Assignment 2

Name :Bagadhi.Harika roll number:320126552002 Branch:CSM(AI&ML) DOS:27-07-2022

In [ ]:
```
Specification:
Find the dataset in the Moodle Unit -4 . find iris dataset in Dataset folder.

Analyze the Dataset Iris and visualize the data and display different plots for ge

display following plots using iris data

i)scatter

ii)Bar

iii) Hist

iv)pie

v)Density

Observations are important.
```

Importing required modules

In [7]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as ply
import seaborn as sns
```

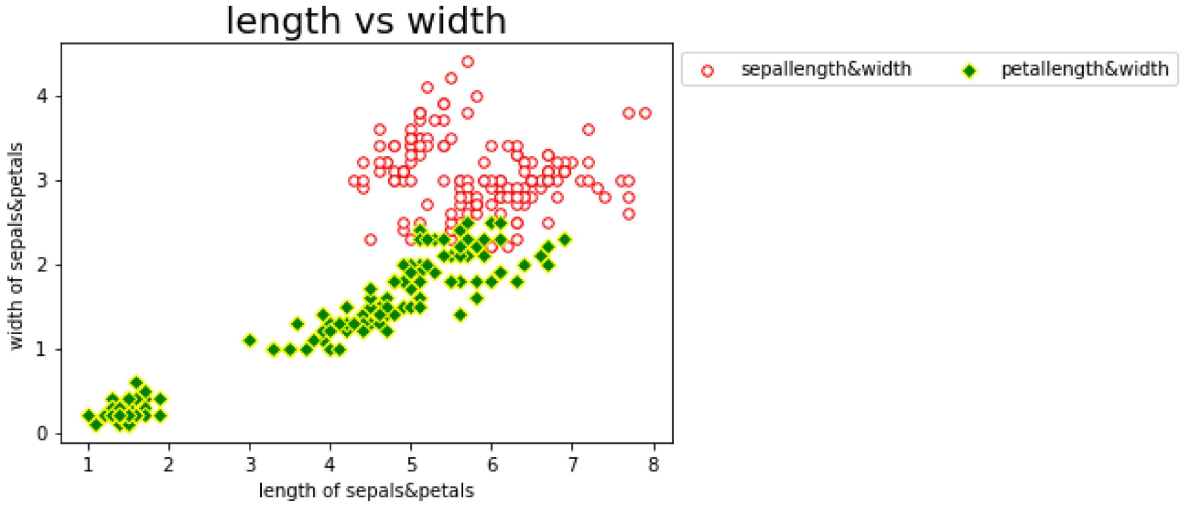
Importing the csv file

In [8]:
```
df=pd.read_csv("Iris.csv")
print(df.head(5))
print(df.shape)
print(df.size)
print(df.isna().sum().sum())
print(df.columns)
print(df.info())
print(df["Species"].value_counts())
```

```
   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm      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
(150, 6)
900
0
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
<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
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
None
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
Name: Species, dtype: int64
```

Scatter plot

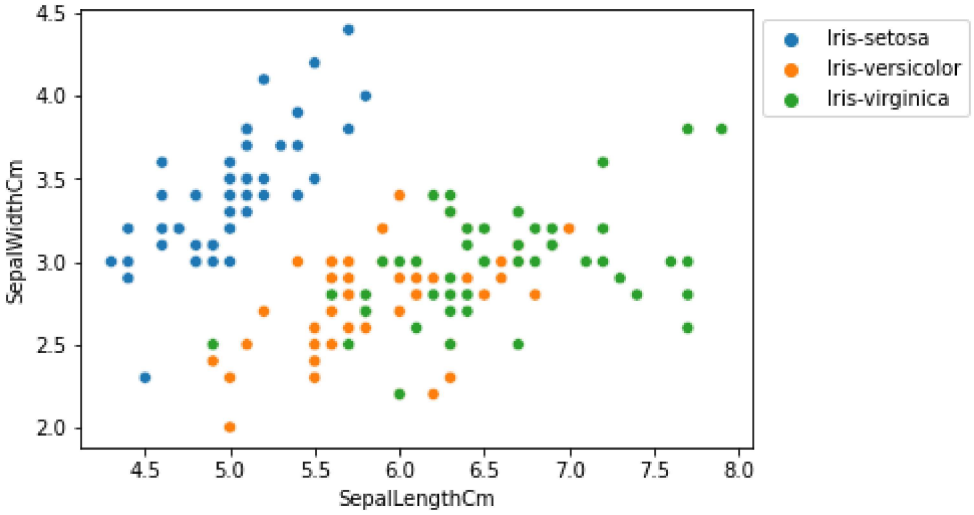
In [90]:
```
ply.scatter(df["SepalLengthCm"],df["SepalWidthCm"]  ,s=30.5,c="white",marker="o",e
ply.scatter(df["PetalLengthCm"],df["PetalWidthCm"],color="green",marker="D",edgeco
ply.xlabel("length of sepals&petals")
ply.ylabel("width of sepals&petals")
ply.title("length vs width",fontsize=20)
ply.legend(bbox_to_anchor =(1,1), ncol = 2,loc=2)
ply.show()
```



1)-->From the above given scatter plot we can infer that as the length of the sepals and petals increasing corresponding width of sepals and petals is also increasing. 2)-->The width of sepals and petals is much denser in between the range of 5-6 and sparse in the range of

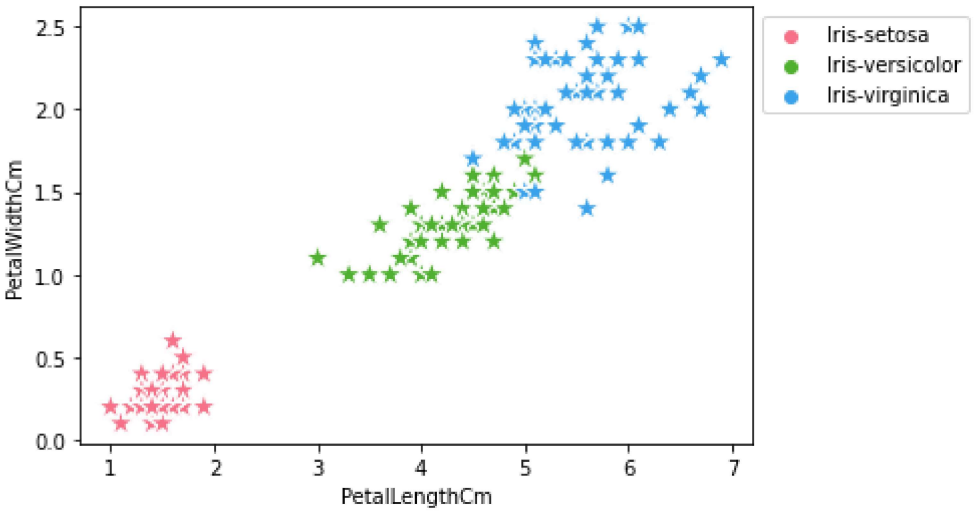
2-4. 3)-->The length of sepals and petals is much denser in between the range of 6-7 and sparse in the range of 7-8.

In [83]:
```
sns.scatterplot(x='SepalLengthCm', y='SepalWidthCm',
                hue='Species', data=df, )

ply.legend(bbox_to_anchor=(1, 1), loc=2)

