Python Pandas Conditional Selection of Rows and Columns

Fan Wang

2020-11-18

Contents

1 Select Rows and Columns from Dataframe

Go to the **RMD**, **PDF**, or **HTML** version of this file. Go back to fan's Python Code Examples Repository (bookdown site).

```
import numpy as np
import pandas as pd
import random as random
import string as string
```

1.1 Generate a Testing Dataframe

Generate a testing dataframe for selection and other tests.

```
# Seed
np.random.seed(999)
random.seed(999)
# Numeric matrix 3 rows 4 columns
mt_numeric = np.random.randint(10, size=(5, 4))
st_rand_word_block = ''.join(random.choice(string.ascii_lowercase) for ctr in range(5*5*3))
mt_string = np.reshape([st_rand_word_block[ctr: ctr + 5].capitalize() for ctr in range(0, len(st_rand_wo
mt_data = np.column_stack([mt_numeric, mt_string])

# Matrix to dataframe
df_data = pd.DataFrame(data=mt_data,
                       index=[ 'r' + str(it_col) for it_col in np.array(range(1, mt_data.shape[0]+1))],
                       columns=[ 'c' + str(it_col) for it_col in np.array(range(1, mt_data.shape[1]+1))]

# Replace values
df_data = df_data.replace(['Zvcss', 'Dugei', 'Ciagu'], 'Zqovt')

# Print table
print(df_data)
```

```
##     c1 c2 c3 c4     c5      c6      c7
## r1  0  5  1  8  Zqovt   Rppez   Ukuzu
## r2  1  9  3  0  Zqovt   Sbwyi   Mzhum
```

```
## r3  5  8  8  0  Zqovt  Qgfvk  Fcrto
## r4  5  2  5  7  Wxlev  Upoax  Bhdxu
## r5  4  6  2  7  Hmziq  Lbyfo  Dntrz
```

1.2 Select Rows Based on Column/Variable Values

There is a dataframe with many rows, select a subset of rows where a particular column/variable's value is equal to some value.

```python
# Concatenate to matrix
df_data_subset = df_data.loc[df_data['c5'] == 'Zqovt']
# Print
print(df_data_subset)
```

```
##     c1 c2 c3 c4     c5     c6     c7
## r1  0  5  1  8  Zqovt  Rppez  Ukuzu
## r2  1  9  3  0  Zqovt  Sbwyi  Mzhum
## r3  5  8  8  0  Zqovt  Qgfvk  Fcrto
```

See How to select rows from a DataFrame based on column values.