



Máster en Business Intelligence y Big Data

Proyecto Fin de Máster

Modelos analíticos para la empresa El Catador

**El
Catador**

Autores

Helwing Alejandro Peña de Jesús

Juan José García Villalona

Julio Alberto Peña Ortega

Octavio Antonio Cruz Vásquez

Tutor

Ruben Díaz Sanchidrian

Índice

| | |
|--|----|
| 1. Introducción..... | 5 |
| 1.1 Cifras | 5 |
| 1.2 Distribución..... | 5 |
| 1.3 Principales distribuidores..... | 6 |
| 2. Definición del problema | 7 |
| 2.1 Contexto de la empresa | 7 |
| 2.2 Planteamiento del problema..... | 8 |
| 3. Investigación y toma de datos..... | 10 |
| 3.1 Validación de hipótesis | 10 |
| 3.1.1 Identificación de hipótesis a validar..... | 10 |
| 3.1.2 Proceso de validación..... | 11 |
| 3.1.3 Análisis y conclusiones de las encuestas y entrevistas | 11 |
| 3.2 Análisis preliminar de datos..... | 13 |
| 3.3 Análisis entorno | 14 |
| 3.3.1 Factores político - jurídicos..... | 14 |
| 3.3.2 Factores culturales | 14 |
| 3.3.3 Factores económicos..... | 15 |
| 3.3.4 Factores sociodemográficos..... | 16 |
| 3.3.5 Factores tecnológicos..... | 16 |
| 3.3.6 Factores medio ambientales | 17 |
| 3.4 Análisis competitivo | 18 |
| 4. Plan estratégico | 20 |
| 4.1 Análisis y diagnóstico..... | 20 |
| 4.1.1 Análisis DAFO | 21 |
| 4.2 Definición modelo de negocio | 23 |
| 4.2.1 Propuesta de valor | 24 |
| 4.2.2 Socios claves | 25 |
| 4.2.3 Recursos clave | 26 |
| 4.2.4 Relación con clientes | 26 |
| 4.2.5 Segmento de clientes | 27 |
| 4.2.6 Canales de distribución..... | 27 |
| 4.2.7 Actividades clave..... | 28 |
| 4.2.8 Estructura de costes | 28 |
| 4.2.9 Fuentes de ingresos | 29 |
| 4.3 Plan de acción..... | 30 |
| 4.3.1 Definición del alcance del proyecto: objetivos y métricas | 30 |
| 4.3.2 Análisis de actividades: modelo lógico - arquitectura técnica..... | 32 |

| | |
|--|----|
| 4.3.3 Análisis de recursos: talento humano y recursos físicos | 46 |
| 4.3.4 Gestión del tiempo (cronograma) | 47 |
| 5. Optimización de resultados | 51 |
| 5.1 Beneficios tangibles | 51 |
| 5.1.1 Generación de ingresos..... | 51 |
| 5.1.2 Reducción de costes | 54 |
| 5.2 Beneficios intangibles..... | 55 |
| 5.3 Beneficios estratégicos | 56 |
| 5.4 Análisis de rentabilidad | 56 |
| 6. Conclusión | 59 |
| 7. Bibliografía | 60 |
| Anexo 1. Resultados de las encuestas | 61 |
| Anexo 3. Entrevista en El Catador (Levantamiento de Información) | 65 |
| Anexo 3. Código Modelos Analíticos en Python | 67 |

Índice de figuras

| | |
|---|----|
| Figura 1: Diagrama Ishikawa. Descripción de la problemática desde perspectivas diferentes de negocio. | 9 |
| Figura 2: Tipos de bebidas alcohólicas consumidas en República Dominicana | 15 |
| Figura 3: Canvas modelo de negocio | 24 |
| Figura 4: Propuesta de valor..... | 24 |
| Figura 5: Secuencia de trabajo para la propuesta de valor | 25 |
| Figura 6: Segmento de clientes | 27 |
| Figura 7: Canales de distribución | 27 |
| Figura 8: Actividades clave | 28 |
| Figura 9: Ingresos por ventas de segmento de clientes anual | 29 |
| Figura 10: Ingreso por ventas de segmento de clientes mensual | 30 |
| Figure 11 :Síntesis del proceso | 31 |
| Figura 12: Modelo lógico de la solución propuesta (1) | 33 |
| Figura 13: Modelo lógico de la solución propuesta (2)..... | 33 |
| Figura 14: Flujo de proceso de captura de datos..... | 35 |
| Figura 15: Estructura de almacenamiento de datos..... | 36 |
| Figura 16: Gráfica con resultados del método del codo para identificación de clústeres | 39 |
| Figura 17: Representación gráfica del proceso de clusterizado de los productos..... | 40 |
| Figura 18: Cuadro de mando clientes y recomendaciones..... | 41 |
| Figura 19: CRM EL Catador 1.0 con listado de recomendaciones | 42 |
| Figura 20: Bloques de la Inteligencia de Negocios..... | 42 |
| Figura 21: Arquitectura de Kimball | 43 |
| Figura 22: Cuadrante Mágico de Gartner en 2018 para herramientas Business Intelligence y Business Analytics | 45 |
| Figura 23: Arquitectura técnica de la solución propuesta..... | 45 |
| Figura 24: Organigrama Proyecto Modelos Analíticos El Catador | 46 |
| Figura 25: Cronograma estimativo del proyecto | 48 |
| Figura 26: Ventas vinos por segmento 2018 | 52 |
| Figura 27: Ingresos estimados | 53 |
| Figura 28: Costes estimados | 55 |

| | |
|---|----|
| Figura 29: Cash Flow a 5 años | 58 |
| Figura 30: Indicadores rentabilidad | 58 |

Índice de tablas

| | |
|---|----|
| Tabla 1: Estructura de costes | 28 |
| Tabla 2: Tabla de indicadores | 32 |
| Tabla 3: Ventas vinos por segmento 2018 | 51 |
| Tabla 4: Ventas vino por recomendaciones 2018 | 52 |
| Tabla 5: Comportamiento esperado indicadores | 52 |
| Tabla 6: Tabla de ingresos | 53 |
| Tabla 7: Tabla de costes | 54 |
| Tabla 8: Análisis de rentabilidad | 57 |

1. Introducción

El vino es una bebida alcohólica con beneficios para la salud y que ha acompañado al hombre desde hace siglos. La República Dominicana tiene uno de los mercados de vino más grandes, y a la vez, más nuevo de Centroamérica y el Caribe. El continuo crecimiento del consumo de esta bebida se debe a la influencia de los medios televisivos e impresos, en los cuales presentan el consumo de vino como algo muy a la moda y mencionan los beneficios para la salud. También se debe a la gran influencia cultural de los dominicanos en el exterior. Cada vez que visitan a sus familiares en el país, traen nuevas costumbres, consumo de vino siendo una de ellas. Y por último está el impacto del creciente flujo turístico.

De acuerdo con un estudio realizado La República Dominicana cuenta con una red de distribuidoras e importadores de vinos que provienen de diferentes países como son: España, Italia, Argentina, Alemania, Portugal, Chile, Estados Unidos, Francia, Australia, Nueva Zelandia, Canadá y Sudáfrica. Dentro de los principales distribuidores de vinos se encuentran: Vinos, S.A., Álvarez & Sánchez, El Catador y Manuel González; que cuentan con una gran variedad de vinos muy reconocidas a nivel mundial.

1.1 Cifras

En el 2008, el país consumió más de 500 millones de litros de bebidas alcohólicas, el 1.4% corresponde al consumo de vino con tendencias alcistas. El mercado se valoriza en más de US\$14 millones (RD\$490 millones). Es un mercado muy competitivo y dependiente casi en lo absoluto de las importaciones. Los países de origen de los vinos importados son los siguientes, con sus captaciones del mercado:

- ❖ España 40%
- ❖ Chile 25%
- ❖ Estados Unidos 12%
- ❖ Francia 9%
- ❖ Argentina 6%
- ❖ Italia 5%
- ❖ Resto del mundo 3%

Debido a la diversidad de los consumidores de vinos en República Dominicana, existen varios canales de distribución. Un 40 a 50% de los vinos suelen comercializarse en los supermercados e hipermercados, especialmente los vinos procedentes desde Chile y España. Las botellas de vino que se venden por este canal de distribución tienen precios que oscilan entre RD\$175 y RD\$245 (US\$5 y US\$7), lo que evidencia un precio relativamente asequible para un gran segmento de los consumidores dominicanos.

1.2 Distribución

Otra ventaja que existe en la distribución de los vinos por este canal es la expansión continua de las cadenas de supermercados e hipermercados por toda la geografía nacional.

Otro canal de distribución suelen ser los resorts todo incluido, los cuales captan el 40% de la distribución de vinos en el país. La República Dominicana tiene el mayor sector hotelero en todo el Caribe y Centroamérica, con más de 60 mil habitaciones y en franco crecimiento; además que

recibe más de 4 millones de turistas por año. Debido a que el sector turístico está en fase de diversificación, con el desarrollo del turismo de alto lujo, el turismo de golf y el turismo del medio ambiente o mejor conocido como ecoturismo; los consumidores turísticos son cada vez más exigentes en el tipo de vino que demandan. Por tanto, cada vez son más los vinos finos que se comercializan por esta vía.

Los dos canales de distribución restantes son los restaurantes, los cuales captan el 20% de los vinos, y los hoteles de negocios, en los cuales se comercializa el 2%. En los restaurantes suelen vender vino de todo tipo y precios, aunque la modalidad suele ser por copa. Los vinos chilenos son los que más se venden debido a su buena calidad-precio, ligeramente seguido por los vinos españoles. En los hoteles de negocios, el tipo de vino que se comercializa suele ser de bajo costo, usualmente servido en recepciones matrimoniales y en banquetes.

1.3 Principales distribuidores

El sector cuenta con 5 importadores principales, los cuales son también los distribuidores dentro del país:

- ❖ **El Catador, S.A.**; fundada en 1976, ha sido la empresa líder en el desarrollo del mercado de vinos en la República Dominicana. En la actualidad es el distribuidor exclusivo de marcas de todas las regiones de vinos en el mundo, incluyendo Robert Mondavi, St Francis, Au Bon Climat, Sonoma Cutrer, Fetzer, ZD entre otras.
- ❖ **Marcas Premium, S.A.**; establecida en 1993, actualmente es el representante exclusivo de vinos y bebidas alcohólicas procedente desde California, Francia, Escocia, Italia, Chile, Argentina, España y Australia. Algunas de sus marcas son Kendall Jackson, Ferrari Carano, Altas Cumbres Malbec Vintage, Arbor Mist; todas dentro de un portafolio con más de 100 marcas de vino en total.
- ❖ **Manuel González Cuesta, Sucs. C por A**; fundada en 1945. Varias de las marcas que representa en el país son líderes en sus categorías en diversos países. Además, es el distribuidor exclusivo de Ernest and Julio Gallo winery.
- ❖ **Álvarez y Sánchez, C por A**; fundada en 1979. Además de ser un importador y distribuidor de productos de comida y bebidas a nivel nacional, también cuenta con un portafolio con más de 50 marcas de vinos. Dos de sus marcas son Simi y Marimar Torres, y actualmente están diversificando su portafolio con nuevas marcas de Estados Unidos.
- ❖ **Vinos, S.A.**; fundada en 1980 y en 1995 es adquirida por el grupo Brugal. En la actualidad tienen un portafolio de más de 150 marcas de vino, vodka, cognac, gin, brandy, whisky, sidra y agua mineral. Es la única compañía de su clase con dos centros de distribución (uno en Santo Domingo y otro en Santiago). También cuenta con cuatro tiendas especializadas, tres de ellas bajo la marca Punto y Corcho, y La Enoteca. Y también tienen el club de vino Le Connoisseur; con el propósito de educar, informar y promover el consumo de vino en el país. Sus tiendas están localizadas en Santo Domingo, Bávaro-Punta Cana y en la Marina de Chavón en La Romana.

Existe en el mercado una creciente potencial para los vinos blancos, rojos y las champañas. Los vinos con precios que oscilan entre RD\$175 y RD\$280 (US\$5-US\$8) son los que van a tener el mayor crecimiento. Además, el crecimiento del turismo, de la clase media dominicana, de los restaurantes y de los supermercados e hipermercados, ayudará bastante al futuro desarrollo del mercado de vino en la República Dominicana.

2. Definición del problema

2.1 Contexto de la empresa

El Catador es una empresa dominicana con más de 30 años de experiencia en el mercado de la importación y distribución de las más importantes marcas de vinos de todo el mundo.

Durante todos estos años, la empresa ha mostrado un sólido liderazgo, tanto en los puntos de consumo (hoteles y restaurantes), como en los puntos de ventas (supermercados y mayoristas), manteniendo el más alto posicionamiento en la mente de los consumidores dominicanos.

El Catador posee un portafolio conformado por las casas con mayor prestigio de cada región productora de vinos, y es administradora por la familia Bonarelli. La excelente calidad de sus vinos sumada con sus estrategias de ventas, permiten que nuevos productos puedan ser lanzados en el mercado dominicano y que alcancen un buen posicionamiento frente a sus competidores.

El Catador además se involucra mucho en el área educativa y de difusión de conocimiento. Durante los últimos 15 años ha impartido cursos de vinos a más de 4,000 clientes particulares, personal de hoteles y restaurantes en todo el país, su propósito es desarrollar la cultura del vino en la República Dominicana y como resultado incrementar su consumo y capitalizar su liderazgo.

En el área de Servicio al Cliente El Catador se ha preocupado por mantener un personal altamente cualificado y con un excelente dominio del catálogo de productos que ofertan a sus clientes, es la responsabilidad de este personal el de asesorar a los clientes en el momento de la compra, y de verificar el perfil del cliente para realizar recomendaciones ajustadas al mismo.

El mercado de consumo de bebidas alcohólicas en República Dominicana se encuentra en constante cambio y crecimiento. Por otra parte, el mercado internacional está siendo impactado por nuevas regulaciones y desarrollándose a nivel tecnológico, esto supone para El Catador una necesidad de evolucionar y adaptarse a los nuevos cambios.

Algunos de los retos que la empresa deberá superar son:

- ❖ Adaptarse a las cargas impositivas que los productos vinícolas sufran producto de las guerras comerciales entre los mayores mercados del planeta.
- ❖ Facilitar la compra de vinos a través de aplicaciones móviles y el uso de las opciones más populares de servicio a domicilio (Uber Eats y Glovo en el mercado dominicano).
- ❖ Desarrollo de nuevas aplicaciones y páginas web para la difusión de cursos sobre los vinos y la industria.
- ❖ Implementación de tecnologías Big Data para aumentar la producción y calidad de los vinos.
- ❖ Introducción de nuevos productos que el mercado aún no ha experimentado.

2.2 Planteamiento del problema

En la actualidad la empresa está tardando bastante tiempo y empleando recursos adicionales para poder introducir nuevos productos al mercado, lo que se suma a la problemática de no contar con campañas de mercadeo altamente efectivas. Esto provoca una conducta en sus clientes quienes apuestan por lo seguro y compran solo aquellos productos que ya conocen y les gusta, perdiéndose de la oportunidad de conocer nuevos que podrían ser similares a los que ya consumen y otros totalmente nuevos para ellos pero que de igual forma podrían encantarles.

Por otro lado, la empresa cuenta con un personal de ventas en sus tiendas que hacen recomendaciones a los clientes que los visitan basados en su experiencia y conocimiento previo de los productos. Con la introducción de nuevos productos al mercado, los mismos son recomendados a los clientes que visitan las tiendas sin previamente saber sus patrones de compras y gustos. Este método trae riesgos a la empresa debido a que no se conoce con certeza cómo será la reacción del cliente antes los nuevos productos y a este factor se puede sumar la pérdida de la confianza del cliente ante las recomendaciones del personal que le atiende.

El Catador se preocupa de hacer llegar a sus clientes la información de los nuevos productos que introducen al mercado, para lo cual realizan campañas de mercadeo que en la actualidad conllevan un gran coste económico y que no están siendo tan efectivas como deberían, pues no se focalizan en un público concreto, sino que se hace llegar a todos los clientes por igual.

Estas campañas llegan a un número grande de consumidores que al final no les interesa el nuevo producto, y que podrían mejorarse con la introducción de campañas focalizadas a perfiles de clientes específicos y con comportamientos de compras y gustos similares.

Esta situación trae como consecuencia que el departamento de Mercadeo asuma altos costos por los gastos de la campaña, también afecta al departamento de Logística porque mantiene nuevos productos en el inventario por mucho tiempo y afecta las ventas generales de la empresa ya que muchas veces estos productos no se venden de manera esperada y hay que realizar promociones extras para darle crecimiento.

Todos estos departamentos (Mercadeo, Logísticas, Ventas y Servicio al Clientes) están a la espera de una solución mejor optimizada a la hora de mover estos nuevos productos eficientemente a sus clientes.

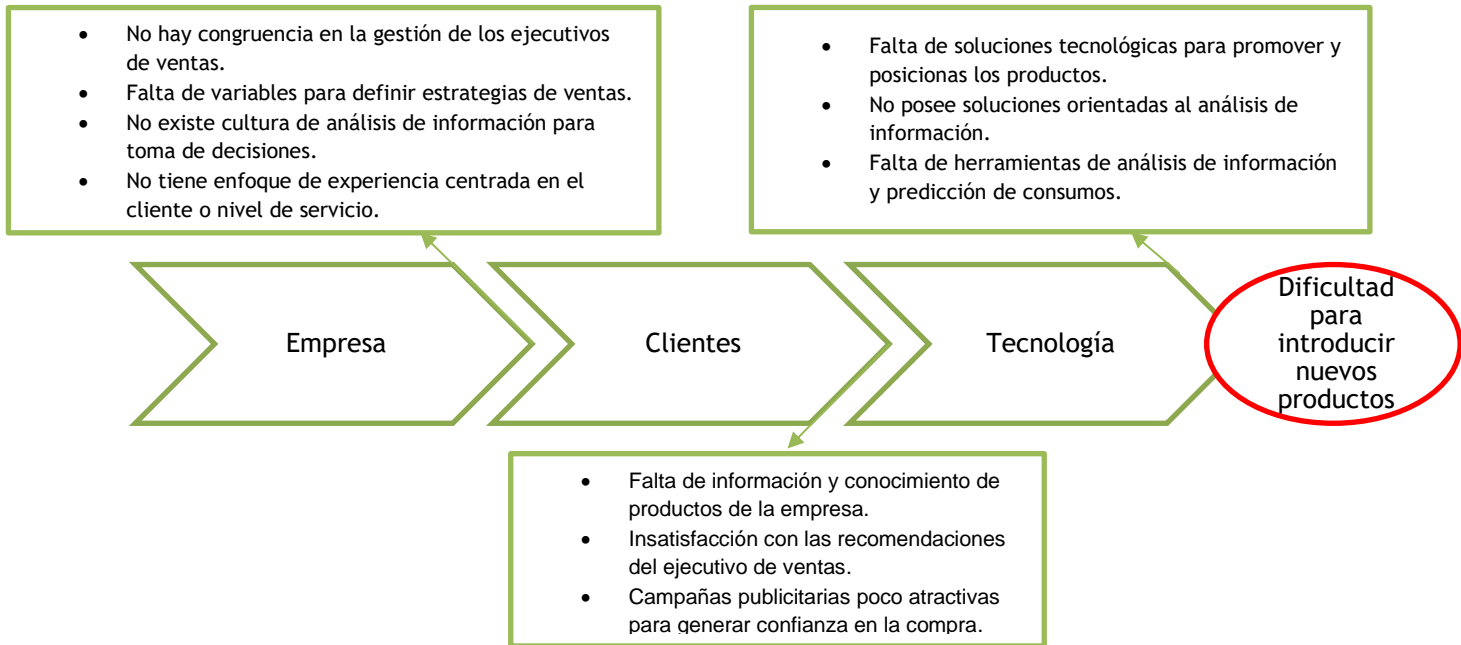


Figura 1: Diagrama Ishikawa. Descripción de la problemática desde perspectivas diferentes de negocio.

3. Investigación y toma de datos

3.1 Validación de hipótesis

3.1.1 Identificación de hipótesis a validar

A través de las siguientes hipótesis planteadas, validaremos nuestro modelo de negocios, y podremos pasar de transformar la información de ideas hacia hechos.

1. El cliente final que compra vinos generalmente compra el mismo producto que conoce y le gusta.
2. El cliente final está abierto a recibir recomendaciones de nuevos productos a comprar.
3. Creo que mi cliente utilizará mi servicio porque aumentará sus ventas introduciendo nuevos productos al mercado y asegurando sus ventas.
4. Creo que mi cliente además aprecia el valor añadido de: disminuir el riesgo de fracaso ante el lanzamiento de nuevos productos.
5. Creo que mi cliente apreciará nuestro servicio ya que simplificamos el proceso de posicionamiento de nuevos productos, pasando de un proceso empírico a un proceso automatizado basado en datos históricos validados.

Establecidas estas hipótesis podemos clasificarlas en diferentes grupos de acuerdo con su origen.

Hipótesis de cliente:

- ❖ El cliente final generalmente compra el mismo producto que conoce y le gusta.
- ❖ El cliente final le gustaría recibir recomendaciones sobre nuevos productos.

Hipótesis de problema:

- ❖ El cliente apreciará mi servicio ya que mejorará el proceso empírico de recomendaciones a un proceso automatizado basado en data histórica.

Hipótesis de producto/solución:

- ❖ A través de mi servicio el cliente aumentará sus ventas introduciendo de manera eficaz nuevos productos al mercado.
- ❖ Del mismo modo apreciará el valor añadido de disminuir el riesgo de fracaso en la venta de esos nuevos productos.

3.1.2 Proceso de validación

Para ejecutar nuestro proceso de validación de las hipótesis del modelo de negocio, nos apoyamos en la realización de varias entrevistas al personal de los Departamentos de TI, Mercadeo y Servicio al cliente de la empresa El Catador.

Así como también la recopilación de datos a través de varias encuestas dirigidas a clientes finales consumidores de vino, esto con el objetivo de conocer desde este punto de vista cómo se sentirá con la solución planteada.

3.1.3 Análisis y conclusiones de las encuestas y entrevistas

Con los fines de validar las hipótesis anteriormente planteadas, hemos realizado entrevistas a los directivos del Catador y un total de 35 encuestas a clientes consumidores de vino, estos resultados se detallan en el anexo.

A efectos de lo observado en las encuestas hemos identificado que un 40% de las personas generalmente no consume la misma marca de vino. Tenemos también que el 100% indicó que puede darse la oportunidad de probar una nueva marca de vino, y bajo los criterios a tomar en cuenta para probar esta marca se detallan: precio y presentación del producto con un 5.7% para ambos mientras que la recomendación se queda con el 88.6%.

Algo importante es que, del total de los encuestados el 82.9% indicó que le gustaría recibir recomendaciones de vinos considerando sus preferencias habituales de consumo. Esto último nos confirma la viabilidad del proyecto.

En cuanto a las entrevistas realizadas, pudimos identificar varias situaciones detalladas por el personal de El Catador, la empresa recientemente estuvo trabajando un proyecto de segmentación de clientes B2C y B2B; donde para los clientes de tipo persona la segmentación es: Collector / Aficionado / Entusiasta / Casual, y estos se basan en diversos factores como lo son antigüedad, valor de compra, tipo de vino, etc. Para el caso de clientes de tipo Empresas se categorizan en Empresa / Corporativo, basados en el volumen de ventas que posee cada cliente.

La segmentación de los clientes se divide en dos grandes grupos:

- ❖ **Grupo 1:** incluye a los segmentos Collector, Aficionado y Corporativo.
- ❖ **Grupo 2:** incluye a los segmentos Entusiasta, Casual y Empresa.

El objetivo de esta segmentación persigue establecer estrategias de mercadeo y publicidad en base a los diferentes grupos de clientes existentes, esto le permite a la empresa tener un detalle más claro y específico del perfil al cual el cliente pertenece y así poder realizar campañas en función al historial de consumo, categoría económica y uno que otro rasgo característico del cliente en cuestión.

La estructura de ventas de la empresa está conformada por una cartera de vendedores quienes atienden a los perfiles de clientes segmentados y una cartera de vendedores de piso quienes están ubicados en cada una de las tiendas físicas.

Se valida que la recomendación de nuevos productos a los clientes la realizan de manera empírica; basados en el conocimiento del vendedor y los hábitos de consumo del cliente se le recomiendan productos similares orientados al segmento al cual el cliente pertenece.

Poseen un CRM donde los vendedores visualizan los vinos favoritos del cliente (información recopilada de las últimas compras), en función de esto el vendedor realiza la recomendación con vinos afines.

La información sobre el cliente, el vendedor la conoce al momento de la facturación no previamente, según nos indican están trabajando en el desarrollo de un App con la cual puedan revisar el perfil del cliente antes del proceso de facturación.

Los vinos favoritos del cliente se eligen en función a los que mayor peso económico tienen en las últimas compras del cliente, dado que según nos explican la cantidad comprada no es un indicador relevante.

Con respecto a los productos, tienen una segmentación definida por el precio de este y se catalogan en 6 segmentos:

- ❖ Categoría Luxury para vinos mayores de 5,500 pesos.
- ❖ Categoría Ultra Premium para vinos mayores de 3,000 y menores de 5,500 pesos.
- ❖ Categoría Super Premium para vinos entre 1,500 y 3,000 pesos.
- ❖ Categoría Premium para vinos entre 750 y 1,500 pesos.
- ❖ Categoría Popular Premium para vinos entre 400 y 750 pesos.
- ❖ Categoría Popular para vinos menores de 400 pesos.

La empresa cuenta con estos principales canales de ventas: Hoteles-restaurantes / Distribuidores / Personas.

Según lo levantado la necesidad de El Catador consiste en que desarrollemos un algoritmo que, en función de los clientes y productos, suministre al vendedor de acuerdo con el cliente consultado:

- ❖ Producto de introducción (recomendación de producto nuevo orientado a su perfil).
- ❖ Producto de variedad (recomendación de productos existentes en el catálogo que se ajusten al hábito de compra del cliente, pero con la peculiaridad de que pueda exceder hasta el 5% del precio habitual de compra del cliente).
- ❖ Producto upsale (recomendación de un producto de un segmento superior al que habitualmente el cliente compra).

Por el lado de las estrategias de ventas, nos indican que estas dependerán del perfil del cliente, y pudieran ser las siguientes:

- ❖ Introducción de productos en carta de vino de restaurantes.
- ❖ Coordinación de eventos para dar a conocer productos.
- ❖ Ofertas en tiendas físicas.
- ❖ Gestión directa (llamadas de ventas y seguimiento).
- ❖ Emails, redes sociales.

A resumidas cuentas consideramos que contamos con suficientes datos para afirmar que el proyecto planteado es viable, además de que nuestras hipótesis quedan altamente confirmadas

mediante los resultados recabados en las entrevistas y los datos obtenidos de los clientes finales en las encuestas.

3.2 Análisis preliminar de datos

La empresa El Catador nos ha facilitado un dataset preliminar en el cual hemos podido identificar varias variables importantes para nuestro análisis. También nos darán acceso VPN a su base de datos donde obtendremos cualquier información necesaria para la creación de nuestros algoritmos.

Las variables que tenemos actualmente son de los clientes, transacciones y productos. A continuación, presentaremos las más importantes de cada renglón.

VARIABLES CLIENTE

- ❖ Número de cliente
- ❖ Perfil del cliente
- ❖ Fecha de creación del cliente

VARIABLES TRANSACCIONES

- ❖ Fecha de la transacción
- ❖ Producto comprado
- ❖ Cantidad de producto comprado
- ❖ Precio de la transacción

VARIABLES PRODUCTO

- ❖ Segmento del producto
- ❖ Tipo y subtipo de producto
- ❖ País de producto
- ❖ Región de producto
- ❖ Uva de producto

Ya con todos estos datos a gran escala podemos empezar a ver cuales de estas variables pueden ser importantes para el entrenamiento y configuración de nuestros modelos analíticos. Con estos datos podemos agrupar clientes y productos por segmento y así entregarle a El Catador las recomendaciones que están esperando.

Este dataset es muy valioso no solo para el departamento de servicio al cliente, sino que también los departamentos de mercadeo y operaciones pueden usarlos para aportar valor a sus departamentos ya sea moviendo el inventario que se tiene en almacén o introduciendo campañas de mercadeo directas basadas en los datos obtenidos.

Para generar aún más valor a El Catador luego de agrupar nuestros productos y clientes por segmentos y grupos de comportamiento similar, podríamos añadir una variable más al dataset indicando cual es el producto más parecido de cada uno de los diferentes segmentos previamente definidos que se le pudieran recomendar a cada cliente.

Teniendo esta información a mano al momento de hablar con cada cliente final, la empresa podría saber de antemano que productos nuevos ofrecer, ver como esta el inventario de cada uno de esos

productos y tomar decisiones de como atacar cada segmento de cliente con campañas publicitarias.

Viendo en detalle los datos que posee la empresa y el alto grado de comunicación que tenemos con la misma tenemos confianza que el proyecto se puede desarrollar y esperamos un alto nivel de éxito, tanto para nuestro grupo como para El Catador.

3.3 Análisis entorno

3.3.1 Factores político - jurídicos

En la República Dominicana existen diversas legislaciones que regulan el comercio, producción y expendio de las bebidas alcohólicas, entre las que podemos citar:

- ❖ Ley General de Alcoholes (Ley 243 del 19 de diciembre del año 1977) que trata sobre la producción, envase y tributación para las bebidas alcohólicas
- ❖ Ley de Prevención y lucha contra el consumo excesivo de bebidas alcohólicas (10 de enero del año 2008)
- ❖ Ley 42-01, Ley General de Salud, que trata sobre el consumo de bebidas alcohólicas en el Capítulo VI, “De las Bebidas Alcohólicas”.
- ❖ Ley 11-92, Código Tributario de la República Dominicana, que en su artículo 361 establece el impuesto que gravará la transferencia de bienes de producción nacional a nivel del fabricante, su importación y la prestación y locación de los servicios descritos en la misma ley, entre estos los productos de alcohol, en la cual el vino de uva tiene una tasa ad valorem de un 20%.
- ❖ Decreto número 308-06 del 24 de julio del año 2006 que regula el expendio de bebidas alcohólicas a partir de las 12 de la noche de domingos a jueves y a partir de la dos de la madrugada los viernes y sábados.

Existen otras mociones y propuestas de Leyes que se encuentran en discusión actualmente, cuya finalidad es controlar el expendio de bebidas alcohólicas en lugares públicos y privados del país.

Partiendo de lo anterior podemos considerar que la fabricación, importación, venta y consumo de bebidas alcohólicas en la República Dominicana cuenta con una base legislativa que regula tácitamente todo lo relacionado a este mercado, así como una fuerte tributación que puede llegar a ser un factor que impacte directamente la producción e importación de bebidas alcohólicas del país, sobre todo el vino.

3.3.2 Factores culturales

El vino es consumido mayormente en el hogar, seguido por eventos especiales (bodas, actividades sociales, entre otras), lo que representa aproximadamente un 23% de los consumidores de bebidas alcohólicas. Solo un 29% de los encuestados en el estudio de “El consumo de bebidas alcohólicas de los dominicanos” realizado por la empresa de investigación Dichter & Neira, expresaron consumir vino de manera exclusiva.

La Cervezas es la categoría más consumida en República Dominicana y a su vez tiene la mayor cantidad de consumidores que no están eligiendo otro tipo de bebida.

De los 398 encuestados, El 35% indicó consumir bebidas alcohólicas mayormente en su hogar, mientras que el 18% consume en colmados, 16% consume en bars o discotecas y el 13% consume en drink store's.

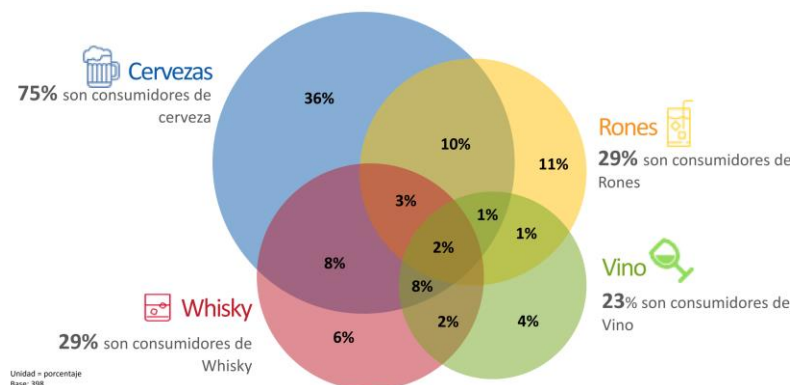


Figura 2: Tipos de bebidas alcohólicas consumidas en República Dominicana

En la República Dominicana no existen restricciones para el consumo de alcohol en las vías públicas y parques, sin embargo, sí tiene restricciones parciales respecto a la publicidad y el patrocinio relacionado al consumo de alcohol y a pesar de que por Ley existe un horario restrictivo para el expendio de bebidas alcohólicas, este se cumple parcialmente.

3.3.3 Factores económicos

Acorde al estudio de “El consumo de bebidas alcohólicas de los dominicanos” realizado por la empresa de investigación Dichter & Neira, a juzgar por la cantidad de alcohol puro reportado por los productores y comercializadores de alcohol a la DGII, el consumo global de alcohol en el país creció un 2.3%.

En 2015 la Dirección General de Impuestos Internos de República Dominicana (DGII) recibió reportes de ventas de 30.4 millones de litros de alcohol puro en bebidas alcohólicas. De esa cantidad, 16.8 millones (55.2%) fue alcohol utilizado en cervezas, que es la bebida con menor grado de esa sustancia en cada litro.

En el mismo año la DGII registró recaudaciones por RD\$16,794.3 millo es por concepto de impuestos a los productos derivados del alcohol. De esa cantidad, RD\$11,974 millones (71.3%) fue recaudada solamente por los impuestos a las cervezas. El restante 28.7% pesa sobre otras bebidas como el ron, vino, vodka y whisky. Lo anterior indica que la cerveza es la bebida por excelencia de los dominicanos.

El Artículo 361 de la Ley 11-92 (Código Tributario), establece un impuesto al consumo de los bienes y servicios incluidos en el Capítulo IV de dicha Ley. El Artículo 362 del Código Tributario, establece que este impuesto gravará la transferencia de algunos bienes de producción nacional a nivel del fabricante, su importación y la prestación y locación de los servicios descritos en la referida Ley.

La base imponible de este impuesto está considerada en 2 partes:

- Cuando los **bienes** transferidos a nivel de fabricante sean alcoholes, cervezas, o productos del tabaco, se tomará como base el precio de venta al por menor, antes de ser aplicado este impuesto.
- En el caso de las importaciones se tomará como base el precio de venta al pormenor, atendiendo al valor definido para la aplicación de los impuestos arancelarios, todos los tributos a la importación o con motivo de ella con excepción del Impuesto sobre las Transferencias de Bienes Industrializados y Servicios (ITBIS).

Los vinos presentan un 20% de impuesto AD VALOREM.

3.3.4 Factores sociodemográficos

República Dominicana establece en 18 años, como la edad mínima requerida para el consumo de alcohol, sin embargo, estudios a nivel mundial señalan que el primer contacto de los adolescentes con el alcohol se produce cada vez a edades más tempranas, en torno a los 12 o 13 años.

República Dominicana consume 6,9 litros de alcohol puro per cápita anual lo que hace que ocupe el puesto número 11 entre los países de América Latina con mayor porcentaje de consumo de alcohol puro per cápita por año, acorde al estudio de la OMS en el que recoge las estimaciones sobre porcentajes de consumo de alcohol en el mundo, basados en las estadísticas proporcionadas por los propios países.

En una encuesta llevada a cabo en 2003 entre alumnos de 13 a 17 años en cuatro países centroamericanos y en República Dominicana, se evidenció que las adolescentes mujeres en República Dominicana (38%) muestran las tasas más elevadas de consumo de alcohol, inclusive en estado de embarazo.

Acorde al estudio realizado por Dichter & Neira, los hombres son los mayores consumidores de bebidas alcohólicas, principalmente en edades comprendidas entre los 25 a 49 años. Este porcentaje está concentrado en su mayoría en Santo Domingo

Ante la alta tasa de mortalidad que conllevan los accidentes de tránsito en República Dominicana debido a la ingesta de alcohol, las autoridades se han visto en la necesidad de establecer fuertes normativas regulatorias que contribuyan a la prevención de estos eventos, así como a la conservación general de la salud.

De cada cien accidentes de tránsito, un 70% tiene como causa la ingesta de alcohol, provocando la segunda tasa de mortalidad más alta del mundo por accidentes de tránsito terrestre relacionados con el consumo de alcohol, a razón de 42 muertos/100 mil habitantes.

Del 2012 al 2013, los médicos de la Unidad de Terapia Intensiva del hospital José M. Cabral y Báez, brindaron asistencia a 125 pacientes con traumatismo severo a causa de accidentes de tránsito. De esa cifra fallecieron 50, es decir, un 40% del total.

Dado que pasaron en el hospital un promedio de 35 días y el costo promedio por paciente fue de 1.3 millones de pesos, la asistencia médica de los 125 accidentados tuvo un costo global de 162 millones de pesos para el hospital.

De los 125 traumatizados graves, 82 habían consumido alcohol de manera excesiva y 20 fueron arrollados por conductores que conducían mientras consumían alcohol.

3.3.5 Factores tecnológicos

La evolución y el desarrollo de las TICs supone un mayor acceso a la información que propicia la aparición de nuevos hábitos sociales y de consumo, lo que ha provocado el desarrollo de herramientas que permitan medir las tendencias de consumo, para brindar más capacidad de negociación y de elección para los clientes.

Cerca de 80% de las implementaciones tecnológicas empresariales son el resultado de las solicitudes (y presión) de los consumidores y no de iniciativas per se de los equipos de TI. La tecnología corporativa es un seguidor de la tecnología de consumo.

Herramientas como inteligencia de negocios, machine learning y big data, permiten ahorrar tiempo, definir estrategias que permitan establecer mejores precios, lograr diferenciación de los

productos que se comercializan, desarrollar productos y ofertas enfocadas en los modelos de consumo de los clientes, así como romper barreras relacionadas a la gestión de ventas tradicional.

Estudiar las necesidades de cada consumidor permite prácticamente garantizar que se produzca la venta.

Las comunidades de usuarios (foros y redes sociales principalmente) también permiten obtener información o conocimientos, sobre las experiencias de los clientes con los productos o servicios que han adquirido.

Si ya ha sido una auténtica revolución para las compañías adaptarse a este nuevo entorno, el gran reto llegará de la mano de la Generación Z, o los también conocidos nativos digitales (los nacidos a partir del año 1996).

Las empresas deben adaptarse a las exigencias de estos consumidores que son aún más tecnológicos y sociales que nunca y que poseen una gran cantidad de información, lo que hace más difícil la gestión de venta y mantener la competitividad. Sin conectividad e información, hoy es imposible competir.

3.3.6 Factores medio ambientales

Un estudio de la Universidad de Washington en Seattle reveló que ningún nivel de consumo de alcohol es seguro. El alcohol ha sido la causa del 12% de muertes en hombres entre 15 y 49 años, además de que solamente un consumo de cero alcoholes minimiza el riesgo general de pérdida de la salud.

Recientemente se encontró correlaciones claras y convincentes entre el consumo de alcohol y las muertes prematuras, el cáncer y los problemas cardiovasculares.

El consumo de alcohol se ha asociado con 23 problemas relacionados con la salud, incluyendo enfermedades cardiovasculares, diferentes tipos de cáncer, cirrosis, diabetes, epilepsia, pancreatitis, infecciones respiratorias y tuberculosis, entre otros. Sin embargo, estudios y científicos de todo el mundo hacen referencia a los descubrimientos de lo bueno que conlleva el consumo moderado de una copa de vino diaria, frente a otro tipo de bebidas.

El consumo moderado de vino, especialmente de vino tinto, tiene beneficios para la salud cardiovascular. Otros beneficios para la salud son:

- Fortalece los huesos
- Reduce el riesgo cardíaco
- Retarda el envejecimiento
- Es antibacteriano
- Es un aliado para bajar peso pues disminuye el colesterol
- Potenciador para nuestro cerebro
- Protege contra las quemaduras solares graves
- Combate el cansancio
- Aumenta nuestras endorfinas
- Limpia nuestro paladar
- Disminuye las posibilidades de cáncer de colon

Uno de los últimos trabajos que ha demostrado estos beneficios es el realizado por el Centro de Investigación Biomédica en Red-Fisiopatología de la Obesidad y la Nutrición, CIBERObn y publicado en American Journal of Clinical Nutrition. El objetivo de la investigación ha sido evaluar los efectos del etanol (alcohol etílico) y los polifenoles (componentes no alcohólicos como el resveratrol) del vino en el proceso de inflamación de las arterias en pacientes con alto riesgo cardiovascular.

Tras analizar a 67 pacientes, los investigadores han demostrado que tanto el etanol como los polifenoles tienen efectos beneficiosos sobre las moléculas inflamatorias causantes de la aterosclerosis, especialmente en sus estadios tempranos, mientras que el exceso de alcohol en el organismo aumenta la presión arterial, lo que favorece la aparición de hipertensión.

En el país, Salud Pública estima que existe una relación causal entre el consumo excesivo de alcohol y una serie de trastornos mentales y enfermedades no transmisibles que han venido presentándose en los últimos años.

3.4 Análisis competitivo

El mercado del Big Data en República Dominicana se encuentra en constante crecimiento. A partir del año 2017 el gobierno dominicano ha mostrado su apoyo al desarrollo de la industria y a la formación de profesionales en el área de Big Data y Business Analytics a través del proyecto República Digital. Un claro ejemplo es este mismo máster auspiciado por el Ministerio de Educación de la República Dominicana y la EOI.

Aun así, es común observar que las soluciones de Big Data que se encuentra en operación por lo general se construyen a lo interno de las organizaciones, reclutando personal cualificado y emprendiendo proyectos de digitalización para dar respuesta a necesidades de las empresas. Tal es el caso de El Catador, que de manera sencilla han adoptado la herramienta Microsoft Power BI para la creación de cuadro de mando de seguimiento de ventas.

A modo general se pueden identificar en el sector empresas independientes que proporcionan servicios de Big Data y Business Analytics a otras empresas de sectores diversos. Sus servicios van desde la implementación de las soluciones hasta la consultoría técnica y de negocio.

Algunas de las empresas identificadas son:

Dell República Dominicana: la empresa Dell proporciona servicios de consultoría a empresas de todos los tamaños y sectores, ayudándoles a obtener información que les permita una toma de decisiones más rápida y efectiva.

Analytica: es una empresa enfocada en prestar servicios de consultoría económica y de estrategias al sector empresarial, instituciones multinacionales y al Estado. Su principal objetivo es aportar en el proceso de Toma de Decisiones, mediante el diseño e implementación de estrategias de negocio que aumenten el valor de sus empresas clientes. Sus productos aportan análisis de indicadores económicos a nivel nacional e internacional. Sus soluciones se dirigen a los sectores: Financiero, Gremios Empresariales, Industrial, Retail, Vehículos, Seguros y Salud y Telecomunicaciones.

Multicomputos: es una empresa que brinda soporte a las organizaciones para el desarrollo y fortalecimiento de su infraestructura de TI así como de los sistemas que la componen.

Soluciones globales: es una empresa de consultoría tecnológica que principalmente da servicios a clientes del sector financiero y de las telecomunicaciones.

Kio Networks: es una empresa que desarrolló soluciones de Data Analytics para sectores de Retail, Servicios Financieros y Telecomunicaciones. Estos se encargan de habilitar modelos de analítica

que ayuden a las empresas a mejorar su eficiencia operativa, tomar decisiones de crecimiento y de nuevos negocios, reducir riesgos y acercarse a sus clientes.

Entre las principales herramientas que utilizan se encuentran:

- ❖ Data Warehouse/Data Mart: SAP Hana, Oracle, Microsoft Sql Server e IBM DB2.
- ❖ Arquitectura Open Source: Apache Spark, Hadoop, Cloudera y HortonWorks.
- ❖ Data Transfer: SAP BODS, Oracle Data Integrator, Microsoft SQL Server Integration Services, IBM DataStage y Talend*.
- ❖ Analytics: R, Alteryx, SAS y Python.
- ❖ Visualizadores: Tableau, QlikView, SAP Business Objects y Power BI.

4. Plan estratégico

4.1 Análisis y diagnóstico

Para hacer un estudio de la situación y estrategia del proyecto usaremos la herramienta DAFO. Considerando que el DAFO es una herramienta que, respondiendo a cuatro enfoques o preguntas de los cuales se determina su propio nombre los cuales son Debilidades, Amenazas, Fortalezas y Oportunidades.

Tomaremos en cuenta los antes mencionados factores para determinar a partir de ellos las cosas que se encuentran dentro de la empresa y podrían impedir el éxito de nuestro proyecto o que lo hagan terminar siendo una herramienta ineficaz a la hora de presentar a la empresa con la información que necesita para mejorar y optimizar su proceso de ventas y así mejorar su posición ante sus clientes y enfrente de sus rivales, los factores que se engloban en este punto son los que consideramos Debilidades.

Luego procederemos a identificar aquellos factores que estando fuera de la empresa, factores externos, pueden limitar el funcionamiento correcto de nuestra solución o de alguna forma debilitar su impacto al momento de intentar mejorar la posición de la empresa en el mercado, estos factores serán listados como Amenazas.

Después de analizar aquellos factores que podrían, de alguna manera ser considerados como negativos, procederemos a iniciar con un análisis de otros factores que podríamos considerar como positivos a nuestro proyecto, y de la misma forma se pueden clasificar como factores externos e internos, iniciando con aquellos que denominaremos como Fortalezas, que son aquellos factores internos que permiten a la empresa sentar bases sólidas en el mercado, que la hace ser competitiva en el mercado y frente a sus competidores directos, entre estos factores podemos incluir sus recursos humanos y materiales que se esfuerzan por lograr ese factor diferenciador de la empresa.

Finalmente analizaremos los factores externos que pueden tener un impacto positivo en el modelo de negocio de la empresa en este marco veremos los factores políticos, económicos y otros que pueden ayudar a la empresa a mejorar sus ventas o abrir nuevos mercados, así como encontrar nuevas oportunidades de negocio que vengan impulsadas por cualquier cambio regulatorio del marco legal que impacte el negocio de la empresa para la cual estamos desarrollando nuestra propuesta de mejora de proceso. Entre las oportunidades también consideraremos las tendencias sociales, culturales que podrían permitir a la empresa a abrir nuevos nichos de mercado que antes no habían sido tomados en consideración como posibles entornos de negocio para la misma.

4.1.1 Análisis DAFO

Debilidades

Falta de variables para definir estrategias de ventas efectivas.

No existe cultura de análisis de información para toma de decisiones.

No tienen enfoque de experiencia centrada en el cliente o nivel de servicio.

La empresa cuenta con un personal de ventas que hace recomendaciones basadas en sus consideraciones sobre el producto.

Alto costos para campañas publicitarias que no motivan al consumo de nuevos productos por parte de los clientes.

No poseen soluciones tecnológicas orientadas al análisis de la información.

Amenazas

Integración de la herramienta de análisis con el sistema CRM actual que utiliza la empresa.

Que el ETL no se realice de manera automática desde el CRM, debido a la posible incompatibilidad entre los sistemas.

No uso de la herramienta de análisis como insumo para la definición de estrategias comerciales y de mercadeo.

No se ha definido el proceso de implementación por parte de la empresa.

Fortalezas

La empresa cuenta con la infraestructura necesaria para implementar el proyecto.

Tienen personal capacitado para recolectar, analizar y realizar propuestas en base a los datos obtenidos.

La alta gerencia de la empresa está interesada en la implementación del proyecto.

Los departamentos de Mercadeo, Logística, Ventas y Servicio al Cliente están a la espera de una herramienta que les permita gestionar sus procesos de manera efectiva.

Oportunidades

Implementación de cuadros de mando analíticos con la información recopilada.

redefinición de la segmentación de clientes B2C / B2B basada en el consumo, para la elaboración de estrategias comerciales y de mercadeo basadas en los modelos analíticos que provea la herramienta.

Elaboración de estrategias y recomendaciones de productos personalizadas, en base al histórico de consumo y las recomendaciones realizadas por la herramienta.

Desarrollar una aplicación móvil que se apoye en la herramienta de análisis para facilitar el proceso de compra a los clientes.

Campañas comerciales focalizadas en perfiles de clientes.

Debilidades

Dentro de las debilidades encontradas vemos que la organización no posee un entendimiento claro de aquellas variables que desea analizar con miras a poder definir estrategias que le permitan fortalecer el proceso de ventas lo cual puede venir dado por el hecho de que no existe una cultura que motive el análisis de información con miras a la mejora de la gestión. La empresa posee herramientas tecnológicas que le permiten administrar la transacción, pero no posee herramientas que le permitan convertir estos datos transaccionales en inteligencias que sirva como soporte a la gestión de la operación.

Debido a que la empresa actualmente basa sus recomendaciones a los clientes en la experiencia de los vendedores no puede proveerle a estos con una experiencia estandarizada, sino que dependiendo del punto de vista y otros factores humanos los vendedores pueden realizar recomendaciones las cuales, quizás no están alineadas con la estrategia de ventas que pueda tener la empresa en un momento determinado.

Las campañas publicitarias se ven afectadas en esta falta de conocimiento puntual de sus clientes y por lo tanto la empresa debe realizar campañas muy grandes para intentar poder posicionar sus productos, nuevos y aquellos que sus ventas hayan mermado pero que aún se encuentren en stock.

