

Proyecto Final - Empresa Aliada

Ruben Tamez

Introducción

Contexto del proyecto

Analizar el desempeño histórico de ventas de productos de consumo para identificar patrones, oportunidades de crecimiento y apoyar la toma de decisiones estratégicas.

Relevancia

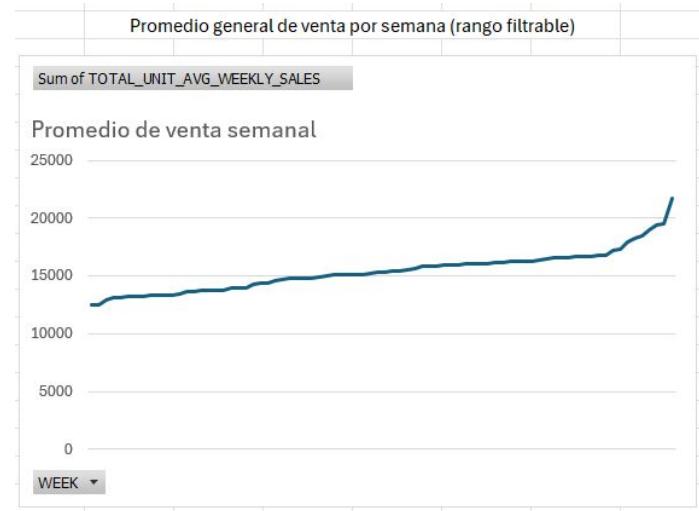
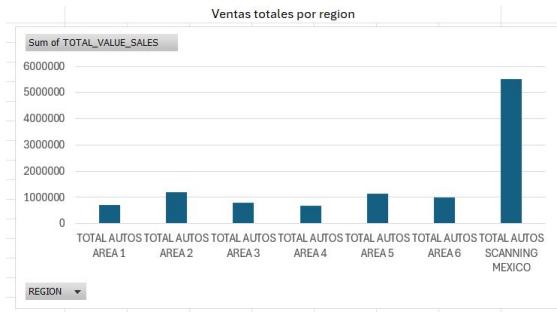
Permite entender qué productos, regiones y períodos generan mayor impacto en ventas y anticipar el comportamiento futuro del mercado.

Herramientas

Excel, SQL Server, Python (Pandas, Matplotlib, Seaborn), Looker Studio y AWS (S3, Glue, Athena).

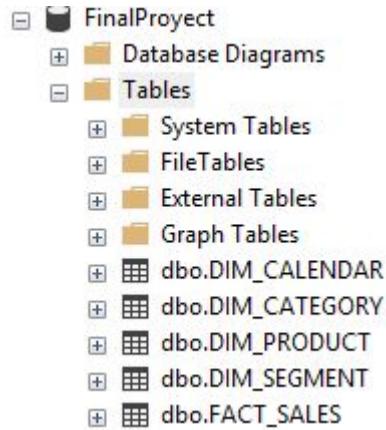
Excel – Exploración

Ventas Totales por ITEM_CODE		
Row Labels		Sum of TOTAL_UNIT_SALES
75000622	\$	48,445.15
75000615	\$	46,461.84
75000639	\$	28,334.42
7501025402051	\$	15,613.21
7501025405090	\$	14,031.41
7501071900143	\$	11,129.23
7501025405151	\$	10,968.69
7501025450212	\$	10,079.70
7501025405212	\$	9,667.08
7501058714596	\$	8,059.99
7501071900129	\$	8,041.87
741120000005	\$	7,171.51
7501058717023	\$	7,024.09
7891035040276	\$	6,988.17
7501025401801	\$	6,166.92



Se utilizaron tablas dinámicas para identificar patrones de ventas por categoría, región y tiempo. Las visualizaciones permitieron detectar tendencias generales y posibles outliers.

