



Mejor en el proceso de entrega de producto

Eliana Pinto

Visit Our Website
olistcompany.com

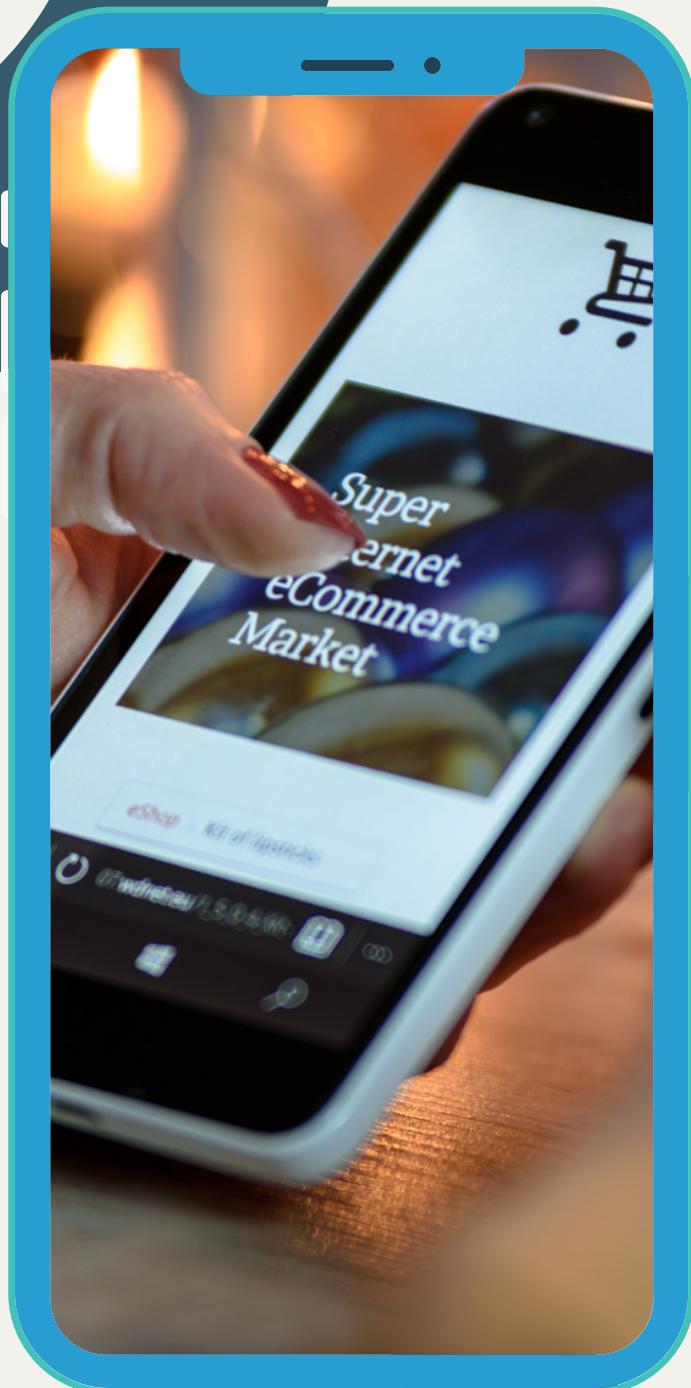


Objetivo

Es una plataforma donde los negocios pueden hacer el proceso de venta de sus productos a través de múltiples plataformas. Esta plataforma asiste a las empresas en el manejo de catálogo, manejar órdenes, marketing y servicio al cliente.

Factores a mejorar en las empresas para ofrecer un mejor servicio.





Mejor Modelo para predecir Score en la entrega de productos.



Aprobacion de producto



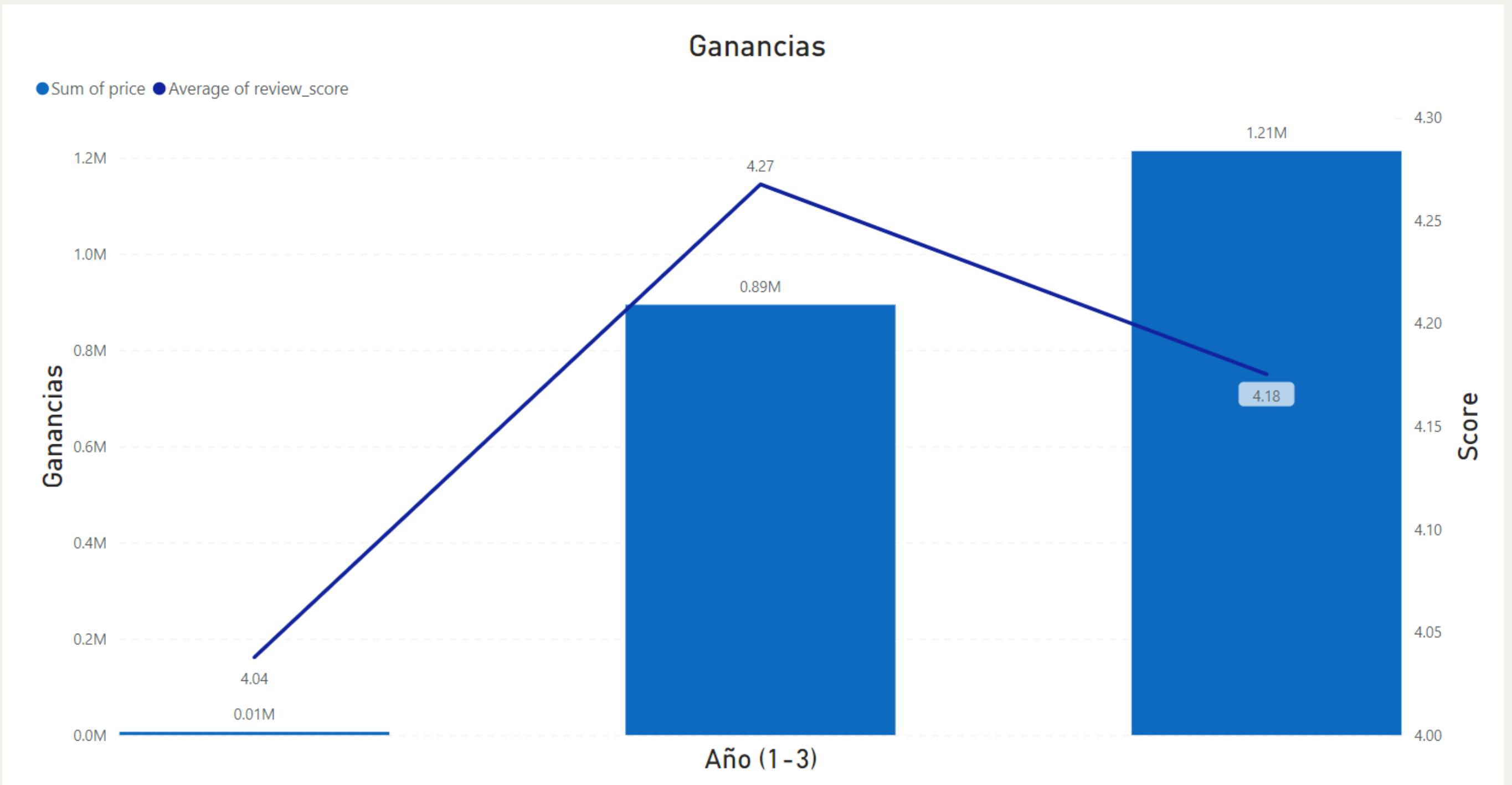
Envio



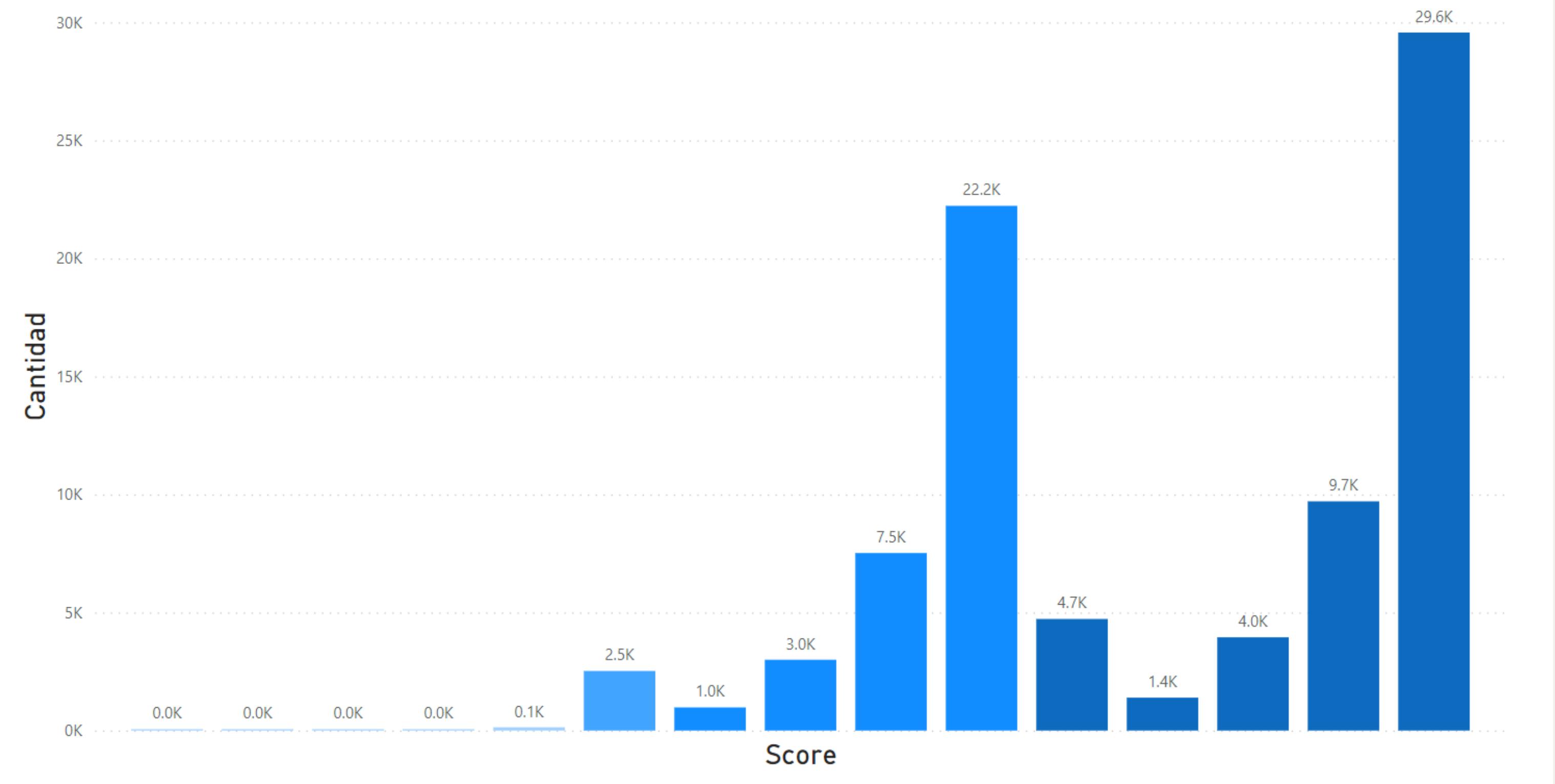
Entrega



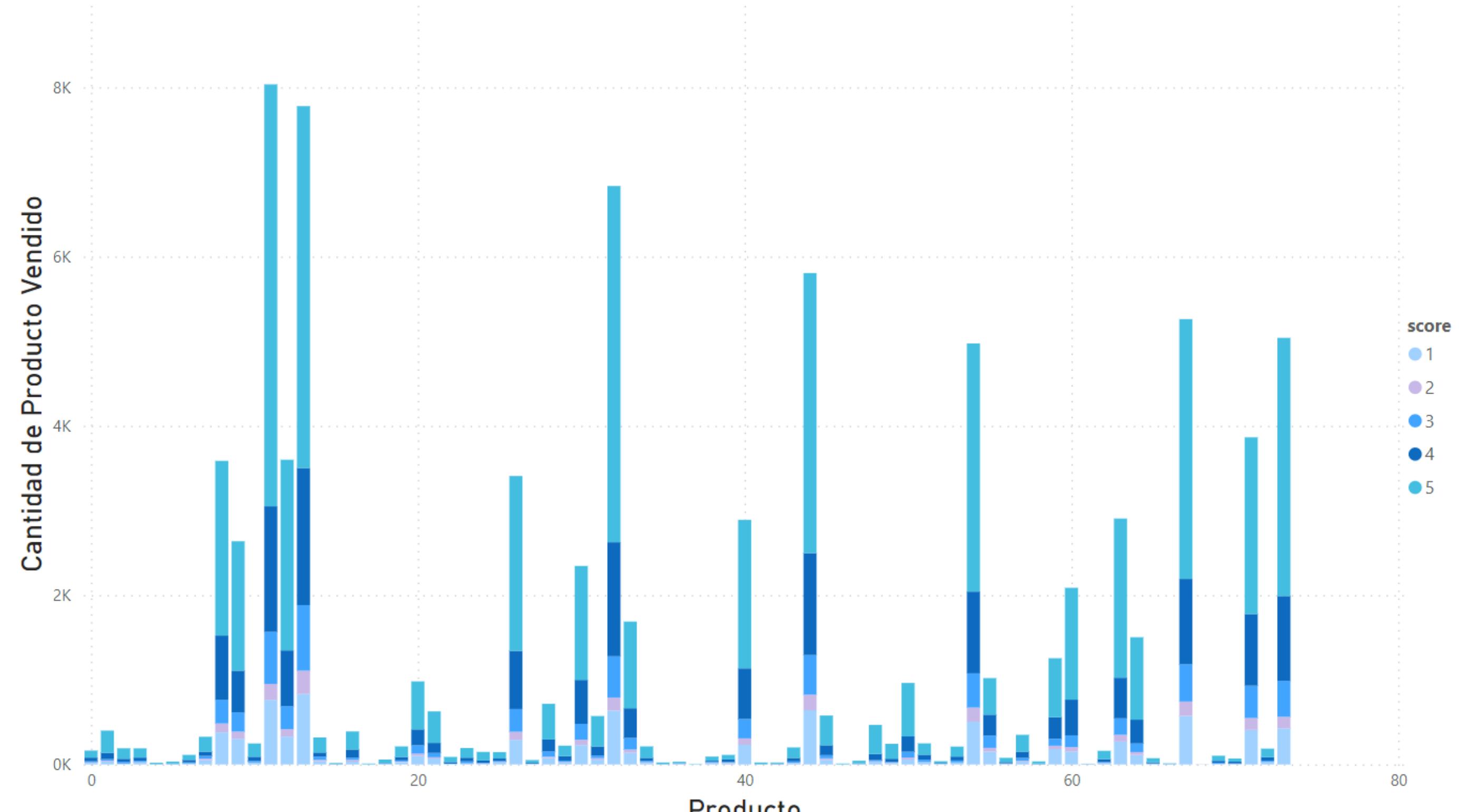
Calificacion del servicio

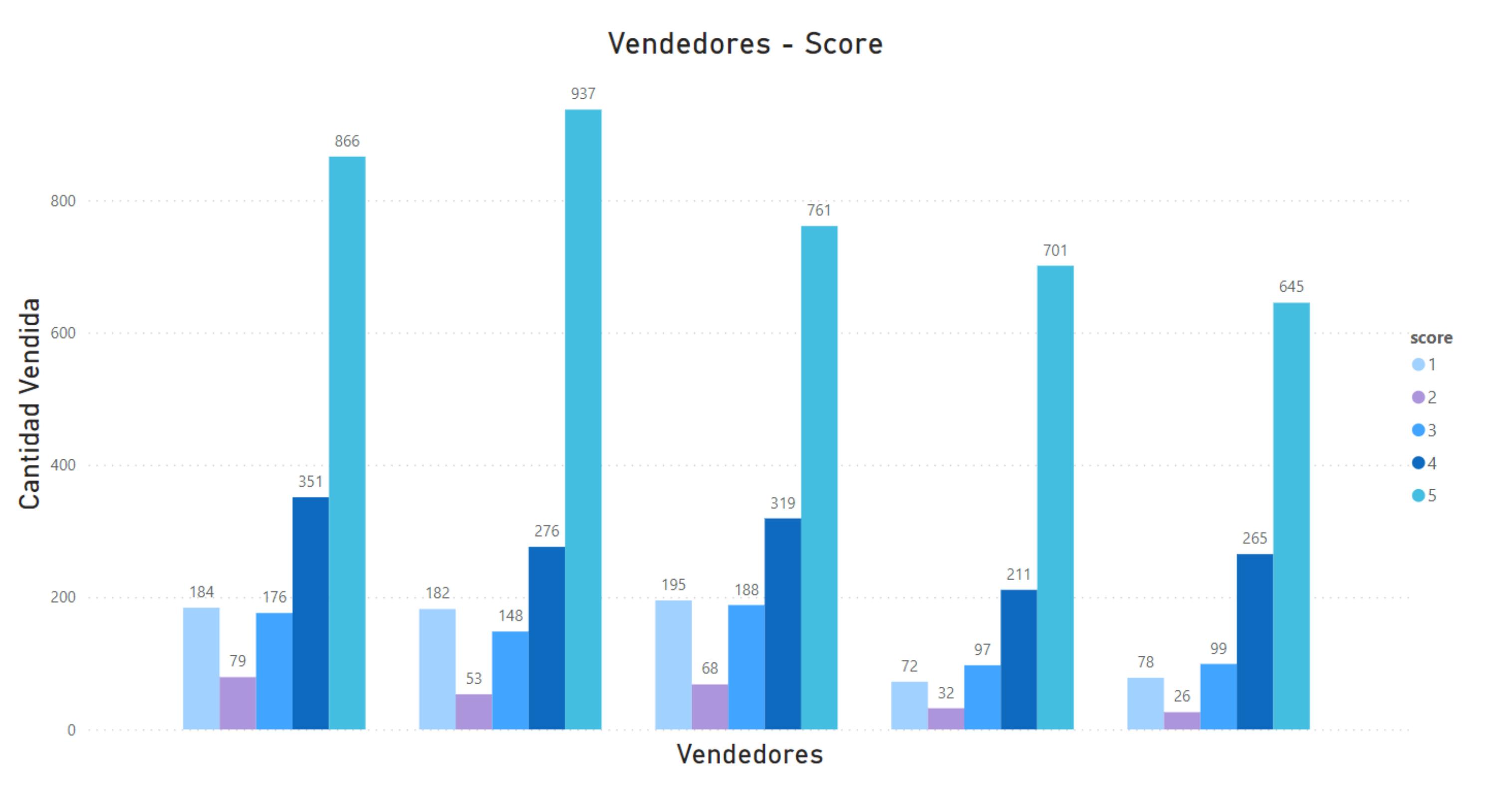


Orders - Review



Producto - Cantidad





Datos

7 datasets, concatenados por claves primarias. Dataset final con 35 columnas y 112.371 filas. 22 Objeto, 9 con decimal y 4 eran enteros



Vendedor

0 **seller_id**
1 ~~seller_zip_code_prefix~~
2 seller_city
3 seller_state



Cliente

0 **customer_id**
1 ~~customer_unique_id~~
2 ~~customer_zip_code_prefix~~
3 customer_city
4 customer_state



Info, entrega

Producto

0 **order_id**
1 ~~order_item_id~~
2 **product_id**
3 **seller_id**
4 ~~shipping_limit_date~~
5 price
6 freight_value



Score

0 ~~review_id~~
1 **order_id**
2 **review_score**
3 ~~review_comment_title~~
4 ~~review_comment_message~~
5 ~~review_creation_date~~
6 ~~review_answer_timestamp~~



