Imagen que contiene Texto

Descripción generada automáticamente

Modelo de marketing para entidad bancaria con metodología crisp-dm

Por:

Daniel Esteban González Zuluaga

David Sequera Lievano

Profesor a cargo:

Jorge Andrés Alvarado Valencia

Pontificia Universidad Javeriana

Maestría de Analítica para la Inteligencia de Negocios

Métodos y aplicaciones de analítica1

2024

1. **Business Understanding**
   1. **Background**

Este documento ofrece un análisis enfocado en el comportamiento del mercado en Estados Unidos en relación con los alimentos saludables, utilizando como punto de partida los datos recopilados de la plataforma Instacart. Teniendo en cuenta que se va a tener un enfoque en comida saludable, se presenta el siguiente análisis del entorno previo. Recuerde que el análisis es **previo**, es decir puede que alguna información no tenga directa relación o tampoco se la haga mención en las conclusiones precisamente porque se realiza antes de conocer la información. Esto es bueno ya que no fuerza una “coincidencia” y busca evitar el sesgo. Las relaciones que tengan sentido serán mencionadas más adelante, la interpretación de los datos adicionales queda a gusto del lector.

1. Political Factor

En los estados unidos la entidad gubernamental que regula la calidad de los productos y el bienestar de las personas es la FDA (Food and Drug Administration). Se encarga de regular los productos del mercado y obligar a los productores a seguir ciertas políticas y estándares como las etiquetas nutricionales, y a su vez educar a los consumidores. Es importante entender que esta entidad no provee una definición exacta de comida saludable, quizás  porque el concepto es muy amplio y puede que atienda más a una dieta como conjunto de productos y rutinas, sin embargo, para efectos prácticos usaremos como punto de partida una definición acotada, proveniente de la normativa que indica que se puede utilizar comercialmente el término “saludable” siempre y cuando siga ciertas condiciones especificadas en la sección 21CFR 101.65(d) del código de regulación federal ​(OFR NARA; GPO, 2024)​

Definición propia y punto de partida. Un producto puede ser considerado saludable si

* Tiene presencia de componentes beneficiosos como frutas, verduras y cereales integrales.
* Tiene límites de componentes negativos como azúcares añadidos, sodio y grasas no saludables.

Adicionalmente como un detalle del cual no profundizaremos dejamos una gráfica de la escuela de medicina de Harvard sobre lo que significa un plato saludable.

Diagrama

Descripción generada automáticamente

*Ilustración 1 Plato de comida saludable ​*(Harvard University, 2023)*​*

Para nuestro contexto (una empresa como Instacart) es bueno tener un conocimiento de la forma en la que se define la comida saludable ya que nos puede ayudar a atraer un publico que le interesa el tema y a su vez a profundizar en los productos disponibles. No solo poder relacionar estos productos a nivel de sistemas de recomendación, sino también luego de conocer estos productos poder comunicar de manera adecuada al consumidor que busca una dieta para mejorar su calidad de vida.

1. Economic Factor

Mapa

Descripción generada automáticamente A map of the world

Description automatically generated

|  |  |
| --- | --- |
| *Ilustración 2 Costo diario de una dieta saludable ​*(Herforth, 2022)*​* | *Ilustración 3 Población incapaz de suplir una dieta saludable ​*(Ritchie, Rosado, & Roser, Our World in Data, 2023)*​* |

En cuanto al poder económico para la adquisición de productos saludables estados unidos es un país el cual tiene un poder adquisitivo capaz sostener una dieta saludable. A continuación, se presenta para la primera ilustración el costo en dólares de una comida saludable en cada país; para la segunda imagen el porcentaje de la población que no puede costear una comida saludable; la tercera ilustración muestra el porcentaje que no se puede costear una dieta calórica suficiente, curiosamente estados unidos tiene un indicador muy alto en esta ilustración, lo cual podría ser una oportunidad para ofrecer un producto en este sector de bajo costo; finalmente la última muestra los ingresos medios respecto al costo de la comida saludable como podemos observar es bastante superior el indicador de los estados unidos.

Imagen que contiene Gráfico

Descripción generada automáticamente A map of the world with different colored countries/regions

Description automatically generated

|  |  |
| --- | --- |
| *Ilustración 4* Personas *incapaces de suplir una dieta calóricamente suficiente ​*(Ritchie, Our World in Data, 2021)*​* | *Ilustración 5 Costo de una dieta contra el ingreso medio por país ​*(Ritchie, Our World in Data, 2021)*​* |

De las ilustraciones se puede asumir que el país cuenta con la capacidad económica para suplir una dieta alimenticia y que además contrario a lo que se piensa, puede tener dificultades nutricionales en áreas inesperadas las cuales pueden ser una muy buena oportunidad para cautivar a un mercado con un buen poder adquisitivo.