Amenazas

Como amenazas al proyecto hemos encontrado que pueden existir elementos no evaluados para completar la integración de las herramientas que sean desarrolladas con el CRM que actualmente posee la empresa, estas herramientas de apoyo a la gestión transaccional de la empresa son, usualmente, cerradas y no permiten de manera transparente que se le integren soluciones de terceros.

Otra posible amenaza es relacionada a los errores técnicos que puedan provocar que el proceso de carga y transformación de la información entren en fallo y dejen de funcionar, este proceso es la base fundamental para que las herramientas que sean desarrolladas por nosotros para nuestro cliente generen los resultados que serán esperados de las mismas. Es posible que errores no manejados durante el proceso de análisis y desarrollo puedan surgir y ante un fallo no esperado el cliente pueda percibir la herramienta como no confiable.

Desde el punto de vista humano puede ser que los usuarios no incorporen el uso de la información provista por nuestra herramienta como parte de su operación diaria y que la misma caiga paulatinamente en desuso por estos. El proceso de implementación de cualquier nueva herramienta de información trae consigo la posible reacción negativa al cambio, sobre todo cuando esta nueva herramienta viene cambia una rutina que los vendedores ven como propia como es la de sugerencias al cliente.

Al momento actual en el que nos encontramos realizado este análisis el cliente no tiene un procedimiento claro de implementación de las herramientas que puedan resultar del proceso de análisis y desarrollo que estaremos llevando con ellos con la finalidad de lograr una mejora en su proceso de ventas. Esta falta de procedimientos puede ser un elemento que detenga parcial o definitivamente la implementación de las soluciones que sean desarrolladas para ellos.

Fortalezas

Nuestro proyecto cuenta con el apoyo de la gerencia de la empresa, así como de los departamentos que se verían envueltos en el mismo, los líderes de los departamentos comprenden que les es necesario contar con una herramienta como la que le proponemos con la finalidad de eficientizar su proceso de ventas y colocación de nuevos productos.

La empresa también posee la infraestructura tecnológica requerida inicialmente, así como el personal con las capacidades necesarias para poder llevar a cabo la implementación y mantenimiento futuro de las herramientas que puedan surgir a partir del proceso de desarrollo en el que nos veremos envueltos juntos con ellos en lo adelante.

Oportunidades

Partiendo del análisis de la información podemos detectar una serie de oportunidades todas ellas dirigidas a la optimización de los procesos de ventas, posicionamiento, experiencia del cliente, mejora de las recomendaciones entre otras que pueden beneficiar al personal de ventas y mercadeo.

Podemos preparar la empresa para en el futuro crear nuevos cuadros de mando, reportes, predicciones y otros tipos de análisis basados en las herramientas que estamos proponiendo que lleven siempre a ajustar o mejorar los procesos establecidos de operación. Este desarrollo puede ser extendido a otros niveles de la empresa y proveer a ellos de información para ayudar en la toma de decisiones de gestión.

Puede ser extendido, el desarrollo propuesto, a una herramienta móvil que permita llevar los resultados a la mano en cualquier momento para facilitar su consulta y no limitarla a que sea realizada mediante el CRM de la empresa.

Entre otras cosas la organización puede mejorar la forma en que realiza las campañas de promoción de productos basando las mismas en los resultados que puedan ser obtenidos del análisis de la información que provean las herramientas que proponemos.

4.2 Definición modelo de negocio

Nuestro proyecto Modelos analíticos para la empresa EL Catador nace como proyecto final del “Máster en Business Intelligence y Big Data” y se enfoca en ayudar a la empresa a pasar de hacer recomendaciones de productos basadas en experiencia a recomendaciones de productos basadas en datos.

Este proyecto de optimización empresarial llega para ayudar a EL Catador a aprovechar y utilizar los datos que ya están colectando para generar nuevos ingresos y reducir gastos innecesarios.

A continuación, se define nuestro modelo de negocio partiendo de la propuesta de valor y relacionando la misma con los recursos, segmentos y actividades clave de la empresa. También se presenta como esperamos que nuestro proyecto ayude a El Catador a generar ingresos y cuales costes están considerados. Para esto utilizaremos el Modelo Canvas y así concretizar la idea y visión global de nuestro proyecto y ver como se interconectan los distintos elementos de este.

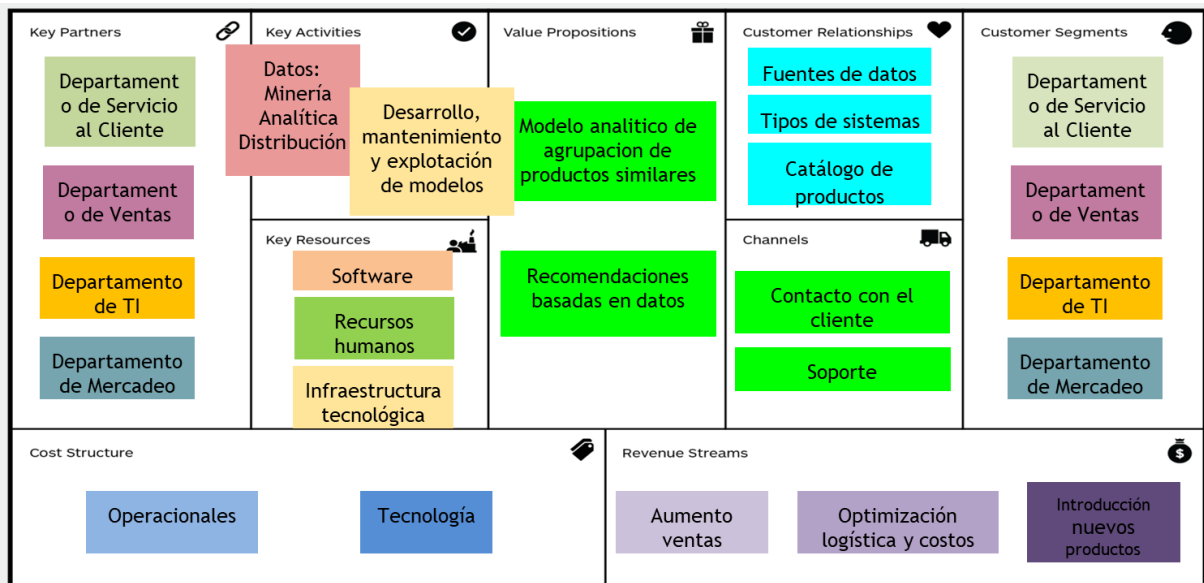


Figura 3: Canvas modelo de negocio

4.2.1 Propuesta de valor

La propuesta de valor de nuestro proyecto consta de modelos analíticos que agruparán los productos ofrecidos por la empresa de acuerdo con su similitud. Con esta agrupación de productos similares de diferentes segmentos, haremos recomendaciones a los clientes finales basadas en compras anteriores.

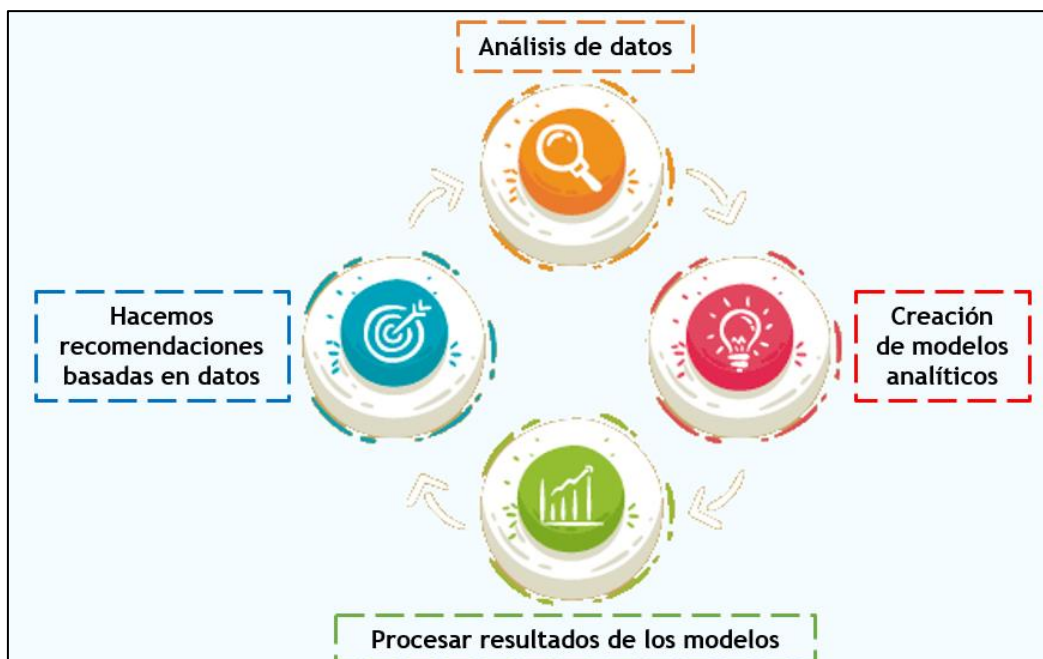


Figura 4: Propuesta de valor

Nuestra propuesta de valor está directamente relacionada con una necesidad de la empresa El Catador, que es tratar de vender productos diferentes y en distintos segmentos a los clientes que

compran de manera frecuente en sus tiendas. Nuestros modelos analíticos permitirán a la empresa conocer mejor a sus clientes y acercarse más a ellos usando productos que van a ser bien recibidos por los mismos.

Un beneficio adicional y con gran valor añadido es que al saber que productos son similares, cuando El Catador quiera introducir un producto nuevo al mercado pues solo tendría que agregar este producto y sus características a la base de datos que alimentará nuestros modelos y reentrenarlos con esta nueva información. Como resultado ya sabrá cuales son los productos existentes que son similares y entonces se los podrá ofrecer a clientes con alta probabilidad de aceptación.

Para lograr esto nuestra secuencia de trabajo que formaría nuestra propuesta de valor es:



Figura 5: Secuencia de trabajo para la propuesta de valor

4.2.2 Socios claves

Al ser un proyecto de optimización en una empresa existente, nuestros socios claves serán los distintos departamentos con los que estaremos interactuando para desarrollar el proyecto. Estos departamentos y su propósito en nuestro proyecto son:

- ❖ **Departamento de tecnología (TI):** obtener información de los sistemas CRM de la empresa que servirán de base para los modelos analíticos, implementar y administrar el modelo analítico.
- ❖ **Departamento de ventas:** proveer información sobre los clientes, volumen e historial de ventas (por clientes, por tiendas, por productos, por temporadas, etc), ciclo de vida de los productos, rotación de ventas, etc.
- ❖ **Departamento de mercadeo:** proveer la estrategia comercial actual de la empresa, suministrar información de tendencias del mercado, proveer información sobre el consumo de los productos (temporadas, ofertas, etc) y los clientes de la empresa (empresariales, individuales, etc).

- ❖ **Departamento de servicio al cliente:** informaciones sobre reclamaciones, principales situaciones posventa, niveles de satisfacción de los clientes, nivel de servicio (cumplimiento de SLAs), información proceso retención de clientes, etc.

4.2.3 Recursos clave

En nuestra propuesta de valor el recurso más importante son los datos de clientes, productos y transacciones que nos proporciona EL Catador. Con estos datos crearemos nuestros modelos analíticos que son la base de nuestro valor añadido. Aparte de los datos, también tenemos varios recursos claves en las siguientes categorías:

Software

- ❖ **Portal web:** Donde en un futuro la empresa tendrá ventas digitales.
- ❖ **Sistemas del cliente:** Gestión de recomendaciones y productos.
- ❖ **Lenguajes de programación:** Con los cuales crearemos y se mantendrán los modelos.

Infraestructura tecnológica

- ❖ **Bases de datos de clientes:** Fuente de datos para nuestros modelos.
- ❖ **Base de datos productos:** Fuente de datos para nuestros modelos.
- ❖ **Base de datos transacciones:** Fuente de datos para nuestros modelos.
- ❖ **Clúster de servidores:** Donde alojaremos nuestros modelos una vez implementados.

Recursos humanos

- ❖ **Encargado departamento de TI:** Nos provee los datos y espacio en los servidores para implementar los modelos.
- ❖ **Encargado departamento de Mercadeo:** Se asegura de hacer promociones basadas en las recomendaciones.
- ❖ **Encargado departamento de Servicio al Cliente:** Entrena a los vendedores para hacer recomendaciones que provean los modelos.
- ❖ **Analistas de ventas:** Revisan y monitorean el comportamiento de las recomendaciones hechas por los modelos.
- ❖ **Analistas de datos:** Siguen el comportamiento de los datos luego de la implementación de los modelos.
- ❖ **Representante de Servicio al Cliente:** Encargado de hacer la recomendación final a los clientes.

4.2.4 Relación con clientes

Con cada uno de estos departamentos tendremos relaciones muy específicas y que nos aportaran distintos tipos de información como podemos ver a continuación.

Departamento de TI

- ❖ Conocer las diferentes fuentes de datos a utilizar para la creación de los modelos.
- ❖ Tipo de Sistemas Transaccionales utilizados para el registro de las compras.
- ❖ Tipos de software CRM utilizados para la administración de las relaciones con los clientes.

Departamento de Mercadeo

- ❖ Conocer los criterios para la creación de las diferentes campañas de mercadeo.
- ❖ Conocer los parámetros considerados para realizar la segmentación de los clientes.

Departamento de Servicio al cliente/Ventas

- ❖ Conocer el catálogo con las distintas variedades de los productos y su precio.

4.2.5 Segmento de clientes

Al presentar una propuesta de optimización de procesos a la empresa, la segmentación del cliente se puede realizar en estos grupos fundamentales dentro de la empresa los cuales describimos a continuación:



Figura 6: Segmento de clientes

4.2.6 Canales de distribución

Nuestra propuesta de valor se distribuye/comunica utilizando tanto medios físicos como electrónicos, aprovechando cada recurso para establecer una comunicación efectiva con nuestro cliente.

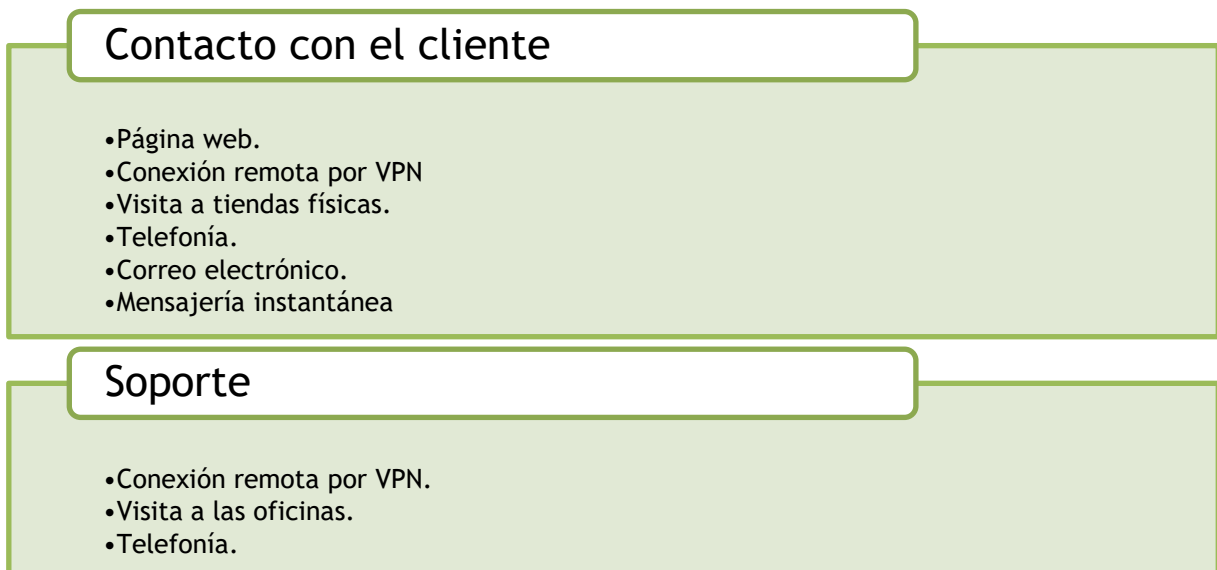


Figura 7: Canales de distribución

4.2.7 Actividades clave

Para brindar modelos con buena calidad y que aporten valor al negocio hemos identificado las siguientes actividades clave.

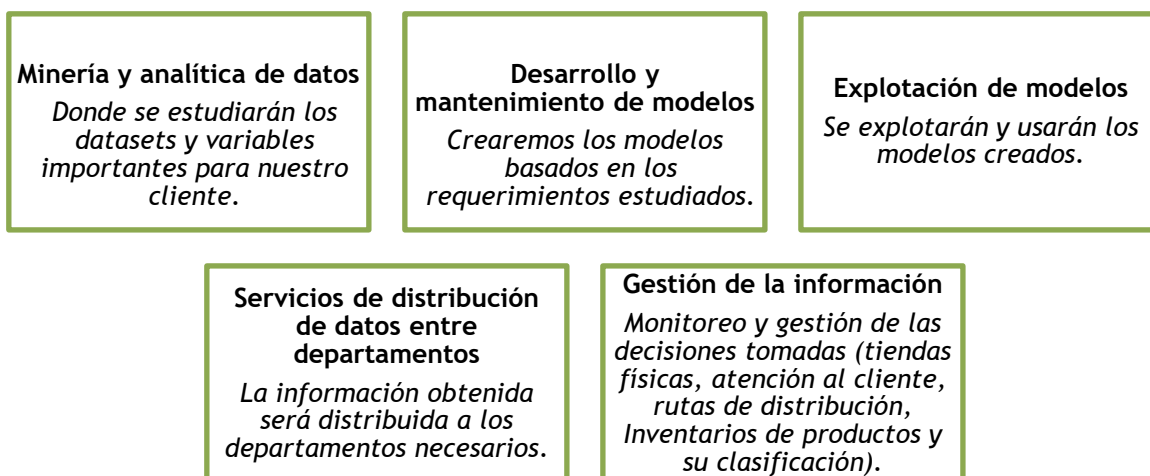


Figura 8: Actividades clave

4.2.8 Estructura de costes

Los costes de los modelos analíticos están centrados en las siguientes estructuras:

| Operacionales | Tecnología |
|---|--|
| Recursos Humanos que interactuarán en la creación de modelos (nómina, cargas sociales, etc.). | Infraestructura on-site, servidores y terminales necesarios para la creación de los modelos. |
| | Servicios IaaS & Comunicación. |

Tabla 1: Estructura de costes

Los costes operacionales en nuestro proyecto estarán enfocados en el personal técnico de El Catador que se encargará de implementar y mantener los modelos analíticos una vez en integrados. Los salarios mensuales del personal serán:

- ❖ Líder de proyecto: RD\$80,000 /mes
- ❖ Arquitecto de datos: RD\$60,000 /mes
- ❖ Arquitecto de procesos ETL: RD\$45,000 /mes
- ❖ Analista de requerimientos: RD\$30,000 /mes
- ❖ Desarrollador de modelos analíticos: RD\$55,000 /mes

Los costes de tecnología no serán tan altos ya que aprovecharemos la infraestructura actual de la empresa y agregaremos un servidor para nuestros modelos y el servicio Amazon Web Services para backup.

- ❖ Servidor: HP Proliant DL380 Gen 10: RD\$335,800.11
- ❖ Mantenimiento servidores: RD\$150,000 /año
- ❖ Amazon Web Services: RD\$30,000 /mes

4.2.9 Fuentes de ingresos

Los ingresos y beneficios de nuestro proyecto y el uso de los modelos analíticos se obtendrán a través de:

- ❖ Implementación de segmentación de los clientes corporativos e individuales.
- ❖ Implementación de estrategias de mercadeo, promoción y ventas por segmentos.
- ❖ Introducción de nuevos productos por segmentos.
- ❖ Optimización del proceso de logística e inventario.
- ❖ Mejora de la experiencia del cliente, lo cual aumenta las ventas.
- ❖ Optimización de costos por eficientización de los procesos comerciales.

Los ingresos actuales de El Catador en los últimos 3 años (2017,2018 y lo que va de 2019) por segmento de clientes, nos muestran un aumento en la mayoría de los segmentos entre 2017 y 2018. Luego de implementar los modelos esperamos que en el 2019 y 2010 el aumento sea aún mayor.

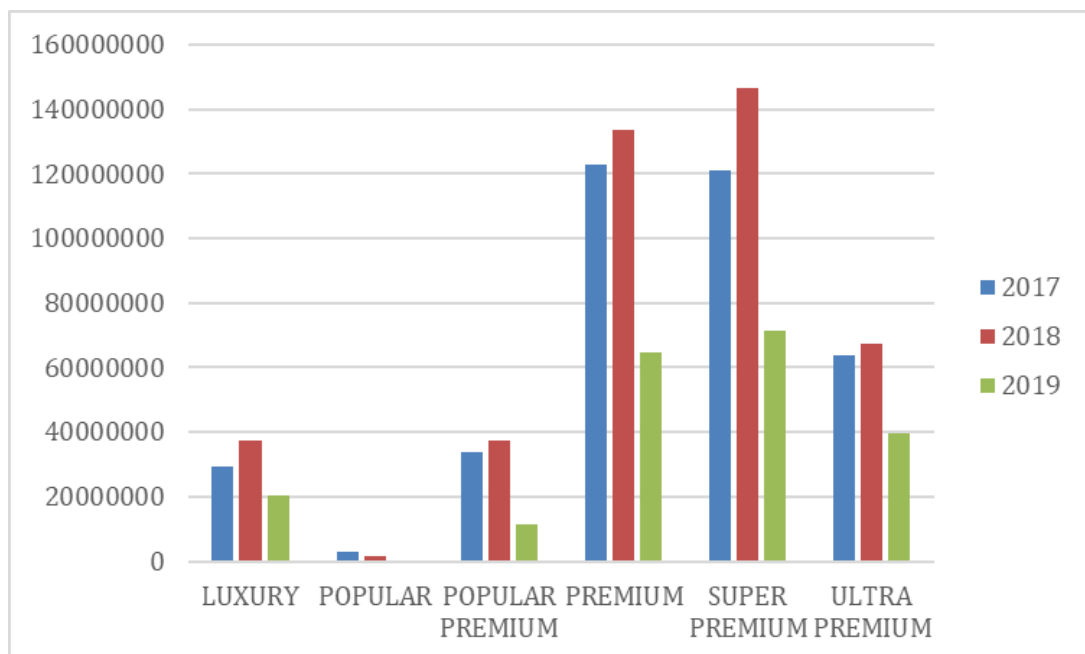


Figura 9: Ingresos por ventas de segmento de clientes anual

En una vista más detallada mes a mes, vemos que las actividades de fin de año son la parte primordial del enfoque de ventas de la empresa. Ya que si vemos los ingresos por mes notamos que diciembre es el mes en el cual todos los segmentos aumentan las ventas significativamente. Esto quiere decir que, al implementar los modelos analíticos, este mes es la fecha clave de poder aprovechar los resultados de las recomendaciones para aumentar aún más los ingresos.

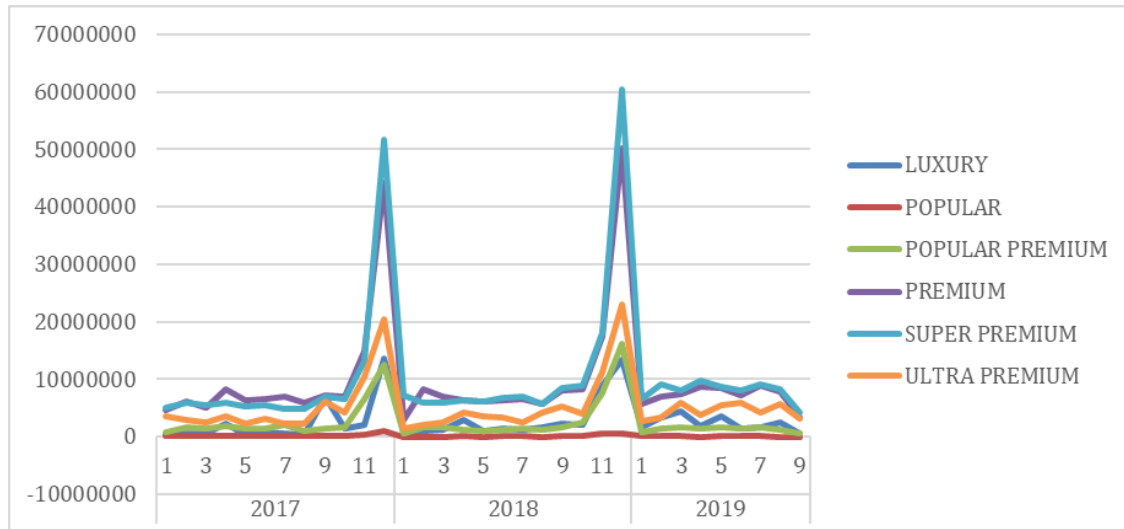


Figura 10: Ingreso por ventas de segmento de clientes mensual

4.3 Plan de acción

A continuación, describiremos los puntos en los que detallaremos el plan de acción para el desarrollo de nuestra propuesta de negocio. Con estos buscamos describir la necesidad e importancia del proyecto.

