Financial Data Analysis Report

Overview

This report analyzes the financial dataset provided in the Jupyter notebook (finance.ipynb). The dataset contains sales and profit information across different segments, countries, and products.

Key Findings

Data Cleaning

- Currency symbols ($) and commas were removed from monetary columns
- Missing values in the Profit column were handled by converting to NaN
- Date column was converted to datetime format and sorted

Outlier Analysis

- Calculated IQR (Interquartile Range) for Profit column: 21,286.15
- Identified 93 outlier records (13.3% of total data)
- Outliers primarily consist of high-profit transactions for products like Paseo and Carretera

Profit Distribution

- Q1 (25th percentile): 3,875.85
- Q3 (75th percentile): 25,162.00
- Median profit: Approximately 12,000 (estimated)

Product Performance

- High-profit outliers concentrated in certain products:
    - Carretera
    - Paseo
    - Montana
    - Velo

Time Trends

- Data spans from 2013 to 2014
- Highest profits often occur in November and December (holiday season)

Recommendations

1. **Investigate High-Profit Products**: Focus on products like Paseo and Carretera that generate exceptional profits to understand drivers of success.
2. **Seasonal Strategy**: Capitalize on year-end profit spikes by increasing inventory and marketing in Q4.

3. **Outlier Analysis**: Examine whether high-profit outliers represent sustainable business or one-time events.
4. **Data Quality Improvements**:
    - Standardize currency formatting at data entry
    - Implement validation rules for profit calculations
    - Document units for all monetary values
5. **Further Analysis**:
    - Break down profits by segment and country
    - Analyze relationship between discounts and profits
    - Examine manufacturing costs vs. sale prices

Conclusion

The dataset reveals significant variations in profitability across products and time periods. The presence of numerous high-profit outliers suggests opportunities to optimize pricing and product mix. Further segmentation analysis would provide additional insights for strategic decision-making.

Appendix

Technical Details

- Python libraries used: pandas, numpy, matplotlib, seaborn
- Data cleaning steps:
    - Removed special characters from monetary columns
    - Handled missing values
    - Converted data types appropriately
- Visualization: Box plots used to identify outliers