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

```
df=pd.read_csv('/content/ecommerce_customer_data_custom_ratios.csv')
```

```
df
```

	Customer ID	Purchase Date	Product Category	Product Price	Quantity	Total Purchase Amount	Payment Method	Customer Age	Returns	Customer Name	Age	Gender	Churn
0	46251	2020-09-08 09:38:32	Electronics	12	3	740	Credit Card	37	0.0	Christine Hernandez	37	Male	0
1	46251	2022-03-05 12:56:35	Home	468	4	2739	PayPal	37	0.0	Christine Hernandez	37	Male	0
2	46251	2022-05-23 18:18:01	Home	288	2	3196	PayPal	37	0.0	Christine Hernandez	37	Male	0
3	46251	2020-11-12 13:13:29	Clothing	196	1	3509	PayPal	37	0.0	Christine Hernandez	37	Male	0
4	13593	2020-11-27 17:55:11	Home	449	1	3452	Credit Card	49	0.0	James Grant	49	Female	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...

```
df.head(4)
```

	Customer ID	Purchase Date	Product Category	Product Price	Quantity	Total Purchase Amount	Payment Method	Customer Age	Returns	Customer Name	Age	Gender	Churn
0	46251	2020-09-08 09:38:32	Electronics	12	3	740	Credit Card	37	0.0	Christine Hernandez	37	Male	0
1	46251	2022-03-05	Home	468	4	2739	PayPal	37	0.0	Christine	37	Male	0

```
df.tail(4)
```

	Customer ID	Purchase Date	Product Category	Product Price	Quantity	Total Purchase Amount	Payment Method	Customer Age	Returns	Customer Name	Age	Gender	Churn
249996	48835	2021-11-23 01:30:42	Home	27	1	3615	Credit Card	42	1.0	Jeremy Rush	42	Female	1
249997	21019	2020-07-02	Home	17	5	2466	Cash	41	0.0	Tina	41	Male	0

```
df.shape
```

```
(250000, 13)
```

```
df.shape[0]
```

```
250000
```

```
df.shape[1]
```

```
13
```

```
df.isnull().sum()
```

```
Customer ID                0
Purchase Date              0
Product Category           0
Product Price              0
Quantity                   0
Total Purchase Amount      0
Payment Method             0
Customer Age               0
Returns                47596
```

```
Customer Name               0
Age                         0
Gender                      0
Churn                       0
dtype: int64
```

(df.isnull().sum()/df.shape[0])*100

```
df.isnull().sum().sum()
```

```
47596
```

```
df.notnull().sum()
```

```
Customer ID              250000
Purchase Date            250000
Product Category         250000
Product Price            250000
Quantity                 250000
Total Purchase Amount    250000
Payment Method           250000
Customer Age             250000
Returns                  202404
Customer Name            250000
Age                      250000
Gender                   250000
Churn                    250000
dtype: int64
```

```
df.notnull().sum().sum()
```

```
3202404
```

```
df.dtypes
```

```
Customer ID                int64
Purchase Date             object
Product Category          object
Product Price              int64
Quantity                   int64
Total Purchase Amount      int64
Payment Method            object
Customer Age               int64
Returns                  float64
Customer Name             object
Age                        int64
Gender                    object
Churn                      int64
dtype: object
```

```
df['Churn'] = df['Churn'].astype(bool).replace({0:True,1:False})
```

```
df.dtypes
```

```
Customer ID                int64
Purchase Date             object
Product Category          object
Product Price              int64
Quantity                   int64
Total Purchase Amount      int64
Payment Method            object
Customer Age               int64
Returns                  float64
Customer Name             object
Age                        int64
Gender                    object
Churn                       bool
dtype: object
```

```
df['Gender'] = df['Gender'].astype('category').replace({'Male': 'Male', 'Female': 'Female'})
```

```
df.dtypes
```

```
Customer ID                int64
Purchase Date             object
Product Category          object
Product Price              int64
Quantity                   int64
Total Purchase Amount      int64
Payment Method            object
```

```
Customer Age               int64
Returns                  float64
Customer Name             object
Age                     category
Gender                  category
Churn                       bool
dtype: object
```

```
df
```

	Customer ID	Purchase Date	Product Category	Product Price	Quantity	Total Purchase Amount	Payment Method	Customer Age
0	46251	2020-09-08 09:38:32	Electronics	12	3	740	Credit Card	37
1	46251	2022-03-05 12:56:35	Home	468	4	2739	PayPal	37
2	46251	2022-05-23 18:18:01	Home	288	2	3196	PayPal	37
3	46251	2020-11-12 13:13:29	Clothing	196	1	3509	PayPal	37
4	13593	2020-11-27 17:55:11	Home	449	1	3452	Credit Card	49
...	...	...	...	...	...	...	...	...

```
df.isnull().sum()
```

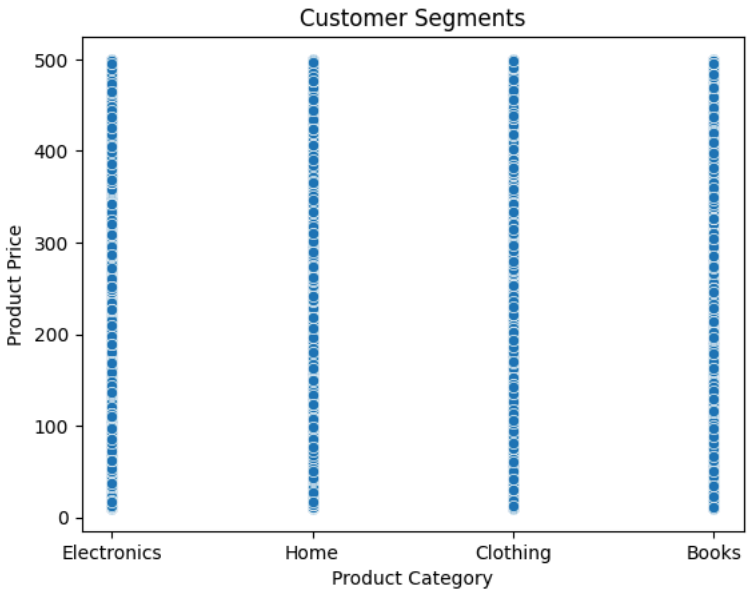
```
Customer ID                 0
Purchase Date               0
Product Category            0
Product Price               0
Quantity                    0
Total Purchase Amount       0
Payment Method              0
Customer Age                0
Returns                 47596
Customer Name               0
Age                         0
Gender                      0
Churn                       0
dtype: int64
```

```
df = df.dropna()
```

```
df.isnull().sum()
```

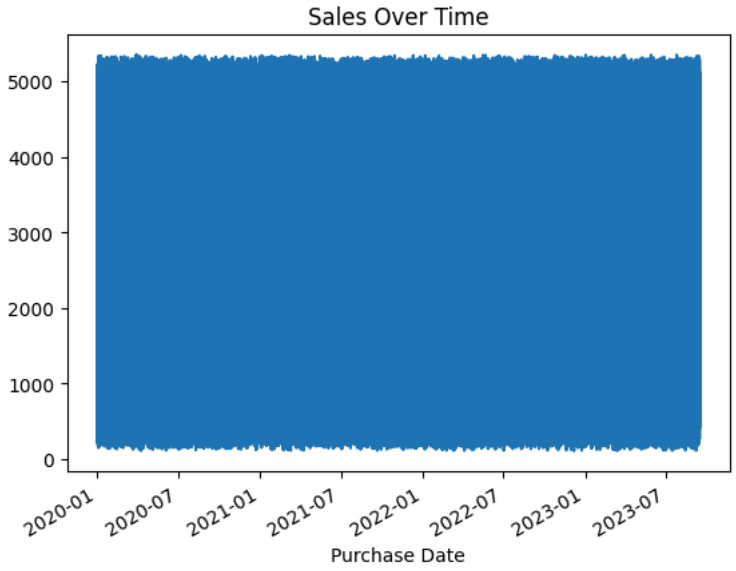
```
Customer ID             0
Purchase Date           0
Product Category        0
Product Price           0
Quantity                0
Total Purchase Amount   0
Payment Method          0
Customer Age            0
Returns                 0
Customer Name           0
Age                     0
Gender                  0
Churn                   0
dtype: int64
```

```
sns.scatterplot(x='Product Category', y='Product Price', data=df)
plt.title('Customer Segments')
plt.show()
```

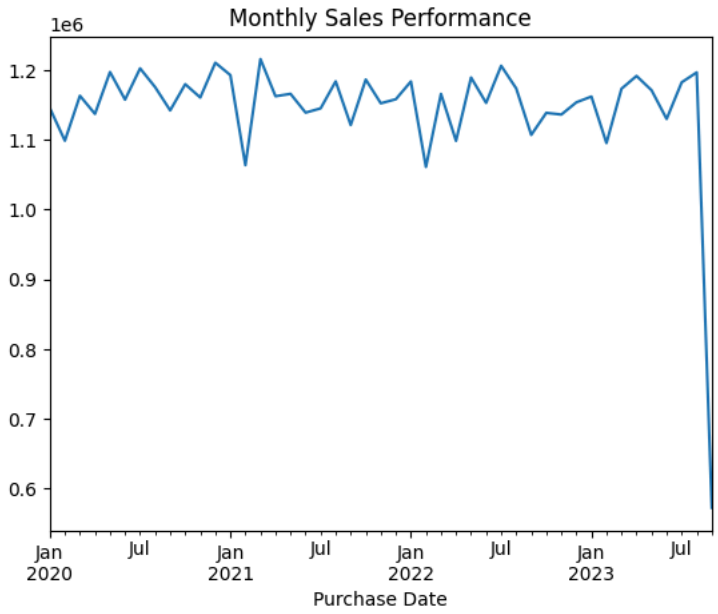



```
# Line Chart
df['Purchase Date'] = pd.to_datetime(df['Purchase Date'])
df.set_index('Purchase Date', inplace=True)
df['Total Purchase Amount'].plot()
plt.title('Sales Over Time')
plt.show()
```

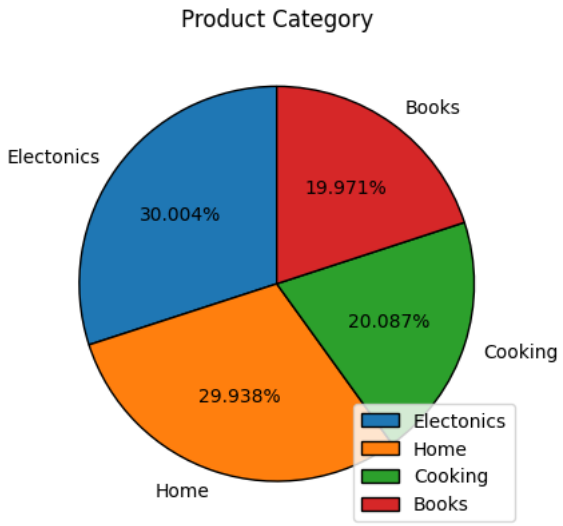
```
<ipython-input-76-d03869d8a7ce>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/u
  df['Purchase Date'] = pd.to_datetime(df['Purchase Date'])
```



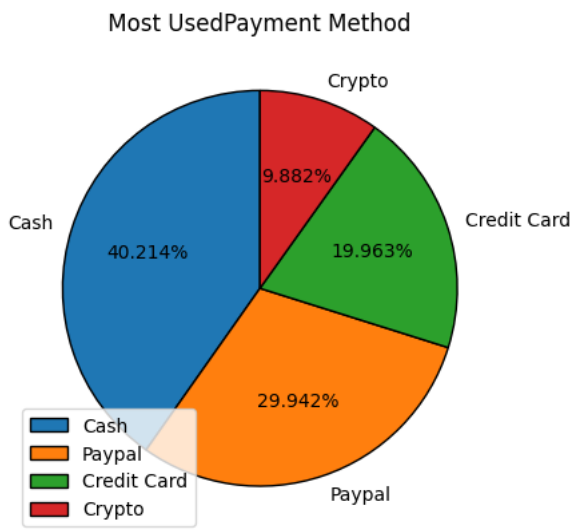

```
# Line Chart of Sales Performance
df.resample('M')['Product Price'].sum().plot()
plt.title('Monthly Sales Performance')
plt.show()
```




```
plt.pie(x=df['Product Category'].value_counts(),labels=['Electonics', 'Home', 'Cooking', 'Books'] ,startangle=90,autopct="%0.3f%%",coun
plt.legend(loc="lower right")
plt.title("Product Category")
plt.show()
```




```
plt.pie(x=df['Payment Method'].value_counts(),labels=['Cash', 'Paypal', 'Credit Card', 'Crypto'] ,startangle=90,autopct="%0.3f%%",count
plt.legend(loc="lower left")
plt.title(" Most UsedPayment Method ")
plt.show()
```



```
plt.hist(df['Product Price'], bins=30, edgecolor='black')
plt.title('Histogram of Product Category')
plt.xlabel('Product Category')
plt.ylabel('Product Price')
plt.grid(True)
plt.show()
```