

How can I fix an input that contains NaN, infinity, or a value too large for dtype('float64')?

Authored by
stats writer

June 27, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I fix an input that contains NaN, infinity, or a value too large for dtype('float64')*?. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=154573>

If an input contains values such as NaN, infinity, or a value that is too large for the dtype('float64'), there are several steps that can be taken to fix it. Firstly, check the source of the input and see if any errors were made during data collection or processing. If so, correct those errors and re-input the data. Alternatively, if the data is accurate and the values cannot be changed, consider using a different data type that can handle larger values, such as dtype('float128'). Additionally, you can use functions like `isfinite()` and `isnan()` to identify and handle these special values in your code. Properly handling these types of inputs will ensure accurate and reliable data analysis.

Fix: Input contains NaN, infinity or a value too large for dtype('float64')

One common error you may encounter when using Python is:

ValueError: Input contains infinity or a value too large for dtype('float64').

This error usually occurs when you attempt to use some function from the scikit-learn module, but the DataFrame or matrix you're using as input has NaN values or infinite values.

The following example shows how to resolve this error in practice.

How to Reproduce the Error

Suppose we have the following pandas DataFrame:

```
import pandas as pd
import numpy as np

#create DataFrame
df = pd.DataFrame({'x1': ,
'x2': ,
'y': })
```

```
#view DataFrame
print(df)
```

```
x1 x2 y
0 1 1.0 NaN
1 2 3.0 78.0
2 2 3.0 85.0
3 4 5.0 88.0
4 2 2.0 72.0
5 1 2.0 69.0
6 5 1.0 94.0
7 4 inf 94.0
8 2 0.0 88.0
9 4 3.0 92.0
10 4 4.0 90.0
```

Now suppose we attempt to fit a using functions from :

```
from sklearn.linear_model import LinearRegression
```

```
#initiate linear regression model
```

```
model = LinearRegression()
```

```
#define predictor and response variables
```

```
X, y = df, df.y#fit regression model
```

```
model.fit(X, y)
```

```
#print model intercept and coefficients
```

```
print(model.intercept_, model.coef_)
```

ValueError: Input contains infinity or a value too large for dtype('float64').

We receive an error since the DataFrame we're using has both infinite and NaN values.

How to Fix the Error

The way to resolve this error is to first remove any rows from the DataFrame that contain infinite or NaN values:

```
#remove rows with any values that are not finite
```

```
df_new = df
```

```
#view updated DataFrame  
print(df_new)
```

```
x1 x2 y  
1 2 3.0 78.0  
2 2 3.0 85.0  
3 4 5.0 88.0  
4 2 2.0 72.0  
5 1 2.0 69.0  
6 5 1.0 94.0  
8 2 0.0 88.0  
9 4 3.0 92.0  
10 4 4.0 90.0
```

The two rows that had infinite or NaN values have been removed.

We can now proceed to fit our linear regression model:

```
from sklearn.linear_model import LinearRegression
```

```
#initiate linear regression model  
model = LinearRegression()
```

```
#define predictor and response variables
```

```
X, y = df_new], df_new.y#fit regression model  
model.fit(X, y)
```

```
#print model intercept and coefficients  
print(model.intercept_, model.coef_)
```

```
69.85144124168515
```

Notice that we don't receive any error this time because we first removed the rows with infinite or NaN values from the DataFrame.

The following tutorials explain how to fix other common errors in Python: