

# SPPU-TE-COMP-CONTENT - KSKA Git

Q1) For the Iris Dataset, List down the Features and their types.

ANS. The Iris Dataset comprises 150 instances of Iris plant, each characterized by four continuous Feature.

1. Sepal Length.

The length of sepal in centimeters (cm's)

2. Sepal Width

The width of the sepal in centimeters (in cm's)

3. Petal Length.

The length of the Petal in centimeters.

4. Petal Width.

The width of the petal in centimeters.

The Iris dataset is one of the most iconic datasets in the world of Machine Learning and Data Science.

The Dataset contains information about three species of Iris Flower.

Q2) Write a code to create a Histogram for each Feature.

ANS. CODE:-

```
import numpy as np
```

```
import pandas as pd
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
dataset = sns.load_dataset('Iris')
```

```
dataset.head()
```

```
fig, axes = plt.subplots(1, 2, figsize = (10, 9))
```

```
sns.histplot(dataset['sepal length'], ax = axes[0,0])
```



# SPPU-TE-COMP-CONTENT - KSKA Git

```
sns.histplot (dataset ['sepal-width'], ax = axes [0,2])  
sns.histplot (dataset ['petal-length'], ax = axes [2,0])  
sns.histplot (dataset ['petal-width'], ax = axes [1,1])  
-- END
```

Q3.) Write a code to create a boxplot for each feature.

ANS.

```
CODE:-  
import numpy as np  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
dataset = sns.load_dataset ('Iris')  
fig, axes = plt.subplots (2, 2, figsize = (16, 9))  
sns.boxplot (y = 'petal-length', x = 'species', data = dataset, ax = axes [0,0])  
sns.boxplot (y = 'petal-width', x = 'species', data = dataset, ax = axes [0,1])  
sns.boxplot (y = 'sepal-length', x = 'species', data = dataset, ax = axes [1,0])  
sns.boxplot (y = 'sepal-width', x = 'species', data = dataset, ax = axes [1,1])  
-- END
```

Q4.) Identify the outliers from the boxplot drawn for Iris Dataset.

ANS.

- To identify the outliers, following are required.
- 1. IQR: It refers to the InterQuartile Range. The box represents the range between the 25th percentile ( $Q_1$ ) and the 75th percentile ( $Q_3$ ).
- 2. Whiskers: The lines extending from the box show the range of data within  $[Q_1 - 1.5 * IQR, Q_3 + 1.5 * IQR]$ .



# SPPU-TE-COMP-CONTENT - KSKA Git

0 STEPS TO IDENTIFY THE OUTLIERS:-

1. Calculate the IQR i.e. the difference between the  $Q_3$  and  $Q_1$
2. Define Bounds.
  - Lower Bounds :  $Q_1 - 1.5 \times IQR$
  - Upper Bounds :  $Q_3 + 1.5 \times IQR$
3. IF Any data is outside these bounds, those will be considered as outliers.