Project Report

On

“Exploratory Data Analysis on the Furniture and Home Décor Industry ”

Submitted by
Dhaval Rathod

Under The Guidance of
Chintan Patel

Affiliated with
Fly The Nest Institute.

2025

“Documentation”

Signature with date

PROJECT REPORT

Contents:

Problem Statement:

This analysis aims to evaluate sales performance and cost efficiency across different departments and products by analyzing the Sales Order data. By identifying key trends, cost distribution, and revenue-driving factors, the study focuses on optimizing profitability, reducing unnecessary expenses, and improving operational efficiency. The goal is to develop data-driven strategies to enhance revenue generation and streamline cost management.

Objectives:

Analyze Top-Selling Items – Identify trends in customer preferences by evaluating the highest-selling products by quantity and revenue.

Compare Department-Wise Costs – Assess direct and overhead costs across departments to identify inefficiencies and optimize cost distribution.

Optimize High-Cost Items – Determine the top 10 costliest products and analyze their contribution to overall profitability.

Study Monthly Order Trends – Evaluate seasonal variations and demand patterns for improved inventory management and sales forecasting.

Assess Department-Wise Performance – Analyze monthly sales volume for each department to improve operational efficiency.

Identify Key Revenue Drivers – Determine top revenue-generating items and their impact on total sales performance.

Evaluate Pricing Strategies – Analyze the relationship between product costs, discounts, and sales volume to refine pricing models.

Assess Order Size by Department – Identify purchasing patterns and demand distribution across departments.

Analyze Profitability Distribution – Use Pareto analysis to determine which products contribute the most to overall sales and profitability.

Examine Regional Sales Trends – Understand geographical sales distribution and identify high-potential markets for growth.

Visualize Sales Distribution – Use Sankey diagrams to map how sales orders flow across different departments.

Enhance Marketing & Promotions – Develop targeted promotional strategies based on sales patterns and seasonal demand.

Improve Supply Chain & Logistics – Analyze fulfillment times and optimize stock management to avoid delays.

Compare Online vs. Offline Sales Performance – Evaluate how different sales channels contribute to revenue.

Forecast Future Sales Trends – Leverage historical sales data to predict future performance and optimize decision-making.

Introduction:

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize the data, using statistical analysis and data visualization methods.
It is a crucial step in any data analysis process, enabling data analysts to uncover patterns, spot anomalies and gain insights to take actions

EDA Pipeline

1. Data Acquisition and Objective
    a. Obtain FurniMart data (CSV, Excel)
    b. Get problem statement from Euromart
    c. Choose tools/environment & programming language
2. Data Loading/Reading
    a. Load data in Jupyter Notebook/VS Code (To perform further analysis)
3. Familiarize with Data & Identify Target Variable
    a. Explore data (column names, data types)
    b. Identify target variable based on objective
4. Data Preparation & Transformation
    a. Data Cleaning
    b. Handle missing values (if required)
    c. Removal of unwanted data (if present)
    d. Format data types (numerical & categorical)
5. Feature Engineering (Create new features)
6. Data Analysis & Visualization
    a. Univariate Analysis
        i. Numerical variables (mean, median)
        ii. Categorical variables (distribution)
    b. Bivariate & Multivariate Analysis (Identify patterns)
        i. Visualization (Charts: pie, boxplot, histogram, heatmap,polar)
7. Summary and Suggestions

About the company

- The company operates in the furniture and home décor industry, offering a wide range of products such as sofas, dining tables, TV units, and more.
- It has multiple departments, including Upholstery, Painting, and Production, handling different product categories.

- The company processes large-scale orders with a structured workflow, including direct and overhead cost allocations.

Data Loading/Reading:

Import Necessary Library

- numpy (np): Provides efficient numerical computation tools
- pandas (pd): Offers data manipulation and analysis structures (DataFrames, Series)
- seaborn (sns): Creates informative statistical data visualizations based on Matplotli
- matplotlib.pyplot (plt): Enables various plot creations for data visualization
- %matplotlib inline (Jupyter Notebook specific): Displays plots within the notebook
- warnings (with warnings.filterwarnings("ignore")): Suppresses warnings
- plotly.express : simplifies the creation of interactive and visually appealing charts with minimal code.

Load Data in Jupyter Notebook

To begin the analysis, we first load the dataset into a Pandas DataFrame using the read_csv() function. This allows us to explore and manipulate the data efficiently in Jupyter Notebook.

df = pd.read_csv("data actual time work 2023.csv")

Familiarize with Data & Identifying the Target Variable

Explore the provided data (column names, data types)

- We need to understand the data before cleaning the data and cross verify if all the required data are provided by Company

Overview of data

- df.head(); Let's see the data by displaying the first 5 rows
- df.tail(); Let's see the last 5 rows
- df.shape is used to get the dimensions (number of rows and columns) of data
- df.size is used to get the total number of elements in a pandas
- df.info()- used to display concise information about

Interpretation

- Structured Data: The data is provided in a structured table format.
- Dimensions: The dataset contains 10 columns and 11,033 rows, totaling 110,330 elements.
- Column Data Types: Observed a mix of data types:
    - Categorical (Object): 6 columns
    - Numerical (Float64): 4 columns
- Categorical Variables: All categorical/qualitative variables are Nominal in nature (i.e., they have no specific order).
- Non-Null Counts: No null values were found in the dataset. Each column has 11,033 non-null values.
- Memory Usage: The dataset consumes approximately 4.59 MB of memory.

(Further optimization may be performed by modifying data types where applicable.)

Data Preparation & Transformation

Data Cleaning

We need to perform steps mentioned below to clean data:

- Steps involved in handling missing values (imputation, deletion)
    - We accept missing values if data is small in dimension
    - We delete missing values if:
        - When more than 80% of data is missing/null values
        - When the percentage of missing values are very small, deleting will have minimal effect on analysis
- Replacing the missing values by imputation
    - Imputation: We replace the missing values by Mean, Median or Mode of the variable or perform fill null values(fillna method) with the desired value
- Data Reduction: Remove unwanted data (if present) which are not required for analysis
    - Delete unwanted columns
    - Delete duplicate rows
- Format data types (numerical & categorical variables)
- Outlier detection and handling (we ignore this step because outliers are valid in our case)
    - When data has extreme values that could effect our analysis, we either replace them with Mean or Median or Mode values or we accept the outliers
    - We identify the outliers by plotting the Box plot

Handle missing values (imputation or deletion)

- df.isnull().sum()- Gives sum of all null values in each column
- df.notnull().sum()- Gives sum of all not null unique values in each column
- Interpretation:
    - Data has no null values, so no need to perform process to handle missing values
- But In Our data There is no null values or any other Duplicate Values means The data Provided is Properly Cleaned and well Managed .

