



**INSTACART\_DB**

**PROYECTO 1**

Data Mining

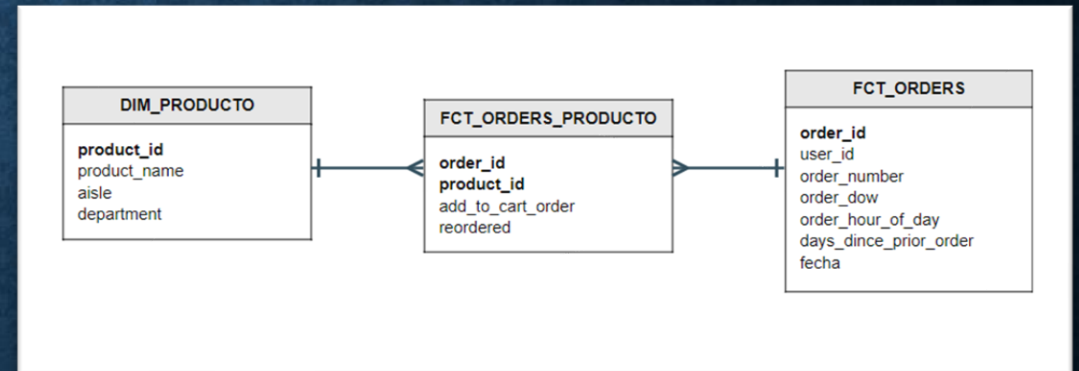
# PROCESO DE LIMPIEZA

- Conversión de datos a valores correctos
  - ❖ `days_since_prior_order` y `add_to_cart_order` son variables cuantitativas discretas. Por otro lado `reordered` es una variable booleana.
- Tratamiento de valores nulos (NaN)
  - ❖ Nombres nulos en `product_name` cambiados a “Unknown Product”.
  - ❖ Valores nulos de `add_to_cart_order` cambiados a 1 al confirmar la hipótesis de que son casos cuando existe una adición simultanea de varios productos al carrito.
  - ❖ Valores nulos de `days_since_prior_order` a -1 haciendo referencia a que es la primera compra del usuario y no tiene ordenes previas.
- Tratamiento de valores duplicados
  - ❖ Eliminar las filas duplicadas en `instacart_orders`. Se valido que la información de cada fila con id duplicado sea exactamente la misma, a excepción de la columna id que se añadió para evitar corrupciones en MySQL.
- Cambios finales
  - ❖ Eliminar la columna id de toda la tabla `instacart_orders`.
  - ❖ Añadir la columna fecha a `instacart_orders`. La fecha sería la fecha de inserción en el caso actual y simularía una columna de fecha de creación de las ordenes.



# MODELADO DE DATOS

- Esquema estrella: consultas rápidas. No existen muchos atributos que ameriten romper la dimensión en otras dimensiones y usar un esquema Snowflake
- Tablas del modelado
  - ❖ dim\_producto (dimesión de productos, nombre de producto, departamentos)
  - ❖ fct\_orders (dimesión de ordenes, relación con el usuario, día de compra, fecha)
  - ❖ fct\_orders\_producto (dimensión de ordenes, relación con cada producto en la orden, orden de adición del producto y si fue reordenado o no).
- Se eliminaron aisle\_id y department\_id ya que se desnormalizaron las relaciones.



# DIM\_PRODUCTO

| PRODUCT_ID                 | PRODUCT_NAME                                      | \ |
|----------------------------|---|---|
| 1                          | Chocolate Sandwich Cookies                        |   |
| 2                          | All-Seasons Salt                                  |   |
| 3                          | Robust Golden Unsweetened Oolong Tea              |   |
| 4                          | Smart Ones Classic Favorites Mini Rigatoni Wit... |   |
| 5                          | Green Chile Anytime Sauce                         |   |
|                            |   |   |
| AISLE                      | DEPARTMENT  |   |
| cookies cakes              | snacks  |   |
| spices seasonings          | pantry  |   |
| tea                        | beverages   |   |
| frozen meals               | frozen  |   |
| marinades meat preparation | pantry  |   |

# FCT\_ORDERS\_PRODUCTO

| ORDER_ID | PRODUCT_ID | ADD_TO_CART_ORDER | REORDERED |
|----------|------------|-------------------|-----------|
| 4        | 10054      | 5                 | True      |
| 4        | 17616      | 10                | True      |
| 4        | 21351      | 6                 | True      |
| 4        | 22598      | 7                 | True      |
| 4        | 25146      | 11                | True      |

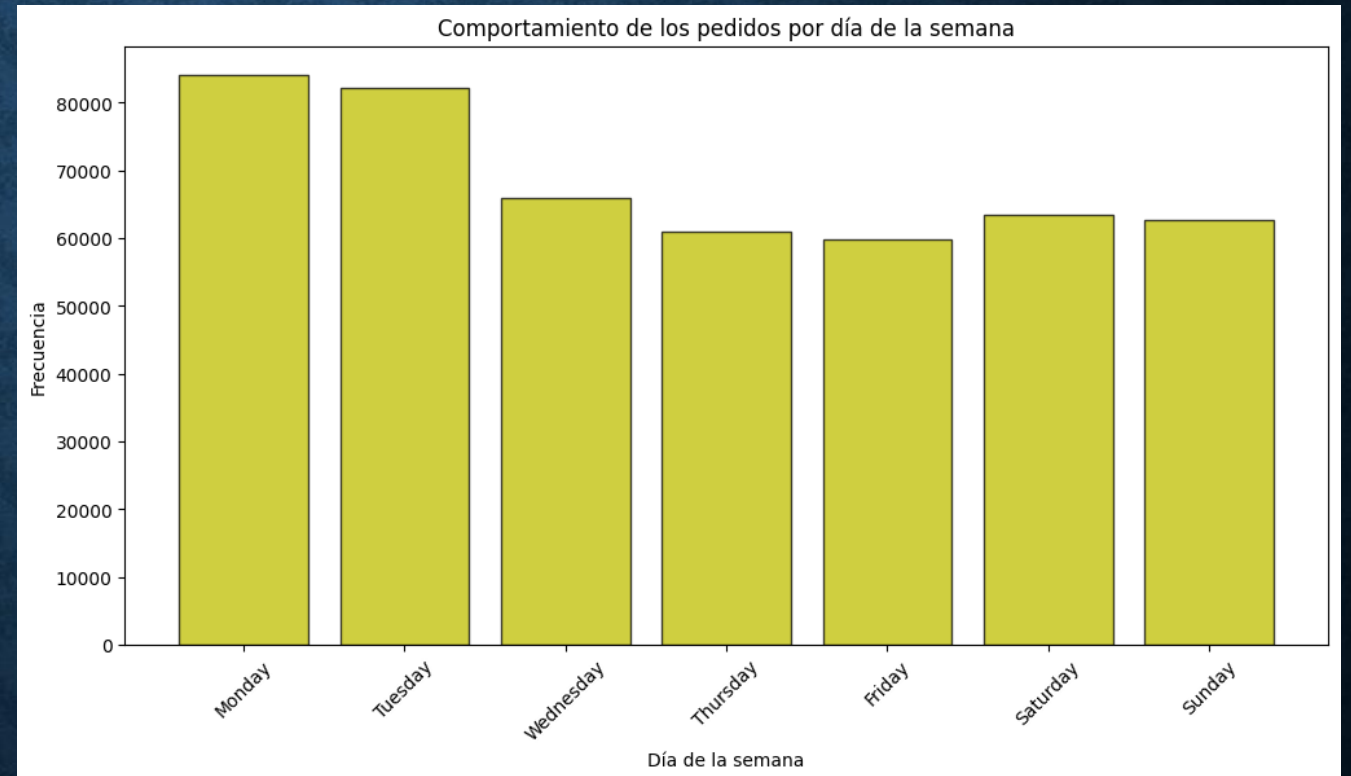
# FCT\_ORDERS

| ORDER_ID               | USER_ID | ORDER_NUMBER | ORDER_DOW | ORDER_HOUR_OF_DAY | \ |
|------------------------|---------|--------------|-----------|-------------------|---|
| 1515936                | 183418  | 11           | 6         | 13                |   |
| 1690866                | 163593  | 5            | 5         | 12                |   |
| 1454967                | 39980   | 4            | 5         | 19                |   |
| 1768857                | 82516   | 56           | 0         | 20                |   |
| 3007858                | 196724  | 2            | 4         | 12                |   |
|                        |         |              |           |                   |   |
| DAYS_SINCE_PRIOR_ORDER | FECHA   |              |           |                   |   |
|                        | 30      | 2025-02-10   |           |                   |   |
|                        | 9       | 2025-02-10   |           |                   |   |
|                        | 2       | 2025-02-10   |           |                   |   |
|                        | 10      | 2025-02-10   |           |                   |   |
|                        | 17      | 2025-02-10   |           |                   |   |

# RESULTADOS

## COMPORTAMIENTO DE COMPRA SEGÚN DÍA DE LA SEMANA

Se evidencia un mayor índice de compra en los días Lunes y Martes

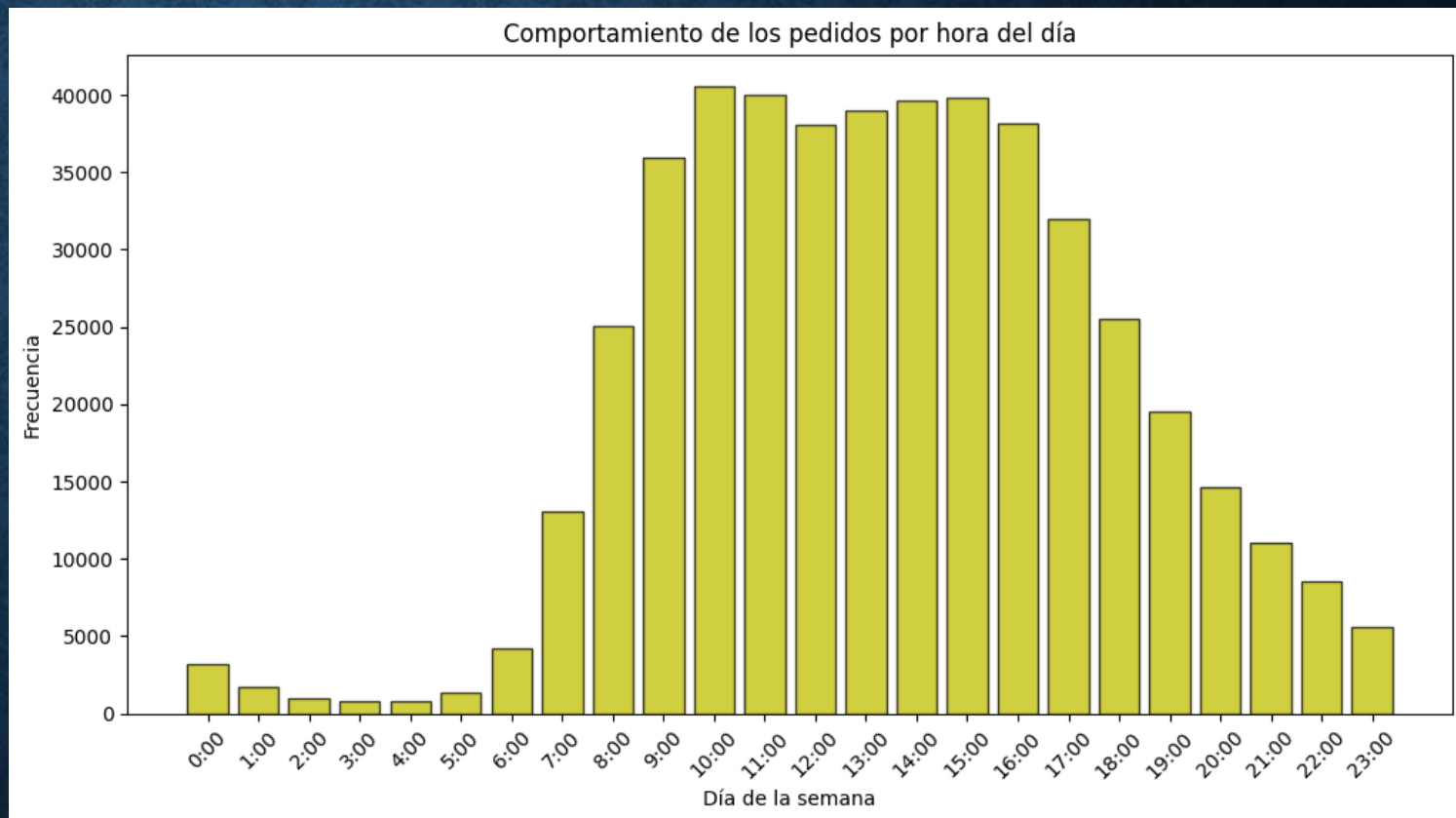




# RESULTADOS

## COMPORTAMIENTO DE COMPRA SEGÚN HORA DEL DÍA

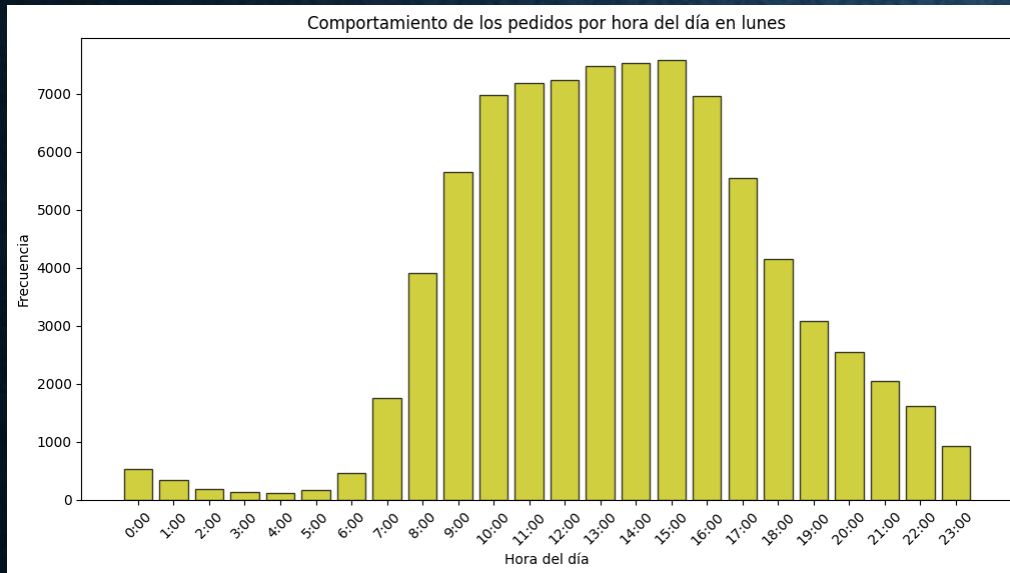
Se evidencia un mayor índice de compra en las horas donde la mayor parte de la gente está activa (de 9:00 a 16:00).



# RESULTADOS

## COMPORTAMIENTO SEGÚN HORA DEL DÍA Y DÍA DE LA SEMANA

### LUNES



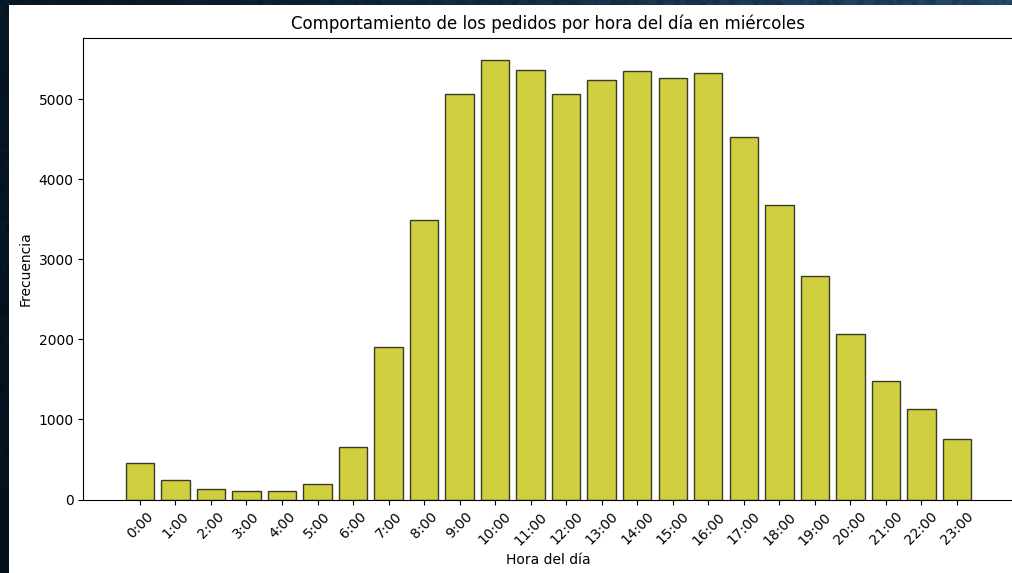
### MARTES



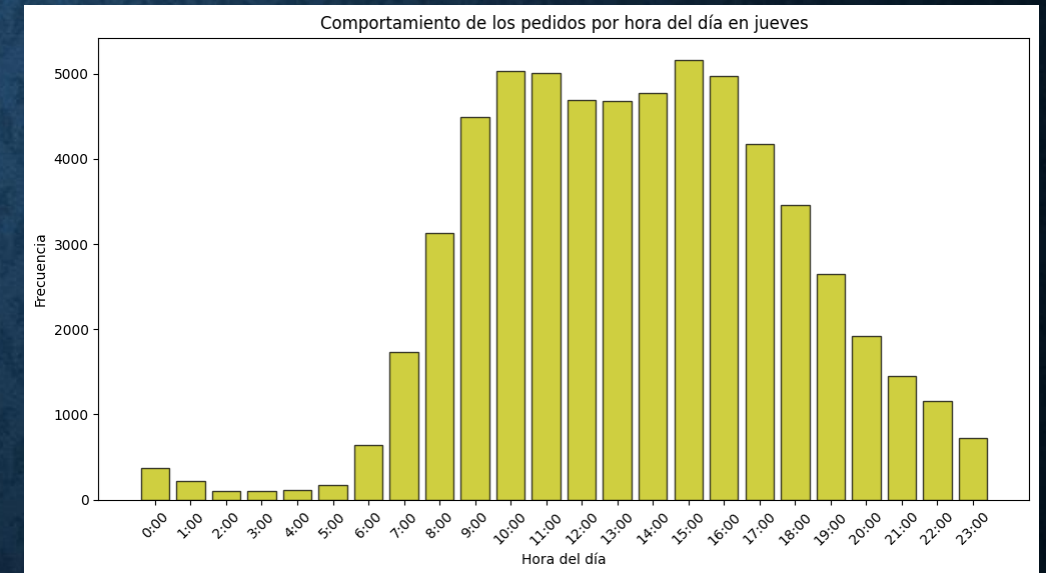
# RESULTADOS

## COMPORTAMIENTO SEGÚN HORA DEL DÍA Y DÍA DE LA SEMANA

### MIÉRCOLES



### JUEVES

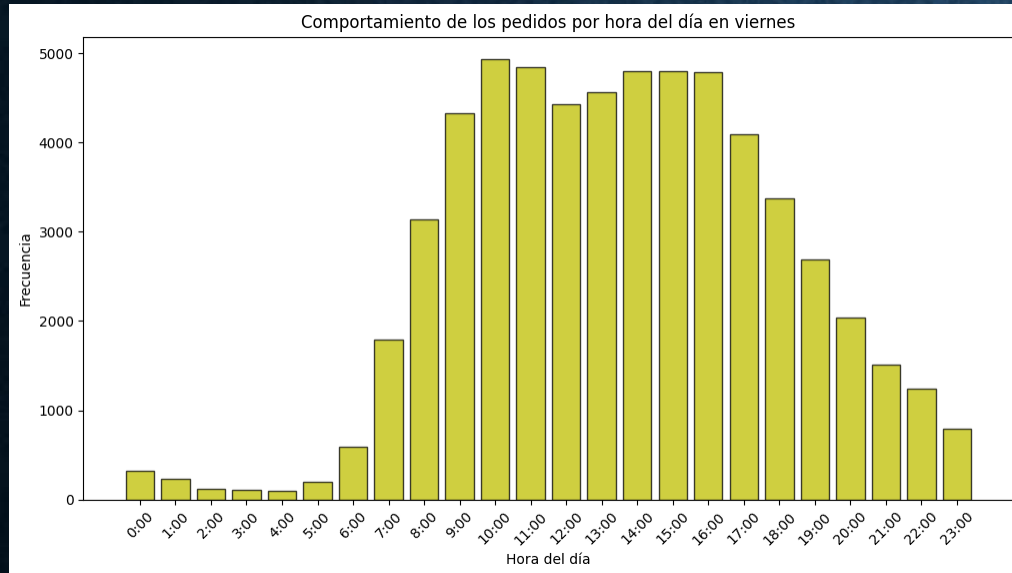




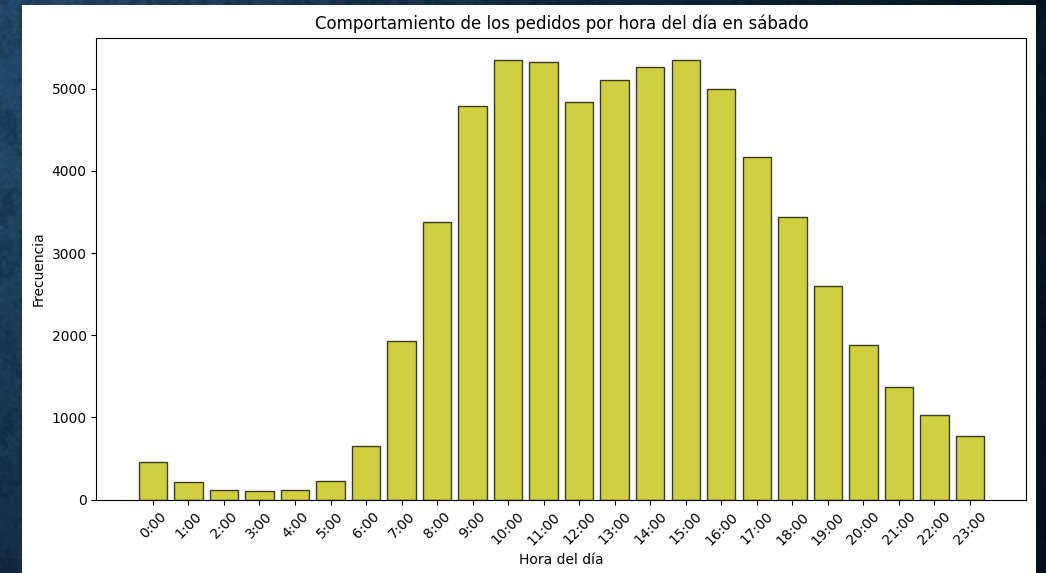
# RESULTADOS

## COMPORTAMIENTO SEGÚN HORA DEL DÍA Y DÍA DE LA SEMANA

### VIERNES



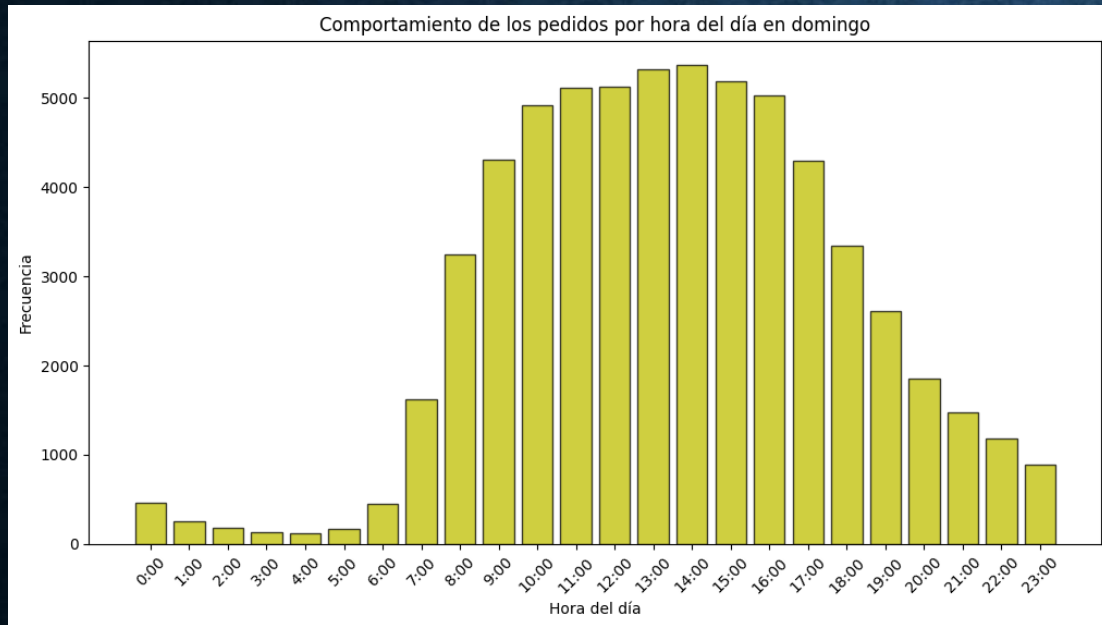
### SÁBADO



# RESULTADOS

## COMPORTAMIENTO SEGÚN HORA DEL DÍA Y DÍA DE LA SEMANA

### DOMINGO



### CONCLUSIONES

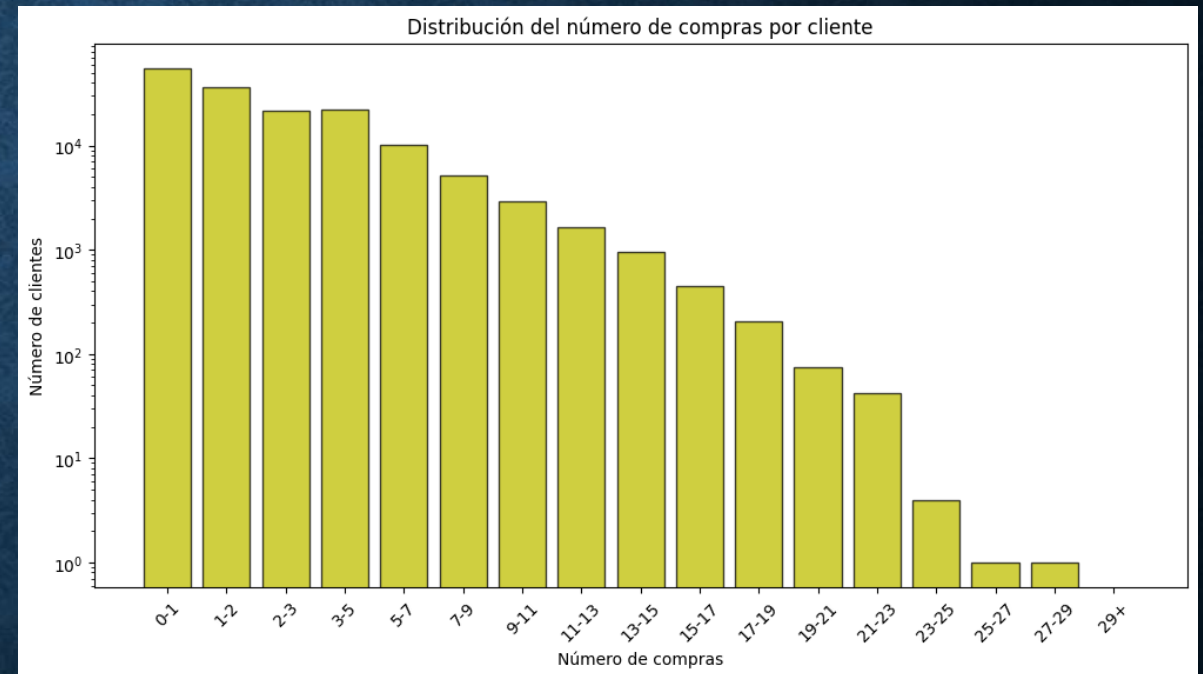
- Las distribuciones de compras en cada día siguen una misma distribución respecto a la distribución general de horas.
- A pesar de que la distribución de compras por hora sea la misma, los niveles de compra son mucho más altos a esas horas en los días lunes y martes.
- Existe una leve tendencia de realizar pedidos en horas más tempranas el día Martes.

# RESULTADOS

## DISTRIBUCIÓN DE LAS ÓRDENES HECHAS POR LOS CLIENTES

Promedio de compras por cliente: 3.042181952145938  
Mayor número de compras por cliente: 28  
Menor número de compras por cliente: 1

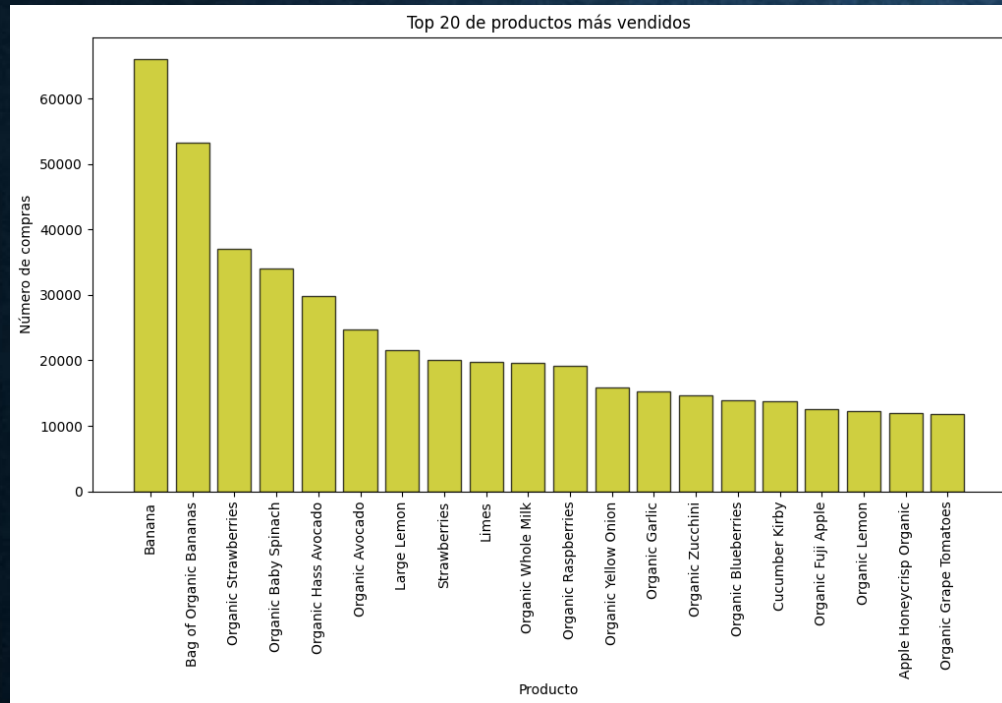
**Los clientes realizan en promedio solamente 3 ordenes. Es necesario incentivar y enganchar a los clientes para que realicen más ordenes.**





# RESULTADOS

## TOP 20 PRODUCTOS MÁS FRECUENTES



Número de ordenes de cada producto de los 20 más vendidos:

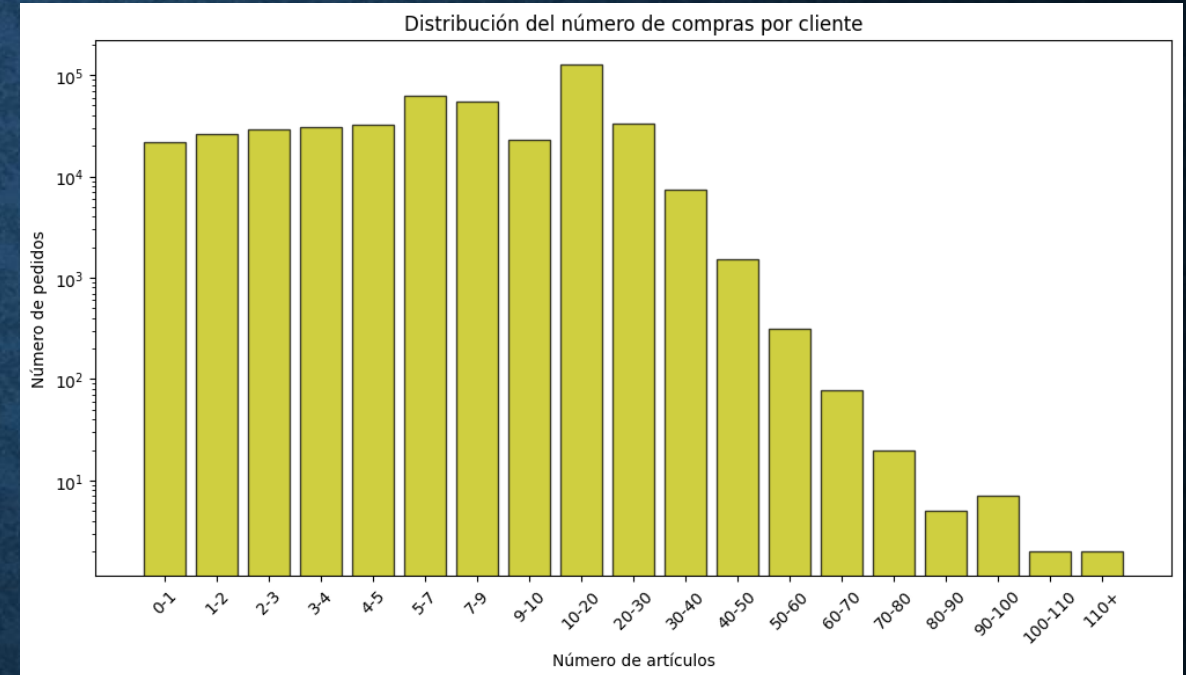
|                          |       |
|--------------------------|-------|
| Banana                   | 66050 |
| Bag of Organic Bananas   | 53297 |
| Organic Strawberries     | 37039 |
| Organic Baby Spinach     | 33971 |
| Organic Hass Avocado     | 29773 |
| Organic Avocado          | 24689 |
| Large Lemon              | 21495 |
| Strawberries             | 20018 |
| Limes                    | 19690 |
| Organic Whole Milk       | 19600 |
| Organic Raspberries      | 19197 |
| Organic Yellow Onion     | 15898 |
| Organic Garlic           | 15292 |
| Organic Zucchini         | 14584 |
| Organic Blueberries      | 13879 |
| Cucumber Kirby           | 13675 |
| Organic Fuji Apple       | 12544 |
| Organic Lemon            | 12232 |
| Apple Honeycrisp Organic | 11993 |
| Organic Grape Tomatoes   | 11781 |

