Exploratory Data Analysis Report: Uncovering Insights to Improve Sales Forecasting

**Introduction**

* **Objective:** Understand sales trends, customer behavior, and profitability.
* **Dataset:** Contains sales transactions, customer details, and product information.

**Definition of EDA:**

Explain what Exploratory Data Analysis (**EDA**) is. Emphasize that it's an approach to analyzing datasets to summarize their main characteristics, often with visual methods. Mention that EDA helps to gain a better understanding of the data, identify patterns, spot anomalies, and check assumptions.

**Dataset Description:**

This dataset contains **51290 records** and **24 columns**, covering various aspects of **sales transactions**. It includes details about **customer orders**, including **sales, discounts, profits, shipping costs, and product categories**. Additional features such as **season, weekday, economic indicators, and promotions** are also present.

**Purpose:**

The dataset is designed for Exploratory Data Analysis (EDA) to uncover insights into sales performance, profitability, customer behavior, and the impact of discounts and shipping costs on business revenue. This analysis is crucial within the context of sales forecasting as it allows us to understand the nature of sales data (e.g., trends, seasonality), identify factors that may influence sales (e.g., promotions, seasonality, external events), and check data quality for potential issues like missing values or outliers. Ultimately, EDA informs the feature engineering and model selection process, helping businesses optimize pricing, improve marketing strategies, enhance operational efficiency, and gain insights that lead to better, data-driven decisions.

**Key Insights and EDA charts:**

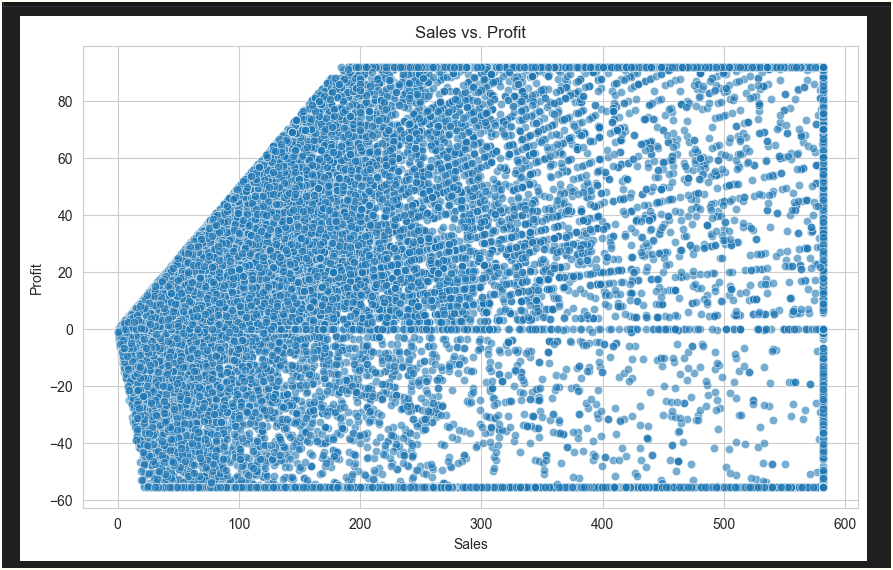
1. **Sales vs. Profit** (Scatter Plot)
   * There is a **positive correlation** between sales and profit.
   * Higher sales generally lead to higher profits, but some **high-sales transactions have low or negative profits**, indicating possible discounting issues or cost inefficiencies.
2. **Total Sales by Category** (Bar Chart)
   * Some product categories contribute significantly more to total sales than others.
   * The **top-performing categories** should be prioritized in marketing and inventory management.
3. **Discount vs. Profit** (Scatter Plot)
   * A **negative correlation** exists between discount percentage and profit.
   * Higher discounts tend to reduce profitability, suggesting that discounting strategies should be carefully optimized.
4. **Sales Distribution by Season** (Box Plot)
   * Sales tend to fluctuate across different seasons, with peaks in specific quarters.
   * This insight can help businesses plan **seasonal promotions** and **stock adjustments**.
5. **Total Sales by Ship Mode** (Bar Chart)
   * Some shipping methods generate higher sales than others.
   * Businesses may want to **incentivize cost-effective shipping options** that maintain customer satisfaction.
6. **Quantity vs. Sales** (Scatter Plot)
   * There is a pattern indicating that **bulk purchases lead to higher total sales**.
   * This can guide strategies for bulk pricing or wholesale promotions.
7. **Region-wise Profit** (Bar Chart)
   * Some regions are more profitable than others.
   * Companies can adjust their **regional sales strategies** and **marketing efforts** accordingly.
8. **Shipping Cost vs. Sales** (Scatter Plot)
   * No strong correlation, meaning **high shipping costs do not necessarily result in higher sales**.
   * Businesses should evaluate if shipping cost reductions could improve customer demand.
9. **Average Sales by Discount and Profit Bins** (Heatmap)