1. Social Factor

Respecto al ámbito social el cual nos habla del interés del mercado, podemos ver de un reporte de la FMI como la confianza de los consumidores sube respecto a la FDA es bastante fuerte, la gente está interesada en entender que contiene los productos que consume.

A graph of sales growth

Description automatically generated with medium confidenceA chart of food content

Description automatically generated with medium confidence

|  |  |
| --- | --- |
| *Ilustración 6 Confianza en entidades de las personas a la hora de comprar ​*(FMI, 2017)*​* | *Ilustración 7 Afirmaciones de productos que buscan los compradores cuando compran un producto alimenticio ​*(FMI, 2017)*​* |

A graph with blue rectangles

Description automatically generated

*Ilustración 8 Encuesta sobre la definición de “ser saludable”​* (Barbalova & Cazin, 2024)*​*

**1.2 Business goal**

1. Optimización de inventarios y compras: Utilizar el análisis de datos para predecir la demanda de productos y así optimizar los niveles de inventario para garantizar la disponibilidad de productos frescos y así también minimizar las perdidas por exceso de inventario, así mismo observar los días de compra para ofrecer beneficios.
2. Expansión en la oferta de productos: Analizar los datos en la tendencia de compra y preferencias de los clientes para identificar oportunidades de expansión que satisfagan la necesidad y deseos de los clientes.

**1.3 Business success criteria**

1. Optimización de inventarios y compras

* Eficiencia en gestión de inventario: Este KPI se calcula mediante la observación y seguimiento de la cantidad de productos excedentes y aumento en la compra de los días menos comunes.

1. Expansión en la oferta de productos:

* Tasa de aceptación en combos o productos nuevos: Este KPI evalúa la aceptación de los nuevos productos introducidos en la tienda por parte de los clientes. Se calcula como el porcentaje de ventas totales representadas por estos nuevos productos durante un período de tiempo específico.

**1.4 Data mining goal**

1. Identificar dentro de las compras de los usuarios, el departamento que tenga un buen support

y además contenga dentro de sí productos saludables.

1. Identificar patrones de compra de los clientes y desarrollar un modelo para seleccionar y recomendar productos con un alto potencial de venta (support), maximizando simultáneamente el lift de las recomendaciones personalizadas en nuestra plataforma de comercio electrónico y local.

**1.5 Data mining success criteria**

1. Identificar correctamente el departamento que ofrece un buen soporte al cliente con una alta precisión. Esto se puede medir mediante la tasa de aciertos en la clasificación de las transacciones de compra.
2. El KPI para este objetivo se centra en evaluar el impacto de las recomendaciones personalizadas en las ventas totales. Este KPI abarca dos métricas clave: el lift, que refleja la eficacia de la venta conjunta de dos o más productos, y el support, que muestra la proporción de las ventas de uno o más productos con respecto a las ventas totales. El objetivo primordial es maximizar tanto el lift como el support, garantizando así que las recomendaciones personalizadas generen un impacto relevante en las ventas totales.

**2. Data Understanding**

**2.1 Describe data**

En la base de datos se identificaron aproximadamente 33.8 millones de productos comprados y 3.4 millones de ordenes en la plataforma de instacart. La base de datos sin modificar viene con las siguientes tablas:

**Aisle**: Esta tabla contiene dos columnas ‘aisle\_id’ y ‘aisle’, representan los pasillos físicos o de la aplicación donde se encuentran ubicados los productos. En total hay 134 “pasillos” distintos.

**Departments**: Esta tabla se organiza de la misma forma que Aisle con las columnas ‘department\_id’ y ‘department’, representan los departamentos que existen en la plataforma. En total hay 21 departamentos distintos.

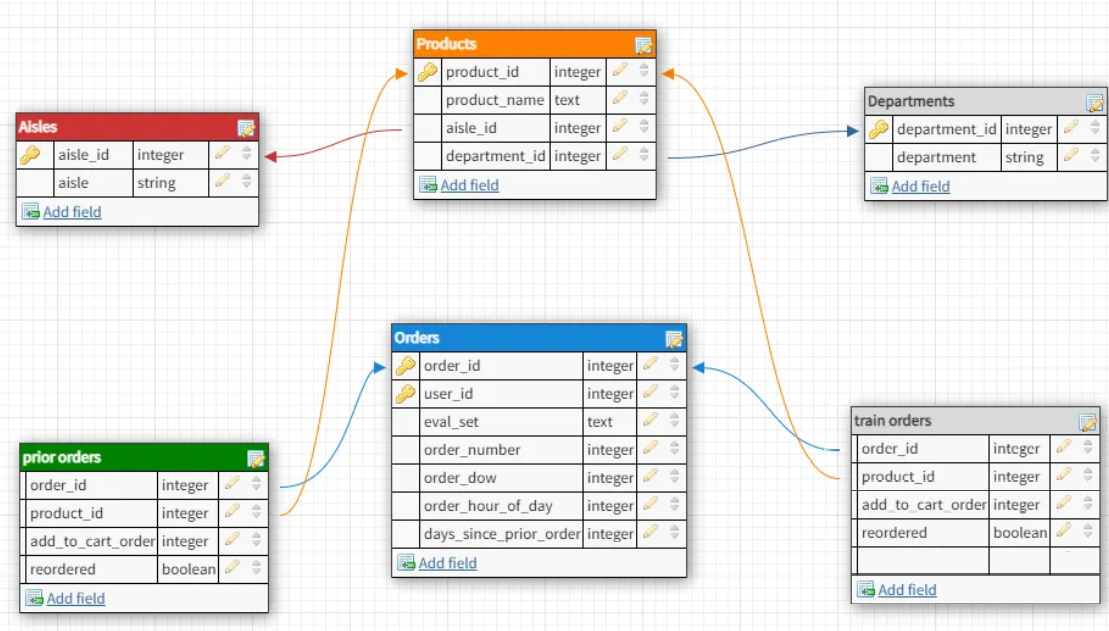
**Order\_products\_\***: Esta parte se compone de dos tablas ‘order\_products\_prior’ y ‘order\_products\_train’, estas dos tablas se componen de las mismas columnas, comenzando con ‘order\_id’ y ‘product\_id’, estas columnas se entienden por sí solas, pero hay otras dos columnas:

* Add\_to\_cart\_order: Muestra el orden en el que se agregó el producto al carrito durante ese pedido específico.
* Reordered: Es un indicador binario que muestra si el producto se ha pedido previamente por el cliente.

**Orders:** Esta es una tabla con varias columnas donde como id solo tenemos dos columnas ‘order\_id y ‘user\_id’, las otras columnas son las siguientes:

* Eval\_set: Indica a qué conjunto de datos pertenece el pedido ("prior", "train" o "test").
* Order\_number: Representa el número de pedido para un cliente específico. Este número aumenta con cada nuevo pedido realizado por el mismo cliente.
* Order\_dow: Indica el día de la semana en que se realizó el pedido (0 para domingo, 1 para lunes, etc.).
* Order\_hour\_of\_day: Indica la hora del día en que se realizó el pedido.
* Day\_since\_prior\_order: Representa el número de días desde el último pedido del mismo cliente. El campo está vacío para el primer pedido de cada cliente

**Sample\_submission:** Esta es una tabla no contiene datos reales, es utilizada para competiciones de modelados predictivos.

****

*Gráfico 1. Descripción de la base de datos (Srinidhi Karjol, 2021).*

**2.2 Explore data**

Para abordar el análisis descriptivo de la base de datos de Instacart disponible en Kaggle, se llevó a cabo un examen de las características clave del conjunto de datos. Este análisis se realizó con el propósito de responder a los objetivos planteados, proporcionar información precisa y confiable sobre el comportamiento de los usuarios en la plataforma de Instacart.

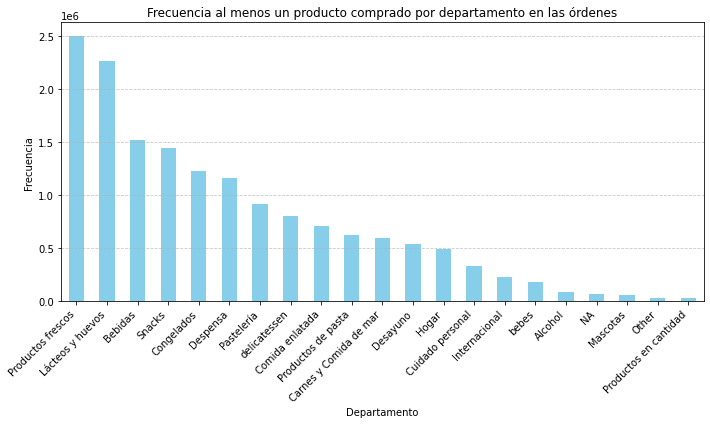
El análisis se ejecutó sobre las tablas de ‘departments’, ’order\_products\_\*’, ’orders’ y ‘products’. Nos enfocamos en explorar la diversidad de productos en cada departamento, ya que esto podría ayudarnos a identificar los departamentos con una mayor cantidad de productos, posiblemente indicando una mayor demanda en esos departamentos. Sin embargo, en el Gráfico 2 observamos que una amplia variedad de productos en un departamento no necesariamente implica que sea el departamento con más compras en las órdenes, se puede corroborar más adelante.

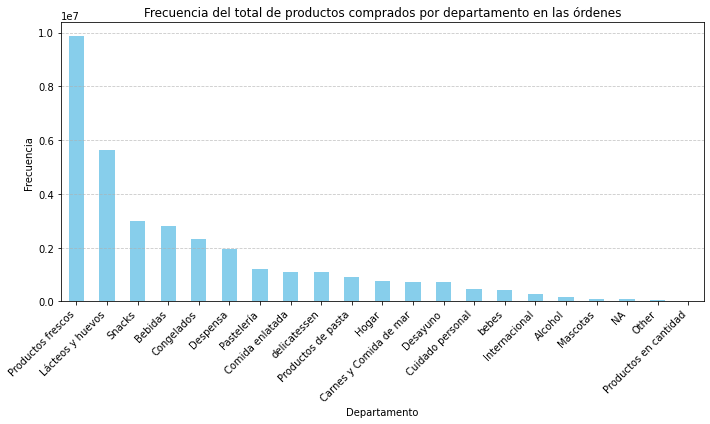
Gráfico, Histograma

Descripción generada automáticamente

*Gráfico 2. Cantidad de productos por departamento.*

También pudimos examinar qué departamentos contenían la gran mayoría de productos comprados. Para ello, realizamos un conteo de cuántas veces se compró al menos un ítem de un departamento en el total de órdenes. Además, observamos cuántas compras en total se realizaron por departamento en el conjunto total de órdenes, lo que se reflejó en el Gráfico 3. La distinción entre estos dos gráficos radica en observar qué tan populares son los departamentos en las compras, independientemente o no de la cantidad de productos del mismo departamento en una sola orden.





*Gráfico 3. Frecuencia de compra de cada departamento en todas las ordenes.*

Es importante destacar que, aunque se generaron gráficos y tablas para respaldar el análisis, no se incluyen todos en este informe introductorio, ya que el enfoque se centra en los resultados clave y las conclusiones más relevantes. Más adelante se verán más gráficos adaptados a la propuesta.

**2.3 Verify data quality**

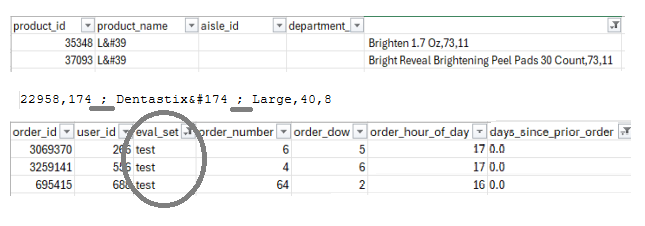
El proceso de verificar la calidad de los datos es fundamental en cualquier análisis o investigación que se base en información cuantitativa. En este contexto, es imprescindible identificar y abordar posibles datos erróneos o falsos. Por consiguiente, se analizarán los datos que podrían afectar nuestro modelo:

**Test data**: En la tabla de orders se puede observar que contiene datos falsos en la columna de eval\_set por lo que es importante tenerlo en cuenta a la hora de usar esta información para nuestro modelo. Estos datos están demarcados por la palabra “Test”.

**Productos mal etiquetados**: Hay que observar algunos productos tienen caracteres especiales que al momento de leerlo pueden lanzar un error, como por ejemplo: si un objeto tiene ‘;’ dentro del nombre del producto puede ser visto como un delimitador, así mismo hay productos que tienen ‘,’ dentro del nombre del producto.

**Day since prior order:** Más que simplemente una cuestión de calidad de datos es crucial reconocer que cuando un cliente es nuevo en la tienda, la columna 'Day\_since\_prior\_order' obtiene el valor de N/A. Por lo tanto, si deseamos utilizar esta columna para un análisis, es fundamental tener en cuenta esta particularidad.

A continuación, en la imagen 1 podemos observar algunas de las particularidades mencionadas anteriormente.



*Imagen 1. Errores encontrados.*

**3. Data preparation**

**3.1 Clean data**

El proceso de garantizar la calidad de los datos es fundamental para cualquier análisis. En el apartado de verificación de la calidad de los datos, se llevó a cabo un análisis para identificar posibles problemas y deficiencias en los datos recopilados. Este análisis dio lugar a una serie de procesos de limpieza de datos, los cuales se realizaron con el objetivo de remover, o estudiar los datos según fuera necesario para mejorar su calidad y confiabilidad. Estos procesos se llevaron a cabo, asegurando que cada acción realizada estuviera respaldada por una sólida fundamentación.

**Test data**: Con el propósito de obtener solo datos de ordenes reales, optamos por eliminar todas las ordenes que tengan como valor **test** en la columna ‘eval\_set’.

La limpieza de datos que se describe a continuación está centrada en el objetivo principal de nuestro proyecto: establecer una tienda con un enfoque saludable. Uno de los departamentos más saludable y principal es 'produce' (Betterhealth, 2022). Sin embargo, también buscamos identificar qué otros departamentos están estrechamente relacionados con este. En el cuarto gráfico, se presenta el porcentaje de productos comprados de cada departamento en relación con las compras de productos de 'produce'. Por ejemplo, se observa que el 45% de las compras de 'Alcohol' ocurren cuando dentro de la orden se compra al menos un producto de 'produce'.

Gráfico, Gráfico de barras

Descripción generada automáticamente

*Gráfico 4. Frecuencia de compra de productos de otros departamentos cuando se compra ‘produce’.*

Se puede observar que la mayoría de los productos se adquieren junto con los productos de la sección de 'produce', lo que sugiere una alta compatibilidad entre 'produce' y prácticamente todos los departamentos. Este hecho constituye un indicador muy positivo para considerar a 'produce' como producto principal para la venta. Sin embargo, dado que actualmente estamos enfocados en la creación de una tienda que únicamente ofrece productos saludables, como frutas y verduras, hemos optado por realizar una depuración de los departamentos y mantener únicamente 'produce'. No obstante, dejamos abierta la posibilidad de ampliar nuestra oferta de productos en otros departamentos en el futuro.

**3.2 Construct data**

En esta sección, exploraremos cómo hemos transformado nuestros datos existentes en nuestra base de datos final. Estas modificaciones son esenciales para mejorar la precisión y utilidad de nuestros modelos. Hay que tener en cuenta las eliminaciones que ya se hicieron en el punto anterior.

**Unión de tablas**: Las tablas de ‘order\_products\_train’ y ‘order\_products\_prior’ contienen información real por lo que se decidió manejarla como una sola tabla, haciendo una concatenación entre ellas.

**Day\_since\_prior\_order**: Esta columna contiene valores vacíos por lo que para futuros análisis y preservación de toda la columna como entero se opta por dejar estos valores en -1.

**3.3 Dataset description**

El conjunto de datos final para realizar el modelo de negocio es el siguiente:

**Departments**: Esta tabla solo contendrá el departamento con el que se quiere abrir la nueva tienda: ‘Produce’.

**Order\_products:** Esta tabla es la unificada entre prior y train, además solo contendrá las órdenes que contengan al menos un producto del departamento escogido.

**Orders:** Esta tabla ya no tiene las filas que contengan datos Test y además se filtraran los datos por la ordenes que estén en order\_products

**Products:** Esta tabla al igual que las otras se filtrarán los productos por los departamentos deseados.

**4. Modeling**

**4.1 Select modeling techniques**

Para modelar los datos de la base de datos de Instacart Kaggle, se empleará el algoritmo de reglas Apriori. Esta técnica es idónea para identificar patrones de asociación entre productos, lo cual es fundamental para comprender las preferencias y hábitos de compra de los clientes. Dado que se busca crear una tienda basada en productos de 'produce' la aplicación de reglas Apriori resulta apropiada, ya que permite analizar la frecuencia de compra de estos productos y descubrir relaciones de compra comunes entre ellos. Asimismo, esta técnica cumple con el supuesto de que los productos adquiridos conjuntamente con frecuencia reflejan preferencias o necesidades similares por parte de los clientes, lo que facilita la toma de decisiones sobre el surtido de productos a ofrecer en la tienda. Por lo tanto, la elección de utilizar reglas Apriori está fundamentada en su capacidad para revelar patrones de compra significativos en la base de datos de Instacart Kaggle, lo que contribuirá a la eficacia del proceso de toma de decisiones en el desarrollo de la tienda.

**4.2 Generate test design**

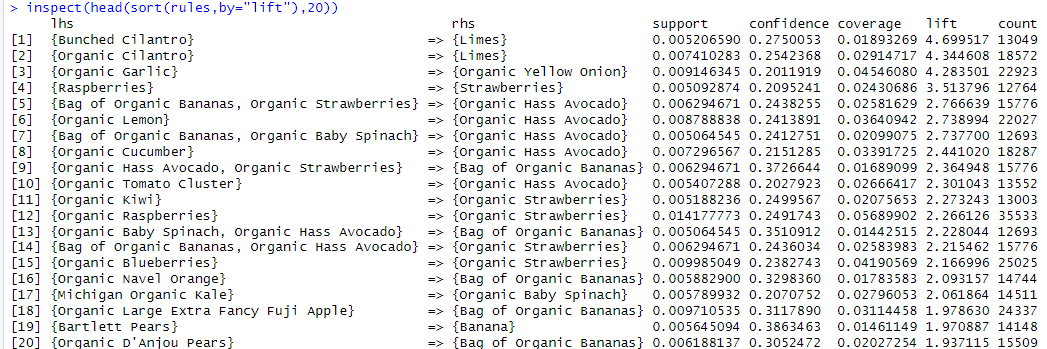
Durante la aplicación de la técnica de reglas a priori en una base de datos extensa, es posible obtener reglas inesperadas o conocidas como reglas espurias. Estas últimas hacen referencia a relaciones entre productos que, si bien pueden parecer fuertes en términos numéricos, carecen de una base sólida en la relación causal.

Al ordenar por lift hay algunas reglas que llegan o superan un lift de 1000, podría sugerirse una conexión perfecta entre esos productos; sin embargo, este valor numérico carece de significado real. Además, es importante destacar que el support de esta asociación es notablemente bajo.  
  
Para solucionar estas reglas sin sentido hemos llegado a 3 conclusiones que tienen sus pros y contras, pero aun así resultan mejor que obtener estos resultados sin sentido:

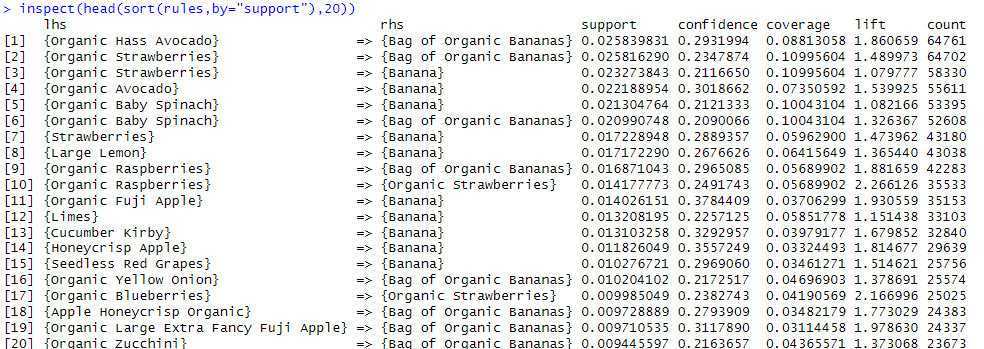
* **Muestreo aleatorio de la base de datos:** En este escenario, es crucial considerar que seleccionar un porcentaje muy pequeño de la base de datos podría potencialmente distorsionar los resultados y las reglas esperadas. Por otro lado, un muestreo adecuado nos permitirá capturar las reglas importantes y obtener valores más apropiados para las reglas al disponer de una base de datos más representativa. Aunque en los resultados no vemos una mejora significativa para estas reglas.
* **Reducción de productos en la regla:** En esta situación, es importante tener en cuenta que al asignar un número menor de productos a una regla, mejoraremos significativamente las reglas existentes. Sin embargo, también existe el riesgo de perder posibles reglas relevantes.
* **Reducción de confianza y aumento del support:** Al reducir la confianza, también observamos una mejora significativa en el lift, pero esto implica una disminución en la confianza de las reglas. Esto indica una probabilidad baja de que ocurra la compra de un producto dado otro. Lo cual es una de las consecuencias de manejar una gran base de datos y cantidad de productos. Incrementar el support también contribuye a reducir los lifts extremadamente altos. Sin embargo, solicitar un support demasiado alto en esta base de datos puede resultar en la ausencia de cualquier regla.

**4.3 Build model**

Por último, observando las conclusiones mencionadas en la anterior sección se decidió por usar un support de al menos un 0.5%, una confianza de 0.2, un máximo de productos por regla de 3 y la base de datos de produce completa sin muestreo. Dándonos un resultado de 64 reglas, en el siguiente gráfico se puede observar las reglas más importantes ordenada por support y lift.



