

REPORTE 01

Omar Nayib García Sabag

Diciembre 2021

1 Introducción

LifeStore es una tienda virtual que maneja una amplia gama de artículos, recientemente, la Gerencia de ventas, se percató que la empresa tiene una importante acumulación de inventario. Asimismo, se ha identificado una reducción en las búsquedas de un grupo importante de productos, lo que ha redundado en una disminución sustancial de sus ventas del último trimestre.

En el siguiente trabajo se expondrá un análisis sobre las ventas utilizando *Python*.

2 Definición del Código

Primero, cargamos los datos de LifeStore que se encuentran en el archivo *lifestore_file.py*.

2.1 Librerías

Cargamos las librerías que vamos a utilizar para hacer el análisis de los datos.

```
1 import pandas as pd
2 import plotly.express as px
```

La librería *Pandas* es con la que vamos a manipular los datos, la de *Plotly* va a ser la librería para graficar los resultados.

2.2 Preparación de los datos

Ahora lo que hacemos es transformar las listas de LifeStore a formato DataFrame para poder trabajar con ellas de una forma más sencilla.

La primera en ser procesada es la lista de productos *lifestore_products*.

```
1 df_lifestore_products = pd.DataFrame(lifestore_products, columns = ['
    id_product', 'name', 'price', 'category', 'stock'])
2 df_lifestore_products
```

Obtenemos un DataFrame de 96 filas y 5 columnas con toda la información de los productos que hay en la tienda.

	id_product	name	price	category	stock
0	1	Procesador AMD Ryzen 3 3300X S-AM4, 3.80GHz, Q...	3019	procesadores	16
1	2	Procesador AMD Ryzen 5 3600, S-AM4, 3.60GHz, 3...	4209	procesadores	182
2	3	Procesador AMD Ryzen 5 2600, S-AM4, 3.40GHz, S...	3089	procesadores	987
3	4	Procesador AMD Ryzen 3 3200G con Gráficos Rade...	2209	procesadores	295
4	5	Procesador Intel Core i3-9100F, S-1151, 3.60GH...	1779	procesadores	130
...
91	92	Gettech Audifonos con Micrófono Sonority, Alá...	149	audifonos	232
92	93	Ginga Audifonos con Micrófono GI18ADJ01BT-RO, ...	160	audifonos	139
93	94	HyperX Audifonos Gamer Cloud Flight para PC/PS...	2869	audifonos	12
94	95	logear Audifonos Gamer GHG601, Alámbrico, 1.2 ...	999	audifonos	2
95	96	Klip Xtreme Audifonos Blast, Bluetooth, Inalám...	769	audifonos	2

96 rows × 5 columns

La siguiente lista en procesar es la del historial de ventas *lifestore_sales*. Para ésta agregamos una columna para el día, mes, año y cuatrimestre de la venta para poder hacer un análisis más específico en cada periodo de tiempo.

```

1 df_lifestore_sales = pd.DataFrame(lifestore_sales, columns = ['id_sale',
2                               'id_product', 'score', 'date', 'refund'])
3 df_lifestore_sales = pd.concat([df_lifestore_sales, df_lifestore_sales["
4                               date"].str.split("/", expand = True)], axis = 1)
5 df_lifestore_sales.columns = ['id_sale', 'id_product', 'score', 'date',
6                               'refund', 'day', 'month', 'year']
7 quarter = []
8 for m in df_lifestore_sales['month']:
9     if m in ['01', '02', '03']:
10         q = 'Q1'
11     elif m in ['04', '05', '06']:
12         q = 'Q2'
13     elif m in ['07', '08', '09']:
14         q = 'Q3'
15     elif m in ['10', '11', '12']:
16         q = 'Q4'
17     quarter.append(q)
18 df_lifestore_sales['quarter'] = pd.Series(quarter)
19 df_lifestore_sales

```

El resultado es un DataFrame con 283 filas y 9 columnas.

	id_sale	id_product	score	date	refund	day	month	year	quarter
0	1	1	5	24/07/2020	0	24	07	2020	Q3
1	2	1	5	27/07/2020	0	27	07	2020	Q3
2	3	2	5	24/02/2020	0	24	02	2020	Q1
3	4	2	5	22/05/2020	0	22	05	2020	Q2
4	5	2	5	01/01/2020	0	01	01	2020	Q1
...
278	279	84	5	05/05/2020	0	05	05	2020	Q2
279	280	85	5	05/05/2020	0	05	05	2020	Q2
280	281	85	5	28/04/2020	0	28	04	2020	Q2
281	282	89	3	06/01/2020	0	06	01	2020	Q1
282	283	94	4	10/04/2020	0	10	04	2020	Q2

283 rows × 9 columns

Por último, convertimos la lista de búsquedas *lifestore_searches* a DataFrame.

```
1 df_lifestore_searches = pd.DataFrame(lifestore_searches, columns = ['  
    id_search', 'id_product'])  
2 df_lifestore_searches
```

	id_search	id_product
0	1	1
1	2	1
2	3	1
3	4	1
4	5	1
...
1028	1029	94
1029	1030	94
1030	1031	95
1031	1032	95
1032	1033	95

1033 rows × 2 columns

2.3 Análisis

2.3.1 Productos Vendidos

Empezamos el análisis obteniendo los productos más vendidos y menos vendidos de la tienda (sin contar las devoluciones).

```
1 df_sales_product = pd.DataFrame(df_lifestore_sales[df_lifestore_sales['  
    refund'] == 0].groupby('id_product').size()).reset_index()  
2 df_sales_product.columns = ['id_product', 'sales']  
3 df_sales_product = df_sales_product.merge(df_lifestore_products, on = '  
    id_product')  
4 df_sales_product['profit'] = df_sales_product['sales'] *  
    df_sales_product['price']  
5 df_sales_product = df_sales_product[['id_product', 'name', 'sales', '  
    price', 'profit']]  
6 df_sales_product['id_product'] = df_sales_product['id_product'].astype(  
    str)  
7 df_sales_product = df_sales_product.sort_values(by = ['sales', 'profit'],  
    ascending = False).reset_index(drop = True)  
8 df_sales_product['price'] = df_sales_product['price'].map('${:,.2f}').  
    format()  
9 df_sales_product['profit'] = df_sales_product['profit'].map('${:,.2f}').  
    format()  
10 df_sales_product.style.hide_index()
```

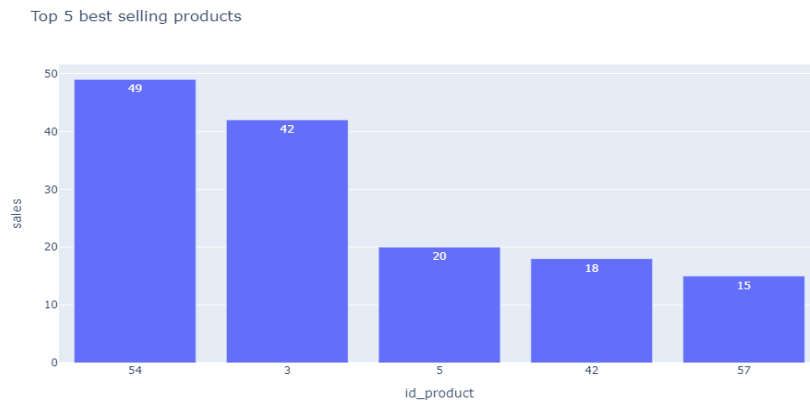
Obtenemos la siguiente tabla, está ordenada del producto más vendido al menos vendido.

