

Store Sales Dataset Analysis

1. Introduction

For this analysis, an open-source dataset from **Kaggle** was used, which contains detailed transactional sales data. The dataset is publicly available for educational and research purposes. The dataset includes multiple tables detailing customer orders, product categories, discounts, and sales rep assignments.

2. Data Collection

Data Sources

The data for this analysis was collected from internal customer relationship management (CRM) and sales databases. The dataset included the following fields:

- **Customer Information:** Custkey, customer_name, Customer Segment
- **Sales Data:** Sales Amount, DateKey, Item Class, Discount Amount, Sales Quantity
- **Product Information:** Item, Item Number, Item Category
- **Sales Rep Information:** Sales Rep, Region
- **Delivery Data:** Promised Delivery Date, Actual Delivery Date
- **Financials:** Sales Margin, Sales Cost, Sales Price

Data Description

The dataset comprises multiple tables detailing customer orders, item categories, applied discounts, sales transactions, and sales rep assignments. The primary objective was to derive actionable insights from this data, facilitating better decision-making on pricing, product focus, and customer segmentation.

3. Data Cleaning and Preparation

Data cleaning and preparation were crucial to ensure accurate analysis. The following steps were taken to prepare the data:

- **Handling Missing Data:** Missing values, especially in Sales Amount, Customer Name, and Sales Margin, were handled either by imputation or exclusion, depending on their relevance to the analysis.

- **Date Formatting:** DateKey and Invoice Date were initially stored as strings. These fields were converted into valid date formats using Tableau's DATEPARSE function to ensure accurate time-based analysis.
- **Outlier Detection:** Significant outliers, especially in Sales Amount and Sales Margin, were identified and reviewed. Extreme values were evaluated to ensure they were not erroneous before inclusion in the analysis.
- **Data Structuring:** Data was aggregated by key dimensions such as Item Class, Sales Rep, and Customer Segment to enable meaningful visualizations and comparisons across different categories.

4. Analytical Questions and Visualizations

Sales Performance Analysis

Question: What are the total sales over time (daily, weekly, monthly)?

- **Visualization:** *Line Chart* – Sales trends over time.
- **Insight:** The visualization revealed notable peaks in sales during key periods such as the end-of-year holiday season, suggesting seasonality in customer purchasing behavior.

Question: Which products are generating the highest revenue?

- **Visualization:** *Bar Chart* – Top 10 highest revenue-generating products.
- **Insight:** Certain product categories consistently outperformed others, contributing significantly to overall sales, thus highlighting opportunities for targeted promotions.

Question: How do sales fluctuate by product category and day of the week?

- **Visualization:** *Bar Chart and Day of Week Sales Trend*.
- **Insight:** Sales were found to peak on specific days of the week, with certain product categories being more popular on weekends.

Profitability Analysis

Question: What are the total profit margins across different products?

- **Visualization:** *Bar Chart* – Profit margin by product category.
- **Insight:** Identified high-margin products that consistently delivered strong profits, while some products with lower margins indicated opportunities for cost optimization.

Question: How do discounts affect profitability?

- **Visualization:** *Scatter Plot* – Discount Amount vs. Profit Margin.

- **Insight:** The analysis showed a negative correlation between high discount rates and profitability, particularly in certain product categories, indicating the need to review discount strategies.

Question: What is the Return on Investment (ROI) for each product?

- **Visualization:** *ROI Bar Chart*.
- **Insight:** Certain products with high sales volumes but low margins were identified, requiring adjustments to either pricing or cost management strategies to improve profitability.

Customer Analysis

Question: Who are our top customers based on sales volume?

- **Visualization:** *Bar Chart – Top Customers by Sales Volume*.
- **Insight:** High-value customers were identified, forming the basis for a targeted loyalty program aimed at further strengthening relationships with these key accounts.

Question: What is the customer retention rate? Are certain customers repeatedly purchasing?

- **Visualization:** *Retention Analysis – Repeat Purchases by Customer*.
- **Insight:** The data revealed a core group of loyal customers who made frequent repeat purchases, indicating strong brand loyalty. These customers could be targeted for future up-sell or cross-sell opportunities.

Sales Rep Performance

Question: Which sales reps are bringing in the most revenue?

- **Visualization:** *Bar Chart – Sales Rep Performance*.
- **Insight:** The analysis identified top-performing sales reps, suggesting a need to replicate successful sales strategies across the team.

Question: How does sales performance vary across regions?

- **Visualization:** *Heatmap – Regional Sales Performance*.
- **Insight:** Significant differences were found in sales rep performance across regions, highlighting opportunities to realign sales strategies for underperforming areas.

Product and Inventory Analysis

Question: What are the sales trends for each product over time?

- **Visualization:** *Line Chart – Product Sales Trends*.

- **Insight:** By tracking sales trends, high-demand products were identified, allowing for better inventory management and product placement.

Question: Which items are frequently discounted, and how do discounts affect their sales?

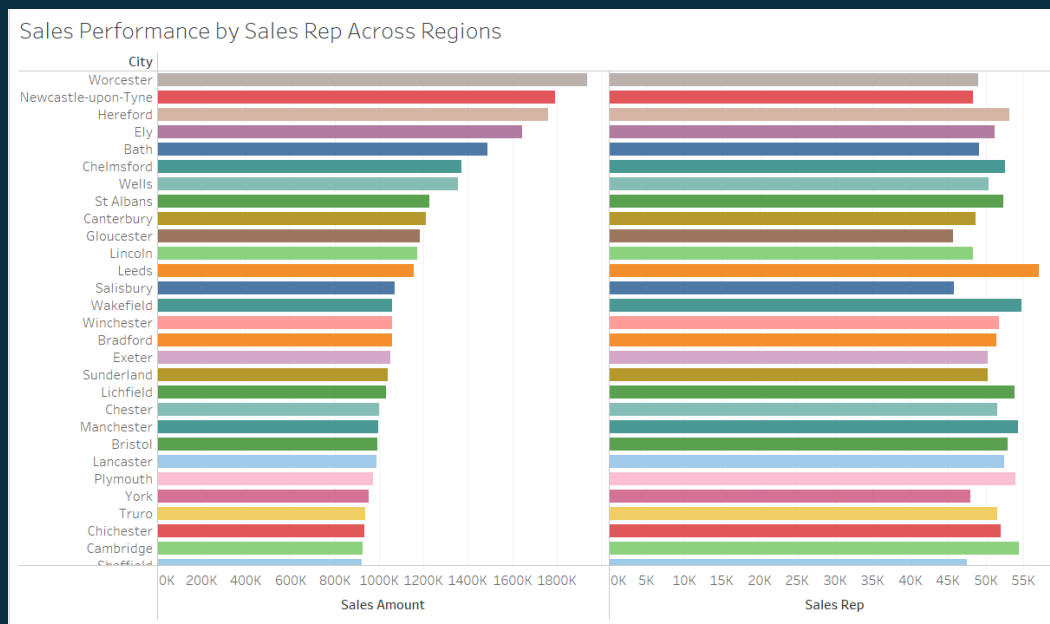
- **Visualization:** *Bar Chart* – Discounted Items and Sales Impact.
- **Insight:** Discounted products saw a notable increase in sales, though the impact on overall profitability varied.

5. Conclusion and Recommendations

Key Insights:

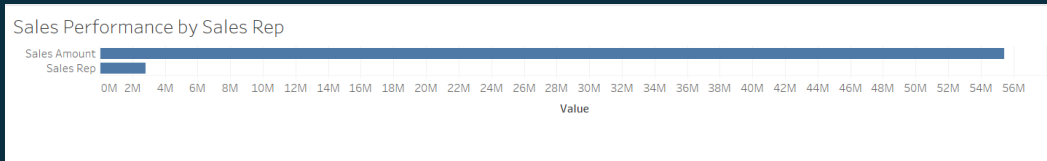
Sales Performance by Sales Rep Across Regions

This chart visualizes the sales performance across different regions and cities. It highlights which regions and sales representatives are contributing the most to the overall sales. Worcester, Newcastle, and Hereford show the highest sales performance, suggesting a concentration of business activity in these areas.



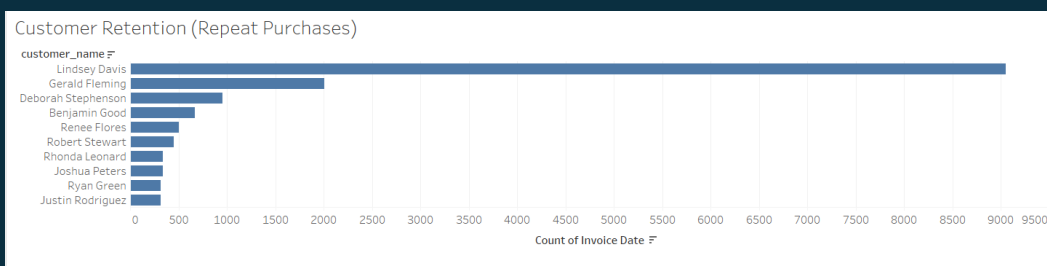
Sales Performance by Sales Rep

This chart shows the performance of individual sales reps. It helps identify top performers and areas where sales performance can be improved. The total sales amount achieved by the sales team exceeds 50M, with top performers contributing the most to the overall sales.



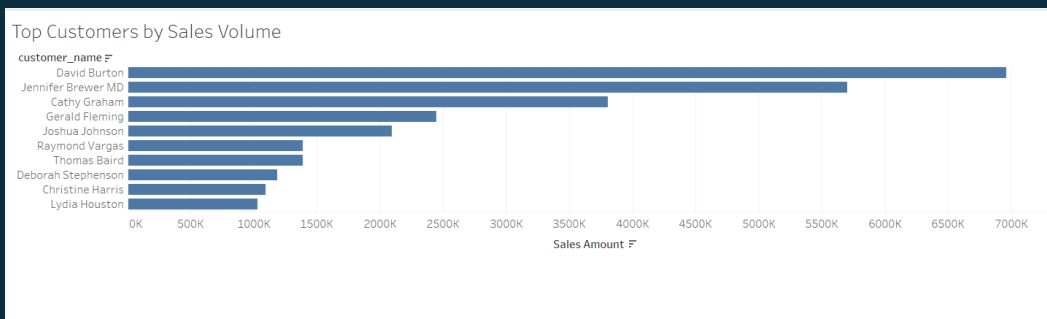
Customer Retention (Repeat Purchases)

This chart highlights the top customers based on repeat purchases. Lindsey Davis, Gerald Fleming, and Deborah Stephenson are among the most loyal customers. These insights can help with designing retention programs targeting high-repeat customers to maximize long-term value.



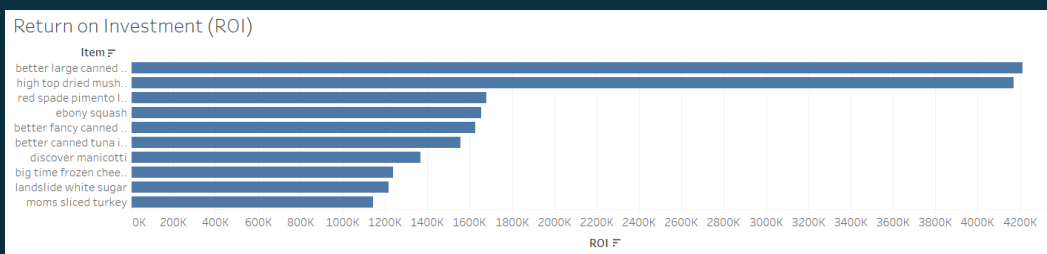
Top Customers by Sales Volume

This chart ranks customers by total sales volume. David Burton, Jennifer Brewer, and Cathy Graham contribute the most revenue, making them key accounts for maintaining strong business relationships.