*Gráfico 5. Reglas ordenadas por lift.*

**

*Gráfico 6. Reglas ordenadas por support.*

**5. Evaluation**

**5.1 Assess model**

Para introducir el análisis de los resultados y evaluación del modelo Apriori en el contexto de la minería de datos, es fundamental comprender su desempeño, robustez y calidad general. En este sentido, los informes detallados que presentaremos proporcionarán una evaluación exhaustiva del rendimiento del modelo Apriori, incluyendo métricas clave como el soporte, la confianza y la lift representadas en gráficas para cada regla de asociación descubierta. Además, compararemos estos resultados con los objetivos previamente establecidos de minería de datos y los Indicadores Clave de Desempeño (KPI) definidos para el proyecto. Este análisis nos permitirá identificar las reglas de asociación más relevantes y significativas generadas por el modelo, así como evidenciar su impacto en el logro de los objetivos comerciales. A continuación, presentaremos una serie de gráficos que complementarán esta evaluación, ofreciendo una representación visual de los patrones y relaciones descubiertos por el modelo Apriori.

Comenzaremos nuestro análisis con un scatter plot que representa las 64 reglas de asociación generadas por el modelo Apriori. Este gráfico nos permitirá visualizar de manera clara y concisa las reglas que se distinguen por tener un alto soporte y lift en comparación con otras.

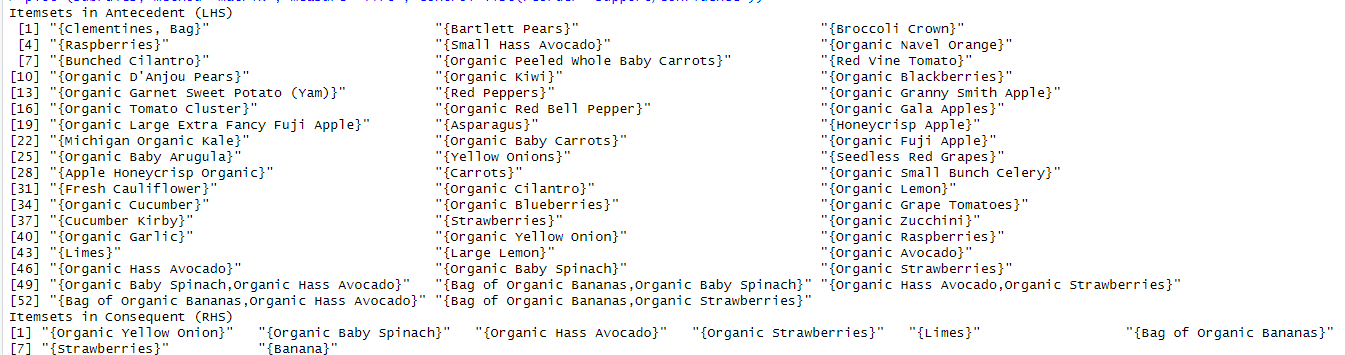
**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

*Gráfico 7. Scatter plot de las reglas.*

En el gráfico, podemos identificar ciertas reglas que exhiben un confidence relativamente alto en comparación con otras. Estas reglas, en su mayoría, muestran un lift significativo junto con un soporte que no supera el 2.5%. Este hallazgo era previsible, dado el gran volumen de productos y pedidos asociados al departamento 'produce'. La presencia de estas reglas destacadas sugiere una relación robusta entre los elementos y ofrece una valiosa perspectiva sobre las características distintivas de este departamento en particular. Ahora observaremos los productos que están en todas las reglas.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

*Gráfico 8. Scatter plot de las reglas.*

Observamos que hay 8 productos que se presentan en todas las reglas de asociación, entre los cuales destacan la 'banana' y el 'bag of organic bananas' como los productos más recurrentes. Además, se destaca la presencia de ciertos productos que muestran un lift excepcional con otros, como es el caso del 'lime' con el 'bunched cilantro'. Estos hallazgos resaltan la importancia y la influencia significativa de ciertos productos dentro del conjunto de datos, lo que sugiere oportunidades estratégicas para la gestión de inventario y la promoción de ventas cruzadas.

**5.2 Produce final report**

Este informe gerencial tiene como objetivo presentar estrategias sugeridas para los productos identificados como clave para un inicio exitoso en las ventas y la apertura de la tienda. A continuación, detallaremos aspectos importantes a tener en cuenta para lograr resultados óptimos y cumplir con los objetivos de negocio establecidos.