id_product	name	sales	price	profit
54	SSD Kingston A400, 120GB, SATA III, 2.5", 7mm	49	\$259.00	\$12,691.00
3	Procesador AMD Ryzen 5 2600, S-AM4, 3.40GHz, Six-Core, 16MB L3 Cache, con Disipador Wraith Stealth	42	\$3,089.00	\$129,738.00
5	Procesador Intel Core i3-9100F, S-1151, 3.60GHz, Quad-Core, 6MB Cache (9na. Generación - Coffee Lake)	20	\$1,779.00	\$35,580.00
42	Tarjeta Madre ASRock Micro ATX B450M Steel Legend, S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD	18	\$1,779.00	\$32,022.00
57	SSD Adata Ultimate SU800, 256GB, SATA III, 2.5", 7mm	15	\$889.00	\$13,335.00
29	Tarjeta Madre ASUS micro ATX TUF B450M-PLUS GAMING, S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD	13	\$2,499.00	\$32,487.00
4	Procesador AMD Ryzen 3 3200G con Gráficos Radeon Vega 8, S-AM4, 3.60GHz, Quad-Core, 4MB L3, con Disipador Wraith Spire	13	\$2,209.00	\$28,717.00
2	Procesador AMD Ryzen 5 3600, S-AM4, 3.60GHz, 32MB L3 Cache, con Disipador Wraith Stealth	12	\$4,209.00	\$50,508.00
47	SSD XPG SX8200 Pro, 256GB, PCI Express, M.2	11	\$1,209.00	\$13,299.00
12	Tarjeta de Video ASUS NVIDIA GeForce GTX 1660 SUPER EVO OC, 6GB 192-bit GDDR6, PCI Express x16 3.0	9	\$6,619.00	\$59,571.00
48	SSD Kingston A2000 NVMe, 1TB, PCI Express 3.0, M2	9	\$2,559.00	\$23,031.00
7	Procesador Intel Core i7-9700K, S-1151, 3.60GHz, 8-Core, 12MB Smart Cache (9na. Generación Coffee Lake)	7	\$8,559.00	\$59,913.00
44	Tarjeta Madre MSI ATX B450 TOMAHAWK MAX, S-AM4, AMD B450, 64GB DDR4 para AMD	6	\$2,759.00	\$16,554.00
18	Tarjeta de Video Gigabyte NVIDIA GeForce GT 1030, 2GB 64-bit GDDR5, PCI Express x16 3.0	5	\$2,199.00	\$10,995.00
8	Procesador Intel Core i5-9600K, S-1151, 3.70GHz, Six-Core, 9MB Smart Cache (9na. Generación - Coffee Lake)	4	\$5,399.00	\$21,596.00
6	Procesador Intel Core i9-9900K, S-1151, 3.60GHz, 8-Core, 16MB Smart Cache (9na. Generación Coffee Lake)	3	\$11,809.00	\$35,427.00
11	Tarjeta de Video ASUS AMD Radeon RX 570, 4GB 256-bit GDDR5, PCI Express 3.0	3	\$7,399.00	\$22,197.00
49	Kit SSD Kingston KC600, 1TB, SATA III, 2.5, 7mm	3	\$3,139.00	\$9,417.00
51	SSD Kingston UV500, 480GB, SATA III, mSATA	3	\$2,399.00	\$7,197.00
31	Tarjeta Madre AORUS micro ATX B450 AORUS M (rev. 1.0), S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD	3	\$2,229.00	\$6,687.00
52	SSD Western Digital WD Blue 3D NAND, 2TB, M.2	2	\$5,659.00	\$11,318.00
25	Tarjeta de Video Sapphire AMD Pulse Radeon RX 5500 XT Gaming, 8GB 128-bit GDDR6, PCI Express 4.0	2	\$5,529.00	\$11,058.00
21	Tarjeta de Video MSI AMD Mech Radeon RX 5500 XT MECH Gaming OC, 8GB 128-bit GDDR6, PCI Express 4.0	2	\$5,159.00	\$10,318.00

Con la función *Head* obtenemos el top 5 de productos más vendidos y los graficamos.

```

1 top_sales = 5
2 fig = px.bar(df_sales_product.head(top_sales), x = "id_product", y = "
   sales", text = 'sales',
3             hover_name = "name", hover_data=["id_product", "sales", "
   price", "profit"],
4             title = "Top " + str(top_sales) + " best selling products")
5 fig.show()
```

