handling_missing_data_replace (4)

January 12, 2020

##

Handling Missing Data - replace method

[1]:
```
import pandas as pd
import numpy as np
df = pd.read_csv("weareplace_data.csv")
df
```

[1]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	-99999	7	Sunny
2	1/3/2017	28	-99999	Snow
3	1/4/2017	-99999	7	0
4	1/5/2017	32	-99999	Rain
5	1/6/2017	31	2	Sunny
6	1/6/2017	34	5	0

Replacing single value

[2]:
```
new_df = df.replace(-99999, value=np.NaN)
new_df
```

[2]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/2/2017	NaN	7.0	Sunny
2	1/3/2017	28.0	NaN	Snow
3	1/4/2017	NaN	7.0	0
4	1/5/2017	32.0	NaN	Rain
5	1/6/2017	31.0	2.0	Sunny
6	1/6/2017	34.0	5.0	0

Replacing list with single value

[3]:
```
new_df = df.replace(to_replace=[-99999,-88888], value=0)
new_df
```

[3]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain

1	1/2/2017	0	7	Sunny
2	1/3/2017	28	0	Snow
3	1/4/2017	0	7	0
4	1/5/2017	32	0	Rain
5	1/6/2017	31	2	Sunny
6	1/6/2017	34	5	0

Replacing per column

[6]:
```
new_df = df.replace({
        'temperature': -99999,
        'windspeed': -99999,
        'event': '0'
    }, np.NaN)
new_df
```

[6]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/2/2017	NaN	7.0	Sunny
2	1/3/2017	28.0	NaN	Snow
3	1/4/2017	NaN	7.0	NaN
4	1/5/2017	32.0	NaN	Rain
5	1/6/2017	31.0	2.0	Sunny
6	1/6/2017	34.0	5.0	NaN

Replacing by using mapping

[7]:
```
new_df = df.replace({
        -99999: np.nan,
        'no event': 'Sunny',
    })
new_df
```

[7]:

	day	temperature	windspeed	event
0	1/1/2017	32.0	6.0	Rain
1	1/2/2017	NaN	7.0	Sunny
2	1/3/2017	28.0	NaN	Snow
3	1/4/2017	NaN	7.0	0
4	1/5/2017	32.0	NaN	Rain
5	1/6/2017	31.0	2.0	Sunny
6	1/6/2017	34.0	5.0	0

Regex

[8]:
```
# when windspeed is 6 mph, 7 mph etc. & temperature is 32 F, 28 F etc.
new_df = df.replace({'temperature': '[A-Za-z]', 'windspeed': '[a-z]'},'',␣
 ↪regex=True)
new_df
```

[8]:

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	-99999	7	Sunny
2	1/3/2017	28	-99999	Snow
3	1/4/2017	-99999	7	0
4	1/5/2017	32	-99999	Rain
5	1/6/2017	31	2	Sunny
6	1/6/2017	34	5	0

Replacing list with another list

One way to deal with Categorical data

[9]:
```
df = pd.DataFrame({
    'score': ['exceptional','average', 'good', 'poor', 'average',␣
 ↪'exceptional'],
    'student': ['rob', 'maya', 'parthiv', 'tom', 'julian', 'erica']
})
df
```

[9]:

	score	student
0	exceptional	rob
1	average	maya
2	good	parthiv
3	poor	tom
4	average	julian
5	exceptional	erica

[10]:
```
df.replace(['poor', 'average', 'good', 'exceptional'], [1,2,3,4])
```

[10]:

	score	student
0	4	rob
1	2	maya
2	3	parthiv
3	1	tom
4	2	julian
5	4	erica

[ ]:
```

```