Cleaning Column Labels

Use `all_alpha_08.csv` and `all_alpha_18.csv`

In [19]:
```
import pandas as pd
from IPython.display import display
# load datasets

df_08 = pd.read_csv('all_alpha_08.csv')
df_18 = pd.read_csv('all_alpha_18.csv')
```

In [20]:
```
# view 2008 dataset
df_08.head(1)
```

Out[20]:

	Model	Displ	Cyl	Trans	Drive	Fuel	Sales Area	Stnd	Underhood ID	Veh Class	Air Pollution Score	FI Cal App
0	ACURA MDX	3.7	(6 cyl)	Auto-S5	4WD	Gasoline	CA	U2	8HNXT03.7PKR	SUV	7	Dr

In [21]:
```
# view 2018 dataset
df_18.head(1)
```

Out[21]:

	Model	Displ	Cyl	Trans	Drive	Fuel	Cert Region	Stnd	Stnd Description	Underhood ID	G
0	ACURA RDX	3.5	6.0	SemiAuto-6	2WD	Gasoline	FA	T3B125	Federal Tier 3 Bin 125	JHNXT03.5GV3	

Drop Extraneous Columns

In [22]:
```
# drop columns from 2008 dataset
df_08.drop(['Stnd', 'Underhood ID', 'FE Calc Appr', 'Unadj Cmb MPG'], ax
is=1, inplace=True)

# confirm changes
df_08.head(1)
```

Out[22]:

	Model	Displ	Cyl	Trans	Drive	Fuel	Sales Area	Veh Class	Air Pollution Score	City MPG	Hwy MPG	Cmb MPG	Greenh Gas S
0	ACURA MDX	3.7	(6 cyl)	Auto-S5	4WD	Gasoline	CA	SUV	7	15	20	17	

In [23]:
```
# drop columns from 2018 dataset
df_18.drop(['Stnd', 'Stnd Description', 'Underhood ID', 'Comb CO2'], axis=1, inplace=True)

# confirm changes
df_18.head(1)
```

Out[23]:

	Model	Displ	Cyl	Trans	Drive	Fuel	Cert Region	Veh Class	Air Pollution Score	City MPG	Hwy MPG	Cmb MPG	G
0	ACURA RDX	3.5	6.0	SemiAuto-6	2WD	Gasoline	FA	small SUV	3	20	28	23	

Rename Columns

In [24]:
```
# rename Sales Area to Cert Region
df_08.rename(columns={'Sales Area':'Cert Region'}, inplace=True)

# confirm changes
df_08.head(1)
```

Out[24]:

	Model	Displ	Cyl	Trans	Drive	Fuel	Cert Region	Veh Class	Air Pollution Score	City MPG	Hwy MPG	Cmb MPG	Greer Gas
0	ACURA MDX	3.7	(6 cyl)	Auto-S5	4WD	Gasoline	CA	SUV	7	15	20	17	

In [25]:
```
# replace spaces with underscores and lowercase labels for 2008 dataset
df_08.rename(columns=lambda x: x.strip().lower().replace(" ", "_"), inplace=True)

# confirm changes
df_08.head(1)
```

Out[25]:

	model	displ	cyl	trans	drive	fuel	cert_region	veh_class	air_pollution_score	city_mpg
0	ACURA MDX	3.7	(6 cyl)	Auto-S5	4WD	Gasoline	CA	SUV	7	15

In [26]:
```
# replace spaces with underscores and lowercase labels for 2018 dataset
df_18.rename(columns=lambda x: x.strip().lower().replace(" ","_"), inplace=True)

# confirm changes
df_18.head(1)
```

Out[26]:

	model	displ	cyl	trans	drive	fuel	cert_region	veh_class	air_pollution_score	city_
0	ACURA RDX	3.5	6.0	SemiAuto-6	2WD	Gasoline	FA	small SUV	3	

In [27]:
```
# confirm column labels for 2008 and 2018 datasets are identical
df_08.columns == df_18.columns
```

Out[27]:
```
array([ True,  True,  True,  True,  True,  True,  True,  True,  True,
        True,  True,  True,  True,  True], dtype=bool)
```

In [28]:
```
# make sure they're all identical like this
(df_08.columns == df_18.columns).all()
```

Out[28]:
```
True
```

In [29]:
```
# save new datasets for next section
df_08.to_csv('data_08_v1.csv', index=False)
df_18.to_csv('data_18_v1.csv', index=False)
```

In [ ]:
```
df
```