Python for Business Analytics - Spring 2024, Final Project Report
Instructor: Alec Malstrom
Team: Jeremy Perez and Shajibul Islam

Prompt:

I am your VP of Sales and have presented you with the following question: I am looking to improve this year's performance – ***what suggestions do you have for success?***

Data Source:

Conducted data analysis on largest of data using the database provided that contained the following tables:

CustomerData:

```
1   SELECT *
2   FROM CustomerData
3
```

Results Messages

	column1	Customer_City	Customer_Country	Customer_Email	Customer_Fname	Customer_Id
3	2	San Jose	EE. UU.	XXXXXXXXX	Gillian	19491
4	3	Los Angeles	EE. UU.	XXXXXXXXX	Tana	19490
5	4	Caguas	Puerto Rico	XXXXXXXXX	Orli	19489
6	5	Tonawanda	EE. UU.	XXXXXXXXX	Kimberly	19488
7	6	Caguas	Puerto Rico	XXXXXXXXX	Constance	19487
8	7	Miami	EE. UU.	XXXXXXXXX	Erica	19486
9	8	Caguas	Puerto Rico	XXXXXXXXX	Nichole	19485
10	9	San Ramon	EE. UU.	XXXXXXXXX	Oprah	19484
11	10	Caguas	Puerto Rico	XXXXXXXXX	Germane	19483
12	11	Freeport	EE. UU.	XXXXXXXXX	Freya	19482
13	12	Salinas	EE. UU.	XXXXXXXXX	Cassandra	19481
14	13	Caguas	Puerto Rico	XXXXXXXXX	Natalie	19480

⚠ 0 Ln 3, Col 1 Spaces: 4 UTF-8 LF 20,652 rows MSSQL 00:00:01 johndroescher.com : Fall_2023 (60)

DataDictionary:

```
1   SELECT *
2   FROM DataDictionary
3
```

Results Messages

	FIELDS	DESCRIPTION
1	Type	: Type of transaction made
2	Days for shipping (real)	: Actual shipping days of the purchased product
3	Days for shipment (scheduled)	: Days of scheduled delivery of the purchased produ…
4	Benefit per order	: Earnings per order placed
5	Sales per customer	: Total sales per customer made per customer
6	Delivery Status	: Delivery status of orders: Advance shipping , Lat…
7	Late_delivery_risk	: Categorical variable that indicates if sending is…
8	Category Id	: Product category code
9	Category Name	: Description of the product category
10	Customer City	: City where the customer made the purchase
11	Customer Country	: Country where the customer made the purchase
12	Customer Email	: Customer's email
13	Customer Fname	: Customer name

Results grid

⚠ 0 Ln 2, Col 20 Spaces: 4 UTF-8 LF 52 rows MSSQL 00:00:00 johndroescher.com : Fall_2023 (60)

MainTable:

```
1  SELECT *
2  FROM MainTable
3
```

Results Messages

	column1	Category_Id	Customer_Id	Department_Id	Order_Customer_Id	Order_Id	Order_Ite
1	0	73	20755	2	20755	77202	1360
2	1	73	19492	2	19492	75939	1360
3	2	73	19491	2	19491	75938	1360
4	3	73	19490	2	19490	75937	1360
5	4	73	19489	2	19489	75936	1360
6	5	73	19488	2	19488	75935	1360
7	6	73	19487	2	19487	75934	1360
8	7	73	19486	2	19486	75933	1360
9	8	73	19485	2	19485	75932	1360
10	9	73	19484	2	19484	75931	1360
11	10	73	19483	2	19483	75930	1360
12	11	73	19482	2	19482	75929	1360

⚠ 0 Ln 2, Col 15 Spaces: 4 UTF-8 LF 180,519 rows MSSQL 00:00:05 johndroescher.com : Fall_2023 (60)

OrderData:

```
1  SELECT *
2  FROM OrderData
3
```

Results Messages

	column1	Benefit_per_order	Order_City	Order_Country	Order_Customer_Id	order_date_DateOrd
1	0	91.25	Bekasi	Indonesia	20755	2018-01-31 22:5
2	1	-249.08999633789062	Bikaner	India	19492	2018-01-13 12:2
3	2	-247.77999877929688	Bikaner	India	19491	2018-01-13 12:0
4	3	22.860000610351562	Townsville	Australia	19490	2018-01-13 11:4
5	4	134.2100067138672	Townsville	Australia	19489	2018-01-13 11:2
6	5	18.579999923706055	Toowoomba	Australia	19488	2018-01-13 11:0
7	6	95.18000030517578	Guangzhou	China	19487	2018-01-13 10:4
8	7	68.43000030517578	Guangzhou	China	19486	2018-01-13 10:2
9	8	133.72000122070312	Guangzhou	China	19485	2018-01-13 10:0
10	9	132.14999389648438	Guangzhou	China	19484	2018-01-13 09:3
11	10	130.5800018310547	Tokio	Japón	19483	2018-01-13 09:1
12	11	45.689998626708984	Manado	Indonesia	19482	2018-01-13 08:5

⚠ 0 Ln 2, Col 15 Spaces: 4 UTF-8 LF 180,519 rows MSSQL 00:00:08 johndroescher.com : Fall_2023 (60)

ProductData:

```
1  SELECT *
2  FROM ProductData
3
```

Results Messages

	column1	Category_Id	Category_Name	Product_Card_Id	Product_Category_Id	Product_Descripti
1	0	73	Sporting Goods	1360	73	NULL
2	1	17	Cleats	365	17	NULL
3	2	29	Shop By Sport	627	29	NULL
4	3	24	Women's Apparel	502	24	NULL
5	4	13	Electronics	278	13	NULL
6	5	12	Boxing & MMA	249	12	NULL
7	6	9	Cardio Equipment	191	9	NULL
8	7	41	Trade-In	917	41	NULL
9	8	37	Electronics	828	37	NULL
10	9	29	Shop By Sport	642	29	NULL
11	10	37	Electronics	818	37	NULL
12	11	37	Electronics	825	37	NULL

⚠ 0 Ln 2, Col 17 Spaces: 4 UTF-8 LF 118 rows MSSQL 00:00:00 johndroescher.com : Fall_2023 (60)

ShippingData:

```
1  SELECT *
2  FROM ShippingData
3
```

Results Messages

	column1	Days_for_shipping_real	Days_for_shipment_scheduled	Delivery_Status	Late_delivery_ris
1	0	3	4	Advance shipping	0
2	1	5	4	Late delivery	1
3	2	4	4	Shipping on time	0
4	3	3	4	Advance shipping	0
5	4	2	4	Advance shipping	0
6	5	6	4	Shipping canceled	0
7	6	2	1	Late delivery	1
8	7	2	1	Late delivery	1
9	8	3	2	Late delivery	1
10	9	2	1	Late delivery	1
11	10	6	2	Shipping canceled	0
12	11	5	2	Late delivery	1

⚠ 0 Ln 2, Col 18 Spaces: 4 UTF-8 LF 65,752 rows MSSQL 00:00:01 johndroescher.com : Fall_2023 (60)

StoreData:

```
1  SELECT *
2  FROM StoreData
3
```

Results Messages

	column1	Department_Id	Department_Name	Latitude	Longitude
1	0	2	Fitness	18.251453399658203	-66.03705596923828
2	1	2	Fitness	18.279451370239258	-66.03706359863281
3	2	2	Fitness	37.292232513427734	-121.88127899169922
4	3	2	Fitness	34.125946044921875	-118.291015625
5	4	2	Fitness	18.253768920898438	-66.03704833984375
6	5	2	Fitness	43.01396942138672	-78.87906646728516
7	6	2	Fitness	18.242538452148438	-66.03705596923828
8	7	2	Fitness	25.928869247436523	-80.16287231445312
9	8	2	Fitness	18.23322296142578	-66.03705596923828
10	9	2	Fitness	37.773990631103516	-121.96662902832031
11	10	2	Fitness	18.28284454345703	-66.03705596923828
12	11	2	Fitness	40.65486526489258	-73.58707427978516
13	12	2	Fitness	36.67633819580078	-121.65651702880086