SQL – Análisis



```
2 | select *  
3 | from DIM_CALENDAR  
4 |  
5 | select *  
6 | from DIM_CATEGORY  
7 |  
8 | select *  
9 | from DIM_PRODUCT  
10 |  
11 | select *  
12 | from FACT_SALES  
13 |  
14 | select *  
15 | from DIM_SEGMENT
```

Results

WEEK	YEAR	MONTH	WEEK_NUMBER	DATE
04-21	2021	1	4	2021-01-31

ID_CATEGORY	CATEGORY
1	FABRIC TREATMENT and SANIT

MANUFACTURER	BRAND	ITEM	ITEM_D
INDS. ALEN	CLORALEX	0000075000592	CLORAL

WEEK	ITEM_CODE	TOTAL_UNIT_SALES	TOT/
34-22	7501058716422	0.021999999807907	1.48

CATEGORY	ATTR1	ATTR2	ATTR3	FORMAT
1	SAFE...	FABR...	PODER O...	GEL

```
1 | SELECT  
2 | c.CATEGORY,  
3 | SUM(f.TOTAL_VALUE_SALES) AS total_sales  
4 | FROM FACT_SALES f  
5 | JOIN DIM_PRODUCT p  
6 | ON f.ITEM_CODE = p.ITEM  
7 | JOIN DIM_CATEGORY c  
8 | ON p.CATEGORY = c.ID_CATEGORY  
9 | GROUP BY c.CATEGORY;
```

Results

CATEGORY	total_sales
AIR CARE	1109246.92982756
FABRIC TREATMENT and SANIT	7478343.72615999
LAVATORY CARE & BRC	16595.1380126047
LAVAJAILLAS	243539.229876287
MEGA SUPERFICIES	2195134.86878494

Se diseñó un esquema dimensional y se realizaron consultas con joins para analizar ventas por categoría y región.

Python – Limpieza

```
# Carga de Datos
base_path = r"C:\Users\Rubén\Documents\DATA ANALIST\git\TareasEBACmx\ProyectoFinal-EmpresaAliada\DATA\csv"
dim_calendar = pd.read_csv(f"{base_path}\DIM_CALENDAR.csv")
dim_category = pd.read_csv(f"{base_path}\DIM_CATEGORY.csv")
dim_product = pd.read_csv(f"{base_path}\DIM_PRODUCT.csv")
dim_segment = pd.read_csv(f"{base_path}\DIM_SEGMENT.csv")
fact_sales = pd.read_csv(f"{base_path}\FACT_SALES.csv")
fact_sales.head()
```

```
# Eliminar duplicados
fact_sales.drop_duplicates(inplace=True)

# Convertir fechas
dim_calendar['DATE'] = pd.to_datetime(dim_calendar['DATE'], format='%d/%m/%Y')

# Asegurar tipos numéricos
fact_sales['TOTAL_VALUE_SALES'] = pd.to_numeric(
    fact_sales['TOTAL_VALUE_SALES'], errors='coerce'
)

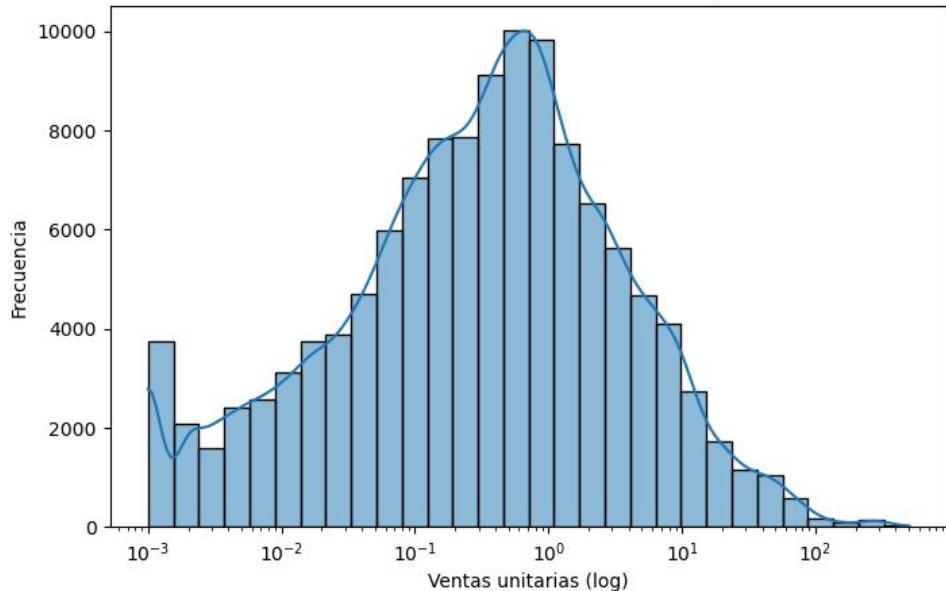
fact_sales['TOTAL_UNIT_SALES'] = pd.to_numeric(
    fact_sales['TOTAL_UNIT_SALES'], errors='coerce'
)
```

```
✓ df = fact_sales \
    .merge(dim_product, left_on='ITEM_CODE', right_on='ITEM', how='left') \
    .merge(dim_category, left_on='CATEGORY', right_on='ID_CATEGORY', how='left') \
    .merge(dim_calendar, on='WEEK', how='left')
df.head()
```

