

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns

In [3]: df=pd.read_csv("Technohack task 3 Remove duplication.csv")

In [5]: df.head()
```

Out[5]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	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

```
In [6]: df.shape
```

Out[6]:

(150, 6)

```
In [7]: df.drop_duplicates()
```

Out[7]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	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
...
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	6.5	3.0	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

```
In [12]: df.drop_duplicates(subset=['Id'])
```

Out[12]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	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
...
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	6.5	3.0	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

```
In [13]: df.drop_duplicates(subset=['Species'])
```

Out[13]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
50	51	7.0	3.2	4.7	1.4	Iris-versicolor
100	101	6.3	3.3	6.0	2.5	Iris-virginica

```
In [20]: df.drop_duplicates(subset=['Petal Length Cm','Petal Width Cm'])
```

Out[20]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	Species
0	1	5.1	3.5	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
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	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	6.5	3.0	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

102 rows × 6 columns

```
In [16]: df.drop_duplicates(subset=['Sepal Length Cm','Sepal Width Cm'])
```

Out[16]:

	Id	Sepal Length Cm	Sepal Width Cm	Petal Length Cm	Petal Width Cm	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
...
136	137	6.3	3.4	5.6	2.4	Iris-virginica
137	138	6.4	3.1	5.5	1.8	Iris-virginica
138	139	6.0	3.0	4.8	1.8	Iris-virginica
143	144	6.8	3.2	5.9	2.3	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica

116 rows × 6 columns

For the data analyzation and visualization, drop the id column (unique) in this datasets. Try to identify the unique column.

In []: