Data Importing Exercises

In [ ]:

Importing Libraries

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

import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
sns.set()
```

In [ ]:

Q. **Import** the csv-file **cars_raw.csv** with the appropriate pandas method and **inspect** the data!

In [2]:
```
cars = pd.read_csv('cars_raw.csv')
cars
```

Out[2]:

					Welcome	to	the	cars	Dataset!
Feel	**free**	**to**	**analyze**	**and**	clean	the	messy	Dataset	!
18.0	**8**	**307.0**	**130.0 hp**	**3504**	12.0	70	United States	chevrolet chevelle malibu	NaN
15.0	**8**	**350.0**	**165.0 hp**	**3693**	11.5	70	United States	buick skylark 320	NaN
18.0	**8**	**318.0**	**150.0 hp**	**3436**	11.0	70	United States	plymouth satellite	NaN
16.0	**8**	**304.0**	**150.0 hp**	**3433**	12.0	70	usa	amc rebel sst	NaN
...	**...**	**...**	**...**	**...**	...	...	...	...	...
27.0	**4**	**101.0**	**83.0 hp**	**2202**	15.3	76	europe	renault 12tl	NaN
17.0	**6**	**250.0**	**100.0 hp**	**3329**	15.5	71	usa	chevrolet chevelle malibu	NaN
14.5	**8**	**351.0**	**152.0 hp**	**4215**	12.8	76	usa	ford gran torino	NaN
25.0	**6**	**181.0**	**110.0 hp**	**2945**	16.4	82	usa	buick century limited	NaN
Thanks	**for**	**analyzing**	**this**	**Dataset**	!	NaN	NaN	NaN	NaN

331 rows × 5 columns

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

Out[3]:

					Welcome	to	the	cars	Dataset!
Feel	**free**	**to**	**analyze**	**and**	clean	the	messy	Dataset	!
18.0	**8**	**307.0**	**130.0 hp**	**3504**	12.0	70	United States	chevrolet chevelle malibu	NaN
15.0	**8**	**350.0**	**165.0 hp**	**3693**	11.5	70	United States	buick skylark 320	NaN
18.0	**8**	**318.0**	**150.0 hp**	**3436**	11.0	70	United States	plymouth satellite	NaN
16.0	**8**	**304.0**	**150.0 hp**	**3433**	12.0	70	usa	amc rebel sst	NaN

In [4]:
```
cars.tail()
```

Out[4]:

					Welcome	to	the	cars	Dataset!
27.0	**4**	**101.0**	**83.0 hp**	**2202**	15.3	76	europe	renault 12tl	NaN
17.0	**6**	**250.0**	**100.0 hp**	**3329**	15.5	71	usa	chevrolet chevelle malibu	NaN
14.5	**8**	**351.0**	**152.0 hp**	**4215**	12.8	76	usa	ford gran torino	NaN
25.0	**6**	**181.0**	**110.0 hp**	**2945**	16.4	82	usa	buick century limited	NaN
Thanks	**for**	**analyzing**	**this**	**Dataset**	!	NaN	NaN	NaN	NaN

In [5]:
```
cars.info()
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 331 entries, ('Feel ', ' free', ' to', ' analyze', ' and') to ('Thanks',
' for ', ' analyzing', ' this', ' Dataset')
Data columns (total 5 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Welcome    331 non-null    object
 1    to        330 non-null    object
 2    the       330 non-null    object
 3    cars      330 non-null    object
 4    Dataset!  1 non-null      object
dtypes: object(5)
memory usage: 35.8+ KB
```

In [ ]:

In [ ]:

In [ ]:

Q. Use appropriate parameters in the pd.read_csv() method to clean the format. The following issues need to be solved:

- **Remove** the **first row(s)** containing nonsense content.
- **Remove** the **last row(s)** containing nonsense content.
- Define that there are **no appropriate column labels/headers** in the data.
- **Set** the following **column labels/headers**:

labels = ['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration', 'model year', 'origin', 'name']

In [6]:
```
labels = ['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleration
```

In [8]:
```
cars = pd.read_csv('cars_raw.csv',
                   skiprows=2,
                    skipfooter=1,
                    header=None,
                    names=labels
                   )
cars
```

```
C:\Users\alhef\AppData\Local\Temp\ipykernel_11696\1796019620.py:1: ParserWarning: Fa
lling back to the 'python' engine because the 'c' engine does not support skipfoote
r; you can avoid this warning by specifying engine='python'.
  cars = pd.read_csv('cars_raw.csv',
```

Out[8]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	
0	18.0	8	307.0	130.0 hp	3504	12.0	70	United States	ch c
1	15.0	8	350.0	165.0 hp	3693	11.5	70	United States	
2	18.0	8	318.0	150.0 hp	3436	11.0	70	United States	ply s
3	16.0	8	304.0	150.0 hp	3433	12.0	70	usa	am
4	17.0	8	302.0	140.0 hp	3449	10.5	70	usa	T
...	...	...	...	...	...	...	...	...	
324	12.0	8	429.0	198.0 hp	4952	11.5	73	usa	m m bro
325	27.0	4	101.0	83.0 hp	2202	15.3	76	europe	
326	17.0	6	250.0	100.0 hp	3329	15.5	71	usa	ch c
327	14.5	8	351.0	152.0 hp	4215	12.8	76	usa	for
328	25.0	6	181.0	110.0 hp	2945	16.4	82	usa	c

329 rows × 9 columns

In [9]:
```
cars.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 329 entries, 0 to 328
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   mpg           329 non-null    float64
 1   cylinders     329 non-null    int64  
 2   displacement  329 non-null    float64
 3   horsepower    329 non-null    object 
 4   weight        329 non-null    int64  
 5   acceleration  329 non-null    float64
 6   model year    329 non-null    int64  
 7   origin        329 non-null    object 
 8   name          329 non-null    object 
dtypes: float64(3), int64(3), object(3)
memory usage: 23.3+ KB
```

In [ ]:

Q. Once you are happy with the import, export and save cars as new csv-file (cars_new.csv). Do not export any RangeIndex!

In [11]:
```
cars.to_csv('cars_new_edit.csv', index=False)
```

In [12]:
```
pd.read_csv('cars_new_edit.csv')
```

Out[12]:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	
0	18.0	8	307.0	130.0 hp	3504	12.0	70	United States	ch c
1	15.0	8	350.0	165.0 hp	3693	11.5	70	United States	
2	18.0	8	318.0	150.0 hp	3436	11.0	70	United States	ply s
3	16.0	8	304.0	150.0 hp	3433	12.0	70	usa	am
4	17.0	8	302.0	140.0 hp	3449	10.5	70	usa	T
...	...	...	...	...	...	...	...	...	
324	12.0	8	429.0	198.0 hp	4952	11.5	73	usa	m m bro
325	27.0	4	101.0	83.0 hp	2202	15.3	76	europe	
326	17.0	6	250.0	100.0 hp	3329	15.5	71	usa	ch c
327	14.5	8	351.0	152.0 hp	4215	12.8	76	usa	for
328	25.0	6	181.0	110.0 hp	2945	16.4	82	usa	c

329 rows × 9 columns

==========

GOOD LUCK!