



23 DE JUNIO DE 2024

BUSINESS CASE - DATA ANALYST

BRAZILIAN E-COMMERCE PUBLIC DATASET BY OLIST

JOAQUIN VETTORAZZI

Business Case: Sellers Growth

This Business Case aims to assess the skills of a Data SSR Analyst to process and extract insights from a public Olist dataset available on Kaggle. The dataset includes information on orders, sellers, products, and order items.

Objectives

- Demonstrate the analyst's ability to clean, transform, and analyze complex datasets.
- Evidence the ability to extract valuable and actionable insights from data.
- Communicate findings clearly and concisely to a non-technical audience.

Data

[Olist_order_items_dataset](#)

[Olist_sellers_dataset](#)

[Olist_orders_dataset](#)

[Olist_products_dataset](#)

Guiding Questions

Identify top-performing sellers:

- Which sellers have experienced the highest sales growth?
- What are the characteristics of these top-performing sellers (e.g., location, product categories, average order value)?
- What strategies have these sellers implemented to achieve their growth?

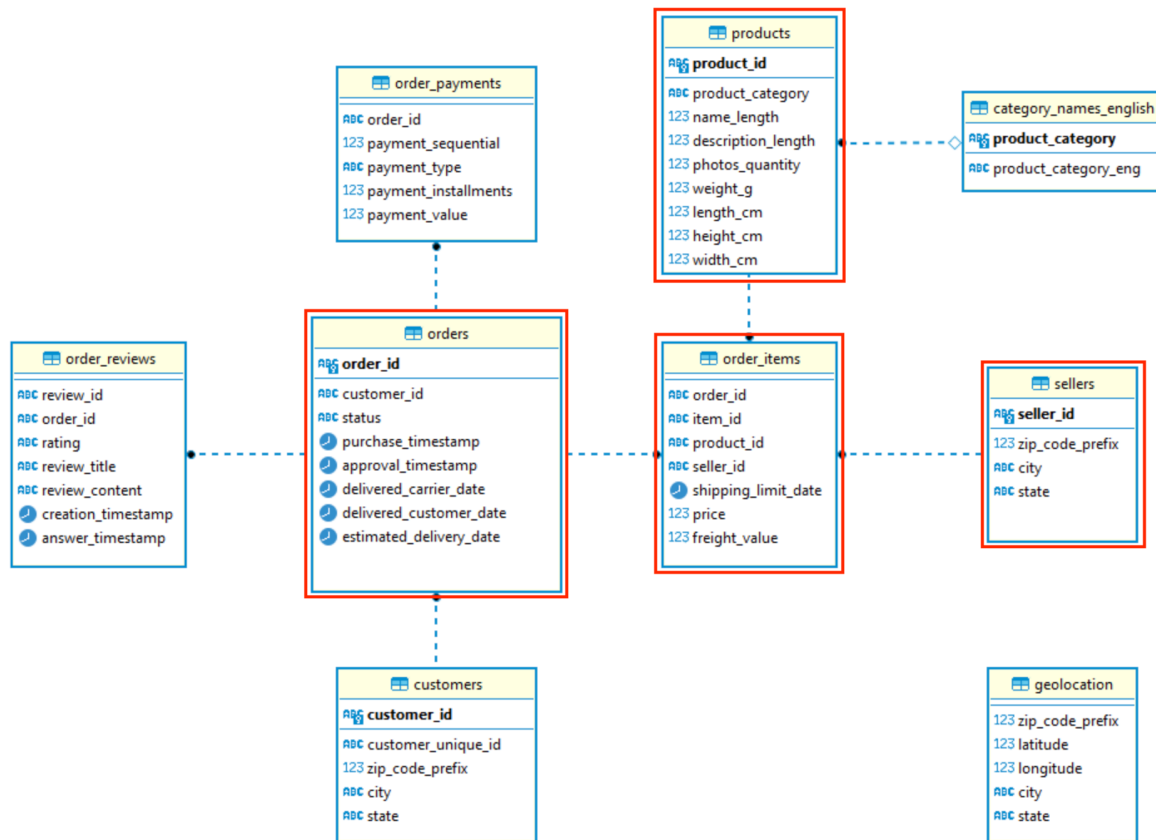
Analyze seller performance trends:

- How have sales trends for different sellers changed over time?
- Are there any seasonal or cyclical patterns in seller sales?
- What factors might be driving these trends and patterns?

Explore seller growth opportunities:

- What are the potential opportunities for sellers to grow their sales?
- What new products or categories could sellers add to their offerings?
- How can sellers improve their service efforts?

DATASET: Brazilian E-Commerce Public



Metodología utilizada

Para abordar el análisis y extracción de insights del dataset de Olist, se emplearon herramientas y técnicas de análisis de datos. A continuación, se detallan los pasos seguidos:

Para ver este apartado en profundidad se recomienda ver script de código en el siguiente link:

<https://colab.research.google.com/drive/1rYq2kbq6dukEvyEmrlr49bBs2DwXyl1D?usp=sharing>

Entorno de Trabajo:

- Se utilizó Google Colab como entorno de desarrollo para la escritura y ejecución de scripts en Python.

Carga de Datos:

- Los cuatro datasets proporcionados (olist_order_items_dataset.csv, olist_sellers_dataset.csv, olist_orders_dataset.csv, y olist_products_dataset.csv) fueron cargados utilizando la librería Pandas.

Limpieza y Transformación de Datos:

- Limpieza: Se manejaron valores nulos, se eliminaron duplicados y se corrigieron formatos de datos incorrectos.
- Transformación: Se crearon nuevas columnas necesarias para el análisis, como la duración de la entrega y el retraso en la entrega. Se utilizó la librería datetime para manejar y transformar los datos de fechas y horas.
- Agregación: Se agregaron datos por diferentes dimensiones (vendedor, año, mes) para facilitar el análisis posterior.

Análisis Exploratorio de Datos (EDA):

- Se emplearon las librerías matplotlib y seaborn para la visualización de datos, lo que permitió identificar patrones, tendencias y outliers.
- Se realizaron análisis descriptivos para obtener estadísticas resumidas y visualizaciones clave, como histogramas de tiempos de retraso y gráficos de barras de ventas por vendedor.

Exportación de Datos:

- Los datos transformados y consolidados fueron exportados a archivos CSV, listos para ser importados en Power BI.
-

Visualización en Power BI:

- El archivo CSV exportado fue importado en Power BI.
- Se diseñaron y desarrollaron dashboards interactivos que permiten explorar la información de manera dinámica y comprender mejor el desempeño de los vendedores, las tendencias de ventas y las oportunidades de crecimiento.
- Los dashboards incluyen visualizaciones como gráficos de líneas para tendencias de ventas, gráficos de barras para comparaciones entre vendedores y mapas geográficos para analizar la distribución regional de los vendedores. Se detalla más sobre este punto a continuación.

Data | Brazilian E-Commerce Public Dataset by Olist

El dataset con el que se trabajó contiene información de 100,000 pedidos realizados en el periodo 2016-2018 en múltiples marketplaces de Brasil. Sus características permiten

visualizar un pedido desde muchas dimensiones, en este caso, nos centraremos en 4 datasets:

- [Olist_order_items_dataset](#)
- [Olist_sellers_dataset](#)
- [Olist_orders_dataset](#)
- [Olist_products_dataset](#)

Insights – Dashboard Power BI

Para poder realizar un análisis completo, creamos un archivo en Power bi que contiene 4 pestañas:

- 1) Ventas y sellers en general
- 2) Órdenes y tendencias
- 3) Panorama Geográfico
- 4) Productos y categorías

Para esta parte del documento se recomienda ir visualizando los Dashboards en Power BI.

Ventas

En la sección Ventas se puede observar toda la información relevante a los sellers y su performance. A la izquierda, en el gráfico de barras, se encuentran los top sellers del mercado brasileiro de e-commerce.

Si seleccionamos cualquiera de los ‘top sellers’ podremos ver información de las órdenes correspondientes al seller, en la tabla de la derecha gráfico. Por ejemplo, si seleccionamos el mejor seller, podremos observar en la tabla de la derecha todas las órdenes que realizó, el estado, la ciudad, la categoría de productos de cada orden y su status. A su vez, de ese mismo vendedor podremos ver cuanto vendió en total, cuál fue el precio final en fletes, cuantas órdenes realizaron sus clientes y cuál es el precio promedio de todas sus órdenes. En caso de no seleccionar un vendedor, los indicadores tomarán en cuenta a todos los sellers.

En este dashobard interactivo, al igual que todos los demás, podremos analizar varios elementos a la vez, ya que cada visualización está relacionada con las demás y va cambiando según nuestra selección. Por ejemplo, podemos seleccionar un año o un rango de fecha específico y analizar todas las métricas del dashboard únicamente para ese período específico. Lo mismo para analizar todas las métricas para una categoría específica o para un estado o ciudad en particular.

Por lo que podemos analizar en este tablero, en los 3 años de operación, considerando “price” como lo que pagó el cliente en cada orden, hubo ventas con un valor de **13.4 millones**.

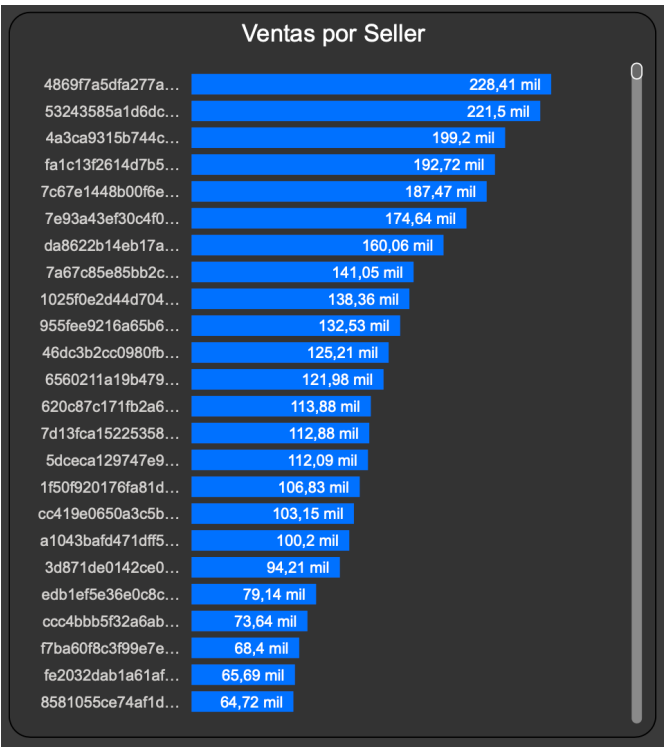
En cuanto al **precio del flete**, se le cobró al cliente un total de **2.2 millones**. Se realizaron **97.700 órdenes** con un **precio promedio** de **120.4**.

Se puede observar, además, las ventas por cada año independiente. En el **2018** fue el año que más se vendió, con un valor de **7.74 millones**, lo que corresponde a un **73% más** comparado con **2017**. En el año **2016** se registraron ventas de **41.000**, un monto casi despreciable comparado con los años siguientes.



Top Sellers

En el tablero creado, se podrá observar los 'top sellers' de la compañía a lo largo de los 3 años analizados y la tabla con el detalle de cada orden:



Seller	Ciudad	Estado	Categoría de producto	Status
fffe41c64501cc87c801fd61db3f6244	ibitinga	SP	cama_mesa_banho	delivered
fffe18544ffabc95dfada21779c9644f	sao paulo	SP	informatica_acessorios	delivered
fffce4705a9662cd70adb13d4a31832d	curitiba	PR	esporte_lazer	delivered
ffcd46ef2263f404302a634eb57f7eb	sao paulo	SP	informatica_acessorios	delivered
fffc94f6ce00a00581880bf54a75a037	itajai	SC	utilidades_domesticas	delivered
fffb3b5462987e66fb49b1c5411df2	salto	SP	casa_construcao	delivered
fffb9224b6fc7c43ebb0904318b10b5f	sao goncalo	RJ	relogios_presentes	delivered
fffb2ef8874127f75b52b643880fd7e0	brusque	SC	cama_mesa_banho	delivered