ply.show()
```



1)-->The above scatter plot gives the relation between sepal length and sepal width of different species mentioned in the data set. 2)-->The plot gives us inference that iris setosa marked with blue as the highest sepal width for lowest sepal length 3)-->We can conclude that as the sepal length increases sepal width decreases 4)-->The sepal length and width is least for the specie iris versicolor

In [111…
```
sns.scatterplot(data=df,x='PetalLengthCm', y='PetalWidthCm',
                hue='Species',s=200,c=.9,palette="husl",marker="*" )
ply.legend(bbox_to_anchor=(1, 1), loc=2)
ply.gri()
```

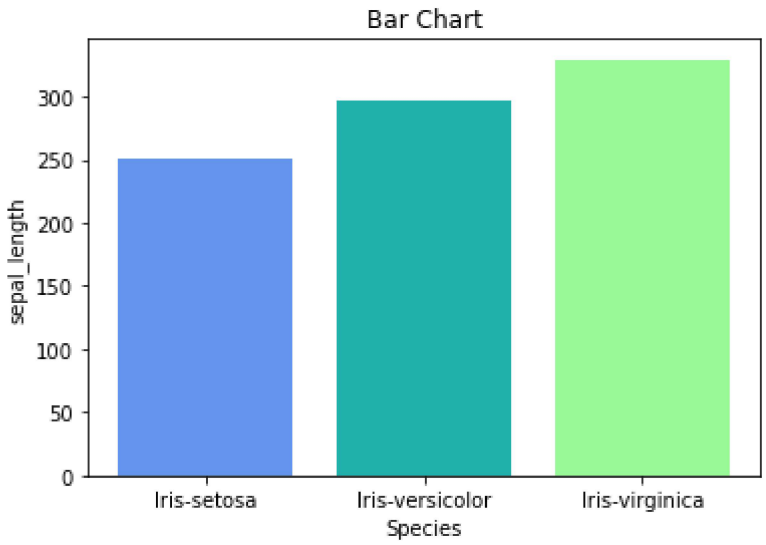


1)-->The above scatter plot gives the relation between petal length and petal width of different species mentioned in the data set. 2)-->The plot gives us the inference that as the petal length increases petal width also increases 3)-->The petal length and petal width is

higher for the specie iris-viriginica 4)-->The dense region is in the range of 4.5-6.5 5)-->The petal length and petal width is lower for the specie iris-setosa

Bar plot

In [131…
```
df1 = df.groupby('Species')["SepalLengthCm"].sum().to_frame().reset_index()
ply.bar(df1['Species'],df1["SepalLengthCm"],color = ['cornflowerblue','lightseagree
ply.title('Bar Chart')
ply.xlabel('Species')
ply.ylabel('sepal_length')
ply.show()
```



1)-->The above plotted bar graph gives us the information about total sum of sepal length for three different species plotted in the form of bars. 2)-->The sepal length is greater for the specie iris-verginica 3)-->The total sepal length of iris-setosa is 200cm 4)-->The total sepal length of iris-versicolor is 300cm 5)-->Thus the type of specie can be identified using the sepal length.

In [186…
```
s=df.groupby('Species')["PetalLengthCm"].sum()
y=df.groupby('Species')["PetalWidthCm"].sum()
z=df.groupby('Species')["SepalWidthCm"].sum()
w=df.groupby('Species')["SepalLengthCm"].sum()
df2=pd.DataFrame([s,y,z,w])
#df2.plot(x="Species",kind="bar",stacked=False,title="Species Specification")
```

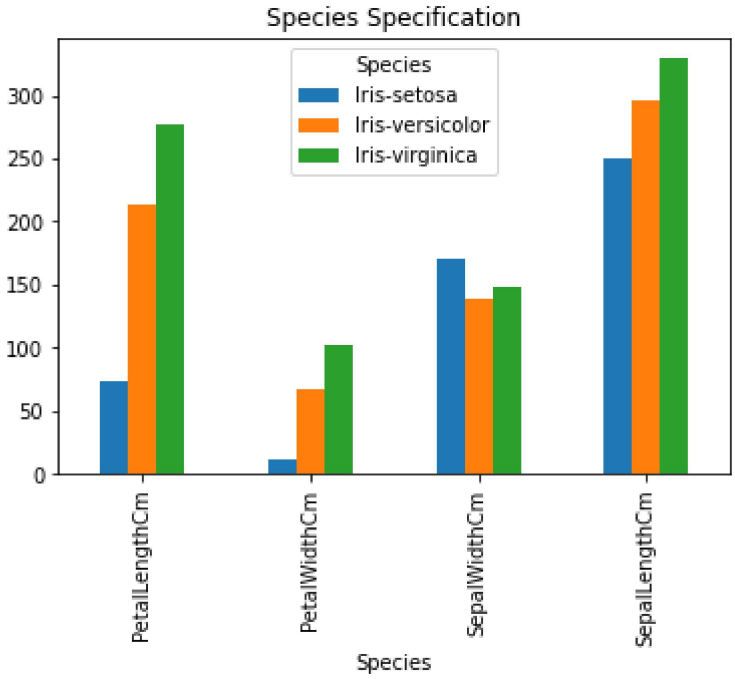
In [219…
```
df2
```

Out[219]:

Species	Species	Iris-setosa	Iris-versicolor	Iris-virginica
PetalLengthCm	PetalLengthCm	73.2	213.0	277.6
PetalWidthCm	PetalWidthCm	12.2	66.3	101.3
SepalWidthCm	SepalWidthCm	170.9	138.5	148.7
SepalLengthCm	SepalLengthCm	250.3	296.8	329.4

In [220…
```
df2.plot(x="Species",kind="bar",stacked=False,title="Species Specification")
```

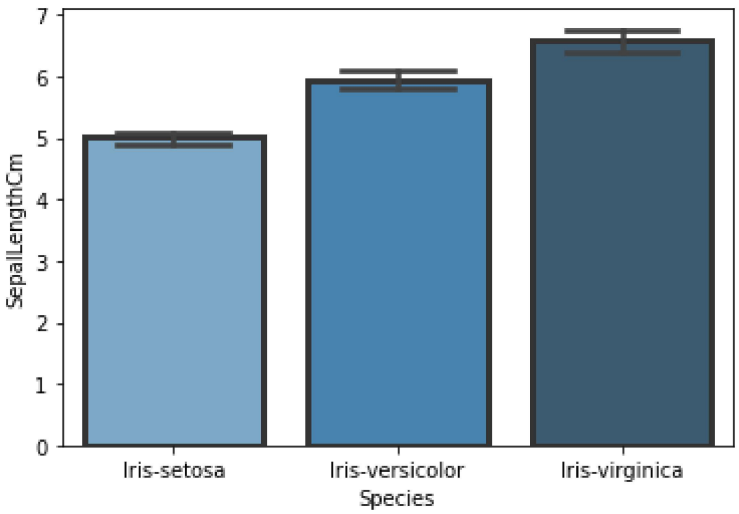
Out[220]:
```
<AxesSubplot:title={'center':'Species Specification'}, xlabel='Species'>
```



1)-->The above graph is the plot of multiple bar graphs 2)-->Here all the dimensions of the species along x-axis 3)-->All the four dimensions are greater for the specie iris-verginica 4)-->The opposite case with specie iris-setosa 5)-->The values of the specie iris-versicolor range in between the other two species

In [244…

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.barplot(x='Species',y='SepalLengthCm', data=df,capsize=.5,palette="Blues_d",li
plt.show()
```

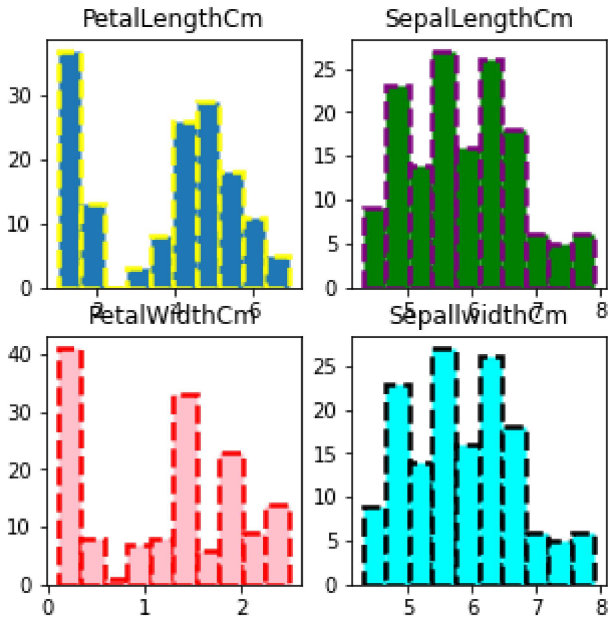


1)-->The above plotted bar graph gives us the information about total sum of sepal length for three different species plotted in the form of bars. 2)-->The sepal length is greater for the specie iris-verginica as indicated by the bar 3)-->The total sepal length of iris-setosa is less when compared to other species 4)-->The total sepal length of iris-versicolor range in between the two species 5)-->Thus the type of specie can be identified using the sepal length.

Histogram

In [11]:
```
fig= plt.subplots(2, 2, figsize=(5,5))
ply.subplot(2,2,1)
ply.hist(df["PetalLengthCm"],bins=10,edgecolor="yellow",linewidth=2.5,linestyle="da
ply.title("PetalLengthCm")
ply.subplot(2,2,2)
ply.hist(df["SepalLengthCm"],bins=10,color="green",edgecolor="purple",linewidth=2.5
ply.title("SepalLengthCm")
ply.subplot(2,2,3)
ply.hist(df["PetalWidthCm"],bins=10,color="pink",edgecolor="red",linewidth=2.5,line
ply.title("PetalWidthCm")
ply.subplot(2,2,4)
ply.hist(df["SepalLengthCm"],bins=10,color="cyan",edgecolor="black",linewidth=2.5,l
ply.title("SepallwidthCm")
```

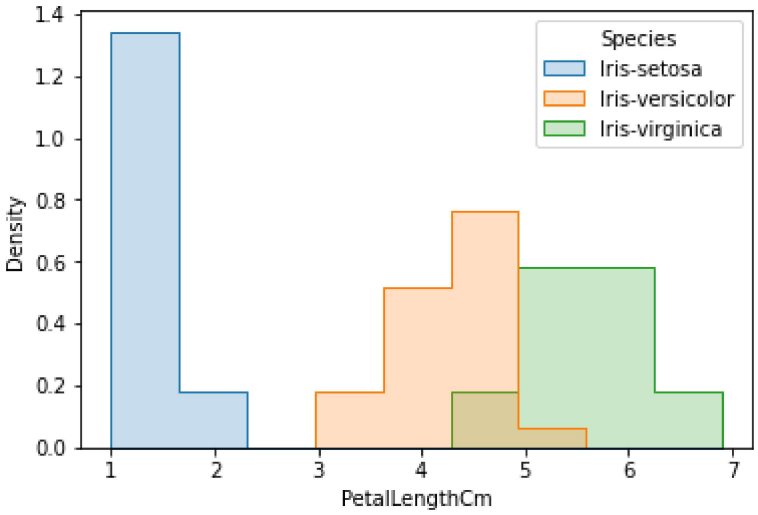
Out[11]:
```
Text(0.5, 1.0, 'SepallwidthCm')
```



1)-->The above graph is a combined plot of all the dimensions of the given specie 2)-->From the above graph we can conclude that the sepal length and width values are higher when compared to petal length and width values 3)-->Thus by plotting the paired plots we can compare one dimension with the other dimensions also have comparision amond the same dimension

In [270…
```
sns.histplot(data=df, x="PetalLengthCm", hue="Species",element="step",stat="density
```

Out[270]:
```
<AxesSubplot:xlabel='PetalLengthCm', ylabel='Density'>
```

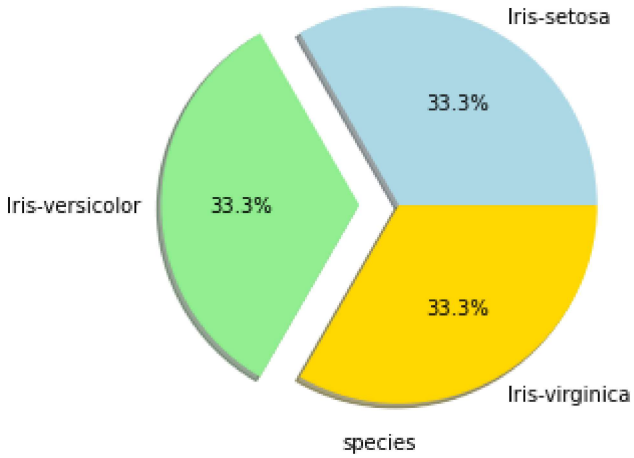


1)-->The above given histogram is the plot petal lengths of various species 2)-->The petal length is greater for the specie iris-setosa 3)-->The petal length is lesser for the specie iris-virginica 4)-->The petal length of specie iris-versicolor range in between the petal length of other two species 5)-->Also we can infer that petal length values of iris-versicolor and virginica merges in between the range of 4-5

Pie chart

In [273…

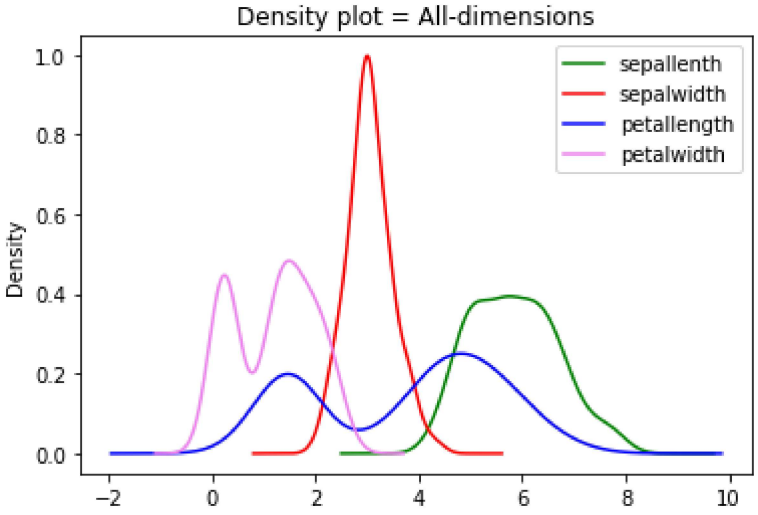
```
a= df['Species'].value_counts()
species = a.index
count = a.values
colors= ['lightblue','lightgreen','gold']
explode = (0,0.2,0)
plt.pie(count, labels=species,shadow=True,
        colors=colors,explode = explode, autopct='%1.1f%%')
plt.xlabel('species')
plt.axis('equal')
plt.show()
```



1)-->The above depicted pie chart gives us the count of difference species on an all 2)-->We can speculate that the count of three different species is same in the given data set 3)-->The percentage of each specie is 33.3 out of 100

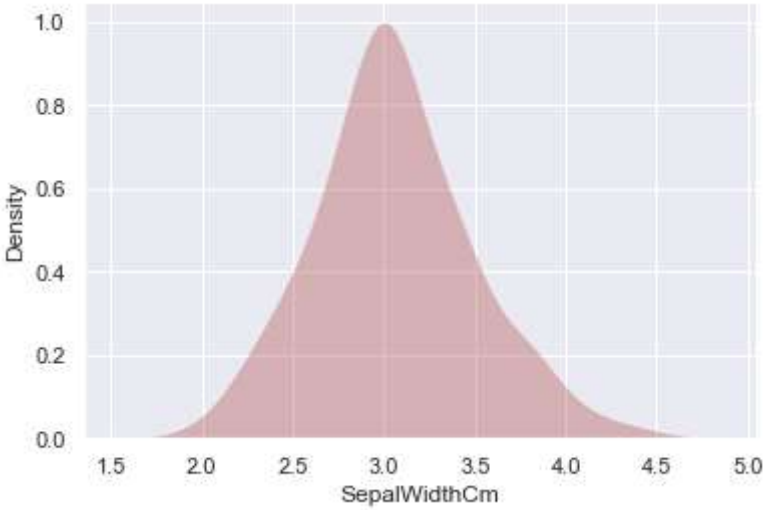
Density plot

In [10]:
```
import matplotlib.pyplot as plt
df.SepalLengthCm.plot.density(color='green',label="sepallenth")
df.SepalWidthCm.plot.density(color='red',label="sepalwidth")
df.PetalLengthCm.plot.density(color='blue',label="petallength")
df.PetalWidthCm.plot.density(color='violet',label="petalwidth")
plt.title('Density plot = All-dimensions')
plt.legend()
plt.show()
```



1)-->From the above depicted density plot we can summarize the following point 2)-->The sepal width is highest among all other dimensions 3)-->The petal length is lowest among all the other dimensions 4)-->Petal width and sepal length lie in between densities of all dimensions

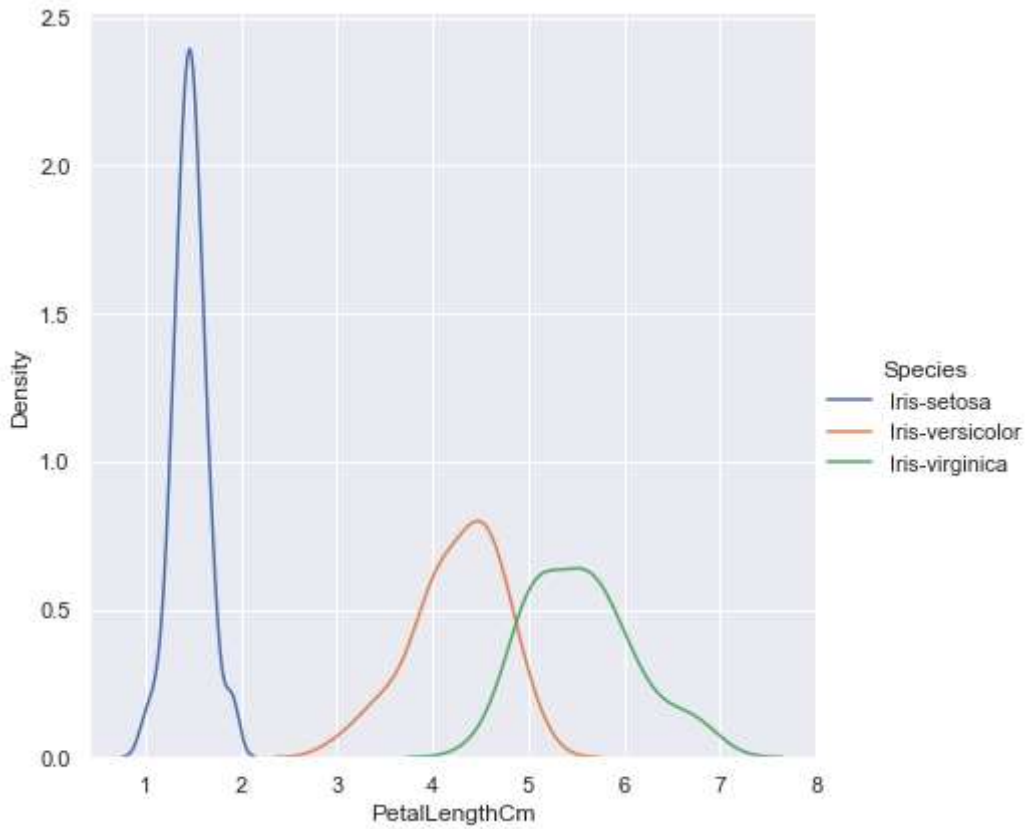
In [296…
```
sns.set(style="darkgrid")
sns.kdeplot(df["SepalWidthCm"],color="brown", fill=True, alpha=.3, linewidth=0)
plt.show()
```



1)-->The above graph is a density plot plotted using seaborn module functionalities 2)-->The density of sepal width is higher for the value 3 3)-->The density of sepal width steeps up until it reaches the value 3 4)-->Then it sees a downfall till the value 4.5

In [293…
```
sns.FacetGrid(df, hue="Species", height=6) \
   .map(sns.kdeplot, "PetalLengthCm") \
   .add_legend()
```

Out[293]: <seaborn.axisgrid.FacetGrid at 0x193f9f41730>



1)-->The above density plot gives the relation between densities of petal length of various species 2)-->Thus we can surmise that the density of petal length of iris-setosa specie has a greater value of 2.4 3)-->The densities of petal lengths of iris-versicolor and iris-verginica range in the region of 0.5 to 1.0 4)-->As the petal lengths increases,the densities decreases is the inference that we can get from the plot

THE END