H&M Retail Performance and Profitability Analysis

Introduction

In the context of a highly dynamic and competitive fashion retail industry, data-driven decision-making is essential for optimizing profitability and customer engagement. This project focuses on analyzing a dataset from H&M comprising 10,000 retail transaction records. The primary objective was to extract insights related to product profitability, seasonal trends, customer segmentation, and pricing anomalies. A comprehensive Tableau dashboard was developed to visualize and filter these insights for strategic decision-making.

Abstract

The project utilized transaction-level data from H&M, including variables such as product type, selling price, profit, customer age, membership status, and season. Python was used to clean the dataset, conduct exploratory data analysis (EDA), and detect outliers, while Tableau was employed to build an interactive dashboard. The analysis revealed that bottomwear and outerwear were among the most profitable product groups. Spring emerged as the best season in terms of volume, while autumn generated the highest estimated profit due to strong outerwear performance. The dashboard includes filters for product group, season, and article ID, enabling business users to examine performance drivers across multiple dimensions.

Tools Used

- Python (Pandas, NumPy, Matplotlib, Seaborn) – for data preprocessing, analysis, and visualization
- Google Colab – for executing Python code
- Tableau Public – for building and publishing the interactive dashboard
- Microsoft Excel – for data validation and cross-checks (where required)

Steps Involved in Building the Project

1. Data Cleaning

- The dataset was imported into Python and checked for missing values.
- The fashion_news_frequency column had 40% null values, which were excluded.
- The t_dat column was converted to datetime format to support seasonal analysis.
- Age data was binned into groups: <18, 18–30, 31–45, 46–60, and 60+.

2. Profitability Analysis

- Profit was analyzed using the estimated_profit field.
- Data was grouped by product_group_name and product_type_name to determine average profitability.

- A boxplot was generated to assess profit distribution across categories.
- Bottomwear was identified as the highest contributor to estimated profit, followed by outerwear.

3. Seasonal Analysis
- Seasonal trends were evaluated using the season variable.
- Spring had the highest number of items sold (4,000 units).
- Autumn generated the highest estimated profit, driven primarily by outerwear sales.

4. Customer Segmentation
- Customers were segmented based on age and club membership status.
- The largest segment consisted of active members in the 18–30 age group, indicating strong engagement among younger consumers.

5. Outlier Detection
- Price outliers in the price_x column were identified using the Interquartile Range (IQR) method.
- Approximately 2,000 records were flagged as price outliers, accounting for 20% of the dataset.
- These anomalies may represent special promotions, premium-priced items, or potential data entry issues.

6. Tableau Dashboard Development
The Tableau dashboard included the following components:
- Total Estimated Profit: ₹80,970
- Number of Items Sold: 10,000
- Best Season by Sales Volume: Spring, with 4,000 items sold
- Estimated Profit by Product Group: Highest for bottomwear
- Sales by Season and Product Group: Autumn yielded the most profit due to outerwear
- Filters: Product Group Name, Season, and Article ID

The dashboard allows users to interactively explore seasonal performance, category-wise trends, and item-level details.

Conclusion

This project effectively demonstrates the value of integrating Python-based analysis with Tableau dashboards to derive business insights from retail data. The findings indicate that bottomwear and outerwear are the most profitable product categories. Spring and autumn are critical seasons for H&M, each dominating in terms of volume and profit respectively. The identification of customer engagement patterns and pricing outliers further supports strategic decision-making. Overall, the dashboard and supporting analysis provide a robust foundation for inventory planning, targeted marketing, and pricing optimization.