

MPG Cars

Introduction:

The following exercise utilizes data from [UC Irvine Machine Learning Repository](https://raw.githubusercontent.com/a-forty-two/COG_GN22CDBDS001_MARCH_22/main/cars1.csv).

Step 1. Import the necessary libraries

```
In [41]: import pandas as pd
import random
```

Step 2. Import the first dataset [cars1](#) and [cars2](#).

```
In [42]: data1=pd.read_csv('https://raw.githubusercontent.com/a-forty-two/COG_GN22CDBDS001_MARCH_22/main/cars1.csv')
data2=pd.read_csv('https://raw.githubusercontent.com/a-forty-two/COG_GN22CDBDS001_MARCH_22/main/cars2.csv')
print(data2.head())
data1.head()
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	\
0	33.0	4	91	53	1795	17.4	76	
1	20.0	6	225	100	3651	17.7	76	
2	18.0	6	250	78	3574	21.0	76	
3	18.5	6	250	110	3645	16.2	76	
4	17.5	6	258	95	3193	17.8	76	

	origin	car
0	3	honda civic
1	1	dodge aspen se
2	1	ford granada ghia
3	1	pontiac ventura sj
4	1	amc pacer d/l

Out[42]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	NaN	NaN	NaN	NaN
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	NaN	NaN	NaN	NaN
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite	NaN	NaN	NaN	NaN
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst	NaN	NaN	NaN	NaN
4	17.0	8	302	140	3449	10.5	70	1	ford torino	NaN	NaN	NaN	NaN

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

```
In [43]: cars1=data1
cars2=data2
cars1
```

Out[43]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	Unnamed: 9	Unnamed: 10	Unnamed: 11	Unnamed: 12
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	NaN	NaN	NaN	NaN
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	NaN	NaN	NaN	NaN
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite	NaN	NaN	NaN	NaN
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst	NaN	NaN	NaN	NaN
4	17.0	8	302	140	3449	10.5	70	1	ford torino	NaN	NaN	NaN	NaN
...
193	24.0	6	200	81	3012	17.6	76	1	ford maverick	NaN	NaN	NaN	NaN
194	22.5	6	232	90	3085	17.6	76	1	amc hornet	NaN	NaN	NaN	NaN
195	29.0	4	85	52	2035	22.2	76	1	chevrolet chevette	NaN	NaN	NaN	NaN
196	24.5	4	98	60	2164	22.1	76	1	chevrolet woody	NaN	NaN	NaN	NaN
197	29.0	4	90	70	1937	14.2	76	2	vw rabbit	NaN	NaN	NaN	NaN

198 rows × 14 columns

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

```
In [44]: Cars1=cars1.iloc[:,[0,1,2,3,4,5,6,7,8]]
Cars1
```

Out[44]:

	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
...
193	24.0	6	200	81	3012	17.6	76	1	ford maverick
194	22.5	6	232	90	3085	17.6	76	1	amc hornet
195	29.0	4	85	52	2035	22.2	76	1	chevrolet chevette
196	24.5	4	98	60	2164	22.1	76	1	chevrolet woody
197	29.0	4	90	70	1937	14.2	76	2	vw rabbit

198 rows × 9 columns

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

```
In [45]: Cars1.count()
```

Out[45]:

mpg	198
cylinders	198
displacement	198
horsepower	198
weight	198
acceleration	198
model	198
origin	198
car	198
dtype:	int64

```
In [46]: cars2.count()
```

Out[46]:

mpg	200
cylinders	200
displacement	200
horsepower	200
weight	200
acceleration	200
model	200
origin	200
car	200
dtype:	int64

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

```
In [47]: CARS=pd.concat([Cars1,cars2])
CARS
```

Out[47]:

	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 [48]: owners = []
for i in range(0,398):
    n = random.randint(15000,73000)
    owners.append(n)
print(owners)
```

```
[33644, 37489, 63479, 30304, 47214, 52478, 64923, 68120, 43419, 60401, 57715, 65653, 44611, 70027, 42147, 36263, 24199, 36114, 59717, 23648, 36639, 72991, 40317, 54029, 22802, 27080, 71097, 63131, 19219, 38952, 65091, 64561, 38476, 49484, 32129, 69815, 68237, 19367, 30424, 39892, 28020, 69423, 26117, 64722, 54218, 61594, 49031, 61557, 25435, 34878, 30458, 51774, 42798, 52220, 29488, 60089, 15991, 28274, 30558, 32910, 64577, 51606, 50629, 62181, 58491, 48656, 25940, 26956, 52848, 53146, 28314, 15373, 30476, 16479, 26412, 16115, 72811, 62175, 49459, 58827, 28341, 21738, 30944, 43952, 70627, 70978, 17659, 22091, 62140, 24055, 48757, 32280, 67361, 61734, 24411, 69976, 54384, 33937, 42333, 51519, 63606, 40603, 34649, 33400, 60966, 47303, 69847, 56699, 56340, 29198, 21934, 67853, 44023, 47130, 63549, 34496, 45734, 56842, 35907, 67992, 69838, 71081, 45777, 56828, 62006, 23600, 57791, 26251, 33053, 52563, 62060, 31358, 33980, 26324, 60740, 20676, 64590, 37302, 36501, 52184, 40787, 16697, 43095, 49870, 16421, 39166, 27666, 51246, 37242, 72028, 68190, 62807, 64136, 34764, 23660, 69371, 32382, 51512, 23565, 23076, 26452, 61686, 35366, 52143, 21186, 53008, 21087, 30894, 35612, 70180, 70709, 27737, 64011, 58209, 31711, 17921, 18989, 36507, 50790, 72816, 18856, 36866, 51398, 24094, 21721, 31910, 38137, 43300, 32258, 41659, 65325, 63889, 72130, 35399, 64447, 66457, 18044, 61378, 46815, 51988, 23473, 55252, 27246, 70984, 26323, 63252, 31498, 61642, 56663, 61365, 22558, 38293, 26243, 59793, 22601, 59240, 18074, 54572, 21779, 30219, 71761, 65310, 51847, 21630, 52915, 52367, 23965, 23817, 58211, 30820, 32289, 55651, 19267, 62307, 62610, 19596, 26791, 68804, 18484, 59362, 58325, 39656, 51279, 28194, 17064, 59935, 51058, 71693, 48101, 45938, 47564, 52048, 35722, 66248, 28826, 55378, 24599, 32495, 72075, 72458, 22219, 27119, 38649, 16089, 69402, 41016, 49706, 44935, 20900, 23782, 29929, 51142, 31120, 52554, 46389, 31491, 67636, 72193, 51168, 58950, 33156, 21245, 23267, 67820, 41772, 54273, 58970, 54841, 66993, 42729, 46791, 46625, 62986, 15532, 43900, 55704, 69174, 36996, 52413, 56796, 40087, 32099, 32686, 49084, 67438, 51945, 31036, 66785, 38177, 53254, 40675, 44104, 57891, 67273, 42802, 60993, 69555, 47575, 58372, 49350, 30890, 26933, 43494, 64578, 62489, 15722, 19831, 16563, 46566, 21835, 27201, 70960, 54588, 39087, 72115, 29873, 25313, 56843, 45688, 43360, 60795, 57888, 66533, 52230, 28915, 52054, 49814, 47296, 17398, 46592, 47139, 28489, 21133, 46505, 44911, 61850, 68464, 50107, 49288, 25310, 53222, 60223, 34079, 38283, 19785, 46572, 55337, 53441, 35459, 58569, 60048, 60293, 19340, 55196, 33759, 46687, 19334, 71573, 68834, 67540, 20540, 59290, 62217, 34576, 34758, 28602, 23327, 35976, 36938, 56315, 35705, 16468, 42860, 37177, 46852, 48146, 70269, 31858]
```

Step 8. Add the column owners to cars

```
In [49]: CARS['owners']=owners
```

```
In [50]: CARS
```

Out[50]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model	origin	car	owners
0	18.0	8	307	130	3504	12.0	70	1	chevrolet chevelle malibu	33644
1	15.0	8	350	165	3693	11.5	70	1	buick skylark 320	37489
2	18.0	8	318	150	3436	11.0	70	1	plymouth satellite	63479
3	16.0	8	304	150	3433	12.0	70	1	amc rebel sst	30304
4	17.0	8	302	140	3449	10.5	70	1	ford torino	47214
...
195	27.0	4	140	86	2790	15.6	82	1	ford mustang gl	37177
196	44.0	4	97	52	2130	24.6	82	2	vw pickup	46852
197	32.0	4	135	84	2295	11.6	82	1	dodge rampage	48146
198	28.0	4	120	79	2625	18.6	82	1	ford ranger	70269
199	31.0	4	119	82	2720	19.4	82	1	chevy s-10	31858

398 rows × 10 columns

```
In [ ]:
```