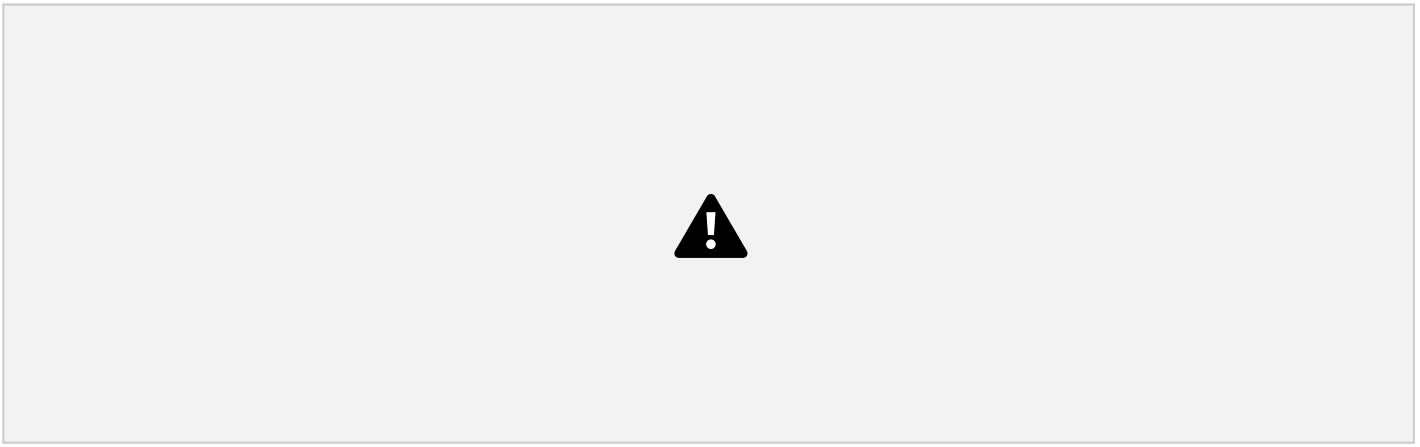
ASSIGNMENT 1

Name : Abhishek Badgujar

Class : BE - A

Roll no : 49

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



Print the Iris Dataset

In [61]:

```
print(iris_read.head())
```

```
  Id Sepal Length Sepal Width Petal Length Petal Width Species  0 1 5.1
3.5 1.4 0.2 Iris-setosa  1 2 4.9 3.0 1.4 0.2 Iris-setosa  2 3 4.7 3.2
1.3 0.2 Iris-setosa  3 4 4.6 3.1 1.5 0.2 Iris-setosa  4 5 5.0 3.6 1.4
0.2 Iris-setosa
```

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

[62]:

```
print
( "Number of features in the dataset :"
    ,              len ( iris_read .
columns )) print ("Type of features present in the dataset:" )
           iris_read . dtypes
```

```
Number of features in the dataset : 6 Type
of features present in the dataset:
```

Out[62]:

In

```
Id int64
Sepal Length float64
Sepal Width float64
Petal Length float64
Petal Width float64
Species object
dtype: object
```

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

In [63]:

```
iris_read.describe()
```

Out[63]:

	Id	Sepal Length	Sepal Width	Petal Length	Petal Width
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000

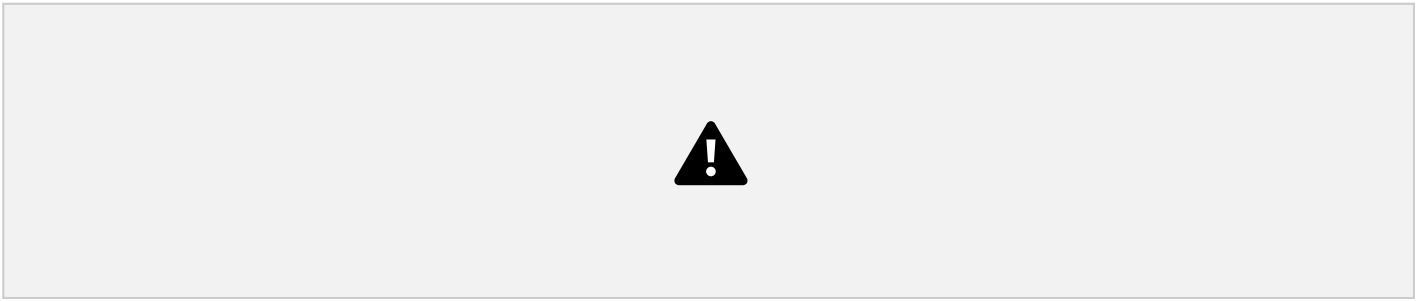
	Id	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	

Iris-virginica

148 149 6.2 3.4 5.4 2.3 Iris-virginica In

149 150 5.9 3.0 5.1 1.8 Iris-virginica **150** rows × 6 columns

In



	Id	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
0	1	5.1	3.5	1.4	0.2	1
1	2	4.9	3.0	1.4	0.2	1
2	3	4.7	3.2	1.3	0.2	1
3	4	4.6	3.1	1.5	0.2	1
4	5	5.0	3.6	1.4	0.2	1
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	3
146	147	6.3	2.5	5.0	1.9	3
147	148	6.5	3.0	5.2	2.0	3
148	149	6.2	3.4	5.4	2.3	3
149	150	5.9	3.0	5.1	1.8	3

150 rows × 6 columnsIn [47]:

Out[47]:

```
Id int64
Sepal Length float64
Sepal Width float64
Petal Length float64
Petal Width float64
Species int64
dtype: object In [21]:
```

```
print("Variance is:\n",iris_read.var())
```

```
Variance is:
```

```
 Id 1887.500000
Sepal Length 0.685694
Sepal Width 0.188004
Petal Length 3.113179
In
Petal Width 0.582414
Species 0.223714
dtype: float64 [36]:
```

```
print("Percentile is :"
,np.percentile(iris_read,
95))
```

```
Percentile is : 105.04999999999995 Out[36]:
```

	Id	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
0	1	5.1	3.5	1.4	0.2	1
1	2	4.9	3.0	1.4	0.2	1
2	3	4.7	3.2	1.3	0.2	1
3	4	4.6	3.1	1.5	0.2	1
4	5	5.0	3.6	1.4	0.2	1
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	1
146	147	6.3	2.5	5.0	1.9	1
147	148	6.5	3.0	5.2	2.0	1
148	149	6.2	3.4	5.4	2.3	1
149	150	5.9	3.0	5.1	1.8	1

150 rows × 6 columns

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

```
[44]:
```

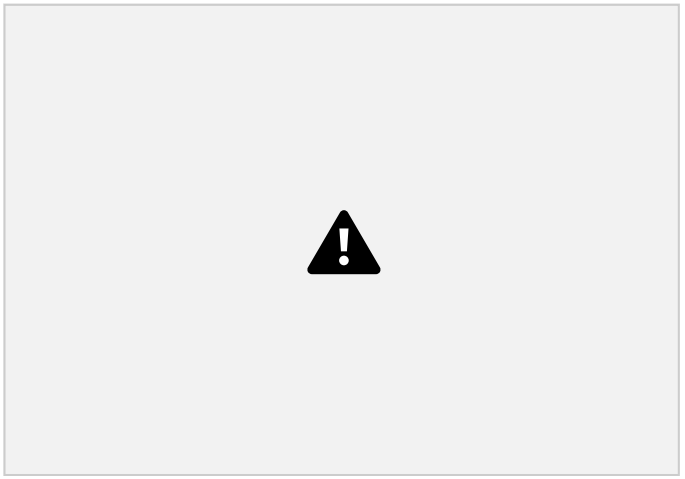
```
print("Histogram") sns.displot(iris_read['Sepal
Length'],kde = False) sns.displot(iris_read['Sepal
Width'],kde = False) sns.displot(iris_read['Petal
Length'],kde = False) sns.displot(iris_read['Petal
Width'],kde = False)
sns.displot(iris_df['Species'],kde = False)
```

```
Histogram
```

```
Out[44]:
```

```
<seaborn.axisgrid.FacetGrid at 0x151c9ef8>
```

In



4. Create a boxplot for each feature in the dataset. All ofthe boxplots should be combined into a single plot. Compare distributions and identify outliers. [54]:

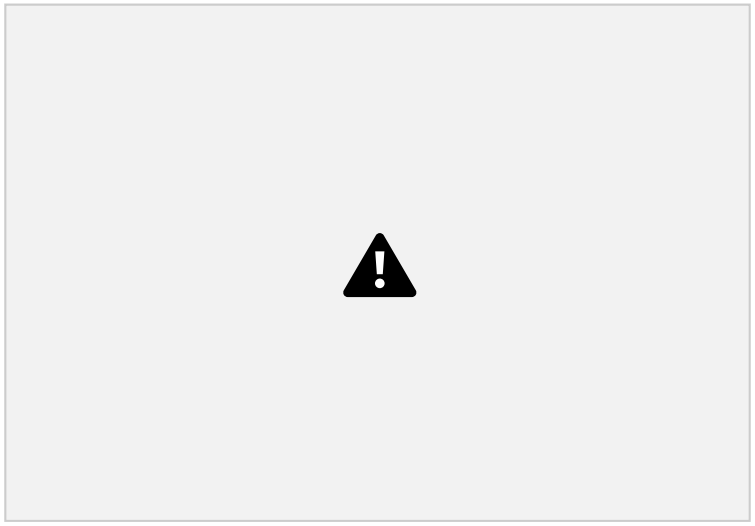
```
plot = sns.countplot(x = 'Sepal Length',data = iris_read)
sns.boxplot(x = 'Sepal Length', y = 'Sepal Width', data = iris_read)
plot.set_xticklabels(plot.get_xticklabels(),rotation=40)
```

Out[54]:

```
[Text(0, 0, '4.3'),
 Text(1, 0, '4.4'),
 Text(2, 0, '4.5'),
 Text(3, 0, '4.6'),
 Text(4, 0, '4.7'),
 Text(5, 0, '4.8'),
 Text(6, 0, '4.9'),
 Text(7, 0, '5.0'),
 Text(8, 0, '5.1'),
 Text(9, 0, '5.2'),
 Text(10, 0, '5.3'),
 Text(11, 0, '5.4'),
 Text(12, 0, '5.5'),
 Text(13, 0, '5.6'),
 Text(14, 0, '5.7'),
 Text(15, 0, '5.8'),
 Text(16, 0, '5.9'),
 Text(17, 0, '6.0'),
 Text(18, 0, '6.1'),
 Text(19, 0, '6.2'),
 Text(20, 0, '6.3'),
 Text(21, 0, '6.4'),
 Text(22, 0, '6.5'),
 Text(23, 0, '6.6'),
 Text(24, 0, '6.7'),
```

In

```
 Text(25, 0, '6.8'),
 Text(26, 0, '6.9'),
 Text(27, 0, '7.0'),
 Text(28, 0, '7.1'),
 Text(29, 0, '7.2'),
 Text(30, 0, '7.3'),
 Text(31, 0, '7.4'),
 Text(32, 0, '7.6'),
 Text(33, 0, '7.7'),
 Text(34, 0, '7.9')]
```



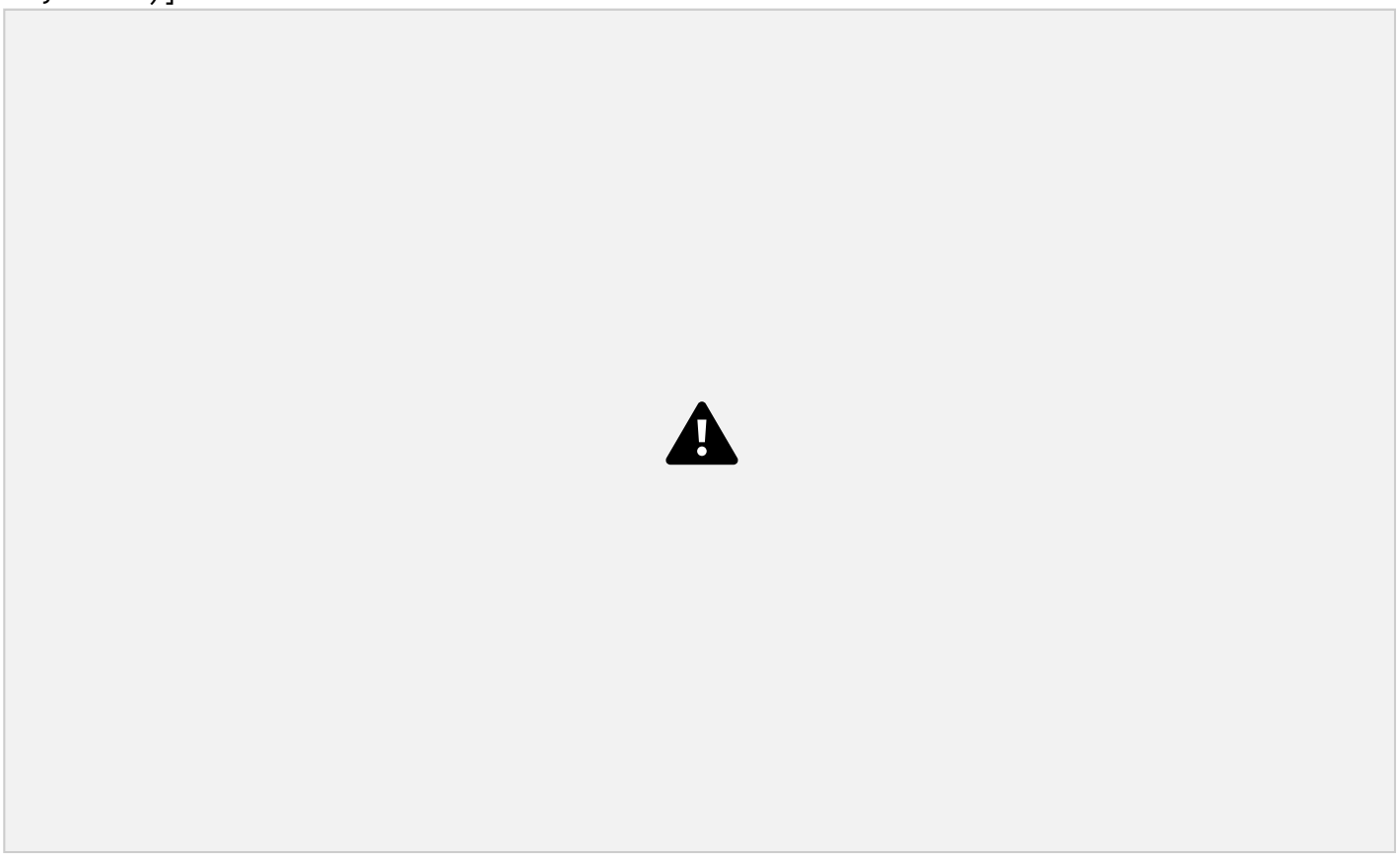
[55]:

```
plot = sns.countplot(x = 'Petal Length',data = iris_read)
sns.boxplot(x = 'Petal Length', y = 'Petal Width', data = iris_read)
plot.set_xticklabels(plot.get_xticklabels(),rotation=
40)
```

Out[55]:

```
[Text(0, 0, '1.0'),
 Text(1, 0, '1.1'),
 Text(2, 0, '1.2'),
 Text(3, 0, '1.3'),
 Text(4, 0, '1.4'),
 Text(5, 0, '1.5'),
 Text(6, 0, '1.6'),
 Text(7, 0, '1.7'),
 Text(8, 0, '1.9'),
 Text(9, 0, '3.0'),
 Text(10, 0, '3.3'),
 Text(11, 0, '3.5'),
 Text(12, 0, '3.6'),
 Text(13, 0, '3.7'),
 Text(14, 0, '3.8'),
```

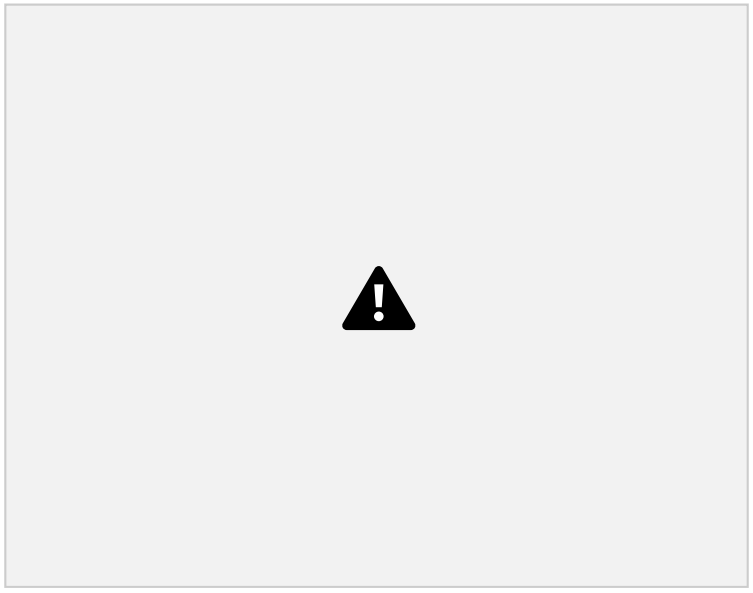
```
 Text(15, 0, '3.9'),
 Text(16, 0, '4.0'),
 Text(17, 0, '4.1'),
 Text(18, 0, '4.2'),
 Text(19, 0, '4.3'),
In
 Text(20, 0, '4.4'),
 Text(21, 0, '4.5'),
 Text(22, 0, '4.6'),
 Text(23, 0, '4.7'),
 Text(24, 0, '4.8'),
 Text(25, 0, '4.9'),
 Text(26, 0, '5.0'),
 Text(27, 0, '5.1'), Text(28, 0, '5.2'),
 Text(29, 0, '5.3'),
 Text(30, 0, '5.4'),
 Text(31, 0, '5.5'),
 Text(32, 0, '5.6'),
 Text(33, 0, '5.7'),
 Text(34, 0, '5.8'),
 Text(35, 0, '5.9'),
 Text(36, 0, '6.0'),
 Text(37, 0, '6.1'),
 Text(38, 0, '6.3'),
 Text(39, 0, '6.4'),
 Text(40, 0, '6.6'),
 Text(41, 0, '6.7'), Text(42,
 0, '6.9')]
```



```
[Text(0, 0, 'Iris-setosa'),
```

```
 Text(1, 0, 'Iris-versicolor'), Text(2,
 0, 'Iris-virginica')]
```

In [70]:

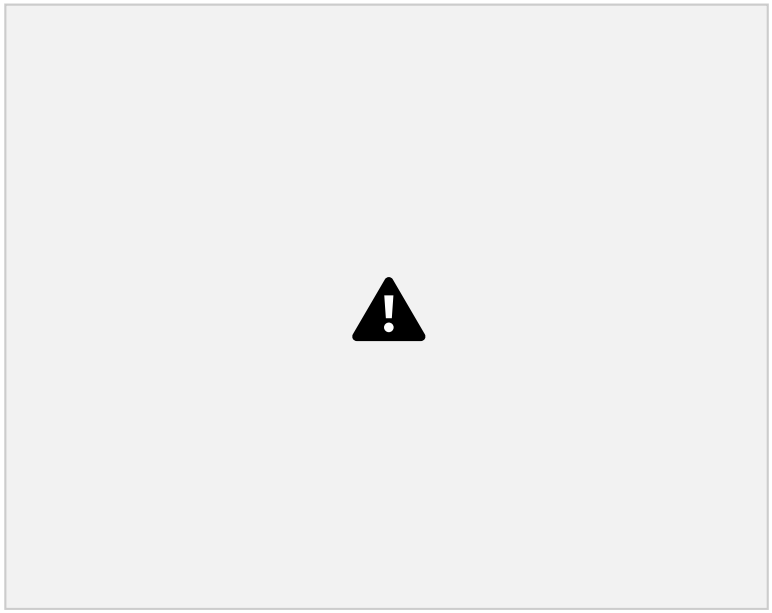


```
plot = sns.countplot(x = 'Species',data = iris_df)  sns.boxplot(x = 'Species', y = 'Sepal Width', data = iris_df)

plot.set_xticklabels(plot.get_xticklabels(),rotation = 40 )
```

Out[70]:

```
[Text(0, 0, 'Iris-setosa'), Text(1,
 0, 'Iris-versicolor'),
 Text(2, 0, 'Iris-virginica')]
```

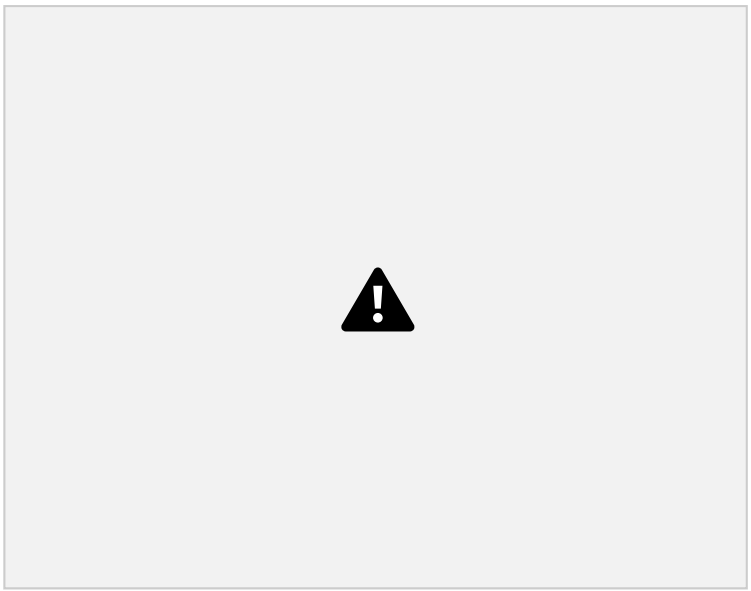


In [71]:

```
plot = sns.countplot(x = 'Species',data = iris_df)
sns.boxplot(x = 'Species', y = 'Petal Length', data
= iris_df)
plot.set_xticklabels(plot.get_xticklabels(),rotation=40)
```

Out[71]:

```
[Text(0, 0, 'Iris-setosa'),
 Text(1, 0, 'Iris-versicolor'), Text(2,
 0, 'Iris-virginica')]
```



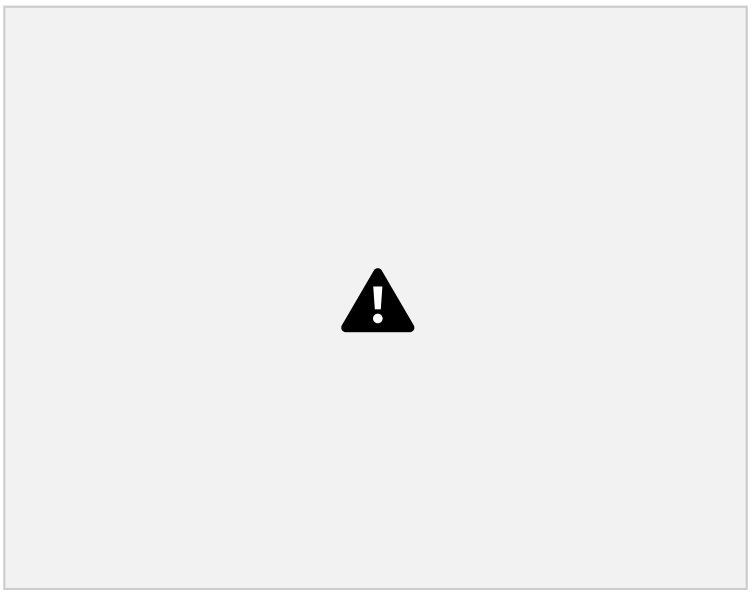
In [72]:

```
plot = sns.countplot(x = 'Species',data = iris_df)
sns.boxplot(x = 'Species', y = 'Petal Width', data
= iris_df)
plot.set_xticklabels(plot.get_xticklabels(),rotation=40)
```

Out[72]:

```
[Text(0, 0, 'Iris-setosa'),
```

```
Text(1, 0, 'Iris-versicolor'), Text(2,
0, 'Iris-virginica')]
```



In [ ]: