

## Fixing cyl Data Type

- 2008: extract int from string
- 2018: convert float to int

Load datasets `data_08_v2.csv` and `data_18_v2.csv`. You should've created these data files in the previous section: *Filter, Drop Nulls, Dedupe*.

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

# load datasets

df_08 = pd.read_csv('data_08_v2.csv')
df_18 = pd.read_csv('data_18_v2.csv')
```

```
In [2]: df_08.head()
```

Out[2]:

	model	displ	cyl	trans	drive	fuel	veh_class	air_pollution_score	city_mpg	hwy_mpg
0	ACURA MDX	3.7	(6 cyl)	Auto- S5	4WD	Gasoline	SUV	7	15	20
1	ACURA RDX	2.3	(4 cyl)	Auto- S5	4WD	Gasoline	SUV	7	17	22
2	ACURA RL	3.5	(6 cyl)	Auto- S5	4WD	Gasoline	midsize car	7	16	24
3	ACURA TL	3.2	(6 cyl)	Auto- S5	2WD	Gasoline	midsize car	7	18	26
4	ACURA TL	3.5	(6 cyl)	Auto- S5	2WD	Gasoline	midsize car	7	17	26

```
In [3]: df_18.head(3)
```

Out[3]:

	model	displ	cyl	trans	drive	fuel	veh_class	air_pollution_score	city_mpg	hwy_n
0	ACURA RDX	3.5	6.0	SemiAuto- 6	2WD	Gasoline	small SUV	3	20	
1	ACURA RDX	3.5	6.0	SemiAuto- 6	4WD	Gasoline	small SUV	3	19	
2	ACURA TLX	2.4	4.0	AMS-8	2WD	Gasoline	small car	3	23	

```
In [4]: # check value counts for the 2008 cyl column
df_08['cyl'].value_counts()
```

```
Out[4]: (6 cyl)      409
        (4 cyl)      283
        (8 cyl)      199
        (5 cyl)       48
        (12 cyl)     30
        (10 cyl)     14
        (2 cyl)       2
        (16 cyl)      1
        Name: cyl, dtype: int64
```

Read [this \(https://stackoverflow.com/questions/35376387/extract-int-from-string-in-pandas\)](https://stackoverflow.com/questions/35376387/extract-int-from-string-in-pandas) to help you extract ints from strings in Pandas for the next step.

```
In [5]: # Extract int from strings in the 2008 cyl column
df_08['cyl'] = df_08.cyl.str.extract('(\d+)').astype(int)
```

```
In [6]: # Check value counts for 2008 cyl column again to confirm the change
df_08.cyl.value_counts()
```

```
Out[6]: 6      409
        4      283
        8      199
        5       48
        12     30
        10     14
        2       2
        16      1
        Name: cyl, dtype: int64
```

```
In [8]: # convert 2018 cyl column to int
df_18['cyl'] = df_18.cyl.astype(int)
```

```
In [9]: df_08.to_csv('data_08_v3.csv', index=False)
df_18.to_csv('data_18_v3.csv', index=False)
```

```
In [ ]:
```