Python for Business Analytics - Spring 2024, Final Project
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 the analysis using the database provided that contained the following tables, CustomerData, DataDictionary, MainTable, OrderData, ProductData, ShippingData and StoreData. Each with thousands of rows of data.
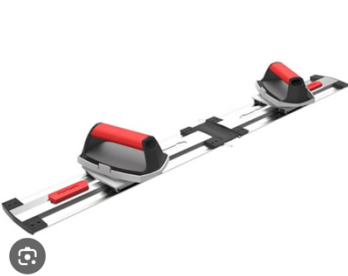
Data Limitations:

- For 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. Notice a drastic change in sales from 2017 to 2018
2. Compared the sales data

Year	Total Sales
2015	$11,089,543.55
2016	$11,055,996.76
2017	$10,610,910.01
2018	$297,952.05
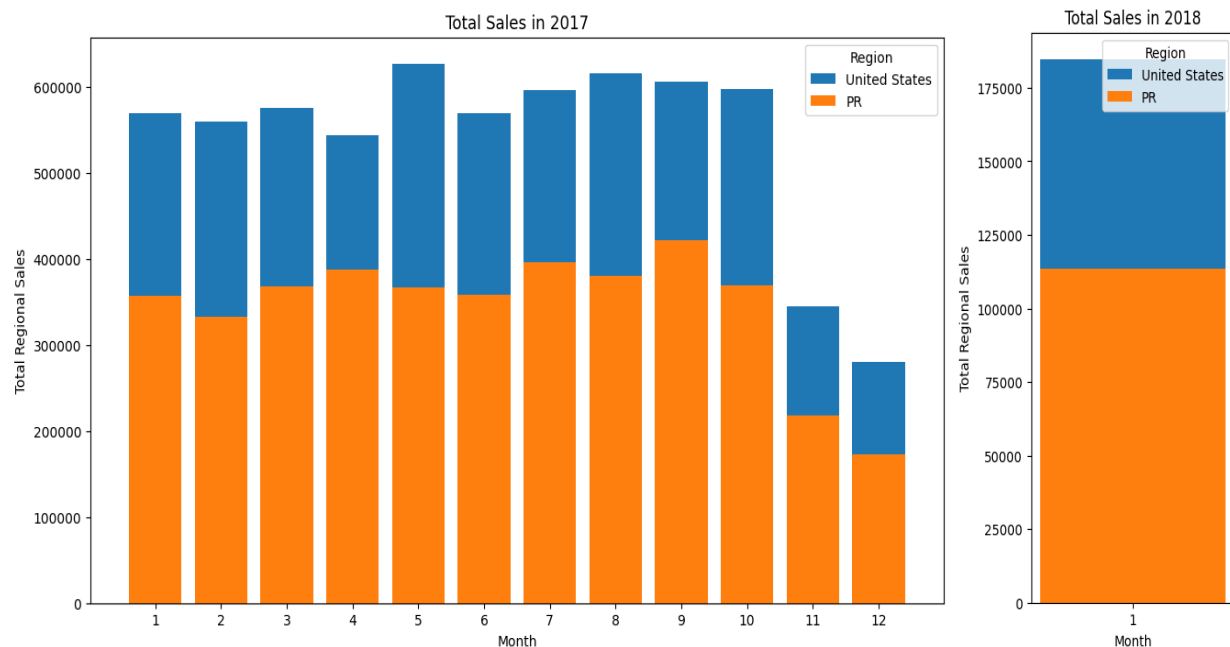
- Sales in the United States are significantly higher than those in Puerto Rico throughout the year. (USA is a country while Puerto Rico is territory owned by the United State,s compared to each state Puerto Rico is significantly higher)

- 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.

Top 10 Best Selling Products by Year
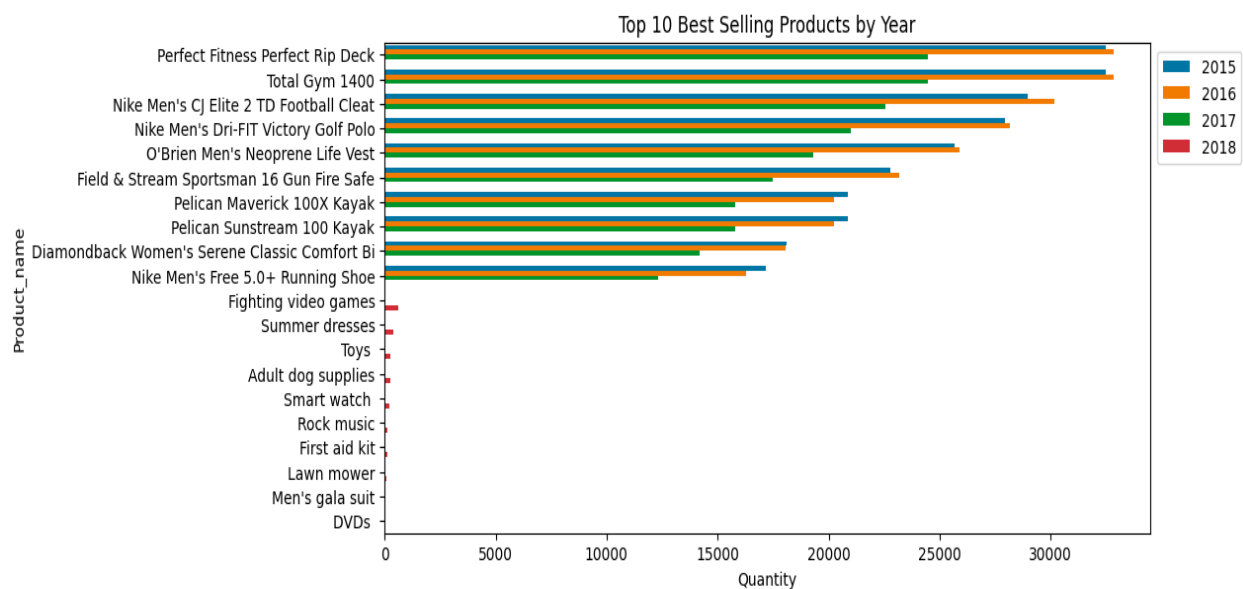
- From 2015 to 2016 the quantity of the top selling products slightly increased and plummeted in 2017.
- The decrease in sales for 2017, can be attributed to the rise of competitors that sell a variety of products such as Costco and Amazon.

4. The product distribution is based on the product's category sold for each year:
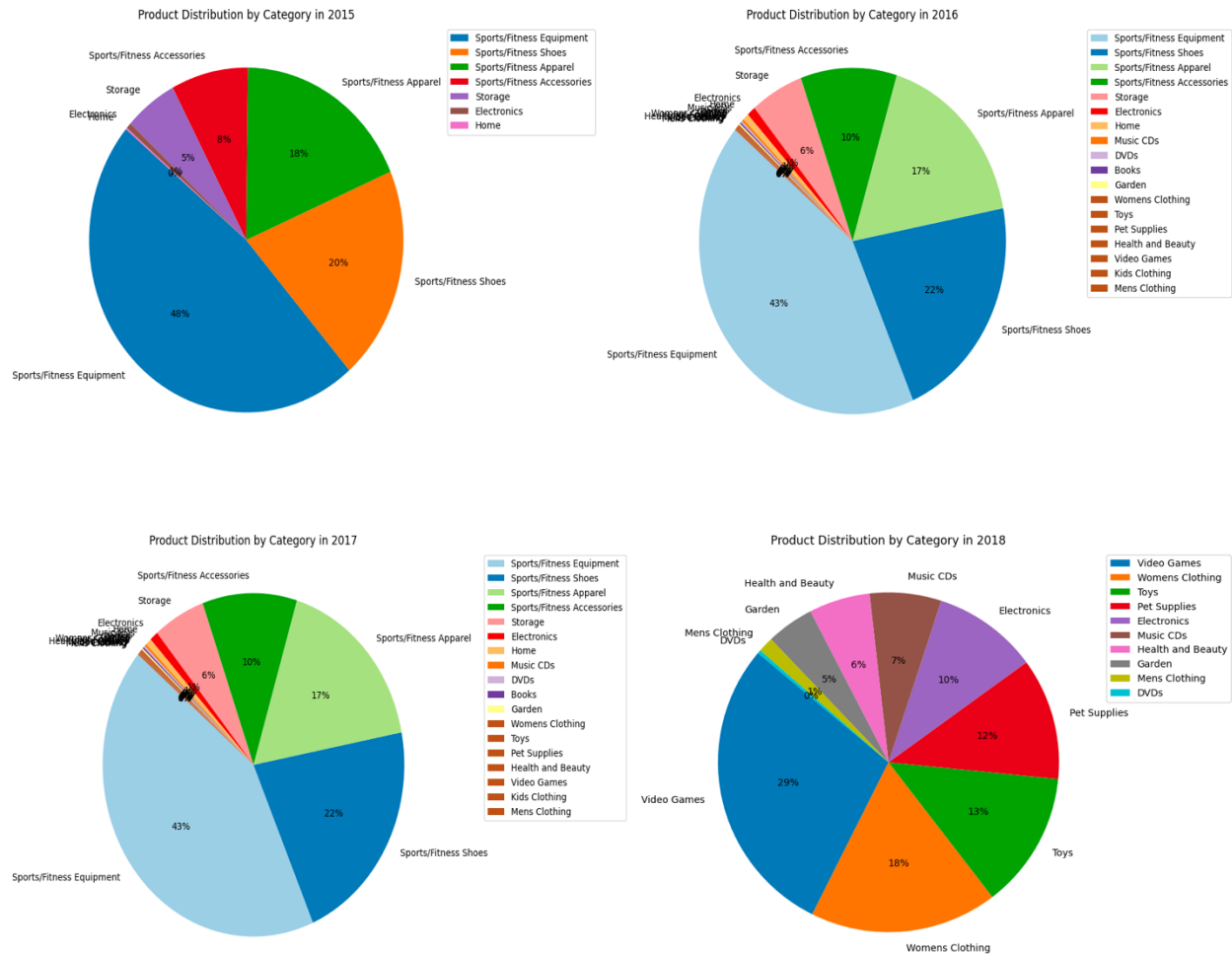
- In 2016, the company transitioned from only selling sports/fitness-related goods to offering new and different product types.
- The product distribution from 2016 was the same for 2017.
  - The new products are represented in the obscure clustered labels with low percentages meaning low sales were 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.

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.

Total Profits from 2015 to 2018
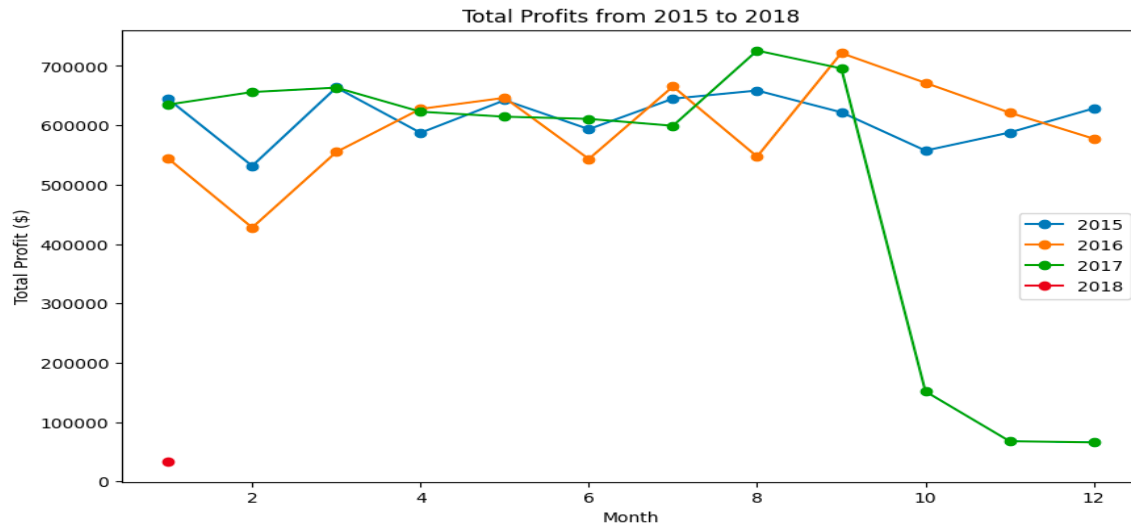
Why?

- 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