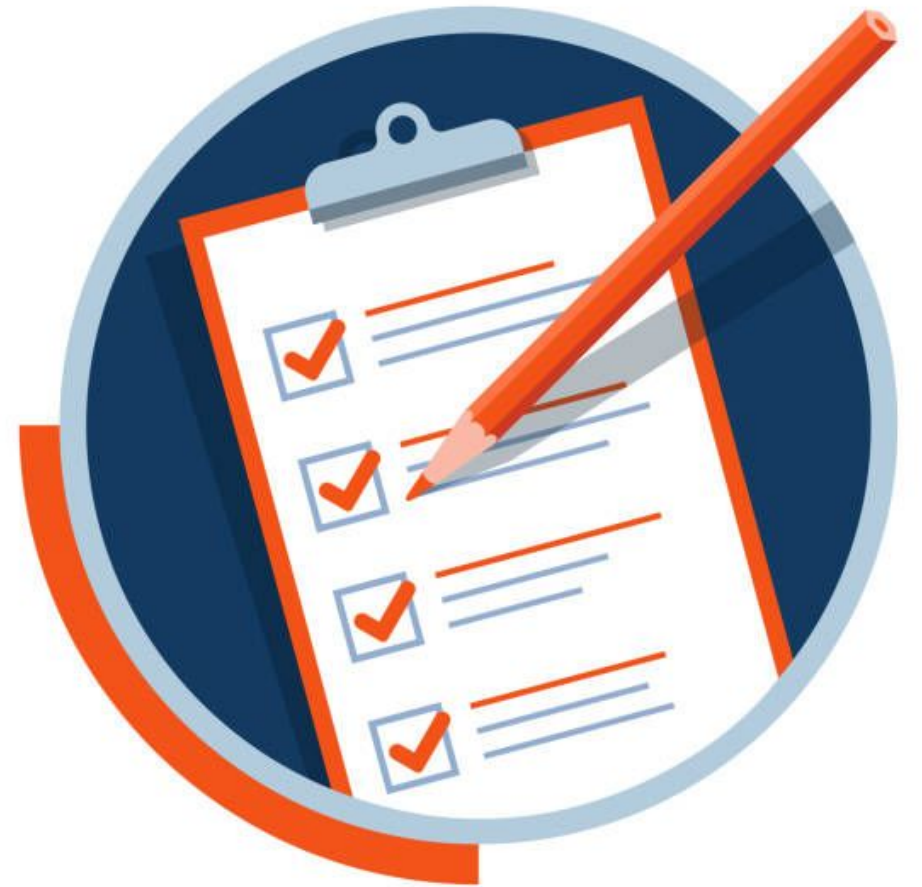


MODELADO PREDICTIVO EN ÓRDENES DE VENTA POR INTERNET

Federico Moscardo
Agustín Walker
Santiago Yacob

AGENDA

- 01** | Contexto, Negocio y Audiencia
- 02** | Resumen de metadata
- 03** | Hipótesis/Preguntas de Interés
- 04** | Análisis Exploratorio
- 05** | Aplicación de Algoritmos de ML
- 06** | Insights y Next Steps



CONTEXTO

- **Incremento** exponencial de personas con **acceso a internet** en los últimos años
- **Globalización** y la reciente **pandemia**, han conformado el escenario ideal para situar a las ventas en línea como reales protagonistas de la economía mundial
- **Las personas están cambiando sus hábitos de consumo:** menos compras presenciales y más compras online

NEGOCIO

- Importante **empresa de comercio electrónico de Estados Unidos** ha recopilado información sobre las órdenes de compra generada por clientes en forma online
- **El objetivo de la empresa** es contratar a un equipo de Data Science para que ayude a su equipo gerencial a **reducir costos operativos y maximizar márgenes de utilidad**

AUDIENCIA DEL ESTUDIO

- **Equipo gerencial de la compañía**

RESUMEN DE METADATA

Dataset Original:

order_id: id de la orden de compra
order_date: fecha de la orden de compra
status: estado de la orden
item_id: id producto comprado
qty_ordered: cantidad ordenada del producto
price: precio del producto
value: cantidad por precio
discount_amount: monto de descuento
total: precio final de compra
category: categoría del producto
payment_method: método de pago
cust_id: id del cliente
Gender: género del cliente
Age: edad del cliente

Customer_since: fecha de alta del cliente

County: condado

City: localidad

State: estado

Zip: código postal

Region: región

Discount_Percent: % descuento asociado

Dataset API:

date: fecha del valor de las acciones

open: valor apertura acción

high: valor máximo acción

low: valor mínimo acción

close: valor cierre acción

volume: volumen de acciones

PREGUNTAS DE INTERÉS

Son las preguntas que el cliente realizó al equipo de Data Science, en busca de respuestas que le permitan maximizar sus resultados y así tener una ventaja competitiva en el mercado

¿Cuál es la región con mayor participación en ventas online?

¿La región/estado con mayor incremento en ventas online?

¿Cuáles son las categorías con tendencia a incrementar o disminuir su cantidad de ordenes online?

¿Cómo es la situación de las órdenes de compra canceladas?

