Importing necessary modules

In [61]:
```
import pandas as pd
import statistics as st
import matplotlib.pyplot as plt
from scipy.stats import skew
from scipy.stats import kurtosis
import seaborn as sns
```

In [12]:
```
data=pd.read_csv(r"K:\Desktop\NIIT\tables\DS1_C4_S5_Students_Scores_Data_Practice.csv")
data
```

Out[12]:

	Name	Statistics	Python	Tableau
0	David	62	89	56
1	James	47	87	86
2	Robert	55	67	77
3	Thomas	74	55	45
4	Steven	31	47	73
5	Paul	77	72	62
6	Gary	85	76	74
7	Justin	63	79	89
8	Patrick	42	44	67
9	Tyler	32	99	67
10	Peter	71	99	97
11	Bryan	63	69	68

Task1: To use measures of central tendency to find which technology the students have performed really well

In [13]:
```
mean=[]
median=[]
mode=[]
for item in data.iloc[:,1:].columns:
    mean.append(data[item].mean())
    median.append(data[item].median())
    mode.append(st.mode(data[item]))

cols=["Statistics","Python","Tableau"]
measures=pd.DataFrame([mean,median,mode],columns=[cols],index=["Mean","Median","Mode"])
print("""The below table says that python has the highest of all central measures students have performed well in python""")
measures
```

```
The below table says that python has the highest of all central measures students have performed well in python
```

Out[13]:

	Statistics	Python	Tableau
Mean	58.5	73.583333	71.75
Median	62.5	74.000000	70.50
Mode	63.0	99.000000	67.00

Task2: To identify the measures of variablity to identify best technology the students learn

In [14]:
```
std=[]
var=[]
mean=[]
for item in data.iloc[:,1:].columns:
    mean.append(data[item].mean())
    std.append(data[item].std())
    var.append(data[item].var())
var_measures=pd.DataFrame([mean,std,var],columns=cols,index=["Mean","Standard dev","Variance"])

var_measures

```

Out[14]:

	Statistics	Python	Tableau
Mean	58.500000	73.583333	71.750000
Standard dev	17.500649	18.436418	14.429295
Variance	306.272727	339.901515	208.204545

Task3: To identify outliers and skewness

In [21]:
```
outliers=[]
sk=[]
for item in data.iloc[:,1:].columns:
    q1=data[item].quantile(0.25)
    q3=data[item].quantile(0.75)
    IQR=q3-q1
    upp=1.5*IQR + q3
    low=q1-1.5*IQR
    sk.append(skew(data[item]))
    outs=[]
    for val in data[item]:
        if(val>upp or val<low):
            outs.append(val)
    outliers.append(outs)
print("From below table its clear that statistics has the most skewness and is left skewed")
pd.DataFrame([sk,outliers],columns=cols,index=["Skewness","Outliers"])


```

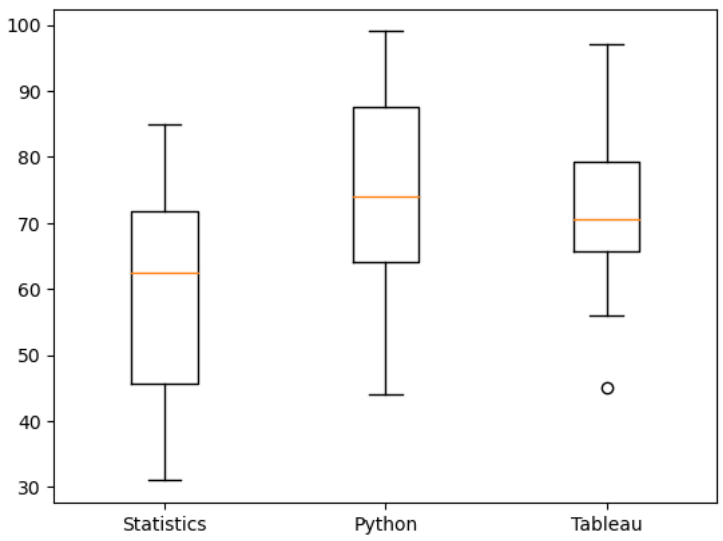
```
From below table its clear that statistics has the most skewness and is left skewed
```

