US Regional Sales Data Analysis

Introduction

This analysis explores US regional sales data to understand the performance of different sales channels, product popularity, customer buying patterns, warehouse efficiency, and pricing strategies. Using US_Regional_Sales_Data.csv, we derive insights through data cleaning, feature engineering, and various visualizations.

Dataset Overview

The dataset contains **7,991 rows** and **16 columns** including details such as Order Number, Sales Channel, Warehouse Code, Order Date, Ship Date, Delivery Date, Discount Applied, Unit Cost, Unit Price, and more. The primary objectives of this analysis include:

1. **Sales Channel Performance** - Analyzing revenue and discounts across sales channels.
2. **Trend Analysis** - Monitoring monthly order volume.
3. **Customer Segmentation** - Grouping customers by average order quantity and sales channel.
4. **Product Performance** - Identifying top products by revenue.
5. **Warehouse Efficiency** - Examining lead time to improve logistics.
6. **Pricing Strategy Optimization** - Assessing the impact of discounts on revenue.
7. **Order Fulfillment** - Understanding delays across the order fulfillment stages.
8. **Regional Interactions** - Visualizing sales channels and warehouse regions interaction.

Data Preparation and Cleaning

1. **Data Loading**: The dataset was loaded using pandas and checked for dimensions and initial rows.
2. **Data Transformation**:
    - The Unit Cost and Unit Price columns contained commas which were removed and then converted to float types for calculations.
3. **Feature Engineering**:
    - **Revenue Calculation**: Created a Revenue column as Unit Price * Order Quantity.

    - **Profit Margin**: Calculated as (Unit Price - Unit Cost) * Order Quantity.
    - **Lead Time Calculation**: Derived from the difference between Ship Date and Order Date.

Key Analysis and Findings

1. Sales Channel Performance Analysis

- **Objective**: To assess revenue generated by different sales channels (In-Store, Online, Distributor, and Wholesale) and identify the most profitable channels.
- **Methodology**: Sales data was aggregated by channel, calculating metrics such as total revenue, average discount applied, and total order count. A bar plot visually represents revenue by channel.
- **Results**:
    - The **In-Store** channel showed strong performance in terms of revenue with moderate discount rates, appealing to direct customers.
    - The **Online** and **Wholesale** channels offered high accessibility, contributing significantly to revenue but often involved higher discounts.
    - **Distributor sales** had a balanced performance with steady revenue contributions.
- **Insight**: The Online channel, though widely accessible, could be optimized by balancing discounts to increase profitability.

2. Monthly Order Volume Trend

- **Objective**: To uncover patterns in order volumes over time and identify peak months or seasonal trends.
- **Methodology**: Monthly resampling of the order data was performed to compute the total order count per month, which was then plotted as a line trend.
- **Results**:
    - A clear **seasonal trend** emerged, with peak orders in the summer months, suggesting higher demand in mid-year.
    - Some troughs appeared in winter, indicating periods of low sales volume, possibly due to reduced seasonal demand.
- **Insight**: These insights could guide inventory management and marketing, allowing stock increases or promotional events during peak months.

3. Customer Segmentation by Sales Channel and Order Quantity

- **Objective**: To categorize customers based on their average order quantity within each sales channel.

- **Methodology**: Customers were grouped by sales channel, and their average order quantity was calculated. A box plot was used to depict variations and outliers across channels.
- **Results**:
    - **Wholesale customers** tend to place larger orders, while **Online customers** had smaller, more frequent orders.
    - **Distributor** customers had moderate order quantities but higher consistency in purchase frequency.
- **Insight**: Understanding customer purchase patterns across channels helps tailor strategies for customer retention and targeted promotions.

4. Product Performance Analysis

- **Objective**: To identify top-performing products by revenue and assess product demand.
- **Methodology**: Products were ranked based on total revenue generated and order frequency. The top 10 products were then highlighted.
- **Results**:
    - The top products contributed to a substantial portion of total revenue, indicating a high reliance on these items.
    - Certain products had a consistently high order volume, suggesting they are popular among regular customers.
- **Insight**: This analysis helps in deciding on product prioritization for promotions, stocking, and bundling.

5. Warehouse Efficiency Analysis

- **Objective**: To measure the average lead time (Order Date to Ship Date) for each warehouse and identify inefficiencies.
- **Methodology**: Lead time was calculated as the difference between the Order Date and Ship Date. Each warehouse's average lead time was computed, and the results were plotted in a bar chart.
- **Results**:
    - Warehouses with shorter lead times performed more efficiently, enabling quicker order fulfillment.
    - A few warehouses exhibited longer lead times, indicating potential bottlenecks in the supply chain.
- **Insight**: Optimizing lead time in certain warehouses could reduce fulfillment delays and enhance customer satisfaction.

6. Pricing Strategy Optimization: Discount and Revenue Impact

- **Objective**: To analyze the relationship between discount levels and revenue across different channels, informing discount strategies.

- **Methodology**: The average discount applied in each sales channel was plotted against total revenue using a scatter plot.
- **Results**:
    - Higher discounts correlated with increased revenue in Online and Wholesale channels, though at the cost of lower profit margins.
    - Lower discounts in the In-Store channel did not significantly impact revenue, showing that customers value the in-store experience over discounts.
- **Insight**: By adjusting discount strategies, especially in Online and Wholesale channels, it may be possible to achieve a better revenue-to-discount ratio.

7. Customer Lifetime Value (CLV) Analysis

- **Objective**: To calculate the lifetime value of customers by examining order frequency and total revenue per customer.
- **Methodology**: Customer CLV was calculated as the total revenue generated by each customer, correlated with their order frequency. A scatter plot was used for visual representation.
- **Results**:
    - Certain customers displayed high revenue with frequent purchases, making them valuable, repeat customers.
    - Low-frequency customers contributed sporadic revenue, indicating room for engagement initiatives.
- **Insight**: The company could target high-CLV customers for loyalty programs while devising strategies to increase order frequency among infrequent buyers.

8. Customer Buying Patterns Analysis

- **Objective**: To understand buying patterns by examining total order quantities and average order sizes for each customer.
- **Methodology**: Distribution analysis was performed on total quantities ordered by customers, with histogram visualization.
- **Results**:
    - A small segment of customers made very large purchases, while most customers had moderate order sizes.
    - Average order size was relatively consistent, indicating stable purchasing behavior.
- **Insight**: Large-purchase customers represent a potential area for bulk discounts or special incentives, while the majority could be encouraged with product bundles or upselling.

9. Order Fulfillment Analysis: Delays and Bottlenecks

- **Objective**: To assess delays in different stages of order fulfillment, identifying bottlenecks.

- **Methodology**: Three stages (Procure-to-Order, Order-to-Ship, Ship-to-Delivery) were analyzed for average delays. Delays were then visualized to show areas with the longest lead times.
- **Results**:
    - The **Order-to-Ship** stage had the longest delays, highlighting a need for process improvements in warehouse operations.
    - **Ship-to-Delivery** delays were minimal, indicating reliable delivery logistics.
- **Insight**: Improvements in the Order-to-Ship process could reduce overall lead times, improving customer satisfaction and operational efficiency.

10. Sales Channel and Regional Interaction Analysis

- **Objective**: To explore the interaction between sales channels and warehouse regions by analyzing order quantities across regions.
- **Methodology**: A pivot table was created, with sales channels and warehouse codes forming a matrix to sum order quantities. The results were visualized as a heatmap.
- **Results**:
    - Certain warehouse regions (e.g., WARE-UHY1004) serviced more orders for specific channels like Distributor and Wholesale.
    - A few regions had balanced interactions across channels, indicating robust regional sales performance.
- **Insight**: Warehouse placements could be optimized based on these insights, with channel-focused stocking to reduce inter-regional order transfers.