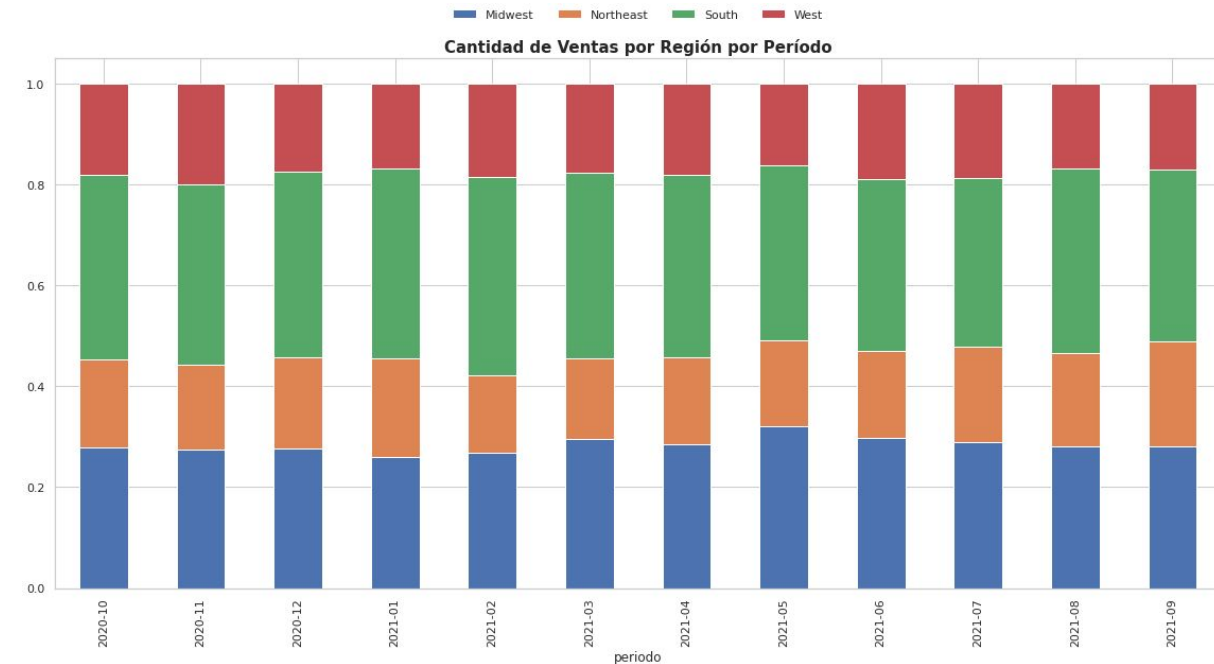
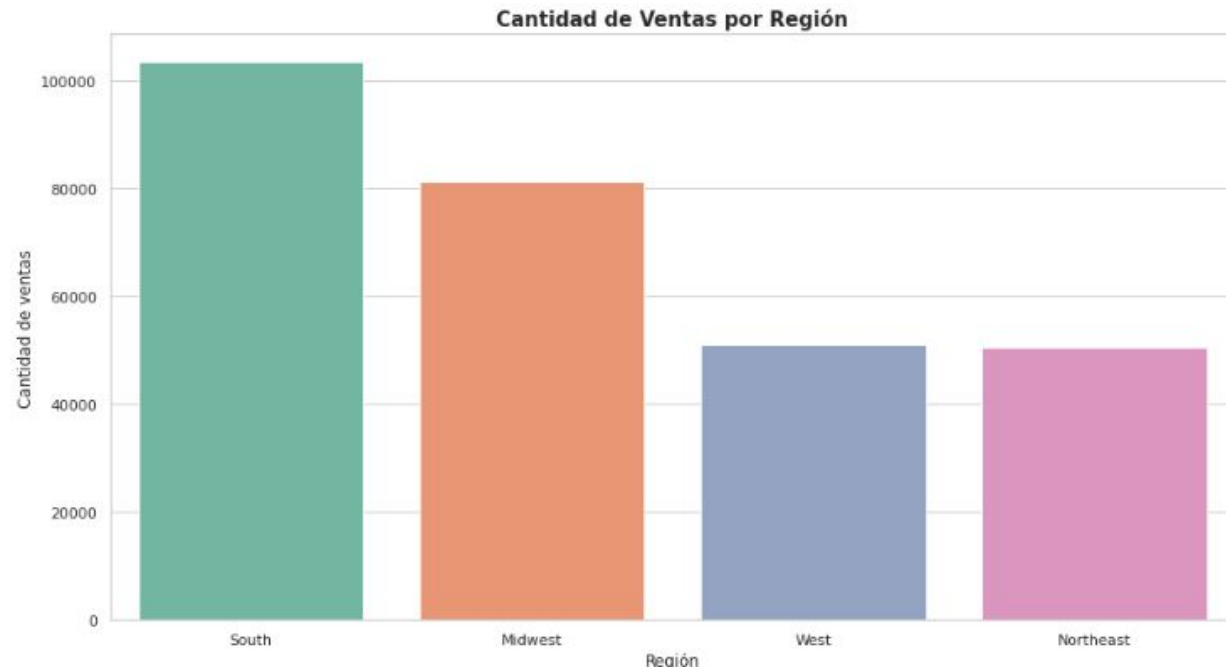
¿Se podría identificar algún patrón asociado a la cancelación de órdenes de compra?

¿Existen factores externos que podrían incidir en la cancelación de órdenes de venta?

