

```
# level 1 --project 1

#Exploratory Data Analysis (EDA) on Retail Sales Data

#importing necessary libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns
```

### #1. data loading & cleaning

```
file_path = 'retail_sales_dataset.csv'

df = pd.read_csv(file_path)


# adjusting display options for our dataframe

pd.set_option('display.max_rows', 20)

pd.set_option('display.max_columns', None)

pd.set_option('display.width', None)

pd.set_option('display.max_colwidth', None)

print("Last few rows of the dataframe:")

print(df.tail())

# displaying the information about the dataframe

print("\nDataFrame Info:")

df.info()


# printing the count of null values in each column for checking of completeness

print("\nnull counts for each column:")

print(df.isnull().sum())

print("Total sales : {}".format(df["Quantity"].sum()))

print("Total profit : {}".format(df["Total Amount"].sum()))

#for counting the no of duplicated rows in our dataframe

print(df.duplicated().sum())
```

output ----

Last few rows of the dataframe:

	Transaction ID	Date	Customer ID	Gender	Age	Product Category	Quantity	Price per Unit	Total Amount
995	996	2023-05-16	CUST996	Male	62	Clothing	1	50	50
996	997	2023-11-17	CUST997	Male	52	Beauty	3	30	90
997	998	2023-10-29	CUST998	Female	23	Beauty	4	25	100
998	999	2023-12-05	CUST999	Female	36	Electronics	3	50	150
999	1000	2023-04-12	CUST1000	Male	47	Electronics	4	30	120

DataFrame Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1000 entries, 0 to 999

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
0	Transaction ID	1000 non-null	int64
1	Date	1000 non-null	object
2	Customer ID	1000 non-null	object
3	Gender	1000 non-null	object
4	Age	1000 non-null	int64
5	Product Category	1000 non-null	object
6	Quantity	1000 non-null	int64
7	Price per Unit	1000 non-null	int64
8	Total Amount	1000 non-null	int64

dtypes: int64(5), object(4)

memory usage: 70.4+ KB

null counts for each column:

Transaction ID 0

Date 0

```
Customer ID    0
Gender         0
Age            0
Product Category  0
Quantity       0
Price per Unit  0
Total Amount    0
dtype: int64
Total sales : 2514
Total profit : 456000
0
```

## #2.descriptive statistics

#it means numerical measures that summarize the main features of a dataset  
(mean,median,mode,standard deviation, etc.)

```
print(df.describe())
```

output--

	Transaction ID	Age	Quantity	Price per Unit	Total Amount
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	500.500000	41.39200	2.514000	179.890000	456.000000
std	288.819436	13.68143	1.132734	189.681356	559.997632
min	1.000000	18.00000	1.000000	25.000000	25.000000
25%	250.750000	29.00000	1.000000	30.000000	60.000000
50%	500.500000	42.00000	3.000000	50.000000	135.000000
75%	750.250000	53.00000	4.000000	300.000000	900.000000
max	1000.000000	64.00000	4.000000	500.000000	2000.000000