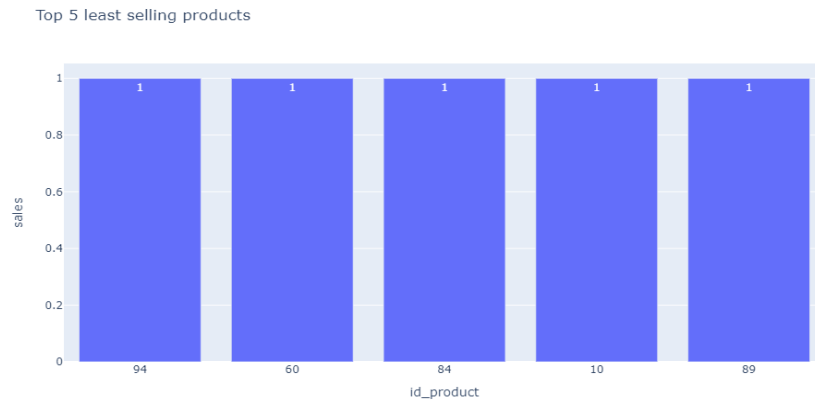


Repetimos el proceso pero ahora con la función *Tail* para obtener los 5 productos menos vendidos.

```

1 top_sales = 5
2 fig = px.bar(df_sales_product.tail(top_sales), x = "id_product", y = "
    sales", text = 'sales',
3             hover_name = "name", hover_data=["id_product", "sales", "
    price", "profit"],
4             title = "Top " + str(top_sales) + " least selling products"
5             )
6 fig.show()

```



2.3.2 Productos Buscados

De manera similar a la sección anterior vamos a obtener el análisis de los productos buscados.

```

1 df_searches_products = pd.DataFrame(df_lifestore_searches.groupby('
    id_product').size()).reset_index()
2 df_searches_products.columns = ['id_product', 'searches']
3 df_searches_products = df_searches_products.merge(df_lifestore_products,
4             on = 'id_product')
5 df_searches_products = df_searches_products[['id_product', 'name', '
    searches']]
6 df_searches_products['id_product'] = df_searches_products['id_product'].
    astype(str)
7 df_searches_products = df_searches_products.sort_values(by = ['searches'
    ], ascending = False).reset_index(drop = True)
8 df_searches_products.style.hide_index()

```

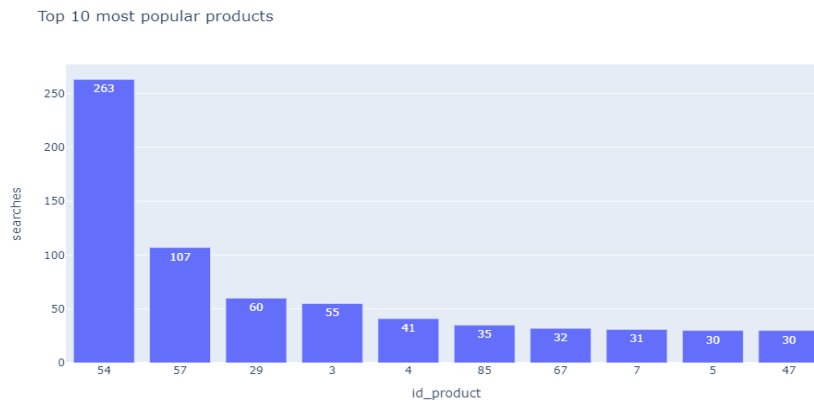
id_product	name	searches
54	SSD Kingston A400, 120GB, SATA III, 2.5", 7mm	263
57	SSD Adata Ultimate SU800, 256GB, SATA III, 2.5", 7mm	107
29	Tarjeta Madre ASUS micro ATX TUF B450M-PLUS GAMING, S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD	60
3	Procesador AMD Ryzen 5 2600, S-AM4, 3.40GHz, Six-Core, 16MB L3 Cache, con Disipador Wraith Stealth	55
4	Procesador AMD Ryzen 3 3200G con Gráficos Radeon Vega 8, S-AM4, 3.60GHz, Quad-Core, 4MB L3, con Disipador Wraith Spire	41
85	Logitech Audífonos Gamer G635 7.1, Alámbrico, 1.5 Metros, 3.5mm, Negro/Azul	35
67	TV Monitor LED 24TL520S-PU 24, HD, Widescreen, HDMI, Negro	32
7	Procesador Intel Core i7-9700K, S-1151, 3.60GHz, 8-Core, 12MB Smart Cache (9na. Generación Coffee Lake)	31
5	Procesador Intel Core i3-9100F, S-1151, 3.60GHz, Quad-Core, 6MB Cache (9na. Generación - Coffee Lake)	30
47	SSD XPG SX8200 Pro, 256GB, PCI Express, M.2	30
48	SSD Kingston A2000 NVMe, 1TB, PCI Express 3.0, M2	27
44	Tarjeta Madre MSI ATX B450 TOMAHAWK MAX, S-AM4, AMD B450, 64GB DDR4 para AMD	25
2	Procesador AMD Ryzen 5 3600, S-AM4, 3.60GHz, 32MB L3 Cache, con Disipador Wraith Stealth	24
42	Tarjeta Madre ASRock Micro ATX B450M Steel Legend, S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD	23
8	Procesador Intel Core i5-9600K, S-1151, 3.70GHz, Six-Core, 9MB Smart Cache (9na. Generación - Coffee Lake)	20
66	TCL Smart TV LED 55S425 54.6, 4K Ultra HD, Widescreen, Negro	15
21	Tarjeta de Video MSI AMD Mech Radeon RX 5500 XT MECH Gaming OC, 8GB 128-bit GDDR6, PCI Express 4.0	15
12	Tarjeta de Video ASUS NVIDIA GeForce GTX 1660 SUPER EVO OC, 6GB 192-bit GDDR6, PCI Express x16 3.0	15
51	SSD Kingston UV500, 480GB, SATA III, mSATA	11
18	Tarjeta de Video Gigabyte NVIDIA GeForce GT 1030, 2GB 64-bit GDDR5, PCI Express x16 3.0	11
40	Tarjeta Madre Gigabyte XL-ATX TRX40 Designare, S-sTRX4, AMD TRX40, 256GB DDR4 para AMD	10

Nuevamente graficamos el top 10 de productos más buscados y menos buscados.

```

1 top_searches = 10
2 fig = px.bar(df_searches_products.head(top_searches), x = "id_product",
3             y = "searches", text = 'searches',
4             hover_name = "name", hover_data=["id_product", "searches"],
5             title = "Top " + str(top_searches) + " most popular products")
6 fig.show()

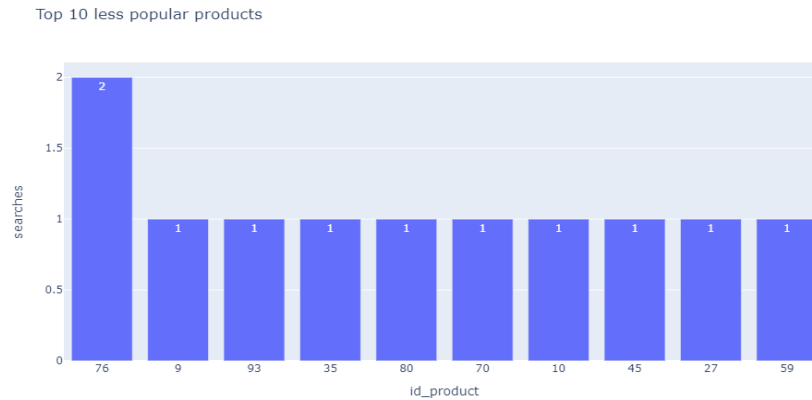
```



```

1 top_searches = 10
2 fig = px.bar(df_searches_products.tail(top_searches), x = "id_product",
3             y = "searches", text = 'searches',
4             hover_name = "name", hover_data=["id_product", "searches"],
5             title = "Top " + str(top_searches) + " less popular
           products")
6 fig.show()

```



2.3.3 Calificación de Productos

Repetimos los pasos anteriores para ahora analizar las calificaciones de los productos vendidos.

```

1 df_rated_product = pd.DataFrame(df_lifestore_sales[['id_product', 'score']]
2                               .groupby('id_product').mean()).reset_index()
3 df_rated_product = df_rated_product.merge(df_lifestore_products, on = '
4     id_product')
5 df_refund = pd.DataFrame(df_lifestore_sales.groupby('id_product').sum()
6                          .reset_index())
7 df_rated_product = df_rated_product.merge(df_refund[['id_product', '
8     refund']], on = 'id_product')
9 df_rated_product['id_product'] = df_rated_product['id_product'].astype(
10    str)
11 df_rated_product = df_rated_product.merge(df_sales_product[['id_product',
12    'sales']], on = 'id_product')
13 df_rated_product = df_rated_product[['id_product', 'name', 'sales', '
14    refund', 'score']]
15 df_rated_product = df_rated_product.sort_values(by = ['score', 'sales'],
16          ascending = False).reset_index(drop = True)
17 df_rated_product.style.hide_index()

```

id_product	name	sales	refund	score
7	Procesador Intel Core i7-9700K, S-1151, 3.60GHz, 8-Core, 12MB Smart Cache (9na. Generación Coffee Lake)	7	0	5.000000
8	Procesador Intel Core i5-9600K, S-1151, 3.70GHz, Six-Core, 9MB Smart Cache (9na. Generación - Coffee Lake)	4	0	5.000000
6	Procesador Intel Core i9-9900K, S-1151, 3.60GHz, 8-Core, 16MB Smart Cache (9na. Generación Coffee Lake)	3	0	5.000000
11	Tarjeta de Video ASUS AMD Radeon RX 570, 4GB 256-bit GDDR5, PCI Express 3.0	3	0	5.000000
49	Kit SSD Kingston KC600, 1TB, SATA III, 2.5, 7mm	3	0	5.000000
1	Procesador AMD Ryzen 3 3300X S-AM4, 3.80GHz, Quad-Core, 16MB L2 Cache	2	0	5.000000
21	Tarjeta de Video MSI AMD Mech Radeon RX 5500 XT MECH Gaming OC, 8GB 128-bit GDDR6, PCI Express 4.0	2	0	5.000000
25	Tarjeta de Video Sapphire AMD Pulse Radeon RX 5500 XT Gaming, 8GB 128-bit GDDR6, PCI Express 4.0	2	0	5.000000
52	SSD Western Digital WD Blue 3D NAND, 2TB, M.2	2	0	5.000000
85	Logitech Audifonos Gamer G635 7.1, Alámbrico, 1.5 Metros, 3.5mm, Negro/Azul	2	0	5.000000
22	Tarjeta de Video MSI NVIDIA GeForce GTX 1050 Ti OC, 4GB 128-bit GDDR5, PCI Express x16 3.0	1	0	5.000000
28	Tarjeta de Video Zotac NVIDIA GeForce GTX 1660 Ti, 6GB 192-bit GDDR6, PCI Express x16 3.0	1	0	5.000000
40	Tarjeta Madre Gigabyte XL-ATX TRX40 Designare, S-sTRX4, AMD TRX40, 256GB DDR4 para AMD	1	0	5.000000
50	SSD Crucial MX500, 1TB, SATA III, M.2	1	0	5.000000
60	Kit Memoria RAM Corsair Dominator Platinum DDR4, 3200MHz, 16GB (2x 8GB), Non-ECC, CL16, XMP	1	0	5.000000
66	TCL Smart TV LED 55S425 54.6, 4K Ultra HD, Widescreen, Negro	1	0	5.000000
67	TV Monitor LED 24TL520S-PU 24, HD, Widescreen, HDMI, Negro	1	0	5.000000
84	Logitech Audifonos Gamer G332, Alámbrico, 2 Metros, 3.5mm, Negro/Rojo	1	0	5.000000
57	SSD Adata Ultimate SU800, 256GB, SATA III, 2.5", 7mm	15	0	4.866667
3	Procesador AMD Ryzen 5 2600, S-AM4, 3.40GHz, Six-Core, 16MB L3 Cache, con Disipador Wraith Stealth	42	0	4.809524
12	Tarjeta de Video ASUS NVIDIA GeForce GTX 1660 SUPER EVO OC, 6GB 192-bit GDDR6, PCI Express x16 3.0	9	0	4.777778
54	SSD Kingston A400, 120GB, SATA III, 2.5", 7mm	49	1	4.720000

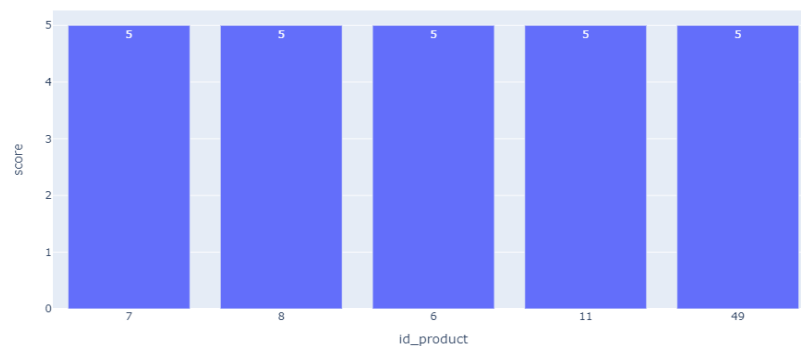
Obtenemos las gráficas de los productos con mejor calificación y los de peor calificación.

```

1 top_rated = 5
2 fig = px.bar(df_rated_product.head(top_rated), x = "id_product", y = "
   score", text = 'score',
3             hover_name = "name", hover_data=["id_product", "sales", "
   refund", "score"],
4             title = "Top " + str(top_searches) + " best rated products"
5             )
fig.show()

```

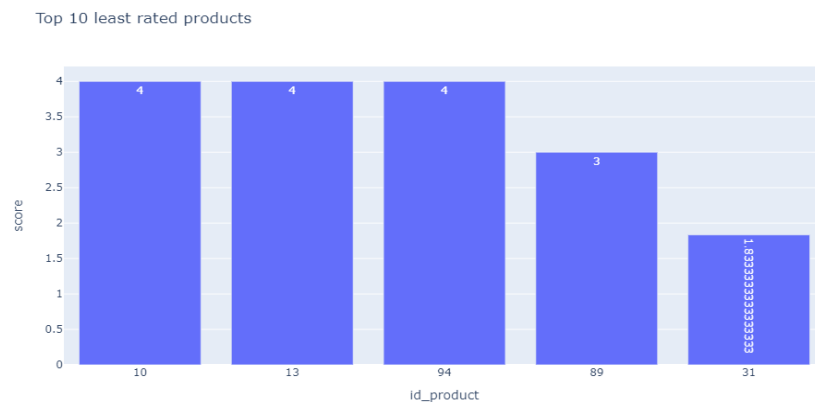
Top 10 best rated products




```

1 top_rated = 5
2 fig = px.bar(df_rated_product.tail(top_rated), x = "id-product", y = "
   score", text = 'score',
3             hover_name = "name", hover_data=["id-product", "sales", "
   refund", "score"],
4             title = "Top " + str(top_searches) + " best rated products"
5             )
6 fig.show()

```



2.3.4 Ganancias Anuales

Por último, obtenemos las ganancias que se tuvieron cada mes.

```

1 df_sales = pd.DataFrame(df_lifystore_sales[df_lifystore_sales['refund']
   == 0])
2 df_sales = df_sales.merge(df_lifystore_products, on = 'id-product')
3 df_sales_month = pd.DataFrame(df_sales[['month', 'price']].groupby('
   month').sum()).reset_index()
4 df_sales_month.columns = ['month', 'profit']
5 fig = px.line(df_sales_month, x = "month", y = "profit", text = "profit"
6              )
7 fig.show()

```



3 Solución al Problema

Gracias al análisis hecho podemos notar que Abril ha sido el mejor mes hasta hoy, de ahí ha venido una baja en las ventas. Mi propuesta es sacar de la venta la *Tarjeta Madre AORUS micro ATX B450 AORUS M (rev. 1.0), S-AM4, AMD B450, HDMI, 64GB DDR4 para AMD* (id 31) ya que de tres ventas las tres han sido devueltas y es el producto con peor calificación.

También promover más la venta de *Procesador AMD Ryzen 5 2600, S-AM4, 3.40GHz, Six-Core, 16MB L3 Cache, con Disipador Wraith Stealth* (id 3), porque es uno de los productos más vendidos y es el que más ganancia nos ha generado.

4 Conclusión

La programación hoy en día es una herramienta cada vez más utilizada. Python es un programa que brinda muchísimos paquetes para diferentes cosas, desde desarrollo web a Inteligencia Artificial, por lo que creo yo aprender este lenguaje es una gran ventaja y le podemos sacar mucho provecho.