



Tiempos recomendados de visita al cliente:

Análisis y modelado para recomendar la duración óptima de visitas de vendedores según la región, utilizando datos comerciales y generando un ranking regional de prioridad.

CENTRAL

cliente_id	canal	score	minutos recomendados
c31b7363...	AUTOSERVICIO	1,00	256,93
3cfa4e46f...	AUTOSERVICIO	0,99	255,16
f90ea870c...	AUTOSERVICIO	0,98	250,53
b453557e...	AUTOSERVICIO	0,97	250,27
ddc3dd80...	AUTOSERVICIO	0,96	245,58

LITORAL

cliente_id	canal	score	minutos_recomendados
4feda32ae8...	AUTOSERVICIO	1,00	382,76
97a3e6ef2f0...	AUTOSERVICIO	0,89	339,85
afb4aa1504...	AUTOSERVICIO	0,81	309,37
09a2779cb8...	AUTOSERVICIO	0,78	299,23
719c9d3e25...	AUTOSERVICIO	0,74	281,57

NEA

cliente_id	canal	score	minutos_recomendados
0719921e7...	KIOSCOS/MAXI...	1,00	0,06
08a0d1ca1...	KIOSCOS/MAXI...	1,00	0,06
08ddde490...	KIOSCOS/MAXI...	1,00	0,06
0999610ea...	KIOSCOS/MAXI...	1,00	0,06
0a78a8660...	KIOSCOS/MAXI...	1,00	0,06

GBA MINORISTAS

cliente_id	canal	score	minutos_recomendados
169357b19...	AUTOSERVICIO	1,00	243,02
0d88a0117...	AUTOSERVICIO	0,97	234,92
da55b0005...	AUTOSERVICIO	0,97	234,88
c17ad82f4...	AUTOSERVICIO	0,97	234,52
5884a7572...	AUTOSERVICIO	0,96	222,52

COSTA ATLÁNTICA

cliente_id	canal	score	minutos recomendados
6e88b7ff63...	AUTOSERVICIO	1,00	318,03
1cc7b69a5b...	AUTOSERVICIO	0,00	218,67
11af841178...	AUTOSERVICIO	0,00	218,36

NOA

cliente_id	canal	score	minutos recomendados
c36652f64e...	AUTOSERVICIO	1,00	255,34
670c94ea9...	AUTOSERVICIO	0,88	225,22
0b59350b6...	AUTOSERVICIO	0,88	225,18
dfbc6bb41...	AUTOSERVICIO	0,88	224,60
6a5b20004f...	AUTOSERVICIO	0,86	220,27

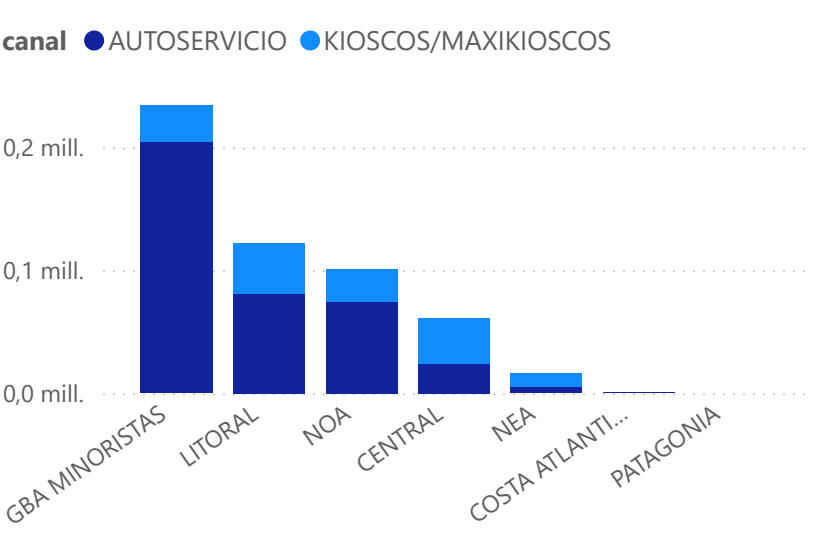
PATAGONIA

cliente_id	canal	score	minutos_recomendados
388308f0923...	AUTOSERVICIO	0	16,24

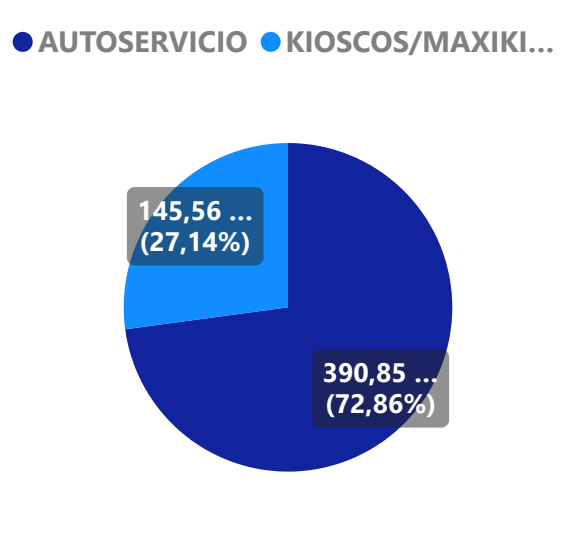
Tiempos recomendados de visita al cliente:

Análisis Exploratorio de los datos

VENTAS POR REGION

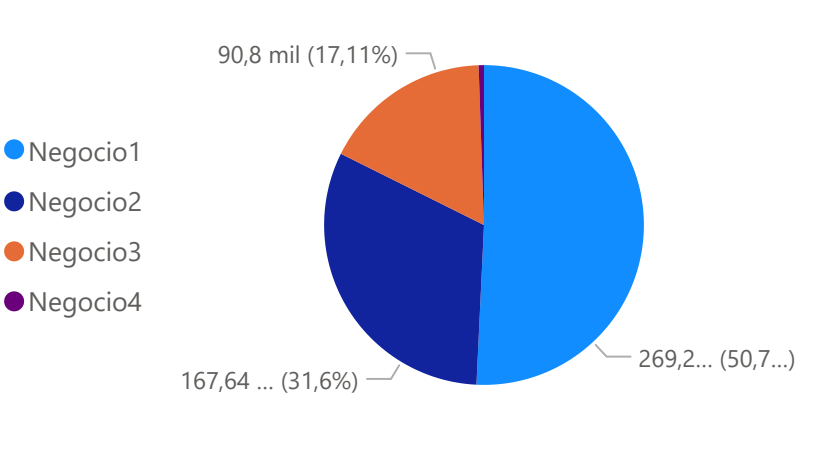


VENTAS POR CANAL



RANKING DE CLIENTES

RankingVentas	cliente_id	Venta total
1	4287a574ace354de2bdb68b16cff63df	7.088,26
2	a8d7a94c59bd58456b330d1d496227ae	5.756,95
3	7c74aa66e32a6ab05b043b02cee501bb	5.610,49
4	cfdd0ec5a9c983616e7813cf9607ba92	4.317,68
5	4deed4cc9373a0c4fb3e07c8dc0e57a2	4.233,58
6	5e36317c7dabe9ba2764f	4.178,96



Venta Total (HI)
536 mil

Venta Negocio 1 (HI)
269 mil

Cientes Únicos
14034

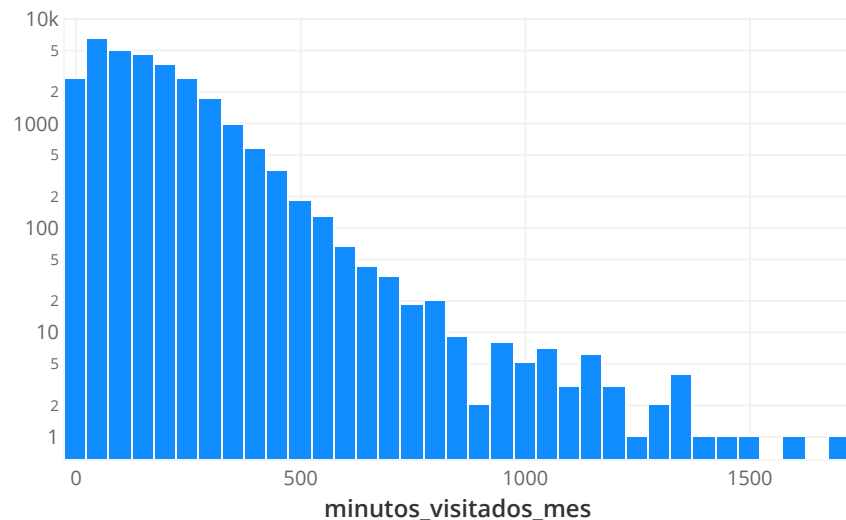


Tiempos recomendados de visita al cliente:

Análisis Exploratorio de los datos

Limpieza del Dataframe:

- Eliminación de duplicados
- Se eliminan datos del primer mes registrado (falta de datos históricos)
- Estandarizan variables categóricas
- Reemplazan placeholders (-999999 / 999999) por NaN
- Se eliminan filas con NaN en el target
- Se imputan negativos que son aproximadamente 0 por 0
- Se suman datos de clientes que se registran dos veces en un mismo mes



Se crean nuevas features comerciales:

- **venta_por_heladera** : venta total en función de la cantidad de heladeras
- **ratio_negX** : venta del negocio X sobre la venta total
- **productos_por_compra** : cantidad de productos/cantidad de compras
- **venta_promedio_x_compra**: venta total / cantidad compras total
- **flag_ventas**: 0 si la suma es 0, y si no

Se observan outliers en ventas y en la variable a predecir



Grandes valores en ventas y en target -> Clientes estratégicos
Ventas negativas -> devoluciones ?
Muchos minutos en clientes con pocas ventas ?



Tiempos recomendados de visita al cliente:

Modelo de Machine Learning

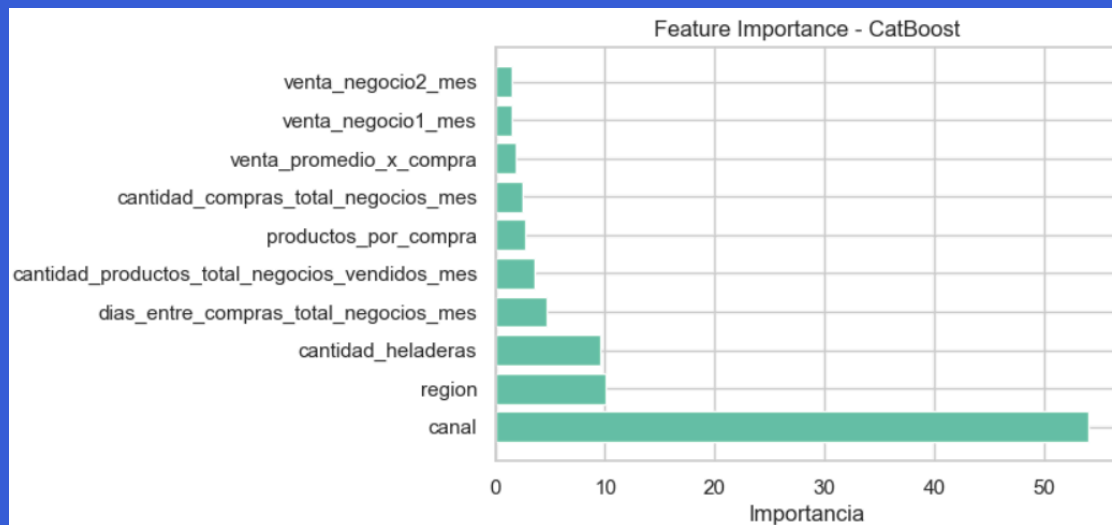


MODELO PROPUESTO: CatBoost

CatBoost es un algoritmo de boosting de árboles basado en gradiente que maneja muy bien variables categóricas y es robusto frente a datos desbalanceados y outliers.

Datos y problema

- Variable objetivo: minutos visitados por mes.
- El target presenta una distribución sesgada con valores extremos (outliers).
- Baseline: predecir la media



Métrica	Valor_train	Baseline_train	Valor_valid	Baseline_valid	Valor_test	Baseline_test
R2	0,58	0,00	0,68	0,00	0,62	0,00
MAE	26,15	50,90	22,56	46,85	26,37	51,70
RMSE	51,13	79,15	38,67	68,39	46,00	74,94

MAE : unidad minutos

Un ahorro en tiempo del 50 % comparado con el modelo Baseline



Tiempos recomendados de visita al cliente:

Conclusiones

- El modelo reduce el error absoluto (MAE) en aproximadamente un 50% respecto al baseline.
- El coeficiente de determinación R^2 indica que el modelo explica entre 58% y 68% de la variabilidad, una mejora significativa.
- El error absoluto promedio es alrededor de 26 minutos, razonable considerando que la media del target es 57 minutos.
- El modelo muestra robustez frente a outliers, pero su desempeño puede mejorar con transformaciones o manejo específico de valores extremos.

Técnicas probadas sin mejora significativa

- Aplicación de transformación logarítmica al target y a features numéricas muy sesgadas.
- Uso de clustering previo simple (KMeans) para segmentar datos antes del modelado.
- Otras técnicas para mitigar outliers (segmentación por percentiles).

Estas técnicas no mejoraron significativamente el desempeño del modelo en el tiempo disponible



Tiempos recomendados de visita al cliente:

Conclusiones

Limitaciones y próximos pasos

- El target y las features presentan valores extremos que afectan la estabilidad y precisión del modelo.
- Para futuras mejoras se propone:
 - Generar un nuevo modelo que prediga prioridad de visita en función de las ventas totales y los minutos
 - Explorar transformaciones más avanzadas o adaptativas del target y features.
 - Implementar técnicas más sofisticadas para detección y manejo de outliers.
 - Experimentar con segmentación y modelos especializados por grupos.
 - Afinar hiperparámetros y arquitectura del modelo con mayor profundidad
 - Entrenar distintos modelos sectorizando a los clientes según sus ventas históricas

Conclusión Final

Con 4 días de desarrollo se logró un modelo sólido que mejora notablemente el baseline trivial, con buen desempeño en validación y test. El trabajo sienta bases para mejoras futuras orientadas a manejar valores extremos y optimizar la predicción.

