

Assignment-2 [PANDAS & SKLEARN]

```
In[6] : import pandas as pd
```

```
In[7] : data = pd.read_excel('test_data.xlsx')
```

```
In[8] : data.head()
```

	f ₁	f ₂	f ₃	f ₄
0	1	2	3.0	4.0
1	5	6	NaN	7.0
2	0	6	9.0	NaN

```
In[9] : data.isnull().sum()
```

f ₁	0
f ₂	0
f ₃	1
f ₄	1

```
dtype: int64
```

```
In[10] : data.dropna()
```

	f ₁	f ₂	f ₃	f ₄
0	1	2	3.0	4.0

```
In[11] : data.dropna(axis=1)
```

	f ₁	f ₂
0	1	2
1	5	6
2	0	6

```

In[18]: import numpy as np
from sklearn.impute import SimpleImputer
imr = SimpleImputer(missing_values=np.nan, strategy="mean")
imr = imr.fit(data)
imputed_data = imr.transform(data)
print(data)
print(imputed_data)

```

	f_1	f_2	f_3	f_4
0	1	2	3.0	4.0
1	5	6	NaN	7.0
2	0	6	9.0	NaN

```

[[ 1.  2.  3.  4.]
 [ 5.  6.  6.  7.]
 [ 0.  6.  9.  5.5.]]

```

```

In[19]: data = pd.read_csv('iris.csv')
data.head()

```

	5.1	3.5	1.4	0.2	Iris-setosa.
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa

```
In[26]: data = pd.read_csv('iris.csv', header=None)
        data.head()
```

	0	1	2	3	4
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

```
In[27]: data.columns = ['sepal length', 'sepal width', 'petal length', 'petal width',
                        'class']
```

```
data.head()
```

	sepal length	sepal width	petal length	petal width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

```
In[28]: np.unique(data['class']).map(mapping)
        data.head()
```

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

```

In[29]: mapping = {'Iris-setosa': 0,
                  'Iris-versicolour': 1,
                  'Iris-virginica': 2}
data['class'] = data['class'].map(mapping)
data.head()

```

	sepal length	sepal width	petal length	petal width	class
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```

In[33]: from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
data['class'] = le.fit_transform(data['class'])
data.head()

```

	sepal length	sepal width	petal length	petal width	class
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0