



# Segmentación de Clientes BLA Concept

Elaborado por **Ana Kozameh** (Leg. N° 58737) y **Felipe Gazzotto** (Leg. N° 59769)

Tutora: **Romina Alpiggiani**

Trabajo Final presentado para la obtención del título  
de Licenciado en Analítica Empresarial y Social

BUENOS AIRES  
SEGUNDO CUATRIMESTRE 2021

## **Resumen Ejecutivo**

El presente informe tiene como objetivo introducir el proyecto realizado para la empresa de indumentaria BLA Concept.

Actualmente, las ventas en la página web de BLA presentan una tasa de conversión en el checkout menor al promedio del sector, y menor a la esperada por la marca. La finalidad de esta investigación es la de crear un plan de implementación de mejoras para revertir esto, y mejorar el porcentaje de ventas.

En una primera etapa, se realizó una recolección de la información disponible, y en función de estos resultados un posterior análisis exploratorio de los datos. Este análisis permitió detectar cómo se caracterizan los clientes de BLA.

Luego de haber efectuado esto, se detectó una problemática que afectaba directamente la solución propuesta. Se vio que no estaba hecha la conexión entre la página web y Google Analytics a nivel cliente. Esto hizo que se cambie el enfoque de cómo se abordó la solución ante el problema planteado. De esta manera, se llegó a la conclusión de que el plan de mejora óptimo para BLA Concept es la de un trabajo de pre-media.

Para realizar esto, se realizó una segmentación y clusterización de los clientes de la marca, permitiendo obtener grupos de clientes con características marcadas.

Por último, se creó un plan de implementación y de futuras medidas que BLA Concept debe llevar a cabo para poder aumentar su tasa de conversión.

## Índice

BLA Concept	4
Industria del Retail de la Moda	4
Estructura actual de BLA Concept	5
Descripción del Problema	7
Objetivo del Proyecto	8
Medición de Valor y KPIs a Impactar	9
Plan de Trabajo	10
Gantt del Proyecto	10
Matriz de Riesgos del Proyecto	11
Datos a Utilizar	13
Herramientas a Utilizar	13
Análisis Exploratorio de Datos	14
Plataformas	14
Análisis Realizado	16
Planteo de Hipótesis	31
Business Case	31
Cambio de Enfoque	32
Business Case actualizado	35
Investigación	35
Base de datos utilizada	35
Clusterización	37
Resultados obtenidos	41
Conclusiones	46
Medición del Proyecto	47
Next challenges	48

## BLA Concept

BLA Concept es una marca de indumentaria argentina. En el año 2017, luego de haber adquirido una vasta experiencia en la industria textil, Verónica Wojcicki se propuso abordar el camino del emprendedurismo. Gracias a sus 26 años trabajando en el rubro, conocía todas las aristas que el negocio necesitaba, prometiendo acercar una propuesta de valor diferente en el mercado argentino.

Desde su fundación, BLA se diferenció por su distintivo diseño. Presentando básicos atemporales con complementos de cada temporada. Fiel a su estilo y a lo que buscaba transmitir, dio nombre a BLA Concept. El origen de su nombre viene de la combinación “Buenos Aires” y “Los Angeles”, el complemento de dos ciudades. Ambas reconocidas mundialmente y embebidas de arte y distinguidas por el estilo y prendas autóctonas. Wojcicki había logrado acercar a clientes argentinos prendas únicas, las cuales llamaron la atención inclusive de diseñadores renombrados.

### Industria del Retail de la Moda

En la industria de la indumentaria, las empresas respetan la llamada “Pirámide de Productos”, en donde se ubican las diferentes prendas que tiene en la colección y muestra cómo estas influyen en las ventas.

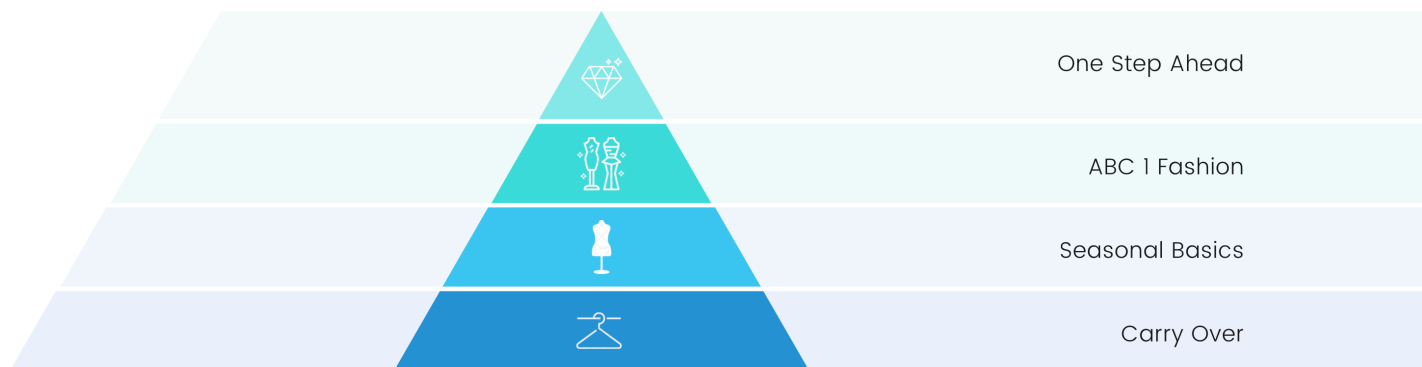


Imagen 1: Pirámide de productos de una empresa de indumentaria.

Los productos **Carry Over** se encuentran en la base de la pirámide, y son aquellos básicos atemporales que acompañarán a la marca sin importar el momento. Son aquellos productos que se producen en mayor volumen y tienen un menor precio. Luego, se encuentran los **Seasonal Basics**. Estas prendas son aquellos Carry Over ajustados al momento del año. Un claro ejemplo es la camisa insignia de la marca, confeccionada en lino para la temporada de verano. Estos dos tipos de productos son los que representan la mayor fuente de ingresos para el negocio.

En tercer lugar, se ubican los **ABC 1 Fashion**. Estas son aquellas prendas que siguen completamente la moda del momento.



Por último, en la punta de la pirámide los **One Step Ahead**, que pocas marcas los tienen como productos. Estas son prendas que se diferencian de la moda del momento y son un anticipo de lo que va a venir. Se producen en un menor volumen pero su precio es mucho mayor.

Diseñar y confeccionar colecciones que sigan esta regla es lo natural en las empresas de indumentaria. Sin embargo, desde el 2017 hasta la actualidad se generó un fenómeno distintivo en BLA. Las prendas que fueron pensadas, por su diseño, para ser One Step Ahead son las que se terminaron convirtiendo en las más demandadas, diferenciándose totalmente del resto de las empresas argentinas. Este es el sello distintivo de BLA: traer al mercado diseños innovadores y deseados por los consumidores.

### Estructura actual de BLA Concept

- **Locales físicos:** en cuanto a la infraestructura, BLA cuenta con un local propio ubicado en el barrio de Retiro, en la Ciudad Autónoma de Buenos Aires.

Además, cuenta con tiendas que venden sus productos en Perú y Miami. Y, próximamente, tendrá presencia en Uruguay.

- **Locales digitales:** tiene una página web en TiendaNube en donde publican y venden los productos. Desde esta página web se puede comprar tanto en Argentina como desde el resto del mundo, ya que la plataforma brinda la opción de presentar los productos en inglés y con el precio en dólares. Una imagen de su landing page se muestra a continuación.

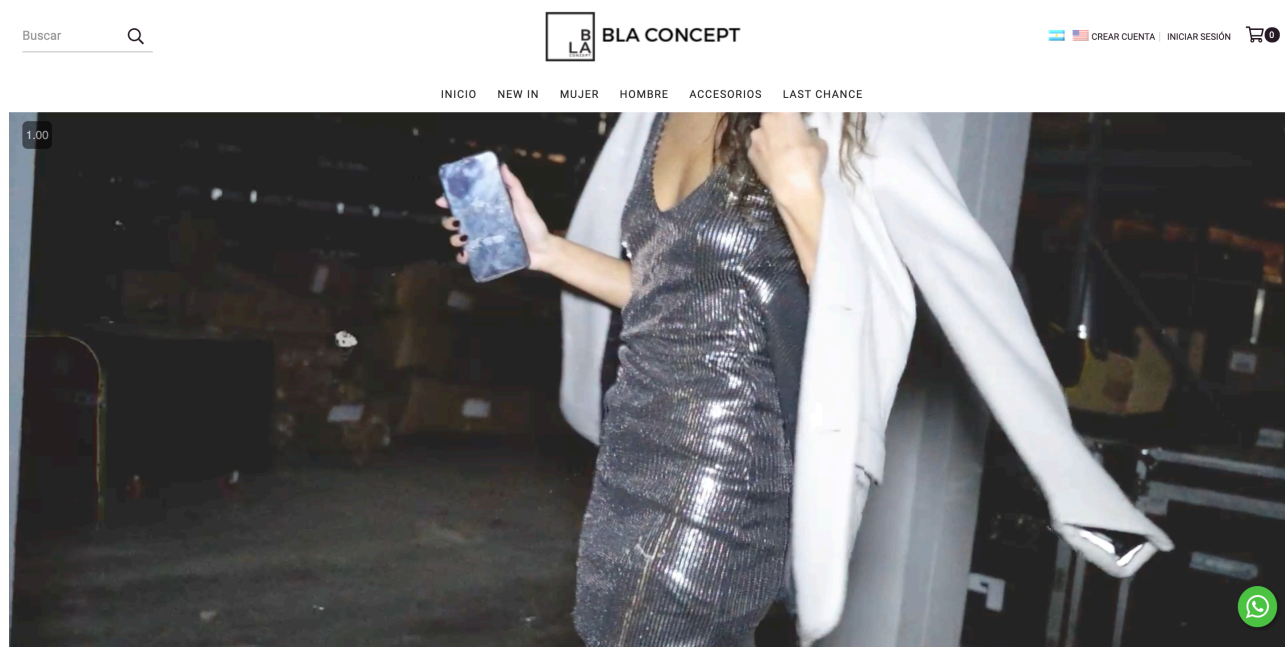


Imagen 2: TiendaNube de BLA Concept.

- **Redes sociales:** su cuenta principal se encuentra en Instagram, llegando a poco menos de los 20.000 seguidores. Además, recientemente, se unió a TikTok, en donde busca encarar al segmento de clientes más jóvenes.

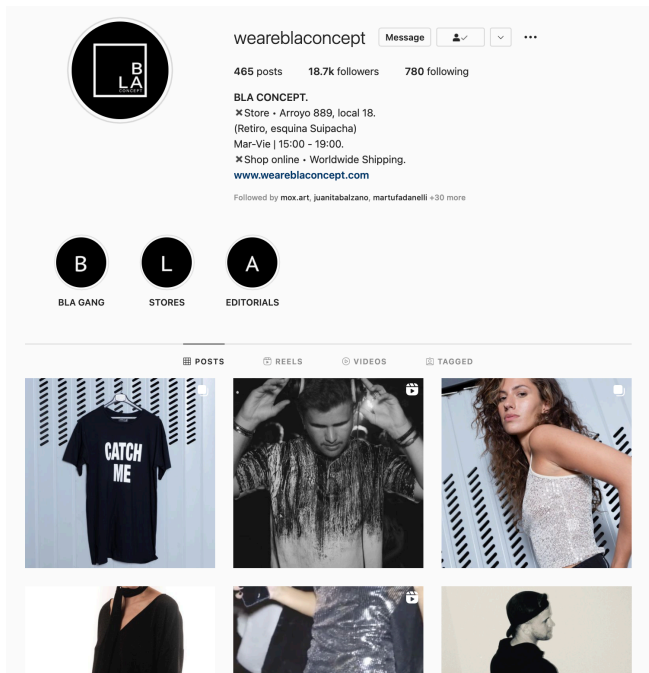


Imagen 3: Feed de Instagram de BLA.

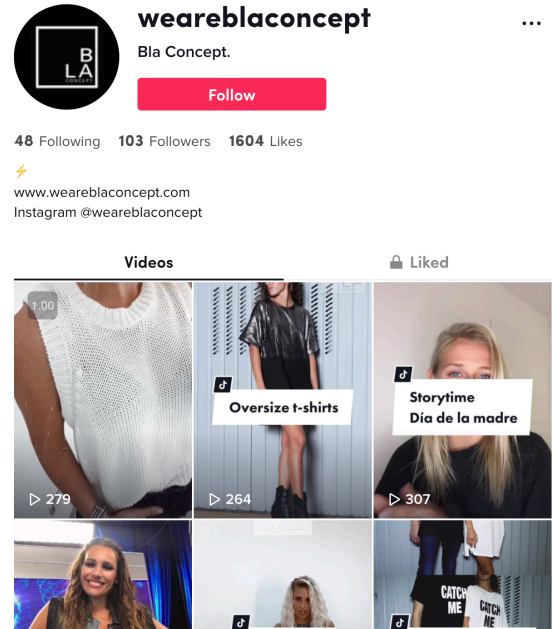


Imagen 4: Feed de TikTok de BLA.

## Descripción del Problema

Actualmente, el mayor porcentaje de las ventas de BLA Concept se realiza a través de su página web, en TiendaNube. Esta plataforma brinda a los usuarios varias métricas y estadísticas sobre cómo se comportan los pedidos de los clientes.

Una sección clave para el negocio es la de **Conversión del proceso de compra**. En esta, se muestra todo el proceso actual desde que un cliente comenzó con la intención de compra cargando productos en el carrito<sup>1</sup>. A continuación se muestra el funnel de ventas actual de BLA.

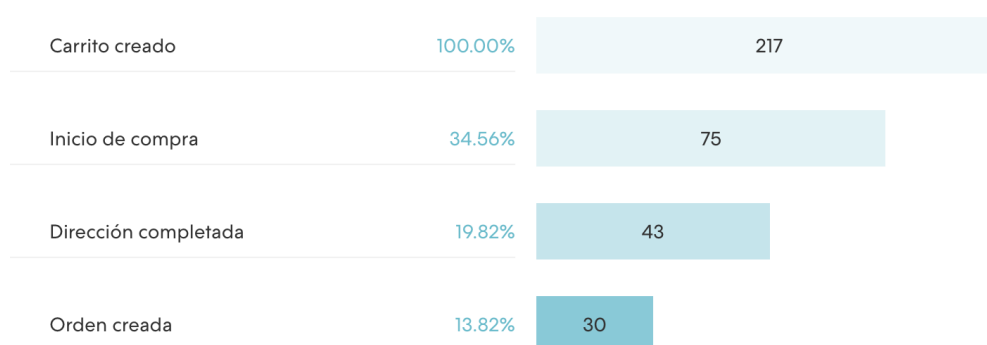


Imagen 5: Funnel de ventas de BLA en la plataforma Tiendanube.

Los estadios dentro de este funnel se componen de la siguiente manera:

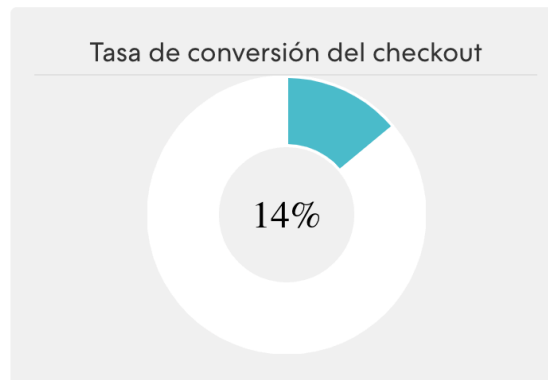
- **Carrito creado:** se crea en el momento que el usuario incluye el primer producto a su carrito.
- **Inicio de compra:** una vez que el usuario seleccionó todos los productos y selecciona en el botón de "iniciar compra", que lo lleva a una nueva ventana en que le pide completar sus datos. En esta etapa sólo se cuenta cuando el usuario selecciona la opción de iniciar la compra, no más de ese momento.
- **Dirección completada:** cuando el usuario, luego de haber iniciado la compra, completa sus datos personales.
- **Orden creada:** es en el momento que, luego de haber completado sus datos y el medio de pago, presiona el botón de iniciar compra. Cabe remarcar que en este momento se crea la orden de compra, pero no se verifica que el pago se haya aceptado, por lo que puede verse cancelada en un futuro si esto no se llega a concretar.

A partir de este funnel, se presenta un indicador clave: la **tasa de conversión del checkout**. Esta métrica el porcentaje de ordenes creadas sobre el total de carritos creados. Los carritos que fueron creados pero no terminaron como orden creada son los **carritos abandonados**. Es decir, carritos que los usuarios comenzaron pero nunca terminaron en una compra.

---

<sup>1</sup> Un carrito es una plataforma dentro de la página web en donde los usuarios agregan los productos de su pedido, antes de realizar la compra.