⚠ 0 Ln 2, Col 15 Spaces: 4 UTF-8 LF 41,956 rows MSSQL 00:00:00 johndroescher.com : Fall_2023 (60)

From these tables, I queried the data to find patterns or inconsistency:

- Shows the total sales, profits and the quantity of each product sold from every order made from 2015 to 2018 by the product's name and its category:

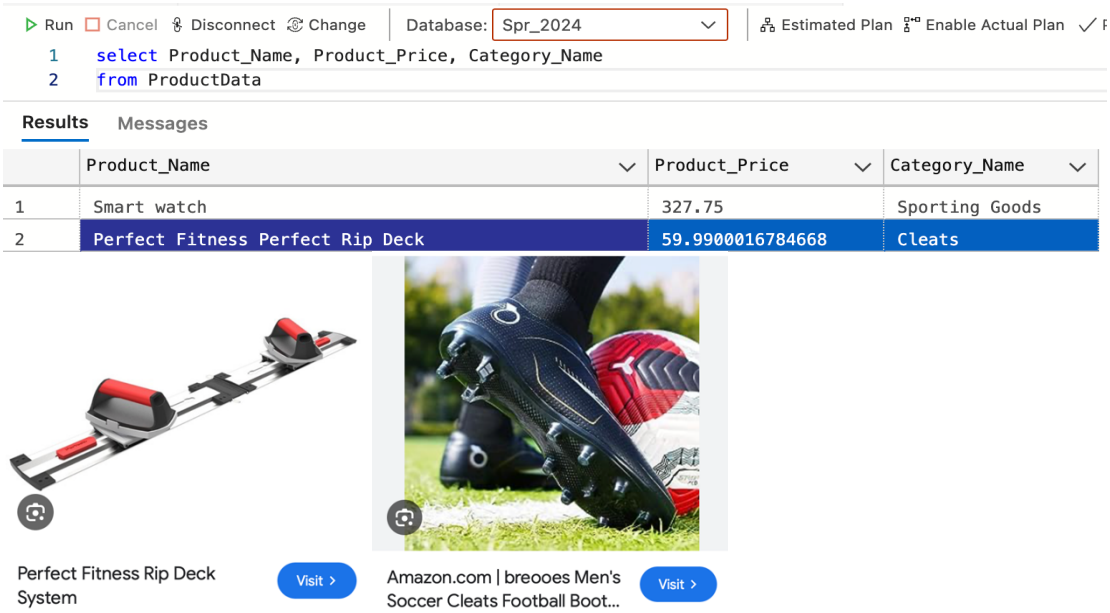
```
select CONVERT (DATE, order_date_DateOrders) as Date_of_order,
       Product_name, Category_Name,
       sum(Sales) as sum_sales, Order_Profit_Per_Order, Order_Item_Quantity

from ProductData

-- Used INNER/LEFT JOIN to attach the fields/columns from MainTable to ProductData table using the same column, Category_Id
left join dbo.MainTable on dbo.ProductData.Category_Id = dbo.MainTable.Category_Id

-- Used OUTER JOIN to  attach the fields/columns from OrderData to MainTable using the same column, Order_Id
full outer join dbo.OrderData on dbo.MainTable.Order_Id = dbo.OrderData.Order_Id

GROUP by Product_Name,Category_Name, order_date_DateOrders,Order_Profit_Per_Order, Order_Item_Quantity

ORDER by order_date_DateOrders desc
```

Results Messages

	Date_of_order	Product_name	Category_Name	sum_sales	Order_Profit_Per_Order	Order_Item_Quantity
1	2018-01-31	Fighting video games	Video Games	39.75	9.789999962	1
2	2018-01-31	Toys	Toys	11.539999961853027	1.529999971	1
3	2018-01-31	Smart watch	Sporting Goods	327.75	91.25	1
4	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-126.5599976	1
5	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-107.9599991	1
6	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	103.5899963	1
7	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	70.51000214	1
8	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	21.14999962	1
9	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-48.77999878	1
10	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-51.79999924	1
11	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	71.76000214	1
12	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	71.37999725	1
13	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-301.0700073	1
14	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	-20.22999954	1
15	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	17.09000015	1
16	2018-01-31	Summer dresses	Women's Clothing	215.82000732421875	89.26000214	1

Data Restrictions:

- Upon querying the data, I discovered that for the year 2018, the data is only limited to the month of January to quantify yearly performance and compare it to previous years. Whereas for previous years, 2015 to 2017 had data for all months of the year.
- The data in certain columns do not match. For instance, in the table ProductData, most of the products in the column Product_Name were categorized incorrectly with the values in the column Category_Name. Which leads to questioning the accuracy of the entire dataset. This leaves room for error if we were to calculate the total number of products and group them by category.



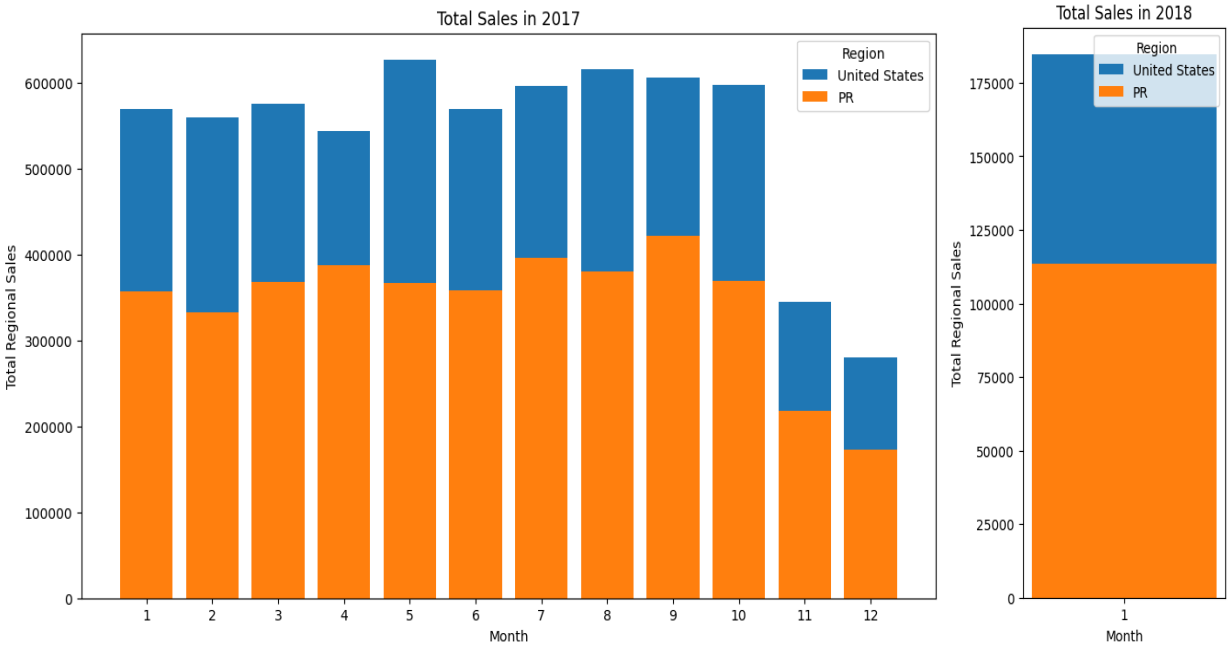
Data Preparation:

- I attempted to clean the data within the database but did not have permission granted to do so. Therefore, using python's Panda library to manipulate the data, I used the loc function to update the Category name based on the condition; if a product's name is present in the data frame, product_type, change the value in the column Category_Name for that product name.
- Using the SQL database, I queried the necessary data for the analysis. Then converted the queried data into a csv file and utilizing pandas, a python library, read the csv file into a pandas Data Frame. Using pandas, I performed functions to modify the columns, such as merge and drop columns.
- Lastly, I utilized python to make additional aggregations and create visualizations (bar graphs, pie charts and line graphs) using the matplotlib and seaborn python libraries.

Data Analysis:

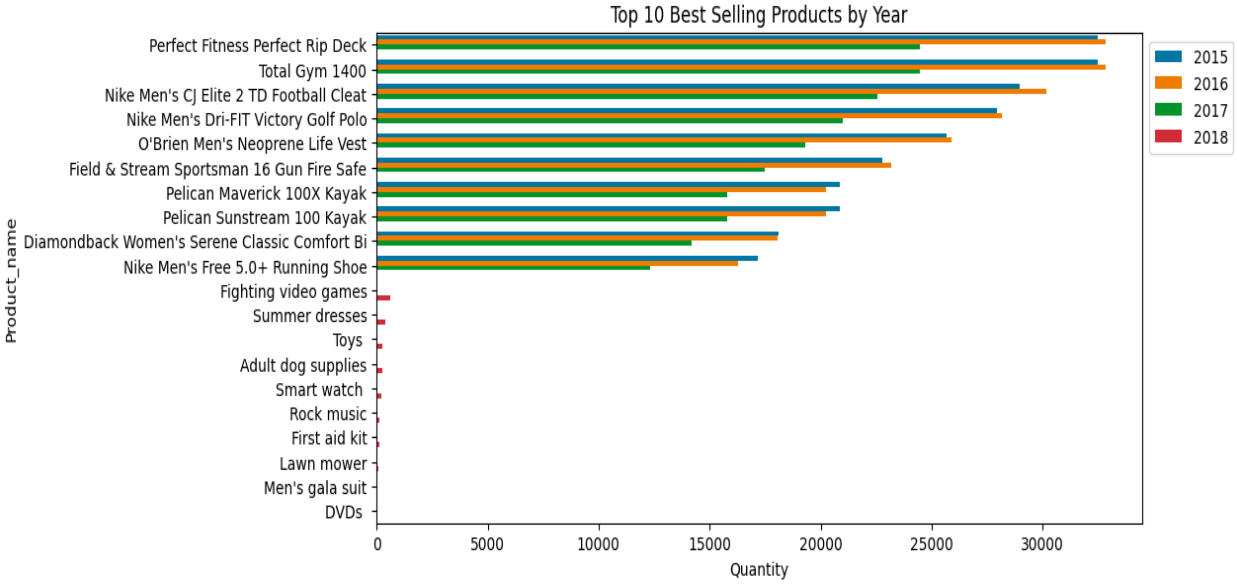
1. **Compared the sales data. Notice a drastic change in sales from 2017 to 2018**

Year	Total Sales
2015	$11,089,543.55
2016	$11,055,996.76
2017	$10,610,910.01
2018	$297,952.05



