

Fix city_mpg, hwy_mpg, cmb_mpg datatypes

2008 and 2018: convert string to float

Load datasets `data_08_v4.csv` and `data_18_v4.csv`. You should've created these data files in the previous section: *Fixing Data Types Pt 2*.

```
In [1]: # load datasets
import pandas as pd

df_08 = pd.read_csv('data_08_v4.csv')
df_18 = pd.read_csv('data_18_v4.csv')

In [2]: # convert mpg columns to floats
mpg_columns = ['city_mpg', 'hwy_mpg', 'cmb_mpg']
for c in mpg_columns:
    df_18[c] = df_18[c].astype(float)
    df_08[c] = df_08[c].astype(float)
```

Fix greenhouse_gas_score datatype

2008: convert from float to int

```
In [3]: # convert from float to int
df_08['greenhouse_gas_score'] = df_08.greenhouse_gas_score.astype(int)
```

All the datatypes are now fixed! Take one last check to confirm all the changes.

```
In [4]: df_08.dtypes
```

```
Out[4]: model                object
displ                float64
cyl                  int64
trans                object
drive                object
fuel                 object
veh_class            object
air_pollution_score float64
city_mpg             float64
hwy_mpg              float64
cmb_mpg              float64
greenhouse_gas_score int64
smartway             object
dtype: object
```

```
In [5]: df_18.dtypes
```

```
Out[5]: model                object
displ                float64
cyl                  int64
trans                object
drive                object
fuel                 object
veh_class            object
air_pollution_score float64
city_mpg             float64
hwy_mpg              float64
cmb_mpg              float64
greenhouse_gas_score int64
smartway             object
dtype: object
```

```
In [6]: df_08.dtypes == df_18.dtypes
```

```
Out[6]: model                True
displ                True
cyl                  True
trans                True
drive                True
fuel                 True
veh_class            True
air_pollution_score True
city_mpg             True
hwy_mpg              True
cmb_mpg              True
greenhouse_gas_score True
smartway             True
dtype: bool
```

```
In [7]: # Save your final CLEAN datasets as new files!
df_08.to_csv('clean_08.csv', index=False)
df_18.to_csv('clean_18.csv', index=False)
```

In []: