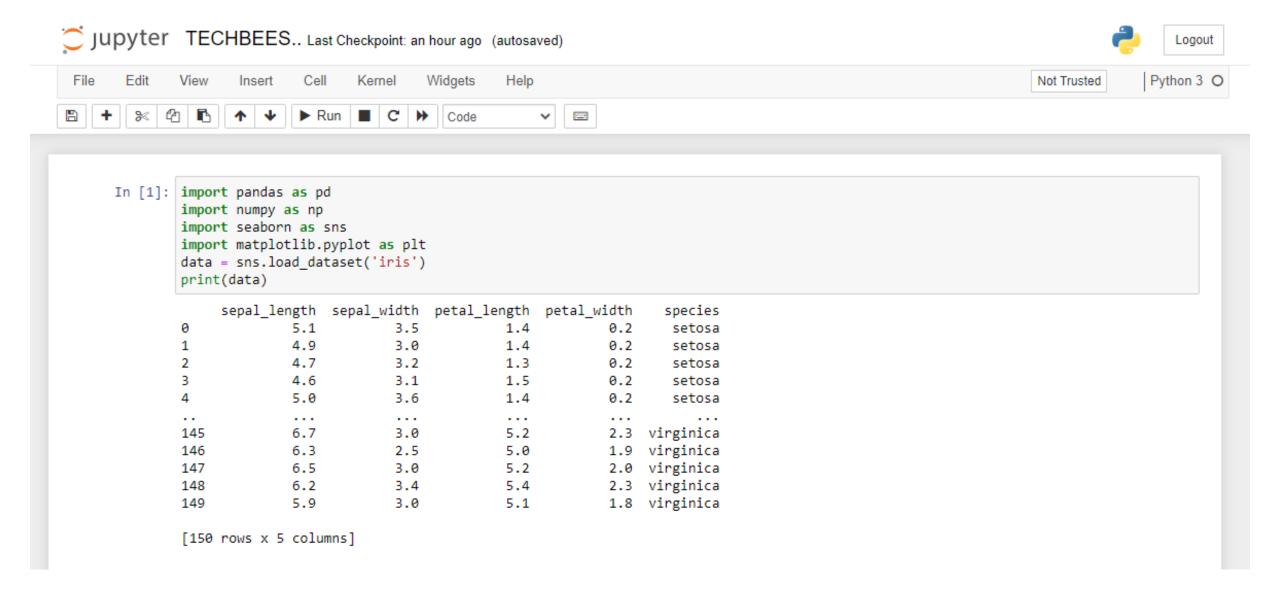
### NAME – Sanchita Agrawal

EMAIL ID – <u>sanchitaagrawal1304@gmail.com</u>

WEEK 1 - Data Analysis of the Iris Dataset

#### **WEEK 1 ASSIGNMENT SOLUTIONS:**



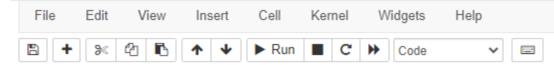




Not Trusted

Logout

Python 3 O



#### Print the first five records of the dataset for each 'species'

#### Out[33]:

|     | sepal_length | sepal_width | petal_length | petal_width | species    |
|-----|--------------|-------------|--------------|-------------|------------|
| 0   | 5.1          | 3.5         | 1.4          | 0.2         | setosa     |
| 1   | 4.9          | 3.0         | 1.4          | 0.2         | setosa     |
| 2   | 4.7          | 3.2         | 1.3          | 0.2         | setosa     |
| 3   | 4.6          | 3.1         | 1.5          | 0.2         | setosa     |
| 4   | 5.0          | 3.6         | 1.4          | 0.2         | setosa     |
| 50  | 7.0          | 3.2         | 4.7          | 1.4         | versicolor |
| 51  | 6.4          | 3.2         | 4.5          | 1.5         | versicolor |
| 52  | 6.9          | 3.1         | 4.9          | 1.5         | versicolor |
| 53  | 5.5          | 2.3         | 4.0          | 1.3         | versicolor |
| 54  | 6.5          | 2.8         | 4.6          | 1.5         | versicolor |
| 100 | 6.3          | 3.3         | 6.0          | 2.5         | virginica  |
| 101 | 5.8          | 2.7         | 5.1          | 1.9         | virginica  |
| 102 | 7.1          | 3.0         | 5.9          | 2.1         | virginica  |
| 103 | 6.3          | 2.9         | 5.6          | 1.8         | virginica  |
| 104 | 6.5          | 3.0         | 5.8          | 2.2         | virginica  |

#### Find out which sepal\_length value appears the most.

#### Find out the mean sepal\_width group by species.

```
In [28]: # mean of sepal_width of setosa
    data.loc[(data['species']=='setosa')].mean()['sepal_width']

Out[28]: 3.428000000000001

In [29]: # mean of sepal_width of versicolor
    data.loc[(data['species']=='versicolor')].mean()['sepal_width']

Out[29]: 2.77000000000000005

In [30]: # mean of sepal_width of virginica
    data.loc[(data['species']=='virginica')].mean()['sepal_width']

Out[30]: 2.9739999999999998
```

# Find out which species has the maximum and which species has the minimum petal\_length.

```
In [9]: # virginica has maximum petal length
         data.loc[data['petal_length'].idxmax()]
 Out[9]: sepal length
                              7.7
         sepal width
                               2.6
         petal_length
                               6.9
         petal width
                               2.3
         species
                         virginica
         Name: 118, dtype: object
In [10]: # setosa has minimum petal length
         data.loc[data['petal_length'].idxmin()]
Out[10]: sepal_length
                            4.6
         sepal width
                            3.6
         petal_length
                         1.0
                            0.2
         petal width
         species
                         setosa
         Name: 22, dtype: object
```

#### Find out the median petal\_width for the entire dataset.

```
In [11]: data["petal_width"].median()
# 1.3 is the median of petal_width for the entire dataset
Out[11]: 1.3
```

## Print the first 20 records of the dataset, in the ascending order of sepal\_length.

```
In [31]: newdata = data.head(20)
In [32]: newdata.sort_values(by = 'sepal_length',ascending = True)[['sepal_length']]
Out[32]:
               sepal_length
           13
                       4.3
            8
                       4.4
            3
                       4.6
                       4.6
                       4.7
           12
                       4.8
           11
                       4.8
                       4.9
                       4.9
            4
                       5.0
                       5.0
           17
                       5.1
                       5.1
           19
                       5.1
           16
                       5.4
                       5.4
                      5.4
           10
           18
                       5.7
                       5.7
           14
                       5.8
```

## Print the last 20 records of the dataset, in the descending order of sepal\_width.

In [14]: newdata.sort\_values(by = 'sepal\_width',ascending = False)[['sepal\_width']] Out[14]: sepal\_width 15 4.4 14 4.0 3.9 5 3.9 3.8 3.8 3.7 3.6 17 3.5 3.5 3.4 3.4 3.4 3.2 3.1 3.1 3.0 12 3.0 13 3.0 2.9

### Display the species and their characteristics where the petal\_length is less than 3.

```
In [19]: df = data[['petal_length', 'species']]
In [20]: df.loc[(df['petal_length']<3)]</pre>
Out[20]:
              petal_length species
                    1.4 setosa
                     1.4 setosa
                    1.3 setosa
                    1.5 setosa
                    1.4 setosa
                    1.7 setosa
                    1.4 setosa
                     1.5 setosa
                    1.4 setosa
                     1.5 setosa
          11
                    1.6 setosa
                    1.4 setosa
          13
                    1.1 setosa
          14
                    1.2 setosa
          15
                    1.5 setosa
                    1.3 setosa
          17
                    1.4 setosa
                    1.7 setosa
          19
                     1.5 setosa
                    1.7 setosa
          21
                     1.5 setosa
          22
                    1.0 setosa
          23
                    1.7 setosa
          24
                    1.9 setosa
          25
                     1.6 setosa
          26
          27
                    1.5 setosa
          28
                    1.4 setosa
          29
                    1.6 setosa
                     1.6 setosa
          31
                     1.5 setosa
                    1.5 setosa
```

| 33 | 1.4 | setosa |
|----|-----|--------|
| 34 | 1.5 | setosa |
| 35 | 1.2 | selosa |
| 36 | 1.3 | setosa |
| 37 | 1.4 | setosa |
| 38 | 1.3 | setosa |
| 39 | 1.5 | setosa |
| 40 | 1.3 | setosa |
| 41 | 1.3 | setosa |
| 42 | 1.3 | setosa |
| 43 | 1.6 | setosa |
| 44 | 1.9 | setosa |
| 45 | 1.4 | setosa |
| 46 | 1.6 | setosa |
| 47 | 1.4 | setosa |
| 48 | 1.5 | setosa |
| 49 | 1.4 | setosa |

## Count the number of records for the Virginica species where petal\_length is more than 5.

```
In [22]: df = data[['petal_length','species']]
In [23]: df1 = df.loc[(df['petal_length']>5)]
In [24]: (df1['species']=='virginica').count()
# 42 records for the Virginica species where petal_length is more than 5
Out[24]: 42
```

### Find out what is the maximum median petal\_length.

```
In [25]: data.loc[(data['species']=='setosa')].median()
Out[25]: sepal_length
                         5.0
         sepal width
                         3.4
         petal length
                        1.5
         petal width
                         0.2
         dtype: float64
In [26]: data.loc[(data['species']=='versicolor')].median()
Out[26]: sepal length
                         5.90
         sepal_width
                         2.80
         petal_length
                         4.35
         petal_width
                         1.30
         dtype: float64
        data.loc[(data['species']=='virginica')].median()
In [27]:
Out[27]: sepal_length
                         6.50
         sepal width
                         3.00
         petal_length
                         5.55
         petal width
                         2.00
         dtype: float64
In [ ]: # 5.55 is the maximum median petal_length of virginica species
```