## Exercises based on workbooks 4

Now switch back to using the Car\_Insurance\_Claim.csv

Include a Screen of each of the results below please.

Merge

It just didn’t make sense to merge a dataset to car insurance.

Instead put a screen of the results below

<https://realpython.com/pandas-merge-join-and-concat/>

#### Inner Join

A screenshot of a computer

Description automatically generated

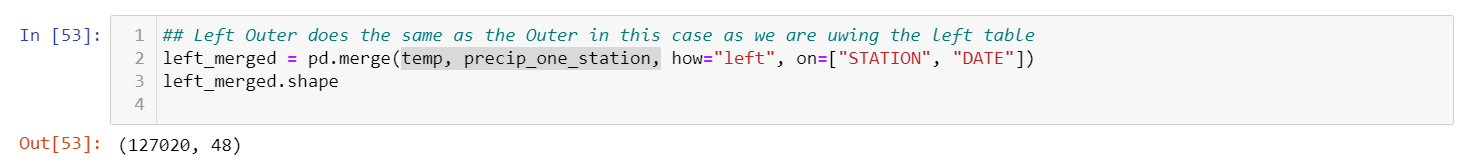
#### 

#### Outer Join

A screenshot of a computer

Description automatically generated

#### Left Join



#### Right Join

#### A screenshot of a computer code Description automatically generated

**Concatenation**

This doesn't make sense either, however it’s one way to practise concatenation!  
Concatenate the 4 csv files we have been working with together into a new data file



Do the above, but add one parameter of your choosing to a difference concatenation command

Here’s the documentation to help

<https://pandas.pydata.org/docs/reference/api/pandas.concat.html>

A screenshot of a computer

Description automatically generated

Go back to the original car insurance and find which records have nans in.  
Remove those records. How many remain? 8149

A screenshot of a computer

Description automatically generated

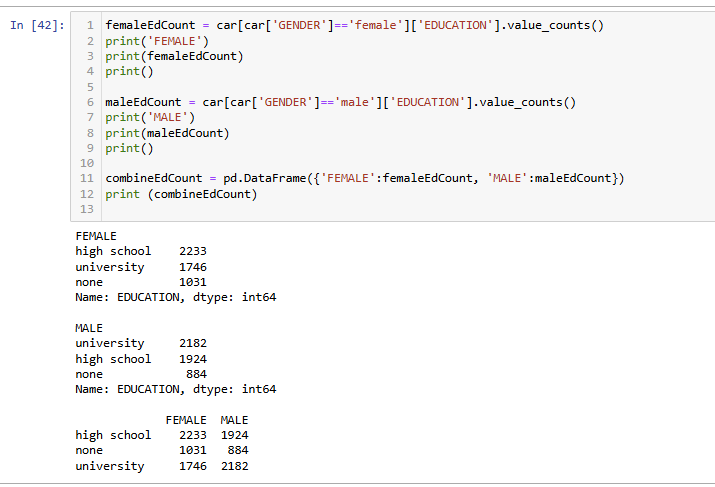
Find the value counts for income, then comment about the strange nature of this population in the dataset.

A screenshot of a computer

Description automatically generated

It could be argued that the upper class may have out layers that incorrectly weight the mean

Compare the value counts of women’s education compared to men’s education in this dataset.  
This will likely be two different commands.

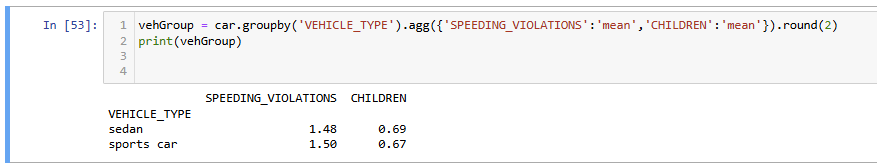


Do the same for ‘sports car’ and sedan and speeding violations. Just take a screen of the results for now

A screenshot of a computer

Description automatically generated

Now perform a group by of vehicle type, and include an aggregate mean of speeding violations and children. Which vehicle type has the highest of each?



**Pivot Tables**

Note, I’ve found a resource that adds further complexity to this with formula I prefer:

<https://www.geeksforgeeks.org/how-to-create-a-pivot-table-in-python-using-pandas/>

Create a pivot table which uses DRIVING EXPERIANCE as an index, has the values ‘Updated Risk’ and the aggfun as mean. What do experience levels indicate for risk rating according to the risk rating we created?

(If you’ve lost the risk rating we created, use PAST\_ACCIDENTS instead)

A white rectangular box with text

Description automatically generated with medium confidence

Create a pivot table with the two indexes INCOME and GENDER, make the values CREDIT\_ SCORE, and include the aggregate function Median, Mean and Min

A screenshot of a computer

Description automatically generated

Create any Pivot table of your choice that I haven’t asked for that generates an insight of some description about our drivers. What are the results and your insight

A screenshot of a computer program

Description automatically generated