Ordenes

0 **order_id**
1 **customer_id**
2 order_status
3 order_purchase_timestamp
4 order_approved_at
5 order_delivered_carrier_date
6 order_delivered_customer_date
7 order_estimated_delivery_date



Informacion

Producto

0 **product_id**
1 product_category_name
2 ~~product_name_length~~
3 ~~product_description_length~~
4 ~~product_photos_qty~~
5 product_weight_g
6 product_length_cm
7 product_height_cm
8 product_width_cm

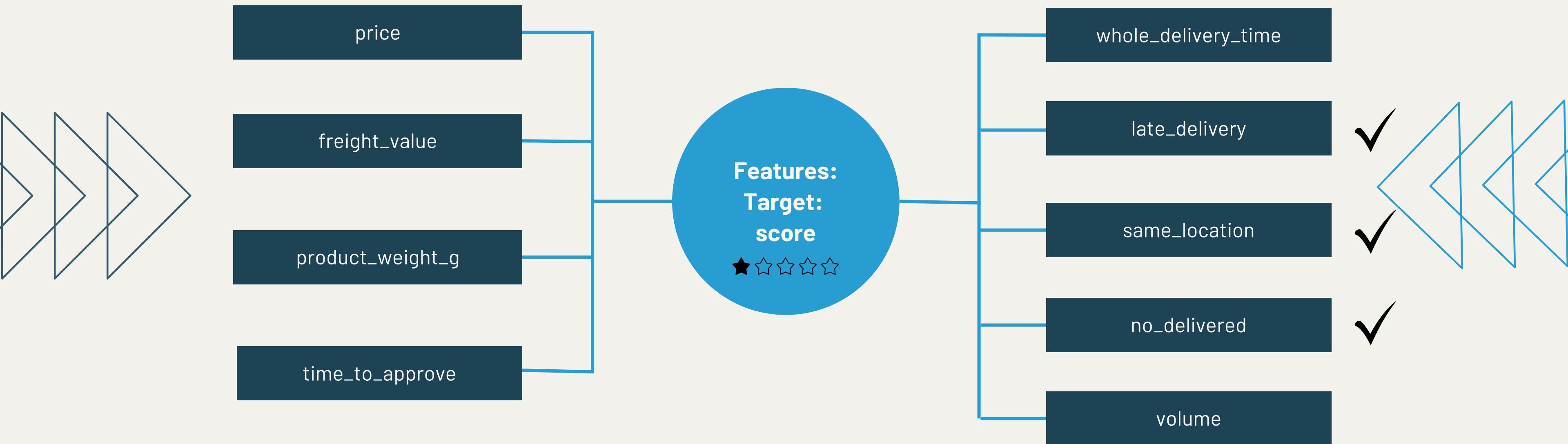
Procesamiento datos



- 1 Formato de las celdas.
8 columnas a datetime
- 2 Creación de Columnas
 - Tiempo de aprobacion - time_to_approve
 - Tiempo completo de entrega - whole_delivery_time
 - Entrega a tiempo - late
- 3 Entrega
 - Order_Status: no_delivered
 - Mismo estado - Location
 - Volumen
- 4 Grupo de muestra a trabajar
Eliminan ordenes que sean mas
de un producto
DATASET A TRABAJAR
- 5 Limpieza, más Nans
time_to_approve, volume, whole_delivery_time,



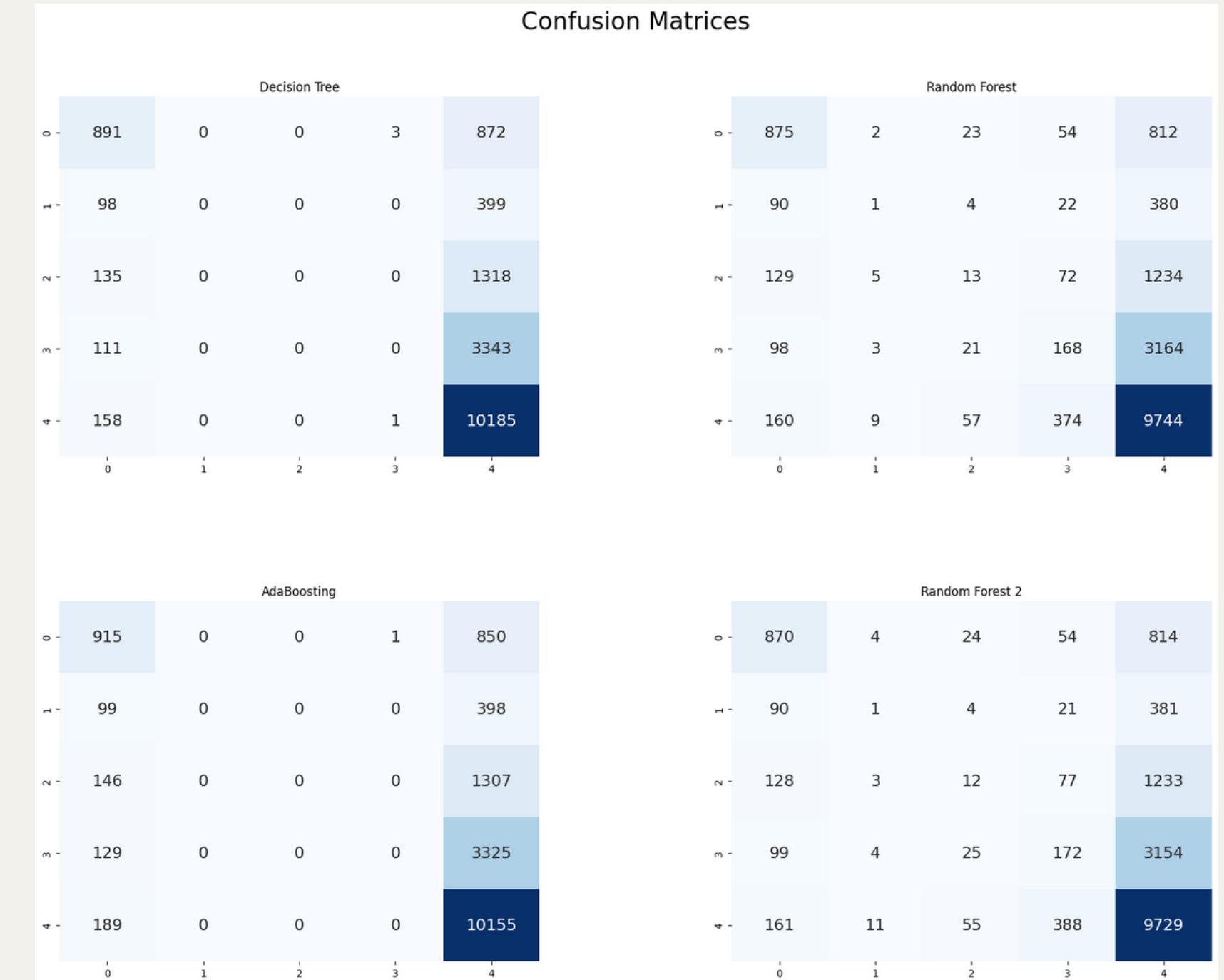
Features y Target



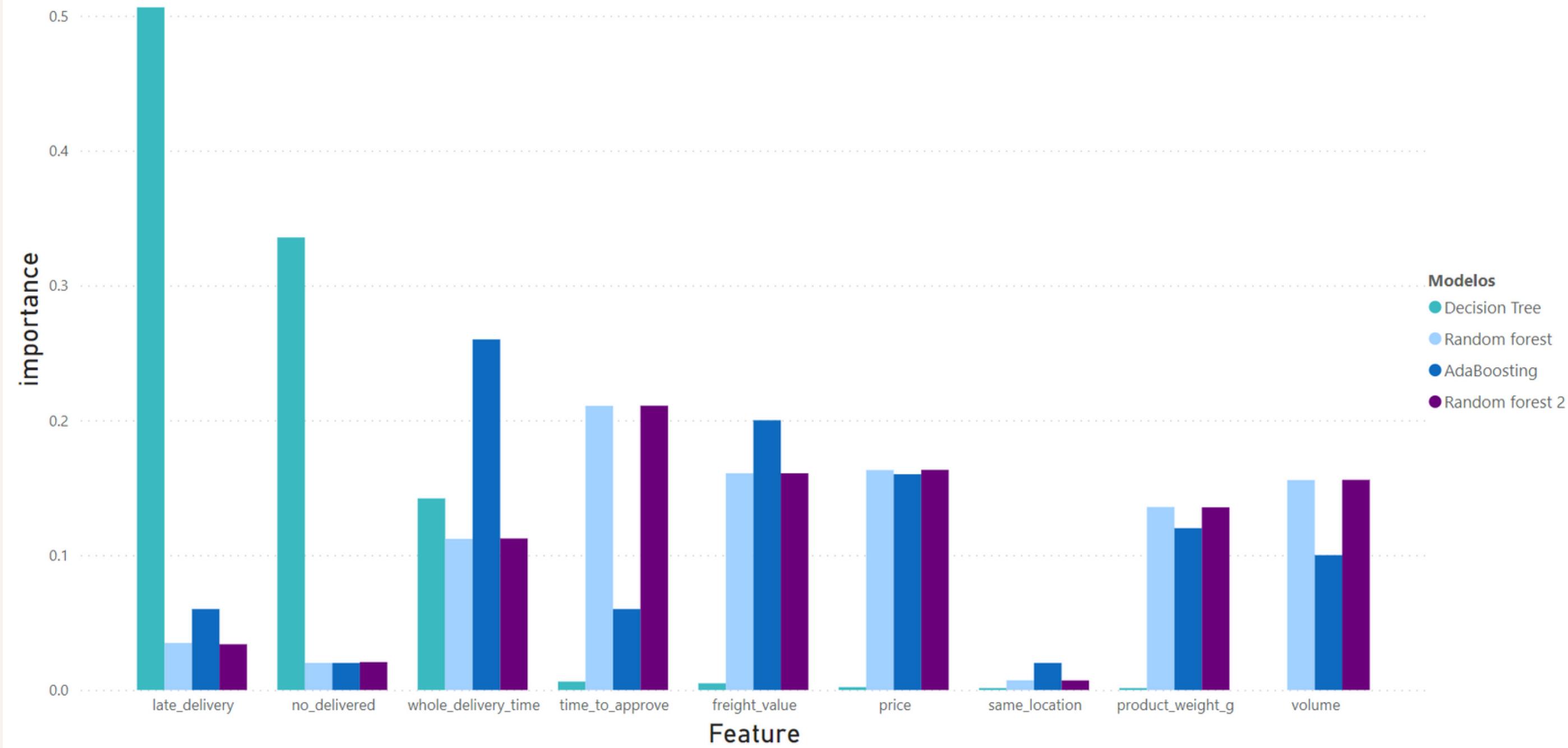
Preparación de los datos

Modelos

Modelo	Decision Tree	Random Forest	Ada Boosting	Random Forest 2	Logistic Regression
score_train	0.634578	0.634578	0.633450	0.999900	0.633550
accuracy_score	0.632408	0.616707	0.632066	0.615736	0.632465



Feature Importance



Feature importance

Feature importance	Decision Tree	Random forest	AdaBoosting	Random forest 2
price	0.002090	0.163072	0.16	0.163230
freight_value	0.004915	0.160740	0.20	0.160718
product_weight_g	0.001420	0.135665	0.12	0.135441
time_to_approve	0.006206	0.210716	0.06	0.210832
whole_delivery_time	0.142084	0.112079	0.26	0.112373
late_delivery	0.506267	0.034837	0.06	0.033898
same_location	0.001425	0.007171	0.02	0.007024
no_delivered	0.335593	0.020031	0.02	0.020667
volume	0.000000	0.155689	0.10	0.155818

Coeficientes de correlacion

Regression 1	Regression 2	Regression 3	Regression 4	Regression 5
1.784650	-0.444046	-2.020498	-0.091580	0.771473
-0.405112	-0.129421	-0.824936	0.216030	1.143439
0.264565	0.255749	0.276681	-0.164999	-0.631997
-0.685802	-0.023496	0.233677	0.119127	0.356493
3.940106	2.917765	2.740916	-1.053713	-8.545074
1.535394	0.723003	-0.282217	-0.981409	-0.994770
0.050075	0.050805	0.025443	-0.024418	-0.101904
4.881268	2.948037	1.382735	-1.135953	-8.076086
-0.137185	0.037741	0.008421	0.001755	0.089268

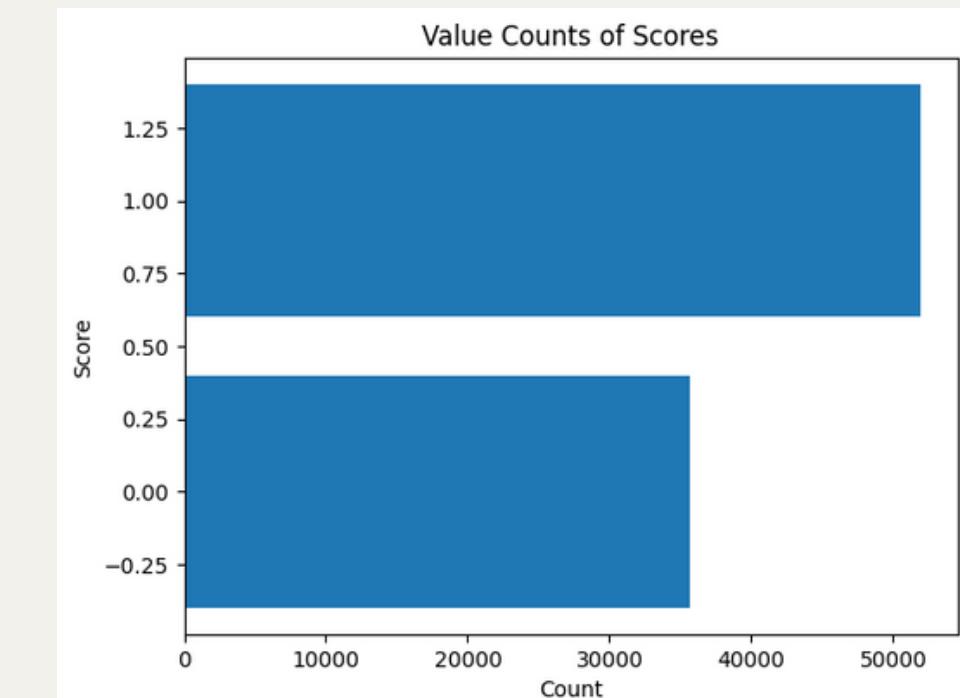
Nuevas columnas

Features

price
freight_value
product_weight_g
volume
time_to_approve
whole_delivery_time
no_delivered
late_delivery
same_location
closer
freight_higher_mean
priceless
heavy
month
day
product_category_name
region

Traget

Satisfaction



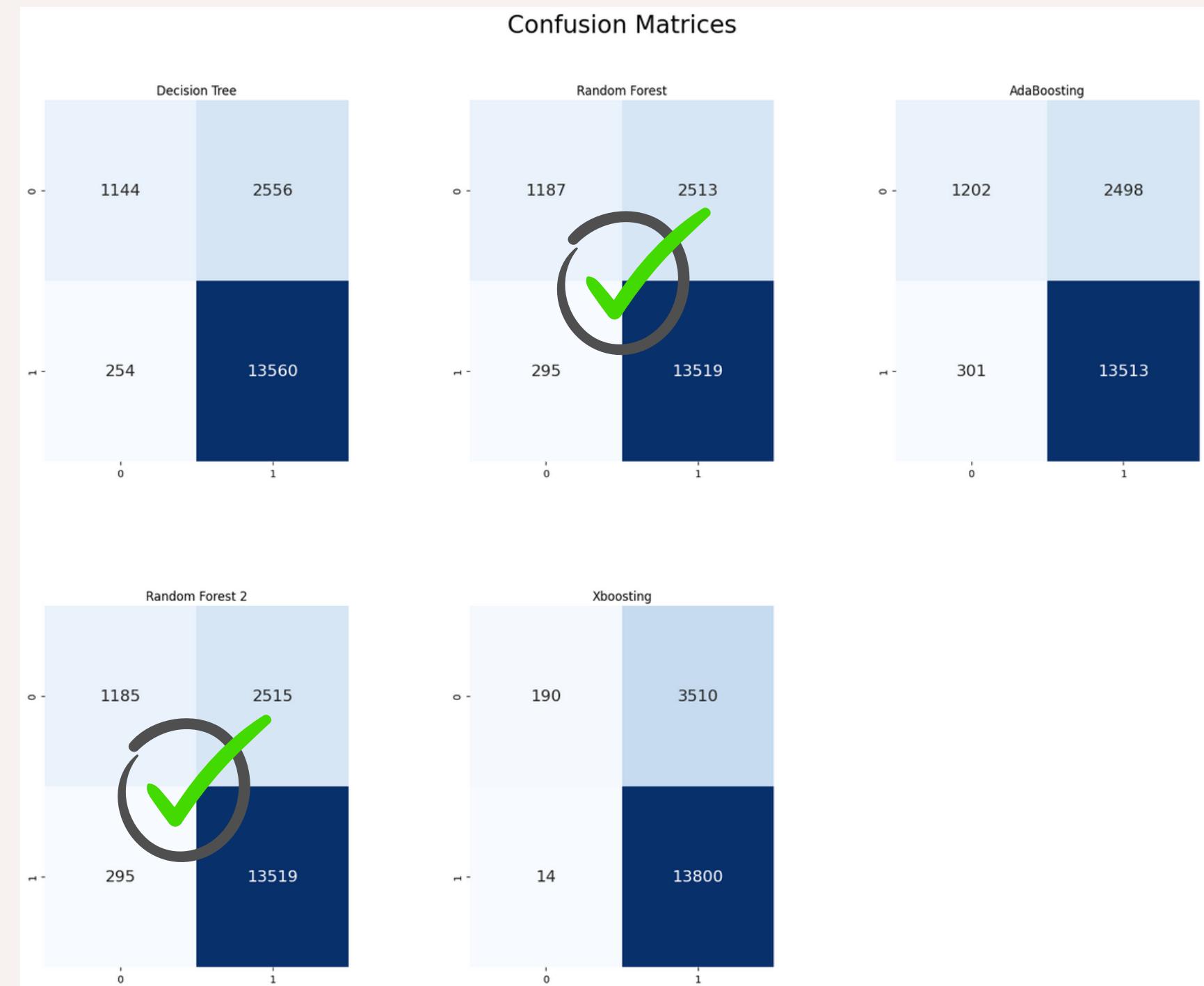
Modelo	Decision Tree	Random forest	AdaBoosting	Random forest 2	Xboosting	Regression 1
time_to_approve	0.003963	0.133163	0.04	0.133242	0.002535	0.555111
price	0.000770	0.104898	0.02	0.105041	0.003125	0.064616
freight_value	0.003044	0.106610	0.14	0.106354	0.003614	1.324615
product_weight_g	0.000739	0.089741	0.06	0.089765	0.003415	-0.695901
product_category_name	0.000000	0.060754	0.08	0.060897	0.004607	• -0.017774
whole_delivery_time	0.087440	0.125914	0.36	0.124046	0.016623	-9.186200
late_delivery	0.554069	0.082402	0.08	0.083812	0.756363	-1.805577
same_location	0.002439	0.007255	0.02	0.007231	0.009655	-0.074310
no_delivered	0.339198	0.029699	0.04	0.029689	0.185606	-7.276604
volume	0.004790	0.100218	0.14	0.100339	0.004283	-0.258297
freight_higher_mean	0.000000	0.005475	0.00	0.005395	0.000000	0.062851
priceless	0.000000	0.006074	0.00	0.006006	0.000000	-0.078247
heavy	0.000000	0.006135	0.00	0.006131	0.000000	0.017961
month	0.000935	0.073676	0.02	0.073502	0.003082	0.217445
day	0.002611	0.063639	0.00	0.064182	0.003637	-0.013886
closer	0.000000	0.004348	0.00	0.004370	0.003454	-0.075424

Importancias bajas casi todos los modelos. Logistic Regression

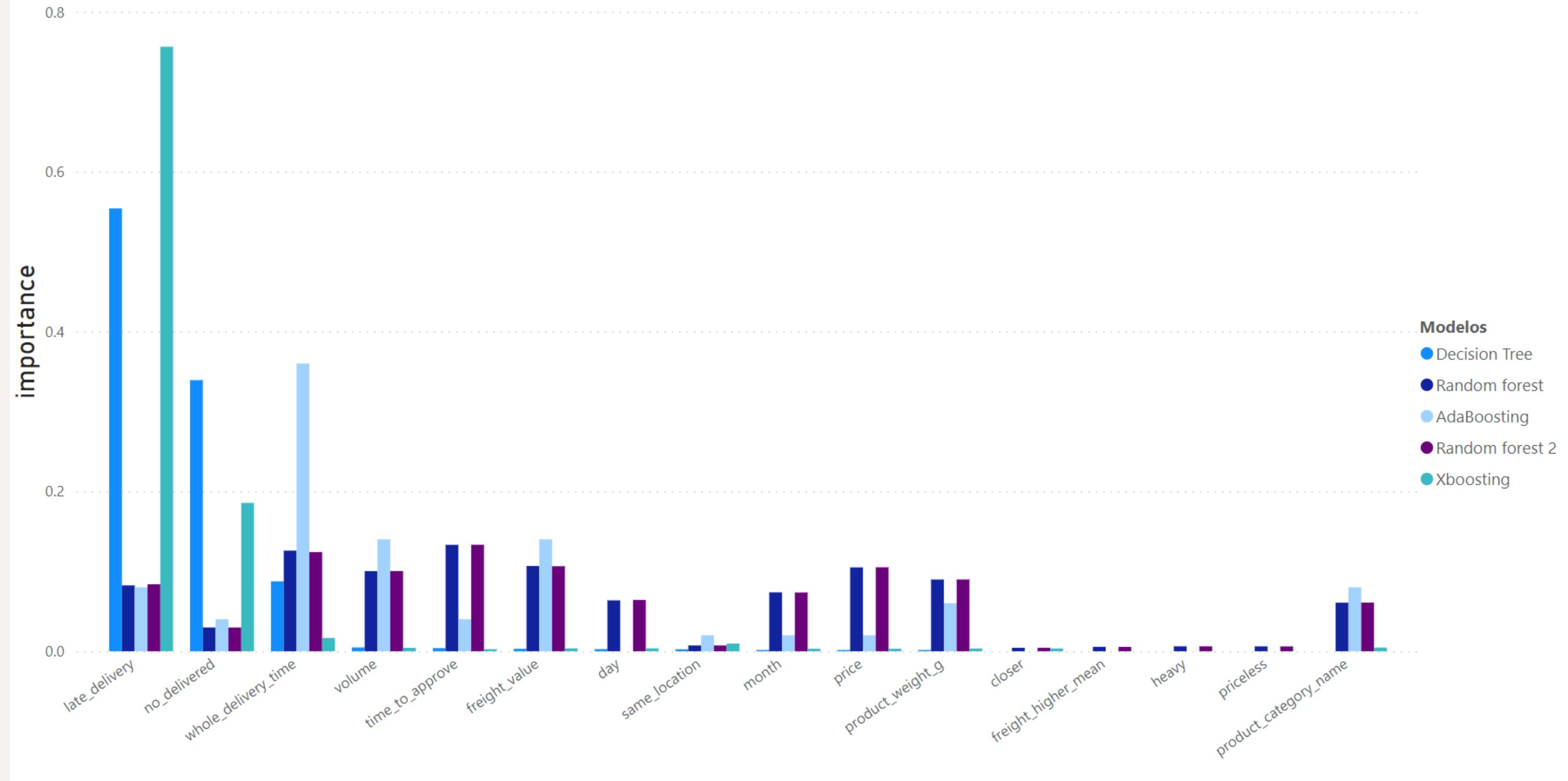
time_to_approve, price, freight_value, weight, whole delivery time, late_delivery, no_delivered, volume

Modelos

Model	Decision Tree	Random Forest	AdaBoosting	Random Forest 2	Logistic Regression	XGB
score_train	0.841627	0.841627	0.840285	1.000000	0.840156	0.840156
accuracy_score	0.839557	0.839671	0.840185	0.839557	0.840413	0.798790



Feature Importance



Conclusiones

Modelo	DecisionTree	RandomForest	AdaBoosting	Random Forest 2	Logistic Regression	XGB
score_train	0.841627	0.841627	0.840285	1.000000	0.840156	0.840156
accuracy_score	0.839557	0.839671	0.840185	0.839557	0.840413	0.798790

