DATA SALES

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings
        warnings.filterwarnings('ignore')
```

```
In [2]: data =pd.read_csv('C:/Users/Oooba/Desktop/Analysis with python/intern2grow/sales_data.csv')
        data
```

Out[2]:

	date	product	category	price	quantity	revenue
0	01/01/2022	Smartphone	Electronics	600.0	10.0	6000.0
1	01/01/2022	Laptop	Electronics	1200.0	5.0	6000.0
2	02/01/2022	T-Shirt	Clothing	20.0	50.0	1000.0
3	03/01/2022	Headphones	Electronics	100.0	20.0	2000.0
4	04/01/2022	T-Shirt	Clothing	20.0	25.0	500.0
...	...	...	...	...	...	...
364	27/12/2022	Watch	Accessories	150.0	5.0	750.0
365	28/12/2022	Coat	Clothing	100.0	5.0	500.0
366	29/12/2022	Headphones	Electronics	100.0	10.0	1000.0
367	30/12/2022	Smartphone	Electronics	600.0	11.0	6600.0
368	31/12/2022	Hoodie	Clothing	40.0	30.0	1200.0

369 rows × 6 columns

Data cleaning

```
In [3]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 369 entries, 0 to 368
Data columns (total 6 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   date      369 non-null    object 
 1   product   369 non-null    object 
 2   category  369 non-null    object 
 3   price     367 non-null    float64
 4   quantity  368 non-null    float64
 5   revenue   368 non-null    float64
dtypes: float64(3), object(3)
memory usage: 17.4+ KB
```

```
In [6]: data.isna().sum()
```

```
Out[6]: date        0
        product     0
        category    0
        price       2
        quantity    1
        revenue     1
        dtype: int64
```

```
In [7]: x=data['price'].mean()
        y=data["quantity"].mean()
        z=data['revenue'].mean()
        x ,y,z
```

```
Out[7]: (211.22615803814713, 14.565217391304348, 2060.679347826087)
```

```
In [8]: data['price']= data['price'].fillna(x)
        data['quantity']= data['quantity'].fillna(y)
        data['revenue']= data['revenue'].fillna(z)
        data.isna().sum()
```

```
Out[8]: date        0
        product     0
        category    0
        price       0
        quantity    0
        revenue     0
        dtype: int64
```

```
In [9]: data.duplicated().sum()
```

```
Out[9]: 1
```

```
In [10]: data=data.drop_duplicates()
```

```
In [11]: data.describe()
```

Out[11]:

	price	quantity	revenue
count	368.000000	368.000000	368.000000
mean	211.691447	14.523275	2063.018150
std	226.848748	8.557794	1910.401151
min	20.000000	3.000000	300.000000
25%	50.000000	8.000000	800.000000
50%	100.000000	12.000000	1200.000000
75%	258.419619	20.000000	2400.000000
max	1200.000000	50.000000	7200.000000

DATA ANALYSIS

Q1:What was the total revenue generated by the company over the course of the year?

```
In [12]: Total_Revenue=data['revenue'].sum()
         print("The Total Revenue =",Total_Revenue,"$")
```

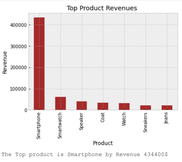
```
The Total Revenue = 759190.679347826 $
```

Q2: Which product had the highest revenue? How much revenue did it generate?

```
In [13]: Proudct_Rev =data.groupby('product')['revenue'].sum().sort_values(ascending=False).head(7)
         Proudct_Rev
```

```
Out[13]: product
         Smartphone    434400.000000
         Smartwatch     60960.479348
         Speaker        39680.000000
         Coat           33800.000000
         Watch          32700.000000
         Sneakers       21600.000000
         Jeans          20650.000000
         Name: revenue, dtype: float64
```

```
In [14]: from matplotlib import style
         style.use("bmh")
         plt.figure(facecolor="white")
         Proudct_Rev.plot(kind="bar",color='brown')
         plt.title("Top Product Revenues ")
         plt.xlabel("Product")
         plt. ylabel("Revenue")
         plt.grid(True)
         plt.show()
         print("The Top product is Smartphone by Revenue 434400$")
```




```
The Top product is Smartphone by Revenue 434400$
```

Q3:What was the average price of a product sold by the company?

```
In [15]: Price_AVG =data.groupby('product')['price'].mean().sort_values(ascending=False).reset_index()
         Price_AVG.columns=['Product','AVG Price']
         Price_AVG
```

Out[15]:

	Product	AVG Price
0	Laptop	1200.000000
1	Smartphone	600.000000
2	Tablet	400.000000
3	Smartwatch	200.000000
4	Watch	152.783007
5	Coat	100.000000
6	Headphones	100.000000
7	Sneakers	80.000000
8	Speaker	80.000000
9	Backpack	50.000000
10	Jeans	50.000000
11	Wallet	45.102180
12	Hoodie	40.000000
13	T-Shirt	20.000000

Q4:What was the total quantity of products sold by the company?

```
In [16]: Total_quantity =data['quantity'].sum()
         print("The Total Quantity of Product =", Total_quantity)
```

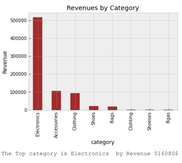
```
The Total Quantity of Product = 5344.565217391304
```

Q5:Which category had the highest revenue? How much revenue did it generate?

```
In [17]: category_Rev =data.groupby('category')['revenue'].sum().sort_values(ascending=False)
         category_Rev
```

```
Out[17]: category
         Electronics    516080.000000
         Accessories    106760.479348
         Clothing        93150.000000
         Shoes           29440.000000
         Bags            19500.000000
         Clothing         1200.000000
         Shoeses           960.000000
         Bgas              900.000000
         Name: revenue, dtype: float64
```

```
In [30]: category_Rev.plot(kind="bar",color='brown')
         plt.title(" Revenues by Category ")
         plt.xlabel("category")
         plt. ylabel("Revenue")
         plt.grid(True)
         plt.show()
         print("The Top category is Electronics  by Revenue 516080$")
```

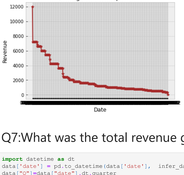



```
The Top category is Electronics  by Revenue 516080$
```

Q6:What was the average revenue per sale?

```
In [8]: Sales_avg_Rev =data.groupby('date')['revenue'].sum().sort_values(ascending=False).reset_index()
```

```
In [19]: plt.plot(Sales_avg_Rev['date'],Sales_avg_Rev['revenue'],marker="*",color="brown")
         plt.title(" Average Revenue per Sale ")
         plt.xlabel("Date")
         plt. ylabel("Revenue")
         plt.grid(True)
         plt.show()
```



Q7:What was the total revenue generated in each quarter of the year? (i.e. Q1, Q2, Q3, Q4)

```
In [26]: import datetime as dt
         data['date'] = pd.to_datetime(data['date'],  infer_datetime_format=True)
         data["Q"]=data["date"].dt.quarter
         Quarter_Rev =data.groupby("Q")["revenue"].sum()
         Quarter_Rev
```

```
Out[26]: Q
         1    211330.000000
         2    189790.679348
         3    149750.000000
         4    209320.000000
         Name: revenue, dtype: float64
```

```
In [29]: Quarter_Rev.plot(kind="bar",color='brown')
         plt.title(" Revenues by  Quarter ")
         plt.xlabel("Quarter")
         plt. ylabel("Revenue")
         plt.grid(True)
         plt.show()
```