**Data Visualization III**

Download the Iris ﬂower dataset or any other dataset into a DataFrame. (e.g., <https://archive.ics.uci.edu/ml/datasets/Iris>). Scan the dataset and give the inference as:

List down the features and their types (e.g., numeric, nominal) available in the dataset.

Create a histogram for each feature in the dataset to illustrate the feature distributions.

Create a box plot for each feature in the dataset. Compare distributions and identify outliers.

In [10]:

# import pandas as pd import numpy as np import seaborn as sns

In [2]: df

**=**

# pd.read\_csv('Iris.csv')

In [ ]:

**pd.**

In [3]:

# df.head()

Out [3]:

**Species**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Id** | **SepalLengthCm** | **SepalWidthCm** | **PetalLengthCm** | **PetalWidthCm** |
| **0** 1 | 5.1 | 3.5 | 1.4 | 0.2 |
| **1** 2 | 4.9 | 3.0 | 1.4 | 0.2 |
| **2** 3 | 4.7 | 3.2 | 1.3 | 0.2 |
| **3** 4 | 4.6 | 3.1 | 1.5 | 0.2 |
| **4** 5 | 5.0 | 3.6 | 1.4 | 0.2 |

Iris- setosa

Iris- setosa

Iris- setosa

Iris- setosa

Iris- setosa

In [4]:

# df.info()

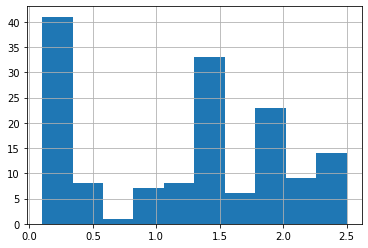
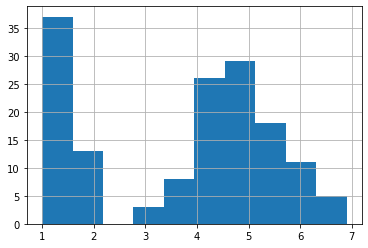
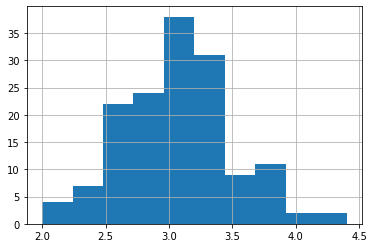
<class 'pandas.core.frame.DataFrame'> RangeIndex: 150 entries, 0 to 149 Data columns (total 6 columns):

# Column Non-Null Count Dtype

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **0** | **Id** | **150 non-null** |  | **int64** |
| **1** | **SepalLengthCm** | **150 non-null** |  | **float64** |
| **2** | **SepalWidthCm** | **150 non-null** |  | **float64** |
| **3** | **PetalLengthCm** | **150 non-null** |  | **float64** |

|  |  |  |
| --- | --- | --- |
| 1. **PetalWidthCm 150 non-null float64** 2. **Species 150 non-null object dtypes: float64(4), int64(1), object(1) memory usage: 7.2+ KB**   **In [16]: df.isnull().sum()**  **Out [16]: Id** **0**  **SepalLengthCm** **0**  **SepalWidthCm** **0**  **PetalLengthCm** **0**  **PetalWidthCm** **0**  **Species** **0**  **dtype: int64**  **In [5]: df['SepalLengthCm'].hist()**  **Out [5]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f2891370b20>**    **In [6]: df.hist()**  **Out [6]: array([[<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f2883dc91c0>,**  **<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f28838fa7c0>], [<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f2883926c70>,**  **<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f28838d3160>], [<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f288388b550>,**  **<matplotlib.axes.\_subplots.AxesSubplot object at 0x7f2883837a00>]], dtype=object)** | | |
| **In [7]:** | **df['SepalWidthCm'].hist()** |  |

Out [7]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f28837c8fa0>



In [8]:

**df['PetalLengthCm'].hist()**

Out [8]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f28836d2400>

In [9]:

**df['PetalWidthCm'].hist()**

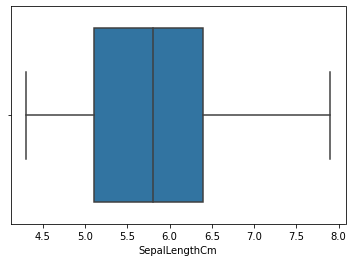
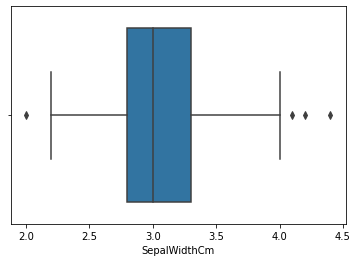
Out [9]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f28836d2700>

In [11]:

**sns.boxplot(df['SepalWidthCm'])**

/home/ihack-pc/.local/lib/python3.8/site- packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid

positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.



warnings.warn(

Out [11]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f2875d5ffa0>

In [12]:

**sns.boxplot(df['SepalLengthCm'])**

/home/ihack-pc/.local/lib/python3.8/site- packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

Out [12]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f2875d3da60>

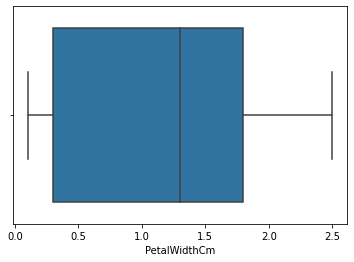
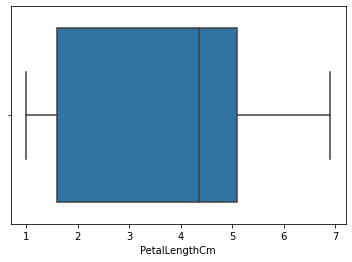
In [13]:

**sns.boxplot(df['PetalLengthCm'])**

/home/ihack-pc/.local/lib/python3.8/site- packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

Out [13]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f2875cd8730>



**from pandas\_profiling import ProfileReport**

**report**

**ProfileReport(df, title**

**"Sample Report")**

**report**

In [14]:

**sns.boxplot(df['PetalWidthCm'])**

/home/ihack-pc/.local/lib/python3.8/site- packages/seaborn/\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(

Out [14]: <matplotlib.axes.\_subplots.AxesSubplot at 0x7f2875d896a0>

In [ ]:

In [15]:

**=**

**=**

Summarize dataset: 0%| | 0/5 [00:00<?, ?it/s] Generate report structure: 0%| | 0/1 [00:00<?, ?it/s] Render HTML: 0%| | 0/1 [00:00<?, ?it/s]

Overview

|  |  |
| --- | --- |
| Dataset statistics |  |
| Number of variables | 6 |
| Number of observations | 150 |
| Missing cells | 0 |
| Missing cells (%) | 0.0% |
| Duplicate rows | 0 |
| Duplicate rows (%) | 0.0% |
| Total size in memory | 7.2 KiB |
| Average record size in memory  Variable types | 48.9 B |
| Numeric | 5 |
| Categorical | 1 |

Alerts

**Id** is highly correlated with **SepalLengthCm** and 2 other ﬁelds (SepalLengthCm, PetalLengthCm, PetalWidthCm)

**High correlation**

**SepalLengthCm** is highly correlated with **Id** and 2 other

**High correlation**

Out [15]:

In [ ]:

In [ ]: