MPG Cars

Introduction:

The following exercise utilizes data from
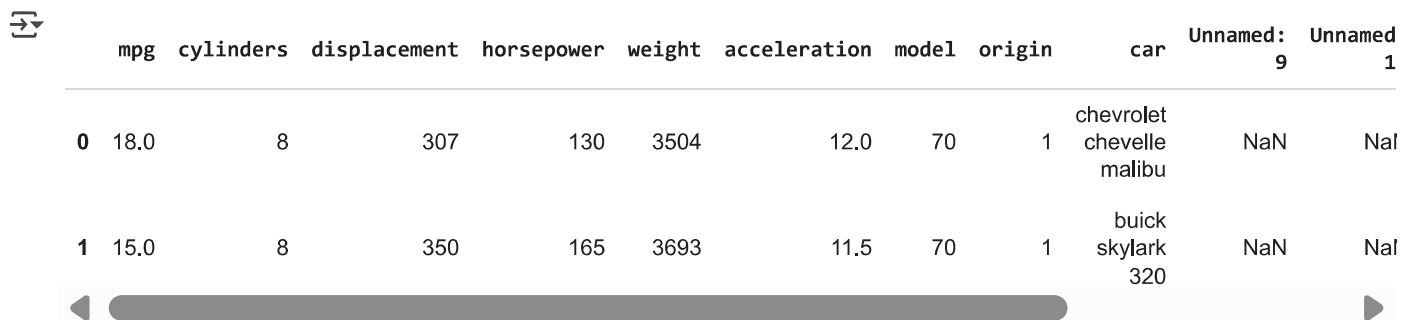
Step 1. Import the necessary libraries

```
import pandas as pd
import numpy as np
```

Step 2. Import the first dataset cars1 and cars2.

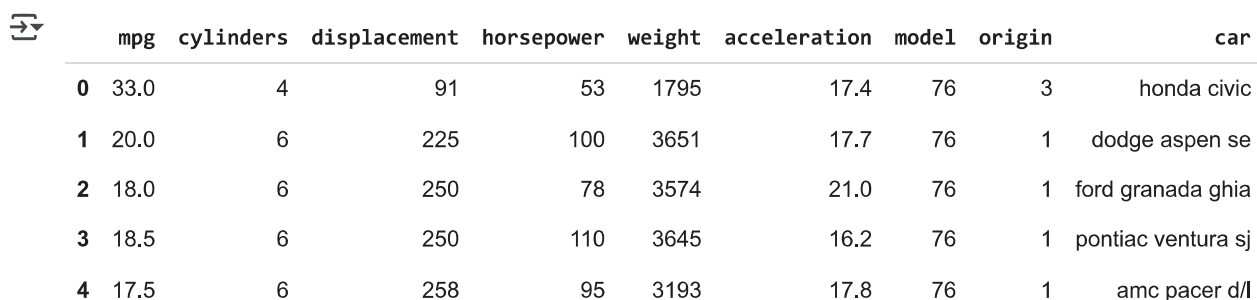### Step 3. Assign each to a variable called cars1 and cars2

```
cars1 = pd.read_csv('https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/cars1.csv')
cars2 = pd.read_csv('https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/cars2.csv')
```

```
cars1.head()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	Unnamed: 9	Unnamed 1
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	NaN	NaN
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	NaN	NaN
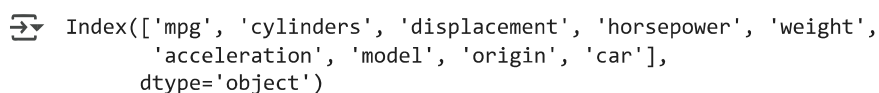
```
cars2.head()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car
0	33.0	4	91	53	1795	17.4	76	3	honda civic
1	20.0	6	225	100	3651	17.7	76	1	dodge aspen se
2	18.0	6	250	78	3574	21.0	76	1	ford granada ghia
3	18.5	6	250	110	3645	16.2	76	1	pontiac ventura sj
4	17.5	6	258	95	3193	17.8	76	1	amc pacer d/l

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

```
cars1 = cars1.loc[:, ~cars1.columns.str.contains('^Unnamed')]
cars1.columns
```

```
Index(['mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'model', 'origin', 'car'],
      dtype='object')
```

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

```
print(cars1.shape[0])
cars2.shape[0]
```

```
198
200
```

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

```
cars = pd.concat([cars1, cars2])
cars.head()
```

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

```
rand_nums = np.random.randint(15000, high=73001, size=398, dtype='l')
print(rand_nums)
```

```
[72923 52221 66617 56513 53073 26215 57605 72200 29643 35591 16736 42917
 40066 60097 40119 32051 56751 24847 40911 20428 65322 42042 61149 32686
 37616 67414 35059 35565 48157 54999 34798 35556 58486 57466 59317 69312
 42609 22070 56980 18088 46715 18336 62035 66863 49732 27594 57883 60907
 30576 31318 61165 51201 63072 57457 28193 30678 55624 30065 19773 18252
 29345 56396 16375 52483 29405 69203 61424 21120 59231 59377 41563 56399
 22949 49687 46957 61356 28315 21661 50734 53702 58316 40361 66421 63464
 67023 48209 70911 66897 47815 44613 46056 60855 29256 71574 21108 22916
 51555 31177 29053 22627 51570 44341 54293 72328 53644 52776 36142 25515
 43530 48642 28739 15789 49644 18686 32959 69362 46825 64271 42650 61719
 47397 19040 36280 44557 35646 60909 31376 30047 66209 41303 66367 48602
 69382 53449 60978 33125 37866 36461 48079 66576 21855 16539 39321 29347
 23818 53547 15649 30702 49138 48351 72583 49945 60510 28535 49416 57476
 30841 22204 43804 21551 69307 34467 52265 35513 31917 34012 45389 43200
 18206 50865 63586 59727 23680 59866 20262 52185 61315 30571 47282 39026
 19349 35540 44820 28918 38143 59543 17648 38282 46685 34359 70067 36989
 34346 37809 37174 38237 33185 42505 70908 45144 33920 60961 41016 32515
 17715 67277 23696 57631 19808 39943 37609 18684 36055 65519 31703 58470
 60535 52262 64982 39793 39698 66316 71614 29355 59335 21278 39858 21647
 29659 37246 66367 22959 33809 44634 67541 66889 25259 28283 63298 19996
 37988 57830 55779 27147 54146 18301 43132 63422 27005 25629 34088 69659
 38429 34109 27697 44760 47996 56858 44112 33156 36532 63080 33365 69478
 27815 42057 52606 60584 25605 18879 19422 59506 33808 15952 70201 47133
 31271 34604 49731 53748 17361 42112 18954 34841 62983 67563 65244 29095
 32117 23475 21662 21309 54574 35412 21747 25090 55474 37529 36286 50218
 22259 38555 39672 55442 27770 65946 32776 41485 64657 22374 27646 63111
 57933 27303 62792 50164 45819 72135 44453 53938 44083 20910 57216 59910
 45437 24506 37766 72719 70985 66675 57588 39344 56241 57793 30468 33314
 58806 19743 16504 57573 30153 16499 27001 44551 27463 57040 26034 45894
 57164 45440 36129 33744 33387 30522 67273 36893 26507 50621 27942 60979
 68794 15430 47931 25628 29163 64268 67222 33616 22617 37575 44863 29690
 16865 18136 53377 17730 63617 42609 52181 67169 36788 39173 40118 61447
 54135 39120 46134 48646 53896 71202 37737 39651 64434 51487 66187 24281
 46584 54126]
```

Step 8. Add the column owners to cars

```
cars['owners'] = rand_nums
cars.head()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	owners
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	72923
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	52221
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite	66617
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst	56513
4	17.0	8	302	140	3449	10.5	70	1	ford torino	53073