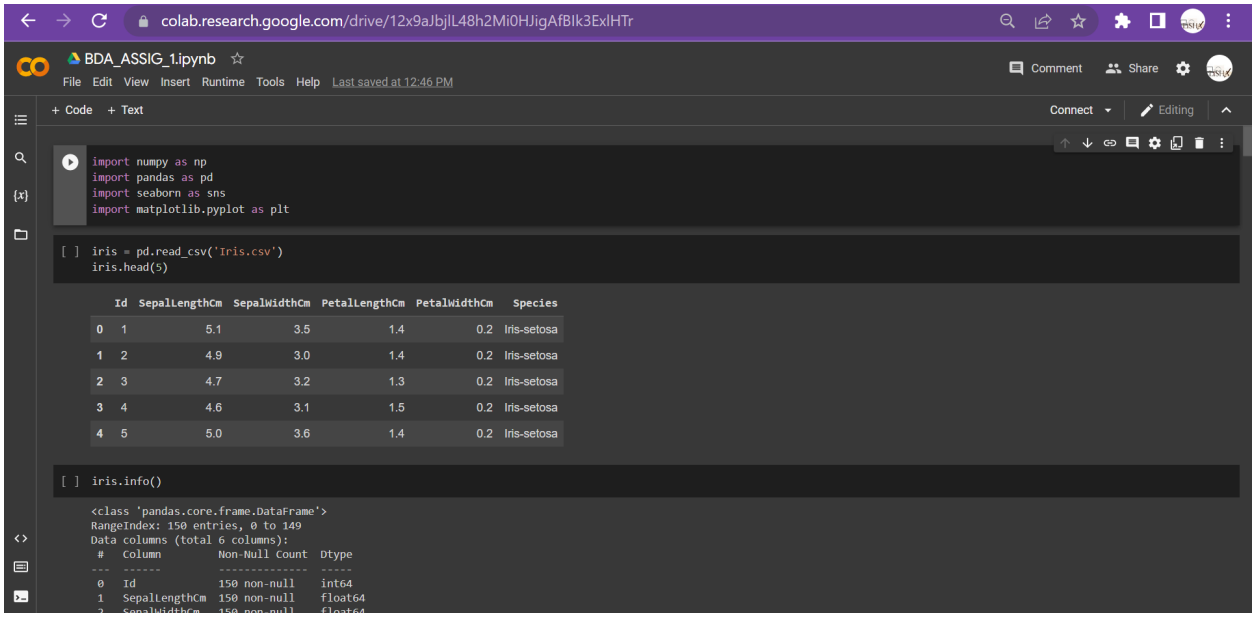
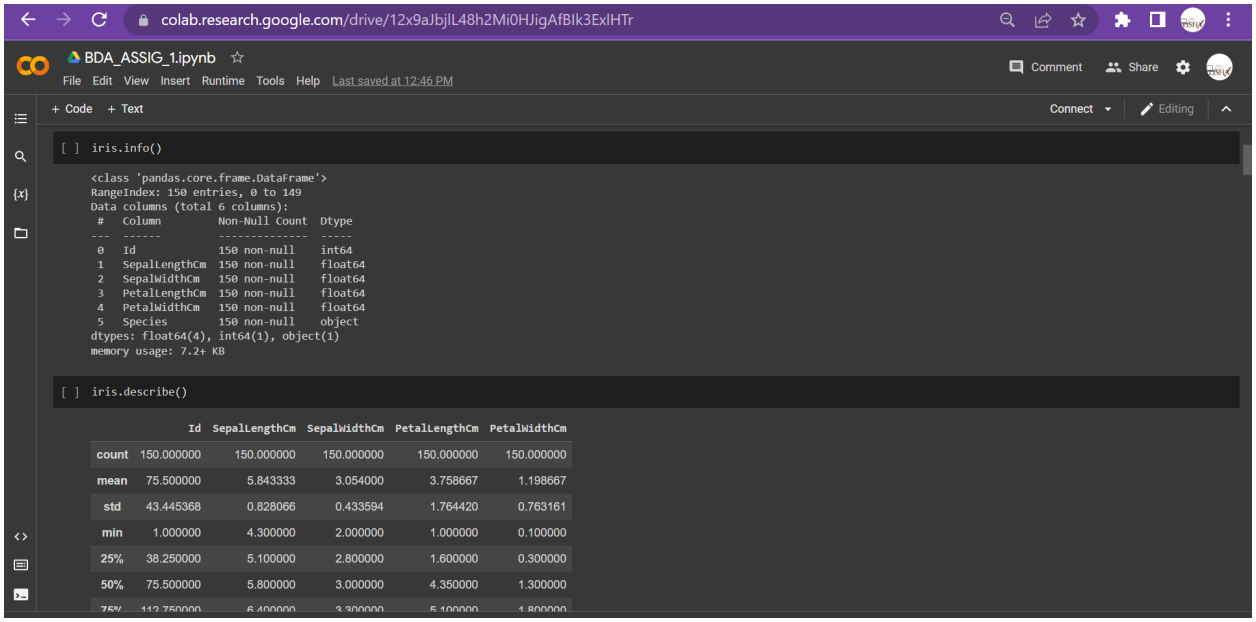
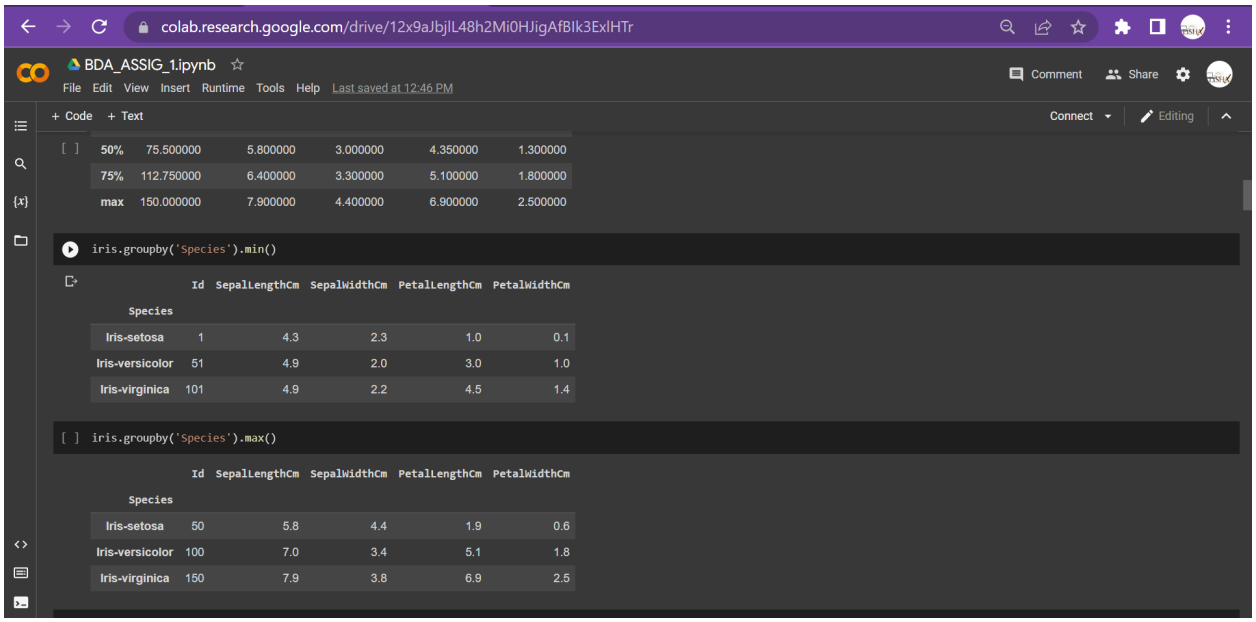
Name: Lisha Ingawale **Roll No:2193140** **Class: LY CSE IS 1**

Assignment 1:

1. Getting data to work with: Download the sample dataset locally for any application (Kaggle)
2. Setting up the working directory.
3. Unpacking the data. Decompress the file locally.
4. Looking at the data. Display the top (10) and bottom (10) of the file.
5. Measuring the length of the data set. Count the number of lines in the file.
6. Encode the categorical data
7. Plot a graph and give your insights for the application selected cases.




colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr
BDA_ASSIG_1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:46 PM
Comment
Share
+ Code + Text
Connect
Editing
[ ] iris.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
4 PetalWidthCm 150 non-null float64
5 Species 150 non-null object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
[ ] iris.describe()
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
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


colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr
BDA_ASSIG_1.ipynb
File Edit View Insert Runtime Tools Help Last saved at 12:46 PM
Comment
Share
+ Code + Text
Connect
Editing
50% 75.500000 5.800000 3.000000 4.350000 1.300000
75% 112.750000 6.400000 3.300000 5.100000 1.800000
max 150.000000 7.900000 4.400000 6.900000 2.500000
iris.groupby('Species').min()
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
Species
Iris-setosa 1 4.3 2.3 1.0 0.1
Iris-versicolor 51 4.9 2.0 3.0 1.0
Iris-virginica 101 4.9 2.2 4.5 1.4
[ ] iris.groupby('Species').max()
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
Species
Iris-setosa 50 5.8 4.4 1.9 0.6
Iris-versicolor 100 7.0 3.4 5.1 1.8
Iris-virginica 150 7.9 3.8 6.9 2.5

colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

Comment Share Connect Editing

+ Code + Text

Iris-virginica	150	7.9	3.8	6.9	2.5

```
[ ] iris.isnull().any()
```

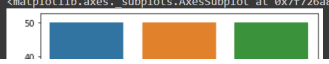
```
Id              False
SepalLengthCm   False
SepalWidthCm    False
PetalLengthCm   False
PetalWidthCm    False
Species         False
dtype: bool
```

```
iris.groupby('Species').count()
```

Species	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Iris-setosa	50	50	50	50	50
Iris-versicolor	50	50	50	50	50
Iris-virginica	50	50	50	50	50

```
[ ] sns.countplot(x=iris['Species'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f726a8daed0>
```



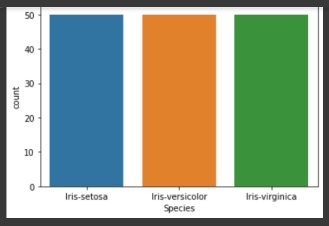
colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

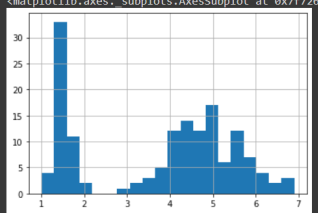
Comment Share Connect Editing

+ Code + Text




```
[ ] iris['PetalLengthCm'].hist(bins=20)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f726a359ed0>
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

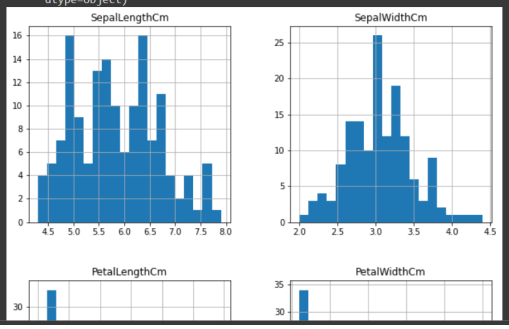
Comment Share

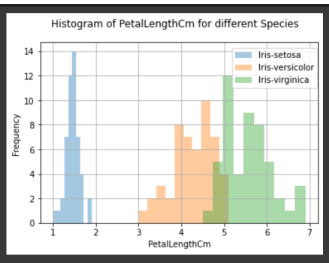
+ Code + Text

Connect Editing

```
[ ] iris.drop('Id',axis=1).hist(bins=20,figsize=(10,10))
```

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f7276bc9610>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f726a25a510>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f726a212b10>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f726a1d6150>]],
      dtype=object)
```






```
[ ] iris.groupby('Species').PetalWidthCm.plot.hist(alpha=0.4)
    plt.xlabel('PetalWidthCm')
    plt.suptitle('Histogram of PetalWidthCm for different Species')
    plt.legend(loc=(0.69,0.75))
    plt.grid()
```

colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

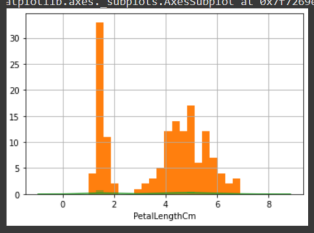
Comment Share

+ Code + Text

Connect Editing

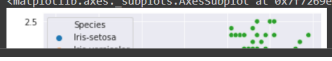
```
plt.hist(data=iris,x='PetalLengthCm',bins=20)
iris['PetalLengthCm'].hist(bins=20)
sns.distplot(iris['PetalLengthCm'],bins=20)
```

```
sr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your cod
warnings.warn(msg, FutureWarning)
atplotlib.axes._subplots.AxesSubplot at 0x7f7269e2b150>
```




```
[ ] sns.set_style('darkgrid')
    sns.scatterplot(data=iris,x='PetalLengthCm',y='PetalWidthCm',hue='Species')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7269e13110>
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

Comment Share

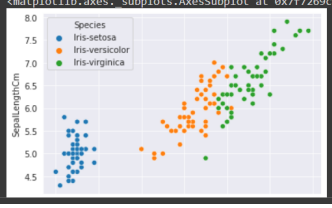
+ Code + Text

Connect Editing

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7269e13110>
```


```
[ ] sns.scatterplot(data=iris,x='PetalLengthCm',y='SepalLengthCm',hue='Species')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7269c351d0>
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

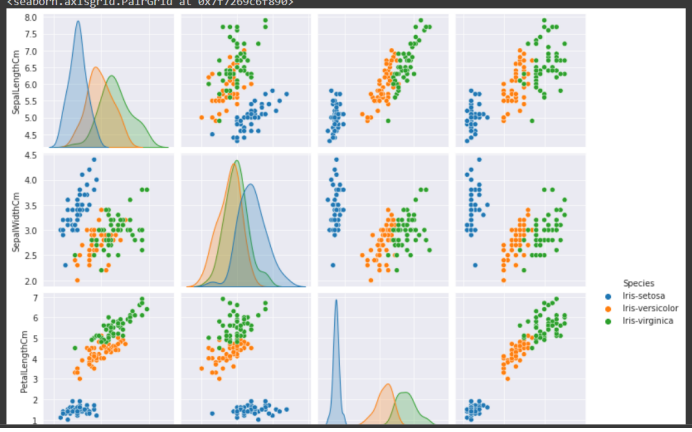
Comment Share

+ Code + Text

Connect Editing

```
sns.pairplot(data=iris.drop('Id',axis=1),hue='Species')
```

```
<seaborn.axisgrid.PairGrid at 0x7f7269c6f890>
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

Comment Share

+ Code + Text

Connect Editing

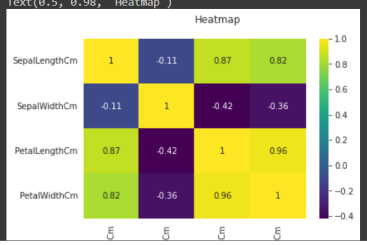
```
[ ] iris.drop('Id',axis=1).corr()
```

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
SepalLengthCm	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.817954	-0.356544	0.962757	1.000000

+ Code + Text

```
sns.heatmap(iris.drop('Id',axis=1).corr(),cmap='viridis',annot=True,)
plt.suptitle('Heatmap')
```

```
Text(0.5, 0.98, 'Heatmap')
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

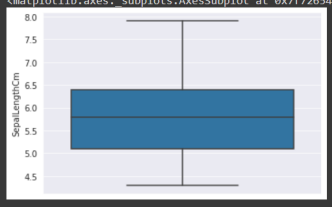
Comment Share

+ Code + Text

Connect Editing

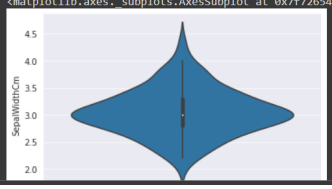
```
sns.boxplot(data=iris,y='SepalLengthCm')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f72654d8810>
```




```
[ ] sns.violinplot(data=iris,y='SepalWidthCm')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f726544db90>
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr

BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

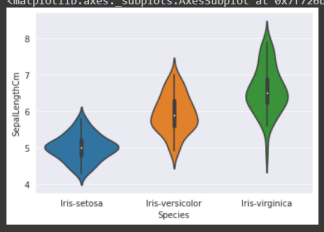
Comment Share

+ Code + Text

Connect Editing

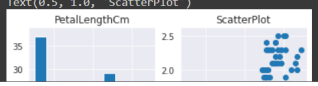
```
[ ] sns.violinplot(data=iris,x='Species',y='SepalLengthCm')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f726b148950>
```




```
[ ] fig,axes = plt.subplots(ncols=2,nrows=1)
    axes[0].hist(iris['PetalLengthCm'])
    axes[0].set_title('PetalLengthCm')
    axes[1].scatter(iris['PetalLengthCm'],iris['PetalWidthCm'])
    axes[1].set_title('ScatterPlot')
```

```
Text(0.5, 1.0, 'ScatterPlot')
```



colab.research.google.com/drive/12x9aJbjlL48h2Mi0HJigAfBIk3ExlHTr#scrollTo=kZWPHr7EJ9C_

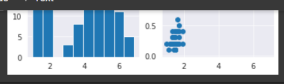
BDA_ASSIG_1.ipynb

File Edit View Insert Runtime Tools Help Last saved at 12:46 PM

Comment Share

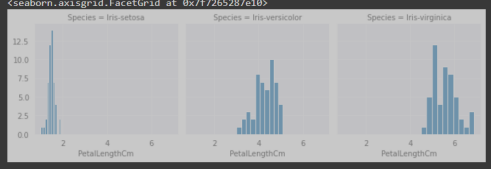
+ Code + Text

Connect Editing



```
[ ] g=sns.FacetGrid(data=iris,col='Species')
    g.map(plt.hist,'PetalLengthCm')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f7265287e10>
```




```
[ ] h=sns.FacetGrid(data=iris,col='Species')
    h.map(plt.scatter,'PetalLengthCm','PetalWidthCm',color='r')
```

```
<seaborn.axisgrid.FacetGrid at 0x7f72651f2110>
```