E-COMMERCE SALES DATA ANALYSIS
PYTHON PROJECT BY VIRMANAND

[20]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

[26]:
```
df=pd.read_csv("Sample - Superstore.csv",encoding ='latin-1')
```

[28]:
```
df
```

[28]:

hip Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Di
/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600	2	
/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400	3	
/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	
/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	
/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
/23/2014	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	33180	South	FUR-FU-10001889	Furniture	Furnishings	Ultra Door Pull Handle	25.2480	3	
3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	FUR-FU-10000747	Furniture	Furnishings	Tenex B1-RE Series Chair Mats for Low Pile	91.9600	2	

```
df.head()
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category	Produ Nan
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-BO-10001798	Furniture	Bookcases	Bu Somers Collecti Bookca
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	...	42420	South	FUR-CH-10000454	Furniture	Chairs	Hon Delu Fab Upholster Stacki Chairs
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	...	90036	West	OFF-LA-10000240	Office Supplies	Labels	Se Adhesi Addre Labels f Typewrite
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	FUR-TA-10000577	Furniture	Tables	Bretfo CR45 Series Sl Rectangul Tab
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	33311	South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fc 'N Roll Ca Syste

5 rows × 21 columns

```
df.tail()
```

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	...	Postal Code	Region	Product ID	Category	Sub-Category
9989	9990	CA-2014-110422	1/21/2014	1/23/2014	Second Class	TB-21400	Tom Boeckenhauer	Consumer	United States	Miami	...	33180	South	FUR-FU-10001889	Furniture	Furnishings
9990	9991	CA-2017-121258	2/26/2017	3/3/2017	Standard Class	DB-13060	Dave Brooks	Consumer	United States	Costa Mesa	...	92627	West	FUR-FU-10000747	Furniture	Furnishings

36]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 21 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Row ID         9994 non-null   int64
 1   Order ID       9994 non-null   object
 2   Order Date     9994 non-null   object
 3   Ship Date      9994 non-null   object
 4   Ship Mode      9994 non-null   object
 5   Customer ID    9994 non-null   object
 6   Customer Name  9994 non-null   object
 7   Segment        9994 non-null   object
 8   Country        9994 non-null   object
 9   City           9994 non-null   object
 10  State          9994 non-null   object
 11  Postal Code    9994 non-null   int64
 12  Region         9994 non-null   object
 13  Product ID     9994 non-null   object
 14  Category       9994 non-null   object
 15  Sub-Category   9994 non-null   object
 16  Product Name   9994 non-null   object
 17  Sales          9994 non-null   float64
 18  Quantity       9994 non-null   int64
 19  Discount       9994 non-null   float64
 20  Profit         9994 non-null   float64
dtypes: float64(3), int64(3), object(15)
memory usage: 1.6+ MB
```

50]:
```
 row,columns=df.shape
```

52]:
```
row
```

52]:
```
9994
```

54]:
```
columns
```

54]:
```
21
```

58]:
```
df.isnull().sum()
```

58]:
```
Row ID          0
Order ID        0
Order Date      0
Ship Date       0
Ship Mode       0
Customer ID     0
Customer Name   0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
```

[66]:
```
df['Order ID'].nunique()
```

[66]:
```
5009
```

[68]:
```
df['Country'].nunique()
```

[68]:
```
1
```

[70]:
```
df['City'].nunique()
```

[70]:
```
531
```

[78]:
```
df['Region'].nunique()
```

[78]:
```
4
```

[74]:
```
df.describe()
```

[74]:

	Row ID	Postal Code	Sales	Quantity	Discount	Profit
count	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000	9994.000000
mean	4997.500000	55190.379428	229.858001	3.789574	0.156203	28.656896
std	2885.163629	32063.693350	623.245101	2.225110	0.206452	234.260108
min	1.000000	1040.000000	0.444000	1.000000	0.000000	-6599.978000
25%	2499.250000	23223.000000	17.280000	2.000000	0.000000	1.728750
50%	4997.500000	56430.500000	54.490000	3.000000	0.200000	8.666500
75%	7495.750000	90008.000000	209.940000	5.000000	0.200000	29.364000
max	9994.000000	99301.000000	22638.480000	14.000000	0.800000	8399.976000

CLEANING OF DATA

[82]:
```
#DROP THE ROW ID COLUMN
df.drop(columns=['Row ID'],inplace =True)
```

[84]:
```
df.head(1)
```

[84]:

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category
0	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798	Furniture	Bookcases

```
#change the data type object to date time
df['Order Date']=pd.to_datetime(df['Order Date'])
```

```
df['Order Month']=df['Order Date'].dt.month
```

```
df.head(1)
```

	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	...	Region	Product ID	Category	Sub-Category	Product Name
0	CA-2016-152156	2016-11-08	11/11/2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	...	South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase

1 rows × 21 columns

```
df['Order year']=df['Order Date'].dt.year
```

```
df['Order day']=df['Order Date'].dt.dayofweek
```

```
df.head(1)
```

Ship Mode	Customer ID	Customer Name	Segment	Country	City	State	...	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit	Order Month	Order year
Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	...	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96	2	0.0	41.9136	11	2016

visualization

monthly sales analysis

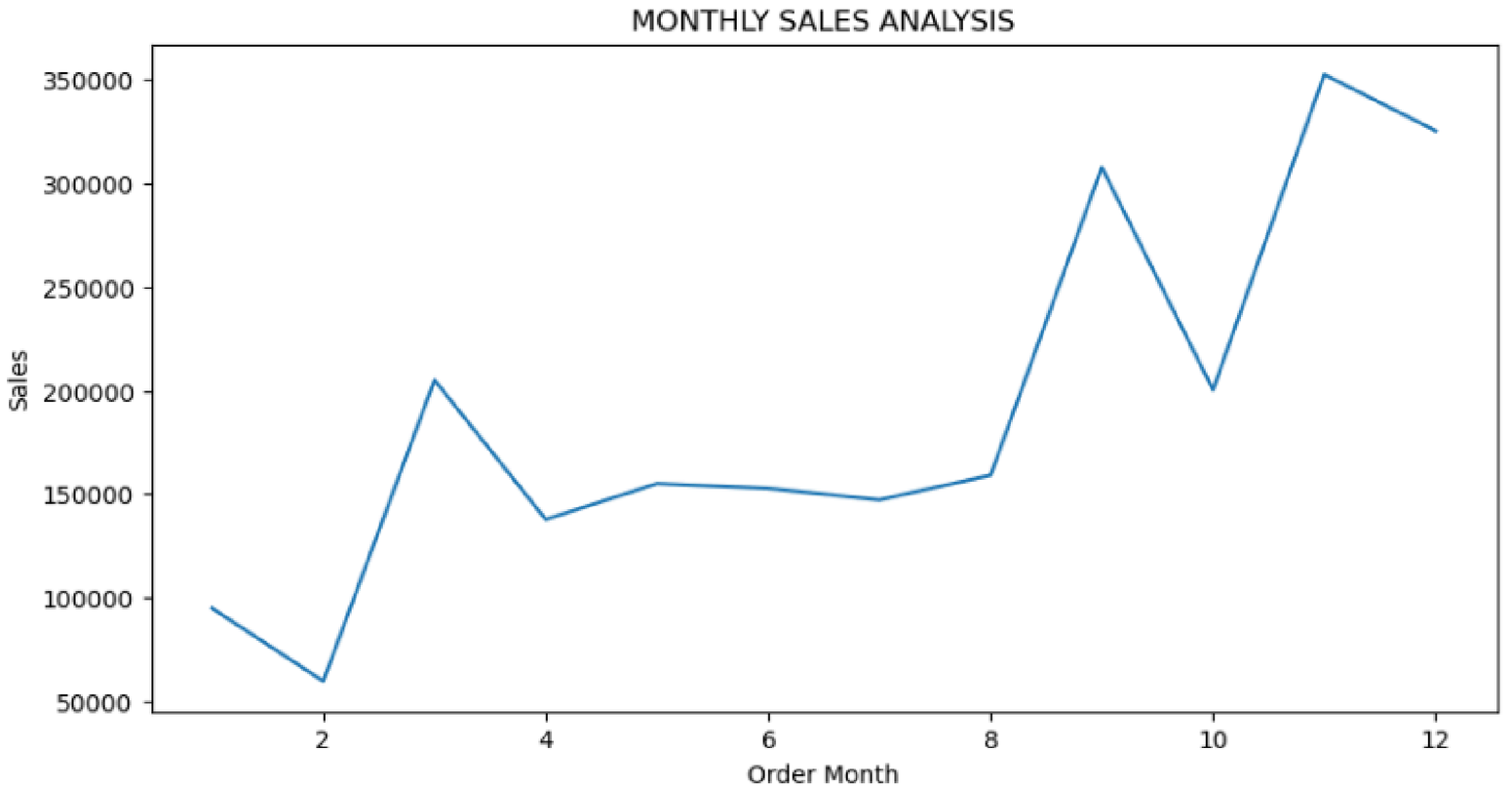
```
month=df.groupby(['Order Month'],as_index=False)['Sales'].sum().reset_index().sort_values(by='Sales',ascending=False)
```

```
month
```

	index	Order Month	Sales
10	10	11	352461.0710
11	11	12	325293.5035
8	8	9	307649.9457

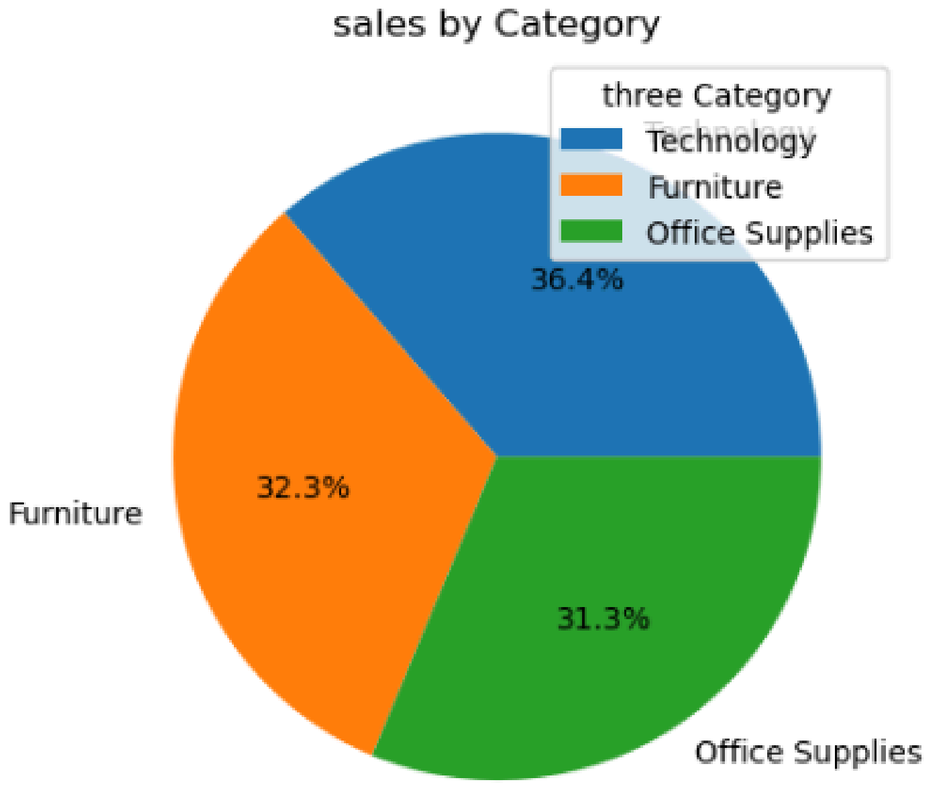
9	9	10	200322.9847
7	7	8	159044.0630
4	4	5	155028.8117
5	5	6	152718.6793
6	6	7	147238.0970
3	3	4	137762.1286
0	0	1	94924.8356
1	1	2	59751.2514

[151]:
```
plt.figure(figsize=(10,5))
plt.title('MONTHLY SALES ANALYSIS')
sns.lineplot(y='Sales',x='Order Month',data=month)
plt.show()
```



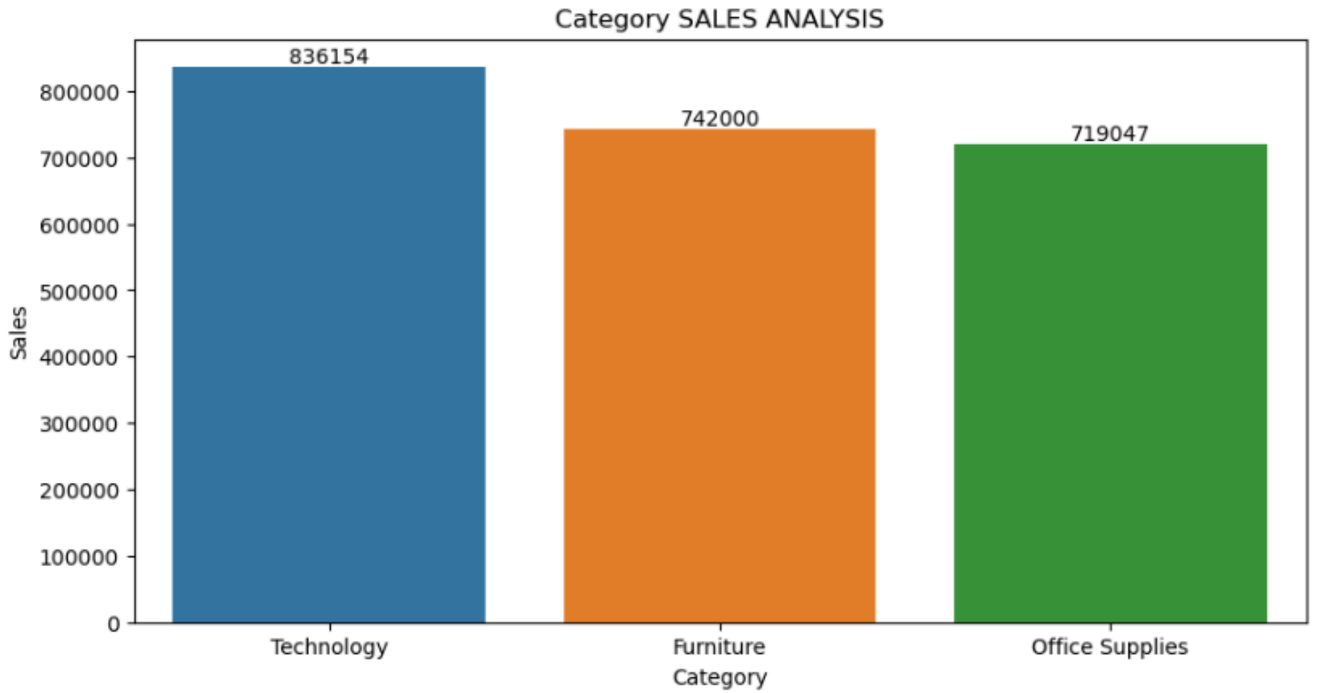
[143]:
```
ctgy=df.groupby(['Category'],as_index=False)['Sales'].sum().reset_index().sort_values(by='Sales',ascending=False)
```

[195]:
```
plt.pie(ctgy['Sales'],labels=ctgy['Category'],autopct='%1.1f%%')
plt.legend(title='three Category')
plt.title('sales by Category')
plt.show()
```



[181]:
```
plt.figure(figsize=(10,5))
plt.title('Category SALES ANALYSIS')
x=sns.barplot(y='Sales',x='Category',data=ctgy)
for i in x.containers:
    x.bar_label(i)

plt.show()
```



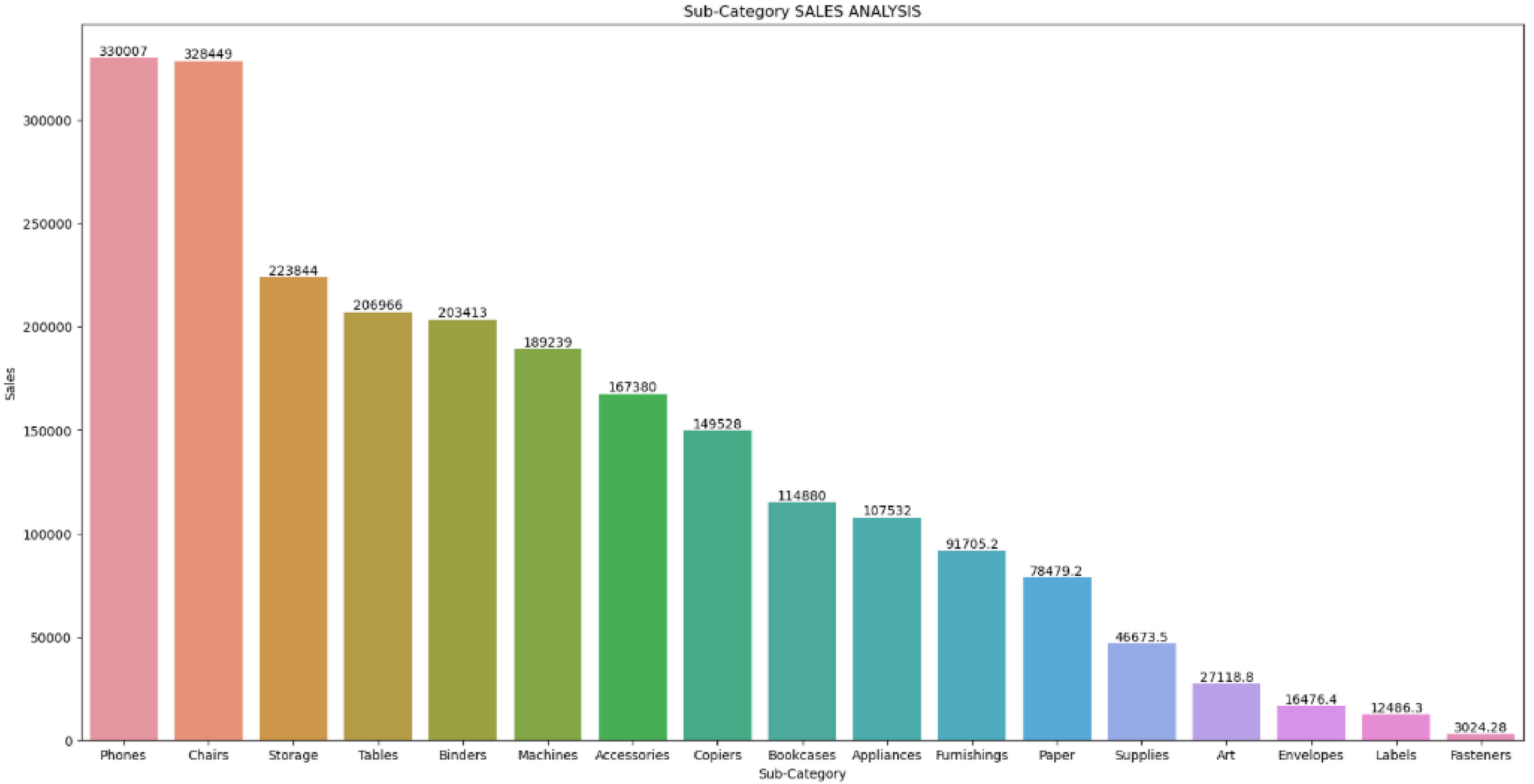
[183]:
```
sc=df.groupby(['Sub-Category'],as_index=False)['Sales'].sum().reset_index().sort_values(by='Sales',ascending=False)
```

[185]:
```
sc
```

[185]:

	index	Sub-Category	Sales
13	13	Phones	330007.0540
5	5	Chairs	328449.1030
14	14	Storage	223843.6080
16	16	Tables	206965.5320

```
]: plt.figure(figsize=(20,10))
plt.title('Sub-Category SALES ANALYSIS')
x=sns.barplot(y='Sales',x='Sub-Category',data=sc)
for i in x.containers:
    x.bar_label(i)
plt.show()
```

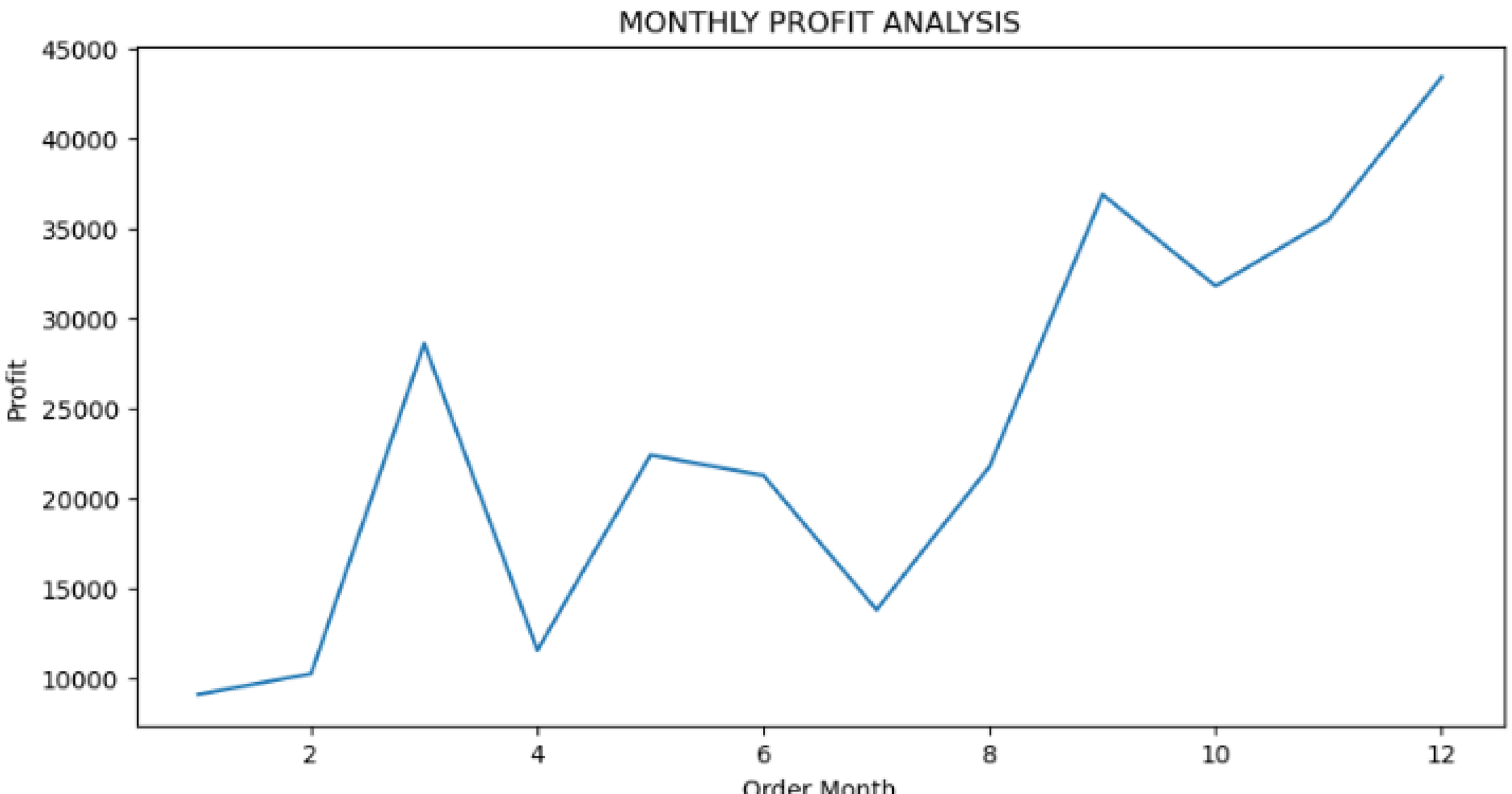


```
month=df.groupby(['Order Month'],as_index=False)['Profit'].sum().reset_index().sort_values(by='Profit',ascending=False)
```

```
month
```

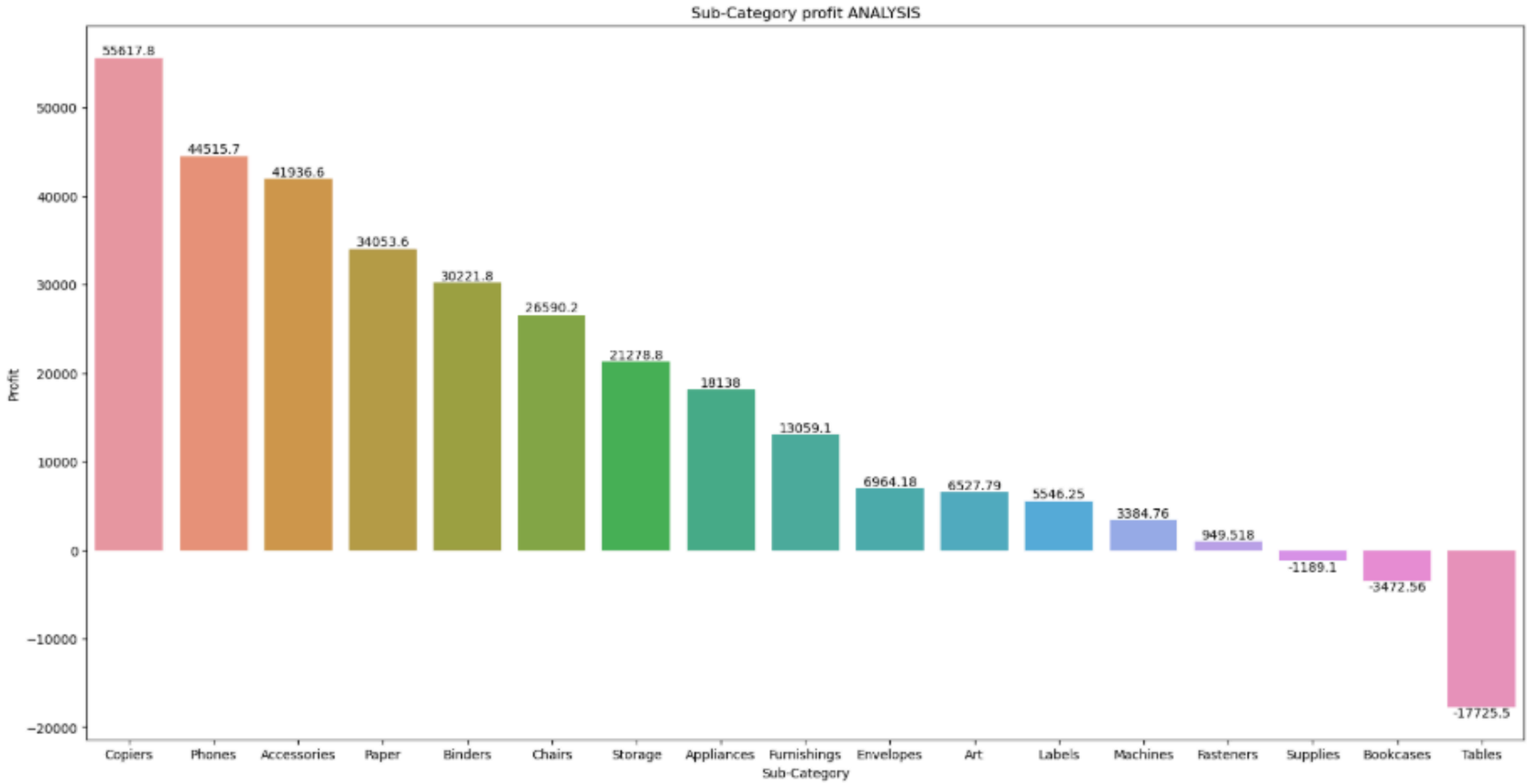
	index	Order Month	Profit
11	11	12	43369.1919
8	8	9	36857.4753
10	10	11	35468.4265
9	9	10	31784.0413
2	2	3	28594.6872
4	4	5	22411.3078
7	7	8	21776.9384
5	5	6	21285.7954
6	6	7	13832.6648
3	3	4	11587.4363
1	1	2	10294.6107
0	0	1	9134.4461

```
plt.figure(figsize=(10,5))
plt.title('MONTHLY PROFIT ANALYSIS')
sns.lineplot(y='Profit',x='Order Month',data=month)
plt.show()
```



2	2	Art	6527.7870
10	10	Labels	5546.2540
11	11	Machines	3384.7569
8	8	Fasteners	949.5182
15	15	Supplies	-1189.0995
4	4	Bookcases	-3472.5560
16	16	Tables	-17725.4811

[243]:
```
plt.figure(figsize=(20,10))
plt.title('Sub-Category profit ANALYSIS')
x=sns.barplot(y='Profit',x='Sub-Category',data=sp)
for i in x.containers:
    x.bar_label(i)
plt.show()
```



analyse sales-to-profit ratio

300]:
```
sales_profit_by_segment = df.groupby('Segment').agg({'Sales': 'sum', 'Profit': 'sum'}).reset_index()
sales_profit_by_segment['Sales_to_Profit_Ratio'] = sales_profit_by_segment['Sales'] / sales_profit_by_segment['Profit']
print(sales_profit_by_segment[['Segment', 'Sales_to_Profit_Ratio']])
```

```
       Segment  Sales_to_Profit_Ratio
0     Consumer               8.659471
1    Corporate               7.677245
2  Home Office               7.125416
```

[ ]: