In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")
%matplotlib inline

print("Libraries imported successfully!")
```

```
Libraries imported successfully!
```

In [2]:
```
import os

os.chdir(r"C:\Users\somus\Downloads")

print("Current folder:", os.getcwd())
```

```
Current folder: C:\Users\somus\Downloads
```

In [3]:
```
import os
print(os.listdir())
```

```
['Adhaar.pdf', 'aes-encryption-decryption.png', 'Airline_Reservation_System_SQL_P
roject_With_Triggers.pdf', 'amount-1.png', 'Anaconda3-2024.02-1-Windows-x86_64.ex
e', 'android-studio-2024.1.1.11-windows (1).exe', 'androidparty.png', 'ATS Friend
ly Resume.pdf', 'AWS_Cloud_Intern_Interview_QA.pdf', 'AWS_Core_Services_QA.pdf',
'Bank_Management_System_SQL_Explained.pdf', 'bg.jpg', 'biryani.jpeg', 'blue bg.jp
g', 'boxes-1.png', 'burger.jpeg', 'canceled-orders-1.png', 'canvas-background.pn
g', 'ChromeSetup.exe', 'Cleaning_log.xlsx', 'CLG ID.pdf', 'Data Visualization and
Storytelling using Power BI.pptx', 'DATAFLOWDIAGRAM.png', 'DBMS LAB MANUAL.pdf',
'debug.log', 'desktop.ini', 'dice_images', 'dice_images.zip', 'DOT TECHNOLOGY.pd
f', 'E-commerce Dataset.csv', 'ecommerce_dataset_10000.csv', 'familyrestaurant.jp
eg', 'flowchart.png', 'forest.jpg', 'frenchfries.jpeg', 'friedrice.jpeg', 'Git-2.
51.1-64-bit.exe', 'HARSHI DOCUMENTATION - MBTS.doc', 'icecream.jpeg', 'IMG2025011
4190253.jpg', 'IMG20250114190317.jpg', 'IMG20250114190330.jpg', 'IMG_20250216_122
254.jpg', 'IMG_20250803_165204_921.webp', 'IMG_20250808_194447.jpg', 'indian.jpe
g', 'IPL2025Batters.csv', 'JAVA 1.7.rar', 'java all concepts.pdf', 'JAVA LAB Prog
rams-20240702.zip', 'javascript-interview-questions-and-answers-pdf-free.pdf', 'j
ava_solutions_detailed_inline_comments.pdf', 'jdk-8u144-windows-x64.exe', 'Junior
_Associate_Consultant_Interview_QA.pdf', 'kadiyam.Soma Sekhar Resume.pdf', 'lan
d', 'lecture-sql.ppt', 'LockDownBrowserOEMSetup (1).exe', 'LockDownBrowserOEMSetu
p.exe', 'LTIMINDTREE FORM.pdf', 'Mall_Customers_Cleaned.xlsx', 'Mall_Customers_Ra
wData.xlsx', 'manchurian.jpeg', 'master chef.png', 'mongodb-windows-x86_64-8.2.1-
signed.msi', 'mongosh-2.5.8-x64.msi', 'More_Interview_Questions_Junior_Associate_
Consultant.pdf', 'mysql-5.1.44-win32.msi', 'mysql-connector-odbc-8.0.43-win32.ms
i', 'mysql-essential-5.0.67-win32.msi', 'mysql-installer-community-8.0.43.0.msi',
'Mysql_for_python.msi', 'netbeans-8.1-windows.exe', 'noodles.jpeg', 'onefor.jpg',
'OOSE SAMPLE PROJECT1 (2) (1).docx', 'OOSE SAMPLE PROJECT1 (2).docx', 'Ordinary P
erson Ringtone - MobCup.Com.Co.mp3', 'PBIDesktopSetup_x64.exe', 'PD_2385351040
(3).pdf', 'PD_2385351040.doc', 'photo.jpg', 'pizza.jpeg', 'ppt unit-1.pptx', 'ppt
unit-2.pptx', 'ppt unit-3.pptx', 'ppt Unit-4.pptx', 'PRIVACY-PRESERVING_MULTI-KEY
WORD_RANKED_SEARCH_OVER_ENCRYPTED_CLOUD_DATA[1] (2).pdf', 'PRIVACY-PRESERVING_MUL
TI-KEYWORD_RANKED_SEARCH_OVER_ENCRYPTED_CLOUD_DATA[1].pdf', 'profit-pct-1.png',
'profits-1.png', 'python-3.12.4-amd64.exe', 'README.md', 'Record 1.jpg', 'Record
2.jpg', 'Record 3.jpg', 'Record 4.jpg', 'Record_2025-01-15-17-34-58.mp4', 'sabtoo
l_win8_10_11_v170530.zip', 'sai (1).docx', 'sai.docx', 'samosa.jpeg', 'Sample Res
ume new.docx', 'Sample Resume.pdf', 'sample-chocolate-data.xlsx', 'Screenshot_202
5-05-03-21-29-06-70.jpg', 'Screenshot_2025-08-20-10-04-14-40.jpg', 'Screenshot_20
25-08-20-10-44-53-94.jpg', 'Screenshot_2025-10-02-12-58-28-50.jpg', 'shipments-1.
png', 'Shopping-web-Jsp-Servlet-master.zip', 'SMS Encryption System.pdf', 'SMS EN
CRYPTION SYSTEM.rar', 'SOMA SEKHAR KADIYAM RESUME.docx', 'SOMA-SEKHAR-KADIYAM-Res
ume.docx', 'Soma_Sekhar_Kadiyam_Resume.docx', 'stored procedure SQL.pdf', 'SuperS
tore Sales DataSet.xlsx', 'takewayrestaurant.jpeg', 'Top 115 Java Interview Quest
ions & Answers.pdf', 'Top_30_Non_IT_Interview_QA_Freshers.docx', 'Top_30_Non_IT_I
nterview_QA_Freshers.pdf', 'TS Recommended Apps', 'TypingMaster11India.exe', 'Ult
raViewer_setup_6.6.113_en.exe', 'unit-2 ch-1 notes.pdf', 'unit-2 ch-3 notes.pdf',
'unit-3 ch-1 notes.pdf', 'VC_redist.x64.exe', 'veg meals.jpeg', 'VSCodeUserSetup-
x64-1.104.3.exe', 'WavesAudio.MaxxAudioProforDell2022_fh4rh281wavaa!App', 'Webyog
_SQLyog_6.56_Enterprise', 'Webyog_SQLyog_6.56_Enterprise.rar', 'WhatsApp Image 20
24-06-27 at 15.23.05_6d2a1a28.jpg', '~$reenshot_2025-09-30-12-23-57-77.jpg', '~$_
2385351041.doc', '~WRD1556.tmp', '~WRL0004.tmp']
```

In [4]:
```
df=pd.read_csv("E-commerce Dataset.csv")

df.head()
```

Out[4]:

	customer_id	first_name	last_name	gender	country	product_id	product_name	
0	CUST2353	Erica	Oliver	Female	Canada	PROD108	Fitbit Versa 3	El
1	CUST4463	Christopher	White	Male	China	PROD103	Levi's Jeans	
2	CUST4512	Spencer	Foster	Male	Germany	PROD111	Lego Star Wars Set	
3	CUST5711	Jessica	Harris	Male	France	PROD107	Dyson Vacuum	
4	CUST1296	Amy	Johnson	Female	Brazil	PROD105	Adidas Running Shoes	

In [5]:
```
# Basic structure and summary
df.info()
df.describe(include='number').T
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 15 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   customer_id     119 non-null    object
 1   first_name      119 non-null    object
 2   last_name       119 non-null    object
 3   gender          119 non-null    object
 4   country         119 non-null    object
 5   product_id      119 non-null    object
 6   product_name    119 non-null    object
 7   category        119 non-null    object
 8   quantity        119 non-null    int64 
 9   unit_price      119 non-null    int64 
 10  order_id        119 non-null    object
 11  order_status    119 non-null    object
 12  payment_method  119 non-null    object
 13  rating          119 non-null    int64 
 14  review_text     119 non-null    object
dtypes: int64(3), object(12)
memory usage: 14.1+ KB
```

Out[5]:

	count	mean	std	min	25%	50%	75%	max
quantity	119.0	2.815126	1.377583	1.0	2.0	3.0	4.0	5.0
unit_price	119.0	251.084034	277.300259	25.0	99.0	149.0	229.0	999.0
rating	119.0	2.773109	1.446427	1.0	1.5	3.0	4.0	5.0

In [6]:
```
print("Gender counts:")
print(df['gender'].value_counts(), "\n")
```

```
Gender counts:
gender
Female    51
Other     36
Male      32
Name: count, dtype: int64
```

In [7]:
```
print("Country counts:")
print(df['country'].value_counts().head(10), "\n")

print("Product categories:")
print(df['category'].value_counts(), "\n")

print("Payment methods:")
print(df['payment_method'].value_counts(), "\n")
```

```
Country counts:
country
China        17
Germany      16
Japan        15
France       14
Brazil       12
Canada       11
India        11
UK            9
USA           9
Australia     5
Name: count, dtype: int64

Product categories:
category
Electronics       36
Apparel           29
Toys              14
Sports            14
Books             14
Home & Kitchen    12
Name: count, dtype: int64

Payment methods:
payment_method
Credit Card         49
Cash on Delivery    40
PayPal              30
Name: count, dtype: int64
```

In [8]:
```
df['total_amount'] = df['quantity'] * df['unit_price']
df.head()
```

Out[8]:

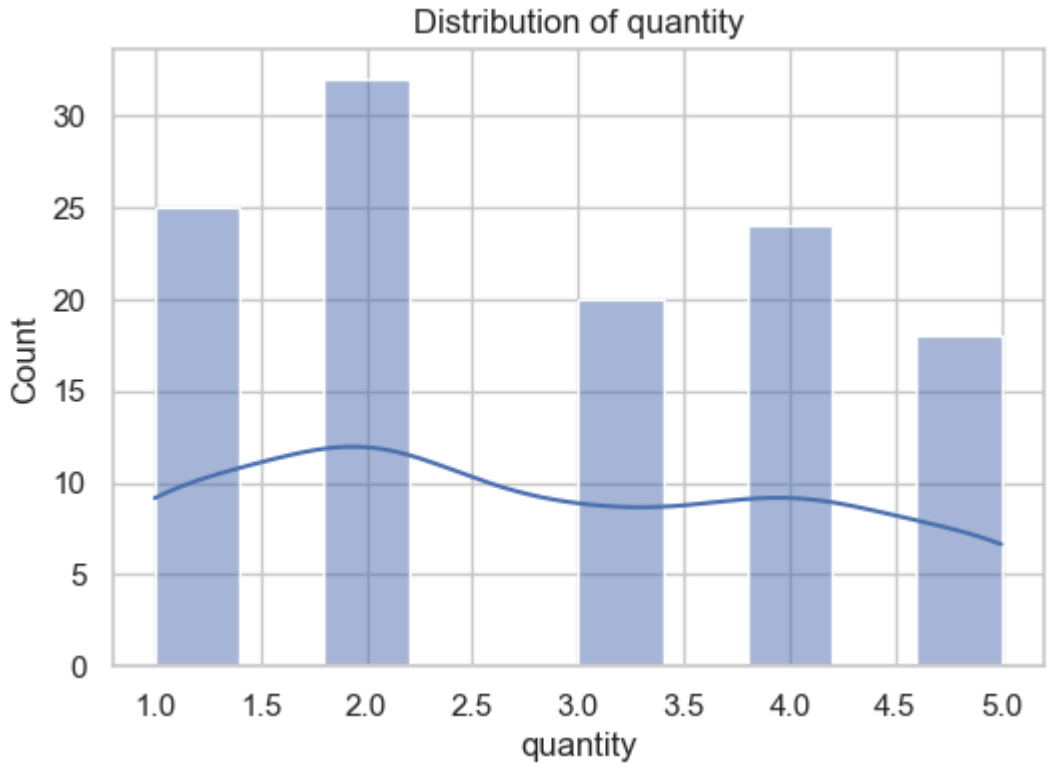
	customer_id	first_name	last_name	gender	country	product_id	product_name	
0	CUST2353	Erica	Oliver	Female	Canada	PROD108	Fitbit Versa 3	El
1	CUST4463	Christopher	White	Male	China	PROD103	Levi's Jeans	
2	CUST4512	Spencer	Foster	Male	Germany	PROD111	Lego Star Wars Set	
3	CUST5711	Jessica	Harris	Male	France	PROD107	Dyson Vacuum	
4	CUST1296	Amy	Johnson	Female	Brazil	PROD105	Adidas Running Shoes	

In [10]:
```
print("Missing values per column:")
print(df.isnull().sum())

