

Exploratory Data Analysis Report

File: sample_retail_dataset.csv

Generated on: 2025-09-04 12:16:37

Summary:

- Rows: 200
- Columns: 9
- Numeric columns: 4
- Categorical columns: 5

DataFrame .info() (excerpt):

```
<class 'pandas.core.frame.DataFrame'>
```

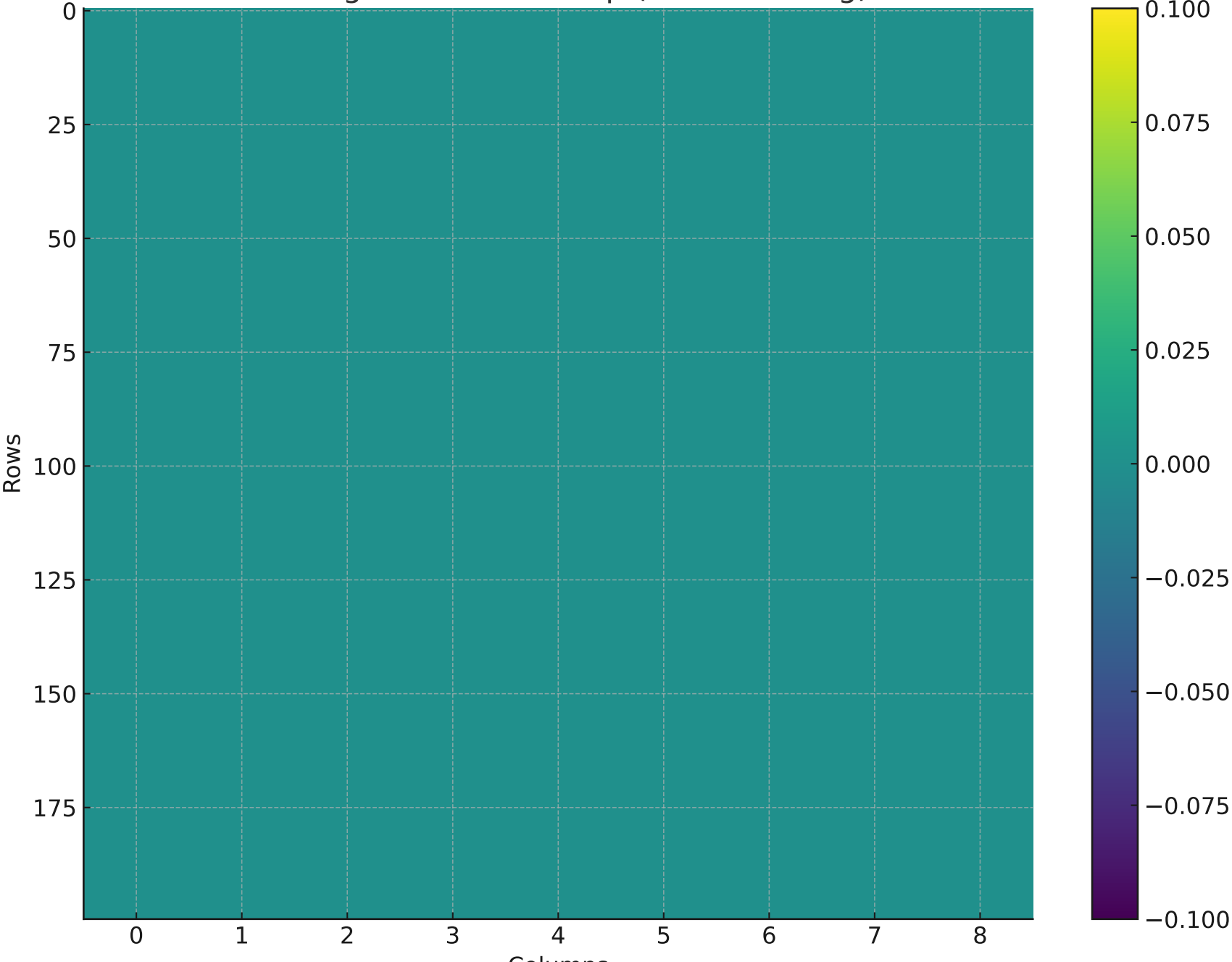
RangeIndex: 200 entries, 0 to 199

Data columns (total 9 columns):

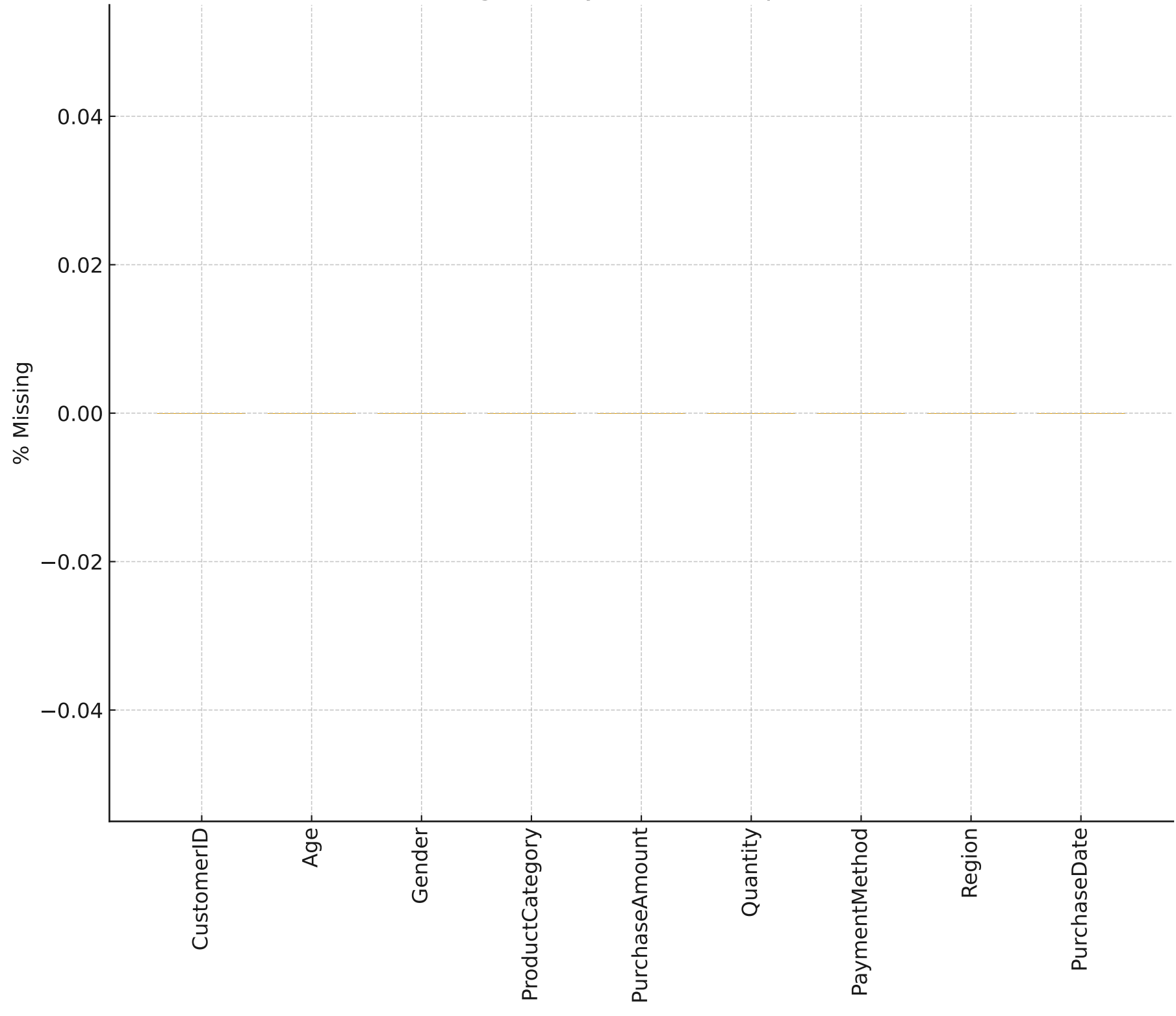
#	Column	Non-Null Count	Dtype
0	CustomerID	200 non-null	int64
1	Age	200 non-null	int64
2	Gender	200 non-null	object
3	ProductCategory	200 non-null	object
4	PurchaseAmount	200 non-null	float64
5	Quantity	200 non-null	int64
6	PaymentMethod	200 non-null	object
7	Region	200 non-null	object
8	PurchaseDate	200 non-null	object

dtypes: float64(1), int64(3), object(5)
memory usage: 14.2+ KB

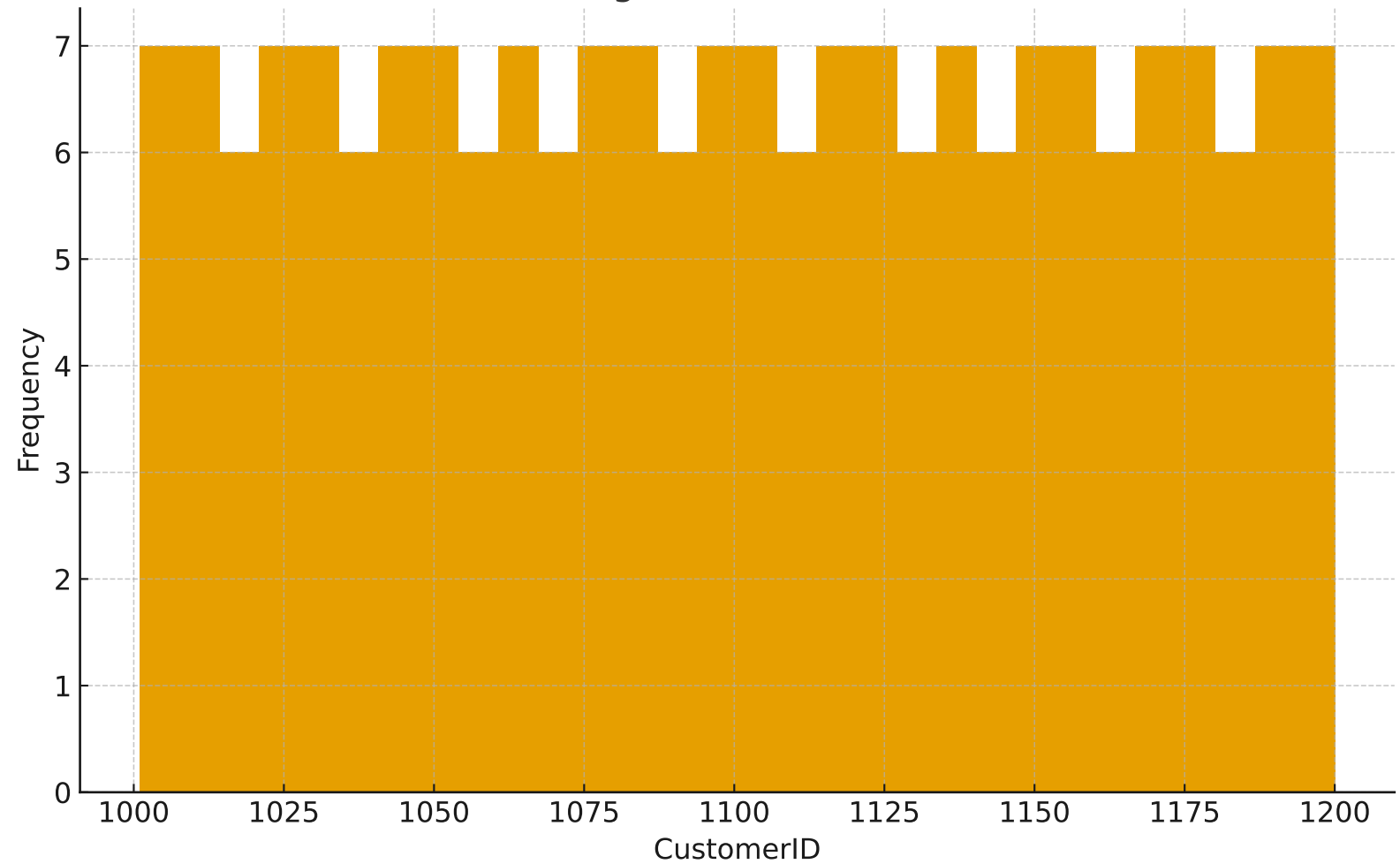
Missing Values Heatmap (True=missing)



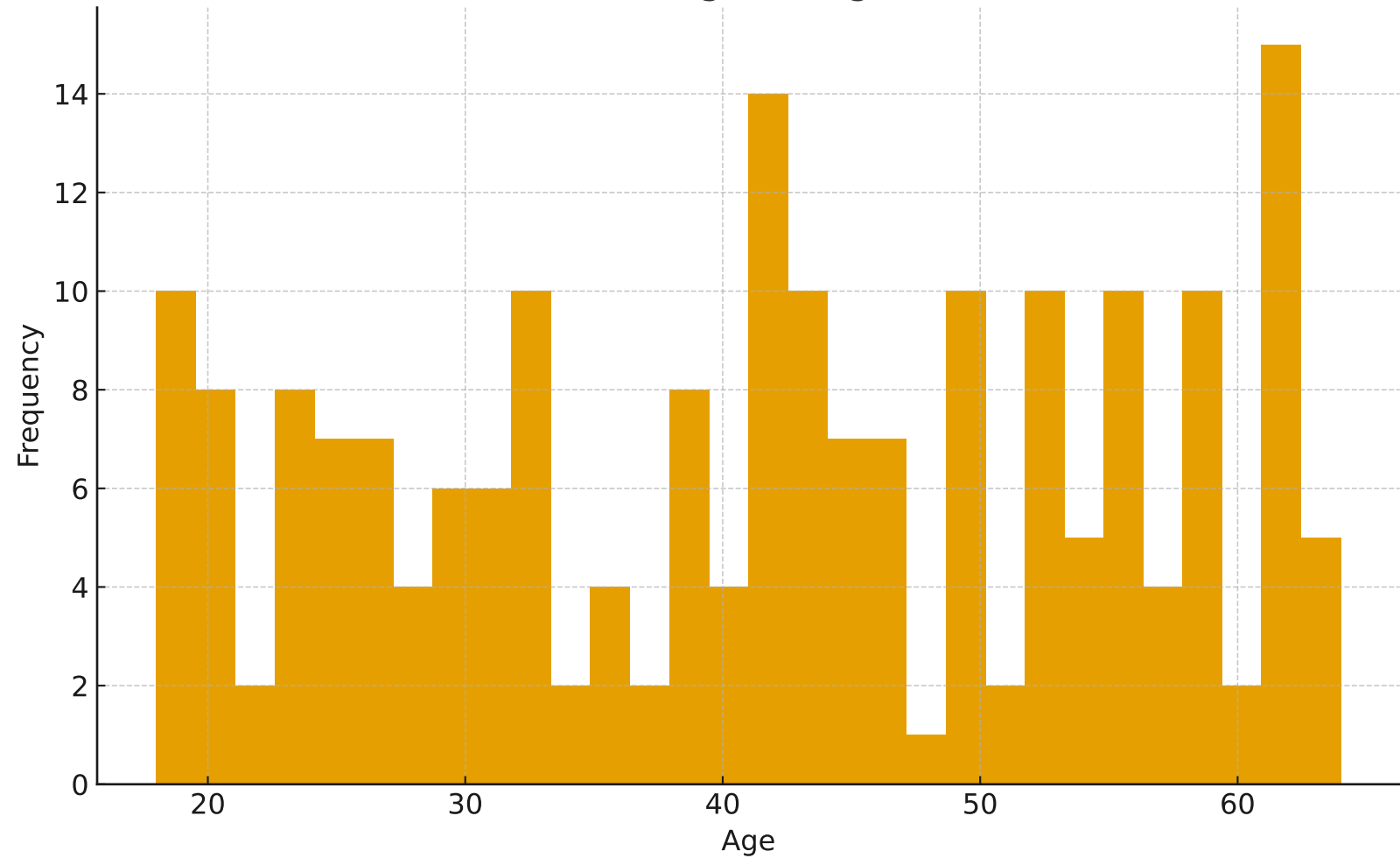
Missingness by Column (Top 20) - %



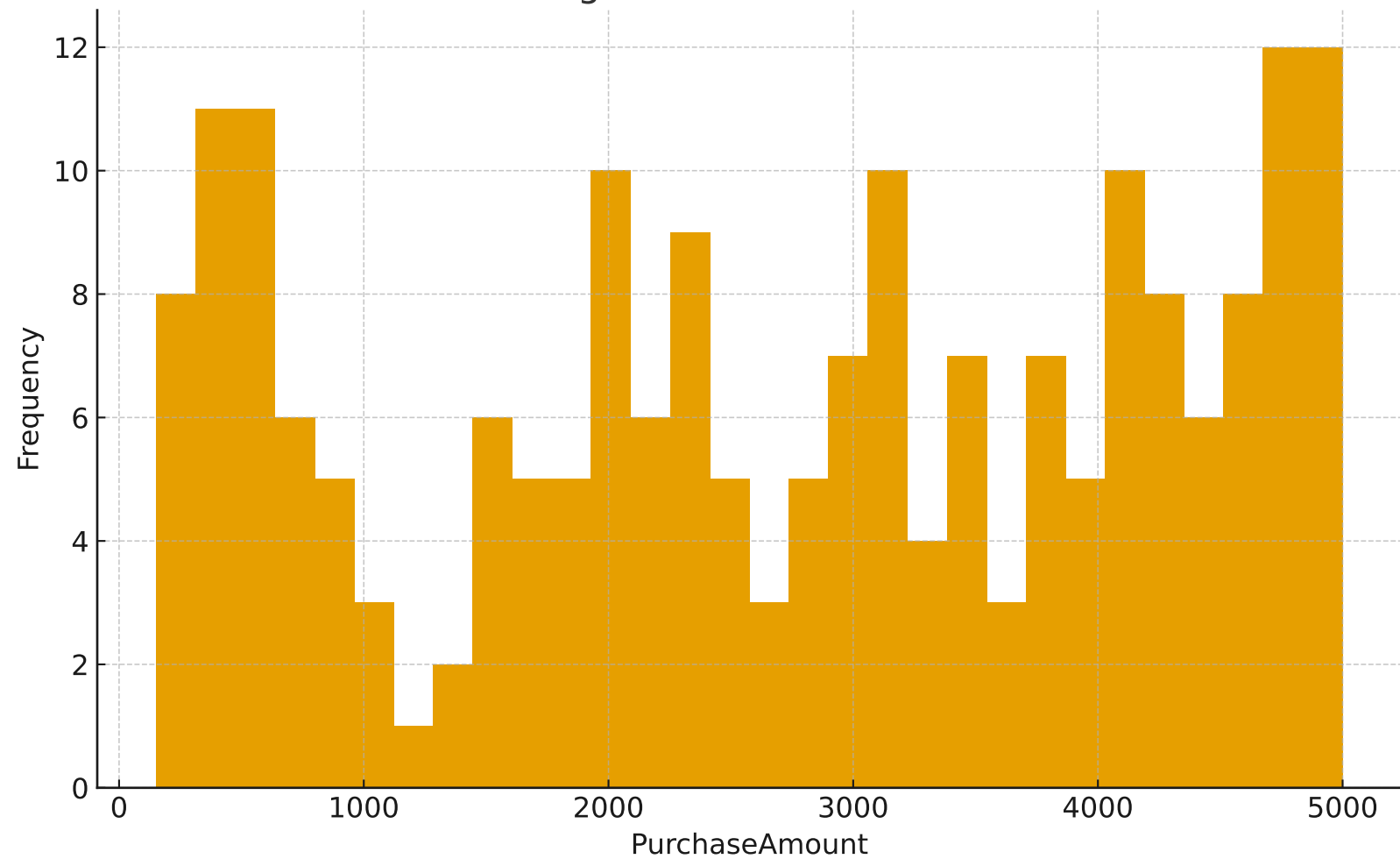
Histogram: CustomerID



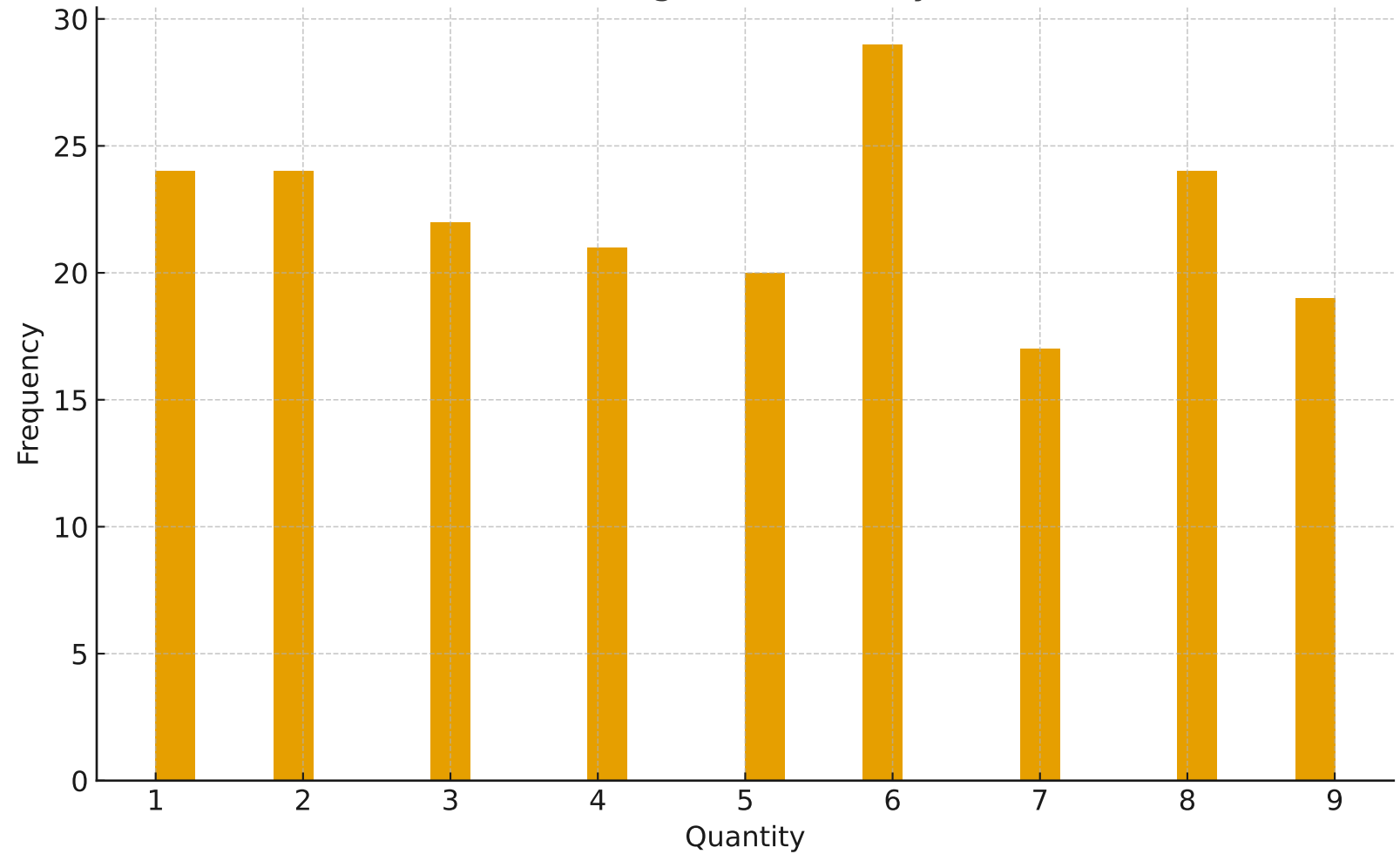
Histogram: Age



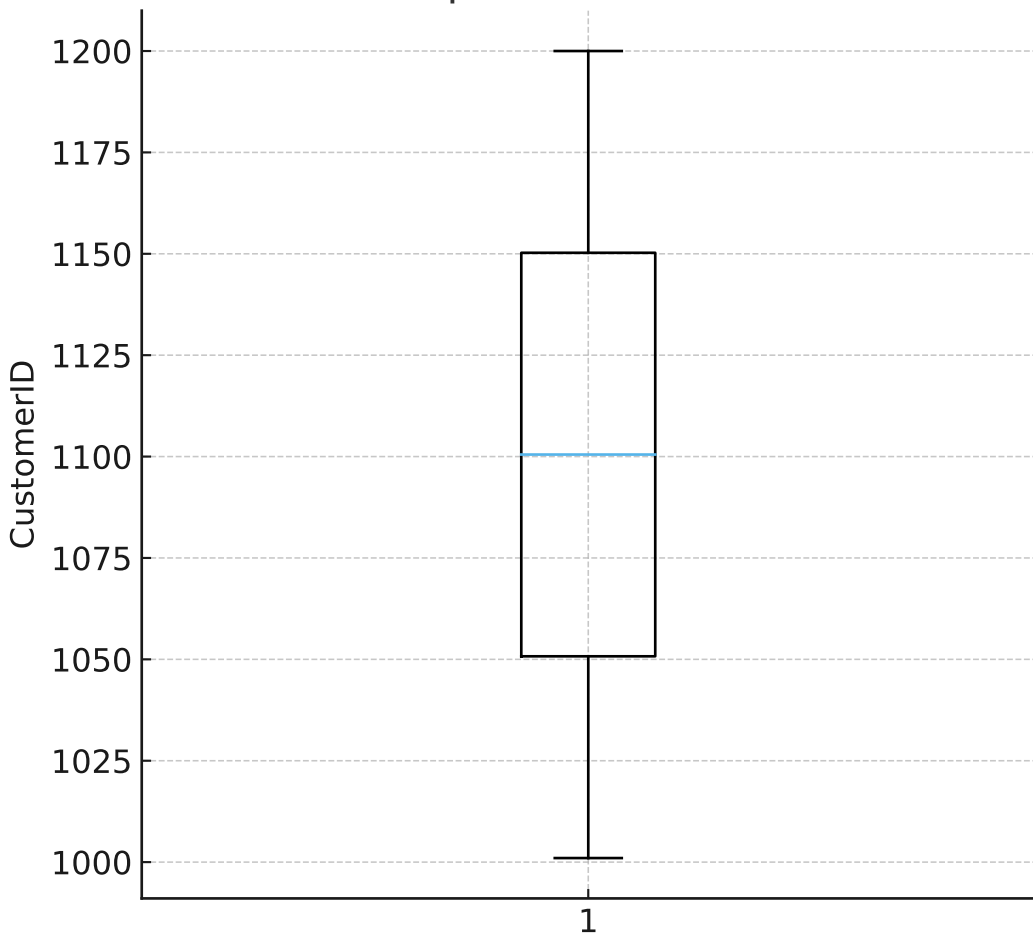
Histogram: PurchaseAmount



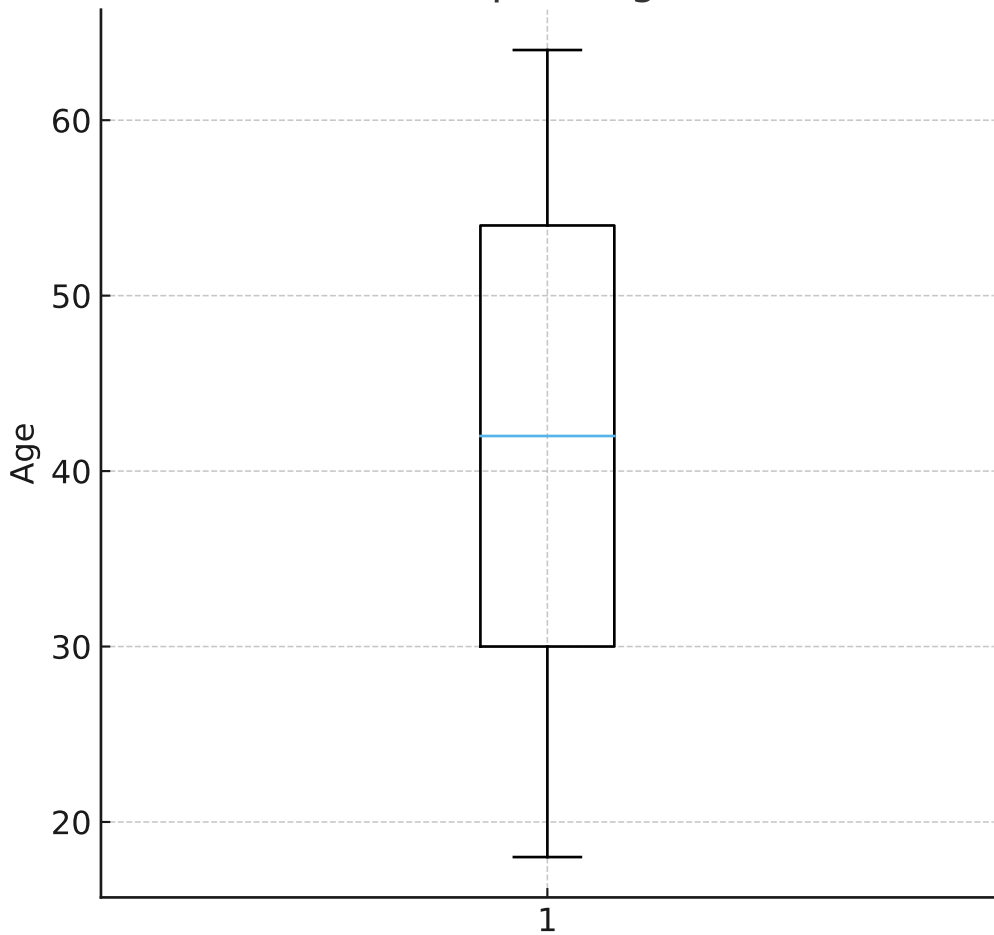
Histogram: Quantity



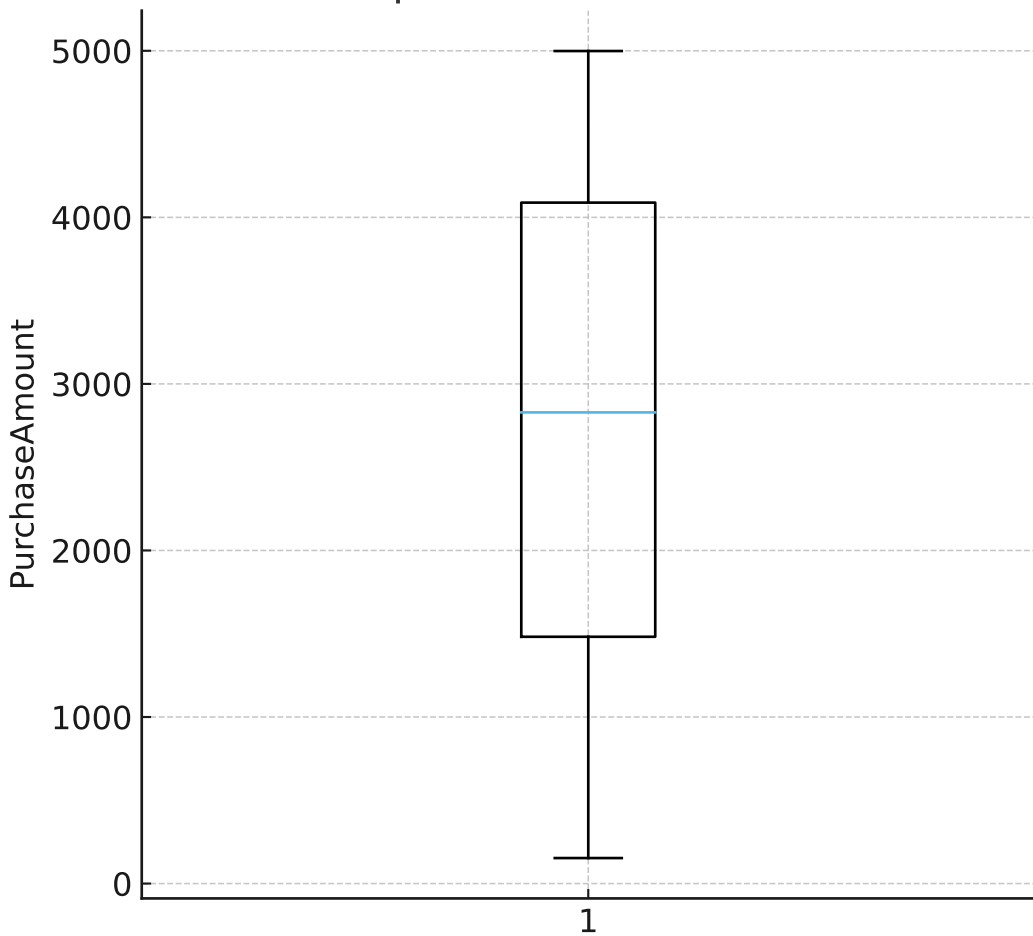
Boxplot: CustomerID



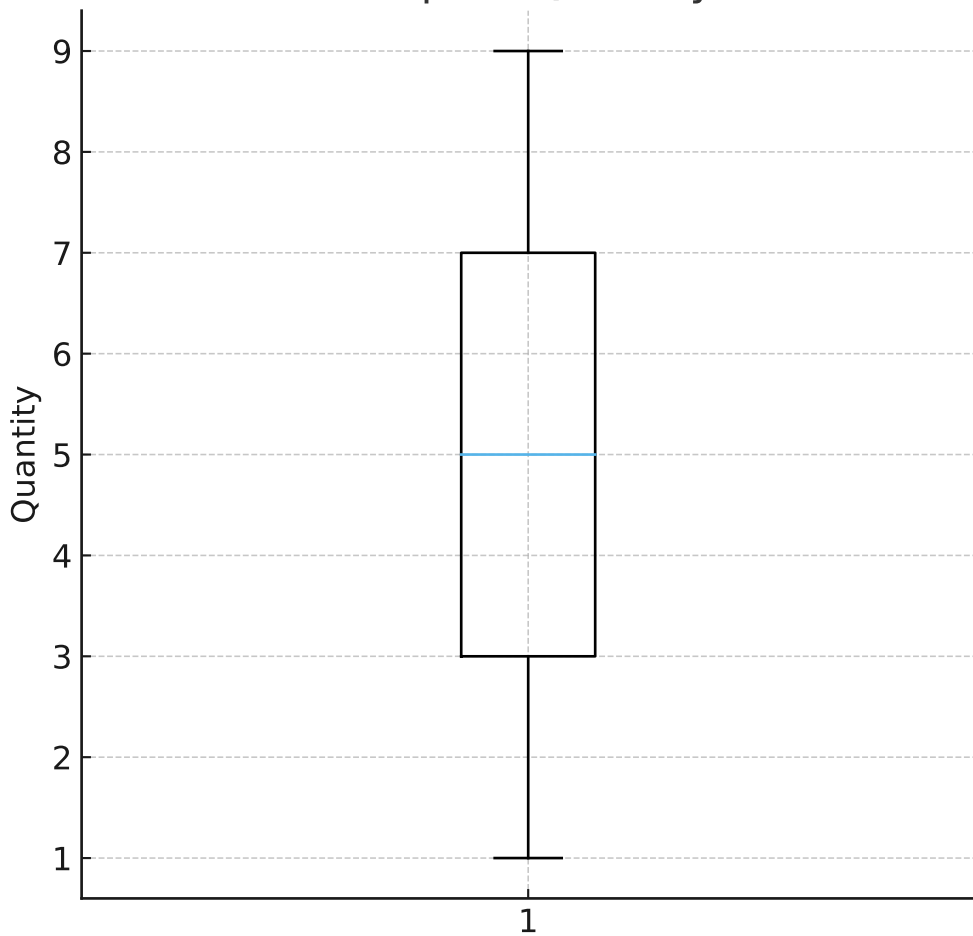
Boxplot: Age



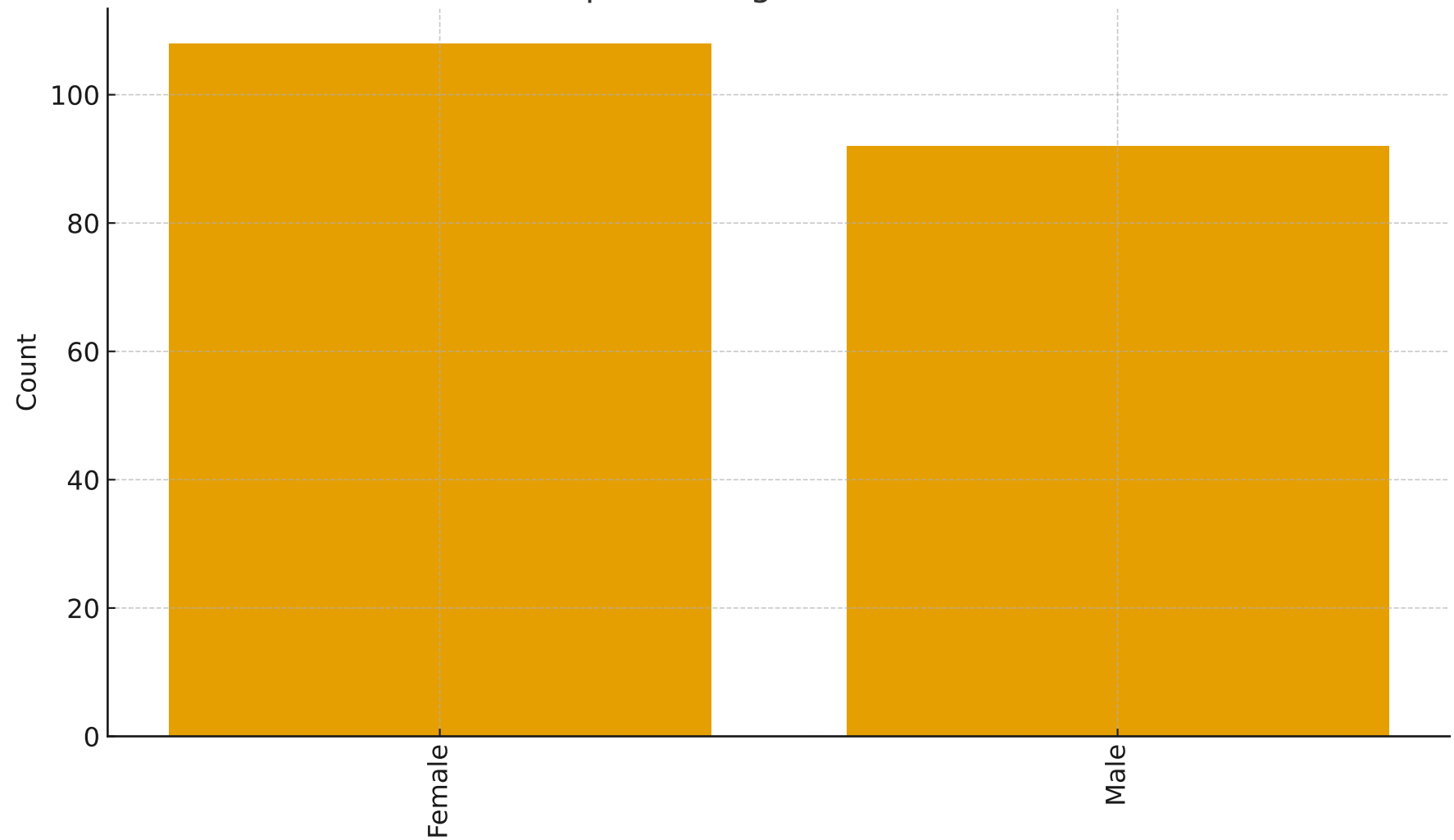
Boxplot: PurchaseAmount



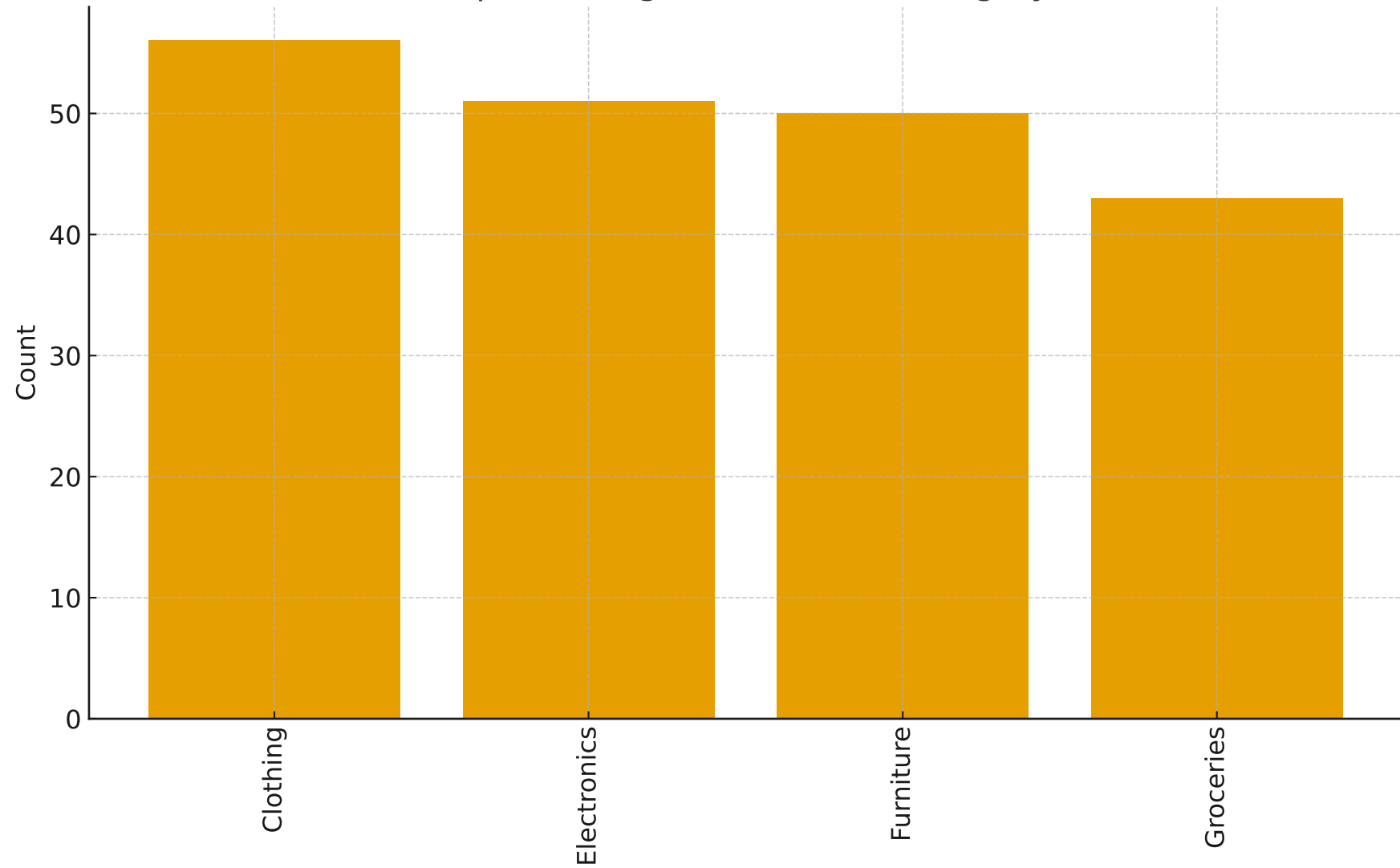
Boxplot: Quantity



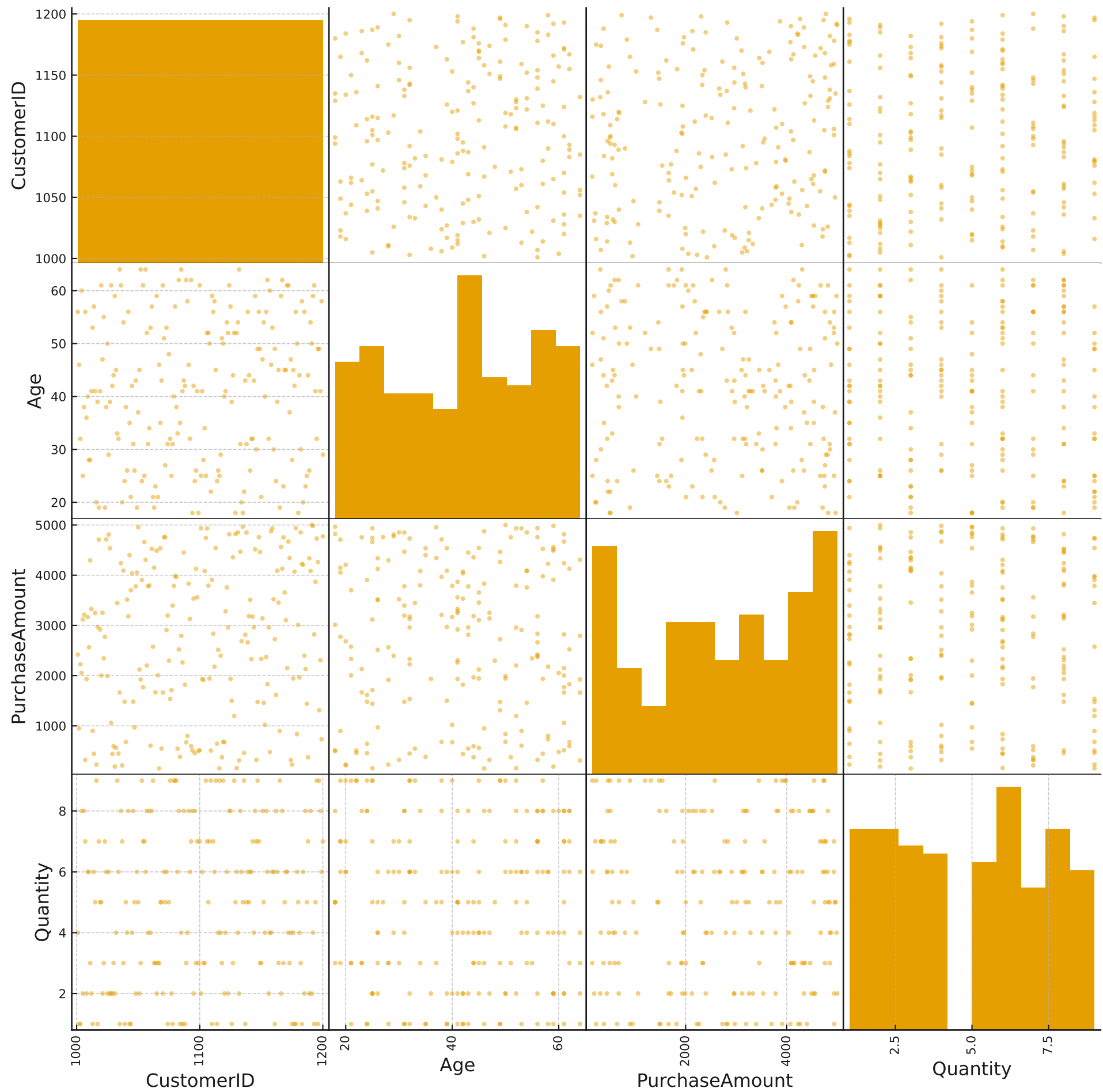
Top 15 categories: Gender



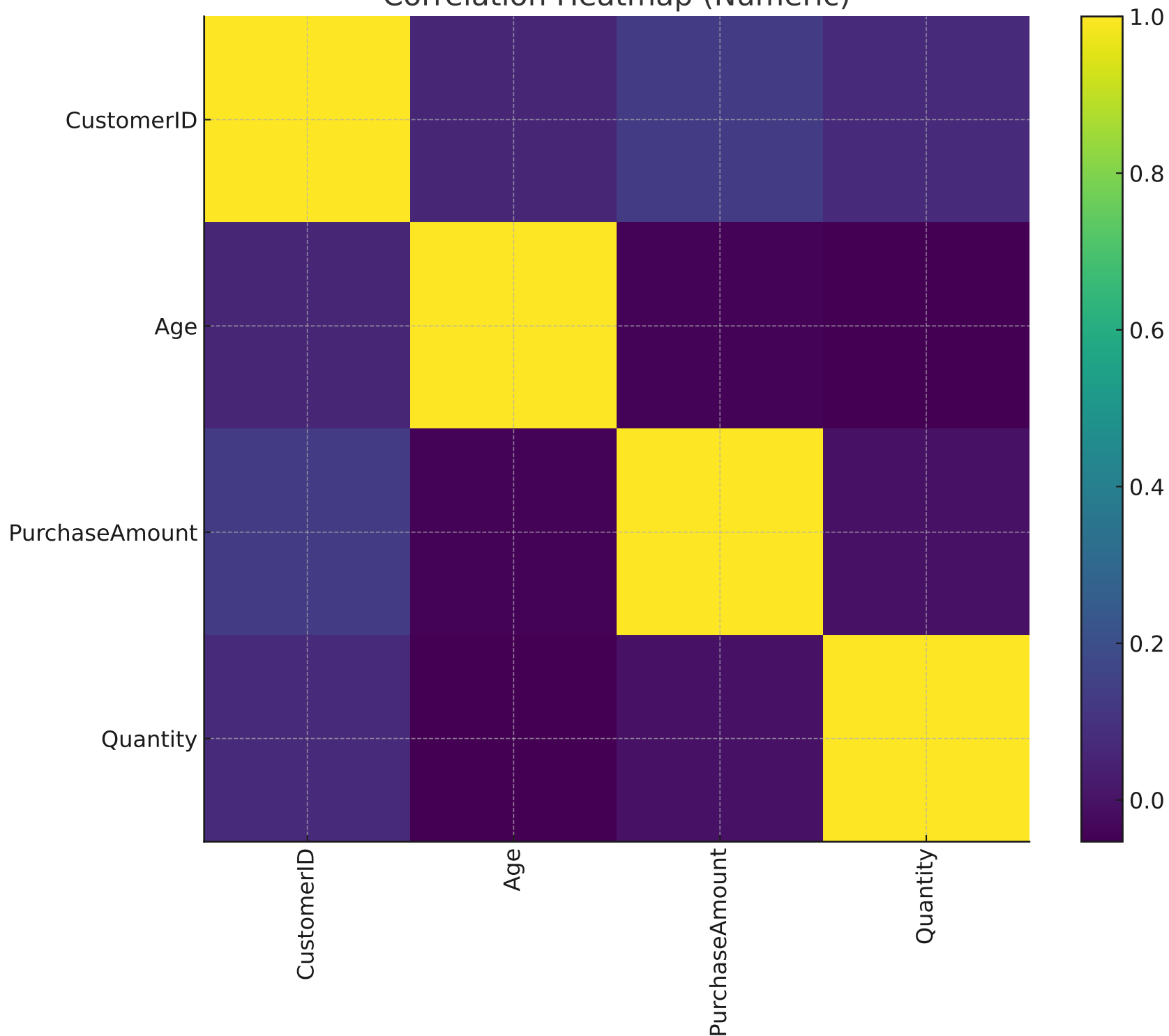
Top 15 categories: ProductCategory



Scatter Matrix (first 5 numeric columns)



Correlation Heatmap (Numeric)



Key Textual Summaries

Skewness (numeric):

CustomerID	0.000000	Age	-0.090408	PurchaseAmount	-0.167656	Quantity	0.025167
------------	----------	-----	-----------	----------------	-----------	----------	----------

Outlier counts by column (IQR method):

CustomerID	0	Age	0	PurchaseAmount	0	Quantity	0
------------	---	-----	---	----------------	---	----------	---