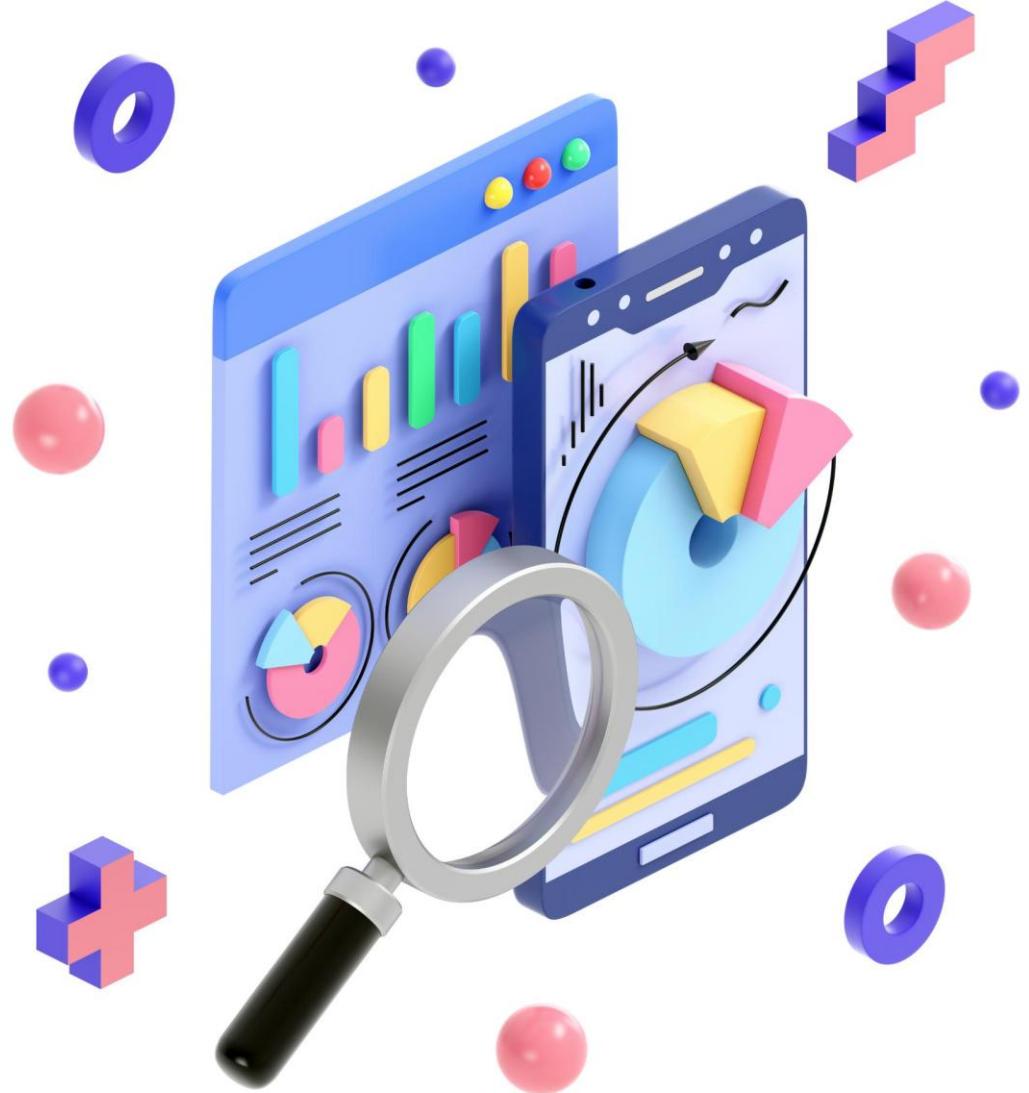




Presentación de la Práctica Final del Curso de Ciencia de Datos

*Análisis de datos para la empresa
Reckitt 2021-2022*

Paulo Andrés Ossa Benítez.



Programa de la Presentación

Introducción del Proyecto

Exploración y Análisis Inicial con Excel

Análisis de Datos con SQL

Limpieza y Transformación de Datos con Python

Análisis Exploratorio de Datos (EDA) y Visualizaciones

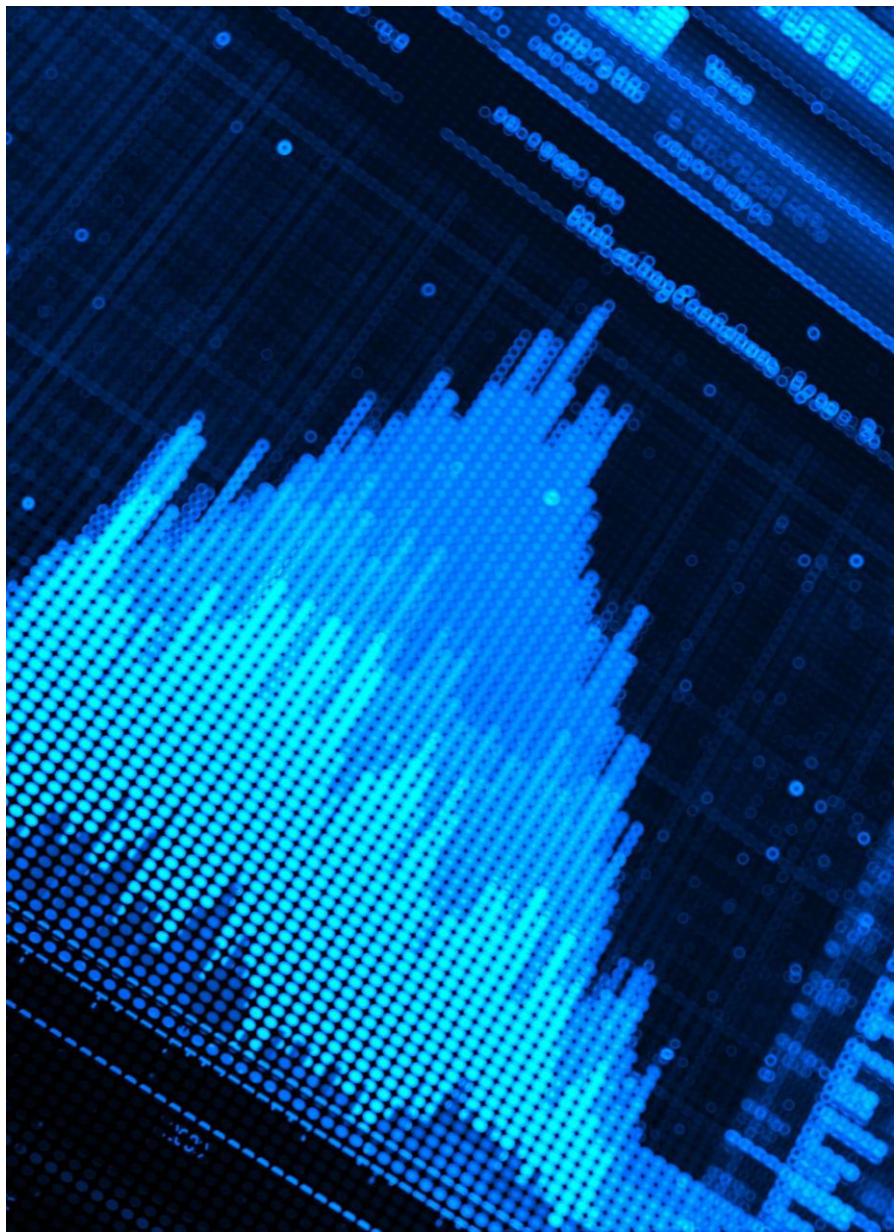
Creación de un Dashboard en Looker Studio

Predicción de Ventas con Machine Learning

Ánalysis en AWS con Cloud Computing

Conclusiones y Recomendaciones

Introducción del Proyecto



Contexto del proyecto

Análisis de Datos de Ventas

El proyecto se enfoca en examinar un conjunto de datos de ventas para descubrir patrones y tendencias que pueden influir en la toma de decisiones.

Identificación de Patrones

Identificar patrones en los datos es esencial para entender mejor el comportamiento de los consumidores y ajustar estrategias comerciales.

Mejora de Estrategias Comerciales

El análisis de datos permitirá mejorar las estrategias comerciales para maximizar los resultados y optimizar el rendimiento del negocio.



Relevancia del análisis

Competitividad Empresarial

Comprender los factores que afectan las ventas permite a la empresa destacarse en un mercado competitivo.

Toma de Decisiones Basadas en Datos

El análisis proporciona información valiosa que ayuda a la toma de decisiones informadas para estrategias futuras.

Información Valiosa

El análisis de ventas revela tendencias y patrones que son esenciales para el crecimiento empresarial.

Herramientas utilizadas



Exploración de datos con Excel

Excel se utilizó para la exploración inicial de datos, ofreciendo herramientas eficaces para el análisis básico y la organización de datos.

Consultas y extracción con SQL

SQL Server fue utilizado para realizar consultas y extracción de datos, permitiendo acceder a información específica de manera eficiente.

Limpieza y análisis con Python

Python se utilizó para la limpieza y análisis de datos, aprovechando bibliotecas poderosas para manipular y analizar datos complejos. Igualmente fue utilizado en la predicción de ventas con Machine Learning.

Visualizaciones con Looker Studio

Looker Studio se utilizó para crear visualizaciones finales, facilitando la presentación de datos de manera clara y efectiva.

Análisis en AWS con Cloud Computing

Se utilizaron varias herramientas dentro de la nube de AWS como S3, Glue, Athena y RDS

Exploración y Análisis Inicial con Excel



Visualizaciones iniciales

Identificación de Patrones

Las visualizaciones iniciales nos permitieron identificar patrones significativos en los datos de ventas, lo que es crucial para el análisis.

Análisis Profundo

Identificamos áreas que requerían un análisis más profundo, facilitando decisiones informadas basadas en datos.

Base para Siguientes Pasos

Las visualizaciones iniciales establecieron una base sólida para el siguiente paso del análisis, guiando nuestras acciones estratégicas.

Contexto exploración en Excel

ANALISIS DE VENTAS 2022-2023

OBJETIVO

- Comportamiento de ventas por categoría
- Comportamiento de ventas por Región
- Comportamiento de ventas por periodo de tiempo

POR PAULO ANDRES OSSA BENITEZ

ACTIVIDAD DE CLASE

- Las tablas originales no se tocaron, en cambio se trabajaron todos los datos en conjunto en una hoja nueva creada llamada MASTER
- Las tablas dinámicas y análisis de datos se hicieron especialmente en la hoja llamada ANALISIS1
- El Dashboard con las gráficas y la información mas entendible esta en la hoja llamada DASHBOARD
- Finalmente en la hoja de Insights hay 3 secciones que son unas consideraciones a tener en cuenta sobre la data, los hallazgos mas importantes basados en el análisis y las recomendaciones que se dan a partir de los hallazgos

< > INDEX ACTIVIDADES FACT_SALES DIM_PRODUCT DIM_CATEGORY DIM_CALENDAR MASTER ANALISIS1 DASHBOARD INSIGHT

Accesibilidad: es necesario investigar

Pestaña Index

Contiene el objetivo del análisis en Excel, contemplando que es el inicio de curso y no se tiene mucho conocimiento sobre el EDA.

En esta práctica se visualizan todas las fuentes de datos proporcionadas, para entender bien la data, y posteriormente se decide hacer una tabla “MASTER”, uniendo catálogos con la data de ventas, para tener un solo archivo maestro para el análisis.

Luego se procede a hacer un análisis de los datos basados en la tabla creada anteriormente , para posteriormente graficar datos relevantes y finalmente poder generar algunos insights con la información obtenida en este primer análisis.

Tabla “Master” en Excel

A	B	C	D	E	F	G	H	I	J	K	L
WEEK	YEAR	MONTH	ID_WEEK	ITEM_CODE	ITEM_DESCRIPTION	BRAND	ID_CATEGORY	CATEGORY_NAME	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES
2	34-22	2022	8	34 7501058792808BP	VANISH OXI ACTION ROSA DADO PACK 120GR+MMCRYSTALWHITE BOTE 900GR PVP 7501058792808BP2	VANISH	1	FABRIC TREATMENT and SANIT	0,006 \$	0,14	1
3	34-22	2022	8	34 7501058715883	VANISH OXI ACTION GOLD QUITAMANCHAS BOLSA 1.8KG 7501058715883	VANISH	1	FABRIC TREATMENT and SANIT	0,487 \$	116,52	2,916
4	34-22	2022	8	34 7702626213774	VANISH OXI ACTION ROSA QUITAMANCHAS DOYPACK 240 GR 7702626213774	VANISH	1	FABRIC TREATMENT and SANIT	1,391 \$	68,45	5,171
5	34-22	2022	8	34 7501058716422	VANISH OXI ACTION GOLD QUITAMANCHAS AHORRO DEL 30% BOT 400GR 7501058716422	VANISH	1	FABRIC TREATMENT and SANIT	0,022 \$	1,48	1,833
6	34-22	2022	8	34 7501058784353	VANISH INTELLIGENCE POLVO BOTE 450 GR NAL 7501058784353	VANISH	1	FABRIC TREATMENT and SANIT	2,037 \$	182,84	5,375
7	34-22	2022	8	34 7501058716064	VANISH OXI ACTION ROSA QUITA/MANCHAS P/ROPA BOT 810GR 7501058716064	VANISH	1	FABRIC TREATMENT and SANIT	0,005 \$	0,68	1,25
8	34-22	2022	8	34 7501058794638BP	VANISH OXI ACTION QUITAMANCHAS GOLDSABOT 900GR+MMCRYSTALWHITE BOT 150GR PVP 7501058794638BP1	VANISH	1	FABRIC TREATMENT and SANIT	0,004 \$	0,55	1
9	34-22	2022	8	34 7501058792778BP	VANISH OXI ACTION ROSA QUITAMANCHAS GOLDSABOT 900GR+MMCRYSTALWHITE BOT 120GR PVP 7501058792778BP1	VANISH	1	FABRIC TREATMENT and SANIT	0,001 \$	0,13	1
10	34-22	2022	8	34 750105877630	VANISH PODER 02 PRE-LAVADOR BOT.ATOM. 650 ML NAL 750105877630	VANISH	1	FABRIC TREATMENT and SANIT	0,716 \$	71,02	4,287
11	34-22	2022	8	34 7501058753441	VANISH PODER 02 DOYPACK 120 GR 7501058753441	VANISH	1	FABRIC TREATMENT and SANIT	2,704 \$	61,23	11,22
12	34-22	2022	8	34 7702626204208BP	VANISH PODER 02 INTELLIGENCE SIN CLORO SOBRE 30 G IMP 7702626204208BP1	VANISH	1	FABRIC TREATMENT and SANIT	2,188 \$	23,79	17,934
13	34-22	2022	8	34 7501058784346	VANISH PODER 02 INTELLIGENCE POLVO BOT 900 GR NAL 7501058784346	VANISH	1	FABRIC TREATMENT and SANIT	0,811 \$	146,76	3,755
14	34-22	2022	8	34 7501058796868	VANISH OXI ACTION GOLD SUPER BARRA WHITE JUMBO 110 GR 7501058796868	VANISH	1	FABRIC TREATMENT and SANIT	3,507 \$	55,86	28,282
15	34-22	2022	8	34 7891035040276	VANISH SUPER BARRA WHITE QUITA MANCHAS 68GR 7891035040276	VANISH	1	FABRIC TREATMENT and SANIT	6,539 \$	80,17	23,025
16	34-22	2022	8	34 7501058796851	VANISH OXI ACTION GOLD SUPER BARRA JUMBO 110 GR 7501058796851	VANISH	1	FABRIC TREATMENT and SANIT	2,547 \$	40,34	20,54
17	34-22	2022	8	34 7501058714596	VANISH SUPER BARRA QUITA MANCHAS 68GR 7501058714596	VANISH	1	FABRIC TREATMENT and SANIT	6,774 \$	86,10	21,505
18	34-22	2022	8	34 7891035040191	VANISH QUITAMANCHAS PRELAVADOR BARRA 75GR 7891035040191	VANISH	1	FABRIC TREATMENT and SANIT	0,008 \$	0,10	1,6
19	34-22	2022	8	34 7501058751188	VANISH WHITE REMOVEDOR DE MANCHAS BOT 1700 ML 7501058751188	VANISH	1	FABRIC TREATMENT and SANIT	0,997 \$	93,12	4,279
20	34-22	2022	8	34 7501058716095	VANISH WHITE LIQUIDO QUITAMANCHAS+BLANQ.S/CLORO BOT 3.6LT 7501058716095	VANISH	1	FABRIC TREATMENT and SANIT	0,078 \$	12,95	2,167
21	34-22	2022	8	34 7501058751461	VANISH WHITE REMOVEDOR D/MANCHAS Y BLANQ.S/CLORO LIQ.BOT 4000M 7501058751461	VANISH	1	FABRIC TREATMENT and SANIT	0,351 \$	66,05	3,311
22	34-22	2022	8	34 7501058757463	VANISH WHITE LIQ. REMOVEDOR DE MANCHAS BOT 435ML NAL 7501058757463	VANISH	1	FABRIC TREATMENT and SANIT	0,004 \$	0,16	2
23	34-22	2022	8	34 7501058757555	VANISH WHITE LIQUIDO DOY PACK 600ML NAL 7501058757555	VANISH	1	FABRIC TREATMENT and SANIT	1,428 \$	51,06	7,933
24	34-22	2022	8	34 7501058751195	VANISH WHITE BOTELLA 900 ML 7501058751195	VANISH	1	FABRIC TREATMENT and SANIT	2,532 \$	129,98	7,791
25	34-22	2022	8	34 0046500022692	SHOUT ADVANCED LAUNDRY STAIN REMOVER ULTRA CONC GEL BOT 246GR 0046500022692	SHOUT	1	FABRIC TREATMENT and SANIT	0,024 \$	2,04	3,429
26	34-22	2022	8	34 0046500729355	SHOUT LAUNDRY STAIN REMOVER BOTELLA C/ATOMIZADOR 650 ML 0046500729355	SHOUT	1	FABRIC TREATMENT and SANIT	0,001 \$	0,05	1
27	34-22	2022	8	34 7501268200001	LA VALENCIANA BOT.PLAST. 1000ML NAL 7501268200001	LA VALENCIANA	1	FABRIC TREATMENT and SANIT	0,184 \$	1,93	8,364
28	34-22	2022	8	34 7501268200506	LA VALENCIANA BLANQUEADOR REGULAR BOT PLAS 2000 ML 7501268200506	LA VALENCIANA	1	FABRIC TREATMENT and SANIT	0,003 \$	0,06	1,5
29	34-22	2022	8	34 7501268200339	LA VALENCIANA BLANQUEADOR REGULAR BOT PLAS 4000 ML 7501268200339	LA VALENCIANA	1	FABRIC TREATMENT and SANIT	0,041 \$	1,31	4,556
30	34-22	2022	8	34 7501268200025	LOS PATITOS SUPER.CONCENT.BOTELLA 1000ML NAL 7501268200025	LOS PATITOS	1	FABRIC TREATMENT and SANIT	0,013 \$	0,14	13
31	34-22	2022	8	34 7501268200049	LOS PATITOS BLANQUEADOR REGULAR BOT PLAST 2 LT NAL 7501268200049	LOS PATITOS	1	FABRIC TREATMENT and SANIT	0,682 \$	12,81	8,525
32	34-22	2022	8	34 7501268200070	LOS PATITOS BOT.PLAST. 4000 ML NAL 7501268200070	LOS PATITOS	1	FABRIC TREATMENT and SANIT	0,243 \$	8,65	4,585
33	34-22	2022	8	34 7501268200261	LOS PATITOS BLANQUEADOR REGULAR BOT PLAST 6LT NAL 7501268200261	LOS PATITOS	1	FABRIC TREATMENT and SANIT	0,005 \$	0,23	5
34	34-22	2022	8	34 7501268200513	LOS PATITOS BLANQUEADOR P.ROPA BOT 520ML 7501268200513	LOS PATITOS	1	FABRIC TREATMENT and SANIT	0,406 \$	4,43	5,884
35	34-22	2022	8	34 0757037517123	OXI CLEAN LIQ.BOT. 655 ML 0757037517123	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,001 \$	0,04	1
36	34-22	2022	8	34 7501080900585	OXI CLEAN SANITIZER QUITAMANCHAS ROPA AROMA FRESCO BOT.13KG 7501080900585	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,072 \$	18,88	1,309
37	34-22	2022	8	34 7501080954786	OXI CLEAN ULTRA QUITAMANCHAS GEL BOT 200ML 7501080954786	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,854 \$	19,88	3,211
38	34-22	2022	8	34 7501080954793	OXI CLEAN ULTRA QUITAMANCHAS GEL BOT 600ML 7501080954793	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,246 \$	16,53	2,828
39	34-22	2022	8	34 7501080955103	OXI CLEAN ULTRA QUITAMANCHAS GEL BOT PLAS 2LT 7501080955103	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,175 \$	32,30	3,365
40	34-22	2022	8	34 7501080955486	OXI CLEAN REMOVEDOR DE MANCHAS BOLSA 900ML 7501080955486	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,708 \$	53,76	2,926
41	34-22	2022	8	34 7501080900165	OXI CLEAN QUITAMANCHAS VERSATIL SOBRE 110 GR 7501080900165	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,921 \$	21,66	5,829
42	05-22	2022	2	5 07411200090141	OXI CLEAN QUITAMANCHAS VERSATIL SOBRE 420 GR 7501080900141	OXI CLEAN	1	FABRIC TREATMENT and SANIT	4,476 \$	247,58	3,034
43	05-22	2022	2	5 7501080952973	OXI CLEAN POLVO BOTE 420 GR 7501080952973	OXI CLEAN	1	FABRIC TREATMENT and SANIT	0,004 \$	0,22	1,333
44	05-22	2022	2	5 7501080900158	OXI CLEAN QUITAMANCHAS VERSATIL SOBRE 860 GR 7501080900158	OXI CLEAN	1	FABRIC TREATMENT and SANIT	2,897 \$	315,68	3,18
45	05-22	2022	2	5 0741120000746	CLARASOL GEL PODER CONCENTRADO BOT PLAS 1L 0741120000746	CLARASOL	1	FABRIC TREATMENT and SANIT	0,011 \$	0,20	3,667
46	05-22	2022	2	5 0741120000505	CLARASOL BLANQUEADOR REGULAR BOT PLAST 1LT NAL 0741120000505	CLARASOL	1	FABRIC TREATMENT and SANIT	71,192 \$	599,42	109,191
47	05-22	2022	2	5 074112000036	CLARASOL BLANQUEADOR REGULAR BOT PLAS 1850 ML 074112000036	CLARASOL	1	FABRIC TREATMENT and SANIT	9,867 \$	180,70	19,655
48	05-22	2022	2	5 074112000029	CLARASOL REGULAR SIN AROMA GARRAFON 3.785 LT 074112000029	CLARASOL	1	FABRIC TREATMENT and SANIT	13,463 \$	371,02	32,598
49	05-22	2022	2	5 074112000012	CLARASOL BLANQUEADOR REGULAR BOT PLAST 500ML NAL 074112000012	CLARASOL	1	FABRIC TREATMENT and SANIT	0,167 \$	0,97	11,929

En esta tabla ya integramos los datos relevantes para hacer el análisis de datos, con una segmentación de la fecha y la marca la información de los productos.

Importante: No se integra la categoría ya que se valida que toda la data pertenece a una sola categoría y justo en el archivo de categorías solo viene un solo item.

Tabla Dinámica por Región

Column Labels		TOTAL AUTOS AREA 1		TOTAL AUTOS AREA 2		TOTAL AUTOS AREA 3		TOTAL AUTOS AREA 4		TOTAL AUTOS AREA 5		TOTAL AUTOS AREA 6		TOTAL AUTOS SCANNING MEXICO		Total TOTAL DE VENTAS		Total % DE VENTAS		
MARCA	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS	TOTAL DE VENTAS	% DE VENTAS
CLORALEX	\$ 378.642,36	3,43%	\$ 651.756,56	5,90%	\$ 375.082,23	3,40%	\$ 327.217,67	2,96%	\$ 479.772,42	4,34%	\$ 484.338,10	4,39%	\$ 2.696.809,17	24,42%	\$ 5.393.618,51	48,84%				
VANISH	\$ 112.386,93	1,02%	\$ 207.487,35	1,88%	\$ 174.925,75	1,58%	\$ 134.621,68	1,22%	\$ 309.934,98	2,81%	\$ 144.788,17	1,31%	\$ 1.084.144,67	9,82%	\$ 2.168.289,53	19,64%				
CLOROX	\$ 85.153,79	0,77%	\$ 139.240,37	1,26%	\$ 121.365,83	1,10%	\$ 97.546,24	0,88%	\$ 131.012,56	1,19%	\$ 111.557,15	1,01%	\$ 685.875,79	6,21%	\$ 1.371.751,72	12,42%				
BLANCATEL	\$ 82.805,61	0,75%	\$ 54.665,12	0,50%	\$ 38.731,16	0,35%	\$ 17.952,13	0,16%	\$ 16.189,47	0,15%	\$ 115.898,52	1,05%	\$ 326.242,00	2,95%	\$ 652.484,02	5,91%				
OXI CLEAN	\$ 27.537,77	0,25%	\$ 39.782,01	0,36%	\$ 40.591,21	0,37%	\$ 26.190,00	0,24%	\$ 59.094,66	0,54%	\$ 25.289,76	0,23%	\$ 218.485,32	1,98%	\$ 436.970,72	3,96%				
DR.BECKMANN	\$ 11.322,87	0,10%	\$ 24.891,29	0,23%	\$ 23.159,07	0,21%	\$ 17.630,28	0,16%	\$ 49.635,84	0,45%	\$ 13.220,32	0,12%	\$ 139.859,58	1,27%	\$ 279.719,23	2,53%				
BANDERA	\$ 9.734,29	0,09%	\$ 38.825,62	0,35%	\$ 16.938,03	0,15%	\$ 19.371,73	0,18%	\$ 21.706,80	0,20%	\$ 28.392,05	0,26%	\$ 134.968,38	1,22%	\$ 269.936,90	2,44%				
LOS PATITOS	\$ 34,51	0,00%	\$ 106,47	0,00%	\$ 2.573,08	0,02%	\$ 5.741,46	0,05%	\$ 60.865,25	0,55%	\$ 2.052,08	0,02%	\$ 71.372,81	0,65%	\$ 142.745,66	1,29%				
CLARASOL	\$ 82,76	0,00%	\$ 390,89	0,00%	\$ 1.056,01	0,01%	\$ 19.240,69	0,17%	\$ 2.105,89	0,02%	\$ 47.629,86	0,43%	\$ 70.506,08	0,64%	\$ 141.012,18	1,28%				
LYSOL	\$ 3.846,48	0,03%	\$ 5.787,08	0,05%	\$ 4.645,36	0,04%	\$ 4.515,48	0,04%	\$ 12.581,66	0,11%	\$ 3.902,62	0,04%	\$ 35.278,66	0,32%	\$ 70.557,35	0,64%				
CLORTEX		0,00%	\$ 18.170,59	0,16%		0,00%	\$ 287,15	0,00%		0,00%		0,00%	\$ 18.457,68	0,17%	\$ 36.915,42	0,33%				
CUELLOS PUNOS Y MANCHITAS	\$ 751,73	0,01%	\$ 2.405,14	0,02%	\$ 1.699,16	0,02%	\$ 1.399,34	0,01%	\$ 3.660,86	0,03%	\$ 437,87	0,00%	\$ 10.348,08	0,09%	\$ 20.696,17	0,19%				
SANYTOL	\$ 484,22	0,00%	\$ 544,05	0,00%	\$ 1.112,90	0,01%	\$ 1.029,58	0,01%	\$ 3.108,99	0,03%	\$ 724,08	0,01%	\$ 7.003,80	0,06%	\$ 14.007,62	0,13%				
IBERIA	\$ 415,59	0,00%	\$ 1.010,11	0,01%	\$ 1.021,79	0,01%	\$ 803,21	0,01%	\$ 1.654,01	0,01%	\$ 650,90	0,01%	\$ 5.555,56	0,05%	\$ 11.111,18	0,10%				
OTHERS MARCA UNIF.	\$ 517,25	0,00%	\$ 2.214,77	0,02%	\$ 277,94	0,00%	\$ 553,87	0,01%	\$ 1.001,70	0,01%	\$ 907,38	0,01%	\$ 5.472,86	0,05%	\$ 10.945,76	0,10%				
CLORALUZ		0,00%		0,00%		0,00%		0,00%		0,00%		0,00%	\$ 3.687,39	0,03%	\$ 3.687,39	0,03%	\$ 7.374,79	0,07%		
LA VALENCIANA	\$ 1,52	0,00%	\$ 9,80	0,00%	\$ 9,92	0,00%	\$ 3.029,81	0,03%	\$ 190,54	0,00%	\$ 239,91	0,00%	\$ 3.481,39	0,03%	\$ 6.962,89	0,06%				
TIDE	\$ 337,68	0,00%	\$ 130,78	0,00%	\$ 268,55	0,00%	\$ 202,78	0,00%	\$ 649,09	0,01%	\$ 171,06	0,00%	\$ 1.759,93	0,02%	\$ 3.519,87	0,03%				
SHOUT	\$ 122,53	0,00%	\$ 500,53	0,00%	\$ 143,08	0,00%	\$ 64,13	0,00%	\$ 122,29	0,00%	\$ 28,16	0,00%	\$ 980,72	0,01%	\$ 1.961,45	0,02%				
PATITO		0,00%	\$ 801,34	0,01%		0,00%		0,00%		0,00%		0,00%		0,00%	\$ 801,34	0,01%	\$ 1.602,68	0,01%		
BCOLORHADA	\$ 35,20	0,00%	\$ 58,38	0,00%	\$ 54,07	0,00%	\$ 44,81	0,00%	\$ 48,51	0,00%	\$ 42,19	0,00%	\$ 283,12	0,00%	\$ 566,28	0,01%				
PURO SOL	\$ 36,85	0,00%		0,00%	\$ 0,19	0,00%		0,00%		0,00%		0,00%	\$ 37,04	0,00%	\$ 74,08	0,00%				
LA JAPONESA		0,00%	\$ 17,92	0,00%		0,00%		0,00%		0,00%		0,00%	\$ 17,92	0,00%	\$ 35,84	0,00%				
HARPIC	\$ 0,04	0,00%		0,00%		0,00%		0,00%		0,00%		0,00%	\$ 0,04	0,00%	\$ 0,07	0,00%				
Grand Total	\$ 714.249,98	6,47%	\$ 1.188.796,15	10,77%	\$ 803.655,34	7,28%	\$ 677.436,00	6,13%	\$ 1.153.335,54	10,44%	\$ 983.957,57	8,91%	\$ 5.521.429,32	50,00%	\$ 11.042.859,89	100,00%				

Tabla Dinámica con filtros y Estadística Descriptiva

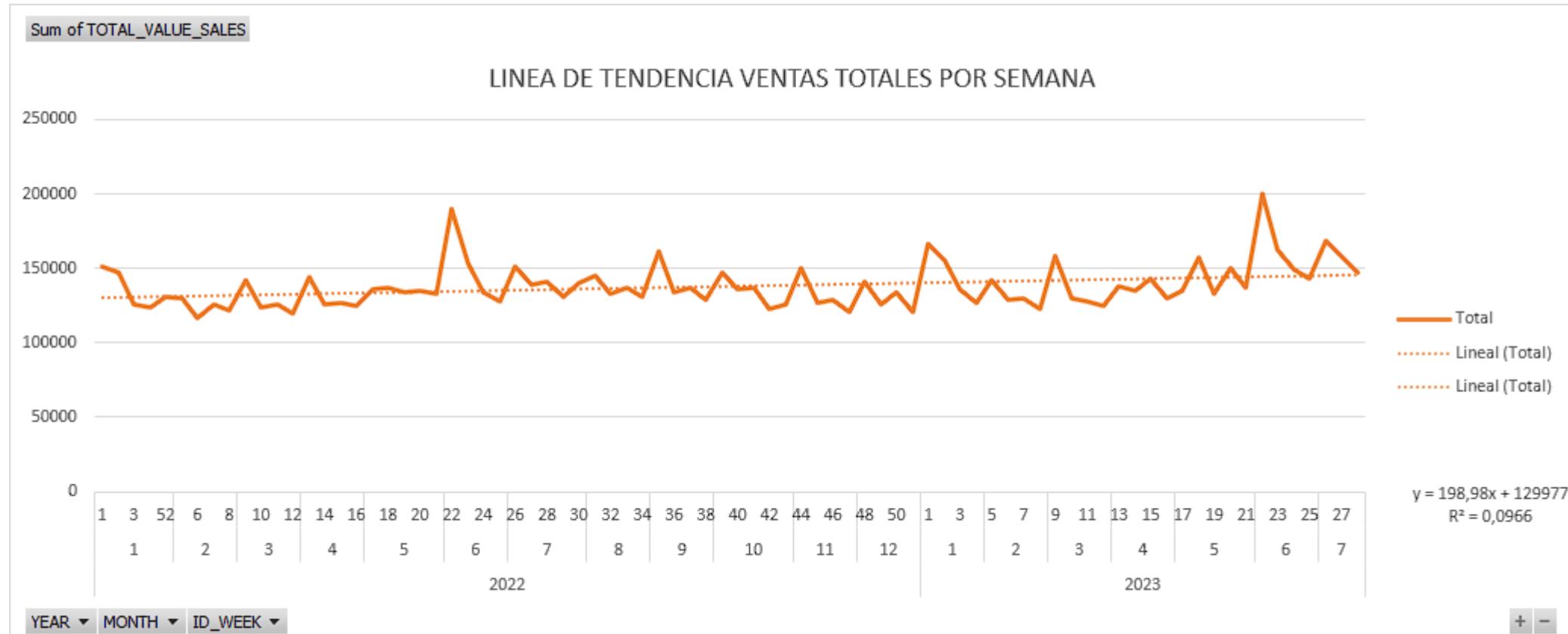
CC	VENTAS	UNID PROM.	S % VENTAS ANUALES	Ticket Prom. X
2022				
①	\$ 678.440,71	10,75	9,68%	\$ 67.751,41
②	\$ 494.535,23	10,09	7,06%	\$ 52.913,93
③	\$ 510.052,35	10,12	7,28%	\$ 53.333,81
④	\$ 522.004,23	9,60	7,45%	\$ 53.799,74
⑤	\$ 674.949,56	10,02	9,63%	\$ 69.506,02
⑥	\$ 603.800,02	11,17	8,62%	\$ 53.630,52
⑦	\$ 702.062,39	10,31	10,02%	\$ 67.651,87
⑧	\$ 545.441,05	10,17	7,78%	\$ 53.639,74
⑨	\$ 561.226,13	10,65	8,01%	\$ 55.158,53
⑩	\$ 668.951,12	9,80	9,54%	\$ 71.121,06
⑪	\$ 526.050,11	9,54	7,51%	\$ 56.867,96
⑫	\$ 521.106,15	9,35	7,44%	\$ 57.253,68
2023				
①	\$ 584.020,63	10,26	14,48%	\$ 59.900,98
②	\$ 523.571,53	9,01	12,98%	\$ 58.954,81
③	\$ 541.149,85	9,77	13,41%	\$ 59.911,75
④	\$ 546.329,64	9,65	13,54%	\$ 61.171,23
⑤	\$ 711.883,29	9,68	17,65%	\$ 77.639,06
⑥	\$ 654.159,28	11,25	16,22%	\$ 60.539,58
⑦	\$ 473.126,63	10,76	11,73%	\$ 46.012,45
Grand Total	\$ 11.042.859,89	10,10	#####	#####

	VENTAS	UNID PROM.	S % VENTAS ANUALES	Ticket Prom. X
MEDIA	\$ 581.203,15	\$ 10,10	\$ 0,11	\$ 59.829,37
MEDIANA	\$ 546.329,64	\$ 10,09	\$ 0,10	\$ 58.954,81
VARIANZA	5563200450	0,348286098	0,001024035	57715491,13
DESV. TIPICA	74586,86513	0,590157689	0,032000543	7597,07122

YEAR
2022
2023

MONTH
1
2
3
4
5
6
7
8
9
10
11
12

Gráfico tendencia de ventas totales por semana



VENTAS MENSUALES POR REGION

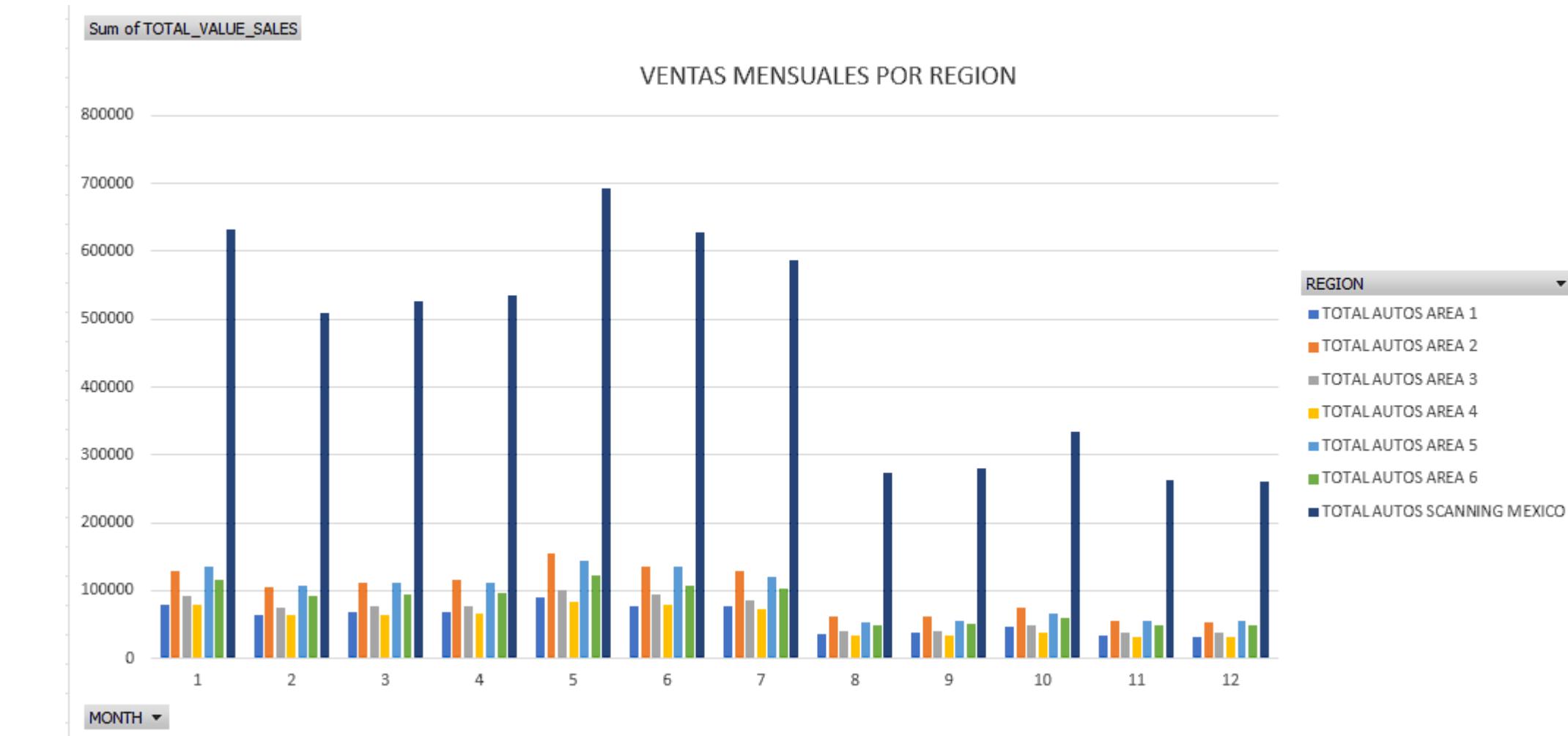


Gráfico ventas por marca

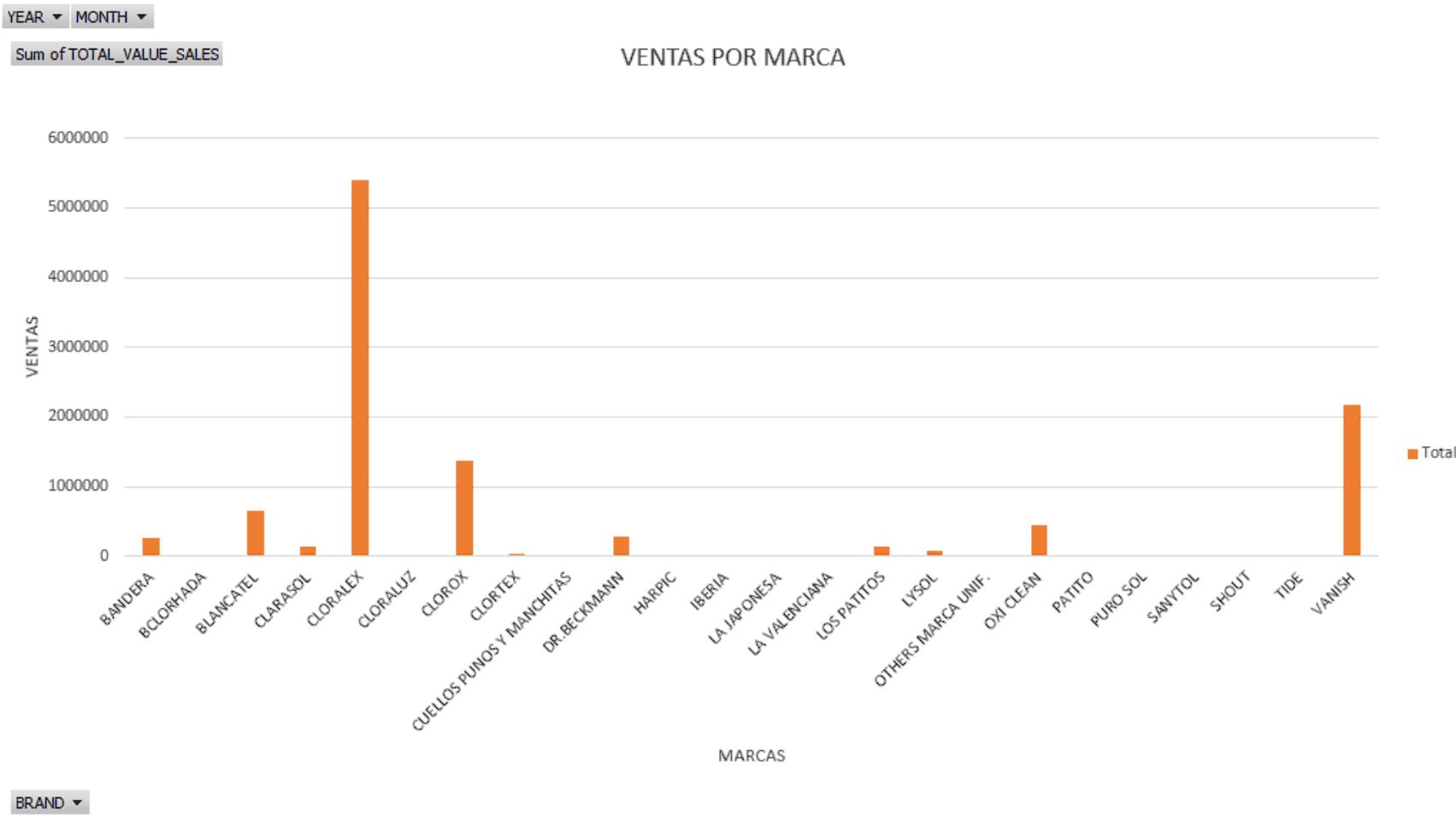


Gráfico correlación unidades vs ventas

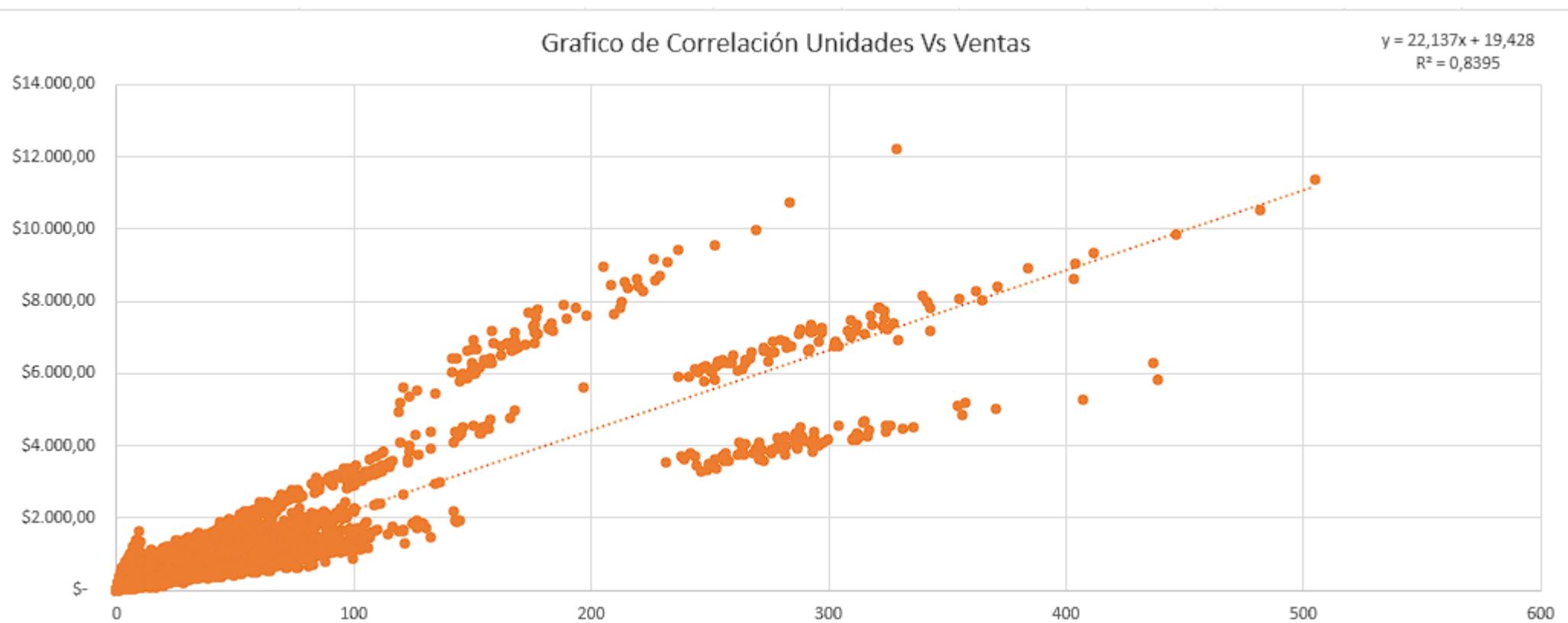
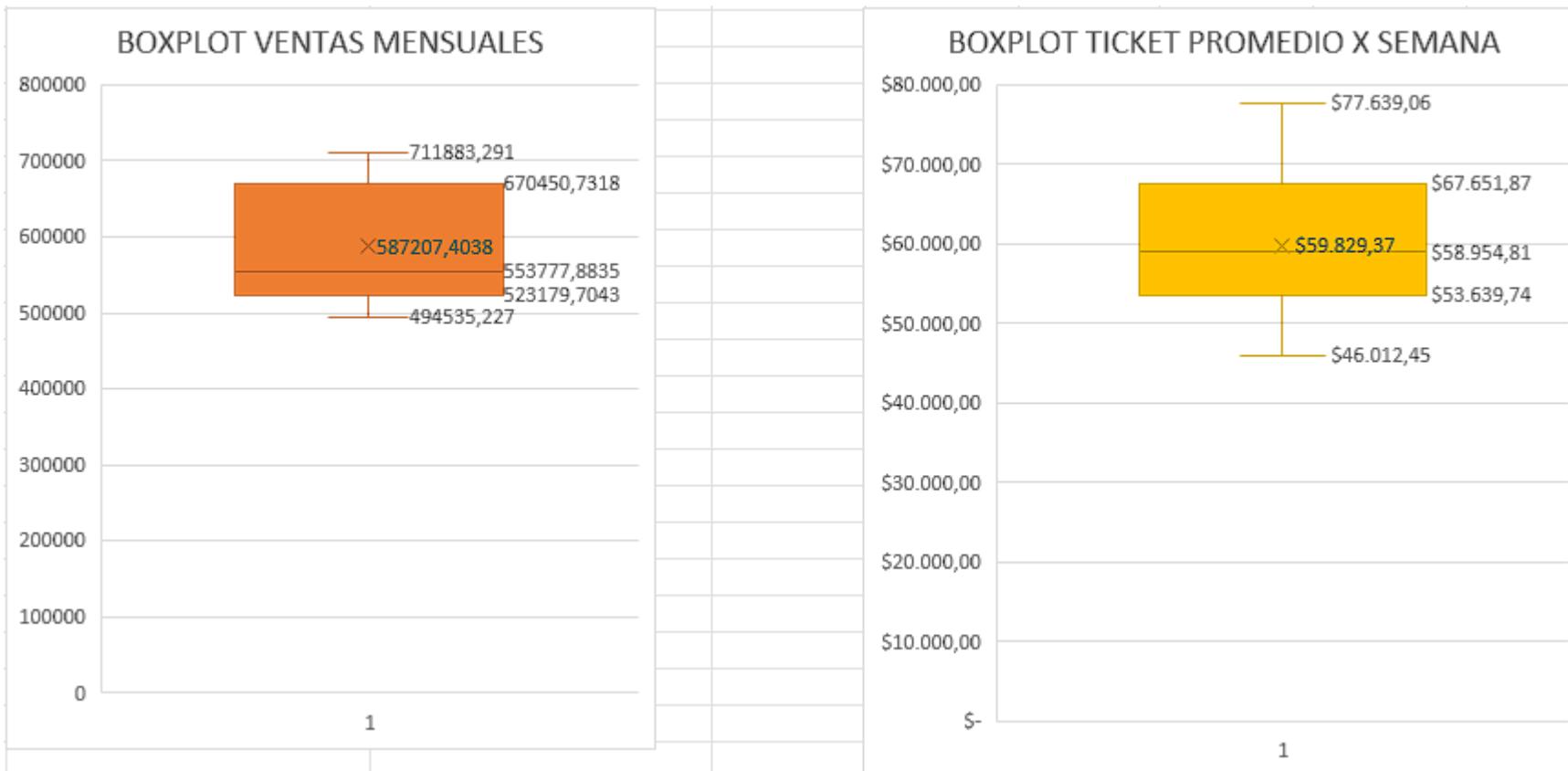


Gráfico Boxplot



Consideraciones y Hallazgos

RESULTADOS DEL ANALISIS DE DATOS DE VENTA

CONSIDERACIONES

La data suministrada tiene ventas desde enero de 2022 hasta julio de 2023, por lo que no se puede hacer un comparativo entre dos años de manera completa, y el mes de Julio de 2023 solo tiene datos de 3 semanas por lo que tambien puede sesgar un poco la información.

Los datos no muestran en que unidades estan ya que se tienen casos de ventas de unidades de 0,006 lo cual pueden ser cientos, miles, o millones. Como no se especificó se proceso la información como estaba, y se debe contemplar hacer la conversión para los datos finales.

Todos los datos suministrados de ventas perteneces a una sola categoría llamada FABRIC TREATMENT and SANIT por lo que no hay forma de hacer una segmentación por esta variable, pero al interior de la tabla de Productos se encuentra la MARCA, lo cual puede ser una segmentación importante para visualizar el comportamiento de ventas, por lo que se estará usando como un instrumento de medida adicional.

HALLAZGOS

Se reportan 24 marcas vendidas, y solo 3 de ellas hacen el "PARETO" (80/20) en las ventas. Estas marcas son Cloralex, Vanish, Clorox. En su mismo orden son la gran representación del 80% de todas las ventas, y aún más importante sólo Cloralex representa casi el 50% de las ventas. Este dato es consistente no sólo en ventas totales, sino en cada una de las regiones de manera independiente.

La región Scanning México vende el 50% y lo demás está distribuido en las otras 6 zonas.

Las siguientes regiones con mejor desempeño son las Areas 2,5 y 6 en su respectivo orden aportando así un 30,12% de las ventas.

Estos son los mejores 5 meses de ventas en su orden:

Mayo de 2023

Julio de 2022

Enero de 2022

Mayo de 2022

Octubre de 2022

En particular en Mayo de 2022 igualmente se tiene un ticket promedio mas alto y una cantidad de productos mas baja, lo cual indica que se vendieron los productos a mejor precio o hubo venta de productos especialmente que son mas caros.

El número de unidades vendidas en promedio por semana es bastante consistente todos los meses.

Los meses con mejor ticket promedio coinciden con los mejores meses de ventas, con la particularidad que antes se mencionó ya que en Mayo de 2023 además de mejores ventas y ticket promedio, la cantidad de productos no fue tan alta.

No se detectan valores atípicos en la mayoría de KPIs de rendimiento, por lo general el comportamiento de ventas es muy estandar, y no hay mucha desviación en ventas, cantidad de productos y ticket promedio. Ademas de las tablas las gráficas de Boxplot nos permiten confirmar esto claramente.

La línea de tendencia de ventas durante los 18 meses va al alza levemente, lo que indica que las ventas en general van subiendo, pero de una manera muy lenta.

El gráfico de correlación entre unidades vendidas y ventas es uno de los hallazgos mas importantes en este análisis.

Hay una gran correlación entre productos con menos ventas y menos cantidad vendida, pero en la medida que sube el precio y la cantidad de productos vendidos, logramos visualizar 3 líneas de tendencia bien marcadas.

El coeficiente de correlación es bastante alto y R2 nos da un 83,95% de confiabilidad en esta tendencia.

Recomendaciones

RECOMENDACIONES

pero a la vez es peligroso porque algun problema con una de ellas puede ser un desatre, por lo cual es importante en un momento donde todo anda bien consolidado, poder hacer un esfuerzo con marcas que pueden hacer un aporte importante a las ventas, y con esto tratar de lograr un curva mas acelerada de crecimiento. Solo estas 5 marcas hoy aportan el 16,14%, es decir casi el otro porcentaje restante del total, por lo hacerlas crecer puede resultar muy ventajoso para la compañía.

BLANCATEL

OXI CLEAN

DR. BECKMANN

BANDERA

LOS PATITOS

Son varias las razones por las que una región puede ser buena en ventas o no, y puede ser la disponibilidad de producto, la situación financiera a nivel geográfico, el posicionamiento local, la segmentación, el branding, etc. Se tienen 3 regiones que no son representativas en las ventas y lo que se puede hacer es un estudio que determine las variables para hacerla mas rentable, o no invertirle mas.

Por otro lado estan las 3 regiones (areas 2, 5 y 6) que conforman ayudan a scanning Mexico que representa el 50% de las ventas a llegar al 80% (Pareto) de ventas. Junto con la recomendación anterior se puede hacer un mix para incentivar en especial estas 4 areas que estan mejor posicionadas para hacer una mejor incursión con las 5 marcas propuestas anteriormente.

Se logro determinar un comportamiento atípico, pero importante en el mes de mayo de 2023 donde quizás pudo ser un mes muy rentable. Fueron las mejores ventas en 18 meses, pero ademas vendiendo menos cantidad de productos y por ende con un ticket de venta mucho mayor. Vale la pena investigar de cerca este caso especial, ya que puede que una estrategia de venta sin promocion haya dado buen resultado, o bien se hayan vendido productos que no suelen venderse que son de mas alto valor. Como sea, identificar cual fue la causa real de este suceso puede ayudar a generar estrategias para repetir esta situación de manera exponencial, generando asi mas ventas con menos esfuerzo.

Descubrir que los KPIs de rendimientos son muy estables son buenos para entender que se tiene una compañía sólida, pero tambien denota en que no hay un crecimiento importante en ventas, y por esto se hace necesario revisar las estrategias si este es el objetivo de la compañía, ya que hemos podido detectar varias oportunidades en cuanto a producto y a regiones de venta.

Como se mencionaba en los hallazgos la correlación entre el número de unidades vendidas y las ventas es estupendamente alto, y es congruente con el comportamiento de ventas, pero lo que se nota en la gráfica nos da idea de 3 nichos que se pueden explorar de manera independiente y poder sacar el mejor partido. Esto quiere decir que se forman algo así como 3 líneas de correlación altamente probable, por lo que al identificar cuales son esos productos, sus nichos, su segmentación, su rango de precio; se pueden generar estrategias comerciales que atiendan esas condiciones y se puede subir paralelamente el numero de unidades aumentando los ingresos de manera más exponencial.

Esta es una estrategia que se debe de investigar y unir a las ideas que se puedan realizar con los productos y regiones a desarrollar.

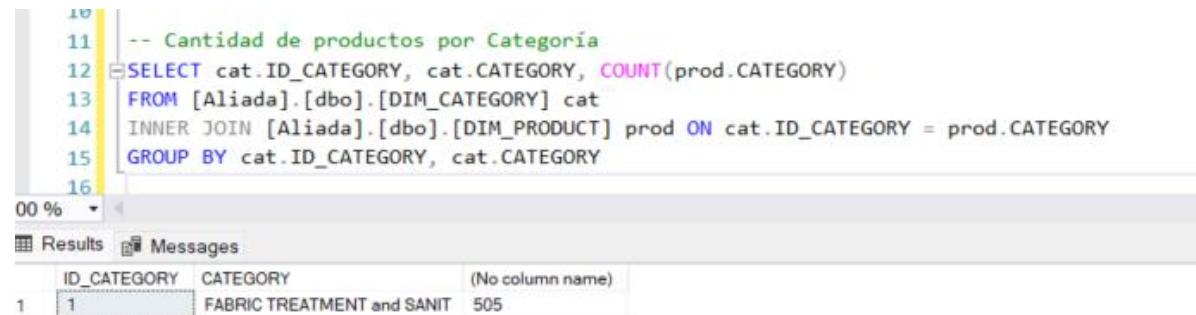
Análisis de Datos con SQL

Creación de la base de datos

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left displays the database structure for 'localhost (SQL Server 16.0.1000.6 - Asus-PauloOssa\paob2)'. The 'Tables' node under 'Aliada' is expanded, showing various fact and dimension tables. The 'Results' tab in the center displays the output of a SQL query. The query selects top 1000 rows from the 'FACT_SALES' table in the 'Aliada' database, specifically targeting columns related to weeks, item codes, total unit sales, total value sales, average weekly sales, and region.

WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION
1	34-22	7501058792808BP2	0,00600000005215406	0,13899998569489	TOTAL AUTOS AREA 5
2	34-22	7501058715883	0,48699988555908	116,518997192383	TOTAL AUTOS AREA 5
3	34-22	7702626213774	1,39100003242493	68,4530029296875	TOTAL AUTOS AREA 5
4	34-22	7501058716422	0,021999998807907	1,48099994659424	TOTAL AUTOS AREA 5
5	34-22	7501058784353	2,03699994087219	182,839004516602	5,375
6	34-22	7501058716064	0,004999998824129	0,679000020027161	TOTAL AUTOS AREA 5
7	34-22	7501058794963BP1	0,0040000001898881	0,552999973297119	TOTAL AUTOS AREA 5
8	34-22	7501058792778BP1	0,00100000004749745	0,128000006079674	TOTAL AUTOS AREA 5
9	34-22	7501058757630	0,716000020503998	71,0179977416992	4,28700017929077
10	34-22	7501058753441	2,703999961853	61,2290000915527	TOTAL AUTOS AREA 5
11	34-22	7702626204208BP1	2,18799996376038	23,7919998168945	17,9340000152588
12	34-22	7501058784346	0,81099988509583	146,759002685547	3,7550001444092
13	34-22	7501058796868	3,50699996948242	55,8559989929199	28,281999580127
14	34-22	7891035040276	6,53900003433228	80,1689987182617	TOTAL AUTOS AREA 5
15	34-22	7501058796851	2,54699993133545	40,3359985351563	20,5400009155273
16	34-22	7501058714596	6,77400016784668	86,1039962768555	21,5049991607666
17	34-22	7891035040191	0,0080000037997961	0,104000002145767	1,60000002384186
18	34-22	7501058751188	0,99699979019165	93,1240005493164	4,27899980545044
19	34-22	7501058716095	0,078000016093254	12,9530000686646	2,16700005531311
20	34-22	7501058751461	0,35100010967255	66,0530014038086	3,31100010871887
21	34-22	7501058757463	0,0040000018998981	0,159999996423721	2
22	34-22	7501058757555	1,42799997329712	51,0620002746582	7,93300008773804
23	34-22	7501058751195	2,53200006484985	129,975997924805	7,79099988937378
24	34-22	0046500022692	0,024000002086163	2,03999996185303	3,42899990081787
25	34-22	0046500729355	0,0010000004749745	0,04600000089407	1
26	34-22	7501268200001	0,18400000357628	1,92499995231628	8,36400032043457
27	34-22	7501268200506	0,0030000002607703	0,06199999886721	1,5
28	34-22	7501268200339	0,0410000011324883	1,31200003623962	4,55600023269653
29	34-22	7501268200025	0,013000002682209	0,14300000667572	13
30	34-22	7501268200049	0,68199981403351	12,8050003051758	8,52499961853027
31	34-22	7501268200070	0,24300000715256	8,65100002288818	4,5850003814697
32	34-22	7501268200261	0,0049999988824129	0,23399997377396	5
33	34-22	7501268200513	0,40599998831749	4,42500019073486	5,88399982452393
34	34-22	0757037517123	0,0010000004749745	0,041999994337559	1
35	34-22	7501080900585	0,0719999969005585	18,8799991607666	1,30900001525879
36	34-22	7501080954786	0,85399972343445	19,8810005187988	3,2109996566772

Consulta por categoría



The screenshot shows a SQL query being run in a database environment. The code is as follows:

```
10 -- Cantidad de productos por Categoría
11
12 SELECT cat.ID_CATEGORY, cat.CATEGORY, COUNT(prod.CATEGORY)
13 FROM [Aliada].[dbo].[DIM_CATEGORY] cat
14 INNER JOIN [Aliada].[dbo].[DIM_PRODUCT] prod ON cat.ID_CATEGORY = prod.CATEGORY
15 GROUP BY cat.ID_CATEGORY, cat.CATEGORY
16
```

The results pane shows one row of data:

ID_CATEGORY	CATEGORY	(No column name)
1	FABRIC TREATMENT and SANIT	505

Este dato es importante, porque el campo categoría pierde relevancia, dado todos los productos están asociados a la misma y única categoría existente.

Consulta número de productos por fabricante

Esta información nos permite ver que los 505 productos se distribuyen en 20 Fabricantes, y que solo 5 de ellos (Reckitt, Inds. Alen, Clorox, Church & Dwight y Jabonera la Espuma) suman 407 correspondiente a un poco mas del 80% de todos los productos. Es decir que la cantidad mas relevante de productos son atendidas por $\frac{1}{4}$ de los fabricantes.

Es importante ahora si determinar si las ventas están en estos productos o en otros para poder entender la importancia de los demás fabricantes.

```
17 -- Cantidad de productos por Fabricante
18 SELECT MANUFACTURER, COUNT(MANUFACTURER) AS "Nproductos"
19 FROM [Aliada].[dbo].[DIM_PRODUCT]
20 GROUP BY MANUFACTURER
21 ORDER BY Nproductos DESC
22
23
```

100 %

Results Messages

	MANUFACTURER	Nproductos
1	RECKITT	128
2	INDS. ALEN	127
3	CLOROX	94
4	CHURCH & DWIGHT	31
5	JABONERA LA ESPUMA	27
6	DELTA PRONATURA	22
7	OTHERS FABRICANTE UNIF.	21
8	QUIMICA GONCAL	13
9	BLANQUEADORA MEXICANA	9
10	CLARASOL	6
11	IBERIA	5
12	SC JOHNSON	5
13	P&G	4
14	INDS. H24	3
15	GRUPO AC MARCAS	3
16	FAB. DE JABON LA REINERA	2
17	LA CORONA	2
18	LA JAPONESA	1
19	WHITE MAGIC	1
20	ARTIC WHITE	1

Consulta por marcas

En cuanto al análisis de productos por marcas, tenemos un total de 33 marcas, las cuales el pareto lo tenemos con solo 7 de ellas es decir con el 21% de las marcas.

```
22 -- Cantidad de productos por Marca
23 SELECT BRAND, COUNT(BRAND) AS "Nproductos"
24 FROM [Aliada].[dbo].[DIM_PRODUCT]
25 GROUP BY BRAND
26 ORDER BY Nproductos DESC
27
```

100 %

Results Messages

	BRAND	Nproductos
1	CLORALEX	115
2	VANISH	103
3	CLOROX	94
4	OXI CLEAN	31
5	BANDERA	25
6	DR.BECKMANN	22
7	OTHERS MARCA UNIF.	21
8	VANISH PODER 02	10
9	BLANCATEL	10
10	CLOREX	8
11	LYSOL	8
12	LOS PATITOS	6
13	CLARASOL	6
14	IBERIA	5
15	SHOUT	5
16	TIDE	4
17	VANISH CRYSTAL WHITE	4
18	SANYTOL	3
19	PATITO	3
20	LA VALENCIANA	3
21	CUELLOS PUNOS Y MANCHITAS	3
22	HARPIC	2
23	CLORO 10	2
24	BCOLORHADA	2
25	MARIPOSA	2
26	PURO SOL	1

Query executed successfully.

Análisis ventas por semana

```
34  
35 -- Ventas por semanas  
36 SELECT WEEK, COUNT(ITEM_CODE) as NProductos, SUM(TOTAL_VALUE_SALES) as Total_Ventas_Semana  
37 FROM Aliada.dbo.FACT_SALES  
38 GROUP BY WEEK  
39
```

100 %

Results Messages

WEEK	NProductos	Total_Ventas_Semana
57	21-23	1449
58	13-23	1469
59	40-22	1529
60	18-23	1445
61	23-23	1442
62	44-22	1502
63	16-23	1456
64	03-22	1595
65	36-22	1549
66	06-23	1468
67	15-23	1461
68	51-22	1443
69	10-22	1622
70	20-22	1607
71	46-22	1461
72	50-22	1464
73	11-22	1615
74	35-22	1549
75	10-23	1475
76	21-22	1607
77	49-22	1449
78	38-22	1514
79	03-23	1455
80	25-23	1430

Query executed successfully.

localhost (16.0 RTM) | ASUS-PAULOSSA\paob2 (55) | Aliada | 00:00:00 | 80 rows

En este resultado podemos ver 80 resultados de semanas cuando en la tabla original de semanas tenemos 156 registros. Esto significa que podemos tener un periodo mucho mas reducido de data, y es mucho mejor entenderlo por meses y años.

```
31  
32 -- Total de registros de la tabla  
33 SELECT COUNT(*) as registros FROM Aliada.dbo.FACT_SALES
```

100 %

Results Messages

registros	
1	122002

Tenemos un total de 122002 registros de ventas en la tabla.

Consulta por mes

```
40  
41  -- Ventas por meses  
42  SELECT cal.YEAR, cal.MONTH, COUNT(fact.ITEM_CODE) as NProductos, SUM(fact.TOTAL_VALUE_SALES) as Total_Ventas_Semana  
43  FROM Aliada.dbo.DIM CALENDAR cal  
44  RIGHT JOIN Aliada.dbo.FACT SALES fact ON cal.WEEK = fact.WEEK  
45  GROUP BY cal.YEAR, cal.MONTH  
46  ORDER BY cal.YEAR, cal.MONTH  
47
```

100 %

Results Messages

	YEAR	MONTH	NProductos	Total_Ventas_Semana
1	2022	1	7861	678440,707814707
2	2022	2	6463	494535,226700404
3	2022	3	6465	510052,354504012
4	2022	4	6469	522004,232771203
5	2022	5	8056	674949,557003791
6	2022	6	6365	603800,023925685
7	2022	7	7842	702062,391573925
8	2022	8	6226	545441,046858912
9	2022	9	6140	561226,130288479
10	2022	10	7577	668951,122720982
11	2022	11	5914	526050,105095102
12	2022	12	5827	521106,150401151
13	2023	1	5868	584020,628877795
14	2023	2	5869	523571,528388259
15	2023	3	5891	541149,853695675
16	2023	4	5852	546329,6366596079
17	2023	5	7241	711883,291287403
18	2023	6	5780	654159,278336024
19	2023	7	4206	473126,625821702

Con este query podemos visualizar que tenemos datos de ventas de Enero de 2022 hasta el mes de Julio de 2023, por tanto no es posible hacer un comparativo total de años pero si entre meses con un año de diferencia (Enero-Julio)

Comportamiento mensual por región

```
47 -- Ventas por mes por Region
48
49 SELECT fact.REGION, cal.YEAR, cal.MONTH, COUNT(fact.ITEM_CODE) as NProductos, SUM(fact.TOTAL_VALUE_SALES) as Total_Ventas_Mes
50 FROM Aliada.dbo.DIM_CALENDAR cal
51 RIGHT JOIN Aliada.dbo.FACT_SALES fact ON cal.WEEK = fact.WEEK
52 GROUP BY fact.REGION, cal.YEAR, cal.MONTH
53 ORDER BY fact.REGION, cal.YEAR, cal.MONTH
```

100 %

Results Messages

	REGION	YEAR	MONTH	NProductos	Total_Ventas_Mes
1	TOTAL AUTOS AREA 1	2022	1	1021	42964,722009358
2	TOTAL AUTOS AREA 1	2022	2	840	31327,186153153
3	TOTAL AUTOS AREA 1	2022	3	833	32942,2650199542
4	TOTAL AUTOS AREA 1	2022	4	831	33573,3929293384
5	TOTAL AUTOS AREA 1	2022	5	1037	42691,0660845153
6	TOTAL AUTOS AREA 1	2022	6	833	37365,1599637098
7	TOTAL AUTOS AREA 1	2022	7	1025	45754,1261318726
8	TOTAL AUTOS AREA 1	2022	8	812	36600,2128210431
9	TOTAL AUTOS AREA 1	2022	9	794	39289,6179476769
10	TOTAL AUTOS AREA 1	2022	10	971	47230,4338481767
11	TOTAL AUTOS AREA 1	2022	11	769	34674,6108213094
12	TOTAL AUTOS AREA 1	2022	12	746	32806,9179484688
13	TOTAL AUTOS AREA 1	2023	1	757	36394,6499412082
14	TOTAL AUTOS AREA 1	2023	2	750	33359,8910011295
15	TOTAL AUTOS AREA 1	2023	3	755	34768,4199012271
16	TOTAL AUTOS AREA 1	2023	4	754	34908,5578704063
17	TOTAL AUTOS AREA 1	2023	5	942	46842,1799952583
18	TOTAL AUTOS AREA 1	2023	6	759	40361,6752342517
19	TOTAL AUTOS AREA 1	2023	7	552	30394,8930187463
20	TOTAL AUTOS AREA 2	2022	1	1183	68856,508914439
21	TOTAL AUTOS AREA 2	2022	2	980	50794,2850963016
22	TOTAL AUTOS AREA 2	2022	3	985	53745,3050115798
23	TOTAL AUTOS AREA 2	2022	4	982	56586,4100441253
24	TOTAL AUTOS AREA 2	2022	5	1218	75980,2490729424
25	TOTAL AUTOS AREA 2	2022	6	961	66200,4399444573

Esta tabla nos permite el comportamiento mes a mes de cada Region.

De acuerdo a esta información podemos hacer el siguiente comparativo

Comportamiento mensual por región

REGIÓN	PROM PRODUCTOS MENSUALES	PROMEDIO VENTAS MENSUALES	MEJOR MES	PEOR MES
ÁREA 1	830.57	37592.10	OCTUBRE/22	JULIO/23
ÁREA 2	973.94	62568.22	MAYO/23	FEBRERO/22
ÁREA 3	868.21	42297.65	MAYO/23	JULIO/23
AREA 4	904.31	35654.53	JULIO/22	JULIO/23
AREA 5	849.21	60701.83	ENERO/22	JULIO/23
AREA 6	890.15	51787.24	MAYO/23	JULIO/23
SCANNING MÉXICO	1104.74	290601.54	MAYO/23	JULIO/23

Ya con esta construcción vamos deduciendo que definitivamente su mejor Regio de ventas es Scanning México seguido del Area 2, y por otro lado la peor área es el Area 4 y le sigue el area 3. Mirando la frecuencia de mejor y peor mes, podemos hacer una deducción rápida y de gran certeza como que un gran mes de ventas fue definitivamente mayo de 2023 y el peor fue Julio de 2023, esto quizas causado por las grandes ventas dos meses atras.

Productos más vendidos 2022

```
55 -- Top Ventas de productos por año 2022
56 SELECT TOP 10 cal.YEAR, fact.ITEM_CODE, prod.ITEM_DESCRIPTION, SUM(fact.TOTAL_VALUE_SALES) as Total_Ve
57 FROM Aliada.dbo.DIM_CALENDAR cal
58 RIGHT JOIN Aliada.dbo.FACT_SALES fact ON cal.WEEK = fact.WEEK
59 LEFT JOIN Aliada.dbo.DIM_PRODUCT prod ON fact.ITEM_CODE = prod.ITEM
60 WHERE cal.YEAR = 2022
61 GROUP BY cal.YEAR, fact.ITEM_CODE, prod.ITEM_DESCRIPTION, prod.MANUFACTURER, prod.BRAND
62 ORDER BY cal.YEAR, Total_Ventas_Mes DESC
```

100 %

Results Messages

	YEAR	ITEM_CODE	ITEM_DESCRIPTION	Total_Ventas_Mes	MANUFACTURER	BRAND
1	2022	0000075000622	CLORALEX EL RENDIDOR BOT.PLAST. 2000ML NAL 000007500...	713986,508239746	INDS. ALEN	CLORALEX
2	2022	0000075000639	CLORALEX EL RENDIDOR B <small>Click to select the whole column</small>	689826,967529297	INDS. ALEN	CLORALEX
3	2022	0000075000615	CLORALEX EL RENDIDOR BOT.PLAST. 950ML NAL. 000007500...	417912,376678467	INDS. ALEN	CLORALEX
4	2022	7501025402051	BLANCATEL CONCENTRADO BOT PLAST 3750ML NAL 7501025...	373008,626571655	INDS. ALEN	BLANCATEL
5	2022	7501071900143	CLOROX REGULAR CONCENTRADO BOTPLAST.3800ML 75010...	218818,344268799	CLOROX	CLOROX
6	2022	7501025450212	CLORALEX EL RENDIDOR BOT 1900+100ML= 2LT 7501025450...	159701,056968689	INDS. ALEN	CLORALEX
7	2022	7501025405090	CLORALEX ULTRAGEL CLORO GEL CONCENTRADO BOT 600 M...	158175,850921631	INDS. ALEN	CLORALEX
8	2022	7501025405212	CLORALEX ULTRAGEL CLORO EN GEL CONCENTRADO 950ML ...	147634,200134277	INDS. ALEN	CLORALEX
9	2022	7501025405151	CLORALEX CLORO EN GEL BOT 950ML 7501025405151	122768,55255127	INDS. ALEN	CLORALEX
10	2022	7503002319529	VANISH LIQUIDO BOTELLA 925ML NAL. 7503002319529	104776,628974915	RECKITT	VANISH

Esta tabla nos da información de mucho valor. El primer producto vende casi 7 veces que el décimo así pues las características de producto son bastante relevantes. Al inicio de este análisis vimos como el fabricante Reckitt tiene la mayor cantidad de productos junto con Inds. ALEN. La gran diferencia aca es que quien vende más es el fabricante Inds. ALEN y Reckitt solo logra colocar un producto en el décimo lugar de ventas.

