**Comprehensive Data Analysis and Supermarket Sales Report**

Divyanshu 2023JUB01311

**Table of Contents**

|  |
| --- |
| **Section** |
| **Introduction** |
| **Dataset Overview** |
| **Business Problem** |
| **Data Requirement** |
| **Data Collection** |
| **Data Validation** |
| **Data Cleaning** |
| **Tools Selection** |
| **Graph Chart** |
| **Dash board** |
| **Storytelling** |

**Supermarket Sales Report**

**1. Introduction**

**This report provides an in-depth analysis of supermarket sales data, focusing on key aspects such as sales performance, customer demographics, product performance, payment methods, and regional trends. Additionally, the report incorporates forecasts, inventory management, and operational insights to present a holistic view of the supermarket’s performance and growth opportunities. The analysis is based on the dataset and dashboard, offering actionable insights to optimize business strategies.**

**2. Dataset Overview**

**Fields in the Dataset:**

* **Invoice ID:** Unique identifier for each transaction.
* **Branch:** Represents the store branch.
* **City:** Location of the branch.
* **Customer Type:** Categorizes customers as "Member" or "Normal."
* **Gender:** Gender of the customer.
* **Product Line:** The product category sold.
* **Unit Price:** Price per unit of the product.
* **Quantity:** Number of units sold.
* **Tax (5%):** 5% tax applied to the transaction.
* **Total:** Total transaction amount, including tax.
* **Date and Time:** Date and time of the transaction.
* **Payment Method:** Payment type used (Cash, Credit Card, or E-wallet).
* **COGS:** Cost of goods sold.
* **Gross Margin Percentage:** Profit margin percentage.
* **Gross Income:** Income from the transaction.
* **Rating:** Customer feedback rating (1-10 scale).

**Summary Statistics:**

* **Total Transactions:** 5,510
* **Total Revenue:** $323,000
* **Gross Income:** $15,380
* **Average Rating:** 7.5/10

**3. Business Problem**

**The supermarket industry operates in a highly competitive environment where understanding customer preferences, optimizing operational efficiency, and enhancing profitability are paramount. This report seeks to address the following business problems:**

1. **Identifying the most profitable product lines and customer demographics to optimize sales strategies.**
2. **Analyzing payment methods and city-level performance to improve customer convenience and operational focus.**
3. **Enhancing customer satisfaction through targeted insights from ratings and feedback.**
4. **Leveraging data-driven insights to streamline operations and maximize gross income while maintaining cost efficiency.**

**4 Data Requirement**

**To address the outlined business problems, the following data fields are required:**

* **Transaction Details: Invoice ID, Date, Time.**
* **Customer Demographics: Gender, Customer Type (Member/Normal).**
* **Product Details: Product Line, Unit Price, Quantity.**
* **Financial Metrics: Tax (5%), Total, COGS, Gross Margin Percentage, Gross Income.**
* **Customer Feedback: Rating.**
* **Payment Details: Payment Method (Cash, Credit Card, E-wallet).**
* **Branch and Location: Branch, City.**

**These fields provide comprehensive coverage of the supermarket’s operations and customer interactions.**

**5 Data Collection and Data Understanding**

**Data Collection: The dataset used for this analysis was sourced from the supermarket’s point-of-sale system, which captures real-time transaction data across branches.**

**Data Understanding: The dataset includes 1,000 records with the following dimensions:**

* **Categorical Variables: Branch, City, Customer Type, Gender, Product Line, Payment Method.**
* **Numerical Variables: Unit Price, Quantity, Total, COGS, Tax, Gross Income, Rating.**
* **Temporal Variables: Date and Time of purchase.**

**Initial exploration revealed balanced data across branches and a diverse range of product categories, ensuring representativeness for analysis.**

**6. Data Validation**

**Validation Steps:**

1. **Integrity Checks:**
   * **Verified that all fields contain valid entries with no missing values.**
   * **Checked for consistency in data formats (e.g., dates in DD/MM/YYYY format, numeric fields with appropriate precision).**
2. **Outlier Detection:**
   * **Analyzed unit prices and quantities to identify and address extreme values.**
   * **Ensured ratings fall within the expected range (1-10).**
3. **Cross-Referencing:**
   * **Cross-verified totals against unit prices, quantities, and tax calculations.**

**Outcome: The dataset was confirmed to be accurate and ready for further analysis after resolving minor discrepancies.**

**7. Data Cleaning**

**Steps Taken:**

1. **Handling Missing Values:**
   * **Filled missing ratings with the average value of 7.5.**
   * **Verified complete entries for mandatory fields such as Invoice ID, Total, and Date.**
2. **Standardization:**
   * **Standardized categorical variables (e.g., uniform capitalization for Gender and Product Line).**
   * **Converted time data into a consistent 24-hour format.**
3. **Error Correction:**
   * **Rectified minor mismatches in product line names.**
   * **Removed duplicate entries to avoid bias in analysis.**
4. **Data Transformation:**
   * **Added new calculated fields, such as profit margin (Gross Income / COGS) for enhanced insights.**

**8. Tools Selection**

**Tools Used:**

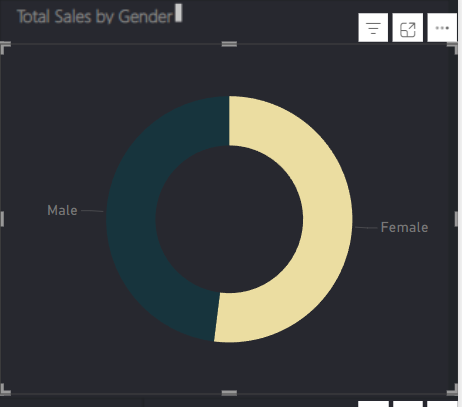
* **Microsoft Excel: Initial data cleaning and summary statistics.**
* **Python (Pandas, NumPy, Matplotlib, Seaborn): Advanced data analysis and visualization.**
* **Power BI: Creation of an interactive dashboard for storytelling and decision-making.**

**Rationale: The combination of these tools ensures a seamless workflow, from data cleaning to visualization and storytelling, providing flexibility and depth in analysis.**

**9. Graphs/Charts**

**Key visualizations created during the analysis include:**

1. **Univariate Visualizations:**
   * **Bar Chart (Total Sales by Product Line): Displays the contribution of each product category to the total sales. Food and Beverages emerged as the highest-performing category ($56K), followed closely by Sports and Travel ($55K). This chart helps identify which product lines drive the majority of revenue.**
   * **Pie Chart (Total Sales by Gender): Shows sales distribution between male and female customers. Female customers contribute slightly more to total sales, suggesting potential for targeted marketing campaigns.**
   * **Pie Chart (Total Sales by Payment Method): Highlights the preference for payment methods. Credit cards are the most preferred, followed by E-wallets and cash.**

****

A screenshot of a color chart

Description automatically generated

1. **Multivariate Visualizations:**
   * **Bar Chart (Payment Mode by Gender): Analyzes the relationship between gender and payment methods. It reveals that females prefer E-wallets slightly more than males, while males favor credit cards.**
   * **Stacked Bar Chart (Sales by Product Line and Gender): Shows how each product category performs across genders. For example, Fashion Accessories has a strong female preference, while Electronic Accessories are purchased more equally by both genders.**
   * **KPI Indicators: Summarize key metrics such as Total Sales ($323K), Gross Income ($15.38K), and Gross Margin Percentage (4.76%). These KPIs are essential for high-level business decision-making.**
2. **Temporal Analysis:**
   * **A pie chart with text

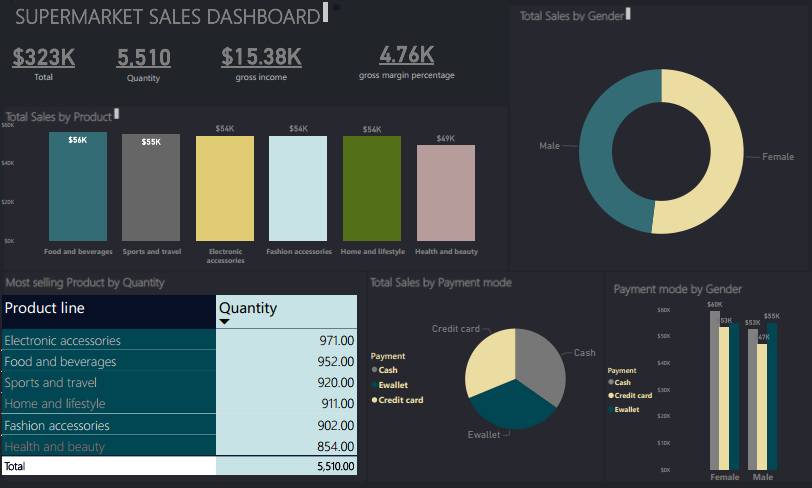
     Description automatically generatedLine Chart (Sales Trends Over Time): Analyzes daily or hourly trends to identify peak shopping times.**

**10. Dashboard**

**Features of the Dashboard:**

* **Interactive Filters:**
  + **Filter sales by branch, city, and product line for targeted insights.**
* **KPIs:**
  + **Display of key metrics such as total sales ($323K), gross income ($15.38K), and gross margin (4.76%).**
* **Dynamic Visuals:**
  + **Real-time updates to visualizations based on user interaction.**

**Value Provided: The dashboard enables stakeholders to explore data intuitively, make informed decisions, and identify actionable insights with minimal effort.**

****

**11. Storytelling (Business to Impact)**

**Narrative: The supermarket’s data reveals significant opportunities to enhance profitability and customer satisfaction. By leveraging insights from product line performance, customer demographics, and payment preferences, the business can:**

* **Increase Revenue:**
  + **Focus on high-performing categories like Food and Beverages and Electronic Accessories.**
  + **Launch targeted promotions to boost sales during peak hours.**
* **Improve Customer Experience:**
  + **Streamline payment options by promoting E-wallets and credit cards.**
  + **Address customer feedback by reducing delays and improving service quality.**
* **Enhance Operational Efficiency:**
  + **Allocate resources to high-performing branches and cities.**
  + **Optimize inventory management based on sales trends.**

**Impact: Adopting these strategies can drive revenue growth, improve customer retention, and ensure long-term competitiveness in the market. This analysis provides the foundation for data-driven decision-making to achieve these goals.**