**Los productos más vendidos son los orgánicos. Se recomienda expandir la gama de productos orgánicos para atraer a más clientes y potencializar las compras de los actuales**

# RESULTADOS

## ¿CUÁNTOS ARTÍCULOS SE COMPRAN GENERALMENTE EN UN PEDIDO?

Promedio de artículos por pedido: 10.098983215049127  
Mayor número de artículos por pedido: 127  
Menor número de artículos por pedido: 1



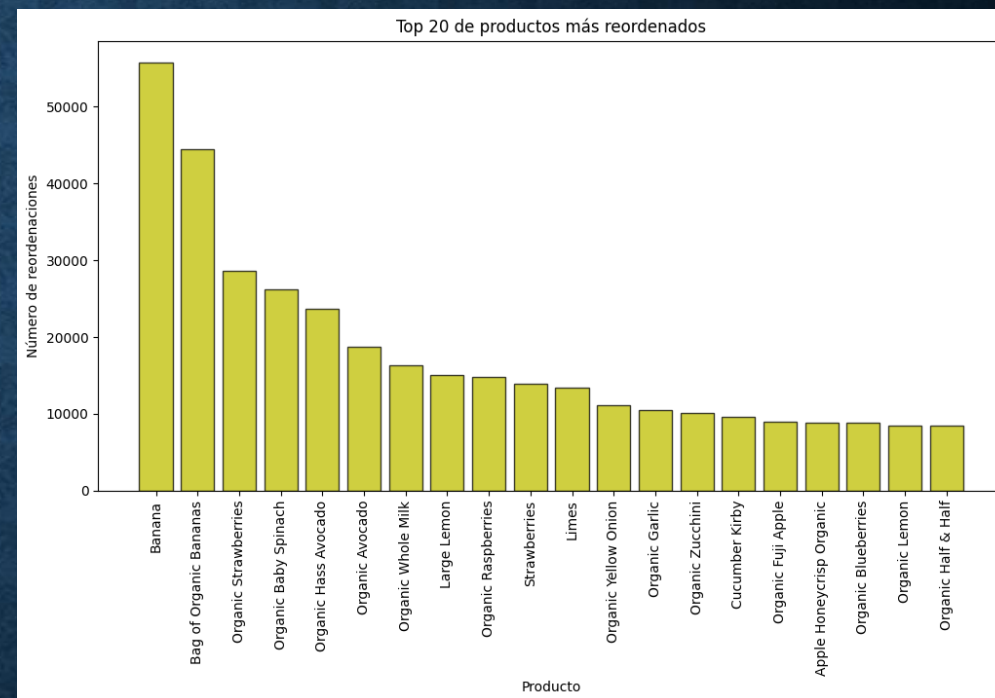
**Una gran cantidad de clientes únicamente compran 1 artículo. Se recomienda ofrecer descuentos y ofertas al seleccionar distintos productos, además también es buena opción ampliar la gama de artículos actualmente disponibles.**

# RESULTADOS

## TOP 20 ARTÍCULOS QUE SE VUELVEN A PEDIR CON MÁS FRECUENCIA

Número de reordenaciones de cada producto de los 20 más reordenados:

|                          |       |
|--------------------------|-------|
| Banana                   | 55763 |
| Bag of Organic Bananas   | 44450 |
| Organic Strawberries     | 28639 |
| Organic Baby Spinach     | 26233 |
| Organic Hass Avocado     | 23629 |
| Organic Avocado          | 18743 |
| Organic Whole Milk       | 16251 |
| Large Lemon              | 15044 |
| Organic Raspberries      | 14748 |
| Strawberries             | 13945 |
| Limes                    | 13327 |
| Organic Yellow Onion     | 11145 |
| Organic Garlic           | 10411 |
| Organic Zucchini         | 10076 |
| Cucumber Kirby           | 9538  |
| Organic Fuji Apple       | 8989  |
| Apple Honeycrisp Organic | 8836  |
| Organic Blueberries      | 8799  |
| Organic Lemon            | 8412  |
| Organic Half & Half      | 8389  |



**Algunos de estos artículos no son los más pedidos. Se recomienda publicitar más los artículos que evidencian un gusto por los clientes en general**



# RESULTADOS

## PROPORCIÓN DE PEDIDOS QUE SE VUELVEN A PEDIR PARA CADA PRODUCTO

Proporción de reordenaciones de cada producto respecto al total de pedidos:

| PRODUCT_ID |          |
|------------|----------|
| 1          | 0.564286 |
| 2          | 0.000000 |
| 3          | 0.738095 |
| 4          | 0.510204 |
| 7          | 0.500000 |
|            | ...      |
| 49690      | 0.800000 |
| 49691      | 0.430556 |
| 49692      | 0.416667 |
| 49693      | 0.440000 |
| 49694      | 0.333333 |

**Sería recomendable tomar acción y difundir más los productos con tasa de reordenación menor a 0.5 o en caso de que su calidad sea mala quitar esos productos de los disponibles para futuras compras.**

# RESULTADOS

## PROPORCIÓN DE PRODUCTOS PEDIDOS QUE SE VUELVEN A PEDIR PARA CADA CLIENTE

|                    | PRODUCT_NAME                                  | COUNT | COUNT_REO | PROP_REO |
|--------------------|---|-------|-----------|----------|
| Sleepytime Vanilla | 15 Grain Whole Grain Bread                    | 1     | 1         | 1.000000 |
|                    | Whole Milk                                    | 3     | 2         | 0.666667 |
|                    | Cauliflower                                   | 1     | 1         | 1.000000 |
|                    | Organic Red Potatoes                          | 1     | 1         | 1.000000 |
|                    | Caffeine Free Herbal Tea Bags                 | 1     | 0         | 0.000000 |
|                    | Golden Raisins                                | 1     | 1         | 1.000000 |
|                    | Decaf Mountain Chai                           | 1     | 0         | 0.000000 |
|                    | Fat Free Smooth & Creamy Plain Organic Yogurt | 1     | 0         | 0.000000 |
|                    | Unknown Product                               | 1     | 1         | 1.000000 |
|                    | Half & Half                                   | 3     | 2         | 0.666667 |
|                    | Whole Wheat Pita Bread Loaves                 | 1     | 1         | 1.000000 |
|                    | Leek  | 1     | 0         | 0.000000 |
|                    | Bunched Carrots                               | 2     | 1         | 0.500000 |
|                    | Organic Pumpkin Puree                         | 1     | 0         | 0.000000 |
|                    | Chick Peas Garbanzos                          | 1     | 0         | 0.000000 |
|                    | Coconut Milk- Classic                         | 1     | 1         | 1.000000 |
|                    | Super Greens Salad                            | 1     | 1         | 1.000000 |
|                    | Red Onion                                     | 1     | 0         | 0.000000 |

**Sería recomendable sugerir a cada usuario por medio de publicidad los productos con una tasa de reordenación individual mayor a 0.6 para aumentar las ordenes de ese usuario. Por otro lado, se debe dejar de sugerir productos con poca o nula tasa de reordenación.**

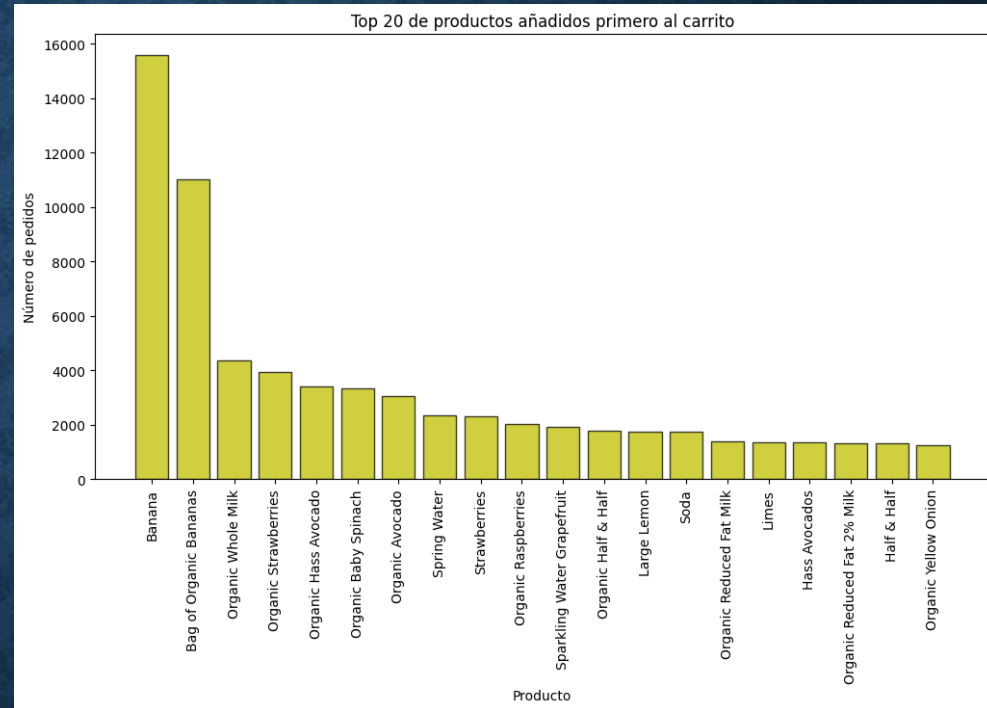


# RESULTADOS

## TOP 20 ARTÍCULOS QUE LA GENTE PONE PRIMERO EN EL CARRITO

Top 20 de productos añadidos primero al carrito:

|                             |       |
|-----------------------------|-------|
| Banana                      | 15566 |
| Bag of Organic Bananas      | 11028 |
| Organic Whole Milk          | 4363  |
| Organic Strawberries        | 3950  |
| Organic Hass Avocado        | 3391  |
| Organic Baby Spinach        | 3337  |
| Organic Avocado             | 3044  |
| Spring Water                | 2336  |
| Strawberries                | 2309  |
| Organic Raspberries         | 2027  |
| Sparkling Water Grapefruit  | 1915  |
| Organic Half & Half         | 1797  |
| Large Lemon                 | 1737  |
| Soda                        | 1733  |
| Organic Reduced Fat Milk    | 1397  |
| Limes                       | 1371  |
| Hass Avocados               | 1340  |
| Organic Reduced Fat 2% Milk | 1311  |
| Half & Half                 | 1309  |
| Organic Yellow Onion        | 1247  |



**Únicamente los dos primeros productos tienen una tasa alta de ser los primeros en ponerse en el carrito. Se recomienda difundir más en general todos los productos que normalmente se añaden primero al carrito ya que representan necesidades comunes de los clientes.**



# LINK DEL REPOSITORIO DE TODO EL ANÁLISIS

- <https://github.com/Ericson3100/Proyecto1DM>

