Student Alcohol Consumption

Introduction:

This time you will download a dataset from the UCI.

Step 1. Import the necessary libraries

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

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called df.

In [2]:
```
csv_url = 'https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/04_Apply/Students_Alcohol_Consumption/student-mat.csv'
df = pd.read_csv(csv_url)
df.head()
```

Out[2]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10

5 rows × 33 columns

Step 4. For the purpose of this exercise slice the dataframe from 'school' until the 'guardian' column

In [3]:
```
stud_alcoh = df.loc[: , "school":"guardian"]
stud_alcoh.head()
```

Out[3]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother
1	GP	F	17	U	GT3	T	1	1	at_home	other	course	father
2	GP	F	15	U	LE3	T	1	1	at_home	other	other	mother
3	GP	F	15	U	GT3	T	4	2	health	services	home	mother
4	GP	F	16	U	GT3	T	3	3	other	other	home	father

Step 5. Create a lambda function that will capitalize strings.

In [4]:
```
capitalizer = lambda x: x.capitalize()
```

Step 6. Capitalize both Mjob and Fjob

In [5]:
```
stud_alcoh['Mjob'].apply(capitalizer)
stud_alcoh['Fjob'].apply(capitalizer)
```

Out[5]:
```
0       Teacher
1         Other
2         Other
3      Services
4         Other
         ...
390    Services
391    Services
392       Other
393       Other
394     At_home
Name: Fjob, Length: 395, dtype: object
```

Step 7. Print the last elements of the data set.

In [6]:
```
stud_alcoh.tail()
```

Out[6]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian
390	MS	M	20	U	LE3	A	2	2	services	services	course	other
391	MS	M	17	U	LE3	T	3	1	services	services	course	mother
392	MS	M	21	R	GT3	T	1	1	other	other	course	other
393	MS	M	18	R	LE3	T	3	2	services	other	course	mother
394	MS	M	19	U	LE3	T	1	1	other	at_home	course	father

Step 8. Did you notice the original dataframe is still lowercase? Why is that? Fix it and capitalize Mjob and Fjob.

In [7]:
```
stud_alcoh['Mjob'] = stud_alcoh['Mjob'].apply(capitalizer)
stud_alcoh['Fjob'] = stud_alcoh['Fjob'].apply(capitalizer)
stud_alcoh.tail()
```

Out[7]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian
390	MS	M	20	U	LE3	A	2	2	Services	Services	course	other
391	MS	M	17	U	LE3	T	3	1	Services	Services	course	mother
392	MS	M	21	R	GT3	T	1	1	Other	Other	course	other
393	MS	M	18	R	LE3	T	3	2	Services	Other	course	mother
394	MS	M	19	U	LE3	T	1	1	Other	At_home	course	father

Step 9. Create a function called majority that returns a boolean value to a new column called legal_drinker (Consider majority as older than 17 years old)

In [8]:
```
def majority(age):
    if age>17:
        return 1
    else:
        return 0
```

In [9]:
```
stud_alcoh["legal_drinker"] = stud_alcoh["age"].apply(majority)
stud_alcoh
```

Out[9]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	legal_drinker
0	GP	F	18	U	GT3	A	4	4	At_home	Teacher	course	mother	1
1	GP	F	17	U	GT3	T	1	1	At_home	Other	course	father	0
2	GP	F	15	U	LE3	T	1	1	At_home	Other	other	mother	0
3	GP	F	15	U	GT3	T	4	2	Health	Services	home	mother	0
4	GP	F	16	U	GT3	T	3	3	Other	Other	home	father	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
390	MS	M	20	U	LE3	A	2	2	Services	Services	course	other	1
391	MS	M	17	U	LE3	T	3	1	Services	Services	course	mother	0
392	MS	M	21	R	GT3	T	1	1	Other	Other	course	other	1
393	MS	M	18	R	LE3	T	3	2	Services	Other	course	mother	1
394	MS	M	19	U	LE3	T	1	1	Other	At_home	course	father	1

395 rows × 13 columns

Step 10. Multiply every number of the dataset by 10.

I know this makes no sense, don't forget it is just an exercise

In [10]:
```
stud_alcoh.applymap(lambda x: x*10 if type(x) is int else x)
```

Out[10]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	legal_drinker
0	GP	F	180	U	GT3	A	40	40	At_home	Teacher	course	mother	10
1	GP	F	170	U	GT3	T	10	10	At_home	Other	course	father	0
2	GP	F	150	U	LE3	T	10	10	At_home	Other	other	mother	0
3	GP	F	150	U	GT3	T	40	20	Health	Services	home	mother	0
4	GP	F	160	U	GT3	T	30	30	Other	Other	home	father	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
390	MS	M	200	U	LE3	A	20	20	Services	Services	course	other	10
391	MS	M	170	U	LE3	T	30	10	Services	Services	course	mother	0
392	MS	M	210	R	GT3	T	10	10	Other	Other	course	other	10
393	MS	M	180	R	LE3	T	30	20	Services	Other	course	mother	10
394	MS	M	190	U	LE3	T	10	10	Other	At_home	course	father	10

395 rows × 13 columns