

Retail – Data Set

About Dataset

Welcome to the Retail Sales and Customer Demographics Dataset! This synthetic dataset has been meticulously crafted to simulate a dynamic retail environment, providing an ideal playground for those eager to sharpen their data analysis skills through exploratory data analysis (EDA). With a focus on retail sales and customer characteristics, this dataset invites you to unravel intricate patterns, draw insights, and gain a deeper understanding of customer behaviour.

****Dataset Overview:**

This dataset is a snapshot of a fictional retail landscape, capturing essential attributes that drive retail operations and customer interactions. It includes key details such as Transaction ID, Date, Customer ID, Gender, Age, Product Category, Quantity, Price per Unit, and Total Amount. These attributes enable a multifaceted exploration of sales trends, demographic influences, and purchasing behaviours.

Why Explore This Dataset?

- **Realistic Representation:** Though synthetic, the dataset mirrors real-world retail scenarios, allowing you to practice analysis within a familiar context.
- **Diverse Insights:** From demographic insights to product preferences, the dataset offers a broad spectrum of factors to investigate.
- **Hypothesis Generation:** As you perform EDA, you'll have the chance to formulate hypotheses that can guide further analysis and experimentation.
- **Applied Learning:** Uncover actionable insights that retailers could use to enhance their strategies and customer experiences.

Questions to Explore:

- How does customer age and gender influence their purchasing behaviour?
- Are there discernible patterns in sales across different time periods?
- Which product categories hold the highest appeal among customers?
- What are the relationships between age, spending, and product preferences?
- How do customers adapt their shopping habits during seasonal trends?
- Are there distinct purchasing behaviours based on the number of items bought per transaction?
- What insights can be gleaned from the distribution of product prices within each category?

◇ Sales & Revenue Analysis

1. Which time periods (daily, weekly, monthly, seasonal) generate the highest revenue?
2. Are there certain days of the week or times of the month when sales spike (e.g., payday effect)?

3. Which product categories contribute the most to total revenue, and which contribute the least?
 4. What is the distribution of high-ticket vs. low-ticket sales?
 5. What is the average revenue per customer, and how does it vary across age groups?
-

◇ Customer Segmentation

6. Can we segment customers into groups (e.g., young-budget, mid-age-premium, seniors) based on their spending habits?
 7. Which age group has the highest repeat purchase frequency?
 8. Do male and female customers differ in terms of basket size (average number of items purchased)?
 9. Who are the top 5% of customers contributing to revenue (Pareto Principle / 80-20 rule)?
 10. Which demographics are more likely to purchase from premium categories?
-

◇ Product Insights

11. Are certain product categories more popular among younger vs. older customers?
 12. Do discounts or lower prices significantly increase sales volume in specific categories?
-

◇ Customer Behavior & Loyalty

13. How many unique customers purchase more than once, and what are their characteristics?
14. Do younger customers tend to try more categories compared to older customers (variety-seeking behavior)?