

Análisis de datos

Recordemos el planteamiento del problema: Una empresa de E-Commerce de Argentina, está evaluando la posibilidad de expandirse al Brasil, y para ello necesita entender como es el mercado de E-Commerce de allá.

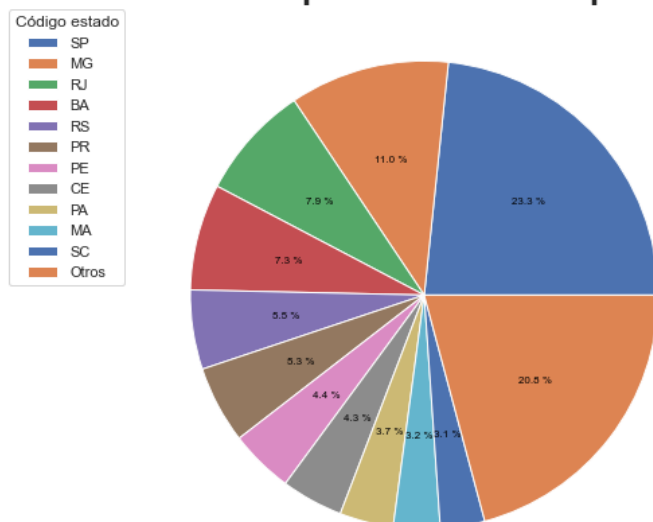
La información con la que contamos es con Olist, líder de E-Commerce en Brasil.

Entendamos la distribución de Brasil

Según datos extraídos por el Instituto Brasileño de Geografía y Estadística, la población de Brasil se distribuye de la siguiente manera



Distribución poblacional por estado



Más o menos, la mitad de la población se encuentra contenida en los 4 estados con mayor población (São Paulo, Minas Gerais, Río de Janeiro, Bahía).

Podría pensarse en realizar la expansión a Brasil, comenzando con estos 4 estados; sin embargo, aún no es suficiente la información para realizar dicha sugerencia.

Entendamos el mercado

En este punto haremos uso de KPI's propios generados con base en los datos proporcionados para entender al negocio (Olist) y cómo es su operación en Brasil. No vale la pena hacer una expansión si el mercado está totalmente dominado y saturado por la principal competencia.

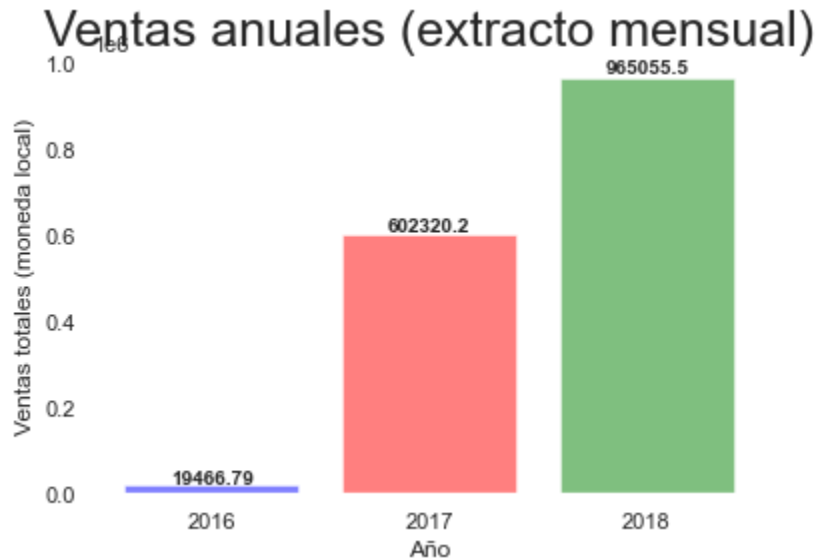
Recordemos lo que es un KPI.

El término KPI, por sus siglas en inglés, significa Key Performance Indicator; es decir, Indicador Clave de Desempeño o Medidor de Desempeño, en otras palabras, nos dan una medida clave sobre algún aspecto del negocio que se traduce en una referencia confiable de las áreas de oportunidad y poder definir una acción futura.

A continuación, veremos el mercado que cubre Olist



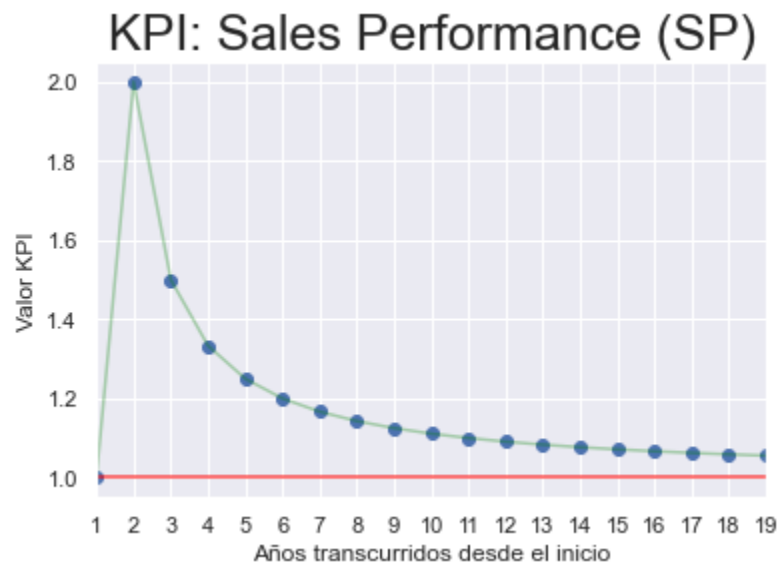
Las ventas anuales han ido en aumento; sin embargo, en la información obtenida, no contamos con todos los meses de todos los años, por lo que se hará el extracto promedio anual mensual para una mejor representatividad.



En promedio, cada mes de cada año, las ventas totales han ido en aumento

Definamos un KPI que nos indique, de manera anual, si realmente existe un crecimiento o no, lo definiremos como 'Sales Performance (SP)'

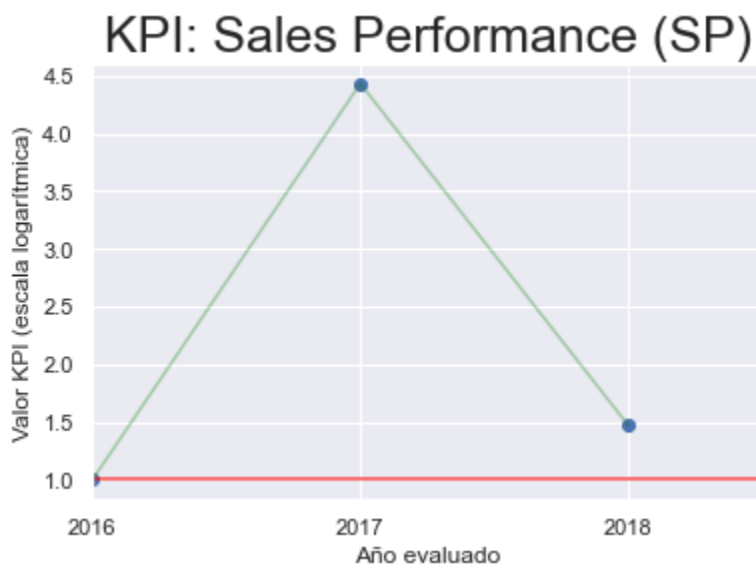
$$KPI_{SP}(year) = \frac{\text{Total Sale (monthly extract)}}{\text{Previous year Total Sale (monthly extract)}}$$



Observamos que asintóticamente, el valor del KPI, tiende a 1 y tiene una subida muy pronunciada. La información que nos brinda el KPI es que, teniendo un crecimiento anual, mensual promedio,

lineal eventualmente tenderemos a saturar el mercado (las compras del año anterior, en dinero, son iguales que las del presente año y serán iguales, en dinero, a las del próximo año).

Veamos qué sucede con Olist



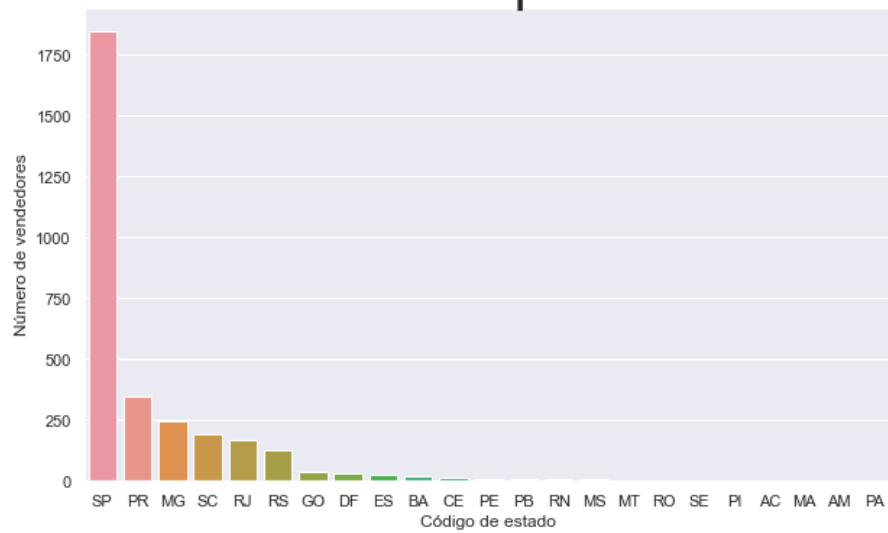
Parece muy estrepitosa la gráfica y podría pensarse que ya no es tiempo de invertir en Brasil, por el contrario, la gráfica nos indica que tuvo un crecimiento brutal el cual se va a desacelerar y tomará varios años para que sature el mercado; es decir, confirmamos que el mercado es lo suficientemente joven para aceptar un nuevo E-Commerce.

Veamos la distribución de clientes(compradores) y vendedores

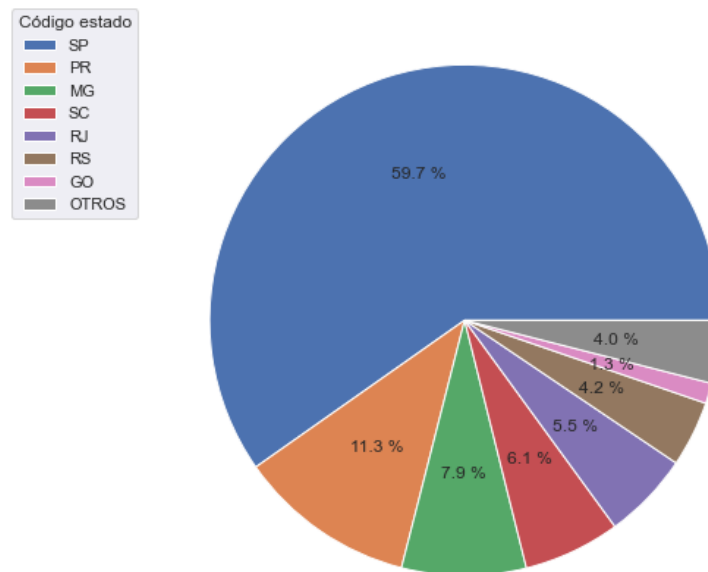
Vendedores

Según los datos proporcionados, los vendedores se distribuyen de la siguiente manera.

Vendedores por estado



Y su representatividad sobre el país

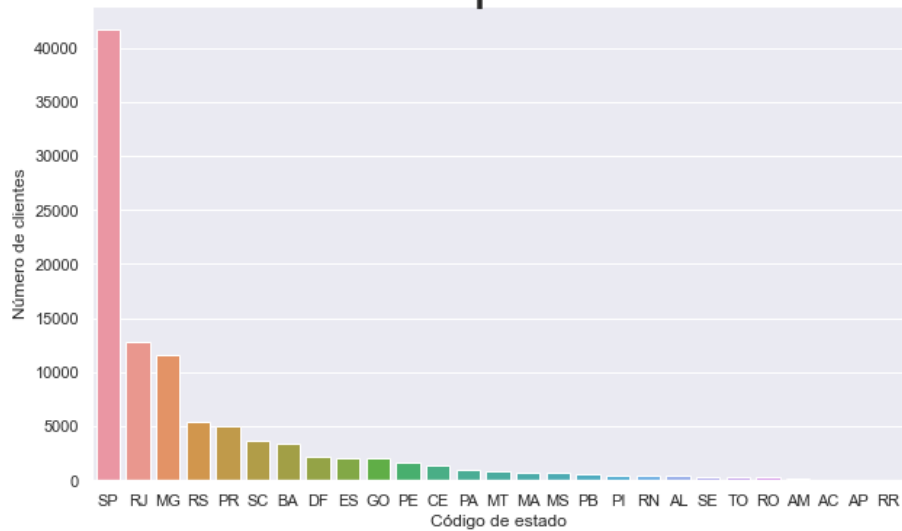


Como lo pensamos en un inicio, la mayoría de vendedores se encuentran en los estados con mayor población

Compradores/clientes

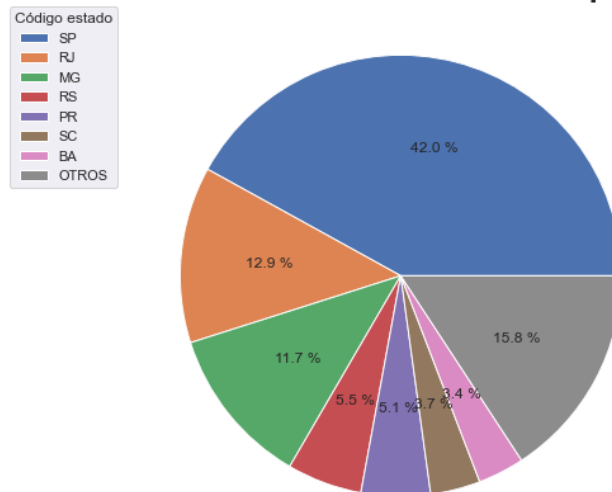
Los compradores/clientes se distribuyen de la siguiente manera

Cientes por estado



Y su distribución comparando con los demás estados

Distribución de los clientes por estado



Como era de esperarse, los estados con mayor cantidad de clientes se encuentran en el Top de los estados más poblados.

Punto importante: A pesar de que Santa Catarina es el 11° estado con mayor cantidad de población está en el Top 4 de vendedores y Top 6 de clientes, más adelante podremos concluir algo más

Dado que la población de Brasil es extensa, vale la pena pensar quiénes son los clientes potenciales a los que podemos llegar. Según la CIA ([CIA World Factbook](#)) Los rangos etarios en Brasil son:

- **0-14 años:** 21.89%
- **15-24 años:** 16.29%
- **25-54 años:** 43.86%
- **55-64 años:** 9.35%

- **65 años y más:** 8.61%

Nuestro rango de alcance para clientes lo definiremos de 15 a 54 años, que sería un 60.15% del total de la población y para vendedores de 25 a 54 años, que sería un 43.86%.

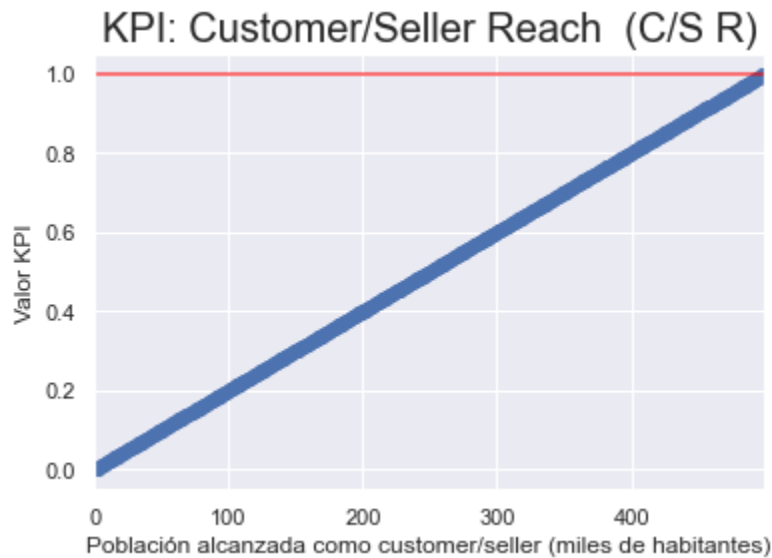
Definamos 2 nuevos KPI:

- Customer Reach (CR)

$$KPI_{CR}(state) = \frac{Total\ customers\ (state)}{Customer\ targetable\ population\ (state)}$$

- Seller Reach (SR)

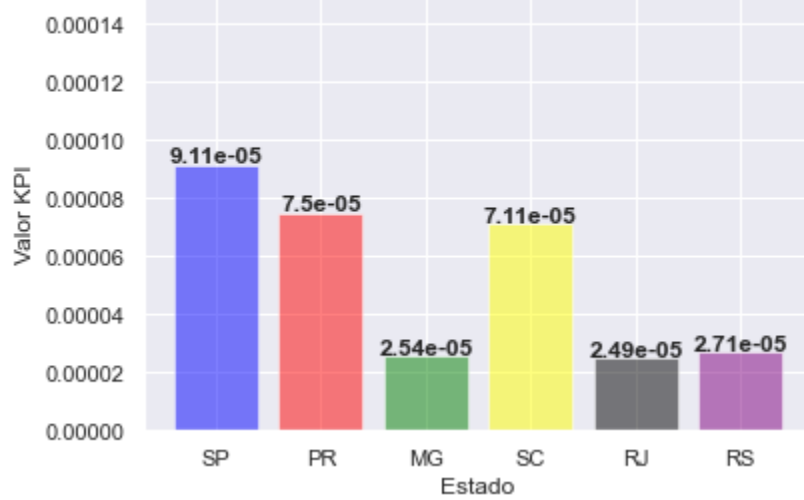
$$KPI_{SR}(state) = \frac{Total\ sellers\ (state)}{Seller\ targetable\ population\ (state)}$$



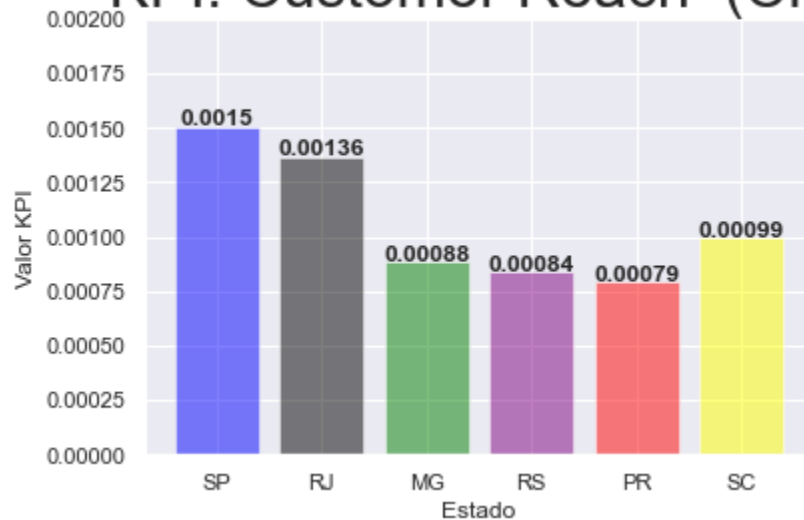
El objetivo del KPI es brindarnos una referencia acerca de la disponibilidad que tiene cierta población para aceptar o no nuevos vendedores/compradores. El objetivo será llevar el valor del KPI lo más cercano a 1 donde, de llegar, nuestro público objetivo habría sido totalmente alcanzado.

Veamos en qué punto se encuentra Olist, por lo antes mencionado tomaremos el Top 6 de los estados

KPI: Seller Reach (SR)



KPI: Customer Reach (CR)



Lo primero a destacar es que Brasil es un mercado con un gran potencial de crecimiento debido a los órdenes de los valores de los KPI son de una magnitud 3 o mayor.

Un caso importante es lo que sucede en el estado SC (Santa Catarina) por su baja población (comparada con el Top 5) no se hubiera esperado encontrarlo dentro del Top de KPI C/S R; sin embargo, aparece. Nos indica que la población tiene las capacidades tecnológicas, de rango de edad y poder adquisitivo para integrarse a Olist como vendedor o cliente. Por lo tanto, concluimos con las siguientes recomendaciones:

1. Brasil es un mercado para E-Commerce con gran potencial y ha demostrado su crecimiento (analíticamente).
2. Se sugiere incursionar, de comienzo, en los estados del Top 6 de mayor población
3. Se sugiere agregar presencia en el estado de Santa Catarina (SC)

Acabamos, vámonos a Brasil... No tan pronto, aún queda más información por analizar

Entraremos a Brasil, pero... y ¿Cómo?

Aunque pareciera tentadora la entrada al mercado brasileño, es importante mencionar el Cómo o al menos, no hacerlo “aventando toda la carne al asador” sigamos entendiendo el mercado brasileño y optimicemos recursos con las siguientes sugerencias.

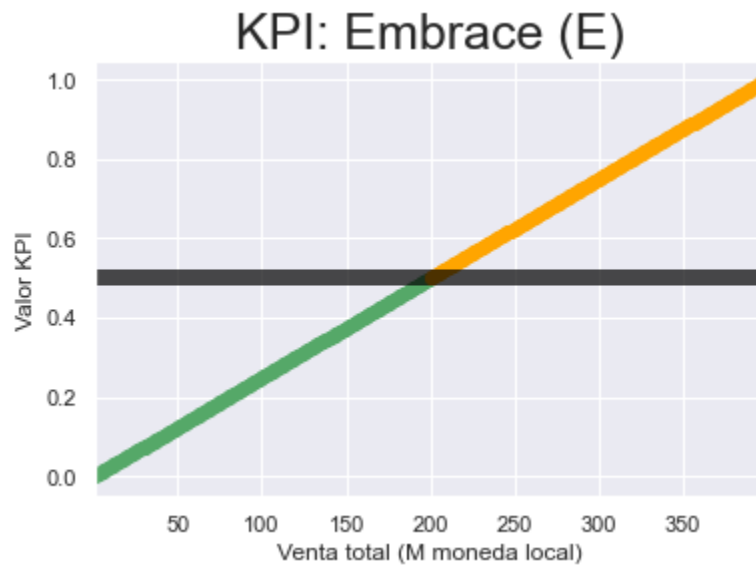
¿En qué gastan los brasileños?

No todos los mercados son iguales, las necesidades, estilo de vida, cultura, etc. Crean las tendencias de mercado, sería incorrecto pensar en utilizar todos los recursos en categorías de productos que no son afines a la población en general, por ello necesitamos tomar la experiencia de Olist, así que veamos cómo se distribuyen sus ventas.

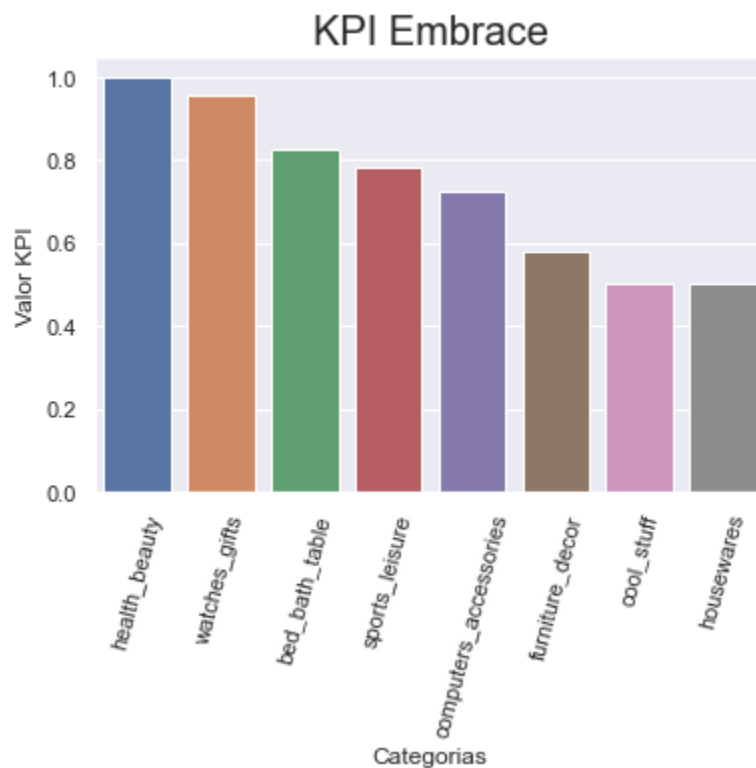


Es claro que el mercado brasileño destina la mayor parte de dinero (a través de E-Commerce) a los productos de Salud y Belleza (health_beauty) seguido muy de cerca por Relojes y regalos. Surge entonces la pregunta, hasta dónde abarcar y tener una “seguridad” de retorno de inversión. Para ello utilizaremos el siguiente KPI Embrace (E):

$$KPI_E = \frac{\text{Total Sale per category}}{\text{Max}(\text{total sale per category})}$$



Lo que nos dice el KPI es que por debajo de la línea negra (línea verde), cualquier categoría necesitaría al menos, vender el doble para ser tan rentable como la categoría más alta. El margen negro puede ser movido, pero su valor de 0.5 es ideal para el comienzo en un nuevo mercado. Aplicándolo a los datos obtenidos tenemos que:



Para maximizar las transacciones y disminuir los costos de desarrollo sugerimos plantear 9 categorías:

- Health_beauty

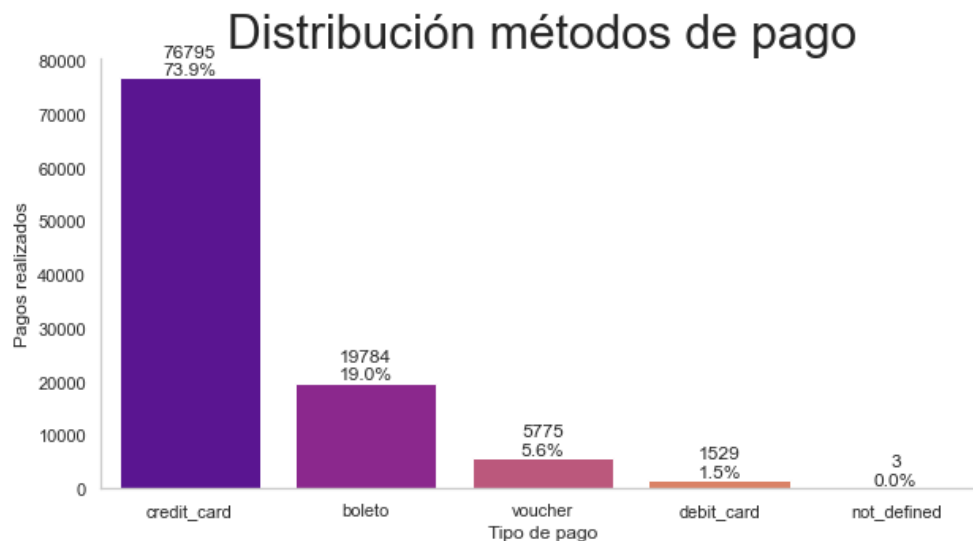
- Watches_gifts
- Bed_bath_table
- Sports_leisure
- Computers_accesories
- Forniture_decor
- Cool_stuff
- Housewares
- Otros

Donde Otros contendrá los artículos de categorías diferentes a las 8 listadas al principio.

Con tal sugerencia garantizamos cobertura de la tendencia del mercado brasileño.

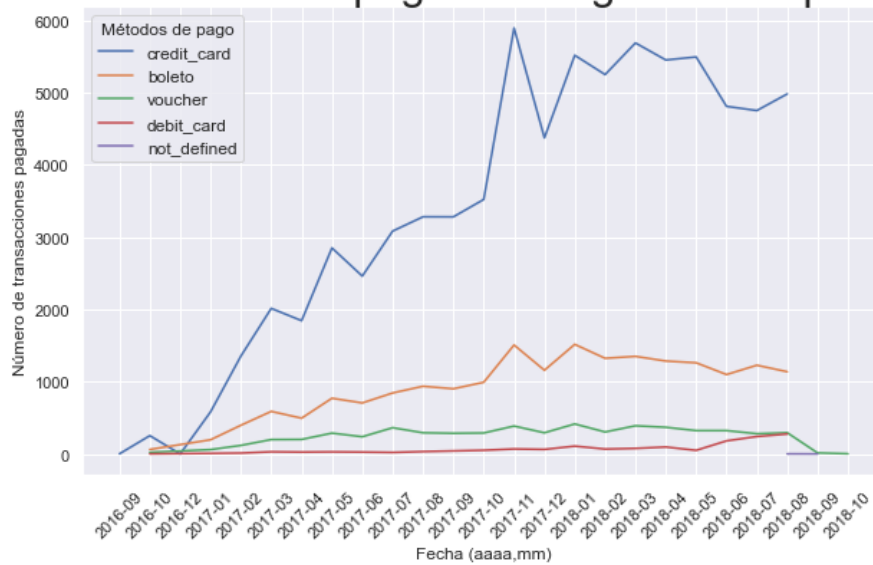
¿Cómo pagan los brasileños?

También es relevante la información de los métodos que se utilizan para comprar artículos, veamos cómo se distribuyen los métodos de pago.



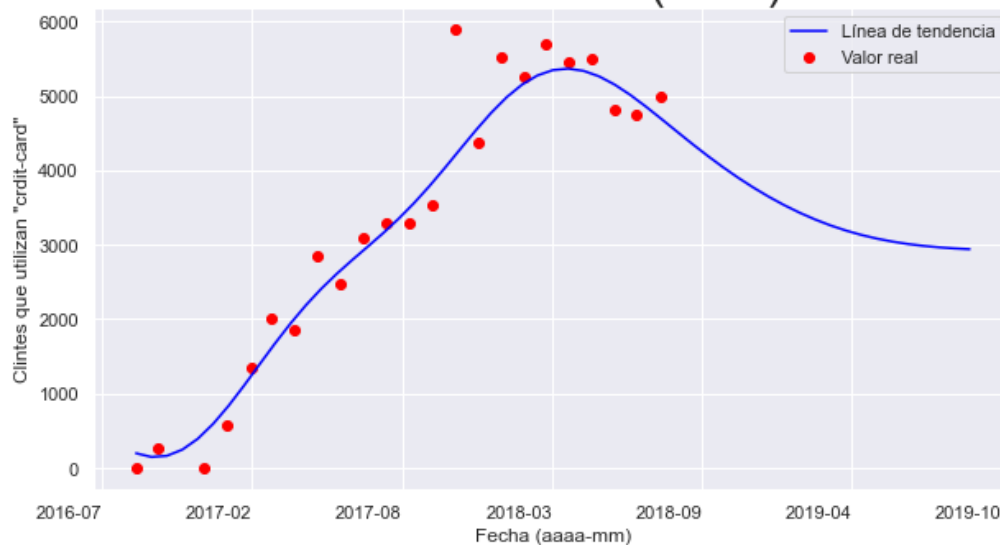
Contundentemente, los pagos con tarjeta de crédito son superiores a cualquier otro método utilizado. A priori, podría pensarse en destinar la mayor parte de los recursos en los pagos con tarjeta de crédito, pero antes de brindar una sugerencia veamos cómo han cambiado a lo largo de tiempo los métodos de pago.

Métodos de pago a lo largo del tiempo



Utilizando un modelo de ML para encontrar una tendencia, se eligió una SVR y se obtuvo

Modelo de ML (SVR)



De no suceder un evento catastrófico, durante los siguientes meses habrá una bajada del uso de tarjeta de crédito que se estabilizará con el tiempo y no hay motivo para pensar, ni información suficiente, para que los otros métodos de pago aumenten en proporciones descomunales con el tiempo

Por lo tanto, como parecía en un principio, la tendencia del uso de tarjeta de crédito será dominante en el tiempo.

Pagos en más de un solo pago

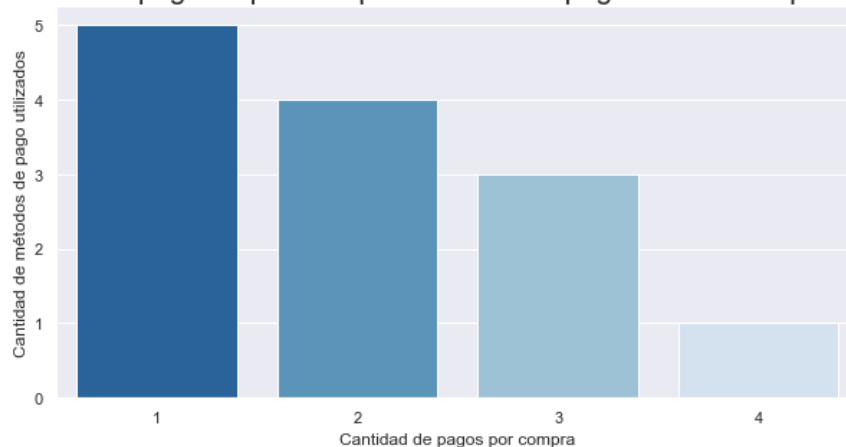
De la información que nos fue proporcionada podemos observar que las compras pueden realizarse con más de un método diferente para cubrir el costo total de los artículos que se quieren comprar, además se puede usar más de una vez el mismo método. Por ejemplo, para adquirir el artículo 'A' de un valor \$A' deseo pagarlo a través de 3 métodos diferentes y realizando 10 transacciones diferentes de la siguiente manera:

- 1 pago con tarjeta de crédito
- 2 pagos con tarjeta de débito
- 7 pagos con voucher

Tal que, la suma de total de todos los pagos sea de \$A'

Veamos si existe un número de pagos máximo diferentes para comprar un artículo.

Métodos de pago empleados por cantidad de pagos realizados por compra



A partir de 4 pagos se utiliza un mismo método de pago para realizarlo y éste es: "voucher"

Realizar múltiples pagos para un mismo artículo consume recursos y análisis.

Por tanto, se brindan las siguientes sugerencias:

- Inversión en la seguridad de transacciones hechas con tarjeta de crédito
- Cobro de comisión por 4 o más métodos para pagar una misma orden

Para este punto contamos ya con un número considerable de sugerencias para una expansión a Brasil; sin embargo, aún no hay que cantar victoria. Aun falta conocer una parte fundamental del proceso de compra-venta y es toda la información que tenga que ver con los clientes, así que evaluaremos ahora qué sugerencias podemos agregar.

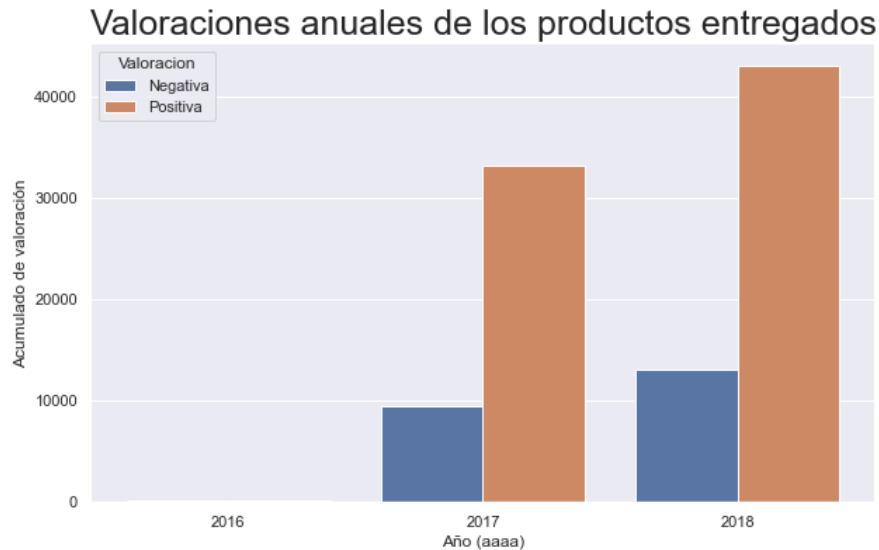
El cliente siempre tiene la razón...

En un proceso de compra-venta, tan importante es el vendedor, como los son los clientes (o compradores), sin clientes no hay mercado y se deben de tener contentos en la medida de lo posible. Afortunadamente Olist tiene un sistema de evaluación, una vez que la orden ha sido entregada, se le envía una encuesta de satisfacción donde el cliente debe o puede:

- Dejar una calificación de 1 a 5 sobre la satisfacción de la orden.
- Dejar un título del mensaje que escribirá sobre su reseña.
- Escribir una reseña sobre la orden pedida.

Notamos que menos de la mitad cuentan con alguna reseña escrita

Veamos cómo han variado las calificaciones a lo largo del tiempo¹.



Separamos las calificaciones de la siguiente manera:

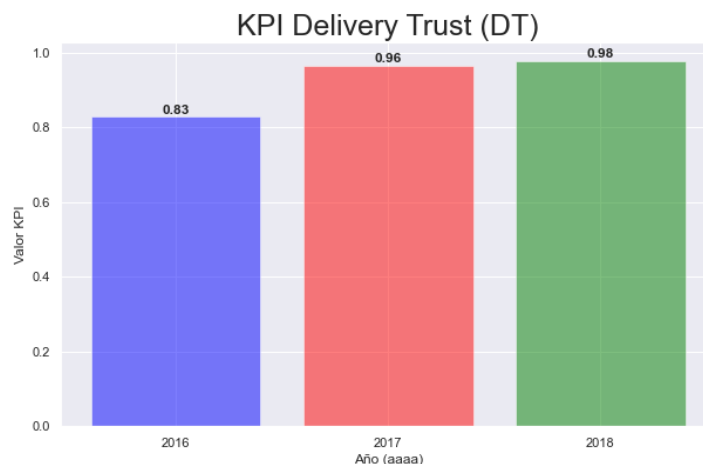
- Positiva: Valor 4 o 5
- Negativa: Valor 1 o 2

Las calificaciones positivas superan a las negativas, pero, ¿Serán representativas estas gráficas?, veámoslo a través de un nuevo KPI, que lo llamaremos Delivery Trust (DT).

$$KPI_{DT} = \frac{\text{Delivered orders}}{\text{Total orders created}}$$

Este KPI nos da información sobre la confianza que existe de que una orden sea entregada si es que se realizó la orden de compra. Entre más cercano sea su valor a 1 significa que las órdenes entregadas tienden a ser el total de órdenes creadas; en otras palabras, si compro una serie de artículos éstos tenderán a ser realmente entregados.

Veamos qué ha sucedido con Olist a través de los años.



Efectivamente, son representativas.

¿Qué influencia una calificación Positiva/negativa?

La respuesta parece obvia, un buen producto tiene una buena calificación y un mal producto una mala, no obstante, una calificación no sólo depende del producto por sí mismo, más aún cuando una orden tiene más de un producto o tipo de producto. La calificación depende de factores como la calidad individual de cada producto o el tiempo que tardó en ser entregado. Con lo segundo sí podemos tener una referencia con la información proporcionada, veamos entonces cómo lo ha hecho Olist.

Logística con Olist

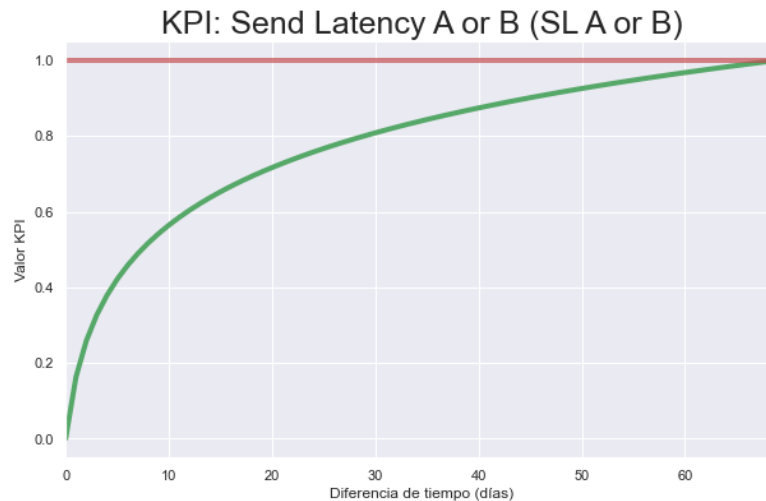
De los datos que nos fueron proporcionados tenemos información del tiempo que le toma al vendedor llevar la orden hacia el operador logístico y una vez lo tiene el operador también contamos con la información del tiempo que tomó entregarlo. Entre más tiempo le tome al vendedor llevar la orden al operador y éste a su destino, el cliente tendrá una percepción prejuizada, en un mal sentido, de la calidad de los productos y/o del proceso en general repercutiendo directamente en la calificación que otorgará.

Definamos entonces los siguientes KPI: Send Latency A (SLA) y KPI: Send Latency B (SLB) como:

$$KPI_{SL\ A\ or\ B} = \frac{\ln(\Delta t(A\ or\ B) + 1)}{\ln(\max(\Delta t(A\ or\ B)))}$$

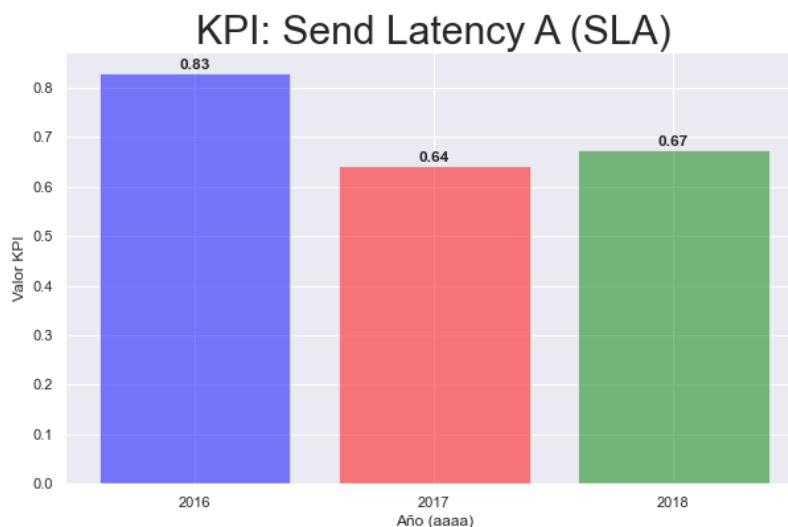
$\Delta t(A)$ = time between sending the order to the logistics operator

$\Delta t(B)$ = time between the shipment from the logistics operator to the customer

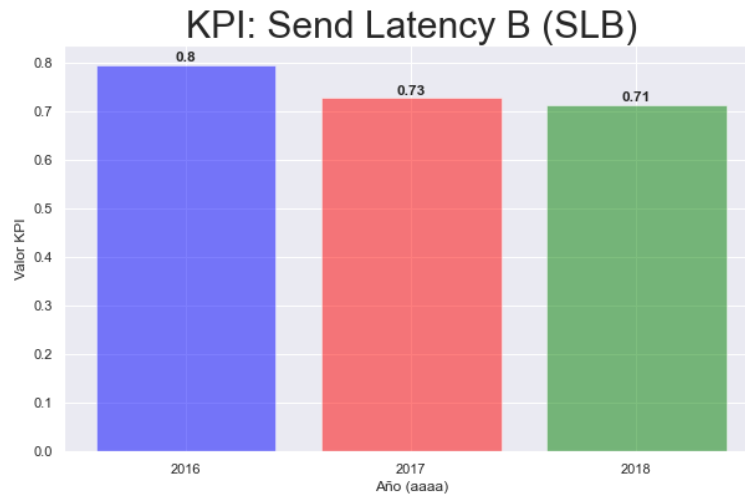


El objetivo es mantener el valor del KPI lo más cercano a 0 puesto que significaría una latencia baja (resultando en un menor tiempo de espera) para el envío de la orden al operador logístico (A) y del operador logístico al cliente (B). Entre más tiempo le tome al vendedor y operador, el valor irá aumentando hasta un valor máximo de 1 representando una muy alta latencia y repercutiendo en la sensación de satisfacción por parte del cliente. La penalización es mayor cercano al 0 y va disminuyendo conforme aumenta esa diferencia en tiempo.

Veamos cómo lo ha hecho Olist



Encontramos una magnífica área de oportunidad, aunque con el paso de los años la latencia de entrega al operador logístico ha disminuido, los valores son muy cercanos a 1; es decir, en general, la mayoría de los vendedores tarda varios días desde que la compra fue aprobada hasta que entregan la orden al operador logístico.



No sólo los vendedores, sino que al operador logístico le toma más de 1 día en poder entregar una orden (ya recibida en sus instalaciones).

Cada orden tiene contenido el costo del envío y costo de producto englobado. Utilizaremos un modelo de ML para calcular el costo del envío en función de las características de cada artículo.

Se utilizó un modelo de Regresión Lineal Múltiple (MLR) y se encontró el que el precio de enviar un artículo depende de las características físicas del producto, tales como su altura, longitud, anchura, peso y distancia hacia donde será enviado.

Se obtuvo la siguiente ecuación:

$$freight\ value = 6.522 + 0.0019x_1 + 0.1039x_2 + 0.1667x_3 + 0.1410x_4 + 1.2668x_5$$

Donde:

$$x_1 = peso\ (g)$$

$$x_2 = largo\ (cm)$$

$$x_3 = alto\ (cm)$$

$$x_4 = ancho\ (cm)$$

$$x_5 = \text{distancia (factor)}$$

Poder predecir el costo del envío tiene diversas ventajas:

- Para el vendedor: Conocer de antemano el precio de enviar cada artículo para poder ajustar precios con tal de ofrecer una mejor alternativa a un mismo producto (comparado con la competencia) o maximizar el volumen del embalaje para envío y poder ofrecer un set de artículos que sea más atractivo a todos los potenciales clientes.
- Para el comprador/cliente: Saber el costo extra por comprar un artículo y elegir aquellos de cuyos vendedores estén más cercanos a su región.
- Para la empresa: Poder brindar la infraestructura para que todo lo anterior sea posible, además de tener la capacidad de buscar un mejor proveedor logístico o, según el área de análisis de negocio, decidir realizar una inversión CapEx para una solución logística propia.

Además, no debemos olvidar el tiempo de entrega. De manera subjetiva, la experiencia de compra de artículos disminuye con respecto al tiempo que dilata la entrega de los mismos y, por el contrario, aumenta con la rapidez de la entrega de los productos. Por lo tanto, una predicción precisa del tiempo de envío impactará en la calidad de experiencia del usuario.

Se utilizó un modelo de Regresión con Bosques Aleatorios (RFR) y se encontró el que el tiempo de envío (una vez en posesión del operador logístico) de un artículo depende de las características físicas del producto, tales como su volumen, peso, precio de envío y distancia hacia donde será enviado.

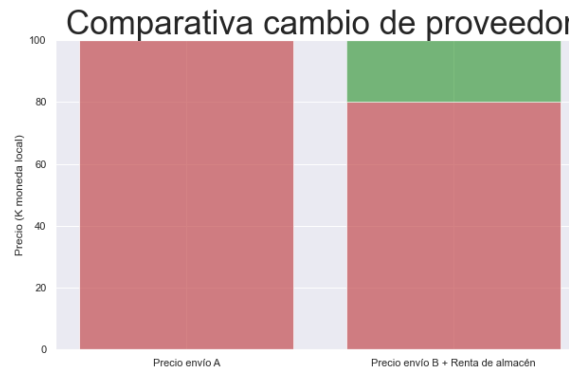
Poder predecir el tiempo de entrega tiene diversas ventajas:

- Para el vendedor: Conocer de antemano el tiempo de entrega podrá, o no, exhibir artículos que pudieran tener fecha de expiración.
- Para el comprador: Poder decidir qué día comprar un artículo y asegurar que alguien se encuentre en el domicilio que pueda recibir la orden.
- Para la empresa: Poder brindar una mejor experiencia de usuario, buscar un mejor proveedor logístico o mejorar la solución propia.

Concluimos con las siguientes sugerencias:

- 1) Opciones al vendedor:
 - a) Sistema de penalización al vendedor por alta latencia de envío de orden al distribuidor logístico o
 - b) Sistema de recompensa por latencia baja de envío de orden al distribuidor o

- c) Renta de almacenes (por estado) para guardar los productos que estarán listos de manera prácticamente inmediata para que el distribuidor logístico pase por ellos siempre y cuando se consiga un operador logístico que disminuya los precios (en relación al modelo antes presentado) y la cantidad de artículos almacenados sea lo suficientemente grande para que el margen de reducción de costos sea igual o mayor al costo de la renta del almacén según la siguiente gráfica.



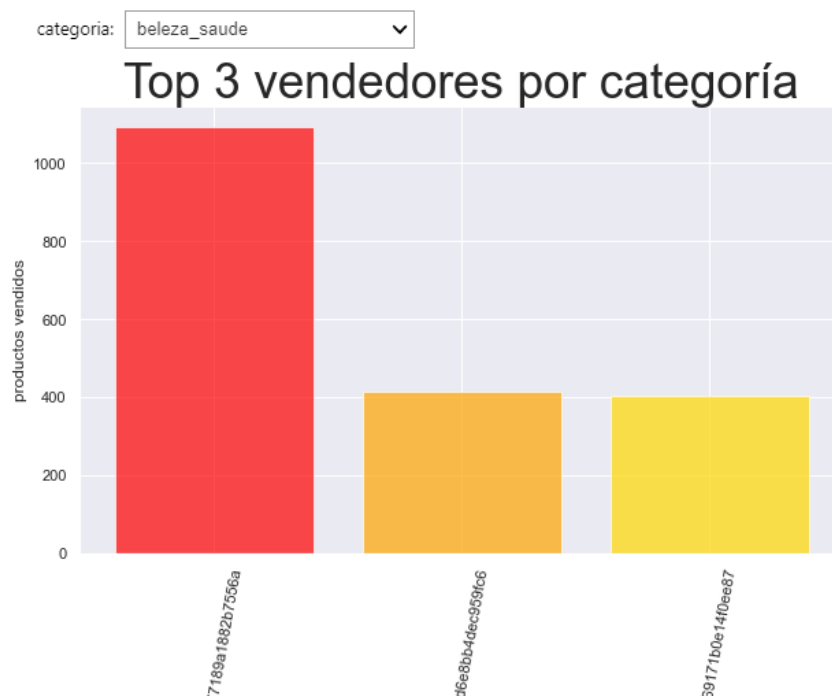
- 2) Opciones al operador logístico:
- a) Sistema de penalización al operador por alta latencia de entrega de orden o
 - b) Sistema de recompensa por latencia baja de entrega de orden o
 - c) Inversión en un sistema de logística propio, que, al no realizar una expansión total al país, los costos no serían tan altos como si se hiciera total.

Se entrega el cliente web app para la predicción de precio de envío y tiempo de entrega con los modelos de ML usados (Streamlit)

Los vendedores también importan

Tan importante es el comprador como el vendedor, sin vendedores no existen los compradores.

Sugerimos un sistema de recompensas para los Top 3 vendedores por categoría



Conclusiones

Entendemos que Brasil es un mercado relativamente joven y abierto para los E-Commerce.

La expansión a Brasil podría darse bajo las siguientes sugerencias:

- La cobertura de estados sugerida es: Sao Paulo (SP), Minas Geras (MG), Río de Janeiro (RJ), Bahía (BA), Río del Sur (RS), Paraná (PR) y Santa Catarina (SC) tanto para vendedores como compradores.

- Sugerimos categorizar los artículos disponibles a su venta en 9 clases:

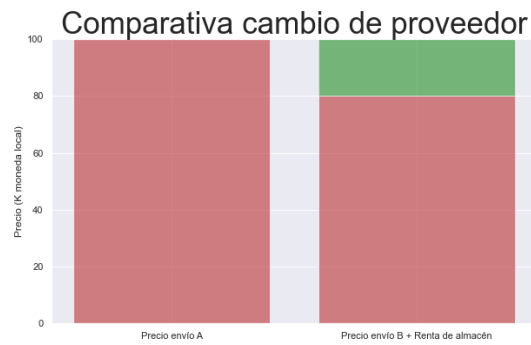
- Health_beauty
- Watches_gifts
- Bed_bath_table
- Sports_leisure
- Computers_accesories
- Forniture_decor
- Cool_stuff
- Housewares
- Otros

- Sugerimos invertir en el área de seguridad para transacciones hechas con tarjeta de crédito.

- Sugerimos la creación de un cobro de comisión por 4 o más métodos para pagar una misma orden (según la empresa mejor convenga).

- Sugerencias a considerar para el vendedor:

- a) Sistema de penalización al vendedor por alta latencia de envío de orden al distribuidor logístico o
- b) Sistema de recompensa por latencia baja de envío de orden al distribuidor o
- c) Renta de almacenes (por estado) para guardar los productos que estarán listos de manera prácticamente inmediata para que el distribuidor logístico pase por ellos siempre y cuando se consiga un operador logístico que disminuya los precios (en relación al modelo antes presentado) y la cantidad de artículos almacenados sea lo suficientemente grande para que el margen de reducción de costos sea igual o mayor al costo de la renta del almacén según la siguiente gráfica.



- Sugerencias a considerar para el operador logístico:

- Sistema de penalización al operador por alta latencia de entrega de orden o
- Sistema de recompensa por latencia baja de entrega de orden o
- Inversión en un sistema de logística propio, que, al no realizar una expansión total al país, los costos no serían tan altos como si se hiciera total.

- Se entrega el cliente web app para la predicción de precio de envío y tiempo de entrega con los modelos de ML usados (Streamlit)

- Sugerimos crear un sistema de recompensas para los Top 3 vendedores por categoría (según la empresa mejor convenga) y entrega de modelo para ubicar dicho top 3 (Streamlit)

-