

## Exercise 1

1. Load any stata file you have into a Pandas DataFrame, rename your file as “Results\_b”, and display the first 4 rows. What is the size of the dataframe?
2. Use the .describe() function to summarize the statistics of a numerical column from a dataset. What insights can you infer from the output?
3. Rename the variables s1q1,s1q2,s1q3y and s3q3 as Sex, Relat\_Hhead, Age, and Health\_insurance respectively.
4. From the original dataframe, select the following variables:  
hhid','province','ur2\_2012','quintile','poverty','Consumption', Sex, Relat\_Hhead ','Age',  
and 'Health\_insurance'
5. Summarize the average Age or Consumption by Province and Urban\_Rural.
6. Which commands that are used to detect outliers for the variable “Age”? run and interpreted the output.
7. Run the percentages of males and females by quintile
8. Analyze the percentages of males and females by province and urban\_rural
9. Document your steps (Provide titles and comments throughout the script to explain each command used and outputs)