Los vendedores que han logrado posicionarse como los mejores del mercado brasileño de e-commerce han adoptado varias estrategias clave:

- **Diversificación de productos:** Los vendedores que ofrecen una amplia gama de productos pueden atraer a diversos perfiles de compradores y abarcar una mayor parte del mercado. La variedad de productos fomenta el cross-selling y el up-selling dentro del mismo negocio. Al estructurar su oferta con la venta cruzada en mente, pueden aumentar el ticket medio por orden, maximizando así sus ingresos.
- **Popularidad del sector o categoría de productos:** Los vendedores que se especializan en nichos muy específicos pueden encontrar una barrera al intentar alcanzar grandes volúmenes de ventas. En cambio, aquellos que venden productos de uso general y alta demanda en el mercado brasileño, como cosméticos, belleza y maquillaje, tienen la posibilidad de vender en grandes cantidades. En el mercado brasileño, estas categorías son populares y pueden ser un derivar en un volumen significativo de ventas.
- **Campañas fuertes para fechas importantes:** Los vendedores que preparan campañas personalizadas para fechas clave como Black Friday, Navidad, el día de la Madre, etc., pueden maximizar el consumo de su público objetivo. Estas campañas, si se ejecutan correctamente, pueden aumentar significativamente las ventas durante estos periodos de alta demanda. Este punto siempre se ve potenciado por buen manejo del marketing digital con microsegmentación por producto y público objetivo.
- **Optimización del tiempo de entrega:** La recurrencia de clientes es vital para cualquier negocio. Para asegurar que un cliente vuelva, es crucial que el servicio y el producto entregados cumplan o superen las expectativas. Un pilar importante del servicio es el tiempo de entrega. Si los productos se entregan en el tiempo prometido o antes, la satisfacción del cliente aumenta, lo que incrementa la probabilidad de que vuelvan a comprar.

ORDENES Y TENDENCIAS

En la pestaña de “ORDENES Y TENDENCIAS” se puede observar lo siguiente:

Un gráfico que muestra cuales son los vendedores que más órdenes recibieron y cuál fue el número total de órdenes recibidas. En este caso, el seller que más recibió órdenes lo hizo con un total de 2.013 órdenes.

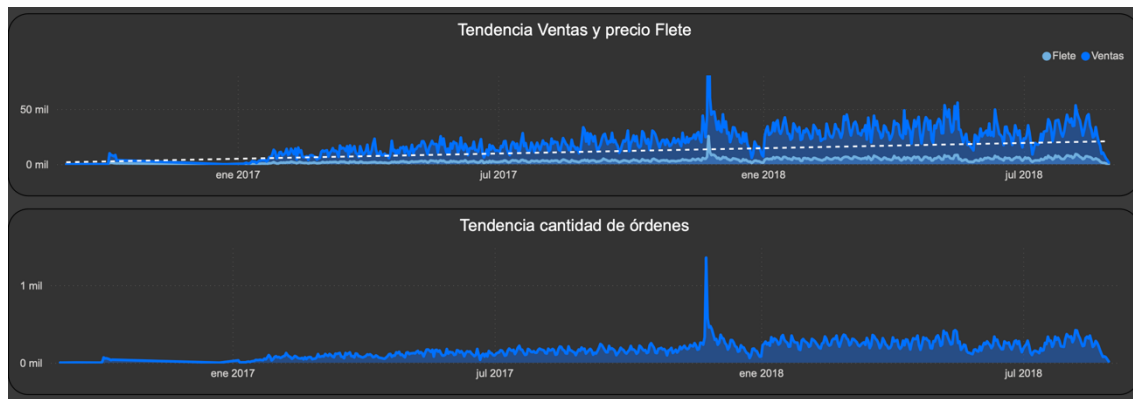


A su vez, en el dashboard se pueden observar 3 tarjetas:

- 1) El tiempo medio estimado de delivery en horas. En este caso el **tiempo promedio estimado** es de **572,4 horas**, es decir **23,8 días**. Creo que este número al día de hoy (2024) debería de estar muy por debajo, ya que el mundo del delivery, logística, e-commerce creció exponencialmente estos últimos años (post pandemia).
- 2) El **tiempo medio real** de delivery en horas. Este indicador marca un valor de **360,1 horas**, es decir, **15 días**. Si lo comparamos con el estimado, el tiempo real de entrega fue menor que el tiempo estimado.
- 3) El tiempo de entrega anticipada. Esto indica el tiempo en horas que entregamos de manera anticipada un producto. En este caso, en los 3 años analizados, el tiempo anticipado fue de **212,2 horas**. Entregar un producto antes de lo esperado puede derivar en una satisfacción aún más alta por parte del cliente.



En el mismo dashboard, se pueden observar dos gráficos que muestran la tendencia del monto de ventas, el precio del flete y la tendencia en cantidad de órdenes, según el momento en el que cliente realiza la compra. Aunque la correlación es muy alta, los gráficos de ventas y de cantidad de órdenes no necesariamente deben coincidir. Esto se debe a que, con la misma cantidad de órdenes, podemos tener precios diferentes, lo que resulta en montos de ventas distintos.



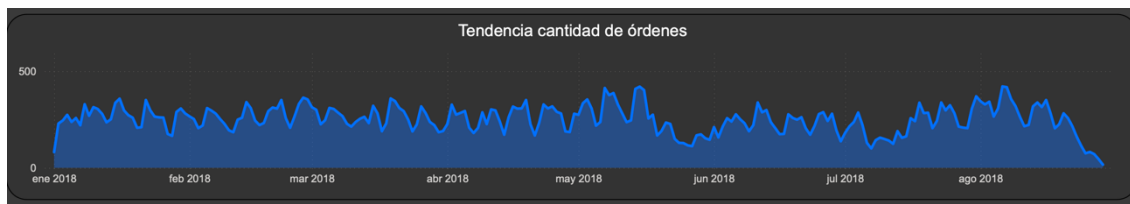
En ambos gráficos podemos observar una tendencia al crecimiento a lo largo de los años. Esto puede atribuirse a varios factores clave. El primero, y posiblemente el más importante, es el cambio en el comportamiento del consumidor. Cada vez más personas buscan la comodidad y la eficiencia en sus compras, lo que ha llevado a la industria del e-commerce a desarrollar nuevas capacidades para satisfacer la demanda de clientes cada vez más exigentes.

Con el cambio generacional, los consumidores han ganado mayor confianza en las compras online, que hoy en día son preferidas por un gran porcentaje de clientes de diversas marcas. Además, a nivel de competencia, las marcas se han visto obligadas a desarrollar sus canales digitales. No hacerlo podría significar perder una parte significativa del mercado.

En el mismo dashboard, podemos seleccionar alguno de los vendedores para analizar la tendencia de sus ventas y órdenes. También podemos filtrarlo por un rango de fechas específico, por la categoría de productos o por el estado (ubicación). A modo de ejemplo, podríamos identificar cuál es el seller que más órdenes recibió en 2018, en la categoría “alimentos_bebidas” en el estado Minas Gerais (MG). Seleccionando al seller, también podríamos ver el total de órdenes que recibió, cómo fue la tendencia en cantidad de órdenes y ventas, y los tiempos promedios de entrega.

Para analizar si se observa un patrón estacional o cíclico en los gráficos, se puede filtrar de forma anual y observar a año completo el comportamiento de las ventas.

A simple vista, no se puede observar un patrón estacional, pero si cíclico. Por ejemplo, en 2018 vemos lo siguiente:



Comportamiento cíclico visto dentro de un mes aleatorio:



Se puede ver la tendencia de todo 2018, como va fluctuando de manera cíclica, formando 4 o 5 picos por mes. Para entender un poco más en profundidad a que se podría deber este comportamiento, filtramos de manera mensual aleatoriamente (lo hacemos con varios meses). Al hacerlo de esta forma, observamos que las ventas suben a principio y mediados de semana y descienden los fines de semanas. Este comportamiento se puede dar por varios factores:

1. Los fines de semana las personas dedican su tiempo a actividades personales, familiares y generalmente fuera de casa. La mayoría de las veces cuando compramos online lo hacemos desde la comodidad de nuestra casa, ya que preferimos hacerlo con tiempo, tranquilos y en lo posible desde una computadora (al menos en 2018). Ahora probablemente esto haya cambiado.
2. Al comienzo de la semana, las personas suelen estar más organizadas y planificadas, lo que incluye la gestión de sus presupuestos. Esto puede llevar a que realicen más compras en los primeros días de la semana.
3. Las tiendas en línea muchas veces lanzan promociones y ofertas al comienzo de la semana para atraer a los consumidores. Estas ofertas pueden impulsar un aumento de ventas.
4. Algunas personas prefieren realizar compras en días laborables cuando la logística y el soporte al cliente son más accesibles, evitando así cualquier inconveniente que pueda surgir durante el fin de semana.

Fechas importantes puntuales

Por otro lado, se pueden identificar algunos hitos importantes en términos de cantidad de órdenes y monto generado por ventas durante el año:

- **Navidad (24 de diciembre):** Se observa un incremento considerable en las semanas previas a esta fecha, ya que es una época en la que se hacen muchos regalos. Los consumidores realizan sus compras con anticipación para tener los regalos listos para Navidad.
- **Black Friday (último viernes de noviembre):** Este es el pico más significativo del año. Las marcas ofrecen descuentos y promociones únicas que incentivan a los

consumidores a comprar. Además, muchas personas aprovechan este día para hacer sus compras navideñas.

- **Día de la madre (segundo domingo de mayo):** Durante las dos primeras semanas de mayo, se observan picos muy altos de ventas. Las personas compran con anticipación para tener listo el regalo para el Día de la Madre.
- **Día del padre (segundo domingo de agosto):** La primera semana de agosto muestra un pico pronunciado en las ventas, ya que los consumidores se preparan para comprar regalos para el Día del padre.
- **Día del niño (12 de octubre):** Aunque el incremento de ventas no es tan acentuado cuando se observan todas las categorías, en la categoría de "brinquedos" se ve claramente un pico durante la primera semana de octubre.
- **Cyber Monday (lunes después de Black Friday):** Este día presenta un comportamiento similar al de Black Friday, con descuentos y promociones que incentivan las compras online.

PANORAMA GEOGRAFICO

En este tablero se puede observar lo siguiente:

Cuáles son los estados con los montos de ventas más altos y la relación proporcional entre ellos. De la misma forma se puede observar el mismo indicador, pero por ciudad.



Con respecto a estos embudos, se puede observar que el estado que más vendió fue **San Pablo (SP)** con **8,6 millones**, seguido de **Estado de Paraná (PR)** con **1,2 millones** y tercero **Minas Gerais (MG)** con **997.000**.

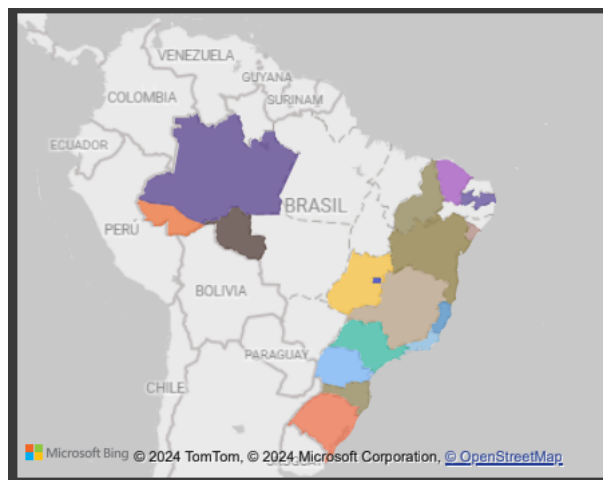
En cuanto a ciudades, podemos identificar a la ciudad de **San Pablo** como la que más vendió con **2.6 millones**. Segundo viene **Ibitinga** con **620.000** y tercero **Curitiba** con **468.000**.

Por debajo de cada gráfico de embudo, se puede ver un Treemap que es útil para visualizar el volumen de órdenes recibidas por Estado (**top 5 estados**) y el volumen de órdenes recibidas por ciudad (**top 5 ciudades**).



Si bien **San Pablo** predomina en **cantidad de órdenes**, podemos ver que **Estado de Paraná** a pesar de ser el segundo en monto de ventas, se posiciona como tercero en volumen de órdenes, quedando **Minas Gerais** en segundo puesto. En cuanto a las ciudades, se respeta el mismo orden que en montos de ventas de las top 3, aunque podemos ver que **Rio de Janeiro** y **Guarulhos** están en el top 5 en monto generado por ventas, pero no así en órdenes recibidas. Esto se debe a que ambas tienen un Ticket medio de orden bastante alto, ya que, con menos órdenes, generan más dinero.

Por último, en el cuadrante inferior izquierdo del dashboard, se puede ver la cantidad total de ciudades y estados. Estos números son relevantes a la hora de expandir el negocio a otras regiones dentro de Brasil. Justo por debajo se encuentra el mapa que localiza las operaciones geográficamente.



Para aprovechar al máximo el dashboard, podemos filtrar las ventas por año, y seleccionar, por ejemplo, un estado. En ese caso, podremos ver los mejores Sellers en ese estado, en qué ciudades se encuentran, la cantidad de ciudades que tiene dicho estado, dónde se ubica en el mapa y los volúmenes de órdenes que se recibieron.

PRODUCTOS Y CATEGORIAS

El dashboard de “PRODUCTOS Y CATEGORIAS” permite hacer un análisis de las categorías y los productos que componen el mercado analizado.

Al igual que con la ubicación geográfica, podemos identificar las categorías que más ventas generaron y las categorías que más órdenes recibieron.



En cuanto al monto generado por ventas, podemos identificar a la categoría **“beleza_saude”** como la que más generó dinero con un valor de **1.2 millones**. Esto es de esperarse, ya que Brasil es uno de los países que más se consume el maquillaje y productos de belleza. A diferencia de otros mercados, principalmente mujeres, a estos productos se le da un uso diario. En segundo puesto, viene la categoría de **relojes** y tercero, productos para el hogar (**“cama_mesa_banho”**).



Si nos recibimos a la cantidad de órdenes recibidas, se comporta de manera similar. El primer puesto lo encabeza **“cama_mesa_banho”**, lo sigue **“beleza_saude”** y tercero lo

ocupa “**esporte_lazer**”. Esta última categoría pertenecería a lo que sería deportes y ocio. Tiene mucho sentido, ya que en Brasil se practica mucho deporte al aire libre, considerando que tienen un clima tropical con temperaturas ideales para salir de casa.

Debajo de los gráficos de categorías, encontramos una tabla que indica los productos más vendidos en función del monto generado por ventas y cantidad de órdenes. Si escogemos una categoría en particular, podremos ver en esa tabla los productos más y menos vendidos por cada categoría. También se puede observar los productos más vendidos por año, por ciudad y por estado.

Productos más vendidos		
Producto	Ventas	Cantidad órdenes
bb50f2e236e5eea0100680137654686c	63.885	187
99a4788cb24856965c36a24e339b6058	42.759	464
25c38557cf793876c5abdd5931f922db	38.907	38
5f504b3a1c75b73d6151be81eb05bdc9	37.734	63
aca2eb7d00ea1a7b8ebd4e68314663af	37.249	427
53b36df67ebb7c41585e8d54d6772e08	37.098	301
e0d64dcfaa3b6db5c54ca298ae101d05	31.787	194
d285360f29ac7fd97640bf0baef03de0	31.100	120
7a10781637204d8d10485c71a6108a2e	30.249	141

En la parte inferior izquierda del gráfico podemos ver tiempo medio estimado de delivery y tiempo medio real de delivery. Si bien son valores que ya vimos en un dashboard anterior, esta vez podemos ver como varían estos números en función de la categoría del producto y los productos en sí mismos.

A modo de ejemplo, si seleccionamos el estado de San Pablo y la ciudad de “Bauru”, podemos observar que la categoría que más genera por ventas (12.000) y más órdenes recibe (56) es “automotivo”. A su vez, seleccionando la categoría, en la tabla de abajo se puede identificar rápidamente los productos que más vendieron en cantidad y en monto de la categoría “automotivo”. Por último, podemos ver los tiempos de delivery de la categoría, y si se quisiera, de un producto en particular.

Oportunidades de crecimiento

- **Expansión de mercado:** Explorar nuevas regiones dentro de Brasil y otros mercados internacionales puede abrir nuevas oportunidades de ventas. Adaptar las estrategias de marketing a las características locales de cada nuevo mercado es esencial para el éxito.
- **Mejora en marketing digital:** Invertir en campañas de marketing digital más efectivas, incluyendo SEO, SEM y publicidad en redes sociales, puede atraer a más clientes potenciales. Las campañas dirigidas y personalizadas pueden mejorar la conversión y fidelización de clientes.
- **Mejora la experiencia del cliente para lograr fidelizarlos:** Entregar el producto es solo una parte de la experiencia de la experiencia de compra. Para lograr que la satisfacción del cliente sea muy alta, los sellers deben estar en todos los

detalles: El marketplace, el proceso de checkout lo más rápido posible, los medios de pago que ofrecen, la velocidad con la que pueden comprar los clientes, el delivery del producto en tiempo y forma, los correos de marketing para incentivar la recurrencia, estrategias para fomentar cross-selling y aumentar el ticket medio por orden, etc. Es necesario que cada seller tenga como pilares de su negocio la conversión, recurrencia y ticket medio para llevar su e-commerce al siguiente nivel.

- **Expandir la gama de productos:** Sin alejarse demasiado del core de su negocio, los vendedores pueden considerar la incorporación de productos complementarios para atraer a un perfil diverso de compradores y fortalecer su imagen de marca. Aunque todos los e-commerce tienen productos estrella, es estratégico añadir artículos básicos y esenciales que incentiven a los clientes a visitar y recurrir a la tienda online con mayor frecuencia.

Optimización del servicio de los vendedores

1. **Optimización del servicio de atención al cliente:** Capacitar al equipo de atención al cliente para responder rápidamente y de manera efectiva a las consultas y problemas puede mejorar significativamente la satisfacción del cliente.
2. **Mejora en la logística y entrega:** Implementar sistemas de gestión de inventario más eficientes y colaborar con servicios de entrega confiables puede asegurar que los productos lleguen a tiempo, mejorando así la experiencia del cliente.
3. **Políticas de devolución y reembolso claras:** Ofrecer políticas de devolución y reembolso claras y justas puede aumentar la confianza del cliente y hacer que se sientan más seguros al realizar compras. Esto impacta en la conversión.
4. **Programas de feedback y mejora continua:** Incluir opciones de pagos flexibles con bajos intereses. Por ejemplo, darle la opción al cliente que pueda comprar los productos en cuotas. Esto llevará a que el cliente esté dispuesto a desembolsar una suma mayor de dinero, por lo que el ticket medio por orden aumentará.
5. **Personalización del servicio:** Utilizar datos de los clientes para personalizar las ofertas y recomendaciones puede hacer que los clientes se sientan más valorados y aumentar la probabilidad de compra.