* El producto con mayor apoyo son las bananas y el paquete de bananas orgánicas. Además, se observó que este producto tiene un alto volumen de compra junto con otros productos, lo que lo convierte en una opción llamativa para el logo o la fruta insignia de la tienda.
* En el comienzo de la tienda se recomienda no empezar con una gran cantidad de productos ya que hay muchos productos que no tienen un buen support o no son comprados con mucha frecuencia, se puede empezar con los productos manejados en las reglas, ya que se tiene una mejor información de estos.
* Para la venta mediante la aplicación, se recomienda tener en cuenta la información proporcionada en esta gráfica, que indica actividad de compra por días en el departamento ‘produce’.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Se observa que los días con menos ventas son de martes a viernes. Por lo tanto, se sugiere aplicar descuentos en la aplicación durante estos días para incentivar las compras y aumentar la actividad en estos períodos.

* En la aplicación móvil, es crucial prestar atención a los lifts de las reglas propuestas, ya que pueden representar recomendaciones que es probable que el cliente desee adquirir. En la tienda física, se pueden colocar los productos con lifts muy altos cerca, para facilitar o incentivar su compra.
* Para aquellos productos con un soporte muy bajo, se sugiere agruparlos con las compras de bananas para evitar su caducidad, o bien, aprovechar la oportunidad para dar a conocer el producto.
* Por último se recomienda ofrecer promociones variadas con los productos que tienen mayor soporte ya que puede aumentar la clientela al ser productos que se consumen más seguido que los demás.

**Bibliografía**

**1.** Kotler-Keller, (2015) Marketing Management. Pearson.  
**2.** Medium. (2021). [Descripción de la base de datos]. Recuperado de <https://medium.com/geekculture/kaggle-instacart-market-basket-analysis-8bfbbf5f2efb>  
**3.** Betterhealth. (2022). [Fruit and vegetables] Recuperado de <https://www.betterhealth.vic.gov.au/health/healthyliving/fruit-and-vegetables>  
**4.** Menifield CE, PhD., Doty N, M.P.H., Fletcher, Audwin, PhD,A.P.R.N., C.F.N.P.-B.C. Obesity in America. Abnf J. 2008;19(3):83-8. https://www.proquest.com/scholarly-journals/obesity-america/docview/218860194/se-2

​​**5.** Barbalova, I., & Cazin, N. (2024, January ). Passport. Retrieved from Megatrends: Wellness - Mapping Strategic Priorities in Health, Beauty and Fashion.

​**6.** FDA. (2022, 02 29). FDA. Retrieved from Nutrient Content Claims; Definition of Term “Healthy” (Proposed Rule) Regulatory Impact Analysis: https://www.fda.gov/about-fda/economic-impact-analyses-fda-regulations/nutrient-content-claims-definition-term-healthy-proposed-rule-regulatory-impact-analysis#:~:text=FDA%20is%20proposing%20to%20redefine%20the%20implied%20nutrient,the%20criteria%20define

​**7.** FMI. (2017). The Food Indtry Association. Retrieved from US Grocery Shopper Trends 2017: https://www.fmi.org/forms/store/ProductFormPublic/u-s-grocery-shopper-trends-2017

​**8.** Food and Agriculture Organization of the United Nations. (2023). Food and Agriculture Organization of the United Nations. Retrieved from TECHNOLOGY AND INNOVATION: A KEY ENABLER FOR AGRIFOOD SYSTEMS TRANSFORMATION UNDER URBANIZATION: https://www.fao.org/3/cc3017en/online/state-food-security-and-nutrition-2023/technology-and-innovation.html

​**9.** Harvard University. (2023, January ). Harvard T.H. Chan School of Public Health. Retrieved from Healthy Eating Plate: https://www.hsph.harvard.edu/nutritionsource/healthy-eating-plate/

​**10.** Herforth, e. a. (2022). Our World in Data. Retrieved from Daily cost of a healthy diet: https://ourworldindata.org/grapher/cost-healthy-diet?time=2021&country=~USA

​**11.** OFR NARA; GPO. (2024, 4 3). Implied nutrient content claims and related label statements. Retrieved from Code of Federal Regulation: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-101/subpart-D/section-101.65

​**12.** Ritchie, H. (2021, July 12). Our World in Data. Retrieved from Three billion people cannot afford a healthy diet: https://ourworldindata.org/diet-affordability

​**13.** Ritchie, H., Rosado, P., & Roser, M. (2023, July). Our World in Data. Retrieved from Food Prices: https://ourworldindata.org/food-prices

​**14.** ROHU4U. (2023). Kaggle. Retrieved from Rohan Instacart: https://www.kaggle.com/code/rohu4u/rohan-instacart