

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('/content/Iris.csv')
df
```



	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
...	...	...	...	...	...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

149 rows × 5 columns

Next steps:

[Generate code with df](#)
[View recommended plots](#)
[New interactive sheet](#)

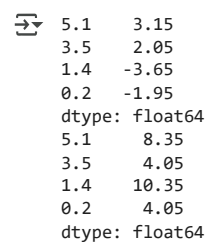
```
# Select numerical columns
numeric_columns = df.select_dtypes(include=['float64', 'int64']).columns
```

## ✓ Detect the outlier using Inter Quantile Range(IQR)

```
# Detect outliers using IQR method
Q1 = df[numeric_columns].quantile(0.25)
Q3 = df[numeric_columns].quantile(0.75)
IQR = Q3 - Q1
```

```
# Define outlier thresholds
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
```

```
print(lower_bound)
print(upper_bound)
```

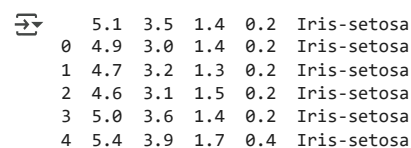


```
5.1    3.15
3.5    2.05
1.4   -3.65
0.2   -1.95
dtype: float64
5.1    8.35
3.5    4.05
1.4   10.35
0.2    4.05
dtype: float64
```

## ✓ remove the outliers from the dataset.

```
# Remove outliers
df_cleaned = df[~((df[numeric_columns] < lower_bound) | (df[numeric_columns] > upper_bound)).any(axis=1)]
```

```
# Print updated dataset
print(df_cleaned.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

Start coding or [generate](#) with AI.