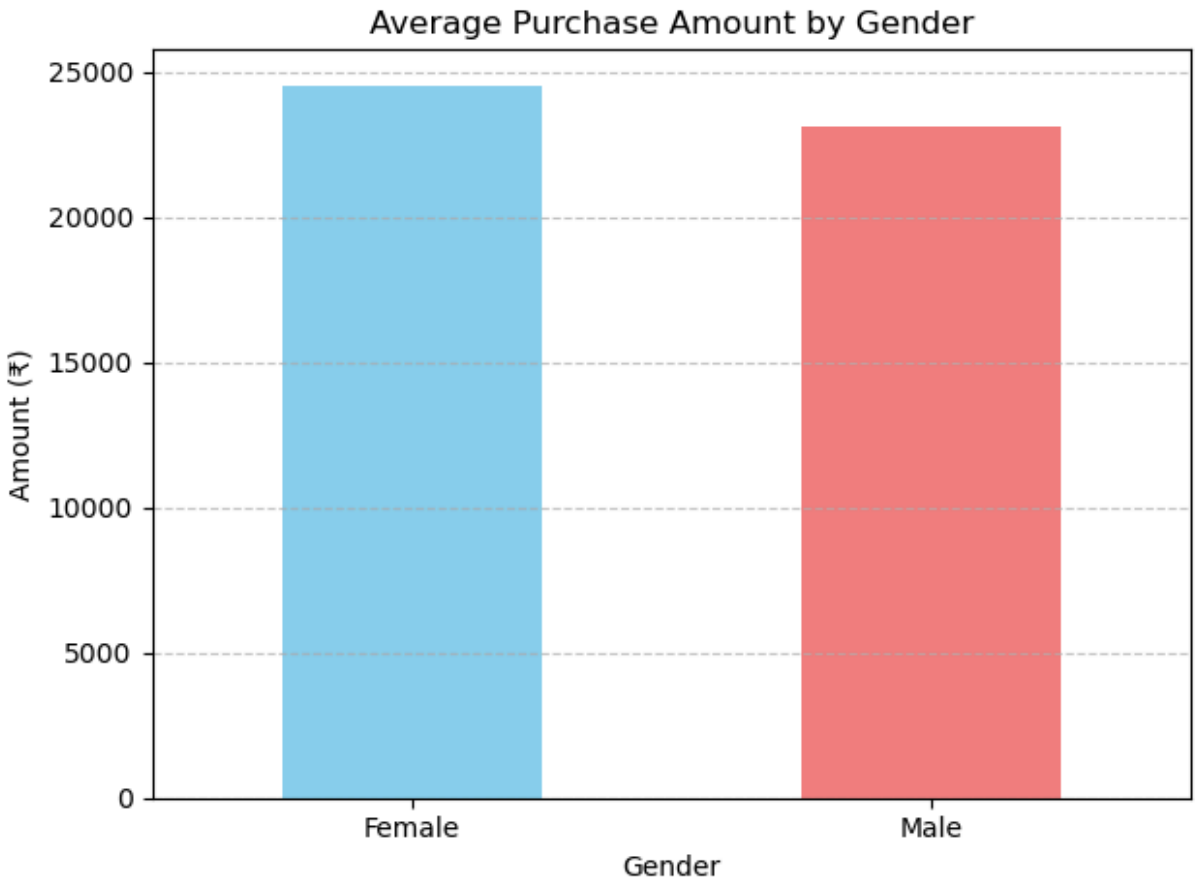
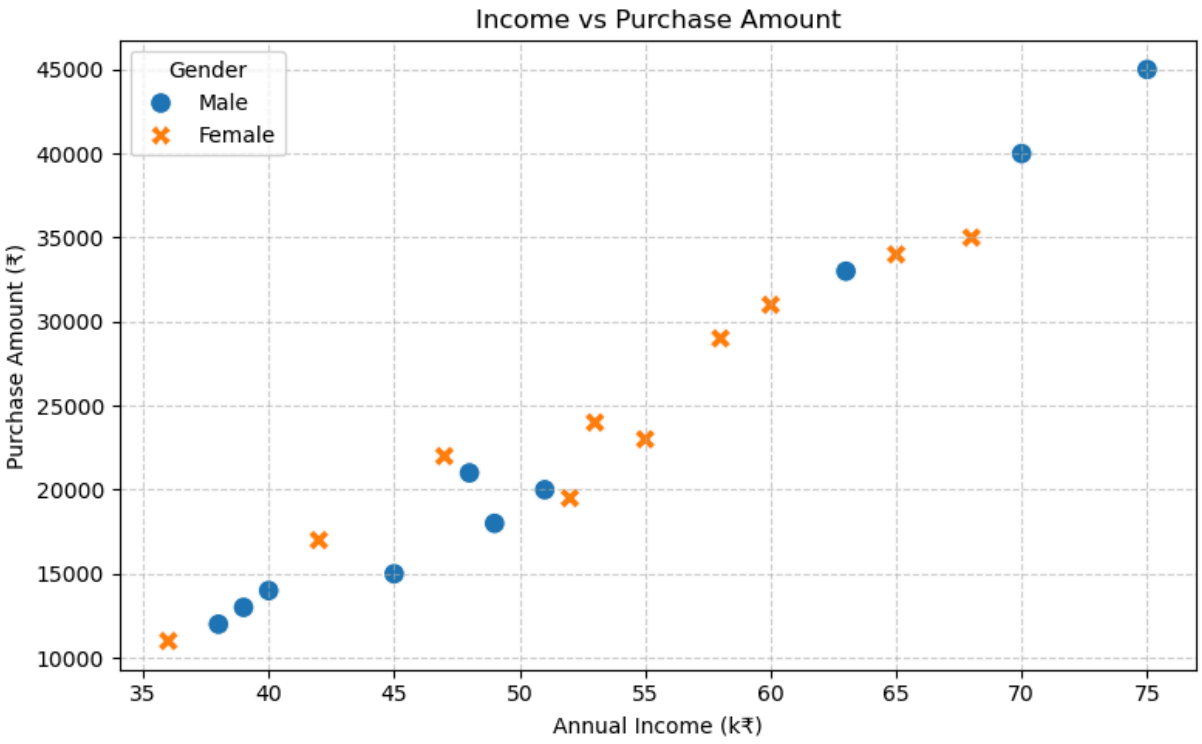
In [3]:
```
import pandas as pd

# Load the dataset
df = pd.read_excel("Ecommerce_Customers.csv.xlsx")

# Show first 5 rows
df.head()

# Basic info
print("\nDataset Info:")
print(df.info())

# Check for missing values
print("\nMissing Values:")
print(df.isnull().sum())

# Rename messy column names for easier handling
df.rename(columns={
    'Annual Income (kâ,¹)': 'Annual Income (k₹)',
    'Purchase Amount (â,¹)': 'Purchase Amount (₹)'
}, inplace=True)

# Confirm new column names
print(df.columns)

gender_purchase = df.groupby("Gender")["Purchase Amount (₹)"].mean()
print(gender_purchase)

import matplotlib.pyplot as plt

gender_purchase.plot(kind='bar', color=['skyblue', 'lightcoral'])
plt.title("Average Purchase Amount by Gender")
plt.ylabel("Amount (₹)")
plt.xlabel("Gender")
plt.xticks(rotation=0)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

import seaborn as sns

plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x="Annual Income (k₹)", y="Purchase Amount (₹)", hue="Gend
plt.title("Income vs Purchase Amount")
plt.xlabel("Annual Income (k₹)")
plt.ylabel("Purchase Amount (₹)")
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()

category_sales = df.groupby("Product Category")["Purchase Amount (₹)"].sum().sort_v
print(category_sales)

plt.figure(figsize=(8, 5))
```

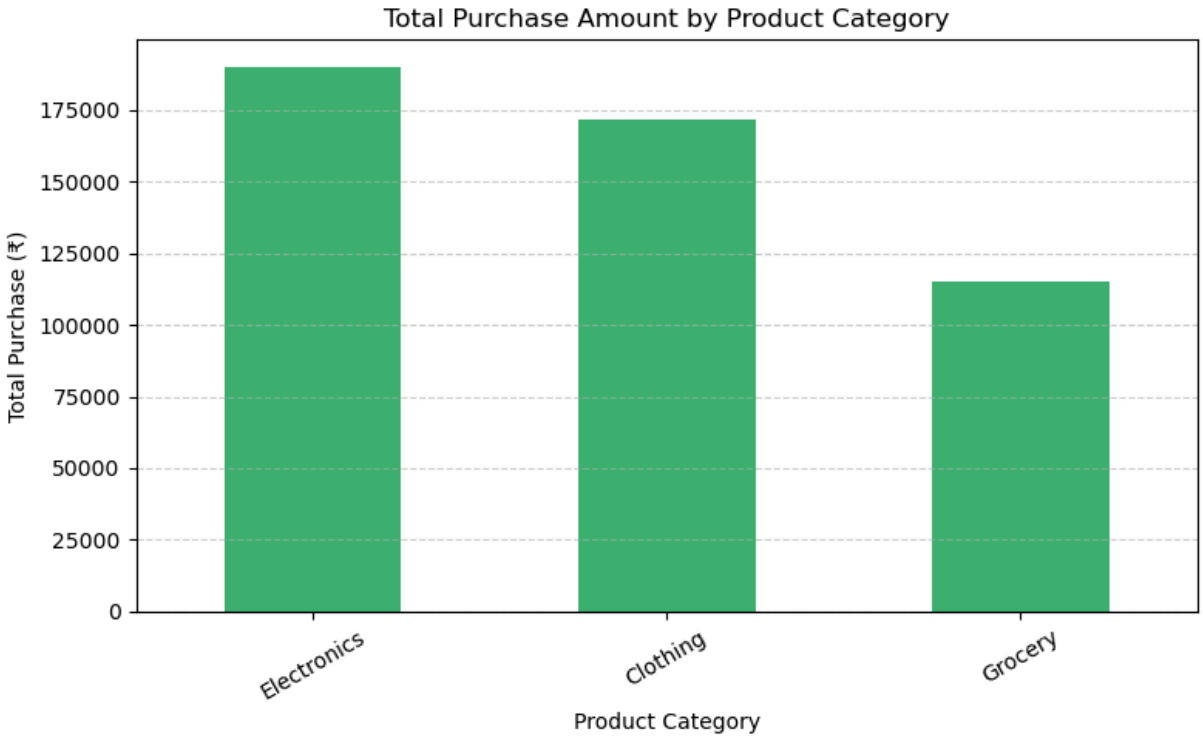
```
category_sales.plot(kind='bar', color='mediumseagreen')
plt.title("Total Purchase Amount by Product Category")
plt.ylabel("Total Purchase (₹)")
plt.xlabel("Product Category")
plt.xticks(rotation=30)
plt.grid(axis='y', linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()


plt.figure(figsize=(8, 5))
sns.scatterplot(data=df, x="Age", y="Spending Score (1-100)", hue="Gender", s=100)
plt.title("Age vs Spending Score")
plt.xlabel("Customer Age")
plt.ylabel("Spending Score (1-100)")
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()


!jupyter nbconvert --to html Customer_Purchase_Analysis_by_YourName.ipynb
```

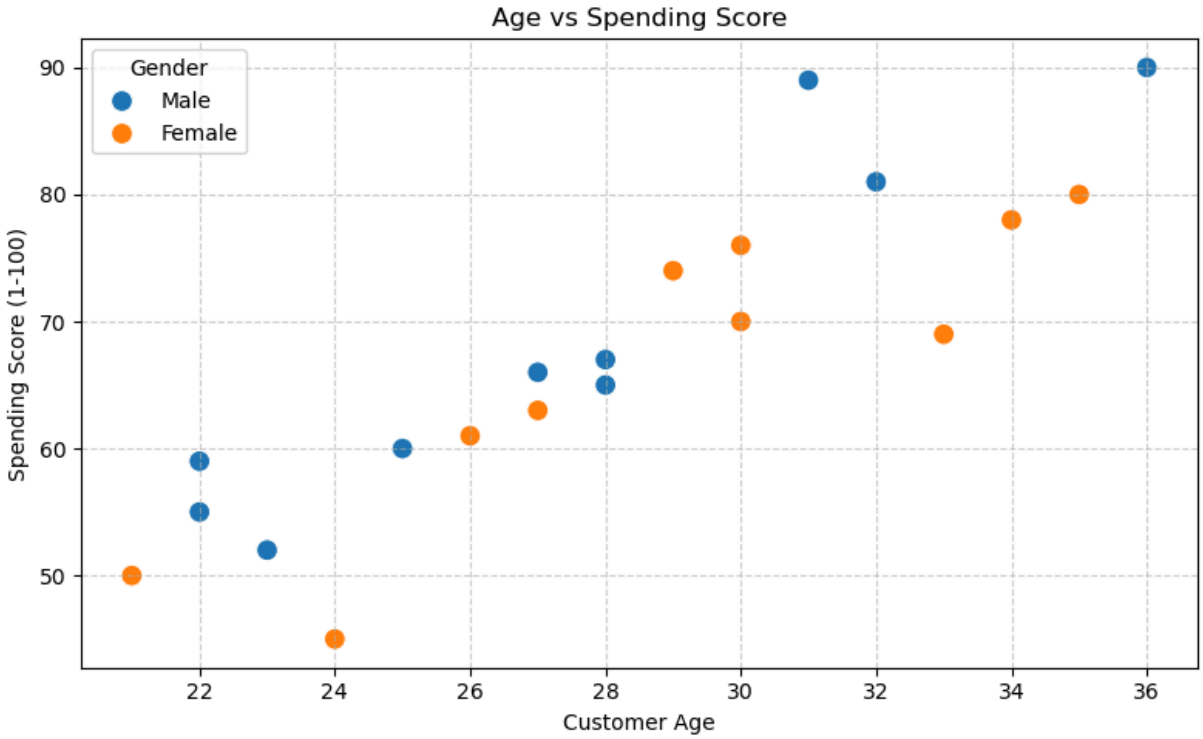
```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 7 columns):
 #   Column                  Non-Null Count  Dtype 
---  ------                  --------------  ----- 
 0   CustomerID              20 non-null     int64 
 1   Gender                  20 non-null     object
 2   Age                     20 non-null     int64 
 3   Annual Income (kâ,¹)    20 non-null     int64 
 4   Spending Score (1-100)  20 non-null     int64 
 5   Purchase Amount (â,¹)   20 non-null     int64 
 6   Product Category        20 non-null     object
dtypes: int64(5), object(2)
memory usage: 1.2+ KB
None

Missing Values:
CustomerID               0
Gender                   0
Age                      0
Annual Income (kâ,¹)      0
Spending Score (1-100)    0
Purchase Amount (â,¹)     0
Product Category         0
dtype: int64
Index(['CustomerID', 'Gender', 'Age', 'Annual Income (k₹)',
       'Spending Score (1-100)', 'Purchase Amount (₹)', 'Product Category'],
      dtype='object')
Gender
Female    24550.0
Male      23100.0
Name: Purchase Amount (₹), dtype: float64
```






```
Product Category
Electronics    190000
Clothing       171500
Grocery        115000
Name: Purchase Amount (₹), dtype: int64
```






```
[NbConvertApp] Converting notebook Customer_Purchase_Analysis_by_YourName.ipynb to h
tml
[NbConvertApp] WARNING | Alternative text is missing on 4 image(s).
[NbConvertApp] Writing 476802 bytes to Customer_Purchase_Analysis_by_YourName.html
```

In [ ]: