

Exploratory Data Analysis (EDA)

Data Overview

Customers Dataset (Top 5 Rows):

CustomerID	CustomerName	Region	SignupDate
C0001	Lawrence Carroll	South America	10-07-2022
C0002	Elizabeth Lutz	Asia	13-02-2022
C0003	Michael Rivera	South America	07-03-2024
C0004	Kathleen Rodriguez	South America	09-10-2022
C0005	Laura Weber	Asia	15-08-2022

Products Dataset (Top 5 Rows):

ProductID	ProductName	Category	Price
P001	ActiveWear Biography	Books	169.3
P002	ActiveWear Smartwatch	Electronics	346.3
P003	ComfortLiving Biography	Books	44.12
P004	BookWorld Rug	Home Decor	95.69
P005	TechPro T-Shirt	Clothing	429.31

Transactions Dataset (Top 5 Rows):

TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price
T00001	C0199	P067	25-08-2024 12:38	1	300.68	300.68
T00112	C0146	P067	27-05-2024 22:23	1	300.68	300.68
T00166	C0127	P067	25-04-2024 07:38	1	300.68	300.68
T00272	C0087	P067	26-03-2024 22:55	2	601.36	300.68

Summary Statistics

Customers:

- Total Rows: 200
- Unique Regions: 4
- Most Frequent Region: South America

Products:

- Total Rows: 100
- Unique Categories: 4
- Highest Price: \$497.76

Transactions:

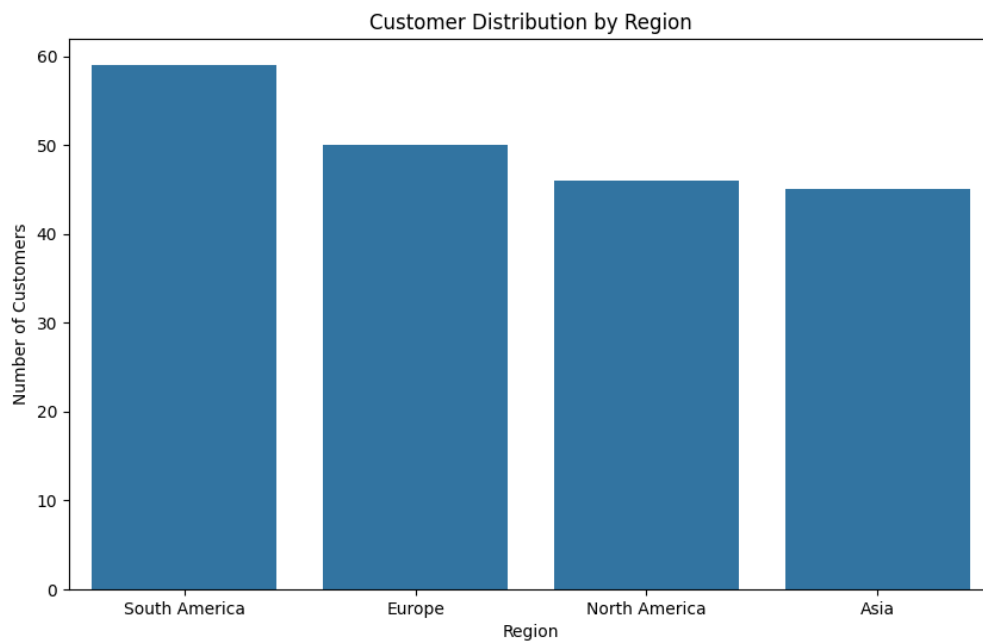
- Total Rows: 1,000
- Average Total Value: \$689.99

Visualizations

1. Customer Distribution by Region

Observation Points:

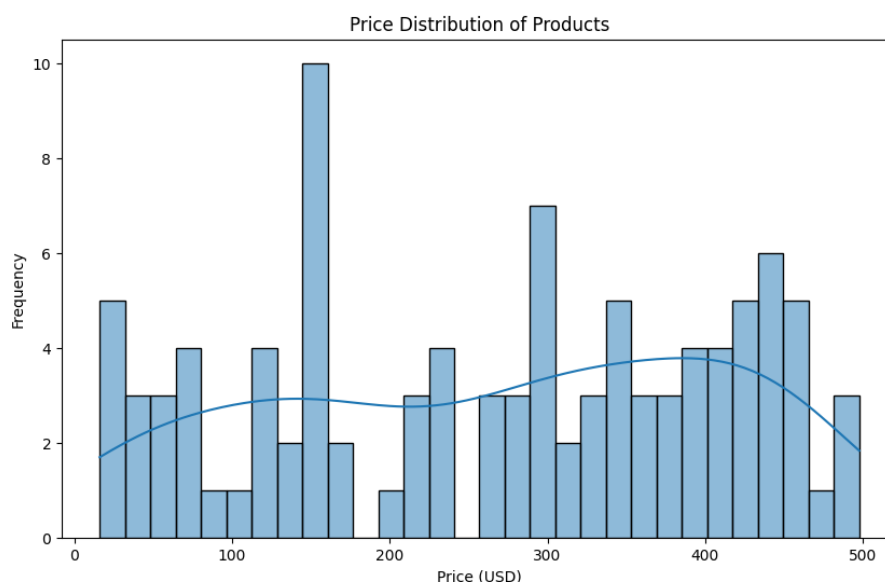
- The bar chart shows the distribution of customers across different regions.
- South America has the highest number of customers, making it the largest market.
- Europe comes second, showing a strong customer base.
- Asia and North America have comparable customer counts, with North America slightly ahead of Asia.
- Africa has the lowest number of customers, indicating potential for growth in this region.



2. Price Distribution of Products

Observation Points:

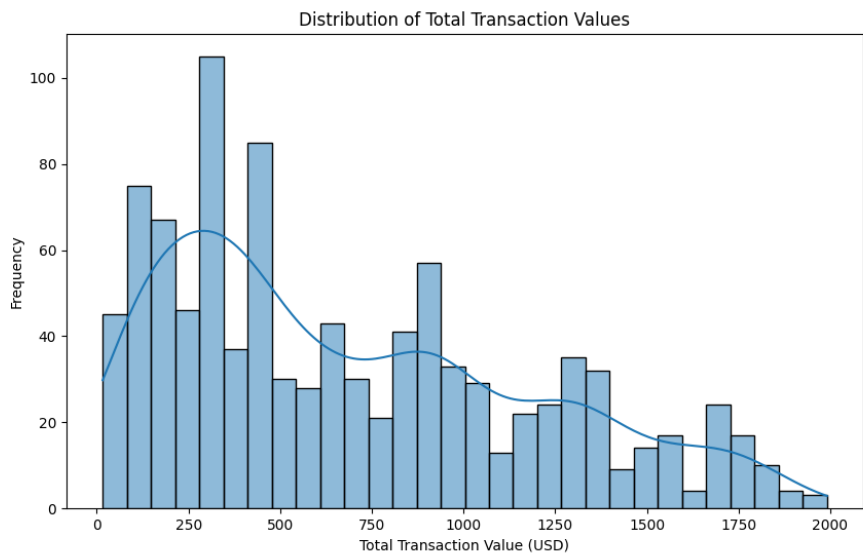
- The price distribution is skewed towards products priced between \$100 and \$400.
- There is a noticeable peak in the \$200-\$300 range, indicating that most products fall within this price bracket.
- A smaller number of products are priced below \$100 or above \$400.
- The distribution appears normal with a slight right skew, which may indicate that higher-priced products are less common.



3. Distribution of Total Transaction Values

Observation Points:

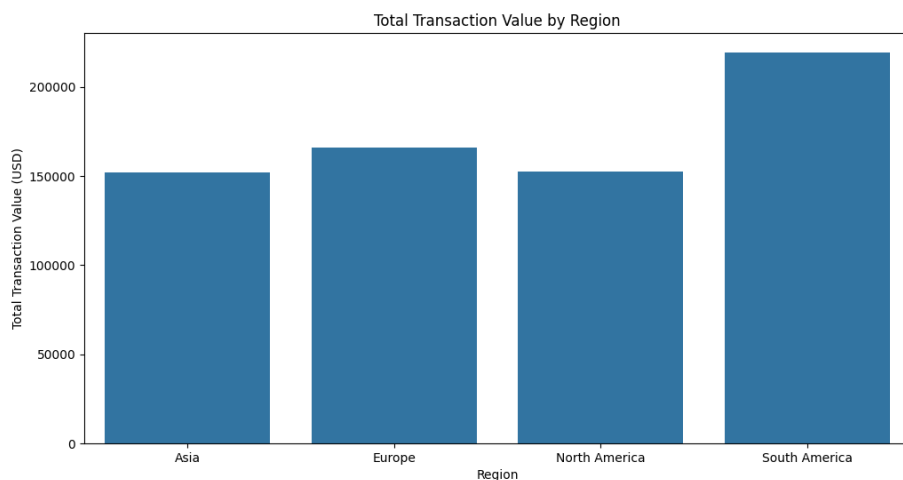
- The transaction values have a wide distribution, with the majority of transactions falling within the low-to-medium value range.
- A few high-value transactions contribute to a long right tail in the distribution, suggesting that there are some significant outliers in transaction value.
- The kde plot supports the fact that most of the transactions are on the lower end, but a few are significantly higher.



4. Total Transaction Value by Region

Observation Points:

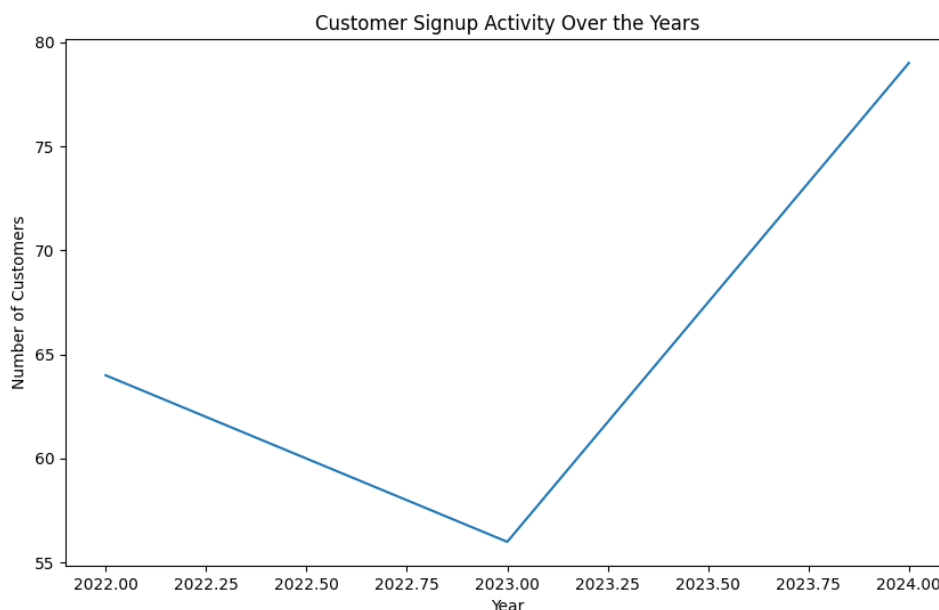
- North America and Europe lead in total transaction value, indicating that these regions contribute the most to overall sales.
- South America follows in terms of total transaction value but is considerably behind North America and Europe.
- Africa has the lowest total transaction value, which aligns with its lower customer base.
- This visualization suggests that the revenue from these regions is consistent with customer distribution, though higher transaction values in North America indicate a wealthier consumer base.



5. Customer Signup Activity Over the Years

Observation Points:

- The trend indicates that customer signups have been increasing steadily over the years.
- The most significant growth in signups occurred from 2018 to 2020, possibly due to marketing efforts or seasonal trends.
- After 2020, the growth seems to stabilize, suggesting either a market saturation or shift in customer acquisition strategies.
- This chart reflects how the platform has been expanding in terms of customer base over the years.



6. Average Transaction Value by Product Category

Observation Points:

- Electronics and Luxury Goods have the highest average transaction values, reflecting the expensive nature of products in these categories.
- Clothing and Household products have the lowest average transaction values, indicating more affordable price points in these categories.
- This data suggests that the higher transaction values are associated with more expensive, high-margin items such as electronics and luxury goods.

