In [152…

```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("retail_sales_data.csv")
```

In [153…

```python
df
```

Out[153…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode
0	1001	2023-01-01	45.0	Clothing	5	3696	NaN
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI
2	1003	2023-01-03	50.0	Furniture	3	1703	NaN
3	1004	2023-01-04	35.0	Electronics	4	166	NaN
4	1005	2023-01-05	45.0	Electronics	1	332	UPI
...	...	...	...	...	...	...	...
305	1006	2023-01-06	25.0	Grocery	4	2256	Card
306	1007	2023-01-07	45.0	Clothing	3	1269	NaN
307	1008	2023-01-08	50.0	Furniture	5	2258	NaN
308	1009	2023-01-09	35.0	Grocery	4	4306	NaN
309	1010	2023-01-10	30.0	Electronics	5	1593	Cash

310 rows × 7 columns

In [154…

```python
# Basic Data Understanding.
```

In [155…

```python
df.head(5)
```

Out[155…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode
0	1001	2023-01-01	45.0	Clothing	5	3696	NaN
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI
2	1003	2023-01-03	50.0	Furniture	3	1703	NaN
3	1004	2023-01-04	35.0	Electronics	4	166	NaN
4	1005	2023-01-05	45.0	Electronics	1	332	UPI

In [156…

```python
df.tail(5)
```

Out[156…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode
305	1006	2023-01-06	25.0	Grocery	4	2256	Card
306	1007	2023-01-07	45.0	Clothing	3	1269	NaN
307	1008	2023-01-08	50.0	Furniture	5	2258	NaN
308	1009	2023-01-09	35.0	Grocery	4	4306	NaN
309	1010	2023-01-10	30.0	Electronics	5	1593	Cash

In [157…

```python
df.shape
```

Out[157…

```
(310, 7)
```

In [158…

```python
df.describe()
```

Out[158…

	Order_ID	Customer_Age	Quantity	Price
count	310.000000	278.000000	310.000000	310.000000
mean	1145.822581	33.053957	3.058065	2550.277419
std	89.107862	10.985417	1.464747	1326.710790
min	1001.000000	18.000000	1.000000	103.000000
25%	1068.250000	25.000000	2.000000	1565.250000
50%	1145.500000	30.000000	3.000000	2560.500000
75%	1222.750000	45.000000	4.000000	3586.500000
max	1300.000000	50.000000	5.000000	4995.000000

In [159…

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 310 entries, 0 to 309
Data columns (total 7 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Order_ID          310 non-null    int64
 1   Order_Date        310 non-null    object
 2   Customer_Age      278 non-null    float64
 3   Product_Category  310 non-null    object
 4   Quantity          310 non-null    int64
 5   Price             310 non-null    int64
 6   Payment_Mode      240 non-null    object
dtypes: float64(1), int64(3), object(3)
memory usage: 17.1+ KB
```

In [160…

```python
df.dtypes
```

Out[160…

```
Order_ID              int64
Order_Date           object
Customer_Age        float64
Product_Category     object
Quantity              int64
Price                 int64
Payment_Mode         object
dtype: object
```

In [161…

```python
df.columns
```

Out[161…

```
Index(['Order_ID', 'Order_Date', 'Customer_Age', 'Product_Category',
       'Quantity', 'Price', 'Payment_Mode'],
      dtype='object')
```

In [162…

```python
df['Order_Date'] = pd.to_datetime(df['Order_Date'], errors='coerce')
df['Order_Date'].dtype
```

Out[162…

```
dtype('<M8[ns]')
```

In [163…

```python
# Check null value .
```

In [164…

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

Out[164…

```
Order_ID             0
Order_Date           0
Customer_Age        32
Product_Category     0
Quantity             0
Price                0
Payment_Mode        70
dtype: int64
```

In [165…

```python
# Fill null value.
```

In [166…

```python
df['Customer_Age'] = df['Customer_Age'].fillna(df['Customer_Age'].median())
```

In [167…

```python
df['Payment_Mode'] = df['Payment_Mode'].fillna('unknown')
```

In [168…

```python
df['Payment_Mode']
```

Out[168…

```
0      unknown
1          UPI
2      unknown
3      unknown
4          UPI
        ...
305       Card
306    unknown
307    unknown
308    unknown
309       Cash
Name: Payment_Mode, Length: 310, dtype: object
```

In [169…

```python
# Check null value.
df.isnull().sum()
```

Out[169…

```
Order_ID            0
Order_Date          0
Customer_Age        0
Product_Category    0
Quantity            0
Price               0
Payment_Mode        0
dtype: int64
```

In [170…

```python
# Check and remove duplicate.
```

In [171…

```python
df.duplicated().sum()
```

Out[171…

```
10
```

In [172…

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

In [173…

```python
df.duplicated()
```

Out[173…

```
0      False
1      False
2      False
3      False
4      False
       ...
295    False
296    False
297    False
298    False
299    False
Length: 300, dtype: bool
```

In [174…

```python
# Create new column.
```

In [175…

```python
df = df.copy()
df['Total_Sales'] = df['Quantity'] * df['Price']
```

In [176…

```python
# Sort Data.
```

In [177…

```python
df.sort_values('Total_Sales',ascending=False)
```

Out[177…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales
120	1121	2023-05-01	22.0	Electronics	5	4811	UPI	24055
57	1058	2023-02-27	22.0	Clothing	5	4727	UPI	23635
155	1156	2023-06-05	18.0	Electronics	5	4712	Card	23560
130	1131	2023-05-11	30.0	Clothing	5	4698	Cash	23490
267	1268	2023-09-25	18.0	Electronics	5	4624	UPI	23120
...	...	...	...	...	...	...	...	...
110	1111	2023-04-21	18.0	Furniture	2	177	Cash	354
92	1093	2023-04-03	45.0	Clothing	1	335	Cash	335
4	1005	2023-01-05	45.0	Electronics	1	332	UPI	332
138	1139	2023-05-19	30.0	Furniture	1	297	UPI	297
72	1073	2023-03-14	30.0	Furniture	1	109	Cash	109

300 rows × 8 columns

In [178…

```python
# Filtering Data
```

In [179…

```python
df[df['Total_Sales']>10000]
```

Out[179…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales
0	1001	2023-01-01	45.0	Clothing	5	3696	unknown	18480
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI	16915
7	1008	2023-01-08	50.0	Furniture	5	2258	unknown	11290
8	1009	2023-01-09	35.0	Grocery	4	4306	unknown	17224
11	1012	2023-01-12	50.0	Grocery	5	4107	Cash	20535
...	...	...	...	...	...	...	...	...
284	1285	2023-10-12	30.0	Furniture	3	3551	unknown	10653
285	1286	2023-10-13	35.0	Furniture	4	4164	Card	16656
292	1293	2023-10-20	25.0	Furniture	5	2906	Card	14530
293	1294	2023-10-21	50.0	Clothing	4	2590	unknown	10360
298	1299	2023-10-26	22.0	Clothing	4	3154	Card	12616

88 rows × 8 columns

In [180…

```python
# Total Sales Sum
```

In [181…

```python
df['Total_Sales'].sum()
```

Out[181…

```
2351325
```

In [182…

```python
# Product category wise total sales.
```

In [183…

```python
df.groupby('Product_Category')['Total_Sales'].sum()
```

Out[183…

```
Product_Category
Clothing       568298
Electronics    673040
Furniture      602477
Grocery        507510
Name: Total_Sales, dtype: int64
```

In [184…

```python
# Which category most revnue.
```

In [185…

```python
revenue = df.groupby('Product_Category')['Total_Sales'].sum()
revenue.sort_values(ascending=False).head(1)
```

Out[185…

```
Product_Category
Electronics    673040
Name: Total_Sales, dtype: int64
```

In [186…

```python
# Payment mode wise sales and max payment.
```

In [187…

```python
df.head(5)
```

Out[187…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales
0	1001	2023-01-01	45.0	Clothing	5	3696	unknown	18480
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI	16915
2	1003	2023-01-03	50.0	Furniture	3	1703	unknown	5109
3	1004	2023-01-04	35.0	Electronics	4	166	unknown	664
4	1005	2023-01-05	45.0	Electronics	1	332	UPI	332

In [188…

```python
payment_mode = df.groupby('Payment_Mode')['Total_Sales'].sum()
```

In [189…

```python
payment_mode.sort_values(ascending=False).head()
```

Out[189…

```
Payment_Mode
UPI        670542
Cash       581896
Card       554733
unknown    544154
Name: Total_Sales, dtype: int64
```

In [190…

```python
# Top 5 Sales Order.
```

In [191…

```python
df.nlargest(5,'Total_Sales')
```

Out[191…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales
120	1121	2023-05-01	22.0	Electronics	5	4811	UPI	24055
57	1058	2023-02-27	22.0	Clothing	5	4727	UPI	23635
155	1156	2023-06-05	18.0	Electronics	5	4712	Card	23560
130	1131	2023-05-11	30.0	Clothing	5	4698	Cash	23490
267	1268	2023-09-25	18.0	Electronics	5	4624	UPI	23120

In [192…

```python
# Monthly sales find.
```

In [193…

```python
df['Month'] = df['Order_Date'].dt.month
df.groupby('Month')['Total_Sales'].sum()
```

Out[193…

```
Month
1     267173
2     199164
3     257901
4     194612
5     260783
6     258305
7     220254
8     265947
9     201128
10    226058
Name: Total_Sales, dtype: int64
```

In [194…

```python
 # Category wise sale.
```

In [195…

```python
df.head(5)
```

Out[195…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales	Month
0	1001	2023-01-01	45.0	Clothing	5	3696	unknown	18480	1
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI	16915	1
2	1003	2023-01-03	50.0	Furniture	3	1703	unknown	5109	1
3	1004	2023-01-04	35.0	Electronics	4	166	unknown	664	1
4	1005	2023-01-05	45.0	Electronics	1	332	UPI	332	1

In [196…

```python
# Category Wise Sale.
```

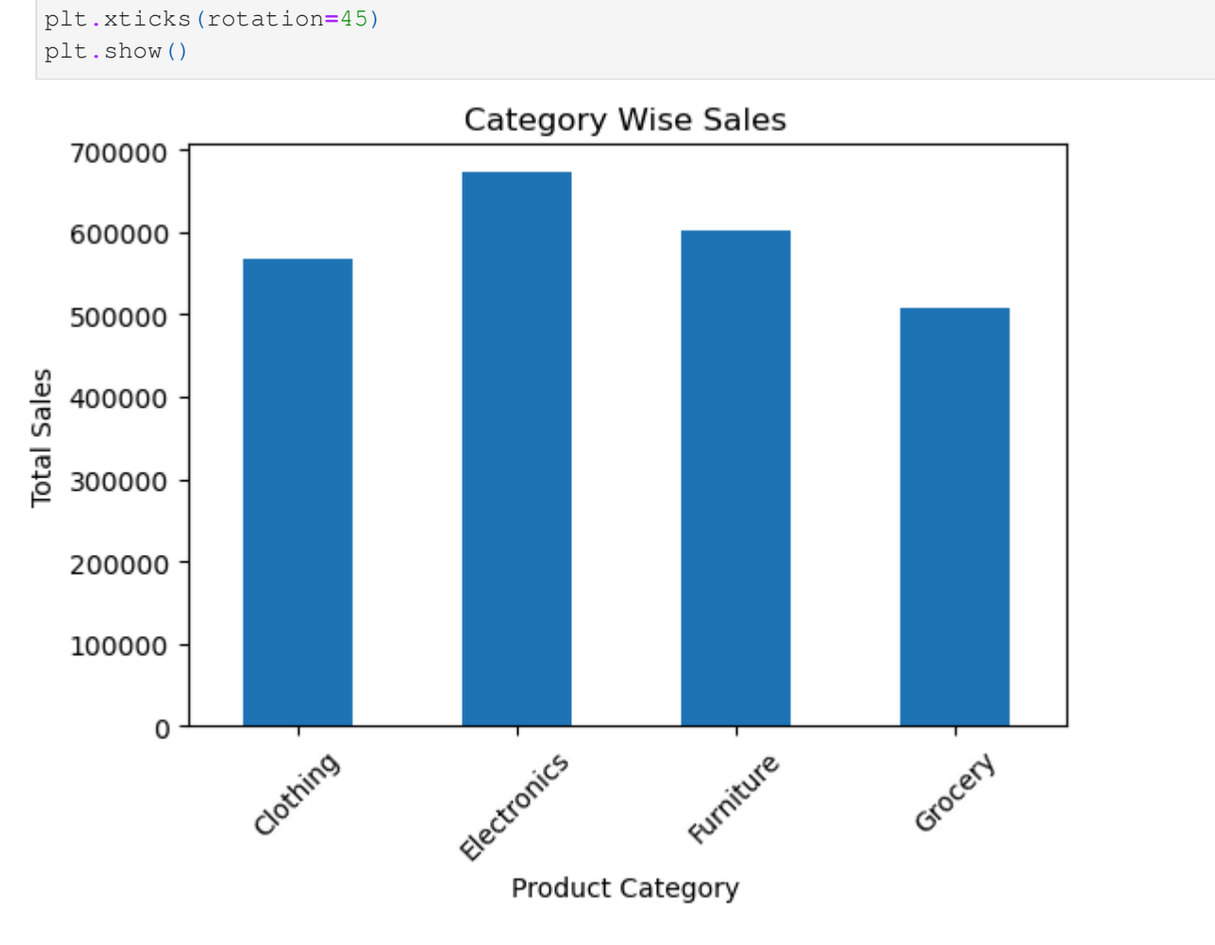
In [243…

