

Case Study: Sales Data Exploratory Analysis

Introduction

In this case study, I share insights from my exploratory data analysis on an annual sales dataset from Kaggle. This dataset contains over 15,000 entries of customer demographics and transaction records from a company, spanning an entire year. My analysis aims to uncover hidden patterns and insights that could inform strategic decisions, particularly in optimizing product offerings and sales strategies. By analyzing this data, I explore critical aspects such as seasonal sales trends, product performance, and target demographics.

Data Analysis Steps

1. Clean & Prepare

Upon exporting the data to Excel, I noticed several missing columns and cells, indicating an incomplete dataset. I addressed these issues using Excel's built-in functions:

1. **Remove Duplicates** - Keeps the first occurrence and deletes any subsequent identical values
2. **TRIM** - Removes extra spacing between words
3. **Spell Check** - Corrects spelling errors
4. **Find & Replace** - Substitutes errors or empty cells with the correct data

2. Transformation

Next, I transformed the dataset in order to perform a more detailed financial analysis of the company's operations. Data transformation is the process of changing the format, structure, or values of data to make it better suited for analysis. This involves tasks like converting data into a standard format or creating new columns based on existing data to reveal hidden insights.

Utilizing existing data elements such as quantity, unit cost, and unit price, I calculated the revenue, cost, and net profit for each transaction.

At this point, I asked the following questions ahead of my analysis:

1. *How does the sales volume change over the different months of the year?*
2. *How can this sales data be used to inform the manufacturer and adjust manufacturing?*
3. *What can the target market be based on the customer demographic data?*

3. Analyze

After cleaning and transforming the data, I searched for patterns or trends using pivot tables. These tables facilitated quick and accurate summarizations of the large dataset,

categorizing, sorting, or filtering the data as needed. I will dive into the details of my findings below.

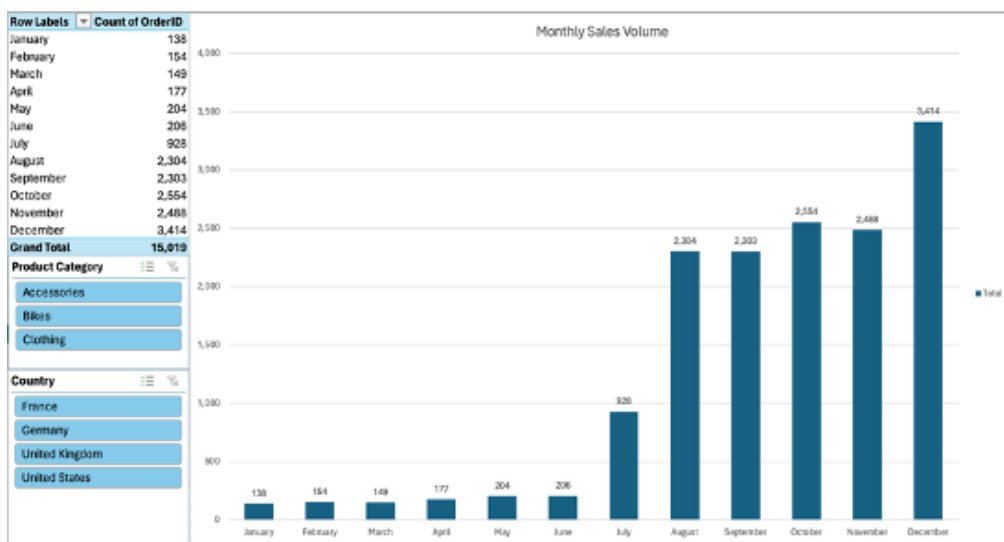
4. Visualize

To visualize my insights, I used pivot charts and graphs in Microsoft Excel. Once my pivot charts were complete, I utilized pivot tables to display the company's monthly and annual financial reports, sales reports, and customer demographics.

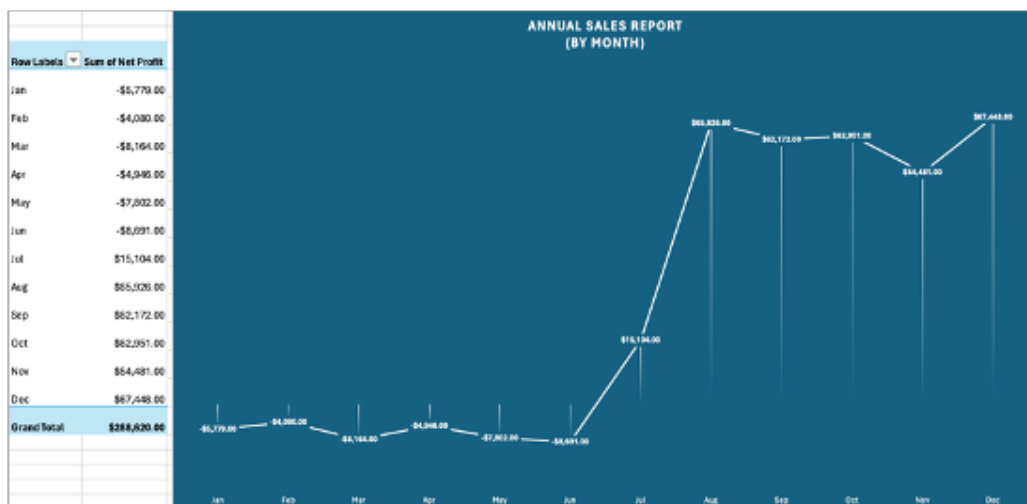
How does the Sales Volume change throughout the year?

Analysis

To understand the yearly sales volume, I selected the data for two Pivot Tables. One that displayed the monthly order volume, and another that displayed monthly net profit.



Adding some slicers helped filter data and get a better understanding. I added two slicers for Product Category and Country.



With the data visualized, a clear seasonal trend emerged. Sales volume consistently increased from July through December, while the company faced losses from January to June. To mitigate these losses, the company could diversify its product lineup and initiate seasonal and off-seasonal promotions, such as sales, discounts, or buy-one-get-one-free offers. Targeted advertising, utilizing detailed customer demographics, could further enhance these efforts.

Insights

The displayed data reveals a distinct seasonal sales pattern, with significant peaks from July to December and troughs from January to June. This fluctuation indicates an opportunity for strategic adjustments in inventory management and marketing efforts to stabilize revenue throughout the year. Specifically, enhancing off-season promotions and tailoring advertising to tap into underutilized customer demographics could mitigate these seasonal dips and improve overall profitability.

How can this data be used to adjust manufacturing and inventory?

Analysis

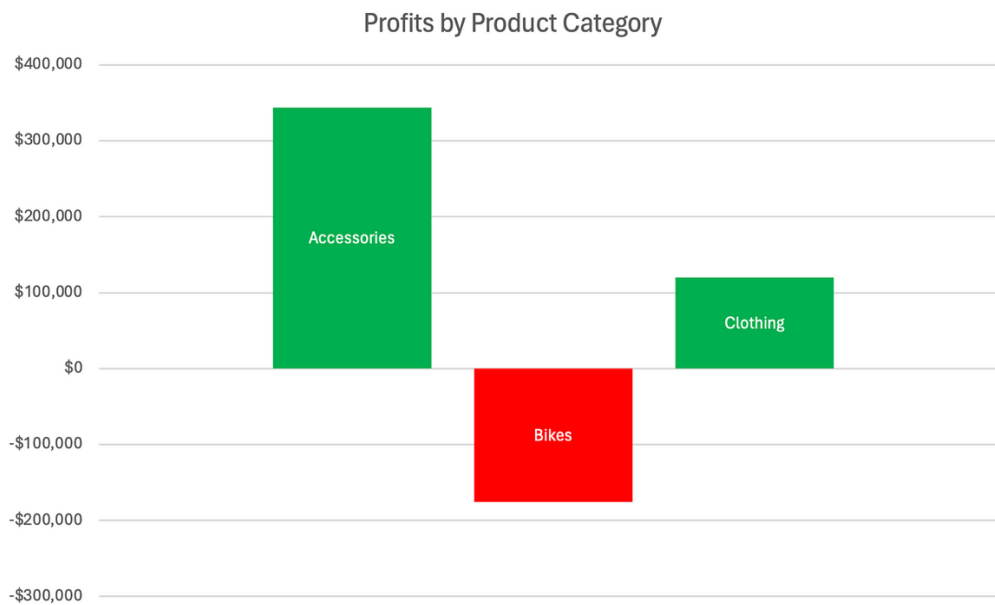
To better understand product performance, I transferred the data into a new pivot table which displayed each category’s monthly profit/loss. The data showed that accessory and clothing sales were non-existent from January through June, suggesting a need for targeted marketing campaigns and promotions to increase market share.

Column Labels 				
	Accessories	Bikes	Clothing	Grand Total
Sum of Net Profit	\$343,518.00	-\$175,334.00	\$120,436.00	\$288,620.00

Column Labels 			
	Accessories	Clothing	Grand Total
Sum of Net Profit	\$343,518.00	\$120,436.00	\$463,954.00

The bike category has consistently underperformed, leading me to simulate the financial impact of ceasing bike sales. I simulated the financial impact of entirely ceasing the sales of bikes. In Figure 3.1, we see bike sales are consistently at a loss each month, totaling up to -\$175,334, which cause the ending profits to remain at \$288,620. In Figure 3.2, the simulated annual profits show ending profit is up by 37.79%, at \$463,954. The best financial strategy for this company would be to inform the manufacturer to discontinue bike production. Reducing bike sales would drastically help optimize the company’s performance.

Insights



Analysis of monthly profit and loss by product category highlights significant inefficiencies, especially the consistent underperformance of the bike category. This persistent loss significantly impacts overall profitability. Consequently, stopping both new sales and manufacturing of bikes could lead to a substantial financial turnaround, increasing annual profits by an estimated 37.79%

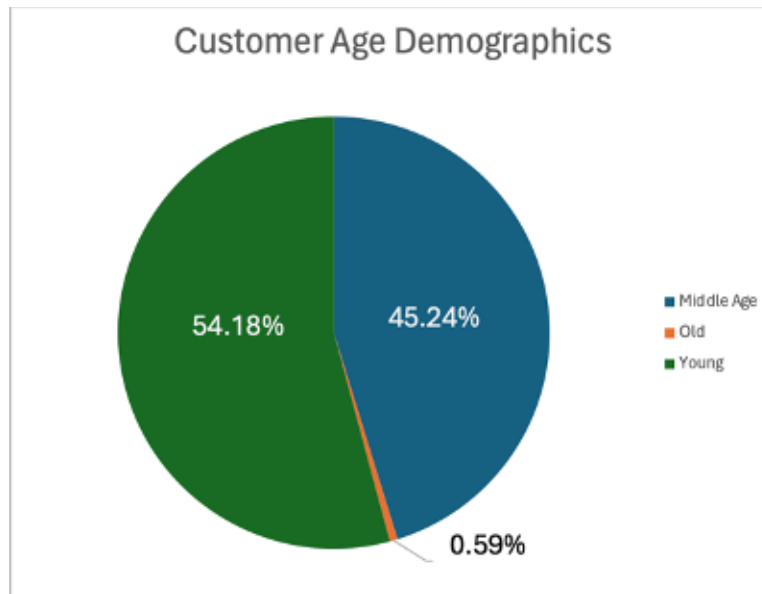
What can the company's target market be based on customer demographics data?

Analysis

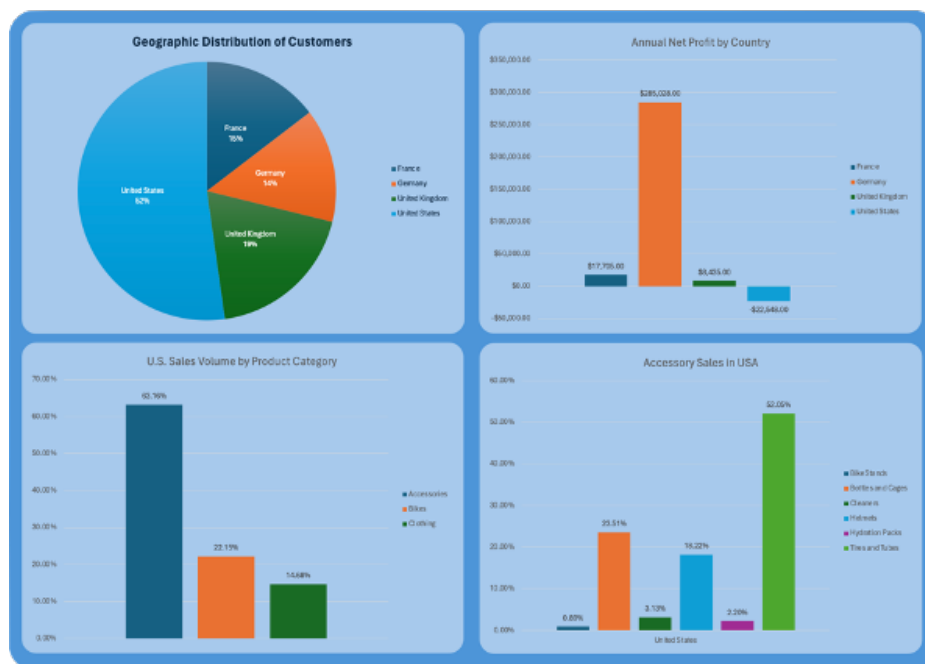
The dataset provides us with plenty of customer information, such as their age, gender, and where they are located. This data allows me to identify a target market, which can then be approached with strategic marketing tactics. I sorted each customer within the following 3 groups: Young (ages 35 and below), Middle Age (ages 36–64), and Senior (ages 65 and up).

Excel Formula

```
=IF([@[Customer Age]]<=35,"Young",IF([@[Customer Age]]<=64,"Middle Age",IF([@[Customer Age]]<90,"Senior")))
```



With the customers sorted in their respective age groups, we can see 54.18% of the client base falls under the 'Young' age group, followed by 45.24% in the 'Middle Age' group.



To gain insight on the geographic distribution of the company's customer base, I utilized another pivot table that used transaction data from every country. This is what we know about the company's target market:

- 52% of the customer base is in USA.
- 63% of USA customers order from Accessories category.
- Tires and Tubes are the top selling accessory in USA.

Insights

The demographic analysis reveals a predominately young customer base, primarily located in the USA, with a strong preference for accessories. Given the strong performance of the accessories category, refocusing marketing and resource allocation toward these high-demand products could enhance profitability.

Conclusion:

Throughout the process of this exploratory data analysis, I shared my workflow and uncovered key business insights that help make data-driven decisions. I started off by explaining how I prepared and analyzed the company data. We went over the company's monthly sales over the span of a year and uncovered that they are experiencing seasonal sales in the months of July to December. I used the insights from the monthly sales analysis to make data-driven decisions about adjusting manufacturing and inventory. Finally, I analyze the customer demographics to get a clear understanding of the company's target market.

Along with the insights, I provided data-driven decisions the business can take to further improve their performance. The data-driven decisions include:

1. Diversify Product Lineup and Promotions
2. Reduce/Eliminate Bike Sales for Increased Net Profit
3. Multi-Channel Marketing Campaigns
4. Targeted Advertising by Demographics