

Name : Prem Bansod
Roll number: 41310

Mock Assignment

Problem Statement:

Download the Iris flower dataset or any other dataset into a DataFrame. (eg <https://archive.ics.uci.edu/ml/datasets/Iris>) Use Python/R and Perform following

How many features are there and what are their types (e.g., numeric, nominal)?

Compute and display summary statistics for each feature available in the dataset. (eg. minimum value, maximum value, mean, range, standard deviation, variance and percentiles

Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram.

Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single plot. Compare distributions and identify outliers.

Code:

```
#!/usr/bin/env python3  
# -*- coding: utf-8 -*-  
"""
```

Created on Wed Dec 16 10:20:36 2020

```
@author: prem  
"""
```

```
import pandas as pd  
import matplotlib.pyplot as plt
```

```
dataset = pd.read_csv('iris.csv')
```

```
dataset.head()
```

```
print(dataset.shape[1]-1)
```

```
print(dataset.dtypes)
```

```
dataset.describe()
```

```
dataset.hist()
```

```
plt.show()
```

```
dataset[['sepal-length', 'sepal-width', 'petal-length', 'petal-width']].plot.hist(bins = 10, title='All features')  
plt.show()
```

```
dataset[['sepal-length']].plot.hist(bins = 10, title='sepal-length')  
plt.show()
```

```
dataset[['sepal-width']].plot.hist(bins = 10, title='sepal-width')  
plt.show()
```

```
dataset[['petal-length']].plot.hist(bins = 10, title='petal-length')  
plt.show()
```

```
dataset[['petal-width']].plot.hist(bins = 10, title='petal-width')  
plt.show()
```

```
dataset.plot.box(title = "All features")  
plt.show()
```

```
dataset[['sepal-length']].plot.box(title = 'sepal-length')  
plt.show()
```

```
dataset[['sepal-width']].plot.box(title = 'sepal-width')  
plt.show()
```

```
dataset[['petal-length']].plot.box(title = 'petal-length')  
plt.show()
```

```
dataset[['petal-width']].plot.box(title = 'petal-width')  
plt.show()
```

output:

```
Activities Spyder Wed 11:18 AM
IPython console
Console 1/A
Python 3.6.9 (default, Oct 8 2020, 12:12:24)
Type "copyright", "credits" or "license" for more information.

IPython 5.5.0 -- An enhanced Interactive Python.
?                -> Introduction and overview of IPython's features.
%quickref        -> Quick reference.
help             -> Python's own help system.
object?         -> Details about 'object', use 'object??' for extra details.

In [1]: import pandas as pd
...: import matplotlib.pyplot as plt
...:

In [2]: dataset = pd.read_csv('iris.csv')
...:

In [3]: dataset.head()
...:
Out[3]:
   sepal-length  sepal-width  petal-length  petal-width  class
0             5.1         3.5           1.4          0.2  Iris-setosa
1             4.9         3.0           1.4          0.2  Iris-setosa
2             4.7         3.2           1.3          0.2  Iris-setosa
3             4.6         3.1           1.5          0.2  Iris-setosa
4             5.0         3.6           1.4          0.2  Iris-setosa

In [4]: dataset.describe()
...:
Out[4]:
   sepal-length  sepal-width  petal-length  petal-width
count  150.000000  150.000000  150.000000  150.000000
mean     5.843333   3.054000   3.758667   1.198667
std     0.828066   0.433594   1.764420   0.763161
min     4.300000   2.000000   1.000000   0.100000
25%     5.100000   2.800000   1.600000   0.300000
50%     5.800000   3.000000   4.350000   1.300000
75%     6.400000   3.300000   5.100000   1.800000
max     7.900000   4.400000   6.900000   2.500000

In [5]: dataset.hist()
...: plt.show()
...:

Permissions: RW End-of-lines: LF Encoding: UTF-8 Line: 17 Column: 1 Memory: 38 %
```