### #3.time series analysis

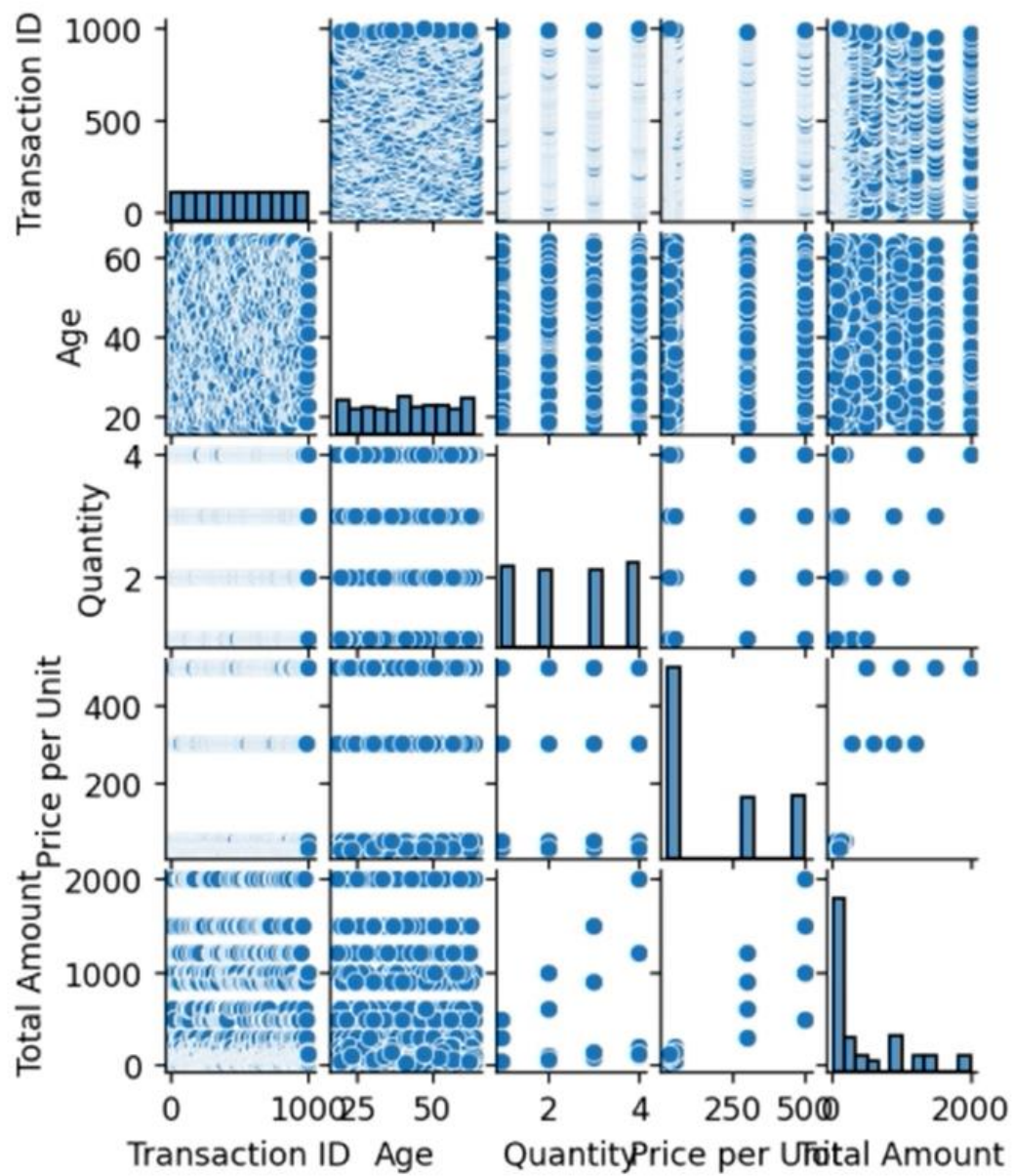
# used to analyze data that is collected, recorded, or observed over time

```
print(sns.pairplot(df))
```

```
plt.show()
```

output--

time series analysis



```
print("-----")
```

#### #4.Customer and product analysis

#here we are analyzing customer demographics and purchasing behavior.

```
print(df["Product Category"].value_counts())
```

output--

Customer and product analysis

Product Category

Clothing    351

Electronics 342

Beauty     307

Name: count, dtype: int64

```
print("-----")
```

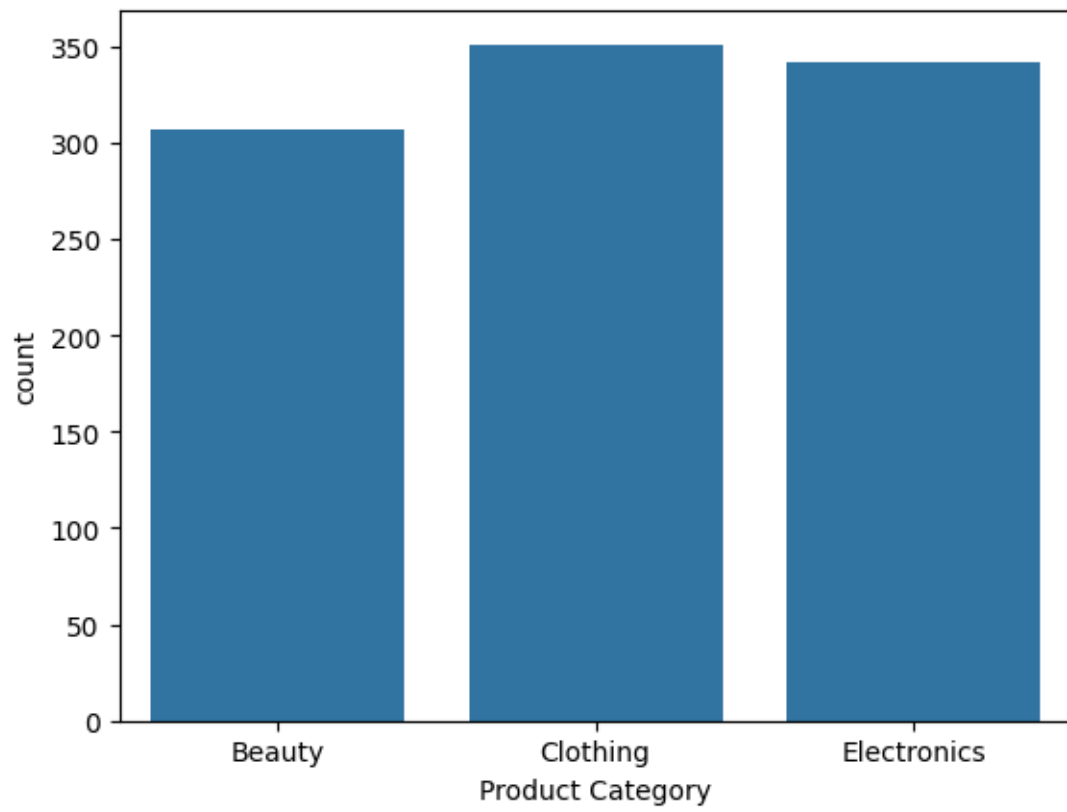
#### #5.Visualization

#here we present the insights through bar charts, line plots, and heatmaps

#bar graphs

```
sns.countplot(x=df["Product Category"])
```

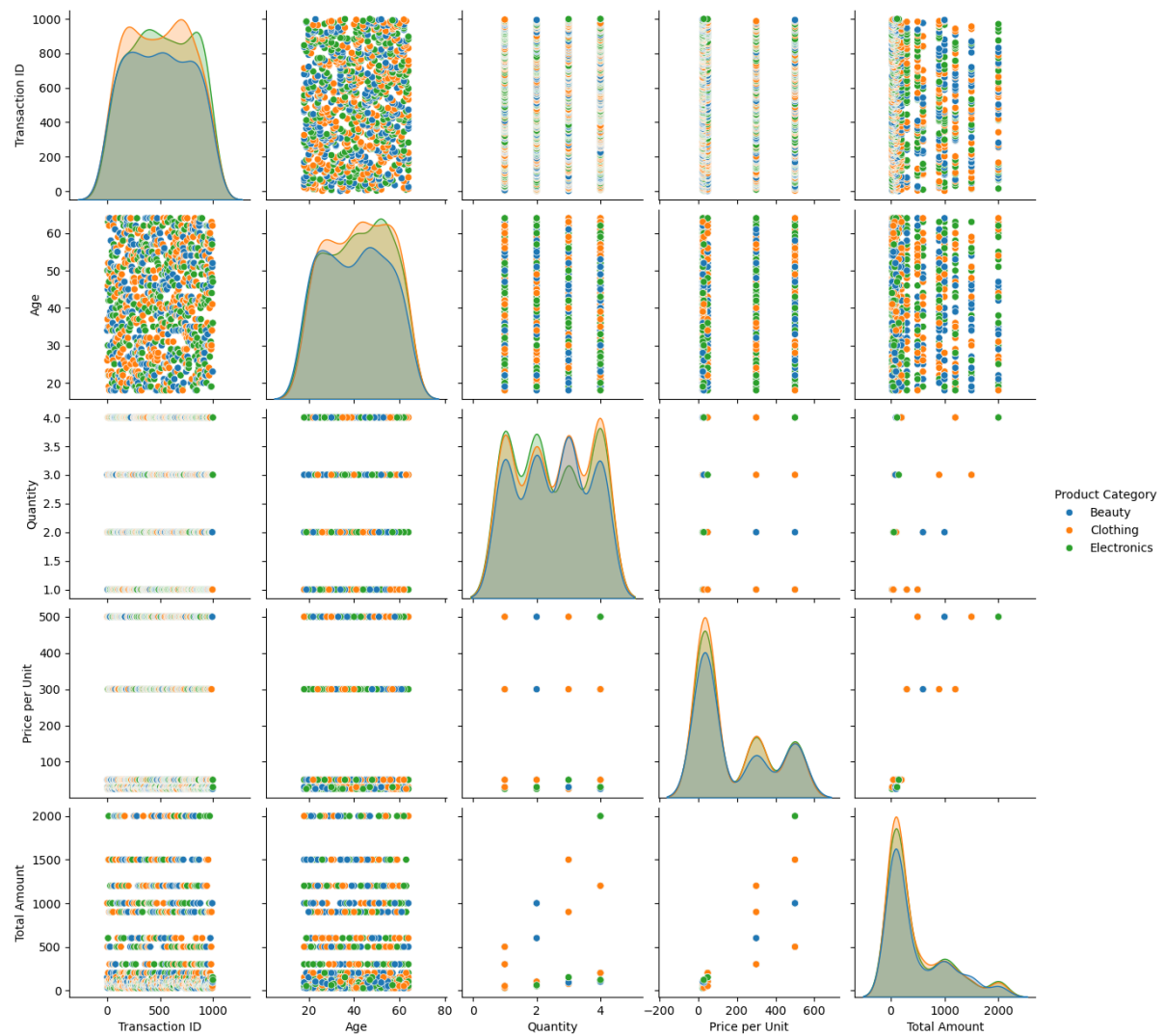
```
plt.show()
```



#heat maps

```
sns.pairplot(df, hue="Product Category")
```

```
plt.show()
```



#printing piecharts

```
cate=df["Product Category"].value_counts()
```

```
explode_list=[0,0.1,0.1]
```

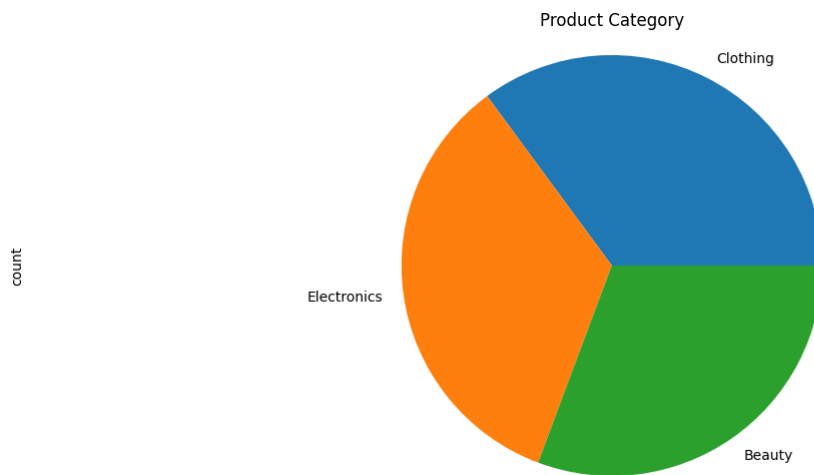
```
color_list=["Red","Blue","Seagreen"]
```

```
cate.plot(kind="pie", figsize=(15,6))
```

```
plt.title("Product Category")
```

```
plt.axis("equal")
```

```
plt.show()
```



## #6.recommendations

1.We can analyze the least performing product categories

and consider some plans to improve their sales, such as marketing campaigns or product diversification.

2.We can monitor changes in sales trends over time to identify upcoming market opportunities according to consumer preferences.