En primer lugar, necesitaremos identificar los datos de entrada que nos servirán para alimentar nuestro modelo analítico, luego de este procesamiento los resultados serán mostrados en los correspondientes cuadros de mando usando herramientas de visualización y que el objetivo de esto sea facilitar a la alta gerencia la toma de decisiones.

4.3.1 Definición del alcance del proyecto: objetivos y métricas

Diseñamos nuestro plan de negocio en función a las hipótesis validadas, con ellas podemos definir los objetivos que queremos conseguir y con base a estos las estrategias a implementar para alcanzarlos.

Nuestro plan de negocios se resume en su misión y visión.

Misión

Proveer de un sistema automatizado de recomendaciones de productos a los clientes que les permita a los vendedores potenciar sus ventas.

Visión

Liderar los procesos de recomendaciones, posicionando nuevos productos en el gusto de los clientes e incentivando la rotación de inventario de las empresas.

Objetivo

Aplicado al caso del Catador, nuestro objetivo consiste en: con el aprovechamiento de las técnicas y tecnologías de Business Intelligence desarrollar para el área de ventas una solución que realice recomendaciones de productos de tipo: introducción, variedad y upsale en función al segmento del cliente y categoría de los productos elegido.

Esto estará conformado por los siguientes objetivos específicos:

- ❖ Aumento de un 10% en las ventas anuales de nuevos productos basados en recomendaciones.
- ❖ Aumento de un 10% en las ventas anuales de productos de variedad, lo que permitirá una mejor rotación de inventario.
- ❖ Aumento de un 5% en las ventas de productos de los segmentos Ultra Premium/Luxury.

4.3.1.1 Alcance del proyecto

Para la toma de decisiones como requisito necesitaremos disponer de los datos. Necesitamos recopilar información en la empresa sobre su operación, como se recogen, almacenan y organizan los datos, en función de esto para nuestra solución se estaría creando un Datamart, nuestra idea consiste en: a través de algoritmos de Machine Learning analizar los datos y suministrar 3 diferentes tipos de productos especificados por Negocios, esto con el propósito de estimular las ventas de los mismos, para este análisis se deberá determinar cuáles variables de productos del dataset de la empresa son las más relevantes que pudieran ser consideradas como fuente de entrada, de esta forma estaríamos creando valor añadido al proceso de recomendación de productos a los clientes ya que pasaría de ser un proceso empírico a uno basado en datos históricos.

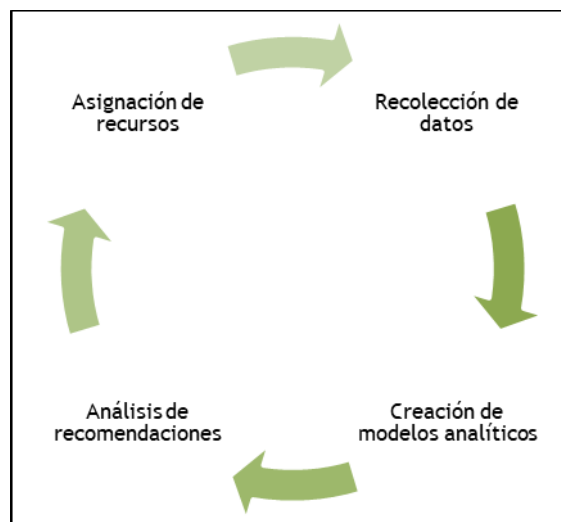


Figure 11 : Síntesis del proceso

4.3.1.2 Métricas

Acompañando el alcance del proyecto, en el siguiente acápite definiremos indicadores claves de desempeño que nos ayudaran a evaluar si los objetivos declarados están siendo alcanzados. De este modo cuantificaremos los objetivos y se reflejara el rendimiento esperado para la toma de decisiones en búsqueda de influir positivamente en el plan estratégico de la empresa.

Dicho esto, enumeramos los siguientes KPI que pueden servirnos de guía y termómetro a nuestro plan de negocio.

- ❖ **Total de ventas por mes**
 - # Histórico de ventas por mes
- ❖ **Porcentaje de Ventas de Nuevos productos basados en recomendaciones**
 - $\text{Monto de ventas de nuevos productos} / \# \text{Monto total venta de productos} \times 100$

- ❖ **Porcentaje de Ventas de productos de variedad**
 - $\text{Monto de ventas de productos de variedad} / \text{Monto total venta de productos} \times 100$
- ❖ **Porcentaje de ventas de productos segmentos superior**
 - $\text{Monto de ventas de productos de segmento Ultra Premium/Luxury} / \text{Monto total venta de productos} \times 100$

A continuación, detallamos cada KPI para entender mejor el comportamiento del negocio.

| Objetivo | Estrategia | Indicador | Forma de cálculo | Frecuencia |
|--|---|---|--|-----------------|
| Aumento de un 10% en las ventas anuales de nuevos productos basados en recomendaciones | Acuerdos con proveedores | Porcentaje de Ventas de Nuevos productos basados en recomendaciones | Monto de ventas de nuevos productos / # Monto total venta de productos x 100 | Mensual / Anual |
| | Reducción de costes de colocación de nuevos productos | | | |
| | Coordinar y ofrecer periódicamente cata de vinos a clientes potenciales | | | |
| Aumento de un 10% en las ventas anuales de productos de variedad | Diseñar un plan de marketing para promocionar los productos en inventario | Porcentaje de Ventas de Productos de variedad | Monto de ventas de productos de variedad / Monto total venta de productos x 100 | Mensual / Anual |
| | Planes de incentivos por ventas de productos | | | |
| Aumento de un 5% en las ventas de productos de los segmentos Ultra Premium/Luxury | Planes de descuento y ofertas de temporadas | Porcentaje de ventas de productos segmentos superior | Monto de ventas de productos de segmento Ultra Premium/Luxury / Monto total venta de productos x 100 | Mensual / Anual |
| | Planes de incentivos por ventas de productos | | | |

Tabla 2: Tabla de indicadores

4.3.2 Análisis de actividades: modelo lógico - arquitectura técnica

Modelo lógico

Nuestra solución tecnológica se plantea como un nuevo proceso agregado a la infraestructura IT de la empresa El Catador, el cual agregará valor dando recomendaciones de productos basadas en datos existentes e históricos de la empresa. Estas recomendaciones podrán ser vistas por distintos departamentos usando las herramientas actuales de visualización disponibles en la empresa.

De forma general, nuestro sistema estará compuesto por 4 bloques principales:

- ❖ Adquisición de los datos
- ❖ Procesamiento y análisis de los datos
- ❖ Creación de modelos analíticos para recomendaciones
- ❖ Visualización de las recomendaciones

Cada uno de estos bloques es parte de un proceso continuo de mejora y entrenamiento de los modelos analíticos y son extremadamente dependientes uno del otro.

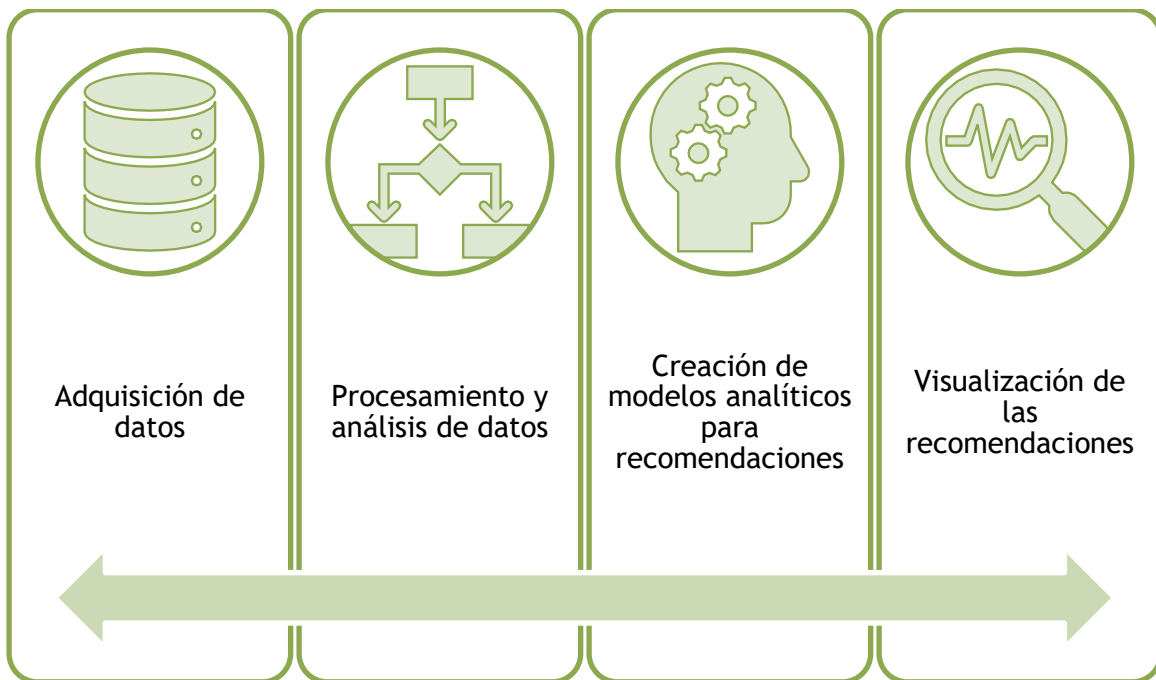


Figura 12: Modelo lógico de la solución propuesta (1)

En un diagrama más detallado podemos ver a continuación como están conformados cada uno de estos bloques principales.

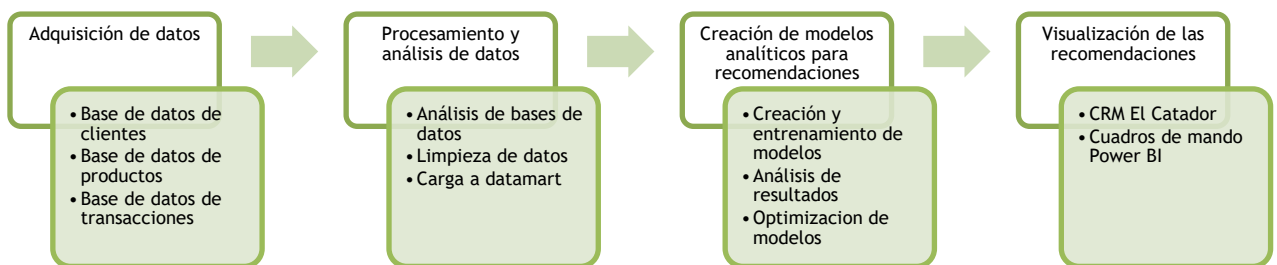


Figura 13: Modelo lógico de la solución propuesta (2)

4.3.2.1 Identificación de los datos

Teniendo en cuenta que uno de los objetivos que tenemos es proveer al cliente con un sistema de recomendaciones que les permitan ofrecer a sus clientes productos basados en su perfil de compra y sabiendo que los sistemas de recomendación han tenido un importante realce en los últimos años sobre todo en empresas de ventas en línea que los utilizan para ofrecer a sus clientes con productos diversos basados, entre otros casos, en sus historiales de búsqueda. Estos sistemas aportan a las empresas la oportunidad de concretizar la venta de un producto determinado que maximiza su valor tanto para el cliente como para el vendedor en un determinado momento.

Estos sistemas toman la información transaccional de un cliente y la convierte en conocimiento accionable recomendando un determinado producto que sea interesante tanto al cliente como para la empresa en cada transacción que realice el primero.

En el caso de la empresa El Catador poseen un sistema CRM en el cual los vendedores pueden visualizar, a partir del histórico de compras de un cliente, un resumen que les permite realizar un análisis de los posibles gustos del cliente. Este análisis es hecho por el vendedor al momento de realizar el acercamiento al cliente para proceder con la facturación de la compra actual.

Se encuentran trabajando actualmente en una aplicación móvil con la finalidad de poder tener acceso a este histórico antes de que el cliente complete una compra, pero estará accediendo a la misma información contenida en las bases de datos del CRM de la empresa. Como se indica en la sección 3.1.3 del presente documento:

Los clientes se encuentran catalogados en este CRM en 2 grandes grupos:

- ❖ Grupo 1
 - Collector
 - Aficionado
 - Corporativo
- ❖ Grupo 2
 - Entusiasta
 - Casual
 - Empresa

Los productos se categorizan, de acuerdo con su precio de venta, en:

- ❖ Luxury
- ❖ Ultra Premium
- ❖ Super Premium
- ❖ Premium
- ❖ Popular Premium
- ❖ Popular

Manejan una base de datos de productos - clientes - transacciones con aproximadamente 17 años de transacciones almacenados en una base de datos Microsoft SQL Server, la misma será la fuente de información para el desarrollo del proyecto.

4.3.2.2 Captura de datos

Esta etapa nos permitirá traer desde las bases de datos del sistema transaccional del cliente a nuestra área intermedia en donde procederemos a realizar las labores de limpieza y normalización de la data previo a su carga al área de almacenamiento final en el datamart desde donde la información servirá de insumo para las siguientes etapas de modelado y visualización.

Este proceso puede ser realizado utilizando las herramientas de integración que ofrece Microsoft: Microsoft SQL Server (MSSQL), que puede fungir como repositorio para las áreas de almacenamiento intermedio, así como para la base de datos que servirá para almacenar la información que pase al datamart.

Microsoft SQL Server Integration Services (SSIS), herramienta que permite la creación de soluciones para la integración y transformación de datos. La misma permite solucionar problemas de negocios complejos como cargar datos a datamarts, limpiar y/o minar datos, así como administrar objetos de SQL Server y datos.

El flujo del proceso de captura de datos es como se muestra en el siguiente gráfico:

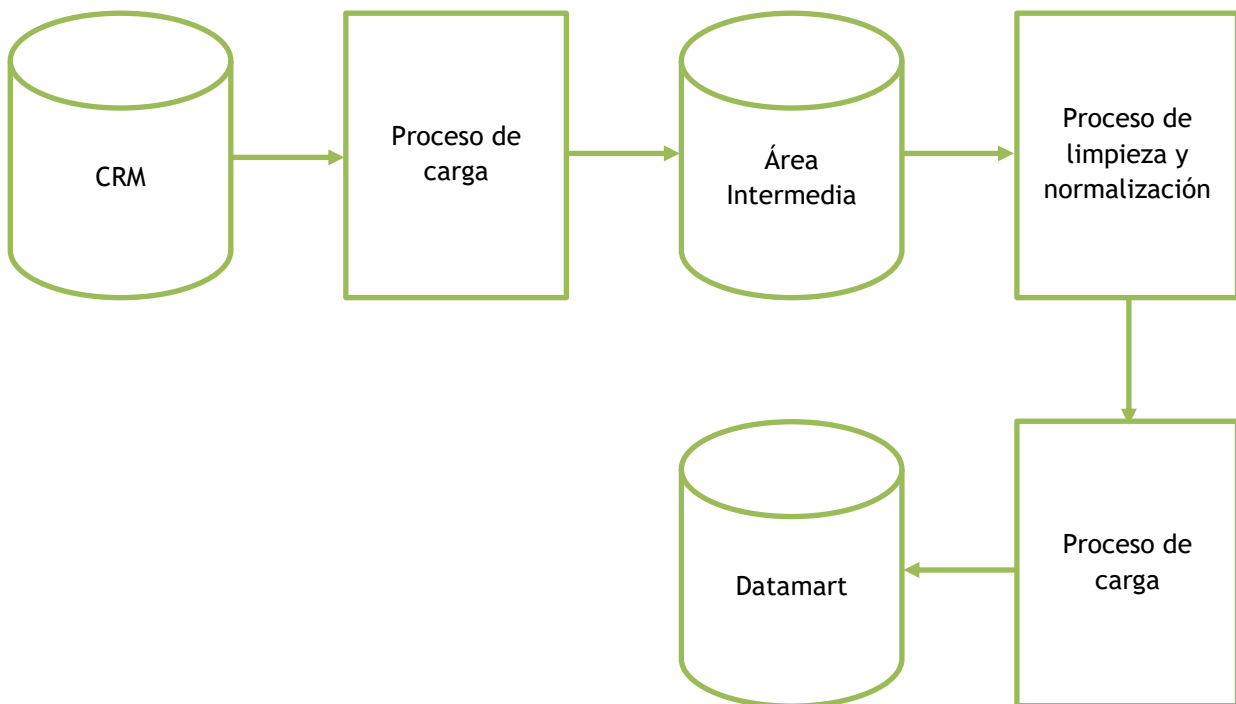


Figura 14: Flujo de proceso de captura de datos

4.3.2.3 Almacenamiento de los datos

Como pudimos ver en el gráfico anterior luego de los datos ser cargados y procesados desde el origen son almacenados en una base de datos desde la cual serán accedidos para su procesamiento final y modelación. La estructura de este almacenamiento es como se muestra en el siguiente diagrama:

DATAMART



Figura 15: Estructura de almacenamiento de datos

4.3.2.4 Transformación y validación de los datos

El negocio ha proporcionado un dataset de una gran envergadura, el cual contiene sub-datasets de productos, clientes y transacciones, como paso previo para la correcta creación de modelos analíticos es necesario analizar qué datos son relevantes y que procesos de depuración deben de realizarse para transformar y enriquecer los datos.

A continuación, se explora cada dataset de forma independiente para exponer la selección de datos y las transformaciones que se requieren:

❖ Dataset de productos

Este dataset contiene el catálogo de productos del negocio. El primer paso es la selección de sólo los vinos para lo cual se emplea la variable “TIPO_BEBIDA” con valor “Vino”.

El siguiente paso se centra en la depuración de las variables que solo son necesarias para el desarrollo del modelo de recomendaciones, dichos datos son:

- ID_ARTICULO: identificador único de producto.
- POPULARIDAD: categoría de negocio de los productos basadas en su precio de venta.
- CODIGO_ABC: identificador literal de la calidad del producto.
- PAIS: país de origen del producto.
- REGION: región de origen del producto.
- UVA: nombre de la uva empleada en la elaboración del producto.
- ID_CASA: nombre de la casa productora.

❖ Dataset de clientes

Este dataset contiene toda la información que el negocio genera de un cliente, es un dataset muy rico en información, pero para el caso que nos atañe nos centramos en la selección de las variables con mayor relevancia para la construcción del sistema de recomendaciones. Estas variables se centran en la identificación de los gustos de compra de los clientes, y son:

- ID_CLIENTE: identificador único de cliente.
- Casa1: principal casa productora de vinos que el cliente frecuenta comprar.
- Catas2: segunda casa productora de vinos que el cliente frecuenta comprar.
- Pais1: principal país de procedencia del vino que el cliente frecuenta comprar.
- Pais2: segundo país de procedencia del vino que el cliente frecuenta comprar.
- Uva1: principal uva del agrado del cliente.
- Uva2: segunda uva del agrado del cliente.

❖ Dataset de transacciones

Este dataset se compone de todas las transacciones de ventas que se producen en el negocio, contiene datos de facturación y devolución de productos. Como primer paso se realiza una selección de las facturas empleando la variable “TIPO_DOCUMENTO” con valor “FACTURA”.

El siguiente paso se centra en la depuración de las variables que solo son necesarias para el desarrollo del modelo de recomendaciones, dichos datos son:

- NUMERO_DOCUMENTO: identificador único de la transacción
- ID_CLIENTE: identificador de cliente
- FECHA: fecha en que se efectuó la transacción
- ID_ARTICULO: identificador de producto
- CANTIDAD: cantidad de productos comprados
- IMPORTE_VENTA: precio total de venta

Se identificó que el precio unitario del producto sería una variable de gran relevancia para la construcción del modelo, por lo cual se procede a su generación empleando el “IMPORTE_VENTA” y la “CANTIDAD” para efectuar una división entre ellas y sacar el resultado que se ha de guardar en la variable “PRECIO_UND”. Por otro lado, la variable de “FECHA” requirió de una transformación de su tipo de dato, de cadena de caracteres a fecha.

❖ Enriquecimiento de los datasets

Variabes puntuales de cada dataset son empleados para enriquecer los datos de cada uno, como es el caso del dataset de transacciones, la variable calculada “PRECIO_UND” se integra al dataset de productos, para de esta manera disponer del precio unitario de cada producto.

Adicionalmente, se crea un dataset de clientes-productos que contiene los identificadores de clientes y productos para realizar una relación entre ambos datasets. Su función es identificar por cada cliente, el producto que este ha comprado con mayor frecuencia en el último año.

4.3.2.5 Descubrimiento y modelado

En el apartado anterior se realizó una descripción de los datasets que se emplean en la construcción del modelo, y se realiza un descubrimiento de las variables con mayor relevancia para la analítica que se desea conseguir con el modelo.

El modelo propuesto se basa en el clusterizado del dataset de productos con la finalidad de enriquecer el dataset con el valor de clúster de cada producto y posteriormente realizar búsquedas avanzadas que servirán de base para el sistema de recomendaciones.

El primer paso que se llevó a cabo fue identificar el número efectivo de clústeres que debían componer el modelo, para lo cual se empleó la Técnica del Codo, un método de evaluación que hace uso de la sumatoria de distancias cuadráticas de los elementos de un clúster para identificar que tan bien agrupados están.

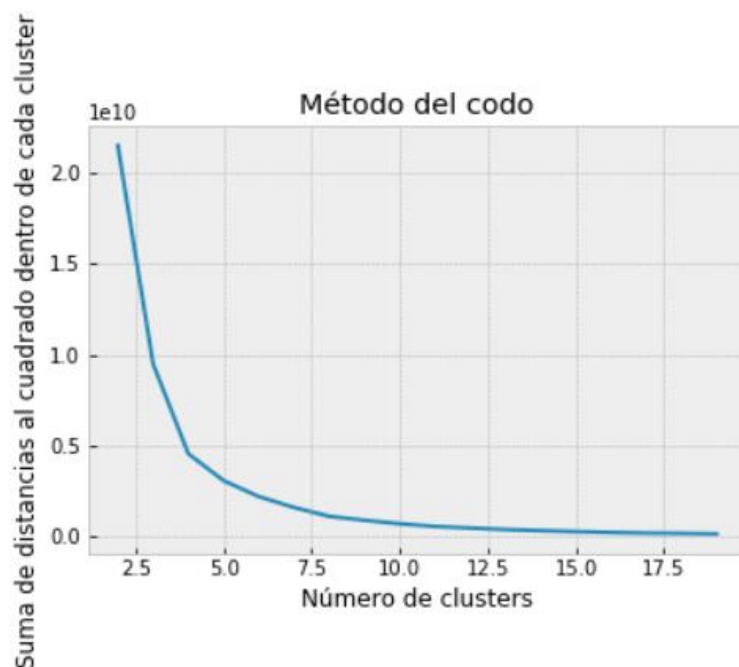


Figura 16: Gráfica con resultados del método del codo para identificación de clústeres

Dado los resultados se optó por la creación del modelo en base a 8 clústeres, debido a que es el número intermedio en el cuadrante 7.5-10 que es donde se identifica la normalización del comportamiento de la tendencia.

❖ Modelo de clusterizado

Para la construcción del modelo se decidió emplear el algoritmo “k-means” por ser uno de los más empleados para resolver este tipo de problemas y ser uno de los más efectivos. Y se utilizó el dataset de productos para su entrenamiento, con la salvedad que las variables categóricas del dataset fueron transformadas a un equivalente numérico, debido a que los modelos de aprendizaje automática son más efectivos si se emplean variables numéricas en vez de categóricas.

Una vez obtenido un modelo entrenado se procede a la predicción de los clústeres, empleando el mismo dataset de productos como entrada del modelo se consigue que para cada registro del dataset se obtenga el número de clúster al cual pertenece.

❖ Resultados del clusterizado

Se obtuvo una lectura del rendimiento del modelo de un 79% de efectividad en la predicción del clúster de nuevos productos. Y como se evidencia en la imagen siguiente, la agrupación de los productos se consigue satisfactoriamente. Como se aprecia los productos cambian de color atendiendo al clúster al que pertenecen y se evidencia una clara separación entre los mismos.



Figura 17: Representación gráfica del proceso de clusterizado de los productos

❖ Sistema de recomendaciones

El objetivo del sistema de recomendaciones es obtener por cliente aquellos productos que mejor se ajusten a sus gustos en tres vertientes, la primera, productos del segmento que frecuenta comprar el cliente, la segunda, productos de un segmento superior al que acostumbra el cliente, y la tercera, productos nuevos que están siendo introducidos a la tienda.

Para conseguir este resultado, se emplea el dataset clientes-productos para identificar el producto cabecera de cada cliente. Una vez conocido el producto, se identifica al clúster que este pertenece. Entonces se procede a extraer todos los productos que están dentro de ese mismo clúster.

