

## Practical-1

```
[503]: import pandas as pd
```

```
[504]: dataset = pd.read_csv("/home/sai/Documents/Iris.csv")
```

```
[505]: dataset.head(n=5)
```

```
[505]:
```

	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

```
[506]: dataset.tail(n=5)
```

```
[506]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
145	146	6.7	3.0	5.2	2.3	
146	147	6.3	2.5	5.0	1.9	
147	148	6.5	3.0	5.2	2.0	
148	149	6.2	3.4	5.4	2.3	
149	150	5.9	3.0	5.1	1.8	

  

	Species
145	Iris-virginica
146	Iris-virginica
147	Iris-virginica
148	Iris-virginica
149	Iris-virginica

```
[507]: dataset.columns
```

```
[507]: Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',  
        'Species'],  
        dtype='object')
```

```
[508]: dataset.index
```

```
[508]: RangeIndex(start=0, stop=150, step=1)
```

```
[509]: dataset.shape
```

```
[509]: (150, 6)
```

```
[510]: dataset.dtypes
```

```
[510]: Id                int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species          object
dtype: object
```

```
[511]: dataset['Species'].unique()
```

```
[511]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

```
[512]: dataset.describe(include='all')
```

```
[512]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
count	150.000000	150.000000	150.000000	150.000000	150.000000	
unique	NaN	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	NaN	
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	
75%	112.750000	6.400000	3.300000	5.100000	1.800000	
max	150.000000	7.900000	4.400000	6.900000	2.500000	

  

	Species
count	150
unique	3
top	Iris-setosa
freq	50
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

```
[513]: dataset.sort_index(axis=1, ascending=False)
```

```
[513]:
```

	Species	SepalWidthCm	SepalLengthCm	PetalWidthCm	PetalLengthCm	\
0	Iris-setosa	3.5	5.1	0.2	1.4	
1	Iris-setosa	3.0	4.9	0.2	1.4	
2	Iris-setosa	3.2	4.7	0.2	1.3	
3	Iris-setosa	3.1	4.6	0.2	1.5	
4	Iris-setosa	3.6	5.0	0.2	1.4	

```

..      ...      ...      ...      ...
145 Iris-virginica      3.0      6.7      2.3      5.2
146 Iris-virginica      2.5      6.3      1.9      5.0
147 Iris-virginica      3.0      6.5      2.0      5.2
148 Iris-virginica      3.4      6.2      2.3      5.4
149 Iris-virginica      3.0      5.9      1.8      5.1

```

```

      Id
0      1
1      2
2      3
3      4
4      5

```

```

..      ...
145 146
146 147
147 148
148 149
149 150

```

[150 rows x 6 columns]

```
[514]: dataset.sort_values(by='SepalLengthCm')
```

```

[514]:      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
13   14           4.3           3.0           1.1           0.1
38   39           4.4           3.0           1.3           0.2
8     9           4.4           2.9           1.4           0.2
42   43           4.4           3.2           1.3           0.2
41   42           4.5           2.3           1.3           0.3
..   ...           ...           ...           ...           ...
117 118           7.7           3.8           6.7           2.2
118 119           7.7           2.6           6.9           2.3
122 123           7.7           2.8           6.7           2.0
135 136           7.7           3.0           6.1           2.3
131 132           7.9           3.8           6.4           2.0

```

```

      Species
13   Iris-setosa
38   Iris-setosa
8     Iris-setosa
42   Iris-setosa
41   Iris-setosa
..   ...
117  Iris-virginica
118  Iris-virginica
122  Iris-virginica

```

```

135  Iris-virginica
131  Iris-virginica

[150 rows x 6 columns]

```

```
[515]: dataset.columns.values
```

```
[515]: array(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm',
        'PetalWidthCm', 'Species'], dtype=object)
```

```
[516]: dataset.describe(include='all')
```

```
[516]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
count	150.000000	150.000000	150.000000	150.000000	150.000000	
unique	NaN	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	NaN	
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	
75%	112.750000	6.400000	3.300000	5.100000	1.800000	
max	150.000000	7.900000	4.400000	6.900000	2.500000	

  

	Species
count	150
unique	3
top	Iris-setosa
freq	50
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

```
[517]: dataset['SepalLengthCm']
```

```
[517]: 0      5.1
      1      4.9
      2      4.7
      3      4.6
      4      5.0
      ...
     145     6.7
```

```

146    6.3
147    6.5
148    6.2
149    5.9
Name: SepalLengthCm, Length: 150, dtype: float64

```

```
[518]: dataset.sort_index(axis=1, ascending=False)
```

```
[518]:
```

	Species	SepalWidthCm	SepalLengthCm	PetalWidthCm	PetalLengthCm	\
0	Iris-setosa	3.5	5.1	0.2	1.4	
1	Iris-setosa	3.0	4.9	0.2	1.4	
2	Iris-setosa	3.2	4.7	0.2	1.3	
3	Iris-setosa	3.1	4.6	0.2	1.5	
4	Iris-setosa	3.6	5.0	0.2	1.4	
..	...	...	...	...	...	
145	Iris-virginica	3.0	6.7	2.3	5.2	
146	Iris-virginica	2.5	6.3	1.9	5.0	
147	Iris-virginica	3.0	6.5	2.0	5.2	
148	Iris-virginica	3.4	6.2	2.3	5.4	
149	Iris-virginica	3.0	5.9	1.8	5.1	

```

      Id
0      1
1      2
2      3
3      4
4      5
..    ...
145  146
146  147
147  148
148  149
149  150

```

```
[150 rows x 6 columns]
```

```
[519]: dataset.iloc[5]
```

```
[519]: Id          6
SepalLengthCm    5.4
SepalWidthCm     3.9
PetalLengthCm    1.7
PetalWidthCm     0.4
Species          Iris-setosa
Name: 5, dtype: object
```

```
[520]: dataset.iloc[0:3]
```

```
[520]:
```

	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

```
[521]: dataset.loc[:, ["SepalLengthCm", "SepalWidthCm"]]
```

```
[521]:
```

	SepalLengthCm	SepalWidthCm
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6
..	...	...
145	6.7	3.0
146	6.3	2.5
147	6.5	3.0
148	6.2	3.4
149	5.9	3.0

[150 rows x 2 columns]

```
[522]: n = 15 # Replace with your desired value
dataset.iloc[:n, :]
```

```
[522]:
```

	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
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa

```
[523]: n = 3 # Replace with your desired value
dataset.iloc[:, :n]
```

```
[523]:
```

	Id	SepalLengthCm	SepalWidthCm
0	1	5.1	3.5
1	2	4.9	3.0

2	3	4.7	3.2
3	4	4.6	3.1
4	5	5.0	3.6
...	...	...	...
145	146	6.7	3.0
146	147	6.3	2.5
147	148	6.5	3.0
148	149	6.2	3.4
149	150	5.9	3.0

[150 rows x 3 columns]

```
[524]: m = 5 # Replace with your desired value
n = 2 # Replace with your desired value
dataset.iloc[:m, :n]
```

```
[524]:   Id  SepalLengthCm
0    1             5.1
1    2             4.9
2    3             4.7
3    4             4.6
4    5             5.0
```

```
[525]: dataset.iloc[3:5, 0:2]
```

```
[525]:   Id  SepalLengthCm
3    4             4.6
4    5             5.0
```

```
[526]: dataset.iloc[[1, 2, 4], [0, 2]]
```

```
[526]:   Id  SepalWidthCm
1    2             3.0
2    3             3.2
4    5             3.6
```

```
[527]: dataset.iloc[[1, 2, 4], [0, 2]]
```

```
[527]:   Id  SepalWidthCm
1    2             3.0
2    3             3.2
4    5             3.6
```

```
[528]: dataset.iloc[1:3, :]
```

```
[528]:   Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
1    2             4.9             3.0             1.4             0.2  Iris-setosa
2    3             4.7             3.2             1.3             0.2  Iris-setosa
```

```
[529]: dataset.iloc[:, 1:3]
```

```
[529]:      SepalLengthCm  SepalWidthCm
0           5.1           3.5
1           4.9           3.0
2           4.7           3.2
3           4.6           3.1
4           5.0           3.6
..          ...           ...
145          6.7           3.0
146          6.3           2.5
147          6.5           3.0
148          6.2           3.4
149          5.9           3.0
```

[150 rows x 2 columns]

```
[530]: dataset.iloc[1, 1]
```

```
[530]: 4.9
```

```
[531]: dataset['SepalLengthCm'].iloc[5]
```

```
[531]: 5.4
```

```
[532]: dataset.head()
```

```
[532]:   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
```

```
[533]: dataset.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

```
[534]: dataset.describe()
```

```
[534]:
```

	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
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
[535]: dataset['Species'].unique()
```

```
[535]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

```
[536]: dataset['Species'].value_counts()
```

```
[536]: Species
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
Name: count, dtype: int64
```

```
[537]: dataset[['SepalLengthCm', 'SepalWidthCm']]
```

```
[537]:
```

	SepalLengthCm	SepalWidthCm
0	5.1	3.5
1	4.9	3.0
2	4.7	3.2
3	4.6	3.1
4	5.0	3.6
..	...	...
145	6.7	3.0
146	6.3	2.5
147	6.5	3.0
148	6.2	3.4
149	5.9	3.0

[150 rows x 2 columns]

```
[538]: dataset.loc[3:7, ['SepalLengthCm', 'PetalLengthCm']]
```

```
[538]:
```

	SepalLengthCm	PetalLengthCm
3	4.6	1.5
4	5.0	1.4

5	5.4	1.7
6	4.6	1.4
7	5.0	1.5

```
[539]: cols_2_4 = dataset.columns[2:4]
print(cols_2_4)
```

```
Index(['SepalWidthCm', 'PetalLengthCm'], dtype='object')
```

```
[540]: data_cols_2_4 = dataset[cols_2_4]
print(data_cols_2_4)
```

	SepalWidthCm	PetalLengthCm
0	3.5	1.4
1	3.0	1.4
2	3.2	1.3
3	3.1	1.5
4	3.6	1.4
..	...	...
145	3.0	5.2
146	2.5	5.0
147	3.0	5.2
148	3.4	5.4
149	3.0	5.1

[150 rows x 2 columns]

```
[541]: selected_data = data_cols_2_4.iloc[4:10]
print(selected_data)
```

	SepalWidthCm	PetalLengthCm
4	3.6	1.4
5	3.9	1.7
6	3.4	1.4
7	3.4	1.5
8	2.9	1.4
9	3.1	1.5

```
[542]: dataset.isnull()
```

```
[542]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
..	...	...	...	...	...	...
145	False	False	False	False	False	False
146	False	False	False	False	False	False

147	False	False	False	False	False	False
148	False	False	False	False	False	False
149	False	False	False	False	False	False

[150 rows x 6 columns]

```
[543]: dataset.isnull().any()
```

```
[543]: Id                False
SepalLengthCm          False
SepalWidthCm           False
PetalLengthCm          False
PetalWidthCm           False
Species                False
dtype: bool
```

```
[544]: dataset.isnull().sum()
```

```
[544]: Id                0
SepalLengthCm          0
SepalWidthCm           0
PetalLengthCm          0
PetalWidthCm           0
Species                0
dtype: int64
```

```
[545]: dataset.isnull().sum(axis=1)
```

```
[545]: 0      0
1      0
2      0
3      0
4      0
..
145    0
146    0
147    0
148    0
149    0
Length: 150, dtype: int64
```

```
[546]: dataset['SepalLengthCm'].isnull().sum()
```

```
[546]: 0
```

```
[547]: dataset.groupby(['Species'])['SepalLengthCm'].apply(lambda x: x.isnull().sum())
```

```
[547]: Species
      Iris-setosa      0
      Iris-versicolor  0
      Iris-virginica   0
      Name: SepalLengthCm, dtype: int64
```

```
[548]: dataset['PetalLengthCm'] = dataset['PetalLengthCm'].astype(int)
      print(dataset['PetalLengthCm'])
```

```
0      1
1      1
2      1
3      1
4      1
..
145    5
146    5
147    5
148    5
149    5
      Name: PetalLengthCm, Length: 150, dtype: int64
```

```
[549]: #b. Data normalization
```

```
[550]: from sklearn import preprocessing
```

```
[551]: dataset.head()
```

```
[551]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1	0.2	Iris-setosa
1	2	4.9	3.0	1	0.2	Iris-setosa
2	3	4.7	3.2	1	0.2	Iris-setosa
3	4	4.6	3.1	1	0.2	Iris-setosa
4	5	5.0	3.6	1	0.2	Iris-setosa

```
[552]: x = dataset[['SepalLengthCm']].values.astype(float)
```

```
[553]: min_max_scaler = preprocessing.MinMaxScaler()
```

```
[554]: x_scaled = min_max_scaler.fit_transform(x)
```

```
[555]: df_normalized = pd.DataFrame(x_scaled)
```

```
[556]: df_normalized
```

```
[556]:
```

	0
0	0.222222
1	0.166667

```

2    0.111111
3    0.083333
4    0.194444
..    ...
145  0.666667
146  0.555556
147  0.611111
148  0.527778
149  0.444444

```

```
[150 rows x 1 columns]
```

```
[557]: #label encoder
from sklearn import preprocessing
```

```
[558]: df = dataset
```

```
[559]: df['Species'].unique()
```

```
[559]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

```
[560]: label_encoder = preprocessing.LabelEncoder()
```

```
[561]: df['Species'] = label_encoder.fit_transform(df['Species'])
```

```
[562]: df['Species'].unique()
```

```
[562]: array([0, 1, 2])
```

```
[ ]:
```

```
[563]: #b. One-Hot Encoding:
```

```
[564]: from sklearn import preprocessing
```

```
[565]: df['Species'].unique()
```

```
[565]: array([0, 1, 2])
```

```
[566]: df['Species'] = label_encoder.fit_transform(df['Species'])
```

```
[567]: features_df = df.drop(columns=['Species'])
```

```
[568]: enc = preprocessing.OneHotEncoder()
enc_df = pd.DataFrame(enc.fit_transform(df[['Species']]).toarray())
```

```
[569]: df_encode = features_df.join(enc_df)
```

```
[570]: df_encode
```

```
[570]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	0	1	\
0	1	5.1	3.5	1	0.2	1.0	0.0	
1	2	4.9	3.0	1	0.2	1.0	0.0	
2	3	4.7	3.2	1	0.2	1.0	0.0	
3	4	4.6	3.1	1	0.2	1.0	0.0	
4	5	5.0	3.6	1	0.2	1.0	0.0	
...	...	...	...	...	...	...	...	
145	146	6.7	3.0	5	2.3	0.0	0.0	
146	147	6.3	2.5	5	1.9	0.0	0.0	
147	148	6.5	3.0	5	2.0	0.0	0.0	
148	149	6.2	3.4	5	2.3	0.0	0.0	
149	150	5.9	3.0	5	1.8	0.0	0.0	

```
2
```

0	0.0
1	0.0
2	0.0
3	0.0
4	0.0
...	...
145	1.0
146	1.0
147	1.0
148	1.0
149	1.0

```
[150 rows x 8 columns]
```

```
[571]: df_encode.rename(columns={0: 'Iris-Setosa', 1: 'Iris-Versicolor', 2: 'Iris-virginica'}, inplace=True)
```

```
[572]: # Step 10: Observe the merged DataFrame
df_encode
```

```
[572]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	\
0	1	5.1	3.5	1	0.2	
1	2	4.9	3.0	1	0.2	
2	3	4.7	3.2	1	0.2	
3	4	4.6	3.1	1	0.2	
4	5	5.0	3.6	1	0.2	
...	...	...	...	...	...	
145	146	6.7	3.0	5	2.3	
146	147	6.3	2.5	5	1.9	
147	148	6.5	3.0	5	2.0	
148	149	6.2	3.4	5	2.3	

```
149 150          5.9          3.0          5          1.8
```

```

      Iris-Setosa  Iris-Versicolor  Iris-virginica
0             1.0             0.0             0.0
1             1.0             0.0             0.0
2             1.0             0.0             0.0
3             1.0             0.0             0.0
4             1.0             0.0             0.0
..           ...             ...             ...
145           0.0             0.0             1.0
146           0.0             0.0             1.0
147           0.0             0.0             1.0
148           0.0             0.0             1.0
149           0.0             0.0             1.0

```

```
[150 rows x 8 columns]
```

```
[ ]:
```

```
[573]: #C. Dummy Variable Encoding
```

```
[574]: df['Species'].unique()
```

```
[574]: array([0, 1, 2])
```

```
[575]: label_encoder = preprocessing.LabelEncoder()
df['Species'] = label_encoder.fit_transform(df['Species'])
```

```
[576]: one_hot_df = pd.get_dummies(df, prefix="Species", columns=['Species'],
↳drop_first=False)
```

```
[577]: one_hot_df
```

```
[577]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species_0 \
0	1	5.1	3.5	1	0.2	True
1	2	4.9	3.0	1	0.2	True
2	3	4.7	3.2	1	0.2	True
3	4	4.6	3.1	1	0.2	True
4	5	5.0	3.6	1	0.2	True
..	...	...	...	...	...	...
145	146	6.7	3.0	5	2.3	False
146	147	6.3	2.5	5	1.9	False
147	148	6.5	3.0	5	2.0	False
148	149	6.2	3.4	5	2.3	False
149	150	5.9	3.0	5	1.8	False

```
Species_1 Species_2
```

0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
..	...	...
145	False	True
146	False	True
147	False	True
148	False	True
149	False	True

[150 rows x 8 columns]

[ ]:

[ ]: