STATISTICAL ANALYSIS and decision making

(group project)

**Table of Contents**

1. **Industry Overview**
2. **Company Profile**
3. **DATA DESCRIPTION**
   * **Dataset Overview**
   * **Rationale for Data Selection**
4. **Descriptive Analysis**

* **Statistical Summary**
* **Profit**
* **Sales**

1. **Interpretation for Operations and Objectives**
2. **VISUALISATION**
   * **Pie Chart**
   * **Bar Chart**
   * **Scatter Plot**

* **Line Fit Plot**

1. **DASHBOARD ANALYSIS**
   * **Key Insights from the Dashboard**
   * **How Different Stakeholders Can Use the Dashboard**
2. **CONCLUSION AND RECOMMENDATIONS**

* **Target Audience Identification**
* **Key Findings Summary**
* **Strategic Recommendations**
* **Usefulness of the Work**

1. **REFERENCES**
2. **DRILLBIT REPORT**

**INDUSTRY OVERVIEW:**

The dataset represents sales operations from the consumer goods and office supplies industry, particularly focused on the sale of furniture, office supplies, and technology products across different regions within the United States. The consumer goods sector is highly competitive, with companies striving to meet evolving consumer demands, which are heavily influenced by trends like ecommerce growth, sustainability, and the shift toward home offices post-pandemic.

Management of supply chain disruptions, ensuring competitiveness while remaining profitable, and ensuring quick and dependable delivery are among the many problems that need to be addressed by the industry. New opportunities in the industry include increased sales of products designed for remote work, enhanced customer experience with personalized offerings, and utilization of analytics for improving efficiency and targeting customers. The dataset addresses these dynamics quite well because it contains data with respect to sales patterns, target cohorts, types of delivery and their profitability, which are essential in dealing with industry challenges and tapping into the market.

**COMPANY PROFILE:**

For this analysis, we can assume the company is a midsized ecommerce firm that specializes in selling furniture, office supplies, and technology products. The company’s mission is to provide high-quality products with fast and efficient shipping, ensuring customer satisfaction. Their vision is to be the leading supplier in the office and consumer goods sectors by leveraging data-driven decisions to improve operational efficiency and customer engagement.

The company’s target market includes businesses and individual consumers within the United States, segmented by categories like "Consumer" and "Corporate," which is reflected in the dataset. The company offers a range of products across furniture (e.g., chairs, tables), office supplies (e.g., labels, storage systems), and technology, catering to different customer needs across various regions.

**DATA DESCRIPTION**

**Dataset Overview:**

The dataset consists of 22 variables, providing comprehensive information regarding sales of orders, including:

* Customer and Order Details: Order ID, Order Date, Ship Date, Customer ID, Customer Name, and Segment.
* Geographical Information: Region, Country, and Postal Code.
* Product Details: Product ID, Category, Sub-Category, and Product Name.
* Sales and Profitability Metrics: Sales, Quantity, Discount, and Profit.

The dataset is relatively large, covering multiple transactions across various years, making it suitable for identifying sales trends, customer behaviour, and profitability analysis. The inclusion of shipping modes also provides insights into logistical aspects, which can help assess operational efficiency.

**Rationale for Data Selection:**

This dataset was selected due to its comprehensive nature, which allows for a detailed analysis of sales, customer segments, and profitability across different product categories. This gives a perfect match of the hypothetical organization's objectives by providing data points that aid in improving the firm's offered products, enhancing shipment efficiency, and targeting the most profitable segments of the customer base.

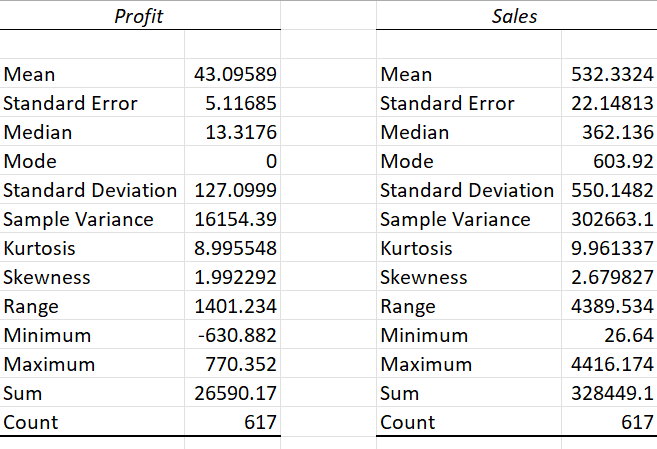
Additionally, the dataset’s sales and profitability metrics provide insights into revenue streams, helping the company identify high-performing products and regions. The detailed customer information enables the company to create tailored marketing strategies, while the order and shipping data can support operational improvements by identifying bottlenecks in the supply chain.

This data can support decision-making processes by informing inventory management, marketing strategies, and customer relationship management, ultimately leading to improved profitability and customer satisfaction.

**DESCRIPTIVE ANALYSIS**

**Statistical Summary**

The statistics provide a descriptive analysis of two variables: Profit and Sales. Let's break down their meanings in the context of an organization's operations and objectives:



1**. Profit**:

* Mean (43.09589): This indicates the average profit across the dataset. The company’s operations are yielding a profit of approximately 43.1 units on average, suggesting that while the business is profitable, profit margins may be relatively low.
* Standard Error (5.11685): A measure of the variability in the mean profit. A smaller value here suggests that the mean is a reliable estimate of the true population mean, and the company’s profitability is relatively stable.
* Median (13.3176): The middle value in the dataset. Half of the profits are above 13.3 units, and half are below. This value being lower than the mean suggests that a few large profits may be skewing the average upward.
* Mode (NA): The mode is the most frequent value. It seems there's no dominant profit figure that appears more than once in the data.
* Standard Deviation (127.0999): This large number indicates high variability in profit values. Some units of the business might be significantly more profitable than others, or some periods may be substantially more successful.
* Skewness (1.992292): Positive skewness indicates that the distribution of profits has a long tail on the right, meaning there are a few instances of very high profits.
* Range (1401.234): This is the difference between the largest and smallest profit values, indicating wide fluctuation in profitability, which could be due to seasonal factors or operational inefficiencies.
* Minimum (630.882): This suggests there are losses in some cases, which may require addressing specific areas of the business to mitigate these losses.
* Maximum (770.352): The maximum profit value, showing the highest profitability achieved, providing a benchmark for optimizing operations.
* Sum (26590.17): This is the total profit accumulated over the 617 instances (perhaps months, projects, or departments).
* Count (617): Number of data points, indicating the breadth of the dataset used for this analysis.

2**. Sales:**

* Mean (532.3324): The average sales per unit/department/project/etc. Sales are relatively high compared to profits, which may suggest either high operational costs or potential pricing issues that impact profitability.
* Standard Error (22.14813): A moderate variability in the sales data, indicating a somewhat consistent sales performance across the dataset.
* Median (362.136): The midpoint sales value, showing that while the mean is high, more than half of the sales instances are below this point, indicating significant variance in sales performance.
* Mode (603.92): The most frequent sales value, showing that a sales figure around 603.92 is common.
* Standard Deviation (550.1482): High variability in sales, suggesting some units or periods are performing much better in sales than others.
* Skewness (2.679827): Positive skewness indicates that the distribution of sales is right skewed, with several large sales instances driving up the mean.
* Range (4389.534): Wide fluctuation in sales performance, suggesting inconsistencies in either market demand or operational capacity.
* Minimum (26.64): Very low sales in some periods, indicating potential inefficiencies or external factors affecting performance.
* Maximum (4416.174): The maximum sales value, representing the peak performance period.
* Sum (328449.1): The total sales achieved across all data points, indicating the overall size of the organization’s operations.
* Count (617): Number of sales data points, matching the profit dataset.

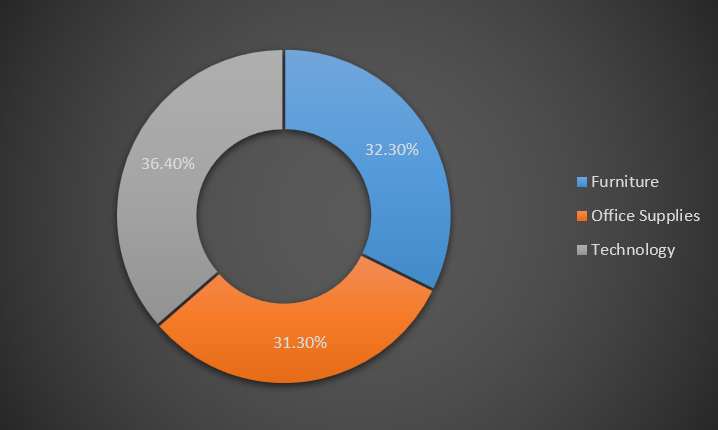
**Interpretation for Operations and Objectives**:

* Variability in Profit and Sales: Both profits and sales show high variability, indicating the organization’s performance may be inconsistent across departments, units, or time periods. Understanding the drivers of this variability can help management focus on stabilizing operations.
* Skewness and Outliers: The positive skewness in both sales and profits suggests that while there are outliers (high sales and profit values), the majority of the data falls below the mean. This could point to a few high performing units or periods, but also highlights that the company should focus on improving the consistency of performance across the board.
* Operational Efficiency: The gap between high sales and relatively lower profits may suggest high operational costs, inefficiencies, or issues with pricing strategies that need to be addressed to improve margins.
* Strategic Decisions: To achieve sustainable growth, the organization might need to focus on reducing the losses (as indicated by the negative minimum profit), possibly by identifying underperforming areas and either improving or discontinuing them. Additionally, leveraging high performing areas could help boost overall profitability.
* Future Objectives: Analysing the maximum values and the total sales/profit provides benchmarks for setting growth and optimization targets. The company may aim to reduce variability (standard deviation) and improve the median performance to bring lower performing units closer to the overall mean.

This analysis can help management prioritize actions that drive more consistent performance and better profitability while addressing operational inefficiencies.

**VISUALISATION**

1. **PIE CHART**



The image displays a pie chart in Excel that shows the distribution of sales across three categories:

* Furniture: 32.30%
* Office Supplies: 31.30%
* Technology: 36.40%

What it reveals:

* Technology has the highest share of sales, accounting for 36.40%.
* Furniture follows with 32.30%.
* Office Supplies contributes the lowest share, at 31.30%.

**Implications:**

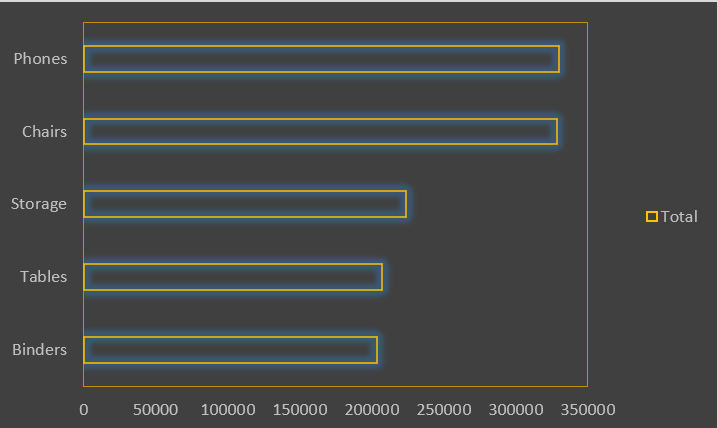
Technology products are driving the largest portion of sales. This may suggest a trend where customers are prioritizing technology over other categories, possibly due to higher demand for gadgets, tools, or tech solutions.

Furniture and Office Supplies have nearly equal sales proportions, indicating steady demand for both categories. However, since Furniture edges out Office Supplies, companies could explore strategies to boost the latter's sales, perhaps through promotions or improving product offerings.

The fairly balanced distribution suggests that the organization has diversified sales across categories, which is good for minimizing risk. If one category faces a downturn, others can help balance overall sales performance.

In summary, the chart shows the importance of focusing on technology to sustain growth while potentially finding ways to expand the share of office supplies

1. **BAR CHART**



The table shows sales data for five product categories:

* Binders:203,412.733
* Tables:206,965.532
* Storage:223,843.608
* Chairs: 328,449.103
* Phones:330,007.054

The total sales across all categories amount to 1,292,678.03. The chart appears to be a horizontal bar chart, which is useful for comparing values across different categories. Each bar represents the total sales for each product.

Insights from the Data:

* Top Product: Phones have the highest sales, contributing 330,007.054, slightly ahead of Chairs (328,449.103).
* Low Performing Product: Binders have the lowest sales at 203,412.733, making it the least profitable category among the five.
* Close Sales Figures: Tables and Binders have similar sales numbers, with Tables slightly higher than Binders
* Mid-Range Performers: Storage falls in the middle, with sales amounting to 223,843.608, performing better than Binders and Tables but significantly lower than Phones and Chairs.

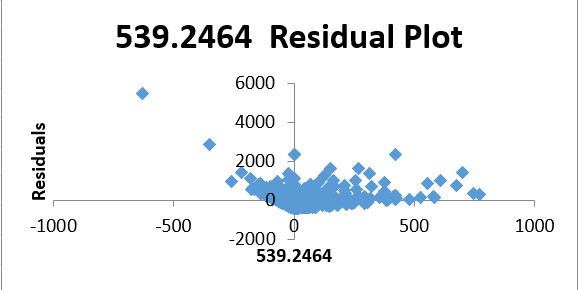
Potential Recommendations:

* Focus on Top Performers: Given that Phones and Chairs lead in sales, increasing inventory or marketing for these categories could drive further revenue.
* Improve Sales for Low Performers: Strategies to increase the sales of Binders (the lowest) and Tables (second lowest) could include discounts, promotions, or reevaluating the product’s pricing and positioning.
* Monitor Mid-Performers: Storage has a moderate performance, so maintaining its current strategy while looking for growth opportunities may be wise.

1. **SCATTER PLOT**

The two key visualizations are a Residual Plot and a Line Fit Plot related to the regression output.

* Residual Plot:

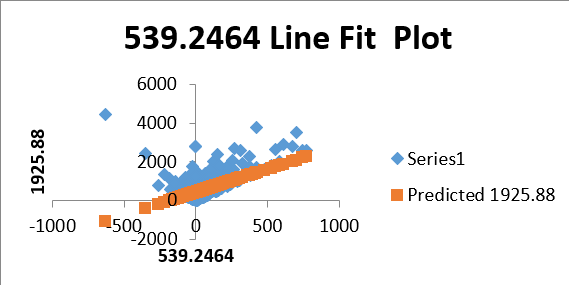


The x-axis represents the predicted value (539.2464), and the y-axis represents the residuals (the difference between the observed and predicted values).

The scatter pattern shows a generally random distribution around the 0 residual line.

Implication: A random scatter suggests that the model's errors are evenly distributed, which is a good sign for model validity. However, there seems to be some heteroscedasticity (nonconstant variance) where residuals increase as the predicted values increase, which could indicate that the model is not handling variance well for higher values.

* **Line Fit Plot:**



This shows how well the regression line fits the actual data. The x-axis is the predicted value (539.2464), while the y-axis represents both the actual and predicted data points.

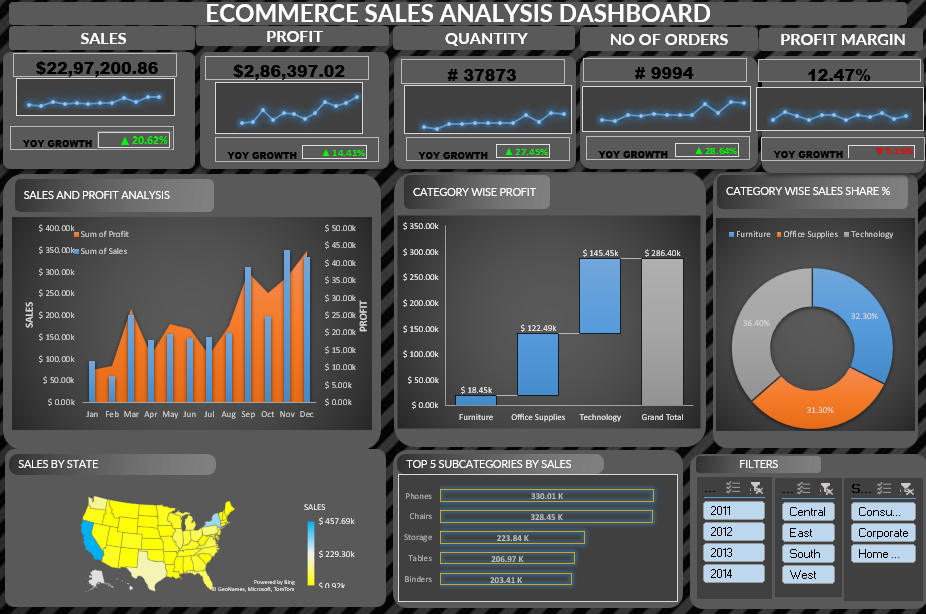
The red line shows the predicted trend (constant at 1925.88), while the blue dots represent actual data points.

Implication: There is a significant variance between the actual data and the predicted trend, particularly as the actual data points scatter widely across different values. This suggests that the model's predictive power is weak, as it does not capture the true variability in the data. The constant prediction may indicate an overly simplistic model (perhaps a linear regression that is not a good fit for this data).

Hence the residuals are somewhat randomly distributed, the variance in the data increases, which suggests the model could struggle with larger values.

The fit of the line to the actual data is poor, indicating that the model may need improvement (perhaps using a more complex or nonlinear approach) to better capture the underlying trends in the data.

**DASHBOARD ANALYSIS**



It shows an Ecommerce Sales Analysis Dashboard, which is a powerful tool to visualize key metrics such as profit, quantity sold, number of orders, and profit margins. It also provides additional insights such as category-wise profit, sales by state, top subcategories, and other filters that help in detailed analysis.

**Key Insights from the Dashboard:**

1. Overall Performance:

* Sales: $2.86M in total profit with a YoY growth of 20.62%.
* Quantity Sold: 37,873 units sold with a YoY growth of 27.45%.
* Number of Orders: 9,994 total orders, showing a strong YoY growth of 28.64%.
* Profit Margin: 12.47%, which is showing a slight decline (1.15% YoY).

These figures suggest that the business is growing in aspect of volume and orders, although there might be some concerns about declining margins.

2. Category-Wise Analysis:

* The dashboard shows category-wise profits in Furniture, Office Supplies, and Technology.
* Technology seems to dominate with the highest profit ($2.86M), while Office Supplies show the least profit ($122.49K).

This suggests that Technology is the strongest segment, while Office Supplies may need more attention or cost optimization.

3. Sales Distribution:

A map highlights sales performance by state, with top-performing states in comparison of sales, ranging from $2.23M to $4.65M.

This data can help identify the regions with high demand and those that may require improved sales strategies or marketing focus.

4. Top 5 Subcategories by Sales:

The top performing product subcategories include Phones ($330.61K), Chairs ($322.65K), Storage ($211.47K), Tables ($206.97K), and Binders ($201.14K).

These figures provide insights into which products are driving revenue and can help focus inventory management and marketing on the most profitable or in demand items.

5. Category-Wise Sales Share:

A pie chart provides a breakdown of sales share, with Technology (36.40%), Furniture (32.08%), and Office Supplies (31.51%) having nearly equal shares in overall sales.

This indicates that while technology is the biggest profit driver, sales are relatively balanced across the other categories, showing a diverse product mix.

**How Different Stakeholders Can Use the Dashboard:**

1. Sales Team:

The sales team can use the insights to identify top performing regions and products and focus their efforts on growing sales in underperforming areas. They can also tailor marketing campaigns around high demand products like phones and chairs.

2. Operations and Supply Chain:

The operations team can use the top selling subcategories and state level data to optimize inventory and logistics. Knowing which regions have the highest sales helps to plan efficient distribution and warehousing strategies.

3.Marketing Team: Marketing can leverage the YoY growth rates and performance data to tailor campaigns to specific categories or states. They may focus promotions on underperforming categories like Office Supplies or high margin items like technology products.

4. Finance Department:

The finance team will be interested in the profit margins and the category wise profit breakdown. The declining profit margin may trigger investigations into cost control, pricing strategies, or vendor negotiations to improve profitability.

5. Management and Executives:

Senior management can use this dashboard for strategic decision making. The overall growth in orders and sales, alongside declining profit margins, could lead to discussions about scaling the business while maintaining profitability, optimizing pricing strategies, and reallocating resources to higher performing segments.

This dashboard provides a comprehensive view of the company's performance across key metrics. By allowing stakeholders to drill down into different areas (categories, regions, and product lines), the dashboard aids in informed decision making that can optimize operations, drive sales, and improve profitability across the business.

**CONCLUSION AND RECOMMENDATIONS**

**1. Target Audience Identification**

The target audience for this ecommerce sales analysis dashboard would likely include:

* Sales Managers: Responsible for tracking performance metrics like profit margins, order quantities, and year over year growth.
* Marketing Teams: To evaluate which categories are performing well and how they can improve customer engagement or market penetration.
* Product Managers: To assess which product categories are driving profits and where strategic focus should be placed.
* Senior Executives/Business Owners: Interested in a higlevel overview of sales performance to make business decisions.
* Financial Analysts: To assess profitability trends and sales distribution across different states and product categories.

**2. Key Findings Summary**

* Sales and Profit Trends: The dashboard indicates positive year-over-year growth in sales (20.62%), order quantities (27.45%), and the number of orders (28.64%). However, there is a slight decline in the profit margin (1.15%), which may require further analysis.
* Category wise Profit: Technology stands out with the highest profit ($286.40K), followed by Office Supplies ($122.49K) and Furniture ($145.45K).
* Sales by State: States with higher sales include California ($4.07M) and Texas ($3.22M), while others have significantly lower sales figures.
* Top 5 Subcategories: The top performing subcategories include storage ($100K), copiers ($321.45K), and binders ($221.84K).

**3. Strategic Recommendations**

* Focus on High Profit Categories: Increase marketing and inventory for Technology and Office Supplies since they generate high profits.
* Improve Profit Margins: Investigate why profit margins are declining despite the growth in sales and orders. This could involve optimizing supply chain efficiency, reducing operational costs, or rethinking pricing strategies.
* State Level Strategy: Implement localized marketing campaigns or partnerships in states with lower sales, like those in the Midwest, to boost engagement and revenue.
* Diversify Subcategories: Explore strategies to replicate the success of high performing subcategories like copiers and binders in other less performing categories.

**4. Usefulness of the Work**

* This analysis helps stakeholders quickly assess both high level performance (overall sales growth, profit trends) and granular details (specific subcategory and state performance).
* It provides actionable insights for adjusting strategies based on real time data, such as areas for improvement in profitability and potential growth areas across different regions or product categories.

Overall, the dashboard supports data driven decision making across departments, contributing to a cohesive strategy for business growth and performance optimization. California ($4.07M) and Texas ($3.22M), while others have significantly lower sales figures.

**REFERENCES**

* <https://datasetsearch.research.google.com/>
* <https://docs.google.com/spreadsheets/u/0/d/1L6aBX0uNlzKiJb7JHdkNUile18s9CI4r/edit?gid=1589100670&pli=1#gid=1589100670>