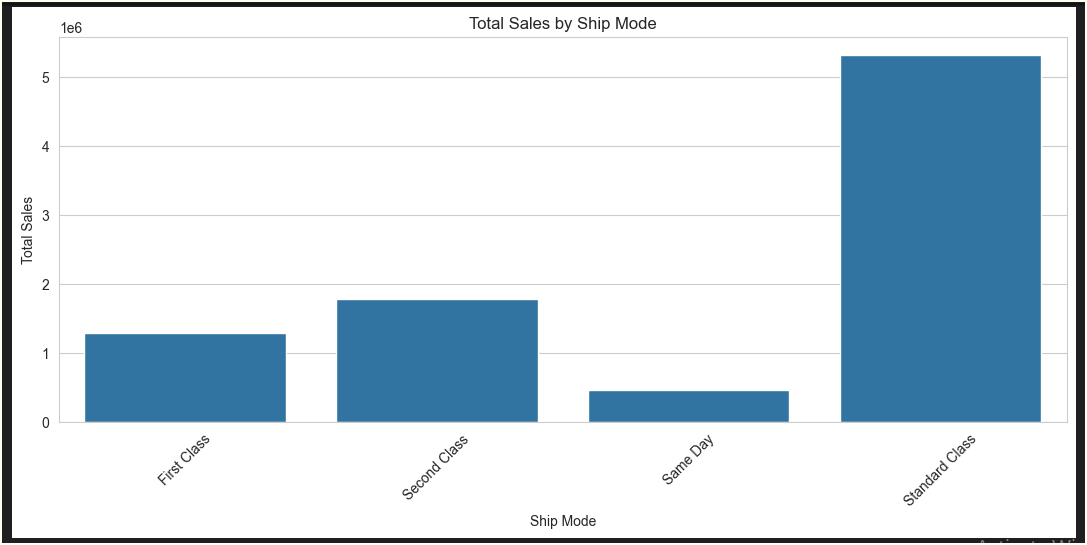
* Show the correlation between decline and profit levels on average sales using a cloud map.
* **What's the benefit?** 
  + 1. It helps you understand the relationship between discounts, profitability, and their impact on sales.
* **You can easily see:**

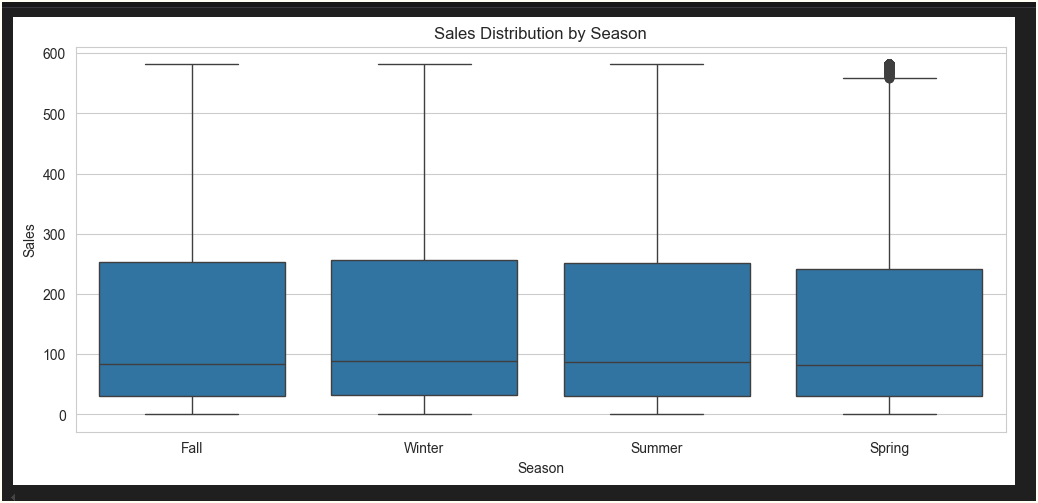
1. Do high discounts lead to higher sales?
2. Is there a certain level of profit or discount where sales are excellent?
3. Are there areas in the table where sales are very low, so you can avoid them?
4. **Understand the frequency of order sizes** (Histogram)

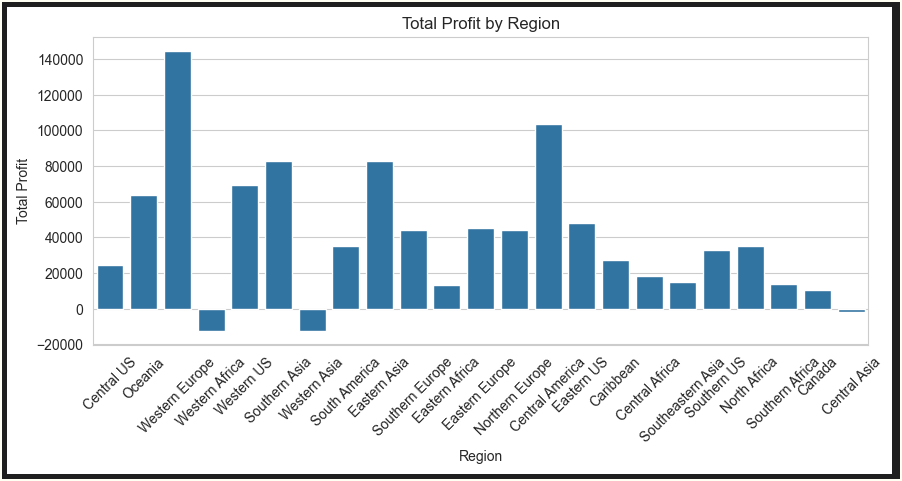
* Understand the frequency of different order quantities — how many times people ordered 1 item, how many times they ordered 5, and so on.
* **Practical Benefits**
  + 1. You can identify the most common quantity customers order. For example, if most orders are for 2 or 3 items, this is a clear pattern.
    2. You can determine whether the distribution is balanced or biased toward one side (i.e., more small or large orders).
    3. You can use the results to: Optimize inventory, Target promotions, Improve future demand forecasting.

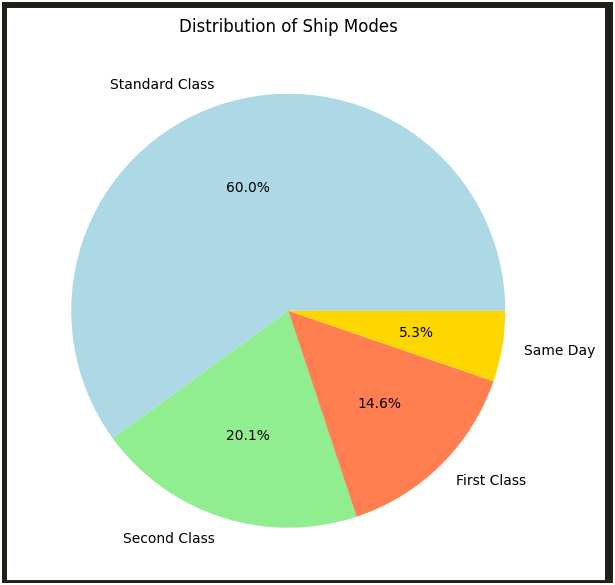
**Graphs:**

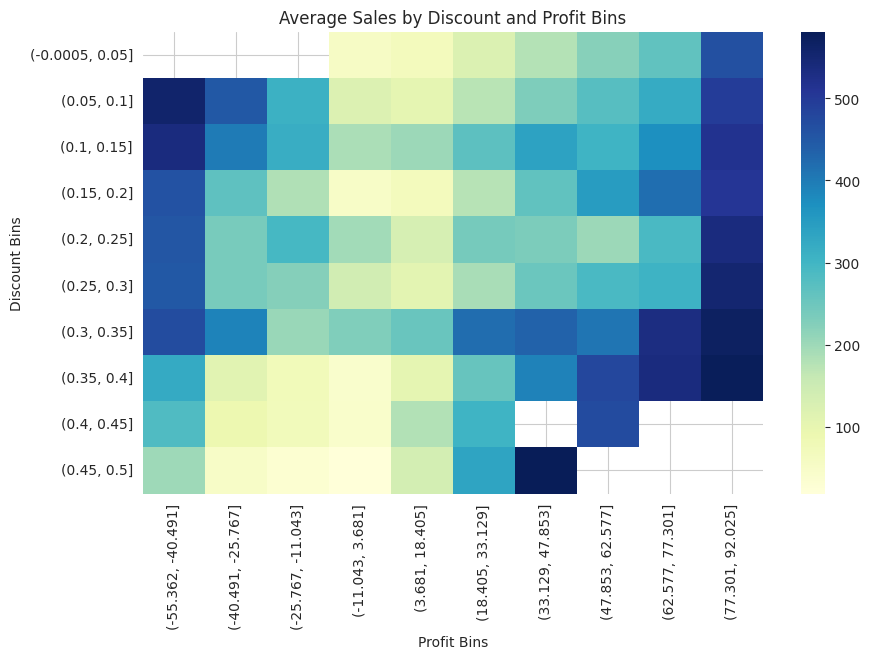
****

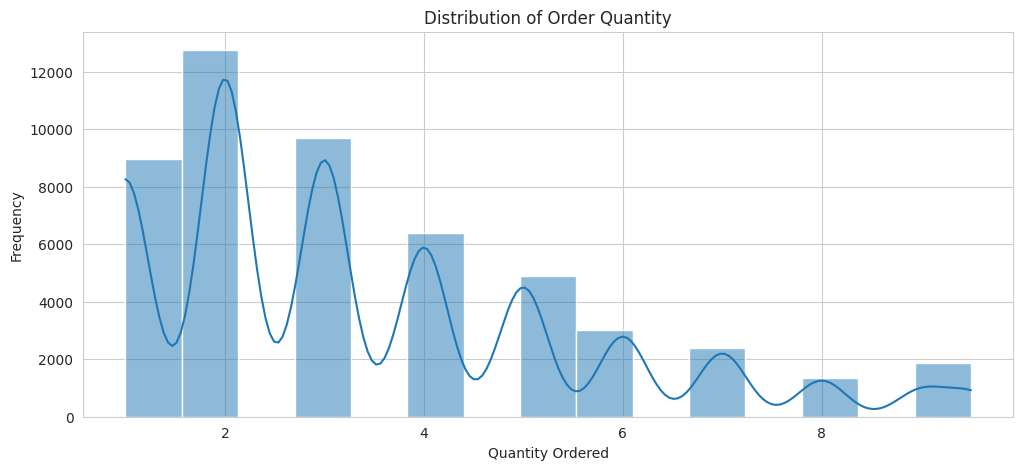
****

****

****

****

****

****

**Recommendations:**

1. **Optimize Discount Strategies**
   * Implement tiered discounting to ensure profitability, avoiding excessive discounts that lead to losses.
   * Analyze which discount levels yield the highest ROI and adjust promotions accordingly.
2. **Reassess Shipping Costs**
   * Investigate whether **high shipping costs correlate with lower sales**.
   * Offer free or discounted shipping on high-value purchases to incentivize more orders.
3. **Enhance Promotions for High-Profit Items**
   * Promotions should focus on **high-margin products** rather than indiscriminately discounting items.
4. **Leverage Seasonal Sales Trends**
   * Allocate marketing budgets for peak seasons where sales naturally increase (e.g., Q4 holiday shopping).
5. **Further Exploration of Customer Segments**
   * Identify high-value customer segments and target them with personalized promotions.
   * Consider customer purchase history to predict future buying behavior and tailor marketing strategies.