

Using the apply method in pandas

February 6, 2021

1 Using the apply() method in pandas

Sometimes, creating a calculated column in pandas is as simple as this:

```
df['difference'] = df['first_column'] - df['second_column']
```

or this:

```
df['date_fixed'] = pd.to_datetime(df['date'])
```

Other times, though, your needs are more complex – you need to take each row of data in your data frame and do *several things* to it. That’s where `apply()` comes in.

Given a function, `apply()` will, uh, *apply* that function to every row in the data frame. A common scenario for doing so would be to create a new column.

An example might make this idea a little more clear. Let’s load up a CSV of Texas death row media witnesses.

```
[6]: import pandas as pd
```

```
[7]: df = pd.read_csv('../data/tx-death-row-media-list.csv',  
    ↳ parse_dates=['execution_date'])
```

Now, let’s say, we want to create a new column with the *month* of the execution. [Given what we know about date objects](#), this should be simple, right?

So this might be my first guess:

```
[8]: df['month'] = df['execution_date'].month
```

```
-----  
AttributeError                                Traceback (most recent call last)  
<ipython-input-8-7f53c16264f0> in <module>  
----> 1 df['month'] = df['execution_date'].month  
  
~/ire/python-mini-bootcamp-2021/env/lib/python3.9/site-packages/pandas/core/  
↳ generic.py in __getattr__(self, name)  
    5460         if self._info_axis.  
↳ _can_hold_identifiers_and_holds_name(name):  
    5461             return self[name]
```

```
-> 5462         return object.__getattr__(self, name)
5463
5464     def __setattr__(self, name: str, value) -> None:
```

```
AttributeError: 'Series' object has no attribute 'month'
```

Womp womp. Looks like we need to create a *function* to do this for us. Then we can *apply* that function to each row.

```
[9]: def get_month(row):
      '''Given a row of data, return the month of the execution date'''
      return row['execution_date'].month
```

... and now we can apply it. We also need to specify *how* it's going to be applied. `axis=0` is the default and attempts to apply the function to each *column*. We want `axis=1`, which applies the function to each *row* of data.

```
[10]: df['month'] = df.apply(get_month, axis=1)
```

```
[11]: df.head()
```

```
[11]:  execution_no execution_date journo_last journo_rest \
0          570      2020-07-08    Graczyk    Michael
1          570      2020-07-08      Brown    Joseph
2          569      2020-02-06    Graczyk    Michael
3          569      2020-02-06      Brown    Joseph
4          569      2020-02-06    Garcia    Imelda

      journo_affiliation inmate_no inmate_last inmate_rest \
0      Associated Press   999137    Wardlow    Billy
1      Huntsville Item   999137    Wardlow    Billy
2      Associated Press   999450      Ochoa    Abel
3      Huntsville Item   999450      Ochoa    Abel
4  Al Dia/Dallas Morning News   999450      Ochoa    Abel

      url  month
0  https://www.tdcj.state.tx.us/death_row/dr_info...    7
1  https://www.tdcj.state.tx.us/death_row/dr_info...    7
2  https://www.tdcj.state.tx.us/death_row/dr_info...    2
3  https://www.tdcj.state.tx.us/death_row/dr_info...    2
4  https://www.tdcj.state.tx.us/death_row/dr_info...    2
```

We could also have dropped in a *lambda expression* for the function – in this case, it's simple enough to be readable:

```
[12]: df['month'] = df.apply(lambda x: x['execution_date'].month, axis=1)
```

```
[16]: df.head()
```

```
[16]: execution_no execution_date journo_last journo_rest \
0          570      2020-07-08   Graczyk   Michael
1          570      2020-07-08    Brown    Joseph
2          569      2020-02-06   Graczyk   Michael
3          569      2020-02-06    Brown    Joseph
4          569      2020-02-06   Garcia    Imelda

      journo_affiliation inmate_no inmate_last inmate_rest \
0      Associated Press   999137   Wardlow    Billy
1      Huntsville Item   999137   Wardlow    Billy
2      Associated Press   999450    Ochoa     Abel
3      Huntsville Item   999450    Ochoa     Abel
4  Al Dia/Dallas Morning News 999450    Ochoa     Abel

                                url  month
0  https://www.tdcj.state.tx.us/death_row/dr_info...    7
1  https://www.tdcj.state.tx.us/death_row/dr_info...    7
2  https://www.tdcj.state.tx.us/death_row/dr_info...    2
3  https://www.tdcj.state.tx.us/death_row/dr_info...    2
4  https://www.tdcj.state.tx.us/death_row/dr_info...    2
```

```
[15]: sorted(df.month.unique())
```

```
[15]: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
```