ANÁLISIS EXPLORATORIO

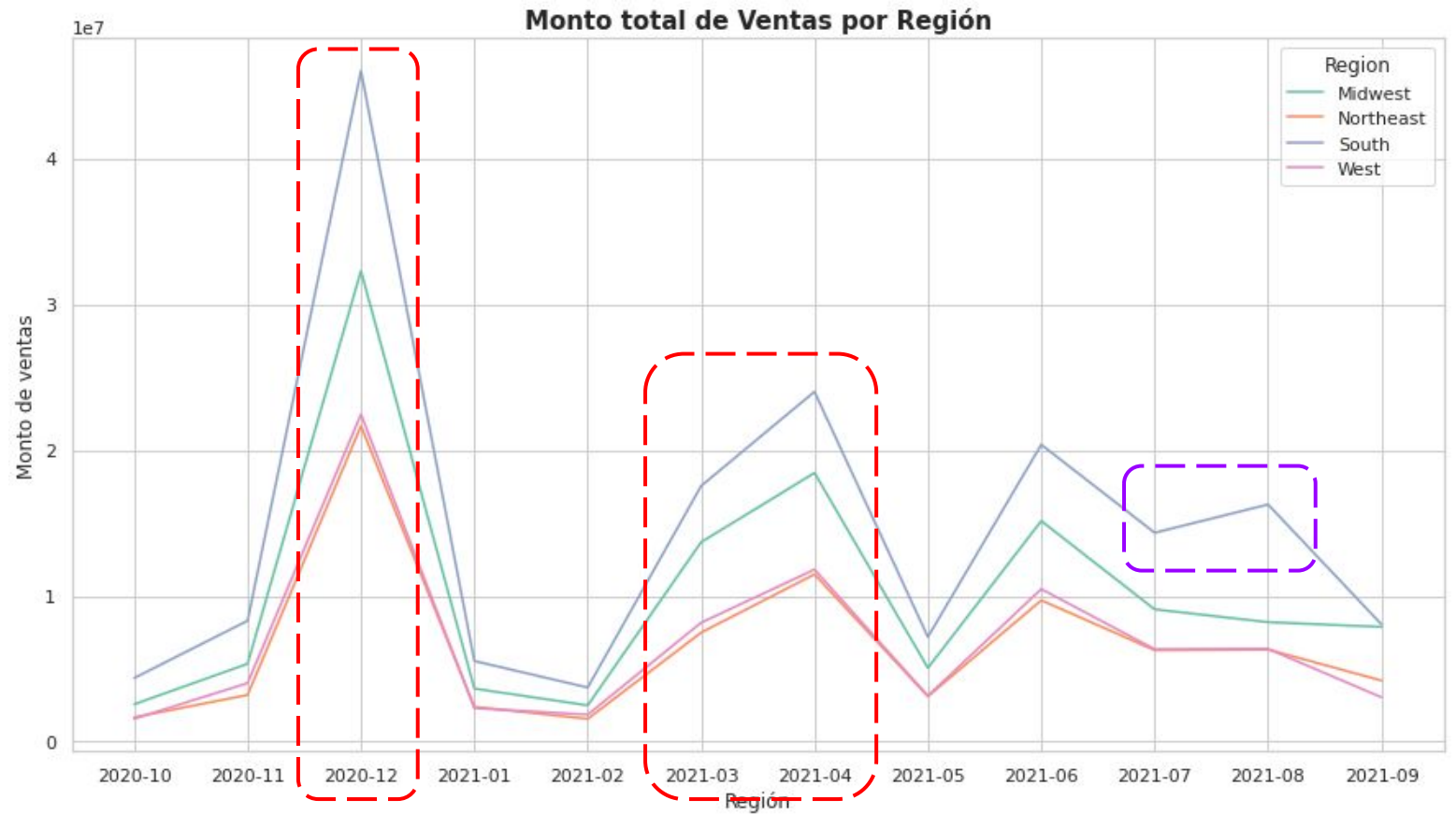
¿CUÁL ES LA REGIÓN CON MAYOR PARTICIPACIÓN EN VENTAS ONLINE?

- **South USA** es la región con mayor participación en órdenes on-line
- **South USA tiene un liderazgo sostenido** a lo largo de los períodos en análisis



¿CUÁL ES LA REGIÓN/ESTADO CON MAYOR INCREMENTO EN VENTAS ONLINE?

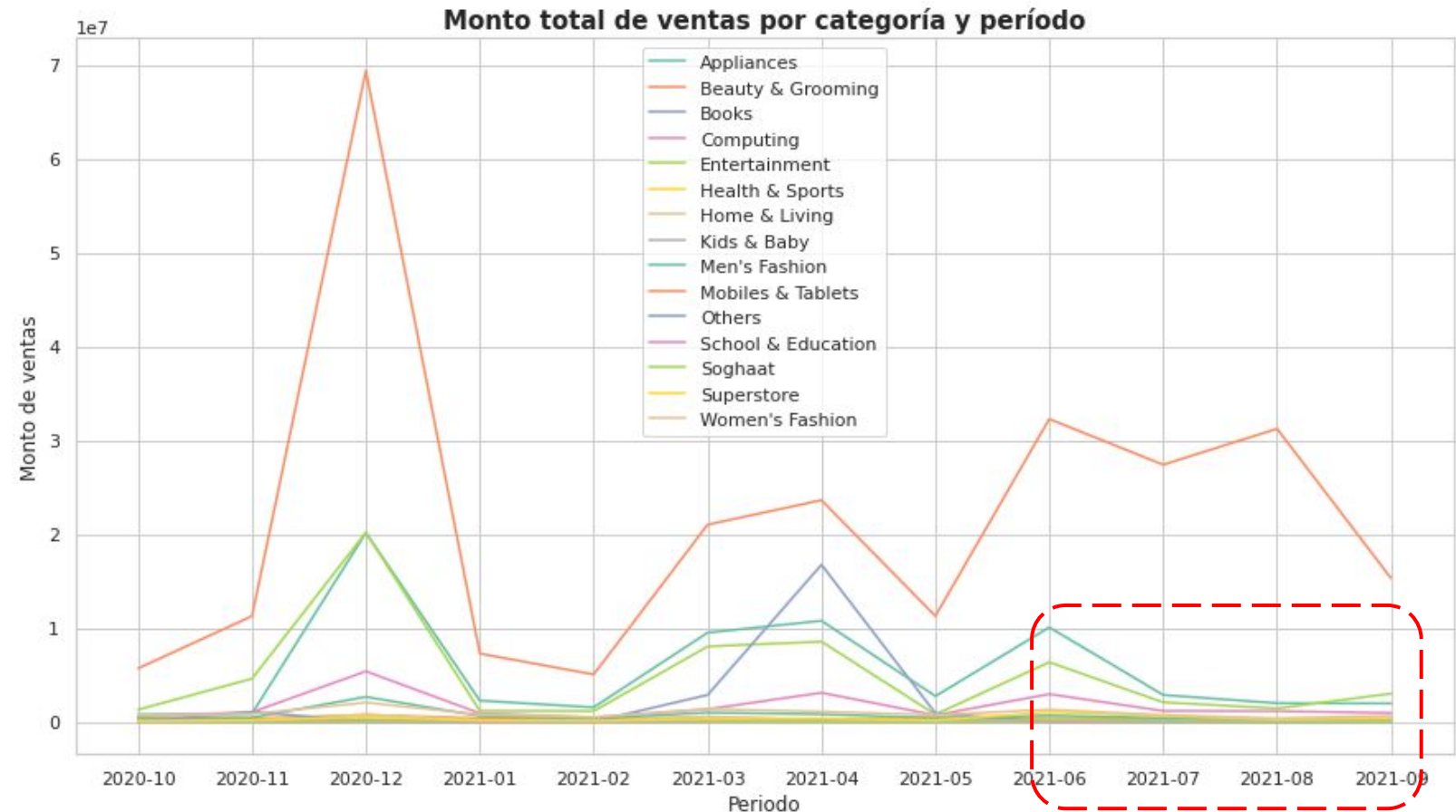
- Existen **2 picos** bien marcados en los meses de **Diciembre 2020** y **Marzo/Abril 2022**
- En la **región South** tenemos un **aumento de las ventas en monto total**, en el mes de **Agosto 2021** que no se observa en otras regiones.



¿CUÁLES SON LAS CATEGORÍAS CON TENDENCIA A INCREMENTAR O DISMINUIR SU CANTIDAD DE ÓRDENES ONLINE?

— **Mobiles & Tablets** es la dominante en sumatoria de ventas en comercio online

● **Books, Entertainment y Men's Fashion** tienen caídas de ventas a partir de 21-Q3

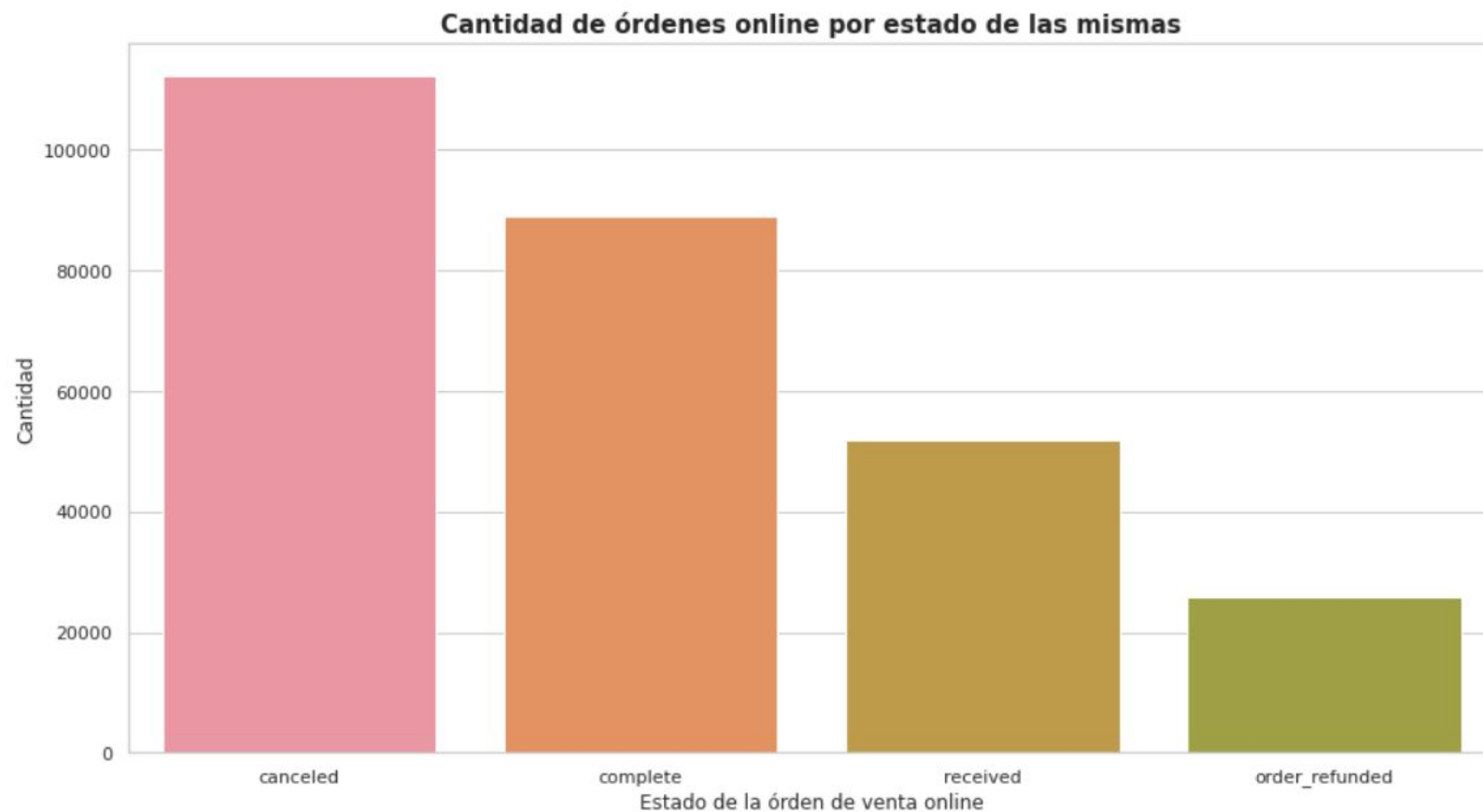


¿CÓMO ES LA SITUACIÓN DE LAS ÓRDENES DE COMPRA CANCELADAS?

El 37,2% de las órdenes están **CANCELADAS**, lo que la convierte en el status TOP1

31,1% están COMPLETAS
(-6,1pp)

