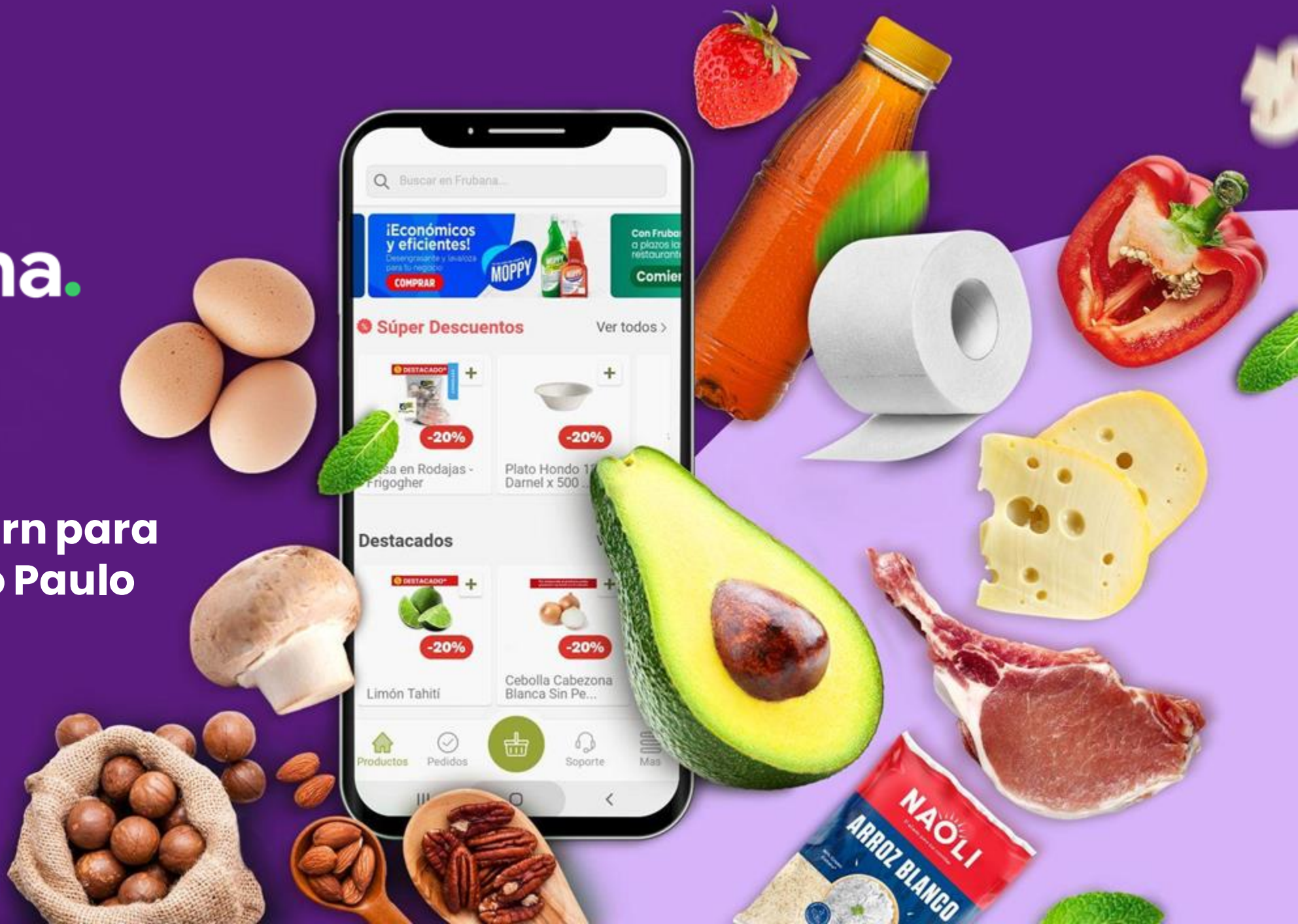




Modelo de Churn para clientes de Sao Paulo



CONTEXTO



Sobre la empresa

Frubana es una empresa fundada en 2018 y actualmente opera en tres países: Colombia, Brasil y México.

Pertenece al sector de alimentos que busca desintermediar la relación entre agricultores y consumidores a través del uso de tecnología y prácticas innovadoras de logística y acompañamiento a personas y empresas.

Su misión

La misión de Frubana es hacer la comida de Latinoamérica más barata y ser el servicio "todo en uno" de los restaurantes de la región.

Los productos

Frubana ofrece productos de diferentes categorías como lo son fruver, abarrotes y proteínas.

PROBLEMÁTICA



La pregunta de negocio que se busca responder en el abordaje de Frubana está basada en encontrar aquella proporción de clientes que tienen la mayor probabilidad de abandonar la compañía



REQUERIMIENTO DE DATOS



REQUERIMIENTOS DE DATOS

Predicción del Churn



Una fase inicial de catalogación de los usuarios a través de etiquetas de cara a su comportamiento de compra, con el fin de encontrar aquellos usuarios de valor significativo que son la población objetivo de la necesidad del negocio



A partir de esta correcta estimación de los grupos, se buscará aplicar dos metodologías comparativas para el modelo de Churn



A: Restaurante que ha comprado en Frubana más de 10 órdenes basada en el algoritmo de Random Forest como técnica de aprendizaje supervisado



B: Uso de un algoritmo de XG Boost, basado en un sistema de ensamblado secuencial de árboles de decisión que identifiquen la población potencial a salir.

ENTENDIMIENTO DE DATOS



ENTENDIMIENTO DE LOS DATOS

frubana

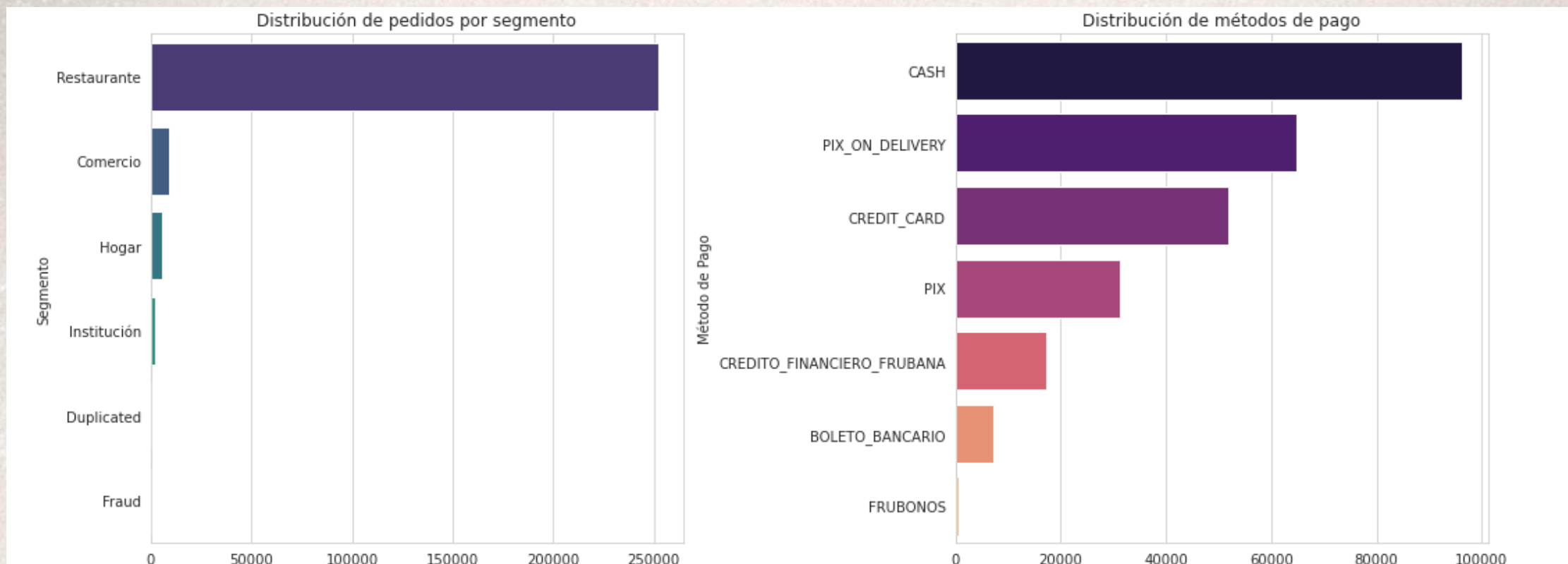


- *Del total de poco más de 645 mil observaciones, se encontraron 269.133 registros válidos, con lo cual se corrigió el problema de duplicidad en la información.*
- La información resultante de 269 mil registros fue objeto de múltiples consultas, uniones, combinaciones y algunas transformaciones internas de la compañía, por lo que se puede afirmar que la información es procesable y será útil para el abordaje de los modelos mencionados.
- Los datos proporcionados son tanto de clientes activos como de clientes que ya se retiraron de la compañía con el propósito de tener información completa para el entrenamiento del modelo

ANÁLISIS EXPLORATORIO DE DATOS



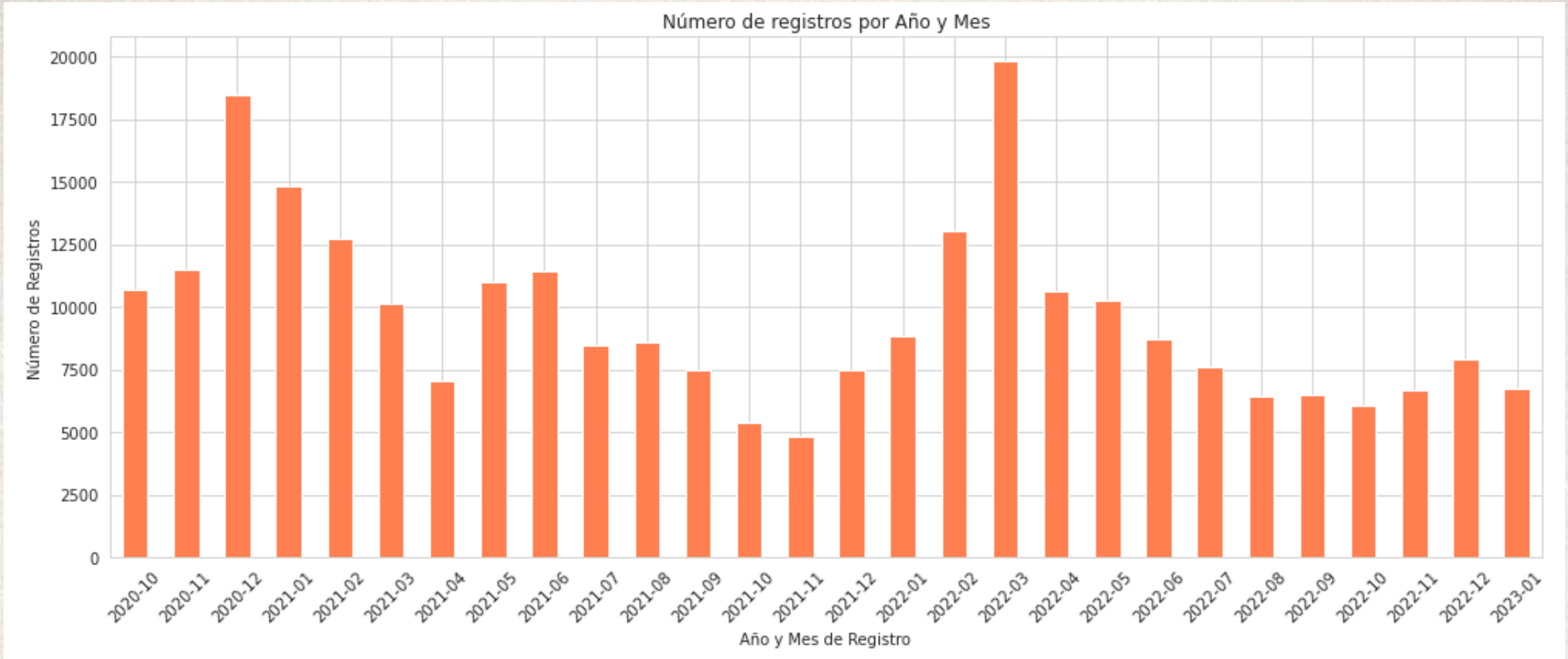
Haciendo el Análisis Exploratorio de Datos



CASH" y "PIX" son los métodos de pago más populares. Otros métodos de pago, como "PIX_ON_DELIVERY" y "CREDIT_CARD", también son comunes, pero no tanto como los dos primeros

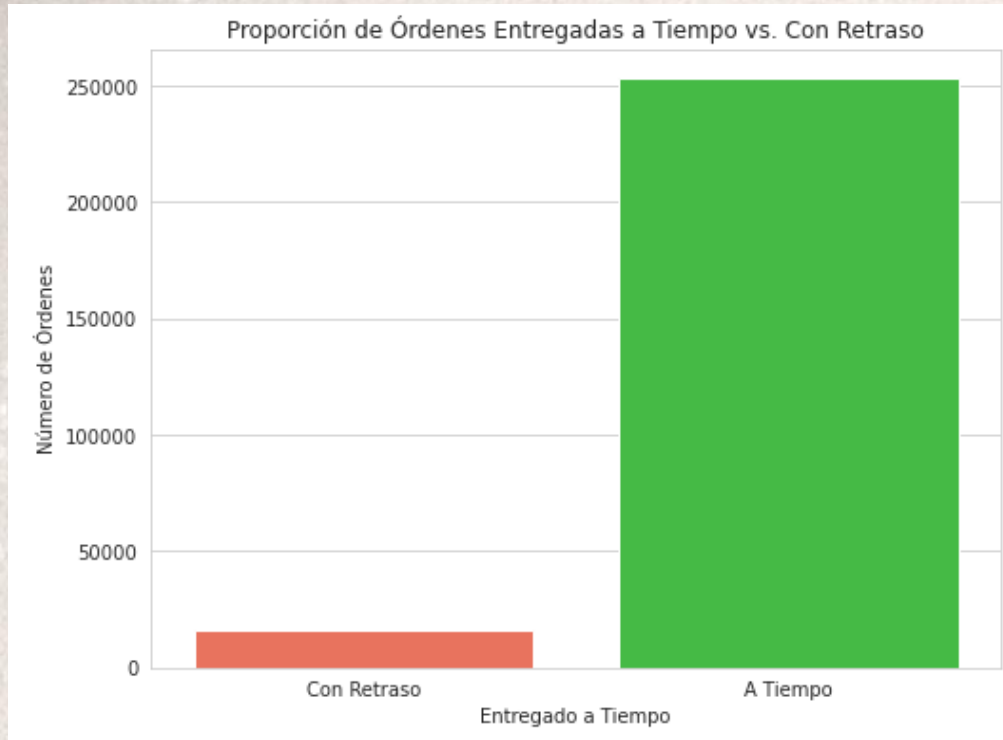
El segmento "Restaurante" es, con mucho, el más popular, seguido por "Super", "Farmacia" y "Otros". Los segmentos "Express" y "Drinks" tienen menos pedidos en comparación

Haciendo el Análisis Exploratorio de Datos

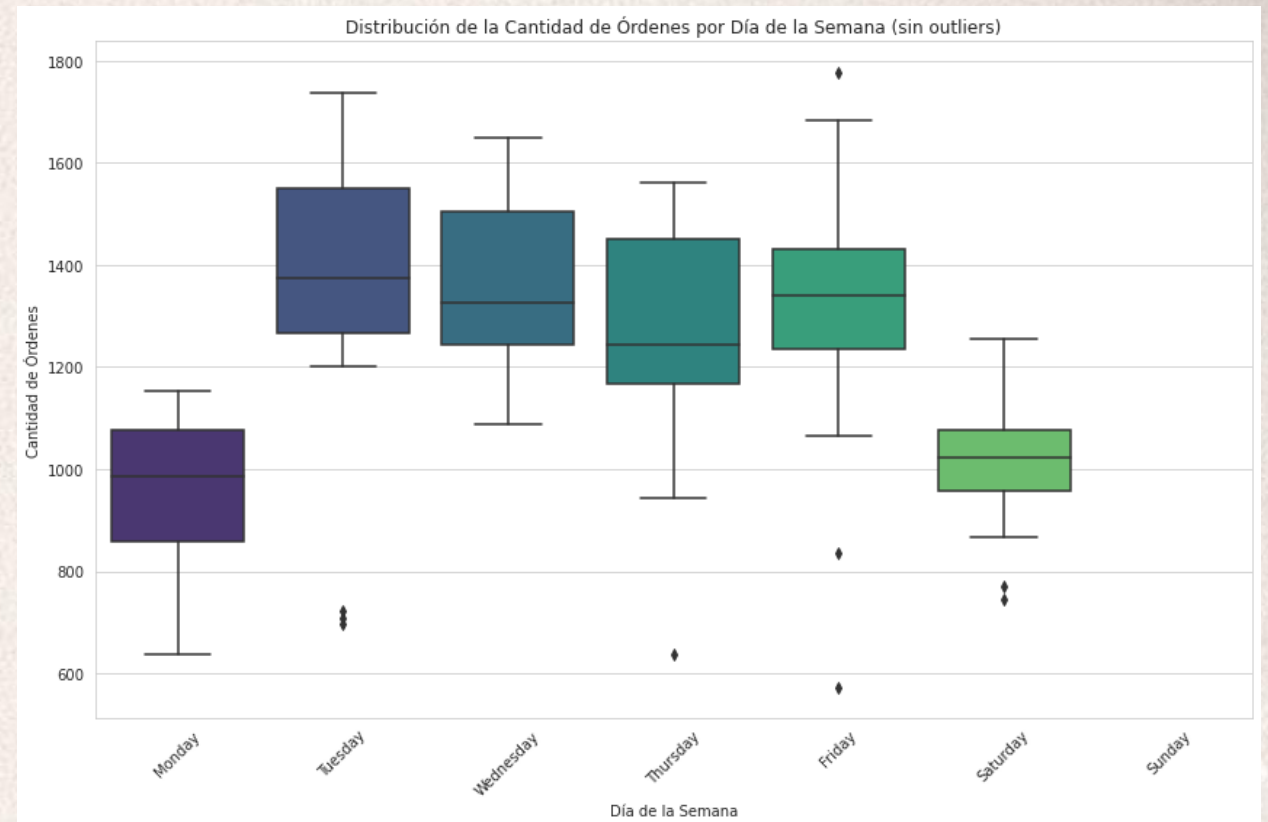


Es posible observar algunos picos de registro ocurridos en los meses de diciembre de 2020 y marzo de 2022, seguidos por un comportamiento particularmente uniforme en el resto de los meses con un leve descenso de registros hacia el último trimestre del año 2021. A priori es posible pensar que hay un descenso en la actividad económica propia de Brasil debido a festividades propias de las fechas

Haciendo el Análisis Exploratorio de Datos

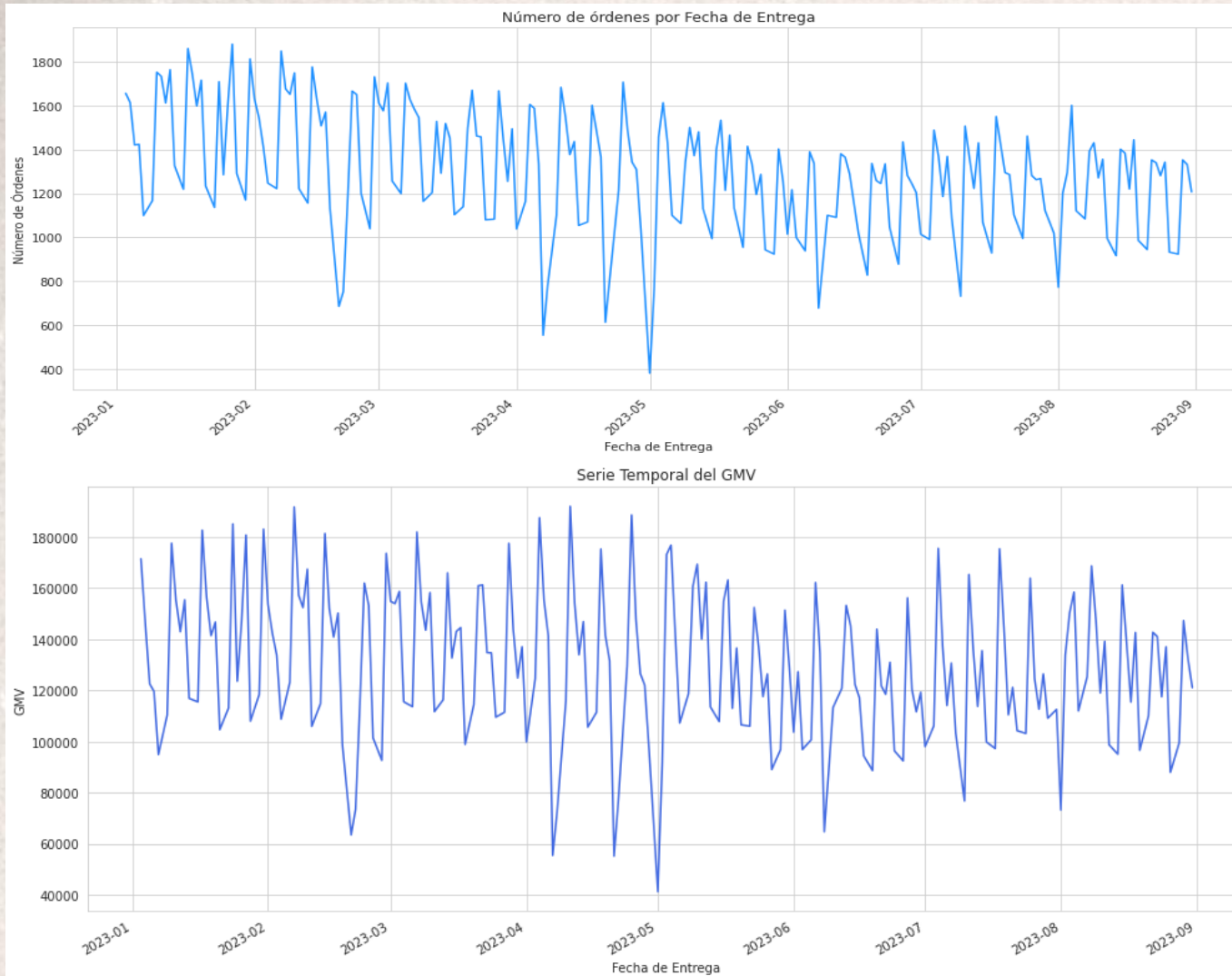


94% de las entregas a tiempo
6% de las entregas con retraso



Los días martes y miércoles hay una mayor cantidad de órdenes, comparadas con los días sábado y lunes en donde baja la actividad

Haciendo el Análisis Exploratorio de Datos



El número de órdenes y el GMV tienen propiedades de Estacionariedad

Es posible generar descomposiciones para capturar tendencias, estacionalidades y predecir valores dadas las propiedades de las series

CONCLUSIONES PRELIMINARES



Del proceso

- La información proporcionada por Frubana contaba con robustez lo que hizo más fácil el alistamiento, procesabilidad y demás virtudes de los datos para poder ejecutar un posterior modelo.
- El mecanismo de recolección de datos a partir del método de entrevista fue crucial para una estimación más estratégica del problema a solucionar y poder responder a preguntas que agreguen valor al negocio.

De nuestros resultados

- Con la integración del modelo RFM es posible ya contestar uno de los dos interrogantes abordados en el documento de cara a la correcta categorización de usuarios.
- El análisis exploratorio de datos arrojó componentes de estacionariedad en las series de tiempo utilizadas, lo que permitirá mejores validaciones teóricas a partir de que se cumple un mayor número de propiedades.
- Es necesario generar un segundo punto de contacto con el negocio para socializar el proceso de etiquetado propuesto y ver si hay algún comentario que mejore o corrija el enfoque generado en este documento.

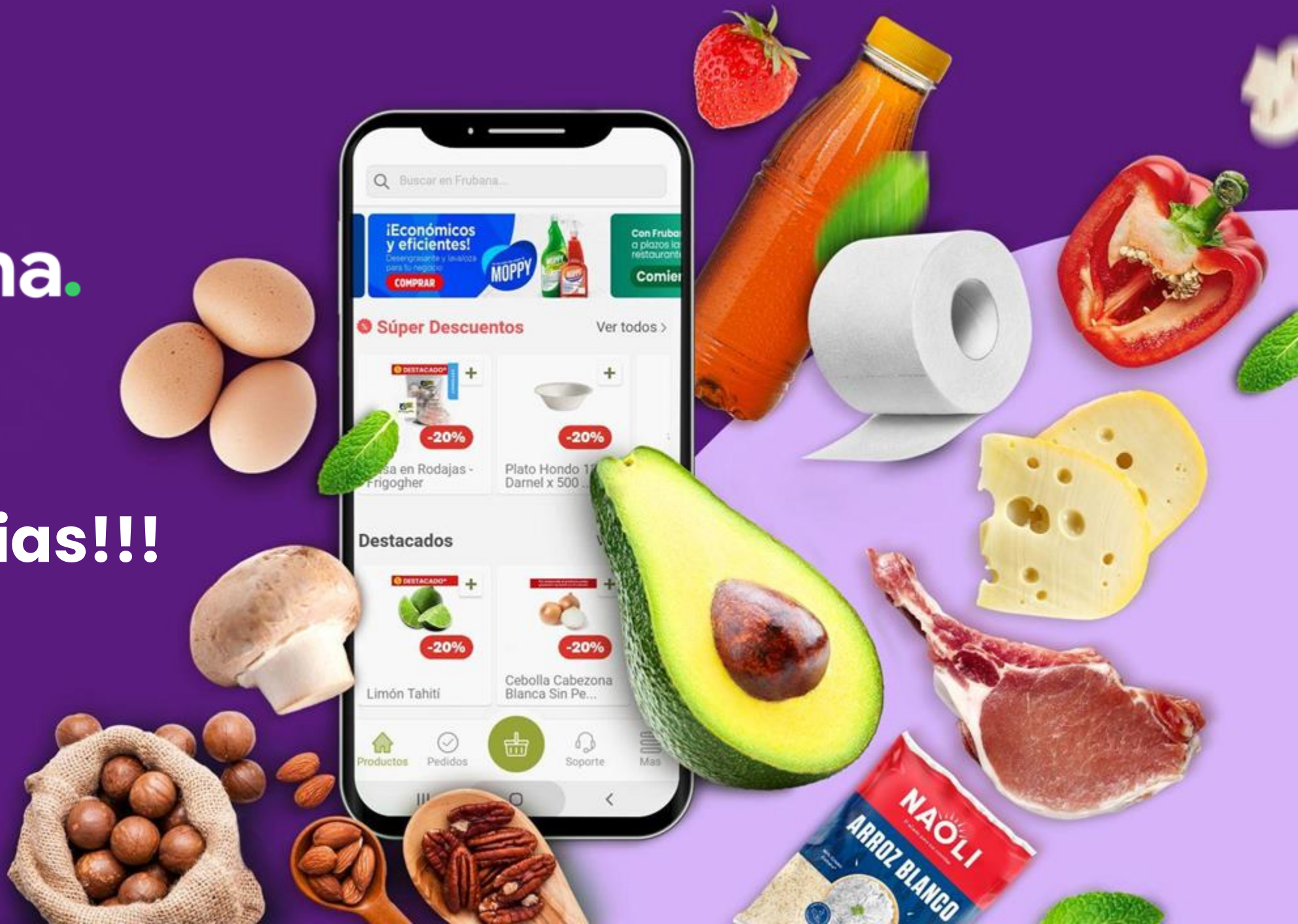
Con base en la información y necesidades proporcionadas por el negocio, se estableció a partir de los datos un Modelo RFM (Recency, Frequency, Monetary) para generar una primera segmentación de usuarios a partir de la recencia de la última compra, la frecuencia de compras y el monto o importe de las compras para cada usuario.

- En promedio, se espera que un cliente aporte US\$97,23u a lo largo de su tiempo como cliente
- El valor promedio de GMV por compra es de aproximadamente US\$116,92.
- En promedio, un cliente realiza aproximadamente 27.67 compras
- La duración promedio de un cliente (tiempo de registro y última orden) es de aproximadamente 175,36 días.

Tipo de Cliente	Recencia	Frecuencia	Monto	Tamaño
VIP	Alto	Alto	Alto	2.048
Leales	Medio-Alto	Alto	Alto	1.989
En Riesgo	Bajo	Medio	Medio	2.126
Posible Churn	Muy bajo	Bajo	Bajo	2.124



!!!Gracias!!!





Andrés Felipe Gutiérrez Huerfano
Jonnatan Mauricio Triana Alvarez
Juan Carlos Avila Villalba
Jenniffer Sofia Escudero Perdomo

