

# Can you please explain why merging on object and int64 columns can be problematic?

Authored by  
**stats writer**

June 30, 2024

## RECOMMENDED CITATION

stats writer (2024). *Can you please explain why merging on object and int64 columns can be problematic?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=160706>

When merging data from two different sources, it is important to ensure that the columns used for merging are compatible. However, when trying to merge on an object column and an int64 column, this can lead to potential problems. This is because object columns contain string values while int64 columns contain numerical values. The mismatch in data types can result in incorrect or incomplete merges, causing errors in the merged dataset. Therefore, it is crucial to properly convert data types before attempting to merge on these columns to avoid any issues.

## **Fix: You are trying to merge on object and int64 columns**

**One error you may encounter when using pandas is:**

**ValueError: You are trying to merge on int64 and object columns.**

**If you wish to proceed you should use `pd.concat`**

**This error occurs when you attempt to merge two pandas DataFrames but the column you're merging on is an object in one DataFrame and an integer in the other DataFrame.**

**The following example shows how to fix this error in practice.**

**How to Reproduce the Error**

**Suppose we create the following two pandas DataFrames:**

```
import pandas as pd
```

```
#create DataFrame
```

```
df1 = pd.DataFrame({'year': ,  
'sales': })
```

```
df2 = pd.DataFrame({'year': ,  
'refunds': })
```

```
#view DataFrames
```

```
print(df1)
```

```
year sales
```

```
0 2015 500
```

```
1 2016 534
```

```
2 2017 564
```

```
3 2018 671
```

```
4 2019 700
```

```
5 2020 840
```

```
6 2021 810
```

```
print(df2)
```

```
year refunds
```

```
0 2015 31
```

```
1 2016 36
```

2 2017 40

3 2018 40

4 2019 43

5 2020 70

6 2021 62

Now suppose we attempt to merge the two DataFrames:

```
#attempt to merge two DataFrames
```

```
big_df = df1.merge(df2, on='year', how='left')
```

**ValueError: You are trying to merge on int64 and object columns.**

If you wish to proceed you should use `pd.concat`

We receive a `ValueError` because the year variable in the first DataFrame is an integer but the year variable in the second DataFrame is an object.

How to Fix the Error

The easiest way to fix this error is to simply convert the year variable in the second DataFrame to an integer and then perform the merge.

The following syntax shows how to do so:

```
#convert year variable in df2 to integer  
df2=df2.astype(int)
```

```
#merge two DataFrames
```

```
big_df = df1.merge(df2, on='year', how='left')
```

```
#view merged DataFrame
```

```
big_df
```

```
year sales refunds
```

```
0 2015 500 31
```

```
1 2016 534 36
```

```
2 2017 564 40
```

```
3 2018 671 40
```

```
4 2019 700 43
```

```
5 2020 840 70
```

```
6 2021 810 62
```

**Notice that we don't receive any ValueError and we are able to successfully merge the two DataFrames into one.**

**Additional Resources**

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