Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

TUTORIAL NO.2

Perform the following operations using Python on the data sets Compute and display summary statistics for each feature available in the dataset. (e.g. minimum value, maximum value, mean, range, standard deviation, variance and percentiles), Create a histogram for each feature in the dataset to illustrate the feature distributions

---

```
[27]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt

[4]: df=pd.read_csv("nba.csv")
```

Read Csv File

```
[5]: df.head()
```

[5]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25	2-Jun	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25	6-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27	5-Jun	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22	5-Jun	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29	10-Jun	231	NaN	5000000.0

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

[7]:

	Number	Age	Weight	Salary
count	455.000000	457.000000	457.000000	4.450000e+02
mean	17.668132	26.938731	221.522976	4.853190e+06
std	15.945315	4.404016	26.368343	5.230408e+06

[7]:

	Number	Age	Weight	Salary
count	455.000000	457.000000	457.000000	4.450000e+02
mean	17.668132	26.938731	221.522976	4.853190e+06
std	15.945315	4.404016	26.368343	5.230408e+06
min	0.000000	19.000000	161.000000	3.088800e+04
25%	5.000000	24.000000	200.000000	1.074169e+06
50%	13.000000	26.000000	220.000000	2.841960e+06
75%	25.000000	30.000000	240.000000	6.500000e+06
max	99.000000	40.000000	307.000000	2.500000e+07

finding minimum Value

```
[8]: df["Number"].min()

[8]: 0.0
```

Finding Maximum Value

```
[9]: df["Number"].max()

[9]: 99.0
```

Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

Finding Mean

```
df["Weight"].mean()
```

```
221.52297592997812
```

Finding The Varience

```
df["Weight"].var()
```

```
695.2894928780854
```

```
df.describe()
```

	Number	Age	Weight	Salary
count	455.000000	457.000000	457.000000	4.450000e+02
mean	17.668132	26.938731	221.522976	4.853190e+06
std	15.945315	4.404016	26.368343	5.230408e+06
min	0.000000	19.000000	161.000000	3.088800e+04
25%	5.000000	24.000000	200.000000	1.074169e+06
50%	13.000000	26.000000	220.000000	2.841960e+06
75%	25.000000	30.000000	240.000000	6.500000e+06
max	99.000000	40.000000	307.000000	2.500000e+07

Finding The Info

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 457 entries, 0 to 456
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      457 non-null    object
 1   Team      457 non-null    object
 2   Number    455 non-null    float64
 3   Position  457 non-null    object
 4   Age       457 non-null    int64
 5   Height    457 non-null    object
 6   Weight    457 non-null    int64
 7   College   373 non-null    object
 8   Salary    445 non-null    float64
dtypes: float64(2), int64(2), object(5)
memory usage: 32.3+ KB
```

Range

```
numericdf.select_dtypes(include=[np.number]).columns#columns in the numeric variable of number
```

```
numeric
```

```
Index(['Number', 'Age', 'Weight', 'Salary'], dtype='object')
```

```
rangedf[numeric].min()-df[numeric].max()
```

```
ran
```

Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

```
[41]: ran=df[numeric].min()-df[numeric].max()

[44]: ran

[44]: Number         -99.0
      Age            -21.0
      Height        -146.0
      Salary   -24969112.0
      dtype: float64

[11]: #bins represent no. of the column
      #alpha means transparent Visible
```

Standard Deviation

```
[50]: np.std(df['Weight'])

[50]: 26.339477437280433
```

Percentile

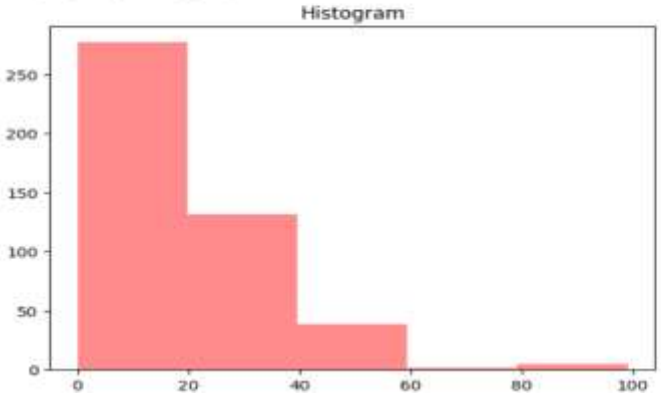
```
[52]: np.percentile(df['Weight'],60)

[52]: 230.0
```

Histogram of The Number

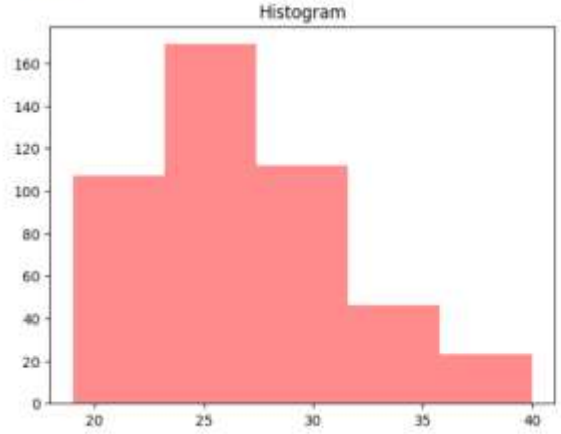
```
[18]: plt.hist(df['Number'],bins=5,alpha=0.45,color="red")
      plt.title("Histogram")
      plt.show()

[18]: Text(0.5, 1.0, 'Histogram')
```



Finding The Age Histogram

```
[23]: plt.hist(df['Age'],bins=5,alpha=0.45,color="red")
      plt.title("Histogram")
      plt.show()
```



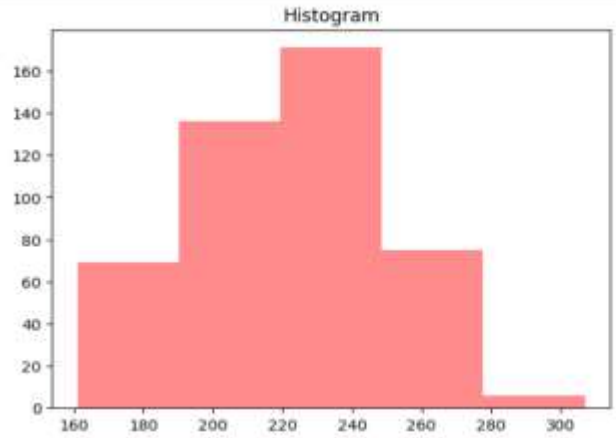
Name: Abhishek Mali
Roll No.: 213070
PRN No.: 22320024

```
df.head()
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25	2-Jun	180	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25	6-Jun	235	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27	5-Jun	205	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22	5-Jun	185	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29	10-Jun	231	NaN	5000000.0

Finding The Weight Histogram

```
plt.hist(df["Weight"],bins=5,alpha=0.45,color="red")
plt.title("Histogram")
plt.show()
```



Finding The Histogram Of Salary

```
plt.hist(df["Salary"],bins=5,alpha=0.45,color="red")
plt.title("Histogram")
plt.show()
```

Finding The Histogram Of Salary

```
plt.hist(df["Salary"],bins=5,alpha=0.45,color="red")
plt.title("Histogram")
plt.show()
```