EDA Report: Synthetic Sales Data

Key Findings:

- The dataset contains 500 entries with 8 features including demographics, income, and purchase behavior.

- Gender is almost equally distributed, and customers are mainly from **South** and **West** regions.

- Most customers are aged between **25 to 45 years**.

- Average **Annual Income** is approximately ₹50,000 with moderate dispersion.

- Purchase amounts show **slight monthly growth**, suggesting seasonal or promotional trends.

Outliers and Patterns Detected:

- A few customers have exceptionally high **Annual Income** values (above ₹100,000), as shown by the boxplot.

- **Spending Score** shows no strong correlation with Age or Income, implying varied customer behavior.

- Some clusters are visible in the **Income vs Spending Score** scatter plot, hinting at potential for customer segmentation.

Suggestions for Further Analysis:

- Segment customers based on Income and Spending Score using **KMeans** or **DBSCAN** for targeted marketing.

- Study **temporal purchase patterns** by grouping data quarterly or by specific campaigns.

- Conduct **region-wise performance analysis** to align promotions and resources better.

- Replace mean imputation with **predictive imputation** for missing values in future for more accuracy.

- Integrate more features like **product categories**, **loyalty status**, or **feedback scores** for deeper insights.