**Sales and Customer Analysis Report**

**Data Overview**

The dataset contains detailed information on transactions, customer demographics, and product performance:

* **Total Records**: 10,187
* **Columns**: 16 (e.g., Transaction ID, Product ID, Category, Revenue, Gender, Age, Region)

**Data Cleaning Process**

1. **Missing Values**: Numerical Missing values were replaced with the median and categorical values were replaced by the most frequent value which is the mode
2. **Duplicates**: There 13 duplicate rows that were dropped to make the dataset more reliable.
3. **Outliers**: Outliers were identified using IQR and then replaced by the mean calculated with the outliers excluded
4. **Timestamp Conversion:** Converted Timestamp to datetime format for time-based analysis.
5. **Scaling:** we made scaling for numeric data to ensure all features contribute equally to model performance.
6. **Label encoding:** To convert categorical data into numeric to be used by machine learning models

**Data Model**

The dataset is a star schema as it is simple and easy

**Analysis**

**1. Sales Performance by Category**

Sales Analysis showed that groceries generated the most total sales



**2. Revenue by Region**

Red regions have the highest revenue and yellow show the least

A map of the world with red dots

Description automatically generated

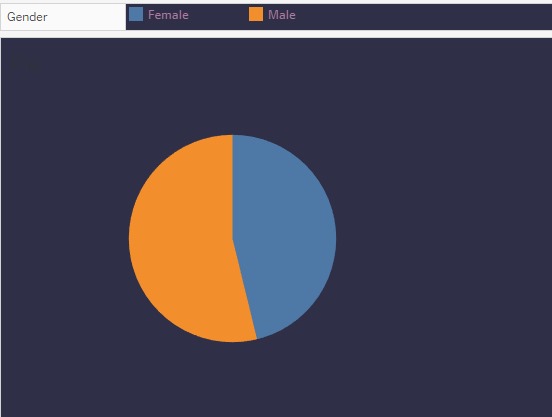
**3. Customer Age Distribution**

* **Average Age**: 43 years
* **Age Range**: 16 to 67 years
* Most customers fall between **40 and 44 years**, representing a middle-aged customer demographic. A graph with a bar chart

  Description automatically generated with medium confidence

**4. Spending Trends by Gender**

* **Female Customers**: total spend of $52,475
* **Male Customers**: total spend of $61,126

Showing that Male customers make the most of the revenue

1. **Revenue Supplier Distribution**

**Highest revenue supplier:** West PLC

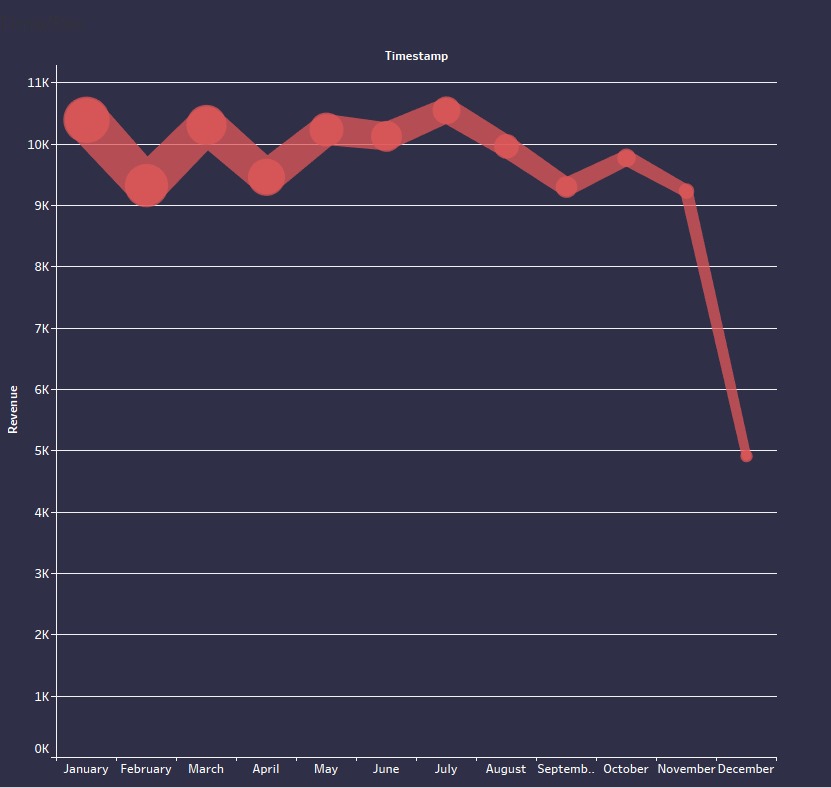
**Lowest revenue supplier:** Bryan LTD

A graph with different colored bars

Description automatically generated with medium confidence

1. **Revenue over Time**

The line graph shows that the revenue stays stable most of the year but meets a sharp decline in december

****

**Key Takeaways**

1. **Top-performing Categories**: Electronics and Groceries dominate revenue generation, indicating strong customer demand in these areas.
2. **Regional Performance**: North and East regions perform significantly better than South and West.
3. **Customer Demographics**: Focus marketing efforts on middle-aged customers (31–55 years).
4. **Gender Insights**: Males have a slightly higher spending average, which could guide targeted promotions.

**Recommendations**

* **Enhance Inventory Planning**: Reduce overstock or understock issues by analysing demand trends.
* **Targeted Promotions**: Tailor campaigns for Electronics and Groceries in North and East regions.
* **Customer Engagement**: Design initiatives to attract younger customers under 31 years old.