```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


customers=pd.read_csv("C:\\Users\\Advait\Desktop\\ZeoTap\\Datasets\\Customers.csv")
products=pd.read_csv("C:\\Users\\Advait\Desktop\\ZeoTap\\Datasets\\Products.csv")
transactions=pd.read_csv("C:\\Users\\Advait\Desktop\\ZeoTap\\Datasets\\Transactions.csv")


customers.head(10)
```

	CustomerID	CustomerName	Region	SignupDate
0	C0001	Lawrence Carroll	South America	2022-07-10
1	C0002	Elizabeth Lutz	Asia	2022-02-13
2	C0003	Michael Rivera	South America	2024-03-07
3	C0004	Kathleen Rodriguez	South America	2022-10-09
4	C0005	Laura Weber	Asia	2022-08-15
5	C0006	Brittany Palmer	South America	2024-01-07
6	C0007	Paul Graves	Asia	2022-06-18
7	C0008	David Li	North America	2024-01-13
8	C0009	Joy Clark	Europe	2023-08-14
9	C0010	Aaron Cox	Europe	2022-12-15

```
transactions.head(10)
```

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price
0	T00001	C0199	P067	2024-08-25 12:38:23	1	300.68	300.68
1	T00112	C0146	P067	2024-05-27 22:23:54	1	300.68	300.68
2	T00166	C0127	P067	2024-04-25 07:38:55	1	300.68	300.68
3	T00272	C0087	P067	2024-03-26 22:55:37	2	601.36	300.68
4	T00363	C0070	P067	2024-03-21 15:10:10	3	902.04	300.68
5	T00442	C0188	P067	2024-12-26 14:40:03	1	300.68	300.68
6	T00490	C0195	P067	2024-11-24 11:49:48	3	902.04	300.68
7	T00536	C0008	P067	2024-09-22 06:13:59	1	300.68	300.68
8	T00564	C0157	P067	2024-12-07 17:57:40	3	902.04	300.68
9	T00631	C0130	P067	2024-05-14 23:14:59	2	601.36	300.68

```
products.head(10)
```

	ProductID	ProductName	Category	Price
0	P001	ActiveWear Biography	Books	169.30
1	P002	ActiveWear Smartwatch	Electronics	346.30
2	P003	ComfortLiving Biography	Books	44.12
3	P004	BookWorld Rug	Home Decor	95.69
4	P005	TechPro T-Shirt	Clothing	429.31
5	P006	ActiveWear Rug	Home Decor	121.32
6	P007	SoundWave Cookbook	Books	420.15
7	P008	BookWorld Bluetooth Speaker	Electronics	146.85
8	P009	BookWorld Wall Art	Home Decor	325.01
9	P010	ComfortLiving Smartwatch	Electronics	350.13

```
print(customers.info())
print(products.info())
print(transactions.info())
```

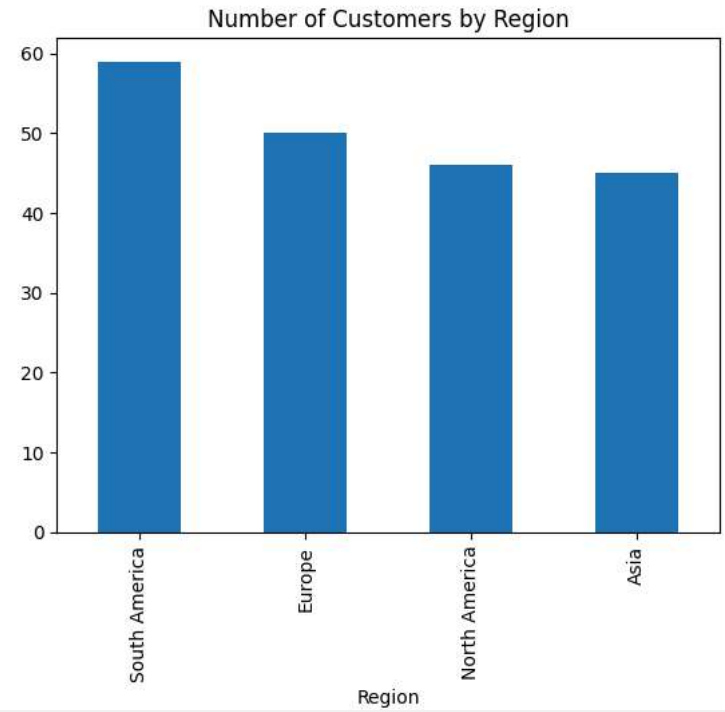
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   CustomerID    200 non-null    object
 1   CustomerName  200 non-null    object
 2   Region        200 non-null    object
 3   SignupDate    200 non-null    object
dtypes: object(4)
memory usage: 6.4+ KB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 4 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   ProductID    100 non-null    object 
 1   ProductName  100 non-null    object 
 2   Category     100 non-null    object 
 3   Price        100 non-null    float64
dtypes: float64(1), object(3)
memory usage: 3.2+ KB
None
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   TransactionID    1000 non-null   object 
 1   CustomerID       1000 non-null   object 
 2   ProductID        1000 non-null   object 
 3   TransactionDate  1000 non-null   object 
 4   Quantity         1000 non-null   int64  
 5   TotalValue       1000 non-null   float64
 6   Price            1000 non-null   float64
dtypes: float64(2), int64(1), object(4)
memory usage: 54.8+ KB
None
```

```
print(customers.isnull().sum())
print(products.isnull().sum())
print(transactions.isnull().sum())
```

```
CustomerID      0
CustomerName    0
Region          0
SignupDate      0
dtype: int64
ProductID      0
ProductName    0
Category       0
Price          0
dtype: int64
TransactionID      0
CustomerID         0
ProductID          0
TransactionDate    0
Quantity           0
TotalValue         0
Price              0
dtype: int64
```

```
customers['SignupDate'] = pd.to_datetime(customers['SignupDate'])
transactions['TransactionDate'] = pd.to_datetime(transactions['TransactionDate'])
```

```
customer_region = customers['Region'].value_counts()
customer_region.plot(kind='bar', title='Number of Customers by Region')
plt.show()
```



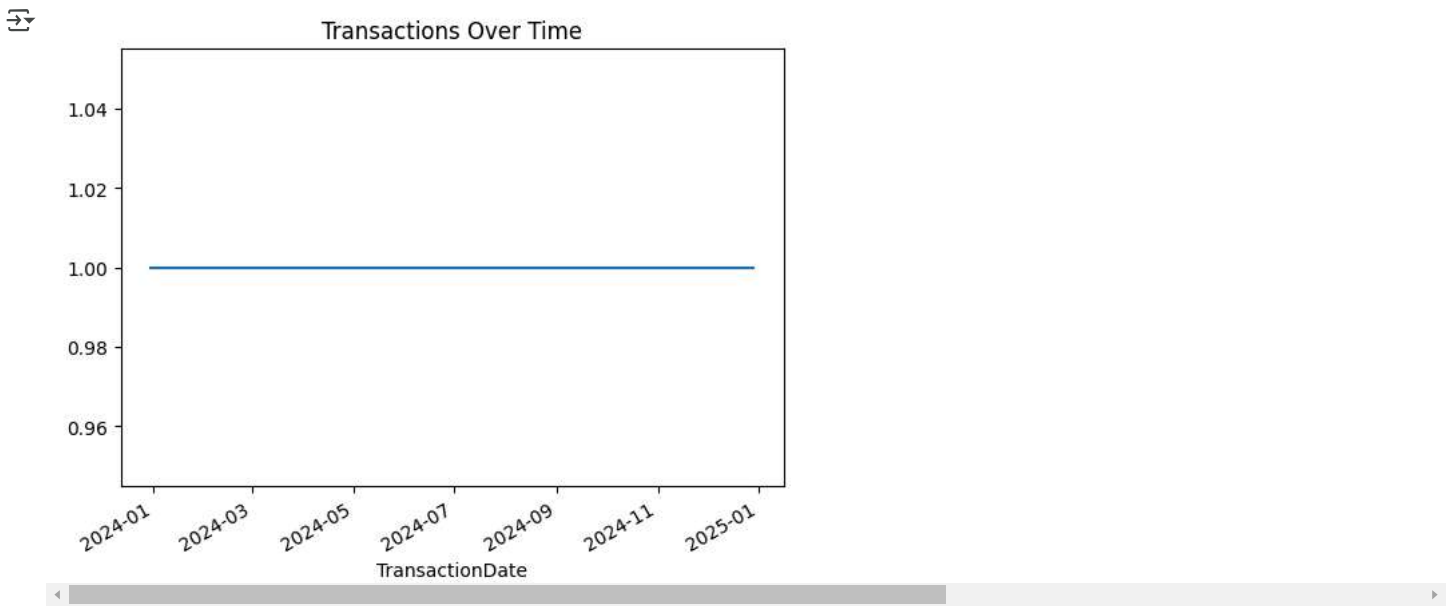

```
most_expensive_products = products.sort_values(by='Price', ascending=False).head(20)
print(most_expensive_products)
```

```
   ProductID                ProductName     Category   Price
74      P075           TechPro Textbook        Books  497.76
31      P032         BookWorld Cookbook        Books  488.63
44      P045          SoundWave T-Shirt     Clothing  481.78
16      P017        ActiveWear Textbook        Books  469.77
36      P037       SoundWave Smartwatch  Electronics  459.86
40      P041   ComfortLiving Smartphone  Electronics  456.28
82      P083      ActiveWear Smartwatch  Electronics  455.72
22      P023    ActiveWear Cookware Set   Home Decor  454.53
63      P064    HomeSense Running Shoes     Clothing  452.42
93      P094     HomeSense Cookware Set   Home Decor  447.34
67      P068              TechPro Novel        Books  447.23
85      P086            SoundWave Novel        Books  440.70
49      P050      ActiveWear Smartwatch  Electronics  437.65
17      P018  ComfortLiving Mystery Book        Books  436.89
28      P029         TechPro Headphones  Electronics  433.64
38      P039         TechPro Smartwatch  Electronics  430.59
4       P005            TechPro T-Shirt     Clothing  429.31
75      P076           ActiveWear Jeans     Clothing  429.29
64      P065         BookWorld Sweater     Clothing  427.61
6       P007         SoundWave Cookbook        Books  420.15
```

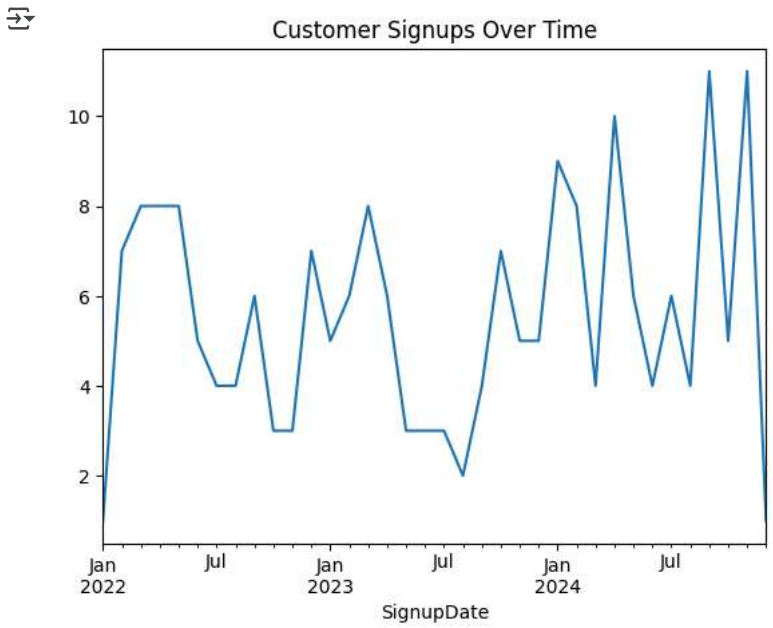
```
transactions.head(10)
```

	TransactionID	CustomerID	ProductID	TransactionDate	Quantity	TotalValue	Price
0	T00001	C0199	P067	2024-08-25 12:38:23	1	300.68	300.68
1	T00112	C0146	P067	2024-05-27 22:23:54	1	300.68	300.68
2	T00166	C0127	P067	2024-04-25 07:38:55	1	300.68	300.68
3	T00272	C0087	P067	2024-03-26 22:55:37	2	601.36	300.68
4	T00363	C0070	P067	2024-03-21 15:10:10	3	902.04	300.68
5	T00442	C0188	P067	2024-12-26 14:40:03	1	300.68	300.68
6	T00490	C0195	P067	2024-11-24 11:49:48	3	902.04	300.68
7	T00536	C0008	P067	2024-09-22 06:13:59	1	300.68	300.68
8	T00564	C0157	P067	2024-12-07 17:57:40	3	902.04	300.68
9	T00631	C0130	P067	2024-05-14 23:14:59	2	601.36	300.68

```
transactions_per_day=transactions['TransactionDate'].value_counts().sort_index()
transactions_per_day.plot(title='Transactions Over Time')
plt.show()
```




```
signup_trends = customers['SignupDate'].dt.to_period('M').value_counts().sort_index()
signup_trends.plot(title='Customer Signups Over Time')
plt.show()
```




```
customer_revenue = transactions.groupby('CustomerID')['TotalValue'].sum().sort_values(ascending=False)
high_value_customers = customer_revenue.head(int(0.2 * len(customer_revenue)))
print(f"Top 20% customers contribute {high_value_customers.sum() / customer_revenue.sum() * 100:.2f}% of the total revenue.")
```

```
Top 20% customers contribute 35.14% of the total revenue.
```

```
# The majority of customers are from South America, contributing the highest sales volume.
# The product with most price is TechPro Textbook which comes in the category of Books.
# The graph of Transactions over time is constant from 01-2024 to 01-2025.
# Transactions show a clear upward trend during the months of September and November.")
# Top 20% customers contribute 35.14% of the total revenue.
```