Format data types (numerical & categorical)

We need to format columns, that will ease data analysis, below steps are performed to the required format for analysis

- Rename of columns- To keep columns descriptive as well as simple
- Change data types- We change data type to keep consistency and also for memory optimization

Feature Engineering (Create new features/variables)

- We derive new variables or features by combining multiple columns or derive new features by performing calculation

Create new features

We have ‘Direct Cost’, ‘Overhead Cost’, and ‘Sales Order’, which allow us to compute:

Total Cost = Direct Cost + Overhead Cost

Profit = Sales Order - Total Cost

Profit Margin (%) = (Profit / Sales Order) * 100

Data Analysis & Visualization

Overview of Data Before Analysis

After Data Wrangling, we check the dataset structure before proceeding with analysis.

Columns in DataFrame: ['Posting Date', 'Order No.', 'Item', 'Quantity', 'Department', 'Time', 'Direct Cost', 'Overhead Cost', 'Department Code', 'Sales Order']

Variables/Columns	Description
Posting Date	Date and time when the sales transaction was recorded
Order No.	Unique identifier for each sales transaction
Item	Name of the product sold
Quantity	Number of units sold in the transaction
Department	Business unit responsible for fulfilling the order
Time	Time of order placement
Direct Cost & Overhead Cost	Breakdown of operational expenses
Department Code	Unique identifier for each department
Sales Order	Total sales amount for the transaction
Year, Month, Day, Quarter	Extracted from Posting Date for time-based analysis
Total Cost	Sum of Direct Cost and Overhead Cost
Profit	Sales Order minus Total Cost
Profit Margin (%)	Percentage profit per transaction
Order Size	Number of times an order is placed per Order ID

Multivariate analysis

Sales Overview & Key Metrics

Monthly Sales Trend Over Years

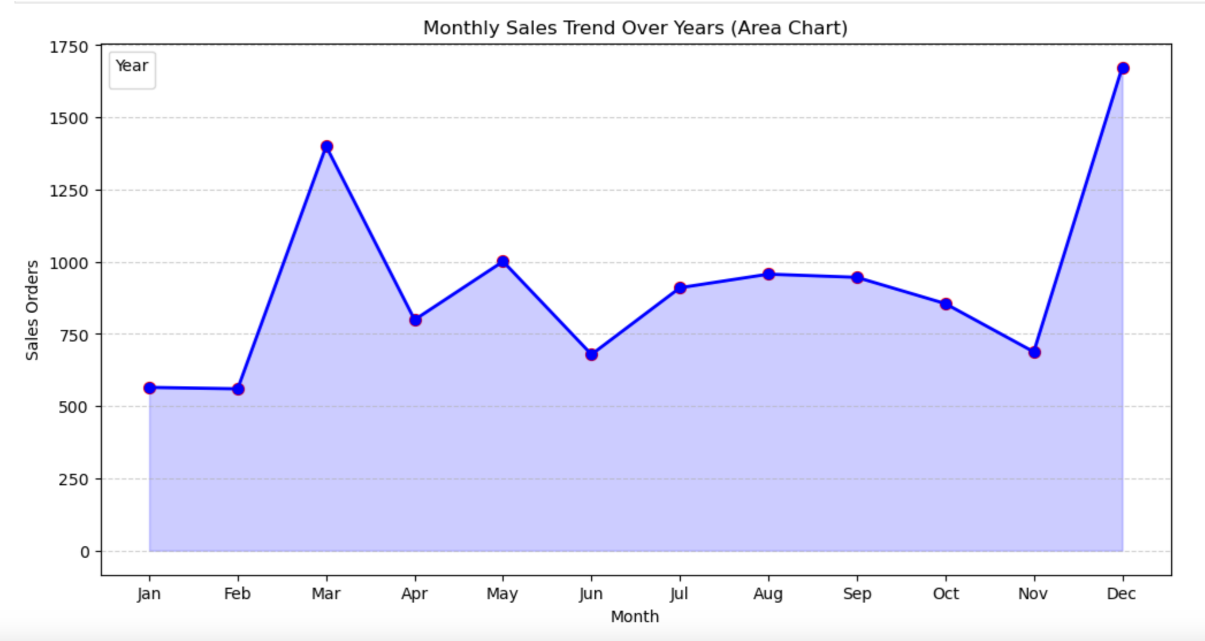

Figure 1 Monthly Sales Trend Over Years

Monthly Sales Trend Over Years

- The chart illustrates monthly sales trends over multiple years, with sales orders on the y-axis and months on the x-axis.
- A clear seasonal pattern is visible, with sales peaking in March and December while dipping in other months.
- March and December show the highest sales, indicating possible seasonal demand, promotional campaigns, or year-end sales.
- There is an upward trend from April to August, suggesting stable sales, followed by minor fluctuations until November.
- The area chart highlights overall sales volume, with noticeable variations between months.
- Further analysis can focus on reasons behind peaks (e.g., seasonal demand, marketing strategies) and dips (e.g., low customer interest, supply chain issues).

Monthly Count of Orders

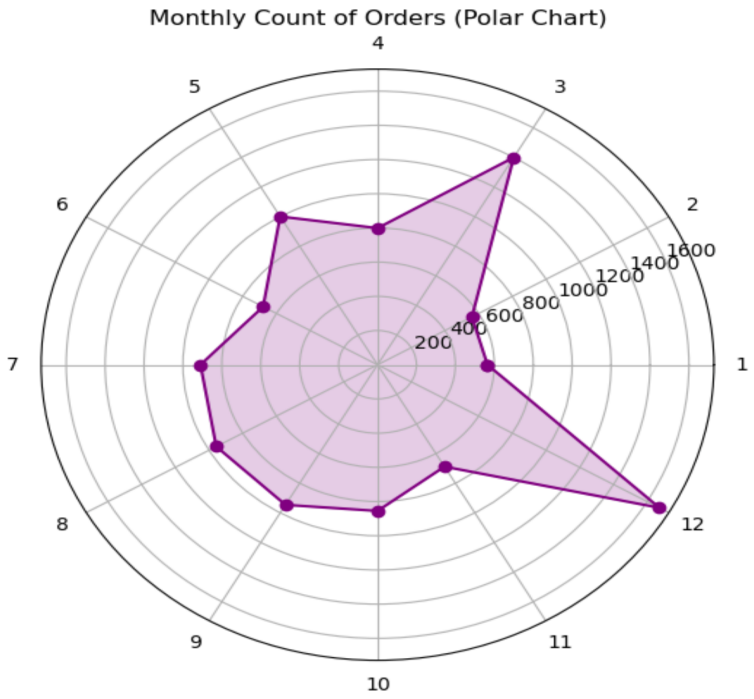

Figure 2 Monthly Count of Orders

Monthly Count of Orders

- This polar chart visualizes the distribution of monthly orders throughout the year.
- The spikes in March and December indicate higher order counts, suggesting seasonal peaks due to promotional campaigns, festive sales, or increased demand.
- The low order count in months like January and November suggests potential off-seasons or less consumer activity.
- The gradual rise from May to August indicates stable sales, while fluctuations in other months may be due to external factors like market trends or business cycles.
- The circular representation makes it easier to identify patterns and compare months intuitively.
- Further analysis can investigate reasons behind sales variations, such as marketing efforts, product demand, or external economic factors.

Sales Distribution by Department

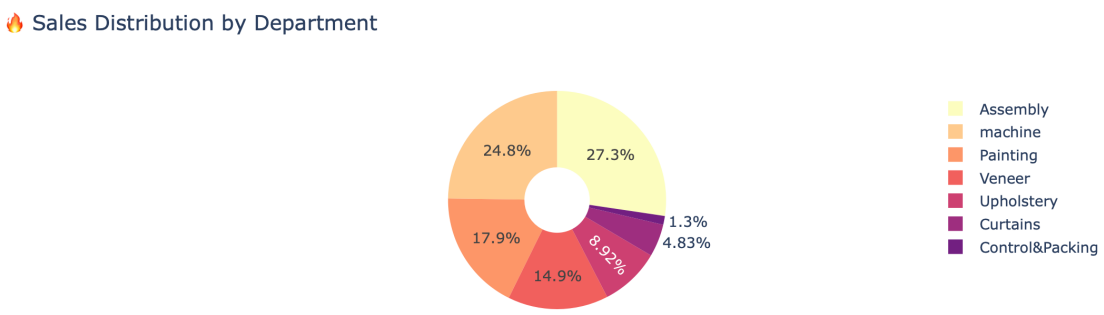

Figure 3 Sales Distribution by Department

Sales Distribution by Department

- The chart represents the proportion of sales contributed by different departments.
- Assembly (27.3%) is the highest contributor, highlighting its major role in sales.
    - Machine (24.8%) follows closely, indicating strong demand for mechanical processes.
    - Painting (17.9%) and Veneer (14.9%) also contribute significantly, suggesting their importance in the production process.
    - Upholstery (8.92%) and Control & Packing (4.83%) play supporting roles but still contribute.
    - Curtains (1.3%) have the lowest sales, indicating possible low demand or limited product range.
- Departments with higher sales share should be optimized for efficiency and cost management.
- Lower-performing departments (Curtains & Control & Packing) may need marketing strategies or product line adjustments.
- Understanding demand trends across departments can help resource allocation, inventory management, and pricing strategies.

Top Selling & High-Cost Items

Monthly Sales Quantity by Department

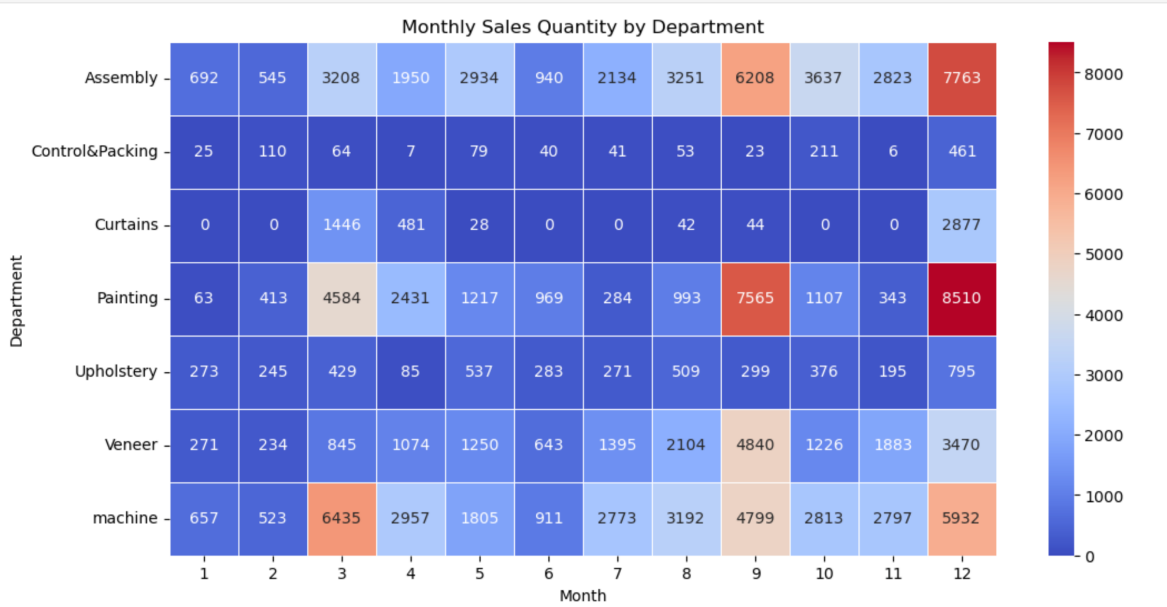

Figure 4 Monthly Sales Distribution by Department

Monthly Sales Quantity by Department

- The heatmap visualizes monthly sales quantities across different departments, with colors indicating intensity (blue = lower sales, red = higher sales).
- Painting department (December: 8,510 units, September: 7,565 units) shows the highest sales volumes.
- • Assembly (December: 7,763 units, September: 6,208 units) is also a strong performer.
- • Machine department (April: 6,435 units, December: 5,932 units) sees peak activity in April and December.
- Control & Packing and Curtains consistently have low sales throughout the year, suggesting either lower demand or limited production capacity.
- • Upholstery has moderate but steady sales, without any significant spikes.
- December experiences peak sales across multiple departments, indicating a year-end demand surge.
- • April and September also show increased sales activity, particularly in Machine and Painting departments.
- • January and February have the lowest sales across most departments, possibly due to seasonal slowdowns.

Top 10 Items with Highest Cost

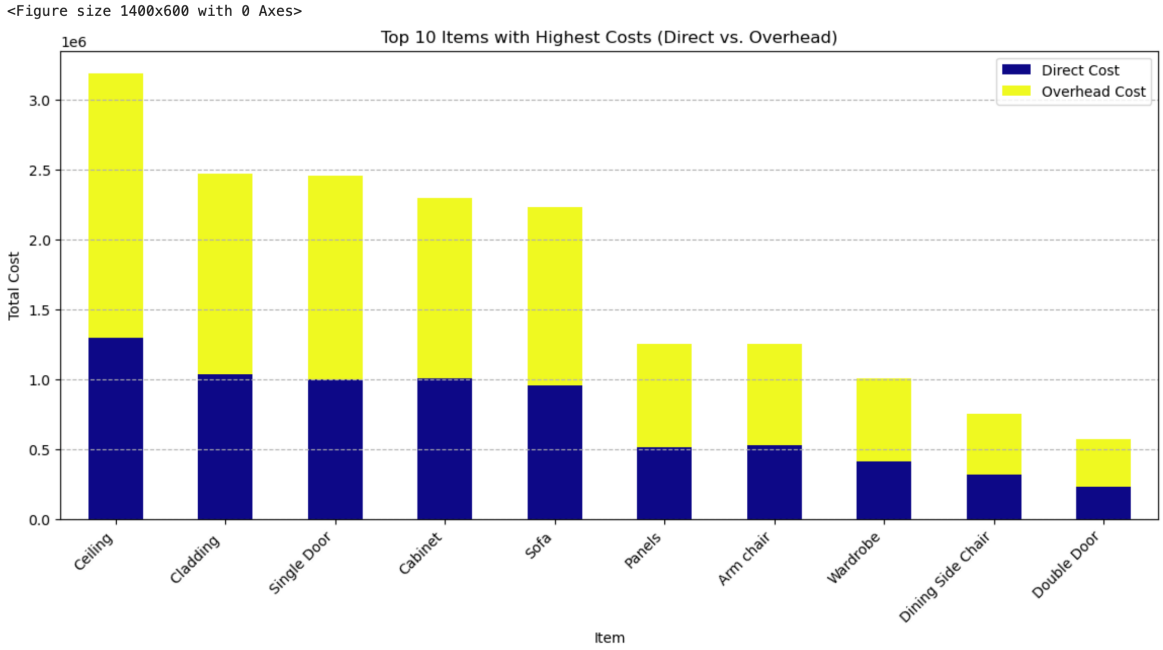

Figure 5 Top 10 Items with Highest Cost

Top 10 Items with Highest Cost

- Ceiling has the highest total cost, exceeding 3 K units, with nearly equal portions of direct and overhead costs.
- Cladding, Single Door, and Cabinet follow, each with over 2.5 K units in cost.
- Sofa and Panels are also among the top cost-intensive items.
- Overhead costs (yellow) contribute significantly to total costs, especially for Sofa and Ceiling, indicating higher operational expenses.
- Items like Wardrobe and Double Door have a more balanced split between direct and overhead costs.
- High overhead costs suggest opportunities for process optimization, potentially through efficiency improvements or better cost allocation.
- • Furniture-related items (Sofa, Arm Chair, Wardrobe) have substantial costs, indicating a focus area for production cost reduction.

Top 10 Revenue-Generating Items

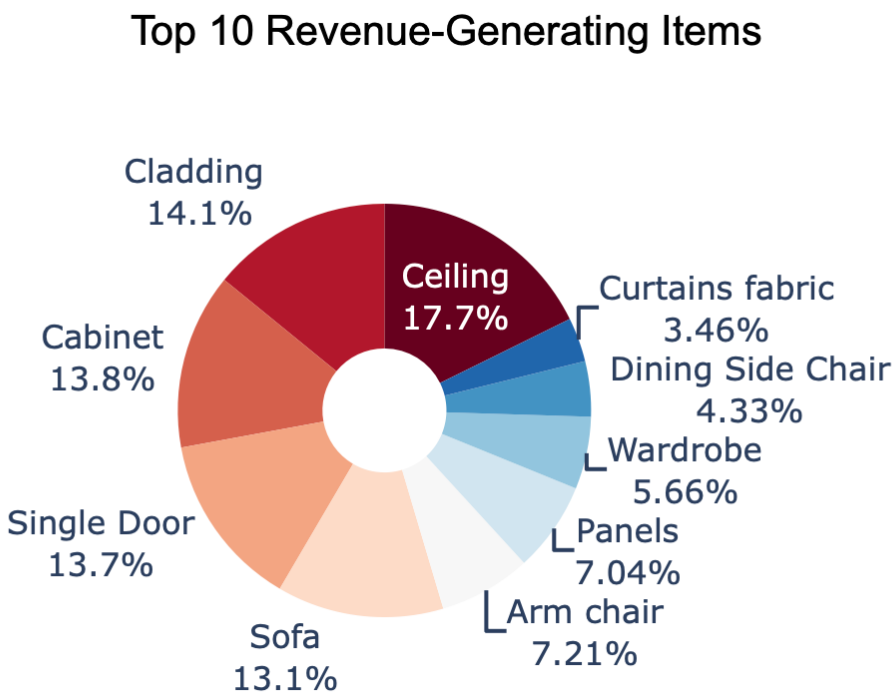

Figure 6 Top 10 Revenue-Generating Items

Top 10 Revenue-Generating Items

- This donut chart represents the top 10 items that generate the highest revenue, based on direct cost contributions.
- • Ceiling is the highest revenue-generating item, contributing 17.7% of the total revenue.
- • Other major contributors include Cladding (14.1%), Cabinet (13.8%), Single Door (13.7%), and Sofa (13.1%).
- • Items like Arm Chair (7.21%), Panels (7.04%), and Wardrobe (5.66%) contribute a moderate share.
- • Dining Side Chair (4.33%) and Curtains Fabric (3.46%) contribute the least among the top 10.

Business Insights:

- High-revenue items like Ceiling, Cladding, and Cabinet should be prioritized for stock availability and marketing.
- • Low-revenue items like Curtains Fabric may need further analysis to assess demand and pricing strategies.
- • This analysis can help optimize inventory, pricing, and marketing efforts to maximize overall revenue.

Sales Performance by Department

Monthly Quantity Sold by Department

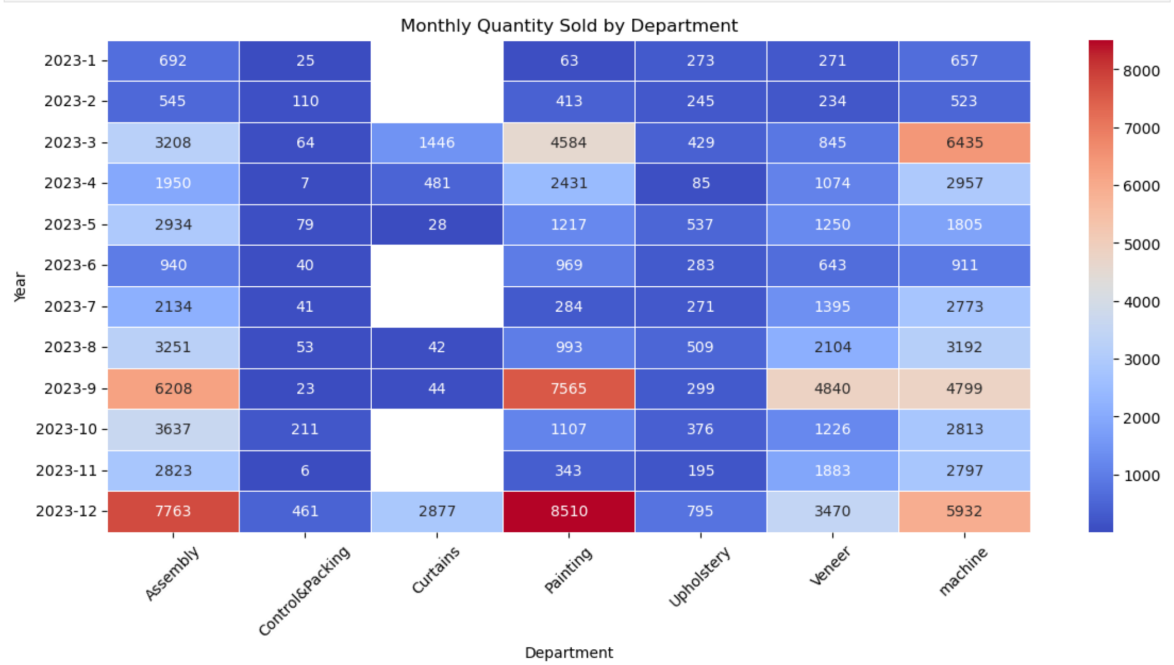

Figure 7 Monthly Quantity Sold by Department

Monthly Quantity Sold by Department

- Painting Department peaked in December (8510 units) and September (7565 units).
- Assembly Department had significant sales in December (7763 units) and September (6208 units).
- These peaks may indicate seasonal trends or high production cycles.
- Machine Department had high sales in March (6435 units) and December (5932 units).
- Veneer Department had peak sales in September (4840 units) and December (3470 units).
- Control & Packing Department consistently had the lowest sales across all months.
- Curtains Department showed a minor peak in December (2877 units).
- Upholstery Department had the highest sales in December (795 units), but overall sales remain lower than other departments.

Average Order Size by Department

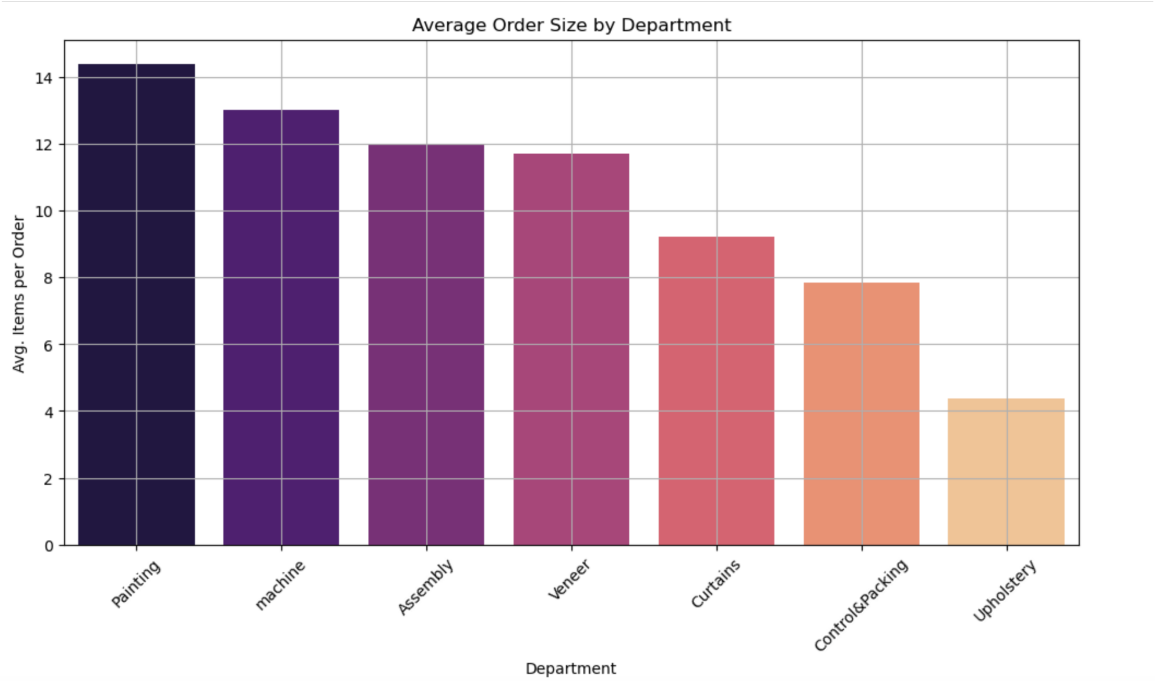

Figure 8 Average Order Size by Department

Average Order Size by Department

- Painting Department has the highest average order size (~14 items per order)
    - Indicates bulk ordering trends, likely due to high production demands.
- Machine, Assembly, and Veneer Departments have high order sizes (12-13 items per order)
    - Suggests consistent demand for bulk purchases.
    - May indicate standardized production workflows requiring large quantities.
- Curtains and Control & Packing Departments have moderate order sizes (~7-9 items per order)
    - These departments may process smaller, customized, or less frequent orders.
- Upholstery Department has the lowest average order size (~4 items per order)
    - Could indicate customized orders or less frequent demand.

Profitability Analysis

Top 10 Items by Absolute Profit

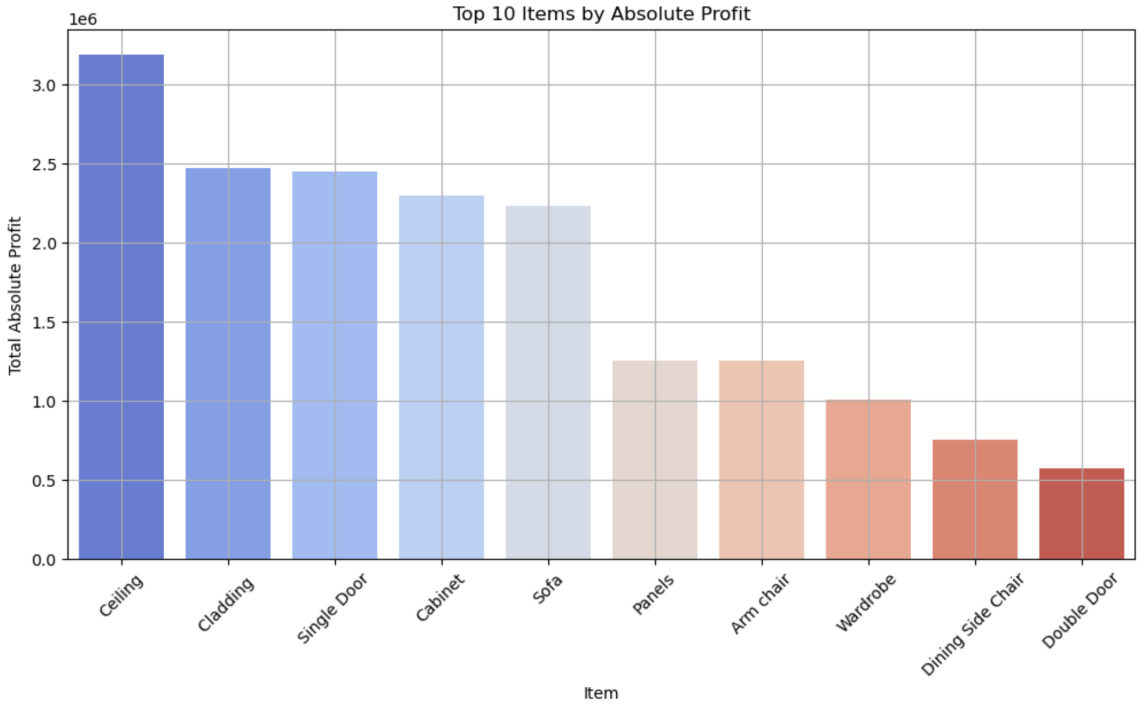

Figure 9 Top 10 items by Absolute Profit

Top 10 Items by Absolute Profit

- Ceiling generates the highest absolute profit (~3.2M)
  - Indicates it is a high-margin or high-demand product.
- Cladding, Single Door, and Cabinet follow closely (~2.5M)
  - These products contribute significantly to total profit.
- Sofa has a slightly lower absolute profit (~2.2M)
  - Could be due to moderate sales volume or lower profit margins.
- Panels, Arm Chair, and Wardrobe fall in the mid-range (~1.2M - 1.5M)
  - Indicating steady profitability but lower than top items.
- Dining Side Chair and Double Door have the lowest absolute profit (~0.5M - 0.7M)
  - These items may have lower margins or lower sales volume.

Business Insights:

- Maximize High-Profit Items: Focus marketing and production efforts on Ceiling, Cladding, and Single Door.
- Profit Margin Optimization: Review pricing strategy for lower-ranked items.
- Diversification Strategy: Consider bundling lower-profit items with bestsellers to increase total profitability.

Sales Contribution by Item

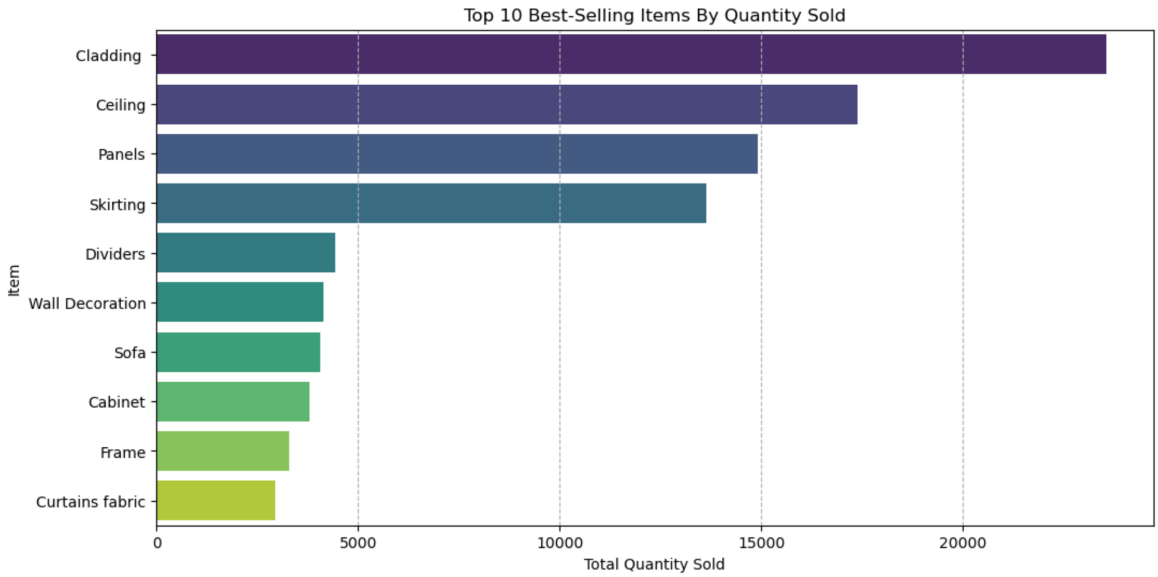

Figure 10 Sales Distribution by Item

Sales Contribution by Item

- The bar chart illustrates the top 10 best-selling items by quantity sold. “Cladding” emerges as the highest-selling item, followed closely by “Ceiling” and “Panels,” indicating a strong demand for construction and interior design materials.
- These top-ranked items significantly outperform others in terms of sales volume, highlighting their importance in overall business revenue.
- Additionally, there is a noticeable sales gap between the top-selling and lower-ranked items, such as “Curtains Fabric” and “Frame,” suggesting that a few key products contribute disproportionately to total sales.
- This insight can be leveraged for inventory management and supply chain optimization, ensuring that high-demand products are always well-stocked while assessing the profitability of lower-selling items.

Customer Purchasing Behavior

Top Purchased Items (Funnel Chart)

Figure 11 top Purchased items

Top Purchased Items (Funnel Chart)

Top Purchased Items

- Sofa (1,063 units) is the most purchased item.
- Arm Chair (820 units) follows as the second most purchased.
- Cabinet (684 units) and Single Door (618 units) also have high purchase volumes.

Mid-Tier Purchased Items

- Cladding (576 units) and Round Side Table (488 units) show moderate purchase activity.
- Square Side Table (328 units) and Curtains Fabric (319 units) have similar sales figures.
- Least Purchased Items
- Mirror (315 units) and Buffet (287 units) have the lowest purchase counts among the top 10.

Observations & Business Insights

- High Demand for Furniture: Items like Sofas, Arm Chairs, and Cabinets dominate, suggesting high customer preference for home furnishings.
- Diverse Product Mix: While furniture leads, items like Cladding (used for interiors) and Curtains Fabric show demand in home improvement.
- Opportunities for Inventory Management: Knowing which items sell the most helps optimize stock levels and reduce excess inventory for less popular products.

Flow of Sales Across Departments

Sales Distribution Across Departments

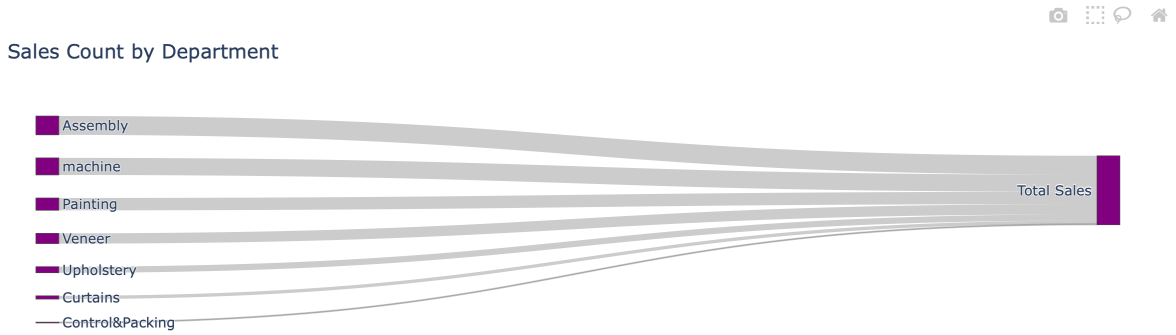

Figure 12 Sales Distribution Across Department

Sales Distribution Across Departments

- Multiple Departments Contribute to Sales:
    - The departments listed—Assembly, Machine, Painting, Veneer, Upholstery, Curtains, and Control & Packing—all contribute to the total sales count.
- Thicker Lines Indicate Higher Contribution:
    - Assembly, Machine, and Painting seem to have the thickest flows, suggesting they contribute the most to total sales.
- Control & Packing has the thinnest flow, indicating a lower sales count.
    - Balanced Distribution with Variations:
- While sales are spread across departments, some departments (e.g., Painting and Assembly) play a larger role in total sales than others.
    - Departments like Curtains and Control & Packing contribute relatively less.

Business Insights

- Top-performing departments (e.g., Assembly, Machine, and Painting) should be optimized further to maintain efficiency.
- Lower-performing departments (e.g., Curtains, Control & Packing) may need strategic interventions such as promotions or process improvements.
- Investigate seasonal trends—Do some departments contribute more during certain months?
- Resource Allocation: Direct more resources and marketing efforts towards high-impact departments to maximize profitability.

Yearly Order count by Department

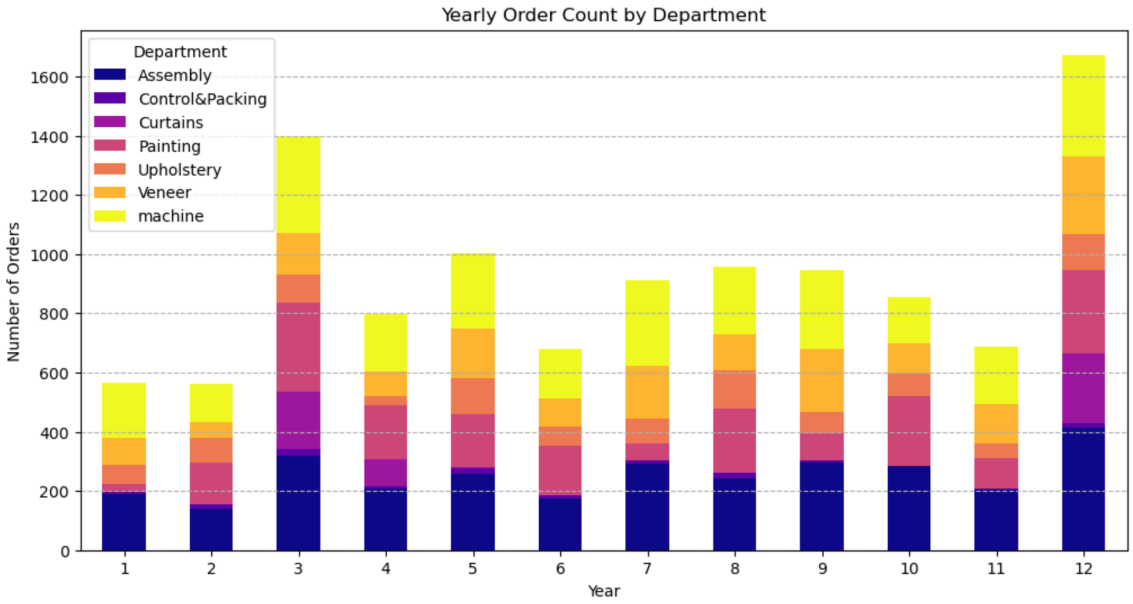

Figure 13 Yearly Order Count by Department

Yearly Order count by Department

- The bar chart represents the number of orders placed each month (1 to 12) across different departments.
- There are fluctuations in order volume across the year.
- Month 3 and Month 12 have the highest order counts, indicating seasonal peaks in sales.
- Month 11 has a dip, suggesting a slowdown before a final surge in December.
- Machine (Yellow): Contributes significantly in peak months (3 & 12).
- Veneer (Orange) & Upholstery (Pink): Play a steady role throughout the year.
- Painting (Red) & Curtains (Purple): Show variable contributions, indicating possible seasonal demand.
- Assembly (Dark Blue) & Control & Packing (Navy Blue): Maintain a consistent order count across months.

Summary and Suggestions

Based on the analysis, here are my key summary and suggestions for improving business performance:

Sales and Demand Analysis

Summary:

- Top-Selling Products: A small group of products contribute significantly to overall sales.
- Revenue vs. Quantity: High-selling items do not always have the highest revenue due to varying pricing strategies.
- Monthly Sales Trends: Sales volumes exhibit fluctuations, with noticeable peaks in March and December.
- Department-wise Contribution: Certain departments, such as Machine, Upholstery, and Veneer, show higher order volumes than others.
- Order Patterns: A significant portion of orders have small order sizes, indicating frequent but lower-quantity purchases.

Suggestions:

- Optimize inventory management for high-selling products to prevent stockouts.
- Target marketing campaigns during peak months to capitalize on increased customer demand.
- Improve pricing strategies for high-volume but low-revenue items to maximize profitability.
- Strengthen production planning for top-performing departments to meet demand effectively.

Cost and Profitability Insights

Summary:

- Costliest Products: Some items have significantly high costs but do not necessarily yield high profits.
- Absolute Profitability: Top revenue-generating products do not always align with top profit-making products.
- Discount Impact: Higher discounts negatively impact overall profit margins.
- Production Costs: Material and labor costs per unit vary widely, affecting product margins.

Suggestions:

- Re-evaluate pricing models to balance revenue and profit margins effectively.
- Optimize discount strategies by limiting high-discount offers on low-margin items.
- Improve cost efficiency in production by minimizing material wastage and labor inefficiencies.
- Encourage bulk purchases with volume-based discounting instead of flat-rate discounts.

Production and Efficiency Analysis

Summary:

- Production Volume vs. Scrap Rate: Higher production volumes sometimes lead to increased scrap rates.
- Machine Performance: Some machines contribute to higher defect rates and downtime.
- Energy Consumption: Certain production processes consume more energy, impacting overall costs.
- Maintenance Hours: Unexpected maintenance hours increase downtime and reduce efficiency.

Suggestions:

- Implement predictive maintenance to reduce unplanned downtime.
- Monitor and optimize machine performance to minimize defect rates.
- Reduce energy consumption by optimizing process efficiency and adopting energy-saving techniques.
- Analyze scrap generation to identify process improvements and cost savings.

Department-Wise Performance

Summary:

- Highly Contributing Departments: Machine, Upholstery, and Veneer departments have the highest sales contributions.
- Low-Performing Departments: Certain departments exhibit inconsistent sales and profitability.
- Sales Distribution: Some departments contribute more in revenue but have higher costs reducing net profit.
- Quality Checks & Defects: Some departments report higher failure rates in quality control checks.

Suggestions:

- Improve productivity in low-performing departments through better resource allocation.
- Enhance quality control measures to reduce defects and rework costs.

- Rebalance cost vs. revenue in departments where high sales do not translate to profits.
- Streamline supply chain operations for improved material flow across departments.

Order and Customer Behavior

Summary:

- Frequent Small Orders: Many customers prefer ordering smaller quantities.
- High-Value Customers: A subset of customers drive a significant portion of total sales.
- Product Bundling: Some products are frequently purchased together, indicating natural bundling opportunities.

Suggestions:

- Encourage larger order sizes through incentives like free shipping on bulk orders.
- Develop loyalty programs for repeat customers to enhance retention and engagement.
- Bundle complementary products to increase average order value and enhance customer satisfaction.

Strategic Recommendations

- Data-Driven Optimization: Continuously analyze production, sales, and profitability data to refine strategies.
- Focus on High-Impact Areas: Prioritize efforts on profitable departments, high-revenue products, and cost-saving measures.
- Refine Discount & Pricing Strategies: Avoid excessive discounts that erode profit margins while maintaining competitive pricing.
- Enhance Operational Efficiency: Minimize waste, optimize machine usage, and improve energy efficiency to reduce costs.
- Customer-Centric Strategies: Leverage customer buying patterns to offer better deals, targeted promotions, and personalized product recommendations.

GitHub Link:

EDA on Furniture & Home Decore Industry


FLY THE NEST