

Python for Business Analytics, Spring 2024, Final Project

Instructor: Alec Malstrom

Team: Jeremy Perez, 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:

We conducted our analysis using the database provided that contained the following tables, CustomerData, DataDictionary, MainTable, OrderData, ProductData, ShippingData and StoreData.

Data Limitations:

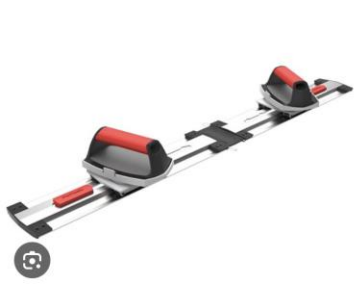
- For 2018, the data is limited to 1 month to quantify yearly performance and compare it to previous years. Whereas previous years, 2015 to 2017 had data from Jan. To Dec.
- The data in certain columns do not match. For instance, in the table ProductData, most of the products in the column Product_Name 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.

Run Cancel Disconnect Change Database: Spr_2024 Estimated Plan Enable Actual Plan ✓ f

```
1 select Product_Name, Product_Price, Category_Name
2 from ProductData
```


Results Messages

	Product_Name	Product_Price	Category_Name
1	Smart watch	327.75	Sporting Goods
2	Perfect Fitness Perfect Rip Deck	59.9900016784668	Cleats



Perfect Fitness Rip Deck System

[Visit >](#)



Amazon.com | breooes Men's Soccer Cleats Football Boot...

[Visit >](#)

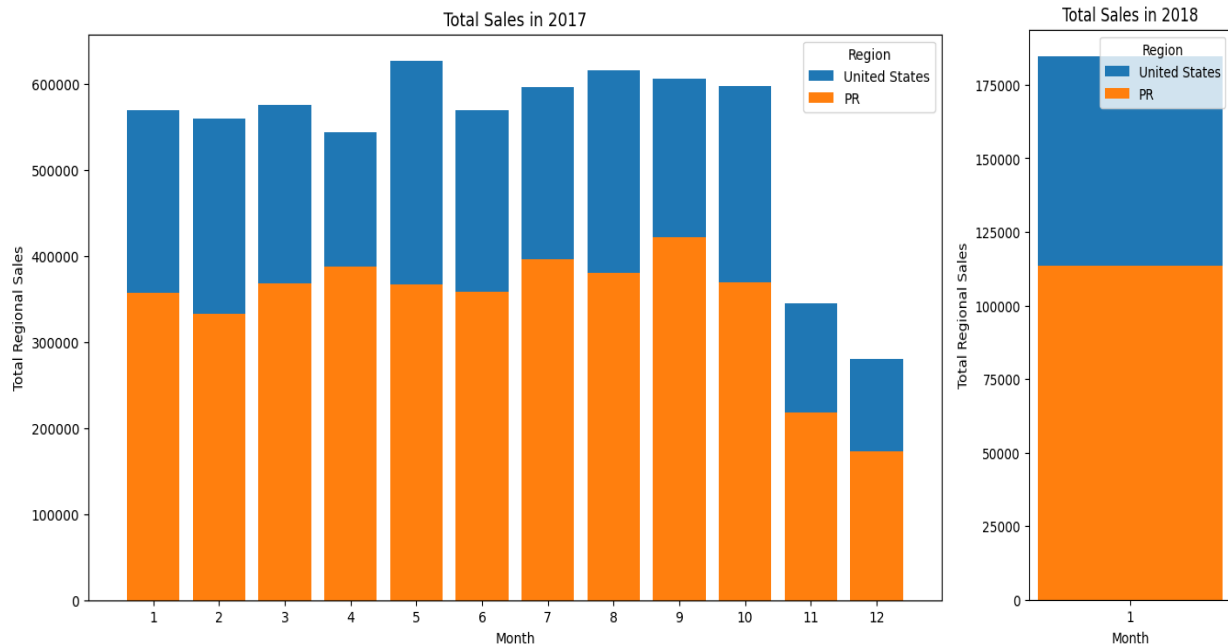
Data Preparation:

- I attempted to clean the data within the database but did not have permission granted to do so. Therefore, using python and panda's library, 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, we conducted queries for the necessary data for our analysis. We 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, we could perform functions to modify the columns, such as merge and drop columns.
- Lastly, we utilized python to make additional aggregations and create visualizations (bar and line graphs) using the matplotlib and seaborn python libraries.

Data Analysis:

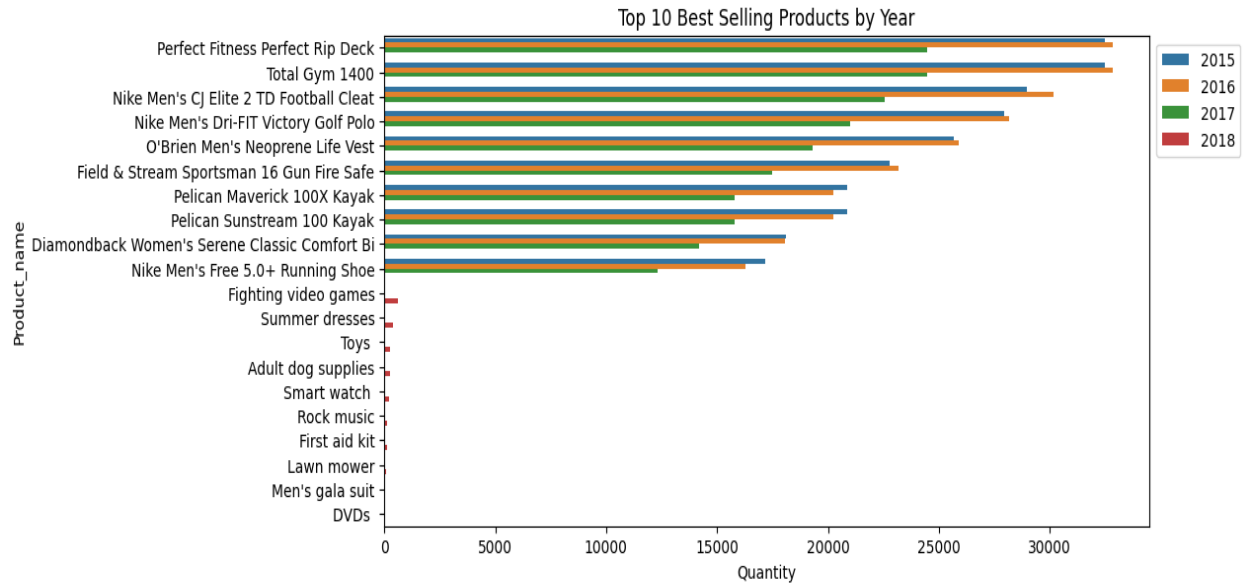
1. We reframed the question: *to give suggestions to improve the company's performance.*
- 2.

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



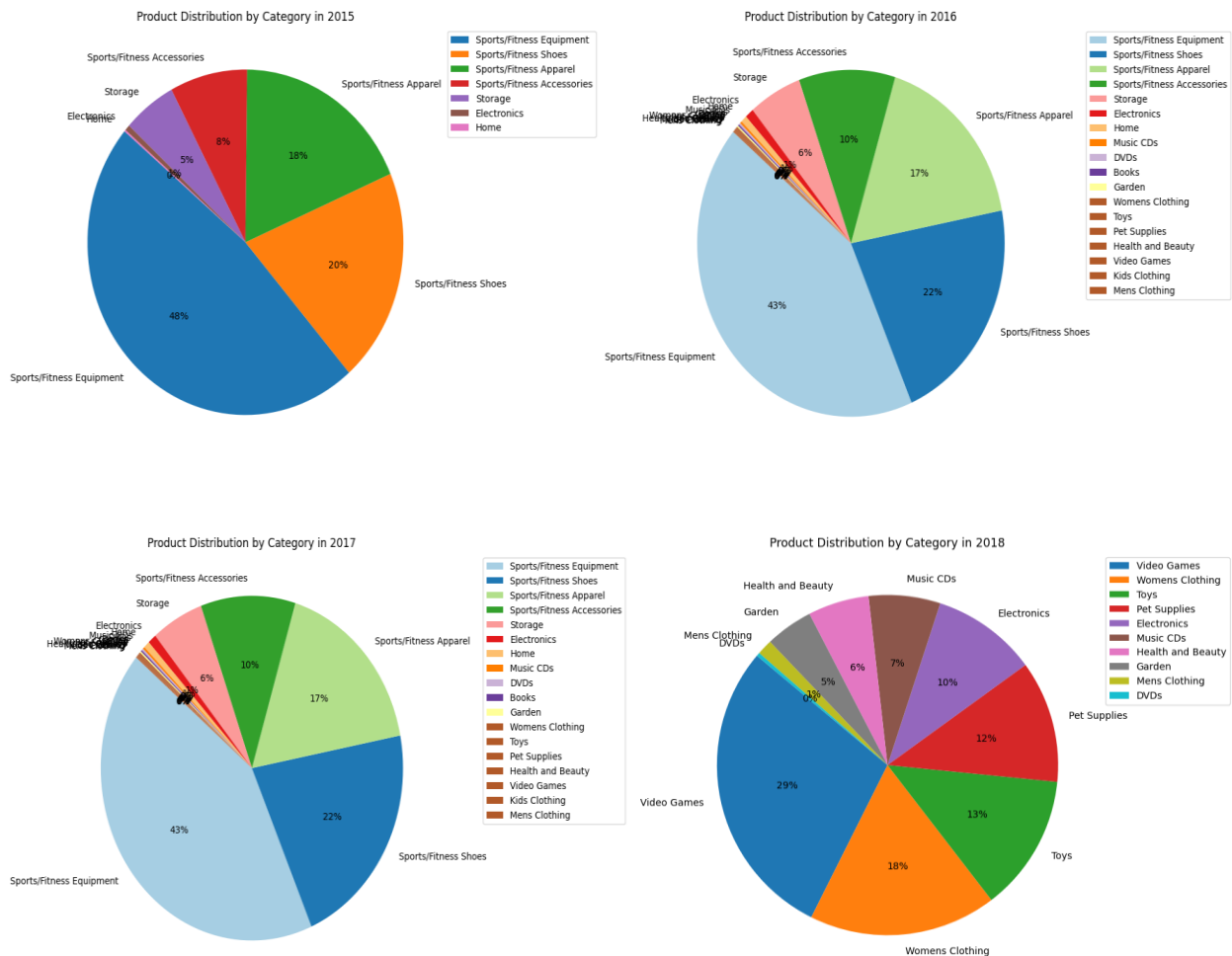
- Sales in the United States are significantly higher than those in Puerto Rico throughout the year. (USA is a country while Puerto Rico is territory owned by the United State, compared to each state Puerto Rico is significantly higher)
- The total sales for both regions combined reach their peak in 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

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



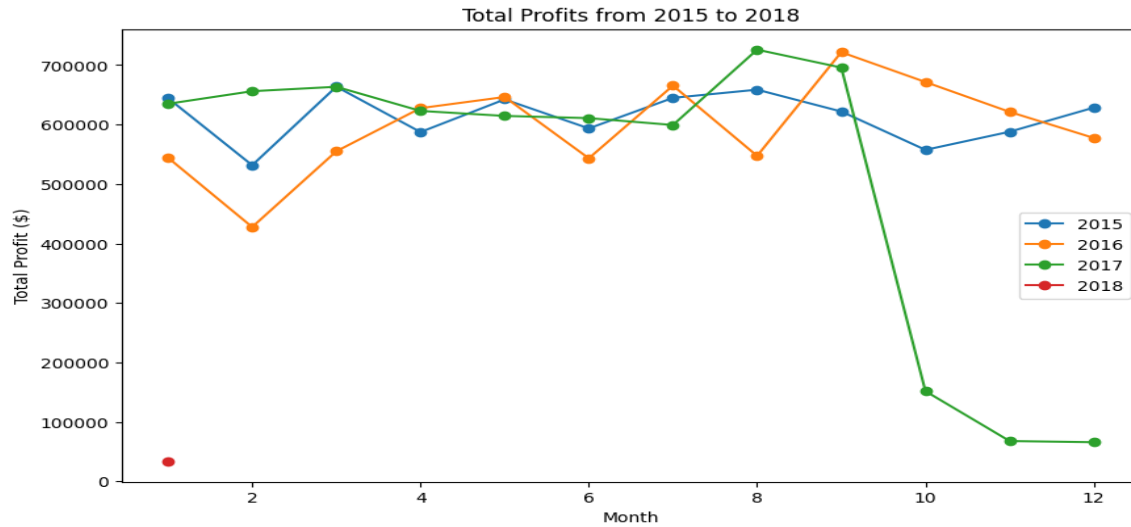
- From 2015 to 2016 the quantity of the top selling products slightly increased and plummeted in 2017.
- The decrease in sales for 2017, could be attributed to the rise of competitors that sell a variety of products such as Costco and Amazon.

4. 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.
 - Most of the new products are represented in the obscure cluster with low percentages meaning very 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.

5. 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.

Solution

The company should start selling sports/fitness products again.

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>

Sale_Year	Sale_Month	Category_Name
2017	11	Garden
2017	11	Crafts
2017	11	Music
2017	11	Health and Beauty
2017	11	DVDs
2017	11	Consumer Electronics
2017	11	Men's Clothing
2017	11	Pet Supplies
2017	12	Computers
2017	12	Crafts
2017	12	Women's Clothing
2017	12	Sporting Goods
2017	12	Children's Clothing
2017	12	Consumer Electronics
2017	12	DVDs
2017	12	Pet Supplies
2017	12	Video Games
2017	12	Cameras
2017	12	Baby
2017	12	Books
2017	12	Toys
2017	12	CDs

Sale_Year	Sale_Month	Category_Name
2018	1	Women's Clothing
2018	1	Sporting Goods
2018	1	Garden
2018	1	Music
2018	1	Health and Beauty
2018	1	Video Games
2018	1	Pet Supplies
2018	1	Men's Clothing
2018	1	Toys
2018	1	DVDs