

Fill valid code/values in place of blanks.

In [1]:

```
# import required libraries
import pandas as pd
import numpy as np
```

In [2]:

```
scores = [29,27,14,23,29,10]

# find the mean of all items of the list 'scores'
np.mean(scores)
```

Out[2]:

22.0

In [3]:

```
# find the median of all items of the list 'scores'
np.median(scores)
```

Out[3]:

25.0

In [4]:

```
from statistics import mode

fruits = ['apple', 'grapes', 'orange', 'apple']

# find mode of the list 'fruits'
mode(fruits)
```

Out[4]:

'apple'

In [5]:

```
from random import sample
data = sample(range(1,100),50) # generating a list 50 random integers

# find variance of data
np.var(data)
```

Out[5]:

837.4416

In [6]:

```
# find standard deviation
np.std(data)
```

Out[6]:

28.93858324106417

**Please download the file "data\_statistics.csv".**

In [9]:

```
# read data_python.csv using pandas
mydata = pd.read_csv("data_statistics.csv")
```

In [10]:

```
# print first few rows of mydata
mydata.head()
```

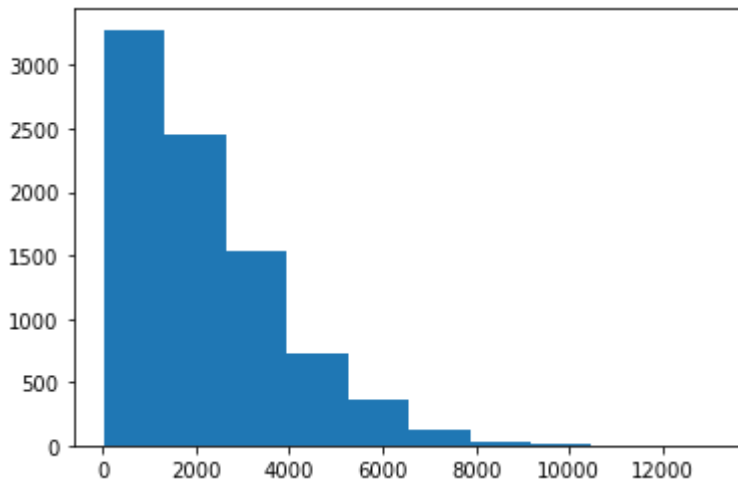
Out[10]:

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_I
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	



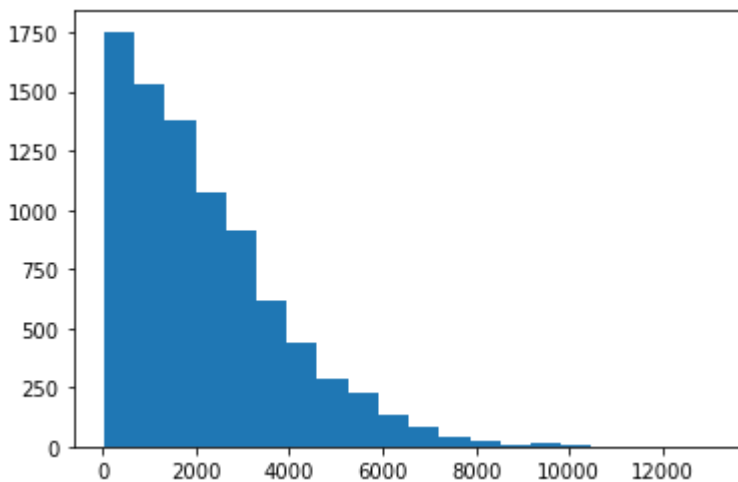
In [12]:

```
# plot histogram for 'Item_Outlet_Sales'
import matplotlib.pyplot as plt
plt.hist(mydata['Item_Outlet_Sales'])
plt.show()
```



In [13]:

```
# increadse no. of bins to 20
plt.hist(mydata['Item_Outlet_Sales'], bins=20)
plt.show()
```



In [14]:

```
# find mean and median of 'Item_Weight'
np.mean(mydata['Item_MRP']), np.median(mydata['Item_MRP'])
```

Out[14]:

(140.9927819781768, 143.0128)

In [15]:

```
# find mode of 'Outlet_Size'
mydata['Outlet_Size'].mode()
```

Out[15]:

```
0    Medium
dtype: object
```

In [16]:

```
# frequency table of 'Outlet_Type'
mydata['Outlet_Type'].value_counts()
```

Out[16]:

```
Supermarket Type1    5577
Grocery Store        1083
Supermarket Type3     935
Supermarket Type2     928
Name: Outlet_Type, dtype: int64
```

In [17]:

```
# mean of 'Item_Outlet_Sales' for 'Supermarket Type2' outlet type
np.mean(mydata['Item_Outlet_Sales'][mydata['Outlet_Type'] == 'Supermarket Type2'])
```

Out[17]:

```
1995.4987392241392
```

In [18]:

```
# mean of 'Item_Outlet_Sales' for 'Supermarket Type3' outlet type
np.mean(mydata['Item_Outlet_Sales'][mydata['Outlet_Type'] == 'Supermarket Type3'])
```

Out[18]:

```
3694.038557647059
```

In [19]:

```
# 2 sample independent t-test
from scipy import stats
stats.ttest_ind(mydata['Item_Outlet_Sales'][mydata['Outlet_Type'] == 'Supermarket Type2'],
```

Out[19]:

```
Ttest_indResult(statistic=-20.442923116350805, pvalue=5.856140005446105e-84)
```

In [ ]:

