



Facultad de Ingeniería
Métodos y aplicaciones de analítica I

Taller Grupal 1 – Primer Semestre 2022

Taller 1 grupal

Análisis Market Basket

Laura Juliana Mora , Ricardo Macias Bohorquez, Johan Sebastian Muñoz

Ana María Beltrán Cortés

*Estudiante de Mestría en Analítica de Datos para la Inteligencia de Negocios
Profesor, Departamento de Ingeniería Industrial
Pontificia Universidad Javeriana, Bogotá, Colombia*

1. ENTENDIMIENTO DEL NEGOCIO

1.1 Situación del Negocio

Tras dos años de pandemia, el sector de supermercados y comestibles en Estados Unidos, ha tenido un aumento del comercio por internet (e-commerce), a través de sus diferentes plataformas y en las mismas tiendas; esta industria para el año 2018 estaba facturando alrededor de 701.188 millones de dólares [1], mientras que para el 2020 estuvo facturando alrededor de 6,2 billones de dólares; sin embargo, a nivel de locales con atención presencial si se vio una disminución, y para esta fecha se tienen aproximadamente 63 mil lugares [2].

Entre los principales supermercados se destacan Walmart, Kroger y Walgreens [2], a pesar que el Estadounidense promedio, prefiere realizar su mercado directamente en lugares físicos, hoy en día se presenta la entrada de nuevos competidores, permitiendo que a través de sus aplicaciones, los usuarios puedan hacer su mercado desde la comodidad de su hogar o su carro y les llegue la comida a estos lugares, facilitando la búsqueda de los productos para los clientes [3]; sin embargo, compañías como Walmart se mantienen a la vanguardia y han venido implementando estos servicios con el paso del tiempo. Adicionalmente, el gobierno de Estados Unidos tras la pandemia, como una iniciativa del departamento de agricultura creó farmers.gov, con el fin de promover la participación de pequeños y grandes productores agrícolas, los cuales se han visto afectados por la pandemia. Uno de los servicios que se presentan a través de este portal, es que los negocios pueden contactar con los agricultores de la región, con el fin de abastecer sus locales, directamente del sitio de producción [4].

Para el año 2019 el sector agrícola de Estados Unidos se posicionaba como uno de los más importantes a nivel mundial; adicionalmente, en época pre pandemia se estimaba que para el 2050 el consumo de alimentos incrementaría en un 60% [5]. Como se

mencionó anteriormente el gobierno de los Estados Unidos tras los últimos años ha venido creando nuevos incentivos, con el fin de apoyar el sector, una de los más destacables es la inserción a mujeres, veteranos de guerra, esto con el fin de tener nuevos agricultores, ya que los antiguos agricultores están envejeciendo, y el país por su alta demanda necesita una nueva generación en este sector [6]. Uno de los proveedores para los agricultores es Acadian®, en razón a que es un producto enfocado en aumentar la calidad y el rendimiento de la fruta, el crecimiento de las raíces, el desarrollo temprano de la planta y ayudan en la resistencia al estrés y la recuperación. [7] Finalmente, el estado, recientemente realizó una encuesta donde se menciona que existían para el 2020 alrededor de 2,02 millones de granjas en EE. UU, con 897 millones de acres de tierra en granjas en 2020, el tamaño promedio de las granjas fue de 444 acres, solo un poco más que los 440 acres registrados en principios de la década de 1970.[8] Adicionalmente, el estado estadounidense también ha creado una guía para sus ciudadanos, con el fin de incentivar la alimentación saludable a través del consumo de alimentos provenientes del sector agrícola como lo son las frutas y las verduras, incluyéndolas en su alimentación diaria a través de diferentes menús [9]. Además con el paso de los años se ha incrementado el consumo de Yogurt en este país, debido a sus componentes nutritivos y a la amplia cantidad de presentaciones que maneja, ya que no necesariamente debe ser derivado de los lácteos, sino que se presentan también como derivados de los frutos secos aportando a una mejor nutrición de sus consumidores [10].

1.2 Objetivos del Negocio

Como consultores de una empresa de yogures, formada en el sector agrícola en USA, y observando el reciente incentivo de consumo de alimentos saludables, se busca conocer con qué tipos de productos se pueden generar alianzas o creación de nuevos productos, según los gustos de nuestros clientes.

- Creación de nuevos productos con base en las preferencias de los consumidores.
- Encontrar alianzas entre productos, para creación de combos promocionales de venta presencial y/o virtual que permitan el crecimiento de las empresas asociadas.
- Entender los hábitos de compra de los usuarios, de manera general y luego particular para el negocio.

1.3 Objetivo Minería de Datos

Teniendo en cuenta el objetivo, como consultores se plantea a través de reglas de asociaciones, investigar cuales son los productos agrícolas, ganaderos y de otros campos se ven asociados con la compra de Yogurt en supermercados.

- Encontrar al menos de 5-10 asociaciones, con productos diferentes al yogurt que puedan desencadenar en nuevos productos o nuevas alianzas.
- Identificar y agrupar las diferentes presentaciones donde la base principal fuera el yogurt, para agruparlas con un mismo nombre, permitiéndonos tener un mejor análisis de las asociaciones.

2. ENTENDIMIENTO DE LOS DATOS

2.1 Descripción de los datos

Para la consultoría nos fueron entregados 7 archivos tipo csv que hacen parte de una base de datos relacional, con información relacionada con órdenes de compra, clientes, departamentos, productos y pasillos, realizadas a la empresa Instacard, sin embargo, se desconoce el proceso de recolección utilizado en estos datos; a continuación se realiza una descripción de cada uno de estos archivos:

- aisles.csv es un master data de pasillos, donde podemos identificar que existen 134 pasillos diferentes.(Ver tabla 1)

Nombre Variable	Tipo de datos	Llave primaria	Descripción
aisle_id	int	Si	Id único para identificar el pasillo
aisle	chr	No	Nombre del pasillo

Tabla 1. Estructura base aisles

- departments.csv es un master data de departamentos, en total existen 21 departamentos diferentes.(Ver tabla 2)

Nombre Variable	Tipo de datos	Llave primaria	Descripción
department_id	int	Si	Id único para identificar el departamento
department	chr	No	Nombre del departamento

Tabla 2. Estructura base departments

- products.csv es un master data de producto, que no solo nos dice que tiene 49.688 productos diferentes, además que nos indica en qué pasillo están y a qué departamento pertenecen.(Ver tabla 3)

Nombre Variable	Tipo de datos	Llave primaria	Descripción
product_id	int	Si	Id único para identificar el producto
product_name	chr	No	Nombre del producto
aisle_id	int	No	Id único para identificar el pasillo
department_id	int	No	Id único para identificar el departamento

Tabla 3. Estructura base products

- orders.csv presenta datos generales de cada una de las órdenes de compra, el total de órdenes es de 3.421.083, de las cuales solo tenemos información de que productos hay en cada orden para (3.421.083), esto debido a que la tabla test no tiene los datos en el formato correcto en 75.000 órdenes , continuando, encontramos que cada orden está asociada a un cliente, tenemos aquí también el nombre del csv donde se encuentra la información de que productos fueron

llevados en esa orden, además tenemos el día de la semana en que se hizo, hora del día y días desde la orden anterior a la que se esté revisando.(Ver tabla 4)

Nombre Variable	Tipo de datos	Llave primaria	Descripción
order_id	int	Si	Id único para identificar la orden
user_id	int	No	Id único del cliente
eval_set	chr	No	Tabla donde se encuentra más información de la orden
order_number	int	No	Numero de orden del cliente
order_dow	int	No	Día de la semana en que se hizo la orden
order_hour_of_day	int	No	Hora en que se hizo la orden
days_since_prior_order	chr	No	Días desde la última orden del cliente respecto a la orden en cuestión.

Tabla 4. Estructura base orders

- order_products_prior.csv, order_products_train.csv y sample_submission.csv .Nos dan información de qué productos y la cantidad de unidades de los mismos, que se llevó en cada orden. Es importante tener en cuenta que la división de los mismos es porque normalmente se entregan de esta manera para las competencias de kaggle. El de prior tiene datos anteriores de órdenes respecto a la última de los usuarios (32.434.489 registros), train es la información de la última orden de los usuarios (1.384.617 registros) y sample_submission es para pruebas de aprendizaje automático (75.000 registros), que en este caso esta última no es relevante.(Ver tabla 5)

Nombre Variable	Tipo de datos	Llave primaria	Descripción
order_id	int	Si	Id único para identificar la orden
product_id	int	No	Id único para identificar el producto
add_to_cart_order	chr	No	Cantidad de unidades el producto en cuestión
reordered	int	No	Si el usuario ha reordenado el producto en otras órdenes.

Tabla 5. Estructura bases prior y train

2.2 Exploración de los datos

En el proceso de exploración encontramos los siguientes descubrimientos:

- El departamento que está en la mayor cantidad de órdenes es “produce” con casi un 16% del total, aunque solo tiene menos del 5% del total de productos distintos, esto se puede deber a que justo los productos dentro de este departamento, son cosas de uso y consumo constante (fresh vegetables,fresh fruits). Por otro lado, “personal care”, es el departamento que tiene más productos distintos con un 13%

del total, pero solo está en el 2% de las órdenes, ya que son productos que no todo el tiempo o no en todas las compras se lleva (soap, hair care). (Ver imagen 1)

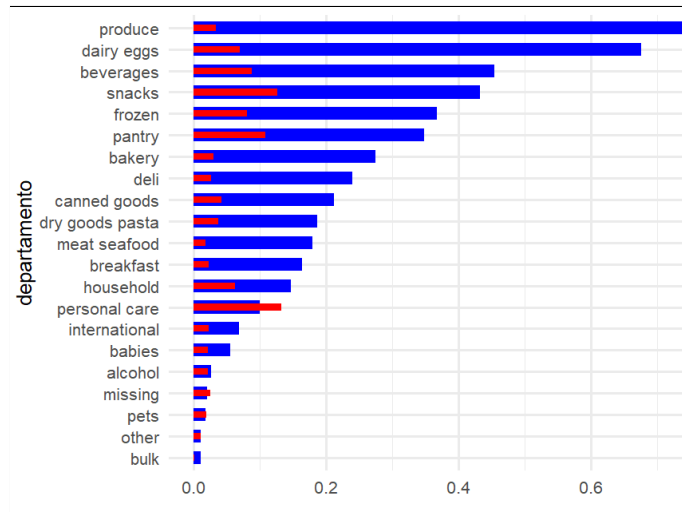


Imagen 1. Porcentaje del total productos y órdenes por departamento

- Dentro del top 10 de productos, medido por la cantidad de órdenes en las que se piden, encontramos que 9 de los 10 hacen parte del departamento “produce” y el otro es de “dairy eggs”. Lo que va muy alineado con el análisis anterior. (Ver imagen 2)

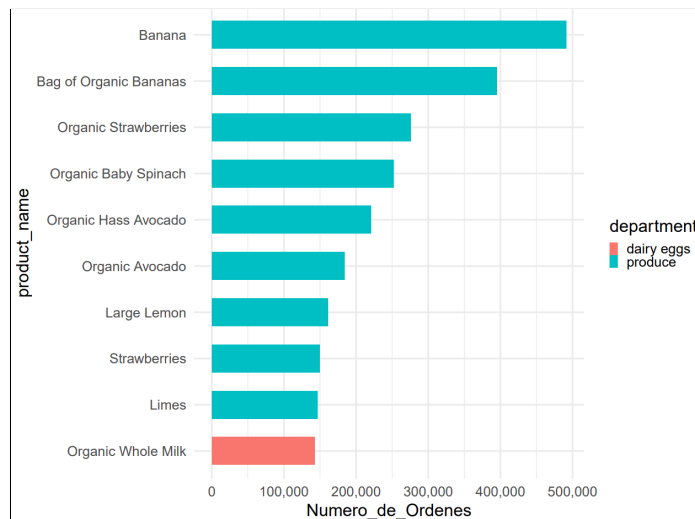


Imagen 2. Top 10 Productos

- Haciendo un análisis de cuál es el producto que más se compra para cada departamento encontramos. (Ver imagen 3)

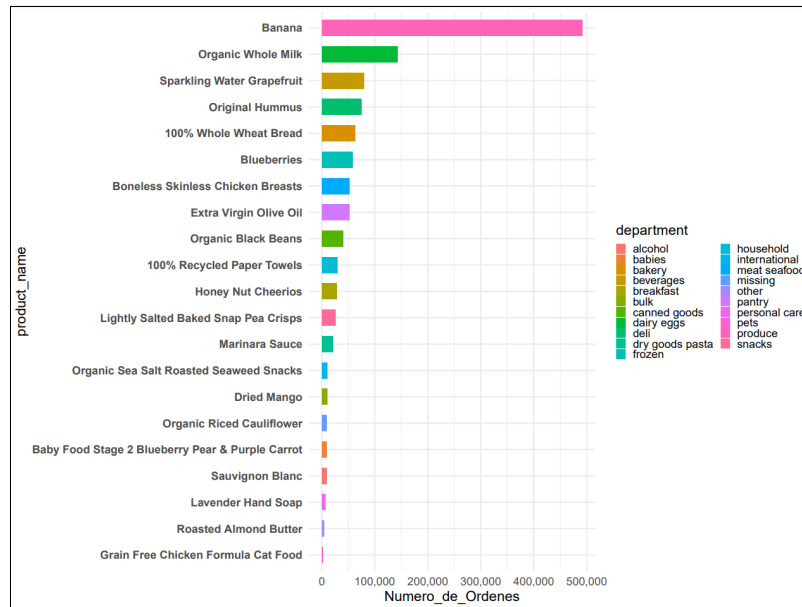


Imagen 4. Producto Top por Departamento

- Pasando a los clientes, recalcamos que todos hacen por lo menos de 4 pedidos. Por otro lado, el número de órdenes por un cliente más grande es de 100 órdenes. El 81% de los clientes tienen entre 4 y 25 órdenes en total.

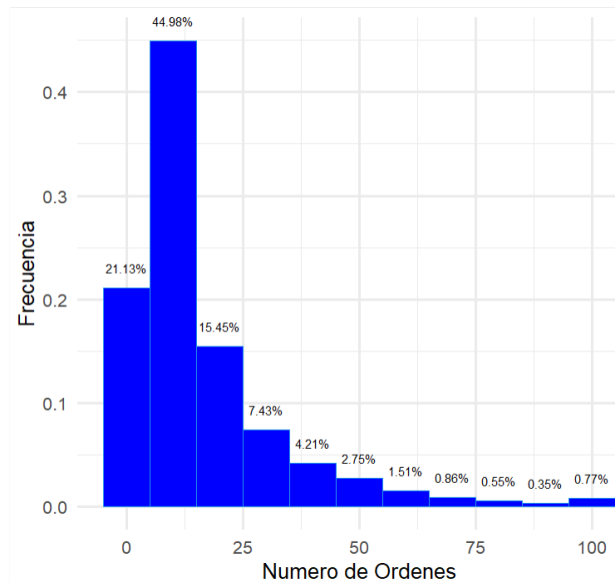


Imagen 5. Porcentaje Frecuencia de compra

- El día con mayores órdenes es el 0 con más de 600.905, y el que menos es el día 4 con 436.339, lamentablemente no tenemos claro cada número a qué día de la semana hace referencia (Lunes, Martes, Miércoles, etc). **(Ver Imagen 6)**



Imagen 6. Número de Órdenes por día de la semana

2.3 Validación de los datos

Tras realizar el análisis de la exploración de los datos y el entendimiento de los mismos, se puede resaltar que la gran mayoría de los datos en los diferentes sets de datos son categóricos, adicionalmente solo se presenta una falta de datos en `sample_submission.csv` ya que como se mencionó anteriormente este set de datos está enfocado para realizar diferentes predicciones de los datos faltantes. Finalmente se destacan clientes que tienen 0 órdenes, sin embargo cabe resaltar que la base de datos final tiene buena calidad para el desarrollo de los objetivos, aunque se asume que algunos datos pueden contener errores como de escritura, sin embargo pueden ser minimizados por el peso de los demás.

3. PREPARACIÓN DE LOS DATOS

Como paso inicial, para la preparación de datos, se realizan diferentes uniones entre los sets de datos con el fin de obtener un solo conjunto de datos en una sola tabla; se inició realizando un `merge` entre la tabla donde están los pasillos y los productos, tomando como referencia el `id` de los pasillos, con el fin de saber en cuál pasillo se encuentra cada producto, posteriormente la tabla resultante de la unión anterior, se une con la de departamento y así se conoce a cuál departamento pertenece cada producto con su pasillo respectivo, ya no solo por el `id` del departamento sino también por su nombre. Por otro lado, se juntan los datos del set de `"order_products__prior"` con el de `"order_products__train"`, de esta forma se crea la segunda tabla resultante, que finalmente se une a la tabla resultante inicial (donde se encuentran los datos de pasillos, productos, departamentos), para esto se utiliza el `id` de productos.

Posteriormente como el enfoque de la empresa es el Yogurt, se crea un set de datos con todos los datos de productos asociados al Yogurt existentes en el sistema, este nuevo dataset se une a la tabla principal donde ya se encuentran todas las órdenes y sus productos; luego se pasa a colocar el nombre "Yogurt" a todos los productos que tienen en su nombre la palabra Yogurt, ya que para este ejercicio no nos interesa el tipo de Yogurt.

Antes de obtener las relaciones entre productos, se pasa a seleccionar los productos que se encuentren en los departamentos de interés, por la exploración de datos se sabe que

el departamento que tiene más presentaciones de Yogurt es el de dairy eggs y encontramos que este departamento tiene asociaciones relevantes para el ejercicio con los departamentos de frozen, pantry y breakfast, para esto se seleccionan todas las órdenes que tengan productos en estos departamentos. Se procede a la creación de una nueva variable “Orden_Producto” donde se une el nombre del departamento con el nombre del producto. Esta variable con la variable “Order_id” son transformadas a formato transaccional. A continuación se presenta la estructura con la que queda la tabla final. (Ver imagen 7)

```
> str(transacciones)
Formal class 'transactions' [package "arules"] with 3 slots
..@ data      :Formal class 'ngCMatrix' [package "Matrix"] with 5 slots
.. .. ..@ i      : int [1:19296912] 771 3190 7623 7836 8226 9045 13520 372 1784 2522 ...
.. .. ..@ p      : int [1:3083850] 0 7 16 22 34 35 43 54 56 62 ...
.. .. ..@ Dim     : int [1:2] 13570 3083849
.. .. ..@ Dimnames:List of 2
.. .. .. ..$ : NULL
.. .. .. ..$ : NULL
.. .. ..@ factors : list()
..@ itemInfo    :'data.frame': 13570 obs. of 1 variable:
.. ..$ labels: chr [1:13570] "& Go! Hazelnut Spread + Pretzel Sticks - pantry" "\\Im Pei-nut
Cookie & Brown Butter Candied Bacon Ice cream Sandwich - frozen" "\\Mies Vanilla Rohe\\" Ice
..@ itemsetInfo:'data.frame': 3083849 obs. of 1 variable:
.. ..$ transactionID: chr [1:3083849] "1" "2" "3" "5" ...
```

