

How can I create a Pandas DataFrame from a dictionary with different lengths?

Authored by
stats writer

June 25, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a Pandas DataFrame from a dictionary with different lengths?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=151587>

To create a Pandas DataFrame from a dictionary with different lengths, one can use the 'from_dict' function in Pandas. This function allows for the creation of a DataFrame from a dictionary, where the keys become the column labels and the values become the data in each column. In the case of a dictionary with varying lengths, the missing values will automatically be filled with 'NaN' (Not a Number) in the DataFrame. This approach allows for the efficient organization and manipulation of data with different lengths in a tabular format.

Pandas: Create DataFrame from dict with Different Lengths

You can use the following basic syntax to create a pandas DataFrame from a dictionary whose entries have different lengths:

```
import pandas as pd
```

```
df = pd.DataFrame(dict())
```

This syntax converts a list of arrays in the dictionary into a list of pandas Series.

This allows us to create a pandas DataFrame and simply fill in NaN values to ensure that each column in the resulting DataFrame is the same length.

The following example shows how to use this syntax in practice.

Example: Create Pandas DataFrame from dict with Different Lengths

Suppose we have the following dictionary that contains entries with different lengths:

```
#create dictionary whose entries have different lengths
```

```
some_dict = dict(A=, B=, C=)
```

```
#view dictionary
```

```
print(some_dict)
```

```
{'A': , 'B': , 'C': }
```

If we attempt to use the `from_dict()` function to convert this dictionary into a pandas DataFrame, we'll receive an error:

```
import pandas as pd
```

```
#attempt to create pandas DataFrame from dictionary
```

```
df = pd.DataFrame.from_dict(some_dict)
```

ValueError: All arrays must be of the same length

We receive an error that tells us all arrays in the

dictionary must have the same length.

To get around this error, we can use the following syntax to convert the dictionary into a DataFrame:

```
import pandas as pd
```

```
#create pandas DataFrame from dictionary
```

```
df = pd.DataFrame(dict())
```

```
#view DataFrameprint(df)
```

```
A B C
```

```
0 2 9.0 4.0
```

```
1 5 3.0 4.0
```

```
2 5 NaN 2.0
```

```
3 7 NaN NaN
```

```
4 8 NaN NaN
```

Notice that we're able to successfully create a pandas DataFrame and NaN values are filled in to ensure that each column is the same length.

If you would like to replace these NaN values with other values (such as zero), you can use the `replace()` function as follows:

```
#replace all NaNs with zeros  
df.replace(np.nan, 0, inplace=True)
```

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

```
A B C  
0 2 9.0 4.0  
1 5 3.0 4.0  
2 5 0.0 2.0  
3 7 0.0 0.0  
4 8 0.0 0.0
```

Notice that each NaN value has been replaced with zero.

The following tutorials explain how to perform other common operations in pandas: