Data Exploration using Seaborn

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as seaborn
import matplotlib.pyplot as plt
```

In [20]:

```
#Import the iris data set
from sklearn.datasets import load_iris
iris = load_iris()
# Assigning iris data set into  data frame using features as column names
data = pd.DataFrame(iris.data, columns=iris.feature_names)
# creating label data frame with the coulmns as Species types
label=pd.DataFrame(list(map(lambda x : iris.target_names[x],iris.target)),column
s=['Species'])
iris=pd.concat([data,label],axis=1)
# Printin top 5 records of the iris data set
print(iris.head())
```

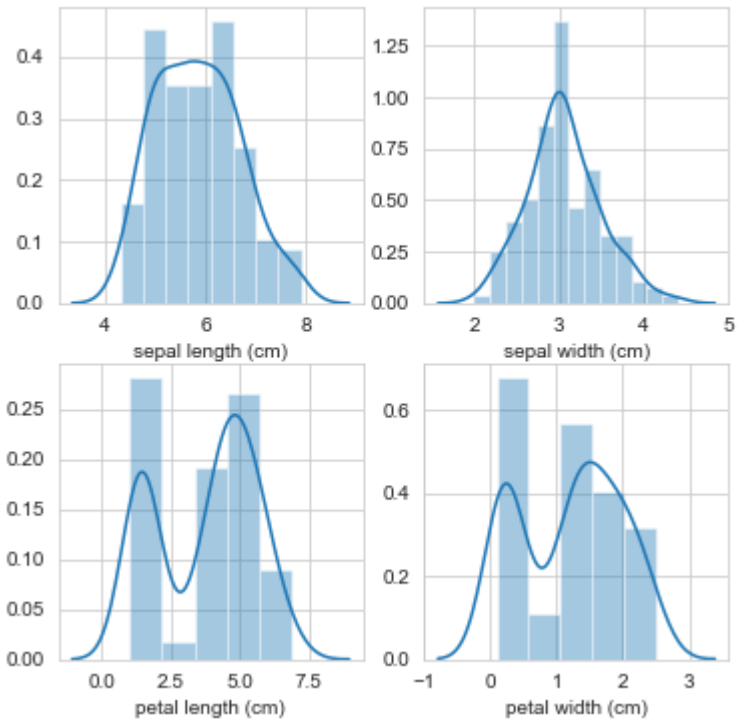
```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal wid
th (cm)  \
0                5.1               3.5                1.4
    0.2
1                4.9               3.0                1.4
    0.2
2                4.7               3.2                1.3
    0.2
3                4.6               3.1                1.5
    0.2
4                5.0               3.6                1.4
    0.2

  Species
0  setosa
1  setosa
2  setosa
3  setosa
4  setosa
```

Problem 1

In [25]:

