

US Regional Sales Data Exploratory Analysis Report

Overview

This report presents the findings from an exploratory data analysis (EDA) of the "US Regional Sales Data". The analysis focuses on understanding sales trends, seasonality patterns, and relationships between various factors that influence sales performance.

Key Findings

1. Sales Channel Analysis

- **In-Store sales** generate the highest revenue at approximately \$29 million (50.5% of total revenue)
- **Online sales** are the second-highest channel at approximately \$21 million (36.5% of total revenue)
- **Wholesale** contributes the least at approximately \$7.8 million (13% of total revenue)

This suggests that while digital transformation is important, the physical retail presence remains the strongest revenue driver for the business.

2. Seasonality Patterns

The heatmap of average revenue by month and day of week reveals several interesting patterns:

- **Highest revenue day:** Thursday in February (\$12,041 average revenue)
- **Strong weekend performance:** Sundays in January, May, and November show high average revenue
- **Weekday patterns:** Thursdays generally perform well across most months
- **Monthly variations:** February, July, and November show strong performance across multiple days

These patterns can inform staffing decisions, inventory management, and promotional timing.

3. Impact of Discounts on Revenue

The analysis of discount rates versus revenue shows:

- A slight negative correlation between discount percentage and revenue
- Higher discounts (30-40%) generally result in lower average revenue than moderate discounts (5-15%)
- Significant variability exists at each discount level, suggesting other factors influence revenue beyond just discount rates.

This indicates that while discounts can drive sales, very high discounts may not be optimal for maximizing revenue.

Also, this does not mean people love to pay more but it may indicate that most of the products have a low discount percentage overall.

4. Order Quantity and Revenue Relationship

There is a strong positive relationship between order quantity and average revenue: - Revenue increases steadily as order quantity increases from 1 to 8 units - The average revenue for 8-unit orders (\$14,507) is more than 7 times higher than for single-unit orders (\$1,982) - The relationship appears to be approximately linear, with each additional unit adding similar incremental revenue

This suggests that strategies to increase order size could significantly boost revenue.

5. Warehouse Distribution

Revenue is not evenly distributed across warehouses:

- **WARE-NMK1003** generates the highest proportion of revenue (35.4%)
- **WARE-PUJ1005** accounts for 19.9% of revenue
- **WARE-UHY1004** contributes 18.4% of revenue
- **WARE-XYS1001** represents 16.9% of revenue
- **WARE-NBV1002** has the smallest share at 9.4% of revenue

This distribution may reflect differences in warehouse location, inventory management, or regional demand patterns.

Recommendations

Based on the analysis, the following strategies could potentially improve sales performance:

1. **Channel Optimization:** While maintaining strong in-store operations, invest in enhancing the online shopping experience to increase its revenue contribution.
2. **Seasonal Planning:** Align inventory levels, staffing, and promotions with the identified seasonal patterns, particularly focusing on high-revenue days like Thursdays.
3. **Discount Strategy:** Optimize discount rates around the 5-15% range where revenue appears to be maximized, and avoid excessive discounts above 30% unless clearing inventory is the primary goal.
4. **Order Size Incentives:** Implement strategies to encourage larger order quantities, such as volume discounts, bundle offers, or free shipping thresholds.
5. **Warehouse Efficiency:** Investigate the performance differences between warehouses to identify best practices from high-performing locations that could be applied across the network.

Additional Project Assets

- **EDA Notebook:**

a notebook with detailed analysis and visualizations can be accessed here:

[US Regional Sales Data EDA Notebook](#)

- **Interactive Dashboard:**

A dashboard summarizing key trends, seasonal patterns, and feature impacts is available here:

[Sales Data Dashboard](#)

- **Cleaned Datasets:**

The dataset has been cleaned, preprocessed, and made ready for forecasting and modeling tasks. And they are available here:

[Cleaned Datasets \(Google Drive Link\)](#)

Key actions included:

- Handling missing values and outliers.
- Encoding categorical variables.
- Creating time-based and promotional features.
- Saving finalized datasets for flexibility in modeling.

Available files:

- **US_Regional_Sales_Data.csv** (original raw dataset)
- **cleaned_sales_data_no_encoding.csv**
(cleaned but without encoding, for custom developer use)
- **cleaned_sales_data.csv** (fully cleaned and encoded, ready for modeling)