

rjqg48lfj

January 23, 2025

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[3]: df = pd.read_csv("C:/Users/91703/OneDrive/Desktop/E-COMMERCE PROJECT.csv")
```

```
[6]: df.head()
```

```
[6]:
```

	User_ID	Product_ID	Category	Price (Rs.)	Discount (%)	\
0	337c166f	f414122f-e	Sports	36.53	15	
1	d38a19bf	fde50f9c-5	Clothing	232.79	20	
2	d7f5f0b0	0d96fc90-3	Sports	317.02	25	
3	395d4994	964fc44b-d	Toys	173.19	25	
4	a83c145c	d70e2fc6-e	Beauty	244.80	20	

	Final_Price(Rs.)	Payment_Method	Purchase_Date
0	31.05	Net Banking	12-11-2024
1	186.23	Net Banking	09-02-2024
2	237.76	Credit Card	01-09-2024
3	129.89	UPI	01-04-2024
4	195.84	Net Banking	27-09-2024

```
[16]: df.columns
df.dtypes
```

```
[16]: User_ID          object
Product_ID         object
Category           object
Price (Rs.)        float64
Discount (%)        int64
Final_Price(Rs.)    float64
Payment_Method      object
Purchase_Date       object
dtype: object
```

```
[18]: df.isnull().sum()
```

```
[18]: User_ID          0
      Product_ID      0
      Category        0
      Price (Rs.)      0
      Discount (%)     0
      Final_Price(Rs.) 0
      Payment_Method   0
      Purchase_Date    0
      dtype: int64
```

```
[19]: df.describe()
```

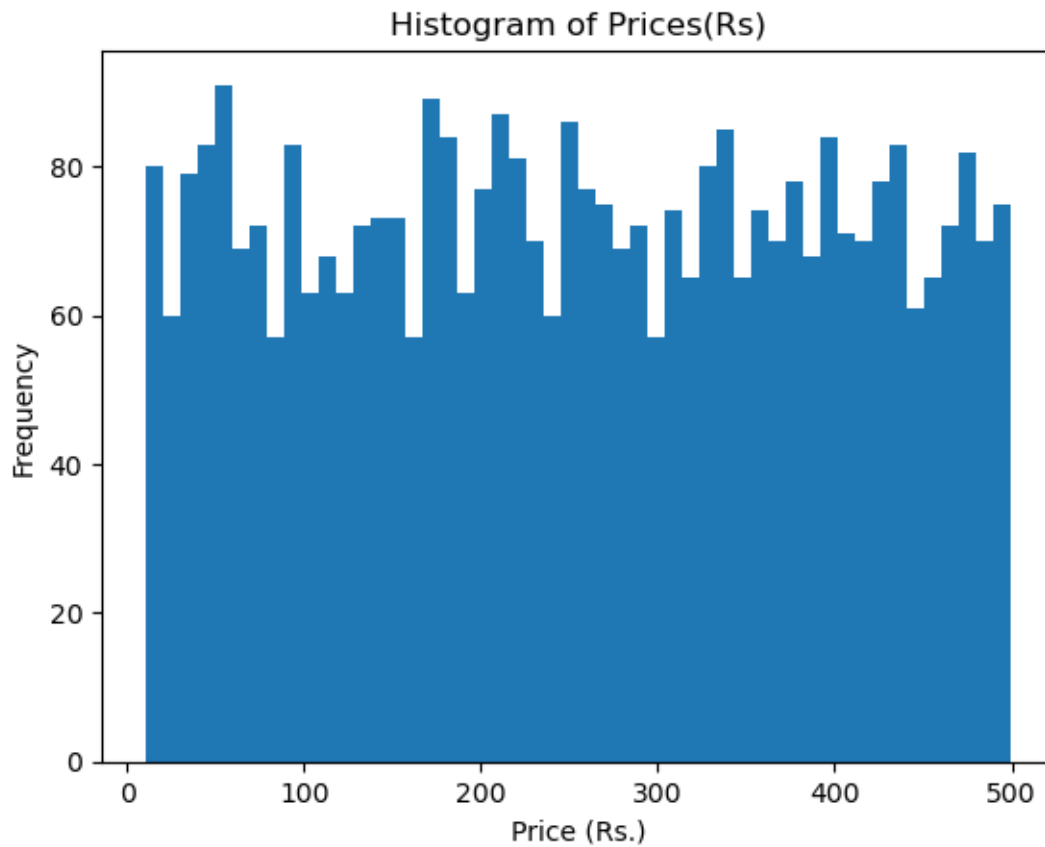
```
[19]:
```

	Price (Rs.)	Discount (%)	Final_Price(Rs.)
count	3660.000000	3660.000000	3660.000000
mean	254.800675	18.825137	206.906579
std	141.682621	14.731338	122.687844
min	10.090000	0.000000	5.890000
25%	134.012500	5.000000	104.512500
50%	253.845000	15.000000	199.185000
75%	377.595000	25.000000	304.117500
max	499.960000	50.000000	496.820000

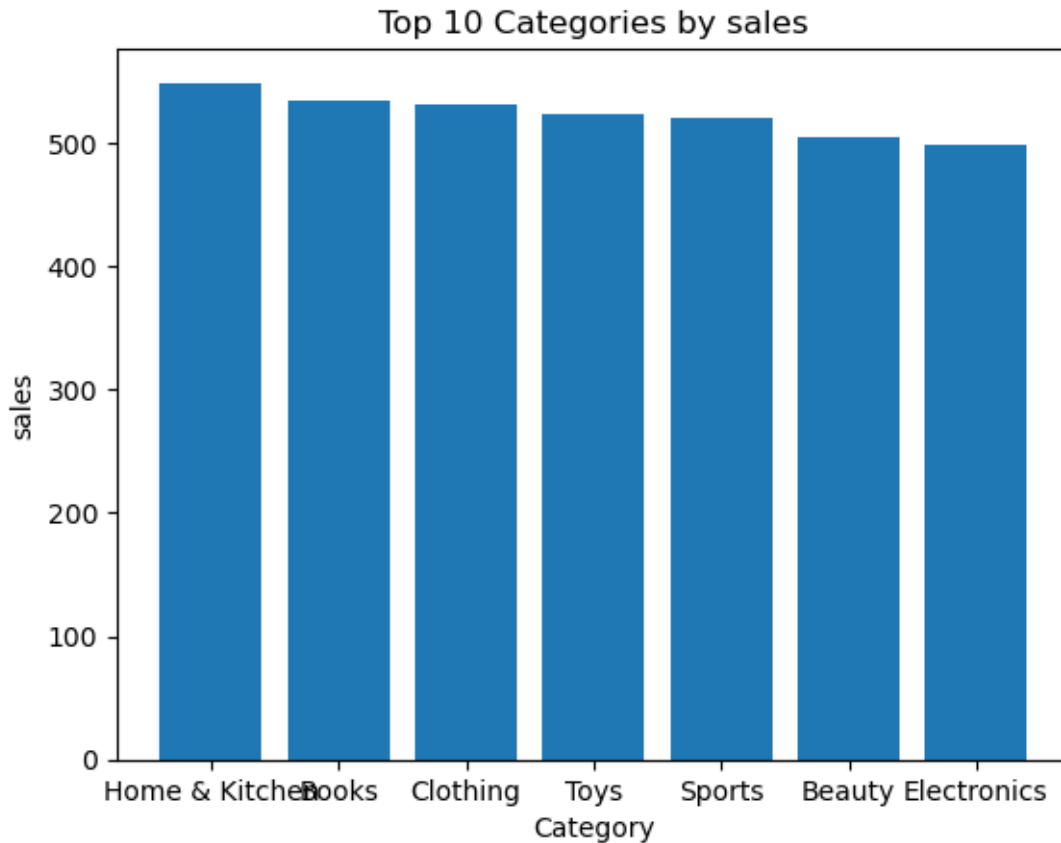
```
[23]: df["Price (Rs.)"].describe()
```

```
[23]: count    3660.000000
      mean     254.800675
      std     141.682621
      min     10.090000
      25%     134.012500
      50%     253.845000
      75%     377.595000
      max     499.960000
      Name: Price (Rs.), dtype: float64
```

```
[25]: plt.hist(df['Price (Rs.)'], bins=50)
      plt.xlabel('Price (Rs.)')
      plt.ylabel('Frequency')
      plt.title('Histogram of Prices(Rs)')
      plt.show()
```



```
[27]: #BAR CHART OF TOP 10 CATEGORIES BY SALES
top_10_categories = df['Category'].value_counts().head(10)
plt.bar(top_10_categories.index,top_10_categories.values)
plt.xlabel('Category')
plt.ylabel('sales')
plt.title('Top 10 Categories by sales')
plt.show()
```



```
[29]: df.head()
```

```
[29]:
```

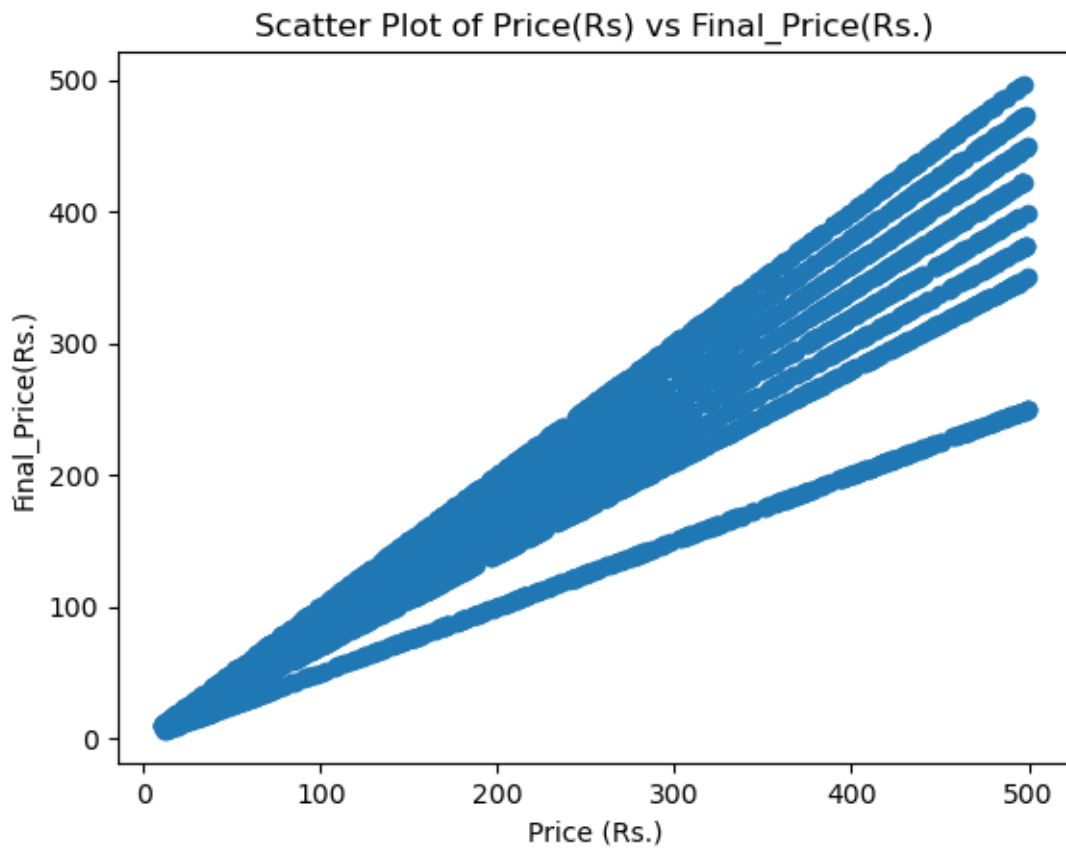
	User_ID	Product_ID	Category	Price (Rs.)	Discount (%)	\
0	337c166f	f414122f-e	Sports	36.53	15	
1	d38a19bf	fde50f9c-5	Clothing	232.79	20	
2	d7f5f0b0	0d96fc90-3	Sports	317.02	25	
3	395d4994	964fc44b-d	Toys	173.19	25	
4	a83c145c	d70e2fc6-e	Beauty	244.80	20	

	Final_Price(Rs.)	Payment_Method	Purchase_Date
0	31.05	Net Banking	12-11-2024
1	186.23	Net Banking	09-02-2024
2	237.76	Credit Card	01-09-2024
3	129.89	UPI	01-04-2024
4	195.84	Net Banking	27-09-2024

```
[31]: #Scatter Plot of Price(Rs) vs Quantity
```

```
plt.scatter(df['Price (Rs.)'], df['Final_Price(Rs.)'])
plt.xlabel('Price (Rs.)')
```

```
plt.ylabel('Final_Price(Rs.)')
plt.title('Scatter Plot of Price(Rs) vs Final_Price(Rs.)')
plt.show()
```



[35]: *#Category-wise Mean Price**

```
category_mean_price = df.groupby('Category')['Price (Rs.)'].mean().reset_index()
print(category_mean_price)
```

	Category	Price (Rs.)
0	Beauty	252.738693
1	Books	259.123052
2	Clothing	263.115913
3	Electronics	251.646867
4	Home & Kitchen	249.255938
5	Sports	258.619212
6	Toys	248.962772

[36]: *#Category-wise Sales Count**

```
category_sales_count = df['Category'].value_counts().reset_index()
category_sales_count.columns = ['Category', 'Sales Count']
print(category_sales_count)
```

	Category	Sales Count
0	Home & Kitchen	549
1	Books	534
2	Clothing	531
3	Toys	523
4	Sports	520
5	Beauty	505
6	Electronics	498

[37]: *#Top 5 Categories by Sales**

```
top_5_categories = category_sales_count.nlargest(5, 'Sales Count')
print(top_5_categories)
```

	Category	Sales Count
0	Home & Kitchen	549
1	Books	534
2	Clothing	531
3	Toys	523
4	Sports	520

[40]: *#Top 5 Categories ki Average Price**

```
top_5_categories = df['Category'].value_counts().nlargest(5).index.tolist()
category_avg_price_dict = {}
for category in top_5_categories:
    category_avg_price_dict[category] = df.loc[df['Category'] == category, 'Price_
↳(Rs.)'].mean()
print(category_avg_price_dict)
```

```
{'Home & Kitchen': 249.25593806921677, 'Books': 259.1230524344569, 'Clothing':
263.1159133709981, 'Toys': 248.96277246653923, 'Sports': 258.6192115384615}
```

[42]: *#most used payment method*

```
print(df['Payment_Method'].value_counts().head(1))
```

```
Payment_Method
Credit Card    760
Name: count, dtype: int64
```

[]: