

SALES DATA ANALYSIS

- The sales dataset contains 10,000 rows and 21 columns (Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Discount, Profit).
- Each **row** in the dataset represents:
- A **single product line** from a particular order.
- If a customer ordered 3 different items in one order, it appears as **3 rows with the same Order ID**.

EXPLORATORY DATA ANALYSIS STEPS:

- **Initial Data Loading and Inspection:** Explain the initial steps of loading the dataset and performing basic inspections using `df.head()`, `df.dtypes`, `df.shape`, and `df.count()` to understand the data's structure, column types, dimensions, and completeness.
- **Data Cleaning - Dropping Columns:** Detail the rationale behind dropping the 'Postal Code' and 'Country' columns, which was done early in the process. Explain how these columns were identified as unnecessary for the subsequent analysis.
- **Data Cleaning - Duplicate Rows Check:** Explain the process of checking for and identifying duplicate rows using `df.duplicated()` and its importance in ensuring data integrity for analysis.
- **Feature Engineering - Date Columns:** Describe the feature engineering steps involving the 'Order Date' column, including converting it to datetime objects and extracting 'Year_Month' and 'Year' components for time-series analysis.
- **Feature Engineering - Profit Margin:** Explain the creation of the 'Profit_Margin' column, which is calculated as 'Profit' divided by 'Sales', and its significance for understanding product profitability.
- **Feature Engineering - Discount Buckets:** Detail the creation of 'Discount_Bucket' by categorizing the 'Discount' column into predefined ranges. Explain how this helps analyze the impact of different discount levels on profit.
- **Feature Engineering - Delivery Time:** Describe the calculation of 'Delivery_Time' by subtracting 'Order Date' from 'Ship Date', and how this new feature enables analysis of shipping efficiency.
- **Final Task:** Summarize the entire EDA process, highlighting its importance in preparing the data for meaningful visualization and analysis, and setting the foundation for the subsequent reports.

This is the step by step process of the changes made in the dataset.

EXPLANATION AND ANALYSIS OF GRAPHS:

- **Yearly Sales and Profit Trends:** This section analyzes the yearly sales and profit trends displayed in the bar charts. The objective is to identify overall business performance patterns, growth, or decline over the years.
- **Monthly Sales and Profit Trends:** This section analyzes the monthly sales and profit trends shown in the line plots. The objective is to discuss seasonality, consistent growth, or any observed decline over the months, providing insight into cyclical patterns.
- **Sales and Profit by Category:** This section explains the pie charts showing the share of total sales and profit by product category. The objective is to identify the best and worst performing categories based on their contribution to overall sales and profit.
- **Profit Distribution by Sub-Category:** This section discusses the box plot of profit distribution by sub-category. The objective is to identify sub-categories with high profitability, those with consistently low or negative profitability, and any significant outliers that might indicate unusual performance.
- **Sales by State and City Treemap:** This section explains the treemap visualizing sales by state and city. The objective is to identify top-performing regions and cities in terms of sales contribution at a glance.
- **Profit by State and City Treemap:** This section explains the treemap visualizing profit by state and city. The objective is to identify top-performing regions and cities in terms of profit, highlighting any discrepancies where a region might have high sales but low profit, or vice versa.
- **Sales and Profit by Customer Segment:** This section discusses the bar charts showing total sales and profit by customer segment. The objective is to identify which customer segments contribute most significantly to sales and profit, informing targeted marketing strategies.
- **Sales by Segment and Category Heatmap:** This section explains the heatmap showing sales by segment and category. The objective is to identify combinations of customer segments and product categories that generate particularly high sales, indicating strong market fit.
- **Top 10 Customers by Sales and Profit:** This section discusses the bar charts identifying the top 10 customers by sales and profit. The objective is to highlight key customers and quantify their significant contribution to the business.
- **Top and Bottom 10 Products by Sales:** This section explains the bar charts showing the top 10 and bottom 10 products by sales. The objective is to identify the best-selling products as well as the least performing ones, guiding inventory and marketing decisions.

- **Top 10 Products by Profit Margin:** This section discusses the bar chart of the top 10 products by profit margin. The objective is to identify products with the highest profitability relative to their sales, which are crucial for overall business health.
- **Sales vs Profit Margin (Product-Level) colored by Category:** This section explains the interactive scatter plot of sales vs. profit margin for each product, colored by category. The objective is to discuss products with high sales but low margin, and vice-versa, identifying underlying trends within different product categories.
- **Discount vs Profit Scatter Plot:** This section discusses the scatter plot of discount vs. profit, colored by category. The objective is to analyze the relationship between discount offered and the resulting profit, especially how this dynamic varies across different product categories.
- **Profit Distribution by Discount Level Buckets:** This section explains the box plot showing profit distribution by different discount ranges. The objective is to analyze how different discount levels impact the spread and median of profit, helping to optimize pricing strategies.
- **Average Profit by Category at Each Discount Range:** This section discusses the bar chart showing average profit by category at different discount ranges. The objective is to identify which categories remain profitable even with higher discounts and which ones suffer significant profit loss under similar conditions.
- **Count of Orders by Shipping Mode:** This section explains the bar chart showing the count of orders by shipping mode. The objective is to identify the most frequently used shipping modes, which can inform logistical planning.
- **Average Profit by Shipping Mode:** This section discusses the bar chart showing average profit by shipping mode. The objective is to identify which shipping modes are most profitable on average, potentially guiding shipping policy improvements.
- **Sales and Profit Share by Region:** This section explains the pie charts showing sales and profit share by region. The objective is to identify the regions that contribute most significantly to overall sales and profit, aiding regional strategic focus.
- **Region vs Category Sales Heatmap:** This section discusses the heatmap showing sales across regions and categories. The objective is to identify strong sales performing region-category combinations, highlighting areas of success.
- **Quantity per Order Distribution:** This section explains the histogram of quantity per order. The objective is to analyze the most common order quantities, which can inform packaging and inventory management.
- **Delivery Time Trend Over Months:** This section discusses the line chart showing the average delivery time trend over months. The objective is to identify any patterns or changes in delivery efficiency over time, such as improvements or seasonal delays.

- **Delivery Time Distribution:** This section explains the histogram of delivery time distribution. The objective is to analyze the typical delivery times and identify any outliers, indicating potential service issues or exceptional performance.
- **Category vs Region Total Profit Heatmap:** This section discusses the heatmap showing total profit across categories and regions. The objective is to identify both profitable and unprofitable category-region combinations, guiding strategic adjustments.
- **Category vs Region Profitability (Bubble Chart):** This section explains the scatter plot showing category vs. region profitability, sized by sales. The objective is to identify regions/categories with high sales but low or negative profit and vice versa, allowing for nuanced performance assessment.
- **Product Category & Sub-Category Profit Heatmap (Negative Areas Highlighted):** This section discusses the heatmap showing profit by product category and sub-category, specifically highlighting negative profit areas. The objective is to identify sub-categories that are frequently loss-making, even at low or zero discounts, indicating potential product or pricing issues.