Sales Performance Analysis: Uncovering Trends and Insights from a Comprehensive Data

Name : Varun Paka **Student ID : 23096047**

Email ID :vp24aav@herts.ac.uk **Git hub Link : Link**

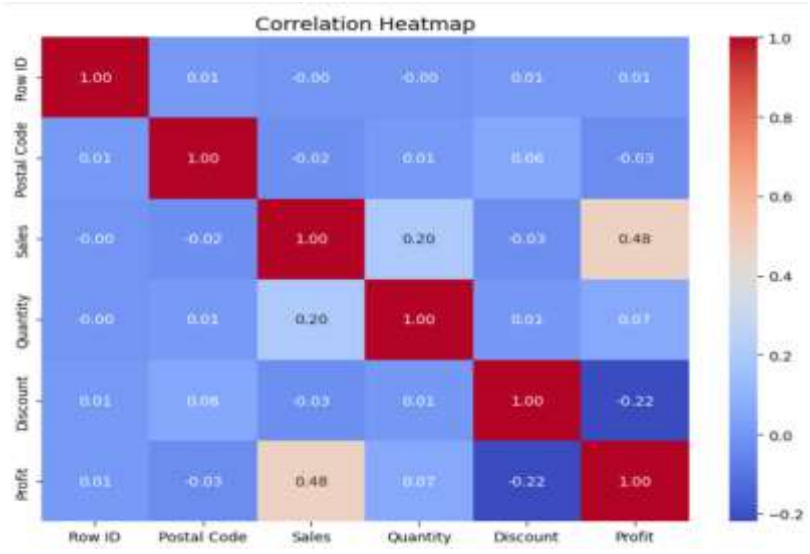
Introduction

Overview This report analyses a wide sales dataset of 9,994 records and 22 attributes, capturing key characteristics including, but not limited to, sales figures, profits, discounts, quantities, and detailed customer and product information. The key metrics include total sales ranging from 0.44 to $22,638.48, with an average of $229.86, and profit ranging from a loss of - $6,599.98 to a gain of $8,399.98, averaging $28.66. The dataset also contains temporal and geographic information, including the dates of order and ship, postal codes, and regional classifications that help reach deeper insights into trend and pattern analyses.

The most important objectives of the present analysis are to find out underlying patterns in sales performance and profitability, understand the effectiveness of discounts and quantity sold on profit margins, and identify regional and categorical trends using higher-order clustering and regression techniques. The statistical measures show large variability in sales and profits, indicating outliers and diverse market behaviours. Visualizations such as histograms, scatter plots, and clustering diagrams will be used in this report to provide actionable insights that will help in strategic decision-making toward enhancing sales effectiveness and optimizing profit margins.

Visualizations and Insights

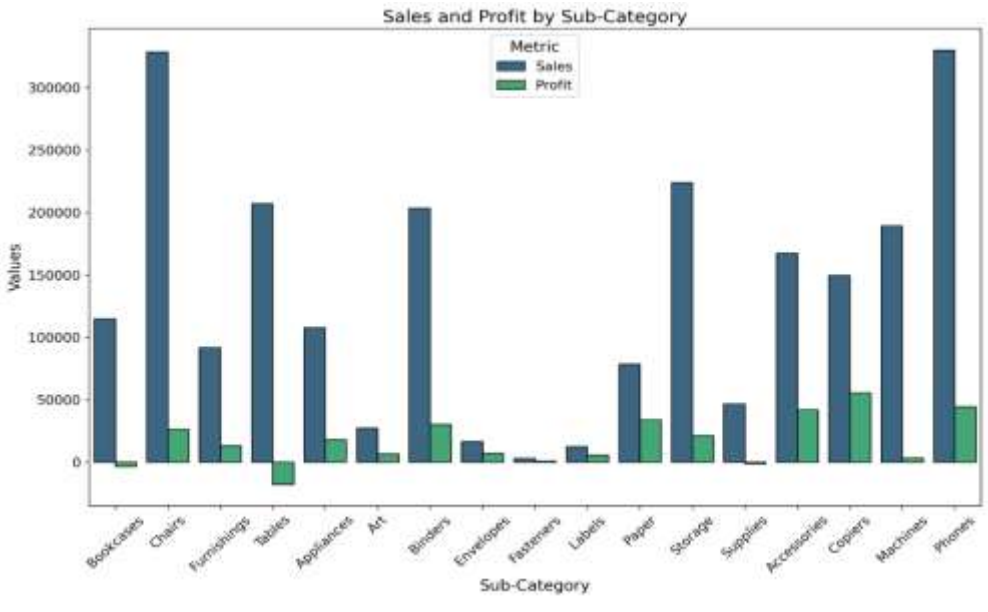
1. **Correlation Heatmap**



The correlation heatmap unveils important relationships between variables:

- Sales and Profit (0.48): a moderate positive correlation, reflecting that with increased sales comes higher profits.
- Discount and Profit (-0.22): The correlation is negative, indicating that the higher the discount, the lower the profit.
- Sales and Quantity (0.20): A moderate positive correlation, which means that the higher the sales, the higher the quantities sold.

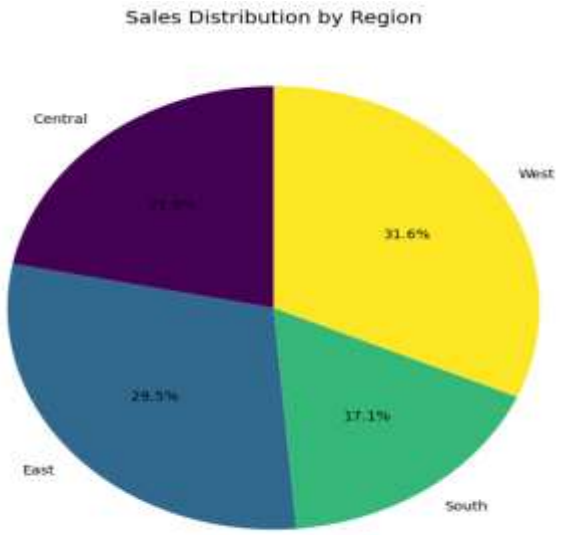
2. Sales and Profit by Sub-Category (Bar Chart)



This bar chart compares the sales and profit across sub-categories:

- High Sales and Profit: Phones and Chairs present good performances.
- Low Sales, High Profit: Copiers are highly profitable despite lower sales
- High Sales and Low Profit: Fasteners and Supplies contribute minimally to overall performance

.



3.Sales Distribution by Region (Pie Chart)

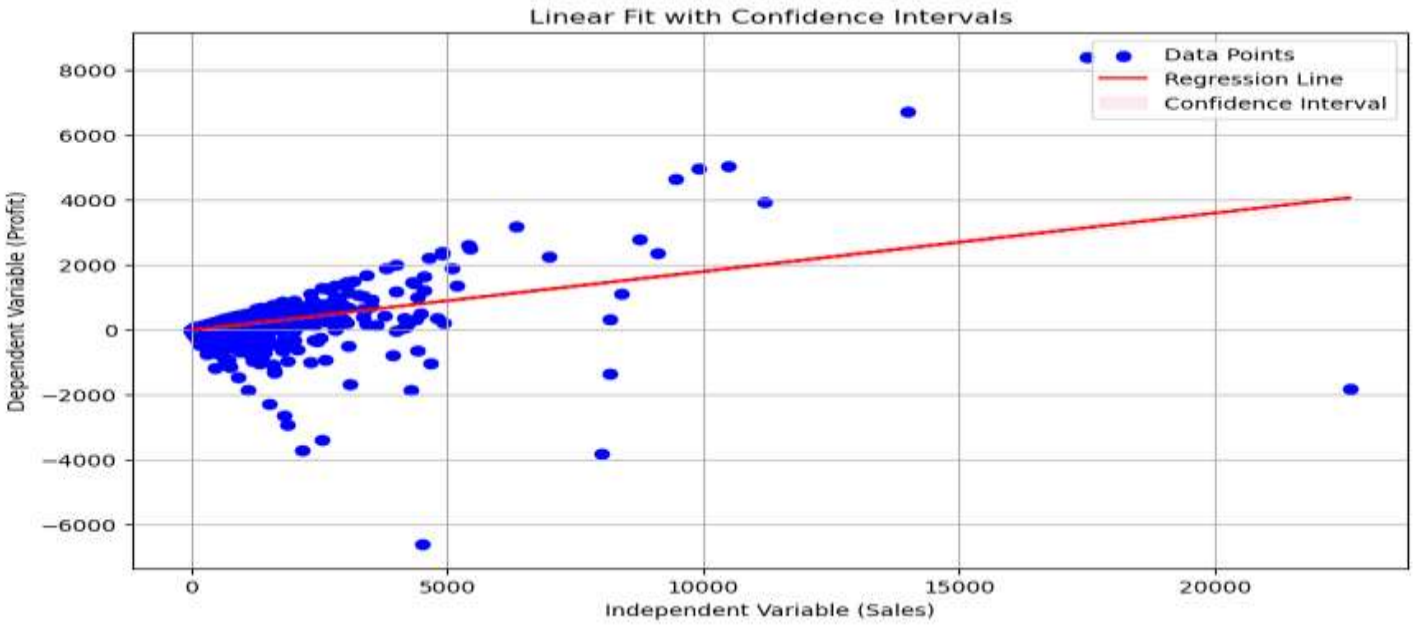
The pie chart below depicts sales by region distribution:

a. West (31.6%), East (29.5%): These are the highest sales percentages, reflecting high performance.

b. Central, 21.8%, and South, 17.1%: Poor sales, indicating opportunities for further growth.

These will enable views on performance trends, correlations, and regional contributions to overall sales, which could inform strategic decisions.

Linear Fit Analysis



The scatter plot "Linear Fit with Confidence Intervals" shows the relationship between Sales (independent variable) and Profit (dependent variable). Key features of the plot include individual data points (blue dots), a regression line (red), and a shaded confidence interval around the line.

Insights:

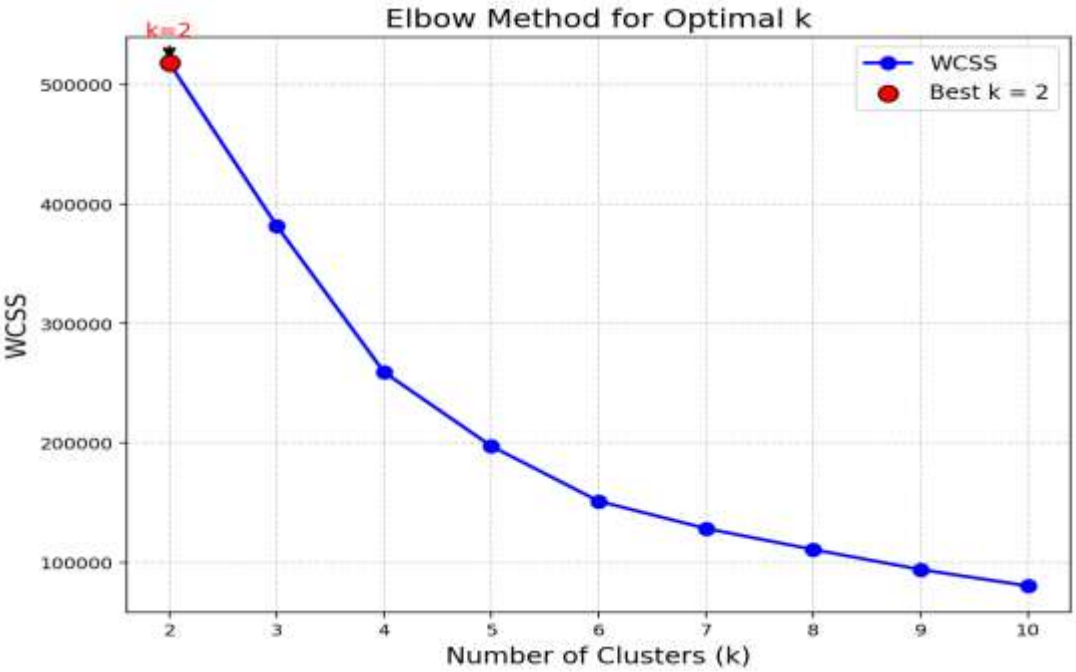
Positive Correlation: The regression line illustrates a positive trend; this means that higher sales are mostly associated with higher profits.

Confidence Interval: The shaded area represents the confidence interval of the model's predictions. The narrower it is, the more confident the predictions; wider intervals indicate variability in the relationship.

This effectively visualizes the relationship between Sales and Profit, showing how each may relate financially, based on changes in sales performance.

Clustering Analysis - Elbow Method and Silhouette Scores

The clustering evaluation combines insights from the **Elbow Method** and **Silhouette Scores** to determine the optimal number of clusters.



Elbow Method Insights:

1. Optimal Number of Clusters: The Elbow Method suggests 2 as the optimal number of clusters, where the Within-Cluster Sum of Squares (WCSS) shows a sharp decline and begins to flatten, indicating diminishing returns with additional clusters.

2. Interpretability : Selecting 2 clusters ensures that segmentation remains simple while effectively capturing key data patterns.

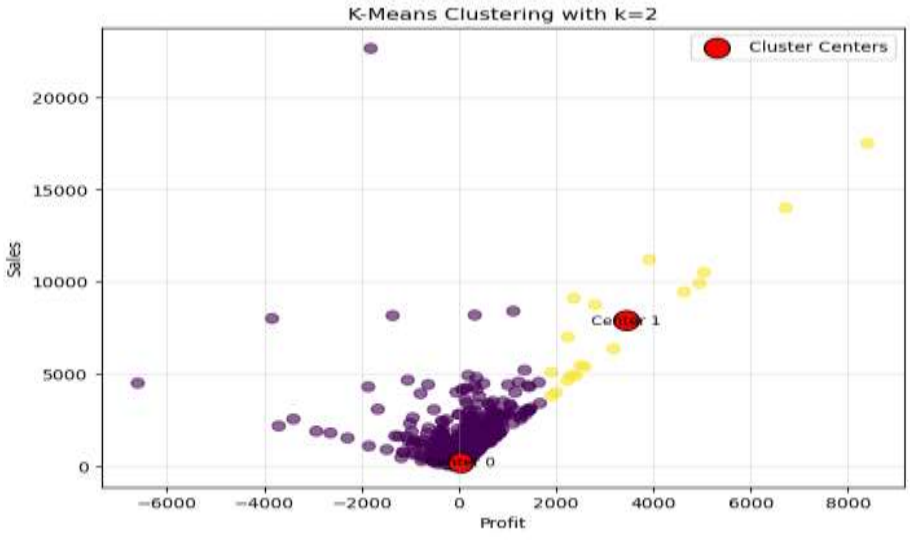
Silhouette Score Insights:

1. Best Cluster Configuration : The highest **silhouette score for 2 clusters is 0.97**, which indicates excellent cluster separation and cohesion.

2. Score Decline : For more than 2 clusters, silhouette scores decrease progressively, again confirming 2 as the optimum number for meaningful segmentation.

Conclusion: The combined evaluation shows that 2 clusters achieve the best balance between interpretability, compactness, and separation to provide actionable insights for data segmentation.

K-Means Clustering Analysis



This scatter plot, "K-Means Clustering with k=2" , shows the clustering of data points based upon their Sales - y-axis, and Profit - x-axis. The K-Means algorithm segments the data into two clear clusters.

Insights:

1. Cluster Separation : The plot highlights two well-separated clusters, represented by purple and yellow data points. This clear demarcation indicates that the algorithm effectively distinguishes between the groups based on the relationship between sales and profit.

2. Cluster Centres : The red circles, labelled "Centre 0" and "Centre 1," mark the mean positions of each cluster. These centres can serve as reference points for understanding the average characteristics of each group.

Conclusion: The statistical and visualization techniques used in this report included **correlation heatmaps, bar charts, pie charts, regression analysis, and clustering**. **Key insights to be derived include the positive correlation of sales with profit, the impact of discounts on profitability**, regional sales distributions, and **optimal segmentation of data using K-Means clustering**. These findings can help drive strategic decision-making by identifying the main drivers of performance, areas for improvement, and targeted actions to improve profitability and efficiency. **Future analyses could explore additional variables, incorporate time-series data, or apply advanced machine learning models for deeper insights and predictive analytics**.