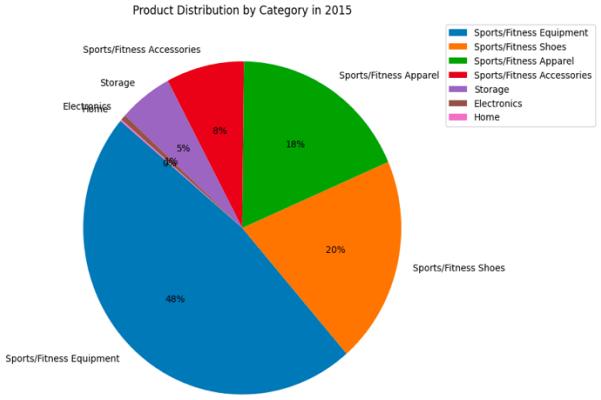
- Sales in the United States are significantly higher than those in Puerto Rico throughout the year. (USA is a country while Puerto Rico is territory owned by the United States compared to each state Puerto Rico is significantly higher)
- The total sales for both regions combined reach their peak in May and remain relatively high in all the other months.
- The lowest combined sales started to occur in November, with a notably sharp decrease, especially in the United States, and continued to decrease till the end of the year and the beginning of the New year.
- Although the total sales values fluctuate monthly, the proportion of sales from Puerto Rico to the United States remains relatively consistent. This indicates that any factors affecting sales trends will impact both regions similarly

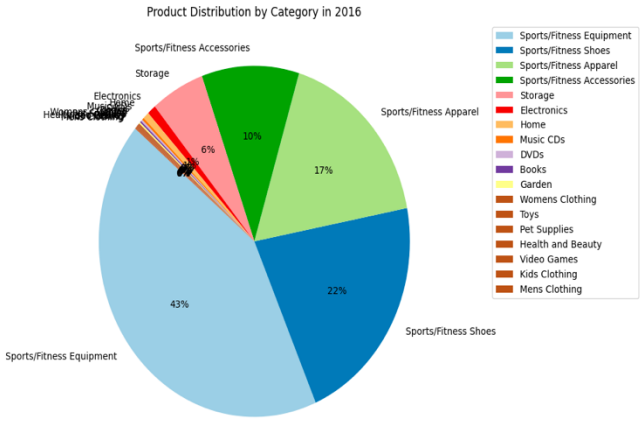
2. **Identified the ten best-selling products for each year in the United States and Puerto Rico.**

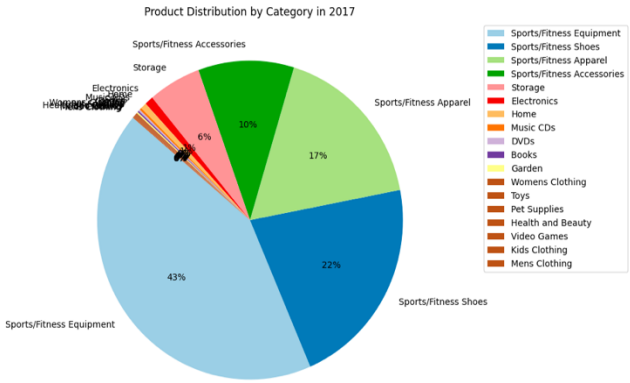


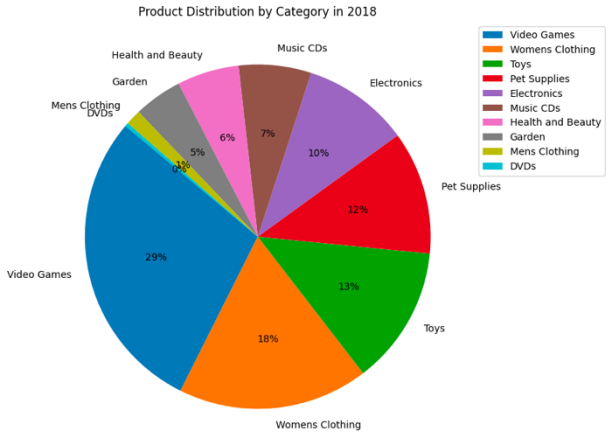
- From 2015 to 2016 the quantity of the top selling products slightly increased and plummeted in 2017.
- The decrease in sales for 2017, can be attributed to the rise of competitors that sell a variety of products such as Costco and Amazon.

3. The product distribution is based on the product's category sold for each year:



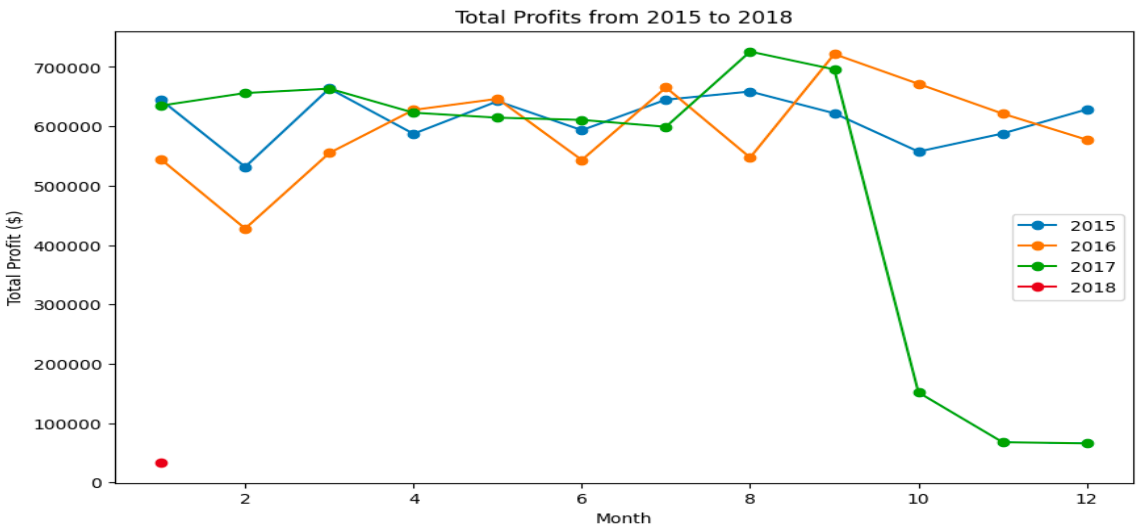






- In 2016, the company transitioned from only selling sports/fitness-related goods to offering new and different product types.
- The product distribution from 2016 was the same for 2017.
    - The new products are represented in the obscure clustered labels with low percentages meaning low sales were made from those products.
- In 2018, the company no longer made any sold sports/fitness related products but continued to sell the new miscellaneous products instead.

4. **The decline in sales in 2017 and external factors that were inevitable greatly affected the company's profits from October 2017 to January 2018.**



Why?

- Puerto Rico was the company's best source of revenue. Although, "Hurricane Irma passed Puerto Rico on September 6, 2017, causing an estimated $1 billion in damages... Two weeks later, on September 20, Hurricane Maria, the largest hurricane to hit Puerto Rico... The storms have had a tremendous impact on the Puerto Rican population beyond loss of life, injuries and destruction. They have caused major disruption in economic life, manufacturing, and research activities" ***(Welton, 2020)***.
- The aftermath affected many sectors, particularly the retail sector, including a noticeable decline in sporting goods sales
- Residents prioritized recovery and rebuilding, diverting spending away from non-essential items such as sporting goods and shifted focus to immediate needs like food, water, and essential needed to help them get through this struggle.
- Since Puerto Rico to the United States sales are relatively consistent, because of the hurricane affecting sales in Puerto Rico it also affected sales performance in the USA.

Suggestion

The company should return to selling sports/fitness products again. Focusing on the products that were most popular in the US. The company can also try to target new customers from nearby Caribbean countries such as the Dominican Republic.

References:

Welton, Michael et al. “Impact of Hurricanes Irma and Maria on Puerto Rico Maternal and Child Health Research Programs.” *Maternal and child health journal* vol. 24,1 (2020): 22-29. doi:10.1007/s10995-019-028242 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7059554/#:~:text=Hurricane%20Irma%20passed%20Puerto%20Rico,3%20fatalities%20in%20Puerto%20Rico.

https://www.rand.org/pubs/research_reports/RR2600.html

https://www.rand.org/hsrd/hsoac/projects/puerto-rico-recovery/hurricanes-irma-and-maria.html