


```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files

# Set styles
sns.set(style='whitegrid')

# Load dataset
df = pd.read_csv('Customer_Behaviour.csv') # or the name of your file

# Check first few rows
df.head()
```




	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

+ Code

+ Text

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



	0
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

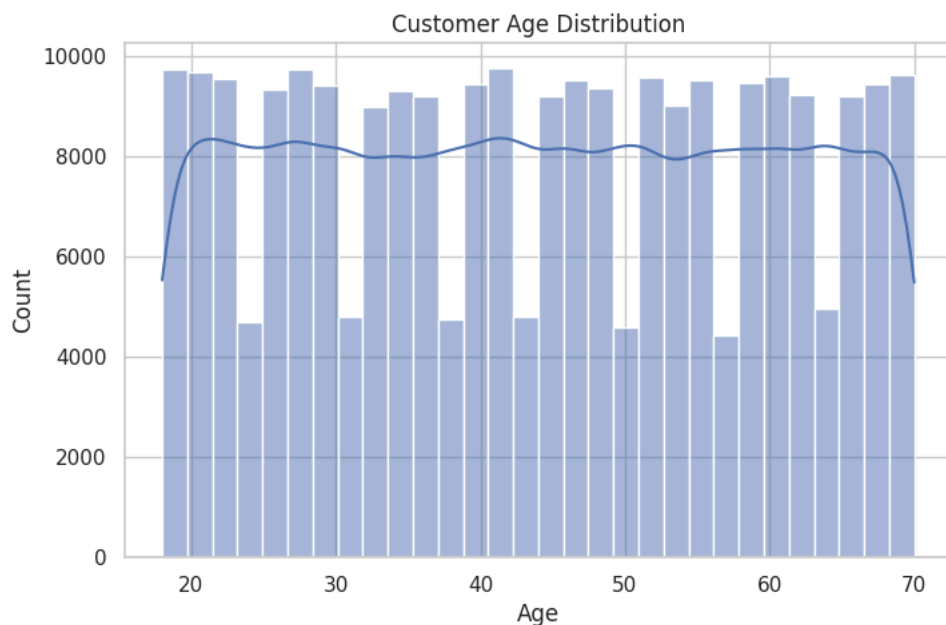
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 250000 entries, 0 to 249999
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Customer ID           250000 non-null int64
1   Purchase Date         250000 non-null object
2   Product Category      250000 non-null object
3   Product Price         250000 non-null int64
4   Quantity              250000 non-null int64
5   Total Purchase Amount 250000 non-null int64
6   Payment Method        250000 non-null object
7   Customer Age          250000 non-null int64
8   Returns               202404 non-null float64
9   Customer Name         250000 non-null object
10  Age                   250000 non-null int64
11  Gender                250000 non-null object
12  Churn                 250000 non-null int64
dtypes: float64(1), int64(7), object(5)
```

memory usage: 24.8+ MB

```
df = df.drop_duplicates()
```

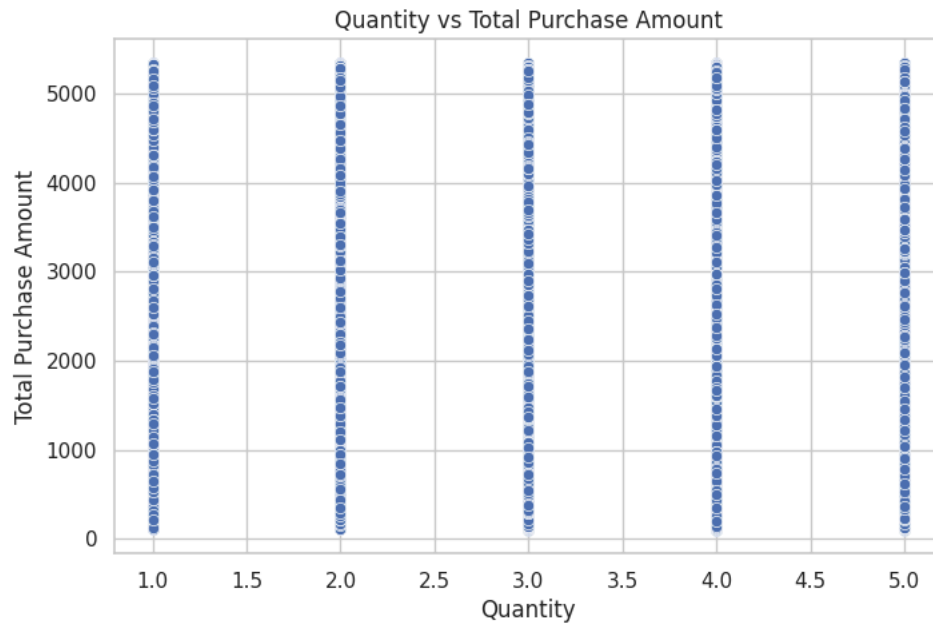
```
# Age distribution
plt.figure(figsize=(8, 5))
sns.histplot(df['Age'], bins=30, kde=True)
plt.title("Customer Age Distribution")
plt.show()
```



```
plt.figure(figsize=(8, 5))
sns.scatterplot(
    x='Product Price',
    y='Total Purchase Amount',
    data=df
)
plt.title("Product Price vs Total Purchase Amount")
plt.xlabel("Product Price")
plt.ylabel("Total Purchase Amount")
plt.grid(True)
plt.show()
```



```
plt.figure(figsize=(8, 5))
sns.scatterplot(
    x='Quantity',
    y='Total Purchase Amount',
    data=df
)
plt.title("Quantity vs Total Purchase Amount")
plt.xlabel("Quantity")
plt.ylabel("Total Purchase Amount")
plt.grid(True)
plt.show()
```



```
print(df.columns.tolist())
```

```
['Customer ID', 'Purchase Date', 'Product Category', 'Product Price', 'Quantity', 'Total Purchase Amount', 'Payment Method', 'Customer A
```

```
# Correlation Heatmap
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```



Correlation Heatmap



```
# Display summary statistics
df.describe()
```



	Customer ID	Quantity	Product Price	Quantity	Total Purchase Amount	Customer Age	Returns	Age	Churn
count	250000.00000	250000.00000	250000.00000	250000.00000	250000.00000	250000.00000	202404.00000	250000.00000	250000.00000
mean	25004.03624	251.659512	141.568577	1.414694	1442.933565	43.940528	0.497861	43.940528	0.199496
std	14428.27959	141.568577	1.414694	1.414694	1442.933565	15.350246	0.499997	15.350246	0.399622
min	1.00000	10.000000	1.000000	1.000000	100.000000	18.000000	0.000000	18.000000	0.000000
25%	12497.75000	132.000000	2.000000	2.000000	1477.000000	31.000000	0.000000	31.000000	0.000000
50%	25018.00000	255.000000	3.000000	3.000000	2724.000000	44.000000	0.000000	44.000000	0.000000
75%	37506.00000	377.000000	4.000000	4.000000	3974.000000	57.000000	1.000000	57.000000	0.000000

```
# Save cleaned data
df.to_csv('cleaned_customer_data.csv', index=False)
```

```
# Download file
files.download('cleaned_customer_data.csv')
```



Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.