Name : SINU S MARIAM

Designation : Data Science & Business Analytics Intern

Organization :The Sparks Foundation

#Task 3 - Exploratory Data Analysis - Retail

Data Set - SampleSuperstore

Problem Statement:

- Perform Explaratory Data Analysis on data set 'SampleSuperstore'.
- As a Business Manager Try to find out the weak areas where we can work to make more profit
- Derive Business problems by Exploring the data
- Create Dashboards for the same

In [1]:
```
#importing necessary libraries
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
#importing libraries for visualisation
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
```

In [2]:
```
#importing Data
data_file=r'C:\Users\sinun\Downloads\SampleSuperstore.csv'
data_frame=pd.read_csv(data_file)
```

In [3]:
```
####  Performing descriptive analysis. Understand the variables and their correspond
```

In [4]:
```
# Understanding the Data Variables
data_frame.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9994 entries, 0 to 9993
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Ship Mode     9994 non-null   object 
 1   Segment       9994 non-null   object 
 2   Country       9994 non-null   object 
 3   City          9994 non-null   object 
 4   State         9994 non-null   object 
 5   Postal Code   9994 non-null   int64  
 6   Region        9994 non-null   object 
 7   Category      9994 non-null   object 
 8   Sub-Category  9994 non-null   object 
 9   Sales         9994 non-null   float64
 10  Quantity      9994 non-null   int64  
 11  Discount      9994 non-null   float64
 12  Profit        9994 non-null   float64
```

```
dtypes: float64(3), int64(2), object(8)
memory usage: 1015.1+ KB
```

In [5]:
```
# Show the top 5 Rows of data
data_frame.head()
```

Out[5]:

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sale
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	261.960
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	731.940
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	14.620
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	957.577
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	22.368

In [6]:
```
# Performing Descriptive Analysis
data_frame.describe().T
```

Out[6]:

	count	mean	std	min	25%	50%	75%	m
Postal Code	9994.0	55190.379428	32063.693350	1040.000	23223.00000	56430.5000	90008.000	99301.0
Sales	9994.0	229.858001	623.245101	0.444	17.28000	54.4900	209.940	22638.4
Quantity	9994.0	3.789574	2.225110	1.000	2.00000	3.0000	5.000	14.0
Discount	9994.0	0.156203	0.206452	0.000	0.00000	0.2000	0.200	0.8
Profit	9994.0	28.656896	234.260108	-6599.978	1.72875	8.6665	29.364	8399.9

In [7]:
```
# Displaying the columns in the dataset
data_frame.columns
```

Out[7]:
```
Index(['Ship Mode', 'Segment', 'Country', 'City', 'State', 'Postal Code',
       'Region', 'Category', 'Sub-Category', 'Sales', 'Quantity', 'Discount',
       'Profit'],
      dtype='object')
```

In [8]:
```
# Checking for null values
data_frame.isnull().sum()
```

Out[8]:
```
Ship Mode       0
Segment         0
Country         0
City            0
State           0
Postal Code     0
Region          0
Category        0
```

```
Sub-Category    0
Sales           0
Quantity        0
Discount        0
Profit          0
dtype: int64
```

- No missing values in any columns

In [9]:
```
#Understand the main categories of products sold by company
data_frame.Category.unique()
```

Out[9]:
```
array(['Furniture', 'Office Supplies', 'Technology'], dtype=object)
```

In [10]:
```
data_frame.rename(columns = {'Sub-Category':'SubCategory'}, inplace = True)#Renaming
```

In [11]:
```
#Understand the main Subcategories of products sold by company
data_frame.SubCategory.unique()
```

Out[11]:
```
array(['Bookcases', 'Chairs', 'Labels', 'Tables', 'Storage',
       'Furnishings', 'Art', 'Phones', 'Binders', 'Appliances', 'Paper',
       'Accessories', 'Envelopes', 'Fasteners', 'Supplies', 'Machines',
       'Copiers'], dtype=object)
```

In [12]:
```
#Understand the different segments of Customers purchasing the products
data_frame.Segment.unique()
```

Out[12]:
```
array(['Consumer', 'Corporate', 'Home Office'], dtype=object)
```

Retail Products company sells Products in 3 main categories:

- Furniture (chairs, Tables,Office Chairs etc)
- Office supplies (Paper, Binders, Fasteners, Envelopes etc)
- Technology (Phones, Accessories, Various machines) ##### -The 3 main type of Segments to which products are sold are Home Office, Corporate and Consumer.

Understanding the Target Audience or End User

- Buisness Manager of the retail products company is the End User/Target Audience. We have to make data visualization regarding sales and Profit for each Category and Subcategory of products SEGMENTwise to help him determine various strategies about which products his company should continue to offer for sale and which products should be discontinued

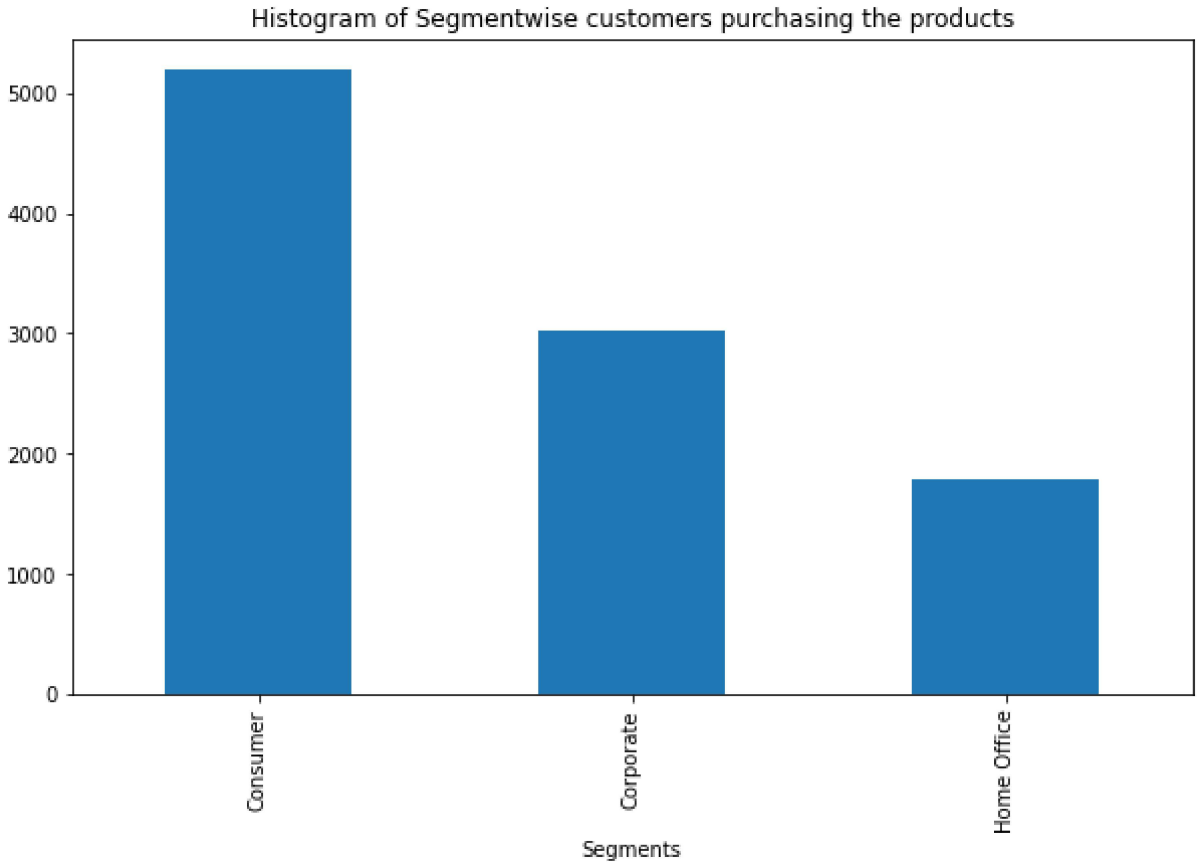
Visually Explore the variables using Histogram

In [13]:
```
#Find the number of each segments of customers
data_frame['Segment'].value_counts()
```

Out[13]:
```
Consumer       5191
Corporate      3020
Home Office    1783
Name: Segment, dtype: int64
```

In [14]:
```
# Plot Histogram of Segment
data_frame['Segment'].value_counts().plot(kind='bar',figsize=(10,6),title="Histogram
```

Out[14]:
```
<AxesSubplot:title={'center':'Histogram of Segmentwise customers purchasing the prod
ucts'}, xlabel='Segments'>
```

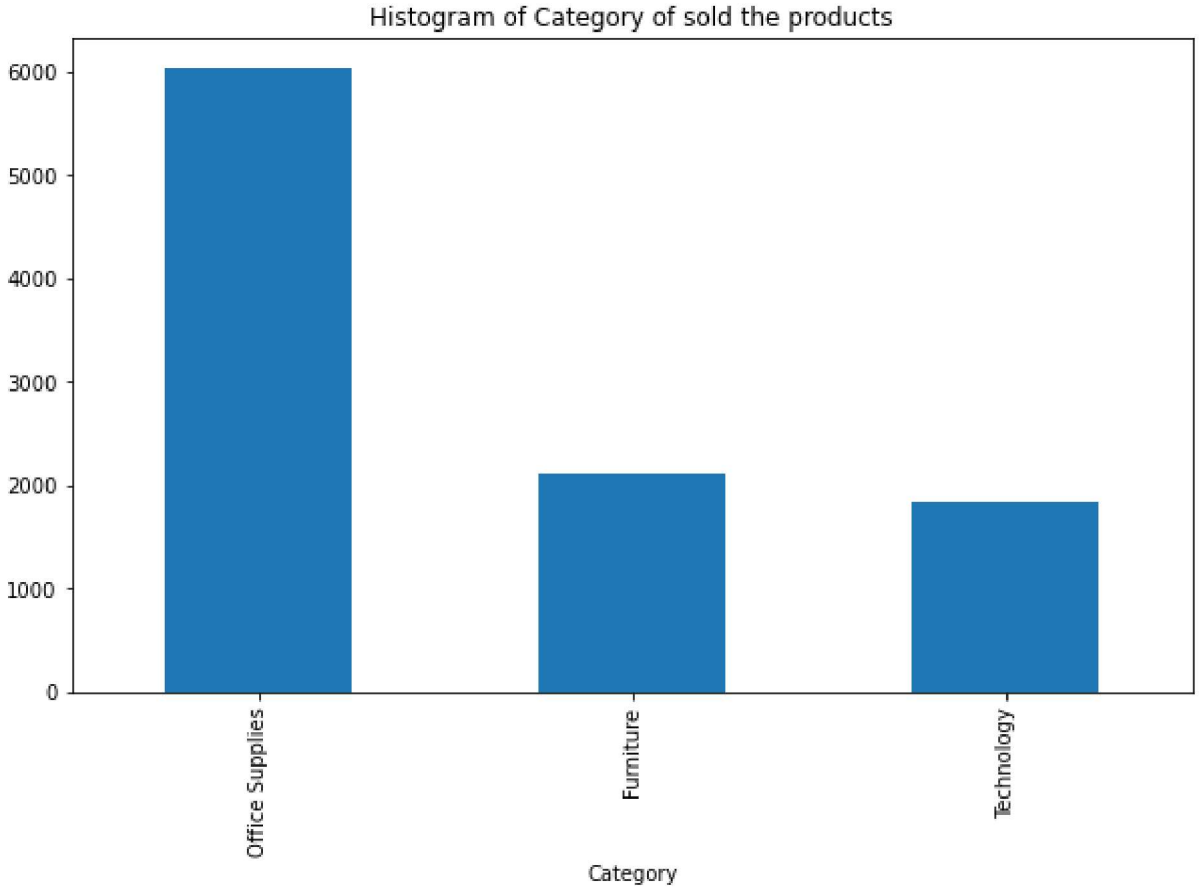


In [15]:
```
#Find the number of each Category of Products
data_frame['Category'].value_counts()
```

Out[15]:
```
Office Supplies     6026
Furniture           2121
Technology          1847
Name: Category, dtype: int64
```

In [16]:
```
# Plot Histogram of Category
data_frame['Category'].value_counts().plot(kind='bar',figsize=(10,6),title="Histogra
```

Out[16]:
```
<AxesSubplot:title={'center':'Histogram of Category of sold the products'}, xlabel
='Category'>
```

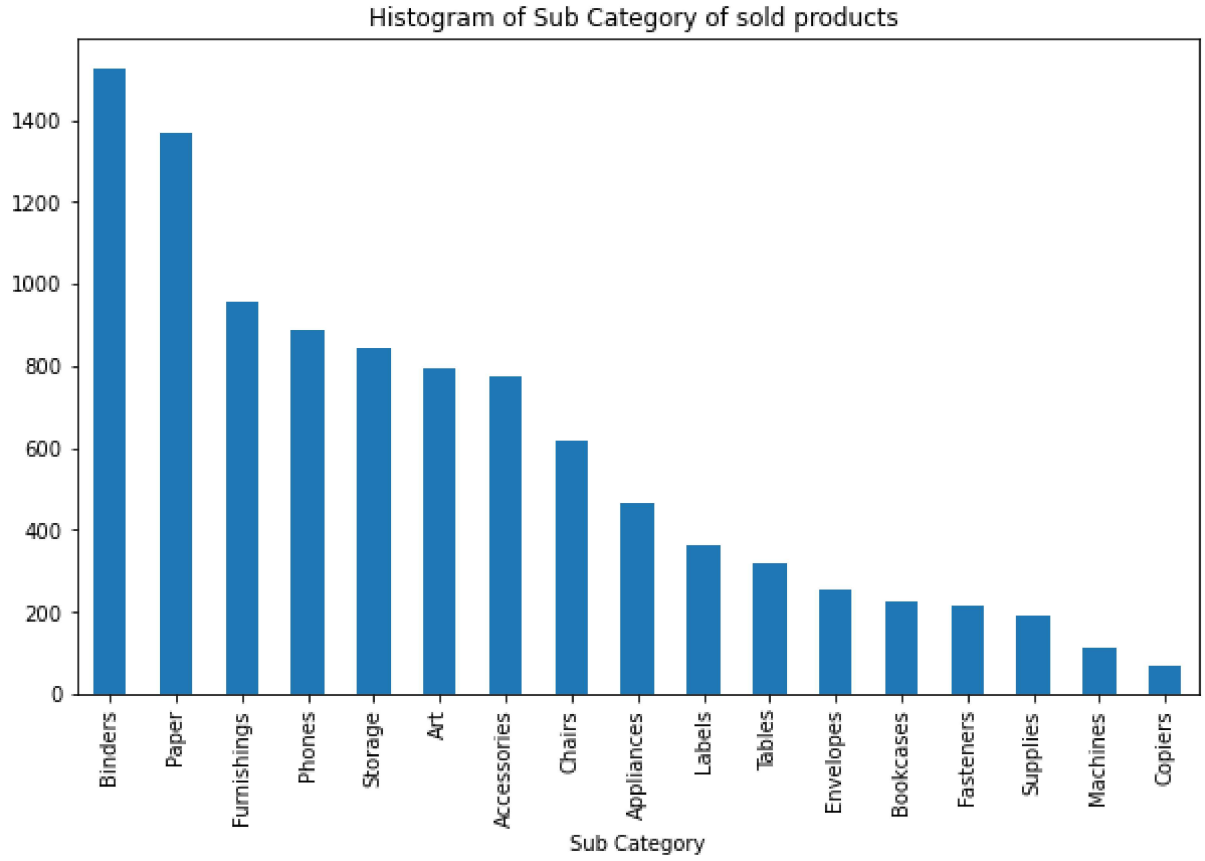


In [17]:
```
# Understand the different values in each SubCategory
data_frame['SubCategory'].value_counts()
```

Out[17]:
```
Binders        1523
Paper          1370
Furnishings     957
Phones          889
Storage         846
Art             796
Accessories     775
Chairs          617
Appliances      466
Labels          364
Tables          319
Envelopes       254
Bookcases       228
Fasteners       217
Supplies        190
Machines        115
Copiers          68
Name: SubCategory, dtype: int64
```

In [18]:
```
# Plot Histogram of SubCategory
data_frame['SubCategory'].value_counts().plot(kind='bar',figsize=(10,6),title="Histo
```

Out[18]:
```
<AxesSubplot:title={'center':'Histogram of Sub Category of sold products '}, xlabel
='Sub Category'>
```



* SALES AND PROFIT DASHBOARD IS CREATED USING TABLEAU