A continuación haremos un deep dive sobre esta situación

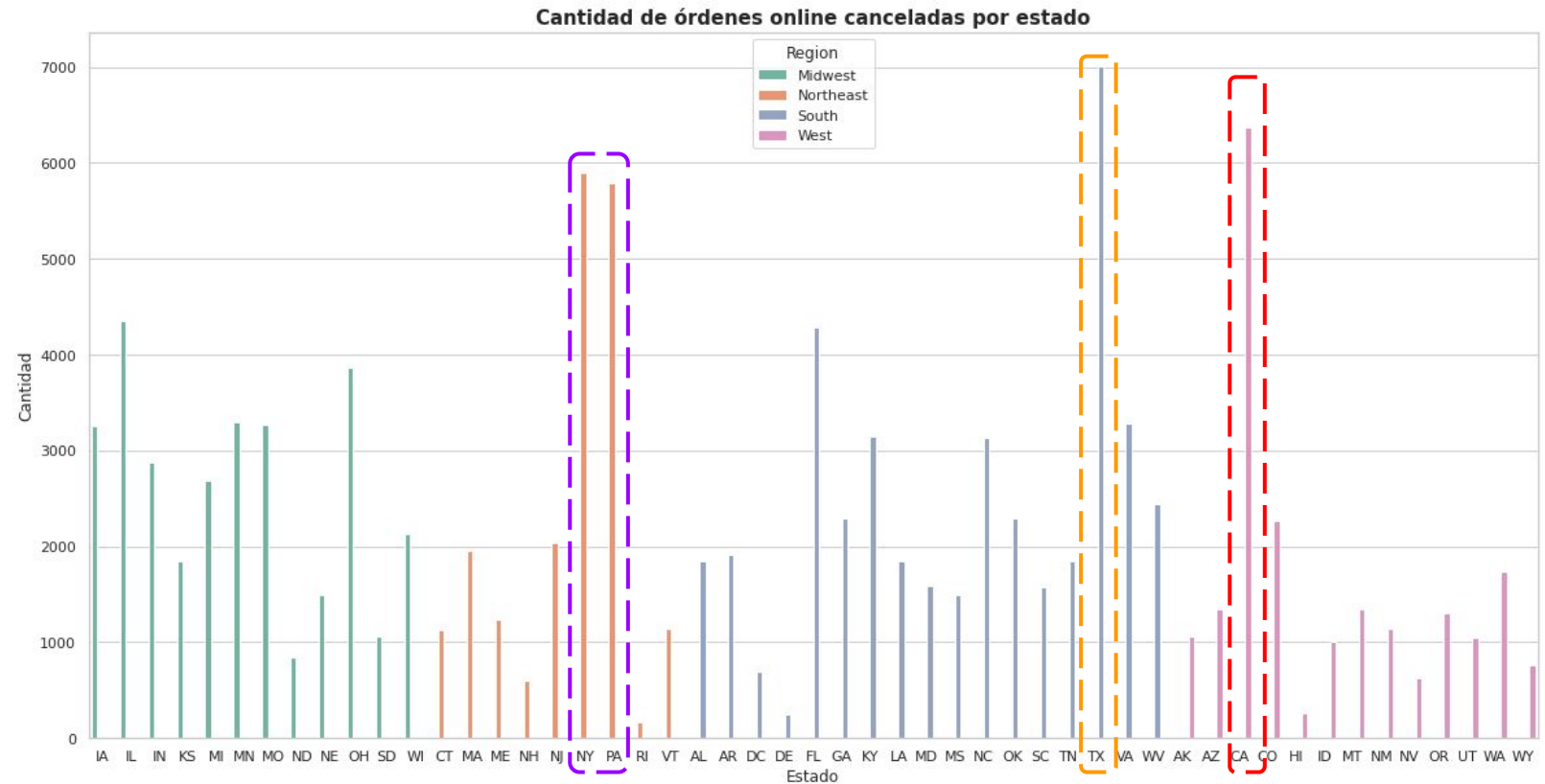


¿CÓMO ES LA SITUACIÓN DE LAS ÓRDENES DE COMPRA CANCELADAS? ¿SE PODRÍA IDENTIFICAR ALGÚN PATRÓN?

Si hacemos un corte por región, **no hay concentración de cancelaciones por región**, por lo que se descarta esta hipótesis

Sí tenemos **concentración de cancelaciones en algunos estados**, tal es el caso de:

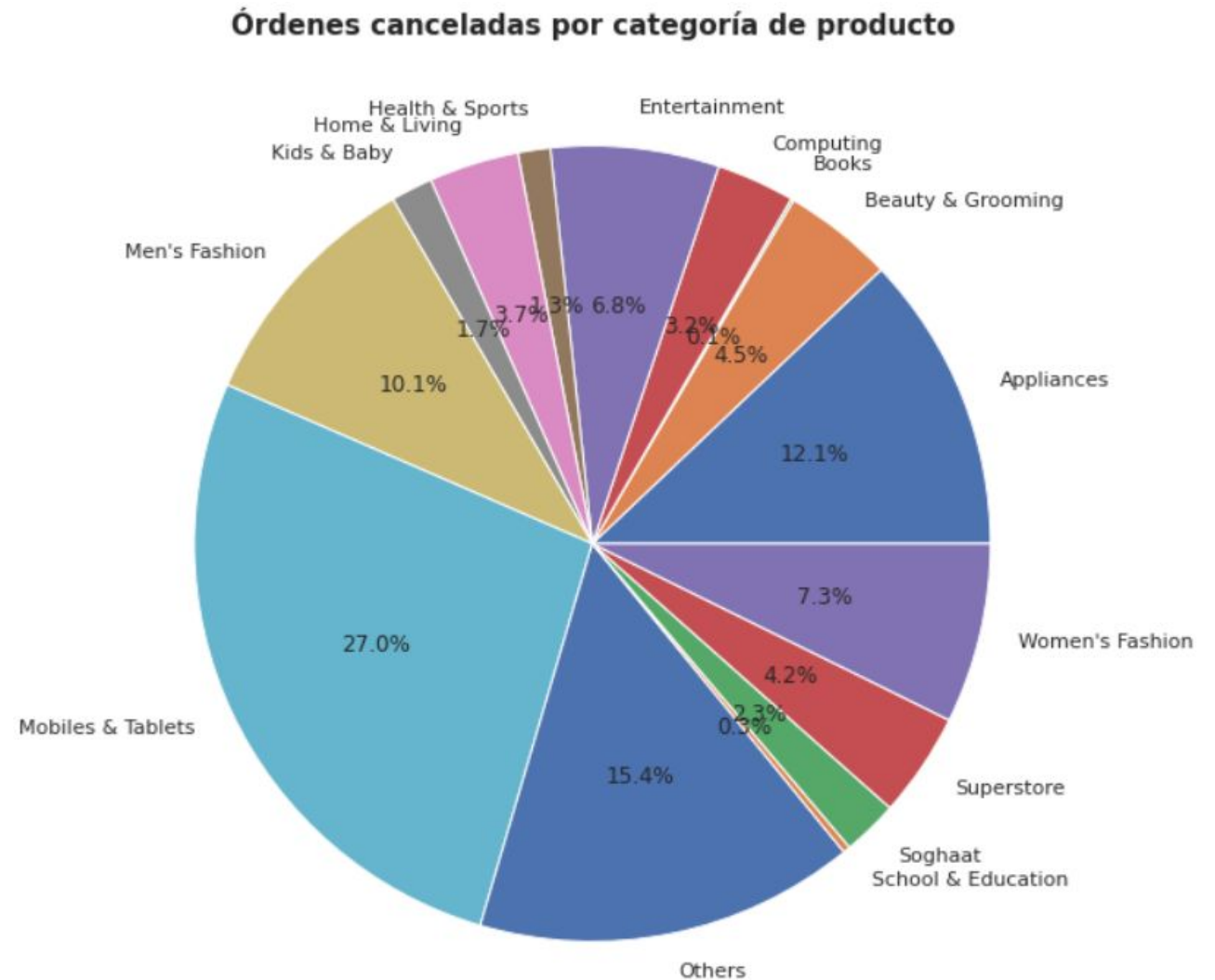
- CA (California - West)
- NY/PA (New York y Pensilvania - Northeast)
- TX (Texas - South)



¿SE PODRÍA IDENTIFICAR ALGÚN PATRÓN AL RESPECTO DE ÓRDENES DE VENTA CANCELADAS?

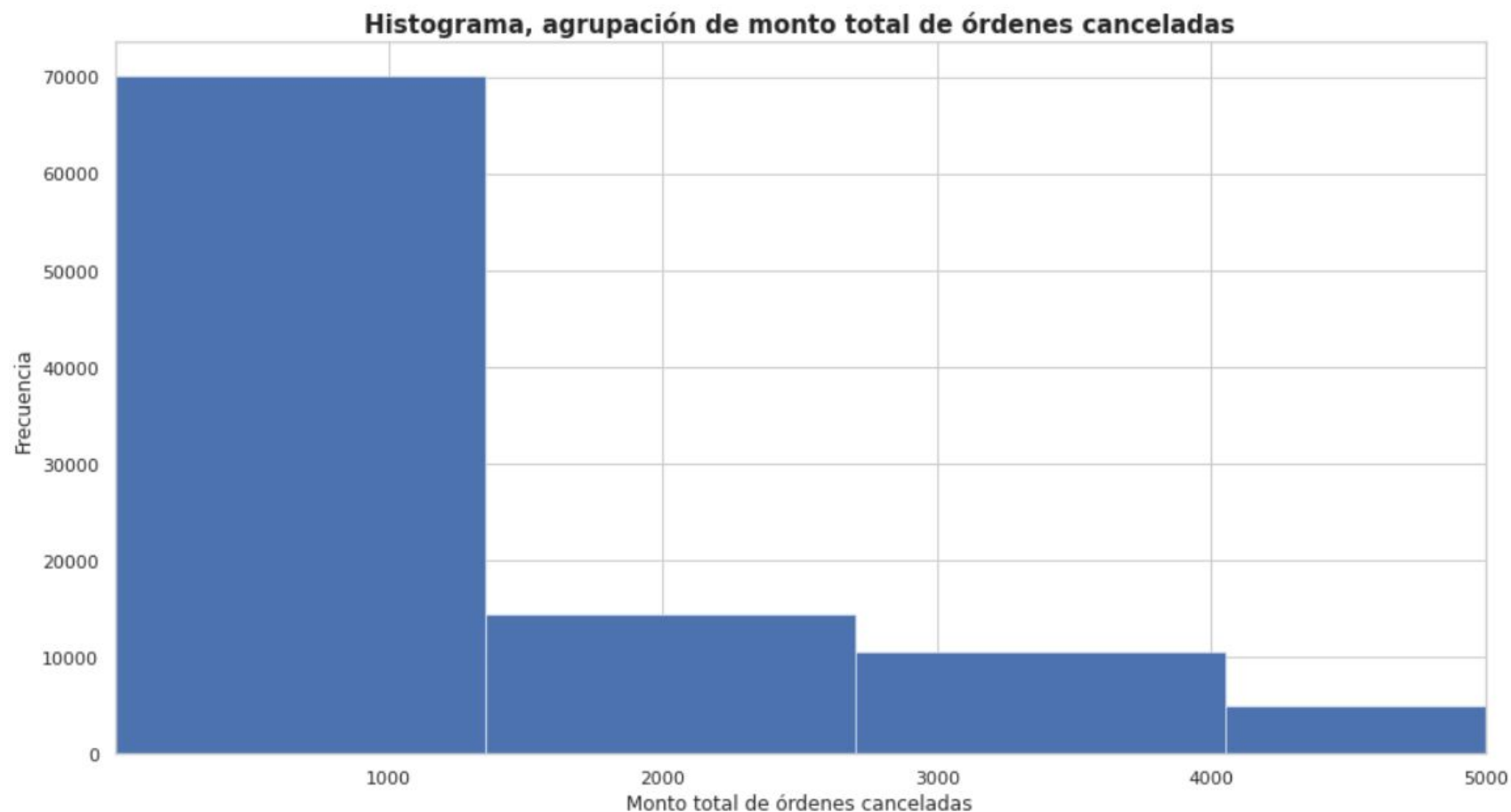
Mobiles & Tablets llega casi a 30% de sus órdenes canceladas

Tiene coherencia, ya que como veíamos antes es la categoría con mayor cantidad de ventas, es un dato importante que tenemos que considerar en el análisis por el tipo de producto



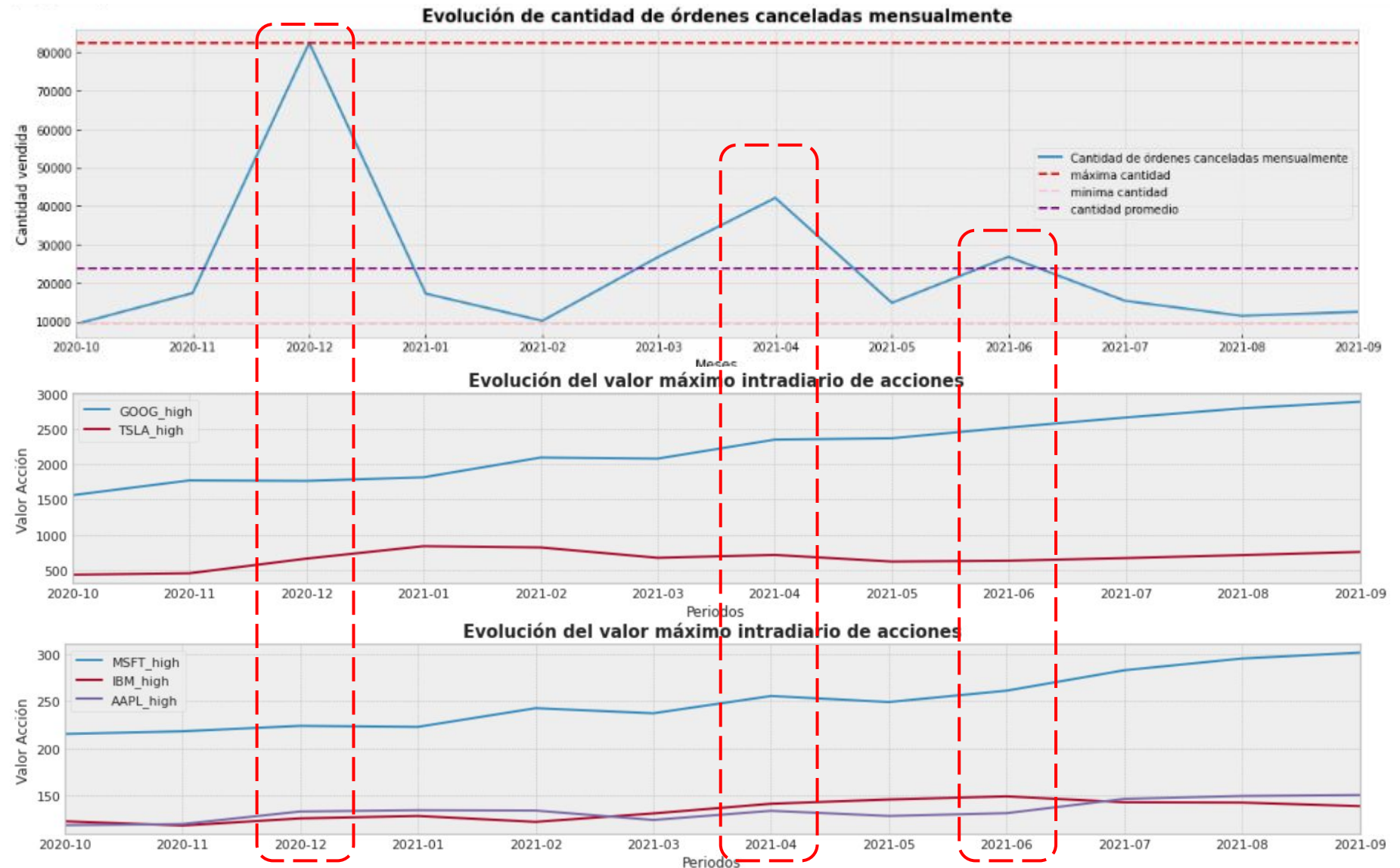
¿SE PODRÍA IDENTIFICAR ALGÚN PATRÓN AL RESPECTO DE ÓRDENES DE VENTA CANCELADAS?

Casi el **90%** de las **órdenes canceladas** fueron órdenes de un **valor inferior a los 2500 USD**



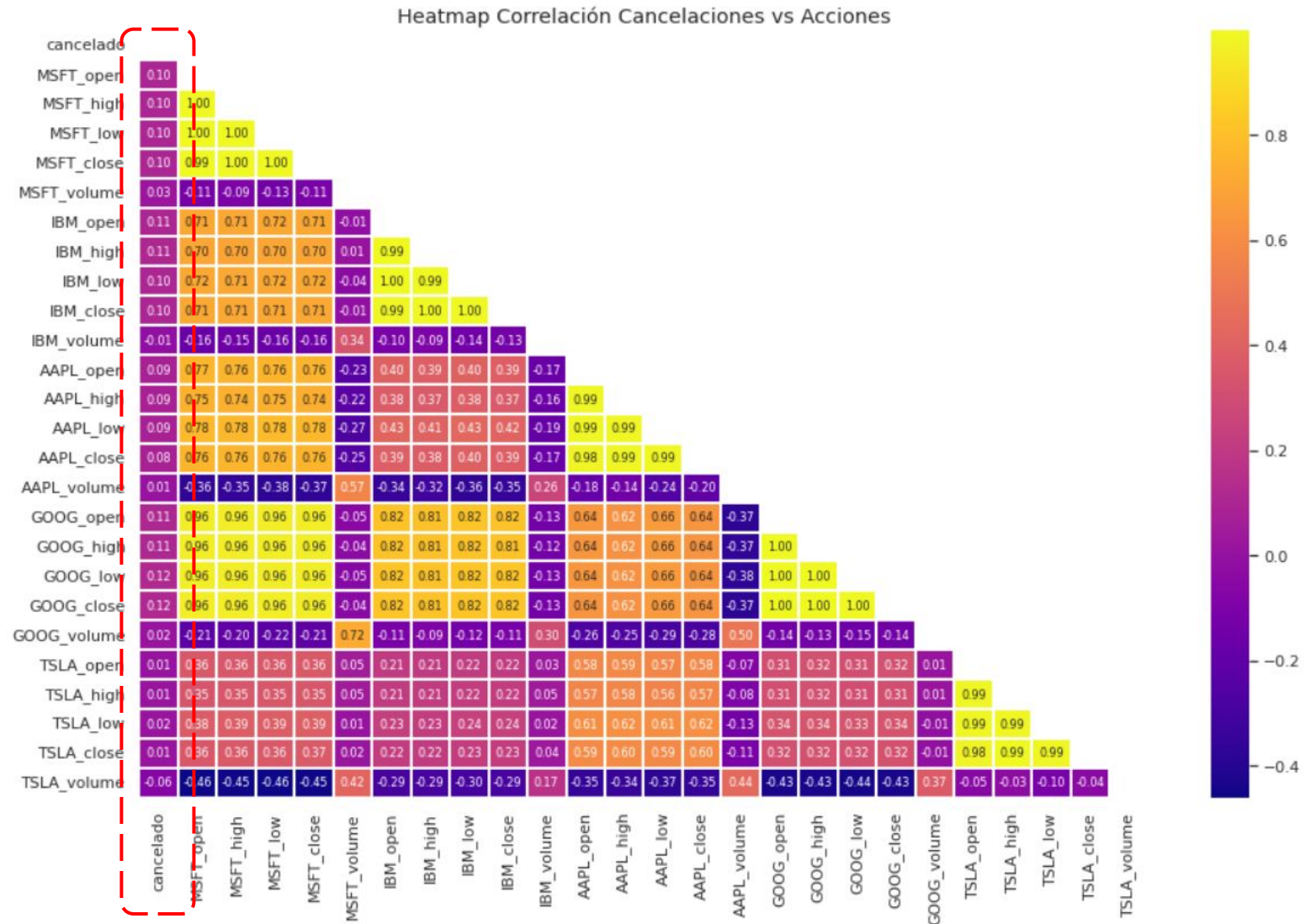
¿EXISTEN FACTORES EXTERNOS QUE PODRÍAN INCIDIR EN LA CANCELACIÓN DE ÓRDENES?

- No se observa una **correlación directa** entre la cantidad de **órdenes canceladas** mensualmente con las variables de referencia promedio de las **acciones de las principales compañías tecnológicas**



¿EXISTEN FACTORES EXTERNOS QUE PODRÍAN INCIDIR EN LA CANCELACIÓN DE ÓRDENES?

- En **el análisis OLS** que se realizó con todas las variables de las acciones (valor de apertura y cierre, máximo y mínimo) **también arrojó resultados negativos**
- En el heatmap vemos gráficamente mismo resultado: no hay correlación



¿CÓMO ES LA SITUACIÓN DE LAS ÓRDENES DE COMPRA CANCELADAS?

Resumen de la situación de cancelación de órdenes

- **No hay concentración de cancelaciones por REGION**
- **Existe una concentración de cancelaciones** en los **ESTADOS** de California, New York, Pensilvania y Texas, se solicitará al equipo comercial de estos estados más información para hacer un deep dive en este insight que queda como hipótesis de esta situación
- **Mobiles & Tablets** llega casi a **30% de sus órdenes canceladas**, también se solicitará más información al equipo comercial para validar o descartar esta hipótesis
- **90% de las órdenes canceladas** fueron órdenes de un **valor inferior a los 2500 USD**
- Factores externos como los **valores de acciones de las principales compañías tecnológicas**, no afectan a las cancelaciones de ventas online

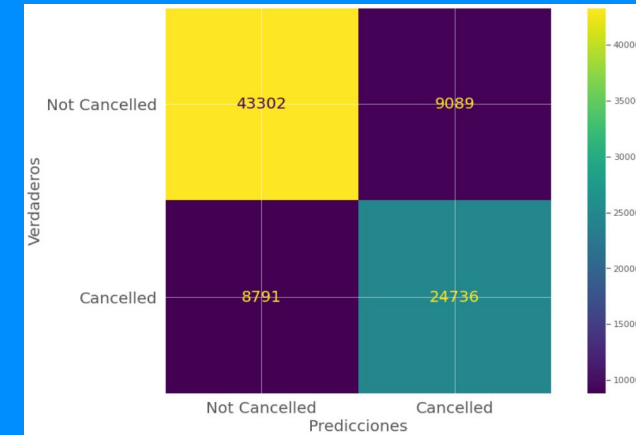
APLICACIÓN DE ALGORITMOS DE ML

DECISION TREE

El **objetivo** de este modelo es **lograr predecir la cancelación o no de una compra**, para que de esta forma la compañía pueda tomar las acciones necesarias de esta casuística para reducir costos operativos y maximizar márgenes de utilidad

Métricas de la performance del modelo:

- Accuracy: 79,18%
- Precision_positiva: 73%
- Precision_negativa: 83%
- Recall_sensibilidad: 74%
- Recall_especificidad: 83%



En función de los datos obtenidos, este modelo se ajusta a las necesidades de la compañía

BAYES NAIVE CLASSIFIER

El **objetivo** de este modelo es **comparar la performance de este con el Decision Tree** y elegir el que mejor se ajuste a las necesidades de la compañía

Métricas de la performance del modelo:

- Accuracy: 68,6%
- Precision_positiva: 62%
- Precision_negativa: 72%
- Recall_sensibilidad: 51%
- Recall_especificidad: 80%

El modelado no presenta una exactitud que consideramos aceptable; todas las métricas presentaron valores inferiores al modelado de clasificación a partir de Decision Tree

Claramente nos encontramos en un caso de underfitting dado que el modelo encuentra dificultad para encontrar patrones y predecir órdenes que a futuro se cancelarán

Se descarta este modelo y optamos por el modelo de Decision Tree para emplear en este caso de estudio

INSIGHTS &
NEXT STEPS

INSIGHTS

Mobiles & Tablets es la categoría dominante en sumatoria de ventas en comercio online

Books, Entertainment y Men's Fashion tienen caídas de ventas a partir de 21-Q3

El 37,2% de las órdenes están CANCELADAS

- **no hay concentración de cancelaciones por región**
- Existe una **concentración de cancelaciones en los ESTADOS de California, New York, Pensilvania y Texas**
- **90% de las órdenes canceladas** fueron órdenes de un **valor inferior a los 2500 USD**
- **Factores externos** como los valores de acciones de las principales compañías tecnológicas, **no afectan a las cancelaciones** de ventas online

NEXT STEPS

Realizar un **deep dive con el equipo comercial**, de lo sucedido en los **picos de ventas de Dic-2020 y Mar/Abr-2022**. Esto podría generar un insight para prepararnos mejor para futuros incrementos de ventas y aprovechar esta demanda estacional

Realizar un **workshop entre equipos comerciales regionales**, para **analizar la situación de órdenes de ventas de la región South y la concentración de cancelaciones en los ESTADOS de California, New York, Pensilvania y Texas**

- Existe una demanda particular en esta región?
- Tenemos diferentes estrategias que nos hacen captar mejor la demanda? Se pueden llevar estas estrategias al resto del país?

Implementar el modelo de Decision Tree para predecir futuras cancelaciones. Esto permitirá tomar las mejores acciones y reducir el impacto en los costos operativos y maximizar márgenes de utilidad de la compañía