



Maestría en Analítica para la Inteligencia de Negocios

Trabajo de Grado Aplicado

Empresa: VIPO Group

Entrega 3:

Entendimiento del negocio

Entendimiento de los datos

Exploración de los datos

Limpieza de datos

Selección de técnicas de modelado

Evaluación del modelo

Evaluación de los criterios de éxito

Presentado por:

Juan Sebastián Martínez

Miguel Omar Ríos Cabra

Ernesto Stiven Roa

Sebastián Quintero Osorio

Junio 2021

Contenido

| | |
|--|----|
| Fase I. Entendimiento del negocio | 4 |
| 1.1. Entendimiento del entorno de la compañía | 4 |
| 1.1.1. La industria de la moda en el mundo | 4 |
| 1.1.2. La tecnología en la industria de la moda | 4 |
| 1.1.3. El rol de los sistemas de recomendación en la industria de la moda | 5 |
| 1.1.4. El mercado en la industria de la moda | 6 |
| 1.1.5. VIPO Group, la nueva propuesta. | 7 |
| 1.1.6. Los retos de VIPO Group | 9 |
| 1.1.7. El servicio de recomendación como diferenciador | 11 |
| 1.1.8. El poder de los clientes | 13 |
| 1.1.9. El rol de la personalización en las ventas | 14 |
| 1.2. Determinación de los objetivos de negocio | 15 |
| 1.2.1. Problema de negocio | 15 |
| 1.2.2. Objetivo de negocio | 16 |
| 1.2.3. Criterio de éxito | 16 |
| 1.3. Valoración de la situación | 16 |
| 1.3.1. Inventario de recursos | 16 |
| 1.3.2. Recursos, supuestos y restricciones | 17 |
| 1.3.3. Terminología | 17 |
| 1.3.4. Costos y beneficios | 18 |
| 1.4. Determinación de los objetivos de minería de datos | 18 |
| 1.4.1. Objetivos de minería de datos | 18 |
| 1.4.2. Criterio de éxito objetivo minería de datos | 18 |
| 1.4.3. Alineación entre objetivos de negocio y objetivos de minería de datos | 18 |
| 1.5. Producción del plan de trabajo | 19 |
| 1.5.1. Resumen de ejecución | 19 |
| 1.5.2. Valoración de herramienta y técnicas | 20 |
| 2. Fase II. Comprensión de los Datos | 22 |
| 2.1. Recopilación de datos iniciales | 22 |
| 2.2. Descripción de los datos | 23 |
| 2.3. Análisis exploratorio de los datos | 24 |

| | | |
|------|---|----|
| 3. | Fase III. Preparación de Datos | 33 |
| 3.1. | Selección de los datos | 33 |
| 3.2. | Limpieza de los datos | 33 |
| 3.3. | Construcción y formateo de los nuevos datos | 35 |
| 4. | Fase IV. Modelado | 36 |
| 4.1. | Selección de técnicas de modelado. | 36 |
| 4.2. | Comprobación del modelo | 36 |
| 4.3. | Construcción del modelo | 37 |
| 4.4. | Evaluación del modelo | 39 |
| 5. | Fase V. Evaluación | 39 |
| 5.1. | Evaluación de los resultados | 39 |
| | Referencias | 43 |

Fase I. Entendimiento del negocio

1.1. Entendimiento del entorno de la compañía

1.1.1. La industria de la moda en el mundo

La industria de la moda en el mundo genera un interés alto en el entorno virtual y en la tecnología, incluso en las redes sociales se discute constantemente opiniones y gustos sobre los estilos, diseños y las prendas. Este sector ha jugado un papel preponderante en el desarrollo industrial y manufacturero de varias economías del planeta; en países como Colombia, el sector textil y de confecciones es pionero en el uso de capital y nuevas tecnologías. En el 2019 el sector textil colombiano representa el 8,2% del PIB industrial del país, el 21% del empleo industrial colombiano y el 9% de las exportaciones manufactureras, según cifras de la ANDI.

Uno de los datos que muestran la relevancia de la industria de la moda en el mundo es que 13 de las 100 marcas en el ranking global pertenecen a la industria de la moda según Interbrand. El sector manufacturero global del textil, la confección, la piel y el calzado genera una cifra de negocio de un billón de dólares y emplea 300 millones de trabajadores en el mundo, según el estudio The Sustainable Fashion Blueprint 2018.

1.1.2. La tecnología en la industria de la moda

Hay dos hechos claros que han dejado la historia del Retail: El primero es que la moda ha estado siempre en un cambio constante y, el segundo es que las máquinas y la tecnología han llegado con fuerza a sectores de la moda textil. La industria de la moda está viviendo constantemente estos cambios y las empresas apuestan por introducir mejoras en este sentido. Cada fase de la cadena de suministro ha sufrido una transformación por la tecnología; diseño, logística, marketing o recursos humanos son las áreas que han tenido modificaciones gracias a diversas tecnologías.

Son muchos los ejemplos de la inclusión de la tecnología en la industria de la moda en los últimos años, Grupo Inditex ha realizado apuestas en procesos como la logística para automatizar sus almacenes logísticos; Amazon instaló en el año 2017 su primer almacén robotizado, sin cajas y sin dependientes; la japonesa Uniqlo también ha apostado por automatizar sus almacenes, por lo que prescindieron del 90% del personal, pero tienen una productividad mayor, ya que las máquinas trabajan las 24 horas del día y el sistema robótico se encarga de la mayor parte del trabajo como entregar mercancía, leer etiquetas electrónicas de los productos, controlar el stock, poner el productos en cajas, etiquetar, entre otros. Otra empresa del sector moda que también es puntera de la tecnología en sus negocios es Zalando, quien se ha aliado con Google para lanzar un chatbot que aconseja a los clientes con base en una batería de preguntas. Otras grandes empresas como Nike, H&M y The North Face utilizan servicios como Alexa o WhatsApp para llegar a un consumidor más joven y acostumbrado con el uso de las tecnologías (Talentiam, 2019).

La presencia de la tecnología en la moda viene tomando más relevancia en procesos que no necesariamente están ligados a la manufactura y que desde hace algunos años se viene trabajando más en potenciar mediante estructuras tecnológicas, procesos de mercadeo y ventas; con esto aparecen modelos de negocio como el e-commerce, Ads de Google, Facebook, Instagram y las mismas iniciativas como la web semántica avanzan en entender mejor a los clientes para mejorar las

ventas o la penetración de las marcas; este tipo de necesidades ha dado pie a que se realicen diferentes investigaciones de sistemas de recomendación. Los sistemas de recomendación usan los comportamientos del consumidor y la información de los productos para identificar las preferencias de los clientes y sugerir proactivamente productos con una probabilidad de compra. Muchos estudios han sido dirigidos para desarrollar dichos sistemas de recomendación y varios sistemas prácticos han sido implementados satisfactoriamente en los negocios. (Choi et al., 2012; Koren, 2009a; Linden et al., 2003; Wei et al., 2016).

1.1.3. El rol de los sistemas de recomendación en la industria de la moda

Los sistemas de recomendación no son una metodología nueva que viene acompañada de todo este boom tecnológico que se ha presentado en los últimos años. Desde su introducción por Marko Balabanovic en 1995, los sistemas de recomendación han sido ampliamente desarrollados debido a la rápida expansión del internet y el Big Data (Zhang, 2019). Entre las esferas más extendidas, estos sistemas se han visto particularmente en tiendas online, películas, videos, música, libros, productos en general y hasta perfiles en redes sociales a quien seguir. Lo que es importante entender, es que los sistemas de recomendación maximizan el valor para el comprador y para el vendedor en un tiempo determinado. Su funcionamiento está determinado por el procesamiento de la información histórica de los usuarios, como sus datos demográficos y transaccionales; información de los productos o contenidos, como las marcas, los precios, características o contenidos similares; y las transforma en insights accionables con el objetivo de predecir qué producto puede ser interesante para el usuario y para la empresa.

Antes de hablar de la industria de la moda, cabe resaltar que los sistemas de recomendación en otras industrias han logrado establecerse dentro de las estrategias de marketing de las compañías. Hay una relación muy estrecha entre la recomendación y la venta cruzada como una de las más rápidas y rentables estrategias para crecer con los clientes actuales. No hay nada igual para una marca que lograr que un cliente existente amplíe el portafolio de necesidades con ellos mismos y por lo general, un sistema de recomendación apunta a este objetivo. Hay diversas publicaciones de noticias que reportaron en el pasado incrementos de ventas de Amazon de un periodo a otro y, algunos analistas indican que ese crecimiento tiene que ver con la forma en que Amazon ha integrado las recomendaciones en casi todas las partes del proceso de compra. Otros analistas apuntan a que el 35% de los ingresos de Amazon es generado por su motor de recomendaciones.

Al abordar específicamente la industria de la moda, el contexto puede ser muy diferente al éxito que han tenido otros productos como los libros, las películas, la música, entre otros. Wikipedia define la moda como un conjunto de prendas de vestir, adornos y complementos basados en gustos, usos y costumbres que se utilizan por una mayoría durante un periodo de tiempo determinado y que marcarán tendencia según la duración de este; aunque también la moda se refiere a algo que se repite muchas veces, en este caso las prendas de vestir. La razón por la cual cambia significativamente el contexto es precisamente porque la moda es reflejo de la vida económica, social y cultural de una sociedad; y dentro de ese entorno son una gran cantidad de factores que influyen de cierta manera la expresión de la moda. Ya no es solamente el relacionamiento de una marca y su cliente, sino también el rol que tienen los diseñadores de moda buscando imponer una tendencia, la libertad de expresión de la sociedad para enviar un mensaje, y las mismas marcas tratando de rentabilizar su negocio y la industria. Es por ello que los sistemas de recomendación han avanzado

más en otro tipo de productos y sectores, porque las preferencias comportamentales de un consumidor y los estilos de moda constituyen los elementos más importantes en toda la cadena de valor de la industria textil y de confecciones.

Según Statista (2018), los productos de moda constituyen la categoría de ventas online de más rápido crecimiento, con una tasa del 19% para el comercio minorista en su conjunto. En el e-commerce el sistema de recomendación juega un rol fundamental y gran parte de las compañías lo proyectan como la clave del éxito en este canal. Sin embargo, en los canales offline también se plantean grandes retos que lleven a trazar una nueva ruta para las tiendas físicas, especialmente en esta coyuntura en la que la crisis mundial generada a causa de la pandemia del COVID-19, incrementó de forma acelerada el crecimiento del canal digital, y en la que algunas empresas han experimentado aumentos de hasta 500% en sus ventas digitales. Pero al enfrentar la nueva realidad y entender cómo serán las tiendas físicas después de finalizar esta crisis por la pandemia, las marcas tendrán el reto de analizar al consumidor y conocer sus deseos, motivaciones y expectativas. Varias hipótesis se esbozan en torno a este contexto, las marcas querrán saber si la experiencia en las tiendas físicas estará asociada a un detalle memorable, o quieren sentirse atendidos de forma personalizada, entre otros aspectos.

1.1.4. El mercado en la industria de la moda

Inbound Cycle es una compañía líder en la generación de estrategias y metodologías de inbound marketing. Ellos plantean un concepto del marketing relacional, como un tipo de marketing que se usa hoy en día para generar valor, donde importan las ventas, pero sobre todo lo que importa es el cliente, buscando que los clientes sean fieles a las marcas, es decir, buscan que un cliente venga, compre, y vuelva una y otra vez. El marketing relacional busca un intercambio de valor, en el cual las marcas le ayudan al cliente para sacar el máximo beneficio posible y ganen ambos en esta relación. La mayoría de las empresas de todos los sectores económicos que están realizando procesos de transformación digital abordan este tipo de marketing para la generación de nuevos planes que aporten diferencial a sus productos y servicios. En la industria de la moda se destaca el punto de venta físico como un factor esencial en la estrategia del marketing relacional; Esther Moreno, Directora Comercial & Marketing de Paco Martínez, planteaba en el 2017 que el mejor CRM son las tiendas, explica que “Las personas que están en la tienda saben escuchar qué es lo que pide el cliente”, Beatriz Tomey con este mismo cargo en Intropia, puntualizaba que “Es clave que al personal del punto de venta le demos las herramientas necesarias para ser capaces de transmitir la información que queremos en cada momento, tienen que saber cuáles son los inputs que necesitan”. Es aquí donde se plantea particularmente en la industria de la moda, en que no hay que olvidarse del offline, y si bien la transformación digital es cada vez más importante para conseguir conectar con un cliente, es en este momento en que es fundamental la conexión con el punto de venta para hacer marketing relacional.

Ahora se plantea una evolución al marketing relacional, y es lo que hoy se conoce como el marketing de contexto. La agencia de marketing digital española Woko define el marketing de contexto como una “técnica que se basa en la personalización de un negocio hacia el cliente potencial enfocado en sus necesidades y gustos, haciendo uso de la inteligencia de datos para generar confianza y ser más relevante”. En otras palabras, esto permitiría incrementar la capacidad de personalización de la oferta, y es aquí donde los sistemas de recomendación encajan muy bien en este aspecto, en la

búsqueda de un producto óptimo para el cliente, en el punto de contacto específico, en un momento de relación de experiencia, y en un estatus en la relación que establece la marca con su cliente.

1.1.5. VIPO Group, la nueva propuesta.

A medida que las industrias van ingresando al mundo digital, la cantidad y la dispersión de información disponible va tomando mayor relevancia, hace que las personas acudan a fuentes de información externas a la hora de tomar decisiones de compra, con el objetivo de obtener recomendaciones que optimicen el uso de los recursos disponibles. Es aquí donde entran los Sistemas de Recomendación que se utilizan para cumplir este objetivo, con un beneficio hacia el comercio y hacia los consumidores finales. Tal como lo plantean varios autores, estos sistemas actúan analizando los perfiles de los consumidores, las preferencias de productos e información, los resultados de búsquedas de atributos o valores, entre otros, para realizar ofertas de acuerdo con el gusto o preferencia de cada consumidor (Benbasat & Zmud, 2003; Chen & Chen, 2005). La interacción que usualmente se da en este tipo de proceso es que un consumidor interactúa con un sistema de recomendación de artículos (recomendación), quien decide ver uno o determinados productos alineados con su necesidad o deseo (watchinglist), para finalmente agregar a una lista de artículos para llevar o comprar (wishinglist).

En concordancia con esto nace VIPO Group, que es una empresa que provee perfiles automáticos basados en tecnología de punta, su promesa de valor está basada en ayudar a las compañías de retail a entender mejor a sus clientes a fin de que estos entreguen una mejor experiencia de marca. VIPO ha estado en el mercado desde el año 2020, nació como una compañía de tecnología emergente basada en la propuesta de analítica que está en auge, encontrando un nicho de negocio novedoso. En busca de lograr esta propuesta de valor, VIPO ha basado su portafolio en tres grandes servicios. Su primera línea, V-Shopper comprende una tecnología de visión por computador para hacer recomendación basada en inteligencia artificial. La metodología de trabajo de esta línea se fundamenta en ser un acompañante de compras enfocándose en la experiencia de cliente sin importar el canal de venta que este elija. De este modo, VIPO ha asegurado una estrategia de recomendación visual suficiente para ser cubierta a través de dos canales de venta que todo retail tiene: el e-commerce y la venta presencial. Para la venta presencial VIPO ha destinado una estrategia basada en un elemento electrónico ubicado en tienda llamado Tótem, el cual, en pocas palabras, es una pantalla táctil de 49 pulgadas con cámaras integradas de alta definición, que además, tiene integrado el Software V-shopper. Parte de la estrategia, es que estos dispositivos se ubican en la tienda de manera estratégica con el fin de recolectar información de los clientes, enfocándose en su estilo de moda y actúa para hacer la recomendación. Como se dijo inicialmente, esta estrategia permite, no solo recomendar, sino entender el perfil de cliente. El segundo producto de VIPO es Sizetool, una herramienta virtual que ayuda a los clientes a encontrar una prenda de acuerdo con sus medidas sin tanto traumatismo, permitiendo una interfaz gráfica amigable, un alto nivel de integración tecnológica y una propuesta intuitiva de alta disponibilidad ante los retos que han llegado con la pandemia. Por último, el tercer servicio de negocio es BI-Store, el cual es un dashboard de información en tiempo real para la gerencia que permite tener información valiosa, disponible, íntegra y suficiente para la toma de decisiones.

VIPO es una compañía emergente que tiene la necesidad de crecimiento, ha puesto su foco estratégico en V-Shopper, pues esta línea tiene un sistema de recomendación robusto, que no tienen

desarrollado los competidores más cercanos del mercado, el cual, es el mostrador para los clientes a fin de iniciar un proyecto de consolidación y expansión. Es por esto que dentro de su visión estratégica, se ha propuesto mejorar la promesa entregada a los clientes a través de una mejor recomendación, pues así, se podrá entregar información más provechosa para sus clientes, aumentando el conocimiento que estos tienen de los comportamientos de los compradores y mejorando la calidad y experiencia del servicio en tienda y en internet. Esto último especialmente porque V-Shopper tiene la capacidad de integrarse con cualquier plataforma de e-commerce como VTEX, creando una oferta de valor para sus clientes de generación de información en los patrones de compra de sus consumidores.

Todo lo mencionado ha permitido que esta empresa plantee un plan de negocio basado en dos pilares fundamentales: el primero su modelo de negocio, donde VIPO ha apostado a dos grandes focos: interacciones, un modelo de precio fijo que permite un mayor control de presupuesto, en donde el cliente tiene mayor margen de maniobra en el control de sus presupuestos, permitiéndole optimizar el uso de sus recursos en otras unidades de negocio. El segundo modelo, es el que depende del porcentaje de conversión, en donde se ha explorado la posibilidad de percibir los ingresos con base en una tarifa por cada venta que se genere gracias a la solución de recomendación entregada por la empresa. De esta manera, es claro mencionar que no se venden o alquilan Tótems, ya que la estrategia base es entregar la solución para su uso, partiendo de contratos piloto con duraciones de 1 a 2 meses para dar a conocer el alcance de la solución y todos los réditos que esta tiene que ofrecer y, luego, en caso de que el piloto sea exitoso generar contratos de al menos un año con el cliente.

A partir de la oferta de VIPO se puede observar la necesidad de mejorar el modelo de recomendación, llevándolo a que aproveche la mayor información disponible. La respuesta a este problema abordado a lo largo del presente documento es un proyecto demasiado amplio. VIPO tiene como realidad actual dos grandes situaciones, por una parte, su necesidad de fidelizar a los clientes con los que ha enlazado una relación comercial y, en segunda, la proyección de crecimiento que se ha trazado en su planeación estratégica. A partir de esto y conforme a lo hablado anteriormente, se han mapeado los “pains” y “gains” de VIPO a través de un análisis de la marca, su recomendador actual y su posición con la competencia, generando a partir de estas variables el perfil actual del negocio y las posibles situaciones a abordar:

Pains:

- El sistema de recomendación actual está basado en tecnología de visión por computador que no considera información de hábitos de cliente ni tipos de clientes.
- Si bien el sistema actual contribuye a las ventas en tienda de los clientes de VIPO, no hay mucha información que permita entender a profundidad las necesidades, expectativas y preferencias de los consumidores finales en la interacción que tienen con el Tótem y de lo que están buscando específicamente.
- Hay una necesidad de VIPO para encontrar mayores beneficios en su producto y evidenciar con sus clientes actuales la manera cómo pueden contribuir a mejorar sus ventas en los puntos físicos.
- Aún no obtiene el financiamiento necesario para crecer a la velocidad definida estratégicamente.

- Sin el nivel de financiación, no tendrá el equipo necesario para realizar la expansión planteada en su plan de negocio.

Gains:

- Su propuesta enlaza la tecnología con la experiencia en tienda, lo cual es innovador y novedoso en el mercado colombiano.
- Su enfoque al cliente está fundamentado en la fidelización, por lo que se asegura de entregar soluciones ajustadas a las necesidades del mercado, pero con la personalización suficiente para mantener la satisfacción del cliente.
- La solución V-Shopper es novedosa y no está consolidada desde la competencia, lo que los hace tener un elemento diferenciador importante en el mercado.
- Su foco de ingreso al mercado es Colombia, un país inmaduro en el desarrollo y despliegue de soluciones de AI en mercados de ropa.
- A través de sus servicios, son capaces de recolectar y transformar información que con el tratamiento adecuado puede llegar a ser valiosa para sus clientes.

De acuerdo con el análisis previo, se ha evidenciado que VIPO tiene un sistema de recomendación novedoso, capaz y que genera una experiencia de compra multicanal, lo cual, lo ubica en una posición importante en el mercado, pero que, de acuerdo con los dolores, no es suficiente. Desde la perspectiva de su situación actual, no le será fácil ingresar y consolidarse en nuevos mercados. El mejoramiento y optimización de su herramienta de recomendación es el único camino de VIPO para lograr los objetivos estratégicos que se ha fijado, resaltando que el nivel de la tecnología de recomendación que ha implementado VIPO es su fortaleza más importante. VIPO en todo el proceso de recolección que lleva a cabo a través del Tótem, realiza la captura de una buena cantidad de información que puede complementar la primera capa de recomendación que actualmente tiene, lo cual, lo pone en una situación de ventaja ante las amenazas y dolores previamente resaltados. De esta manera, el camino del presente proyecto será analizar las estrategias a abordar a través de los datos para obtener una recomendación que, además de tomar lo que tiene puesto, la persona utilice los datos adicionales que recolecta el tótem a fin de entregar una nueva recomendación que se base en sus comportamientos de compra.

1.1.6. Los retos de VIPO Group

VIPO Group es una compañía que cuenta con muy poco tiempo de consolidación, pero su solución permite que medianas y grandes empresas de Retail tengan la oportunidad de conocer a sus consumidores en profundidad y brinden una mejor experiencia de marca; por lo cual es vital el desarrollo que han llevado a cabo de los sistemas de recomendación que ofrezcan una solución real.

Pero el camino que falta por recorrer es arduo, especialmente cuando se ha profundizado en el contexto del mercado y donde se señalan las compañías que compiten directamente con VIPO Group. Empresas como Syte, Donde y Vicenze ya han logrado recaudar aproximadamente USD115.6 millones de dólares en diferentes rondas de inversión durante los últimos 7 años, según cifras consolidadas de Crunchbase al mes de enero de 2021, plataforma líder que conecta las compañías innovadoras con las personas y las oportunidades.

El mercado global y sus proyecciones son lo suficientemente grandes para tener un amplio número de empresas que estén agregando valor con sus diferentes soluciones. Hay empresas con las cuales VIPO Group compite y que están logrando niveles altos de inversión, esto posiciona la empresa en un reto complejo para obtener un producto potenciado que sea más apetecido en el mercado y le permita obtener mayor financiación para apalancar el crecimiento. VIPO Group ha intentado consolidarse en países específicos como Francia, España y Colombia, sin embargo, ya compañías como Donde han manifestado su interés de ingresar al mercado colombiano, y es allí donde VIPO debe luchar por mantener sus clientes y posición de mercado.

Para satisfacer los objetivos de VIPO en este mercado creciente, pero muy competitivo, se ha establecido que el foco del presente proyecto está dado en dos partes fundamentales:

- Diferenciación contra competidores.
- Mejoramiento de la promesa de valor de VIPO.

Más que el aprovechamiento de los datos, VIPO tiene un reto mayor. Su consolidación se basará en la capacidad que este tenga de usar la información que ha recolectado y cómo esta se conjugue con su herramienta actual. Del logro de este se basará la estrategia de la compañía en los próximos tres (3) años y de este dependerá la expansión que se ha propuesto, pues, le permitirá ubicarse en una posición de ventaja con sus competidores, mientras desarrolla nuevas y mejores características en su plataforma y proyecte una mejora sustancial en los servicios que hoy presta.

Por otro lado, la pandemia que inicia en el 2020 obliga a todas las industrias a realizar transformaciones en tiempo récord, y el sector de la moda es uno de ellos, busca adaptarse al entorno, al cliente y al mercado, logrando ser más sostenible, más pequeña y rápida tras la pandemia; así lo plantea la revista especializada Modaes (2020). Muchos expertos consideran que la pandemia ha provocado que la industria de la moda ponga en el foco su cadena de valor, poniendo al consumidor en el centro de las decisiones de la industria de la moda. Anteriormente, el sector tenía otros retos relacionados con la pérdida de la relevancia de la industria en su conjunto, pero ahora los consumidores han dejado de gastar en moda en detrimento de otros sectores. Durante el confinamiento las categorías que más se vendieron fueron las pijamas y el deporte, lo que contribuye al surgimiento de la moda comfy, que incluso gigantes como Hugo Boss han puesto en stand by sus trajes rígidos para adaptarse a esta nueva tendencia. Esta situación ha llevado a que las empresas de moda preparen duros ajustes en sus redes de tiendas y busquen nuevas fórmulas, que permitan disminuir el gasto en actividades y procesos que no generan valor hacia sus productos. El foco de las compañías ha cambiado para ponerse al servicio del comercio electrónico y a la adopción de nuevas tendencias y enseñanzas que les permitan recuperarse de esta crisis. El reto es ahora, entender rápidamente los consumidores, lo que están gastando y lo que quieren cambiar; así lo afirma Luis Lara (2020), un experto en retail muy consultado en España, quien destaca entre los cambios permanentes la “aceleración phygital”, el enfoque al valor tanto por parte de las marcas como por parte de los consumidores y la importancia del small data, es decir, los insights para gestionar los negocios. Es precisamente en estos puntos, donde VIPO Group juega un rol fundamental en la industria de la moda de conformidad con los nuevos retos que le plantea unos desafíos importantes debido al cambio y las nuevas tendencias de la industria, las cuales amenazan la propuesta de valor con la que originalmente surgieron, y los obliga a encontrar mayores beneficios

tangibles para sus clientes que aseguren la continuidad de los mismos y por ende, del negocio de VIPO.

1.1.7. El servicio de recomendación como diferenciador

Se ha definido que existen tres tipos de sistemas de recomendación: el primero es el filtro colaborativo, el segundo es el filtro por contenido, y el tercero es el filtro híbrido, el cual es una combinación del primer y segundo filtro.

El filtro colaborativo usa una gran cantidad de datos de las compras y otros comportamientos para predecir lo que otro cliente está buscando. Pueden existir dos enfoques en este tipo de filtro, un enfoque basado en el usuario que funciona con base en las similitudes de otros usuarios, y otro enfoque basado en los ítems o artículos que recomienda productos basados en cómo los usuarios califican artículos similares.

El filtro por contenido se basa en las preferencias y acciones de los usuarios. Esto quiere decir que, si un usuario manifiesta interés y compra un vestido azul, entonces el sistema de recomendación le mostrará más ítems de color azul relacionados con el interés o la compra que manifiesta. Uno de los grandes problemas de este sistema es que siempre se recomienda en la misma categoría dejando de lado otras categorías en las cuales podría tener algún interés.

Cada uno de estos filtros utilizados de manera independiente tienen sus oportunidades de mejora, y tal como lo manifiesta D. Baier and E. Stüber, (2010) en su propuesta de sistemas de recomendación, la principal debilidad del filtro colaborativo es algo llamado “problema de arranque en frío”, que explica el hecho que este tipo de algoritmos tienden a tener un desempeño deficiente para los clientes y los productos con pocos datos de interacción. Y en contraste, los filtros por contenido tienden a tener un peor desempeño que el filtro colaborativo. Mientras que un sistema de recomendación por filtro híbrido sobrepasa estas limitaciones y usa ambos enfoques para tener una precisión más alta. Los enfoques de recomendación siguen un enfoque estándar en el que cada artículo o prenda es un candidato probable, sin tener en cuenta uno de los determinantes en el momento de compra como lo son las tendencias de la moda y las preferencias junto con gustos de los consumidores.

Una mirada más profunda sobre el mercado permite identificar una cantidad importante de empresas que ofrecen este servicio de manera diferente, pero todos con apoyo de la tecnología y con sistemas tradicionales de recomendación. Lo que es claro, es que la moda tiene un impacto muy alto sobre la vida de las personas en su día a día. Usualmente las personas prestan mucha atención a su forma de vestir y por ello se convierte en un aspecto vital en las culturas y sociedades. Si bien los sistemas de recomendación son tecnologías que se usan desde hace años, en una experiencia de compra presencial las personas buscaban la manera de entender las tendencias de la moda a través de las personas que los asesoraban en las mismas tiendas; ya desde allí parte la necesidad de llevar esta experiencia a un nivel con un mayor impacto.

Entre las compañías que más resaltan actualmente en el sector con este tipo de servicios se pueden evidenciar algunas de las siguientes:

| COMPANY | FOUNDED | LOCATION |
|---|---------|----------|
|  VIPO | 2020 | FRA |
|  syte | 2015 | ISR |
|  donde | 2014 | USA |
|  V i S E N Z E | 2013 | USA |
|  FASHWELL | 2014 | EU |
|  clarifai | 2013 | USA |
|  Slyce | 2012 | USA - CA |
|  catchoom | 2011 | EU |
|  markable.ai | 2015 | USA |
|  WIDE EYES | 2013 | EU |
|  Vue.ai | 2016 | USA |
|  CORTEXICA | 2015 | USA |

La más antigua es “Catchoom” que, siendo la más longeva, apenas cuenta con 10 años en el mercado. Esto nos demuestra que este es un mercado aún en primera fase de explotación, pues su nivel de inmadurez es importante. La propuesta de valor de VIPO y su producto más importante no tiene mayor diferencia a lo que Catchoom ha desarrollado en todo este tiempo. VIPO ha alcanzado una posición importante en el mercado, subsanando las debilidades de sus competidores más importantes, como el tiempo de integración, el nivel de recomendación por tecnología y el entendimiento de comprador y no comprador en tienda.

Todas las compañías que actualmente están en el negocio tienen dentro de su oferta de valor la tecnología y la inteligencia artificial como herramientas esenciales en la generación de sugerencias de compra. Sin embargo, no todas tienen un enfoque exclusivo en la industria de la moda, de hecho, compañías como Clarify, Slyce, Markable AI y Catchoom ofrecen en sus páginas web su portafolio en diversas industrias como la aeronáutica, e-commerce, manufactura, operaciones financieras, turismo y otros. Mientras que otras empresas como Syte, Donde, Vicenze y VIPO están enfocados en la industria de la moda entregando diversas soluciones de Inteligencia Artificial, Computer Vision y Natural Language Processing para lograr su cometido. Algunos de ellas ya han alcanzado cifras millonarias de fondos de inversión para desarrollar aún más sus productos; la más reciente noticia está relacionada con la compañía Syte quienes incrementaron USD\$40 millones de dólares en diciembre pasado a través de una ronda de fondos de inversión adicional, alcanzando un total de un poco más de USD\$71 millones de dólares acumulados de diferentes rondas de fondos de inversión.

El portafolio más diversificado lo puede tener tal vez la compañía Syte, la cual ha logrado conformar su portafolio de servicios especializado en la moda en 4 soluciones: Una aplicación de descubrimiento visual enfocada en soluciones que buscan conectar, por medio de inteligencia artificial, con la experiencia del cliente en las tiendas virtuales y presenciales para lograr una interacción diferente en el proceso de compra, ya sea en búsqueda, recomendación y descubrimiento de productos. En segundo lugar, una plataforma especializada para la búsqueda y clasificación de productos, en la cual realizan procesos de etiquetar con atributos e imágenes y poder aumentar el catálogo. En tercer lugar han desarrollado un marketplace único ofrecido a las marcas

con estrategias de redireccionamiento de tráfico en la web para impulsar las visitas y la compra de productos. Y por último, un carrusel de recomendación basada en la hiper-personalización de los clientes finales usando los datos en tiempo real de los compradores y la inteligencia artificial visual para predecir los productos con mayor probabilidad de conversión de compra. En comparación con VIPO no existe una brecha importante pues, la aplicación de V-Shopper tiene una aplicación de visión por computador e inteligencia Artificial que recomienda en tienda, que busca y clasifica por producto. Dentro de su alcance, VIPO no ha visto necesario establecer un Marketplace propio, pues, su estrategia se basa en la asistencia a las superficies a través de los datos recolectados y sus plataformas propias, tanto físicas como electrónicas y, la diferencia entre el carrusel y el V-Shopper, puede tener pocas aristas, más allá de que se puede afirmar que VIPO ha logrado cubrir ciertas brechas que Syte no ha observado con detenimiento como la integración fácil de sus soluciones a los sistemas de los clientes, un store de analítica para la toma de decisiones (producto que VIPO ofrece dentro de su catálogo principal) y una herramienta complementaria al V-Shopper que se encarga de minimizar la mala experiencia por talla que tienen los clientes llamada SizeTool, que ayuda a encontrar la talla perfecta con pocas preguntas a través de una interfaz muy amigable para todos los clientes sin importar su nivel de manejo de sistemas de información.

Existe otra compañía con un portafolio más reducido en su oferta, llamada Donde, esta es una compañía con soluciones muy enfocadas en el e-commerce y en ofrecer una plataforma que ayuda al comprador a disminuir su tiempo en sus búsquedas de productos. Lo hace a través de 3 enfoques: el primero, a través de una navegación y búsqueda visual de los productos; el segundo, una recomendación personalizada usando tecnología muy similar a Syte, en la cual logran conectar al cliente usando el proceso de etiquetas de los productos y las selecciones de los usuarios para crear una recomendación que encaje con el estilo de la persona; y el tercero, una suite de soluciones que recopila datos de los atributos en el mapa de experiencia del cliente para mostrar estadísticas, vistas y tendencias a las marcas en tiempo real.

Por último, se encuentra la empresa Vicenze que plantea sus soluciones de acuerdo con el canal, pensando en e-commerce, tienda presencial u operaciones de retail. Con soluciones muy similares a Donde, su servicio enfocado en el e-commerce plantea una navegación simple entre los productos, una recomendación inteligente y el uso de los datos en todos los tipos de dispositivos en los que el consumidor navega. Su servicio de experiencia en tienda busca conectar a los compradores con el mundo virtual a través de una experiencia omnicanal, de asistencia personalizada en el punto y la interacción con productos alternos desde lo que manifiestan interés en la tienda. Y su operación de retail se concentra en brindar las estadísticas y el seguimiento de tendencias necesarias para las marcas en búsqueda de una planeación más eficiente en las promociones del punto de venta.

1.1.8. El poder de los clientes

Desde cualquier punto de vista, una variable como las ventas se ha visto usualmente relacionada con el tráfico de personas en un punto de venta o en un e-commerce y, por ende, se crea una relación de conversión entre el tráfico y las ventas generadas. Los comerciantes ahora están evidenciando que el tener mayor tráfico es sólo el primer paso hacia un proceso de conversión, es

aquí donde se genera la necesidad de optimizar la conversión antes de invertir en estrategias que generen mayor volumen de visitas.

Hay comerciantes que cometen el error de utilizar sólo las promociones para mejorar las tasas de conversión, pero hay evidencia suficiente que permite concluir que la personalización de la experiencia de compra contribuye a incrementar las tasas de conversión en el proceso de venta en los puntos presenciales y en un e-commerce. Magento (2016) plantea que una de las principales razones por las cuales los clientes abandonan una compra virtual sin comprar no es por el precio, sino porque no logran encontrar lo que realmente quieren. Para ellos hay dos maneras sencillas de ayudar a los clientes a encontrar los productos correctos: la búsqueda de artículos y la búsqueda personalizada de artículos, donde la primera plantea sólo la necesidad de implementar herramientas encaminadas a la oportunidad de mayores interacciones y la segunda, ayuda a determinar mayores artículos como la categoría, tamaño, color, etc. Y como se ha planteado anteriormente, el servicio de recomendación sería la opción perfecta para VIPO de lograr una personalización de la experiencia de compra en los consumidores finales.

Para el caso de VIPO Group, una vez los consumidores finales visitan las tiendas de sus clientes y además interactúan con el tótem, se generan mayores oportunidades no sólo de información para el entendimiento del consumidor, sino también para la generación de estrategias de merchandising que contribuyan a incrementar la intención de compra de los consumidores finales. Mayores interacciones con la herramienta indudablemente contribuyen a una mayor probabilidad de compra de artículos con los cuales se ha interactuado en la herramienta.

La interacción que muestra esta oportunidad es lo que hemos llamado anteriormente la lista de deseos (“wishinglist”), es el momento en que los consumidores finales han manifestado la mayor intención de adquirir un producto porque lo ha visto y le ha agregado a la lista. Una lista de deseos es la manifestación de lo que los consumidores quieren y tienen mayor probabilidad de comprar, es decir, le permite entender a los clientes de VIPO ver casi en tiempo real cuáles son las tendencias de moda para activar estrategias comerciales con el objetivo de incrementar sus ventas. Una baja cantidad de artículos dentro de la lista de deseos indica una situación negativa, quiere decir que las personas no están reaccionando a lo que estás proponiendo dentro del sistema de recomendación.

Hay compañías que piensan que un sistema de recomendación simplemente consiste en conectar un programa y dejarlo funcionar para que haga su magia, pero lo cierto es que no todas las recomendaciones son iguales; especialmente cuando hablamos de la industria de la moda. Lo que sí es claro es que la mayoría de los sectores económicos que tienen potencial de mercado se han visto retados ante la personalización de los productos y servicios en general, es decir, entre más precisa sea una recomendación ajustado a lo que realmente un consumidor final está buscando, mejores resultados tendrán las empresas en sus ventas y en lograr que ellos efectivamente reaccionan al estímulo.

1.1.9. El rol de la personalización en las ventas

Cuando se habla de recomendación de productos personalizados, se está refiriendo cuando en un sitio se muestra una selección de productos que pueden ser únicos o ajustados a la necesidad del

visitante, basado en su perfil. No se puede hablar de recomendación cuando un individuo está viendo la misma recomendación de la persona que está al lado de ese individuo. Y aunque pueden existir diferentes enfoques de cómo abordar la personalización, desde una base que comprende historial de compras, o desde el perfil o ubicación del cliente, o buscando productos afines, entre otras.

Hay diversos estudios que muestran la fuerte correlación que existe entre las sugerencias de un producto y el impacto sobre las ventas en un determinado lugar. La personalización también ha sido mostrada como un factor que influye sobre el aumento en el nivel de lealtad de los consumidores (Cyber Dialogue, 2001; Srinivasan, Anderson, & Ponnayolu, 2002). Dos profesores en la Universidad de Montreal lograron demostrar que los consumidores son fuertemente influenciados en sus opciones de compra por recomendaciones online; sin embargo, no todas las recomendaciones en línea son influenciadas potencialmente; son los sistemas de recomendación los que tienen un nivel más alto de influencia sobre los consumidores, e incluso mayor influencia se encontró en comparación con las fuentes convencionales de recomendación (Senecal & Nantel, 2004).

Otros acercamientos a la personalización de recomendación de productos muestran que estos sistemas aumentan dramáticamente el promedio del valor de la orden en las tiendas. Una prueba de ello, son estudios que muestran datos reales sobre las experiencias que no tienen ninguna relación con un sistema de recomendación tienen un valor promedio alrededor de USD\$44,41; mientras que con una simple recomendación este número llegó a incrementarse en un 369% (Serrano, 2021). El autor destaca que es claro que entre más personalizada y comprometida sea la recomendación, mayores beneficios tendrán las tiendas en órdenes de compra más grandes. Esto nos lleva a concluir que las recomendaciones de productos personalizados tienen un mayor impacto sobre otras prácticas que no son personalizadas como por ejemplo, mostrar pruebas sociales que indican qué tan popular es un producto, u otras prácticas basadas en reglas de negocio que indican los productos que están bajos en inventarios.

La principal razón por la cual la personalización en los productos de recomendación es importante, se debe a su conveniencia y a la efectividad que tiene ajustarse a las necesidades, preferencias y deseos de los consumidores; no es tanto lo que se muestra sino a quién, cuándo y cómo. Realmente las evidencias muestran cómo la personalización puede retener y convertir clientes, y ejecutar un proceso de información constante que se está alimentando y evolucionando para lograr que las marcas cambien sus estrategias y puedan superar las dificultades que les ha traído los cambios en los últimos años y una pandemia que ha impactado de manera significativa la economía mundial.

1.2. Determinación de los objetivos de negocio

1.2.1. Problema de negocio

A partir del mercado actual y de las necesidades de VIPO en su posición de la industria, el problema de negocio se relaciona directamente con la manera como la empresa debe garantizar una propuesta de valor diferencial en el mercado, que le genere a sus clientes un impacto positivo sobre las ventas, que al final justifique la inversión realizada en V-Shopper y por ende la continuidad de sus contratos. La coyuntura actual está llevando a que los clientes del perfil de VIPO enfoquen sus esfuerzos en otras estrategias diferentes a lo que ofrece la compañía con su oferta de valor, teniendo en cuenta que el foco actual está en la redirección de los recursos a otros proyectos que garanticen

el margen en estos tiempos de crisis derivada por la pandemia. A partir de esto VIPO se ha planteado conservar los contratos actuales buscando acciones que mejoren la calidad de sus servicios y el retorno de la inversión de sus clientes.

1.2.2. Objetivo de negocio

Incrementar el porcentaje de clientes que renuevan el contrato para el uso de V-Shopper, logrando la fidelización y crecimiento del mercado de VIPO Group mediante el aumento en la venta de sus clientes.

1.2.3. Criterio de éxito

El proyecto se puede considerar un éxito si se logra un aumento en un 25% los clientes que renuevan el contrato.

1.3. Valoración de la situación

1.3.1. Inventario de recursos

A continuación, se listan los recursos de minería de datos, dispuestos para el desarrollo del presente proyecto:

Hardware:

- Equipos de escritorio con acceso a internet para el desarrollo de los avances y modelos.

Software:

- Se cuenta con licencias libres y estudiantiles del software R-Studio y compiladores compatibles con el lenguaje de programación Python.

Herramientas de comunicación:

La comunicación interna entre los estudiantes y VIPO Group se establece a través de las siguientes herramientas:

- Google Meet: Herramienta de videoconferencias usada para las reuniones de revisión de avances, dudas y propuestas con VIPO.
- Slack: Herramienta de resolución de inquietudes, transferencia de archivos de análisis y definiciones de la compañía.
- Trello: Herramienta de gestión del proyecto.

Datos:

- Se dispone de información de la interacción de 10.000 clientes con el tótem de VIPO en la tienda Falabella Colina, sin embargo, la idea es que a futuro se pueda seguir trabajando este proyecto con información de pilotos realizados en otras tiendas.

Recursos Humanos:

- Cuatro (4) estudiantes de Maestría de Analítica para la Inteligencia de Negocios.
- CEO VIPO GROUP. William Bernal.
- Equipo de bases de datos VIPO Group: Equipo disponible para la transformación de datos de acuerdo con la necesidad.

1.3.2. Recursos, supuestos y restricciones

Requerimientos:

- Evaluar la situación actual y usar los datos recolectados a través del tótem.

Supuestos:

- Los datos recolectados y la estructura presentada por VIPO es suficiente para la ejecución del proyecto.
- Los datos han sido recolectados de una fuente confiable, segura y consistente que permita establecer los análisis necesarios para edificar el sistema de recomendación.
- Los datos no han sido modificados, ajustados o manipulados por ninguna instancia, por lo que pueden ser minables y analizables de acuerdo con el objetivo del presente proyecto.

Restricciones:

- Los criterios de éxito del modelo son aplicables después de la fase de finalización y despliegue del presente proyecto, por lo que su evaluación dependerá netamente de VIPO Group.

1.3.3. Terminología

- **Carrito de compras:** Es el nombre que recibe la plataforma, módulo o plugin que se encarga de recopilar los pedidos que hacen los usuarios en una página web que cuente con una tienda en línea.
- **Skus:** Formato de identificación de las prendas.
- **Estilo:** Conjunto de rasgos peculiares que caracterizan una cosa, una persona, un grupo o un modo de actuación.
- **“Watching List”:** Listado de elementos vistos por el consumidor o cliente, que no han sido enviados al carrito de compras.
- **“Wishing List”:** Listado de elementos enviados por el cliente al carrito de compras.
- **Sistema de recomendación:** Es una herramienta que, a partir de un conjunto de criterios y valoraciones sobre los datos de los usuarios, sugiere predicciones sobre elementos que puedan ser de utilidad o valor para el mismo.

1.3.4. Costos y beneficios

El proyecto se propone por medio de la modalidad de Trabajo de Grado Aplicado en la Maestría en Analítica para la Inteligencia de Negocios. Dentro de esta modalidad, se ha establecido contacto directo con el CEO de VIPO Group para avanzar y alinear las necesidades de la empresa con el desarrollo metodológico del proyecto como también con alcanzar el objetivo final de negocio y minería de datos. Hasta esta fase, no hay ningún costo asociado con el avance de esta etapa de planteamiento del problema, desarrollo del modelo y evaluación de las oportunidades; sin embargo, poner en marcha el proyecto si requiere de unos costos vinculados a las soluciones tecnológicas que se usan para el desarrollo de los diferentes algoritmos, infraestructura para un almacenamiento y acceso óptimo a los datos, entre otros.

Tabla 1. Escenario A. Costo empresa de desarrollo del proyecto.

| ítem | Mes 1 | Mes 2 | Mes 3 | Mes 4 | Mes 5 |
|-----------------------|----------------|---------------|---------------|---------------|---------------|
| Arrendamiento | \$ 600.000 | \$ 600.000 | \$ 600.000 | \$ 600.000 | \$ 600.000 |
| Internet | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 |
| Licenciamiento | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 |
| Electricidad | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 | \$ 150.000 |
| Teléfono | \$ 70.000 | \$ 70.000 | \$ 70.000 | \$ 70.000 | \$ 70.000 |
| Agua y alcantarillado | \$ 40.000 | \$ 40.000 | \$ 40.000 | \$ 40.000 | \$ 40.000 |
| Suministros | \$ 500.000 | \$ 500.000 | \$ 500.000 | \$ 500.000 | \$ 500.000 |
| Personal | \$ 28.000.000 | \$ 28.000.000 | \$ 28.000.000 | \$ 28.000.000 | \$ 28.000.000 |
| Puestos de trabajo | \$ 60.000 | \$ 60.000 | \$ 60.000 | \$ 60.000 | \$ 60.000 |
| Equipos | \$ 63.333 | \$ 63.333 | \$ 63.333 | \$ 63.333 | \$ 63.333 |
| Total Costos | \$ 29.783.333 | \$ 29.783.333 | \$ 29.783.333 | \$ 29.783.333 | \$ 29.783.333 |
| Costo total proyecto | \$ 148.916.665 | | | | |
| Margen | 15% | | | | |
| Precio | \$ 175.196.076 | | | | |

1.4. Determinación de los objetivos de minería de datos

1.4.1. Objetivos de minería de datos

Generar una recomendación complementar a V-Shopper basada en perfiles de compradores que permita aumentar la posibilidad de conversión.

1.4.2. Criterio de éxito objetivo minería de datos

Se complementa V-Shopper con una segunda recomendación que incluye al menos el 80% de las categorías de interés del cliente y un criterio de similitud de al menos el 50% con la recomendación principal.

1.4.3. Alineación entre objetivos de negocio y objetivos de minería de datos

VIPO Group se encuentra en un proceso de consolidación, donde es necesario garantizar la continuidad de sus clientes actuales para con estos asegurar el bienestar económico de la compañía y apalancar la consecución de nuevos negocios. Por lo cual, necesita consolidar V-Shopper como un acompañante de compras que genera un impacto relevante en la venta de sus clientes. Para

lograrlo se apuesta por robustecer el producto, complementando la recomendación actual, la cual es visual y basada en el estilo del usuario, con una recomendación basada en el perfil de compra, generando así una recomendación más asertiva que toma en cuenta el historial de compra de clientes con las mismas preferencias.

Para entender la relación entre los objetivos de negocio y de minería de datos, vale la pena resaltar dos términos que serán abordados más adelante de una manera más profunda. Estos dos conceptos hablan de similitud y variabilidad, es decir, por un lado la similitud que se detecta desde la primera hasta la segunda recomendación, en la cual se presenta una transición de interacciones de diversos artículos; y por otro lado, la variabilidad que se presenta en la segunda recomendación, y la capacidad con la que contaría el modelo para ofrecer artículos lo más variado posible desde el inventario, pero sin alejarse mucho de la necesidad que inicialmente planteó con los artículos ya vistos. La relación de estos dos conceptos permitirá tener claridad sobre la alineación entre el objetivo de negocio con el de minería de datos.

1.5. Producción del plan de trabajo

1.5.1. Resumen de ejecución

El plan de trabajo definido para el proyecto descrito en el presente documento está basado en tres fases de ejecución asociadas a las entregas de los avances definidos con las partes interesadas a partir de la metodología de ejecución CRISP-DM.

En la primera fase se ejecutarán las actividades de levantamiento, reconocimiento y definición de objetivos del proyecto. A partir de la segunda fase el enfoque estará definido hacia el establecimiento de la línea base de información aportada por VIPO, el entendimiento y exploración de los datos, la preparación y limpieza de estos para poder ejecutar los objetivos de minería y la selección de la técnica de modelado a aplicar de acuerdo con la información, revisión de bibliografía y objetivos descritos en el proyecto. Por último, en la tercera fase se aplicará el modelo, se hará una revisión de brechas de este para su mejoramiento y puesta a punto y, se generará la presentación final del proyecto para su entrega a VIPO. A continuación, se pueden observar las fechas definidas para la ejecución de las fases aquí descritas:

Tabla 3. Resumen del proyecto

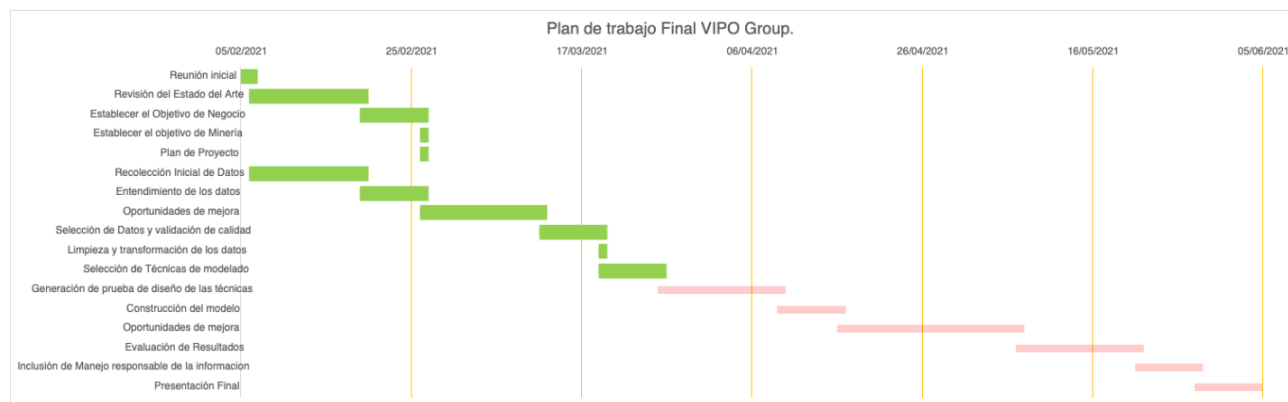
| Entrega | Sección | Tareas | Objetivo final de tarea | % completado | Fecha Inicial | Fecha Fin |
|---------|------------------------|-----------------------------------|--|--------------|---------------|------------|
| 1 | Business Understanding | Reunión inicial | Entender el contexto del negocio y la problemática | 100% | 05/02/2021 | 06/02/2021 |
| 1 | Business Understanding | Revisión del Estado del Arte | Extracción de literatura y validación con respecto a la problemática planteada | 100% | 06/02/2021 | 19/02/2021 |
| 1 | Business Understanding | Establecer el Objetivo de Negocio | Determinar los objetivos de Negocio y sus alcances y métricas | 100% | 19/02/2021 | 26/02/2021 |
| 1 | Business Understanding | Establecer el objetivo de Minería | Determinar los objetivos de Minería y sus alcances y métricas | 100% | 26/02/2021 | 26/02/2021 |
| 1 | Business Understanding | Plan de Proyecto | Construcción Plan de proyecto | 100% | 26/02/2021 | 26/02/2021 |
| 1 | Data understanding | Recolección Inicial de Datos | Obtener la estructura y los datos transaccionales de la empresa | 100% | 06/02/2021 | 19/02/2021 |
| 1 | Data understanding | Entendimiento de los datos | entender el flujo de datos bajo el contexto del negocio | 100% | 19/02/2021 | 26/02/2021 |

| | | | | | | |
|---|--------------------|---|--|------|------------|------------|
| 2 | Data understanding | Oportunidades de mejora | implementar las oportunidades de mejora de Business Understanding | 100% | 26/02/2021 | 12/03/2021 |
| 2 | Data understanding | Selección de Datos y validación de calidad | Extracción de valor de los datos concernientes con los objetivos de minería | 100% | 12/03/2021 | 19/03/2021 |
| 2 | Data Preparation | Limpieza y transformación de los datos | Realizar las transformaciones de datos concernientes con las técnicas vistas en el estado del Arte | 100% | 19/03/2021 | 19/03/2021 |
| 2 | Modeling | Selección de Técnicas de modelado | Determinar las técnicas posibles a implementar para la problemática | 100% | 19/03/2021 | 26/03/2021 |
| 2 | Modeling | Generación de prueba de diseño de las técnicas | presentación de los datos con las posibles técnicas a implementar | 0% | 26/03/2021 | 09/04/2021 |
| 3 | Modeling | Construcción del modelo | construcción del modelo final con sus respectivas ventajas y limitantes | 0% | 09/04/2021 | 16/04/2021 |
| 3 | Evaluation | Oportunidades de mejora | implementar las oportunidades de mejora de Business Understanding | 0% | 16/04/2021 | 07/05/2021 |
| 3 | Evaluation | Evaluación de Resultados | Evaluación del modelo final junto con las métricas planteadas para los objetivos | 0% | 07/05/2021 | 21/05/2021 |
| 3 | Manejo responsable | Inclusión de Manejo responsable de la información | Entender el contexto global del modelo generado bajo el manejo responsable de la información | 0% | 21/05/2021 | 28/05/2021 |
| 3 | Presentación Final | Presentación Final | Presentación Final | 0% | 28/05/2021 | 04/06/2021 |

Fuente: elaboración propia.

En extensión a lo presentado previamente en la tabla, se presenta a continuación el diagrama de secuencias del proyecto:

Ilustración 1. Plan de trabajo proyecto de grado



Fuente: elaboración propia.

1.5.2. Valoración de herramienta y técnicas

La valoración de herramientas se divide en 2 bloques:

Herramientas para el análisis descriptivo de los datos: En la fase de análisis descriptivo hay una amplia gama de herramientas que pueden ser usadas para tal fin, en este caso vamos a usar como herramienta de decisión el cuadrante de Gartner:

Ilustración 2. Cuadrante Gartner Herramientas

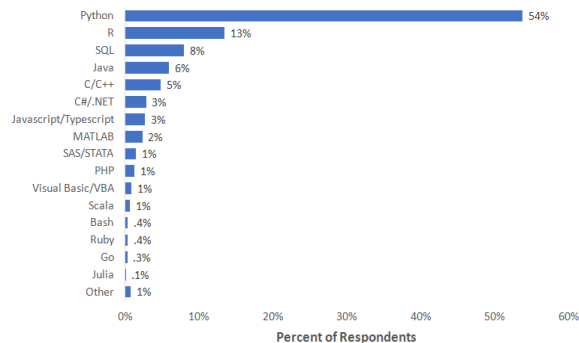


Fuente: Gartner

Como se puede observar Microsoft con su herramienta Power BI es el líder de esta categoría y al realizar la revisión de la herramienta se evidencia que existe una versión desktop que no requiere licencia para su uso y que se ajusta a los requerimientos de esta fase, es por esto que PowerBI será la herramienta que se usará para la exploración de datos y la realización de la descriptiva.

Herramientas para el modelado: En la fase de modelado existen múltiples lenguajes de programación destinados para este tipo de labores, sin embargo, un estudio realizado por Kaggle en 2018 sobre los lenguajes de programación más usados en el entorno de ciencia de datos arrojó los siguientes resultados:

Ilustración 3. ¿Qué lenguaje de programación usa más a menudo?



Note: Data are from the 2018 Kaggle ML and Data Science Survey. You can learn more about the study here: <http://www.kaggle.com/kaggle/kaggle-survey-2018>. A total of 23859 respondents completed the survey; the percentages in the graph are based on a total of 15222 respondents who provided an answer to this question.



Fuente: Kaggle

Lo cual nos lleva a hacer un filtro inicial de los lenguajes dejando solo en el radar a Python y R. Por otro lado, revisando el estudio realizado en 2017 por Analytics Vidhya sobre qué herramienta seleccionar para DataScience podemos encontrar que Python tiene una mejor calificación general que R según los parámetros evaluados en dicho estudio:

Tabla 4. Calificación por herramienta

| Parameter | SAS | R | Python | |
|--|-----|-------------|-------------|-----------|
| Availability/Cost | | 3 | 5 | 5 |
| Ease of Learning | | 4.5 | 2.5 | 3.5 |
| Data Handling Capabilities | | 4 | 4 | 4 |
| Graphical Capabilities | | 3 | 4.5 | 4.5 |
| Advancement in Tools | | 4 | 4.5 | 4.5 |
| Job Scenario | | 4 | 4.5 | 4.5 |
| Customer Service Support and Community | | 4 | 3.5 | 3.5 |
| Deep Learning Support | | 2 | 3 | 4.5 |
| Total | | 28.5 | 31.5 | 34 |

Fuente: Analytics Vidhya

Adicionalmente luego de la revisión de las necesidades futuras planteadas por VIPO para lograr exponer los modelos como servicios que permitan ser consultados mediante otras herramientas tecnológicas se pudo encontrar que Python es la herramienta que más documentación y facilidad ofrece para solucionar dicho requerimiento, es por este motivo que se usará Python como el lenguaje de programación para realizar el modelado.

Técnicas:

Para la selección de las técnicas que van a ser usadas en el desarrollo de este proyecto nos basamos en los 2 objetivos de minería de datos:

- Crear agrupaciones de clientes basados en sus preferencias de compra: Para esto se usarán técnicas de aprendizaje no supervisado como Clustering donde se pretenden evaluar algoritmos comunes K-medias y Clusters Jerárquicos.
- Inferir cual es el artículo de moda más adecuado para sugerirle a un cliente que maximice la probabilidad de tener un cross-selling. Para atacar este objetivo se propone el uso de técnicas como el filtrado colaborativo y los filtros híbridos. Por otro lado, se descarta el uso de algoritmos basados en el análisis de canasta de mercado pues ya se realizó un acercamiento al problema bajo este enfoque.

2. Fase II. Comprensión de los Datos

2.1. Recopilación de datos iniciales

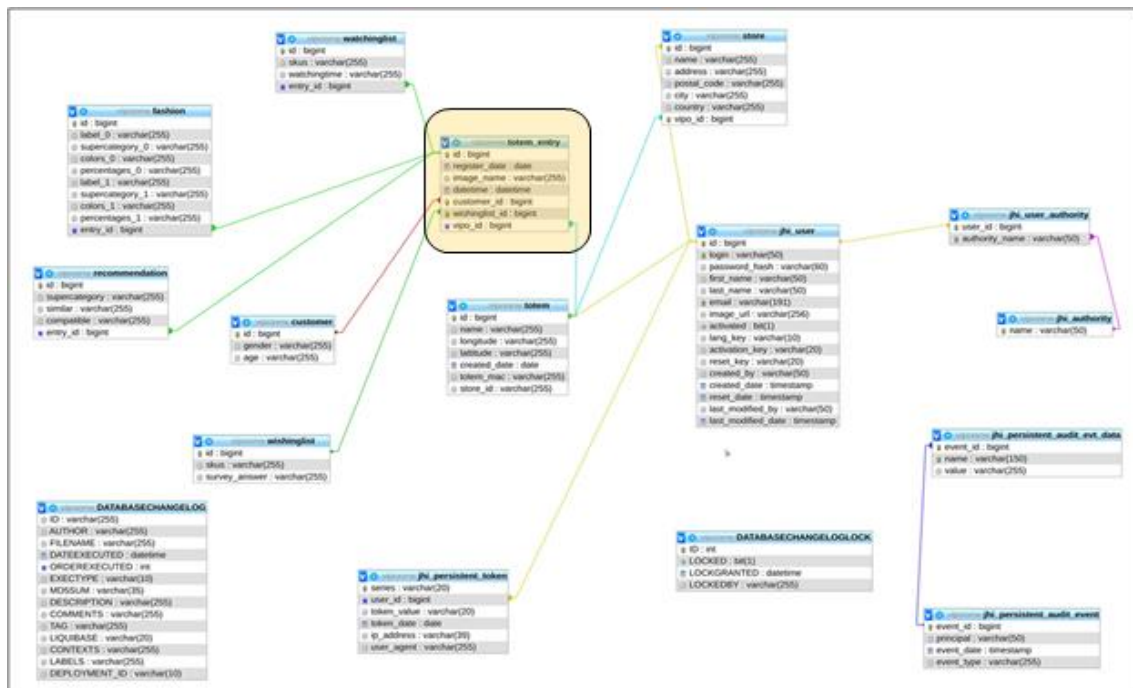
Los datos que se tienen para la exploración inicial son 10.000 registros procedentes de la prueba piloto realizada en Colombia del 31 de Diciembre de 2019 a 30 de enero de 2020, que son el resultado de la interacción de los usuarios con el Tótem en las tiendas de Colina, es necesario tener presente que en estos registros no es posible individualizar a los clientes, puesto que no se siente un ID único por persona sino es por registro; Además también se tiene un Excel con la información

detallada de los ítems que finalizaron en una compra por medio de las recomendaciones al cliente de la tienda dadas por el Tótem.

Adicional a estos registros y teniendo en cuenta el efecto pandemia en donde se ha tenido una reducción de los registros mensuales, se ha solicitado un año de datos históricos para que sean utilizados en los modelos tanto en train, test y validaciones finales, y así evitar tanto el overfitting y mejorar la relación de cantidad de registros por atributos de la tabla transaccional.

2.2. Descripción de los datos

La siguiente imagen es correspondiente al modelo de datos OLTP de VIPO Group:



Fuente: VIPO Group

En donde la tabla transaccional es la Totem_entry (resaltada en amarillo) y cuyas tablas adjuntas son las correspondientes con los datos de la empresa y de la tienda, como lo detallamos a continuación:

- Granularidad: Se tiene que la granularidad es por utilización del Tótem, esto significa que no es posible individualizar a las personas, sino se tiene el registro de lo que “una personas” ha seleccionado en ese momento específico, pues si la misma persona regresa al tótem momentos después haber cerrado la sesión será transparente para la Base de datos y se tomará como otro registro totalmente diferente siendo.
- Temporalidad: Se tiene que la Temporalidad está dada inicialmente y para efectos de exploración del 8 al 20 de diciembre de 2019, sin embargo, teniendo en cuenta el efecto pandemia se tiene previsto que la temporalidad de los datos utilizados en los modelos sean de todo el año 2019 (Teniendo presente las diferentes facetas de train y test).

- Geografía: La geografía de los datos de exploración es de Colombia, sin embargo, como se ha mencionado, la idea es que para futuros proyectos y un mayor alcance, los datos sean utilizados para el entrenamiento y validación de los modelos, no solo de Colombia, sino que también sean recolectados de las tiendas disponibles y en diferentes temporalidades.

Tabla 5. Descripción de las tablas seleccionadas

| Tabla | Atributo | Descripción |
|----------------|--|---|
| Totem Entry | ID | Llave principal de la tabla y representa la interacción de un usuario con el totem. |
| | Register Date | Fecha en la cual se está teniendo interacción con el Totem. |
| | Image_Name | Imagen que es tomada para caracterizar el cliente por medio de computer vision. |
| | Date time | Fecha y Hora del momento exacto en la cual se está teniendo interacción con el Totem. |
| | Customer_ID | Llave foránea procedente de la tabla Customer y que es un incremental del registro. |
| | Wishing List | Llave foránea procedente de la tabla Wishing List y que es un incremental del registro. |
| Totem | Vipo_id | Llave foránea procedente la tabla Totem y que representa el serial del dispositivo. |
| | ID | Llave principal de la tabla y representa el serial del Tótem. |
| | Name | Nombre que puede ser opcional para distinguir al tótem en palabras del negocio. |
| | Longitude | Longitud de la localización del Tótem. |
| | Lattitude | Latitud de la localización del tótem. |
| | Created Date | Fecha en la cual fue instalado el Tótem. |
| Store | Totem Mac | Dirección Mac con la cual el Tótem se conecta a internet. |
| | Store_id | Llave foránea procedente de la tabla Store y la cual es incremental. |
| | ID | Llave principal de la tabla y representa el registro único dado por la empresa VIPO a sus clientes. |
| | name | Nombre que puede ser opcional para distinguir a la tienda en palabras de negocio. |
| | address | Dirección específica de la tienda. |
| | Postal_code | Código postal de la región en donde se encuentra localizada la tienda. |
| Watchinglist | City | Ciudad en la cual se encuentra la tienda. |
| | Country | País en donde se encuentra la tienda. |
| | Vipo_id | Llave foránea procedente de la tabla Totem y que significa el serial del dispositivo. |
| | id | Llave principal de la tabla y es un entero incremental de los ítems presentados al cliente. |
| | skus | Es la llave foránea de los ítems de la tienda. |
| | Watchingtime | Es la fecha en la cual se ha generado este registro. |
| Fashions | entry_id | Llave foránea procedente de la tabla Totem Entry y que representa el registro único. |
| | ID | Llave principal de la tabla y representa las reglas de moda internacionales. |
| | Label V0 | Nombre de la sección de moda. |
| | Super Category 0 | Categoría padre a la cual corresponde la regla. |
| | Colors 0 | El color que representa la regla. |
| | percentages 0 | Valor calculado de extracción de colores de la persona con la similitud de la regla. |
| | Label1 | Nombre de la sección de moda. |
| | Super Category 1 | Categoría padre a la cual corresponde la regla. |
| | Colors 1 | El color que representa la regla. |
| Percentages 1 | Valor calculado de extracción de colores de la persona con la similitud de la regla. | |
| Recommendation | Entry_id | Llave foránea procedente de la tabla Totem Entry y que representa el registro único. |
| | ID | Llave primaria de la tabla y que significa la recomendación dada a la persona que está usando el Totem. |
| | Supercategory | Categoría Padre a la cual pertenece la recomendación. |
| | Similar | Posibles reglas que se pueden aplicar a la recomendación. |
| | Compatible | Recomendación sugerida al cliente según cálculos y reglas de moda. |
| Customer | entry_id | Llave foránea procedente de la tabla Totem Entry y que representa el registro unico. |
| | ID | Llave principal que representa el identificador único del registro. |
| | Gender | Género obtenido por computer vision de la persona que utiliza el totem. |
| Wishing List | Age | Edad obtenida por computer vision de la persona que utiliza el totem. |
| | ID | Llave principal de la tabla y es un entero incremental de los ítems seleccionados por el cliente. |
| | Skus | Es la llave foránea de los ítems de la tienda. |
| | Survey_answer | Respuesta del cliente frente a la sección realizada. |

Fuente: VIPO Group

2.3. Análisis exploratorio de los datos

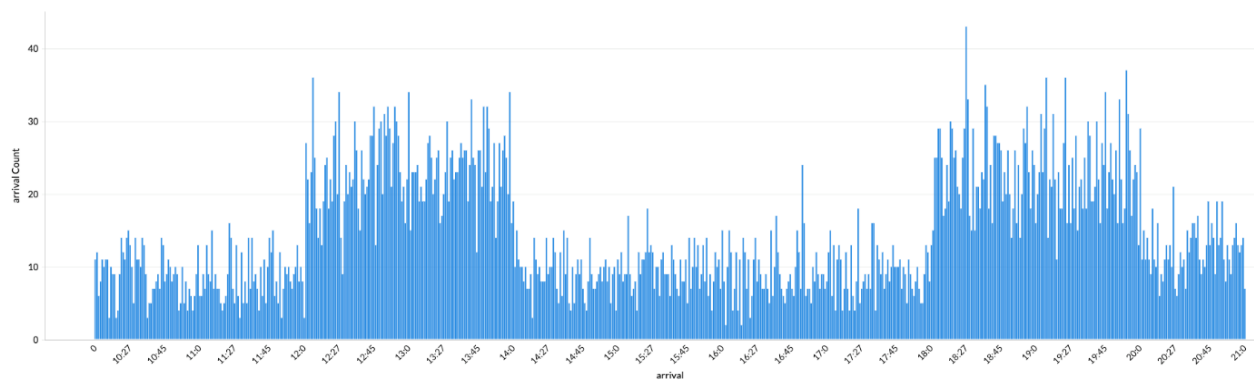
En principio es importante dejar en claro que los datos están distribuidos de manera lógica según su nivel de granularidad basados en categorizaciones standard que permiten una mejor gestión de tienda de los clientes de VIPO, con esto nos referimos a que las tablas que fueron suministradas

están basadas en un diagrama lógico en donde existe un inventario de artículos del cual se puede recomendar a través de la tecnología ya existente, la cual, entrega posteriormente entrega 20 artículos recomendados según estilo y luego a partir de las visualizaciones o pinchazos que dé el cliente final se encontraría una base de vistas (watchinglist) y que a partir de esta, en el mejor de los casos, se genere una lista de compra o deseos de los cuales se pueda ejecutar la venta. Según lo mencionado por VIPO, las personas que llegan a establecer una lista de deseos y están interesados en la compra de los artículos reciben un ticket de sus deseos para pagar en caja y recibir los artículos vistos. A partir de esta información, de la cual partimos y a la cual enfocamos el presente estudio, se establecen definiciones sobre el nivel de conversión del servicio prestado por VIPO y, desde allí, propendemos iniciar el análisis que permita mejorar la cantidad de ventas del cliente de VIPO impactando en la renovación de los contratos que este tiene actualmente con ellos.

A partir de la lógica mostrada previamente, establecemos como se decantan la recomendaciones vistas y deseadas para saber el nivel de estas sobre la recomendación inicial de VIPO, en donde se puede observar claramente que el 40% de las recomendaciones son vistas, y de esas vistas el 43% fueron a la lista de deseos. Desde el enfoque de deseos sobre recomendación el porcentaje fue de 11%, lo cual nos puede indicar que posiblemente si se aumenta la cantidad de recomendaciones basados en información de agrupaciones o vecinos cercanos es que garanticemos una mayor venta. Este insight es fundamental porque liga directamente, de manera relacional, la posibilidad de que el nivel de personalización tome una posición clave dentro del análisis del presente proyecto.

Con base en la intención de encontrar agrupaciones sobresalientes, que permitan mejorar la personalización de la recomendación que actualmente ejecuta VIPO, y que adicionalmente sean suficientes para complementar el servicio de V-shopper de acuerdo a los objetivos definidos previamente, se ha evidenciado que variables como la posición del tótem, el horario de interacción con el mismo, no son variables significativas en esta búsqueda inicial, ya que como se ve a continuación en la Gráfica 1, el comportamiento de los datos de interacción, además de erráticos y poco convencionales, no determina definiciones clave en la búsqueda de establecer posibles grupos de interés a los cuales poder hacer recomendaciones de acuerdo a la similitud de su perfil de compra.

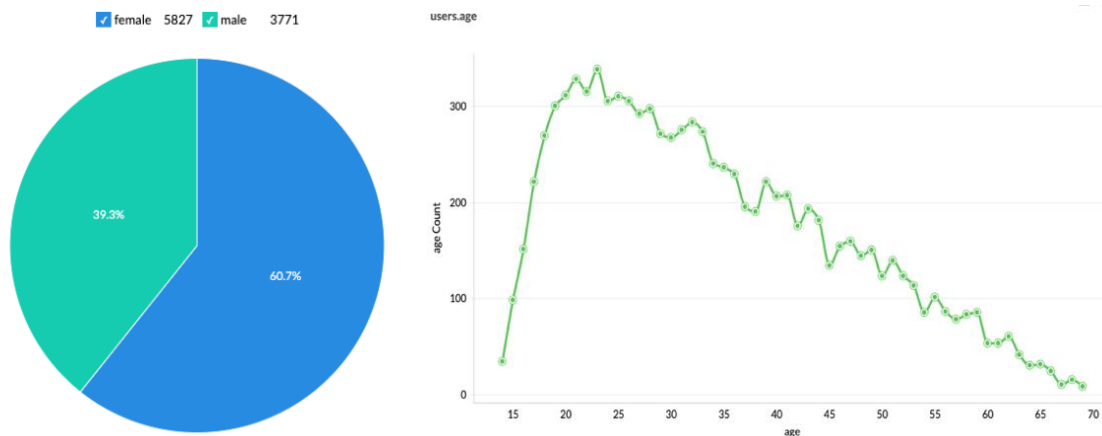
Gráfica 1. Afluencia del Tótem por horario



Fuente: Elaboración propia

Sin embargo, tratando de llevar la revisión de lo más general a lo particular, es preciso mencionar que, en las interacciones realizadas con el tótem hay mayor preponderancia de personas jóvenes, lo cual, estaba en las cuentas, pues se ha partido de la base de que el uso de este tipo de elementos puede verse inclinado a poblaciones con una mayor interacción a las tecnologías de este estilo. Adicionalmente, es preciso mencionar que existe una mayor concentración de mujeres en la muestra entregada por VIPO, tal y como se observa en la Gráfica 2, lo cual, podría establecer algún tipo de causalidad importante sobre la posición del tótem en tienda u otras conjeturas, que lastimosamente no pudieron ser corroboradas con exactitud.

Gráfica 2. Participación de género y edad en la muestra

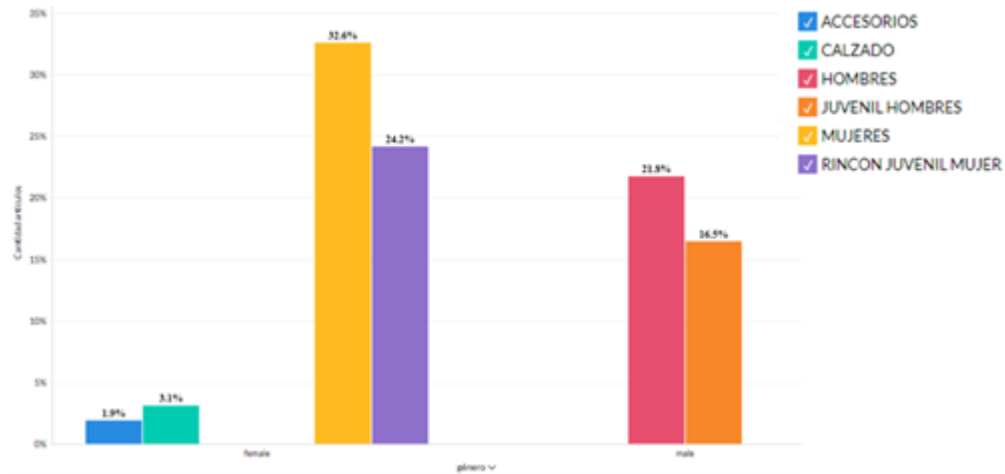


Fuente: Elaboración propia

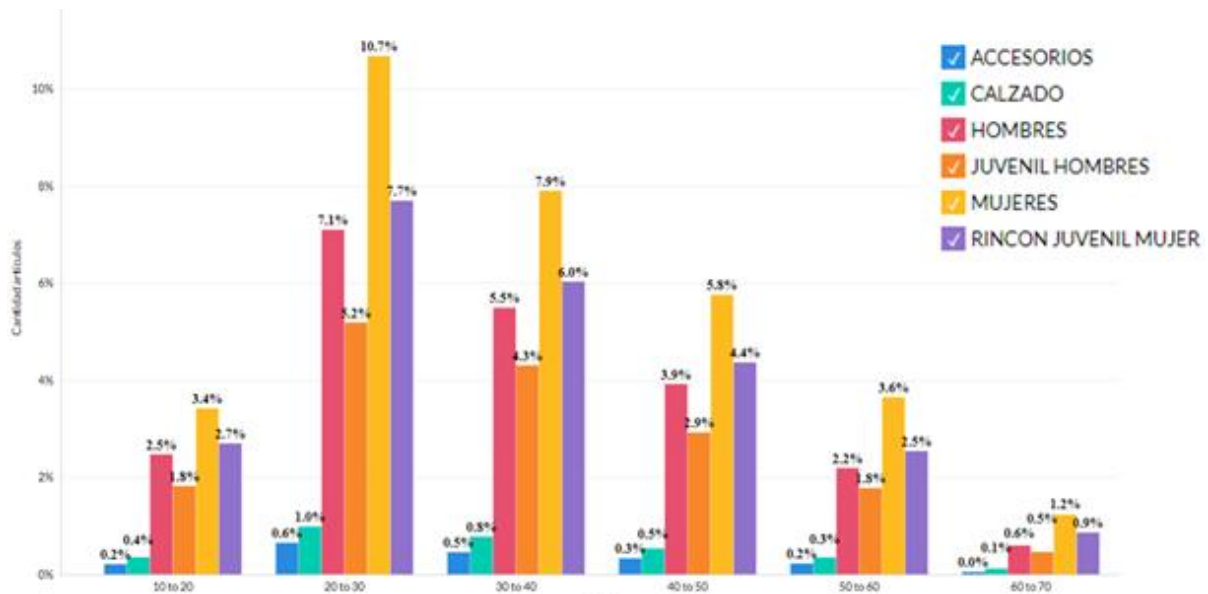
Es por esto que a partir de esta información y continuando con la definición objetivo definida, que las variables con mayor potencial en la búsqueda de establecer segmentos para recomendar podrían fijarse en el tipo de prenda, la edad y el género del comprador, pues en estas, se podría establecer una alta probabilidad de agrupación. Es preciso mencionar, que en la gráfica 3 no se muestra el análisis basado en artículos o clases, pues su nivel de granularidad es bastante extenso. Esto es inconveniente porque no fue posible encontrar un nivel de segmentación visible al ojo que satisficiera las necesidades del problema y que, al relacionarse directamente con otras variables más maleables, como la línea fuera más sencillo de ver. Es por esto que a partir de la gráfica 3 hemos encontrado una primera perspectiva interesante sobre la preponderancia de las variables direccionadas al artículo y su variedad para la construcción del modelo de minería que dirija sus esfuerzos en recomendaciones que mejoren el nivel de personalización. En esta gráfica se observa cómo las mujeres jóvenes mueven el nivel de interacción, por lo que su información sobre las visualizaciones y deseos es de suma importancia en la selección de grupos.

Adicionalmente en la gráfica 3 y 4 se observa como los artículos de mujer, en las edades más críticas están moviendo la serie y enriquecen la visual del modelo hacia la personalización basada en ítems vistos y deseados.

Gráfica 3. Línea por Género



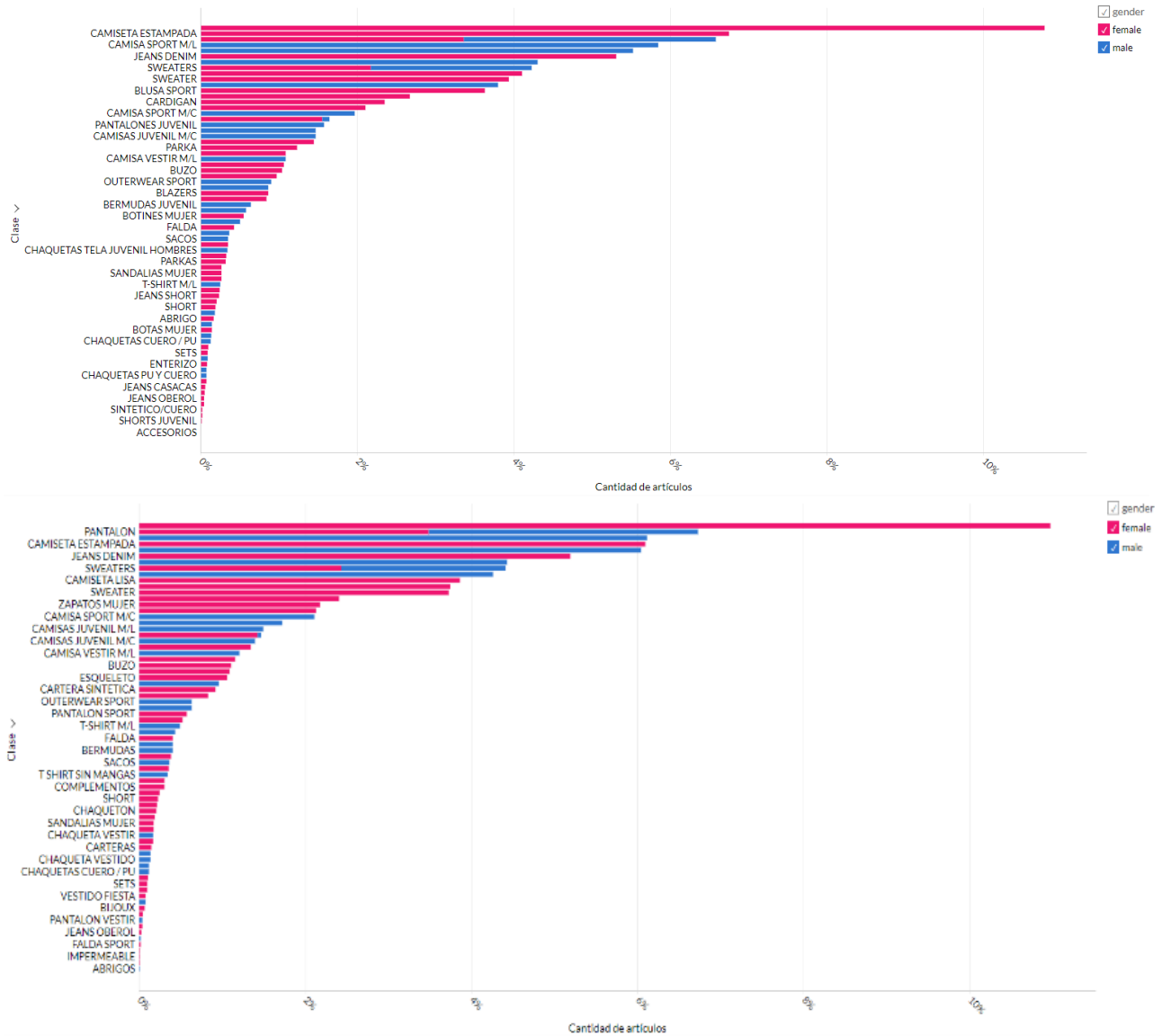
Gráfica 4. Línea por Edad



De otra parte, con el fin de establecer un mejor zoom sobre los artículos, y garantizar que no sea un solo artículo deseado el que mueva el nivel de agrupación, que se genera la gráfica 5 y 6, en donde en los grupos más fuertes por el nivel de interacción se observa que una gran cantidad de artículos, en su mayoría direccionados a mujeres están participando. Esto permite evidenciar una asociación importante sobre las variables críticas y la variedad como tal de elementos que se han visto y

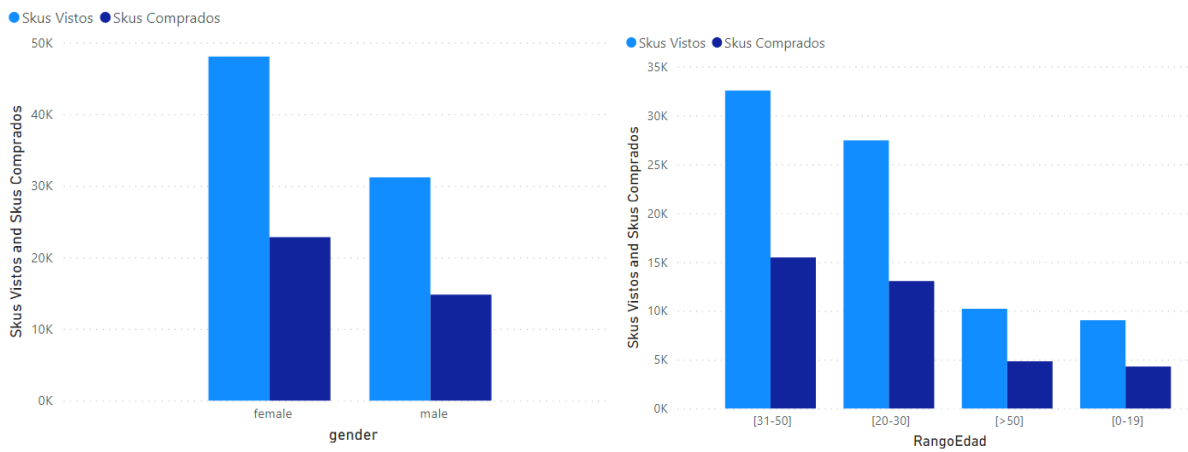
recomendado, pieza clave para que, al seleccionar estas variables en virtud del modelo complementario, entreguen elementos suficientes desde una perspectiva de vecinos más cercanos con una variedad importante para garantizar que no siempre se va a recomendar lo mismo.

Gráficas 5 y 6. Artículos preferidos por rango de edad (5) 20 a 30 años, (6)30 a 40 años



Con la reflexión sobre la variedad de los SKU, es preciso mostrar la cantidad de artículos vistos y deseados desde la perspectiva de género y edad, en donde más allá de corroborar lo mencionado anteriormente se puede ver que un 35% o 40% de los skus vistos han sido enviados a la lista de deseos en cada una de los rangos de edad descritos, por lo que es claro que hay homogeneidad entre variables críticas sin tener pesos que puedan diseminar la posibilidad de entregar un mejor nivel de personalización basados en los datos recopilados. En la gráfica 7, es claro este agrupamiento casi nivelado entre rangos de edad y género.

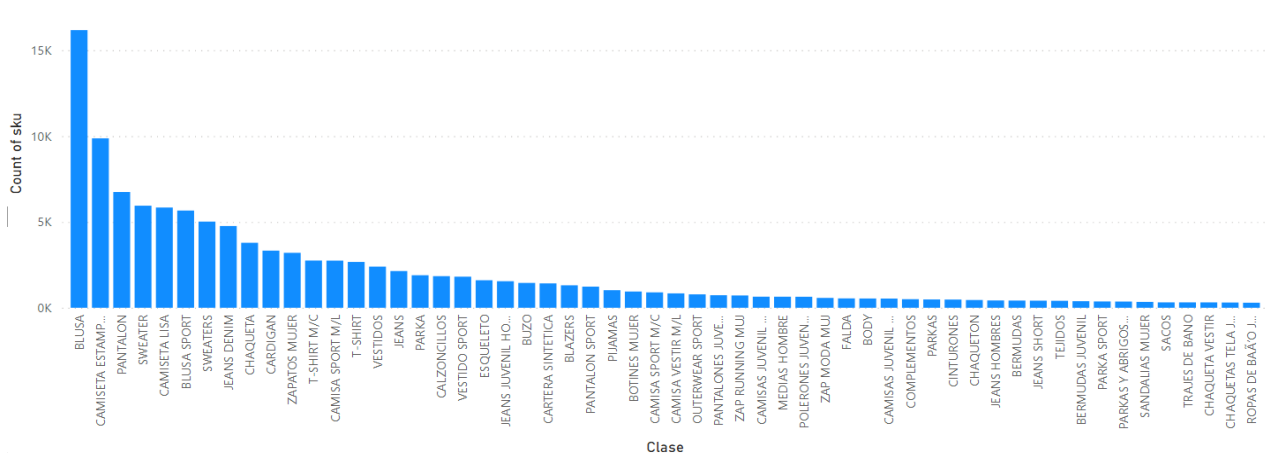
Gráfica 7. SKUs Vistos vs SKUs Comprados por Género y Rango de edad



Fuente: Elaboración propia

Con el fin de revisar el nivel de agrupación por clase, es claro que ciertos artículos, que son de mujer en su mayoría están moviendo la serie de la cantidad de skus descritos por las clases definidas, esto es apenas natural considerando el 70% aproximado de mujeres que interactuaron con el tótem y que, gracias al nivel de variedad que se puede observar podrán entregar una mejor calidad de recomendación basada en su perfil general y principalmente los artículos vistos. En la gráfica 8 se ve una agrupación ordenada de clases y como este nivel de asociación permite evidenciar una mayor concentración según lo definido, lo cual claramente genera una distinción por clase que podría ser complementaria de las líneas y entregar grupos más visibles bajo la búsqueda de un mayor nivel de personalización de la recomendación.

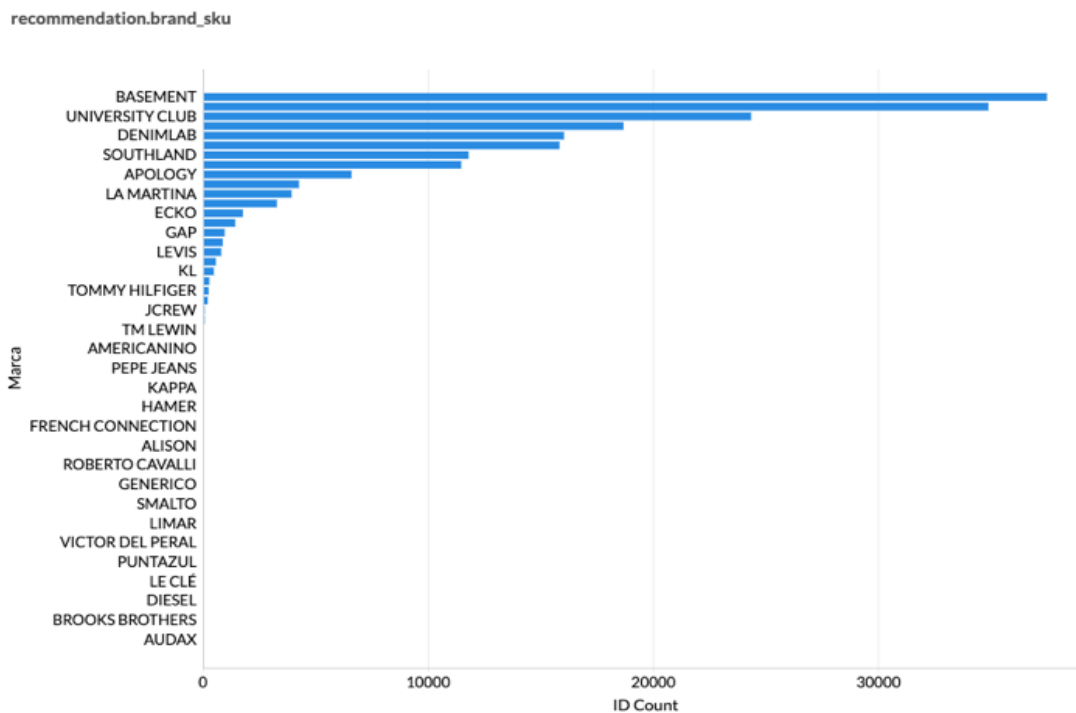
Gráfica 8. Skus por clase



Fuente: Elaboración propia

Continuando con el análisis previo, es preciso mencionar que una de las variables que más interés puede generar en una revisión exploratoria de ropa es la marca, sin embargo, esta variable es complementaria a la línea y la clase previamente analizadas, lo cual indica que el estilo está sesgando el modelo desde esta perspectiva y que con aras de mejorar la calidad de la recomendación ante la posibilidad de sobrecargar el modelo con muchas variables que indican un mismo lugar, hemos decidido dejarla en espera ante la posibilidad de generar ruido ante clasificaciones mejores y con un mayor nivel de agrupación según los perfiles que las describen. En la gráfica 9 se puede observar claramente que 12 marcas condensan el total de artículos recomendados inicialmente y que las cantidades son correspondientes a las visualizaciones dadas en clase y línea.

Gráfica 9. Recomendación por marca



Fuente: Elaboración propia

Un descubrimiento interesante es que las recomendaciones realizadas por el tótem no consideran la talla de la persona, ya que cómo se ha mencionado previamente en el presente documento la recomendación que emite el V-Shopper es meramente de estilo y no considera variables adicionales. De otra parte, desde la perspectiva de negocio, VIPO ofrece otro servicio que es el SizeTool, en donde se realiza un ajuste de las prendas seleccionadas de acuerdo con las medidas de la persona.

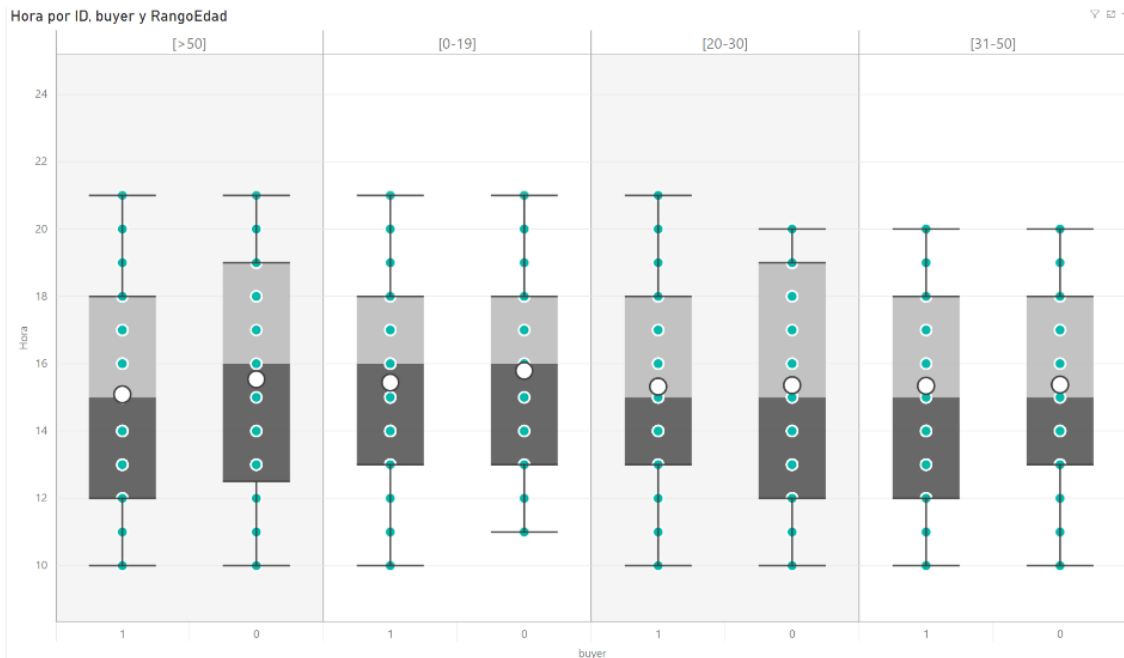
Ilustración 4. Recomendaciones que no distinguen talla

| | |
|-----------|----|
| 4970324 | XS |
| 4311878 | L |
| 881900508 | L |
| 881828558 | S |
| 881910631 | -1 |
| 881898474 | L |
| 881831210 | M |
| 3831241 | 31 |
| 2989656 | 28 |
| 881909945 | M |
| 881995342 | 36 |
| 11396657 | XL |
| 881910049 | L |
| 2894747 | XL |
| 881878127 | M |
| 881836550 | 32 |
| 881829886 | 32 |
| 881831325 | S |
| 11396649 | M |
| 10815112 | 32 |

Fuente: Elaboración propia

Es preciso mencionar que la base involucra los compradores y no compradores, a partir de esto se quiso ver si existieron diferencias entre estos, desde el criterio de visualizaciones, sin embargo, la mayoría de no compradores no tenían una interacción significativa con el tótem en términos de tiempos y terminaban abandonando la estación de interacción. Es por esto que los no compradores fueron retirados de la muestra de exploración y no serán incluidos en el modelo.

Gráfica 10. Compradores y no compradores por rango de edad y hora

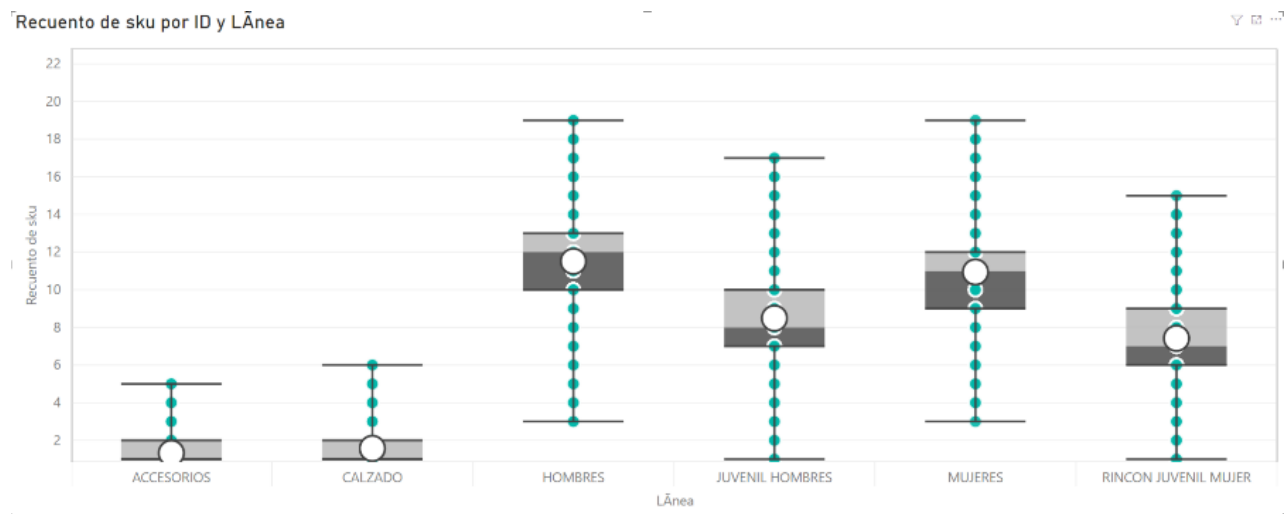


Fuente: Elaboración propia

A partir de lo analizado anteriormente, en la gráfica 11 se puede ver a través de un diagrama de caja que la mediana de hora de uso de la interacción, corresponde principalmente a lo suministrado en el análisis previo y que, a partir de este análisis no existe diferencia significativa entre compradores y no compradores, lo cual nos reafirma el hecho que estos individuos, por variables externas desconocidas, no tuvieron buen relacionamiento con el tótem y eso influyó en su falta de participación de la experiencia completa en venta.

Continuando con las apreciaciones ya mencionadas, se vio una perspectiva muy interesante a partir del comportamiento de la lista de deseos desde una visión por línea de artículo, en la gráfica 12 se establece con claridad que existe una diferencia diciente en la cantidad de productos que se van la lista de deseos por tipo de artículos y que más allá de que estos tengan preponderancia por el artículo per sé, se puede definir un nivel de agrupación o diferencia por el género y grupo objetivo que está enviando el retailer para la venta uno a uno, esto es fundamental porque reafirma el enunciado de que la clase y línea, en su nivel de asociación, y bajo un esquema complementario pueden entregar un buen nivel de agrupación complementario al estilo que aumente las variables consideradas objetivos en el presente proyecto.

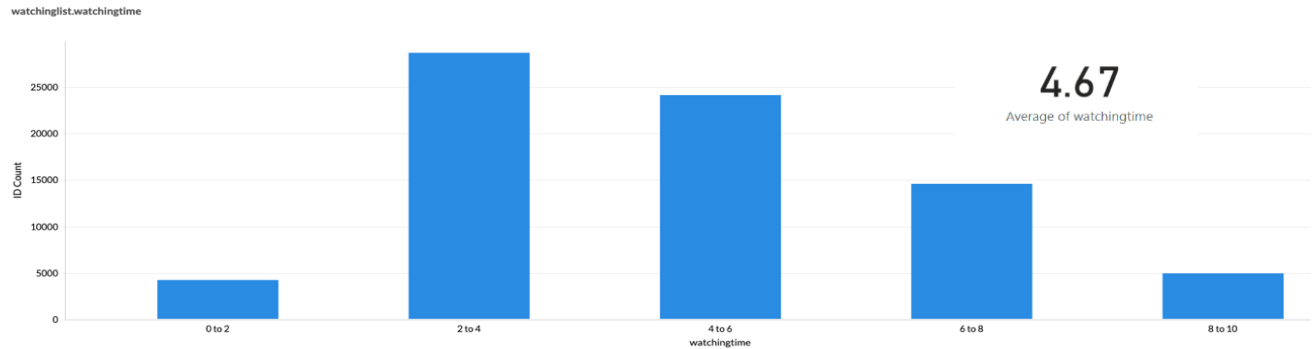
Gráfica 11. Artículos recomendados por línea por usuario



Fuente: Elaboración propia

Finalmente, de la información analizada solo queda revisar la influencia del tiempo de uso del tótem que, en un principio, no es muy relevante, considerando que no supera los 30 segundos, pero que por artículo, tal y como se observa en la gráfica 13 permite establecer el tiempo que le toma a una persona ver una recomendación y descartarla. En promedio este tiempo es de 4,67 segundos, lo cual, por la naturaleza del tótem puede considerarse normal. Adicionalmente, considerando que hasta el momento no se ha podido establecer con los datos una influencia clara entre el tiempo y la selección (que va movida principalmente por el 70% de vistos, pero no comprados).

Gráfica 12. Tiempo de vista por artículo



Fuente: Elaboración propia

3. Fase III. Preparación de Datos

Con base en los objetivos de negocio y minería propuestos, la preparación de los datos estará enfocada en la necesidad de tener data suficiente y seleccionada para realizar las agrupaciones de compradores basadas en sus rutinas y preferencias e inferir los artículos más adecuados para alcanzar un cross-selling. De acuerdo con esto se procede a preparar la información para alcanzar dicho propósito.

3.1. Selección de los datos

Se recibieron las bases de datos en formato JSON, la información venía separada por registro, el primer paso de selección fue la unión de los registros. Adicionalmente, las bases de datos de SKUS vienen en dos grandes archivos demarcados entre hombres y mujeres. Como venían por separado, estas se unen a fin de generar un archivo inicial. Por último, para completar el paso inicial y tener una tabla única para trabajar se realizó la unión de las bases de datos entregadas.

Cabe resaltar que al ser VIPO una empresa de analítica, son muy rigurosos en los procesos de recolección y extracción, por lo que la data entregada en general no tiene muchos elementos de los cuales realizar una limpieza exhaustiva y casi todos los fenómenos que pueden ser fuentes de variación, tienen una explicación de negocio. A continuación, en la fase de limpieza se hará un análisis profundo de la data entregada y que puntos deben ser corregidos para alcanzar la preparación óptima de acuerdo con el objetivo de minería trazado.

3.2. Limpieza de los datos

Con el archivo único se pudo establecer que había una agrupación interna de VIPO que se desconocía, esta información fue solicitada para realizar la homologación correspondiente a data legible. Esta información fue entregada pero no todos los catálogos agrupados en códigos de VIPO eran homologables, por lo que la primera decisión de limpieza tomada fue descartar esta información, que igualmente no generaba valor hacia el objetivo marcado.

Adicionalmente, en el campo de recomendación se encontraron 69 artículos duplicados que fueron retirados de la base de datos, tal y como se observa en la ilustración 6. Por otro lado, cabe resaltar que existía un artículo triplicado.

Ilustración 5. Artículos Duplicados en campo recomendación

```
In [304]: print(groupRecosDuplicados)
          ID      sku  count
112375  5621  881895652     3
34880   1745  881830785     2
45187   2261   8358421     2
57009   2852  881829620     2
98620   4933  881829510     2
...     ...     ...     ...
187986  9403  881835368     2
68311   3417  881836758     2
79141   3959  881727115     2
195931  9800  881929687     2
108237  5414  881832670     2

[69 rows x 3 columns]
```

Fuente: Elaboración propia

Continuando con el análisis de limpieza de datos, se encontraron 15 registros de artículos duplicados en el campo de watchinglist, de estos registros no se evidenció o encontró en el levantamiento con VIPO una explicación de negocio sobre su existencia, por lo cual se procedió a eliminarlos de la base de datos. En la ilustración 7 se puede observar los registros que fueron limpiados para que su perfil único se mantuviera en el modelado posterior.

Ilustración 6. Artículos Duplicados en campo Watchinglist

```
In [305]: print(groupWatchingDuplicados)
          ID      sku  count
55349  6879   8499622     2
78388  9841  881834340     2
14212  1745  881830785     2
68793  8605  881832556     2
60399  7527  10683271     2
71487  8945  881835366     2
38197  4726  881831943     2
27615  3417  881836758     2
9755   1217  4943046     2
52027  6450  881837287     2
59141  7362   9773229     2
61076  7615  881707459     2
61661  7705   4201460     2
27298  3375  881898489     2
4278   520  881835524     2
```

Fuente: Elaboración propia

De la misma manera se encontraron dos (2) artículos duplicados en el wishinglist que igualmente fueron retirados para el análisis:

Ilustración 7. Artículos Duplicados en campo wishinglist

```
In [306]: print(grupWishingDuplicados)
          ID      sku  count
32671  8605  881832556     2
2026   520  881835524     2
```

Fuente: Elaboración propia

3.3. Construcción y formateo de los nuevos datos

Debido a la gran cantidad de SKUS involucrados en el problema y el número limitado de registros, se hace muy complejo obtener similitudes entre usuarios a nivel de artículos, por lo tanto, se evidencia la necesidad de comparar los clientes entre sí bajo una variable diferente al SKU. Para esto existen diferentes agrupadores de los artículos en los datos suministrados como lo son la línea, clase y subclase; sin embargo, al analizar los valores que pueden tomar dichas variables se puede evidenciar que por ejemplo la línea puede tomar solamente 4 valores (hombre, mujer, rincón juvenil hombre, rincón juvenil mujer) lo que no aporta mayor variabilidad para hacer la comparación, por otro lado las variables de clase y subclase tienen una problemática que radica en que no tienen un componente de género, por ejemplo la clase “Camiseta Lisa” existe tanto en hombre como en mujeres y esto podría ser perjudicial para el modelo debido a que podría recomendar prendas de hombre a una mujer o viceversa. Por tal razón se propone construir una nueva variable la cual combine la clase y la línea para poder con esta calcular las similitudes entre los usuarios y garantizar el componente de género, la cual será llamada superclase:

Ilustración 8. Creación de nuevas columnas

```
tbl_Items_seen = pd.DataFrame(columns=['nombre_archivos', 'SKU_SEEN'])
tbl_Items_seen["nombre_archivos"] = np.array(id_user_seen)
tbl_Items_seen["SKU_SEEN"] = np.array(Items_seen)
tbl_Items_seen=pd.merge(tbl_Items_seen,skus,how='left',left_on='SKU_SEEN',right_on='sku')
tbl_Items_seen["SUPERCLASE"]=tbl_Items_seen["Línea"]+"-"+tbl_Items_seen["Clase"]

tbl_Items_1ra_recomendation = pd.DataFrame(columns=['nombre_archivos', 'SKU_recommended'])
tbl_Items_1ra_recomendation["nombre_archivos"] = np.array(id_user_1ra_recomendation)
tbl_Items_1ra_recomendation["SKU_recommended"] = np.array(Items_1ra_recomendation)
tbl_Items_1ra_recomendation=pd.merge(tbl_Items_1ra_recomendation,skus,how='left',left_on='SKU_recommended',right_on='sku')
tbl_Items_1ra_recomendation["SUPERCLASE"]=tbl_Items_1ra_recomendation["Línea"]+"-"+tbl_Items_1ra_recomendation["Clase"]

tbl_Items_2da_recomendation = pd.DataFrame(columns=['nombre_archivos', 'SKU_recommended_2'])
tbl_Items_2da_recomendation["nombre_archivos"] = np.array(id_user_2da_recomendation)
tbl_Items_2da_recomendation["SKU_recommended_2"] = np.array(Items_2da_recomendation)
tbl_Items_2da_recomendation=pd.merge(tbl_Items_2da_recomendation,skus,how='left',left_on='SKU_recommended_2',right_on='sku')
tbl_Items_2da_recomendation["SUPERCLASE"]=tbl_Items_2da_recomendation["Línea"]+"-"+tbl_Items_2da_recomendation["Clase"]
```

Fuente: Elaboración propia.

4. Fase IV. Modelado

4.1. Selección de técnicas de modelado.

De acuerdo con la revisión del estado del arte realizada previamente y, teniendo en cuenta que, tanto el entendimiento del negocio de VIPO Group como los datos generados y analizados por los Tótem, podemos mencionar que actualmente existe un modelo de recomendación de popularidad, teniendo en cuenta la normas internacionales de color y el género; esto quiere decir que, si por medio de tecnologías de visión por computador se determina que el usuario hombre está usando un pantalón negro, se le pueda recomendar una camisa blanca, un cinturón café, una pashmina gris, etc. aunque ha tenido buena aceptación por el mercado además de ser relevante para una recomendación inicial (en frío) es importante saber que se tiene información recolectada de forma explícita (tipo de vestimenta que el usuario desea), como implícita (edad, género, clics, duración de visualización, localización, hora de contacto con el tótem).

Teniendo en cuenta los datos mencionados y obtenidos anteriormente además del conflicto de la dimensionalidad (al ser ineficaz) por la cantidad de ítems en cada una las locaciones, se considera como modelo coadyuvante al sistema de recomendación actual, un sistema por filtrado colaborativo, donde por medio de las interacciones de los usuarios (registros) en el tótem, en conjunto con la base histórica de las interacciones hechas por otros usuarios, es posible crear segmento de clientes teniendo en cuenta los artículos recomendados, vistos (watchinglist) y deseados (wishinglist), así como, el tiempo empleado en cada uno de estos, para poder generar distancias o similitudes entre ellos, ya sea por el método (coseno o euclidiano), para que una vez, sea identificado y segmentado, se le recomiende una prenda aún no vista (nuevo ítem) por medio este método. A partir de este enfoque, se utilizarán dos sistemas de recomendación generando un Sistema de recomendación híbrido a fin de mejorar el rendimiento de cada una de las capas y superar la percepción actual frente al usuario final.

4.2. Comprobación del modelo

Para comprobar la efectividad de este modelo se propone construir dos métricas, las cuales pretenden medir la similitud entre la primera y la segunda recomendación y adicionalmente la variedad de la segunda recomendación, las dos métricas son:

Nivel de personalización: Pretende indicar qué tan ajustada está la recomendación a lo que el cliente está buscando:

$$NP = \frac{\Sigma Superclases_{vistas} \cap Superclases_{segunda\ Recomendación}}{\Sigma Superclases_{vistas}}$$

Variedad: Pretende entender qué tantas opciones diferentes se le están presentando al usuario para que este pueda interesarse por otros tipos de artículos:

$$Variedad = 1 - \frac{\Sigma Superclases_{vistas} \cap Superclases_{segunda\ Recomendación}}{\Sigma Superclases_{vistas} \cup Superclases_{segunda\ Recomendación}}$$

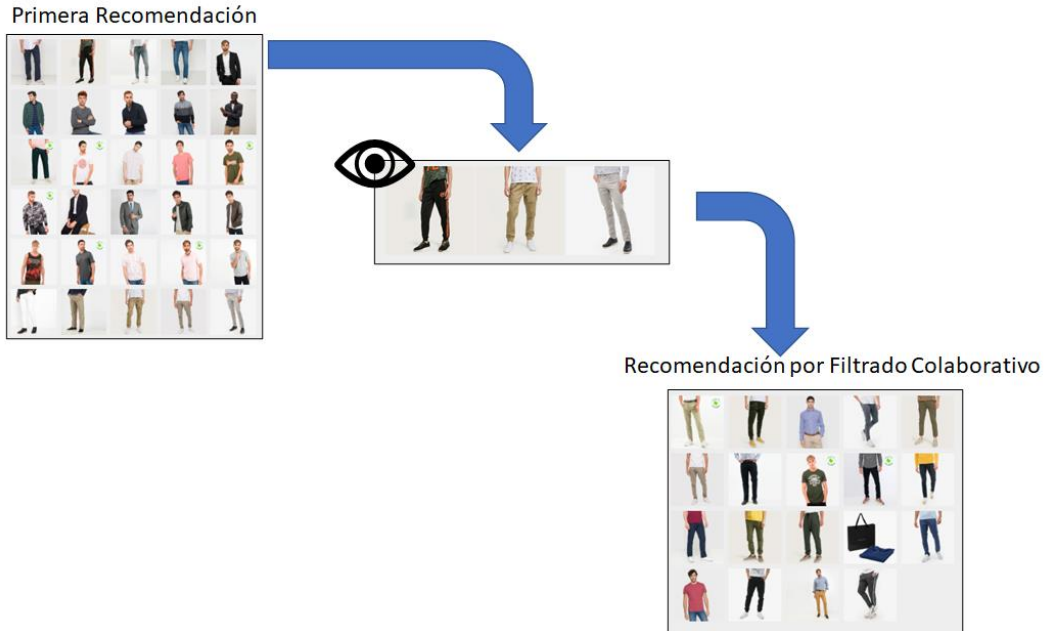
Similitud: Pretende comparar el nivel de semejanza entre la primera recomendación generada por V-Shopper y la generada por el modelo propuesto en este documento. El cálculo para obtener esta métrica es el siguiente:

$$Similitud = \frac{\Sigma Superclases_{primera\ Recomendación} \cap Superclases_{segunda\ Recomendación}}{\Sigma Superclases_{primera\ Recomendación}}$$

4.3. Construcción del modelo

En términos generales la idea del modelo se basa en utilizar los artículos vistos por el usuario, de los que le fueron presentados en la primera recomendación generada por V-Shopper (basada en estilo) y encontrar usuarios que vieron artículos similares, de estos usuarios obtener los productos comprados y con este listado generar una segunda recomendación para V-Shopper basada en el perfil de compra del usuario, a continuación, se ilustra el proceso que se espera que lleve a cabo V-Shopper:

Ilustración 9. Lógica de recomendación del modelo



Fuente: Elaboración propia

Para la construcción del modelo, se divide el dataset entregado en dos bases, entrenamiento (90%) y test (10%). Posteriormente se procede a construir un dataset donde se consigna por usuario los artículos que vio:

```
shown_articles = pd.DataFrame(data['shown_articles'], columns=['sku'])
shown_articles = shown_articles.assign(ID=i)
seen_articles_with_time = pd.DataFrame(data['seen_articles_with_time'], columns=['sku', 'watchingtime'])
seen_articles_with_time = seen_articles_with_time.assign(ID=i)
```

Paso seguido se construye una matriz donde en las filas se tienen los usuarios y en las columnas las superclases, la intersección entre fila y columna contiene la cantidad de artículos vistos por el usuario de esa superclase.

```
train_matrix_sub_class_references = pd.pivot_table(watching_class, values='sku', index='ID',
                                                    columns='superclase', aggfunc='count').fillna(0)
```

Esta matriz que va a ser usada para obtener las distancias y por ende similitudes entre usuarios se le adicionan dos variables a tener en cuenta en el modelo, las cuales son la edad y el género de la persona.

Debido a que el género es una variable categórica esta se va a convertir a dummy para que pueda ser usada dentro del cálculo de distancias.

```
matrix = pd.get_dummies(preMatrix, columns = ["gender"], drop_first = True)
```

Se estandarizan las variables para que todas tengan el mismo peso en el cálculo de distancias:

```
tr = base.append(user, ignore_index=True)
tr = tr.fillna(0)
tr = preprocessing.scale(tr)
```

Para calcular la recomendación de un usuario se corre un modelo algorítmico knn donde se comparan las variables mencionadas anteriormente y se obtiene un número de 10 vecinos más cercanos con el usuario en estudio, este parámetro se configura en 10 para garantizar que la recomendación tenga al menos 10 artículos:

```
k = 11
model_knn = NearestNeighbors(k, 'cosine', n_jobs=-1)
model_knn.fit(tr)
top_k_distances, top_k_users = model_knn.kneighbors(tr, return_distance=True)
```

Posteriormente se obtiene el wishlist de los 10 usuarios más parecidos y la similitud de cada uno, para construir la lista de artículos a recomendar, la cual será retornada en orden importancia basado en dicha similitud:

```
k = 0
for i in similarUsers:
    art = wishBase[wishBase['ID']==i]
    fil = pd.DataFrame(art, columns = ['ID', 'sku'])
    fil = fil.assign(simil=similDist[k])
    wishUser = wishUser.append(fil)
    k = k+1
```

4.4. Evaluación del modelo

En definición con VIPO se determina que los valores esperados para cada una de las métricas son:

- Nivel de personalización: >80% (5 de las 6 superclases)
- Variabilidad: >18% (1 Superclase más) y <50% (3 Superclases más)
- Similitud: >50%

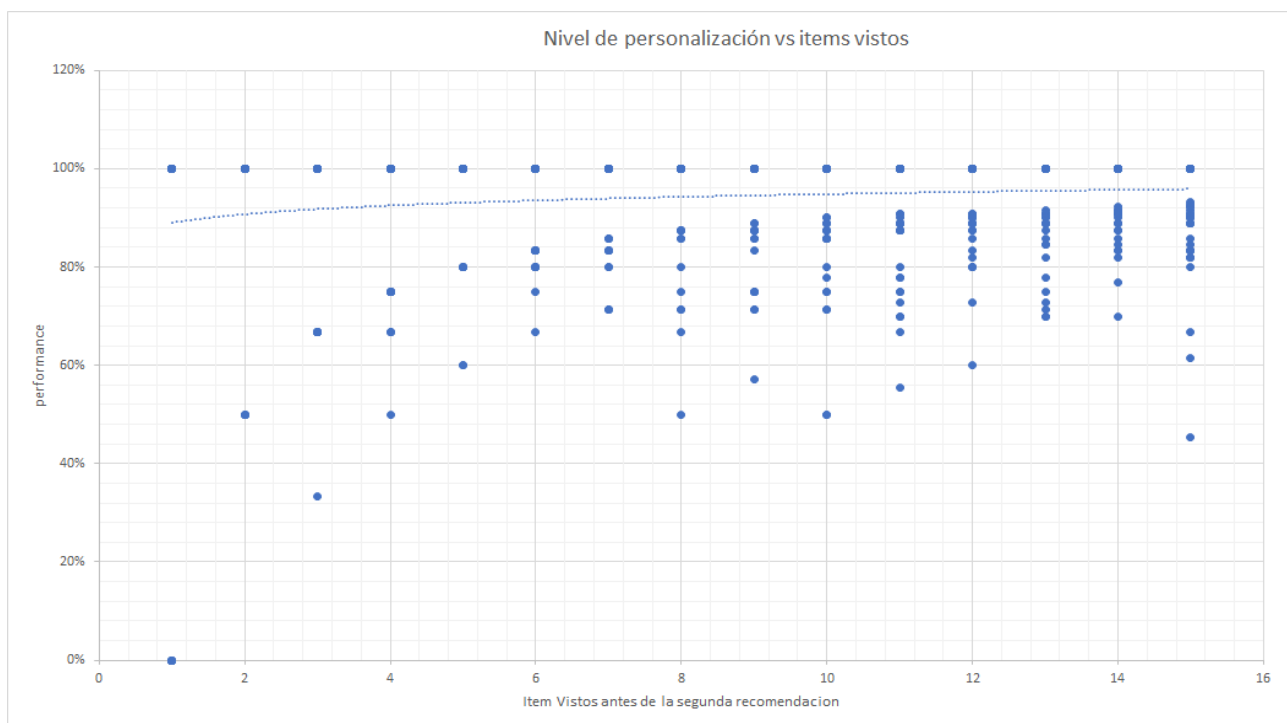
Los valores se calculan tomando como base que en promedio un usuario ve 6 clases en su interacción con V-Shopper.

5. Fase V. Evaluación

5.1. Evaluación de los resultados

Como primer resultado obtenido de la evaluación del modelo se puede observar que el nivel de personalización de la recomendación aumenta a medida que el número de artículos vistos por el usuario es mayor:

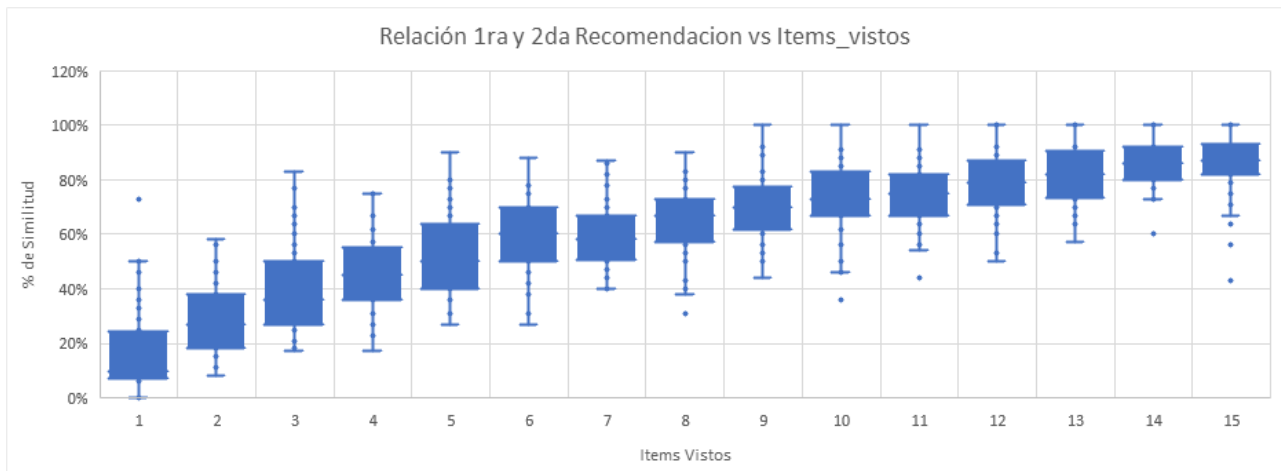
Gráfica 13. Nivel de personalización vs Items Vistos



Fuente: Elaboración propia

Analizando la métrica de similitud entre las recomendaciones se puede inferir que a medida que el número de ítems vistos por un usuario en la interacción con V-Shopper aumenta, el nivel de similitud es mayor, sin embargo, en la mayoría de los casos se garantiza una variación de las clases mostradas al usuario permitiéndole tener más opciones de artículos que generen cierta diferenciación entre las dos recomendaciones:

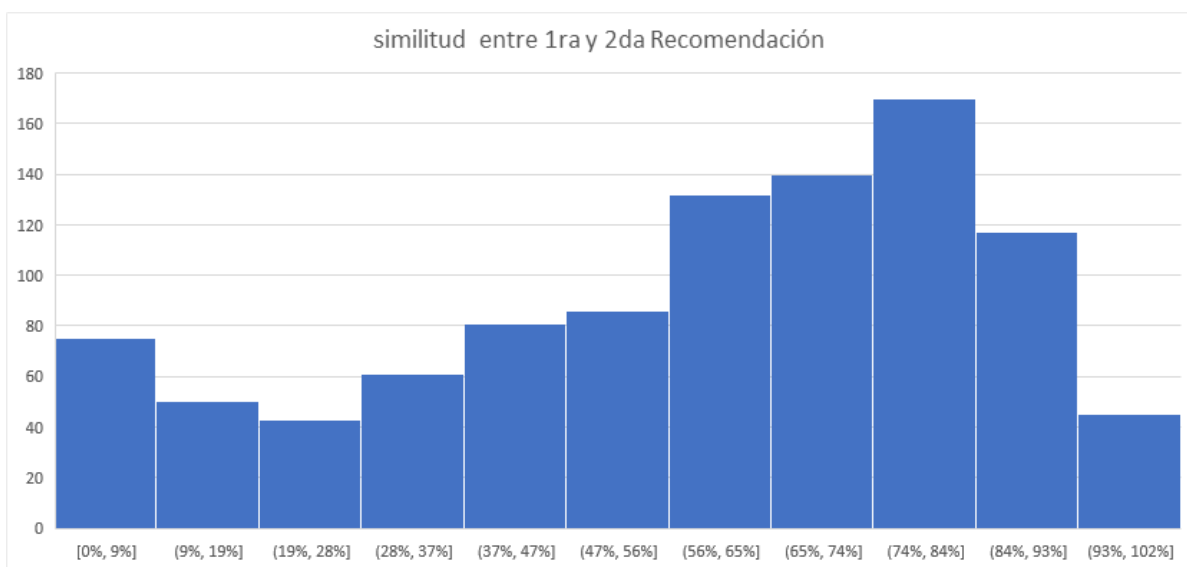
Gráfica 14. Relación 1ra recomendación vs ítems vistos



Fuente: Elaboración propia

Como se puede observar en la gráfica 15, la mayoría de casos de prueba se concentran entre el 56% y el 93% de similitud entre la primera y segunda recomendación:

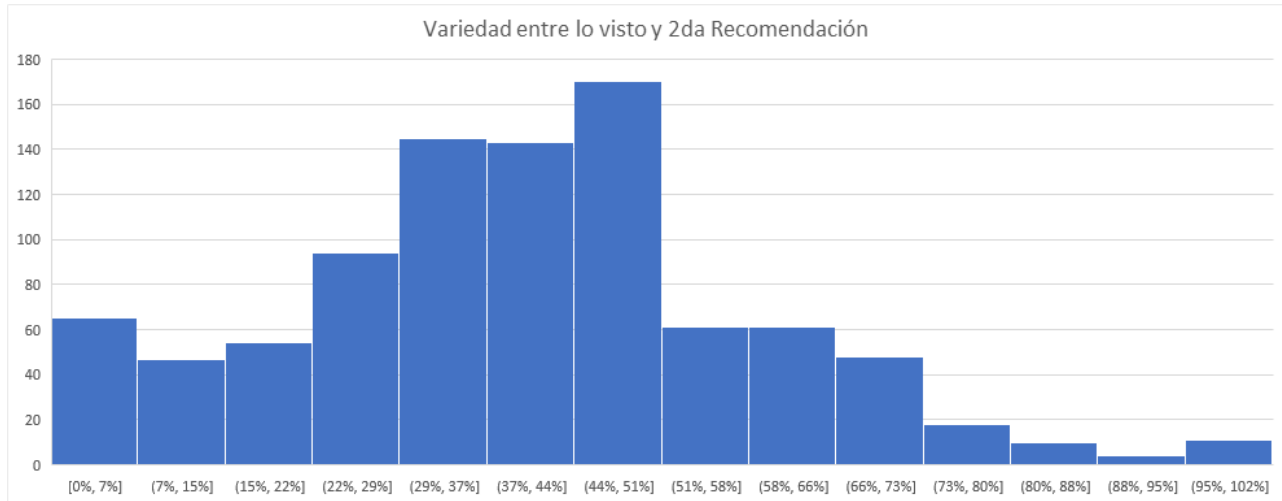
Gráfica 15. Similitud 1ra recomendación vs Segunda Recomendación



Fuente: Elaboración propia

Como se puede observar en la gráfica 16, la mayoría de casos de prueba (65%) se concentran entre el 16% y el 50% de variedad entre las superclases de ítems vistos y las superclases de la segunda recomendación:

Gráfica 16. Variedad entre lo visto y la 2da Recomendación



Fuente: Elaboración propia

Por último, se valida el número de clientes que cumplen con las condiciones establecidas como objetivos de la recomendación se exponen en la siguiente tabla:

| | Variedad | Nivel de personalización |
|--|-----------------|---------------------------------|
| Condiciones | $16% < X < 50%$ | $X > 83%$ |
| Clientes totales | 931 | 931 |
| Clientes que cumplen la condicion | 604 | 815 |
| % clientes que cumplen | 64,88% | 87,54% |
| % clientes que cumplen ambas | 61,87% | |

En conclusión, el modelo cumple con los objetivos planteados, sin embargo, el número de clientes que cumplen con los criterios de variedad no es tan óptimo como el número de clientes que cumplen con el nivel de personalización, debido a esto el porcentaje de clientes que cumplen con ambas condiciones cae al 61.87%, sin embargo es un número aceptable para el modelo.

5.2. Oportunidades de mejora

Una oportunidad de mejora para este modelo, sería poder ligar un poco más la recomendación basada en estilo (primera recomendación) con la recomendación basada en el perfil de compra del cliente (segunda recomendación), esto podría llevarse a cabo si se incluyeran los estilos de los artículos y los colores dentro de la segunda recomendación, sin embargo para poder incluir estas

variables se hace indispensable aumentar la cantidad de registros que se usan en el entrenamiento del modelo y evitar así el problema de la dimensionalidad.

6. Manejo responsable de la información

De acuerdo con la propuesta planteada anteriormente es necesario considerar las formas o modalidad de creación, almacenamiento y organización de la información que se requerirá periódicamente para cumplir con los requisitos mencionados para este trabajo de manera que VIPO pueda cumplir con sus funciones de encargado de las bases y nunca deje de lado la posibilidad de que esto pueda influir negativamente en el modelo de negocio.

Con base en esto es preciso tener en cuenta que como VIPO es una empresa francesa con perfil de introducción global, debe crear una política de protección de datos que tenga como mínimo los siguientes criterios adicionales que complementarían sus políticas actuales:

- Tener en cuenta que los datos comportamentales de cualquier tipo de población se pueden considerar o clasificar como datos sensibles y que, a partir de esta premisa, la analítica derivada de esta lógica se verá muy limitada desde la observación de los mismos y el derecho de las personas sobre disponer de estos aun cuando existan en ficheros automatizados o hayan sido tratados con algoritmos automatizados de gestión de confidencialidad.
- Adicionalmente se entiende que VIPO deja en claro con sus clientes que la seguridad deberá definirla el responsable del fichero, en este caso en específico el retailer. Sin embargo, al ser encargados de datos de sus clientes se deben adherir a la normativa del Marco Europeo de protección de datos y garantizar la correcta transferencia de los datos a los data centers que estos dispongan.
- En complemento a lo anterior es de vital importancia que VIPO resguarde los datos en data centers que garanticen la disponibilidad, integridad y confidencialidad de los datos y, en caso de que tengan data sensible o caracterizada así por sus clientes, esta se resguarde en nubes privadas o servidores propios que minimicen el riesgo de fuga.
- VIPO debe garantizar la ciberseguridad con elementos antivirus que garanticen seguimiento a los datos (DLPS), herramientas antimalware, elementos anti-ataques DDos.
- Las disposiciones de respaldo deben cumplir con lo indicado por normativa francesa con encriptamiento SHA256 en las claves y AES256 en los ficheros.
- Los elementos de ingreso tanto a los servers como a las aplicaciones deben tener elementos de doble autenticación.
- Los procesos de seguridad de Vipo deben estar al nivel de PCI y/o buenas prácticas de la IEC/ISO 27002 e ISO 27001:2015
- Con el alcance del proyecto se puede establecer la sensibilidad ante la fuga de estos ficheros bajo la modalidad de encargado y las responsabilidades definidas bajo el marco de referencia europeo para la protección de datos.

A partir de estas recomendaciones se insta a que VIPO complemente sus políticas y procedimientos de protección de datos y seguridad de la información, con base en un análisis realizado a las propuestas previas en términos de conservación.

Referencias

Albouy, Diane & Adesida, Olabisi (2018). Sustainable Fashion Blueprint Report. (9-10). University of Cambridge MBA. [PowerPoint Presentation \(technofashionworld.com\)](#).

Bakutyte, J. How to use personalized product recommendations (Best practices & examples), 2020.

Benbasat, I. & Zmud, R.W. (2003). "Identity crisis within the IS discipline: defining and communicating the discipline's core properties", in: MIS Quarterly, Vol. 27, No. 1, pp. 183-194.

Daniel B., Eva S., Acceptance of recommendations to buy in online retailing, Journal of Retailing and Consumer Services, Volume 17, Issue 3, 2010, 173-180, <https://doi.org/10.1016/j.jretconser.2010.03.005>.

Davide Di Fatta, Dean Patton, Giampaolo Viglia. The determinants of conversion rates in SME e-commerce websites, Journal of Retailing and Consumer Services, Volume 41, 2018, Pages 161-168, ISSN 0969-6989, <https://doi.org/10.1016/j.jretconser.2017.12.008>.

Dealessandri, M. A game developer's guide to steam wishlists. GamesIndustri.biz Academy, 2020.

Dong, M., Zeng, X., Koehl, L., Zhanga, J. An interactive knowledge-based recommender system for fashion product design in the big data environment. Information Sciences, Volume 540, 2020, 469-488, <https://doi.org/10.1016/j.ins.2020.05.094>.

Keunho Choi, Donghee Yoo, Gunwoo Kim, Yongmoo Suh. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. Electronic Commerce Research and Applications. Volume 11, Issue 4, 2012, 309-317, <https://doi.org/10.1016/j.elerap.2012.02.004>.

Magento Commerce. Ten tips for Turning More Site Visits into Sales: How to increase your conversion rates, 2016.

Schiviani R. Why is Conversion Rate Important for E-commerce, SmartHint Tecnología, 2019.

Sylvain S., Jacques N., The influence of online product recommendations on consumers' online choices, Journal of Retailing, Volume 80, Issue 2, 2004, Pages 159-169, ISSN 0022-4359, <https://doi.org/10.1016/j.jretai.2004.04.001>.

Serrano, S. Personalized product recommendation tips and stats. Barilliance, 2021.

Srinivasan, S. S., Anderson, R., & Ponnayolu, K. (2002). Customer loyalty in E-commerce: An exploration of its antecedents and consequences. Journal of Retailing, 78(1), 41-50.

Tengku M Munawar Chalil, Wirawan Dony Dahana, Chris Baumann. How do search ads induce and accelerate conversion? The moderating role of transaction experience and organizational type, Journal of Business Research, Volume 116, 2020, Pages 324-336, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2020.05.033>.