CLASS ASSIGNMENT

Data Science with Python

MCA-205-[P]

SUBMITTED BY:
Prabhat Singh Bhadoriya
MCAN1CA20015

SUBMITTED TO:
Dr. Sanjay Jain
Associate Professor
Dept. Of CSA/SOET

The following exercise utilizes data from UC Irvine Machine Learning Repository

Step 1. Import the necessary libraries

```
In [1]: import pandas as pd
        import numpy as np
```

Step 2. Import the first dataset cars1 and cars2.

Step 3. Assign each to a variable called cars1 and cars2

```
In [2]: cars1 = pd.read_csv('https://github.com/Drsanjayjainitm/Data-Science-using-Python/blob/main/cars1.csv')
        cars2 = pd.read_csv('https://github.com/Drsanjayjainitm/Data-Science-using-Python/blob/main/cars2.csv')
```

Step 4. Oops, it seems our first dataset has some unnamed blank columns, fix cars1

```
In [3]: cars1 = cars1.dropna(axis = 1)
        cars1.head()
```

Out[3]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302	140	3449	10.5	70	1	ford torino

Step 5. What is the number of observations in each dataset?

In [4]:
```
print(cars1.shape[0])
print(cars2.shape[0])
```

```
198
200
```

Step 6. Join cars1 and cars2 into a single DataFrame called cars

In [5]:
```
cars = pd.concat([cars1, cars2])
cars
```

Out[5]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302	140	3449	10.5	70	1	ford torino
...	...	...	...	...	...	...	...	...	...
195	27.0	4	140	86	2790	15.6	82	1	ford mustang gl
196	44.0	4	97	52	2130	24.6	82	2	vw pickup
197	32.0	4	135	84	2295	11.6	82	1	dodge rampage
198	28.0	4	120	79	2625	18.6	82	1	ford ranger
199	31.0	4	119	82	2720	19.4	82	1	chevy s-10

398 rows × 9 columns

Step 7. Oops, there is a column missing, called owners. Create a random number Series from 15,000 to 73,000.

```
In [6]: owners = pd.Series(np.random.randint(15000, 73000, cars.shape[0]))
        owners
```

```
Out[6]: 0      34677
        1      16818
        2      38446
        3      18431
        4      55646
               ...
        393    17111
        394    46679
        395    15048
        396    49689
        397    41963
        Length: 398, dtype: int64
```

Step 8. Add the column owners to cars

```
In [7]: cars['owners'] = owners
        cars
```

Out[7]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	owners
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	34677
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	16818
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite	38446
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst	18431
4	17.0	8	302	140	3449	10.5	70	1	ford torino	55646
...	...	...	...	...	...	...	...	...	...	...
195	27.0	4	140	86	2790	15.6	82	1	ford mustang gl	38808
196	44.0	4	97	52	2130	24.6	82	2	vw pickup	70484
197	32.0	4	135	84	2295	11.6	82	1	dodge rampage	28310
198	28.0	4	120	79	2625	18.6	82	1	ford ranger	56547
199	31.0	4	119	82	2720	19.4	82	1	chevy s-10	62387

398 rows × 10 columns