Los datos fueron cargados desde archivos CSV utilizando Pandas. Se realizó limpieza eliminando duplicados, corrigiendo tipos de datos y estandarizando fechas. Posteriormente, las tablas de hechos y dimensiones fueron integradas en una única tabla consolidada para facilitar el análisis exploratorio y la visualización.

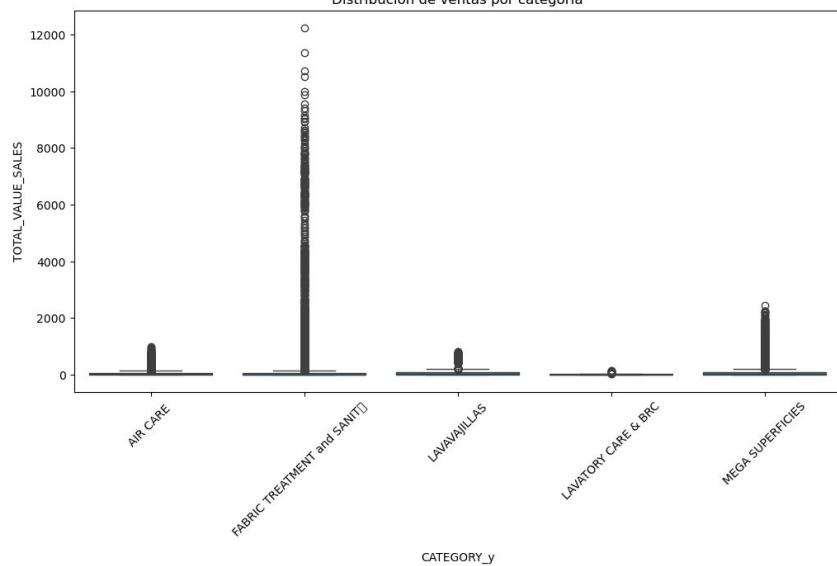
EDA – Visualizaciones

Distribución de ventas unitarias (escala logarítmica)



Debido a la alta dispersión y presencia de valores extremos, se utilizó una escala logarítmica para visualizar adecuadamente la distribución de las ventas unitarias, gracias a esto presenciamos una campana de Gauss.

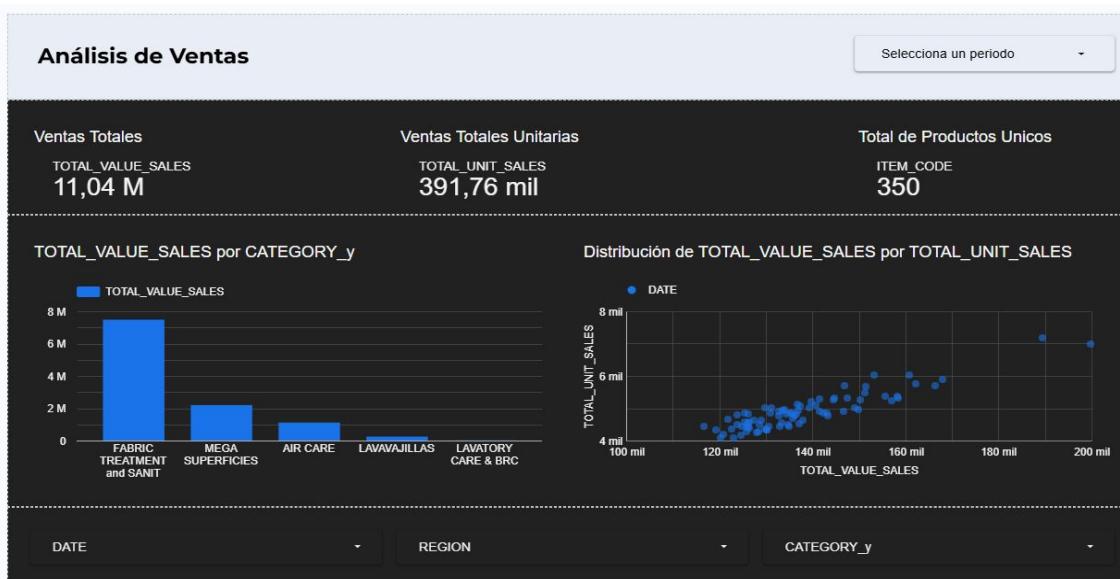
Distribución de ventas por categoría



Notamos una diferencia significativa en la dispersión de 'Fabric Treatment and Sanit', nada raro que observar, simplemente la gran mayoría de productos llevan esta categoría.

Looker Studio – Dashboard

El dashboard de looker studio presenta KPI's imprescindibles a la hora de analizar ventas además de tablas que miden magnitudes de ventas y a su vez, filtros útiles para su posterior uso, fue exportado en formato "Adaptable" por lo que se puede usar a través de teléfonos o tablets sin necesidad de escalar.

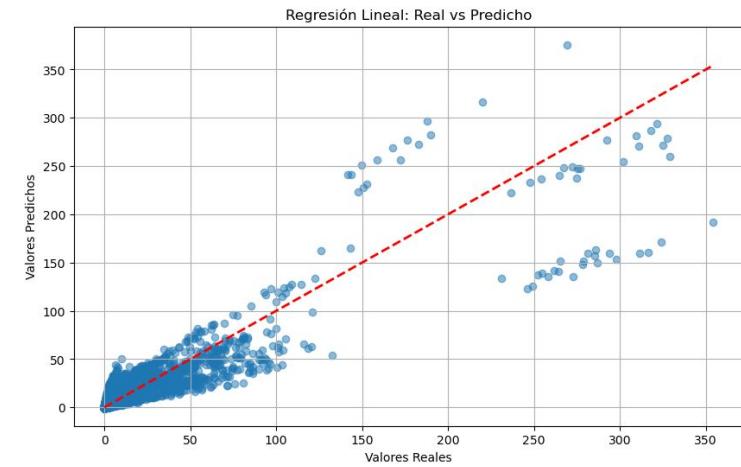
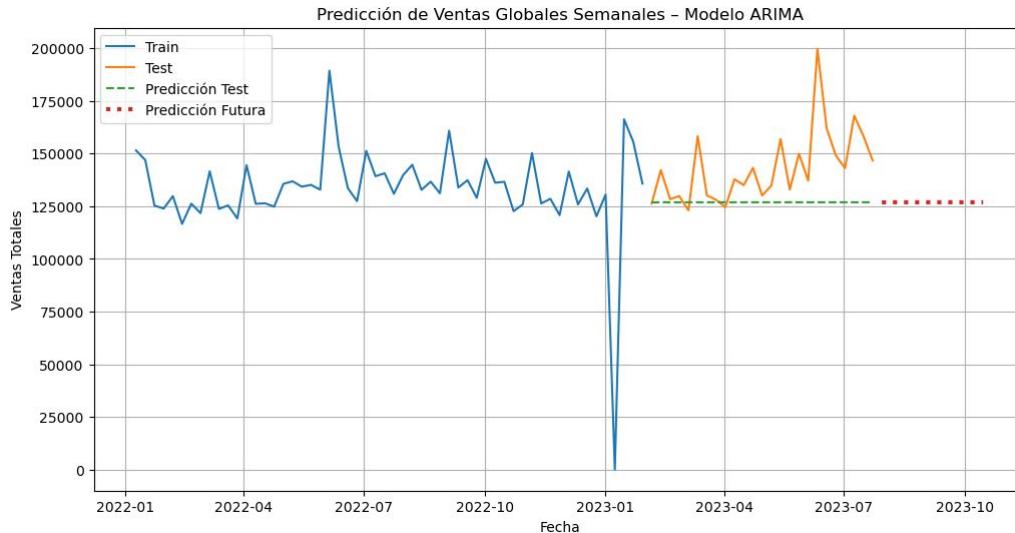


Se puede acceder a través de este link:

<https://lookerstudio.google.com/reporting/c20b6466-fcc8-400d-a916-8a5484a15fc2>

Machine Learning - Predicción de ventas

Por desgracia nuestros modelos no aportan información real futura de ventas, probe múltiples remedios e intentos con IA, pero por desgracia no fue posible la predicción de datos.



AWS - Cloud Computing

Archivos y carpetas (6 total)

	Buscar por nombre	Estado
Nombre		
FACT_SALES_CONSOLIDATED.csv	🔗	Realizado correctamente
DIM CALENDAR.csv	🔗	Realizado correctamente
DIM CATEGORY.csv	🔗	Realizado correctamente
DIM PRODUCT.csv	🔗	Realizado correctamente
DIM SEGMENT.csv	🔗	Realizado correctamente
FACT SALES.csv	🔗	Realizado correctamente

Creación y subida de datos a nuestro bucket “ventasproductos2026”, incluimos nuestra tabla consolidada creada en la interfaz de Python (previamente en excel exportamos todos los archivos a csv para un mejor manejo y uso de datos)

Review and create

Step 1: Set crawler properties

Set crawler properties

Name	Description
Dim.crawl	-

Step 2: Choose data sources and classifiers

Data sources (1) Info
The list of data sources to be scanned by the crawler.

Type	Data source
S3	s3://ventasproducto2026/

Step 3: Configure security settings

Configure security settings

IAM role	Security configuration
AWSGlueServiceRole-Admin	-

Step 4: Set output and scheduling

Set output and scheduling

Database	Table prefix - optional
rekkit-sales.db	-

Creación del crawler que importara nuestras tablas a nuestra db llamada “rekkit-sales”, los datos serán importados a través de nuestro bucket antes creado.

Consulta de visibilidad en Athena, la ETL está lista para futuras consultas.

Origen de datos
AwsDataCatalog

Catálogo
Ningún elemento

Base de datos
rekkit-sales

Tablas y vistas
Crear ▾

Filtrar tablas y vistas

SQL Ln 2, Col 1

```
1 SELECT *
2 FROM "AwsDataCatalog"."rekkit-sales"."archivos" limit 10;
```

Ejecutar de nuevo ▾ Explicar Cancelar Borrar Crear ▾

Volver a utilizar los resultados de la consulta hasta hace 60 minutos

Resultados de la consulta Estado de la consulta

Completo Tiempo en cola: 56 ms Tiempo de ejecución: 448 ms Datos analizados: 614.00 KB

Resultados (10)

Copiar Descargar resultados en formato CSV

#	week	item_code	total_unit_sales	total_value_sales	total_unit_avg_weekly_sales	region	manufacturer	brand	item	item_description
1	34-22	7501058792808BP2	0.006	0.139	1.0	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058792808BP2	VANISH OXI ACTION ROSADO DOYPACK 120GR+MMCRYSTALWHITE BOT 9
2	34-22	7501058715883	0.487	116.519	2.916	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058715883	VANISH OXI ACTION GOLD QUITAMANCHAS BOLSA 1.8KG 75010587
3	34-22	7702626213774	1.391	68.453	5.171	TOTAL AUTOS AREA 5	RECKITT	VANISH	7702626213774	VANISH OXI ACTION ROSA QUITAMANCHAS DOYPACK 240 GR 77026
4	34-22	7501058716422	0.022	1.481	1.833	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058716422	VANISH OXI ACTION GOLD QUITAMANCHA AHORRO DEL 30% BOT 4
5	34-22	7501058784353	2.037	182.839	5.375	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058784353	VANISH INTELLIGENCE POLVO BOTE 450 GR NAL 7501058784353
6	34-22	7501058716064	0.005	0.679	1.25	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058716064	VANISH OXI ACTION ROSA QUITA/MANCHAS P/ROPA BOT 810GR 75
7	34-22	7501058794963BP1	0.004	0.553	1.0	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058794963BP1	VANISH OXI ACTION GOLDROSABOT 900GR+MMCRYSTALWHITEBOT 15
8	34-22	7501058792778801	0.001	0.129	1.0	TOTAL AUTOS AREA 5	RECKITT	VANISH	7501058792778801	VANISH OXI ACTION GOLDROSABOT 600GR+MMCRYSTALWHITEBOT 15

Conclusiones

- Se construyó exitosamente un **proceso ETL** para integrar, limpiar y consolidar datos de ventas a nivel global.
- La estandarización de variables clave (fecha, región, marca y métricas de venta) permitió **análisis consistentes**.
- El dataset final fue optimizado para **visualización en Looker Studio**, habilitando análisis de desempeño y tendencias.
- El análisis temporal mostró **variabilidad en las ventas semanales**, sin patrones estacionales fuertes y consistentes.
- El mayor valor del proyecto reside en la **calidad del dato, la trazabilidad del ETL y el análisis descriptivo**, más que en la predicción