```
from sklearn.ensemble import IsolationForest
import seaborn as sns
import pandas as pd
import numpy as np

data=sns.load_dataset('iris')
```

data

→		sepal_length	sepal_width	petal_length	petal_width	species
	0	5.1	3.5	1.4	0.2	setosa
	1	4.9	3.0	1.4	0.2	setosa
	2	4.7	3.2	1.3	0.2	setosa
	3	4.6	3.1	1.5	0.2	setosa
	4	5.0	3.6	1.4	0.2	setosa
	145	6.7	3.0	5.2	2.3	virginica
	146	6.3	2.5	5.0	1.9	virginica
	147	6.5	3.0	5.2	2.0	virginica
	148	6.2	3.4	5.4	2.3	virginica
	149	5.9	3.0	5.1	1.8	virginica
	150 rd	ows × 5 columns				

```
import seaborn as sns
import pandas as pd
import numpy as np
from scipy.stats import zscore

data = sns.load_dataset('iris')

z_score = np.abs(zscore(data.drop('species', axis=1)))

data
```

}	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
		•••		•••	
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica
150 r	ows × 5 columns				

zscore(data.drop('species', axis=1))

3	sepal_length	sepal_width	petal_length	petal_width
0	-0.900681	1.019004	-1.340227	-1.315444
1	-1.143017	-0.131979	-1.340227	-1.315444
2	-1.385353	0.328414	-1.397064	-1.315444
3	-1.506521	0.098217	-1.283389	-1.315444
4	-1.021849	1.249201	-1.340227	-1.315444
145	1.038005	-0.131979	0.819596	1.448832
146	0.553333	-1.282963	0.705921	0.922303
147	0.795669	-0.131979	0.819596	1.053935
148	0.432165	0.788808	0.933271	1.448832
149	0.068662	-0.131979	0.762758	0.790671
150	rows × 4 columns			

threshold = 3
non_outliers = (z_score < threshold).all(axis=1)</pre>

data_no_outliers = data[non_outliers]

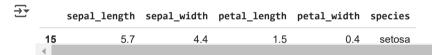
data

3	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica
150 rd	ows × 5 columns				

data_no_outliers

→		sepal_length	sepal_width	petal_length	petal_width	species
	0	5.1	3.5	1.4	0.2	setosa
	1	4.9	3.0	1.4	0.2	setosa
	2	4.7	3.2	1.3	0.2	setosa
	3	4.6	3.1	1.5	0.2	setosa
	4	5.0	3.6	1.4	0.2	setosa
	145	6.7	3.0	5.2	2.3	virginica
	146	6.3	2.5	5.0	1.9	virginica
	147	6.5	3.0	5.2	2.0	virginica
	148	6.2	3.4	5.4	2.3	virginica
	149	5.9	3.0	5.1	1.8	virginica
	149 rd	ows × 5 columns				

outliers = (z_score >= threshold).any(axis=1)
outlier_rows = data[outliers]
outlier_rows



```
import numpy as np
mean = np.mean(data_no_outliers['sepal_length'])
std = np.std(data_no_outliers['sepal_length'])
x = [mean + 3 * std, mean - 3 * std]

from sklearn.ensemble import IsolationForest
import seaborn as sns
import pandas as pd
import numpy as np
from scipy.stats import zscore

data = sns.load_dataset('iris')
```

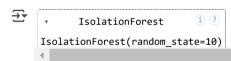
data

_		sepal_length	sepal_width	petal_length	petal_width	species
	0	5.1	3.5	1.4	0.2	setosa
	1	4.9	3.0	1.4	0.2	setosa
	2	4.7	3.2	1.3	0.2	setosa
	3	4.6	3.1	1.5	0.2	setosa
	4	5.0	3.6	1.4	0.2	setosa
	145	6.7	3.0	5.2	2.3	virginica
	146	6.3	2.5	5.0	1.9	virginica
	147	6.5	3.0	5.2	2.0	virginica
	148	6.2	3.4	5.4	2.3	virginica
	149	5.9	3.0	5.1	1.8	virginica
	150 rc	ows × 5 columns				

from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()

data['species']=lb.fit_transform(data['species'])

clf = IsolationForest(random_state=10)
clf.fit(data)



data.head(3)

		sepal_length	sepal_width	petal_length	petal_width	species
	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

y_pred_outliers = clf.predict(data)
y_pred_outliers

https://colab.research.google.com/drive/1cJ1VzSXejs6Rirk2ohjo-D1XIN9Wf6fE#scrollTo=aVqxCYBqeFQ1&printMode=true

idx=np.where(y_pred_outliers==-1)
idx[0]

data.drop(index=[117,131,118],axis=0)

_						
_		sepal_length	sepal_width	petal_length	petal_width	species
	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
	145	6.7	3.0	5.2	2.3	2
	146	6.3	2.5	5.0	1.9	2
	147	6.5	3.0	5.2	2.0	2
	148	6.2	3.4	5.4	2.3	2