

Walmart Data Analysis Report

1. Introduction

1.1 Purpose of the Analysis

The primary objective of this analysis is to gain actionable insights into Walmart's sales data by examining customer behavior, product performance, and sales trends.

Understanding these aspects will aid in making informed business decisions, optimizing marketing strategies, and enhancing inventory management.

1.2 Dataset Overview

The dataset encompasses transaction records from Walmart, detailing various attributes such as User ID, Age Group, Gender, Marital Status, Occupation, Product Category, and Purchase Amount. The data spans multiple years and captures individual purchase instances, providing a comprehensive view of customer interactions and sales dynamics.

2. Data Cleaning Process

To ensure the dataset's integrity and suitability for analysis, a meticulous data cleaning process was undertaken. The steps involved are as follows:

1. Spelling and Header Verification:

- Ensured all column headers were correctly spelled and clearly labeled to prevent confusion during analysis.

2. Data Type Correction and Standardization:

- Corrected data types across various fields. Specifically, the Stay in Current City Years column was standardized by removing the + sign from values representing 4+ years. This adjustment focused on the data within the 0-4 year range and facilitated easier aggregation and calculation.

3. Currency Conversion:

- Converted purchase amounts into a consistent currency format, ensuring all monetary values were standardized for accurate financial analysis.

4. Duplicate Data Considerations:

- Identifying duplicates was challenging due to the dataset's structure. The absence of purchase dates made it difficult to differentiate unique transactions from potential duplicates. Since users could appear multiple times for different purchases and products could be bought by various users, removing duplicates risked eliminating valid transactions. Consequently, no duplicate data was removed to preserve dataset integrity.

5. Handling Missing Values:

- Conducted a thorough check for missing values across all fields. Addressed any gaps to ensure a complete dataset for robust analysis.

6. Data Standardization:

- Converted the Marital Status field from numerical codes to descriptive labels (Single and Married) to enhance data readability and interpretability.

7. Data Sorting:

- Sorted the dataset by User ID to organize data efficiently, facilitating easier tracking of individual users and their transactions during analysis.

8. Transaction-Based Data Structure:

- Observed that the dataset tracks individual transactions rather than unique users. Multiple rows per User ID indicate that the data captures each purchase instance separately, essential for understanding purchasing patterns but not representing unique users.

This structured approach to data cleaning ensures the dataset is reliable, standardized, and primed for insightful analysis.

3. Data Analysis and Findings

3.1 Age and Gender Segmentation Analysis

3.1.1 Objective

To understand how different age groups and genders contribute to overall sales across various product categories.

3.1.2 Methodology

A Pivot Table was created with Product Category as rows and Age Group as columns, segmented further by Gender. The analysis focused on the sum of Purchase Amount to quantify each segment's contribution. A corresponding bar chart was developed to visually represent these findings.

3.1.3 Findings

- **Female Customers (F)** dominate sales with a total purchase amount of \$1,186,232,642.
 - **Age Group Insights:**
 - **26-35:** Highest contribution with \$442,976,233.
 - **36-45:** Significant contributions amounting to \$243,438,963.
 - Lower contributions observed in younger (0-17) and older (55+) age groups.
- **Male Customers (M)** exhibit a higher total purchase amount of \$3,909,580,100.
 - **Age Group Insights:**
 - **26-35:** Leading with \$1,588,794,345.
 - **18-25** and **36-45:** Substantial contributions of \$708,372,833 and \$783,130,921, respectively.
 - **55+:** Notable increase to \$154,984,610, indicating growing spending in this segment.

3.1.4 Recommendations

- **Target Marketing Campaigns:**
 - Focus on the **26-35** age group for both genders, as they represent the highest spenders.

- Develop specialized promotions for the **36-45** and **55+** segments to further boost sales.
 - **Product Category Optimization:**
 - Enhance product offerings in categories popular among high-contributing age groups.
 - Investigate and address the lower engagement in the **0-17** age group to identify potential growth opportunities.
-

3.2 Occupation and Marital Status Segmentation Analysis

3.2.1 Objective

To analyze purchasing behaviors based on customers' occupation levels and marital status.

3.2.2 Methodology

A Pivot Table was constructed with Occupation as rows and Marital Status (Single, Married) as columns, evaluating the sum of Purchase Amount. A pie chart was generated to depict the distribution of purchases across these segments.

3.2.3 Findings

- **Marital Status Influence:**
 - **Single** customers account for \$3,008,927,447 in purchases.
 - **Married** customers contribute \$2,086,885,295.
- **Occupation Insights:**
 - **Occupation Level 0:** Highest purchases among Singles (\$370,825,372) and Married (\$264,581,586).
 - **Occupation Levels 1-17:** Consistent purchasing across both marital statuses, with significant amounts in middle occupations (e.g., Levels 4, 7, 12, 14, 16, 17).
 - **Lower Occupation Levels:** Occupations 8-10 show relatively lower purchase amounts, indicating potential for growth.

3.2.4 Recommendations

- **Customized Marketing Strategies:**
 - Develop targeted campaigns for both Single and Married customers, acknowledging their substantial purchasing power.
 - Focus on high-occupation levels (0, 4, 7, 12, 14, 16, 17) to maximize sales impact.
 - **Occupation-Based Product Offerings:**
 - Tailor product assortments to match the preferences of customers in high-spending occupation levels.
 - Explore opportunities to increase engagement in lower occupation levels through specialized promotions or product introductions.
-

3.3 Product Performance Analysis

3.3.1 Objective

To evaluate the performance of different product categories based on purchase amounts across various segments.

3.3.2 Methodology

A Pivot Table was developed with Product Category as rows and segmented by age groups (0-17, 18-25, etc.). The sum of Purchase Amount was calculated to assess performance. A bar chart was utilized to visualize these results.

3.3.3 Findings

- **Top-Performing Product Categories:**
 - **Category 1:** Leads with \$1,910,013,754 in purchases, indicating strong customer preference.
 - **Category 5:** Substantial sales of \$941,835,229, showcasing consistent demand.
 - **Category 8:** Highest among higher-numbered categories with \$854,318,799.

- **Age Group Influence:**
 - **26-35:** Dominates in most product categories, particularly in Categories 1, 5, and 8.
 - **18-25** and **36-45:** Significant contributors across various categories.
 - **0-17** and **55+:** Lower contributions, suggesting niche markets or areas for potential growth.

3.3.4 Recommendations

- **Inventory Management:**
 - Prioritize stocking and inventory for top-performing categories to meet high demand.
 - Monitor Category 8 closely due to its substantial sales figures.
- **Product Development:**
 - Explore opportunities to expand offerings in high-performing categories.
 - Investigate reasons for lower sales in specific age groups and develop strategies to enhance engagement.
- **Promotional Activities:**
 - Implement targeted promotions for Categories 1, 5, and 8 to sustain and boost sales.
 - Consider introducing new products or variations in underperforming categories to attract a broader customer base.

3.4 Age Groups Product Trend Analysis

3.4.1 Objective

To identify and analyze purchasing trends of different product categories across various age groups over time.

3.4.2 Methodology

A Pivot Table was created with Age Group as columns and Product Category as rows, calculating the sum of Purchase Amount. A line chart was generated to illustrate the trends and patterns observed.

3.4.3 Findings

- **Category 1:**
 - **26-35** age group leads with \$783,813,459, followed by **18-25** with \$362,607,972.
 - Steady growth observed in younger age groups, indicating strong brand loyalty.
- **Category 5:**
 - Dominates the **26-35** age group with \$379,702,493 and **18-25** with \$175,198,782.
 - Significant contributions from mid-age groups suggest targeted marketing success.
- **Category 8:**
 - Highest purchases in the **26-35** age group (\$327,523,339) and **18-25** (\$132,320,061).
 - Consistent upward trend across all age groups, highlighting widespread appeal.
- **Lower Product Categories (e.g., 9, 10, 19):**
 - Minimal contributions across most age groups, indicating niche markets or potential areas for strategic improvement.

3.4.4 Recommendations

- **Trend Monitoring:**
 - Continuously monitor purchasing trends to adapt quickly to changing customer preferences.
 - Utilize predictive analytics to anticipate future trends and adjust inventory accordingly.
- **Age-Specific Marketing:**
 - Develop age-specific marketing strategies to enhance engagement in high-potential categories.

- Leverage the popularity of Categories 1, 5, and 8 among key age groups to introduce complementary products.
 - **Expansion Opportunities:**
 - Explore opportunities to expand product offerings in high-trending categories.
 - Investigate and address the low performance in certain categories to unlock new revenue streams.
-

4. Conclusions and Recommendations

4.1 Key Findings

- **High-Spending Demographics:** The **26-35** age group across both genders and multiple occupation levels represents the highest purchasing power, particularly in top-performing product categories.
- **Gender Influence:** Male customers contribute more significantly to total sales compared to female customers, especially in key age groups.
- **Marital Status Impact:** Both Single and Married customers show substantial purchasing behavior, with Single customers leading in total purchase amounts.
- **Product Category Performance:** Categories 1, 5, and 8 are the top performers, indicating strong customer preference and consistent demand.
- **Trend Insights:** Purchasing trends reveal a steady growth in key product categories among younger age groups, suggesting brand loyalty and potential for future growth.

4.2 Strategic Recommendations

1. Targeted Marketing Campaigns:

- Focus on high-spending demographics, particularly the **26-35** age group, to maximize sales impact.
- Develop gender-specific promotions to leverage the higher purchasing power of male customers.

2. Product Optimization:

- Prioritize inventory and marketing efforts for top-performing product categories (1, 5, 8) to sustain high sales volumes.

- Investigate and enhance underperforming categories to diversify product offerings and capture untapped markets.

3. **Customer Segmentation and Personalization:**

- Utilize detailed customer segmentation (age, gender, occupation, marital status) to personalize marketing messages and product recommendations.
- Implement loyalty programs tailored to high-value segments to foster long-term customer relationships.

4. **Trend Analysis and Predictive Modeling:**

- Continuously analyze purchasing trends to stay ahead of market shifts and consumer preferences.
- Employ predictive analytics to forecast future sales trends and optimize inventory management accordingly.

5. **Enhanced Data Collection:**

- Incorporate additional data points such as purchase dates to improve data accuracy and facilitate more precise duplicate detection.
- Expand data collection to include customer feedback and preferences for a more holistic analysis.

4.3 Future Considerations

- **Expansion of Data Sources:** Integrate data from other channels (e.g., online sales, in-store interactions) to gain a more comprehensive view of customer behavior.
 - **Advanced Analytics Techniques:** Explore machine learning and advanced statistical methods to uncover deeper insights and predictive patterns.
 - **Cross-Functional Collaboration:** Work closely with marketing, sales, and inventory teams to implement data-driven strategies effectively.
-

5. Appendix

5.1 Pivot Tables and Charts

- **Age and Gender Segmentation Analysis:** [Bar Chart]
- **Occupation and Marital Status Segmentation Analysis:** [Pie Chart]
- **Product Performance Analysis:** [Bar Chart]
- **Age Groups Product Trend Analysis:** [Line Chart]

Note: Charts are embedded within the Excel workbook for detailed visualization.