print("\nDuplicate rows:", df.duplicated().sum())
```

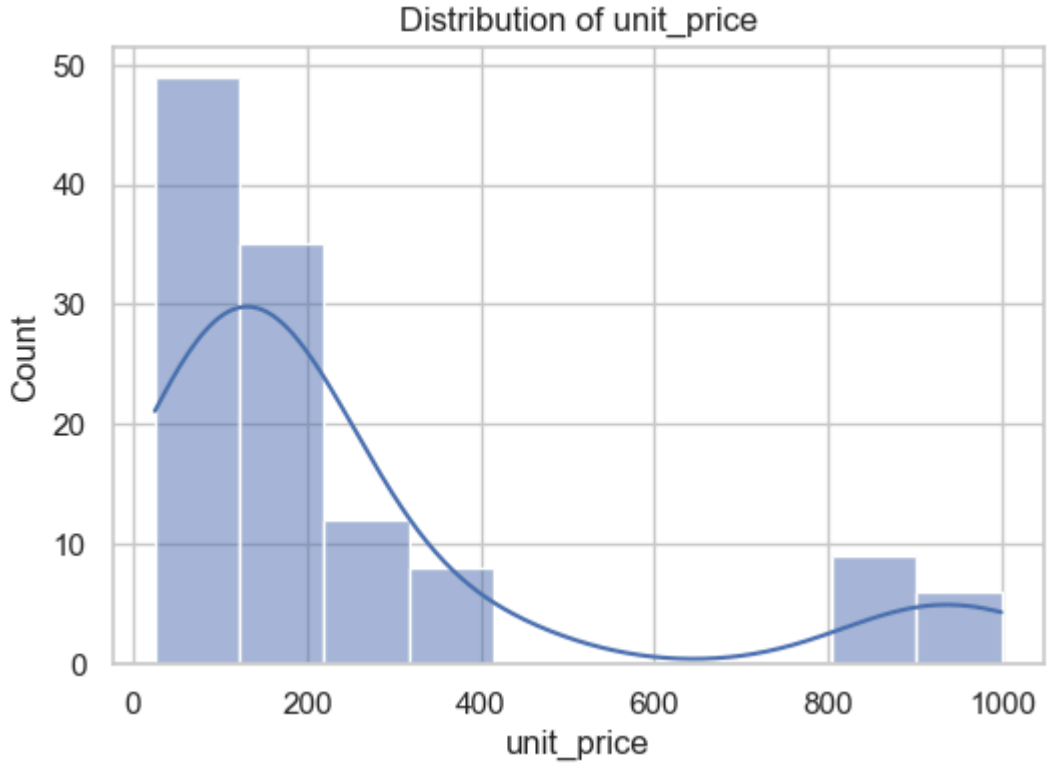
```
Missing values per column:
customer_id       0
first_name        0
last_name         0
gender            0
country           0
product_id        0
product_name      0
category          0
quantity          0
unit_price        0
order_id          0
order_status      0
payment_method    0
rating            0
review_text       0
total_amount      0
dtype: int64

Duplicate rows: 0
```

In [13]:
```
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")
num_cols=['quantity','unit_price','rating','total_amount']

for col in num_cols:
    plt.figure(figsize=(6,4))
    sns.histplot(df[col], bins=10, kde=True)
    plt.title(f"Distribution of {col}")
    plt.show()
```

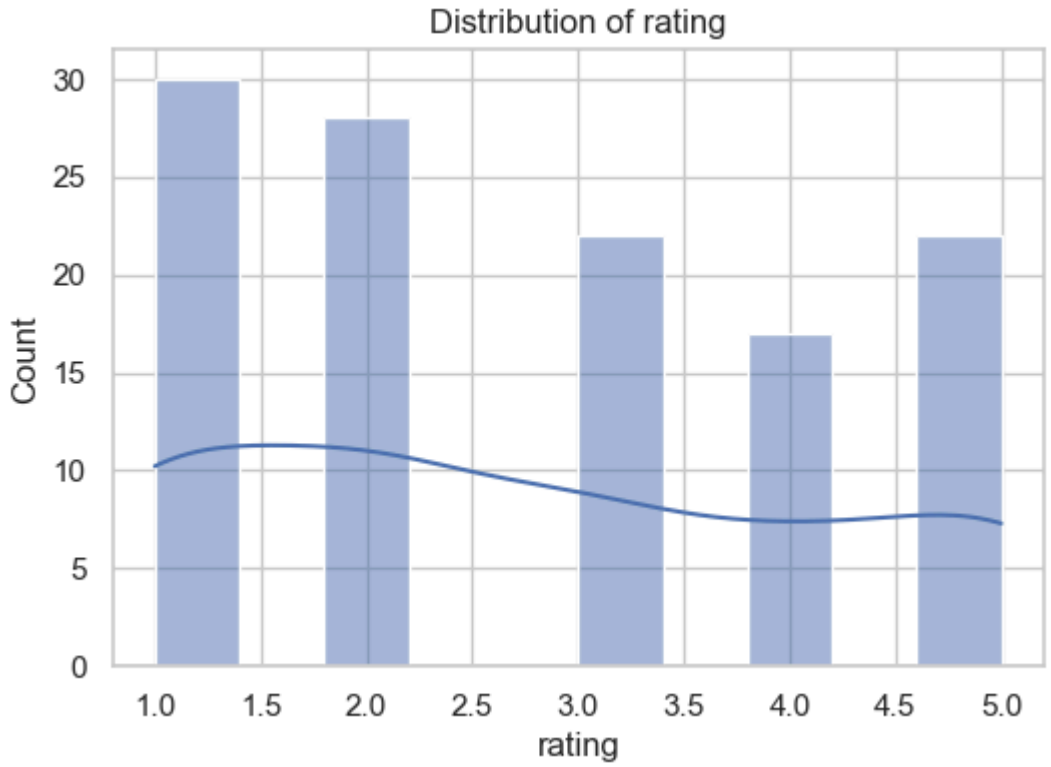
```
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



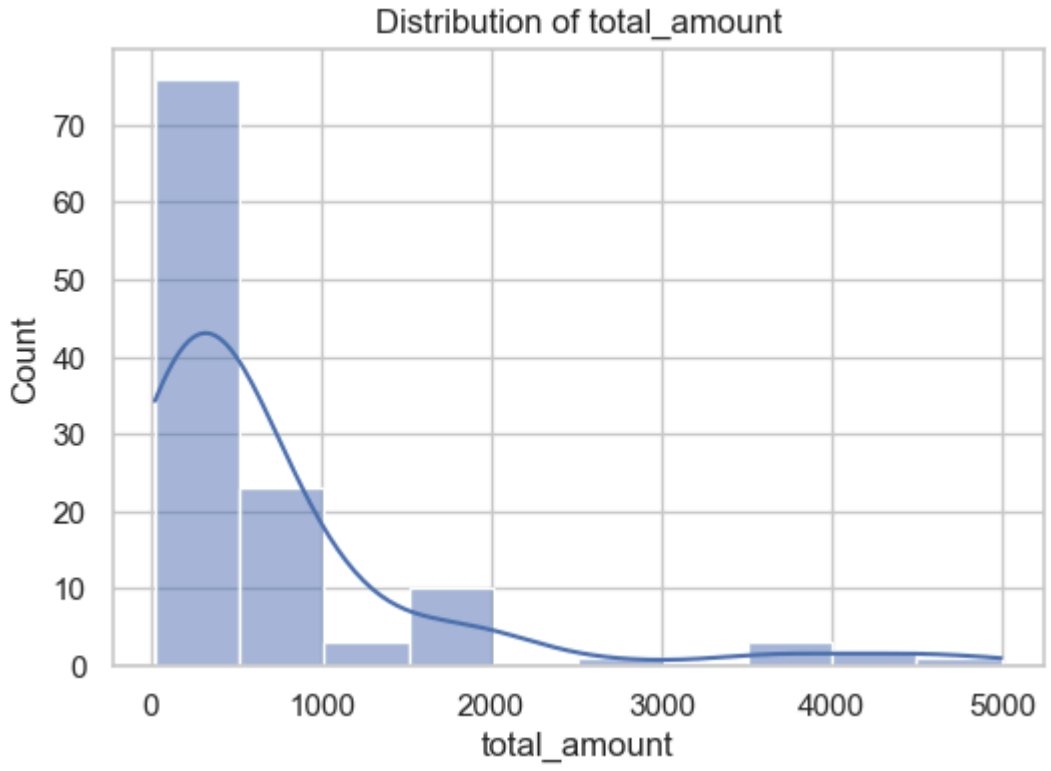

```
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



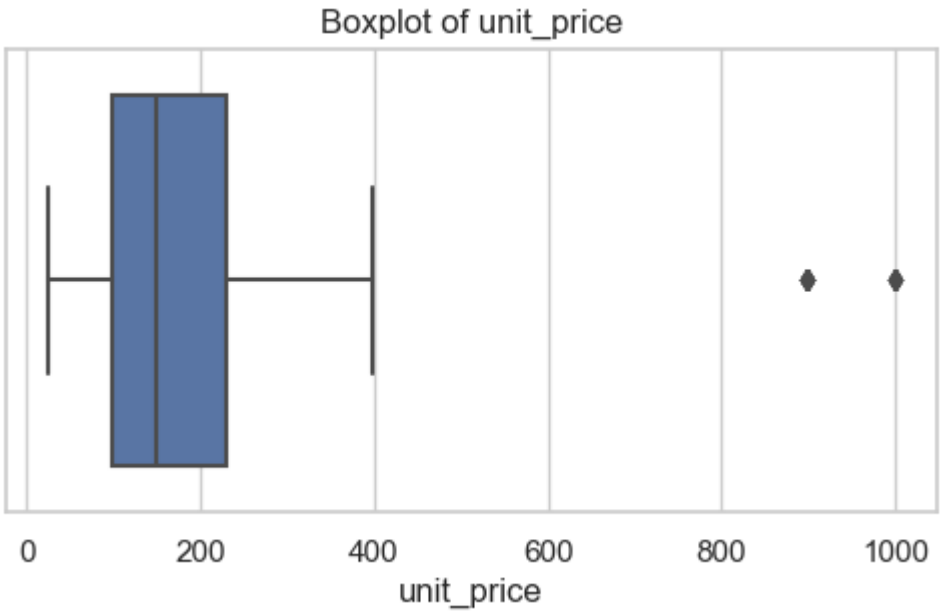

```
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

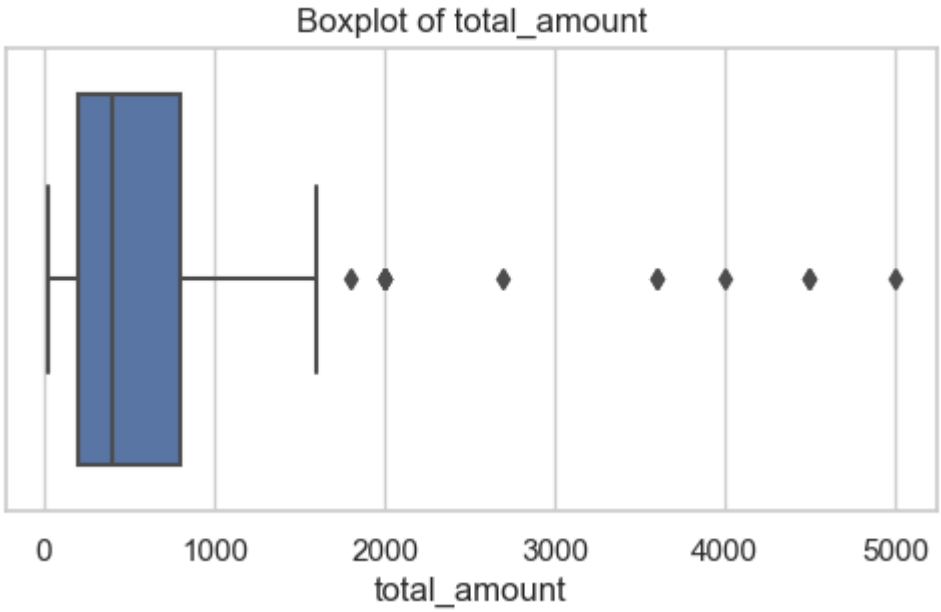



```
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

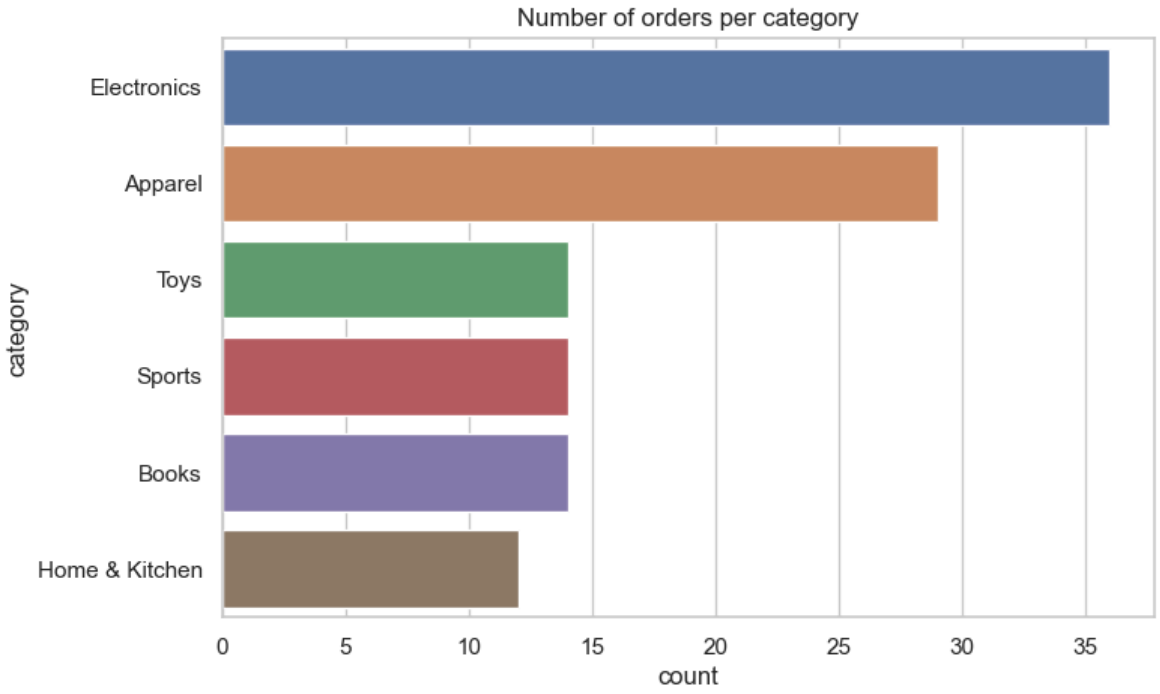


In [14]:
```
for col in['unit_price','total_amount']:
    plt.figure(figsize=(6,3))
    sns.boxplot(x=df[col])
    plt.title(f"Boxplot of {col}")
    plt.show()
```

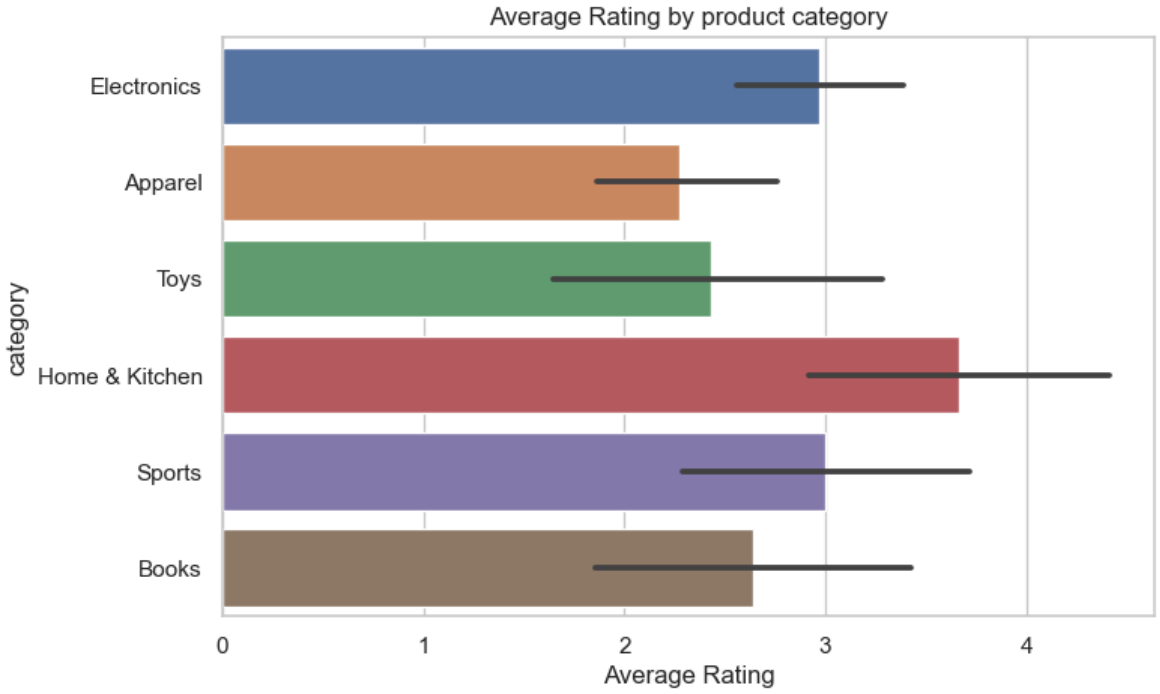




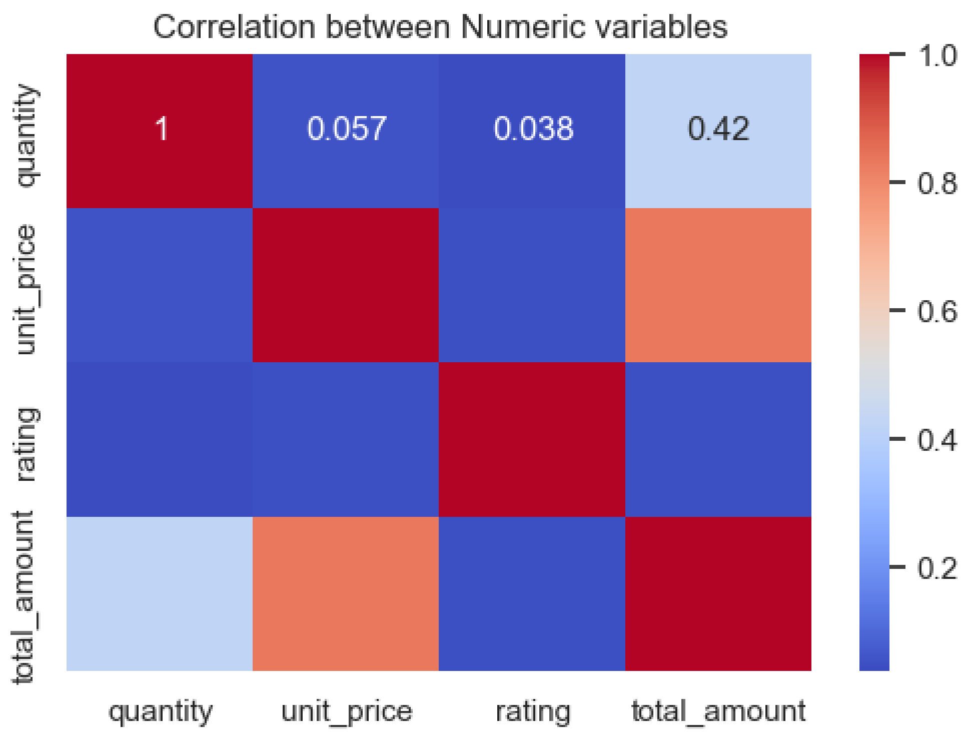
In [15]:
```
plt.figure(figsize=(8,5))
sns.countplot(y='category', data=df, order=df['category'].value_counts().index)
plt.title("Number of orders per category")
plt.xlabel("count")
plt.ylabel("category")
plt.show()
```



In [16]:
```
plt.figure(figsize=(8,5))
sns.barplot(x='rating',y='category', data=df, estimator='mean')
plt.title("Average Rating by product category")
plt.xlabel("Average Rating")
plt.ylabel("category")
plt.show()
```

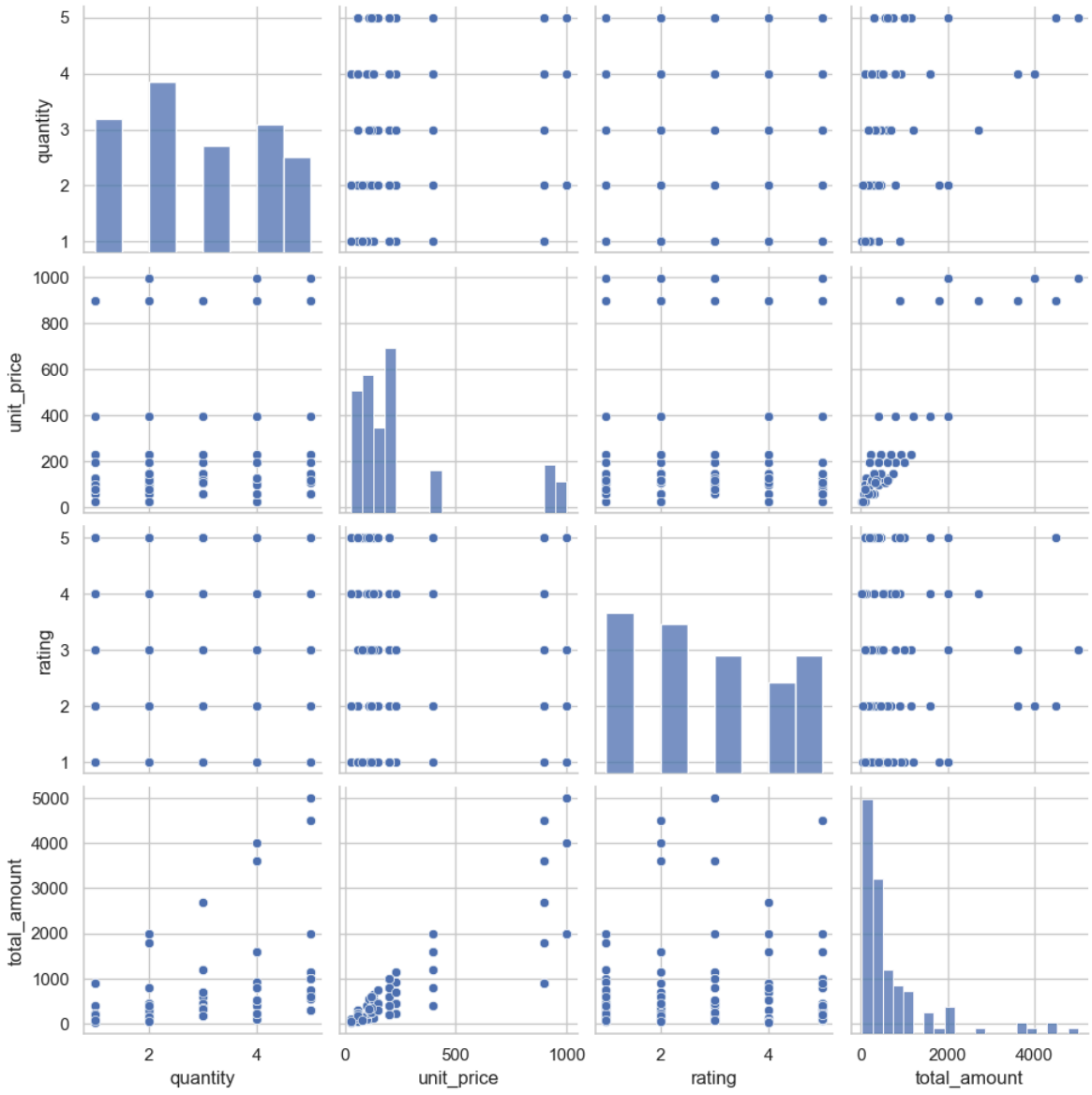


In [17]:
```
plt.figure(figsize=(6,4))
sns.heatmap(df[['quantity','unit_price','rating','total_amount']].corr(), annot=

plt.title("Correlation between Numeric variables")
plt.show()
```



In [18]:
```
sns.pairplot(df[['quantity','unit_price','rating','total_amount']])
plt.show()
```

```
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\somus\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



In [ ]:
```
## Summary of Findings
- Dataset contains 119 transactions and 15 features.
- No missing or duplicate data.
- Electronics and Apparel dominates sales.
- Cash on Delivery is the most common payment method.
- Ratings mostly range 2-5 with average around 3.
- Higher quantity and unit_price lead to higher total_amount.
- A few outliers exist in total_amount (bulk orders).
```