

```
import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
```

```
df="C:/Users/hp/OneDrive/Desktop/New folder/Amazon 2_Raw.xlsx"
pd.read_excel(df)
```

	Order ID	Order Date	Ship Date	
EmailID \				
0	CA-2013-138688	2013-06-13	2013-06-17	DarrinVanHuff@gmail.com
1	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
2	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
3	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
4	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
...	...	...	...	...
3198	CA-2013-125794	2013-09-30	2013-10-04	MarisLaWare@gmail.com
3199	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3200	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3201	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3202	CA-2014-119914	2014-05-05	2014-05-10	ChrisCortes@gmail.com

	Geography	Category \
0	United States, Los Angeles, California	Labels
1	United States, Los Angeles, California	Furnishings
2	United States, Los Angeles, California	Art
3	United States, Los Angeles, California	Phones
4	United States, Los Angeles, California	Binders
...	...	...
3198	United States, Los Angeles, California	Accessories
3199	United States, Costa Mesa, California	Furnishings
3200	United States, Costa Mesa, California	Phones
3201	United States, Costa Mesa, California	Paper
3202	United States, Westminster, California	Appliances

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860

```

7
2                                     Newell 322      7.280
4
3                                     Mitel 5320 IP Phone VoIP phone  907.152
4
4      DXL Angle-View Binders with Locking Rings by S...    18.504
3
...                                     ...      ...
...
3198  Memorex Mini Travel Drive 64 GB USB 2.0 Flash ...    36.240
1
3199  Tenex B1-RE Series Chair Mats for Low Pile Car...    91.960
2
3200                                     Aastra 57i VoIP phone  258.576
2
3201  It's Hot Message Books with Stickers, 2 3/4" x 5"    29.600
4
3202  Acco 7-Outlet Masterpiece Power Center, Wihtou...  243.160
2

```

```

      Profit
0      6.8714
1     14.1694
2      1.9656
3     90.7152
4      5.7825
...      ...
3198  15.2208
3199  15.6332
3200  19.3932
3201  13.3200
3202  72.9480

```

[3203 rows x 10 columns]

```
pd.read_excel(df)
```

```

      Order ID Order Date  Ship Date
EmailID \
0      CA-2013-138688 2013-06-13 2013-06-17  DarrinVanHuff@gmail.com
1      CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
2      CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
3      CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
4      CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
...      ...      ...      ...      ...

```

3198	CA-2013-125794	2013-09-30	2013-10-04	MarisLaWare@gmail.com
3199	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3200	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3201	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3202	CA-2014-119914	2014-05-05	2014-05-10	ChrisCortes@gmail.com

	Geography	Category \
0	United States, Los Angeles, California	Labels
1	United States, Los Angeles, California	Furnishings
2	United States, Los Angeles, California	Art
3	United States, Los Angeles, California	Phones
4	United States, Los Angeles, California	Binders
...	...	...
3198	United States, Los Angeles, California	Accessories
3199	United States, Costa Mesa, California	Furnishings
3200	United States, Costa Mesa, California	Phones
3201	United States, Costa Mesa, California	Paper
3202	United States, Westminster, California	Appliances

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860
7		
2	Newell 322	7.280
4		
3	Mitel 5320 IP Phone VoIP phone	907.152
4		
4	DXL Angle-View Binders with Locking Rings by S...	18.504
3		
...	...	...
...		
3198	Memorex Mini Travel Drive 64 GB USB 2.0 Flash ...	36.240
1		
3199	Tenex B1-RE Series Chair Mats for Low Pile Car...	91.960
2		
3200	Aastra 57i VoIP phone	258.576
2		
3201	It's Hot Message Books with Stickers, 2 3/4" x 5"	29.600
4		
3202	Acco 7-Outlet Masterpiece Power Center, Wihtou...	243.160
2		

	Profit
0	6.8714
1	14.1694
2	1.9656
3	90.7152
4	5.7825
...	...
3198	15.2208
3199	15.6332
3200	19.3932
3201	13.3200
3202	72.9480

[3203 rows x 10 columns]

```
data=pd.read_excel(df)
data.head(10)
```

	Order ID	Order Date	Ship Date	EmailID \
0	CA-2013-138688	2013-06-13	2013-06-17	DarrinVanHuff@gmail.com
1	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
2	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
3	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
4	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
5	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
6	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
7	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
8	CA-2013-161389	2013-12-06	2013-12-11	IreneMaddox@gmail.com
9	CA-2011-167164	2011-05-13	2011-05-15	AlejandroGrove@gmail.com

	Geography	Category \
0	United States,Los Angeles,California	Labels
1	United States,Los Angeles,California	Furnishings
2	United States,Los Angeles,California	Art
3	United States,Los Angeles,California	Phones
4	United States,Los Angeles,California	Binders
5	United States,Los Angeles,California	Appliances
6	United States,Los Angeles,California	Tables
7	United States,Los Angeles,California	Phones
8	United States,Seattle,Washington	Binders
9	United States,West Jordan,Utah	Storage

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860
7		
2	Newell 322	7.280
4		

```

3           Mitel 5320 IP Phone VoIP phone    907.152
4
4   DXL Angle-View Binders with Locking Rings by S...    18.504
3
5           Belkin F5C206VTEL 6 Outlet Surge    114.900
5
6           Chromcraft Rectangular Conference Tables    1706.184
9
7   Konftel 250 ConferenceÂ phoneÂ - Charcoal black    911.424
4
8           Fellowes PB200 Plastic Comb Binding Machine    407.976
3
9           Fellowes Super Stor/Drawer    55.500
2

```

```

Profit
0    6.8714
1   14.1694
2    1.9656
3   90.7152
4    5.7825
5   34.4700
6   85.3092
7   68.3568
8  132.5922
9    9.9900

```

```

df="C:/Users/hp/OneDrive/Desktop/New folder/Amazon 2_Raw.xlsx"
data=pd.read_excel(df)
print(data.shape)

(3203, 10)

```

#Data cleaning

```
data.info
```

```

<bound method DataFrame.info of
Date      EmailID \
0   CA-2013-138688 2013-06-13 2013-06-17  DarrinVanHuff@gmail.com
1   CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
2   CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
3   CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
4   CA-2011-115812 2011-06-09 2011-06-14  BrosinaHoffman@gmail.com
...

```

3198	CA-2013-125794	2013-09-30	2013-10-04	MarisLaWare@gmail.com
3199	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3200	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3201	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3202	CA-2014-119914	2014-05-05	2014-05-10	ChrisCortes@gmail.com

	Geography	Category \
0	United States, Los Angeles, California	Labels
1	United States, Los Angeles, California	Furnishings
2	United States, Los Angeles, California	Art
3	United States, Los Angeles, California	Phones
4	United States, Los Angeles, California	Binders
...	...	...
3198	United States, Los Angeles, California	Accessories
3199	United States, Costa Mesa, California	Furnishings
3200	United States, Costa Mesa, California	Phones
3201	United States, Costa Mesa, California	Paper
3202	United States, Westminster, California	Appliances

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860
7		
2	Newell 322	7.280
4		
3	Mitel 5320 IP Phone VoIP phone	907.152
4		
4	DXL Angle-View Binders with Locking Rings by S...	18.504
3		
...	...	...
...		
3198	Memorex Mini Travel Drive 64 GB USB 2.0 Flash ...	36.240
1		
3199	Tenex B1-RE Series Chair Mats for Low Pile Car...	91.960
2		
3200	Aastra 57i VoIP phone	258.576
2		
3201	It's Hot Message Books with Stickers, 2 3/4" x 5"	29.600
4		
3202	Acco 7-Outlet Masterpiece Power Center, Wihtou...	243.160
2		

	Profit
0	6.8714
1	14.1694
2	1.9656
3	90.7152
4	5.7825
...	...
3198	15.2208
3199	15.6332
3200	19.3932
3201	13.3200
3202	72.9480

```
[3203 rows x 10 columns]>
```

```
data.columns
```

```
Index(['Order ID', 'Order Date', 'Ship Date', 'EmailID', 'Geography',
      'Category', 'Product Name', 'Sales', 'Quantity', 'Profit'],
      dtype='object')
```

```
data.isnull().sum()
```

Order ID	0
Order Date	0
Ship Date	0
EmailID	0
Geography	0
Category	0
Product Name	0
Sales	0
Quantity	0
Profit	0

dtype: int64

```
data.insert(9,"Revenue",data.Profit + data.Sales)
```

```
data.insert(10,"percentage",data.Revenue/data.Profit)
```

```
data["percentage"]=data["percentage"].round(2)
```

```
data
```

	Order ID	Order Date	Ship Date	EmailID \
0	CA-2013-138688	2013-06-13	2013-06-17	DarrinVanHuff@gmail.com
1	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
2	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
3	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com

4	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
---	----------------	------------	------------	--------------------------

...	...	...	...	...
-----	-----	-----	-----	-----

3198	CA-2013-125794	2013-09-30	2013-10-04	MarisLaWare@gmail.com
------	----------------	------------	------------	-----------------------

3199	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
------	----------------	------------	------------	----------------------

3200	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
------	----------------	------------	------------	----------------------

3201	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
------	----------------	------------	------------	----------------------

3202	CA-2014-119914	2014-05-05	2014-05-10	ChrisCortes@gmail.com
------	----------------	------------	------------	-----------------------

		Geography	Category \
0	United States, Los Angeles, California		Labels
1	United States, Los Angeles, California		Furnishings
2	United States, Los Angeles, California		Art
3	United States, Los Angeles, California		Phones
4	United States, Los Angeles, California		Binders

...	...	...
3198	United States, Los Angeles, California	Accessories
3199	United States, Costa Mesa, California	Furnishings
3200	United States, Costa Mesa, California	Phones
3201	United States, Costa Mesa, California	Paper
3202	United States, Westminster, California	Appliances

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860
7		
2	Newell 322	7.280
4		
3	Mitel 5320 IP Phone VoIP phone	907.152
4		
4	DXL Angle-View Binders with Locking Rings by S...	18.504
3		
...	...	...
...		
3198	Memorex Mini Travel Drive 64 GB USB 2.0 Flash ...	36.240
1		
3199	Tenex B1-RE Series Chair Mats for Low Pile Car...	91.960
2		
3200	Aastra 57i VoIP phone	258.576
2		
3201	It's Hot Message Books with Stickers, 2 3/4" x 5"	29.600



```
4
3202 Acco 7-Outlet Masterpiece Power Center, Wihtou... 243.160
2
```

	Revenue	percentage	Profit
0	21.4914	3.13	6.8714
1	63.0294	4.45	14.1694
2	9.2456	4.70	1.9656
3	997.8672	11.00	90.7152
4	24.2865	4.20	5.7825
...	...	...	...
3198	51.4608	3.38	15.2208
3199	107.5932	6.88	15.6332
3200	277.9692	14.33	19.3932
3201	42.9200	3.22	13.3200
3202	316.1080	4.33	72.9480

```
[3203 rows x 12 columns]
```

```
data.describe()
```

	Order Date	Ship Date \
count	3203	3203
mean	2013-05-10 03:06:07.530440192	2013-05-14 01:25:25.195129600
min	2011-01-07 00:00:00	2011-01-09 00:00:00
25%	2012-05-22 00:00:00	2012-05-26 00:00:00
50%	2013-07-22 00:00:00	2013-07-25 00:00:00
75%	2014-05-23 00:00:00	2014-05-27 00:00:00
max	2014-12-31 00:00:00	2015-01-06 00:00:00
std	NaN	NaN

	Sales	Quantity	Revenue	percentage
Profit				
count	3203.000000	3203.000000	3203.000000	3203.00
3203.000000				
mean	226.493233	3.828910	260.342265	inf
33.849032				
min	0.990000	1.000000	-849.995000	-84.00 -
3399.980000				
25%	19.440000	2.000000	24.507400	3.17
3.852000				
50%	60.840000	3.000000	72.696000	4.03
11.166400				
75%	215.809000	5.000000	248.690000	6.26
33.000400				
max	13999.960000	14.000000	20719.940800	inf
6719.980800				
std	524.876877	2.260947	651.268138	NaN
174.109081				

```
data.isnull().sum()
```

```
Order ID      0
Order Date    0
Ship Date     0
EmailID       0
Geography     0
Category      0
Product Name  0
Sales         0
Quantity      0
Revenue       0
percentage    0
Profit        0
dtype: int64
```

```
data.dropna()
```

	Order ID	Order Date	Ship Date	
EmailID \				
0	CA-2013-138688	2013-06-13	2013-06-17	DarrinVanHuff@gmail.com
1	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
2	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
3	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
4	CA-2011-115812	2011-06-09	2011-06-14	BrosinaHoffman@gmail.com
...	...	...	...	...
3198	CA-2013-125794	2013-09-30	2013-10-04	MarisLaWare@gmail.com
3199	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3200	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3201	CA-2014-121258	2014-02-27	2014-03-04	DaveBrooks@gmail.com
3202	CA-2014-119914	2014-05-05	2014-05-10	ChrisCortes@gmail.com

	Geography	Category \
0	United States,Los Angeles,California	Labels
1	United States,Los Angeles,California	Furnishings
2	United States,Los Angeles,California	Art
3	United States,Los Angeles,California	Phones
4	United States,Los Angeles,California	Binders
...	...	...
3198	United States,Los Angeles,California	Accessories

3199	United States,Costa Mesa,California	Furnishings
3200	United States,Costa Mesa,California	Phones
3201	United States,Costa Mesa,California	Paper
3202	United States,Westminster,California	Appliances

	Product Name	Sales
Quantity \		
0	Self-Adhesive Address Labels for Typewriters b...	14.620
2		
1	Eldon Expressions Wood and Plastic Desk Access...	48.860
7		
2	Newell 322	7.280
4		
3	Mitel 5320 IP Phone VoIP phone	907.152
4		
4	DXL Angle-View Binders with Locking Rings by S...	18.504
3		
...	...	...
...		
3198	Memorex Mini Travel Drive 64 GB USB 2.0 Flash ...	36.240
1		
3199	Tenex B1-RE Series Chair Mats for Low Pile Car...	91.960
2		
3200	Aastra 57i VoIP phone	258.576
2		
3201	It's Hot Message Books with Stickers, 2 3/4" x 5"	29.600
4		
3202	Acco 7-Outlet Masterpiece Power Center, Wihtou...	243.160
2		

	Revenue	percentage	Profit
0	21.4914	3.13	6.8714
1	63.0294	4.45	14.1694
2	9.2456	4.70	1.9656
3	997.8672	11.00	90.7152
4	24.2865	4.20	5.7825
...	...	...	...
3198	51.4608	3.38	15.2208
3199	107.5932	6.88	15.6332
3200	277.9692	14.33	19.3932
3201	42.9200	3.22	13.3200
3202	316.1080	4.33	72.9480

[3203 rows x 12 columns]

```
nan_values = data.isna().any()
print(nan_values)
```

Order ID	False
Order Date	False

```

Ship Date      False
EmailID        False
Geography      False
Category       False
Product Name   False
Sales          False
Quantity       False
Revenue        False
percentage     False
Profit         False
dtype: bool

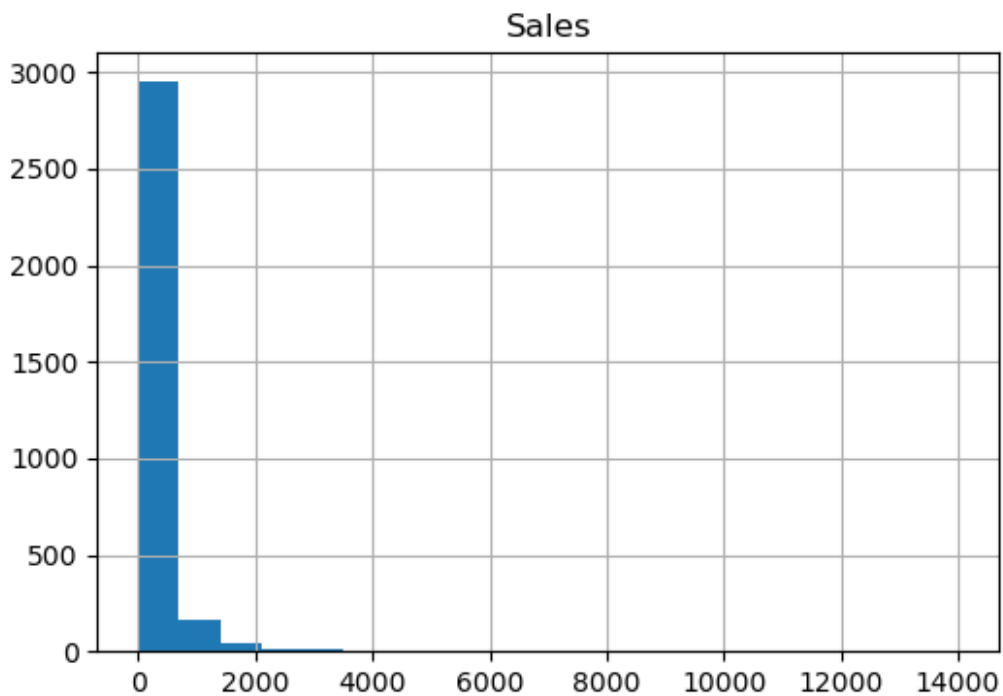
```

```

hist = ["Sales"]
data[hist].hist(bins = 20 ,figsize=(6,4))

array([[<Axes: title={'center': 'Sales'}>]], dtype=object)

```

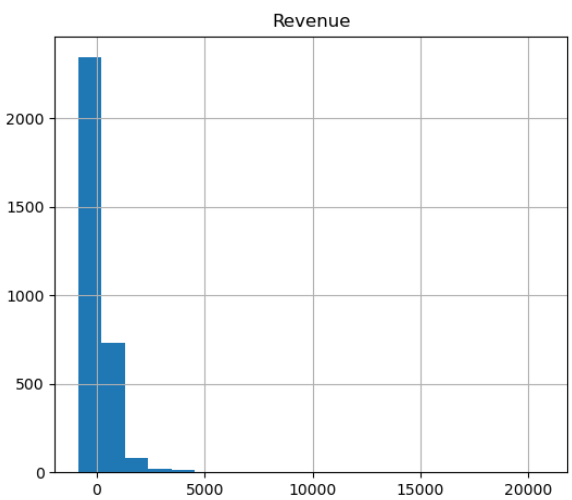
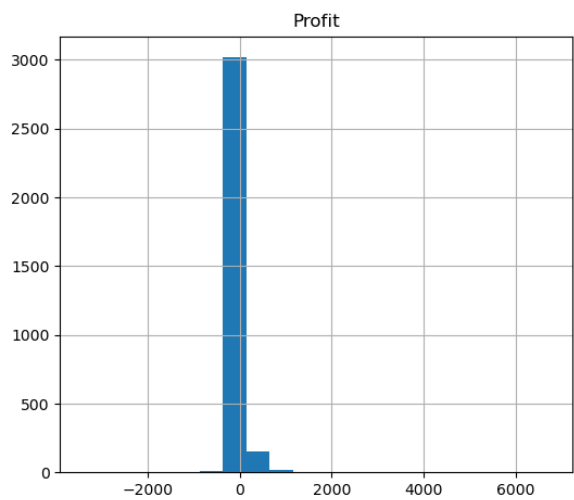
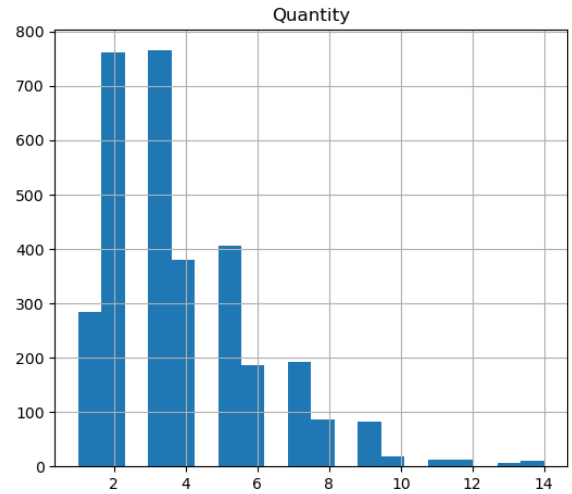
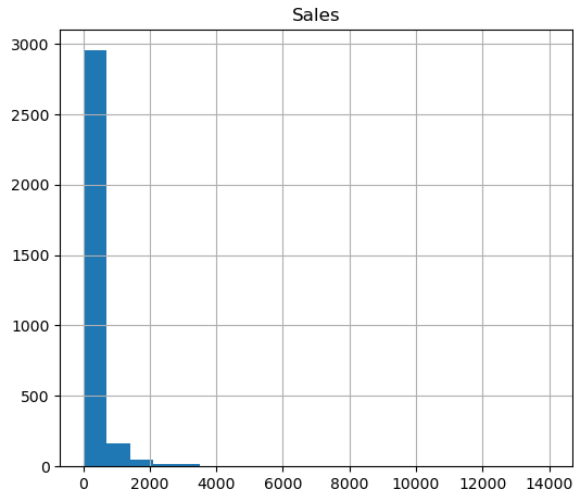


```

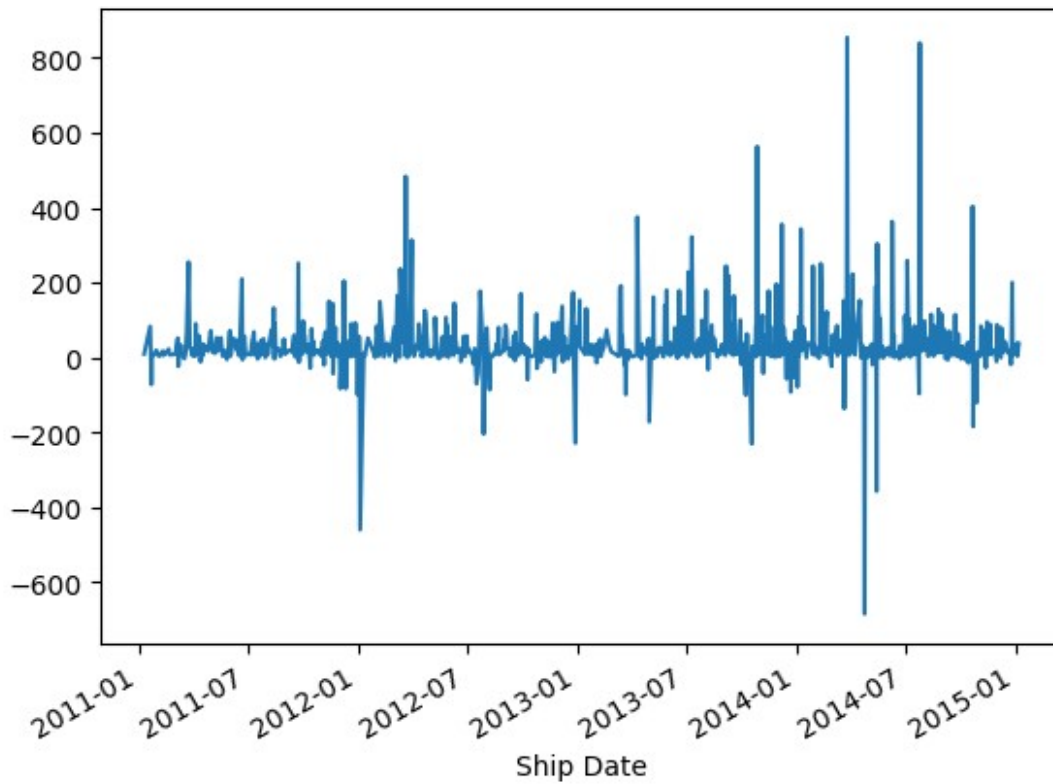
data1 = data[[ 'Sales', 'Quantity', 'Profit','Revenue']]
data1.hist(bins =20 ,figsize = (14,12) )

array([[<Axes: title={'center': 'Sales'}>,
        <Axes: title={'center': 'Quantity'}>],
       [<Axes: title={'center': 'Profit'}>,
        <Axes: title={'center': 'Revenue'}>]], dtype=object)

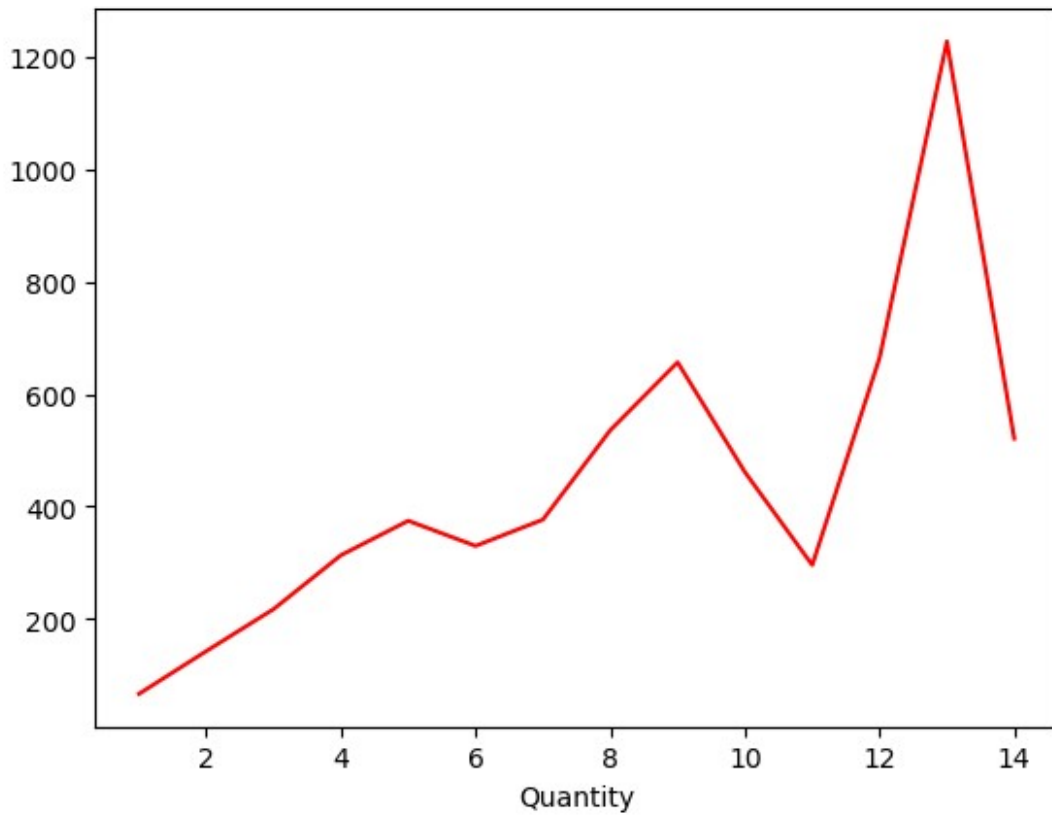
```



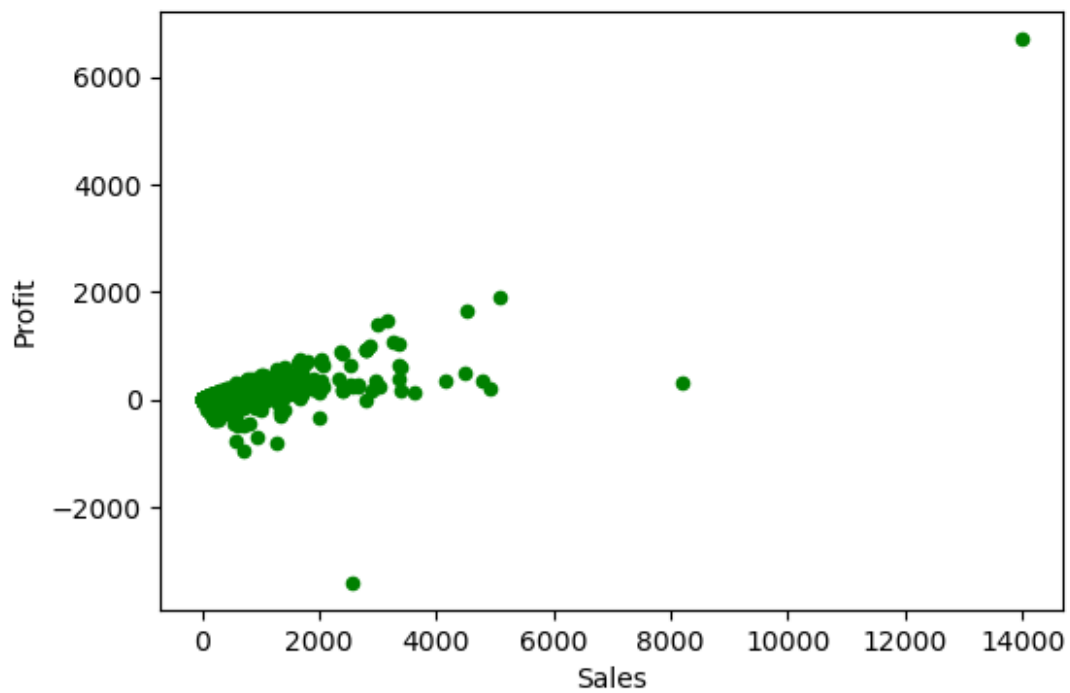
```
data2 = data.groupby('Ship Date')['Profit'].mean()
data2.plot(kind = 'line')
<Axes: xlabel='Ship Date'>
```



```
data3 = data.groupby('Quantity')['Revenue'].mean()  
data3.plot(kind = 'line', color = 'red')  
<Axes: xlabel='Quantity'>
```

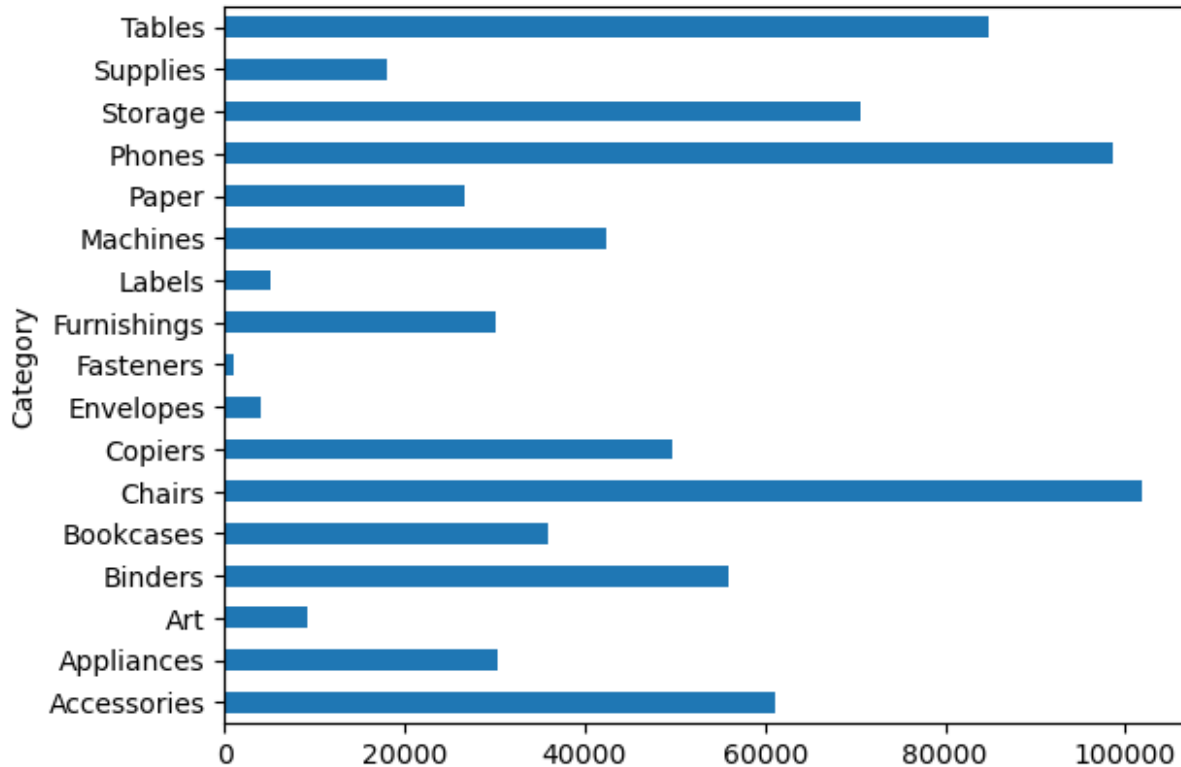


```
data4 = data.plot.scatter(x = 'Sales', y = 'Profit', c = 'green',  
figsize = (6 , 4))
```



```
data5 = data.groupby('Category')['Sales'].sum()
data5.plot.barh(x = 'Sales', y = 'Category',)
```

```
<Axes: ylabel='Category'>
```



```
data.dtypes
```

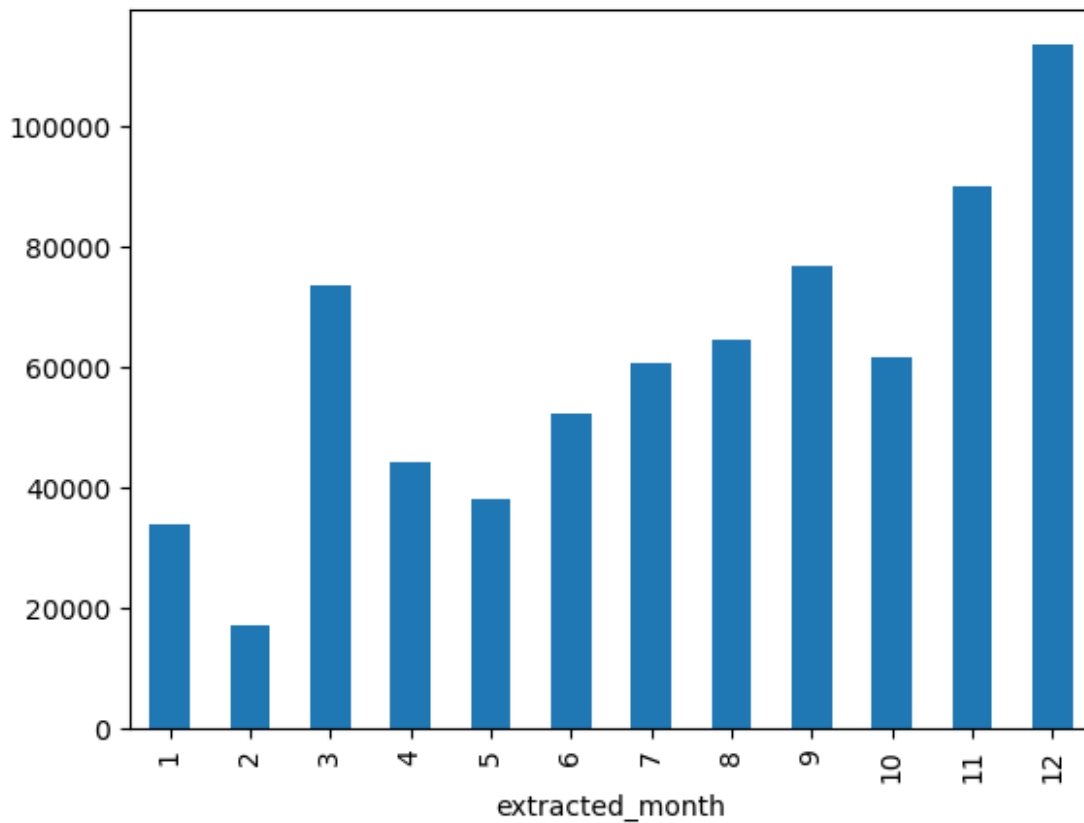
```
Order ID      object
Order Date    datetime64[ns]
Ship Date     datetime64[ns]
EmailID       object
Geography     object
Category      object
Product Name  object
Sales         float64
Quantity      int64
Revenue       float64
percentage    float64
Profit        float64
dtype: object
```

```
data['extracted_month'] = data['Ship Date'].dt.month
```

```
data5 = data.groupby('extracted_month')['Sales'].sum()
data5.plot(kind = 'bar')
```



<Axes: xlabel='extracted\_month'>



data5

extracted\_month

```
1    33734.6810
2    17133.0910
3    73340.4230
4    44188.4990
5    38094.0985
6    52138.6035
7    60533.0220
8    64534.4150
9    76679.2090
10   61527.3915
11   90061.7175
12  113492.6735
```

Name: Sales, dtype: float64

data5.index

```
Index([1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12], dtype='int32',
      name='extracted_month')
```

```
data5.values
```

```
array([ 33734.681 , 17133.091 , 73340.423 , 44188.499 ,
        38094.0985,
         52138.6035, 60533.022 , 64534.415 , 76679.209 ,
        61527.3915,
         90061.7175, 113492.6735])
```

```
data_data = {
    'extracted_month': data5.index,
    'sales': data5.values
}
```

```
data5 = pd.DataFrame(data_data)
```

```
data5
```

	extracted_month	sales
0	1	33734.6810
1	2	17133.0910
2	3	73340.4230
3	4	44188.4990
4	5	38094.0985
5	6	52138.6035
6	7	60533.0220
7	8	64534.4150
8	9	76679.2090
9	10	61527.3915
10	11	90061.7175
11	12	113492.6735

```
index_to_month = {
    1: 'jan', 2: 'feb', 3: 'mar', 4: 'apr', 5: 'may', 6: 'jun', 7: 'jul', 8: 'aug', 9: 'sep',
    10: 'oct', 11: 'nov', 12: 'dec'
}
```

```
data5.extracted_month = data5.extracted_month.map(index_to_month)
```

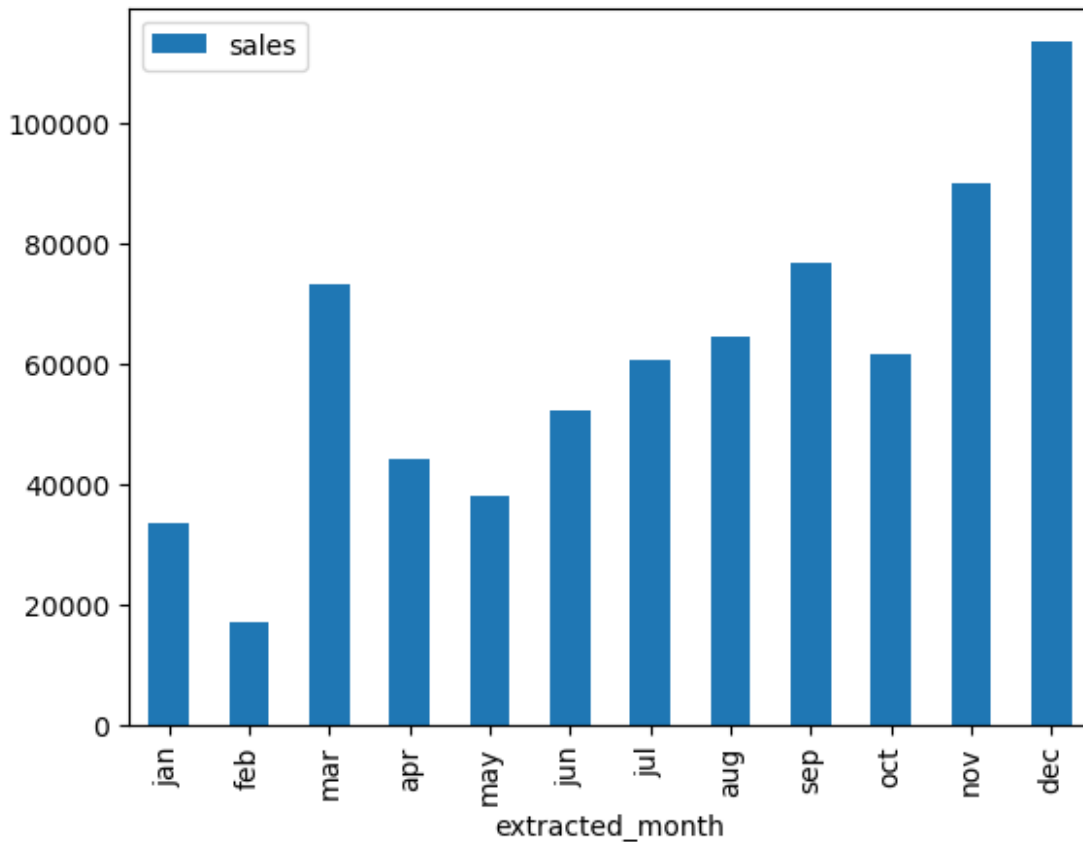
```
data5
```

	extracted_month	sales
0	jan	33734.6810
1	feb	17133.0910
2	mar	73340.4230
3	apr	44188.4990
4	may	38094.0985
5	jun	52138.6035
6	jul	60533.0220
7	aug	64534.4150
8	sep	76679.2090
9	oct	61527.3915

```
10          nov    90061.7175
11          dec   113492.6735
```

```
data5.plot(kind='bar', x='extracted_month', y='sales')
```

```
<Axes: xlabel='extracted_month'>
```



```
data6 = data.groupby('Category')['Profit'].sum().sort_values(ascending = False).head(5)
```

```
data6
```

```
Category
```

```
Copiers      19327.2351
```

```
Accessories  16484.5983
```

```
Binders      16096.8016
```

```
Paper        12119.2364
```

```
Phones       9110.7426
```

```
Name: Profit, dtype: float64
```

```
data6.plot(kind='pie', autopct='%1.1f%%', startangle=90)
```

```
<Axes: ylabel='Profit'>
```