Imagen 7. Tabla transaccional que ingresa al modelo final

4. MODELACIÓN

En un primer momento se aplicó el algoritmo a priori, sin realizar ningún filtro, y enfocado al atributo de departamentos, donde para este caso se utilizó un support=0.1, confidence=0.5. A continuación en la imagen 8 se presentan las 15 primeras relaciones obtenidas con esta prueba.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{canned goods}	=> {pantry}	0.1092716	0.5144522	0.2124039	1.476974	365632
[2]	{bakery, dairy eggs, produce}	=> {frozen}	0.1017437	0.5224251	0.1947528	1.418792	340443
[3]	{dairy eggs, produce, snacks}	=> {frozen}	0.1404529	0.5191417	0.2705483	1.409875	469967
[4]	{dairy eggs, deli}	=> {frozen}	0.1001553	0.5181956	0.1932770	1.407305	335128
[5]	{dairy eggs, deli, produce}	=> {snacks}	0.1044364	0.6072501	0.1719826	1.402527	349453
[6]	{beverages, frozen}	=> {snacks}	0.1127859	0.6025867	0.1871696	1.391756	377391
[7]	{bakery, dairy eggs}	=> {frozen}	0.1151020	0.5112833	0.2251238	1.388533	385141
[8]	{dairy eggs, deli}	=> {snacks}	0.1160960	0.6006717	0.1932770	1.387333	388467
[9]	{dairy eggs, pantry, produce}	=> {frozen}	0.1179923	0.5067331	0.2328490	1.376176	394812
[10]	{deli, produce}	=> {frozen}	0.1038408	0.5034295	0.2062668	1.367204	347460
[11]	{beverages, pantry}	=> {snacks}	0.1003863	0.5896103	0.1702588	1.361785	335901
[12]	{bakery, produce}	=> {frozen}	0.1146887	0.5006066	0.2290995	1.359537	383758
[13]	{canned goods}	=> {frozen}	0.1062421	0.5001892	0.2124039	1.358404	355495
[14]	{deli, produce}	=> {snacks}	0.1205622	0.5844961	0.2062668	1.349973	403411
[15]	{bakery, dairy eggs, produce}	=> {snacks}	0.1128427	0.5794150	0.1947528	1.338238	377581

Imagen 8. Tabla de asociaciones por departamento

Otra de las pruebas iniciales fue enfocada a los productos, donde se utilizó el filtro mencionado anteriormente en la sección de preparación, y el soporte era muy bajo y la confianza como mínimo obtenía un valor del 70%, para este caso como se muestra en la imagen 9 se obtienen valores altísimos en el lift como por ejemplo el caso del lift superior a 20.000.



Imagen 9. Lift de 1 prueba asociaciones por producto

Más adelante, se empezaron a realizar diferentes variaciones de estos atributos, y enfocadas a las relaciones entre productos, donde en su mayoría empezaron a aparecer diferentes reglas ligadas a las bananas, como se puede observar en la siguiente imagen, cabe resaltar que para estos casos ya se destacaban lifts inferiores a los de la segunda prueba.

rules	support	confidence	coverage	lift	count
{Banana - produce,Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs} => {Total 2% Lowfat Greek Strained Yogurt With Blueberry - dairy eggs}	0.00092911	0.31609006	0.00293938	45.7425216	2878
{Banana - produce,Total 2% Lowfat Greek Strained Yogurt With Blueberry - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00092911	0.4522313	0.0020545	45.384208	2878
{Banana - produce,Total 2% Lowfat Greek Strained Yogurt with Peach - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00080127	0.41222388	0.00194377	41.3692162	2482
{Banana - produce,Non Fat Raspberry Yogurt - dairy eggs} => {Icelandic Style Skyr Blueberry Non-fat Yogurt - dairy eggs}	0.00064986	0.48331333	0.00134459	74.9979189	2013
{Banana - produce,Icelandic Style Skyr Blueberry Non-fat Yogurt - dairy eggs} => {Non Fat Raspberry Yogurt - dairy eggs}	0.00064986	0.38009819	0.00170971	69.4502368	2013
{Banana - produce,Total 2% Lowfat Greek Strained Yogurt with Peach - dairy eggs} => {Total 2% Lowfat Greek Strained Yogurt With Blueberry - dairy eggs}	0.00061402	0.31589437	0.00194377	45.7142025	1902
{Banana - produce,Vanilla Skyr Nonfat Yogurt - dairy eggs} => {Icelandic Style Skyr Blueberry Non-fat Yogurt - dairy eggs}	0.00056399	0.37961756	0.00148567	58.9069767	1747
{Banana - produce,Icelandic Style Skyr Blueberry Non-fat Yogurt - dairy eggs} => {Vanilla Skyr Nonfat Yogurt - dairy eggs}	0.00056399	0.3298716	0.00170971	54.0667915	1747
{Banana - produce,Total 2% Greek Strained Yogurt with Cherry 5.3 oz - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00053881	0.3018629	0.00178493	30.2938094	1669
{Bag of Organic Bananas - produce,Total 2% Lowfat Greek Strained Yogurt With Blueberry - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00047585	0.4657188	0.00102176	46.7377619	1474
{Bag of Organic Bananas - produce,Total 2% Lowfat Greek Strained Yogurt with Peach - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00041678	0.40997142	0.0010166	41.1431675	1291
{Organic Strawberries - produce,Total 2% Lowfat Greek Strained Yogurt With Blueberry - dairy eggs} => {Total 2% with Strawberry Lowfat Greek Strained Yogurt - dairy eggs}	0.00041032	0.45506624	0.00090167	45.6687114	1271
{Banana - produce,Non Fat Raspberry Yogurt - dairy eggs} => {Vanilla Skyr Nonfat Yogurt - dairy eggs}	0.0004087	0.30396158	0.00134459	49.8200741	1266

Imagen 10. Prueba 2 productos de Bananas

Con el fin de conocer los productos relacionados con el Yogurt, se aplicó el algoritmo Apriori de reglas de asociación, enfocado en realizar Market Basket Analysis, con un support de 0.001 y una confianza de 0.3, esto debido a que existe una alta variedad de productos dentro de la base de datos, lo que hace necesario que estas variables tomen valores pequeños para mostrar más relaciones, adicionalmente se especifica que el producto Yogurt debe estar en el consecuente generando un total de 768 reglas.

```
> inspect(head(sort(itemsets_filtrado1,by="lift"),n=15))
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Organic Raspberries - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001308106	0.5714690	0.002289023	2.219577	4034
[2]	{Banana - produce, Organic Strawberries - produce, Organic Whole Milk - dairy eggs}	=> {Yogurt - dairy eggs}	0.001270490	0.5659396	0.002244922	2.198101	3918
[3]	{Organic Whole Milk - dairy eggs, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001667397	0.5649929	0.002951182	2.194424	5142
[4]	{Banana - produce, Organic Sticks Low Moisture Part Skim Mozzarella String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001440408	0.5637771	0.002554924	2.189702	4442
[5]	{Organic Strawberries - produce, Whipped Cream Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001204663	0.5610935	0.002146992	2.179279	3715
[6]	{Organic Sticks Low Moisture Part Skim Mozzarella String Cheese - dairy eggs, Organic Strawberries - produce}	=> {Yogurt - dairy eggs}	0.001310051	0.5609553	0.002335393	2.178742	4040
[7]	{Organic Strawberries - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.002544547	0.5593812	0.004548861	2.172628	7847
[8]	{Organic Hass Avocado - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001867147	0.5523261	0.003380516	2.145226	5758
[9]	{Organic Yokiids Very Berry Smoothies - dairy eggs}	=> {Yogurt - dairy eggs}	0.001359016	0.5523195	0.002460561	2.145200	4191
[10]	{Bag of Organic Bananas - produce, Organic Strawberries - produce, Organic Whole Milk - dairy eggs}	=> {Yogurt - dairy eggs}	0.001124244	0.5471043	0.002054900	2.124945	3467
[11]	{Organic Avocado - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.001046420	0.5449173	0.001920327	2.116450	3227
[12]	{Bag of Organic Bananas - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.002636640	0.5422474	0.004862430	2.106081	8131
[13]	{Organic Raspberries - produce, Organic Reduced Fat 2% Milk - dairy eggs}	=> {Yogurt - dairy eggs}	0.001011398	0.5421519	0.001865526	2.105710	3119
[14]	{Banana - produce, Organic Whole String Cheese - dairy eggs}	=> {Yogurt - dairy eggs}	0.002629830	0.5414608	0.004856917	2.103026	8110
[15]	{Banana - produce, Organic Fat Free Milk - dairy eggs}	=> {Yogurt - dairy eggs}	0.001173858	0.5359787	0.002190120	2.081733	3620

Imagen 11. Resultado final asociaciones

5. EVALUACIÓN

Al realizar la evaluación de las reglas de asociación de productos relacionados con el Yogurt, se destaca que la cantidad de órdenes donde este producto se encuentra, como se muestra en la imagen 12 a continuación, son 836.215.

```
> orders_yogurt %>%
+ summarize(count_distinct_orders = n_distinct(order_id))
# A tibble: 1 x 1
  count_distinct_orders
  <int>
1             836215
```

Imagen 12. Total órdenes en las que aparecen productos de yogurt

De igual manera se puede observar que el producto más vendido, tomando los 15 productos con mayor cantidad de órdenes, donde también se pide el Yogurt, son las Bananas, como se mencionó en la sección anterior, que es un producto con alta demanda; al igual que se observa que las compras de Yogurt, van muy ligadas a las compras de diferentes frutas y verduras. **(Ver imagen 13)**

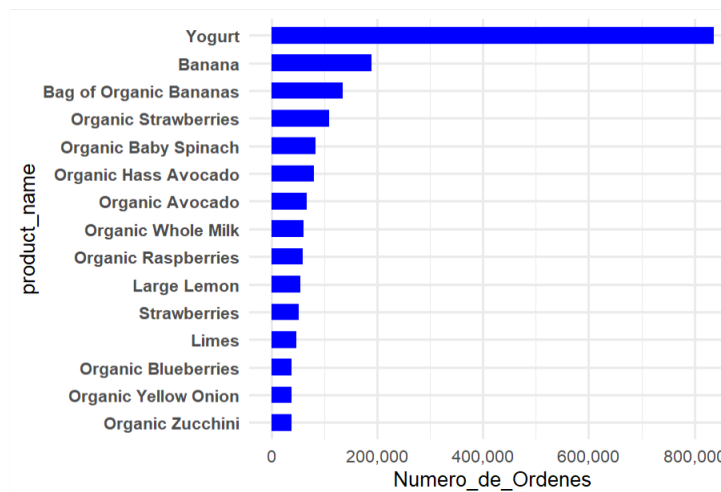


Imagen 13. Lista de 15 productos con la mayor cantidad de órdenes donde también se pide yogurt

Finalmente, de todas las compras, se obtiene que la cantidad de usuarios que consumen yogurt son, como se ve en la imagen 14, 125.178 compradores.

```
> merge_client_yogurt %>%  
+   summarize(count_distinct_users = n_distinct(user_id))  
# A tibble: 1 x 1  
  count_distinct_users  
      <int>  
1             125178
```

Imagen 14. Total usuarios que consumen yogurt

Adicionalmente, en la imagen 15 se muestra el resultado de las reglas obtenidas evidenciando a partir del lift junto con la confianza; en la parte superior izquierda del gráfico y con un rojo más intenso se encuentran las reglas que presentan un alto lift y a su vez una alta confianza, en contraparte de las que se encuentran en la parte inferior izquierda con un color casi blanco lo cual significa que presentan una confianza baja y un lift bajo.

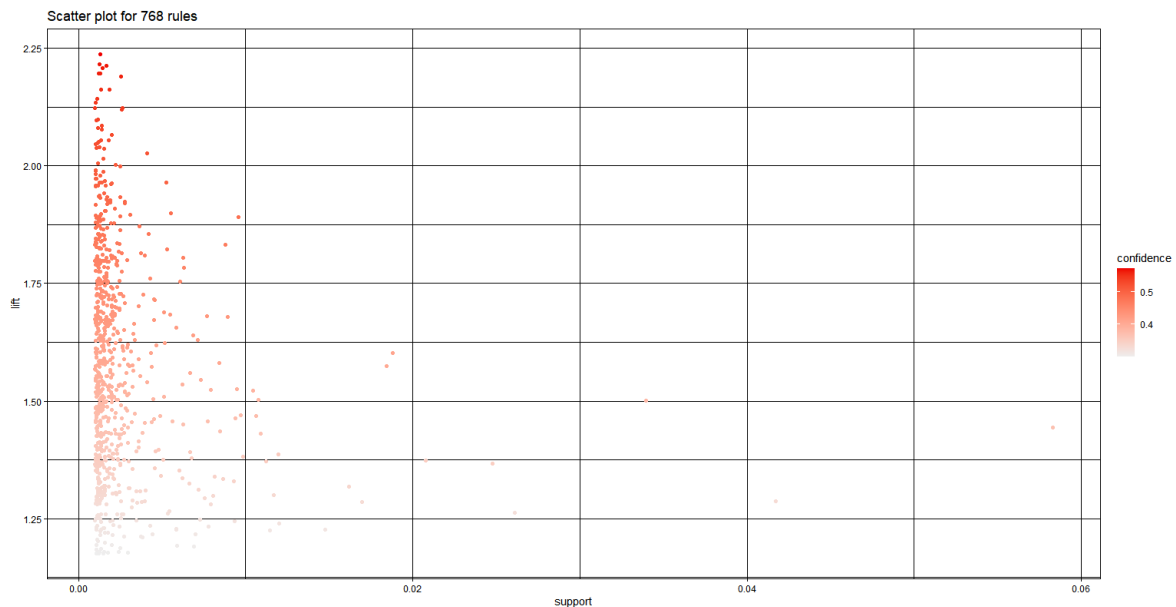


Imagen 15. Gráfica de dispersión de las reglas por soporte, lift y confianza.

6. DESPLIEGUE (RECOMENDACIONES DE NEGOCIO)

Nuestro interés, es que la población norteamericana mejore sus hábitos alimenticios, al incorporar una mayor ingesta de frutas y verduras en su cotidianidad.

Las reglas de asociación nos pueden sugerir:

1. Generación de nuevos productos de yogurt, con mezclas provenientes de productos vegetales tales como el aguacate, el brócoli y la espinaca, los cuales pueden aportar a procesos reposteros, o simplemente incorporar al consumo diario de los nutrientes necesarios en las dietas de los norteamericanos.
2. Establecer productos, con base en yogurt, que estén en mezclas directas para el consumo en tipo Smoothies o Parfait, que sean productos cuyo consumo sea del envase a la boca.
3. Generar un acercamiento con productos agrícolas verdes, como los antes mencionados, generando proximidad en las tiendas físicas o virtuales, y obsequiando un recetario de las salsas que pueden ser usadas como fuente, recetas fáciles y rápidas que complementen la alimentación diaria.
4. Generar cercanía entre las diferentes bayas y el yogurt, estableciendo proximidad física y estrategias pedagógicas de asociación e incorporación de Yogurt, en diferentes recetas que lleven bayas, bananos y Yogurt.
5. En las tiendas físicas se recomienda ofrecer muestras de los nuevos productos generados a partir del yogurt y sus asociados, con el fin de dar a conocer los productos.

Entendemos que por sí mismo, algunas presentaciones del yogurt no son en sí saludables, pese a esto, consideramos que la incorporación saludable de estos productos con un altas cantidades de frutas o verduras, puede hacer que poblaciones como las infantiles, puedan incrementar en sí mismas el consumo de verduras y frutas.

Es una apuesta por mejorar hábitos saludables en la alimentación.

Adicionalmente en los supermercados se recomienda tener los departamentos de dairy eggs, pantry, breakfast y frozen, con el fin de facilitar a los usuarios sus compras, ahorrando tiempo y obteniendo productos asociados entre sí.

BIBLIOGRAFÍA

[1] “FMI | Supermarket Facts.” <https://www.fmi.org/our-research/supermarket-facts> (accessed Mar. 06, 2022).

[2] T. text provides general information S. assumes no liability for the information given being complete or correct D. to varying update cycles and S. C. D. M. up-to-D. D. T. R. in the Text, “Topic: Food retail in the United States,” *Statista*. <https://www.statista.com/topics/1660/food-retail/> (accessed Mar. 06, 2022).

[3] R. Parra, “El foco de la industria de supermercados en 2021: Recuperar, reevaluar, reinventar,” *Abasto*, Jan. 11, 2021.

<https://abasto.com/consejos/el-foco-de-la-industria-de-supermercados-en-2021-recuperar-reevaluar-reinventar/> (accessed Mar. 06, 2022).

[4] "Farmers.gov: Resources for Farmers and Producers," *Farmers.gov*, Jul. 01, 2020. <https://www.farmers.gov/> (accessed Mar. 06, 2022).

[5] ShareAmerica, "Los agricultores de Estados Unidos alimentan al mundo," *ShareAmerica*, Mar. 07, 2019. <https://share.america.gov/es/los-agricultores-de-estados-unidos-alimentan-al-mundo/> (accessed Mar. 12, 2022).

[6] "Farmers | Center For Rural Affairs - Building a Better Rural Future." <https://www.cfra.org/farmers> (accessed Mar. 12, 2022).

[7] "About Us," *Acadian*. <https://acadian-usa.com/about-us/> (accessed Mar. 12, 2022).

[8] "USDA ERS - Farming and Farm Income." <https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/farming-and-farm-income/> (accessed Mar. 12, 2022).

[9] "What is MyPlate? | MyPlate." <https://www.myplate.gov/eat-healthy/what-is-myplate> (accessed Mar. 14, 2022).

[10] "Yogurt y salud | ThinkUSAdairy by the U.S. Dairy Export Council." <https://www.thinkusadairy.org/es/inicio/salud-y-bienestar/yogurt-y-salud> (accessed Mar. 14, 2022).