Data Cleansing using Pandas

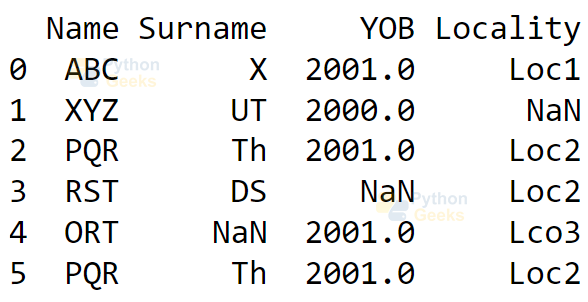
When we are using pandas, we use the data frames. Let us first see the way to load the data frame.  
**Example of loading CSV file as data frame:**

import pandas as pd

data =pd.read\_csv('data.csv')

print(data)

**Output:**

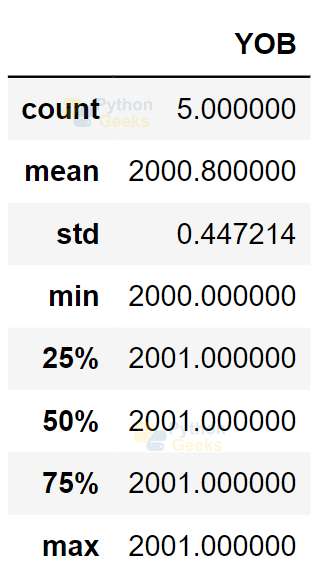
[](https://pythongeeks.org/wp-content/uploads/2021/09/load-csv.png)

Now let us get the information about the data using the describe() and rank() functions.

**Example of describe() function:**

data.describe()

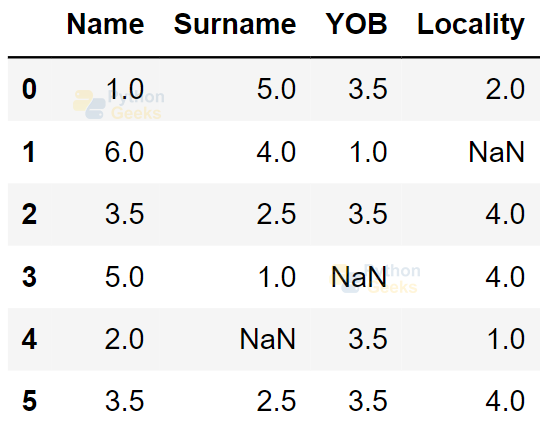
**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/describe.png)

**Example of rank() function:**

data.rank()

**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/rank.png)

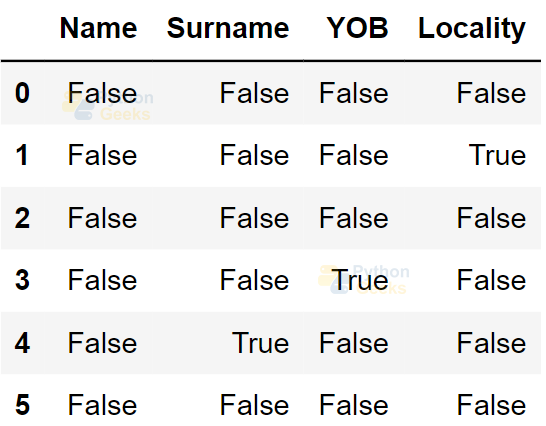
Now let us see different operations we can use on the data frame.

1. Finding and Removing Missing Values

We can find the missing values using isnull() function.  
**Example of finding missing values:**

data.isnull()

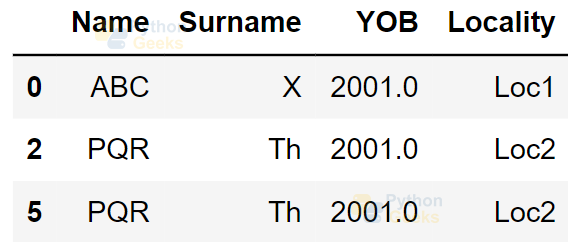
**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/find-missing-values.png)

**Example of removing missing values:**

data.dropna()

**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/remove-missing-values.png)

2. Replacing Missing Values

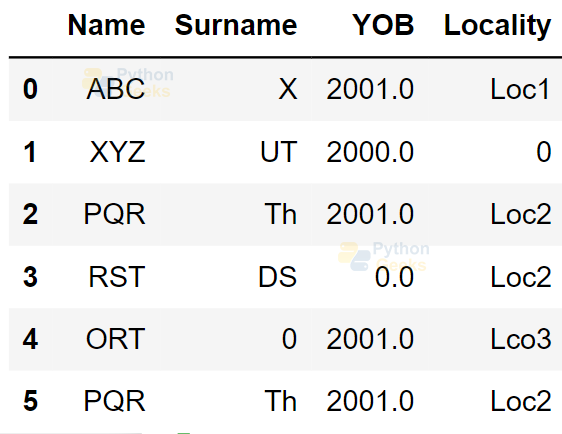
We have different options for replacing the missing values. We can use the replace() function or fillna() function to replace it with a constant value.

**Example of replacing missing values using replace():**

from numpy import NaN

data.replace({NaN:0.00})

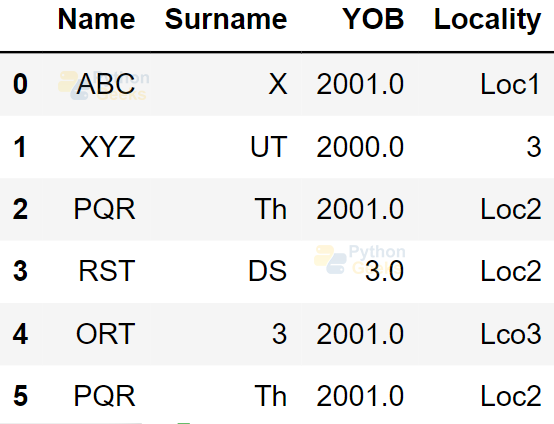
**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/replace-missing-values.png)

**Example of replacing missing values using fillna():**

data.fillna(3)

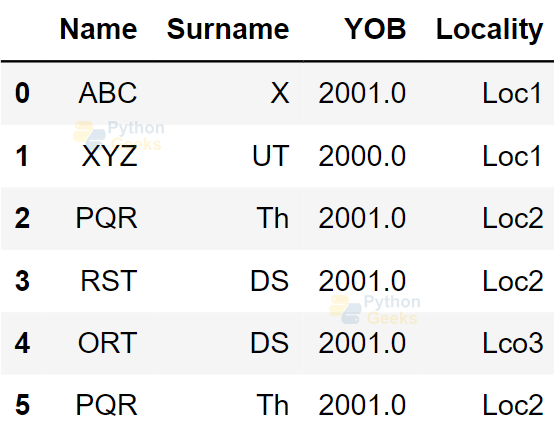
**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/fillna.png)

Using fillna() function, we can fill forward and fill backward as well.  
**Example of replacing missing values by filling forward :**

data.fillna(method='pad')

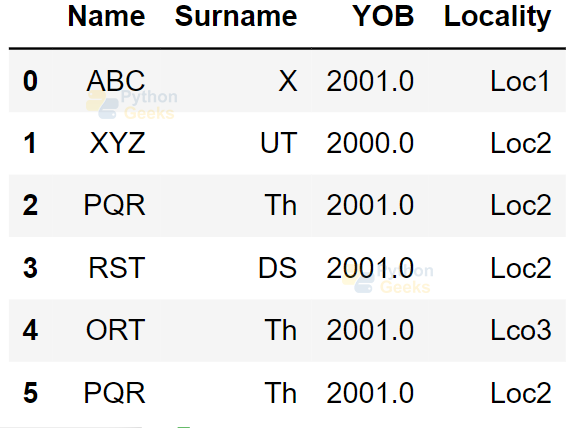
**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/filling-forward.png)

**Example of replacing missing values by filling backward:**

data.fillna(method='backfill')

**Output:**

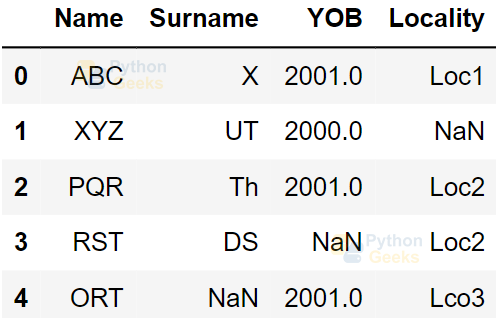
[](https://pythongeeks.org/wp-content/uploads/2021/09/filling-backward.png)

3. Removing Repeated Values

We can remove the repeated values by using the drop\_duplicates() method.  
**Example of removing repeated values:**

data.drop\_duplicates()

**Output:**

[](https://pythongeeks.org/wp-content/uploads/2021/09/remove-repeated.png)

4. Removing Irrelevant Data

We can remove the irrelevant data by using the del method.

**Example of removing irrelevant data:**

del data['YOB']

print(data)

**Output:**

A screenshot of a computer code

Description automatically generated

5. Renaming Columns

We have a function rename() to rename the columns.

**Example of renaming columns:**

print(data.rename(columns={'Name':'FirstName','Surname':'LastName'}))

**Output:**

A close up of a document

Description automatically generated