NAME:KASAGONI JAGRUTHI

COLLEGE:JNTUHCEJ

BRANCH:CSE

PROJECTNAME : Exploratory Data Analysis MINI PROJECT


```
#importing necessary packages

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv('/content/sheet1.csv.csv')
df
```

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs
0	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	1/5/2019	13:08	Ewallet	522.83
1	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.8200	80.2200	3/8/2019	10:29	Cash	76.40
2	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	3/3/2019	13:23	Credit card	324.31
3	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.2880	489.0480	1/27/2019	20:33	Ewallet	465.76
4	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	2/8/2019	10:37	Ewallet	604.17
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	233-67-5758	C	Naypyitaw	Normal	Male	Health and beauty	40.35	1	2.0175	42.3675	1/29/2019	13:46	Ewallet	40.35
996	303-96-2227	B	Mandalay	Normal	Female	Home and lifestyle	97.38	10	48.6900	1022.4900	3/2/2019	17:16	Ewallet	973.80
997	727-02-1313	A	Yangon	Member	Male	Food and beverages	31.84	1	1.5920	33.4320	2/9/2019	13:22	Cash	31.84
998	347-56-2442	A	Yangon	Normal	Male	Home and lifestyle	65.82	1	3.2910	69.1110	2/22/2019	15:33	Cash	65.82
999	849-09-3807	A	Yangon	Member	Female	Fashion accessories	88.34	7	30.9190	649.2990	2/18/2019	13:28	Cash	618.38

```
df.shape
```

```
(1000, 17)
```

```
df.nunique()
```

```
Invoice ID              1000
Branch                     3
City                       3
Customer type              2
```

```
Gender                     2
Product line               6
Unit price               943
Quantity                  10
Tax 5%                   990
Total                    990
Date                      89
Time                     506
Payment                    3
cogs                     990
gross margin percentage    1
gross income             990
Rating                    61
dtype: int64
```

```
df.size
```

```
17000
```

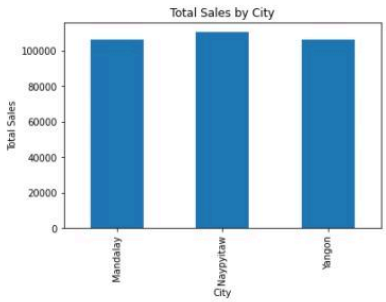
```
df.dtypes
```

```
Invoice ID                object
Branch                    object
City                      object
Customer type             object
Gender                    object
Product line              object
Unit price               float64
Quantity                   int64
Tax 5%                   float64
Total                    float64
Date                      object
Time                      object
Payment                   object
cogs                     float64
gross margin percentage  float64
gross income             float64
Rating                   float64
dtype: object
```

```
df.describe()
```

	Unit price	Quantity	Tax 5%	Total	cogs	gross margin percentage	gross income	Ratin
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.00000	1000.000000	1000.000000	1000.0000
mean	55.672130	5.510000	15.379369	322.966749	307.58738	4.761905	15.379369	6.9727
std	26.494628	2.923431	11.708825	245.885335	234.17651	0.000000	11.708825	1.7185
min	10.080000	1.000000	0.508500	10.678500	10.17000	4.761905	0.508500	4.0000
25%	32.875000	3.000000	5.924875	124.422375	118.49750	4.761905	5.924875	5.5000
50%	55.230000	5.000000	12.088000	253.848000	241.76000	4.761905	12.088000	7.0000
75%	77.935000	8.000000	22.445250	471.350250	448.90500	4.761905	22.445250	8.5000


```
#cities that are biggest contributors tp overall states
import matplotlib.pyplot as plt
df_city_sales = df.groupby('City')['Total'].sum()
df_city_sales.plot(kind='bar')
plt.title('Total Sales by City')
plt.xlabel('City')
plt.ylabel('Total Sales')
plt.show()
```




```
#want to know how many males and females
df.Gender.value_counts()
```

```
Female    501
Male      499
Name: Gender, dtype: int64
```

```
#the average total amount spent by each customer type and gender
df_customer_type = df.groupby('Customer type')['Total'].mean()
print(df_customer_type)
df_customer_type.plot(kind='bar', title = 'Average Total Amount Spent By Customer type ')
```

```
Customer type
Member    327.791305
Normal    318.122856
Name: Total, dtype: float64
<matplotlib.axes._subplots.AxesSubplot at 0x7fc35684f4c0>
```


```
#average unit price for each product line
df_product_line_price = df.groupby('Product line')['Unit price'].mean()
print(df_product_line_price)
```

```
Product line
Electronic accessories    53.551588
Fashion accessories       57.153652
Food and beverages        56.008851
Health and beauty         54.854474
Home and lifestyle        55.316937
Sports and travel         56.993253
Name: Unit price, dtype: float64
```

```
#relationship between the unitprice and quantity of each project
plt.scatter(df['Unit price'], df['Quantity'])
plt.title('Unit Price vs Quantity')
plt.xlabel('Unit Price')
plt.ylabel('Quantity')
plt.show()
```


```
#most profitable product lines in supermarket
product_line_income = df.groupby('Product line')['gross income'].sum()
product_line_income = product_line_income.sort_values(ascending=False)
print('Most profitable product lines:')
print(product_line_income.head(10))
```

```
Most profitable product lines:
Product line
Food and beverages        2673.5640
Sports and travel         2624.8965
Electronic accessories    2587.5015
Fashion accessories       2585.9950
```



```
Home and lifestyle        2564.8530
Health and beauty         2342.5590
Name: gross income, dtype: float64
```