Replace scalar values using .replace()

OPTIMIZING PYTHON CODE WITH PANDAS

Leonidas Souliotis
PhD Candidate

The popular name dataset

Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
2011	FEMALE	ASIAN AND PACIFIC ISLANDER	SOPHIA	119	1
2011	FEMALE	ASIAN AND PACIFIC ISLANDER	CHLOE	106	2

Replace values in pandas

```
start_time = time.time()
names['Gender'].loc[names.Gender=='MALE'] = 'BOY'
print("Replace values using .loc[]: {} sec".format(time.time() - start_time))
```

```
Results from the first method calculated in 0.0311849 seconds
```

Replace values using .replace()

```
start_time = time.time()
names['Gender'].replace('MALE', 'BOY', inplace=True)
print("Time using .replace(): {} sec".fomrat(time.time() - start_time))
```

```
Time using .replace(): 0.0016758441925 sec
```

```
Differerence in speed: 1,704.52411439%
```

Let's do it

OPTIMIZING PYTHON CODE WITH PANDAS

Replace values using lists

OPTIMIZING PYTHON CODE WITH PANDAS

Leonidas Souliotis
PhD Candidate

Replace multiple values with one value

Year of Birth	Gender	Ethnicity	Child's First Name	Count	Rank
2011	FEMALE	WHITE NON HISP	HELENA	97	4

```
start_time = time.time()
names['Ethnicity'].loc[(names["Ethnicity"] == 'WHITE NON HISPANIC') |
(names["Ethnicity"] == 'WHITE NON HISP')] = 'WNH'
print("Results from the above operation calculated in %s seconds" %
 (time.time() - start_time))
```

```
Results from the second method calculated in 0.0276169776917 seconds
```

Replace multiple values using .replace() I

```
start_time = time.time()
names['Ethnicity'].replace(['WHITE NON HISPANIC','WHITE NON HISP'],
'WNH', inplace=True)
print("Time using .replace(): {} sec".format(time.time() - start_time))
```

```
Time using .replace():  0.001447916030884 sec
```

```
Difference in speed: 2160.68681809%
```

```
names['Ethnicity'].replace(['WHITE NON HISP'], 'WHITE NON HISPANIC', inplace=True)
names['Ethnicity'].replace(['BLACK NON HISP'], 'BLACK NON HISPANIC', inplace=True)
```

```
names['Ethnicity'].replace(['BLACK NON HISP','WHITE NON HISP'], ['BLACK NON HISPANIC',
'WHITE NON HISPANIC'], inplace=True)
```

Let's do it

OPTIMIZING PYTHON CODE WITH PANDAS

Replace values using dictionaries

OPTIMIZING PYTHON CODE WITH PANDAS

Leonidas Souliotis
PhD Candidate

Replace single values with dictionaries

```
start_time = time.time()
names['Gender'].replace({'MALE':'BOY', 'FEMALE':'GIRL'},
inplace=True)
print("Time using .replace() with dictionary: {} sec".format(time.time() - start_time))
```

```
Time using .replace() with dictionary: 0.00197792053223 sec
```

```
start_time = time.time()
names['Gender'].replace('MALE', 'BOY', inplace=True)
names['Gender'].replace('FEMALE', 'GIRL', inplace=True)
print("Time using multiple .replace(): {} sec".format(time.time() - start_time))
```

```
Time using multiple .replace(): 0.00307083129883 sec
```

```
Difference in speed: 55.2555448407%
```

Replace multiple values using dictionaries

```
start_time = time.time()
names.replace({'Ethnicity': {'ASIAN AND PACI': 'ASIAN', 'ASIAN AND PACIFIC ISLANDER': 'ASIAN',
                             'BLACK NON HISPANIC': 'BLACK', 'BLACK NON HISP': 'BLACK',
                             'WHITE NON HISPANIC': 'WHITE', 'WHITE NON HISP': 'WHITE'}})

print("Time using .replace() with dictionary: {} sec".format (time.time() - start_time))
```

```
Time using .replace() with dictionary: 0.0028018 sec
```

Let's do it!

OPTIMIZING PYTHON CODE WITH PANDAS