

```
from sklearn.ensemble import IsolationForest
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('spirals.csv')

# Visualize the dataset
plt.figure(figsize=(10, 8))
plt.scatter(df['d1'], df['d2'])
plt.title('Dataset')
plt.xlabel('Característica 1')
plt.ylabel('Característica 2')
plt.show()

# 'contamination' estima a proporção de outliers
model = IsolationForest(contamination = 0.003)
model.fit(df)
outliers = model.predict(df)

# -1 representa um outlier e 1 um ponto normal
outlier = np.sum(outliers == -1) # soma outliers
normal = np.sum(outliers != -1) # soma não outliers
print('Normais = %d \nOutliers = %d' %(normal,outlier))

Normais = 1000
Outliers = 4

#outliers = pd.DataFrame(outliers)
df['Outliers'] = outliers
df1 = df[df['Outliers'] == -1]
df1
```