b8nn8cbeg

April 5, 2024

Data Frame

```python
[8]: # All imports
     import numpy as np
     import pandas as pd
```

Example - 1

Create a Data Frame cars using raw data stored in a dictionary

```python
[9]: cars_per_cap = [809, 731, 588, 18, 200, 70, 45]
     country = ['United States', 'Australia', 'Japan', 'India', 'Russia', 'Morocco',
     ↪'Egypt']
     drives_right = [True, False, False, False, True, True, True]
```

```python
[10]: data = {"cars_per_cap": cars_per_cap, "country": country, "drives_right":
      ↪drives_right}
```

```python
[11]: data
```

```
[11]: {'cars_per_cap': [809, 731, 588, 18, 200, 70, 45],
       'country': ['United States',
        'Australia',
        'Japan',
        'India',
        'Russia',
        'Morocco',
        'Egypt'],
       'drives_right': [True, False, False, False, True, True, True]}
```

```python
[12]: cars = pd.DataFrame(data)

      cars
```

```
[12]:    cars_per_cap        country  drives_right
      0           809  United States          True
      1           731      Australia         False
      2           588          Japan         False
      3            18          India         False
```

```
4            200        Russia        True
5             70       Morocco        True
6             45         Egypt        True
```

[13]: `type(cars)`

[13]: `pandas.core.frame.DataFrame`

Example - 2 (Reading data from a file)

Create a Data Frame by importing cars data from cars.csv

[16]:
```python
# Read a file using pandas

cars_df = pd.read_csv('cars.csv')

cars_df
```

[16]:
```
     USCA    US United States    809   FALSE
0   ASPAC   AUS      Australia  731.0   True
1   ASPAC   JAP          Japan  588.0   True
2   ASPAC    IN          India   18.0   True
3   ASPAC    RU         Russia  200.0  False
4   LATAM   MOR        Morocco   70.0  False
5     AFR    EG          Egypt   45.0  False
6     EUR   ENG        England    NaN   True
```

Example - 3 (Column headers)

Read file - skip header

[35]:
```python
cars_df = pd.read_csv('cars.csv', header=None)

cars_df
```

[35]:
```
        0    1              2      3      4
0    USCA   US  United States  809.0  False
1   ASPAC  AUS      Australia  731.0   True
2   ASPAC  JAP          Japan  588.0   True
3   ASPAC   IN          India   18.0   True
4   ASPAC   RU         Russia  200.0  False
5   LATAM  MOR        Morocco   70.0  False
6     AFR   EG          Egypt   45.0  False
7     EUR  ENG        England    NaN   True
```

Assign Headers

[36]:
```python
# Returns an array of headers
```

```
cars_df.columns
```

[36]: `Int64Index([0, 1, 2, 3, 4], dtype='int64')`

[37]:
```
# Rename Headers

cars_df.columns = ['country code', 'region', 'country', 'cars_per_cap',␣
  ↪'drive_right']
```

[38]:
```
cars_df
```

[38]:
```
  country code region        country  cars_per_cap  drive_right
0         USCA     US  United States         809.0        False
1        ASPAC    AUS      Australia         731.0         True
2        ASPAC    JAP          Japan         588.0         True
3        ASPAC     IN          India          18.0         True
4        ASPAC     RU         Russia         200.0        False
5        LATAM    MOR        Morocco          70.0        False
6          AFR     EG          Egypt          45.0        False
7          EUR    ENG        England           NaN         True
```

Example - 4 (Row index/names)

Read file - skip header and assign first column as index.

[31]:
```
# Index is returned by
cars_df.index
```

[31]: `RangeIndex(start=0, stop=8, step=1)`

[43]:
```
# Read file and set 1st column as index
cars_df = pd.read_csv("cars.csv", header= None, index_col=0)

# set the column names
cars_df.columns = ['region', 'country', 'cars_per_cap', 'drive_right']
cars_df
```

[43]:
```
       region        country  cars_per_cap  drive_right
0
USCA       US  United States         809.0        False
ASPAC     AUS      Australia         731.0         True
ASPAC     JAP          Japan         588.0         True
ASPAC      IN          India          18.0         True
ASPAC      RU         Russia         200.0        False
LATAM     MOR        Morocco          70.0        False
AFR        EG          Egypt          45.0        False
EUR       ENG        England           NaN         True
```

```
[44]:  # Print the new index
       cars_df.index
```

```
[44]:  Index(['USCA', 'ASPAC', 'ASPAC', 'ASPAC', 'ASPAC', 'LATAM', 'AFR', 'EUR'],
       dtype='object', name=0)
```

Rename the Index Name

```
[46]:  cars_df.index.name = 'country_code'
       cars_df
```

```
[46]:                region          country  cars_per_cap  drive_right
       country_code
       USCA              US    United States         809.0        False
       ASPAC            AUS        Australia         731.0         True
       ASPAC            JAP            Japan         588.0         True
       ASPAC             IN            India          18.0         True
       ASPAC             RU           Russia         200.0        False
       LATAM            MOR          Morocco          70.0        False
       AFR               EG            Egypt          45.0        False
       EUR              ENG          England           NaN         True
```

Delete the index name

```
[51]:  cars_df.index.name = None
       cars_df
```

```
[51]:          region          country  cars_per_cap  drive_right
       USCA        US    United States         809.0        False
       ASPAC      AUS        Australia         731.0         True
       ASPAC      JAP            Japan         588.0         True
       ASPAC       IN            India          18.0         True
       ASPAC       RU           Russia         200.0        False
       LATAM      MOR          Morocco          70.0        False
       AFR         EG            Egypt          45.0        False
       EUR        ENG          England           NaN         True
```

Set Hierarchical index

```
[52]:  # Read file and set 1st column as index
       cars_df = pd.read_csv("cars.csv", header= None)

       # set the column names
       cars_df.columns =␣
        ↪['country_code','region','country','cars_per_cap','drives_right']

       cars_df.set_index(['region', 'country_code'], inplace=True)
```

```
[53]: cars_df
```

```
[53]:                         country   cars_per_cap   drives_right
       region country_code
       US     USCA       United States         809.0          False
       AUS    ASPAC          Australia         731.0           True
       JAP    ASPAC              Japan         588.0           True
       IN     ASPAC              India          18.0           True
       RU     ASPAC             Russia         200.0          False
       MOR    LATAM            Morocco          70.0          False
       EG     AFR              Egypt           45.0          False
       ENG    EUR            England           NaN           True
```

```
[ ]:
```

Example - 5 (Write Data Frame to file)

Write cars_df to cars_to_csv.csv

```
[54]: cars_df.to_csv('cars_to_csv.csv')
```

```
[ ]:
```