Data Cleaning fixing Duplicate values

July 28, 2023

```python
[1]: import numpy as np
     import pandas as pd
```

```python
[2]: df=pd.read_csv("C:\\Users\\ASUS\\Desktop\\DS Course\Data Cleaning\\Day19\\iris␣
     ↪- Duplicate.csv")
```

```python
[3]: df.head()
```

```
[3]:       Id  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width Species
     0  1001           5.1          3.5           1.4          0.2  setosa
     1  1002           4.9          3.0           1.4          0.2  setosa
     2  1002           4.9          3.0           1.4          0.2  setosa
     3  1003           4.7          3.2           1.3          0.2  setosa
     4  1004           4.6          3.1           1.5          0.2  setosa
```

```python
[5]: df.duplicated().sum()
```

```
[5]: 3
```

```python
[6]: df.drop_duplicates(inplace=True) #remove duplocates
```

```python
[8]: df.duplicated().sum()
```

```
[8]: 0
```

```python
[9]: df.head()
```

```
[9]:       Id  Sepal.Length  Sepal.Width  Petal.Length  Petal.Width Species
     0  1001           5.1          3.5           1.4          0.2  setosa
     1  1002           4.9          3.0           1.4          0.2  setosa
     3  1003           4.7          3.2           1.3          0.2  setosa
     4  1004           4.6          3.1           1.5          0.2  setosa
     5  1005           5.0          3.6           1.4          0.2  setosa
```

```python
[ ]:
```