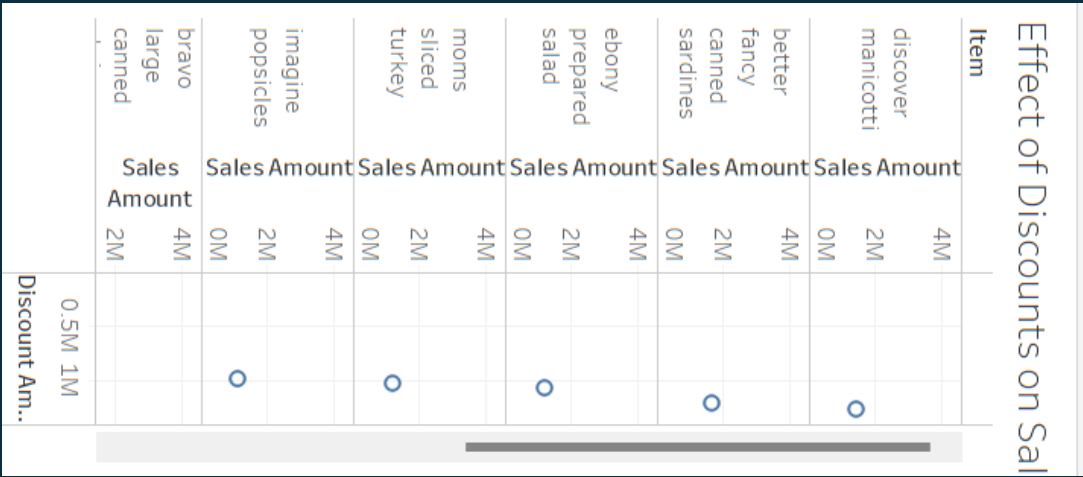
Return on Investment (ROI)

This chart presents the ROI for each product. Products such as "better large canned goods" and "high top dried mushrooms" have the highest ROI, indicating they provide the best return relative to their cost. This insight is crucial for determining where to allocate resources and focus promotional efforts.



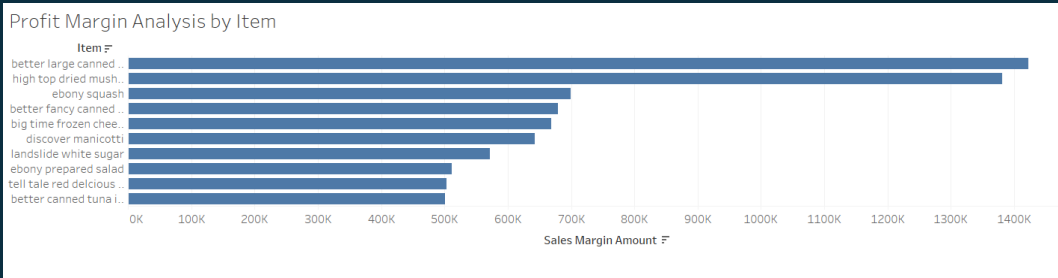
Effect of Discounts on Profitability

This scatter plot shows the relationship between discount amounts and sales. It reveals that some products experience a significant boost in sales when discounted, but profitability might be affected. This data can help businesses balance between boosting sales volume and protecting profit margins.



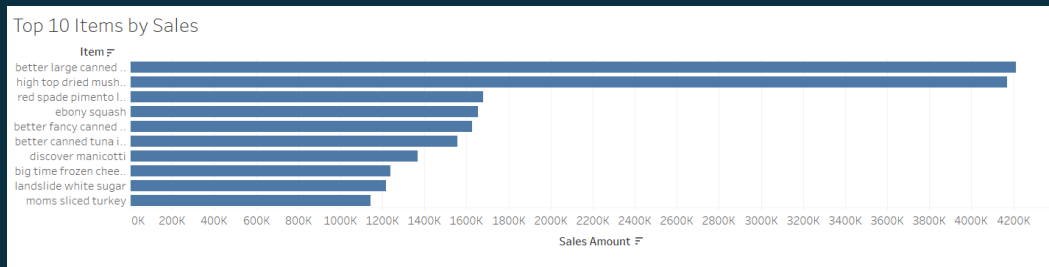
Profit Margin Analysis by Item

This chart displays the profit margin for each product. Products like "better large canned goods" and "high top dried mushrooms" yield the highest margins, while others, such as "better canned tuna," show lower margins. This data informs product pricing strategies and profit maximization efforts.



Top 10 Items by Sales

This chart lists the top 10 items by total sales. High-selling items like "better large canned goods" and "high top dried mushrooms" generate the most revenue. These insights help determine which products to prioritize in marketing and inventory management.



Sales Amount by Date (Time Series)

This time-series chart shows the trend of sales over time. The steady increase in sales from 2017 to 2019 highlights growth patterns that can inform future sales forecasting efforts.