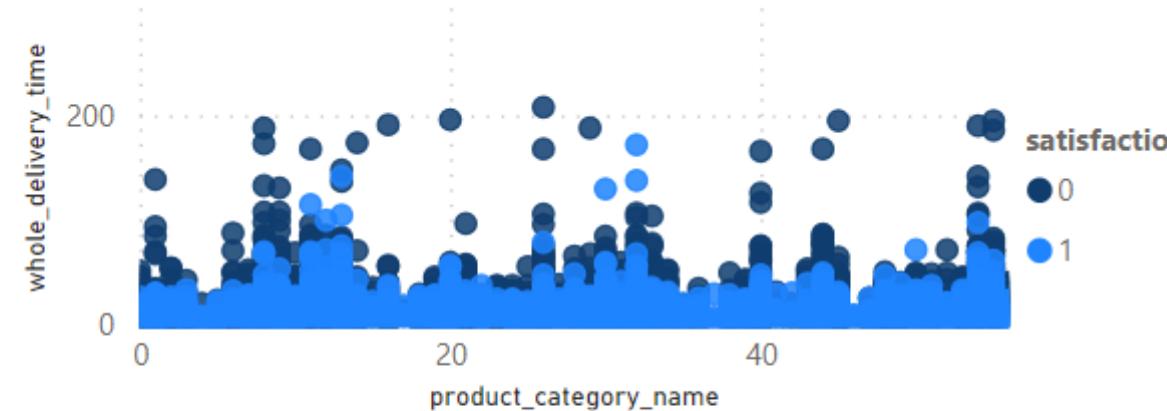
seller_id

All

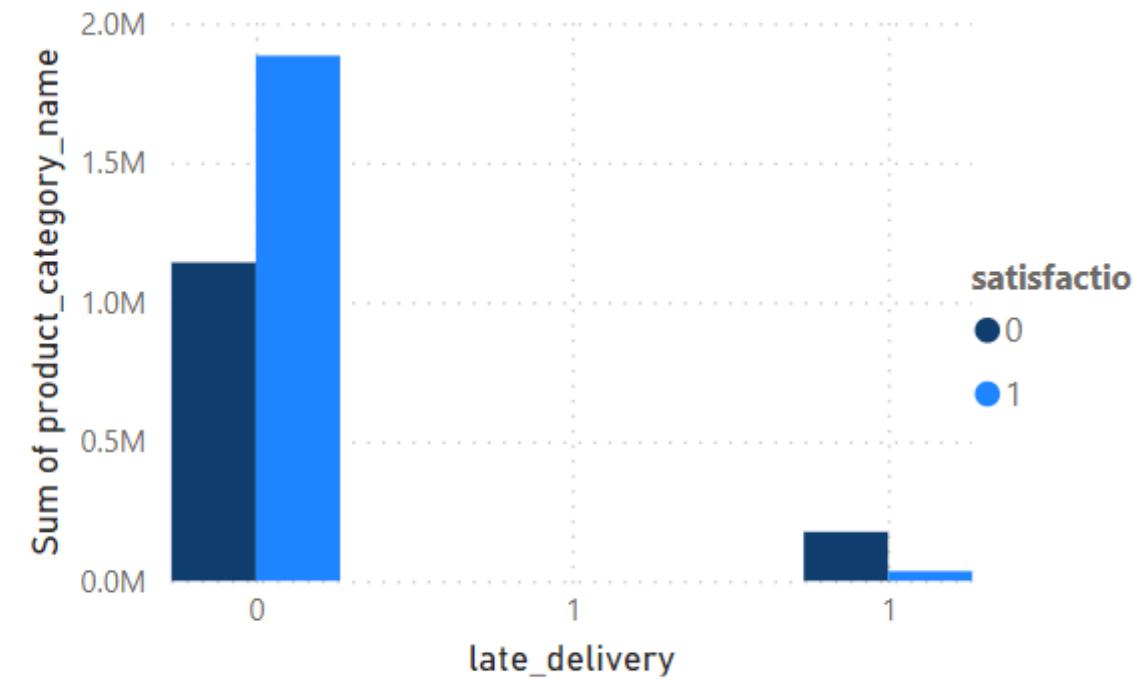
2989

Count of seller_id

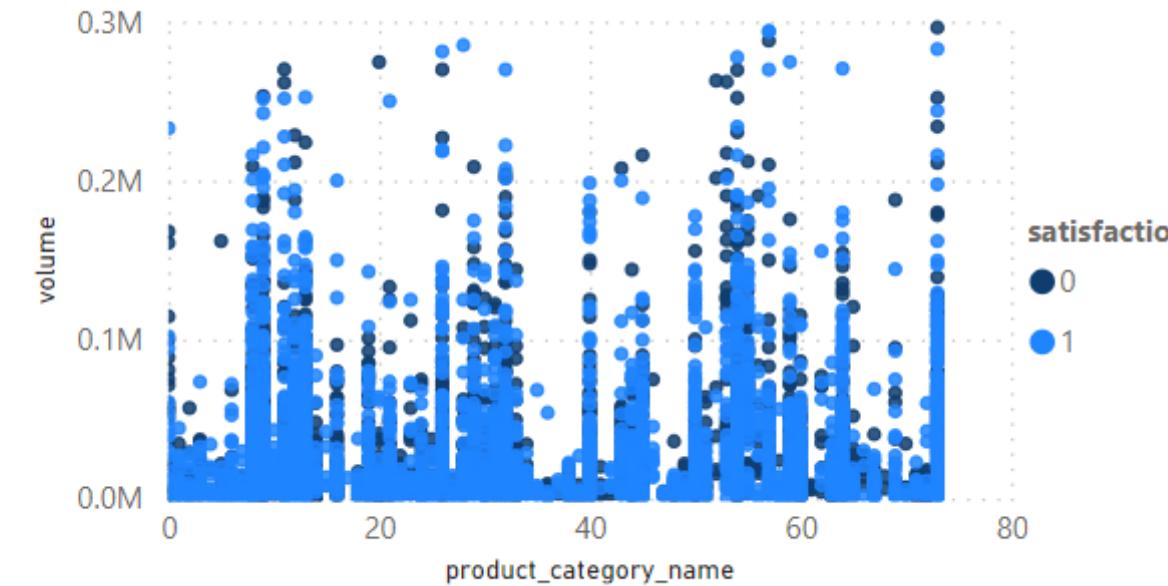
product_category_name and whole_delivery_time



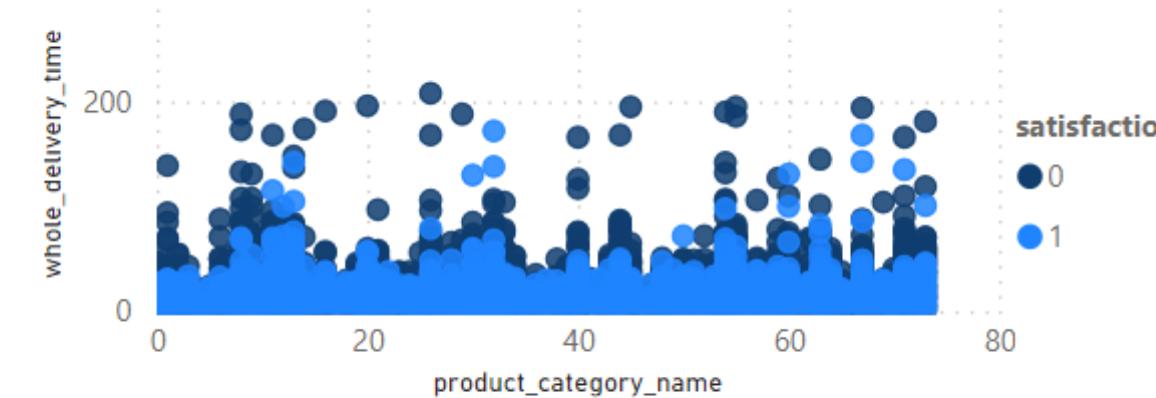
product_category_name by late_delivery and satisfaction



volume and product_category_name



whole_delivery_time and product_category_name



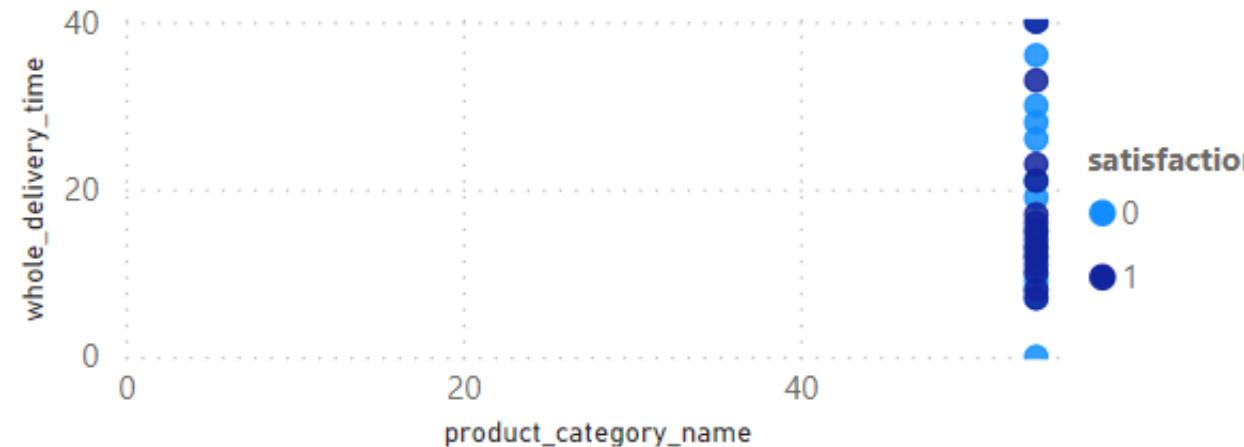
seller_id

002100f778ceb8431b7a1020ff7ab48f

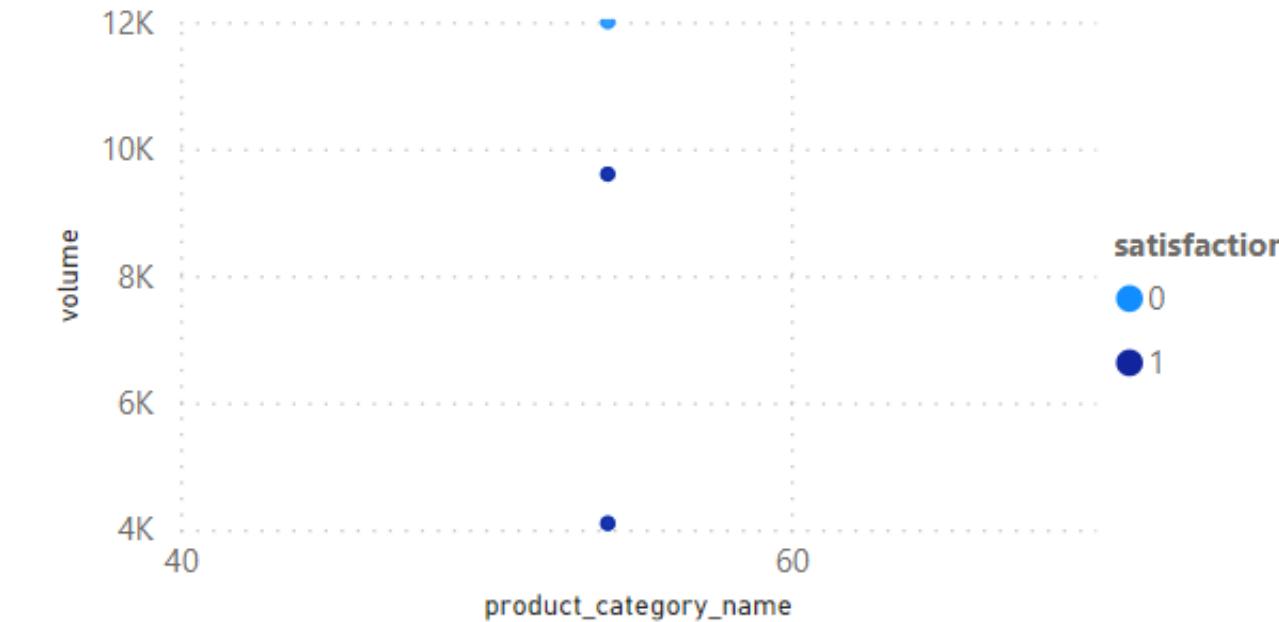
44

Count of order_id

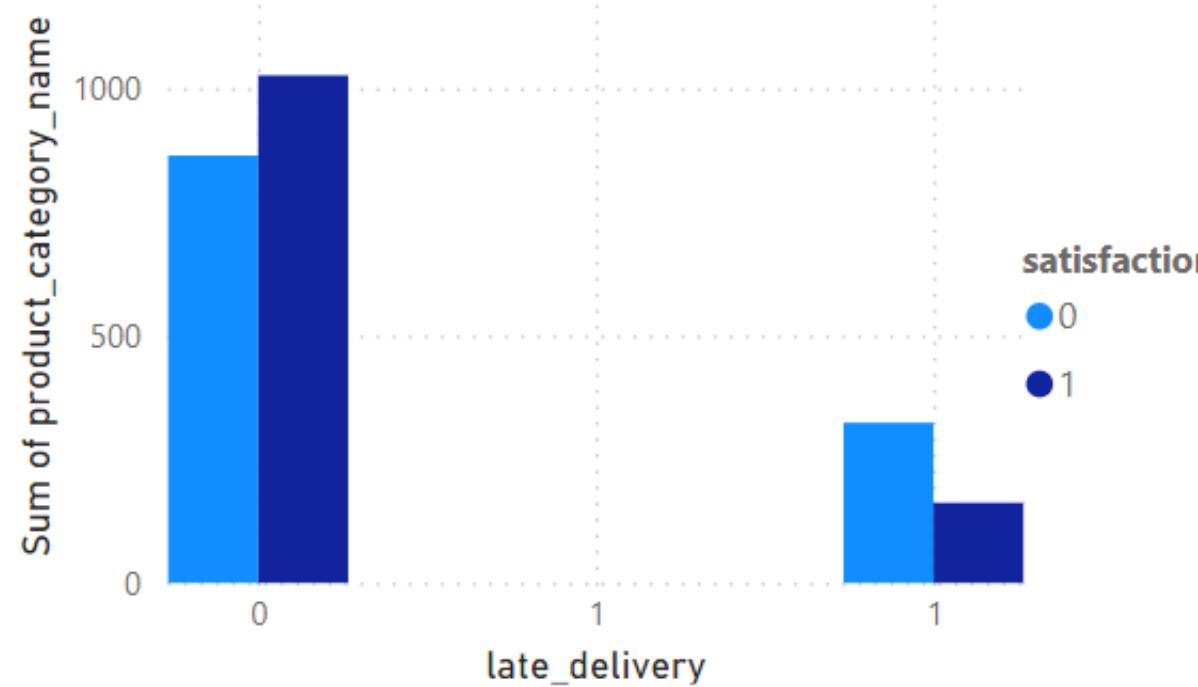
product_category_name and whole_delivery_time



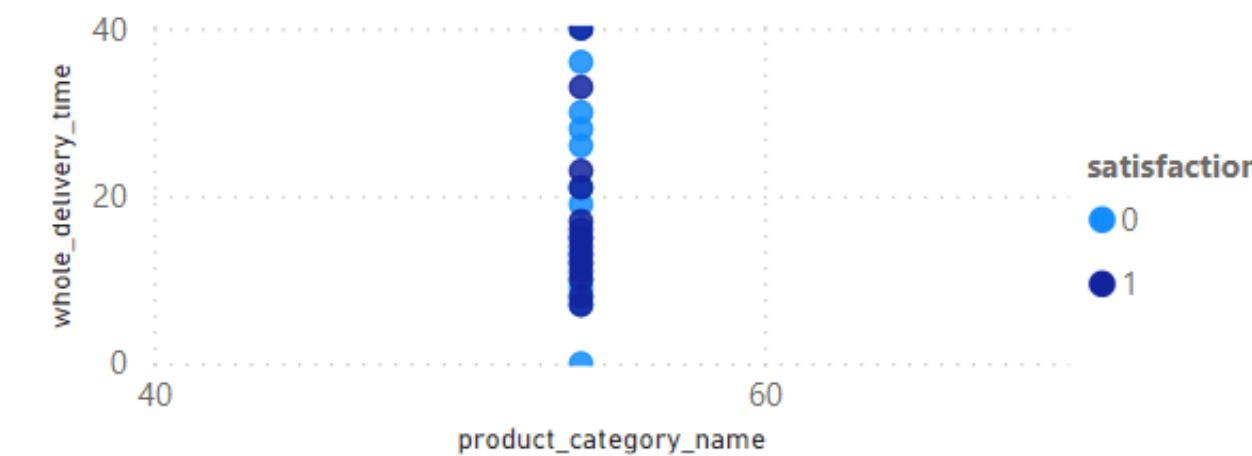
volume and product_category_name



product_category_name by late_delivery and satisfaction



whole_delivery_time and product_category_name



Thank You!