```
# Settin the style as white grid as background
seaborn.set_style("whitegrid");
# making fig size as 6X6 inches with 2 rows and 2 columns
fig, axs = plt.subplots(figsize=(6,6), ncols=2, nrows=2)
# displaying the graphs in each quadrant from 0,0 to 1,1
seaborn.distplot(iris['sepal length (cm)'],ax=axs[0, 0])
seaborn.distplot(iris['sepal width (cm)'], ax=axs[0, 1])
seaborn.distplot(iris['petal length (cm)'], ax=axs[1, 0])
seaborn.distplot(iris['petal width (cm)'], ax=axs[1, 1])
# displaying the Plot
plt.show();
```

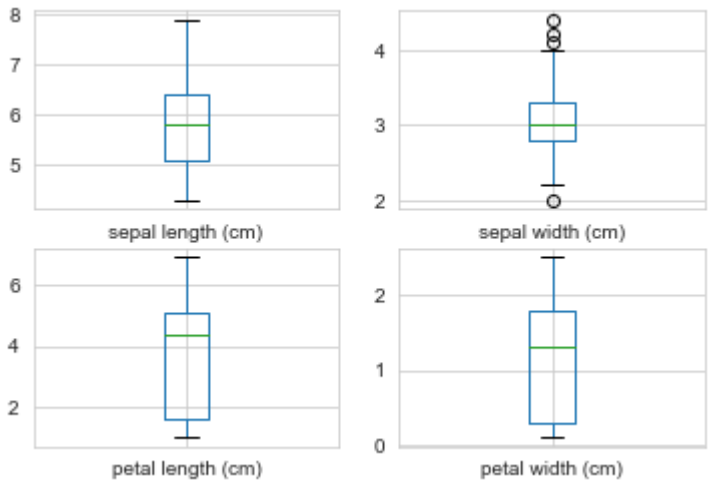
```
/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: Fu
tureWarning: Using a non-tuple sequence for multidimensional indexin
g is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the
future this will be interpreted as an array index, `arr[np.array(se
q)]`, which will result either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumva
l
```



Problem 2

In [26]:

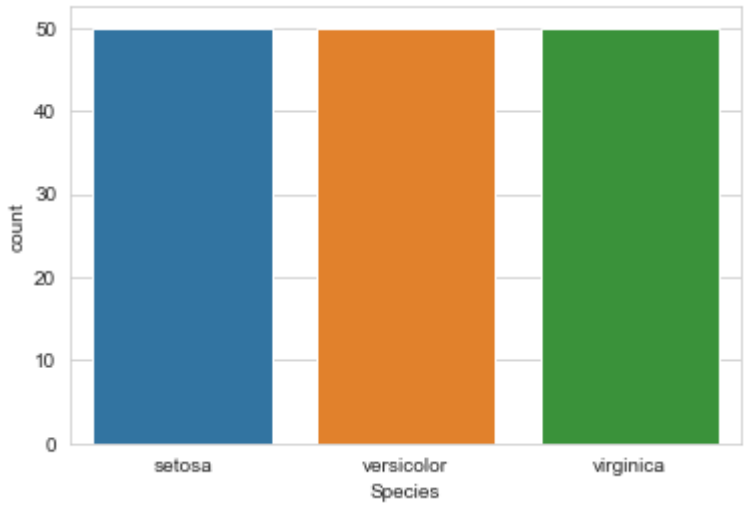
```
# making figure and axes as 2 rows and 2 columns
fig, axes = plt.subplots(2,2)
# displaying plot for all the columns except Species which is last column(-1)
for i,el in enumerate(list(iris.columns.values)[:-1]):
    a = iris.boxplot(el, ax=axes.flatten()[i])
# displaying the Plot
plt.show()
```



Problem 3

In [28]:

```
# making count plot with Species column values
seaborn.countplot(x='Species',data=iris)
# displaying the Plot
plt.show()
```



Problem 4

In [30]:

```
plt.close();
seaborn.set_style("whitegrid");
# creating pair plots using Species as hue
seaborn.pairplot(iris, hue="Species", height=3);
# displaying the Plot
plt.show()
```

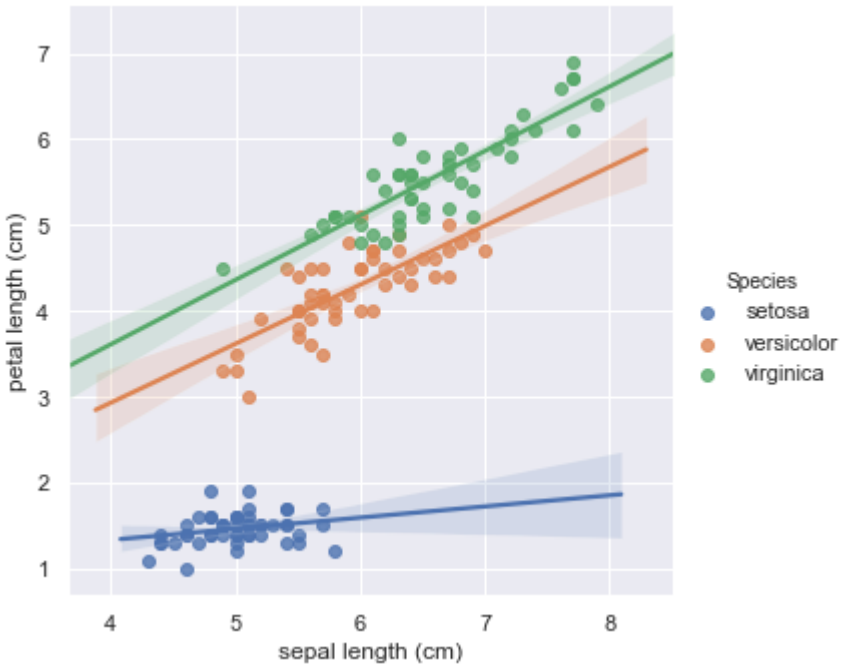
```
/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: Fu
tureWarning: Using a non-tuple sequence for multidimensional indexin
g is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the
future this will be interpreted as an array index, `arr[np.array(se
q)]`, which will result either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumva
l
```

Problem 5

In [32]:

```
plt.close();
seaborn.set_style("whitegrid");
# making color code as true to disply inbuilt color code
seaborn.set(color_codes=True)
# creating implot with sepal length and petal length using Species as hue
seaborn.lmplot(x='sepal length (cm)',y='petal length (cm)',hue='Species',data=ir
is)
# displaying the Plot
plt.show()
```

```
/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: Fu
tureWarning: Using a non-tuple sequence for multidimensional indexin
g is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the
future this will be interpreted as an array index, `arr[np.array(se
q)]`, which will result either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumva
l
```

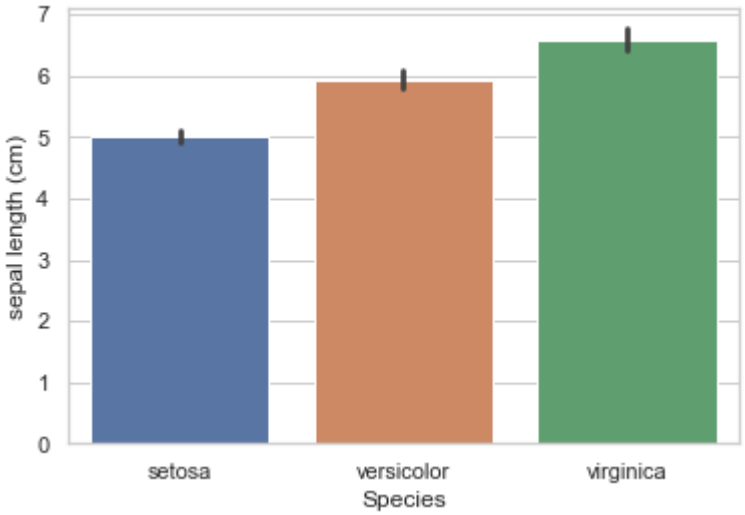


Problem 6

In [34]:

```
# closing the existing plotting
plt.close();
# setting back ground style as white grid
seaborn.set_style("whitegrid");
# creating bar plot with species and sepal length
seaborn.barplot(x='Species',y='sepal length (cm)',data=iris)
# displaying the Plot
plt.show()
```

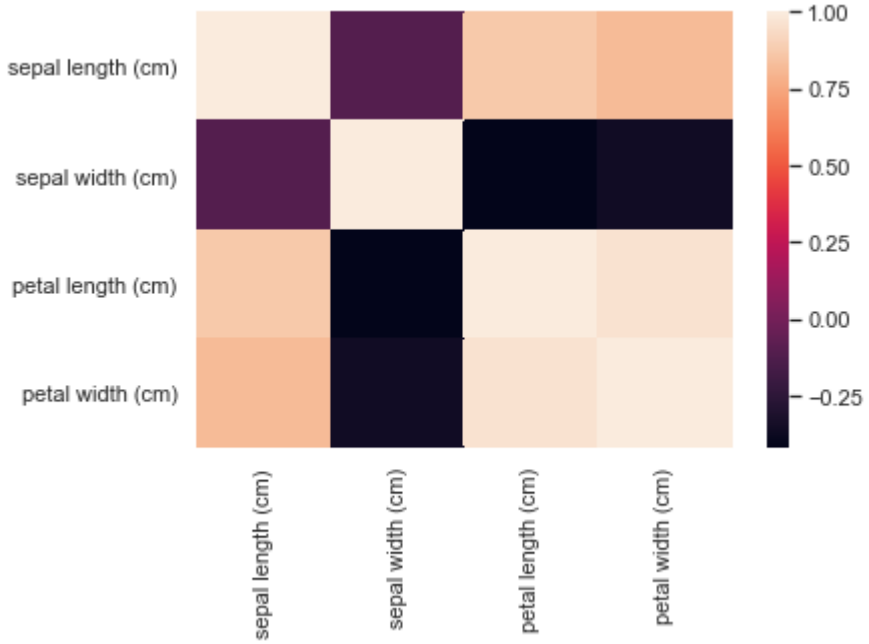
```
/anaconda3/lib/python3.7/site-packages/scipy/stats/stats.py:1713: Fu
tureWarning: Using a non-tuple sequence for multidimensional indexin
g is deprecated; use `arr[tuple(seq)]` instead of `arr[seq]`. In the
future this will be interpreted as an array index, `arr[np.array(se
q)]`, which will result either in an error or a different result.
  return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumva
l
```



Problem 7

In [36]:

```
# calculate the correlation matrix
corr = iris.corr()
# creating heatmap correlation column values
seaborn.heatmap(corr,xticklabels=corr.columns,yticklabels=corr.columns)
# displaying the Plot
plt.show()
```



In [ ]:

```

```