Productos más vendidos 2022

```
64  -- Top Ventas de productos por año 2023
65  SELECT TOP 10 cal.YEAR, fact.ITEM_CODE, prod.ITEM_DESCRIPTION, SUM(fact.TOTAL_VALUE_SALES) as Total_Ventas_Mes, prod.MANUFACTURER, prod.BRAND
66  FROM Aliada.dbo.DIM_CALENDAR cal
67  RIGHT JOIN Aliada.dbo.FACT_SALES fact ON cal.WEEK = fact.WEEK
68  LEFT JOIN Aliada.dbo.DIM_PRODUCT prod ON fact.ITEM_CODE = prod.ITEM
69  WHERE cal.YEAR = 2023
70  GROUP BY cal.YEAR, fact.ITEM_CODE, prod.ITEM_DESCRIPTION, prod.MANUFACTURER, prod.BRAND
71  ORDER BY cal.YEAR, Total_Ventas_Mes DESC |
```

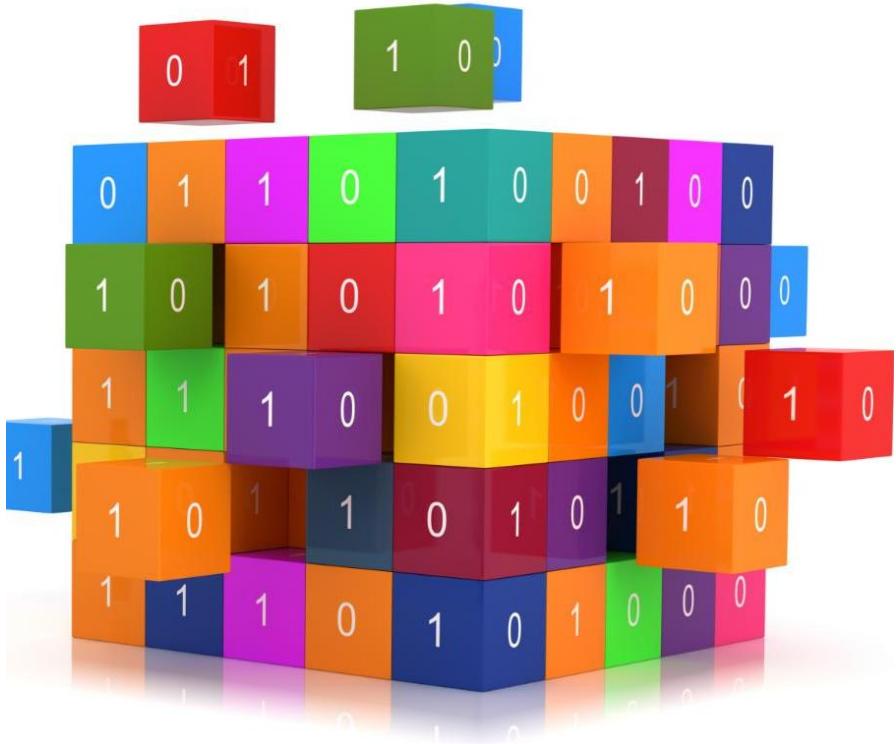
00 %

Results Messages

	YEAR	ITEM_CODE	ITEM_DESCRIPTION	Total_Ventas_Mes	MANUFACTURER	BRAND
1	2023	0000075000639	CLORALEX EL RENDIDOR BOT.PLAST. 3750ML NAL 000007500...	457022.663024902	INDS. ALEN	CLORALEX
2	2023	0000075000622	CLORALEX EL RENDIDOR BOT.PLAST. 2000ML NAL 000007500...	425247.279541016	INDS. ALEN	CLORALEX
3	2023	0000075000615	CLORALEX EL RENDIDOR BOT.PLAST. 950ML NAL 000007500...	240672.1925354	INDS. ALEN	CLORALEX
4	2023	7501071900143	CLOROX REGULAR CONCENTRADO BOTPLAST.3800ML 75010...	181176.434341431	CLOROX	CLOROX
5	2023	7501025402051	BLANCATEL CONCENTRADO BOT PLAST 3750ML NAL 7501025...	118066.604114532	INDS. ALEN	BLANCATEL
6	2023	7501025405106	CLORALEX CLORO EN GEL BP 2000ML 7501025405106	100312.79511261	INDS. ALEN	CLORALEX
7	2023	7501025405212	CLORALEX ULTRAGEL CLORO EN GEL CONCENTRADO 950ML ...	94546.9400177002	INDS. ALEN	CLORALEX
8	2023	7501025450212	CLORALEX EL RENDIDOR BOT 1900+100ML= 2LT 7501025450...	81797.8270492554	INDS. ALEN	CLORALEX
9	2023	7501025405090	CLORALEX ULTRAGEL CLORO GEL CONCENTRADO BOT 600 M...	77608.5357666016	INDS. ALEN	CLORALEX
10	2023	7501025405151	CLORALEX CLORO EN GEL BOT 950ML 7501025405151	77042.900932312	INDS. ALEN	CLORALEX

En general la comparación es muy similar, y por destacar deberíamos decir que Reckitt desaparece del top 10 de ventas para ese año.

Limpieza y Transformación de Datos con Python



Transformaciones

Normalización de Datos

La normalización de datos es un proceso crucial que asegura que los datos estén en un formato coherente y utilizable para el análisis.

Eliminación de Duplicados

Eliminar datos duplicados es esencial para evitar sesgos y garantizar la precisión durante el análisis exploratorio.

Manejo de Valores Nulos

El manejo de valores nulos es vital para mantener la integridad de los datos y evitar errores en las conclusiones del análisis.

Importación de datos

IMPORTACION DE DATOS

```
import pandas as pd

#Generacion de data frames con los 4 archivos dados
df_categorias = pd.read_csv("DIM_CATEGORY.csv")
df_ventas = pd.read_csv("FACT_SALES.csv")
df_calendario = pd.read_excel("DIM_CALENDAR.xlsx")
df_segmentos = pd.read_excel("DIM_SEGMENT.xlsx")
```

Preanálisis Categorías

```
df_categorias.info()
```

[3]

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
---  --  
 0   ID_CATEGORY  5 non-null     int64  
 1   CATEGORY     5 non-null     object 
dtypes: int64(1), object(1)
memory usage: 212.0+ bytes
```

```
df_categorias.describe()
```

[44]

ID_CATEGORY	count	mean	std	min	25%	50%	75%	max
	5.000000	3.000000	1.581139	1.000000	2.000000	3.000000	4.000000	5.000000

```
df_categorias.isnull().sum()
```

[45]

```
... ID_CATEGORY    0
CATEGORY        0
dtype: int64
```

```
#CATEGORIAS
df_categorias.head()
```

[42]

ID_CATEGORY	CATEGORY
0	FABRIC TREATMENT and SANITY\n
1	AIR CARE
2	LAVAJILLAS
3	MEGA SUPERFICIES
4	LAVATORY CARE & BRC

Preanálisis Calendario

```
[49] df_calendario.isnull().sum()
```

```
... WEEK      0
YEAR      0
MONTH     0
WEEK_NUMBER 0
DATE      0
dtype: int64
```

```
[47] df_calendario.info()
```

```
> <class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype  
--- 
 0   WEEK      156 non-null    object 
 1   YEAR      156 non-null    int64  
 2   MONTH     156 non-null    int64  
 3   WEEK_NUMBER 156 non-null  int64  
 4   DATE      156 non-null    datetime64[ns]
dtypes: datetime64[ns](1), int64(3), object(1)
memory usage: 6.2+ KB
```

```
[48] df_calendario.describe()
```

```
...   YEAR      MONTH    WEEK_NUMBER      DATE
count 156.000000 156.000000 156.000000 156
mean 2022.000000 6.487179 26.500000 2022-07-06 20:00:00
min 2021.000000 1.000000 1.000000 2021-01-10 00:00:00
25% 2021.000000 3.750000 13.750000 2021-10-08 06:00:00
50% 2022.000000 6.500000 26.500000 2022-07-06 12:00:00
75% 2023.000000 9.250000 39.250000 2023-04-04 18:00:00
max 2023.000000 12.000000 52.000000 2024-01-01 00:00:00
std 0.819126 3.459419 15.056667          NaN
```

```
[46] #CALENDARIO
df_calendario.head()
```

```
> <class 'pandas.core.frame.DataFrame'>
...   WEEK      YEAR      MONTH    WEEK_NUMBER      DATE
0   01-21  2021      1           1  2021-01-10
1   02-21  2021      1           2  2021-01-17
2   03-21  2021      1           3  2021-01-24
3   04-21  2021      1           4  2021-01-31
4   05-21  2021      2           5  2021-02-07
```

Preanálisis Segmentos

```
[1] df_segmentos.info()
```

df_segmentos.info()

```
[1] <class 'pandas.core.frame.DataFrame'>
RangeIndex: 53 entries, 0 to 52
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype  
--- 
 0   CATEGORY   53 non-null    int64  
 1   ATTR1      53 non-null    object  
 2   ATTR2      53 non-null    object  
 3   ATTR3      52 non-null    object  
 4   FORMAT     53 non-null    object  
 5   SEGMENT    53 non-null    object  
dtypes: int64(1), object(5)
memory usage: 2.6+ KB
```

```
[2] df_segmentos.describe()
```

df_segmentos.describe()

```
[2] [52] ...
```

	CATEGORY
count	53.0
mean	1.0
std	0.0
min	1.0
25%	1.0
50%	1.0
75%	1.0
max	1.0

```
[3] #SEGMENTOS
df_segmentos.head()
```

#SEGMENTOS

```
[3] df_segmentos.head()
```

	CATEGORY	ATTR1	ATTR2	ATTR3	FORMAT	SEGMENT
0	1	CLORO	CLORO	BAMBINO	LIQUIDO	BLEACH
1	1	CLORO	CLORO	GERMICIDA	LIQUIDO	BLEACH
2	1	CLORO	CLORO	MASCOTAS	LIQUIDO	BLEACH
3	1	CLORO	CLORO	MULTIUSOS	GEL	BLEACH
4	1	CLORO	CLORO	MULTIUSOS	LIQUIDO	BLEACH

```
[4] df_segmentos.isnull().sum()
```

df_segmentos.isnull().sum()

```
[4] ...
```

	CATEGORY	ATTR1	ATTR2	ATTR3	FORMAT	SEGMENT
CATEGORY	0					
ATTR1	0					
ATTR2	0					
ATTR3	1					
FORMAT	0					
SEGMENT	0					
dtype: int64						

Preanálisis Ventas

```
[54] #VENTAS
df_ventas.head()

... WEEK ITEM_CODE TOTAL_UNIT_SALES TOTAL_VALUE_SALES TOTAL_UNIT_AVG_WEEKLY_SALES REGION
0 34-22 7501058792808BP2 0.006 0.139 1.000 TOTAL AUTOS AREA 5
1 34-22 7501058715883 0.487 116.519 2.916 TOTAL AUTOS AREA 5
2 34-22 7702626213774 1.391 68.453 5.171 TOTAL AUTOS AREA 5
3 34-22 7501058716422 0.022 1.481 1.833 TOTAL AUTOS AREA 5
4 34-22 7501058784353 2.037 182.839 5.375 TOTAL AUTOS AREA 5
```

```
[55] df_ventas.info()

... <class 'pandas.core.frame.DataFrame'>
RangeIndex: 122002 entries, 0 to 122001
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   WEEK            122002 non-null  object 
 1   ITEM_CODE        122002 non-null  object 
 2   TOTAL_UNIT_SALES 122002 non-null  float64
 3   TOTAL_VALUE_SALES 122002 non-null  float64
 4   TOTAL_UNIT_AVG_WEEKLY_SALES 122002 non-null  float64
 5   REGION          122002 non-null  object 
dtypes: float64(3), object(3)
memory usage: 5.6+ MB
```

Preanálisis Ventas

```
[56] df_ventas.describe()
```

	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES
count	122002.000000	122002.000000	122002.000000
mean	3.211097	90.513761	10.099904
std	14.496009	350.236505	22.650142
min	0.000000	0.001000	0.042000
25%	0.063000	2.662000	2.316000
50%	0.367000	16.812000	3.993500
75%	1.520000	62.961500	8.898000
max	504.681000	12236.759000	794.000000

```
[57] df_ventas.isnull().sum()
```

	WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION	dtype: int64
	0	0	0	0	0	0	

Cambios en tipos de datos

```
#Forzar tipo de datos
df_ventas['TOTAL_UNIT_SALES'] = df_ventas['TOTAL_UNIT_SALES'].astype("float")
df_ventas['TOTAL_VALUE_SALES'] = df_ventas['TOTAL_VALUE_SALES'].astype("float")
df_ventas['TOTAL_UNIT_AVG_WEEKLY_SALES'] = df_ventas['TOTAL_UNIT_AVG_WEEKLY_SALES'].astype("float")
df_ventas['ITEM_CODE'] = df_ventas['ITEM_CODE'].astype("str")

#forzar formato de decimales
df_ventas['TOTAL_UNIT_SALES'] = df_ventas['TOTAL_UNIT_SALES'].round(3)
df_ventas['TOTAL_VALUE_SALES'] = df_ventas['TOTAL_VALUE_SALES'].round(3)
df_ventas['TOTAL_UNIT_AVG_WEEKLY_SALES'] = df_ventas['TOTAL_UNIT_AVG_WEEKLY_SALES'].round(3)

df_ventas.head()
```

	WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION
0	34-22	7501058792808BP2	0.006	0.139	1.000	TOTAL AUTOS AREA 5
1	34-22	7501058715883	0.487	116.519	2.916	TOTAL AUTOS AREA 5
2	34-22	7702626213774	1.391	68.453	5.171	TOTAL AUTOS AREA 5
3	34-22	7501058716422	0.022	1.481	1.833	TOTAL AUTOS AREA 5
4	34-22	7501058784353	2.037	182.839	5.375	TOTAL AUTOS AREA 5

Preparación dataframes para facilitar el análisis

```
# Data frame de ventas que adiciona mes y año al dataframe
# Realizo una copia del dataframe
df_ventas2 = df_ventas

#adiciono las columnas de año y mes
df_ventas2["YEAR"] = df_ventas2['WEEK'].map(df_calendario.set_index('WEEK')[ 'YEAR'])
df_ventas2["MONTH"] = df_ventas2['WEEK'].map(df_calendario.set_index('WEEK')[ 'MONTH'])

#Verifico el nuevo dataframe3
df_ventas2.head(5)
```

	WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION	YEAR	MONTH
0	34-22	7501058792808BP2	0.006	0.139	1.000	TOTAL AUTOS AREA 5	2022	8
1	34-22	7501058715883	0.487	116.519	2.916	TOTAL AUTOS AREA 5	2022	8
2	34-22	7702626213774	1.391	68.453	5.171	TOTAL AUTOS AREA 5	2022	8
3	34-22	7501058716422	0.022	1.481	1.833	TOTAL AUTOS AREA 5	2022	8
4	34-22	7501058784353	2.037	182.839	5.375	TOTAL AUTOS AREA 5	2022	8

Preparación dataframes para facilitar el análisis

```
# Seleccionar las columnas en el orden deseado
df_ventas2 = df_ventas2[['WEEK', 'YEAR', 'MONTH', 'ITEM_CODE',
| | | | | |
| 'TOTAL_UNIT_SALES', 'TOTAL_VALUE_SALES',
| 'TOTAL_UNIT_AVG_WEEKLY_SALES', 'REGION']]
```

[60]

```
# Mostrar las primeras filas para verificar
df_ventas2.head(5)
```

...

	WEEK	YEAR	MONTH	ITEM_CODE	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES	TOTAL_UNIT_AVG_WEEKLY_SALES	REGION
0	34-22	2022	8	7501058792808BP2	0.006	0.139	1.000	TOTAL AUTOS AREA 5
1	34-22	2022	8	7501058715883	0.487	116.519	2.916	TOTAL AUTOS AREA 5
2	34-22	2022	8	7702626213774	1.391	68.453	5.171	TOTAL AUTOS AREA 5
3	34-22	2022	8	7501058716422	0.022	1.481	1.833	TOTAL AUTOS AREA 5
4	34-22	2022	8	7501058784353	2.037	182.839	5.375	TOTAL AUTOS AREA 5

Preparación dataframes para facilitar el análisis

```
> <pre># Finalmente se unen los dataframes de Categoria y Segmentos, ya que no hay claves para hacer un analisis de ventas basaso en
# categorias o segmentos. Aun asi en la practica no hay diferencia porque toda la data es de una sola categoria por tanto no tiene sentido
# añadir ese campo en segmentos.

#Hago copia del dataframe
df_segmentos2 = df_segmentos

#renombro la columna de category para mayor entendimiento
df_segmentos2 = df_segmentos2.rename(columns={'CATEGORY':'ID_CATEGORY'})

#paso el nombre de la categoria al dataset de segmentos
df_segmentos2["CATEGORY"] = df_segmentos2['ID_CATEGORY'].map(df_categorias.set_index('ID_CATEGORY')['CATEGORY'])

#compruebo que todo se haya ejecutado bien
df_segmentos2.head()

[64]</pre>
...
<table border="1">
|  | ID_CATEGORY | ATTR1 | ATTR2 | ATTR3 | FORMAT | SEGMENT | CATEGORY |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | CLORO | CLORO | BAMBINO | LIQUIDO | BLEACH | FABRIC TREATMENT and SANIT\n |
| 1 | 1 | CLORO | CLORO | GERMICIDA | LIQUIDO | BLEACH | FABRIC TREATMENT and SANIT\n |
| 2 | 1 | CLORO | CLORO | MASCOTAS | LIQUIDO | BLEACH | FABRIC TREATMENT and SANIT\n |
| 3 | 1 | CLORO | CLORO | MULTIUSOS | GEL | BLEACH | FABRIC TREATMENT and SANIT\n |
| 4 | 1 | CLORO | CLORO | MULTIUSOS | LIQUIDO | BLEACH | FABRIC TREATMENT and SANIT\n |

</table>
```

Preparación dataframes para facilitar el análisis

```
# Creacion de los archivos con los dos nuevos dataframes  
  
df_ventas2.to_csv('ventas2.csv', index=False)  
df_segmentos2.to_csv('segmentos2.csv', index=False)
```

Análisis Exploratorio de Datos (EDA) y Visualizaciones

Capturas de visualizaciones



Visualizaciones de Datos

Las visualizaciones de datos son herramientas clave que permiten entender patrones y tendencias en los datos analizados.

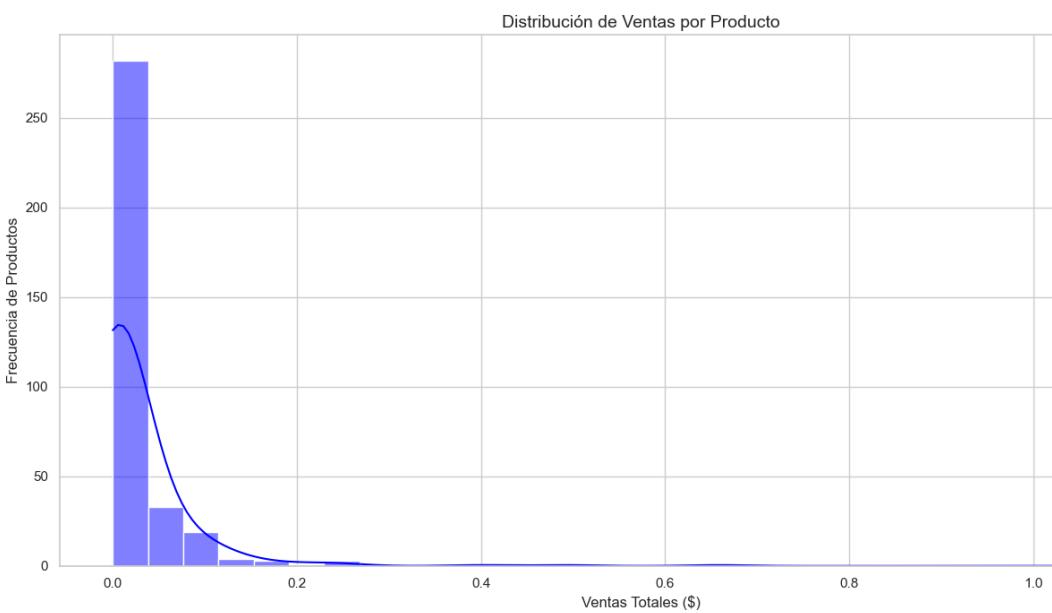
Comunicación de Hallazgos

Las visualizaciones ayudan a comunicar de manera efectiva los hallazgos de la exploración de datos a diferentes audiencias.

Análisis Exploratorio de Datos (EDA)

El EDA es un enfoque para resumir las principales características de los datos, a menudo utilizando visualizaciones gráficas.

Histograma ventas por producto



Partiendo de la información anterior donde sabemos que estamos trabajando con 350 productos, vemos como más de 300 producto están en el primer rango entre 0 y 0,04, y casi toda la totalidad de productos en el rango de 0,12, lo que concentra de gran manera un ticket promedio muy similar y vuelve todos los productos bastante competitivos a nivel precio, entre ellos mismos.

```
# HISTOGRAMA DE VENTAS POR PRODUCTOS

# Agrupar las ventas totales por producto
ventas_por_producto = df_ventas.groupby('ITEM_CODE')[['TOTAL_VALUE_SALES']].sum().reset_index()

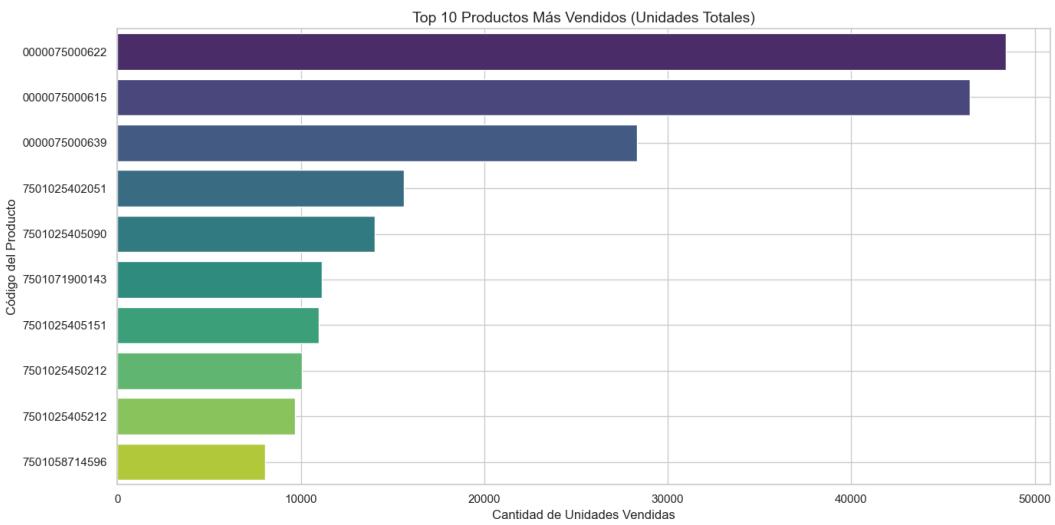
# Configurar el estilo de seaborn
sns.set(style="whitegrid")

# Crear el histograma
plt.figure(figsize=(14, 7))
sns.histplot(data=ventas_por_producto, x='TOTAL_VALUE_SALES', bins=30, kde=True, color='blue')

# Personalizar la gráfica
plt.title('Distribución de Ventas por Producto', fontsize=14)
plt.xlabel('Ventas Totales ($)', fontsize=12)
plt.ylabel('Frecuencia de Productos', fontsize=12)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Distribución de Ventas por Producto.jpg', format='jpg', dpi=300)
plt.show()
```

Top 10 productos más vendidos



Este TOP 10 de producto nos da una idea, de que a pesar de que vimos que todos los productos pueden competir fácilmente dentro de un rango de precio similar, hay productos super ganadores que casi quintuplican las ventas de los puestos 8 al 10 de los 350 productos, Esto quiere decir que probablemente la economía de la empresa se está basando en muy pocos productos que le dan éxito y quizás hay una gran cantidad de productos que no se venden y podrían estar generando costos de producción. Aunque de igual manera depender de un par de productos para el éxito de la compañía podría ser demasiado peligroso, por lo que hacer un análisis incluyendo el tipo de producto, subcategorías y segmentos de compradores podría ser muy valioso para crear más productos ganadores.

```
# AHORA IDENTIFIQUES EL TOP 10 DE LOS PRODUCTOS MAS VENDIDOS

# Agrupar las unidades vendidas totales por producto
ventas_por_producto = df_ventas.groupby('ITEM_CODE')[['TOTAL_UNIT_SALES']].sum().reset_index()

# Ordenar los productos por cantidad de unidades vendidas en orden descendente
ventas_por_producto = ventas_por_producto.sort_values(by='TOTAL_UNIT_SALES', ascending=False)

# Seleccionar el Top 10 de productos
top_10_productos = ventas_por_producto.head(10)

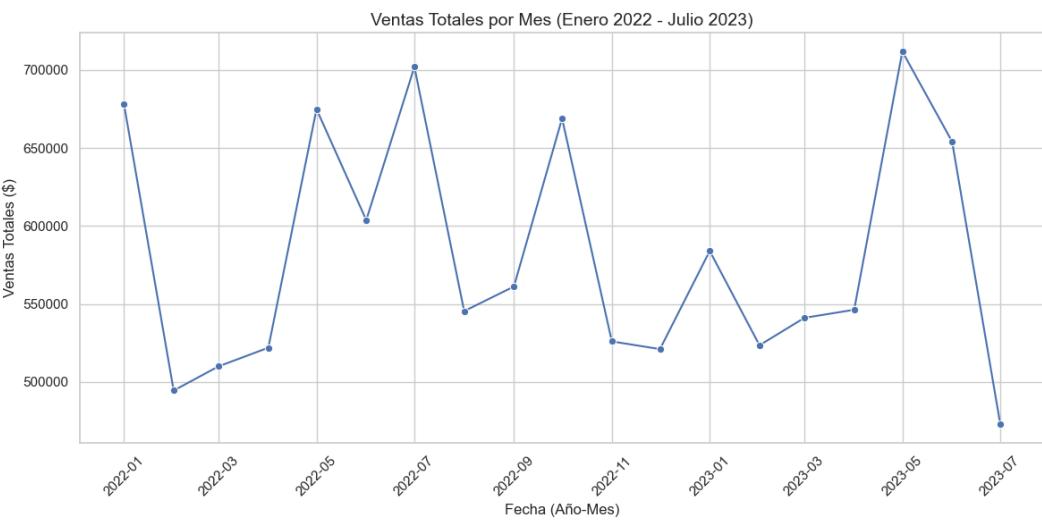
# Configurar el estilo de seaborn
sns.set(style="whitegrid")

# Crear el gráfico de barras
plt.figure(figsize=(14, 7))
sns.barplot(data=top_10_productos, x='TOTAL_UNIT_SALES', y='ITEM_CODE', palette='viridis')

# Personalizar la gráfica
plt.title('Top 10 Productos Más Vendidos (Unidades Totales)', fontsize=14)
plt.xlabel('Cantidad de Unidades Vendidas', fontsize=12)
plt.ylabel('Código del Producto', fontsize=12)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Top 10 Productos Más Vendidos (Unidades Totales).jpg', format='jpg', dpi=300)
plt.show()
```

Ventas por mes



Este grafico de ventas a través del tiempo nos indica varias cosas importantes.

1. Hay una gran variabilidad de ventas mes a mes, lo que es importante para los presupuestos de la empresa, porque seguramente la caja se mueve con rangos muy altos de un mes a otro.
2. Cuando observamos meses iguales de un año a otro, aunque las ventas no son iguales, sí son similares los comportamientos, lo que con mas data a través del tiempo nos podría indicar una constancia en la variabilidad de ventas de un mes a otro con mayor exactitud.
3. Mayo y Julio podrían considerarse meses de mucha importancia para las ventas de la empresa.

```
# TOTAL VENTAS POR MES
# Crear una columna combinada de Año-Mes para un mejor análisis
df_ventas['YEAR_MONTH'] = pd.to_datetime(df_ventas[['YEAR', 'MONTH']].assign(DAY=1))

# Agrupar las ventas totales por mes
ventas_mensuales = df_ventas.groupby('YEAR_MONTH')[['TOTAL_VALUE_SALES']].sum().reset_index()

# Configurar estilo de seaborn
sns.set(style="whitegrid")

# Crear la gráfica
plt.figure(figsize=(12, 6))
sns.lineplot(data=ventas_mensuales, x='YEAR_MONTH', y='TOTAL_VALUE_SALES', marker='o', color='b')

# Personalizar la gráfica
plt.title('Ventas Totales por Mes (Enero 2022 - Julio 2023)', fontsize=14)
plt.xlabel('Fecha (Año-Mes)', fontsize=12)
plt.ylabel('Ventas Totales ($)', fontsize=12)
plt.xticks(rotation=45)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Ventas Totales por Mes (Enero 2022 - Julio 2023).jpg', format='jpg', dpi=300)
plt.show()
```

Unidades vendidas por mes



Esta gráfica de unidades vendidas a través del tiempo confirma varios temas que ya se han venido hablando. Al tener un ticket tan similar entre productos, no hace mucho la diferencia cuando hablamos en cantidades, ya que todos son muy parecidos en precio. Es por esta razón que la gráfica casi que se puede super poner con la anterior porque los números en proporción son muy parecidos entre las unidades vendidas y las ventas totales por mes.

```
# TOTAL UNIDADES VENDADAS POR MES POR MES

# Agrupar las ventas totales por mes
unidades_mensuales = df_ventas.groupby('YEAR_MONTH')[ 'TOTAL_UNIT_SALES'].sum().reset_index()

# Configurar estilo de seaborn
sns.set(style="whitegrid")

# Crear la gráfica
plt.figure(figsize=(12, 6))
sns.lineplot(data=unidades_mensuales, x='YEAR_MONTH', y='TOTAL_UNIT_SALES', marker='o', color='b')

# Personalizar la gráfica
plt.title('Unidades Vendidas por Mes (Enero 2022 - Julio 2023)', fontsize=14)
plt.xlabel('Fecha (Año-Mes)', fontsize=12)
plt.ylabel('Unidades Vendidas', fontsize=12)
plt.xticks(rotation=45)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Unidades Vendidas por Mes (Enero 2022 - Julio 2023).jpg', format='jpg', dpi=300)
plt.show()
```

Ventas mensuales por región

```
import matplotlib.pyplot as plt
import seaborn as sns

# Agrupar las ventas totales por región y mes
ventas_por_region_tiempo = df_ventas.groupby(['YEAR_MONTH', 'REGION'])[['TOTAL_VALUE_SALES']].sum().reset_index()

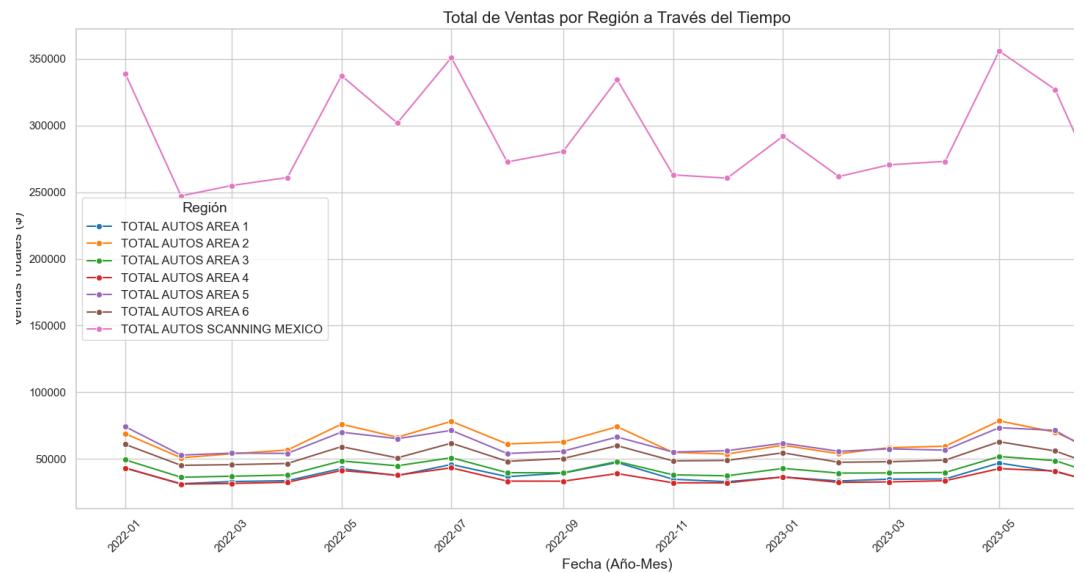
# Configurar estilo de seaborn
sns.set(style="whitegrid")

# Crear la gráfica de líneas
plt.figure(figsize=(16, 8))
sns.lineplot(data=ventas_por_region_tiempo, x='YEAR_MONTH', y='TOTAL_VALUE_SALES', hue='REGION', marker='o', palette='tab10')

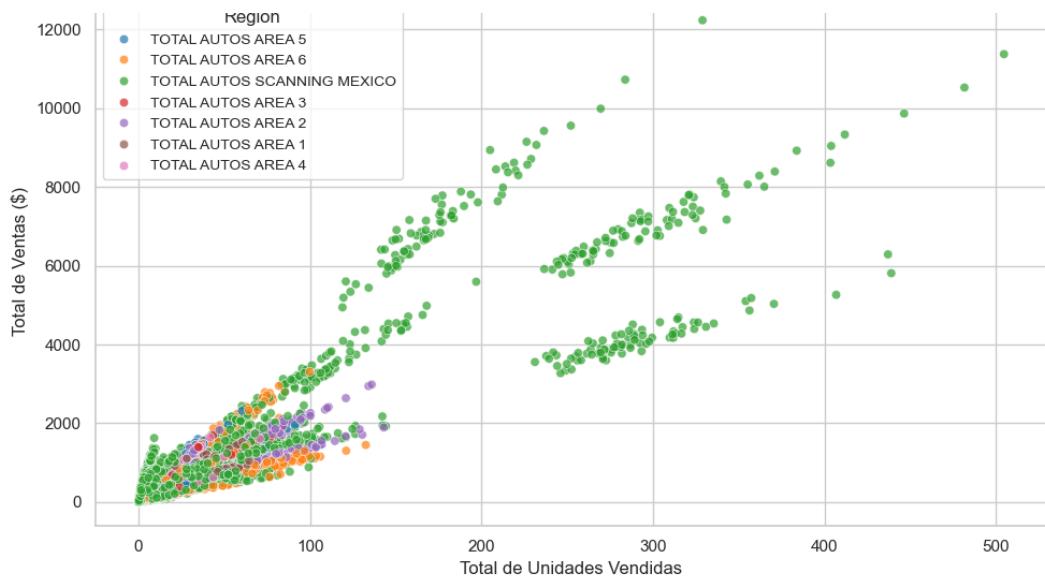
# Personalizar la gráfica
plt.title('Total de Ventas por Región a Través del Tiempo', fontsize=16)
plt.xlabel('Fecha (Año-Mes)', fontsize=14)
plt.ylabel('Ventas Totales ($)', fontsize=14)
plt.xticks(rotation=45)
plt.legend(title='Región', fontsize=12, title_fontsize=14)
plt.grid(visible=True)

# Ajustar el diseño y mostrar la gráfica
plt.tight_layout()
plt.savefig('Total de Ventas por Región a Través del Tiempo.jpg', format='jpg', dpi=300)
plt.show()
```

Las ventas por región, vemos que están distribuidas en 7 regiones, 6 de ellas con un numeral por área y la última llamada scanning México. No es claro si Scanning México sea la totalización de las demás áreas (que es probable) y las demás áreas estén distribuidas a lo largo de todo México. Lo que es cierto es que el comportamiento por área es muy similar, no hay super valores que cambien la tendencia de una manera transcendental. Por otro lado, la gráfica de Scanning México puede ser exponencial porque si es un área importante por ejemplo todo el Estado de México que casi multiplica por 6 las demás áreas. Por eso puede ser un valor total, pero esta información no fue suministrada.



Correlación entre ventas y unidades por región



Estas gráficas de scatterplot son excelentes para ver gráficamente correlaciones, y en definitiva las unidades vendidas van fuertemente atadas a las ventas totales, y esto se debe mucho a que el ticket promedio entre todas las regiones es muy similar de manera positiva como lo indica el desarrollo de la gráfica.

```
sns.set(style="whitegrid")

# Crear el gráfico de dispersión
plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=df_ventas,
    x='TOTAL_UNIT_SALES',
    y='TOTAL_VALUE_SALES',
    hue='REGION',
    palette='tab10',
    alpha=0.7
)

# Personalizar la gráfica
plt.title('Relación entre Unidades Vendidas y Ventas Totales por Región', fontsize=14)
plt.xlabel('Total de Unidades Vendidas', fontsize=12)
plt.ylabel('Total de Ventas ($)', fontsize=12)
plt.legend(title='Región', fontsize=10, title_fontsize=12)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Relación entre Unidades Vendidas y Ventas Totales por Región.jpg', format='jpg', dpi=300)
plt.show()
```

Correlación entre ventas promedio y unidades vendidas

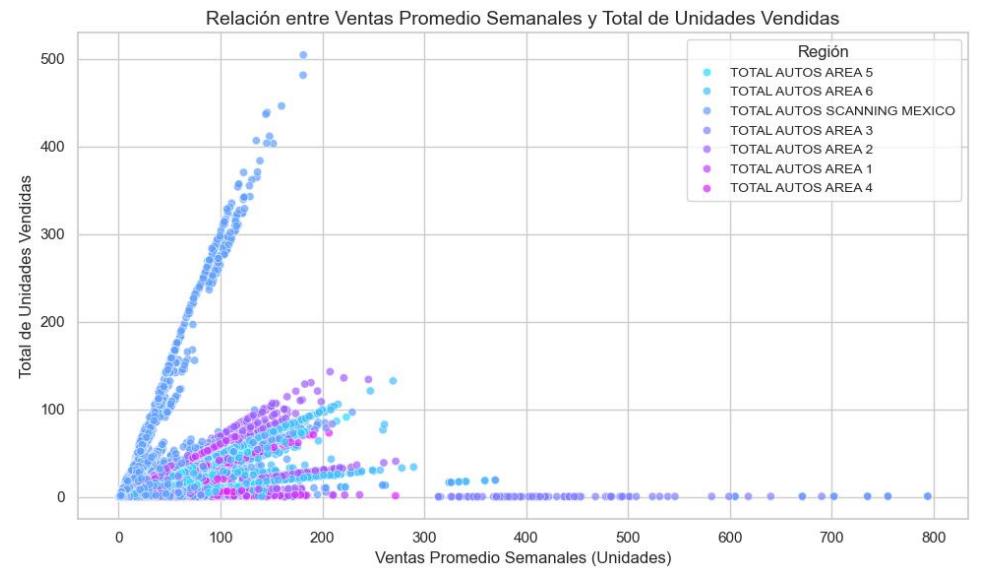
```
# RELACION ENTRE TOTAL_UNIT_AVG_WEEKLY_SALES Y TOTAL_UNIT_SALES POR REGION

# Crear el gráfico de dispersión
plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=df_ventas,
    x='TOTAL_UNIT_AVG_WEEKLY_SALES',
    y='TOTAL_UNIT_SALES',
    hue='REGION',
    palette='cool',
    alpha=0.7
)

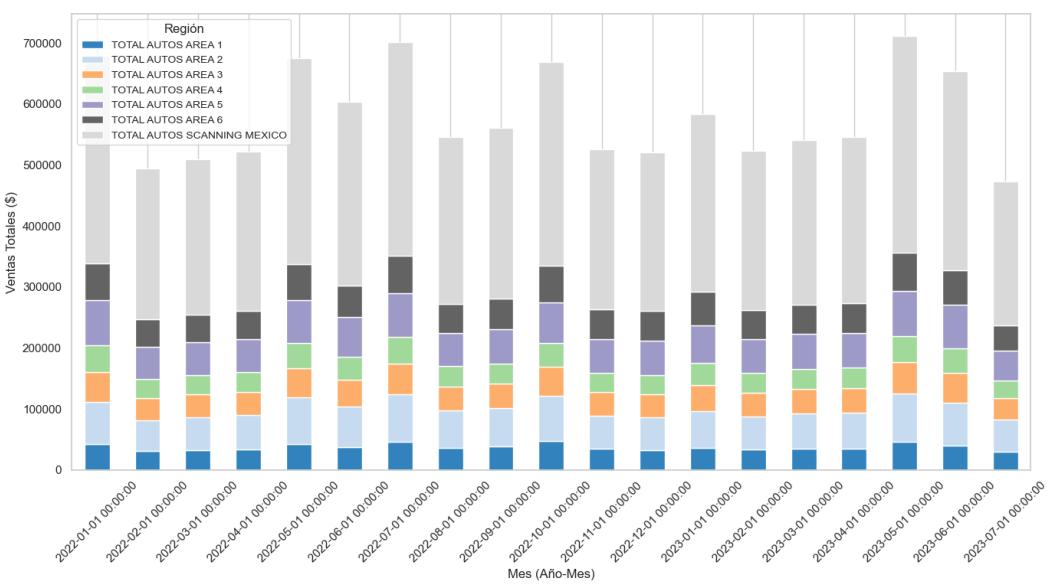
# Personalizar la gráfica
plt.title('Relación entre Ventas Promedio Semanales y Total de Unidades Vendidas', fontsize=14)
plt.xlabel('Ventas Promedio Semanales (Unidades)', fontsize=12)
plt.ylabel('Total de Unidades Vendidas', fontsize=12)
plt.legend(title='Región', fontsize=10, title_fontsize=12)
plt.grid(visible=True)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Relación entre Ventas Promedio Semanales y Total de Unidades Vendidas.jpg', format='jpg', dpi=300)
plt.show()
```

Esta correlación entre ventas promedio semanales y total de unidades vendidas al final debe de ser coincidente porque comparten el comportamiento, pero podemos ver la diferencia que hay entre las áreas y como la correlación puede ser mas positiva que otra.



Visualización geográfica de ventas. (Regiones)



En esta gráfica podemos ver mucho más fácil una combinación muy especial entre uno de los primeros gráficos de línea donde se ve el comportamiento de ventas mes a mes, pero esta vez con una distribución por región, que a la vez nos va confirmando que el Scanning México es la suma de todas las áreas, ya que es el doble de cada una de las medidas.

Visualmente casi podemos asegurar que al área con mayor variabilidad de precios es el área 2 y que puede estar compitiendo con el parea 5 con las mayores ventas por región.

```
# GRAFICO DE BARRAS APLICADAS APRA VENTAS POR REGION Y MES
# Agrupar las ventas totales por región y mes
ventas_por_region_mes = df_ventas.groupby(['YEAR_MONTH', 'REGION'])['TOTAL_VALUE_SALES'].sum()

# Crear el gráfico de barras apiladas
ventas_por_region_mes.plot(kind='bar', stacked=True, figsize=(14, 8), colormap='tab20c')

# Personalizar la gráfica
plt.title('Ventas Totales por Región y Mes', fontsize=16)
plt.xlabel('Mes (Año-Mes)', fontsize=12)
plt.ylabel('Ventas Totales ($)', fontsize=12)
plt.legend(title='Región', fontsize=10, title_fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis='y')

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Ventas Totales por Región y Mes.jpg', format='jpg', dpi=300)
plt.show()
```

Ventas por región

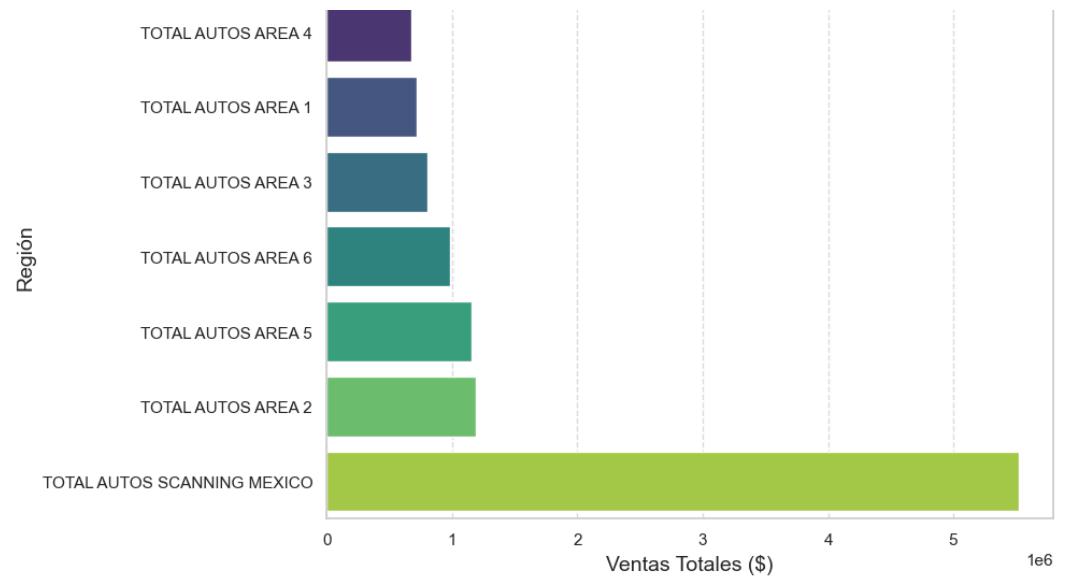
```
ventas_por_region = df_ventas.groupby('REGION')['TOTAL_VALUE_SALES'].sum().sort_values()

# Crear el gráfico de barras horizontales
plt.figure(figsize=(10, 6))
sns.barplot(
    x=ventas_por_region.values,
    y=ventas_por_region.index,
    palette='viridis'
)

# Personalizar la gráfica
plt.title('Ventas Totales por Región', fontsize=16)
plt.xlabel('Ventas Totales ($)', fontsize=14)
plt.ylabel('Región', fontsize=14)
plt.grid(axis='x', linestyle='--', alpha=0.7)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Ventas Totales por Región.jpg', format='jpg', dpi=300)
plt.show()
```

Para confirmar de manera mas clara, el tema de mejores ventas por región, esta gráfica ya nos da una mejor vista, y en efecto el área 2 compite muy de cerca con el área 5 con las mejores ventas en todo México y las dos áreas que menos venden con la 4 y la 1.





Identificación de outliers

Importancia de los outliers

Los outliers pueden indicar problemas o identificar oportunidades significativas en el análisis de datos. Investigar estos valores es crucial.

Investigación de outliers

La investigación de los valores atípicos es esencial para entender su origen y su impacto en el conjunto de datos.

Distribución Top 20 de productos

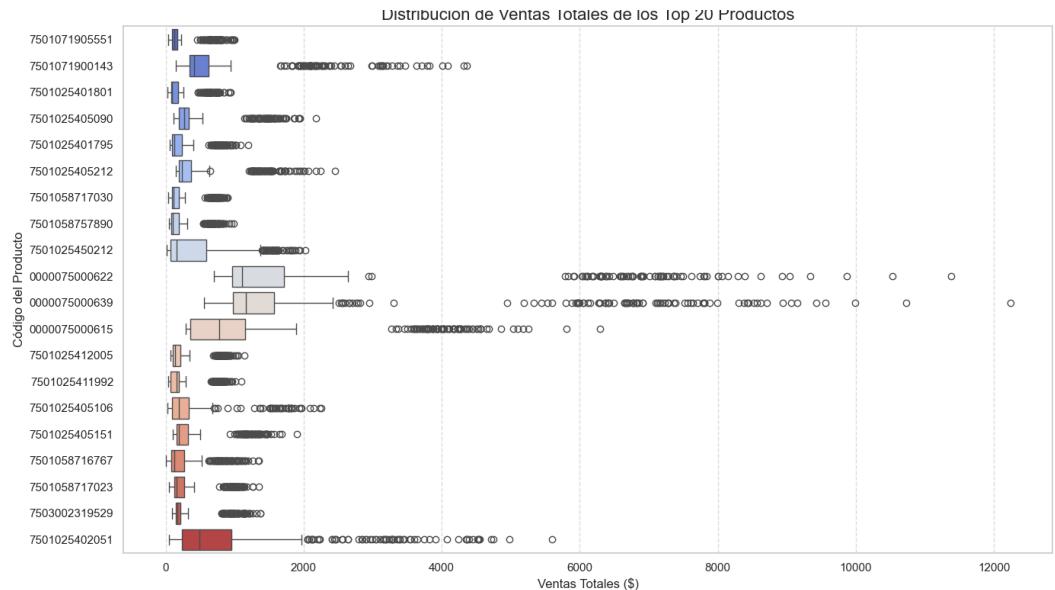
```
# Agrupar las ventas por producto y ordenar por ventas totales
top_productos = df_ventas.groupby('ITEM_CODE')['TOTAL_VALUE_SALES'].sum().sort_values(ascending=False).head(20)

# Filtrar el DataFrame para incluir solo los productos top
df_top_productos = df_ventas[df_ventas['ITEM_CODE'].isin(top_productos.index)]

# Crear el gráfico de caja para los top 20 productos
plt.figure(figsize=(14, 8))
sns.boxplot(
    data=df_top_productos,
    x='TOTAL_VALUE_SALES',
    y='ITEM_CODE',
    palette='coolwarm',
    showfliers=True
)

# Personalizar la gráfica
plt.title('Distribución de Ventas Totales de los Top 20 Productos', fontsize=16)
plt.xlabel('Ventas Totales ($)', fontsize=12)
plt.ylabel('Código del Producto', fontsize=12)
plt.grid(visible=True, axis='x', linestyle='--', alpha=0.7)

# Mostrar la gráfica
plt.tight_layout()
```



Realizar una distribución con 350 productos es realmente bastante difícil, pero cuando hacemos los ranking podemos ver comportamientos bien interesantes. Al inicio vimos que los rangos de precios son bastante similares, y también vimos que solo unos pocos productos son los que mas se venden y realmente generan los números en la compañía.

Viendo de cerca el comportamiento de los 20 productos más vendidos respecto al total de ventas, que tiene un comportamiento típico en general desde 0 a 1000, y unos pocos se suben hasta 2400 con su venta máxima, pero al final del día se generan muchos valores atípicos conocidos como outliers que suben las ventas en gran magnitud. Especialmente estos 3 productos en el medio con valores muy distantes del máximo.

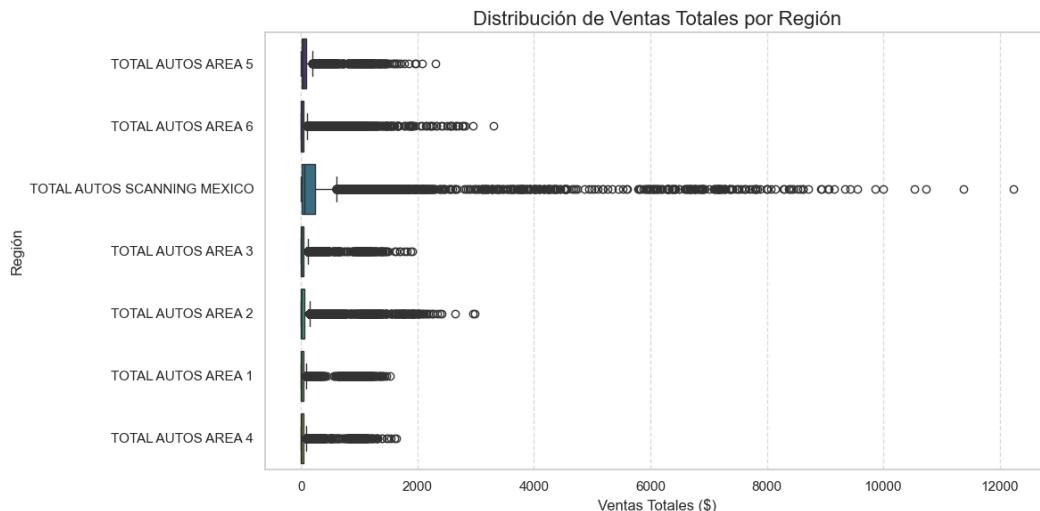
Distribución de ventas totales por región

```
# VENTAS POR REGION

# Crear el gráfico de caja para identificar outliers por región
plt.figure(figsize=(12, 6))
sns.boxplot(
    data=df_ventas,
    x='TOTAL_VALUE_SALES',
    y='REGION',
    palette='viridis',
    showfliers=True
)

# Personalizar la gráfica
plt.title('Distribución de Ventas Totales por Región', fontsize=16)
plt.xlabel('Ventas Totales ($)', fontsize=12)
plt.ylabel('Región', fontsize=12)
plt.grid(visible=True, axis='x', linestyle='--', alpha=0.7)

# Mostrar la gráfica
plt.tight_layout()
plt.savefig('Distribución de Ventas Totales por Región.jpg', format='jpg', dpi=300)
plt.show()
```



Esta gráfica de distribución por Región, tiene un comportamiento similar que podemos leer de la misma forma que el gráfico anterior.

No obstante, obviamente el Scanning México que ahora aceptamos con las ventas totales, vemos ya que entre las Areas, el cuerpo de la gráfica tiene un rango, mínimo, máximo y media muy similares, al igual que los outliers, aunque estos ya tienen un poco más de diferencia.

Creación de un Dashboard en Looker Studio

Visualizaciones interactivas



Facilidad de Exploración de Datos

Las visualizaciones interactivas permiten a los usuarios explorar datos de manera intuitiva, facilitando la navegación a través de diferentes métricas.

Comprendión de Métricas Clave

Con las visualizaciones interactivas, los usuarios pueden comprender mejor las métricas clave y su impacto en el análisis de datos.

Resultados del Análisis

Las visualizaciones ayudan a presentar los resultados del análisis de datos de manera clara, permitiendo una toma de decisiones informada.

Portada del informe



INFORME DE VENTAS 2021 - 2022

OBJETIVO DEL INFORME

Con este informe se pretende tener un mejor conocimiento sobre el comportamiento de las ventas, y al final tener información de valor para generar estrategias para incrementar las ventas.

Para esto se van a visualizar análisis de la siguiente manera:

Ventas en diferentes líneas de tiempo
Análisis de venta de producto
Segmentación de ventas por productos, atributos y geolocalización

Con estos datos se espera identificar productos por segmento de mercado, productos con mayor rendimiento, y productos con mayor oportunidad de crecimiento, así como aquellos productos que no generan mayor relevancia en el mercado.

Finalmente se presentarán algunos insights que ayuden a la toma de decisiones estratégicas para el incremento de ventas, potencializando lo que funciona, mejorando lo que funciona medianamente, y eliminando lo que no funciona.



Ventas por mes-año

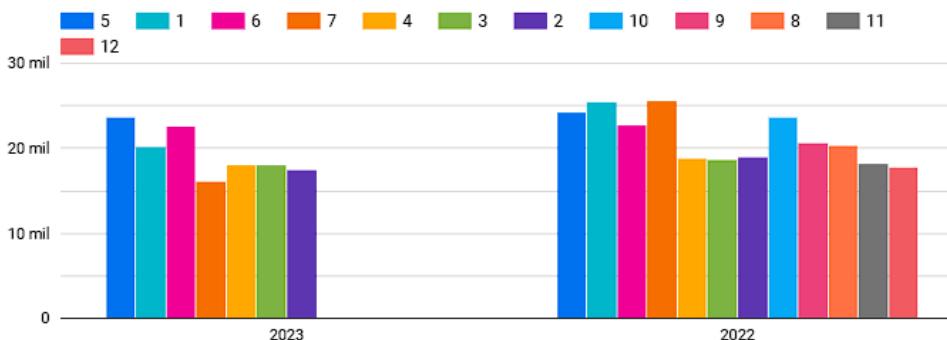
VENTAS POR MES-AÑO

Total Ventas 2021-2022
391.760,23

Promedio Unidades x Semana
3,21

Total ventas 2021-2022
11.042.859,89

Promedio ventas semanal
90,51



MONTH	TOTAL_UNIT_SALES
1. 5	47.905,78
2. 1	45.565,89
3. 6	45.407,43
4. 7	41.754,13
5. 4	36.981,04
6. 3	36.844,49
7. 2	36.469,74
8. 10	23.679,75
9. 9	20.687,91
10. 8	20.337,22

1 - 12 / 12 < >

OBSERVACIONES:

* Vemos que el año 2022 tiene data de los 12 meses, pero 2023 solo hay data hasta el mes de Julio, por lo que los datos por mes no es consistente.
Analizando a nivel de toda la data por meses, y solo filtrando por primer semestre vemos que:
2022: Mejores meses son Enero y Julio / Peores son Marzo y Abril
2023: Mejores meses son Mayo y Junio / Peores son Febrero y Julio

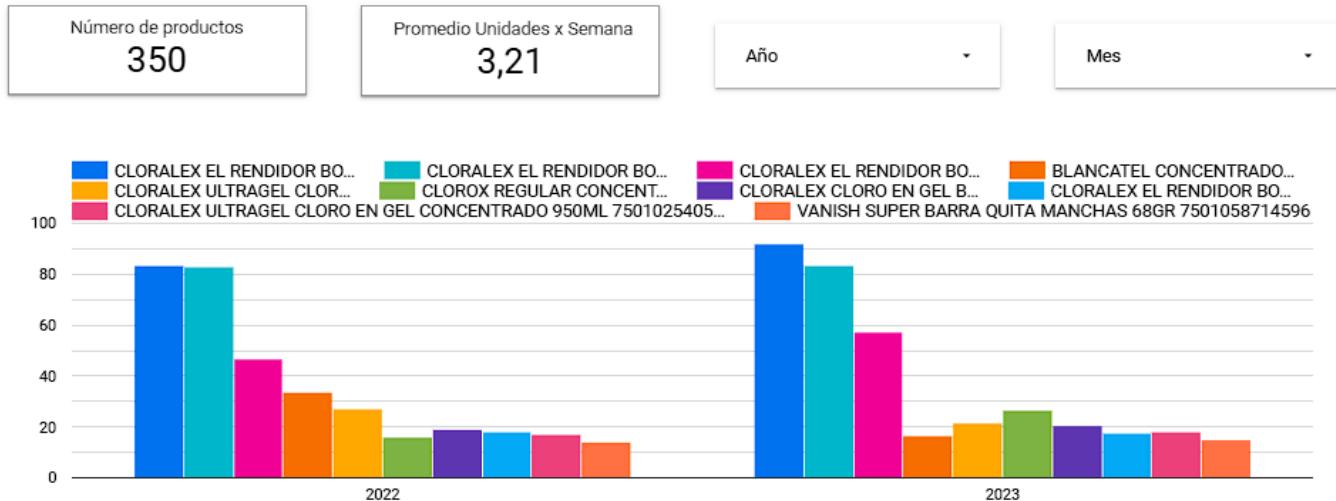
* Evaluando por año con la data existente:
2022 Mejores meses Enero y Julio / Peores Marzo y Diciembre
2023 Mejores meses Mayo y Junio / Peores son Febrero y Julio

* En ventas totales vemos el top 3 con
1. Mayo
2. Enero
3. Junio

En conclusión parecería que el primer semestre es mejor, pero no es tan cierto debido a que no hay data del segundo semestre de 2023.

Ventas por producto

VENTAS POR PRODUCTO



ITEM_DESCRIPTION	TOTAL_UNIT_SA...
1. VANISHOXIACTIONROSAGEDOYPACK500ML+PLANCHYTEXBOTC/ATOM6...	4,65
2. VANISHOXIACTIONROSADOPACK120GR+MMCRYSTALWHITEBOTE900GR...	72,96
3. VANISHOXIACTIONROSABOTE900GR+MMCRYSTALWHITEDOYPACK120GR...	67,46
4. VANISHOXIACTIONGOLDROSABOT900GR+MMCRYSTALWHITEBOT150GR P...	1,87
5. VANISHOXIACTIONGOLDROSABOT150GR+MMCRYSTALWHITEBOT900GR P...	0,15
6. VANISHOXIACTIONGOLDCRYSTALWHITEBOT900GR+MMROSABOT150GR P...	0,15
7. VANISHOXIACTIONGOLDCRYSTALWHITEBOT150GR+MMROSABOT900GR P...	1,87
8. VANISHOXIACTIONCRYSTALWHITEDOYPACK120GR+MMROSABOTE900GR...	67,46
9. VANISHOXIACTIONCRYSTALWHITEBOTE900GR+MMROSADOPACK120GR...	72,96
10. VANISH WHITE REMOVERD D/MANCHAS Y BLANQ.S/CLORO LIQ.BOT4000...	276,9

OBSERVACIONES:

Si observamos los 3 productos mas vendidos en cada año, son los mismo, y de hecho se ve un incremento en ventas en 2023.

Lo interesante en esta vista es que los 3 productos mas vendidos son exactamente el mismo producto, y lo que varía realmente es la presentación por cantidad de mililitros.

Recordando que los datos solo tienen medio 2023 ya podemos visualizar que las ventas son muy buenas con respecto a 2022.



Ventas por región

VENTAS POR AREA

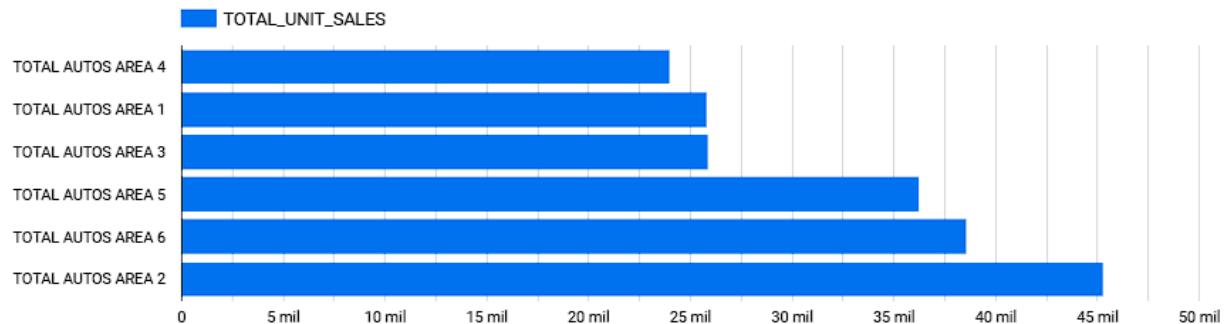


Total de ventas
11.042.859,89

Región

Año

Mes

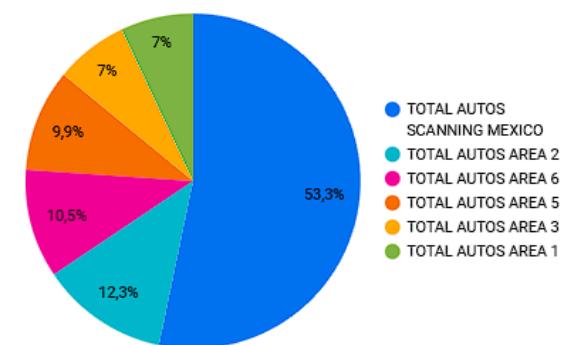


OBSERVACIONES:

Como se puede apreciar el área 2 es la que mejores resultados tiene seguida por la 6 y 5.

Por el otro lado la que menos vende es la 4 seguida de la 1 y la 3.

Es importante determinar el área geográfica, las dimensiones y el volumen demográfico para entender las oportunidades de ventas.





Ventas por marca

VENTAS POR MARCA

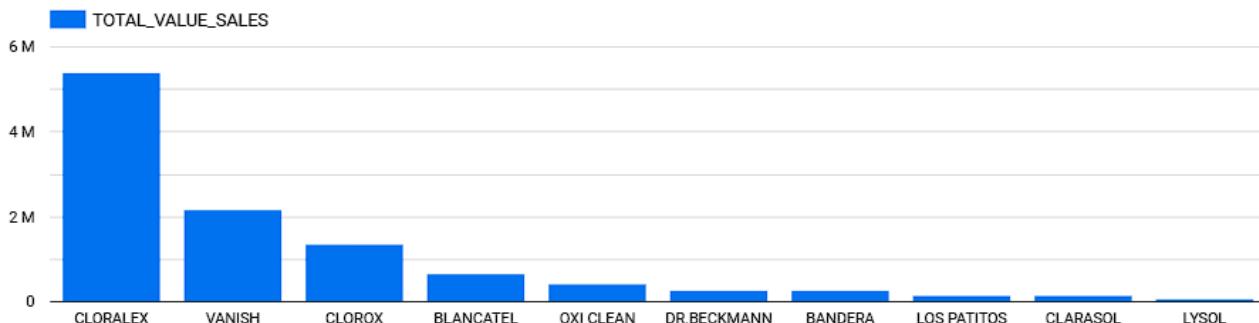


Cantidad de Marcas
24

BRAND

Año

Mes



BRAND	TOTAL_UNIT_SALES
1. CLORALEX	223.300,67
2. VANISH	55.332,94
3. CLOROX	47.881,2
4. BLANCATEL	20.611,3
5. BANDERA	12.123,08
6. CLARASOL	9.821,12
7. LOS PATITOS	6.781,69
8. OXI CLEAN	6.079,37
9. DR.BECKMANN	5.229,32
10. CLORTEX	1.431,94

OBSERVACIONES:

Se confirma la importancia de la marca Cloralex para la compañía que puede vender casi lo mismo que las demás marcas juntas.

Por otro lado las 4 primeras marcas en ventas son congruentes con las posiciones según las unidades vendidas, las siguientes no lo son, es decir que pueden vender mas productos, pero ganar menos dinero, ya que seguramente son productos de muy bajo valor, por tanto no son muy rentables para el negocio.



Análisis bivariable

ANÁLISIS BIVARIABLE



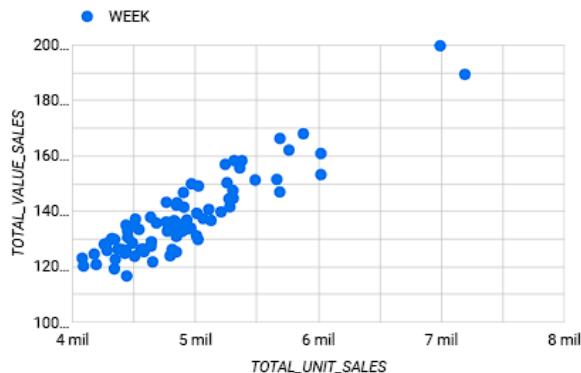
Región ▾

Marca ▾

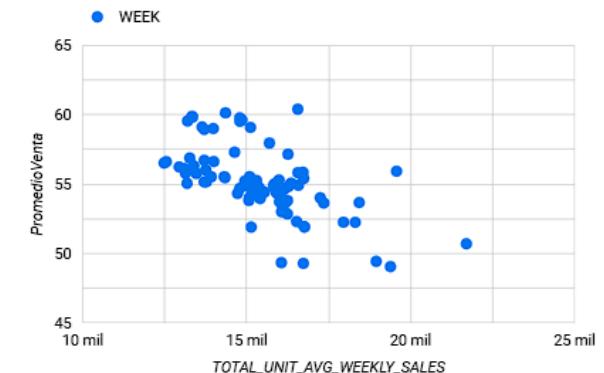
Año ▾

Mes ▾

Distribución Unidades vs Ventas



Distribución Promedio de venta vs promedio unidades ...



OBSERVACIONES:

Vemos un tendencia positiva mucho mas clara en el gráfico de Total de ventas vs Total de unidades vendidas, lo que nos indica algo que puede ser muy obvio y es que a medida que se venden mas unidades, pues igualmente crecen las ventas.

Por otro lado cuando lo miramos a nivel promedio, vemos como no hay una tendencia clara, aunque parece que pudiera ser negativo. Esto puede deberse a lo que se revisó en otros análisis, donde vimos que el ticket de algunos productos es muy bajo, y se necesitan vender muchas unidades, pero el ingreso no es tan alto.

Conclusiones

CONCLUSIONES



CONSIDERACIONES

La data suministrada tiene ventas desde enero de 2022 hasta julio de 2023, por lo que no se puede hacer un comparativo entre dos años de manera completa, y el mes de Julio de 2023 solo tiene datos de 3 semanas por lo que también puede sesgar un poco la información.

Los datos no muestran en qué unidades están ya que se tienen casos de ventas de unidades de 0,006 lo cual pueden ser cientos, miles, o millones. Como no se especificó se proceso la información como estaba, y se debe contemplar hacer la conversión para los datos finales.

Todos los datos suministrados de ventas pertenecen a una sola categoría llamada FABRIC TREATMENT and SANIT por lo que no hay forma de hacer una segmentación por esta variable, pero al interior de la tabla de Productos se encuentra la MARCA, lo cual puede ser una segmentación importante para visualizar el comportamiento de ventas, por lo que se estará usando como un instrumento de medida adicional.

HALLAZGOS

Se reportan 24 marcas vendidas, y solo 3 de ellas hacen el "PARETO" (80/20) en las ventas.

Estas marcas son Cloralex, Vanish, Clorox. En su mismo orden son la gran representación del 80% de todas las ventas, y aún más importante sólo Cloralex representa casi el 50% de las ventas. Este dato es consistente no sólo en ventas totales, sino en cada una de las regiones de manera independiente.

La región Scanning México vende el 50% y lo demás está distribuido en las otras 6 zonas.

Las siguientes regiones con mejor desempeño son las Areas 2,5 y 6 en su respectivo orden aportando así un 30,12% de las ventas.

"Estos son los mejores 5 meses de ventas en su orden:
Mayo de 2023, Julio de 2022, Enero de 2022, Mayo de 2022, Octubre de 2022"

En particular en Mayo de 2022 igualmente se tiene un ticket promedio mas alto y una cantidad de productos mas baja, lo cual indica que se vendieron los productos a mejor precio o hubo venta de productos especialmente que son mas caros.

El número de unidades vendidas en promedio por semana es bastante consistente todos los meses.

Los meses con mejor ticket promedio coinciden con los mejores meses de ventas, con la particularidad que antes se mencionó ya que en Mayo de 2023 además de mejores ventas y ticket promedio, la cantidad de productos no fue tan alta.

No se detectan valores atípicos en la mayoría de KPIs de rendimiento, por lo general el comportamiento de ventas es muy estándar, y no hay mucha desviación en ventas, cantidad de productos y ticket promedio. Además de las tablas las gráficas de Boxplot nos permiten confirmar esto claramente.

La línea de tendencia de ventas durante los 18 meses va al alza levemente, lo que indica que las ventas en general van subiendo, pero de una manera muy lenta.
El gráfico de correlación entre unidades vendidas y ventas es uno de los hallazgos mas importantes en este análisis.

Hay una gran correlación entre productos con menos ventas y menos cantidad vendida, pero en la medida que sube el precio y la cantidad de productos vendidos, logramos visualizar 3 líneas de tendencia bien marcadas.

El coeficiente de correlación es bastante alto y R2 nos da un 83,95% de confiabilidad en esta tendencia.

Recomendaciones

RECOMENDACIONES



"El performance de las 3 principales marcas es estupendamente bueno y representa el 80% de las ventas, pero a la vez es peligroso porque algún problema con una de ellas puede ser un desastre, por lo cual es importante en un momento donde todo anda bien consolidado, poder hacer un esfuerzo con marcas que pueden hacer un aporte importante a las ventas, y con esto tratar de lograr una curva más acelerada de crecimiento. Solo estas 5 marcas hoy aportan el 16,14%, es decir casi el otro porcentaje restante del total, por lo hacerlas crecer puede resultar muy ventajoso para la compañía.

BLANCATEL
OXI CLEAN
DR. BECKMANN
BANDERA
LOS PATITOS

Son varias las razones por las que una región puede ser buena en ventas o no, y puede ser la disponibilidad de producto, la situación financiera a nivel geográfico, el posicionamiento local, la segmentación, el branding, etc. Se tienen 3 regiones que no son representativas en las ventas y lo que se puede hacer es un estudio que determine las variables para hacerla más rentable, o no invertirle más.

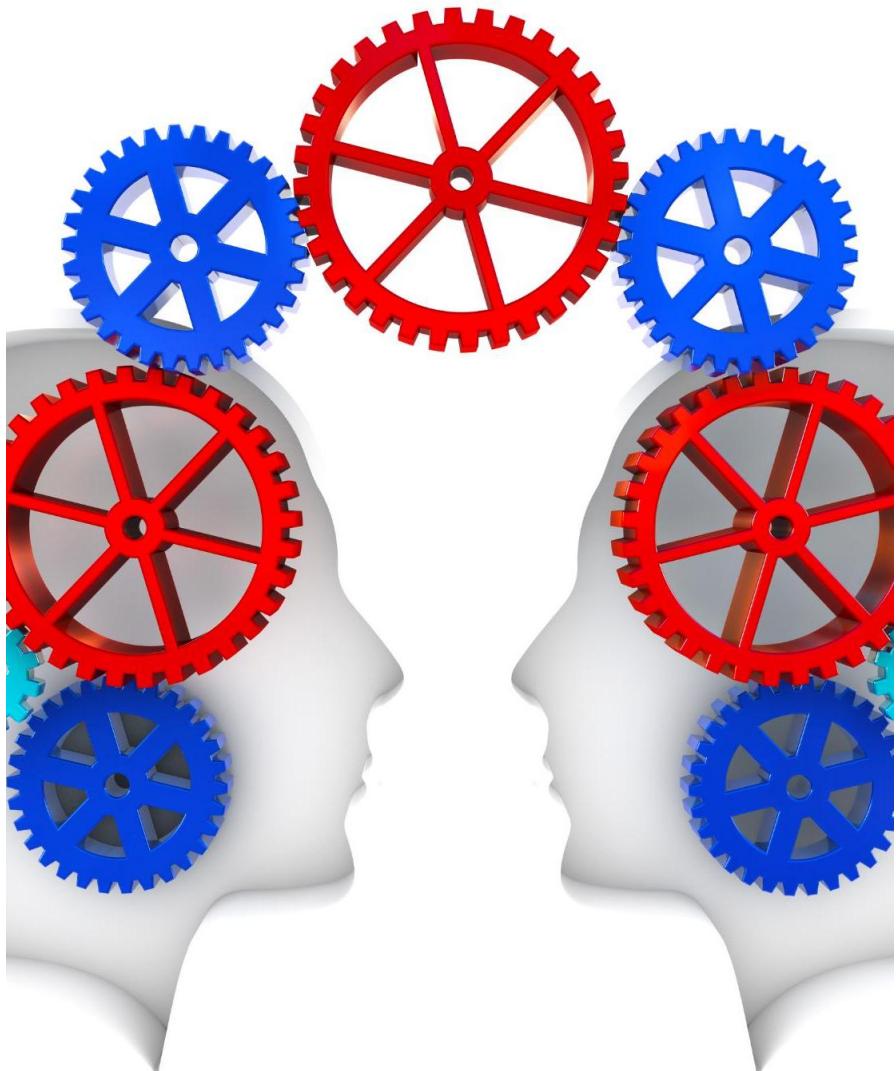
Por otro lado están las 3 regiones (Áreas 2, 5 y 6) que conforman ayudan a scanning México que representa el 50% de las ventas a llegar al 80% (Pareto) de ventas. Junto con la recomendación anterior se puede hacer un mix para incentivar en especial estas 4 áreas que están mejor posicionadas para hacer una mejor inserción con las 5 marcas propuestas anteriormente.
Se logró determinar un comportamiento atípico, pero importante en el mes de mayo de 2023 donde quizás pudo ser un mes muy rentable. Fueron las mejores ventas en 18 meses, pero además vendiendo menos cantidad de productos y por ende con un ticket de venta mucho mayor. Vale la pena investigar de cerca este caso especial, ya que puede que una estrategia de venta sin promoción haya dado buen resultado, o bien se hayan vendido productos que no suelen venderse que son de más alto valor. Como sea, identificar cuál fue la causa real de este suceso puede ayudar a generar estrategias para repetir esta situación de manera exponencial, generando así más ventas con menos esfuerzo.

Descubrir que los KPIs de rendimientos son muy estables son buenos para entender que se tiene una compañía sólida, pero también denota en que no hay un crecimiento importante en ventas, y por esto se hace necesario revisar las estrategias si este es el objetivo de la compañía, ya que hemos podido detectar varias oportunidades en cuanto a producto y a regiones de venta.

Como se mencionaba en los hallazgos la correlación entre el número de unidades vendidas y las ventas es estupendamente alto, y es congruente con el comportamiento de ventas, pero lo que se nota en la gráfica nos da idea de 3 nichos que se pueden explorar de manera independiente y poder sacar el mejor partido.

Esto quiere decir que se forman algo así como 3 líneas de correlación altamente probable, por lo que al identificar cuales son esos productos, sus nichos, su segmentación, su rango de precio; se pueden generar estrategias comerciales que atiendan esas condiciones y se puede subir paralelamente el número de unidades aumentando los ingresos de manera más exponencial. Esta es una estrategia que se debe de investigar y unir a las ideas que se puedan realizar con los productos y regiones a desarrollar.

Predicción de Ventas con Machine Learning



Selección del modelo

Importancia de la selección del modelo

La selección del modelo es esencial para el éxito de cualquier proyecto de Machine Learning. Un modelo adecuado puede mejorar la precisión de las predicciones.

Evaluación de algoritmos

Evaluamos diferentes algoritmos en función de su rendimiento, precisión y capacidad de adaptación a nuestros datos específicos.

Objetivos de predicción

Definir claramente nuestros objetivos de predicción nos ayudó a seleccionar el algoritmo más adecuado para alcanzar esos objetivos.

Filtrado de datos de acuerdo al objetivo de entrenamiento

```
df_filtered['DATE'] = pd.to_datetime(df_filtered['YEAR'].astype(str) + '-' + df_filtered['MONTH'].astype(str) + '-1')

df_grouped = df_filtered.groupby(['DATE', 'BRAND'])[['TOTAL_UNIT_SALES', 'TOTAL_VALUE_SALES']].sum().reset_index()

df_grouped.head()
```

```
C:\Users\paob2\AppData\Local\Temp\ipykernel_25408\4180426011.py:3: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

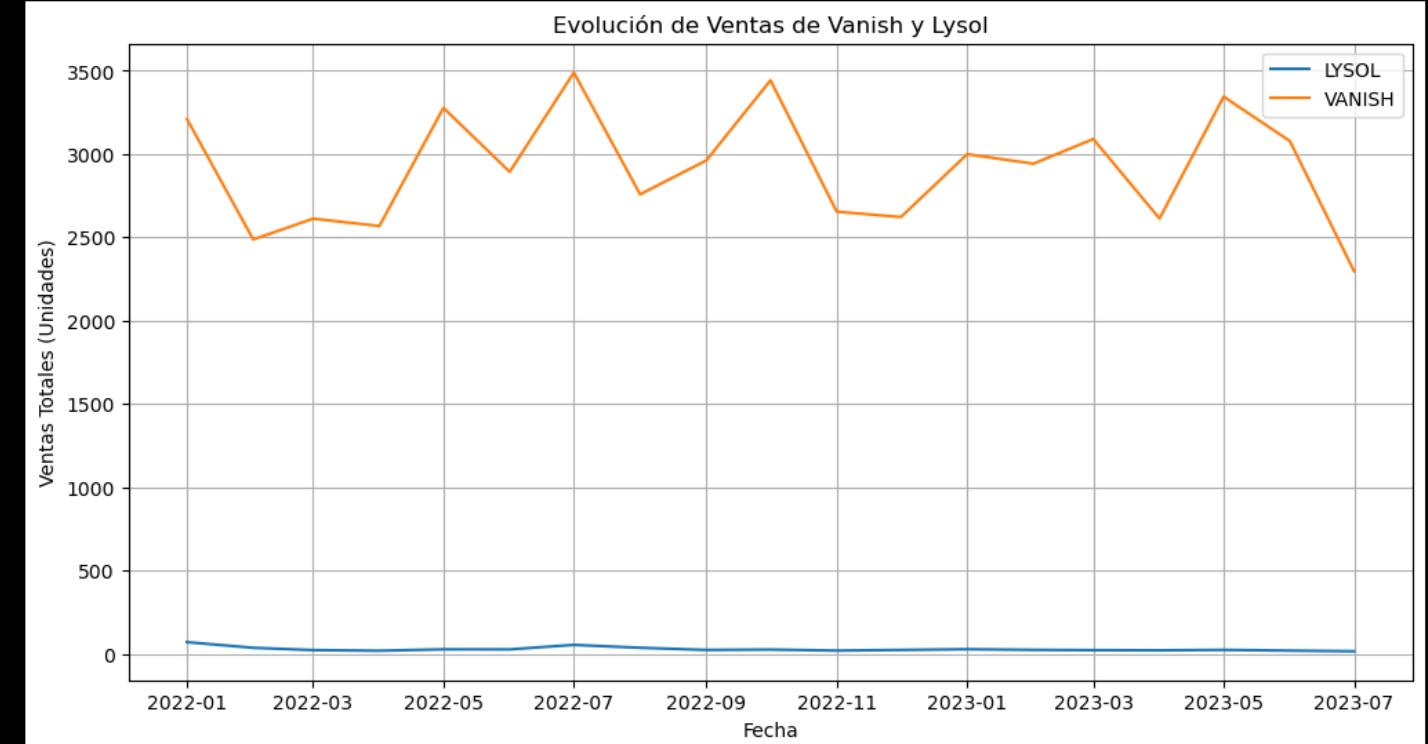
```
df_filtered['DATE'] = pd.to_datetime(df_filtered['YEAR'].astype(str) + '-' + df_filtered['MONTH'].astype(str) + '-1')
```

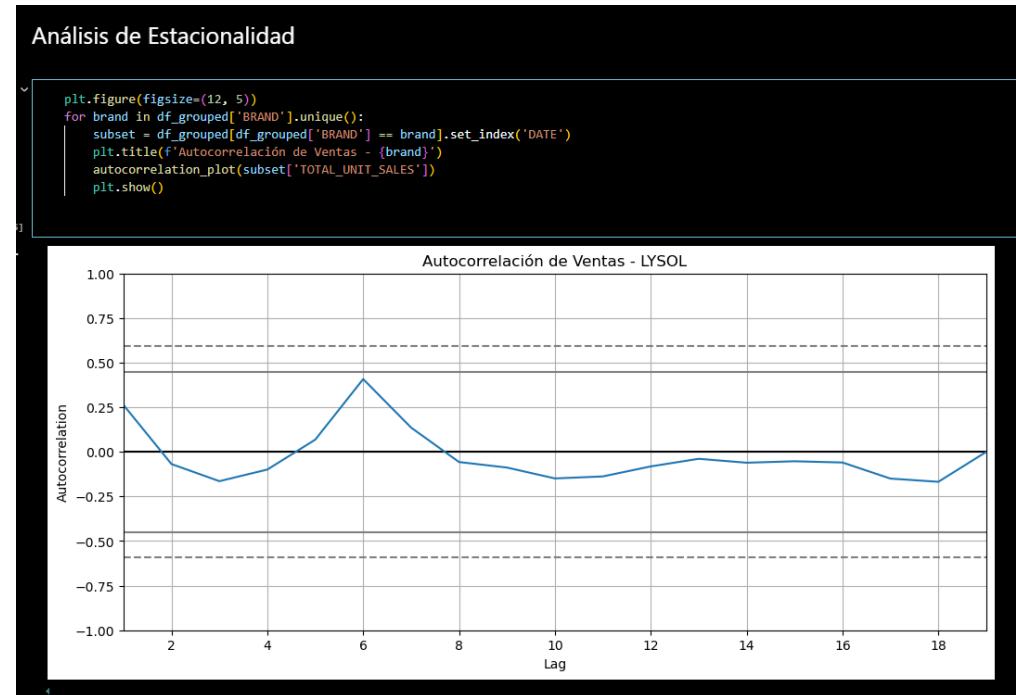
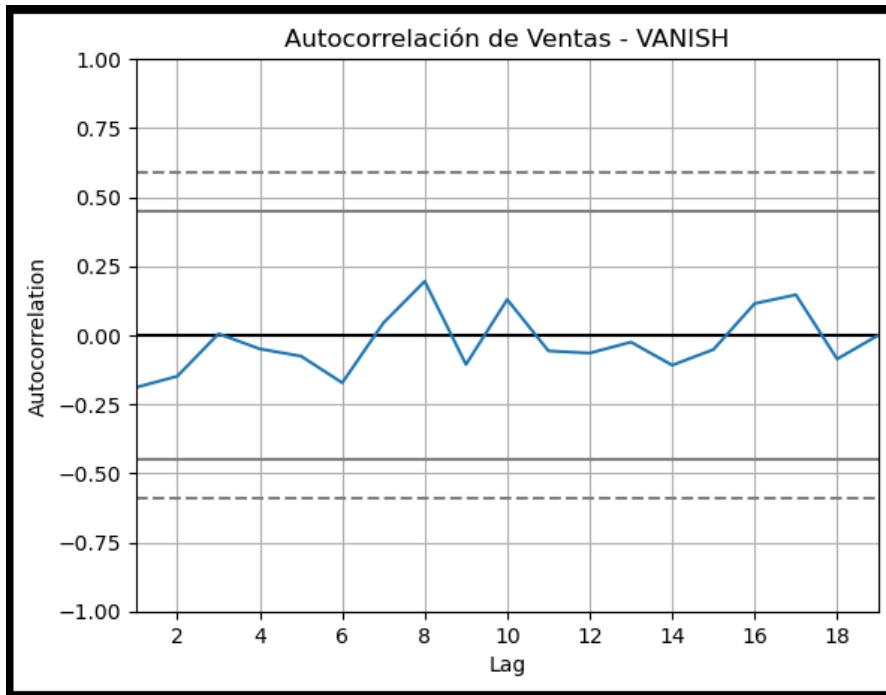
	DATE	BRAND	TOTAL_UNIT_SALES	TOTAL_VALUE_SALES
0	2022-01-01	LYSOL	69.824	9437.448
1	2022-01-01	VANISH	3210.042	125264.422
2	2022-02-01	LYSOL	35.898	5255.695
3	2022-02-01	VANISH	2486.913	90657.308
4	2022-02-01	LYSOL	32.876	3452.520

Selección técnica de modelado

Análisis de tendencias

```
plt.figure(figsize=(12, 6))
for brand in df_grouped['BRAND'].unique():
    subset = df_grouped[df_grouped['BRAND'] == brand]
    plt.plot(subset['DATE'], subset['TOTAL_UNIT_SALES'], label=brand)
plt.xlabel('Fecha')
plt.ylabel('Ventas Totales (Unidades)')
plt.title('Evolución de Ventas de Vanish y Lysol')
plt.legend()
plt.grid()
plt.show()
```





Observaciones

El gráfico muestra la evolución de las ventas mensuales de Vanish y Lysol. Se observan fluctuaciones significativas, lo que sugiere posibles patrones de tendencia y estacionalidad.

Series de tiempo (ARIMA/SARIMA) parece ser la mejor opción, ya que los datos son temporales y muestran fluctuaciones que podrían seguir patrones estacionales. Regresión podría ser útil si encontráramos una fuerte correlación con otras variables, pero como el objetivo es prever ventas futuras en función del tiempo, un modelo de series de tiempo es más adecuado.

Para este caso vamos a usar un modelo de series de tiempo SARIMA

Selección técnica de modelado

Entrenamiento del modelo

Entrenamiento del modelo

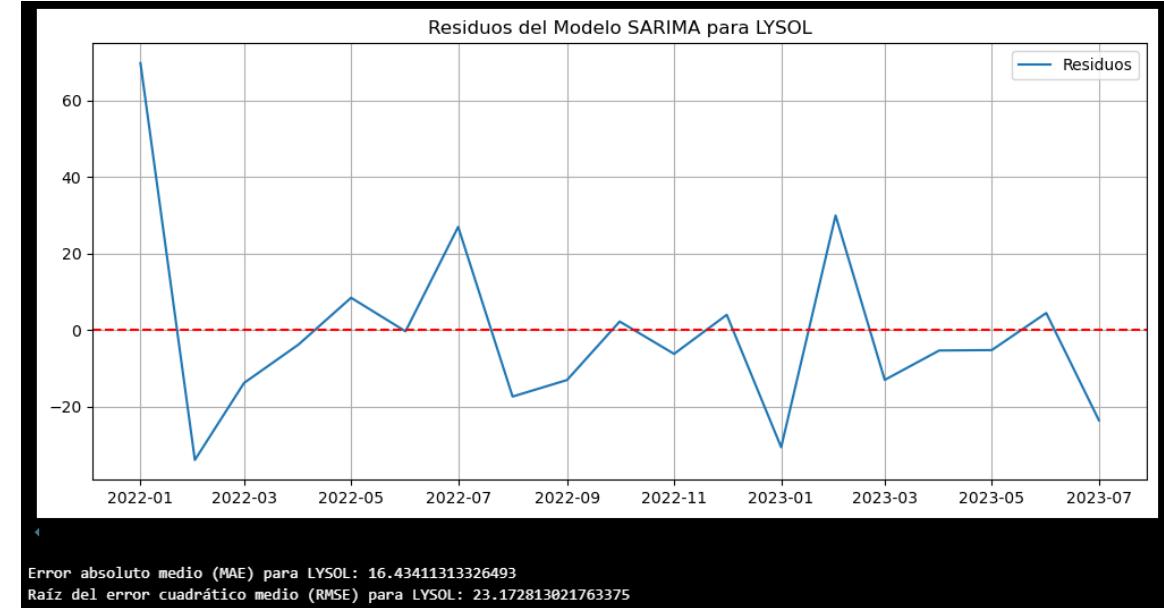
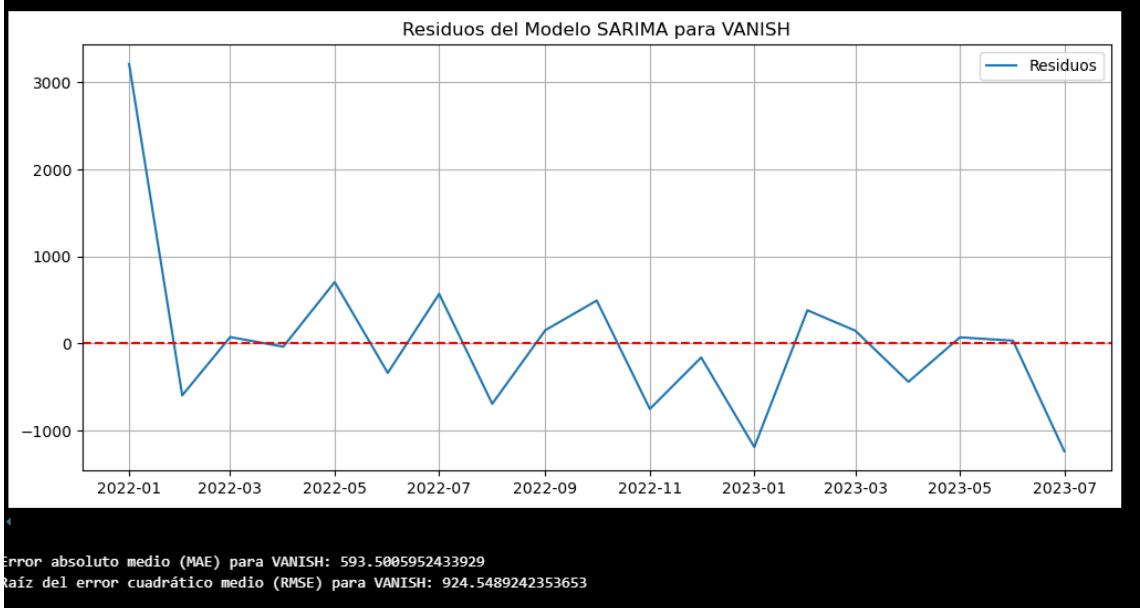
[+ Code](#)[+ Markdown](#)

```
models_dict = {}
for brand in df_grouped['BRAND'].unique():
    subset = df_grouped[df_grouped['BRAND'] == brand].set_index('DATE')
    model = SARIMAX(subset['TOTAL_UNIT_SALES'], order=(1,1,1), seasonal_order=(1,1,1,12))
    sarima_model = model.fit(disp=False)
    models_dict[brand] = sarima_model
```

Observaciones

Los gráficos de autocorrelación muestran una posible estacionalidad en las ventas de Vanish y Lysol, lo que confirma que un modelo SARIMA es adecuado para la predicción.

El error indica que no hay suficientes datos para descomponer la serie con un período de 12 meses, ya que solo hay 19 meses de datos disponibles.



Validación del modelo

Validación del modelo

```
for brand, model in models_dict.items():
    residuals = model.resid
    plt.figure(figsize=(12, 5))
    plt.plot(residuals, label='Residuos')
    plt.axhline(y=0, linestyle='dashed', color='red')
    plt.title(f'Residuos del Modelo SARIMA para {brand}')
    plt.legend()
    plt.grid()
    plt.show()
    print(f'Error absoluto medio (MAE) para {brand}:', mean_absolute_error(df_grouped[df_grouped['BRAND'] == brand]['TOTAL_UNIT_SALES'], model.fittedvalues))
    print(f'Raíz del error cuadrático medio (RMSE) para {brand}:', np.sqrt(mean_squared_error(df_grouped[df_grouped['BRAND'] == brand]['TOTAL_UNIT_SALES'], model.fittedvalues)))
```

Observaciones

Parece que los datos disponibles no son suficientes para dividir adecuadamente en conjuntos de entrenamiento y prueba, lo que impide el ajuste del modelo SARIMA. Esto sugiere que necesitamos más datos históricos o cambiar la estrategia de modelado.

Se intenta entrenar el modelo sin dividir la prueba, de modo que tenga mas datos de entrenamiento.

Ajuste del modelo

```
for brand in df_grouped['BRAND'].unique():
    subset = df_grouped[df_grouped['BRAND'] == brand].set_index('DATE')
    model = SARIMAX(subset['TOTAL_UNIT_SALES'], order=(2,1,2), seasonal_order=(1,1,1,12))
    sarima_model = model.fit(disp=False)
    models_dict[brand] = sarima_model
```

Observaciones

Se han generado predicciones de ventas futuras para los próximos 6 meses utilizando un modelo SARIMA entrenado con toda la serie de datos.

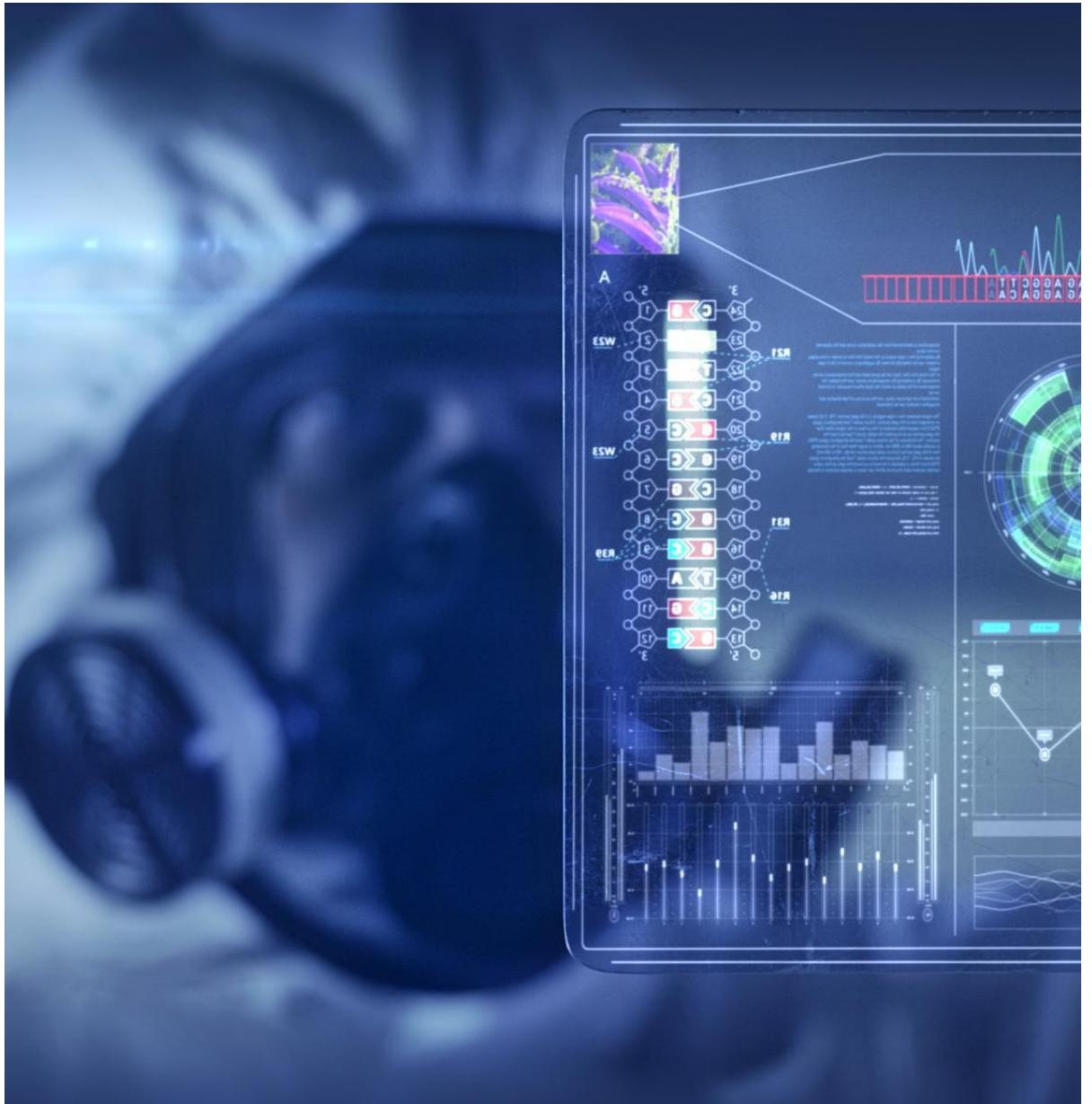
Predicciones futuras

Modelo de Machine Learning

Utilizamos un modelo de Machine Learning para generar predicciones de ventas precisas y basadas en datos históricos.

Importancia de las Predicciones

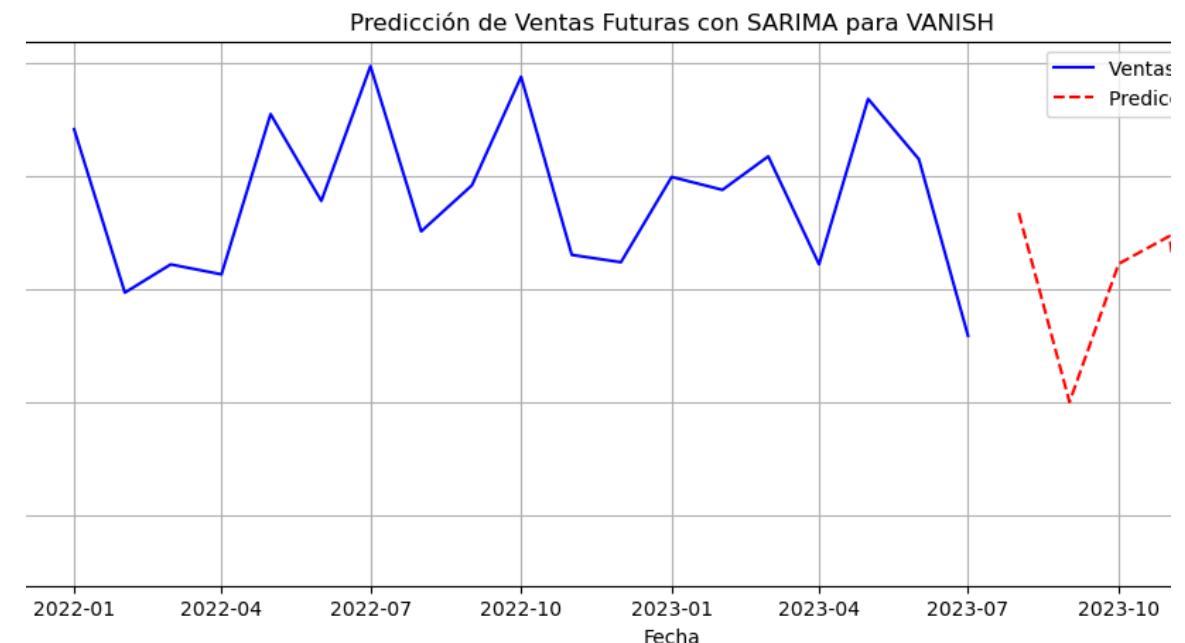
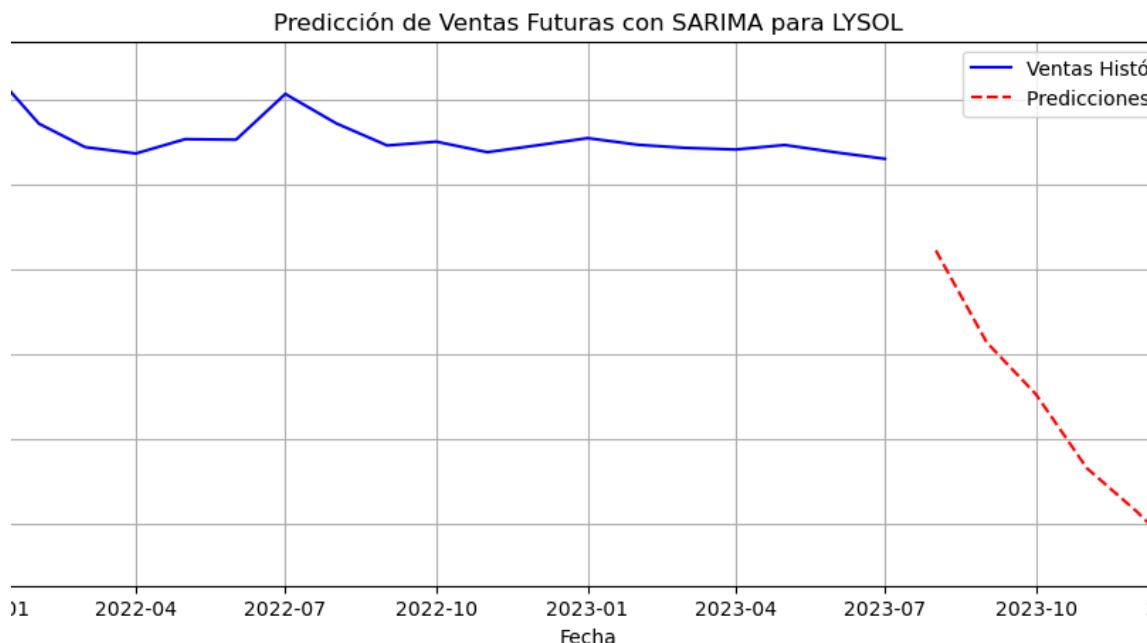
Las predicciones de ventas son fundamentales para la planificación estratégica y decisiones informadas dentro de la empresa.



Predicción de ventas futuras

```
future_months = 6
predictions_dict = {}
for brand, model in models_dict.items():
    subset = df_grouped[df_grouped['BRAND'] == brand].set_index('DATE')
    future_dates = pd.date_range(start=subset.index[-1], periods=future_months + 1, freq='MS')[1:]
    future_preds = model.predict(start=future_dates[0], end=future_dates[-1], dynamic=False)
    predictions_dict[brand] = future_preds
plt.figure(figsize=(12, 5))
plt.plot(subset.index, subset['TOTAL_UNIT_SALES'], label='Ventas Históricas', color='blue')
plt.plot(future_preds.index, future_preds, label='Predicciones Futuras', color='red', linestyle='dashed')
plt.xlabel('Fecha')
plt.ylabel('Ventas Totales (Unidades)')
plt.title(f'Predicción de Ventas Futuras con SARIMA para {brand}')
plt.legend()
plt.grid()
plt.show()
```

Predicción de ventas futuras



Predicción de ventas futuras

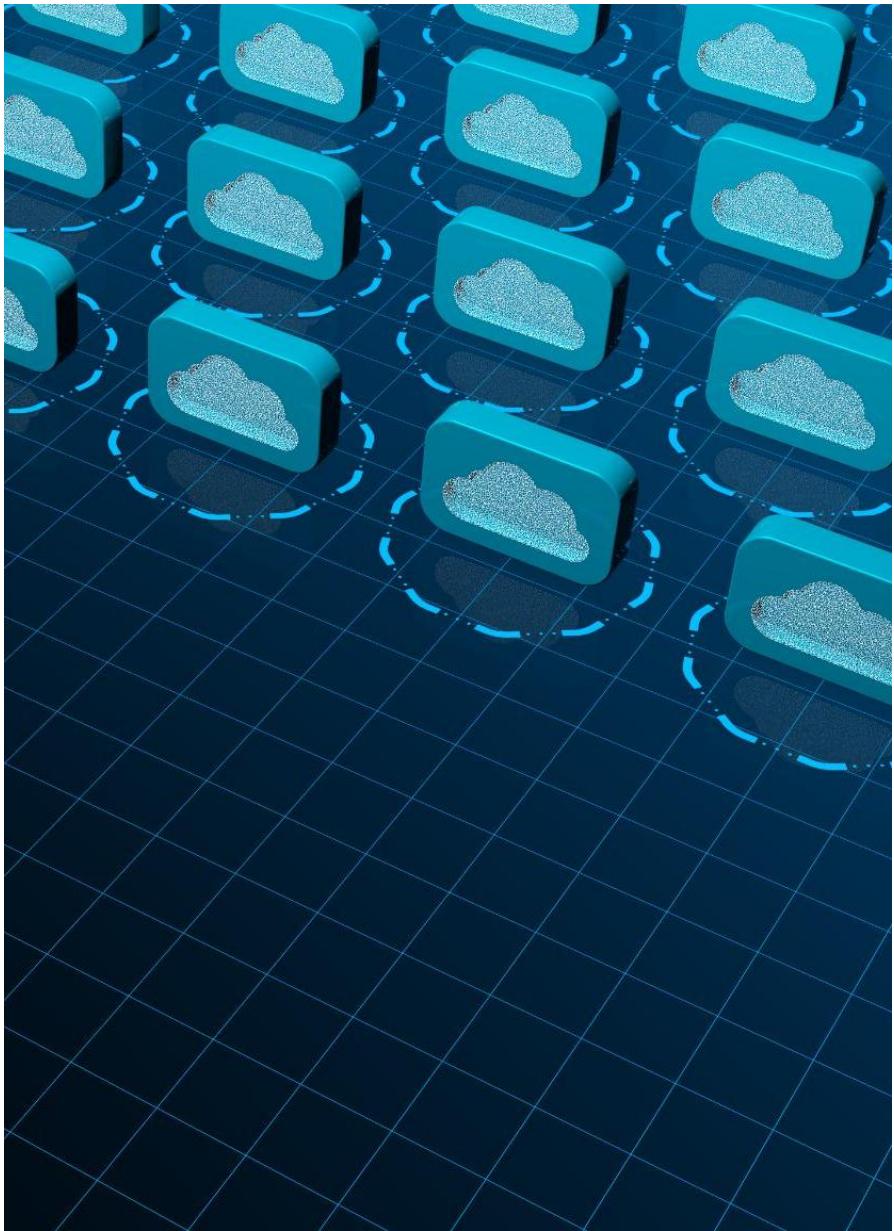
```
future_predictions_df = pd.DataFrame(predictions_dict)  
future_predictions_df.head()
```

	LYSOL	VANISH
2023-08-01	-38.954257	2839.118286
2023-09-01	-93.205003	2002.358469
2023-10-01	-123.895596	2614.157150
2023-11-01	-167.506753	2736.893738
2023-12-01	-192.773165	1299.329911

Observaciones

- Para Vanish, se observa una tendencia de crecimiento en las predicciones futuras.
- Para Lysol, los valores predichos son negativos, lo que indica que el modelo puede no estar capturando correctamente la tendencia de este producto debido a datos insuficientes o ruido en la serie.

Análisis en AWS con Cloud Computing



Configuración del Data Lake en AWS

Almacenamiento Escalable

La configuración del Data Lake en AWS permite el almacenamiento escalable de grandes volúmenes de datos, facilitando su gestión.

Análisis Eficientes

Gracias al Data Lake, podemos realizar análisis eficientes, extrayendo insights valiosos de nuestros datos almacenados.

Obtención de Insights

El uso del Data Lake en AWS nos permitió obtener insights valiosos que impulsan nuestras decisiones estratégicas.

Configuración de un data lake.

Se crea un data lake usando la herramienta de storage de AWS S3, con un bucket llamado ebfac-bucket-reckitt en donde se han cargado los archivos originales proporcionados por el cliente en la carpeta Originales, y en la de Ventas se ha cargado un archivo previamente procesado con la información Master.

The screenshot shows the AWS S3 console interface. At the top, there's a navigation bar with the AWS logo, a search bar, and various service links like Amazon Bedrock, AWS Glue, S3, Athena, QuickSight, RDS, and VPC. The user is signed in as 'Paob2000'. Below the navigation bar, the path 'Amazon S3 > Buckets > ebfac-bucket-reckitt' is visible. The main area displays the 'ebfac-bucket-reckitt' bucket details. The 'Objetos' tab is selected, showing two objects: 'Originales/' and 'Ventas/'. Both are listed as 'Carpeta' (Folder) type objects. The 'Actions' menu is open, showing options like 'Copiar URI de S3', 'Copiar URL', 'Descargar', 'Abrir', 'Eliminar', 'Acciones', 'Crear carpeta', and 'Cargar'. A search bar at the bottom left allows filtering by prefix. The table headers for the object list are 'Nombre', 'Tipo', 'Última modificación', 'Tamaño', and 'Clase de almacenamiento'. The page also includes pagination controls ('< 1 >') and a settings gear icon.

Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
Originales/	Carpeta	-	-	-
Ventas/	Carpeta	-	-	-

Creación de un Crawler en AWS Glue.

El paso siguiente es crear un Crawler llamado ebac-crawler-Reckitt-ventas que va a pasar la información del archivo Master de Ventas, a una base de datos Aurora de AWS. En el proceso se ha creado el nombre de la base de datos: “ebac-db-Reckitt”

The screenshot shows the 'Review and create' step of a crawler setup in AWS Glue. On the left, a vertical navigation bar lists five steps: Step 1 (Set crawler properties), Step 2 (Choose data sources and classifiers), Step 3 (Configure security settings), Step 4 (Set output and scheduling), and Step 5 (Review and create). The Step 5 button is highlighted with a blue circle. The main area contains four sections:

- Step 1: Set crawler properties**: Shows a table with one row for the crawler named "ebac-crawler-reckitt-ventas".

Name	Description	Tags
ebac-crawler-reckitt-ventas	-	-
- Step 2: Choose data sources and classifiers**: Shows a table with one row for an S3 data source.

Type	Data source	Parameters
S3	s3://ebfac-bucket-reckitt/Ventas/master_ventas_empresa...	Recrawl all
- Step 3: Configure security settings**: Shows a table with one row for an IAM role.

IAM role	Security configuration	Lake Formation configuration
AWSGlueServiceRole-ebac	-	-
- Step 4: Set output and scheduling**: Shows a table with one row for the output database.

Database	Table prefix - optional	Maximum table threshold - optional	Schedule
ebac-db-Reckitt	-	-	On demand

At the bottom right are three buttons: "Cancel", "Previous", and a large orange "Create crawler" button.

Ejecución de un Crawler en AWS Glue.

Se ejecuta el crawler de manera exitosa y vemos como se genera la tabla dentro de la base de datos que hemos creados.

Ya queda accesible desde Athena para hacer consultas.

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Crawlers (1) Info		Last updated (UTC) February 22, 2025 at 17:18:00	Action	Run	Create crawler
View and manage all available crawlers.					
Filter crawlers		Last run	Last run timestamp	Log	Table changes from last r...
Name	State	Schedule	Last run	Last run timestamp	Log
ebac-crawler-reckitt-ventas	Ready		Succeeded	February 22, 2025 at 17:18:00	View log
					1 created

Tables

A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

Tables (1)		Last updated (UTC) February 22, 2025 at 17:18:39	Delete	Add tables using crawler	Add table
View and manage all available tables.					
Filter tables		Last data	Data quality	Column statistics	
Name	Database	Location	Classification	Deprecated	View data
master_ventas_empresa_ali	ebac-db-reckitt	s3://ebfac-bucket-reckitt/V	UNKNOWN	-	Table data
					View data quality
					View statistics



Consultas en Athena

Consultas SQL Efectivas

Amazon Athena permite ejecutar consultas SQL en datos almacenados de manera eficiente, facilitando el análisis de grandes volúmenes de información.

Acceso Rápido a Datos

Utilizando Athena, el acceso a los datos se vuelve más ágil, optimizando el tiempo de respuesta en el análisis de datos.

Análisis de Datos Mejorado

El uso de Amazon Athena mejora el proceso de análisis al proporcionar herramientas efectivas y accesibles para los usuarios.

Consultas en Athena.

Consulta de los 5 productos mas vendidos en cada región.

The screenshot shows the AWS Athena console interface. On the left, there's a sidebar with 'Datos' (Data) settings: 'Origen de datos' set to 'AwsDataCatalog', 'Catálogo' set to 'Ningún elemento', and 'Base de datos' set to 'ebac-db-reckitt'. Below this are sections for 'Tablas y vistas' (Tables and views) and 'Tablas (2)' containing 'master_ventas_empresa_aliada_csv' and 'ventas'. There's also a 'Vistas (0)' section. The main area is titled 'Consulta 5' and contains the following SQL code:

```
9 SELECT
10    region,
11    item_description,
12    total_vendido
13  FROM (
14    SELECT
15      region,
16      item_description,
17      total_vendido,
18      ROW_NUMBER() OVER (PARTITION BY region ORDER BY total_vendido DESC) AS rn
19    FROM ventas_por_producto
20  ) ranked
21  WHERE rn <= 5
22  ORDER BY region, total_vendido DESC;
```

Below the code, it says 'SQL Ln 23, Col 1'. There are buttons for 'Ejecutar de nuevo' (Run new), 'Explicar' (Explain), 'Cancelar' (Cancel), 'Borrar' (Delete), and 'Crear' (Create). A checkbox 'Volver a utilizar los resultados de la consulta hasta hace 60 minutos' (Reuse query results for up to 60 minutes) is checked. The results section is titled 'Resultados (35)' and shows a table with columns: '#', 'region', 'item_description', and 'total_vendido'. The data rows are:| # | region | item_description | total_vendido |
| --- | --- | --- | --- |
| 1 | TOTAL AUTOS AREA 1 | CLORALEX EL RENDIDOR BOT.PLAST. 2000ML NAL 0000075000622 | 89383.37800000001 |
| 2 | TOTAL AUTOS AREA 1 | BLANCATEL CONCENTRADO BOT PLAST 3750ML NAL 7501025402051 | 72583.232 |
| 3 | TOTAL AUTOS AREA 1 | CLORALEX EL RENDIDOR BOT.PLAST. 950ML NAL. 0000075000615 | 65106.69200000001 |
| 4 | TOTAL AUTOS AREA 1 | CLORALEX EL RENDIDOR BOT.PLAST. 3750ML NAL 0000075000639 | 62232.00299999999 |
| 5 | TOTAL AUTOS AREA 1 | CLORALEX ULTRAGEL CLORO GEL CONCENTRADO BOT 600 ML 7501025405090 | 22297.630999999998 |

Buttons for 'Copiar' (Copy) and 'Descargar resultados en formato CSV' (Download results in CSV format) are at the bottom of the results table.

Consultas en Athena.

Consulta de ventas totales por mes, ordenada por mayores ventas.

The screenshot shows the AWS Athena Editor interface. On the left, the 'Datos' sidebar is open, showing the 'Origen de datos' set to 'AwsDataCatalog', 'Catálogo' set to 'Ningún elemento', and 'Base de datos' set to 'ebac-db-reckitt'. Under 'Tablas y vistas', there are two tables: 'master_ventas_empresa_aliana_csv' and 'ventas'. The main area displays a query editor with two tabs: 'Consulta 5' and 'Consulta 6'. The 'Consulta 6' tab contains the following SQL code:

```
1 SELECT
2     month,
3     year,
4     SUM(total_value_sales) AS total_sales
5 FROM ventas
6 GROUP BY month, year
7 ORDER BY total_sales DESC;
```

Below the code, the status bar shows 'SQL Ln 8, Col 1'. There are buttons for 'Ejecutar de nuevo', 'Explicar', 'Cancelar', 'Borrar', and 'Crear'. A checkbox 'Volver a utilizar los resultados de la consulta hasta hace 60 minutos' is checked. The results section at the bottom shows the query has completed with a status bar indicating 'Tiempo en cola: 58 ms', 'Tiempo de ejecución: 734 ms', and 'Datos analizados: 22.59 MB'. The results table has columns '#', 'month', 'year', and 'total_sales', with 19 rows of data. The first few rows are:| # | month | year | total_sales |
| --- | --- | --- | --- |
| 1 | 5 | 2023 | 711883.291 |
| 2 | 7 | 2022 | 702062.391000001 |
| 3 | 1 | 2022 | 678440.708 |
| 4 | 5 | 2022 | 674949.5579999996 |
| 5 | 10 | 2022 | 668951.1230000037 |
| 6 | 6 | 2023 | 654159.2789999992 |
| 7 | 6 | 2022 | 603800.0239999985 |
| 8 | 1 | 2023 | 584020.629 |
| 9 | 9 | 2022 | 561226.1299999985 |

Consultas en Athena.

Consulta de ventas por marca.

The screenshot shows the AWS Management Console with the Athena service selected. The main area displays a query editor with two tabs: 'Consulta 5' and 'Consulta 6'. The 'Consulta 6' tab is active, containing the following SQL code:

```
1 SELECT
2     brand,
3     format('$%,.2f', SUM(total_value_sales)) AS total_sales
4 FROM ventas
5 GROUP BY brand
6 ORDER BY SUM(total_value_sales) DESC;
```

Below the code, there are buttons for 'Ejecutar de nuevo' (Run again), 'Explicar' (Explain), 'Cancelar' (Cancel), 'Borrar' (Delete), and 'Crear' (Create). A note indicates that results can be reused for up to 60 minutes. The 'Resultados de la consulta' section shows the query has completed successfully. The results table lists 24 brands with their total sales, ordered from highest to lowest. The top entries include CLORALEX, VANISH, and CLOROX.

brand	total_sales
CLORALEX	\$5,393,618.51
VANISH	\$2,168,289.53
CLOROX	\$1,371,751.72
BLANCATEL	\$652,484.02
OXI CLEAN	\$436,970.72
DR.BECKMANN	\$279,719.23
BANDERA	\$269,936.90
LOS PATITOS	\$142,745.66
CLARASOL	\$141,012.18
LYSOL	\$70,557.35

Conclusiones y Recomendaciones

Resumen de hallazgos clave

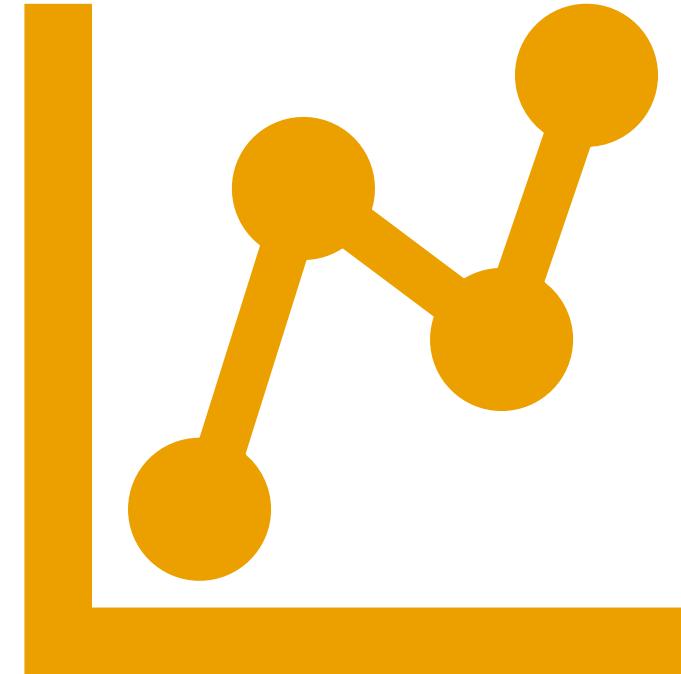
- Se reportan 24 marcas vendidas, y solo 3 de ellas hacen el "PARETO" (80/20) en las ventas.
 - Estas marcas son Cloralex, Vanish, Clorox. En su mismo orden son la gran representación del 80% de todas las ventas, y aún más importante sólo Cloralex representa casi el 50% de las ventas. Este dato es consistente no sólo en ventas totales, sino en cada una de las regiones de manera independiente.
 - La región Scanning México vende el 50% y lo demás está distribuido en las otras 6 zonas.
 - Las siguientes regiones con mejor desempeño son las Áreas 2,5 y 6 en su respectivo orden aportando así un 30,12% de las ventas.

Estos son los mejores 5 meses de ventas en su orden:

 - Mayo de 2023
 - Julio de 2022
 - Enero de 2022
 - Mayo de 2022
 - Octubre de 2022
- En particular en Mayo de 2022 igualmente se tiene un ticket promedio mas alto y una cantidad de productos mas baja, lo cual indica que se vendieron los productos a mejor precio o hubo venta de productos especialmente que son mas caros.
 - El número de unidades vendidas en promedio por semana es bastante consistente todos los meses.
 - Los meses con mejor ticket promedio coinciden con los mejores meses de ventas, con la particularidad que antes se mencionó ya que en Mayo de 2023 además de mejores ventas y ticket promedio, la cantidad de productos no fue tan alta.
 - No se detectan valores atípicos en la mayoría de KPIs de rendimiento, por lo general el comportamiento de ventas es muy estándar, y no hay mucha desviación en ventas, cantidad de productos y ticket promedio. Además de las tablas las gráficas de Boxplot nos permiten confirmar esto claramente.
 - La línea de tendencia de ventas durante los 18 meses va al alza levemente, lo que indica que las ventas en general van subiendo, pero de una manera muy lenta.
 - El gráfico de correlación entre unidades vendidas y ventas es uno de los hallazgos mas importantes en este análisis.
 - Hay una gran correlación entre productos con menos ventas y menos cantidad vendida, pero en la medida que sube el precio y la cantidad de productos vendidos, logramos visualizar 3 líneas de tendencia bien marcadas.
 - El coeficiente de correlación es bastante alto y R2 nos da un 83,95% de confiabilidad en esta tendencia.

Impacto de análisis en la estrategia empresarial

- El análisis revela una dependencia significativa de tres marcas (Cloralex, Vanish y Clorox), que en conjunto representan el 80% de las ventas, con Cloralex aportando casi el 50%. Esta concentración, si bien indica una solidez operativa, también expone a la empresa a riesgos importantes en caso de problemas con alguna de ellas. En este sentido, se hace crucial diversificar el portafolio de marcas, potenciando aquellas con menor participación (como BLANCATEL, OXI CLEAN, DR. BECKMANN, BANDERA y LOS PATITOS) para asegurar una curva de crecimiento más acelerada y reducir la vulnerabilidad ante fluctuaciones del mercado.
- Por otro lado, se destaca la distribución geográfica de las ventas, con áreas de alto rendimiento como la 2, 5 y 6, que suman alrededor del 60,24% de las ventas. Este patrón regional sugiere que, si bien algunas zonas ya están consolidadas, otras requieren un análisis más profundo para identificar oportunidades de mejora o ajustes en la estrategia comercial. La segmentación por región permitirá enfocar los recursos e iniciativas en aquellas áreas con mayor potencial de crecimiento y rentabilidad.
- Finalmente, la estabilidad en los indicadores clave de desempeño y la correlación alta entre unidades vendidas y ventas demuestran un comportamiento consistente y confiable, aunque con un crecimiento moderado en el tiempo. La anomalía observada en mayo de 2023, donde se alcanzó un ticket promedio más alto a pesar de menor volumen de productos, representa una oportunidad para investigar y replicar estrategias que potencien la rentabilidad. En conjunto, estos hallazgos subrayan la necesidad de diversificar la oferta, reforzar la estrategia regional y profundizar en el análisis de nichos de mercado para impulsar un crecimiento sostenido y minimizar riesgos.



Recomendaciones

- Se recomienda diversificar la dependencia actual en las tres marcas líderes (Cloralex, Vanish y Clorox), que aunque generan el 80% de las ventas, representan un riesgo significativo si se llegara a presentar algún inconveniente con alguna de ellas. Es fundamental impulsar el crecimiento de marcas emergentes –como BLANCATEL, OXI CLEAN, DR. BECKMANN, BANDERA y LOS PATITOS– mediante campañas de marketing focalizadas y estrategias comerciales específicas. Esto no solo contribuirá a reducir la vulnerabilidad de la cartera de productos, sino que también permitirá explorar nuevos nichos de mercado y acelerar el crecimiento de la empresa a mediano y largo plazo.
- Además, se debe realizar un análisis profundo del desempeño regional para identificar variables clave que influyen en la rentabilidad de cada zona. Dado que la región Scanning México y las áreas 2, 5 y 6 concentran la mayor parte de las ventas, es importante estudiar las características de estos mercados para replicar las mejores prácticas en regiones con menor rendimiento. Igualmente, la anomalía observada en mayo de 2023 –con un ticket promedio elevado y menor volumen de productos vendidos– sugiere la existencia de estrategias de venta que, de ser identificadas y replicadas, podrían mejorar sustancialmente la rentabilidad. Implementar un seguimiento continuo de los KPIs permitirá ajustar las estrategias de manera proactiva y fomentar un crecimiento sostenible en el futuro.



Futuras mejoras



- Como mejora inicial, es fundamental revisar y estandarizar la calidad de los datos, especialmente en lo que respecta a las unidades de ventas, que actualmente presentan ambigüedad. Implementar un proceso robusto de validación y conversión permitirá que las métricas sean comparables en el tiempo y entre regiones, reduciendo posibles sesgos en el análisis. Además, considerar la integración de fuentes adicionales (por ejemplo, datos de inventario, campañas de marketing o feedback de clientes) enriquecerá el contexto y la capacidad de tomar decisiones basadas en información más completa.
- Otra recomendación es profundizar en el análisis regional y de segmentos de clientes. Realizar estudios detallados en aquellas áreas con bajo rendimiento o con potencial de crecimiento permitirá identificar variables específicas (como disponibilidad de producto, posicionamiento de marca o estrategias locales de venta) que puedan ser optimizadas. Este enfoque segmentado facilitará el diseño de estrategias comerciales personalizadas y la implementación de acciones correctivas oportunas.
- Por último, se sugiere explorar técnicas avanzadas de análisis, como modelos predictivos y machine learning, para anticipar tendencias y comportamientos del mercado. Estas herramientas ayudarán a detectar oportunidades emergentes y ajustar la estrategia de marketing y ventas en tiempo real, permitiendo un monitoreo continuo de los KPIs clave. De esta manera, la empresa podrá responder proactivamente a cambios en el entorno y fomentar un crecimiento sostenido a mediano y largo plazo.