



# ENTREGA FINAL PROYECTO OPAIN

---

Grupo 2

- EL AEROPUERTO EL DORADO, UNO DE LOS MÁS CONCURRIDOS DE AMÉRICA LATINA. CADA AÑO, MÁS DE 25 MILLONES DE PERSONAS PASAN POR SUS PUERTAS





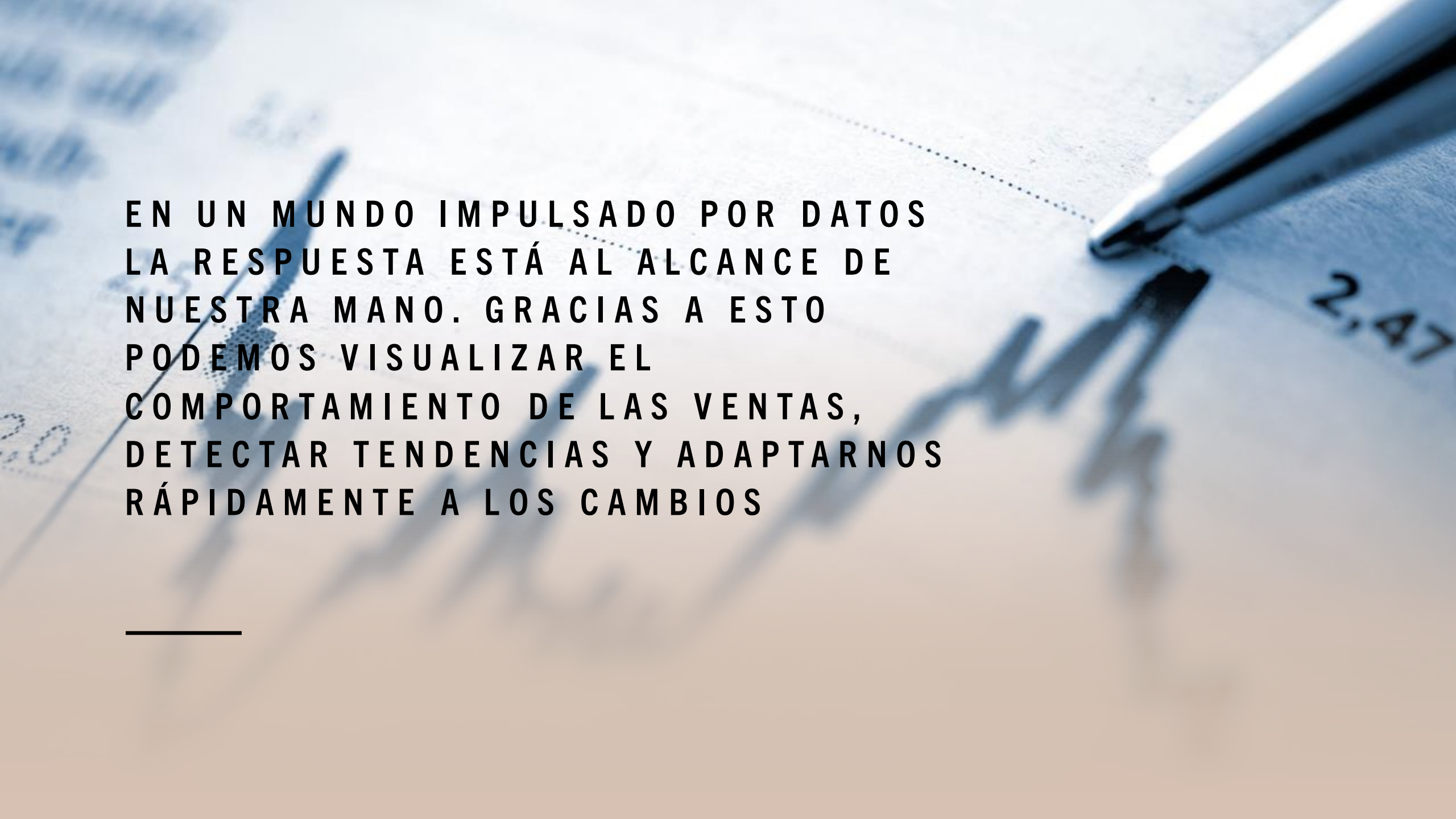


- CADA PERSONA TIENE UNA HISTORIA, UN DESTINO, UNA RAZÓN PARA VIAJAR. PERO MIENTRAS ESPERAN, MUCHOS APROVECHAN PARA HACER UNA PAUSA, COMER ALGO O LLEVARSE UN RECUERDO

- DESDE TIENDAS DE LUJO HASTA LOCALES DE COMIDA, CADA MARCA TIENE SU LUGAR EN EL AEROPUERTO. PERO ¿CUÁLES LA MARCA QUE MÁS VENDE? ¿QUÉ TIENDA TIENE MÁS TRÁFICO? ¿CÓMO SE COMPORTAN LAS VENTAS A LO LARGO DEL DÍA, LA SEMANA, EL AÑO?







EN UN MUNDO IMPULSADO POR DATOS  
LA RESPUESTA ESTÁ AL ALCANCE DE  
NUESTRA MANO. GRACIAS A ESTO  
PODEMOS VISUALIZAR EL  
COMPORTAMIENTO DE LAS VENTAS,  
DETECTAR TENDENCIAS Y ADAPTARNOS  
RÁPIDAMENTE A LOS CAMBIOS

---

# Y NUESTRO OBJETIVO ES MEJORAR LAS VENTAS DEL AEROPUERTO MÁS GRANDE DEL PAÍS

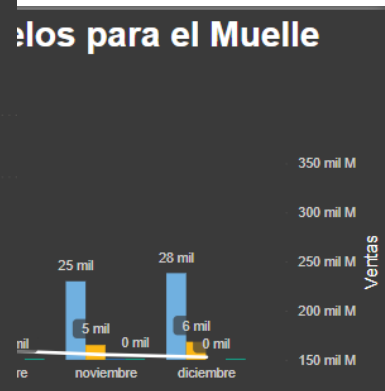
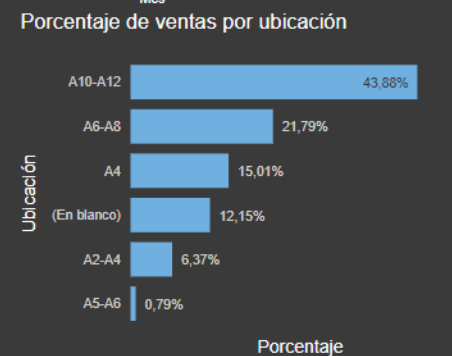
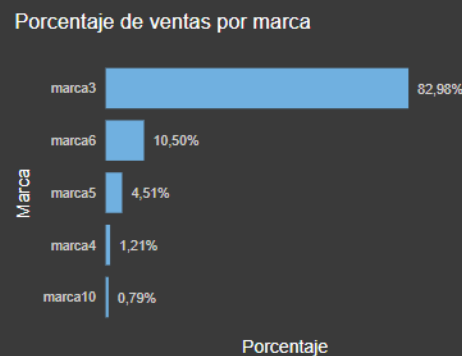
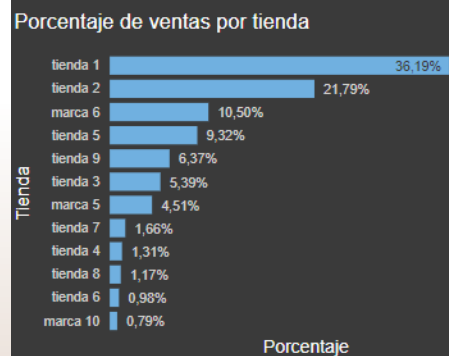


10 marcas

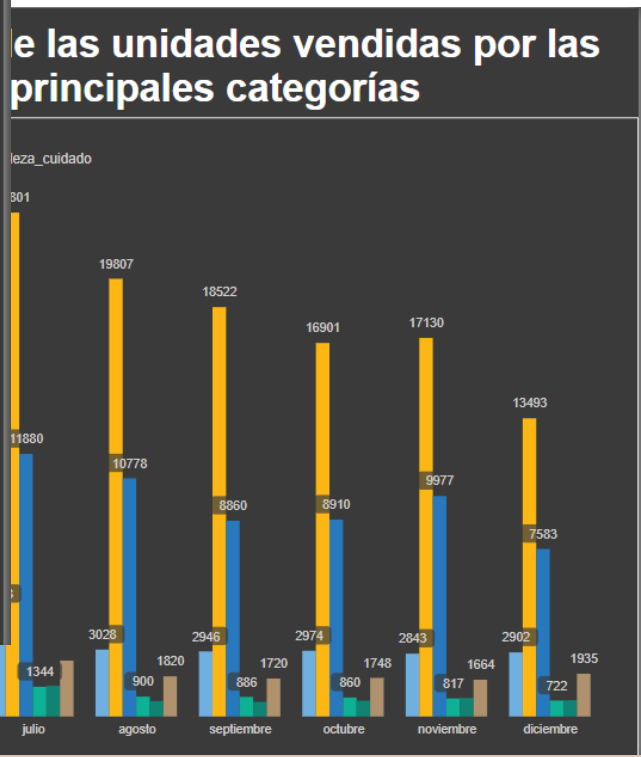
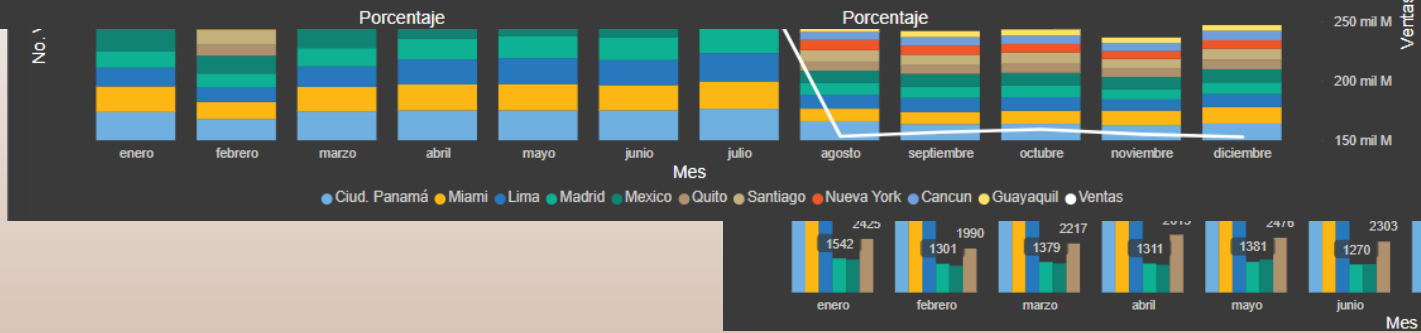


+10 tiendas

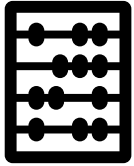
# ¿QUÉ NOS DIJERON LOS DATOS?



- [Link reporte archivo](#)
- [Link reporte Online](#)



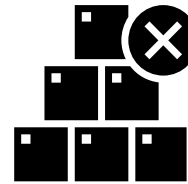
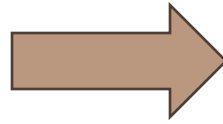
# PROCESO DE LIMPIEZA DE DATOS



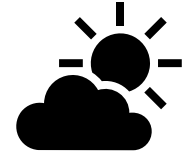
Análisis y limpieza  
datos de ventas



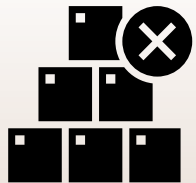
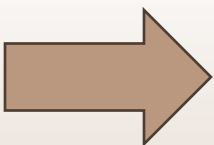
Unificación  
datos de las  
10 marcas



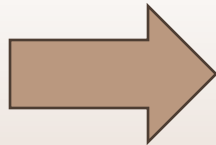
Creación  
columnas  
"dummies"



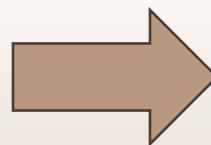
Añadir información  
del clima mediante  
un API



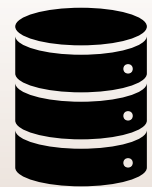
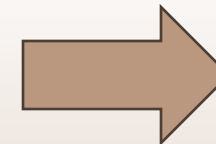
Reclasificar las  
categorías  
descriptivas



Nueva columna  
por ubicación de  
puerta A1 - A14



Limpieza y se  
añaden los datos  
de los vuelos  
comerciales



Dataset final



# DECISIONES SOBRE LOS DATOS FINALES



Se decidió tener en cuenta para el modelo con categorías solo las marcas que tienen información relacionada con las categorías o descripciones de las ventas. (Marca 3,4,5 y 6).



Para el modelo de ventas sin categorías contiene a las 10 marcas sin la información de las categorías.



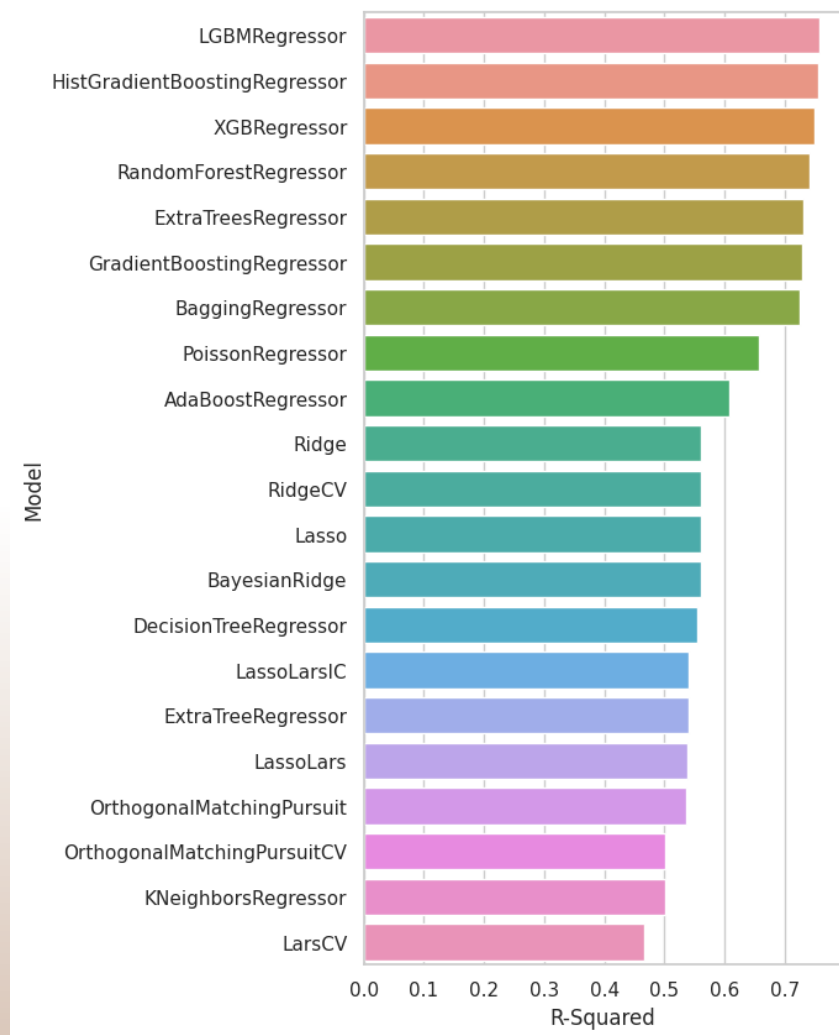
Se decidió solo tener en cuenta los datos del 2022 para el entrenamiento y para los datos de prueba se tuvieron en cuenta los datos del 2023.



Se re agruparon las categorías dadas por las marcas para poder así tener un poco más de significado y relacionar las ventas entre las marcas. Inicialmente teníamos alrededor de 223 categorías ( 120 categorías descriptivas y 103 categorías no descriptivas. Finalmente , se redujo esta cantidad a 29 categorías descriptivas y 103 no descriptivas.

# ¿CÓMO LO HICIMOS?

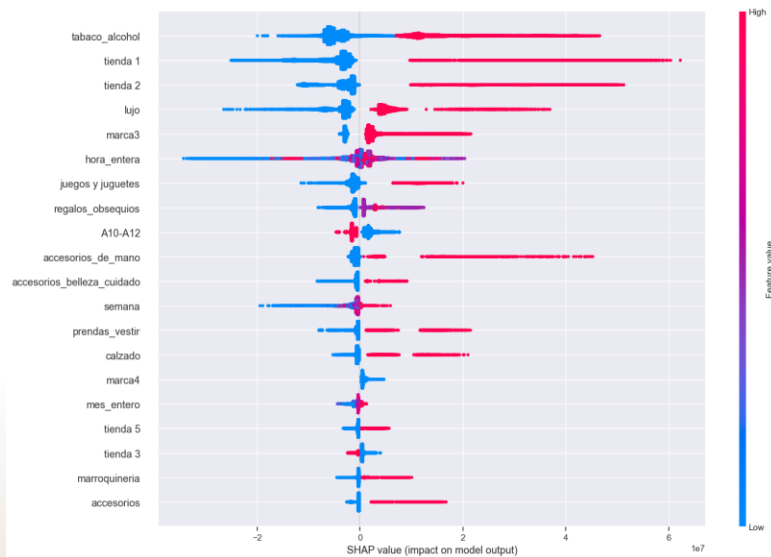
## SELECCIÓN MODELO BASE



# ¿CÓMO LO HICIMOS?

## MODELOS ANALIZADOS (LGMB REGRESSOR)

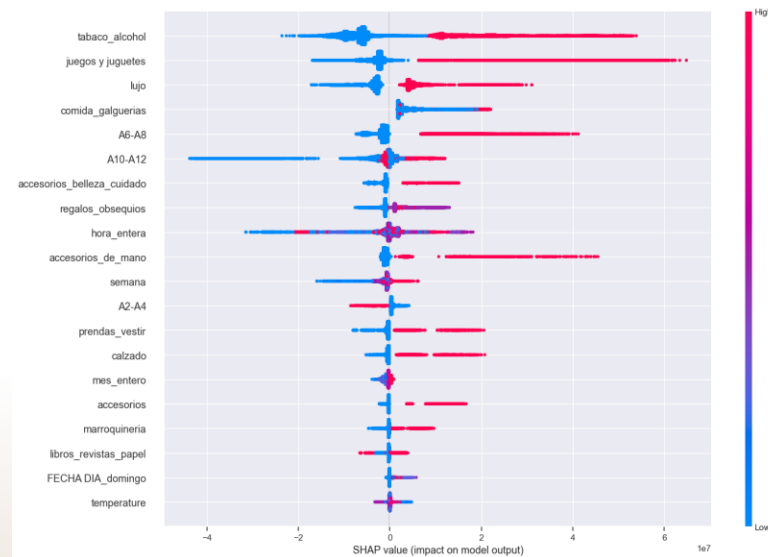
Todas las variables



R2 train=0,84

R2 test=0,75

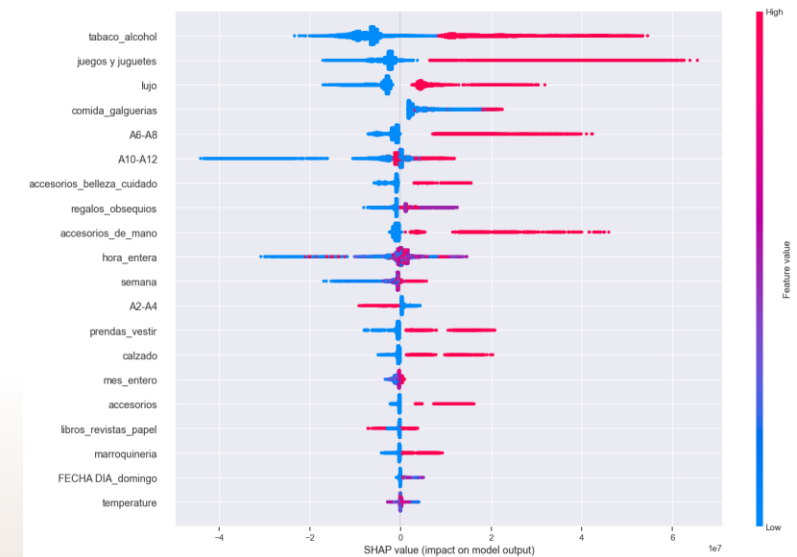
Sin variables de tienda, categorías no especificadas, marcas y aerolíneas



R2 train=0,79

R2 test=0,7

Sin variables de tienda, categorías no especificadas ni marca



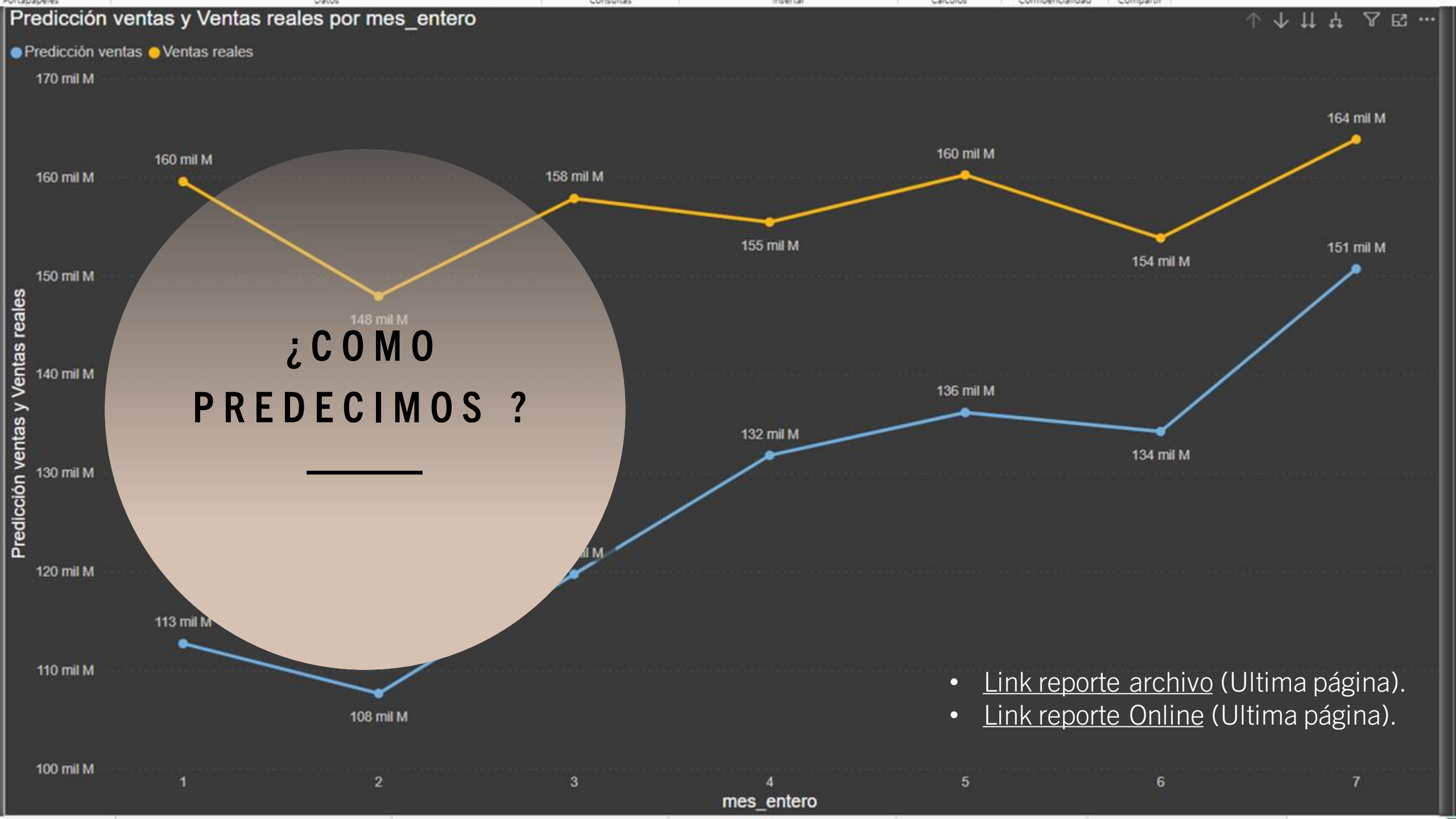
R2 train=0,79

R2 test=0,7



# RESULTADOS Y MÉTRICAS

Modelo	MAE Train	MAE Test	RMSE Train	RMSE Test	R2 Train	R2 Test
Todas las variables	7808426.93	11216316.69	15335847.97	20567723.23	0.84	0.75
Sin variables de tienda, categorías no especificadas ni marca	8896343.74	12128998.44	17769446.63	22354600.02	0.79	0.7
Sin variables de tienda, categorías no especificadas, marcas y aerolíneas	8907448.70	12131249.39	17786849.75	22361158.30	0.79	0.7



# ¿QUÉ CONCLUIMOS A PARTIR DE LOS DATOS Y EL MODELO REALIZADO?



Los años 2022 y 2023 son comparables hasta cierto punto, ya que del 2023 por cuestiones de temporalidad no tendremos todos los datos disponibles.



Las marcas que representan el mayor porcentaje de ventas (aprox. 76%) corresponde a las marcas 3 y 6. Igualmente, las mayores ventas son por el tipo de tienda Duty Free.



Las categorías con mayor influencia en la venta son: tabaco y alcohol, juegos y juguetes, lujo, regalos y obsequios, y accesorios de mano y de la belleza y el cuidado.



Las zonas/ubicación en el muelle internacional que más influyen de forma positiva son: A6-A8 y A10-A12. Mientras que la zona A2-A4 tiene una relación inversa a las ventas.



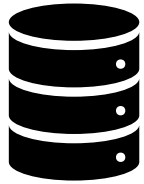
Los vuelos a horas tempranas tienen una relación negativa con las ventas, es decir entre las 0-3 de la mañana. Adicional, se tienen picos de venta entre las 5-8, las 10-15 y las 19-22.



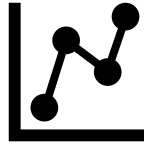
Los días de la semana de mayor venta son: el domingo, el lunes y el sábado.



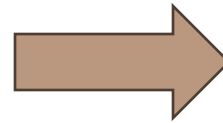
# ¿COMO PREDECIMOS?



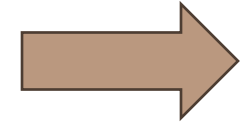
Datos finales



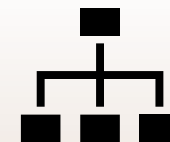
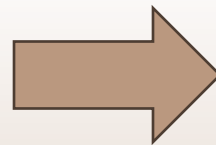
Generación  
modelo final



Predicción  
modelo



Creación API  
modelo escogido



Predicciones  
mostradas en  
aplicativo

# CONCLUSIONES

- Las predicciones y presentadas por el modelo desarrollado son de utilidad para Opain ya que pueden optimizar la cantidad de ventas basados en las tendencias de compra de los viajeros del Aeropuerto El Dorado.
- Este modelo presentado puede servir como un sistema de toma de decisiones para Opain, generando seguridad a la hora de reubicar tiendas, modificar productos y/o dar paso a nuevas alianzas con compañías.



## NEXT STEPS

- Complementar el modelo con mayor información temporal, con el detalle de las categorías que no presentaban una clasificación y nuevas variables que aporten al entendimiento de la dinámica de las ventas dentro del aeropuerto el Dorado.