```python
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))    # 👈 sabse pehle

df.groupby('Product_Category')['Total_Sales'].sum().plot(kind='bar')

plt.title('Category Wise Sales')
plt.xlabel('Product Category')
plt.ylabel('Total Sales')
plt.xticks(rotation=45)
plt.show()
```



In [245…

```python
# Daily sales trends
```

In [247…

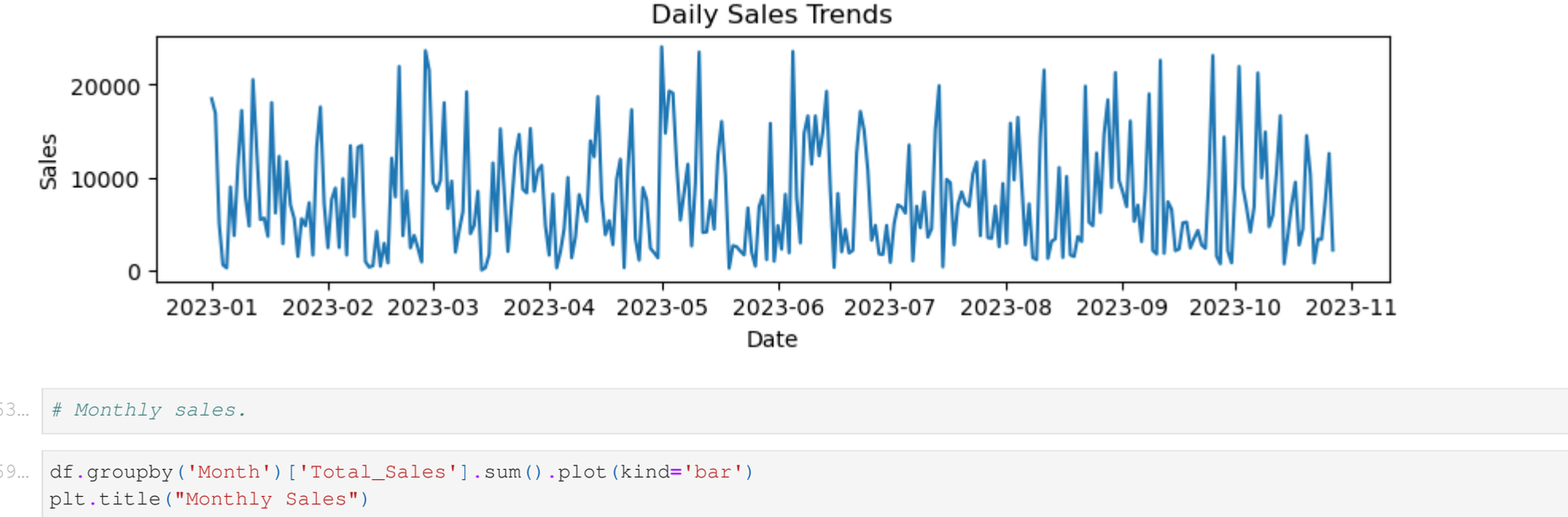
```python
df.head(5)
```

Out[247…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales	Month
0	1001	2023-01-01	45.0	Clothing	5	3696	unknown	18480	1
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI	16915	1
2	1003	2023-01-03	50.0	Furniture	3	1703	unknown	5109	1
3	1004	2023-01-04	35.0	Electronics	4	166	unknown	664	1
4	1005	2023-01-05	45.0	Electronics	1	332	UPI	332	1

In [261…

```python
daily_sales = df.groupby('Order_Date')['Total_Sales'].sum()
plt.figure(figsize=(10,2))
plt.plot(daily_sales)
plt.title('Daily Sales Trends')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.show()
```

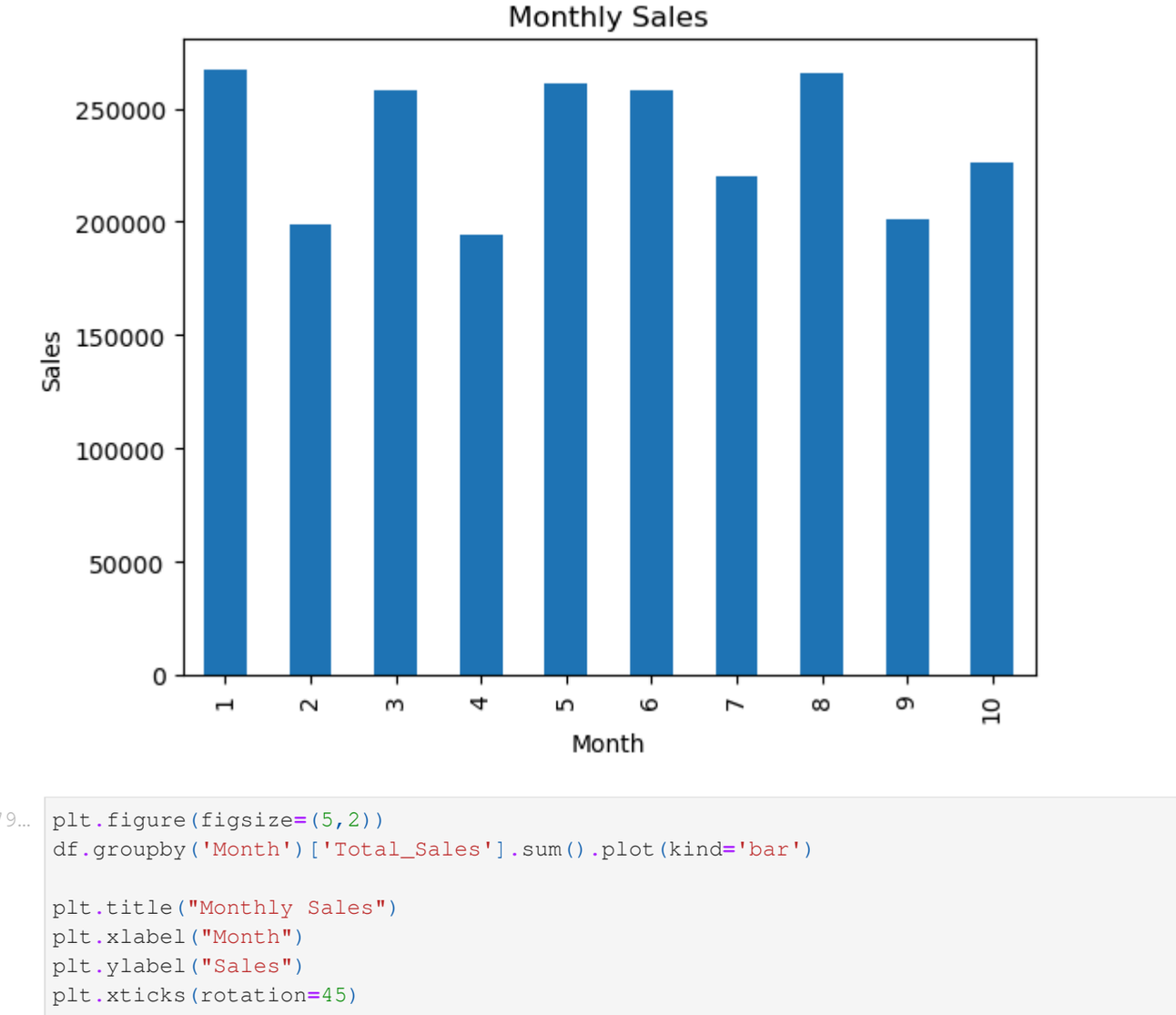


In [263…

```python
# Monthly sales.
```

In [269…

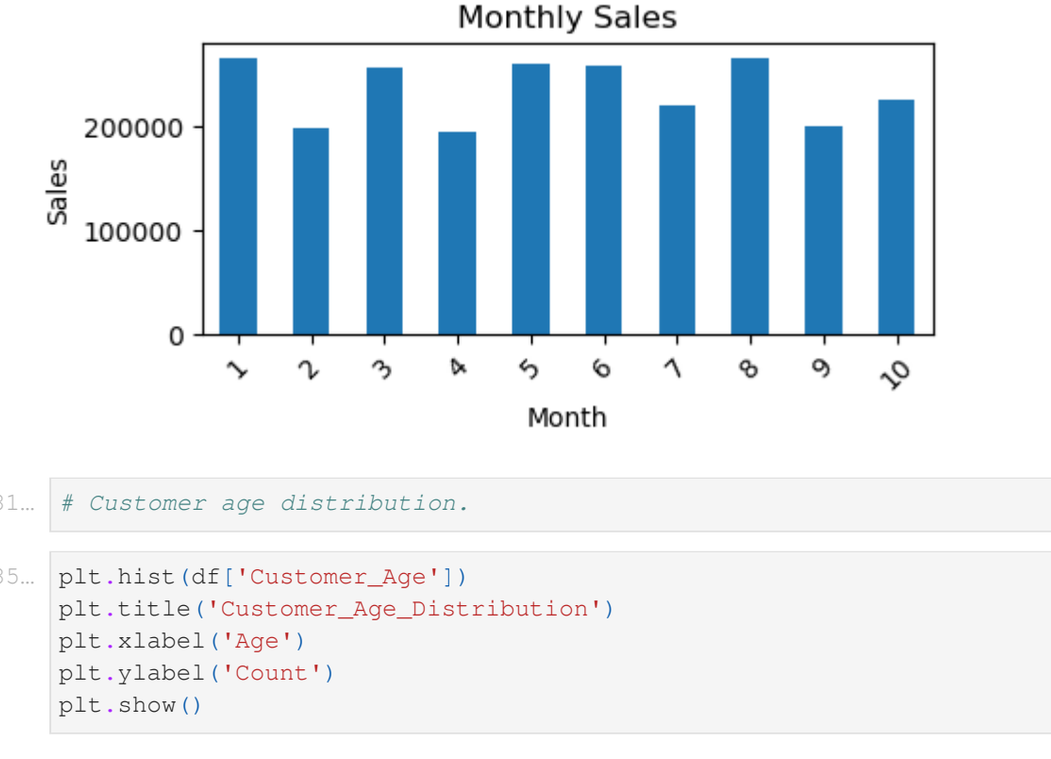
```python
df.groupby('Month')['Total_Sales'].sum().plot(kind='bar')
plt.title("Monthly Sales")
plt.xlabel("Month")
plt.ylabel("Sales")
plt.show()
```



In [279…

```python
plt.figure(figsize=(5,2))
df.groupby('Month')['Total_Sales'].sum().plot(kind='bar')

plt.title("Monthly Sales")
plt.xlabel("Month")
plt.ylabel("Sales")
plt.xticks(rotation=45)
plt.show()
```



In [281…

```python
# Customer age distribution.
```

In [285…

```python
plt.hist(df['Customer_Age'])
plt.title('Customer_Age_Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

In [287…

```python
# Payment mode distribution pie chart.
```

In [291…

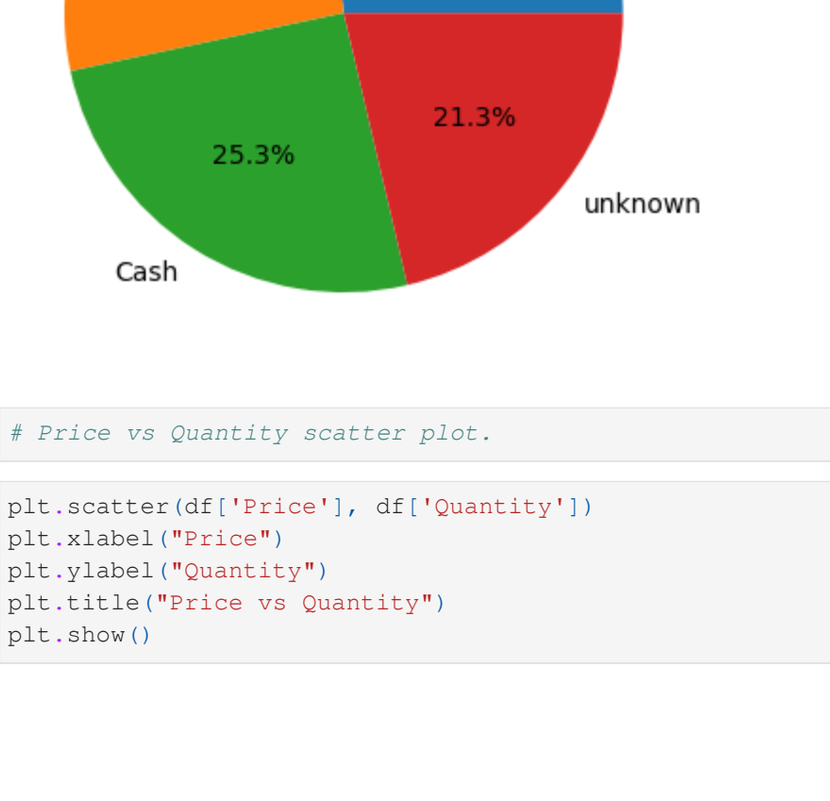
```python
df.head(5)
```

Out[291…

	Order_ID	Order_Date	Customer_Age	Product_Category	Quantity	Price	Payment_Mode	Total_Sales	Month
0	1001	2023-01-01	45.0	Clothing	5	3696	unknown	18480	1
1	1002	2023-01-02	30.0	Electronics	5	3383	UPI	16915	1
2	1003	2023-01-03	50.0	Furniture	3	1703	unknown	5109	1
3	1004	2023-01-04	35.0	Electronics	4	166	unknown	664	1
4	1005	2023-01-05	45.0	Electronics	1	332	UPI	332	1

In [297…

```python
df['Payment_Mode'].value_counts().plot(kind='pie',autopct='%1.1f%%')
plt.title('Payment Mode Distribution')
plt.ylabel("")
plt.show()
```



In [299…

```python
# Price vs Quantity scatter plot.
```

In [301…

```python
plt.scatter(df['Price'], df['Quantity'])
plt.xlabel("Price")
plt.ylabel("Quantity")
plt.title("Price vs Quantity")
plt.show()
```


Price vs Quantity
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
Quantity
0
1000
2000
3000
4000
5000
Price