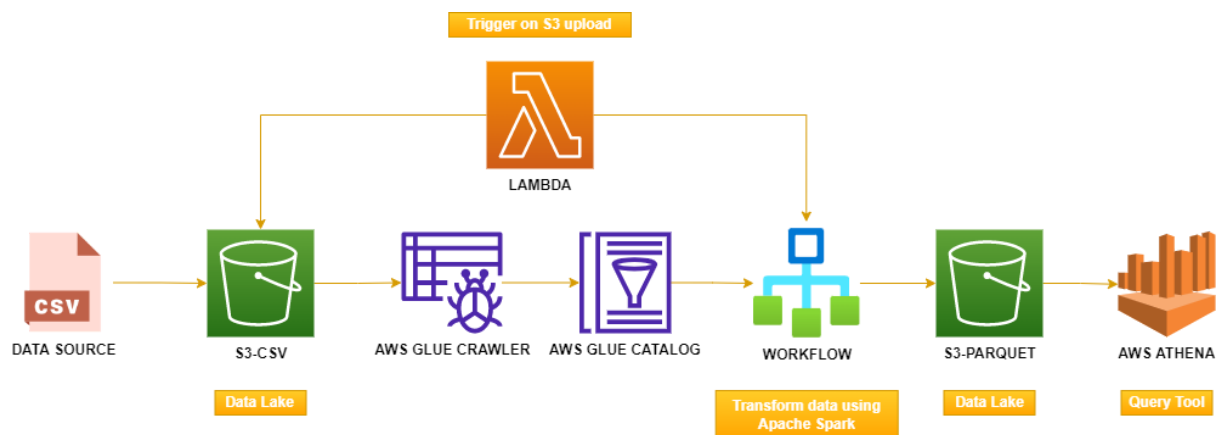
10- MÉTRICAS ANALIZADAS

Las métricas que se analizaron son:

- Evolución de ventas por volumen y cantidad.

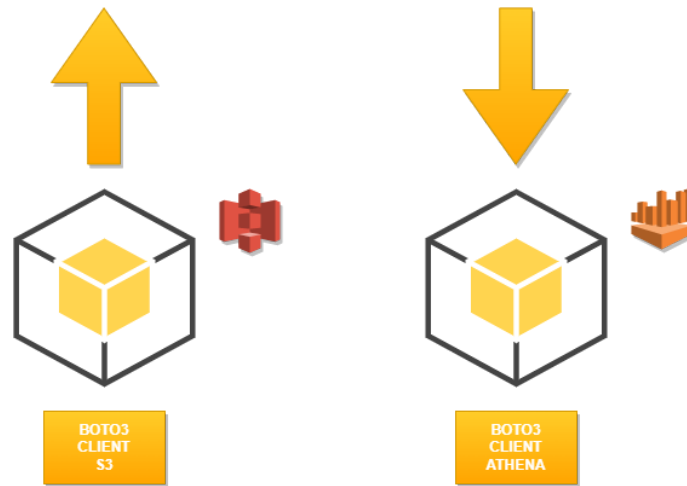
- Estados/ Regiones con mayor cantidad de vendedores y compradores.
- Tiempo de entrega por estado.
- TOP de Productos más vendido (por cantidad) y productos con mejor reputación.
- Medio de pago elegido por región.
- Cantidad de cuotas elegidas por categoría de producto y por monto.
- Origen del registro de leads (vendedores)
- Día de la semana y horario con mayor/menor cantidad de ventas.
- Promedio de costo de envío por región.
- Relación entre el monto de la compra y el medio de pago elegido.
- Cantidad de vendedores por categorías de productos.

11 - PIPELINE IMPLEMENTADO



Automatización

El pipeline implementado contempla la automatización a través de la utilización de la librería BOTO3, la cual permite conectar Python con los servicios de AWS S3 y AWS Athena.



12- CREACIÓN DE DASHBOARD

El dashboard se creó con el framework Streamlit y el deploy del proyecto se realizó utilizando la plataforma Heroku.



ANÁLISIS DE MERCADO

13- OBJETIVO PRINCIPAL:

- Obtener el **15% del mercado** que abarca Olist en el primer año.

Esto implica obtener los siguientes valores en nuestras KPI's:

- Ventas: 1.050.000 R\$
- Compradores: 14.700
- Vendedores: 460
- Productos a vender: 7.500

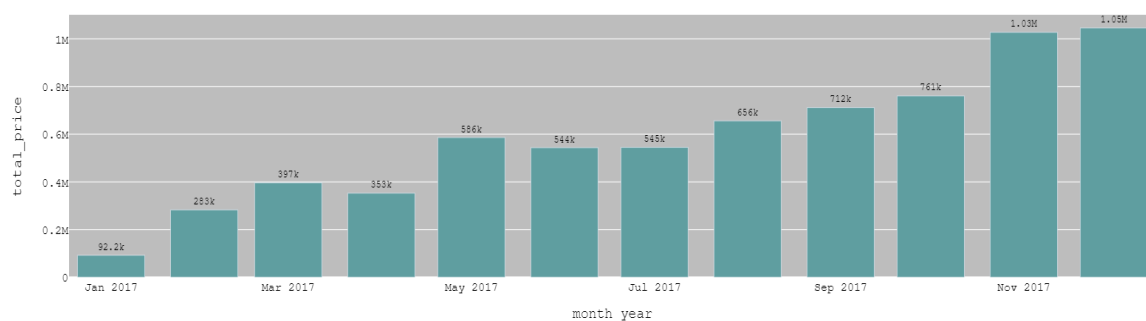
14- ANÁLISIS DE VENTAS

Se utilizó como base del análisis el año 2017, ya que es el único que tiene los datos del año completo.

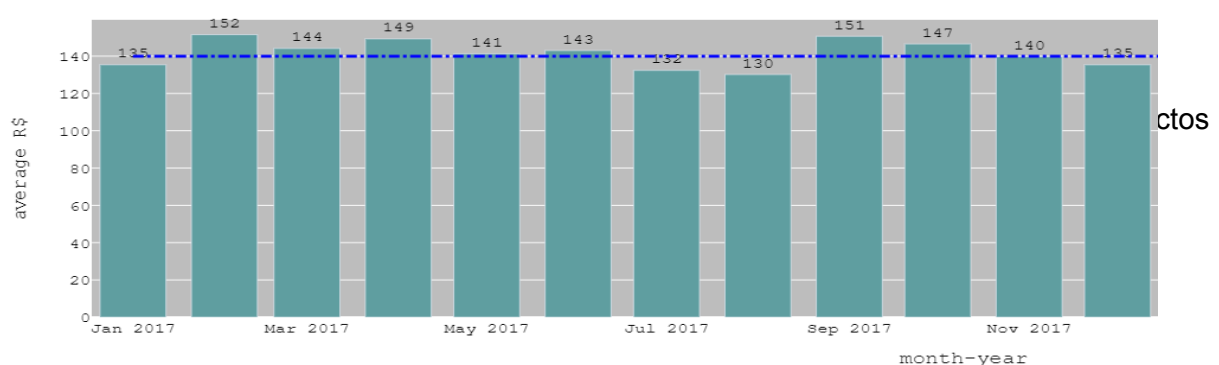
Durante este año, las ventas se han incrementado 1034%, registrando ingresos por 7 millones de R\$.

El ticket promedio observado fue de 140,00 R\$, el cual se mantuvo en ese rango durante todos los meses analizados.

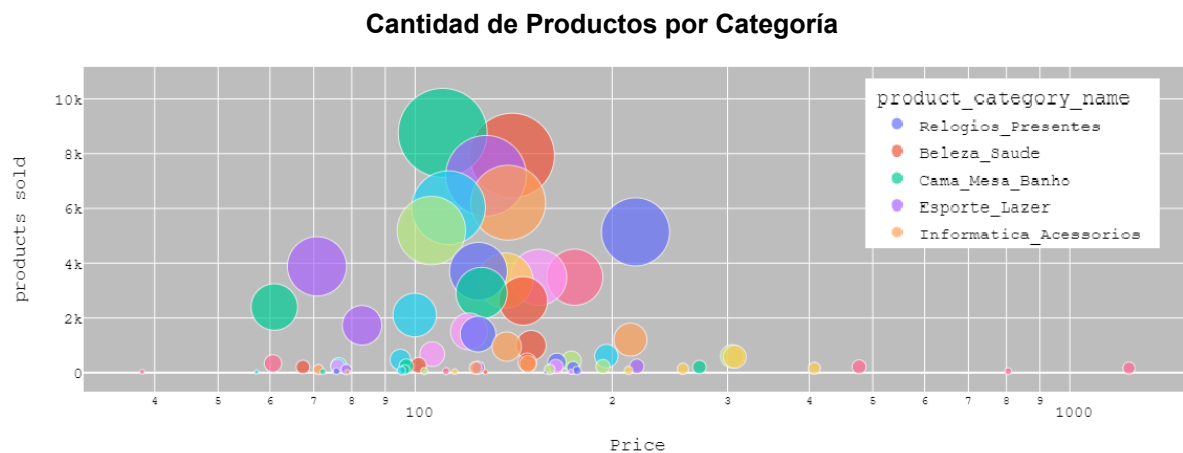
Evolución de ventas por mes



Precio Medio de Productos



Podemos ver que los productos más vendidos se encuentran cercanos al ticket promedio. Podemos concluir que el núcleo del negocio en Olist se encuentra en la venta por cantidad de productos a precios moderados y no en la venta de productos costosos.

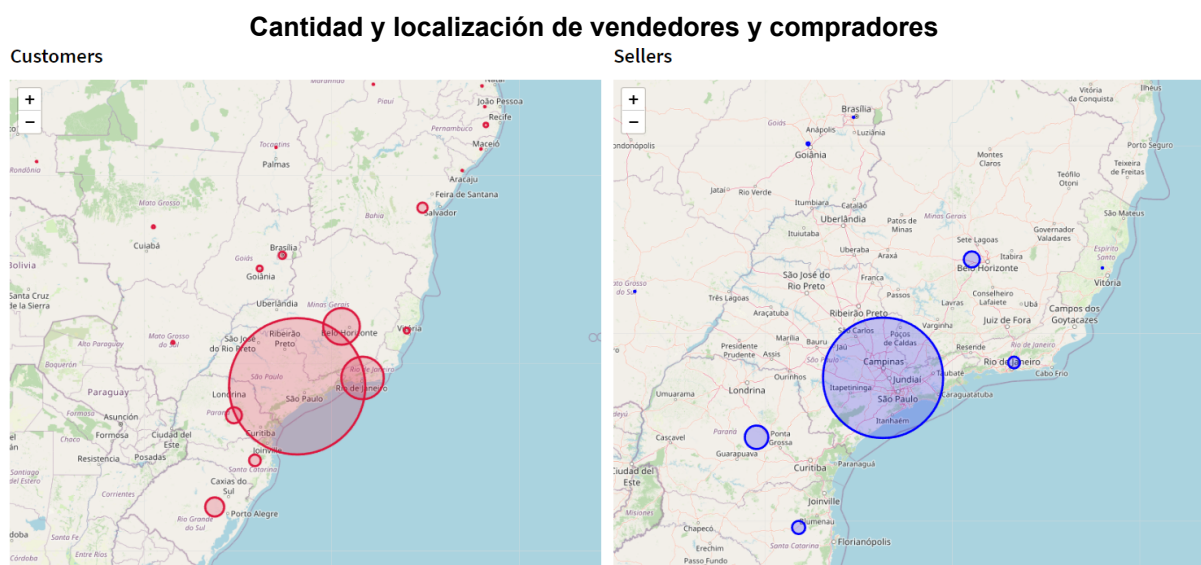


15- ANÁLISIS DE VENDEDORES Y COMPRADORES

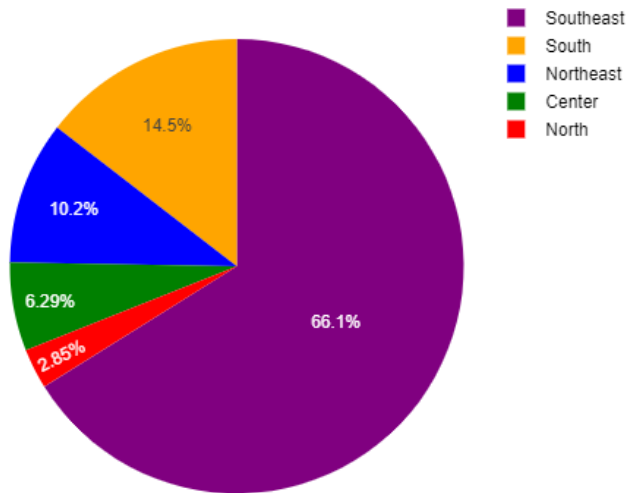
Analizando la distribución de los vendedores, observamos que el 95.47% de los mismos se localizan en las regiones Sur y Sudeste de Brasil.

Los estados con mayor cantidad de vendedores son: Sao Paulo (59.7%), Paraná (11.3%) y Minas Gerais (7.88%) .

En cuanto a los compradores, el 83.55% de ellos se concentran en las regiones Sur y Sudeste. Los estados con mayor cantidad de clientes son: Sao Paulo (41.98%), Rio de Janeiro (12.92%) y Minas Gerais (11.70%).

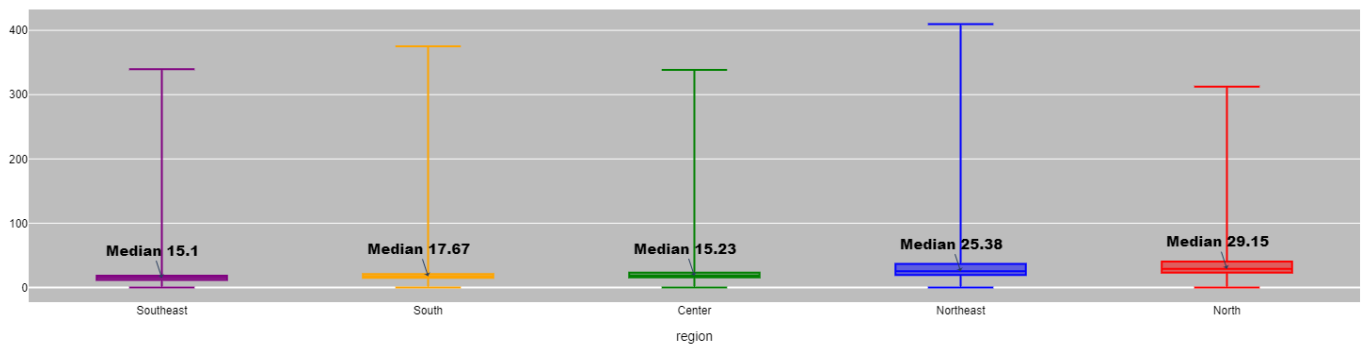


Volumen de Ventas por Región

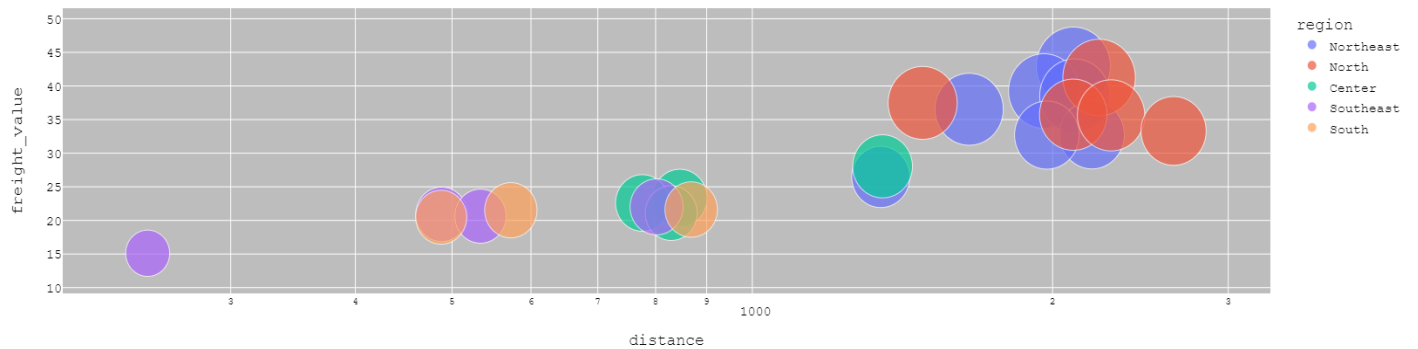


El 66.1% del volumen de ventas se centra en la región Sudeste, seguida por la región Sur con un 14.5%.

Costo de envío medio por región



Distancia/Costo entre compradores y vendedores



Comparando los costos medios de envíos para los clientes de la región Sudeste y los de la región Norte se observa un 66% de diferencia .

Estos costos elevados pueden desalentar las compras para los clientes de la región norte . Por lo cual se recomienda, en una primera etapa concentrarse en captar vendedores y clientes de la zona Sur y Sudeste .

También se visualiza como una oportunidad de negocio, atraer vendedores de la zona Norte y Nordeste para captar un mercado aún no explotado por Olist.

16- ANÁLISIS DE DEMORA DE ENTREGA

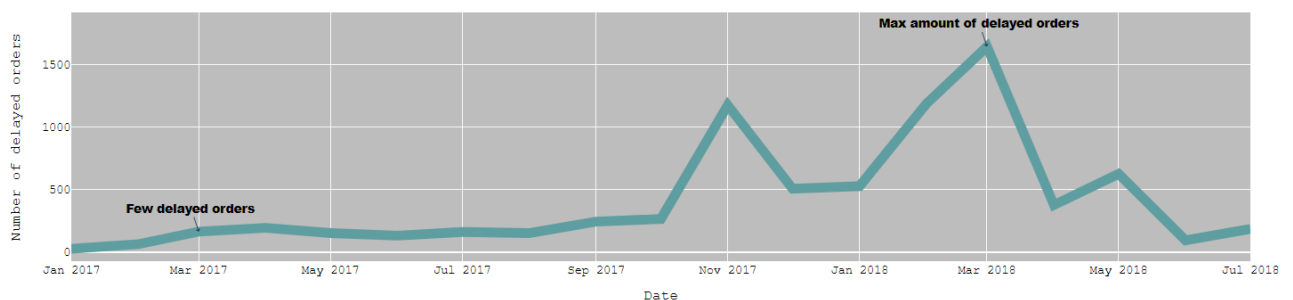
Actualmente Olist cumple con el 89.1% de los plazos de entrega.

Por lo que es recomendable igualar o aumentar en una primera etapa estos plazos, para lograr fidelizar a los clientes.

Analizando el historial de Olist, visualizamos el incremento en el volumen de las ventas que no cumplieron los plazos de entrega, pero se disminuyeron considerablemente los días de demora, evidenciando que con el transcurrir del tiempo ajustaron la logística de entrega de sus productos.

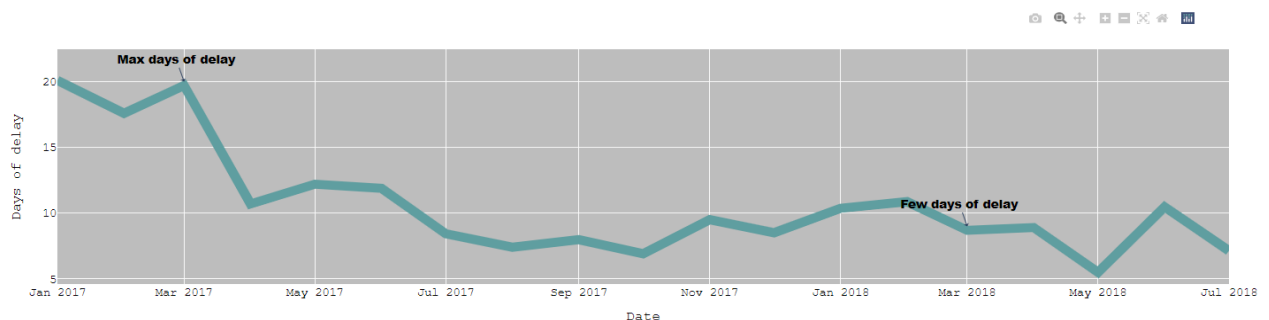
Cantidad de Órdenes con Delay

orders with delays



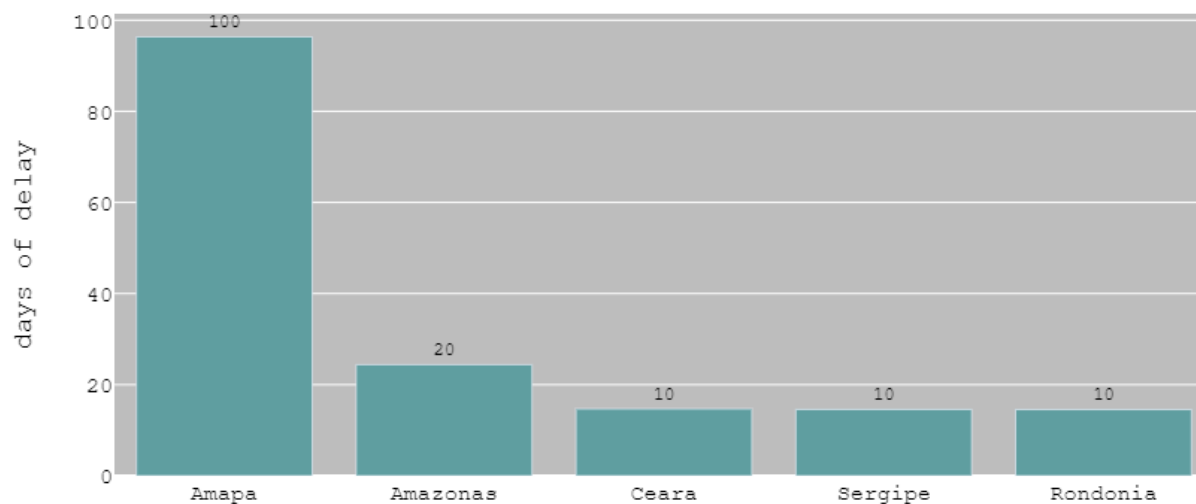
Cantidad de Dias de Delay

Days of delay



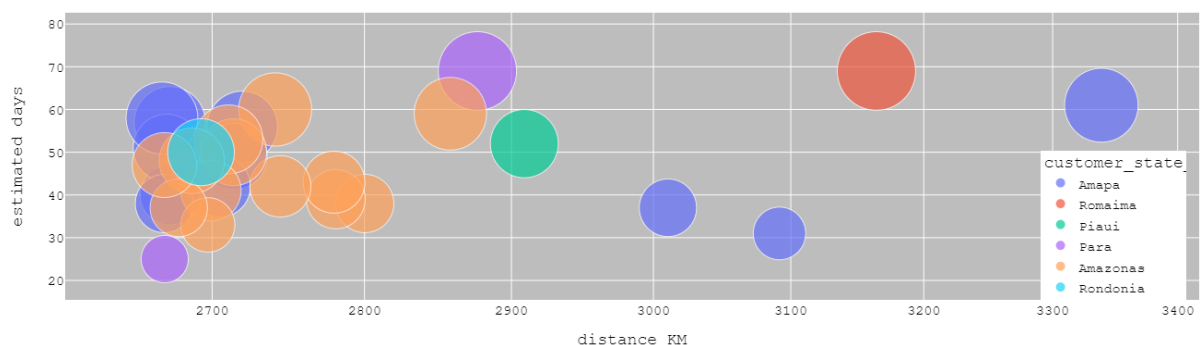
Actualmente los estados de los compradores con mayor cantidad de demora son los siguientes:

Mayor Demora Entre Compradores y Vendedores por Estado



Dichos compradores se encuentran a una distancia mayor de 2.000 km de los vendedores. Se puede afirmar que existe una estrecha relación entre la distancia entre compradores y vendedores, y la demora de entrega.

Días Estimados de Entrega Dependiendo de la Distancia



17- ANÁLISIS DE PRODUCTOS

Visualizando la distribución de las ventas podemos apreciar que el 63.8% de ellas se concentran en las siguientes categorías:

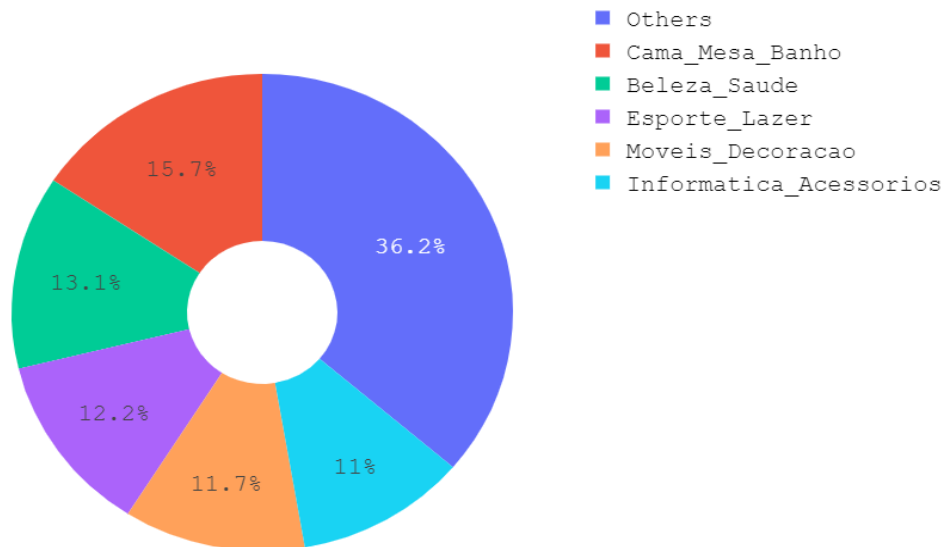
- Blanqueria (Cama Mesa Banho),
- Belleza y Salud (Beleza Saúde),
- Deporte y Ocio (Esporte Lazer),

- Muebles de decoración (Moveis Decoracao).

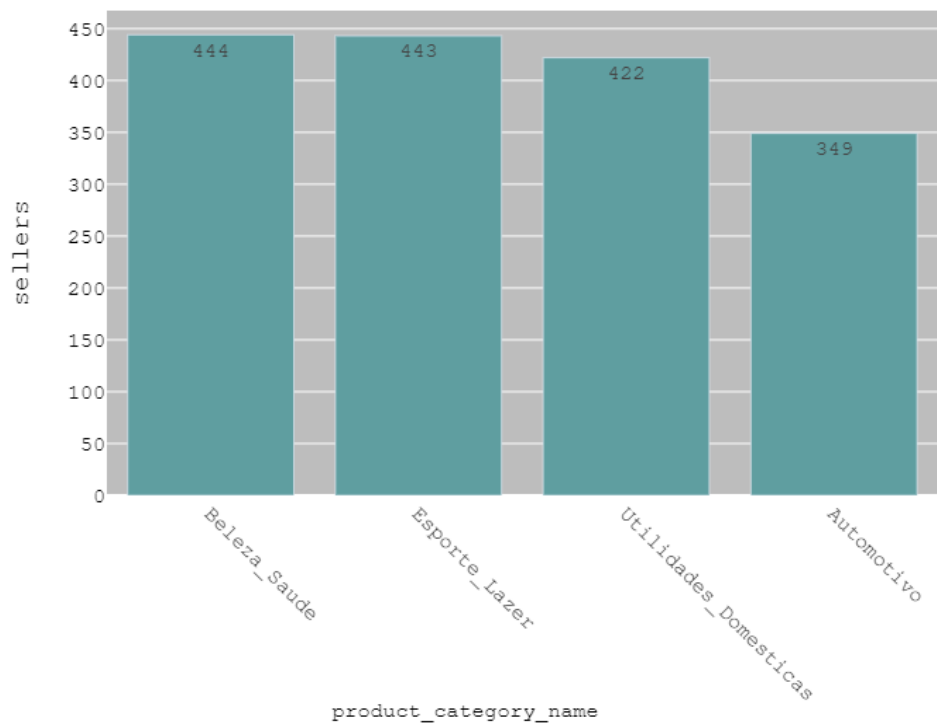
A su vez “Belleza y Salud” y “Deporte y Ocio” también pertenecen al listado de las categorías de productos con más vendedores .

Es recomendable, en un principio, implementar estrategias de marketing en dichas categorías.

Categoría de productos más vendidos



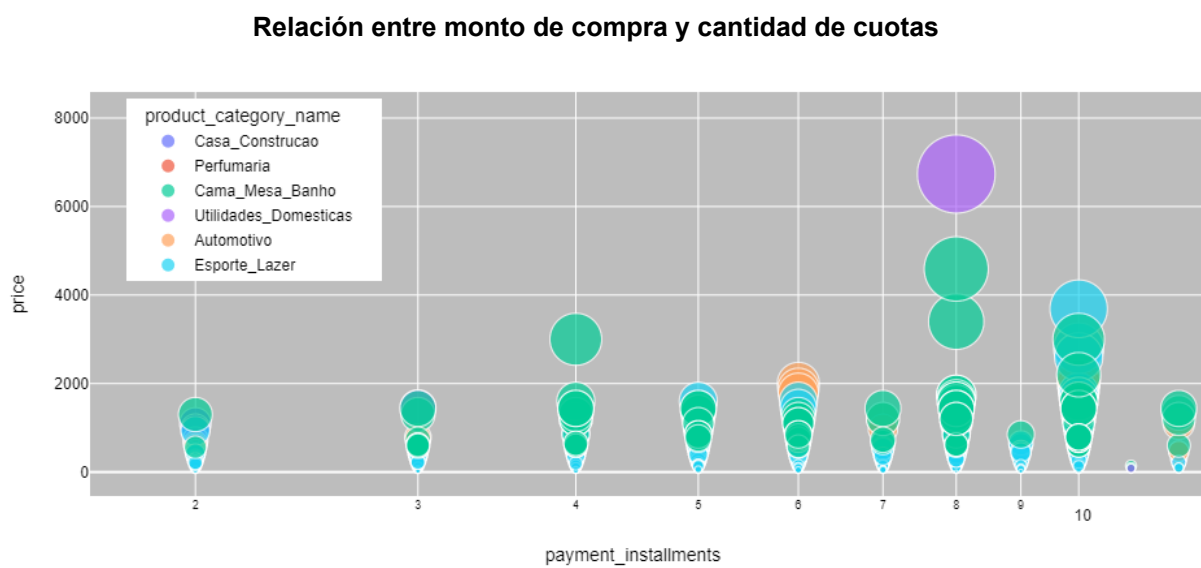
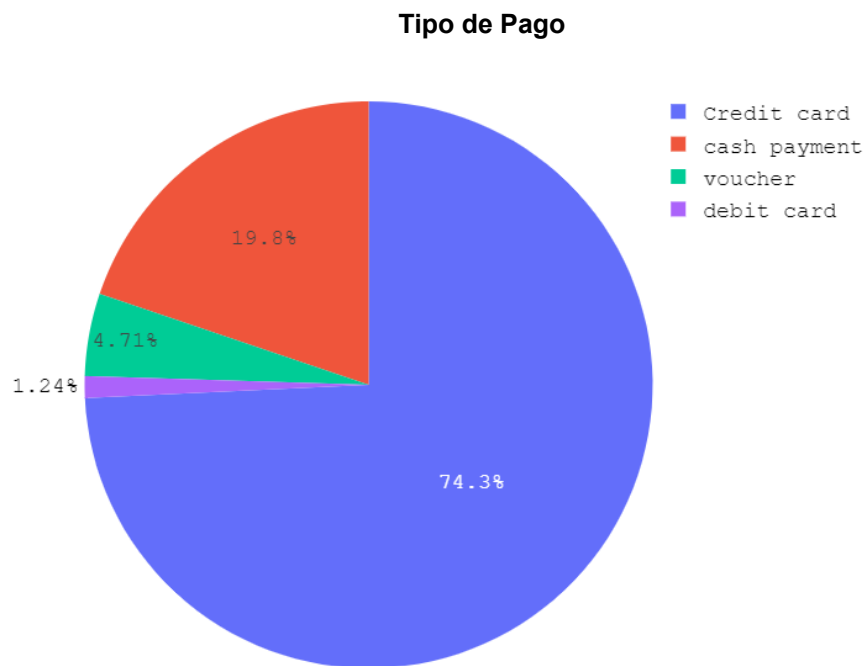
Cantidad de vendedores por categoría



18- ANÁLISIS DE MÉTODOS DE PAGOS

Podemos observar que el 74.3% de los clientes realizan sus compras con tarjeta de crédito, lo cual hace de este tipo de pago el más elegido por los compradores.

A pesar de que la mayoría de las transacciones se realizan en 1 cuota (49%), los productos con precio mayores a 150 R\$, se pagaron entre 8 y 10 cuotas.

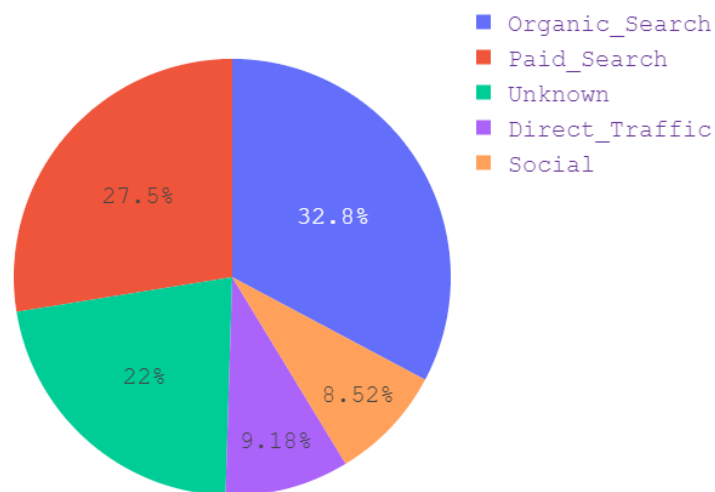


19- ANÁLISIS DE MARKETING

Podemos observar que la mayoría de los vendedores llegaron a las páginas de registro de Olist por medio del tráfico orgánico y tráfico pago.

Es recomendable lanzar una estrategia de marketing de generación de contenidos y SEO combinada con el pago de publicidad direccionada al segmento de vendedores de las mejores categorías.

Origen de los Leads

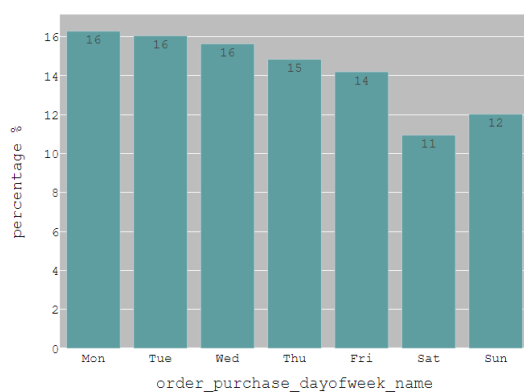


En el siguiente gráfico podemos observar los días y horarios con mayor y menor cantidad de ventas.

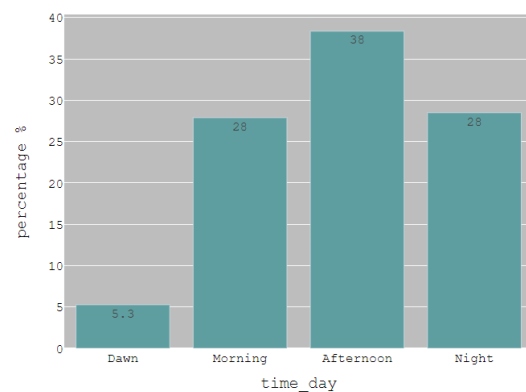
Es recomendable lanzar estrategias de marketing promocional para promover las ventas en los horarios y días con menores transacciones.

Días de la semana y horarios con más ventas

Sales by days of the week



Sales by hours

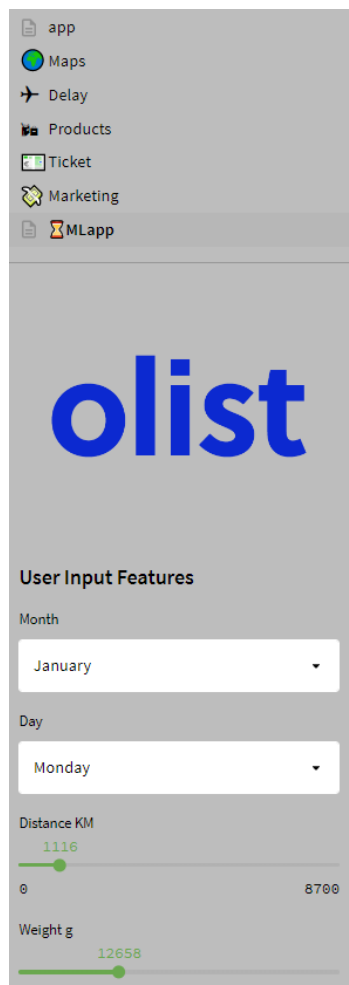


20- ALGORITMOS DE MACHINE LEARNING

Se desarrollaron dos algoritmos de Machine Learning.

Para *predecir el precio de envío*, se desarrolló un algoritmo de *regresión*, que tiene en cuenta 5 características del paquete, que debe ser ingresado por el usuario. La predicción tiene un error de 3-4 reales en el valor predicho.

Para *predecir si el envío va a llegar en el tiempo estimado o no*, se desarrolló un algoritmo de *clasificación* que tiene en cuenta 2 características para realizar la predicción.



Aplicación de Machine Learning



Shipping value prediction

Shipping price

RUN

	purchase_month	purchase_day_of_week	distance	product_volume_cm3	product_weight_g
0	1	1	1116	0	12658

Estimated shipping price

45,48 R\$

Order on time?

On Time