Name : Chaitnaya More

Roll_no :- 01 "B"

Batch :- TB1-B2

Pratical no 10

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:

1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a boxplot for each feature in the dataset.
4. Compare distributions and identify outliers

In [2]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```python
df =pd.read_csv("iris_dataset.csv")
```

In [5]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
 4   species            150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```
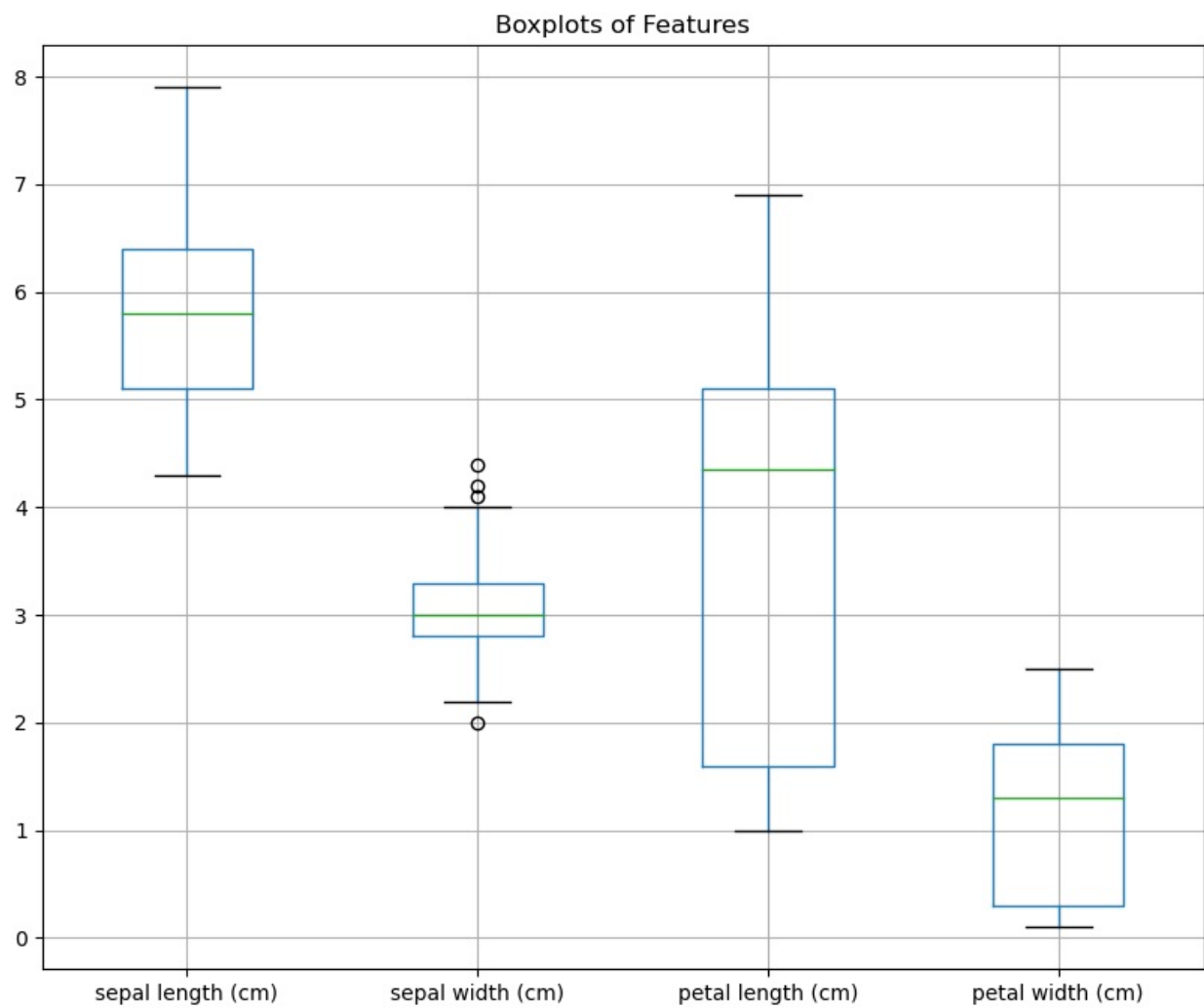
In [7]:
```python
df.hist(figsize=(10,8))
plt.suptitle('Histograms of Features')
plt.show()
```

Histograms of Features
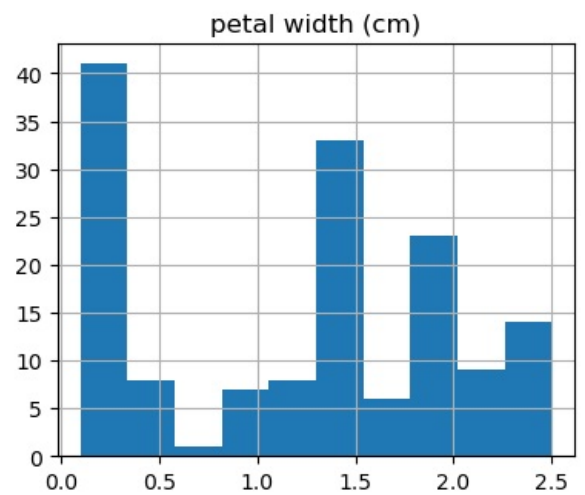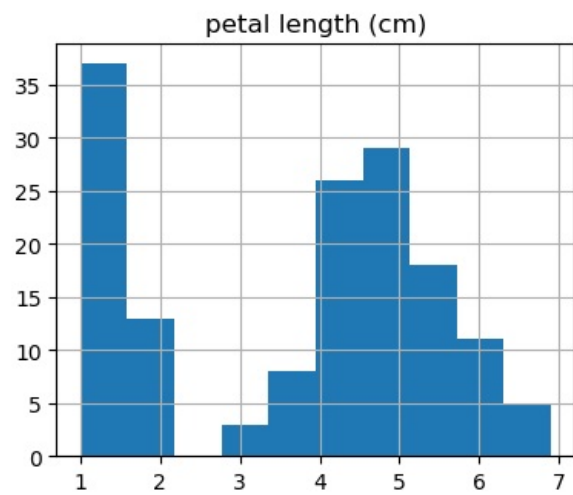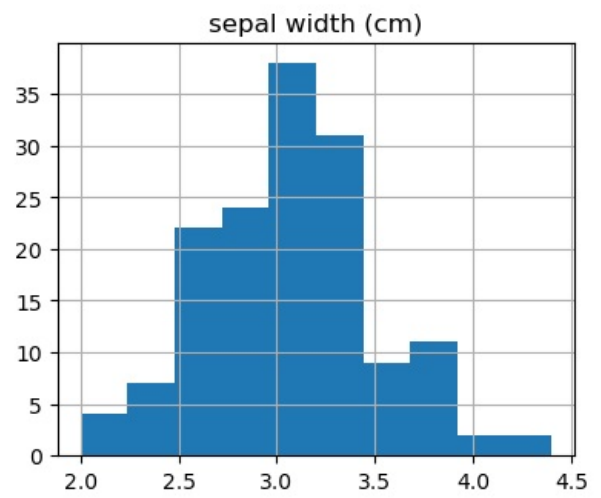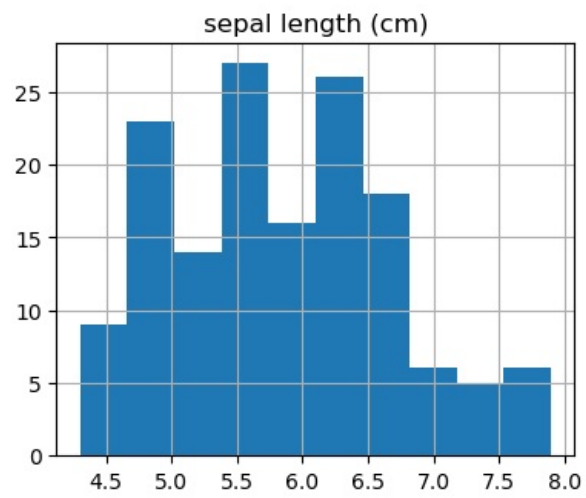

```
In [10]: df.boxplot(figsize=(10,8))
         plt.title('Boxplots of Features')
         plt.show()
```

Boxplots of Features

If we observe closely for the box 2, interquartile distance is roughly around 0.75 hence the values lying beyond this range of (third quartile + interquartile distance) i.e. roughly around 4.05 will be considered as outliers. Similarly outliers withother boxplots can be found.

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js