Obtenidos un grupo de productos relacionados al producto cabecera del cliente se procede a emplear una clasificación propia del negocio albergada en la variable “POPULARIDAD” del dataset de productos y que denota que tan exclusivo es el producto atendiendo a su precio. De esta manera, conociendo el segmento del producto cabecera del cliente, se procede a buscar un producto que comparta el mismo valor de popularidad y otro cuya popularidad sea inmediatamente superior a este.

❖ Variaciones

Las recomendaciones se pueden realizar en dos grados, atendiendo a la necesidad del negocio:

- Recomendaciones no ajustadas

Este tipo de recomendaciones emplean el flujo de comportamiento descrito hasta el momento, es decir, que las recomendaciones se basan en el clúster y popularidad del producto cabecera del cliente.

- Recomendaciones ajustadas

Este tipo de recomendaciones son más específicas para cada cliente, pues se emplean datos adicionales del dataset de clientes que identifican gustos del cliente en cuanto al país y región de procedencia, así como de la uva empleada en la fabricación del vino. En este caso se mantiene como base la identificación de productos en base al clúster y popularidad del producto cabecera del cliente, y se agrega un filtro superior en base a los gustos del cliente.

4.3.2.6 Visualización

Para la visualización de los resultados y recomendaciones estaremos usando 2 soluciones complementarias, pero que se usarán de manera distinta. La primera es la creación de cuadros de mando en Power BI, una herramienta que nuestro cliente ya está usando, donde podemos ver las recomendaciones hechas por los modelos analíticos a cada uno de los clientes.

Esta herramienta nos permitirá ver diversas tablas y gráficas dependiendo de la necesidad del cliente y el departamento que esté interesado en ver los resultados. Podremos ver los resultados en dispositivos móviles, tabletas y ordenadores.

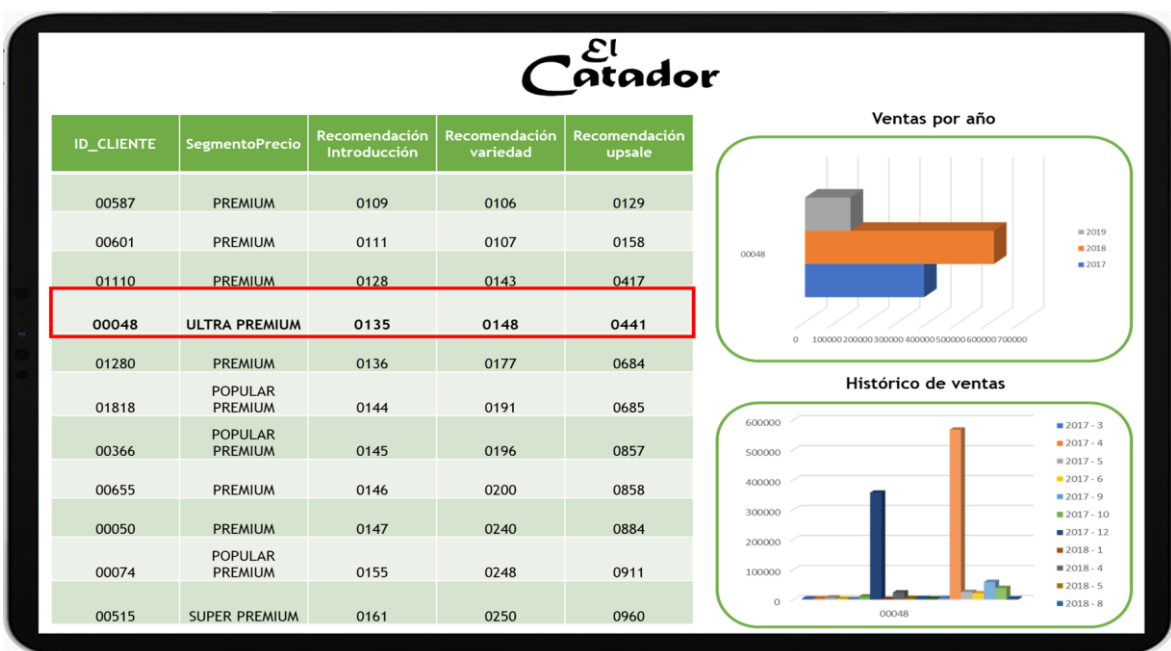


Figura 18: Cuadro de mando clientes y recomendaciones

La segunda solución de visualizar nuestros resultados es el CRM EL Catador 1.0, un programa creado por la empresa para que los vendedores puedan tener de forma fácil y rápida el historial de compra y recomendaciones hechas por los modelos analíticos en una herramienta especializada para ventas. Este programa estará cargado en tabletas en todas las tiendas a nivel nacional.

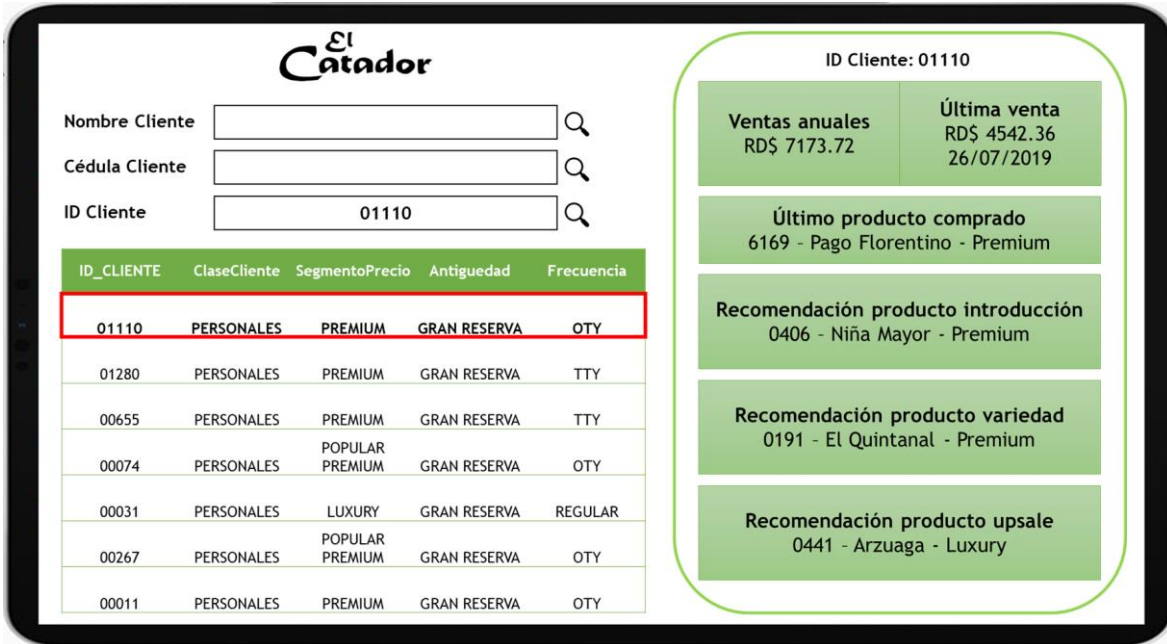


Figura 19: CRM EL Catador 1.0 con listado de recomendaciones

Solución tecnológica: arquitectura técnica

Para cumplir con nuestro objetivo para la empresa El Catador y llevar a cabo la implementación del algoritmo de agrupamiento de productos en función a las variables más representativas necesitaremos desarrollar nuestra solución contando con las siguientes herramientas tecnológicas y segmentando cada etapa de nuestra solución en los siguientes bloques:

- A. Capa de Negocio
- B. Capa de BackOffice
- C. Capa de Presentación



Figura 20: Bloques de la Inteligencia de Negocios

A. Capa de Negocio

Los datos estructurados provistos de fuentes internas como lo son del CRM El Catador v1.0 y ERP Microsoft Dynamics GP son almacenados en una base de datos Microsoft SQL, para los fines de nuestra solución, las tablas a considerar son: Clientes, Artículos y Transacciones.

B. Capa de BackOffice

Identificación, tratamiento y almacenamiento de los datos

Para el proceso de Extracción, Transformación y Carga de los datos (Extract, Transformation and Load o ETL). Una vez identificados los datos, programaremos nuestro sistema para que vaya a las fuentes de datos, extraiga la información, realice las validaciones que consideremos adecuadas, transforme aquellos datos se requieran y los cargue en el repositorio que prepararemos para los fines; un DataMart (adaptándonos al enfoque de arquitectura propuesto por Ralph Kimball y que consiste en la necesidad de establecer un sistema de almacenamiento de datos integrado y estable que garantice la explotación de la información, dando respuesta a todas las preguntas de negocio que surjan) dedicado a nuestra solución, que nos será útil para el análisis.

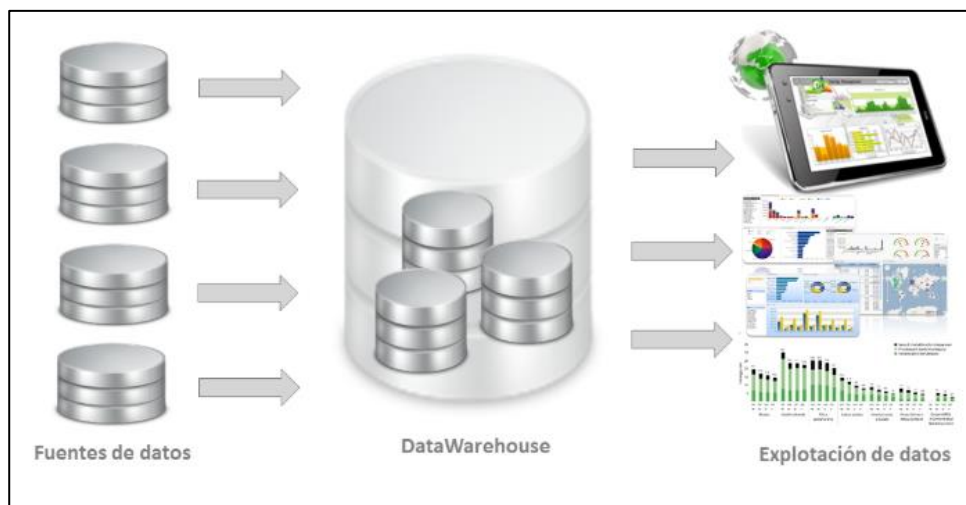


Figura 21: Arquitectura de Kimball

Aprovechando la infraestructura actual que posee nuestro cliente para la configuración y ejecución de los procesos ETLs haremos uso de Microsoft SQL Server Integration Services (SSIS), esta tecnología nos permitirá mover los datos desde el origen hacia el DataMart destino sin modificar los datos del origen, siendo poco invasiva para afectar los menos posible la Base de Datos de Producción y permitiendo hacer iteraciones y cambios de información de la data.

Procesamiento y análisis de los datos

Una vez ubicados los datos dentro de nuestro DataMart, pasaremos a la fase de procesamiento, nuestro enfoque en este punto consiste en utilizar la Minería de datos, que no es más que el conjunto de técnicas que consisten en la creación de programas que a partir de una información o conjunto de datos mediante la aplicación de un algoritmo obtenga una conclusión. En la actualidad esta disciplina se engloba dentro de lo que se denomina Data Science y que incluye Machine Learning para realizar análisis descriptivo y predictivo de forma automática.

Considerando el conocimiento previo y experiencia de los integrantes del grupo de trabajo, además la facilidad con la que se integra con el análisis de datos y el mundo de Business Intelligence el lenguaje de programación que se ha utilizado fundamentalmente es Python por lo que la solución brindada estaría siendo desarrollada bajo el mismo.

Haremos uso de Modelos Analíticos con el algoritmo K-Means de aprendizaje no supervisado utilizando la librería scikit.learn, el mismo puede ser empleado en varias áreas de utilidad como lo son: segmentación de mercados, visión por computadoras, geo-estadística, astronomía y minería de datos en agricultura.

Una vez el modelo este evaluado como bueno y valido, pasaremos a Producción, utilizándolo para la necesidad del negocio, en este caso sería la clasificación de vinos en función a las características proporcionadas.

Decidimos enfocarnos en elegir el aprendizaje no supervisado basándonos en la necesidad del negocio, pues consiste en: a partir de los datos proporcionados buscar grupos de clientes y vinos homogéneos a los cuales se le pueda asignar vinos personalizados en función a sus características, esto será posible mediante la técnica de agrupamiento K-Means.

C. Capa de Presentación

Tomando en cuenta el resultado de los modelos analíticos de Python y en combinación con el módulo de visualización de Power BI daremos visibilidad a los resultados del modelo, dichos resultados consistirán en recomendaciones de:

- ❖ Producto de introducción (producto nuevo orientado a su perfil)
- ❖ Producto de variedad (producto similar del catálogo de productos)
- ❖ Producto upsale (próximo producto de la gama Ultra Premium/Luxury)

A requerimiento del cliente esta información estaría siendo mostrada en una pantalla del CRM El Catador v1.0 para que pueda ser utilizada por la fuerza de ventas que se encuentra tanto en las diferentes sucursales como en los departamentos. Del mismo modo para dar seguimiento al cumplimiento de los objetivos y KPIs se estarán diseñando cuadros de mando en Power BI para que a la alta gerencia se le facilite la toma de decisiones y pueda realizar las optimizaciones que considere en la gestión de ventas.

Esta integración con Power BI permitirá crear, compartir y explorar diferentes cuadros de mando desde navegadores web, tabletas y smartphones. La razón de escoger Power BI para la capa de presentación recae principalmente en que nuestro cliente posee licenciamiento con Microsoft y cuenta con la herramienta, adicionalmente a esto según el estudio “Cuadrante Mágico de Gartner para herramientas de Business Intelligence” se establece a Microsoft PowerBI, Qlik Sense y Tableau como herramientas líderes en el mercado.

Dentro de las características tomadas en consideración en el estudio para dar como líder a Power BI están:

- ❖ Facilidad de uso e interfaz atractiva
- ❖ Costes muy competitivos de las licencias en sus diferentes versiones
- ❖ Visión global de producto (Product vision)
- ❖ Experiencia del cliente (Customer Experience)

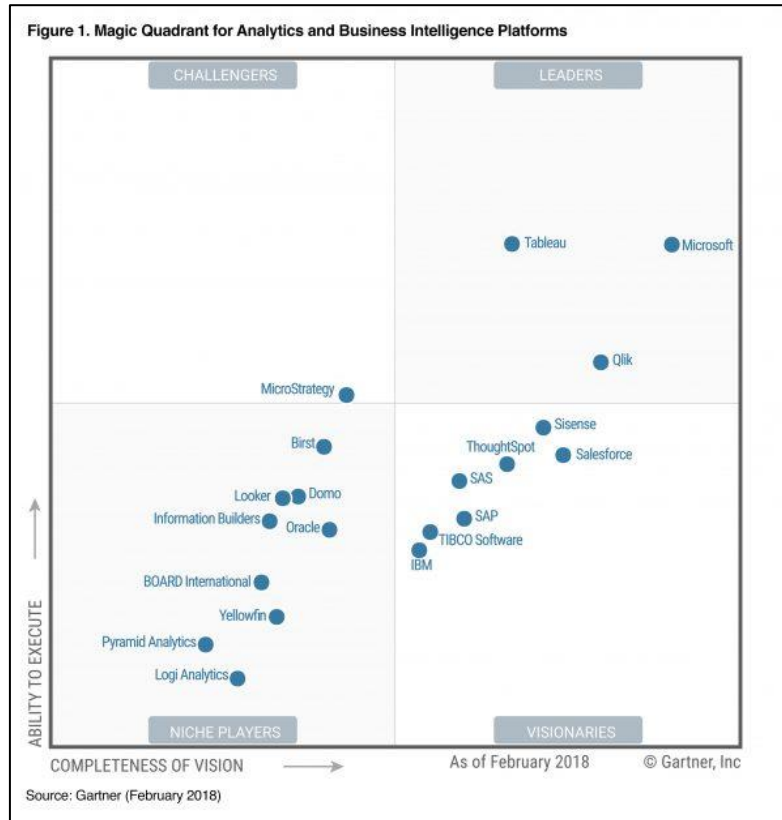


Figura 22: Cuadrante Mágico de Gartner en 2018 para herramientas Business Intelligence y Business Analytics

Teniendo una visión global de todos los procesos a ejecutar en las diferentes fases del proyecto para cumplir con el objetivo de la solución propuesta, nuestra arquitectura técnica quedará de esta forma:

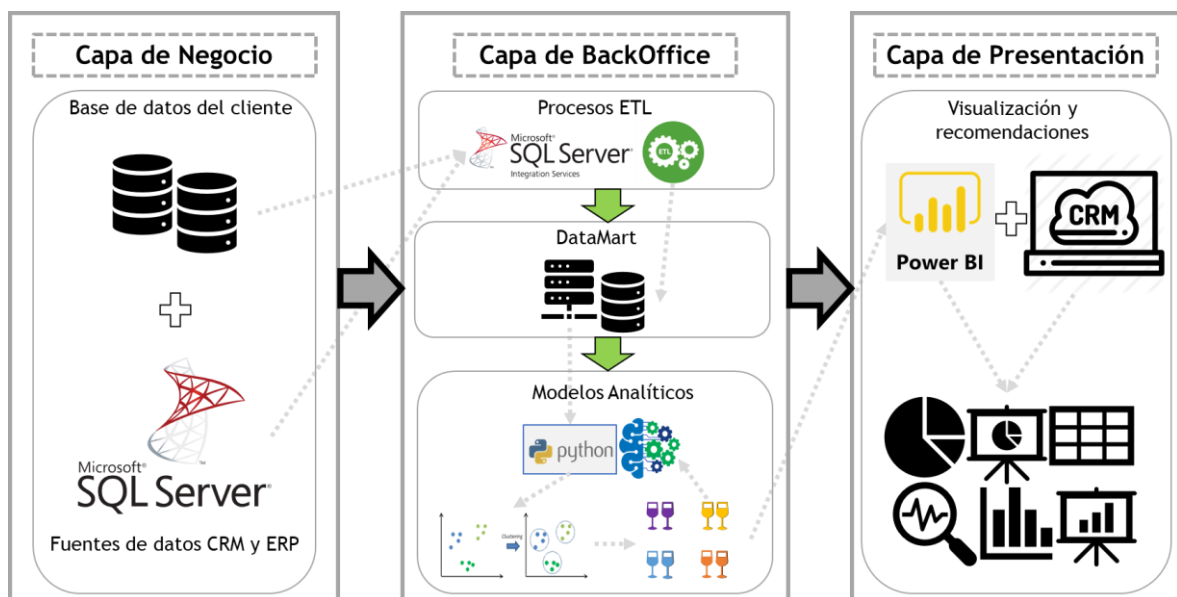


Figura 23: Arquitectura técnica de la solución propuesta

4.3.3 Análisis de recursos: talento humano y recursos físicos

El presente proyecto parte de la base de una implementación de Business Analytics para un cliente en concreto (El Catador), esto supone para los actores principales del proyecto que los esfuerzos realizados irán siempre encaminados a lograr los objetivos propuestos por el cliente (El Catador).

La correcta ejecución del proyecto y consiguiente alcance de resultados supone que en el equipo se involucre tanto personal de negocio y técnico del cliente, así como el personal técnico que se describe a continuación.

4.3.3.1 Análisis del talento humano

De parte de El Catador los actores claves que deben estar presentes son:

- ❖ El Director(a) de Ventas
- ❖ El Director(a) de Marketing
- ❖ El Director(a) de Servicio al Cliente
- ❖ Miembros del Equipo de IT (administradores de bases de datos, administradores de CRM)
- ❖ Miembros del Equipo de Análisis de Ventas.

De parte del Equipo Técnico que lleva a cabo la ejecución de proyecto son claves los roles de:

- ❖ Líder de proyecto
- ❖ Arquitecto de datos
- ❖ Arquitecto de procesos ETL
- ❖ Analista de requerimientos
- ❖ Desarrollador de modelos analíticos

Adicionalmente es esencial que exista una persona que funja como intermediario directo entre el Cliente y el Equipo Técnico, esta persona asume el rol de Dueño de Producto y tiene la responsabilidad de comunicar las necesidades del negocio, así como brindar todo el soporte de negocio que el Equipo Técnico requiere, así como facilitar la toma de decisiones.



Figura 24: Organigrama Proyecto Modelos Analíticos El Catador

4.3.3.2 Análisis de recursos físicos

El producto final de este proyecto será instalado en los servidores del cliente por lo cual será necesario que se configure en la infraestructura TI del cliente un espacio con las siguientes especificaciones:

- ❖ **Ambiente de trabajo**
 - No se requiere el establecimiento de oficinas, el Equipo Técnico realiza su labor de manera remota empleando diversas locaciones que están a su disposición.

- ❖ **Servidor Web**
 - El sistema operativo puede ser un Windows Server o una distribución Linux, por ejemplo: Ubuntu 16.04
 - Al menos 8GB de memoria RAM.
 - El espacio en disco no es una limitante, con 100 GB SSD se asegura el espacio necesario para la instalación de los servicios del servidor.
 - Procesador con al menos 4 núcleos físicos.

- ❖ **Ancho de Banda**
 - La velocidad de conexión a la red está limitada a la disponibilidad del cliente.

- ❖ **Software**
 - Python 3, se recomienda establecer un ambiente virtual independiente del resto del sistema de los servicios externos.
 - Apache2.
 - Módulo mod_wsgi.

- ❖ **Servicios externos**
 - No se especifica la contratación de los servicios externos ya que el cliente proveerá todo el ambiente necesario para la implementación del proyecto.
 - Los servicios de consultoría están dados por el Equipo Técnico.
 - El entrenamiento y desarrollo de los modelos analíticos se realizarán utilizando el equipamiento existente del Equipo Técnico.
 - Los servicios de comunicaciones forman parte de la estructura actual del TI del cliente.

4.3.4 Gestión del tiempo (cronograma)

Para la gestión del tiempo de análisis, desarrollo, puesta en marcha y mantenimiento de este proyecto hemos establecido un período de medición en semanas de trabajo y pretendemos completar el ciclo en un espacio de 20 semanas hasta la puesta en marcha y luego un período de 3 semanas adicionales para validación y mejora del proyecto implementado.

Hemos optado por representar la planificación del proyecto utilizando un Diagrama de Gantt por ser más simple de interpretar gracias a su modelo de filas y columnas, en donde las filas representan las actividades del proyecto y las columnas los períodos asignados a cada actividad, los cuales en nuestro caso serán semanas.

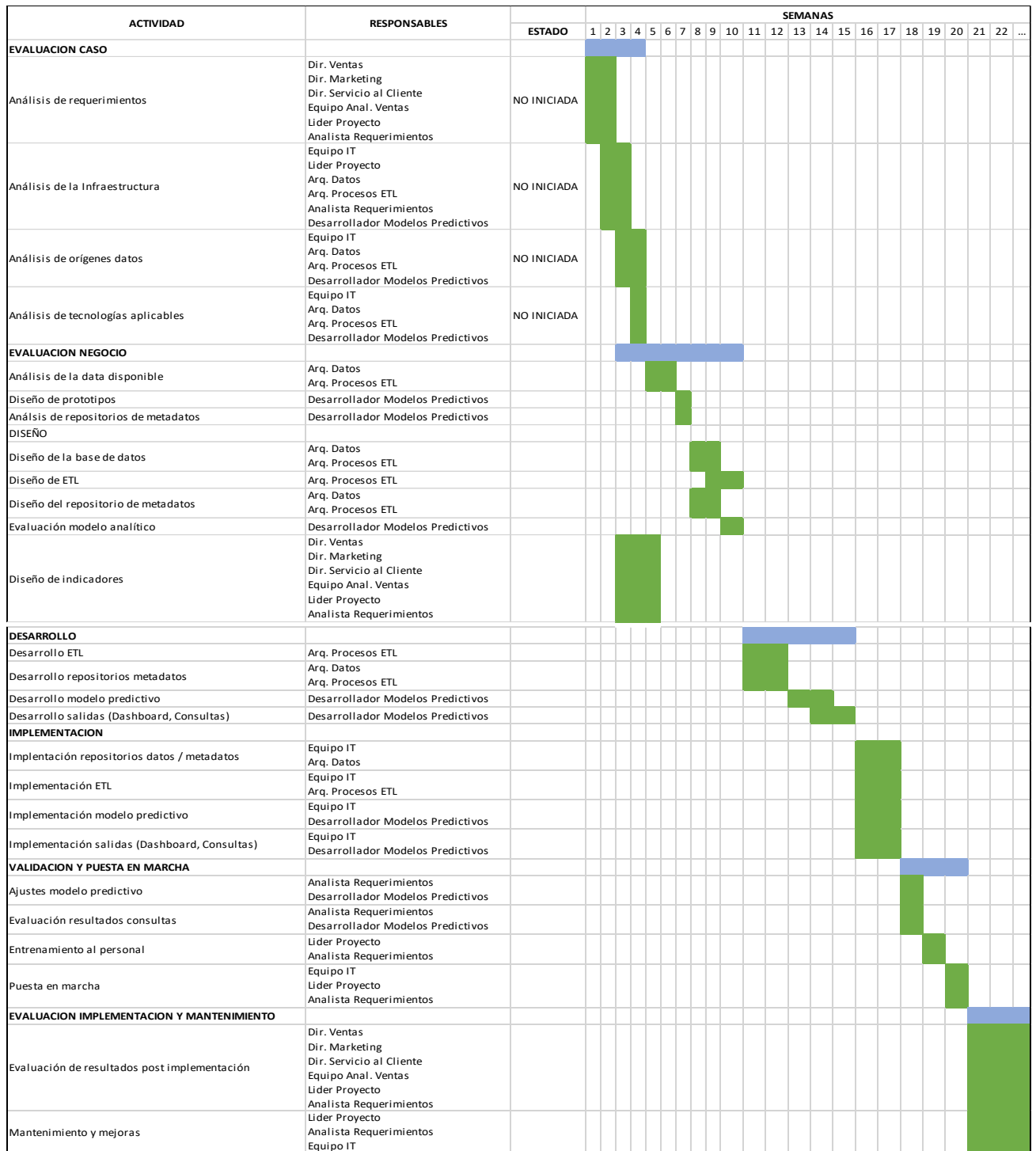


Figura 25: Cronograma estimativo del proyecto

Descripción de etapas

1. **Evaluación Caso**, en esta etapa nos proponemos realizar un levantamiento de la realidad actual de nuestro cliente con miras a poder conocer mejor su situación y el entorno en el que se desenvuelven sus actividades comerciales. Deseamos poder conocer de manera efectiva cuál es su requerimiento para poder realizar una delimitación correcta del alcance de nuestro proyecto, de esa misma forma conocer los recursos tecnológicos también de qué forma está estructurado su departamento de TI para, con esta información, poder inferir que tecnologías pueden ser aplicadas en el desarrollo del proyecto actual.
2. **Evaluación Negocio**, el objetivo en este punto es poder levantar de manera efectiva la información disponible que pueda proveernos el cliente, así como la metadata relacionada a la misma en caso de que la posea con la finalidad de poder recrear su realidad a partir de la información disponible. En esta fase también pretendemos, de mano con el cliente, realizar el diseño de todos los casos de uso y artefactos que sean requeridos para poder preparar el camino que llevará al proceso de desarrollo de la herramienta que proveerá a nuestro cliente con las herramientas de información que hayan sido identificadas como necesarias en la etapa anterior.
3. **Desarrollo**, es en esta etapa en donde se plasma en las herramientas de desarrollo de software todos los componentes que hayan sido diseñado en la etapa anterior, es ahora cuando pasamos del diseño conceptual al físico en donde crearemos las herramientas, de software, que nuestro cliente requiere para poder alcanzar los objetivos que se han propuesto en este proyecto.
4. **Implementación**, luego de haber completado la etapa de desarrollo de los componentes de software que integran el proyecto y haber corregido todos los posibles funcionamientos no esperados que comúnmente suceden al momento de la codificación de herramientas de software, se realiza el proceso de implementación de los objetos resultantes en la infraestructura del cliente. Para completar con este proceso existe más de una forma correcta para hacerlo, en el caso del proyecto actual vamos a cumplir con los requerimientos del cliente y entregaremos las herramientas finales para que el personal de infraestructura de la empresa realice la implementación interna, en caso de que se presente algún problema procederemos a corregirlo y volver a intentar la implementación.
5. **Validación y puesta en marcha**, una vez completado el proceso de implementación se procede a la validación por el cliente del funcionamiento de las herramientas que fueron puesta en funcionamiento, de forma ideal este proceso de validación sucede en un ambiente controlado que comparta las características del ambiente de producción pero que no afecte al mismo, luego de que el cliente realice todas las pruebas al software que considere que deban ser realizadas y los resultados de las mismas sean satisfactorios, el sistema es finalmente puestos en marcha y a disposición de los usuarios finales para que comiencen a beneficiarse en la operación de la empresa de las nuevas herramientas de software.
6. **Evaluación Implementación y Mantenimiento**, esta última etapa es un proceso que debe repetirse mientras las herramientas desarrolladas se encuentren en el ambiente de producción del cliente. En esta etapa las herramientas entran en un proceso de mantenimiento en el que, siempre que se detecte funcionamientos anormales, son corregidas, revalidadas y puestas en producción, los cambios realizados. De la misma

forma el cliente puede detectar la necesidad de crear nuevas funcionalidades a las herramientas ya existentes, por ejemplo, en el proyecto actual, puede ser un nuevo cuadro de mando que presente la información con otro modo de visualización u otro posible punto de vista, estas nuevas funcionalidades, luego de pasar por las etapas de diseño, desarrollo, validación son implementadas en producción para que los usuarios puedan servirse de las mismas en la operación de la organización. Este proceso solo debe terminar cuando el ciclo de vida de la aplicación termine y la misma sea desmontada de la operación del cliente.

5. Optimización de resultados

Como ya se mencionó anteriormente, nuestro proyecto estará enfocado en la empresa El Catador, una empresa ya creada y con varios años de servicio en el mercado. Por esta razón estará enfocado en optimizar la funcionalidad de uno de sus departamentos usando las herramientas de Big Data/Inteligencia de Negocio descritas a lo largo de este documento.

Continuando con esta premisa, para el desarrollo de nuestro proyecto iniciaremos en los próximos puntos el análisis de rentabilidad de este. Esperamos con este análisis confirmar que el proyecto es viable para la empresa y que aportará valor luego de ser implementado.

Para analizar esto estaremos enfocándonos en los diferentes beneficios y costes que tiene el proyecto y luego usaremos varias herramientas para medir la rentabilidad.

5.1 Beneficios tangibles

Para iniciar con el análisis desglosaremos primero los beneficios tangibles, que son los beneficios que podemos medir de forma numérica luego de implementar el proyecto. Estaremos viendo los ingresos y costes esperados para poder hacer un desarrollo correcto de nuestro proyecto.

5.1.1 Generación de ingresos

Los ingresos esperados que generará nuestro proyecto dependerán de las ventas que se produzcan por las recomendaciones hechas por nuestro programa. Los objetivos definidos para las ventas esperadas fueron:

- ❖ Aumento de un 10% en las ventas anuales de nuevos productos basados en recomendaciones.
- ❖ Aumento de un 10% en las ventas anuales de productos de variedad.
- ❖ Aumento de un 5% en las ventas de productos de los segmentos Ultra Premium/Luxury.

Antes que todo debemos conocer las ventas por segmento que tiene El Catador. Para hacer esto tomaremos la data de transacciones que fue proporcionada por la empresa y filtraremos solo las que fueron de producto tipo vino y de ahí las agruparemos por segmento. Usaremos el año 2018 como base, ya que es el último año completo de datos que tenemos.

| AÑO | LUXURY | POPULAR | POPULAR PREMIUM | PREMIUM | SUPER PREMIUM | ULTRA PREMIUM |
|------|-----------------|--------------|-----------------|-----------------|------------------|-----------------|
| 2018 | \$27,076,684.94 | \$887,664.39 | \$24,751,455.62 | \$96,212,008.43 | \$104,313,103.40 | \$53,639,630.08 |

Tabla 3: Ventas vinos por segmento 2018

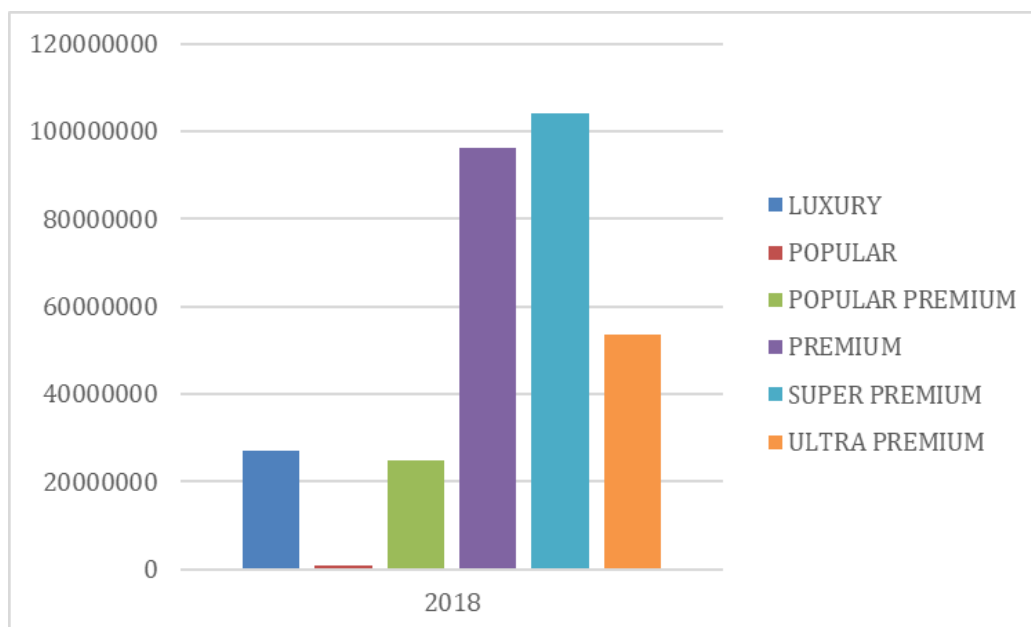


Figura 26: Ventas vinos por segmento 2018

Teniendo las ventas por segmento, nuestro próximo paso es determinar cuántas de estas fueron hechas a través de recomendaciones. Para esto asumiremos que el 2% de las ventas se basan en recomendaciones, por lo que la cantidad total de ventas por recomendaciones en el año 2018 se muestra a continuación. EL total del año 2018 es \$6,137,611.09.

| VENTAS | LUXURY | POPULAR | POPULAR PREMIUM | PREMIUM | SUPER PREMIUM | ULTRA PREMIUM |
|-----------------------------|--------------|-------------|-----------------|----------------|----------------|----------------|
| Recomendaciones 2018 | \$541,533.70 | \$17,753.29 | \$495,029.10 | \$1,924,240.00 | \$2,086,262.00 | \$1,072,793.00 |

Tabla 4: Ventas vino por recomendaciones 2018

Partiendo de estos valores ya podemos calcular ganancias esperadas de nuestro proyecto de acuerdo con los indicadores definidos previamente. El porcentaje esperado en los indicadores será dividido durante 5 años con un crecimiento esperado detallado a continuación:

| Objetivo | Indicador | Comportamiento esperado |
|--|---|-------------------------|
| Aumento de un 10% en las ventas anuales de nuevos productos basados en recomendaciones | Porcentaje de Ventas de Nuevos productos basados en recomendaciones | Aumento de 10% anual |
| Aumento de un 10% en las ventas anuales de productos de variedad | Porcentaje de Ventas de Productos de variedad | Aumento de 10% anual |
| Aumento de un 5% en las ventas de productos de los segmentos Ultra Premium/Luxury | Porcentaje de ventas de productos segmentos superior | Aumento de 5% anual |

Tabla 5: Comportamiento esperado indicadores

Ya sabiendo el comportamiento esperado de cada uno de los indicadores, las siguientes asunciones se hicieron para cada uno de ellos en el primer año.

- ❖ Aumento de un 10% en las ventas anuales de nuevos productos basados en recomendaciones: Asumimos que al introducir el algoritmo las ventas de nuevos productos el primer año serán un 10% adicional a las ventas totales por recomendaciones del 2018.
- ❖ Aumento de un 10% en las ventas anuales de productos de variedad: Asumimos que al introducir el algoritmo las ventas de productos de variedad el primer año serán un 30% adicional a las ventas totales por recomendaciones del 2018.
- ❖ Aumento de un 5% en las ventas de productos de los segmentos Ultra Premium/Luxury: Asumimos que al introducir el algoritmo las ventas de productos de un segmento superior el primer año serán un 20% adicional a las ventas totales por recomendaciones del 2018.

Con esto tenemos la siguiente tabla de ingresos adicionales que traerá el proyecto a la empresa, usando como base los \$6,137,611.09 que se vendieron basados en recomendaciones en 2018.

| Indicador | Año 1 | Año 2 | Año 3 | Año 4 | Año 5 |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Ventas de nuevos productos basados en recomendaciones | \$613,761.11 | \$675,137.22 | \$742,650.94 | \$816,916.04 | \$898,607.64 |
| Ventas de Productos de variedad | \$1,841,283.33 | \$2,025,411.66 | \$2,227,952.83 | \$2,450,748.11 | \$2,695,822.92 |
| Ventas de productos segmentos superior | \$920,641.66 | \$966,673.75 | \$1,015,007.43 | \$1,065,757.81 | \$1,119,045.70 |
| Ingresos totales | \$3,375,686.10 | \$3,667,222.63 | \$3,985,611.20 | \$4,333,421.96 | \$4,713,476.26 |

Tabla 6: Tabla de ingresos

Luego del primer año esperamos que el comportamiento siga de acuerdo con los indicadores definidos y teniendo un crecimiento constante.



Figura 27: Ingresos estimados

5.1.2 Reducción de costes

Los costes definidos de nuestro proyecto se dividen en recursos humanos y tecnología como explicamos en el apartado [4.2.8 Estructura de costes](#). En resumen, tenemos los siguientes costes:

Recursos humanos: RD\$3,240,000.00 pesos al año.

- ❖ Líder de proyecto: RD\$80,000 /mes.
- ❖ Arquitecto de datos: RD\$60,000 /mes.
- ❖ Arquitecto de procesos ETL: RD\$45,000 /mes.
- ❖ Analista de requerimientos: RD\$30,000 /mes.
- ❖ Desarrollador de modelos analíticos: RD\$55,000 /mes.

Tecnología: Inversión inicial de RD\$335,800.11 por servidor más RD\$510,000.00 al año.

- ❖ Servidor HP Proliant DL380 Gen 10: RD\$335,800.11 pesos.
- ❖ Mantenimiento servidores: RD\$150,000 pesos anual.
- ❖ Amazon Web Services: RD\$30,000 pesos mensual.

Para aprovechar los recursos e infraestructura ya existente en la empresa, la reducción de costes la haremos reasignando personal del departamento IT a este nuevo proyecto para que luego de que sea implementado su mantenimiento sea una parte más de su trabajo diario.

Por esta razón las siguientes posiciones de recursos humanos ya estarían dentro de la empresa y no habría que hacer gastos adicionales en nuevo personal:

- ❖ Líder de proyecto
- ❖ Arquitecto de datos
- ❖ Arquitecto de procesos ETL
- ❖ Analista de requerimientos

Con esta reasignación rebajamos los costes anuales de recursos humanos de RD\$3,240,000.00 a RD\$660,000.00 al año.

En relación con los gastos de tecnología la reducción de costes se realizará asignado el mantenimiento de servidores al departamento de TI, así ahorrándonos RD\$150,000 pesos anual.

Teniendo nuestra nueva estructura de costes, podemos calcular que en los 5 años del proyecto nuestros costes serán:

| Coste | Año 1 | Año 2 | Año 3 | Año 4 | Año 5 |
|---------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Servidores HP | \$671,600.22 | - | - | - | - |
| Recursos Humanos | \$660,000.00 | \$673,200.00 | \$686,664.00 | \$700,397.28 | \$714,405.23 |
| Amazon Web Services | \$360,000.00 | \$360,000.00 | \$360,000.00 | \$360,000.00 | \$360,000.00 |
| Consultoría Post-Implementación | \$200,000.00 | - | - | - | - |
| Costes totales | \$1,891,600.22 | \$1,033,200.00 | \$1,046,664.00 | \$1,060,397.28 | \$1,074,405.23 |

Tabla 7: Tabla de costes

De manera gráfica podemos ver que la mayor inversión será realizada el primer año y luego los costes se estabilizan.

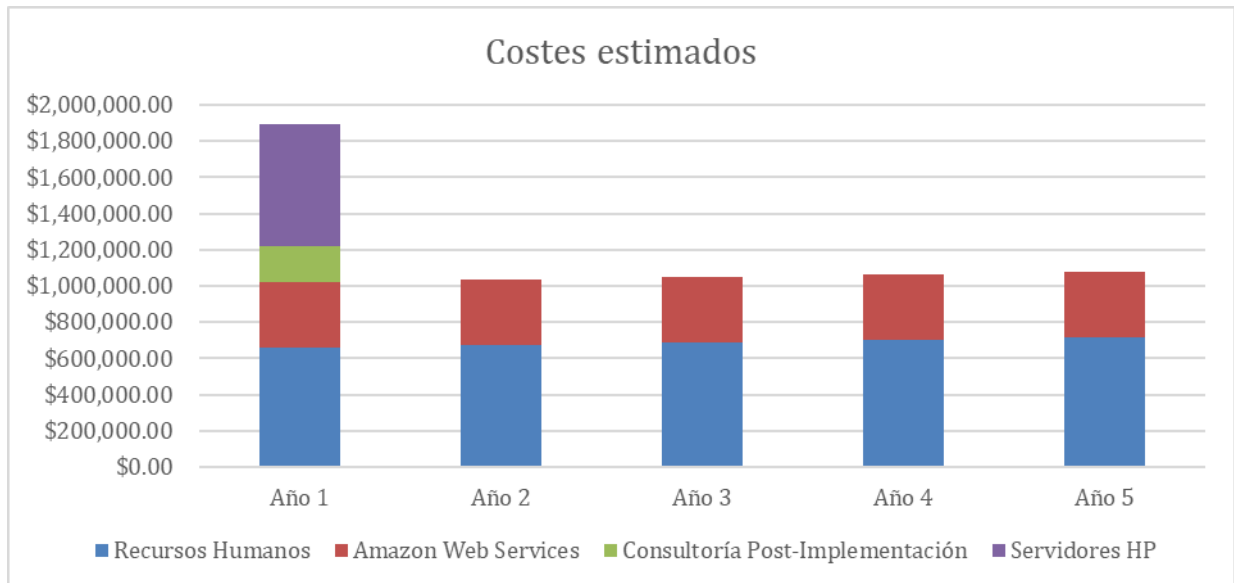


Figura 28: Costes estimados

5.2 Beneficios intangibles

Con la implantación del sistema de recomendación de productos que estamos proponiendo a la empresa El Catador, estaremos dotando a sus departamentos internos con la capacidad de tender una novedosa forma de realizar recomendaciones que pueden llevar a una mejor colocación de sus productos.

Esta mejora en la colocación de nuevos productos, a su vez, puede permitirle a la empresa realizar campañas de mercadeo efectivas basados en la información que estarán generándose a partir de las ventas que se realicen y que tendrán disponible mediante los informes que el departamento de TI les vaya poniendo a su disposición.

Desde el punto de vista que considera la experiencia de los clientes quienes, recibiendo una recomendación basada en sus compras pasadas, podrán tener la capacidad de experimentar nuevos productos, quizás desconocidos para ellos, que les podrán generar una experiencia positiva en el consumo asegurando así su fidelidad para con la empresa.

En otras empresas que han adoptado sistemas de recomendación se ha podido observar una mejora en las ventas puesto que los clientes sienten una mayor cercanía a las empresas luego de que ven que estas toman en consideración sus gustos y pasadas elecciones, los clientes ven estas recomendaciones como una mejora en la calidad de la experiencia de compra.

5.3 Beneficios estratégicos

El Catador se ha posicionado como una empresa líder en su sector, manteniendo un enfoque centrado en el cliente e introducción una gran variedad de nuevos productos al mercado dominicano. Es precisamente en ese enfoque en el cliente, que este proyecto de optimización de su plataforma es tan vital. El sistema de recomendaciones personalizado desarrollado en este proyecto ayudará al Catador a mantener, mejorar y hasta aumentar su interacción con el cliente y la posterior apreciación de este de la calidad del producto y servicio brindado.

Centramos nuestra atención en dos áreas esenciales de una estrategia efectiva de fidelización de clientes y consolidación de ventas. La primera área se refiere al aspecto operativo del personal de ventas en las tiendas y tele vendedores. La segunda área se refiere al aspecto administrativo de las ventas y marketing.

En lo referente a la primera área, el proyecto centra su atención en la mejora constante de la interacción con el cliente, ofreciendo recomendaciones cada vez más apegadas a los gustos de cada consumidor, por lo que el Catador adquiere los siguientes beneficios:

- ❖ Fidelización del cliente.
- ❖ Retroalimentación constante de la apreciación del cliente.
- ❖ Introducción no invasiva de nuevos productos en las tiendas.
- ❖ Diversificación de la cartera de clientes atendiendo a la frecuencia de compra y que tan bien recibidas son las recomendaciones que el sistema realiza.

En lo referente a la segunda área, el proyecto permite que los analistas de ventas del Catador trabajen de manera proactiva sobre la base de una cartera de clientes conocida y la apreciación de compra futura de las clientes basadas en un análisis de las recomendaciones de una porción preseleccionada de la cartera de clientes. El Catador puede entonces establecer estrategias de ventas centradas en segmentos de clientes que comparten los mismos gustos, entre los beneficios encontramos:

- ❖ Campañas de mercadeo dirigidas con un mayor nivel de efectividad de conversión a ventas.
- ❖ Optimización del trabajo de los analistas de ventas.
- ❖ Mayor certeza del momento correcto para introducir un nuevo producto al catálogo y conseguir un segmento de clientes potenciales para adquirirlo.

5.4 Análisis de rentabilidad

Dándole continuidad al desarrollo del proyecto, nos llega la etapa del análisis de la viabilidad económica del mismo, para cumplir con ello estaremos empleando técnicas y fórmulas que nos ayudarán con el análisis económico-financiero.

Mediante el cálculo del ROI (Retorno de la Inversión) pretendemos validar la rentabilidad del proyecto, validando la inversión realizada por el cliente vs los beneficios comerciales generados.

La inversión inicial corresponde a un monto total de **RD\$1,891,600.22** por nuestro producto, relativamente baja con respecto al retorno que se puede llegar a obtener con las recomendaciones personalizadas a los clientes, no solo desde un punto de vista económico, sino también creando valor al servicio que se ofrece y mejorando con esto la percepción del cliente sobre la marca.

Para realizar las proyecciones de los ingresos por concepto de ventas de productos por recomendaciones obtenidas por nuestra solución, para el primer año se emplearon las siguientes consideraciones:

- ❖ Productos nuevos: 10% sobre el monto de ventas por recomendaciones (manuales/empíricas) que fue de \$6,137,611.09.
- ❖ Productos de variedad: aumento de un 30%.
- ❖ Productos del segmento superior: aumento de un 20%

De cara a los sueldos del personal, se aumentó un 2% de manera anual.

ANÁLISIS RENTABILIDAD DEL PROYECTO

| | AÑO 0 | AÑO 1 | AÑO 2 | AÑO 3 | AÑO 4 | AÑO 5 |
|---|-----------------------|-------------|-------------|-------------|-------------|-------|
| INVERSIÓN | | | | | | |
| 2 servidores HP (Proliant DL380 Gen 10) | 671,600.22 \$ | | | | | |
| Amazon Web Services | 360,000.00 \$ | | | | | |
| Consultoría Post Implementación | 200,000.00 \$ | | | | | |
| Total inversiones | 1,231,600.22\$ | - \$ | - \$ | - \$ | - \$ | |

| | INGRESOS/BENEFICIOS | | | | | |
|---------------------------------------|----------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Ventas de nuevos productos | | 613,761.11 \$ | 675,137.22 \$ | 742,650.94 \$ | 816,916.04 \$ | 898,607.64 \$ |
| Ventas de productos variedad | | 1,841,283.33 \$ | 2,025,411.66 \$ | 2,227,952.83 \$ | 2,450,748.11 \$ | 2,695,822.92 \$ |
| Ventas de productos segmento superior | | 920,641.66 \$ | 966,673.75 \$ | 1,015,007.43 \$ | 1,065,757.81 \$ | 1,119,945.70 \$ |
| Total ingresos/beneficios | - \$ | 3,375,686.10 \$ | 3,667,222.63 \$ | 3,985,611.20 \$ | 4,333,421.96 \$ | 4,714,376.26 \$ |

| | GASTOS | | | | | |
|-----------------------------|----------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Sueldos del personal (RRHH) | 660,000.00 \$ | 673,200.00 \$ | 686,664.00 \$ | 700,397.28 \$ | 714,405.23 \$ | 714,405.23 \$ |
| Amazon Web Services | | 360,000.00 \$ | 360,000.00 \$ | 360,000.00 \$ | 360,000.00 \$ | 360,000.00 \$ |
| Total gastos | 660,000.00 \$ | 1,033,200.00 \$ | 1,046,664.00 \$ | 1,060,397.28 \$ | 1,074,405.23 \$ | 1,074,405.23 \$ |

| | | | | | | |
|--------------------------------|-------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| Flujo de caja operativo | -1,891,600.22 \$ | 2,342,486.10 \$ | 2,620,558.63 \$ | 2,925,213.92 \$ | 3,259,016.73 \$ | 3,639,971.03 \$ |
| Valor actual | -1,891,600.22 \$ | 2,129,532.82 \$ | 2,165,750.93 \$ | 2,197,756.51 \$ | 2,225,952.28 \$ | 2,260,135.63 \$ |
| Acumulado | -1,891,600.22 \$ | 237,932.60 \$ | 2,403,683.53 \$ | 4,601,440.05 \$ | 6,827,392.32 \$ | 9,087,527.95 \$ |

Tabla 8: Análisis de rentabilidad

Considerando los datos vistos encontramos los resultados de la rentabilidad financiera de nuestro proyecto, altamente favorables.

Bajo la proyección de 5 años realizada para nuestro proyecto el **Valor Actual Neto** resulta de **RD\$9,696,792.50** al día de hoy, por lo que al ser positivo al descontar los flujos a la tasa de descuento de 15% estaríamos frente a un proyecto rentable.

Por el lado del cálculo de la Tasa Interna de Retorno nos resultó de **128.51%** superior a la tasa de descuento del 15% utilizada para descontar los flujos. Este valor es un importante indicador para tomar la decisión de llevar a cabo el nuevo proyecto, ya que nos permite considerar otras opciones de rentabilidad con menor riesgo y determinar si el proyecto sería viable o no.

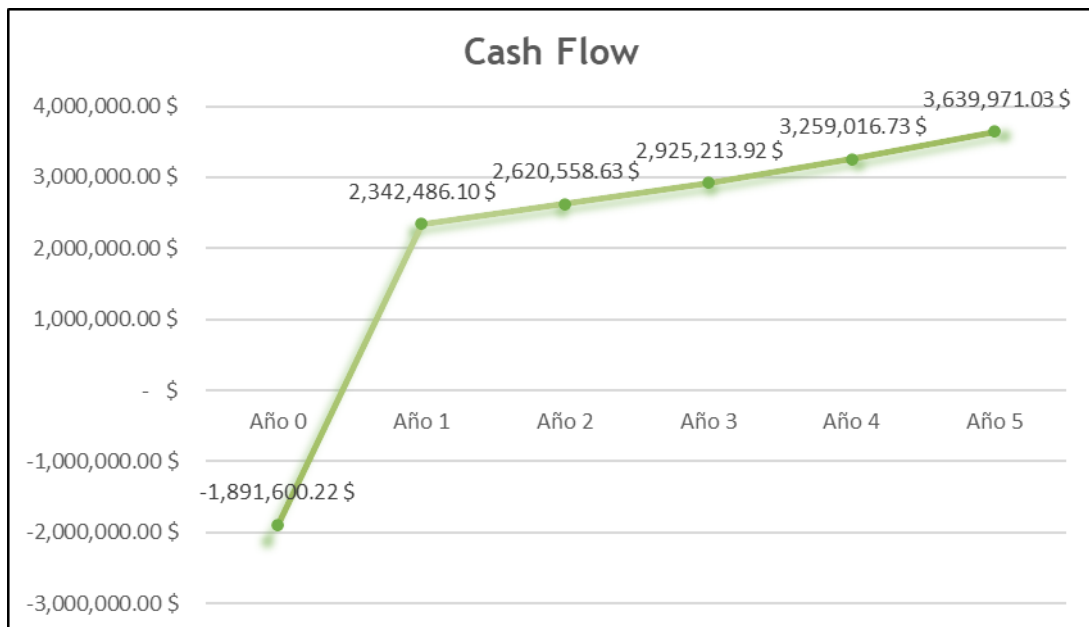


Figura 29: Cash Flow a 5 años

Podemos apreciar un crecimiento constante de los flujos de caja en la proyección que realizamos con un periodo de los 5 años, lo que nos confirma la viabilidad y rentabilidad del proyecto. Mientras que el cálculo resultante del **Retorno de Inversión** nos arrojó un 512%.

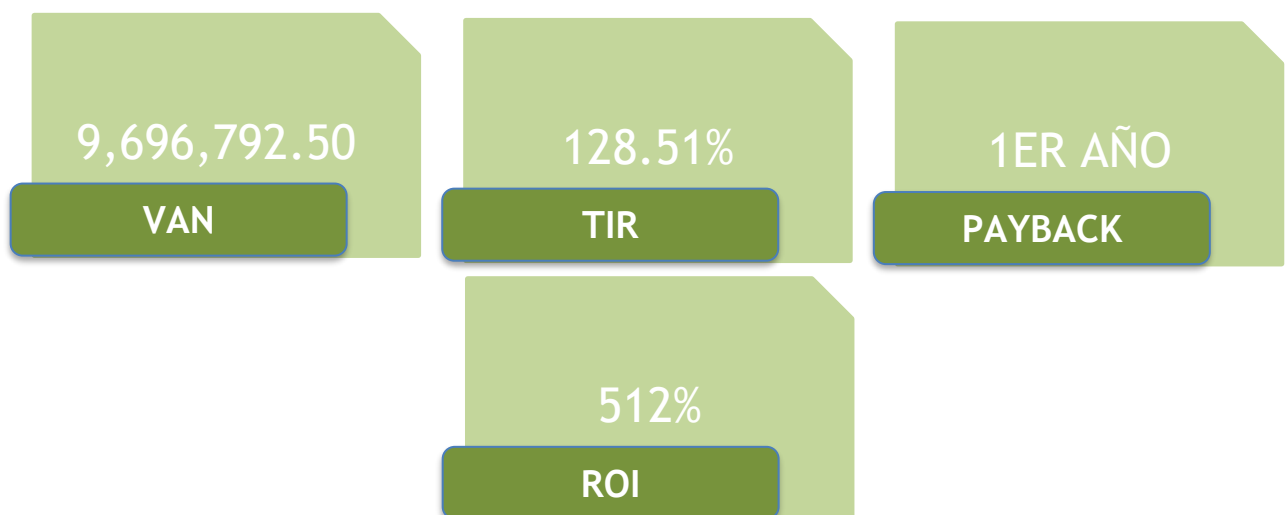


Figura 30: Indicadores rentabilidad

6. Conclusión

A lo largo del desarrollo del proyecto pudimos emplear los conocimientos adquiridos durante el Máster, y de manera especial en el uso de modelos de Machine Learning que bien son empleados para aprender de datos, identificar patrones y tomar decisiones con mínima intervención por parte de un humano; sin dejar a un lado también el uso que le dimos a la capa de visualización y presentación de los datos, que es un aspecto fundamental para que lo procesado mediante los modelos pueda ser visto y explicado de una manera llana al usuario (quien es a final de cuentas que tomará las decisiones relevantes sugeridas), todas estas diferentes actividades realizadas nos han permitido incrementar el conocimiento que teníamos sobre dichos temas y tener una visión clara de cómo poder aplicarlos en nuestro día a día, considerando el cómo nos beneficiarían en las actividades que realizamos en nuestra empresa.

Al haber analizado los resultados de las entrevistas y las encuestas realizadas con el objetivo de validar las hipótesis planteadas al inicio de este proyecto, podemos concluir resaltando la óptima viabilidad del proyecto, tanto a nivel económico como a nivel de mejoras de procesos operativos que actualmente realizan los empleados y departamentos de El Catador, y que les permitirá emplear el tiempo utilizado en otras actividades más relevantes para sus funciones, pues con la presente solución pasaremos de hacer recomendaciones de productos basadas en experiencia a recomendaciones de productos basadas en datos y he aquí la propuesta de valor del proyecto. Mediante el uso de modelos analíticos agruparemos los productos ofrecidos por la empresa de acuerdo con su similitud y otras características relevantes para el modelo, con esta agrupación de productos similares de diferentes segmentos, haremos recomendaciones a los clientes finales basadas en compras anteriores.

Teniendo en cuenta lo anterior planteado, podemos afirmar que estaremos dándole respuesta al problema que existe actualmente con las recomendaciones basadas en experiencia mediante el uso de soluciones de Inteligencia de Negocio aprovechando también tanto la infraestructura tecnológica y recursos humanos existentes en la empresa.

Mediante la definición de algunos indicadores claves de desempeño relacionados a la venta de los productos recomendados por el modelo la alta gerencia podrá evaluar si los objetivos declarados están siendo alcanzados. De este modo también podrán cuantificar los objetivos y validar si el rendimiento esperado está acorde para la toma de decisiones estratégicas.

Para finalizar, podemos asegurar que la solución brindada puede ser escalable a otras empresas en República Dominicana, y entendemos tendremos una excelente tasa de aceptación pues nuestro país está actualmente en una etapa de apertura a la revolución Digital; apoyado por el Gobierno y buscan garantizar el acceso de los dominicanos a las tecnologías de información y comunicación, reduciendo así la brecha tecnológica.

7. Bibliografía

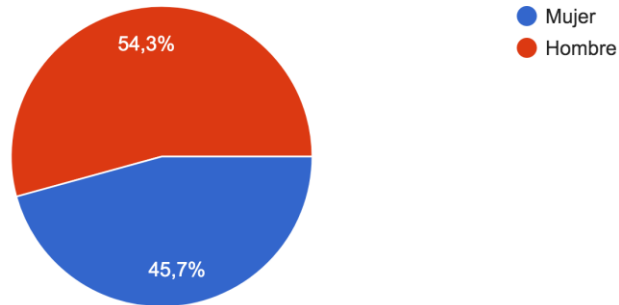
- ❖ El mercado de vino en República Dominicana.
<https://economistadominicano.wordpress.com/2009/07/14/el-mercado-de-vinos-en-republica-dominicana/>
- ❖ Gain report: Dominican Republic Wine Market Brief 2009.
<http://www.calwinexport.com/files/Dominican%20Republic%20Wine%20Brief%202009.pdf>
- ❖ El consumo de bebidas alcohólicas de los dominicanos de Dichter & Neira.
<http://www.dichter-neira.com/wp-content/uploads/2017/02/Mini-Insider-Bebidas-Alcoh%C3%B3licas-RD-V3-1.pdf>
- ❖ Alcohol use and burden for 195 countries and territories.
[https://www.thelancet.com/article/S0140-6736\(18\)31310-2/fulltext](https://www.thelancet.com/article/S0140-6736(18)31310-2/fulltext)
- ❖ Differential effects of polyphenols and alcohol of red wine on the expression of adhesion molecules and inflammatory cytokines related to atherosclerosis: a randomized clinical trial.
<https://academic.oup.com/ajcn/article/95/2/326/4576731>
- ❖ <https://www.linode.com/docs/applications/big-data/how-to-move-machine-learning-model-to-production/>
- ❖ <http://www.b-eye-network.com/view/8360>
- ❖ <http://www.b-eye-network.com/view/8137>
- ❖ <https://docs.aws.amazon.com/machine-learning/latest/dg/pricing.html>
- ❖ De Felipe, Mario. 2019. Introducción a la Inteligencia de Negocio Fundamentos de Business Intelligence, Business Analytics y Big Data.
https://campus.eoi.es/bbcswebdav/pid-259559-dt-content-rid-1086926_1/xid-1086926_1
- ❖ Rochina, Paula. 2016. Python vs R para el análisis de datos. Revista Digital.
<https://revistadigital.inesem.es/informatica-y-tics/python-r-analisis-datos/>
- ❖ Microsoft. 2018. Documentación de SQL. SQL Server Integration Services.
<https://docs.microsoft.com/es-es/sql/integration-services/sql-server-integration-services?redirectedfrom=MSDN&view=sql-server-2017>
- ❖ Dertiano, Victor. 2015. Arquitectura BI (Parte III): El enfoque de Ralph Kimball.
<https://blog.bi-geek.com/arquitectura-el-enfoque-de-ralph-kimball/>
- ❖ López Abellán, Joaquín. 2016. Minería de Datos. Economipedia.
<https://economipedia.com/definiciones/mineria-de-datos.html>
- ❖ Kmedias. Aplicaciones del Algoritmo. En Wikipedia. Recuperado el 06 octubre 2019.
<https://es.wikipedia.org/wiki/K-medias>
- ❖ Efor. 2018. Microsoft Power BI lidera el Cuadrante Mágico de Gartner 2018.
<https://www.lainnovacionnecesaria.com/microsoft-power-bi-lidera-el-cuadrante-magico-de-gartner-2018/>
- ❖ <https://www.elcatador.com.do/>

Anexo 1. Resultados de las encuestas

1- Género (M/F)

Género

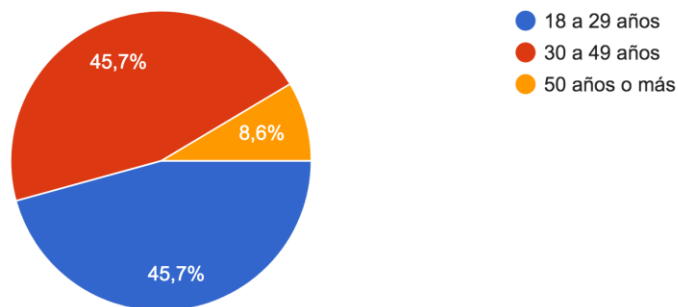
35 respuestas



2- Rango de Edades

Edad

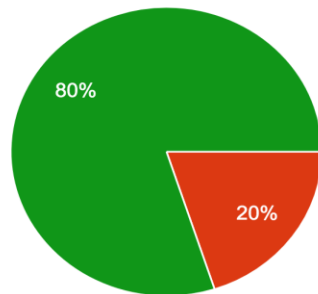
35 respuestas



3- Con qué frecuencia tomas vino

Con que frecuencia tomas vino?

35 respuestas

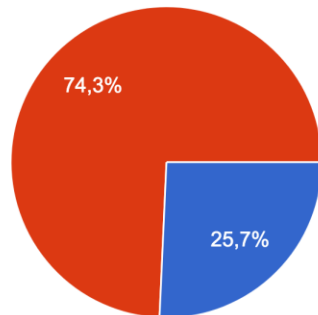


- Diario
- 1 vez a la semana
- 2 o 3 veces a la semana
- Una que otra vez al mes

4- Conocimiento sobre vinos

Posees conocimiento sobre vinos?

35 respuestas

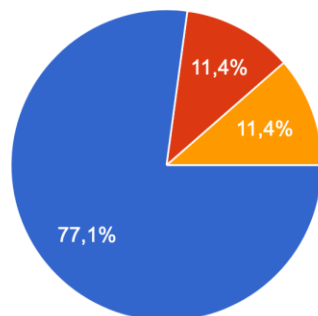


- No conozco sobre vinos
- Conozco un poco
- Domino el tema de los vinos (me pueden decir enólogo)

5-Cuál tipo de vino te gusta

Cual tipo de vino te gusta?

35 respuestas

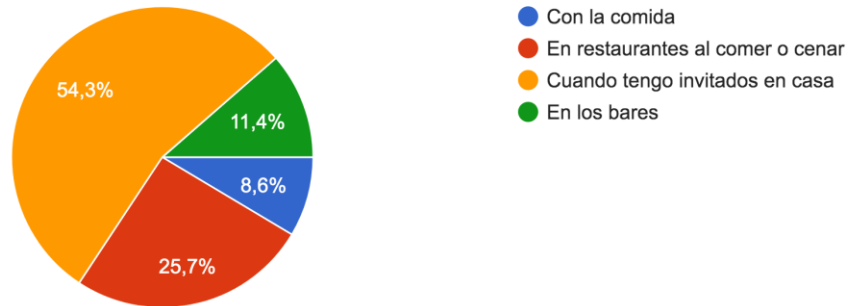


- Tinto
- Rosado
- Blanco

6- En cuáles ocasiones consumes vino

En cuales ocasiones consumes vino?

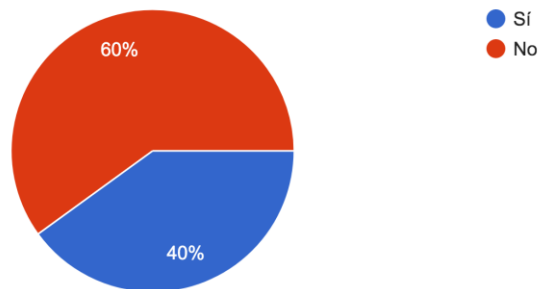
35 respuestas



7- Generalmente consumes la misma marca de vino?

Generalmente consumes la misma marca de vino?

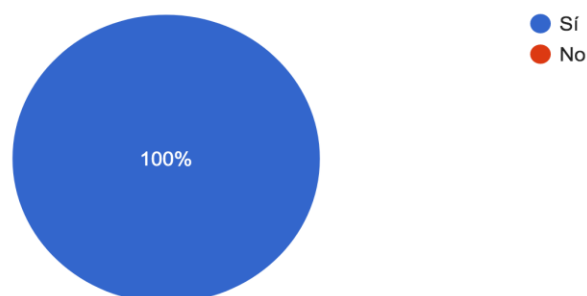
35 respuestas



8- Te darías la oportunidad de probar una nueva marca de vino?

Te darías la oportunidad de probar una nueva marca de vino?

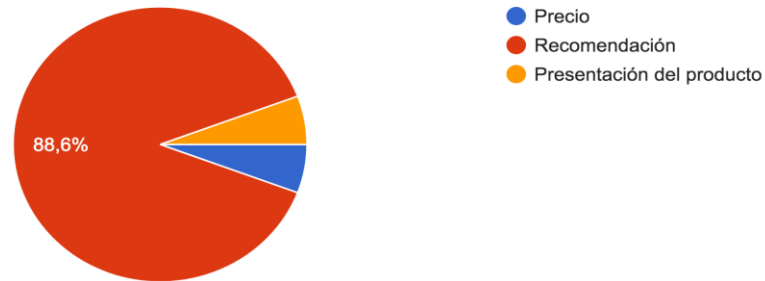
35 respuestas



9- Qué criterio tomarías en cuenta al momento de probar una nueva marca de vino?

Que criterio tomarías en cuenta al momento de probar una nueva marca de vino?

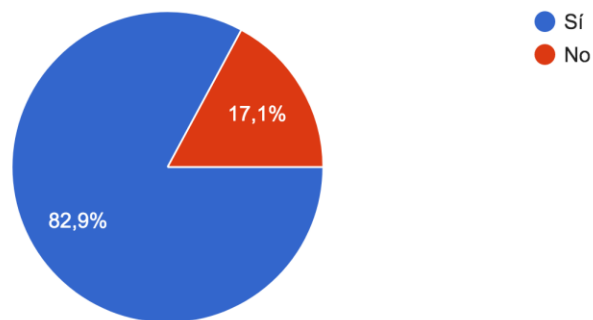
35 respuestas



10- Te gustaría recibir recomendaciones de nuevos vinos considerando tus preferencias de consumo?

Te gustaría recibir recomendaciones de nuevos vinos considerando tus preferencias de consumo?

35 respuestas



Anexo 3. Entrevista en El Catador (Levantamiento de Información)

| | |
|------------------------|--|
| Entrevistado | Santiago Burgos, Director TI |
| Entrevistadores | Julio Peña, Helwing Peña |
| Fecha | 26 de Agosto 2019, 9:00 AM |
| Asunto: | Primer contacto con la Empresa “El Catador”, levantamiento de información |

A continuación, se detalla la información obtenida en el primer contacto que realizamos con la empresa “El Catador”, recibimos el apoyo del Sr. Santiago Burgos, Director del Departamento de Tecnología, quien nos explicó al detalle la situación actual de la empresa y la necesidad puntual que poseen.

La empresa recientemente estuvo trabajando un proyecto de segmentación de clientes B2C y B2B.

La segmentación de los perfiles de los clientes es basada en diversos factores como lo son: antigüedad, valor de compra, tipo de vino, etc.

En el caso de clientes empresariales se categoriza como: Empresa y Corporativo, dicha segmentación es basada primordialmente en el volumen de ventas de cada cliente.

La estructura de ventas de la empresa está conformada por una cartera de vendedores atienden a los perfiles de clientes segmentados y vendedores de piso (ubicados en las tiendas físicas).

Se valida que la recomendación de nuevos productos a los clientes la realizan de manera empírica, basados en el conocimiento del vendedor y los hábitos de consumo del cliente, se le recomiendan productos similares basados en el segmento al cual el cliente pertenece.

Poseen un CRM donde visualizan los vinos favoritos del cliente (información recopilada de las últimas compras), en función de esto el vendedor realiza la recomendación con vinos afines.

La información sobre el cliente el vendedor la conoce al momento de la facturación no previamente, según nos indican están trabajando en el desarrollo de un App con la cual puedan revisar el perfil del cliente antes del proceso de facturación.

Los vinos favoritos del cliente se eligen en función a los que mayor peso económico tienen en las últimas compras del cliente, dado que según nos explican la cantidad comprada no es un indicador relevante.

A partir de lo detallado, se identifican los siguientes agrupamientos relevantes:

Para el caso de perfiles tipo clientes, los tipos se detallan como: Colector, Aficionado, Entusiasta y Casual.

Para el caso de los productos, se realiza una segmentación basada en el precio del producto, donde para: Luxury para vinos mayores de 5,500, Ultra Premium menores de 5,500 y hasta 3,000, Super Premium entre 1,500 y 3,000, Premium entre 750 y 1,500, Popular Premium entre 400 y 750 y por último Popular para vinos menores de 400 pesos.

Se identifican los siguientes canales de ventas: Hoteles-restaurantes, Distribuidores y Personas.

Para la necesidad puntual que se tiene, se nos requiere que se desarrolle un algoritmo que, en función de los clientes y productos, suministre al vendedor de acuerdo al cliente consultado una recomendación de:

- ❖ Un producto de introducción (producto nuevo orientado a su perfil).
- ❖ Un producto de variedad (producto similar del catálogo de productos).
- ❖ Un producto upsale (próximo producto de la gama Ultra Premium/Luxury)

Nos indican que las estrategias de ventas dependerán del perfil del cliente, se detallan las siguientes:

- ❖ Introducción de productos en Carta de vino de restaurantes
- ❖ Coordinación de eventos para dar a conocer productos
- ❖ Ofertas en tiendas físicas
- ❖ Gestión directa (llamadas de ventas y seguimiento)
- ❖ Emails, redes sociales

Los segmentos de clientes se clasifican en dos grandes grupos: Grupo 1 y Grupo 2.

Por el lado de la infraestructura de TI, se describe lo siguiente:

- ❖ Poseen un CRM personalizado de nombre CRM El catador V 1.0 (es su repositorio general de datos)
- ❖ ERP: Microsoft Dynamics GP
- ❖ Cubos: Analysis Services / Reporting services
- ❖ Base de datos MS-SQL
- ❖ BI: Microsoft Power BI
- ❖ Dashboard vendedores de cartera
- ❖ Dashboard vendedores de piso
- ❖ Constan de data de clientes y ventas desde el 2002

Anexo 3. Código Modelos Analíticos en Python

Implementación de un sistema de recomendaciones de vinos

El presente documento muestra el paso a paso para la construcción de las partes que componen el sistema de recomendaciones de vinos que proponemos en este máster.

El mismo se compone de tres partes esenciales:

- 1) El agrupamiento de los vinos dadas las características propias de este y el mayor precio que este ha tenido a lo largo del tiempo.
- 2) La identificación del producto que el cliente más haya comprado en el último año.
- 3) La identificación de productos relacionados a los gustos del cliente en dos vertientes:
 - 3.1) Productos situados en el mismo grupo del producto representativo del cliente.
 - 3.2) Productos situados en el mismo grupo del producto representativo del cliente, además que comportan algún atributo de preferencia del cliente como puede ser el país de origen del vino, la casa productora o la uva que prima en el vino.

Bloque 0: Carga de librerías

En este bloque se importan todas las librerías necesarias para la construcción del sistema de recomendaciones. Así como ajustes de parámetros globales de las mismas.

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from datetime import date, datetime
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans
from sklearn import metrics

%matplotlib inline
plt.style.use('bmh')

# Número de productos por recomendación
no_products_for_recomendation = 1
```

Bloque 1: Carga de datos

En este bloque se importan los datos disponibles para la construcción del sistema de recomendaciones. Específicamente se emplean datos de los productos, de los clientes y de las transacciones efectuadas por los mismos.

In [2]:

```
# Dataframe de productos
products_df = pd.read_csv('./cat_articulos.csv', sep=';', encoding = 'utf-8')
print('Información del dataset de productos:\n')
products_df.info()
```

Dataframe de clientes

```
clients_df = pd.read_csv('./cat_clientes.csv', sep=';', encoding = 'utf-8')
print("\nInformación del dataset de clientes:\n")
clients_df.info()
```

Dataframe de transacciones

```
transactions_df = pd.read_csv('./cat_transacciones.csv', sep=';', encoding = 'utf-8')
print("\nInformación del dataset de transacciones:\n")
transactions_df.info()
```

Información del dataset de productos:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7925 entries, 0 to 7924
Data columns (total 11 columns):
ID_ARTICULO      7925 non-null object
ID_CLASE_ARTICULO 7925 non-null object
POPULARIDAD      7925 non-null object
CODIGO_ABC       7925 non-null object
TIPO_BEBIDA      7925 non-null object
PAIS             7925 non-null object
REGION          7925 non-null object
UVA             7925 non-null object
ESTADO          7925 non-null int64
ID_CASA         7925 non-null object
BODEGA          7925 non-null object
dtypes: int64(1), object(10)
memory usage: 681.1+ KB
```

Información del dataset de clientes:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32153 entries, 0 to 32152
Data columns (total 37 columns):
ID_CLIENTE      32153 non-null object
ISGRUPO         32153 non-null object
CodGrupo        32153 non-null object
ClasfNoVenta    32153 non-null object
ClaseCliente    32153 non-null object
Localidad       32057 non-null object
ClasificacionValorLTM 32153 non-null object
SegmentoPrecio  17283 non-null object
Antiguedad      32149 non-null object
Frecuencia      32149 non-null object
Casa1           12514 non-null object
Catas2          12514 non-null object
Pais1           11746 non-null object
```

```

Pais2          11746 non-null object
Uva1           10879 non-null object
Uva2           10879 non-null object
Vendedor       32153 non-null object
FechaCreacion  32153 non-null object
FechaUltVenta  32153 non-null object
ULTRAUP        32153 non-null float64
SUPERUP        32153 non-null float64
LUXURY         32153 non-null float64
CHAMPAGNE      32153 non-null float64
ULTRAPREMIUM  32153 non-null float64
SUPERPREMIUM  32153 non-null float64
REGALO         32153 non-null float64
DiasNoVenta    32153 non-null int64
UltClasificacionValor 15316 non-null object
FrecuenciaMesesCompralTM 32153 non-null object
MesesCompralTM 32153 non-null int64
TendenciaLTMvsAA 18039 non-null object
Score          32153 non-null float64
SWAG           32153 non-null float64
PerfilPotencial 32153 non-null object
PERFIL         32153 non-null object
PERFIL_ANTERIOR 32153 non-null object
ClasificacionValor 32153 non-null object
dtypes: float64(9), int64(2), object(26)
memory usage: 9.1+ MB

```

Información del dataset de transacciones:

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 263890 entries, 0 to 263889
Data columns (total 12 columns):
TIPO_DOCUMENTO    263890 non-null object
USUARIO_DIGITA    263890 non-null object
NUMERO_DOCUMENTO  263890 non-null int64
ID_CLIENTE        263890 non-null object
ID_VENDEDOR       263890 non-null int64
FECHA             263890 non-null object
ALMACEN_HEADER    263890 non-null object
ALMACEN_LINEA     263890 non-null object
ID_ARTICULO       263890 non-null object
NIVEL_PRECIO      263890 non-null object
CANTIDAD          263890 non-null float64
IMPORTE_VENTA     263890 non-null float64
dtypes: float64(2), int64(2), object(8)
memory usage: 24.2+ MB

```

Bloque 2: Transformación de datos

En este bloque se hace un procesamiento de los datos importados en el bloque 1 con la finalidad de solo dejar aquellos que son útiles para la construcción del sistema y hacer las transformaciones de lugar para un efectivo procesamiento de los mismos.

In [3]:

```
# Del dataframe de productos solo se necesitan los vinos
products_df = products_df[products_df['TIPO_BEBIDA'] == 'Vino']

# Se eliminan las columnas que no son necesarias
products_df = products_df.drop('ID_CLASE_ARTICULO', 1).drop('TIPO_BEBIDA', 1).drop('ESTADO', 1)
).drop('BODEGA', 1)
products_df.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3613 entries, 2 to 7312
Data columns (total 7 columns):
ID_ARTICULO    3613 non-null object
POPULARIDAD    3613 non-null object
CODIGO_ABC     3613 non-null object
PAIS           3613 non-null object
REGION         3613 non-null object
UVA           3613 non-null object
ID_CASA        3613 non-null object
dtypes: object(7)
memory usage: 225.8+ KB
```

In [4]:

```
# Del dataframe de clientes se extraen las variables que albergan un gusto del cliente
clients_reduced_df = clients_df[['ID_CLIENTE', 'Casa1', 'Catas2', 'Pais1', 'Pais2', 'Uva1', 'Uva2']]

# Se eliminan los espacios en el valor del ID del cliente
clients_reduced_df['ID_CLIENTE'] = clients_reduced_df.apply(lambda row: row['ID_CLIENTE'].strip(), axis=1)
/home/japo/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
''''
```

In [5]:

```
# Del dataframe de transacciones solo se necesitan las facturas de venta
transactions_df = transactions_df[transactions_df['TIPO_DOCUMENTO'] == 'FACTURA']

# Se eliminan las columnas que no son necesarias
```

```
transactions_df = transactions_df.drop('USUARIO_DIGITA', 1).drop('ID_VENDEDOR', 1).drop('TIPO_DOCUMENTO', 1).drop('ALMACEN_HEADER', 1).drop('ALMACEN_LINEA', 1).drop('NIVEL_PRECIO', 1)
```

```
# Se calcula el precio unitario de cada producto en cada transacción
```

```
transactions_df['PRECIO_UND'] = transactions_df['IMPORTE_VENTA'] / transactions_df['CANTIDAD']
```

```
# Se transforma el valor de la columna de fecha de una cadena de caracteres a un objeto de fecha
```

```
dates = []
```

```
for row in transactions_df['FECHA']:
```

```
    try:
```

```
        dates.append(datetime.strptime(row, '%Y-%m-%d %H:%M:%S.%f'))
```

```
    except:
```

```
        dates.append(None)
```

```
transactions_df['FECHA'] = dates
```

In [6]:

```
# Se procesan las transacciones para identificar el precio unitario de cada producto
```

```
# y guardarlo en un diccionario para su posterior utilización
```

```
art_prices_df = transactions_df.groupby('ID_ARTICULO')['PRECIO_UND'].max().to_frame().reset_index()
```

```
art_prices_df.columns = ['ID_ART', 'PRECIO_UND']
```

```
art_prices_dict = {}
```

```
for index, row in art_prices_df.iterrows():
```

```
    art_prices_dict[row['ID_ART'].strip()] = row['PRECIO_UND']
```

```
# Se identifica el precio de cada producto y se agrega al dataframe de productos
```

```
prices = []
```

```
for row in products_df['ID_ARTICULO']:
```

```
    try:
```

```
        prices.append(art_prices_dict[row])
```

```
    except:
```

```
        prices.append(0)
```

```
products_df['PRECIO_UND'] = prices
```

Out[6]:

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND |
|---|-------------|-------------|------------|-------|--------|------------|----------------------|------------|
| 2 | 0104 | ULT PREMIU | D | Chile | Apalta | Cabernet S | MONTES | 5254.24 |
| 3 | 0105 | ULT PREMIU | D | Chile | Apalta | Shiraz | VINOS DESCONTINUADOS | 3432.20 |

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND |
|---|-------------|-------------|------------|-------|--------|------------|------------|------------|
| 4 | 0106 | PREMIUM | D | Chile | Apalta | Sauvign B | MONTES | 0.00 |
| 5 | 0107 | PREMIUM | D | Chile | Apalta | Chardonnay | MONTES | 1398.31 |
| 6 | 0108 | SUP PREMIU | D | Chile | Apalta | Cabernet S | SANTA RITA | 0.00 |

In [7]:

```
# Se elijen las transacciones del presente año para identificar por cada cliente
# el producto que este más haya comprado
min_date = pd.Timestamp(date.today().year, 1, 1)
trans_last_year_df = transactions_df[transactions_df['FECHA'] >= min_date]
clients_prods_most_buied = trans_last_year_df.groupby(['ID_CLIENTE', 'ID_ARTICULO'])['CANTIDAD']
].sum().reset_index().sort_values(by=['ID_CLIENTE', 'CANTIDAD'], ascending=False).groupby('ID_CLIENTE').head(1)
clients_prods = {}
for index, row in clients_prods_most_buied.iterrows():
    clients_prods[row['ID_CLIENTE'].strip()] = row['ID_ARTICULO'].strip()
```

Out[7]:

```
{'WBT006': 'WBI003',
'WBT005': 'WBI003',
'WBT004': 'WBI003',
'WBT003': 'WBI003',
'WBT002': 'WBI003',
'WBT001': 'WBI003',
'WBS003': '1204',
'WBS002': 'WBI003',
'WBS001': 'WBI003',
'WBE003': 'WBI003',
'WBE002': 'WB0095',
'WBE001': 'WB0029',
'WBA001': 'WBI003',
...
...
...
...
'SF30466': '4642',
'SF30465': '2467',
'SF30464': '6506',
...}
```


Bloque 3: Modelado

En este bloque se construye el modelo analítico para agrupar los productos por sus atributos principales.

Primeramente, es necesario traducir los atributos categoricos del producto a su equivalente numérico de tal forma que puedan ser empleados para el entrenamiento de un modelo de aprendizaje automatizado.

Bloque 3.1: Preparación del dataset

In [8]:

```
# Se eliminan las variables del producto que no se utilizaran en el modelado
temp = products_df.drop('ID_ARTICULO', 1).drop('PAIS', 1).drop('ID_CASA', 1)

# Se transforman de variables categorias a numericas y se obtiene el dataset de entrenamiento del modelo
products_encoding_df = pd.get_dummies(temp)
products_encoding_df.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 3613 entries, 2 to 7312
Columns: 218 entries, PRECIO_UND to UVA_semillon
dtypes: float64(1), uint8(217)
memory usage: 822.1 KB
```

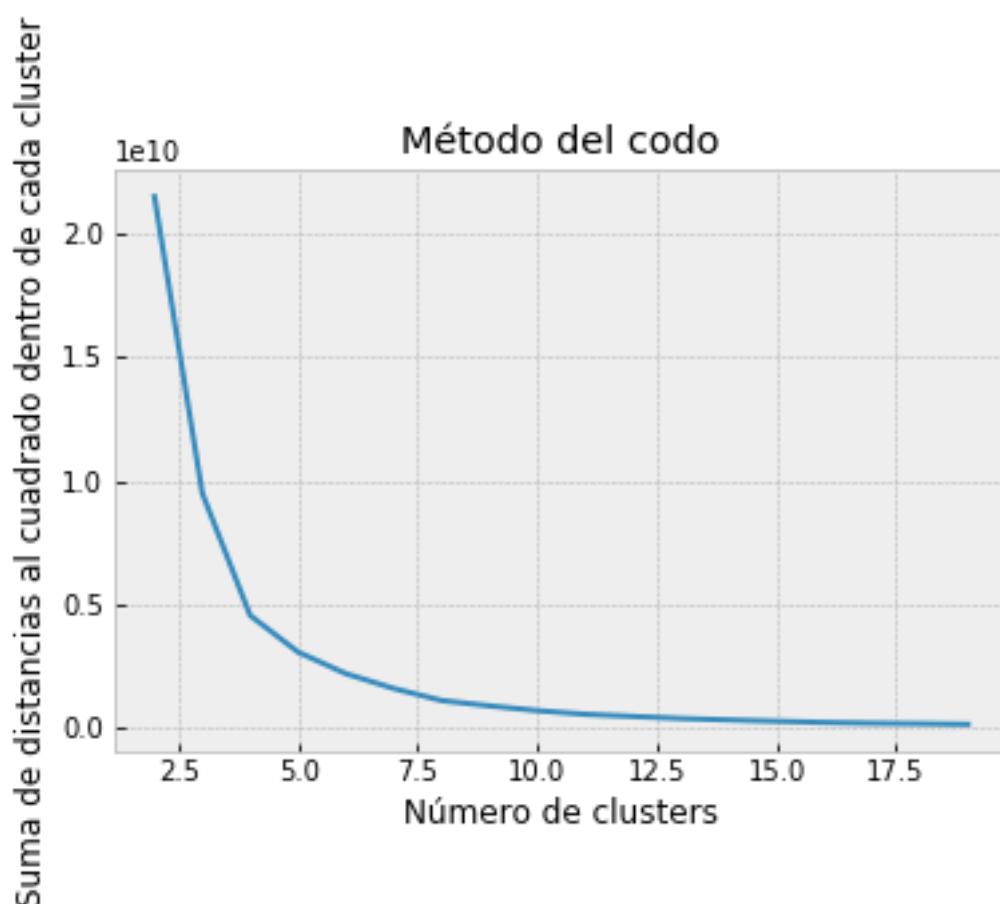
Bloque 3.2: Identificación de clusters

En el área de Machine Learning cuando se enfrenta un problema de clusterizado es muy común no saber que cantidad de clusters son optimos para solucionar el problema, por lo cual se emplea un método que describe la relación entre los datos en cada posible cluster. Este método es conocido como "Método del Codo" y será empleado acontinuación.

In [9]:

```
res = []
N_max=20 # Número máximo de clusters a evaluar
N_min=2 # Número mínimo de clusters a evaluar
for i in range(N_min, N_max):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
    kmeans.fit(products_encoding_df)
    res.append(kmeans.inertia_) # Obtenemos la suma de las distancias al cuadrado dentro de cada cluster

# Gráfica con los resultados del método del codo
plt.plot(range(N_min, N_max), res)
plt.title('Método del codo')
plt.xlabel('Número de clusters')
plt.ylabel('Suma de distancias al cuadrado dentro de cada cluster')
plt.show()
```



Se visualiza pues como en a mayor cantidad de cluster los elementos dentro de estos estan más relacionados por la reducida distancia entre ellos. Para la construcción del modelo se emplearan 8 clusters, debido a que es el número intermedio en el cuadrante 7.5-10 que es donde se identifica la normalización del comportamiento de la tendencia.

In [10]:

```
#El no. de clusters seleccionado
```

```
n_clusters = 8
```

Bloque 3.3: Entrenamiento del modelo

Se ha elegido "k-means" como el algoritmo de clusterizado por ser uno de los más empleados para resolver este tipo de problemas y ser uno de los más efectivos.

In [11]:

```
# Construcción del modelo
```

```
modelProdClustering = KMeans(n_clusters = n_clusters, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
```

```
# Entrenamiento del modelo
```

```
modelProdClustering.fit(products_encoding_df)
```

Out[11]:

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300, n_clusters=8, n_init=10, n_jobs=None, precompute_distances='auto', random_state=0, tol=0.0001, verbose=0)
```

In [12]:

```
# Predicción de los clusters para cada registro deseado
```

```
labels_clusters = modelProdClustering.predict(products_encoding_df)
```

```
# Incorporación del número de cluster al dataset de productos
```

```
products_encoding_df['CLUSTER'] = labels_clusters
```

```
products_df['CLUSTER'] = labels_clusters
```

Bloque 3.4: Visualización del resultado

Una vez se ha obtenido el modelo y se ha empleado para identificar los clusters de los datos de entrada, que en este caso es todo el conjunto de productos de tipo "vino".

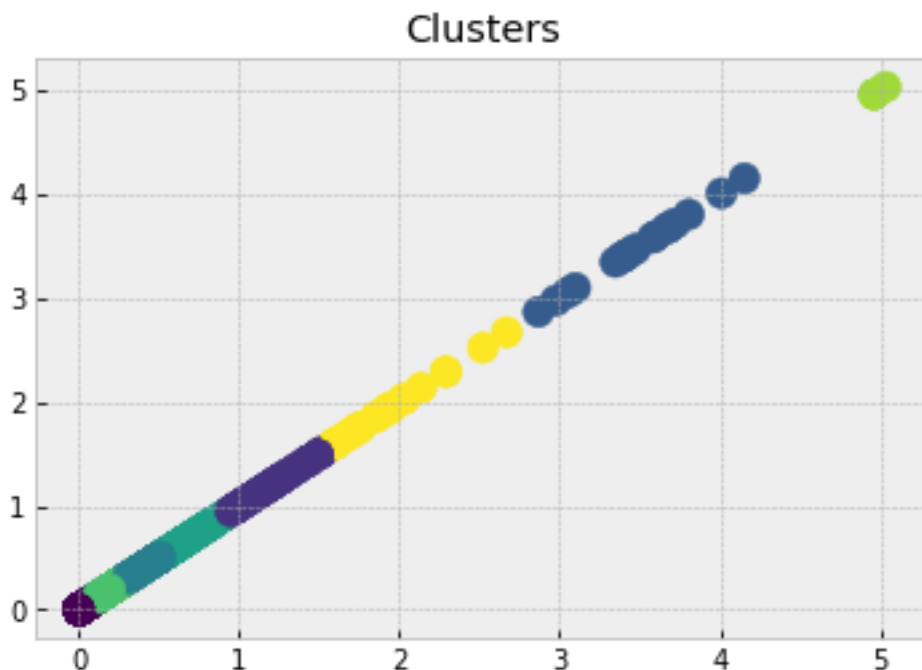
En la visualización de muestra una representación de cada artículo (un círculo) en un cuadrante dado su precio y el mismo con un color que representa el cluster al que pertenece.

In [13]:

```
scale = 0.0001
```

```
plt.title("Clusters")
```

```
plt.scatter(products_df['PRECIO_UND']*scale, products_encoding_df['PRECIO_UND']*scale, c=labels_clusters, s=130);
```



Por otra parte, se cálculo el porcentaje de efectividad del modelo para clasificar de manera correcta nuevos artículos en el futuro.

In [14]:

```
# Cálculo de la efectividad de clasificación del modelo
```

```
labels = modelProdClustering.labels_
```

```
score = metrics.silhouette_score(products_encoding_df, labels, metric='euclidean')
```

```
print('El porcentaje de efectividad del modelo es ', round(score,2)*100,'%')
```

```
El porcentaje de efectividad del modelo es 79.0 %
```

Bloque 4: Desarrollo de funciones

En este bloque se definen todas aquellas funciones que empleará el sistema de recomendaciones para arrojar un resultado de acuerdo a la entrada del usuario.

A nivel de negocio, el cliente cuenta con una clasificación para agrupar sus productos en segmentos de precios, es por ello que las mismas se emplean para arrojar recomendaciones ajustadas a los criterios del cliente, que son:

- 1) Recomendar al cliente un producto del segmento que el suele consumir.
- 2) Recomendar al cliente un producto de un segmento superior al que suele consumir.
- 3) Recomendar al cliente un producto de introducción.

In [15]:

Definición de los segmentos de productos propios del cliente

```
product_prices_segment = {
    'LUXURY':{'gt':5500,'top':1},
    'ULT PREMIU':{'lte':5500,'gt':3000,'top':2},
    'SUP PREMIU':{'lte':3000,'gt':'1500','top':3},
    'PREMIUM':{'lte':1500,'gt':750,'top':4},
    'POP PREMIU':{'lte':750,'gt':400,'top':5},
    'POPULAR':{'lte':750,'top':6},
    'INTRO':{'top':7}
}
```

Función para obtener el segmento de producto dada su posición en el top

```
def getSegmentByTop(top):
    if top < 1:
        top = 1
    elif top > 6:
        top = 6
    for key in product_prices_segment:
        if product_prices_segment[key]['top'] == top:
            return key
    return None
```

In [16]:

Función para buscar en el dataframe de productos, un producto por su ID

```
def getProductById(productId):
    return products_df[products_df['ID_ARTICULO'] == productId]
```

Función para buscar todos los productos perteneciente a un mismo grupo (cluster)

```
def getProductsByCluster(cluster):
    return products_df[products_df['CLUSTER'] == int(cluster)]
```

Función para buscar un producto dentro un mismo segmento

Se recibe un producto que sirve como base y una serie de productos relacionados

```
def getProductSameSegment(origin, products):
    popularity = origin['POPULARIDAD'].values[0]
    products_finded = products[products['POPULARIDAD'] == popularity]
```

```

if products_finded.empty is False:
    return products_finded.sample(n=no_products_for_recomendation)
else:
    None

```

```

# Función para buscar un producto de un segmento superior
# Se recibe un producto que sirve como base y una serie de productos relacionados
def getProductHigherSegment(origin, products):
    popularity = origin['POPULARIDAD'].values[0]
    higher_popularity = getSegmentByTop(product_prices_segment[popularity]['top']-1)
    products_finded = products[products['POPULARIDAD'] == higher_popularity]
    if products_finded.empty is False:
        return products_finded.sample(n=no_products_for_recomendation)
    else:
        None

```

```

# Función para buscar un producto de introducción
# Se recibe un producto que sirve como base y una serie de productos relacionados
def getProductOfIntroduction(origin, products):
    popularity = 'INTRO'
    products_finded = products[products['POPULARIDAD'] == popularity]
    if products_finded.empty is False:
        return products_finded.sample(n=no_products_for_recomendation)
    else:
        None

```

```

# Función para buscar tres recomendaciones para el cliente dado
# Una recomendación de un producto del mismo segmento que el cliente suele consumir
# Otra recomendación de un producto de un segmento superior al que el cliente suele consumir
# Y otra recomendación de un producto de introducción
def getRecomendationsForClient(clientId):
    product_id = clients_prods[clientId]
    if product_id is not None:
        product = getProductById(product_id)
        related_products = getProductsByCluster(product['CLUSTER'])
        productSameSegment = getProductSameSegment(product, related_products)
        productHigherSegment = getProductHigherSegment(product, related_products)
        productOfIntroduction = getProductOfIntroduction(product, related_products)
        return (productSameSegment, productHigherSegment, productOfIntroduction)

```

```

# Función para buscar tres recomendaciones para el cliente dado
# con la particularidad que emplea rasgos del gusto del cliente como:
# el país de origen, la casa producto o la uva del vino.
def getMoreSpecificRecomendationsForClient(clientId):
    client = clients_reduced_df[clients_reduced_df['ID_CLIENTE'] == clientId]
    product_id = clients_prods[clientId]

```

```

if product_id is not None:
    product = getProductById(product_id)
    related_products = getProductsByCluster(product['CLUSTER'])
    check = (related_products['PAIS'] == client['Pais1'].values[0].strip()) | (related_products['PAIS'] == client['Pais2'].values[0].strip()) | (related_products['ID_CASA'] == client['Casa1'].values[0].strip()) | (related_products['ID_CASA'] == client['Casa2'].values[0].strip()) | (related_products['UVA'] == client['Uva1'].values[0].strip()) | (related_products['UVA'] == client['Uva2'].values[0].strip())
    filtered_products = related_products[check]
    productSameSegment = getProductSameSegment(product, filtered_products)
    productHigherSegment = getProductHigherSegment(product, filtered_products)
    productOfIntroduction = getProductOfIntroduction(product, filtered_products)
    return (productSameSegment, productHigherSegment, productOfIntroduction)

```

Bloque 5: Ejemplos de uso del sistema

En este bloque se emplean las funciones desarrolladas en el bloque 4 para mostrar el uso del sistema de recomendación y los resultados del mismo.

Recomendaciones no ajustadas

Las recomendaciones que se obtienen a continuación no emplean los rasgos de gusto del consumo del cliente que han sido recogidas por el negocio a lo largo del tiempo, por ende se tratan de recomendaciones basadas puramente en las características del producto.

In [17]:

```
same_segment, higher_segment, product_of_introduction = getRecomendationsForClient('SF31562')
```

In [18]:

```
#Producto recomendado dentro del mismo segmento
```

```
same_segment
```

Out[18]:

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND | CLUSTER |
|------|-------------|-------------|------------|------|--------|------------|---------|------------|---------|
| 2563 | 3721 | PREMIUM | B | USA | Lodi | Cabernet S | BOGLE | 1675.0 | 5 |

In [19]:

```
#Producto recomendado de un segmento superior
```

```
higher_segment
```

Out[19]:

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND | CLUSTER |
|------|-------------|-------------|------------|---------|---------|------------|-------------|------------|---------|
| 1379 | 2146 | SUPPREMIUM | D | Francia | Borgoña | Pinot Noir | LOUISLATOUR | 2025.42 | 5 |

In [20]:

```
#Producto recomendado de introducción
```

product_of_introduction

Recomendaciones ajustadas

Las recomendaciones que se obtienen a continuación emplean los rasgos de gusto del consumo del cliente que han sido recogidas por el negocio a lo largo del tiempo, adicional al uso de las características del producto.

In [25]:

```
same_segment, higher_segment, product_of_introduction = getMoreSpecificRecommendationsForClient('SF31562')
```

In [26]:

#Producto recomendado dentro del mismo segmento

same_segment

Out[26]:

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND | CLUSTER |
|------|-------------|-------------|------------|---------|------------|-------|-----------------|------------|---------|
| 1259 | 2023 | PREMIUM | C | Francia | Beaujolais | Otras | GEORGES DUBOUFF | 889.83 | 5 |

In [27]:

#Producto recomendado de un segmento superior

higher_segment

Out[27]:

| | ID_ARTICULO | POPULARIDAD | CODIGO_ABC | PAIS | REGION | UVA | ID_CASA | PRECIO_UND | CLUSTER |
|-----|-------------|-------------|------------|-----------|---------|--------|----------------------|------------|---------|
| 623 | 0726 | SUPPREMIUM | D | Australia | Barossa | Shiraz | VINOS DESCONTINUADOS | 1411.02 | 5 |

In [28]:

#Producto recomendado de introducción

product_of_introduction

FIN.