Esta tasa de conversión se encuentra bastante por debajo del promedio de la industria, y es mucho menor que la esperada por el equipo de BLA. Para el mes de agosto tu tasa de conversión fue de tan sólo 14%, lo que significa que más de 85% de los carritos creados fueron abandonados.



*Imagen 6: Tasa de conversión del checkout para el mes de agosto de BLA Concept.*

Esto representa un problema ya que demuestra que se está perdiendo a un gran porcentaje de clientes en algún momento del funnel de ventas, lo que termina incidiendo en la facturación de la marca.

## **Objetivo del Proyecto**

Actualmente, BLA Concept no cuenta con los medios necesarios para detectar qué es lo que causa este porcentaje de carritos abandonados.

El objetivo de este proyecto es encontrar el motivo por el cual la tasa de conversión del checkout es tan baja. A través de un profundo análisis del comportamiento de los clientes y de las ventas de la marca, se buscará identificar en qué parte del proceso de compra es donde se produce esta gran pérdida de futuras ventas. De esta manera, se planteará una implementación de mejora para abordar este problema por parte de BLA.

Se utilizarán todos los datos disponibles y relevantes para la problemática a tratar de TiendaNube y Google Analytics. Además, se pondrá en contacto con Frank Bjelica, actual encargado del manejo de las redes sociales y de la página web de la marca. De esta manera, se abarcarán todas las aristas que pueden estar afectando a la problemática a tratar, y, así, poder encontrar la causa y brindar una solución.

## Medición de Valor y KPIs a Impactar

Como ya se describió, el objetivo del proyecto es identificar las causas de la baja **tasa de conversión del checkout**, calculada como el porcentaje de clientes que finalizan una compra una vez que ya ingresaron algún artículo en su carrito. Las variaciones sobre esta tasa tienen un enorme impacto en el negocio. Esto se debe a que está directamente relacionada con la cantidad de órdenes realizadas a BLA.

Siguiendo este pensamiento, se establecieron que las métricas a medir por el proyecto serán las siguientes:

- **Tasa de conversión del checkout** — este será el principal KPI del proyecto.
- **Facturación total** — brindada por TiendaNube y representa la suma de todas las ventas que realizó BLA (sin incluir las órdenes canceladas).
- **Ticket promedio** — es el resultado de dividir la facturación total por la cantidad de ventas concretadas.
- **Visitas únicas a la página web** — son los usuarios que ingresaron a la tienda por primera vez en el período seleccionado.
- **Cantidad de órdenes pagas** — es la cantidad de órdenes que se crearon y fue aceptado el pago.
- **Funnel del proceso de ventas** — se tendrán en cuenta el comportamiento de los cuatro estadios que tiene un pedido en la plataforma, mencionados anteriormente e ilustrados en la imagen 5.

De esta manera, teniendo en cuenta los lineamientos del proyecto, mientras más exitoso resulte ser el plan de implementación de mejoras que se buscará crear, mayor será la cantidad de órdenes y, por ende, la facturación total.

## Plan de Trabajo

En cuestión a las metodologías utilizadas, el proyecto se realizará en una mayor parte siguiendo una estructura de **Waterfall**. Sin embargo, habrá momentos en los cuales se requerirá de una constante interacción entre todos los involucrados, por lo que se utilizarán en esas instancias **metodologías ágiles**, con reuniones y constante comunicación con el cliente.

## Gantt del Proyecto

A continuación, puede observarse el Gantt con el desarrollo del proyecto.

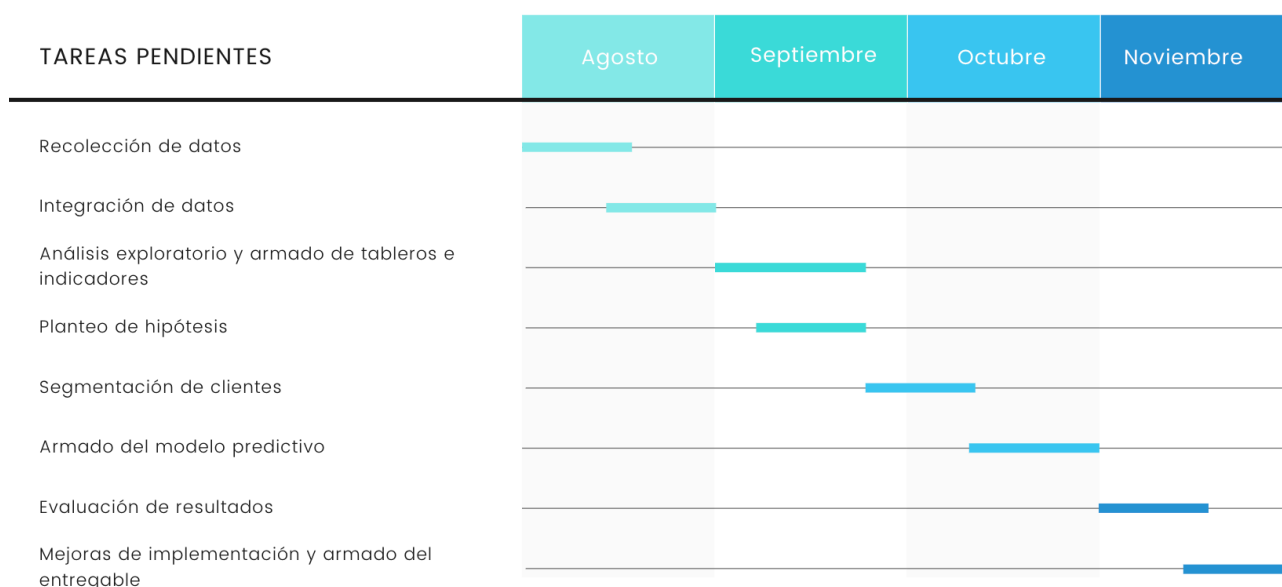


Imagen 7: Gantt del proyecto.

- **Recolección de datos:** en esta primera etapa se buscará recolectar todos los datos disponibles y necesarios para abordar el problema.
- **Integración de datos:** luego, se buscará integrar las diferentes fuentes de datos obtenidas en la etapa anterior para poder efectuar los análisis correspondientes.
- **Análisis Exploratorio y armado de indicadores:** esta etapa busca tener un mejor entendimiento sobre cómo se comportan las distintas variables sobre las cuales se recopila la información de BLA Concept. Se realizarán análisis y estadísticas sobre los datos ya integrados, lo que permitirá tener insights y, en función de los resultados, formular la hipótesis.

- **Planteo de la hipótesis:** luego de haber realizado un análisis y llegado al correcto entendimiento de las variables de la empresa, se planteará la hipótesis a la cual se buscará llegar.
- **Segmentación de clientes:** en esta etapa se buscará segmentar a los clientes según las características que los componen. De esta manera, se podrá identificar cuáles son aquellas personas que recorren todo el funnel de ventas, y realizan la compra y cuáles son las que abandonan el carrito.
- **Armado del modelo predictivo:** una vez que se obtengan los grupos según las características de cada tipo de cliente, se armará el algoritmo del modelo predictivo.
- **Evaluación de los resultados:** en esta etapa se evaluará que los resultados del modelo predictivo sean adecuados.
- **Mejoras de implementación y armado del entregable:** si en la etapa anterior se vio algún problema con los resultados del modelo, se arreglarán en esta etapa. Además, se preparará el informe entregable final.

### Matriz de Riesgos del Proyecto

A continuación, puede observarse la matriz de riesgos identificados para este proyecto.

Ref Nro	Descripción del Riesgo	Probabilidad de ocurrencia	Severidad del Impacto	Evaluación del riesgo	Acción a encarar
1	Tiempo				
	En el caso en que se produzcan retrasos en alguna de las tareas, se podrían demorar otras tareas posteriores, o incluso demorar la finalización del proyecto y su entregable final.	0,3	0,7	0,21	Establecer controles y checkpoints para verificar el cumplimiento del cronograma de implementación y evitar fallas espontáneas que requieran la repetición de alguna tarea, generando demoras en el proyecto.

Ref Nro	Descripción del Riesgo	Probabilidad de ocurrencia	Severidad del Impacto	Evaluación del riesgo	Acción a encarar
2	Datos				
	Los datos pueden llegar a presentar dos tipos de riesgo. Por un lado, no tener la calidad esperada. Por otro lado, no llegar a tener acceso a todos los datos necesarios.	0,4	0,9	0,36	Asegurar la disponibilidad y accesibilidad a las fuentes de datos.
3	Recursos Humanos				
	Potencial falta de alineación entre los integrantes del equipo que podría terminar en productos de menor calidad o un retraso en los plazos del proyecto.	0,5	0,2	0,10	Será fundamental implementar una estrategia de comunicación que acompañe a los integrantes del equipo para que los objetivos sean alcanzables
4	Limitación de hardware				
	En caso de que el hardware actual no tenga el poder suficiente para correr algoritmos complejos o manejar grandes volúmenes de información.	0,1	0,9	0,09	Contar con la posibilidad de acceder a hardware de mayor potencia que pueda enfrentar los desafíos previos.

Tabla 1: Matriz de riesgos del proyecto.



## Datos a Utilizar

La primera etapa del proyecto consiste, como se mencionó anteriormente, en la recolección de los datos que se necesitarán para aplicar los modelos para la resolución del problema. A grandes rasgos, estos serán de dos tipos: datos del negocio y datos del cliente.

Estos datos se obtendrán de los fuentes. Por una lado, se contarán con las bases del datos que brinda TiendaNube, que contiene información de los clientes y de las ventas. Por otro lado, se buscará la conexión de Google Analytics a esta página web. De esta se podrá obtener información más detallada del comportamiento del usuario en la página.

Por otro lado, se investigará la posibilidad de realizar encuestas a los clientes que permitan identificar con mayor facilidad cuál es el problema actual con la tienda online y a qué se debe la enorme cantidad de compras incompletas.

Una vez obtenidos los datos, se realizará una limpieza de formato y obviamente se anonimizarán todos los datos personales de clientes para evitar violar cualquier tipo de regulación sobre seguridad y privacidad de datos. Se dejarán entonces los datos preparados para poder empezar a aplicarle distintos tipos de modelos y algoritmos.

Teniendo ya los datos preparados, se realizarán distintas metodologías de segmentación y clustering para poder identificar grupos similares de clientes, a los que luego se les aplicarán los distintos tipos de algoritmos predictivos para determinar si completaran sus compras e identificar qué es lo que los impide si no lo hiciesen.

## Herramientas a Utilizar

Se estableció que para el desarrollo del proyecto se necesitarán herramientas para tres actividades principales.

1. En un primer lugar, se encuentra la organización y desarrollo del proyecto, lo cual se utilizará el software **Jira**.
2. La segunda actividad, es la del armado del código para la exploración y limpieza de datos junto con aplicación de modelos predictivos. Para esto, se utilizarán dos posibles lenguajes de programación: **R** o **Python**, en un enviroment como Jupyter Notebooks o R Studio ,dependiendo de la elección.
3. Por último, se encuentra la visualización de los datos, que aunque parte de ella probablemente también provenga de los códigos previamente mencionados, para lograr una mayor estética y facilidad de lectura se optó por la utilización de **Tableau**.

## Análisis Exploratorio de Datos

El objetivo del EDA (Exploratory Data Analysis) realizado para este proyecto es el de tener un mejor entendimiento sobre cómo se comportan las distintas variables sobre las cuales se recopila la información de BLA Concept. Como se mencionó anteriormente, las plataformas en las cuales BLA está presente y contienen la información relevante para el proyecto son Tienda Nube y Google Analytics.

En primer lugar, se hará una breve descripción sobre qué es lo que contiene cada una de estas plataformas y qué función cumplen para BLA. En función de estas variables, se introducirá el Análisis Exploratorio de la información que estas páginas contienen. En esta segunda sección se hablará sobre cómo se comportan cada una de las variables introducidas, cómo se relacionan entre sí, y qué nuevos insights fueron encontrados en el proceso.

A continuación, y como resultado del paso anterior, se buscará generar hipótesis que respalden al proyecto.

### Plataformas

#### TiendaNube

TiendaNube es la plataforma de ecommerce que utiliza BLA Concept. La página cuenta con cinco categorías: Mujer, Hombre, Accesorios, Last Chance y Pre-Order. Las dos primeras cuentan con subcategorías según el tipo de prenda, ya sean jeans, abrigos, remeras, etc. En Last Chance se encuentran aquellos productos de los cuales hay poco stock, y en Pre-Order aquellas prendas que todavía no entraron en stock. A continuación se muestra una imagen del encabezado de estas categorías.



Imagen 8: Categorías de la página web de BLA Concept.

Tienda Nube recopila la siguiente información:

#### Datos globales del ecommerce

- Cantidad de **visitas únicas** a la página
- Cantidad de **órdenes** realizadas
- **Facturación**
- Valor promedio del **ticket**

Toda esta información está también segmentada por canal. Esto permite ver si cada una de estas variables fue realizada a través de un smartphone o por la web.

### Datos de los productos

- **Stock** actual de cada producto
- Cantidad de **visitas** que tuvo cada producto
- Cantidad de **ventas** que tuvo cada producto
- Nivel de **facturación** de cada producto

De esta plataforma, se obtuvieron dos bases de datos: la lista de ventas y la lista de clientes.

El listado de los clientes contenía las siguientes variables:

- Nombre y Apellido
- Mail
- DNI
- Teléfono de contacto
- Dirección (Calle y número, Ciudad, Localidad, Provincia y País)
- Cantidad de compras que realizó a través de la página
- Total consumido en todas las compras realizadas (\$)
- Fecha en que se registró
- Fecha de la última compra que realizó
- Si está suscripto o no a email marketing de BLA

El listado de las ventas realizadas por la página web

- Número de orden
- Información del cliente (datos nombrados arriba)
- Fecha en que se concretó
- Estado del pago
- Estado del envío
- Costo de productos
- Costo de envío
- Total
- Medio de envío (envío o retiro en el local)
- Medio de pago
- Fecha de pago
- Fecha de envío
- Productos comprados y su cantidad
- Canal por el cual se realizó la compra (web o móvil)
- Identificador de la orden

## Google Analytics

La TiendaNube de BLA está integrada a Google Analytics. Esta es una herramienta que brinda Google que permite realizar un seguimiento más detallado de todas las interacciones realizadas en la página web.

La información que brinda esta plataforma está dividida en cinco grandes categorías, cada una con sus respectivas subcategorías. La información más relevante para el presente proyecto, y la cual se utilizará en el análisis se detalla a continuación.

Categorías principales utilizadas:

1. **Audiencia:** tiene toda la información con respecto a la audiencia que frecuenta la página web de BLA Concept.
2. **Adquisición:** información relacionada a los canales por los cuales la audiencia llega a la página web.
3. **Comportamiento:** información relacionada a las interacciones que la audiencia hace en la página web.
4. **Conversiones:** toda la información relacionada con las ventas y el armado de carritos en la página.

## **Análisis Realizado**

A continuación, se mostrará el Análisis Exploratorio de Datos ejecutado. Este análisis fue dividido en dos secciones, una para cada fuente de información. En primer lugar, se analizará la información brindada por Google Analytics, y, luego, aquella información brindada por Tienda Nube.

## Google Analytics

En un primer paso se decidió analizar cómo es el comportamiento de aquellas interacciones que terminan en una compra, y aquellas que abandonan en algún momento de la transacción. Además, se decidió evaluarlo según si el visitante es nuevo o no, para ver si esto cambiaba el comportamiento.

Como muestra el gráfico 1, hay un mayor número de visitas por parte de los nuevos visitantes. Sin embargo, son los que vuelven a la página aquellos que mayor interactúan.

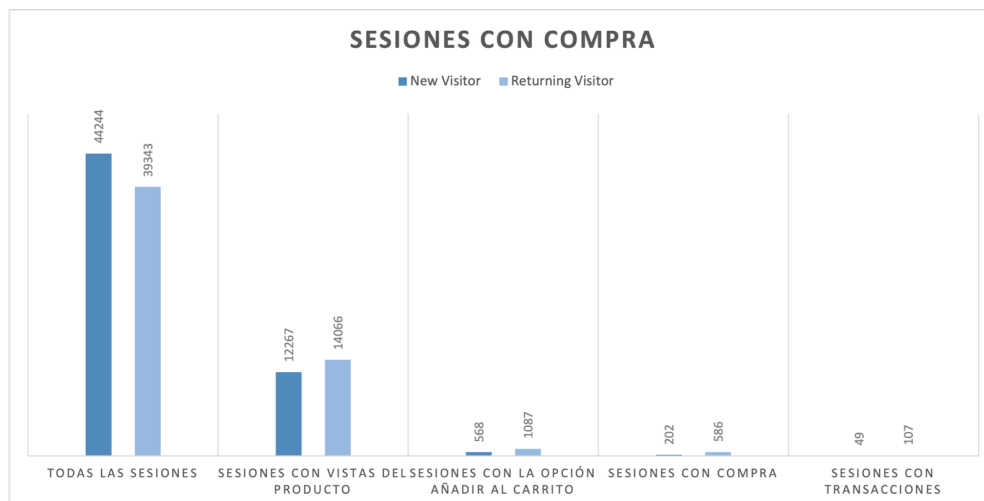


Gráfico 1: Sesiones en Tienda Nube con compra.

El gráfico 2 muestra la evolución de las sesiones, pero esta vez aquellas que terminaron con carritos abandonados. En este caso se puede observar que el mayor número de interacciones lo hacen los nuevos visitantes de la página, a diferencia del caso anterior.

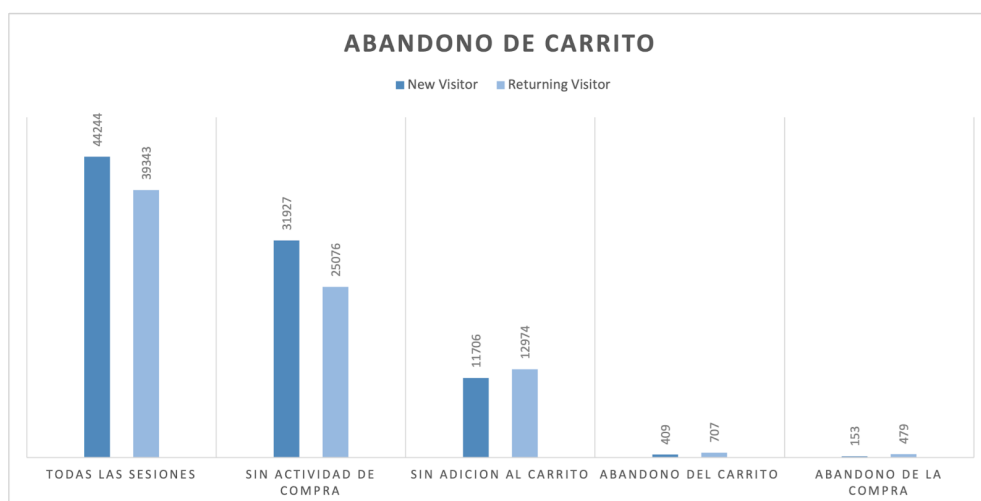


Gráfico 2: Sesiones en Tienda Nube con abandono de carrito.

Como se mencionó en un principio, Tienda Nube cuenta con cuatro pasos para generar una compra:

- Creación del carrito de compras
- Inicio de la compra
- Datos completados
- Creación de la orden

En un segundo momento, lo que se quiso analizar fue cómo era el comportamiento de los visitantes dentro de este funnel. Nuevamente, se hizo una segregación de visitantes nuevos y "returning". Como se observa en el gráfico 3, el número de compras es mayor para aquellos visitantes que ya habían visto la página.

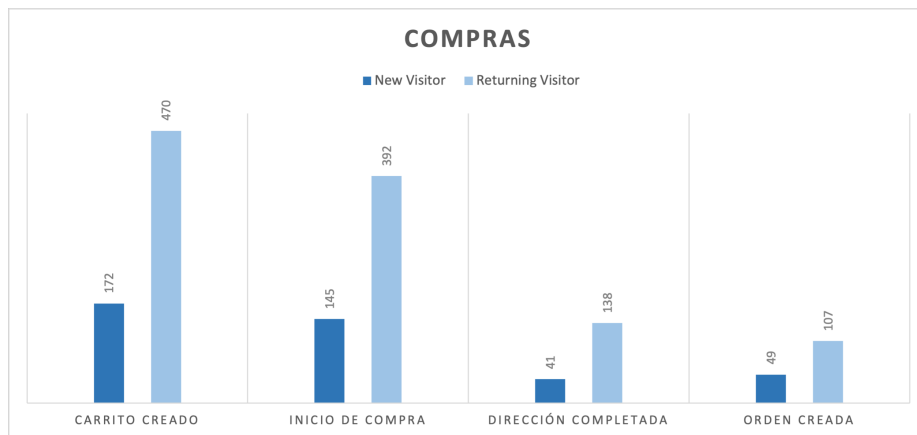


Gráfico 3: Funnel de ventas para compras.

Por otro lado, teniendo en cuenta el objetivo del proyecto, se decidió hacer el mismo análisis pero con aquellos visitantes que abandonaron el proceso en algún momento del funnel. Esto se muestra en el gráfico 4, en donde se puede ver cómo nuevamente son aquellos visitantes que ya estuvieron en la página web que presentan un mayor número de abandono. Además, algo importante a destacar es que en el paso donde la mayoría de los visitantes abandonan el proceso de compra es luego de haber creado el carrito e iniciado la compra, pero antes de terminar de completar sus datos personales.

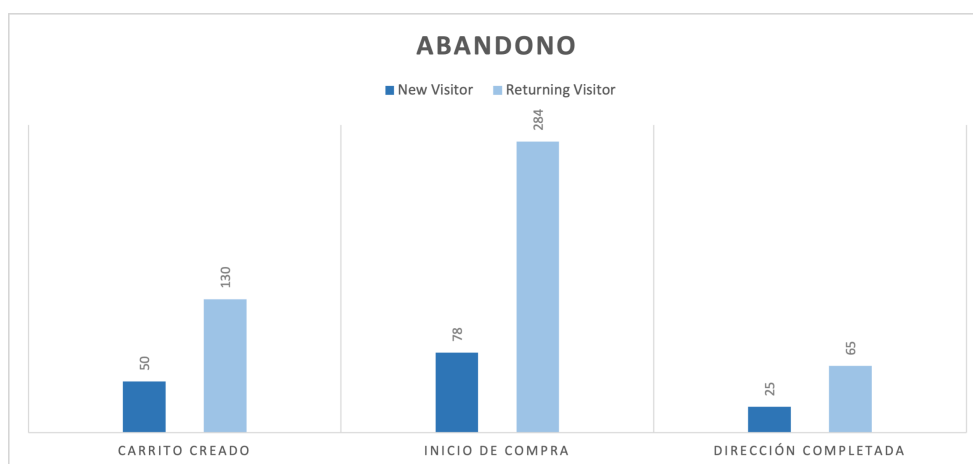


Gráfico 4: Funnel de ventas para abandono de carritos.

Una vez analizados los datos relacionados al funnel de ventas, se decidió analizar cómo es el comportamiento general de la página web de BLA. Para ello, se buscó analizar la cantidad de transacciones que se presentan de manera diaria. Como se puede observar en el gráfico 5, la cantidad de transacciones en el último año se mantuvo entre 1 y 4 por día, con algunas excepciones y un pico presentado el 18 de septiembre del presente año.

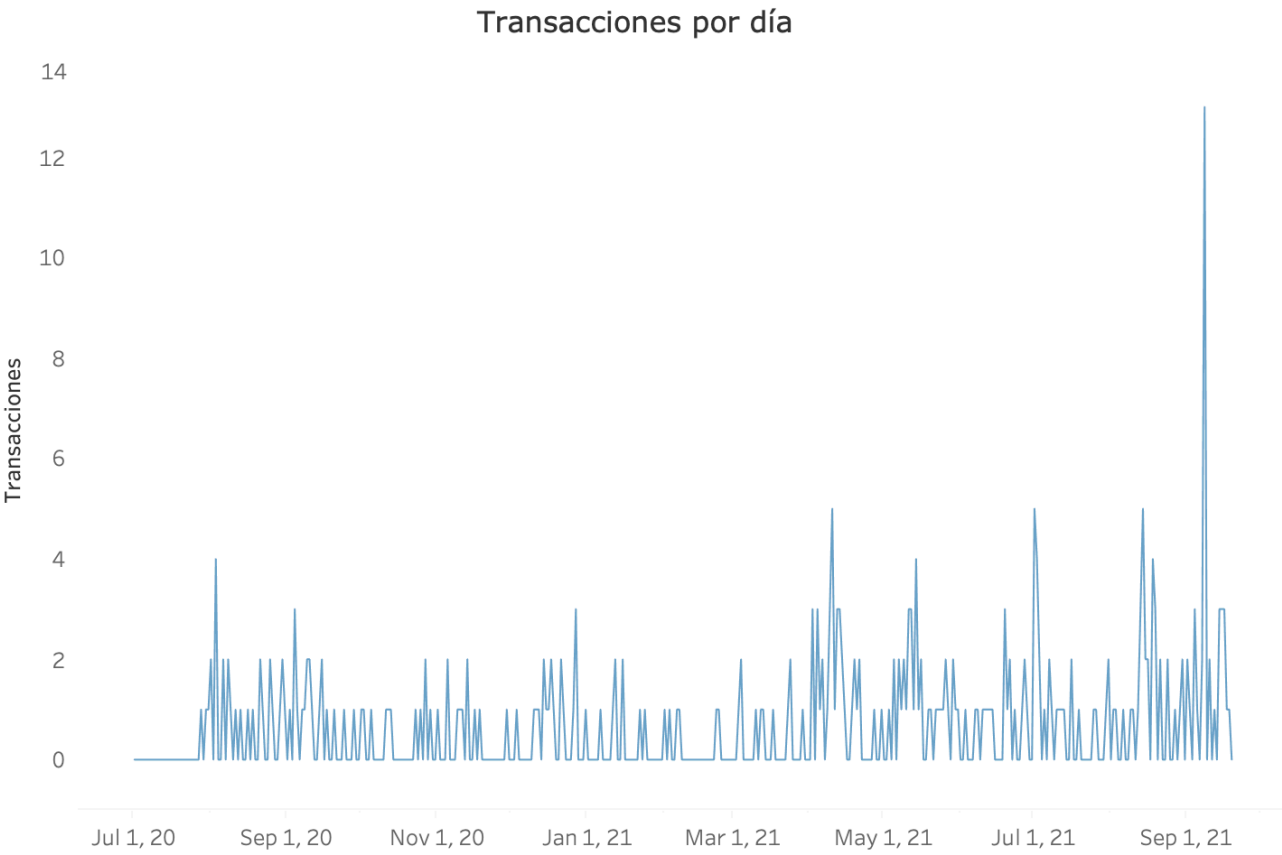
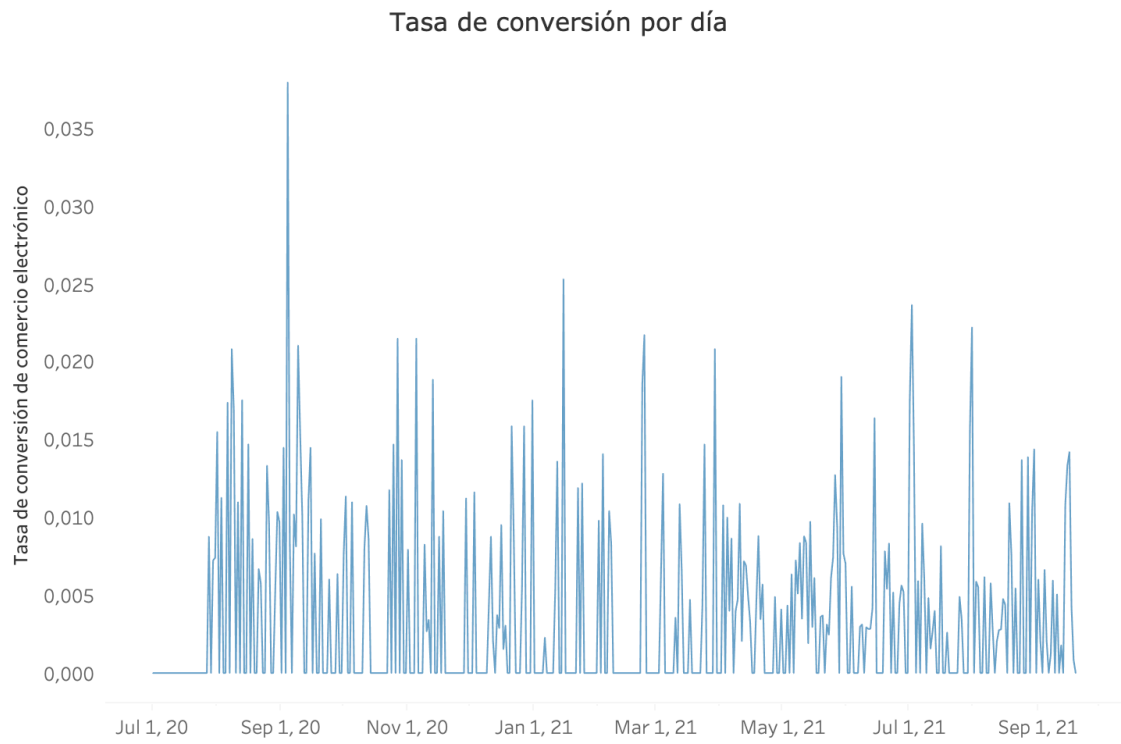


Gráfico 5: Cantidad de transacciones diarias en la página web.

En segundo lugar, y con el foco puesto en el objetivo planteado, se decidió evaluar la tasa de conversión de la página. El gráfico 6 presenta esta tasa de manera diaria en el último año.



*Gráfico 6: Tasa de conversión diaria en la página web.*

Por otro lado, al ser una página web que vende productos, se quiso ver el comportamiento de los mismos. Para ello, se realizó la tabla 1, que muestra el TOP 10 de productos más vendidos en el último año. Se puede observar que hay productos que se repiten, pero se diferencia un cambio en el color del mismo. Además, la venta de estos productos representan aproximadamente el 90% de la facturación. Otro dato importante a remarcar es que se compra una unidad de cada prenda por compra, lo que demuestra que los clientes de BLA son sus consumidores finales.

Producto	Ingresos del producto	Cantidad	Precio medio	Cantidad media
MYSTIC HOODIE	\$ 331200	36	\$ 9200	1,00
FLASH HOODIE dress (Negro)	\$ 133392,71	17	\$ 7846,63	1,00
FLASH HOODIE dress (Plata)	\$ 83893,69	11	\$ 7626,7	1,00
POWER polera (Negro)	\$ 79500	10	\$ 7950	1,00
ATOMIK JKT (Negro)	\$ 66235,77	6	\$ 11039,3	1,00
BUBBLE SWEATSHIRT (Gris)	\$ 66220	7	\$ 9460	1,00
BUBBLE SWEATSHIRT (Negro)	\$ 66220	7	\$ 9460	1,00
MORRISON Mono. (Verde militar, S)	\$ 59600	4	\$ 14900	1,00
BUBBLE SWEATSHIRT (Plata)	\$ 56760	6	\$ 9460	1,00
STAR HOODIE	\$ 51000	5	\$ 10200	1,00
	<b>\$ 2246130,89</b>	<b>416</b>	<b>5399,35</b>	<b>1,01</b>

*Tabla 1: Características del TOP 10 de productos más vendidos.*



En una cuarta parte de este análisis, se decidió analizar a los usuarios de BLA. Para ello, en un primer lugar se buscó identificar a los usuarios activos en la plataforma. El gráfico 7 muestra la evolución de estos y cómo se diferencian en función de la cantidad de días que pasan. Como era de esperar, la diferencia en cantidad de usuarios activos va aumentando en función del tiempo.

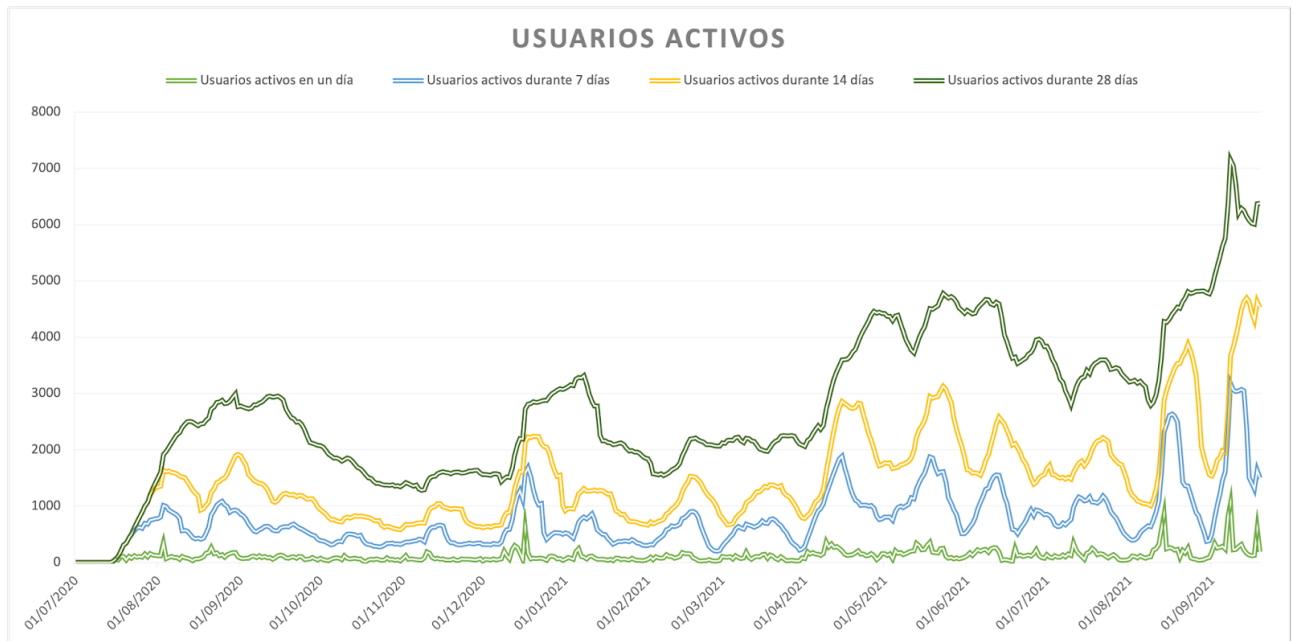


Gráfico 7: Análisis de los usuarios activos.

Luego, se buscó ver cuáles eran las características de estos usuarios. El gráfico 8 muestra la distribución en función del sexo. Se puede observar que un 88% de los usuarios son mujeres y sólo el 12% son hombres, a pesar de que la mayoría de los productos son unisex.

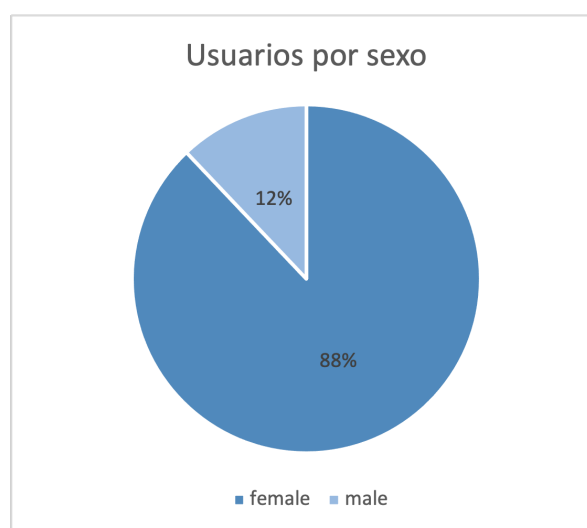


Gráfico 8: Usuarios por sexo.

Además, el gráfico 9 muestra cómo se distribuyen según la edad. En este caso, se puede ver cómo el mayor número de clientes tienen entre 25 y 34 años de edad, seguidos por aquellos que se encuentran entre los 35 y 44 años.

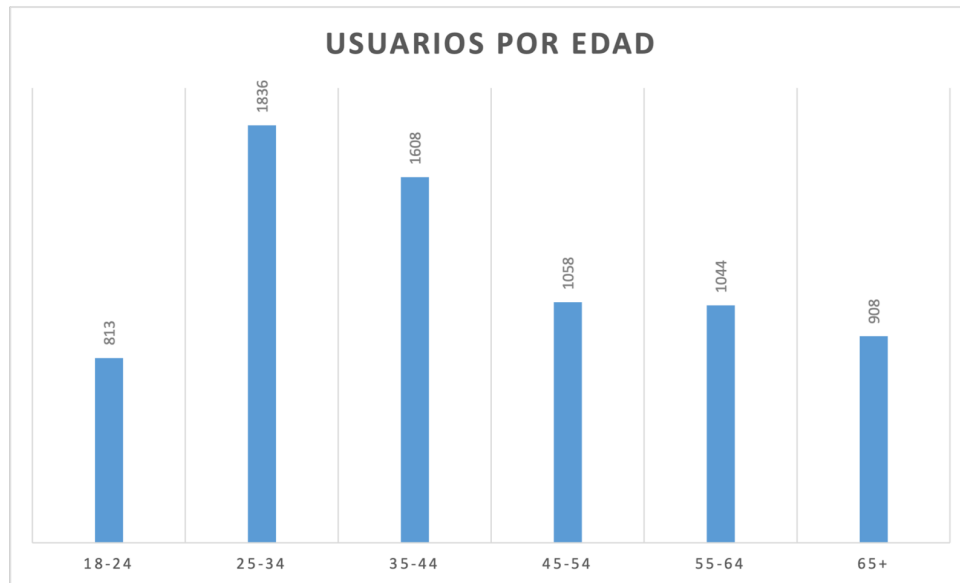


Gráfico 9: Usuarios por edad.

Para poder analizar mejor el comportamiento de estos usuarios, se realizó la siguiente tabla. Se puede ver que el rango entre 25 y 34 representa tanto al mayor número de usuarios como de nuevos usuarios. Sin embargo, los que presentan una mayor tasa de conversión y tienen un mayor número de ventas son aquellos de entre 45 y 54 años. Esto se puede deber a que los verdaderos clientes son sus hijos, pero son los padres quienes pagan.

Edad	Usuarios	Usuarios nuevos	Sesiones	Porcentaje de rebote	Duración media de la sesión	Transacciones	Ingresos	Tasa de conversión
25-34	1836	1672	3739	60,12 %	81,06	7	55286,50	0,19 %
35-44	1608	1472	3096	60,89 %	79,64	9	63513,47	0,29 %
45-54	1058	946	1990	57,59 %	91,46	10	101266,09	0,50 %
55-64	1044	950	2037	55,72 %	95,86	1	18960,00	0,05 %
+ 65	908	825	1865	57,32 %	82,27	3	36141,71	0,16 %
18-24	813	745	1500	59,07 %	75,84	3	16353,00	0,20 %
	<b>7267</b>	<b>6610</b>	<b>14227</b>	<b>58,82 %</b>	<b>83,93</b>	<b>33</b>	<b>291520,77</b>	<b>0,23 %</b>

Tabla 2: Características de los usuarios según su edad.

Por último, una variable relevante que brinda Google Analytics son los segmentos con intención de compra. Para analizar la intención de compra de los usuarios de BLA Concept, se construyó la siguiente tabla. Se puede ver que del TOP 9 de los usuarios de BLA, se presentan 8 que están interesados en todo lo relevante a indumentaria. Y, que el mayor número de usuarios se encuentra en el segmento de indumentaria femenina. Un dato que llamó la atención a la hora de

realizar este análisis, es que uno de los segmentos con mayor intención de compra al cual pertenecen los usuarios de BLA son aquellos interesados en los servicios financieros y bancarios. Y, además, son estos los que presentan la mayor tasa de conversión, con un 0,77% y el segundo número mayor de ingresos.

Segmento con intención de compra	Usuarios	Sesiones	Porcentaje de rebote	Páginas/sesión	Duración media de la sesión	Transacciones	Ingresos	Tasa de conversión de comercio electrónico
Apparel & Accessories/ Women's Apparel	2144	4480	58,62 %	2,75	84,22	7	\$ 76770	0,16 %
Apparel & Accessories	1253	2643	56,79 %	2,82	86,05	1	\$ 11800	0,04 %
Apparel & Accessories/ Shoes	1033	2193	58,91 %	2,74	91,79	1	\$ 6900	0,05 %
Apparel & Accessories/ Outerwear	891	1847	59,66 %	2,75	84,04	3	\$ 29260	0,16 %
Apparel & Accessories/ Shoes/Athletic Shoes	496	978	59,51 %	2,68	78,21	1	\$ 63,53	0,10 %
Apparel & Accessories/ Activewear	494	965	60,10 %	2,71	72,19	2	\$ 19023	0,21 %
Financial Services/Banking Services	316	778	41,00 %	3,68	230,99	6	\$ 46206,5	0,77 %
Apparel & Accessories/Men's Apparel	303	616	52,44 %	3,21	111,17	2	\$ 30410	0,32 %
Apparel & Accessories/ Shoes/Boots	293	651	60,83 %	2,81	89,74	1	\$ 13700	0,15 %
	<b>18366</b>	<b>36668</b>	<b>54,15 %</b>	<b>3,00</b>	<b>117,20</b>	<b>104</b>	<b>\$ 1026658</b>	<b>0,28 %</b>

Tabla 3: Características de los usuarios según segmento.

## Tienda Nube

La primera complicación a la hora de analizar los datos de Tienda Nube es el hecho de que se tuvo que combinar los datos de ventas, clientes y productos manualmente utilizando el email como identificador del usuario y el código de producto como identificador de este. Una vez que se tuvieron los datos combinados, se procedió a analizar la variación de los precios con el tiempo, analizando cómo estos habían sido impactados por la inflación y devaluación del peso argentino.

Número de orden		Email	Fecha	Estado de la orden	Estado del pago	Estado del envío	Moneda	Subtotal de productos	Descuento	Costo de envío	...	País	Última compra	Última compra date	Usuario Registrado	Suscripción al newsletter	Marketing	Marketing (actualización)
0	627	msg1212@hotmail.com	2021-10-17	Abierta	Recibido	No est-empaquetado	ARS	14900	2980.0	600.0	...	Argentina	18/10/2021	2021-10-18 00:00:00	NO	NO	Acepta	15/10/2021
1	626	paularepe@gmail.com	2021-10-16	Abierta	Recibido	No est-empaquetado	ARS	3200	640.0	0.0	...	Argentina	17/10/2021	2021-10-17 00:00:00	NO	NO	Acepta	17/10/2021
2	625	victoriacerator@gmail.com	2021-10-16	Abierta	Recibido	No est-empaquetado	ARS	10900	2180.0	0.0	...	Argentina	17/10/2021	2021-10-17 00:00:00	NO	NO	Acepta	15/10/2021
3	625	victoriacerator@gmail.com	2021-10-16	Abierta	Recibido	No est-empaquetado	ARS	9900	1980.0	0.0	...	Argentina	17/10/2021	2021-10-17 00:00:00	NO	NO	Acepta	15/10/2021
4	624	marielarianati1@gmail.com	2021-10-16	Abierta	Recibido	No est-empaquetado	ARS	11200	2240.0	600.0	...	Argentina	16/10/2021	2021-10-16 00:00:00	NO	NO	Acepta	16/10/2021

Imágen 8: Head de la base utilizada.

A continuación pueden verse la evolución de los precios promedios en pesos.



Gráfico 10: Precio promedio anual en pesos

Para resolver este problema, se recolectaron datos históricos de la cotización del dólar y se llevaron los precios de toda la mercadería de BLA al valor del dólar del día en que se vendió. Esto pudo hacerse ya que prácticamente todos los productos de BLA están hechos con materiales importados que componen casi la totalidad de sus precios. A continuación se puede ver el nuevo gráfico de evolución de precios en dólares.



Gráfico 11: Precio promedio anual en dólares

Es importante destacar que la variación de este gráfico no se explica por un aumento del precio en dólares, sino que está relacionado a que solo está mostrando los precios promedios de los productos vendidos, por lo que un aumento de demanda para un producto más caro podría definitivamente generar ese efecto. Puede verse, de todas formas, una gran estabilidad cuando se lo compara con sus precios en pesos.

Una vez resuelto el problema de los precios, se procedió a eliminar las variables que no aportaban al caso (ej. Identificador de la transacción en el medio de pago) o que contenían un gran porcentaje de valores nulos que los hacían estadísticamente insignificantes (ej. Código de tracking de envío). Terminado ese proceso, se aseguró que todas las variables respetaran sus tipos de datos correspondientes y se pasó al análisis de sus relaciones.



Gráfico 12: Correlación entre variables numéricas.

Observando la correlación entre las variables se descubre lo que ya era de esperarse. Esto es que todas las variables que aportan a los precios (Precio del producto - Total - Total solo producto - Total Dólar) están altamente correlacionadas. Esto se debe a que la mitad de estas fueron creadas justamente a partir de la otra mitad, para poder resolver el problema inflacionario previamente mencionado. Se definió entonces que se descartarán tres de estas variables, dejando únicamente a Total Dólar ya que es la única que muestra los precios de manera estable y no es afectada por factores externos a nuestro proyecto. Fuera de estas variables, no se detectó ningún otro caso de correlación alta entre variables.

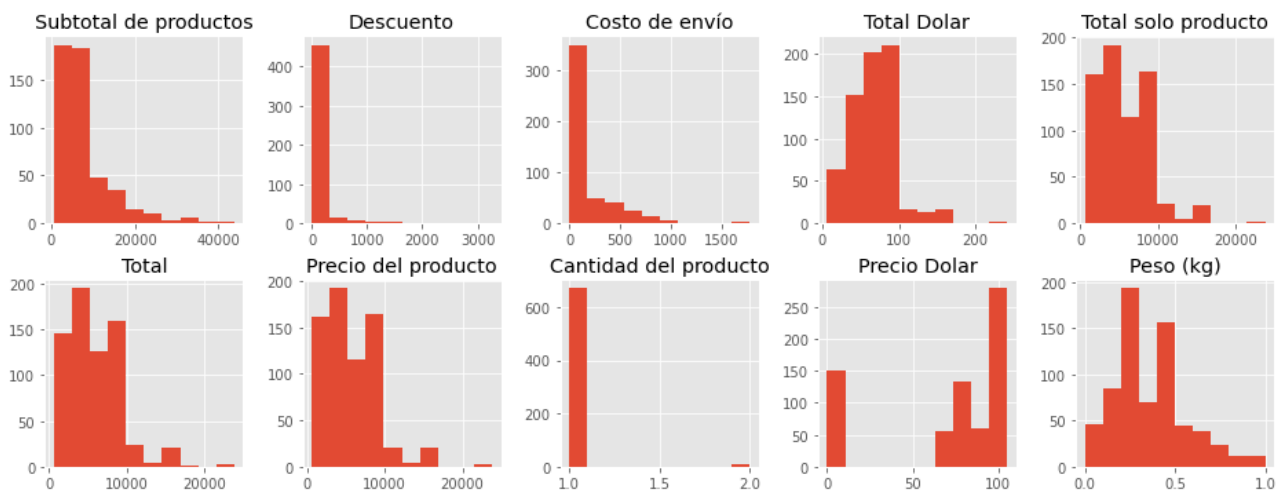


Gráfico 13 Histogramas de variables numéricas.

A partir de la imagen previa, se llegaron a diversas conclusiones que permitieron entender aún más el comportamiento y estado del negocio de BLA. Se concluyeron 3 ideas principales.

En primer lugar, se detectó que todas las variables de precios o totales mostraban una distribución similar. Esto indicaba que existían muchos productos a un precio relativamente similar mientras que otros pocos tenían un precio significativamente más elevado.

En segundo lugar, se reforzó la idea de que BLA no ofrece muchos descuentos ni liquidaciones de sus productos. Esto fue claro cuando se observó que casi todos los registros de descuentos tenían un valor de 0.

Por último, se encontró que muy pocas compras tenían más de un producto, creencia que fue completamente errónea y estaba completamente relacionada a la estructura de la base de datos. Está solo mostraba un producto por registro, por lo que cantidad de producto solo será mayor a 1 si un cliente compra dos unidades de lo mismo (algo que solo se dió para la venta de tapabocas). Al analizar esto nuevamente basándose en el número de orden, el siguiente fue el histograma obtenido.

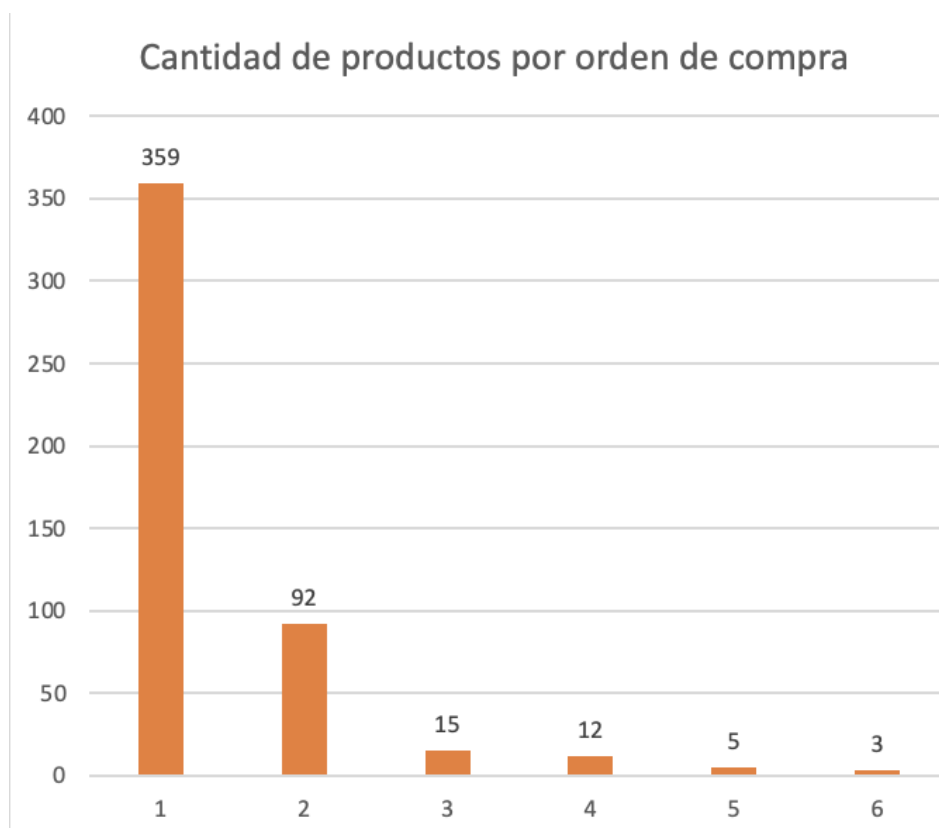


Gráfico 14: Histograma cantidad de productos.

Una vez analizadas las variables numéricas, se pasó a estudiar las frecuencias de las variables categóricas más importantes para nuestro proyecto. Estos son los canales utilizados, las opciones de envíos, los estados de pago y métodos y si el usuario es o no registrado.

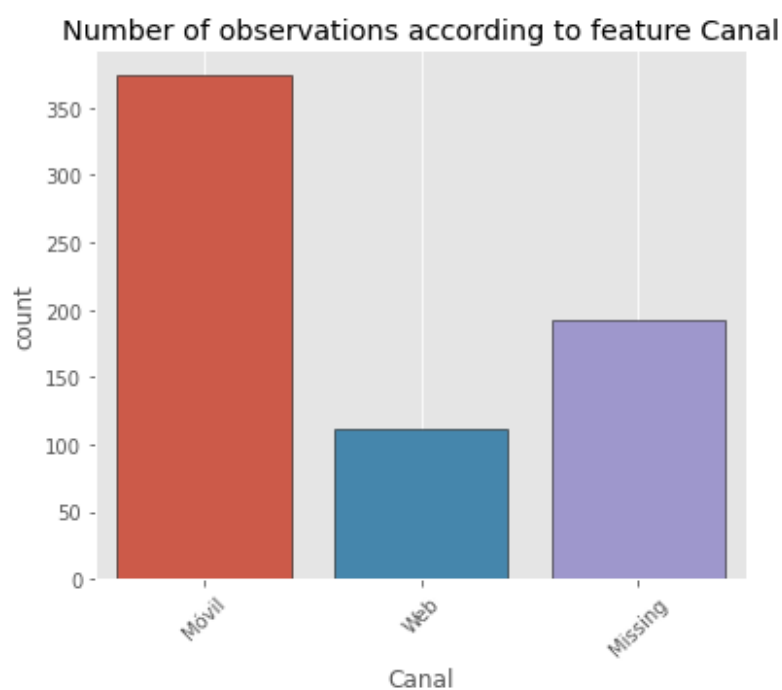


Gráfico 15: Frecuencia de los canales.

Number of observations according to feature Envío sin cargo

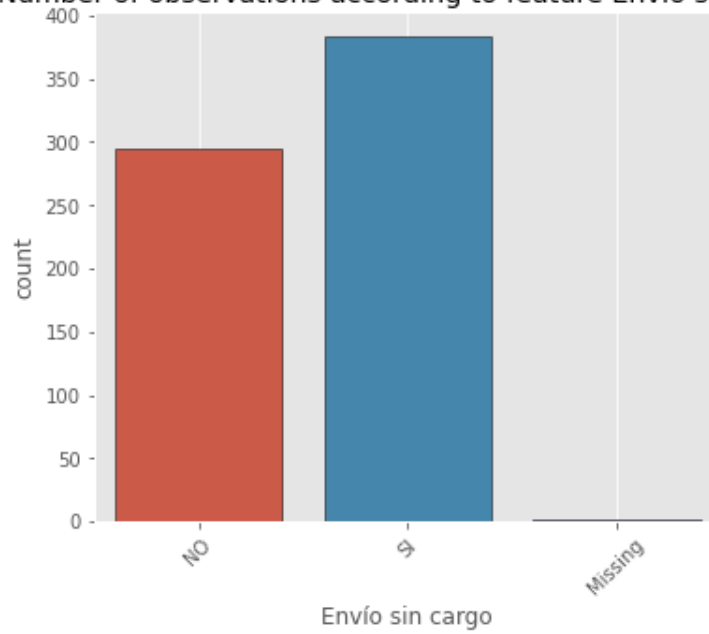


Gráfico 16: Frecuencia de los envíos sin cargo.

Number of observations according to feature Estado del pago

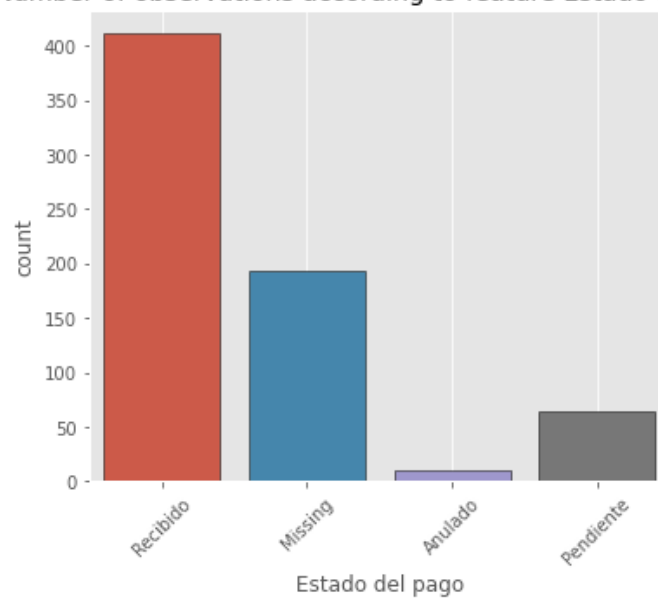


Gráfico 17: Frecuencia del estado de pago.



Number of observations according to feature Medio de pago

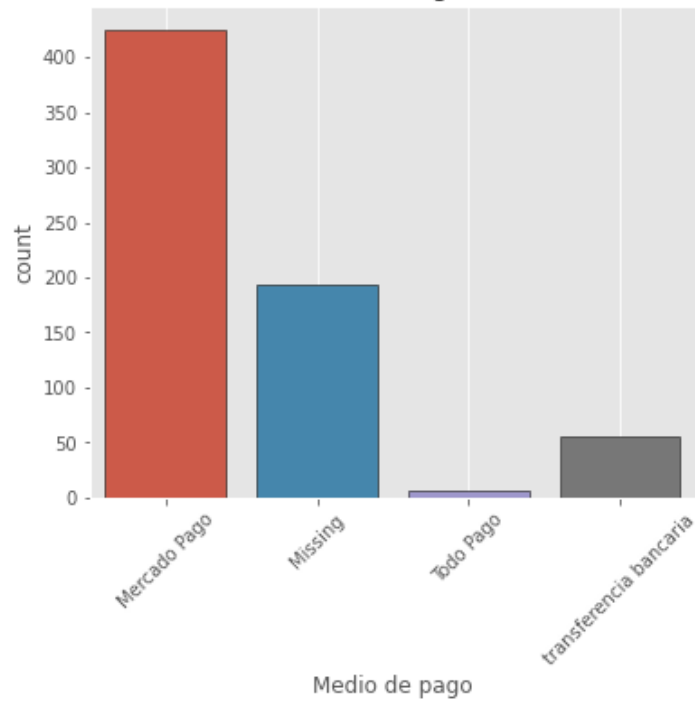


Gráfico 18: Frecuencia del medio de pago.

Number of observations according to feature Usuario Registrado

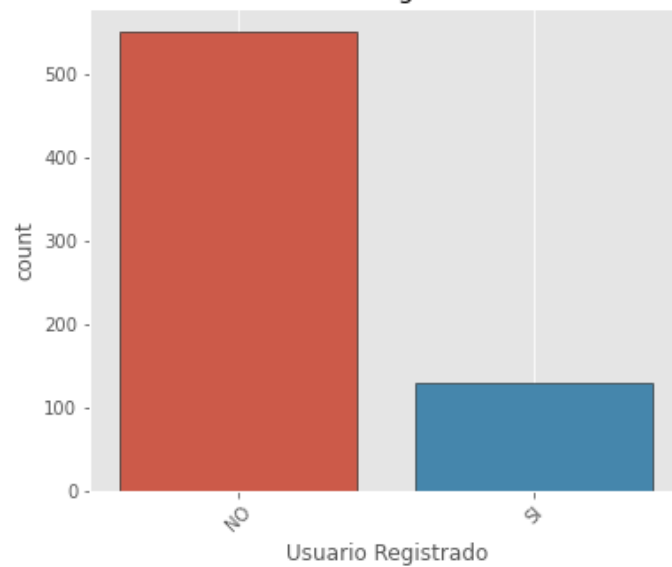
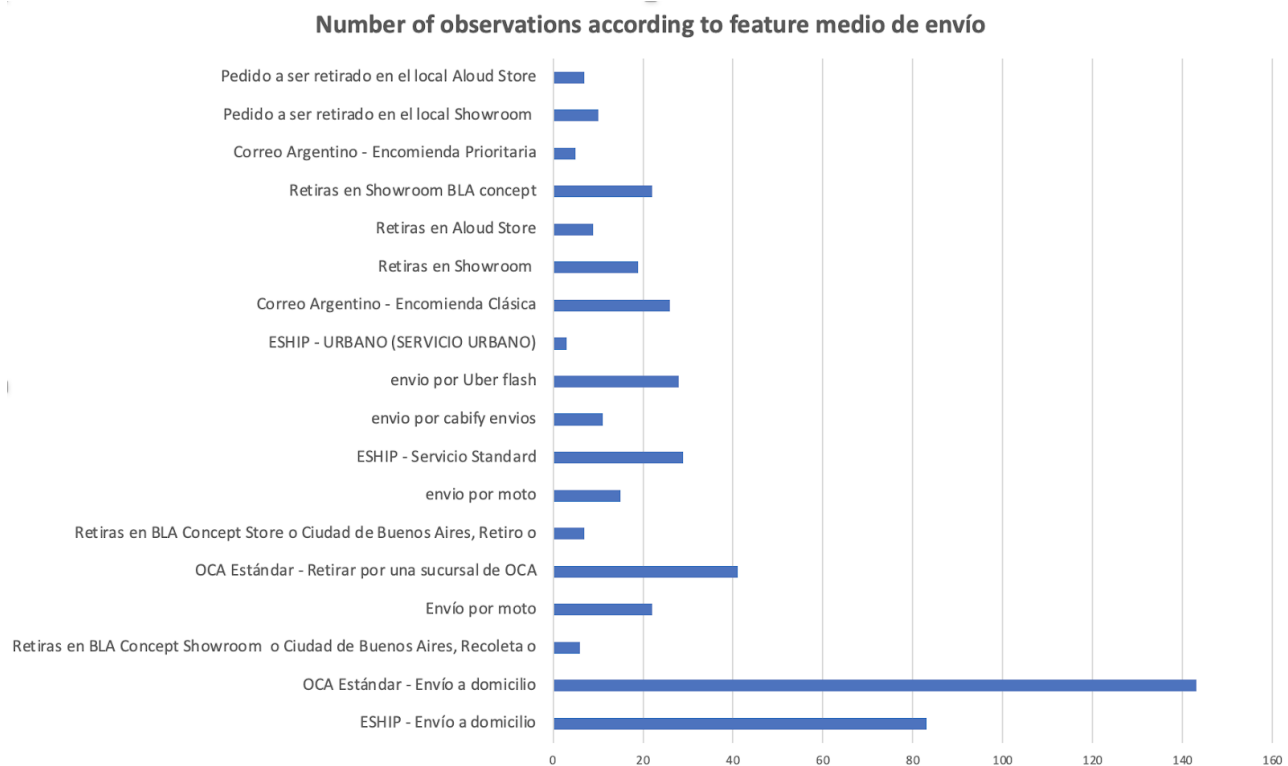


Gráfico 19: Frecuencia de los usuarios registrados.



*Gráfico 20: Frecuencia de los medios de envío.*

A partir de los gráficos de frecuencias, se pudo llegar a diversas conclusiones. La primera, con respecto a los canales (gráfico 15), es que la mayoría de las visitas y compras a BLA se hacen a través de dispositivos móviles, aunque también existen muchos casos en los que no se tiene el canal original. Se espera que este dato esté más completo una vez que se complete la conexión entre TiendaNube y Google Analytics.

Analizando los envíos (gráfico 16), se encontró que no existe tanta diferencia entre las órdenes con envío sin cargo y las que no ofrecen ese servicio. Se esperaba que más gente aprovechara de este beneficio pero la diferencia fue de menos de 25%. También se analizó el medio de envío (gráfico 20) en el cual se encontró que gran parte de las ventas se envían por OCA Estándar e ESHIP a domicilio, aunque existen también otros 16 medios de envío y todos fueron elegidos, al menos, más de 2 veces.

Observando los medios de pago (gráfico 18) y sus estados (gráfico 17), se encontró que la gran mayoría de los usuarios eligen pagar por Mercadopago en vez de transferir el dinero o utilizar Todopago. Con respecto a los estados de pago, se vio que la mayoría de estos estaba recibido, mientras que un grupo menor seguía pendiente, probablemente tratándose de las compras realizadas un período muy corto. Es importante resaltar que en este caso también existen muchos datos sin clasificar y que la unión con Google Analytics servirá para resolver estos problemas.

Por último, se buscó analizar cuántos de los clientes estaban registrados (gráfico 19). Para registrarse, un cliente debe crear una cuenta dentro de la página de BLA, a la que puede acceder más tarde y revisar sus interacciones previas. Observando las frecuencias, se encontró que era ampliamente superior el número de clientes no registrados, algo que no es bueno para BLA ya que limita la cantidad de información que se obtiene y la llegada al cliente.

## Planteo de Hipótesis

Una vez hecho el EDA, surgieron algunas preguntas o hipótesis en función de los resultados:

- ¿Puede ser que el motivo por el cual el mayor número de ventas se realiza a personas de entre 45 y 54 años es que el cliente real son sus hijos, menores de edad y sin tarjeta?
- El idea de BLA Concept es la de crear prendas que, en su mayoría, sean unisex. Sin embargo, el mayor porcentaje de clientes son mujeres.
- Mayor número de abandonos en el proceso de compra durante la etapa 2, antes de completar los datos.

## Business Case

El objetivo del proyecto es poder identificar las características de las personas que abandonan el carrito de compras, y también de aquellas que se convierten en clientes. Esto permitirá a BLA generar estrategias para recuperar esos carritos perdidos, y, al mismo tiempo, retener o atraer a las personas que tienen el perfil de ser consumidor de la marca.

Lo planteado es que, una vez finalizado el proyecto, la tasa de conversión del checkout aumente de un 18% a un 25%. En lo que va del año, fueron creados 1263 carritos, y sólo 233 de ellos terminaron en una compra. Esto se puede observar en la imagen 9.

Teniendo en cuenta que 233 carritos representan un 18% del total de carritos creados, una tasa de conversión del proceso de compra del 25% representaría un total de 316 ventas concretadas. Considerando, además, un ticket promedio anual de \$ 11.193, con el cambio propuesto se aumentaría la facturación en, aproximadamente, **\$ 929.000**, un 40% de la facturación actual.

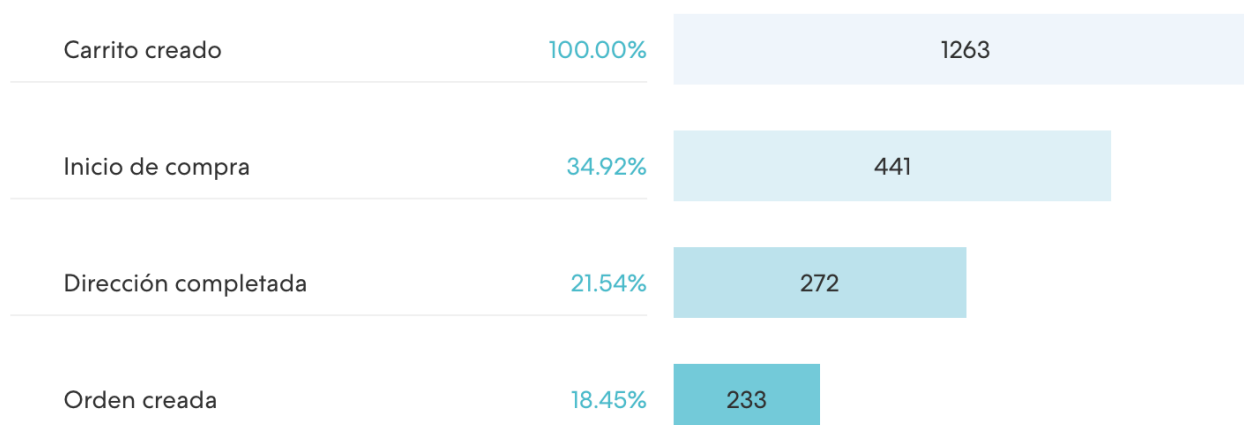


Imagen 9: Conversión del proceso de compra.

## Cambio de Enfoque

Al llegar a la etapa marcada con un círculo negro en la imagen 10, se detectó que faltaba algo esencial para poder llevar a cabo el abordaje de los carritos abandonados: la conexión del ID de los usuarios de TiendaNube en Google Analytics. Sin esto es imposible rastrear al usuario que abandonó el carrito de compras, lo que imposibilita la primera propuesta hecha.

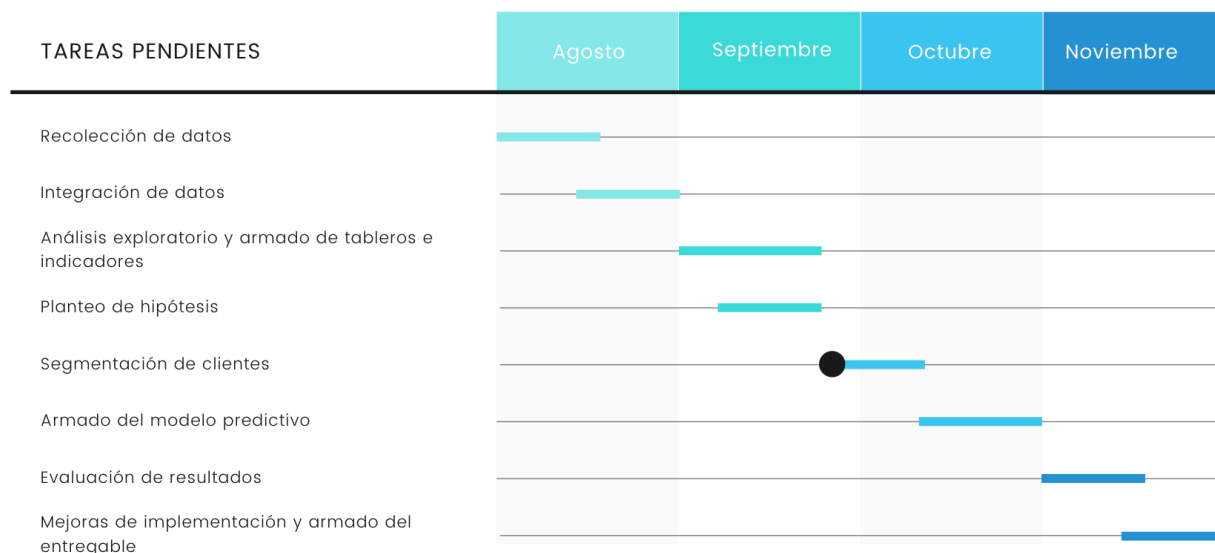


Imagen 10: Primer Gantt del proyecto.

Como solución a este problema, se plantea una segunda y diferente solución: hacer un trabajo enfocado en **pre-media**. De esta manera, no sólo se estará enfocando en la última parte del funnel de ventas, los carros abandonados, sino que se atacará, en consecuencia, a todas las etapas. El nuevo objetivo planteado consiste en darle la posibilidad a BLA de targetear a las personas que realmente son las que más van a convertir en la página.

Para llevar esto a cabo, se harán clusters con aquellas personas que tienen una alta incidencia de compra. Esto va a permitir identificar cuáles son aquellos clientes que le darán a BLA un mayor rendimiento, ya que al utilizar los resultados se verá un aumento en la performance de pre-media.

De esta manera, se presenta el Gantt actualizado para este nuevo enfoque.

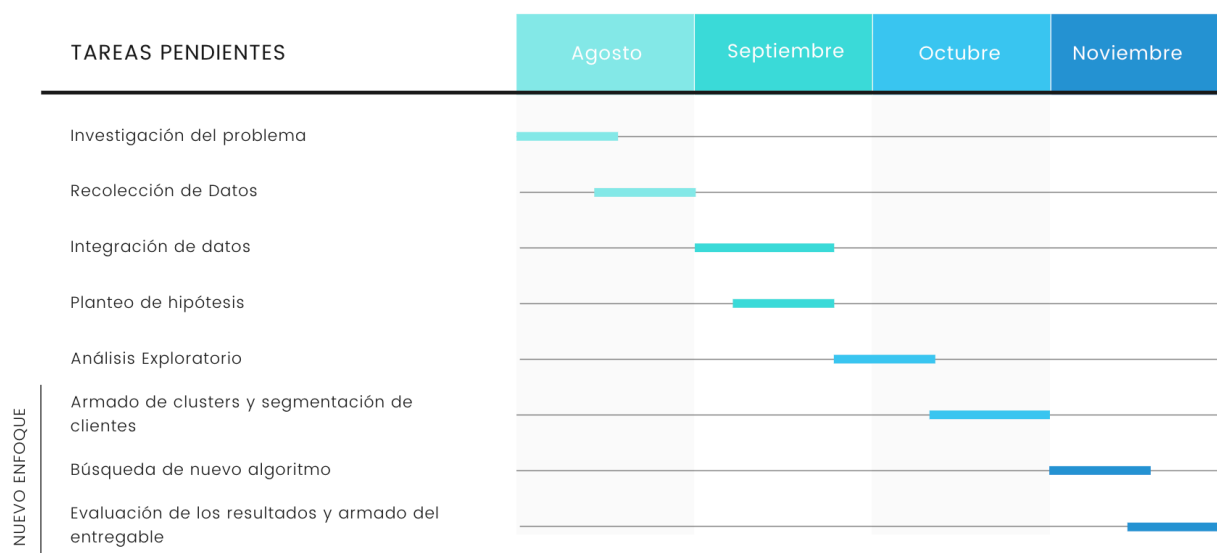


Imagen 11: Gantt actualizado del proyecto.

El procedimiento llevado a cabo en cada etapa se detalla a continuación.

- **Investigación del problema**

En esta primera etapa, de dos semanas de duración, lo que se hizo fue identificar cuál era una problemática actual de BLA Concept a la cual se le podía dar una solución.

- **Planteo de la hipótesis**

Luego, lo que se hizo fue ver cómo se podía llevar a cabo el proyecto para brindar una propuesta rentable para BLA.

- **Recolección de datos**

En esta etapa fue cuando se obtuvo el acceso a la cuenta de TiendaNube de BLA y a Google Analytics. Se obtuvieron y crearon diversas bases de datos con la información brindada por ambas plataformas. Todo esto ocurrió en el plazo de dos semanas.

- **Integración de datos**

La semana siguiente, se procedió a integrar los datos que fueron obtenidos en la etapa de recolección, para así poder llevar a cabo el análisis.

- **Análisis Exploratorio**

En esta etapa, de dos semanas de duración, lo que se hizo fue un análisis exploratorio de los datos ya formados en la etapa anterior. Se pudo ver cómo se relacionaban las variables y cómo era su comportamiento.

Algo importante a remarcar, es que en esta etapa fue cuando se analizó en profundidad los datos brindados por Google Analytics. La conexión entre la página de TiendaNube y Google Analytics fue hecha apenas BLA Concept salió como marca. Sin embargo, se pudo identificar en esta etapa que algo que no se había realizado era la conexión por ID de usuario entre la página web y la plataforma de Google. Esta conexión lo que permite ver es la información a nivel usuario. Es decir, permite ver cuáles fueron los todos movimientos en la página diferenciado para cada persona que ingresó e interactuó con ella.

- **Armado de clusters y segmentación de clientes**

Como primer paso hacia este nuevo objetivo, lo que se hizo y es explicado en este informe es el armado de clusters por clientes. Esto permitirá segmentar a las diferentes personas según su incidencia de compra. Esta etapa tuvo una duración de tres semanas.

- **Búsqueda de nuevo algoritmo**

Luego de no encontrar un algoritmo que pudiera segmentar nuestros registros de manera satisfactoria, se procederá a evaluar algoritmos alternativos que brinden mejores resultados y permitan completar el proceso de segmentación de clientes.

- **Evaluación de los resultados y armado del entregable**

Una vez armados los clusters y poder así segmentar a los clientes, se buscará evaluar qué tan positivo es el impacto de la etapa anterior. Se verificará el rendimiento de los algoritmos formulados para la etapa anterior. Esto, será llevado a cabo en el plazo de dos semanas.

Además, se realizará el informe final con el procedimiento que fue llevado a cabo y los resultados obtenidos.

A continuación, se detalla el paso a paso y los resultados obtenidos en la etapa de **armado de clusters y segmentación de clientes**.

## Business Case actualizado

Dado el cambio de propuesta, también se generó un cambio en el rendimiento que la implementación tendrá para BLA. Al ser un proyecto enfocado en pre-media, el impacto se verá reflejado en la tasa de conversión.

Actualmente, BLA tiene destinado una parte de su presupuesto a pauta publicitaria en redes sociales, sobretodo en Instagram. El resultado de esta nueva propuesta es brindarle a BLA las características de los distintos tipos de clientes que tienen. De esta manera, conociendo mejor sus cualidades, se podrá hacer un mejor targeting en cuanto a las promociones en sus redes. Esto traerá, en consecuencia, más clientes con las mismas características que los actuales, aumentando así la tasa de conversión.

El beneficio que esto trae para BLA es que, manteniendo el mismo presupuesto, podrán aumentar más su tasa de conversión, ya que se apuntará directamente al consumidor de BLA.

## Investigación

Luego de haber realizado el EDA sobre los datos disponibles y haber cambiado el enfoque del proyecto, se buscó la manera de realizar clusters que permitan segmentar a los clientes de BLA.

En un primer lugar, lo que se hizo fue armar una base de datos completa para, en función de ella, realizar los algoritmos de clusterización.

## Base de datos utilizada

En esta sección se explicará cómo se obtuvo la base mencionada, y cómo fue modificada para poder realizar los algoritmos necesarios para cumplir con el objetivo.

Como se mencionó en un principio, TiendaNube ofrece dos datasets disponibles para descargar:

- El total de clientes registrados (pueden haber realizado una compra como no).
- El total de las ventas (incluye a aquellas efectuadas y las canceladas).

Se realizó una única base de datos integrando las dos anteriores. Esto permitió ver todo lo que compró cada cliente de BLA Concept desde el momento en que creó la marca. Brindando información no sólo de la compra, sino también del usuario. La cantidad de variables en esta etapa fue de 47.

Entre las variables que provee TiendaNube se encuentran:

- ID de la compra
- Productos comprados (con su respectivo nombre y la categoría a la cual pertenecen)
- Precio de la compra (por producto, del envío, del descuento, si es que hubo, y el total)
- Medio de pago
- Datos del envío (localidad, costo, formato de envío)
- Quién hizo la compra (nombre y apellido, DNI, dirección)
- Fechas (de la compra, del pago y del envío, del registro del cliente)
- Si el cliente está suscripto (al newsletter de BLA y a marketing)

A partir de estas variables, se pudieron crear otras relevantes para el problema:

En primer lugar, como todas las compras se realizaron en Argentina, siendo su moneda el peso argentino, se tuvo que buscar una forma de comparar los precios de los productos y de los tickets a lo largo de los años, debido a la devaluación del peso. Para esto, se llevó todos los precios al valor de ese momento del dólar. Se buscó el valor histórico del dólar, relacionándolo con la fecha en que se efectuó la compra.

En segundo lugar, se debió normalizar los datos descargados. En diversos campos se encontraba el mismo término escrito de manera diferente, o campos vacíos que por error de TiendaNube no tenían su contenido.

En tercer lugar, se crearon nuevas variables, ya sea dummies o valores relevantes para el proyecto que se podían formular con las variables ya existentes:

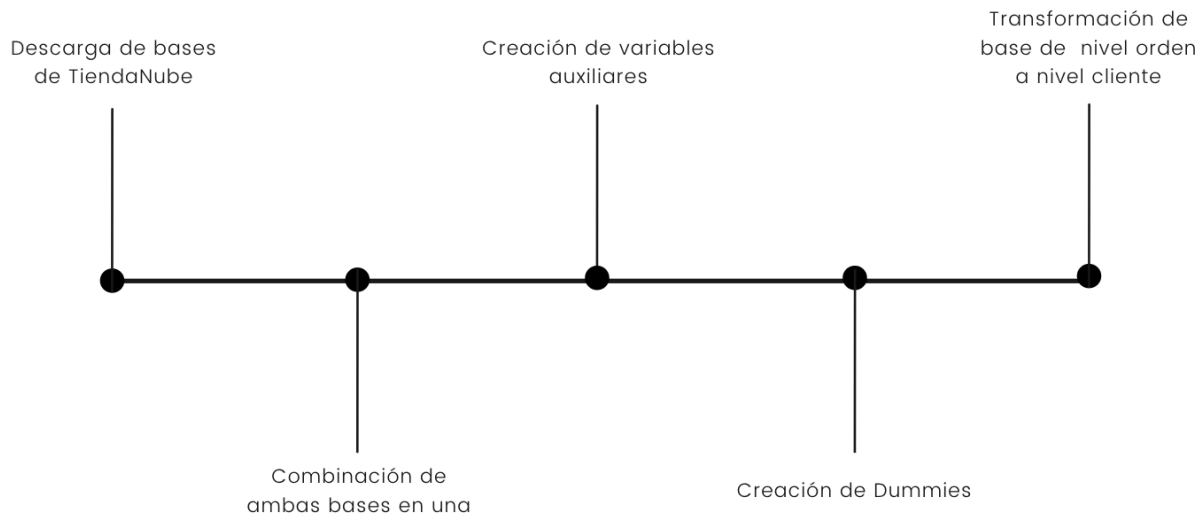
- Precio del producto en dólares.
- Precio del descuento en dólares.
- Valor final del ticket en dólares.
- Cantidad de compras que realizó cada cliente desde su registro.
- Cantidad de dinero invertido en BLA por cliente desde su registro, en dólares.
- Tiempo desde su última compra, en días.
- Variables dummies:
  - Si tuvo descuento
  - Si tuvo envío gratis
  - Medio por el cual se efectuó la compra (móvil - web)
  - Estado de la compra (cancelada - archivada)
  - Género
  - Categoría del producto
  - Tipo de envío
  - Medio de Pago

Estas variables fueron creadas para poder ejecutar los distintos tipos de algoritmos de clusterización, que se explican a continuación.

Por último, se normalizaron las variables no dummies. Para poder así obtener los resultados óptimo.



A continuación, se muestra de manera visual cómo fue el proceso antes detallado.



## Clusterización

Luego de evaluar los resultados obtenidos clusterizando el dataset, se llegó a la conclusión de que este no servía para el análisis. Esto se debía en primer lugar a que el dataset contaba con muchas variables, lo cual complicaba el estudio y análisis de sus resultados. En segundo lugar, también resultó problemático que la mayoría de las variables fueran dummies, ya que al tratarse de 1 y 0 no podían verse gráficamente los resultados de una manera limpia y que permitiera sacar conclusiones.

Al encontrarse estos problemas, se decidió que la mejor manera de proceder era realizar un análisis de componentes principales, esto permitiría reducir la cantidad de variables y evitar que sus valores fueran 1 y 0, facilitando el análisis gráfico de los resultados. Para esto, se realizó el siguiente gráfico para poder determinar cuántos componentes principales conforman nuestro dataset.

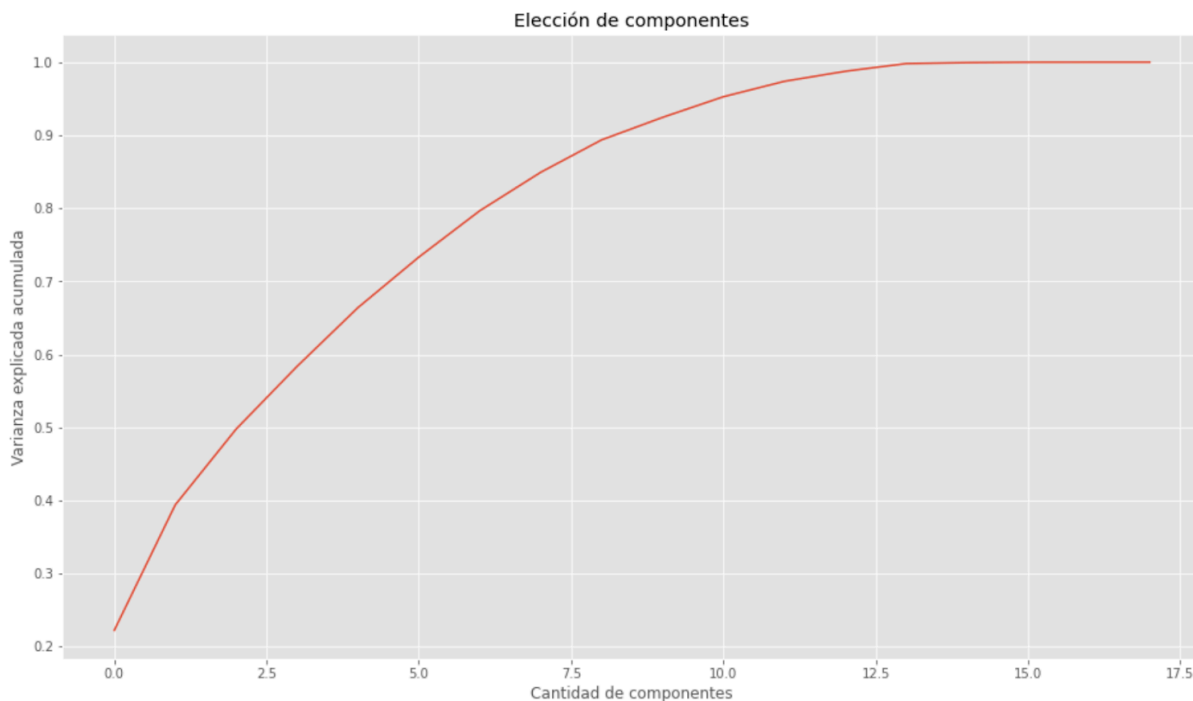


Gráfico 21: Elección de cantidad de componentes principales.

Observando el Gráfico 21, se concluyó que se tomarían 8 componentes. Se decidió esto ya que presentaba una gran disminución en comparación con las 18 variables del dataset original y porque luego de calcularse se determinó que con 8 componentes principales se explicaba el 84,92% de la varianza del dataset, lo cual es más que aceptable.

Una vez que se creó el nuevo dataset con los 8 componentes principales, se procedió a analizar cómo se componía cada componente y cuáles eran sus variables más significativas. Se calcularon entonces los valores absolutos de los autovectores de cada variable para cada componente en la siguiente tabla.

Variable	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
Capital Federal	0,271664	0,698694	0,071428	0,061595	0,067113	0,255777	0,278055	0,315573
Buenos Aires	0,025835	0,068456	0,014139	0,177284	0,075117	0,459772	0,312516	0,171577
GBA	0,216043	0,564613	0,027678	0,177818	0,090279	0,532623	0,285074	0,132445
Retiro en el local	0,230035	0,14052	0,048747	0,215408	0,0393	0,364081	0,075043	0,853561
Pago Virtual	0,055815	0,016216	0,031292	0,123272	0,072587	0,01748	0,072512	0,02258
Canal_movil	0,043107	0,055686	0,255679	0,594498	0,507037	0,073489	0,485007	0,218731
Genero_Mujer	0,085548	0,027083	0,182279	0,190378	0,496883	0,104603	0,215547	0,041166
Categoria_Hoodies	0,610601	0,290634	0,251158	0,009723	0,123704	0,017769	0,071999	0,095419
Categoria_Jeans	0,177653	0,027512	0,788869	0,141467	0,286993	0,021008	0,028792	0,004932
Categoriia_Top	0,360331	0,204242	0,355941	0,55107	0,471951	0,161931	0,030438	0,033155
Envio sin cargo	0,528061	0,170415	0,222821	0,195144	0,272814	0,177561	0,120689	0,136754
Usuario Registrado	0,009589	0,089531	0,171534	0,339755	0,268826	0,4801	0,652054	0,209681
Subscripcion al newsletter	0,011831	0,001522	0,012117	0,034262	0,007092	0,004815	0,050309	0,02551
Marketing	0,008853	0,022486	0,00103	0,053027	0,036297	0,029951	0,042742	0,020492
Total Dolar	0,010784	0,001514	0,013345	0,014018	0,0323	0,011193	0,004726	0,002343
Cantidad de compras	0,001553	1,24E-06	0,008472	0,015892	0,030295	0,012768	0,005927	0,003623
Total promedio por producto	0,010741	0,001221	0,014281	0,01205	0,017365	0,005904	0,006017	0,021974
Fecha desde ultima compra	0,004785	0,000372	0,014705	0,002851	0,013382	0,007174	0,006223	0,018736

Imagen 12: Tabla de autovectores.

Para facilitar el análisis, se formatearon las columnas con colores, viendo entonces que las celdas verdes son las que representan las variables más significativas para el componente y las rojas las menos, ya que su significancia depende del valor absoluto de su autovector. Teniendo esto en cuenta, se llegó a que las 4 variables más importantes para cada componente eran las siguientes.

Componente	Orden por significancia			
	1	2	3	4
CP1	Categoria_Hoodies	Envio sin cargo	Categoriia_Top	Capital Federal
CP2	Capital Federal	GBA	Categoria_Hoodies	Categoria_Top
CP3	Categoria_Jeans	Canal_Movil	Categoria_Top	Categoria_Hoodies
CP4	Categoria_Top	Canal_Movil	Usuario Registrado	Retiro en el local
CP5	Canal_movil	Genero_Mujer	Categoria_Top	Categoria_Jeans
CP6	Usuario Registrado	GBA	Buenos Aires	Retiro en el local
CP7	Usuario Registrado	Canal_Movil	Buenos Aires	GBA
CP8	Retiro en el local	Capital Federal	Canal_Movil	Usuario registrado

Imagen 13: Tabla de significancia de las variables.

Una vez que ya se contaba con el dataset de los componentes principales y se tenía la información como para poder descomponer cada uno, se procedió a avanzar hacia el proceso de clusterización. Se decidió trabajar con el algoritmo k-means.

Para determinar cuántos clusters habría que hacer para que el dataset quede bien segmentado se utilizó el análisis del silhouette score, que indica que tan bien ubicados están los registros dentro de los clusters, es decir, si comparten características con los otros registros ubicados en el mismo. Observando este gráfico, se pudo ver que el número de clusters con el score más alto era entre 5 y 6, así que estos fueron los números de clusters que se utilizaron como parámetros para los algoritmos de k-means.

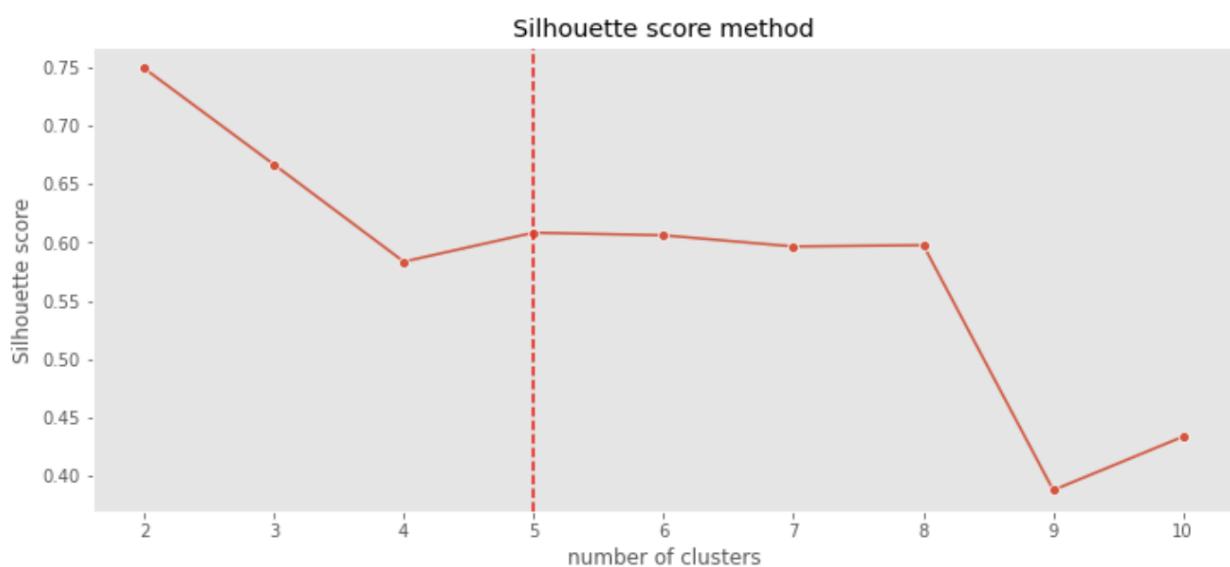


Gráfico 22: Silhouette score k-means.

Se comenzó por clusterizar la base con 5 clusters. Una vez clusterizada la base, se creó una tabla para evaluar cuántos registros habían quedado clasificados en cada cluster.

Cluster	
0	27
1	93
2	123
3	75
4	65

Imagen 3: Cantidad de registros clasificados en los 5 clusters

A diferencia del anterior intento de clusterizar sin realizar el análisis de componentes principales, donde el 65% de los registros formaba parte del mismo cluster, esta vez se ve que los clusters se distribuían los registros de manera más equitativa, algo que si bien no daba indicios sobre su exactitud fue tomado de manera positiva.

Esto se reforzó una vez que se graficaron los clusters, ya que resultaron evidentes las diferencias entre cada uno. A continuación se puede ver un ejemplo de cómo se separaban los clusters a la hora de graficarlos contra los componentes principales.

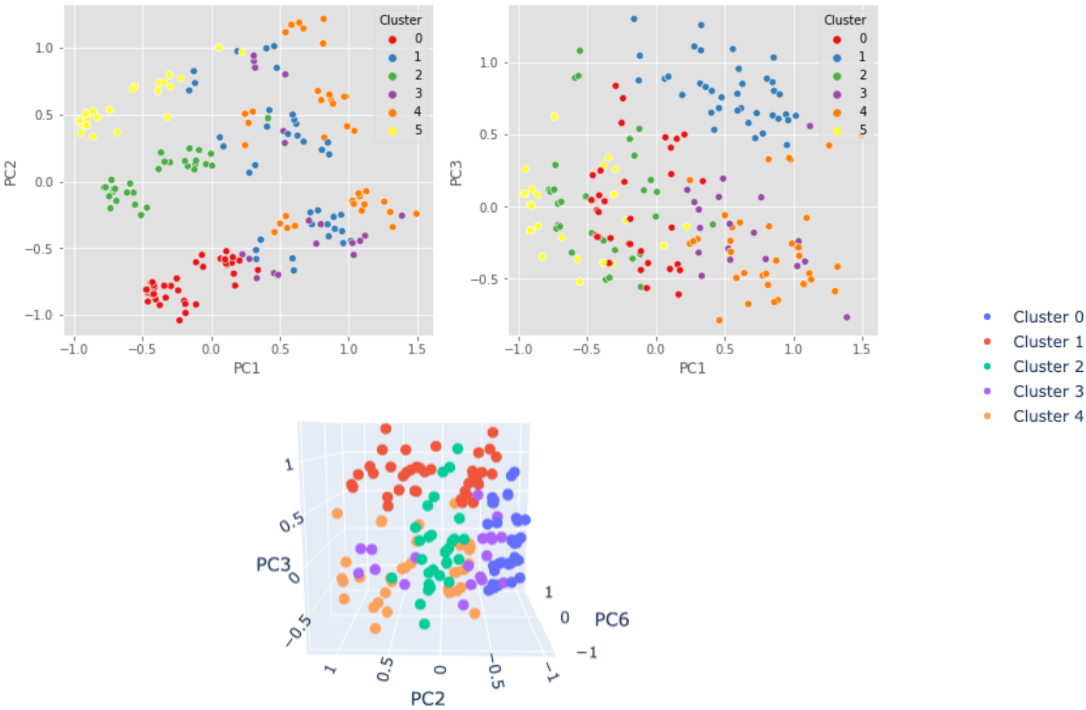


Gráfico 23: Primeros gráficos de los 5 clusters

Una vez obtenidos estos gráficos, se repitió el proceso de clusterización pero para 6 clusters. Los resultados fueron altamente similares, por lo que se decidió descartar esta posibilidad y elegir 5 como el número de clusters óptimos. Esto se debe a que la adición de un cluster no generaba mejores resultados y al tratarse de Bla, una empresa pequeña, esta no cuenta con un presupuesto alto para destinar a pauta publicitaria, ni tiene una enorme cartera de clientes que justifique la necesidad de un cluster adicional.

A continuación se analizó en profundidad los resultados para cada caso utilizando las variables originales del dataset, determinando las características principales de cada cluster.

## Resultados obtenidos

Como se mencionó anteriormente, se obtuvo que la cantidad óptima de clusters es 5. A continuación entonces se estudió la cantidad de clientes en cada uno, como son afectados por las distintas variables, y, también, como se comportan en cada cluster las variables originales de la base.

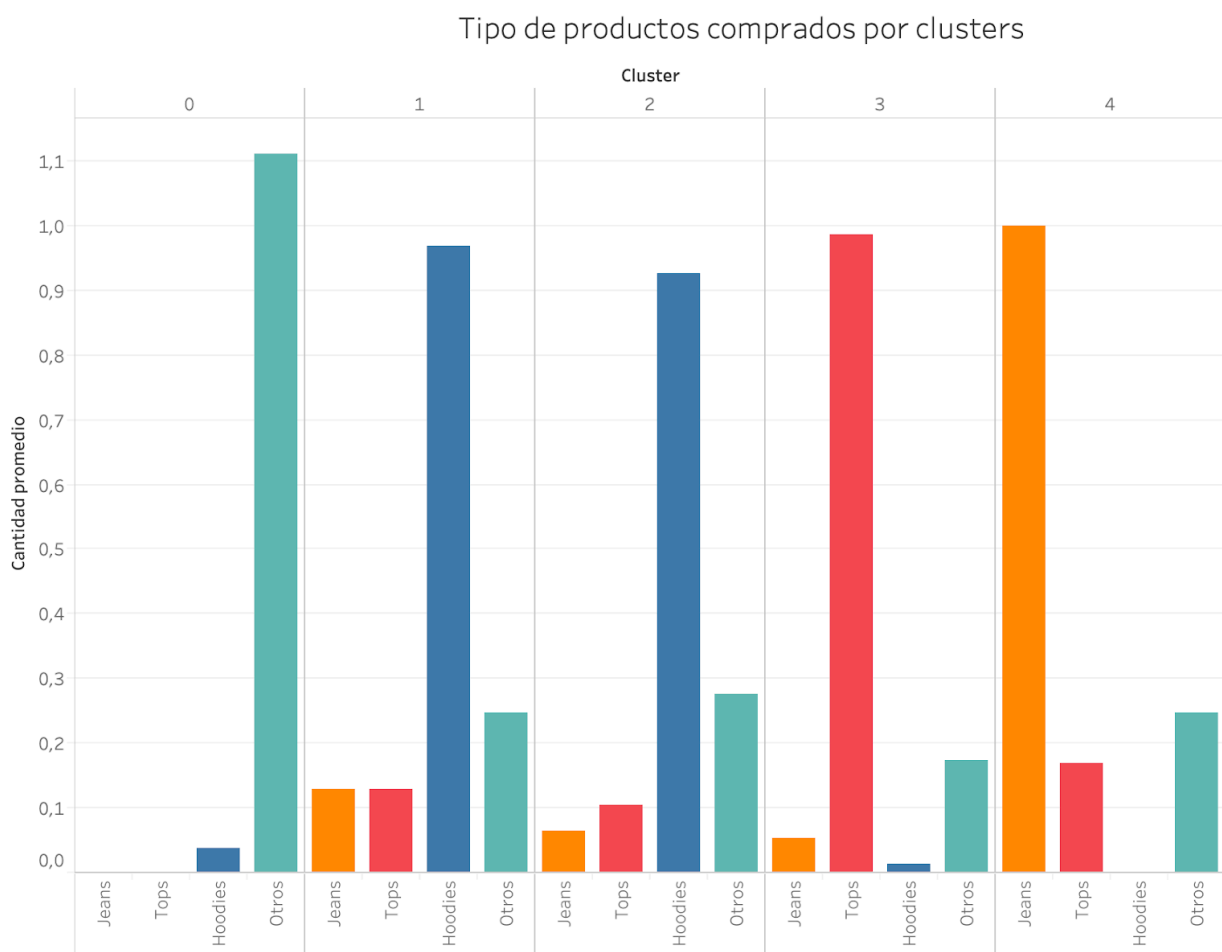


Gráfico 24: Tipo de productos comprados por cluster.

En primer lugar, observando el gráfico 24, se concluyó en que cada cluster representaba la compra de un tipo de producto distinto, con la excepción del 1 y 2 que comparten a los clientes que compran mayoritariamente hoodies. Las conclusiones entonces fueron:

- Cluster 0: Clientes que compran productos de la categoría "otros", estos son principalmente accesorios e incluyen también prendas de outwear como camperas.
- Cluster 1-2: Clientes que compran mayoritariamente productos de la categoría hoodies.
- Cluster 3: Clientes que compran mayoritariamente productos de la categoría tops.
- Cluster 4: Clientes que compran mayoritariamente productos de la categoría jeans.

### Tipo de envío por cluster

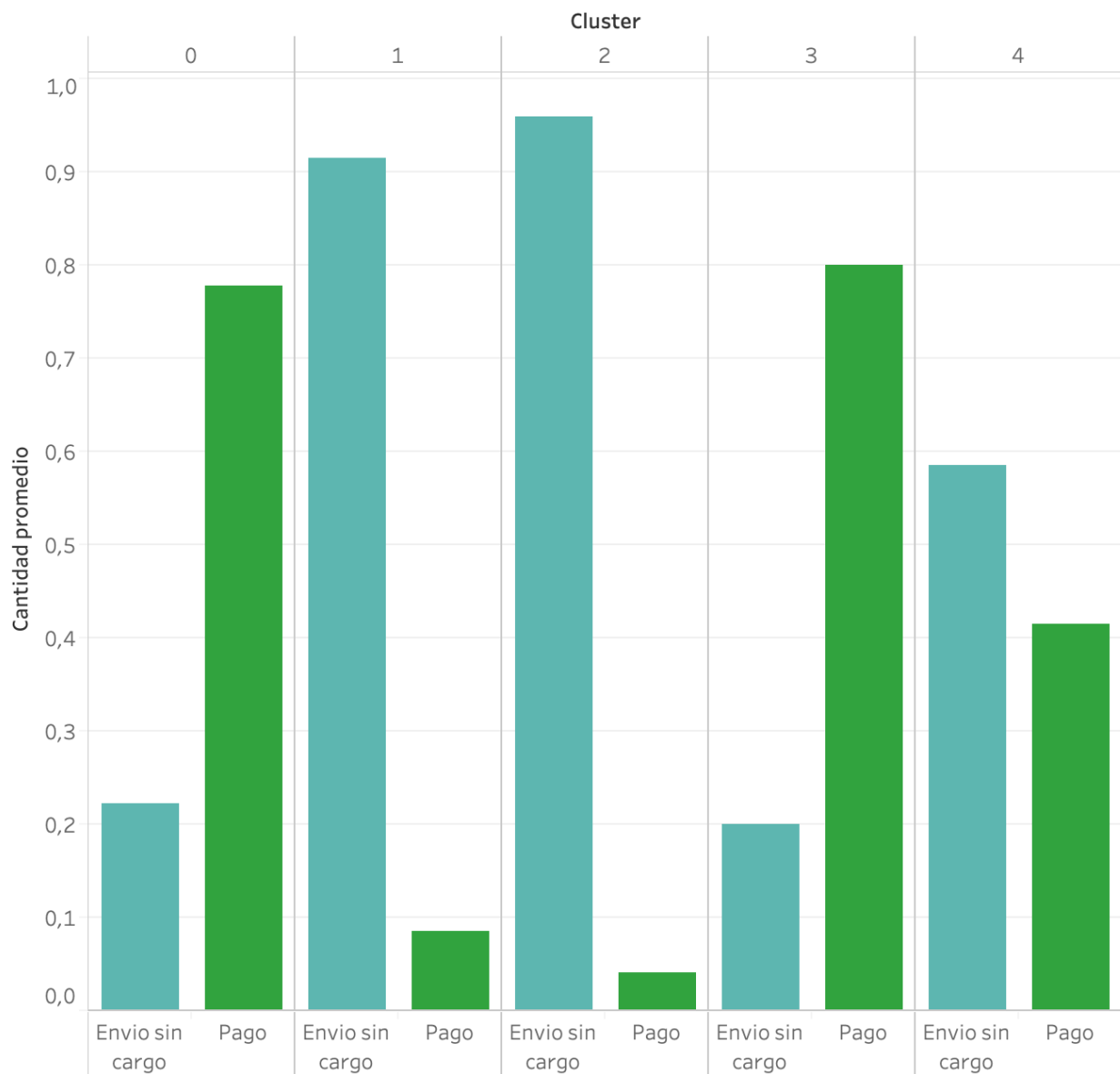


Gráfico 25: Tipos de envío por cluster.

En segundo lugar, el gráfico 25 demuestra que los clusters se comportan de forma distinta en función del tipo de envío, es decir, si es este fue pago o no. Se llegó a las siguientes conclusiones:

- Cluster 0-3: Clientes que, en su mayoría, sus pedidos son con envío pago.
- Cluster 1-2: Clientes que compran mayoritariamente órdenes mayor el precio base de BLA, por lo que se les brinda el envío sin cargo.
- Cluster 4: Estos clientes no presentan una diferencia en cuanto al tipo de envío.

Canal de compra por cluster

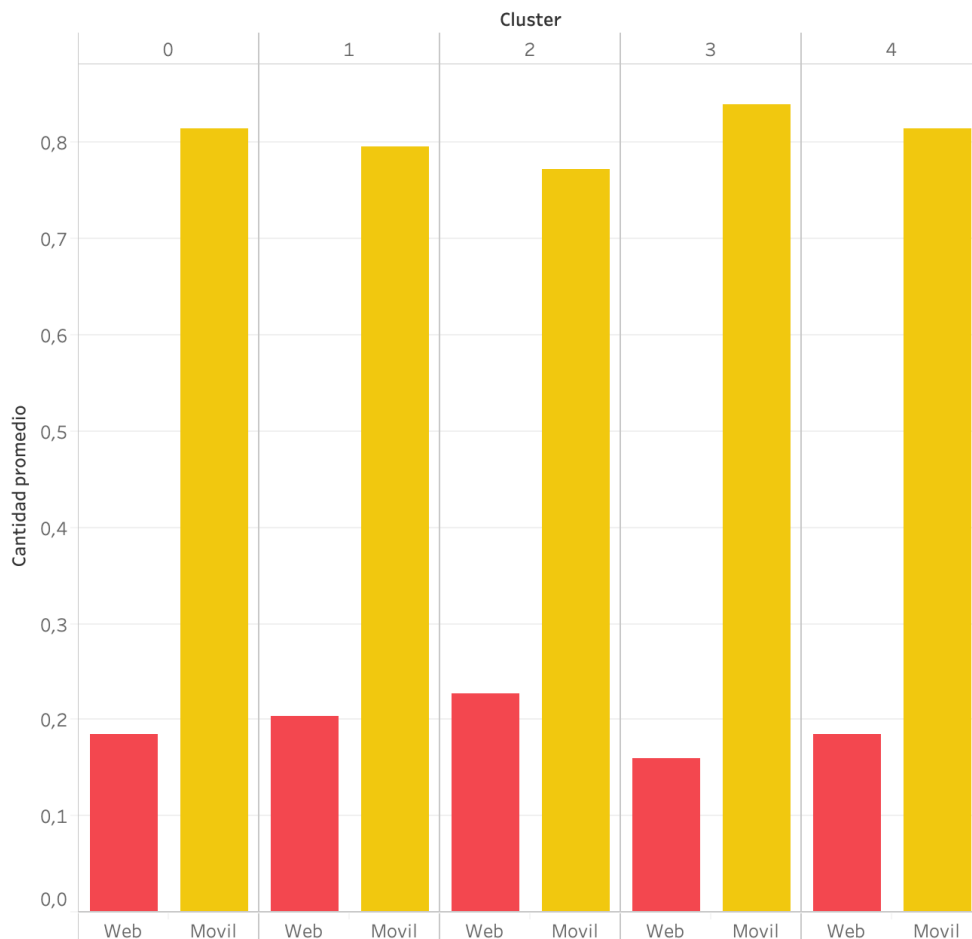


Gráfico 26: Canal de compra por cluster

En tercer lugar, se observó el gráfico 26. Lamentablemente, este no demostró brindar ningún tipo de insight o información que pudiera ser útil para entender las características individuales de los clusters. Lo mismo pasó con las variables de Retiro en el local, Usuario registrado, Suscripción al newsletter, Marketing y medio de pago. Debido a esto estas variables fueron excluidas del análisis.

## Localidades por clusters

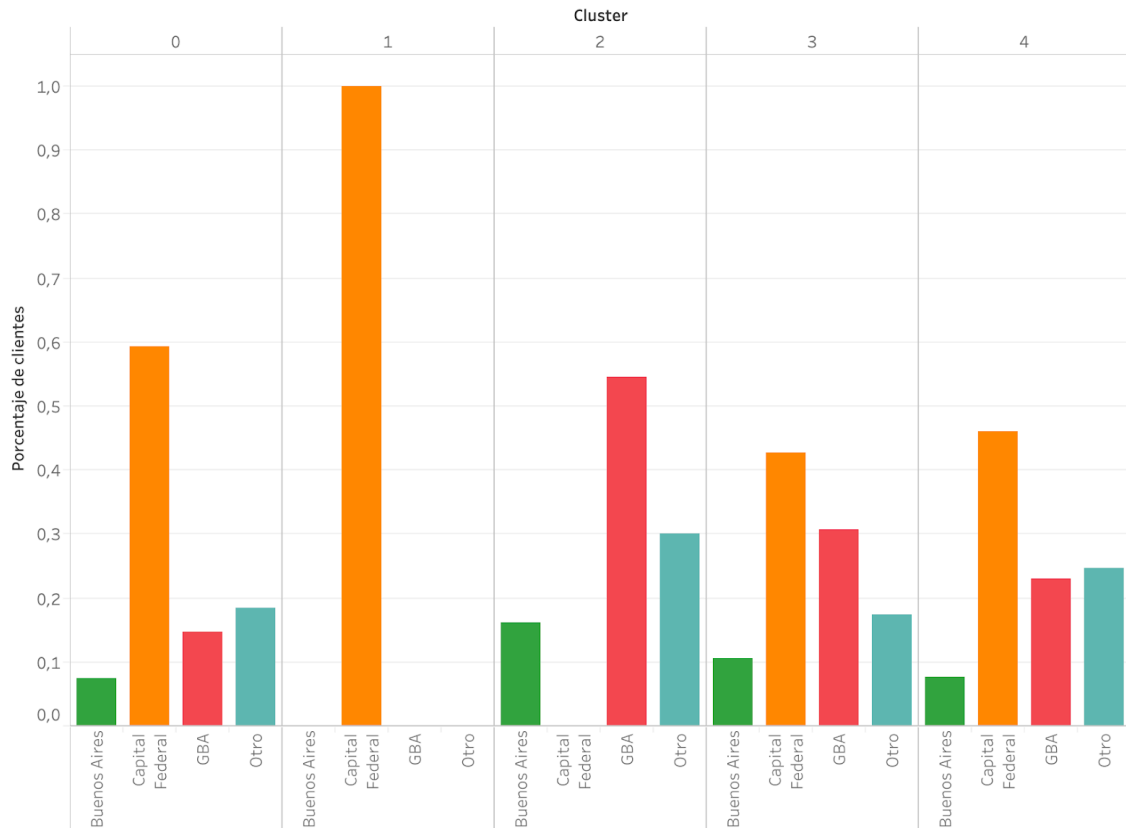


Gráfico 26: Canal de compra por cluster.

En cuarto lugar, se procedió a evaluar el gráfico 26. Este muestra las localidades de los clientes de BLA. Las conclusiones fueron las siguientes:

- Cluster 0: Clientes principalmente de Capital federal (casi 60%), pero que incluye una baja cantidad de clientes de todas las otras localidades.
- Cluster 1: Clientes pura y exclusivamente de la Capital Federal.
- Cluster 2: Clientes principalmente de el Gran Buenos Aires (alrededor de 55%) seguido por un 30% de clientes del interior del país. Único que no incluye Capital Federal.
- Clusters 3-4: Una combinación de clientes de distintas localidades. No hay una que se destaque ni supere la mitad de los clientes del cluster.



## Cantidad de productos promedio

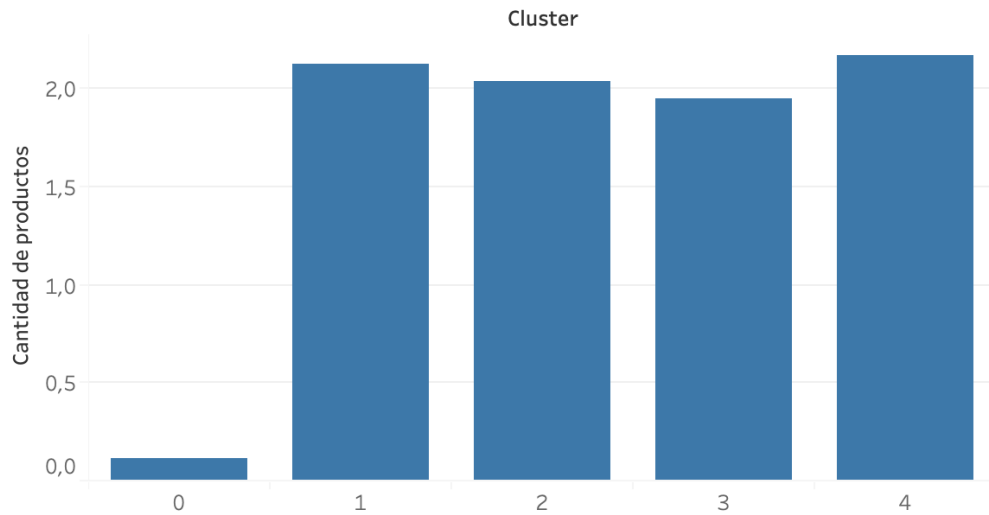


Gráfico 27: Canal de productos promedio por cluster.

El gráfico 27 muestra la cantidad de productos promedio de las compras de cada cluster. Se llegó a las conclusiones:

- Cluster 1-2-4-3: Clientes que compran más de un producto por compra.
- Cluster 0: Clientes que compran pocos productos por compra.

## Cantidad de compras promedio y máximo por cliente



Gráfico 28: Cantidad de compras promedio y máximo de compras por cliente por cluster.

El último gráfico que se analizó describe la cantidad de compras promedio y máximas para cada uno de los clusters. Las conclusiones fueron las siguientes:

- Cluster 0: El promedio de compras más bajo de todos los clusters, con un máximo de tan solo 3 compras de un mismo cliente.
- Cluster 1: Promedio de compras muy alto, con un máximo de 9 compras de un mismo cliente.
- Cluster 2: El promedio de compras más alto de todos los clusters, con un máximo de 22 compras de un mismo cliente.
- Cluster 3-4: Promedio de compras en valores intermedios entre los clusters previamente mencionados, con un máximo de 4 y 5 compras de un mismo cliente respectivamente.

## **Conclusiones**

Se llegó a la conclusión de que los clusters de clientes se caracterizan de la siguiente forma:

### **Cluster 0:**

- Mayoría de clientes de Capital Federal.
- Clientes que compran productos de las categorías accesorios y outwear.
- Pagan por el envío.
- Son aquellos clientes que menos productos por compra presentan.
- Cluster con la menor cantidad de compras por cliente.

### **Cluster 1:**

- Clientes exclusivamente de Capital Federal.
- Compran mayoritariamente productos de la categoría hoodies.
- Reciben envíos sin cargo.
- Suelen presentar más de un producto por compra.
- Cluster con la segunda mayor cantidad de compras por cliente.

### **Cluster 2:**

- Clientes principalmente del Gran Buenos Aires (no hay clientes de Capital Federal).
- Compran mayoritariamente productos de la categoría hoodies.
- Reciben envíos sin cargo.
- Suelen presentar más de un producto por compra.
- Cluster con la mayor cantidad de compras por cliente.

### **Cluster 3:**

- Compran mayoritariamente productos de la categoría tops.
- Clientes que pagan por el envío.
- Presentan más de un producto por compra, en promedio.
- Muchos realizan solo una compra.

### **Cluster 4:**

- Compran mayoritariamente productos de la categoría jeans.
- Tienen más de un item por compra.
- Muchos realizan solo una compra.

## **Medición del Proyecto**

Una vez implementado el plan de acción propuesto, se deben medir los resultados para verificar la eficacia de la solución. Para esto, se propone realizar un Split Testing. Esta, es una técnica que busca enviar distintas versiones del producto a distintos grupos de clientes y medir los resultados. De esta manera, se puede ver cuál opción es más rentable.

En el caso de BLA Concept, la propuesta de medición consiste en realizar un proceso de A/B Testing. Se seleccionarán clusters previamente mencionados y, a una parte de los clientes se les enviará pauta publicitaria destinada específicamente para ellos. Es decir, se buscarán usuarios con las mismas características que aquellos de los clusters, y se les enviará pauta diseñada para ellos. Por otro lado, a otra población de usuarios se les seguirá enviando la pauta de la manera actual.

Una vez realizado esto, se medirá en qué caso hubo una tasa de conversión más alta, y si este crecimiento fue significativo para el negocio.

## **Next challenges**

Como se planteó al principio del informe, se tuvo que realizar un cambio de enfoque a lo planteado inicialmente: el problema actual que tiene BLA Concept con el gran número de carritos abandonados. Esto fue así ya que, actualmente, no se cuentan con los datos necesarios para brindar una solución a esto.

Sin embargo, la primera propuesta a futuro que se cree que traerá buenos resultados para BLA es la de comenzar a trackear el ID por usuario de TiendaNube en Google Analytics. Si se empieza a recolectar esta información ahora, en el plazo de un año ya contarán con la cantidad de datos suficientes como para resolver esta problemática.

Por otro lado, y siguiendo el enfoque actual del proyecto, un paso a implementar luego de la finalización del mismo es la de implementar en el targeting de los clientes los resultados brindados en la última etapa de este proceso. De esta manera, se aumentará la performance de pre-media. Así se logrará atraer más personas que se identifiquen con el perfil del cliente de BLA, aumentando de esta manera las ventas.