KEY_Practice10_Pandas-Intro

July 12, 2019

1 Practice: Intro to Pandas

First we need to import the `pandas` package. Let's give it the same special nickname we used before.

```
[2]: # import the pandas package with the nickname pd

import pandas as pd
```

Now we'll create a `DataFrame`. Modify this cell to save it to a variable called `car_data`.

```
[4]: # save the DataFrame to a variable
car_data = pd.DataFrame({'acceleration': [12.0,11.5,11.0,12.0,10.5,10.0,9.0,8.
 ↪5,10.0,8.5,10.0,8.0,9.5,10.0,15.0,15.5,15.5,16.0,14.5,20.5],
 'cylinders': [8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 4, 6, 6, 6, 4, 4],
 'displacement': [307.0,350.0,318.0,304.0,302.0,429.0,454.0,440.0,455.0,390.
 ↪0,383.0,340.0,400.0,455.0,113.0,198.0,199.0,200.0,97.0,97.0],
 'horsepower': [130.0,165.0,150.0,150.0,140.0,198.0,220.0,215.0,225.0,190.0,170.
 ↪0,160.0,150.0,225.0,95.0,95.0,97.0,85.0,88.0,46.0],
 'model_year': [70,70,70,70,70,70,70,70,70,70,70,70,70,70,70,70,70,70,70,70],
 'mpg': [18.0,15.0,18.0,16.0,17.0,15.0,14.0,14.0,14.0,15.0,15.0,14.0,15.0,14.
 ↪0,24.0,22.0,18.0,21.0,27.0,26.0],
 'name': ['chevrolet chevelle malibu','buick skylark 320','plymouth␣
 ↪satellite','amc rebel sst','ford torino','ford galaxie 500','chevrolet␣
 ↪impala','plymouth fury iii','pontiac catalina','amc ambassador dpl','dodge␣
 ↪challenger se',"plymouth 'cuda 340",'chevrolet monte carlo','buick estate␣
 ↪wagon (sw)','toyota corona mark ii','plymouth duster','amc hornet','ford␣
 ↪maverick','datsun pl510','volkswagen 1131 deluxe sedan'],
 'origin':␣
 ↪['usa','usa','usa','usa','usa','usa','usa','usa','usa','usa','usa','usa','usa','usa','japan
 'weight':␣
 ↪[3504,3693,3436,3433,3449,4341,4354,4312,4425,3850,3563,3609,3761,3086,2372,2833,2774,2587,
```

Now explore the `car_data` DataFrame. View the first few rows:

```
[5]: # view the first 5 rows

car_data.head(5)
```

```
[5]:    acceleration  cylinders  displacement  horsepower  model_year   mpg  \
     0          12.0          8         307.0       130.0          70  18.0
     1          11.5          8         350.0       165.0          70  15.0
     2          11.0          8         318.0       150.0          70  18.0
     3          12.0          8         304.0       150.0          70  16.0
     4          10.5          8         302.0       140.0          70  17.0

                             name origin  weight
     0  chevrolet chevelle malibu    usa    3504
     1          buick skylark 320    usa    3693
     2         plymouth satellite    usa    3436
     3                amc rebel sst   usa    3433
     4                 ford torino    usa    3449
```

```
[6]: # view the first 10 rows

     car_data.head(10)
```

```
[6]:    acceleration  cylinders  displacement  horsepower  model_year   mpg  \
     0          12.0          8         307.0       130.0          70  18.0
     1          11.5          8         350.0       165.0          70  15.0
     2          11.0          8         318.0       150.0          70  18.0
     3          12.0          8         304.0       150.0          70  16.0
     4          10.5          8         302.0       140.0          70  17.0
     5          10.0          8         429.0       198.0          70  15.0
     6           9.0          8         454.0       220.0          70  14.0
     7           8.5          8         440.0       215.0          70  14.0
     8          10.0          8         455.0       225.0          70  14.0
     9           8.5          8         390.0       190.0          70  15.0

                             name origin  weight
     0  chevrolet chevelle malibu    usa    3504
     1          buick skylark 320    usa    3693
     2         plymouth satellite    usa    3436
     3                amc rebel sst   usa    3433
     4                 ford torino    usa    3449
     5             ford galaxie 500   usa    4341
     6             chevrolet impala   usa    4354
     7            plymouth fury iii   usa    4312
     8             pontiac catalina   usa    4425
     9           amc ambassador dpl   usa    3850
```

What does the end of the DataFrame look like? Try viewing the last few rows:

```
[7]: # view the last 5 rows

     car_data.tail(5)
```

```
[7]:     acceleration  cylinders  displacement  horsepower  model_year   mpg  \
     15          15.5          6         198.0        95.0          70  22.0
```

```
16       15.5          6         199.0        97.0         70  18.0
17       16.0          6         200.0        85.0         70  21.0
18       14.5          4          97.0        88.0         70  27.0
19       20.5          4          97.0        46.0         70  26.0

                          name  origin  weight
15             plymouth duster     usa    2833
16                 amc hornet     usa    2774
17               ford maverick     usa    2587
18               datsun pl510   japan    2130
19   volkswagen 1131 deluxe sedan  europe    1835
```

What happens when you call head or `tail` without putting a number between the parentheses?

```python
[8]: # try using head without a number

car_data.head()
```

```
[8]:   acceleration  cylinders  displacement  horsepower  model_year   mpg  \
0            12.0          8         307.0        130.0          70  18.0
1            11.5          8         350.0        165.0          70  15.0
2            11.0          8         318.0        150.0          70  18.0
3            12.0          8         304.0        150.0          70  16.0
4            10.5          8         302.0        140.0          70  17.0

                         name origin  weight
0   chevrolet chevelle malibu    usa    3504
1           buick skylark 320    usa    3693
2          plymouth satellite    usa    3436
3               amc rebel sst    usa    3433
4                 ford torino    usa    3449
```

```python
[9]: # try using tail without a number

car_data.tail()
```

```
[9]:    acceleration  cylinders  displacement  horsepower  model_year   mpg  \
15           15.5          6         198.0        95.0          70  22.0
16           15.5          6         199.0        97.0          70  18.0
17           16.0          6         200.0        85.0          70  21.0
18           14.5          4          97.0        88.0          70  27.0
19           20.5          4          97.0        46.0         70  26.0

                          name  origin  weight
15             plymouth duster     usa    2833
16                 amc hornet     usa    2774
17               ford maverick     usa    2587
18               datsun pl510   japan    2130
19   volkswagen 1131 deluxe sedan  europe    1835
```

What do the rows of the DataFrame represent?

Answers: Different types of cars

What measurements or observations do we have about the data?

Answer: acceleration, cylinder count, engine displacement, horsepower, model year, mpg, name, country of origin, and weight

What types of variables are in each of the columns?

- **acceleration**: float
- **cylinders**: int
- **displacement**: float
- **horsepower**: float
- **model_year**: int
- **mpg**: float
- **name**: string
- **origin**: string
- **weight**: int