Out[21]:

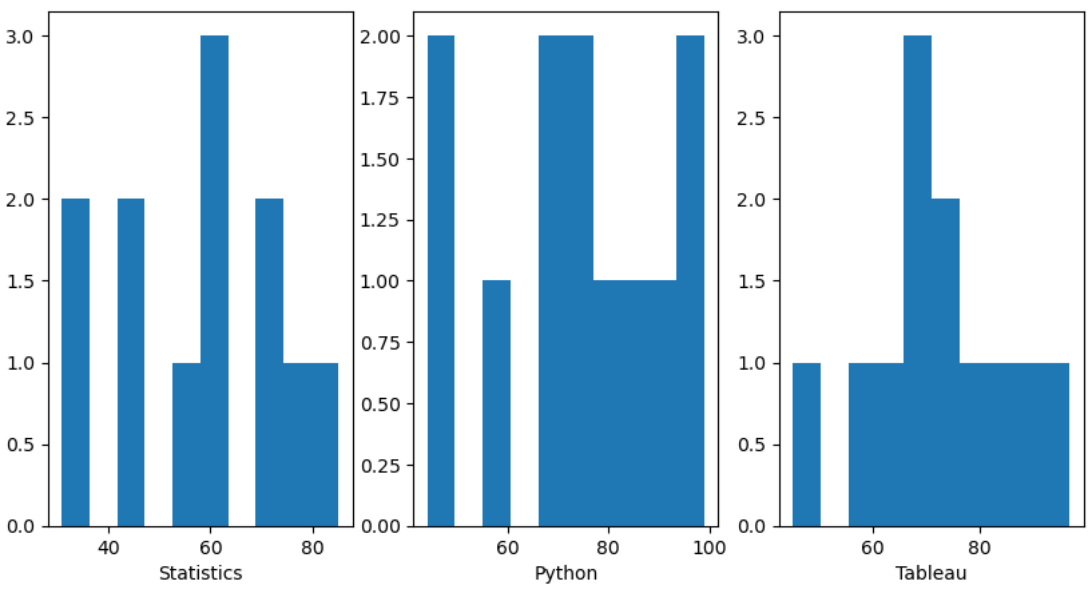
	Statistics	Python	Tableau
Skewness	-0.258974	-0.187711	0.001292
Outliers	[]	[]	[45]

Task4: To visually represet the distribution of scores

In [24]:
```
plt.boxplot(data.iloc[:,1:],labels=["Statistics","Python","Tableau"]);
```



In [44]:
```
fig,ax=plt.subplots(1,3,figsize=(10,5))
ax[0].hist(data["Statistics"])
ax[0].set_xlabel('Statistics')
ax[1].hist(data['Python'])
ax[1].set_xlabel('Python')
ax[2].hist(data['Tableau'])
ax[2].set_xlabel('Tableau')
plt.show()
```



Task5: To show kurtosis and plot the kde plot

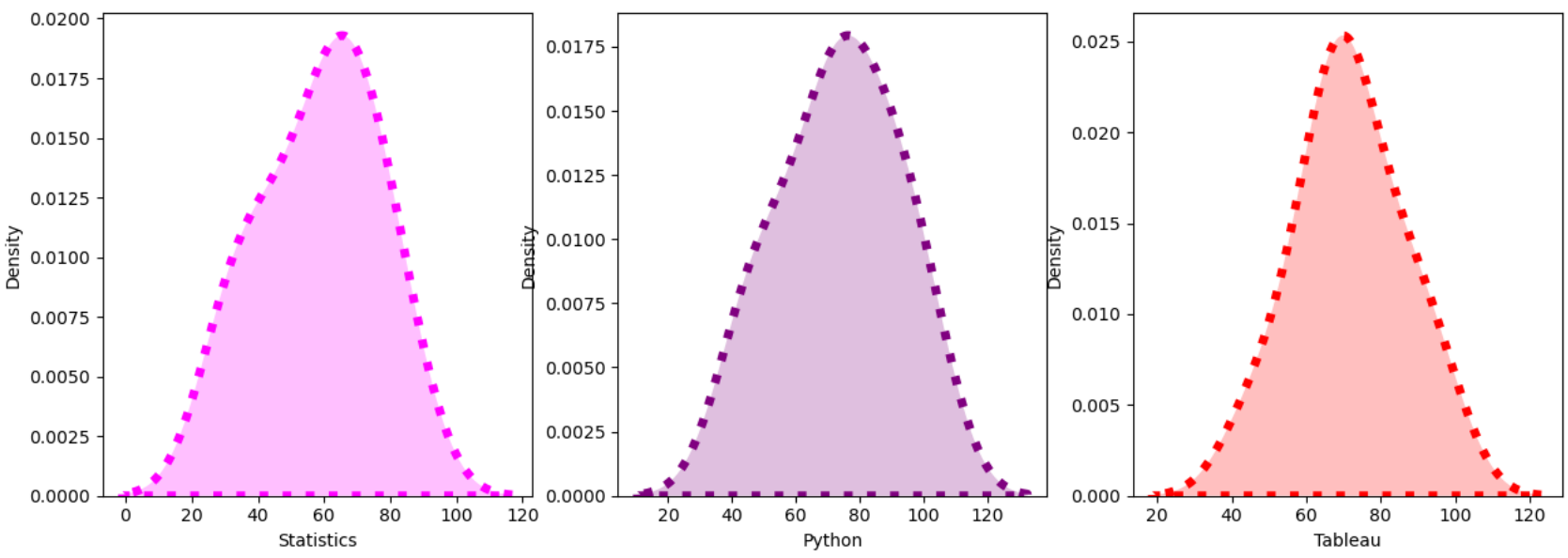
In [64]:
```
kurt=[]
for item in data.iloc[:,1:].columns:
    kurt.append(kurtosis(data[item]))
print("From below table its clear that all these values are platykurtic kurtosis")

pd.DataFrame([kurt],columns=cols,index=["Kurtosis"])
```

```
From below table its clear that all these values are platykurtic kurtosis
```

Out[64]:

	Statistics	Python	Tableau
Kurtosis	-1.043844	-1.015947	-0.435691

In [70]:
```
fig,ax=plt.subplots(1,3,figsize=(15,5))
sns.kdeplot((data.Statistics),color="magenta",fill=True,linestyle="dotted",linewidth=5,label="Actual",ax=ax[0])
sns.kdeplot((data.Python),color="Purple",fill=True,linestyle="dotted",linewidth=5,label="Actual",ax=ax[1])
sns.kdeplot((data.Tableau),color="red",fill=True,linestyle="dotted",linewidth=5,label="Actual",ax=ax[2])
plt.show()
```



Task6: Conclusion

In [71]:
```
"""The data provided has a decent spread in data from the observations of central measures and variation measures and none
of the columns contains except Tableau which has one outlier observable in both boxplot as well as Skewness table from
histogram we can observe that python learners have scored really well most of the data points lie between 75 and 100 and 99
being the mode. The tableau column shows more of a bell curve nature thus has the best distribution amongst all the data """
```

Out[71]:
```
'The data provided has a decent spread in data from the observations of central measures and variation measures and none\nof the columns contains except Tableau which has one outlier obse
rvable in both boxplot as well as Skewness table from \nhistogram we can observe that python learners have scored really well most of the data points lie between 75 and 100 and 99 \nbeing
the mode. The tableau column shows more of a bell curve nature thus has the best distribution amongst all the data '
```