

Abdurakhmonova Mushtariy 201939183

Homework 2. Working with Iris dataset

Using Iris.csv dataset:

- 1) read dataset based on pandas
- 2) dataset description
- 3) calculate missing values on the dataset
- 4) replace missing values
- 5) prepare a bar chart

Reading a dataset based on Pandas

```
In [41]: import pandas as pd
df_iris = pd.read_csv('dataset/Iris.csv')
```

Dataset description

```
In [42]: df_iris.loc[:]
```

Out[42]:

	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	NaN	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	NaN	NaN	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Total number of data - 150 Number of classes - 3

- Iris-setosa
- Iris-versicolor
- Iris-virginica

Number of features - 4

- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm

Calculating missing values on the dataset

Checking if there is a NaN values.

```
In [44]: df_iris.isnull().values.any()
```

```
Out[44]: True
```

```
In [46]: missing_values = df_iris.isnull().sum().sum()
```

```
In [47]: print(f'Total missing values are: {missing_values}')
```

```
Total missing values are: 18
```

Replacing missing values

Replacing NaN values for single column

```
In [49]: df_iris['SepalLengthCm'] = df_iris['SepalLengthCm'].fillna(0)
```

Checking if there is a NaN values in a 'SepalLengthCm' column.

```
In [50]: df_iris['SepalLengthCm'].isnull().values.any()
```

```
Out[50]: False
```

Checking the number of NaN values

```
In [51]: df_iris['SepalLengthCm'].isnull().sum().sum()
```

```
Out[51]: 0
```

Replacing all NaN values with zeros(0)

```
In [52]: df_iris = df_iris.fillna(0)
```

Checking if there is a NaN values.

```
In [53]: df_iris.isnull().values.any()
```

```
Out[53]: False
```

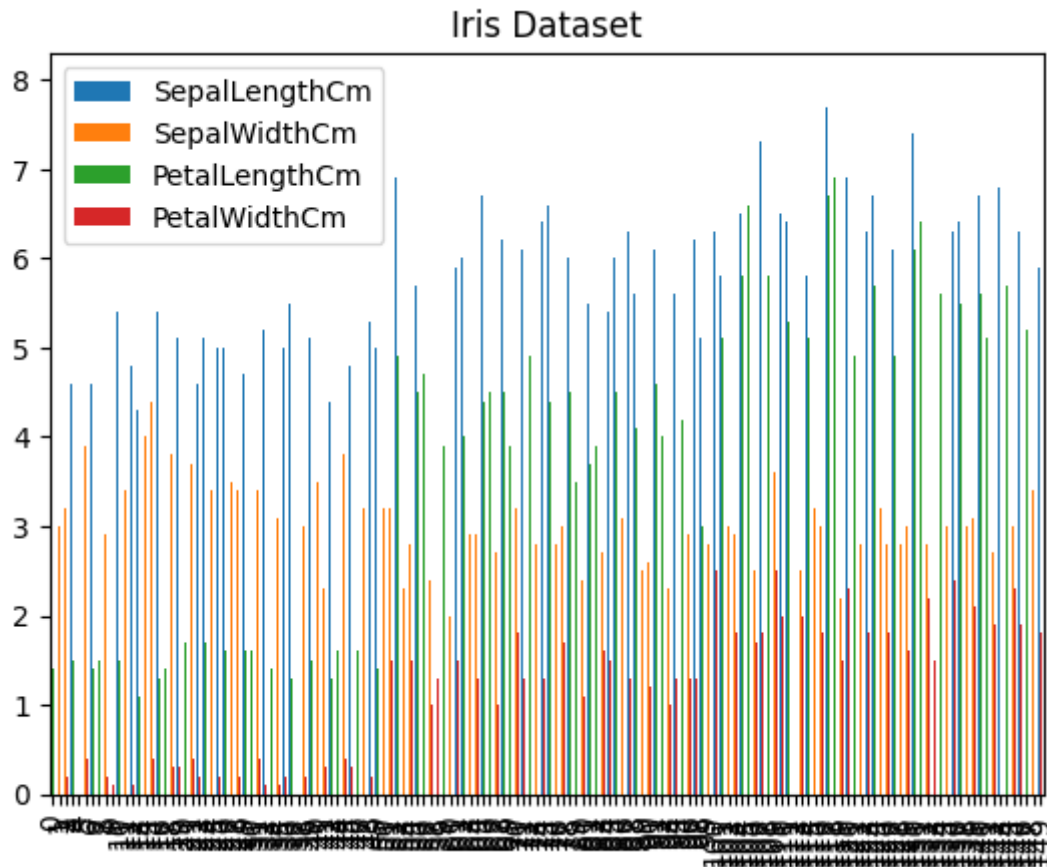
```
In [54]: print(df_iris.to_string())
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	
Species						
0	1	5.1	3.5	1.4	0.2	I
ris-setosa						
1	2	4.9	3.0	1.4	0.2	I
ris-setosa						
2	3	4.7	3.2	1.3	0.2	I
ris-setosa						
3	4	4.6	3.1	1.5	0.2	I
ris-setosa						
4	5	5.0	3.6	0.0	0.2	I
ris-setosa						
5	6	5.4	3.9	1.7	0.4	I
ris-setosa						
6	7	4.6	3.4	1.4	0.3	I
ris-setosa						
7	8	5.0	3.4	1.5	0.2	I
ris-setosa						
8	9	4.4	2.9	1.4	0.2	I
ris-setosa						

Preparing a bar chart

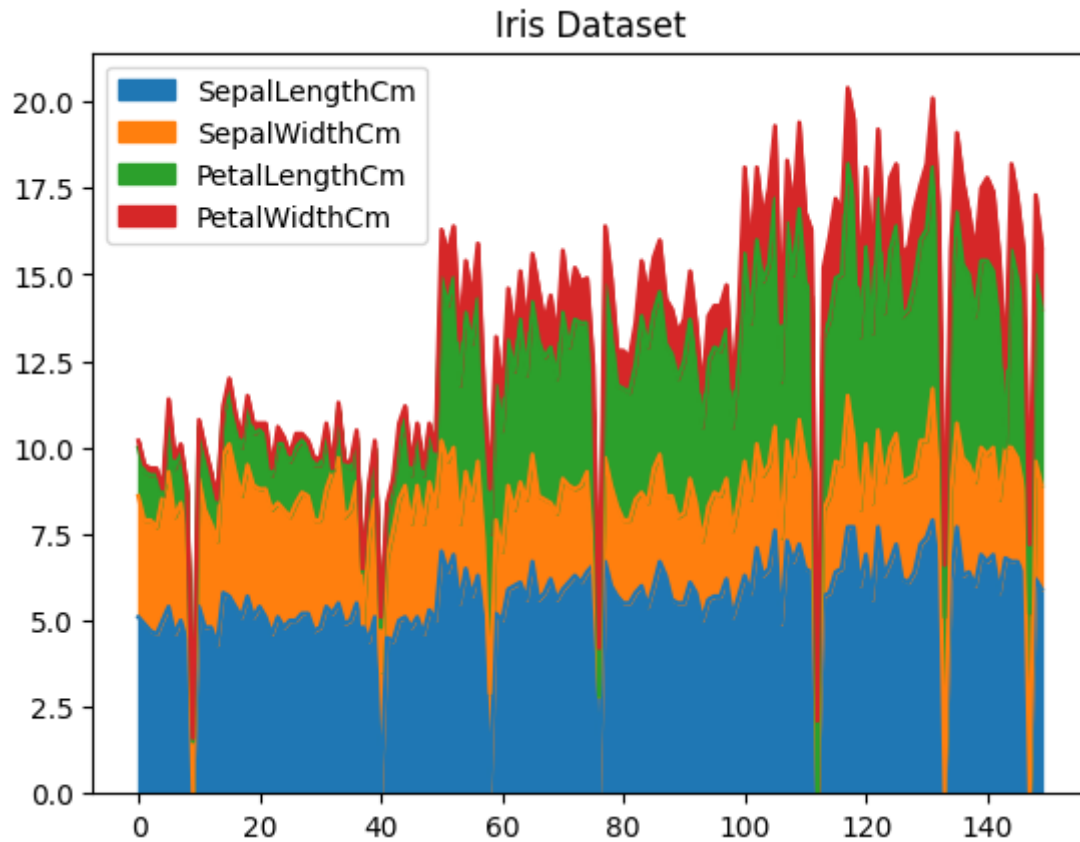
```
In [55]: df_iris.plot(kind = 'bar', y = ["SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", "Species"], title='Iris Dataset')
```

```
Out[55]: <AxesSubplot:title={'center':'Iris Dataset'}>
```



```
In [56]: df_iris.plot(kind = 'area', y = ["SepalLengthCm", "SepalWidthCm", "PetalLeng",  
      "PetalWidthCm", "Species"], title='Iris Dataset')
```

```
Out[56]: <AxesSubplot:title={'center':'Iris Dataset'}>
```



```
In [ ]:
```