Assignment 2 – Exploring the Data

* List the features and identify their types (nominal, numeric, etc.)

A screenshot of a computer

AI-generated content may be incorrect.

🔍 Describing the Dataset

|  |  |  |
| --- | --- | --- |
| Feature List and Data Types | | |
| Feature | Data Type | Description |
| Customer ID | Numeric (Int) | Unique ID for each customer |
| Age | Numeric (Int) | Customer’s age |
| Gender | Nominal (object) | Gender of the customer |
| ProductCategory | Nominal (object) | Product category purchased |
| PurchaseAmount | Numeric (float) | Total amount spent |
| PaymentMethod | Nominal (object) | Method used for payment |
| StoreRegion | Nominal (object) | Region where purchase was made |
| PurchaseDate | DateTime | Date of the purchase |

* Summary Statistics (.describe() output)



**Count**: 5 rows

**Mean Purchase Amount**: 95.58

**Min Purchase Amount**: 13.5

**Max Purchase Amount**: 220.0

**Standard Deviation**: 80.12

**25% Quartile (Q1)**: 45.0

**Median (Q2)**: 78.9

**75% Quartile (Q3)**: 120.5

* Range, Mean and Standard deviation of PurchaseAmount

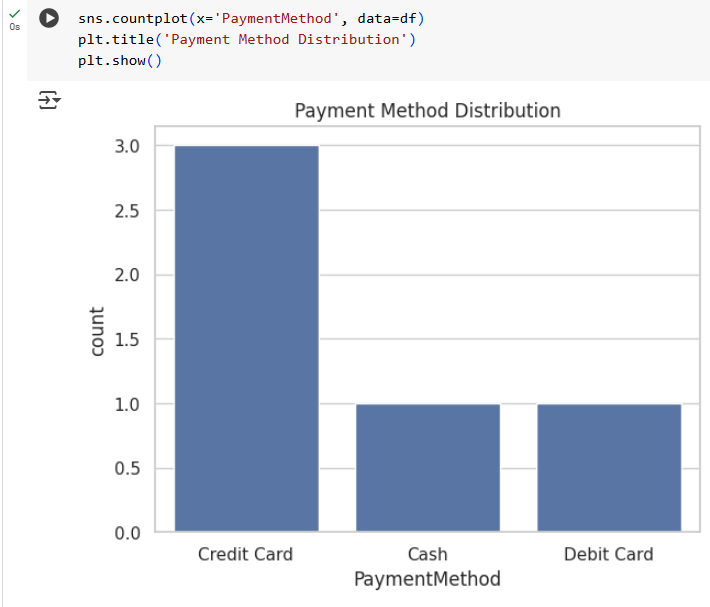


**Range**: 220.0 - 13.5 = 206.5

**Mean**: 95.58 (from .describe())

**Standard Deviation**: 80.12 (from .describe())

📊 Visualize the Data

* Include or describe two charts:
* Chart1: Histogram of PurchaseAmount
  + **Why used**: Shows the distribution of purchase values
  + **Reveals**: Most purchases are clustered in the lower range, with one high outlier
* Chart2: Bar Chart of PaymentMethod
  + 
  + **Why used**: Categorical variable best visualized with bar chart.
  + **Reveals**: Majority of purchases are made with credit card.

🔗 Explore Relationships

* Correlation Heatmap



* Explored correlation between **Age** and **Purchase Amount**
* Correlation Coefficient: 0.22 (weak positive correlation)
* Interpretation
  + There is a slight positive relationship between age and spending. Older customers may tend to spend a bit more.
  + This weak correlation suggests age might be a minor factor in predicting purchase amount, but it shouldn’t be the only feature considered for modeling.