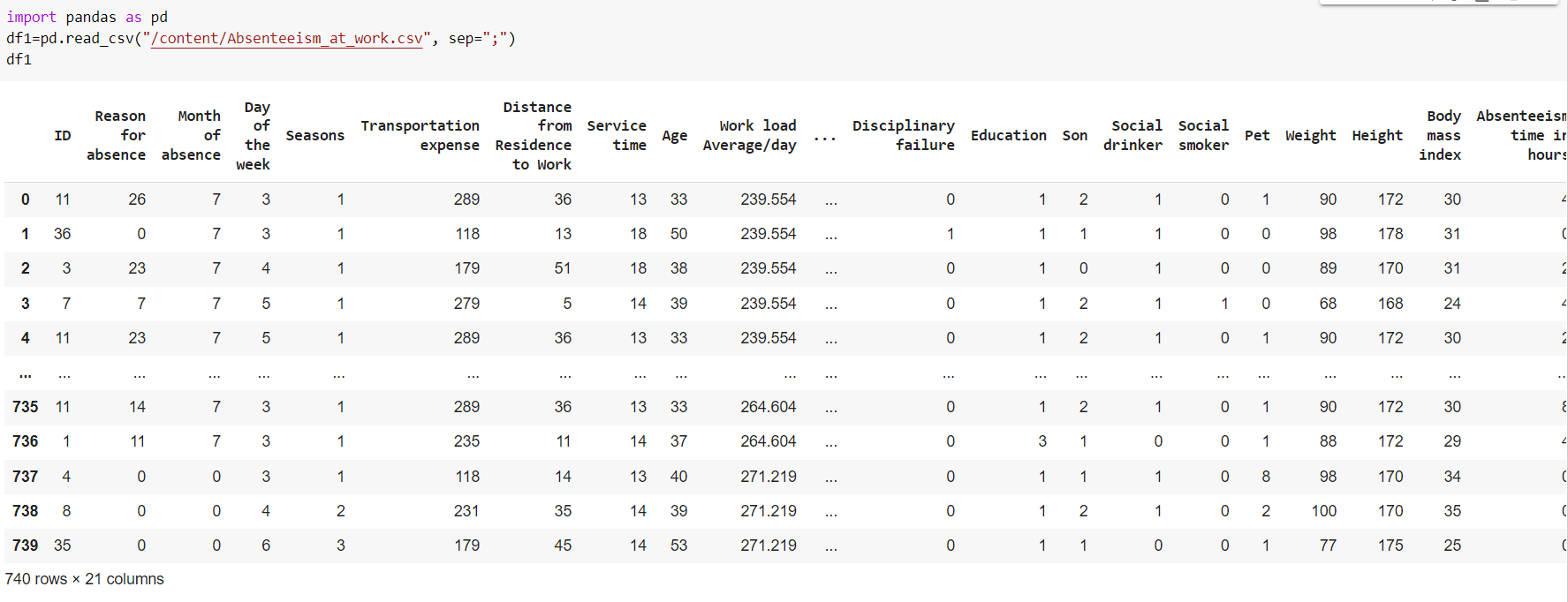
NAME: PARIPOORNA D

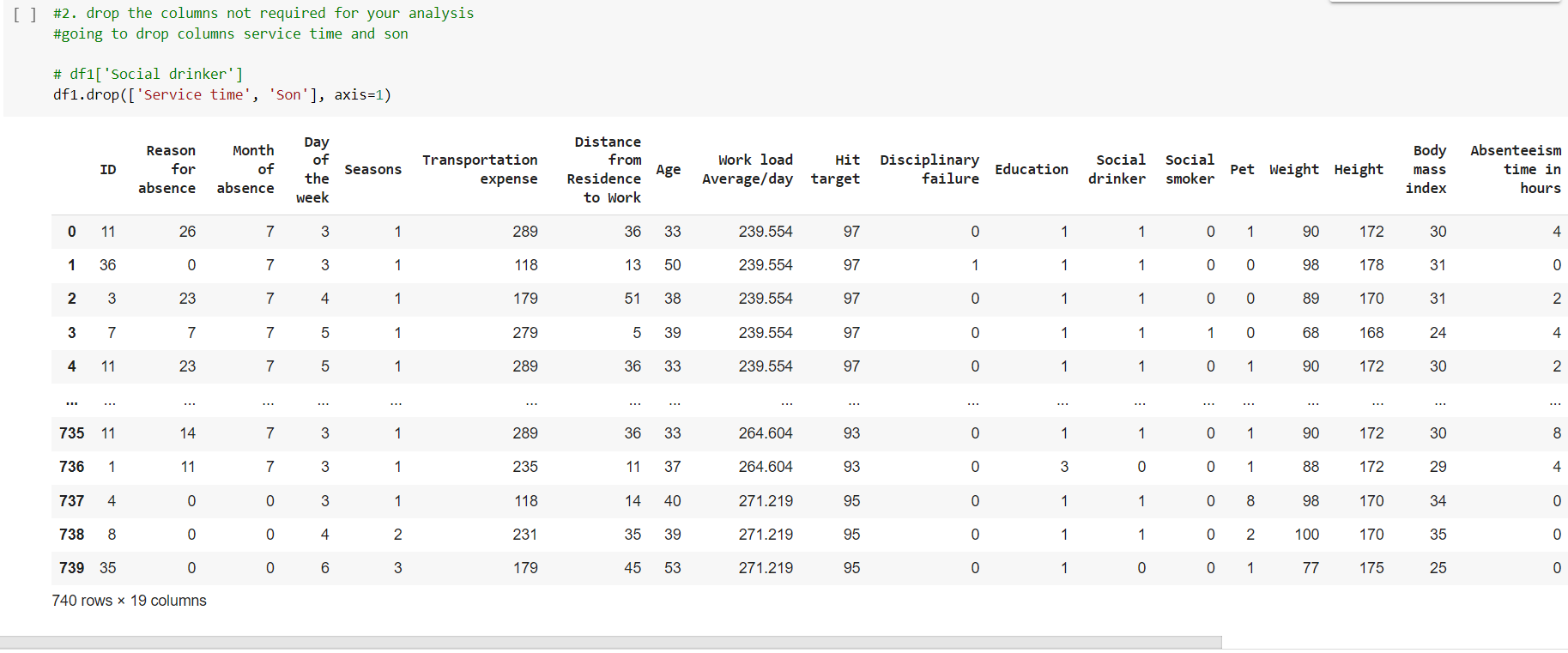
ROLL\_NO: CB.EN.U4CSE20245

SET1:

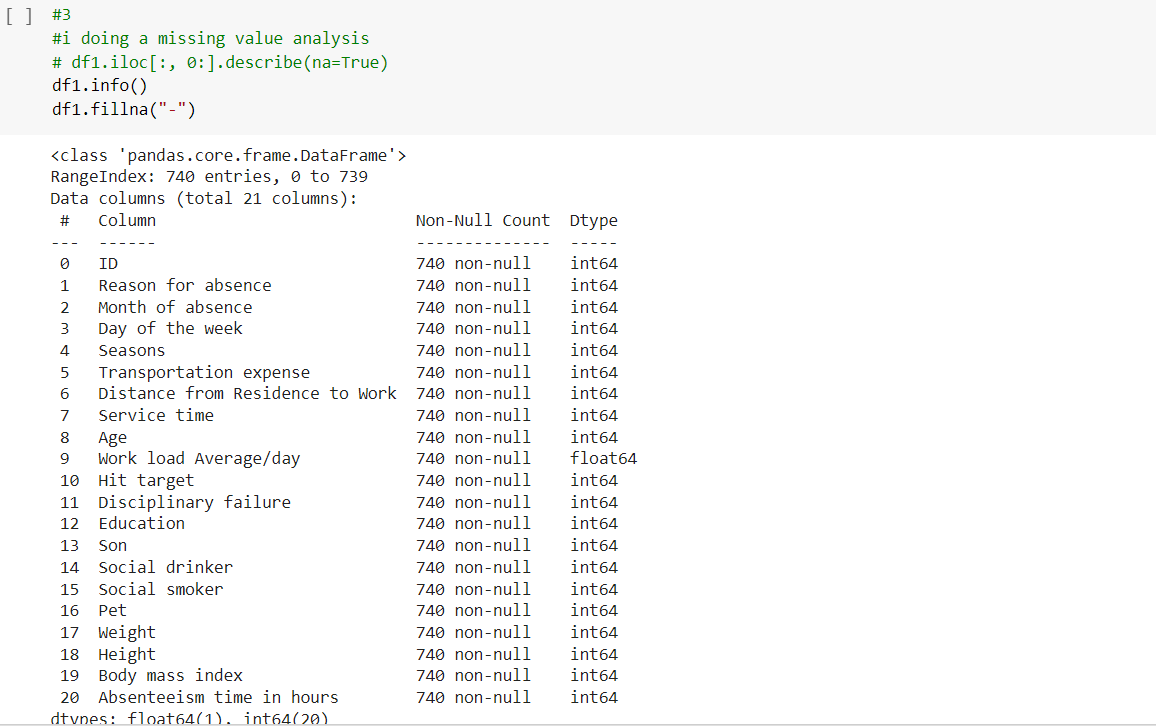
1. Import data and save it in your disk space (1)

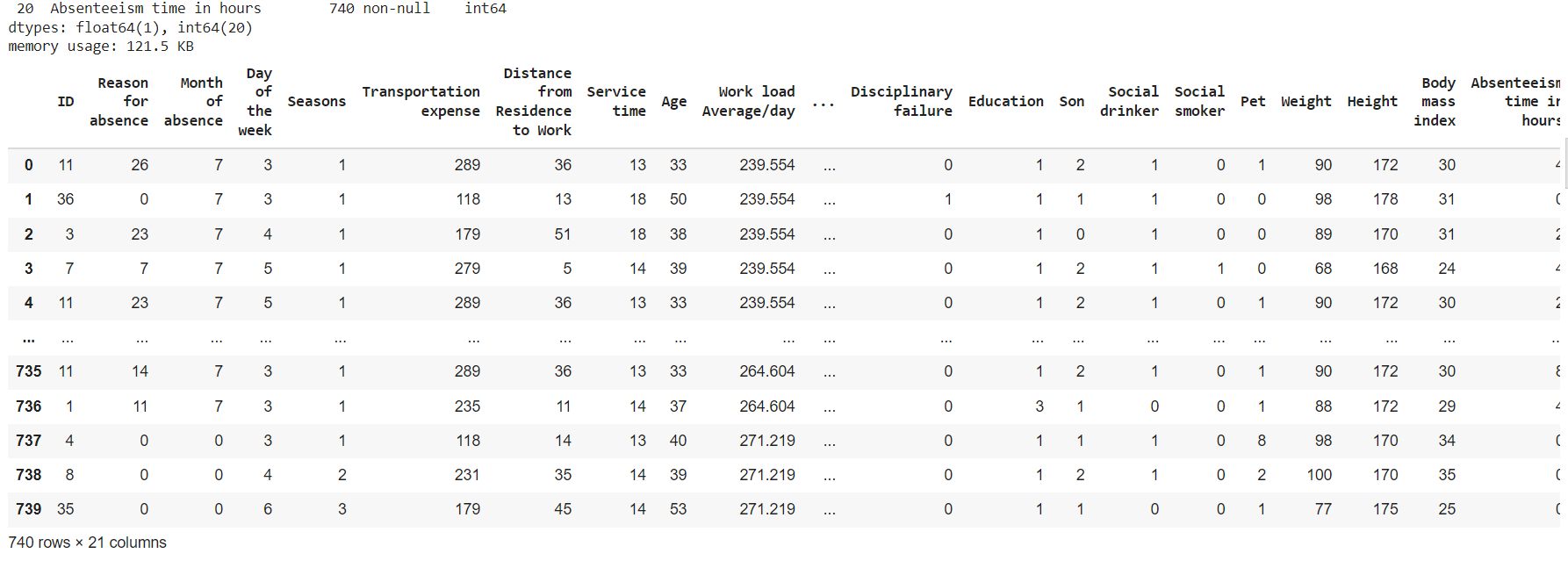


1. Drop the columns not required for your analysis. (1)

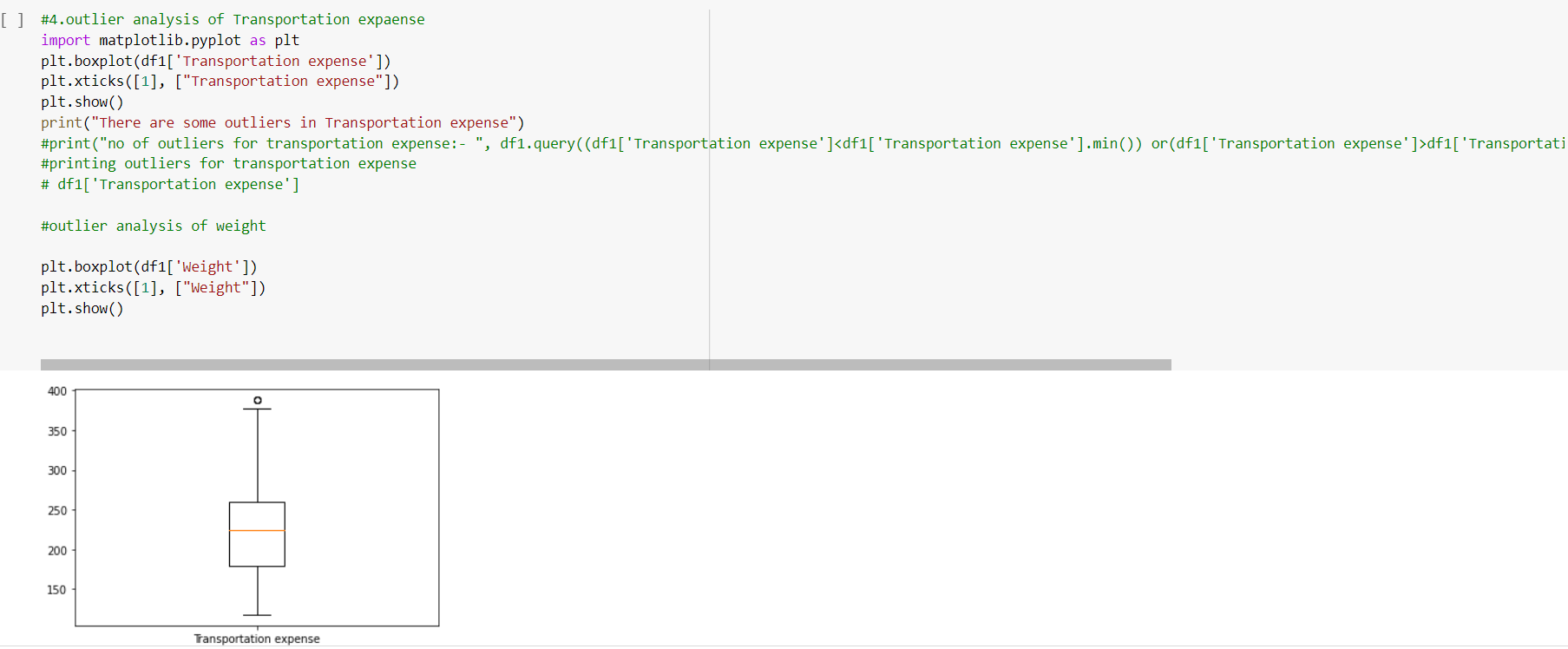


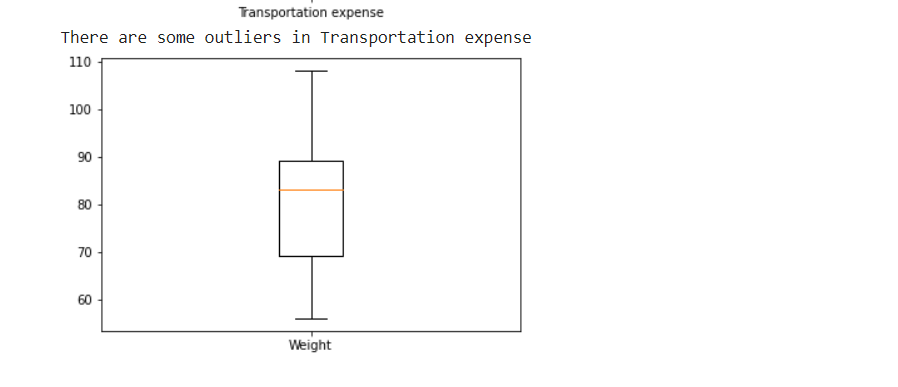
1. Do a missing value analysis and perform appropriate steps to correct them (2)



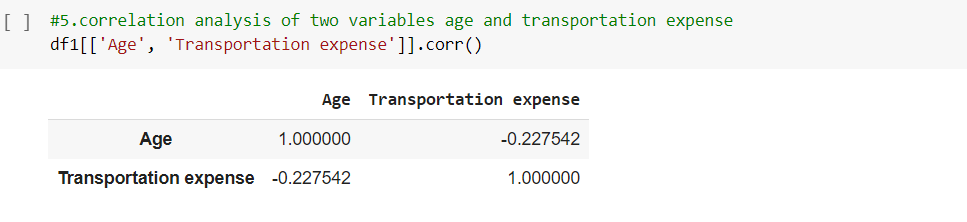


1. Perform an outlier analysis for any two columns (2)

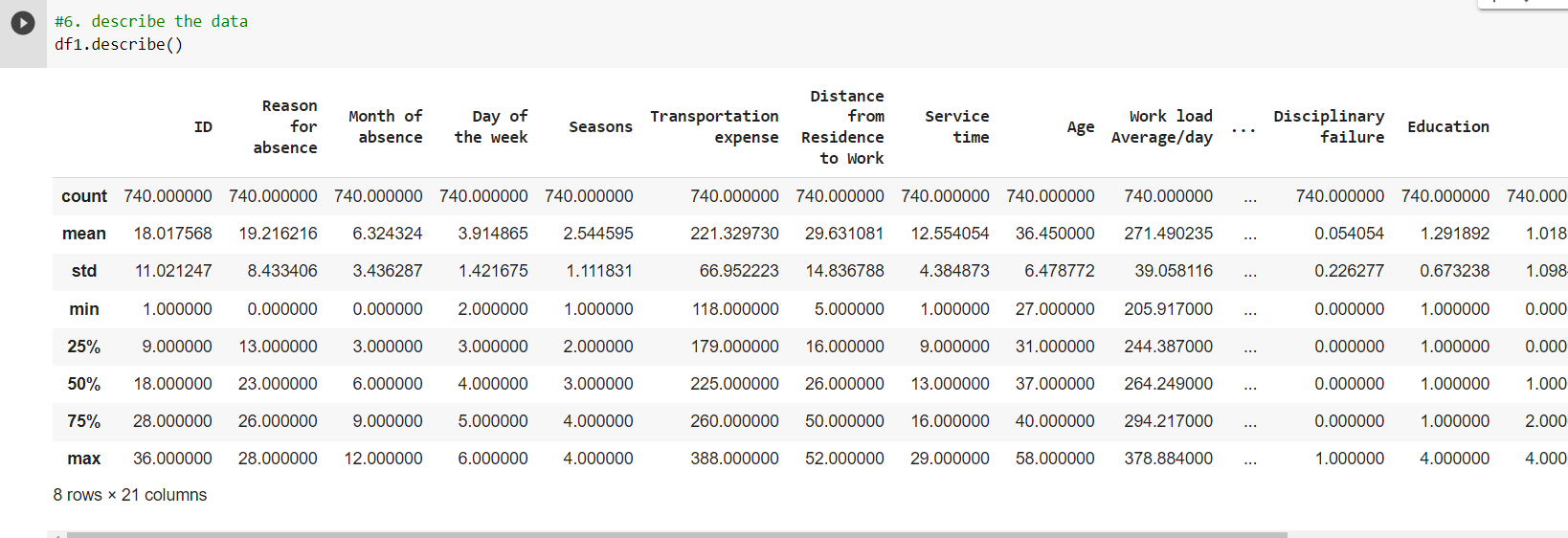




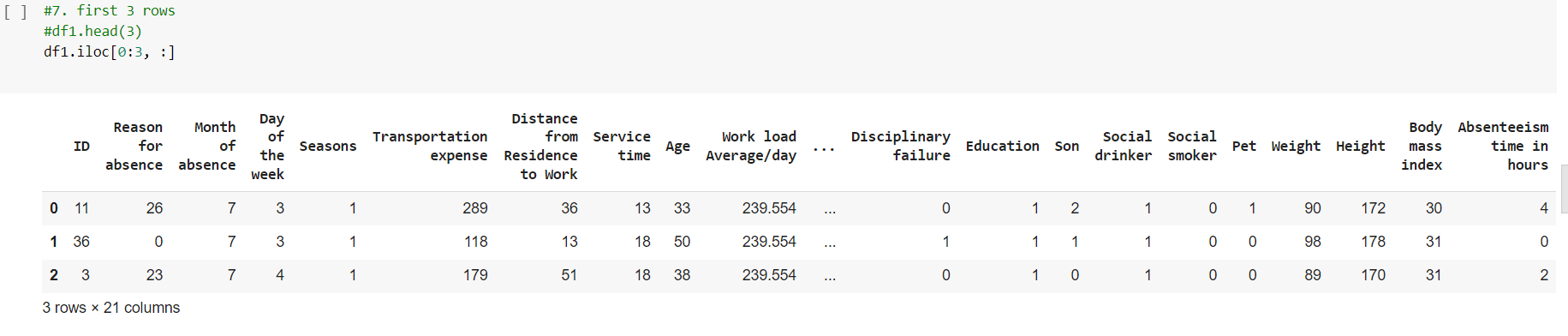
1. Do a correlation analysis of any two variables. (2)



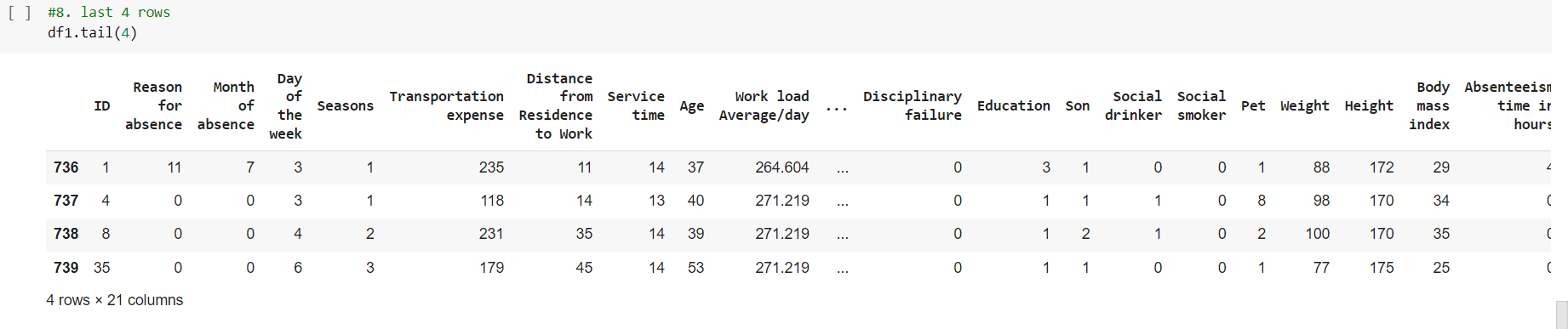
1. Describe the data (1).



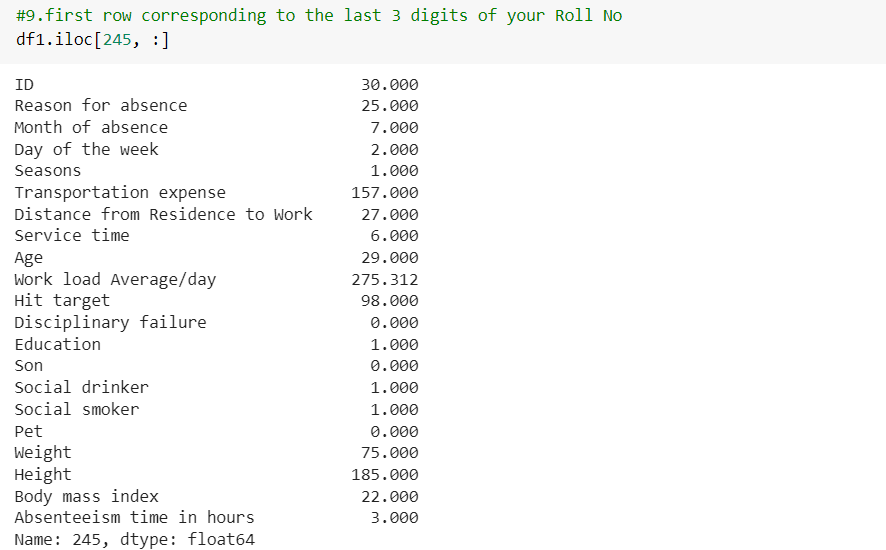
1. Display the first 3 rows (1).



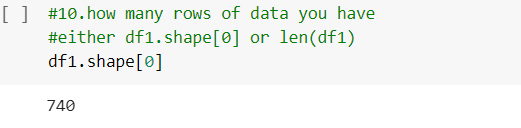
1. Display the last 4 rows(1)



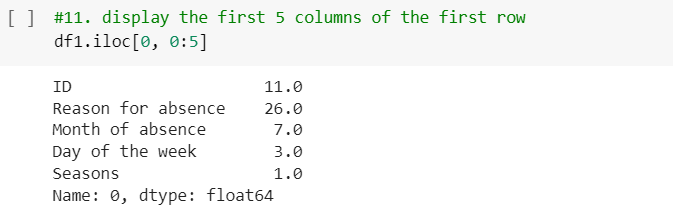
1. Locate the first row corresponding to the last 3 digits of your Roll No (2)



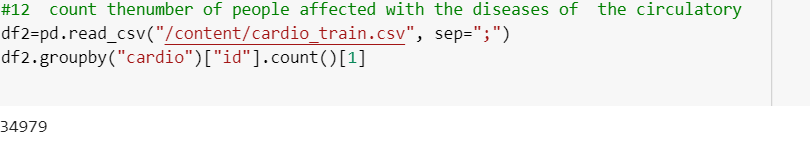
1. How many rows of data do you have? (1)



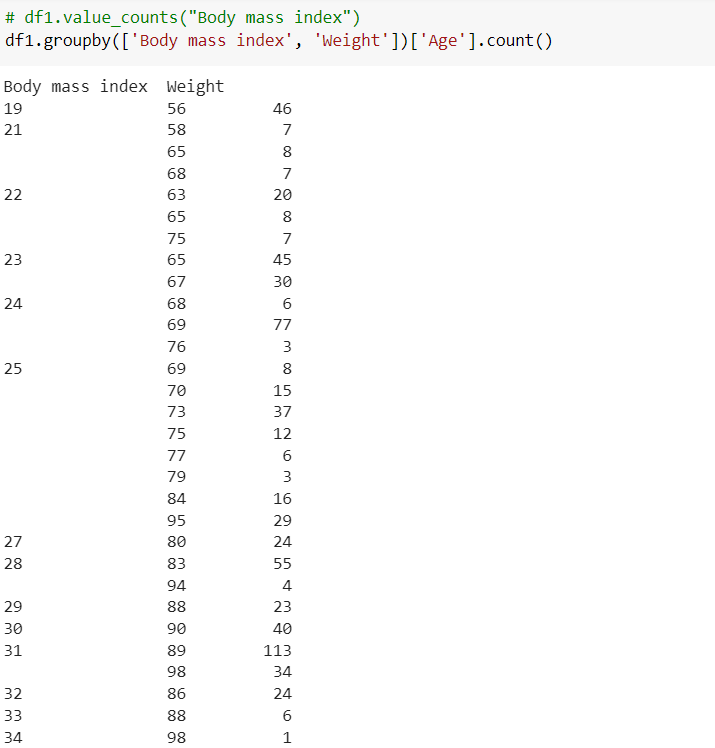
1. Display the first 5 columns of the first row (2)

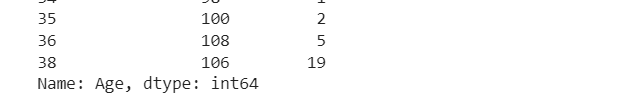


1. Count the number of people affected with the diseases of the circulatory system (2)

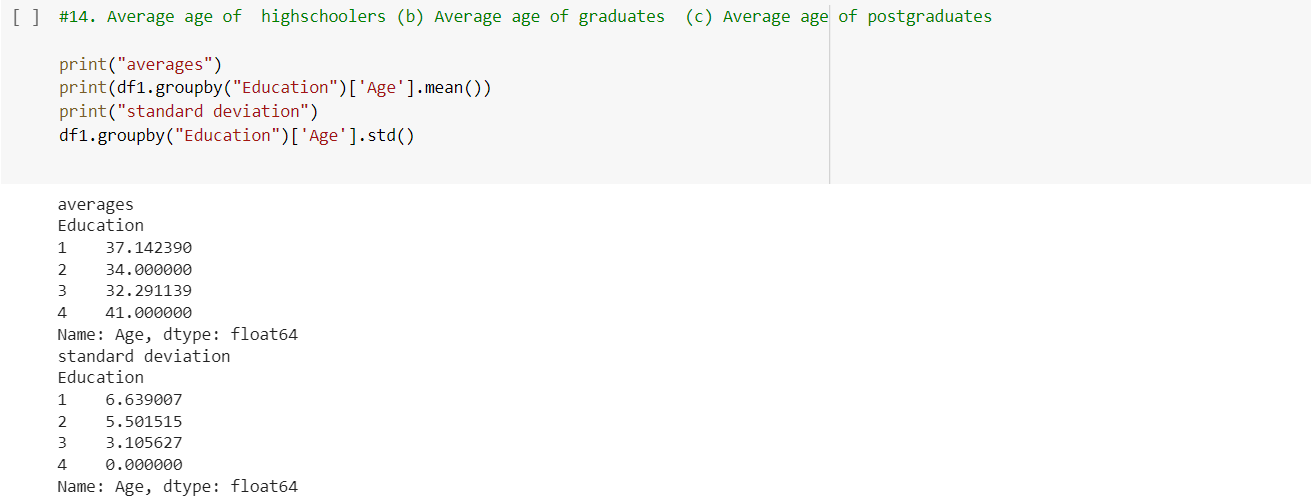


1. Split people according to their body mass index into two groups: overweight (25 to 30) and obese (>30). (2)

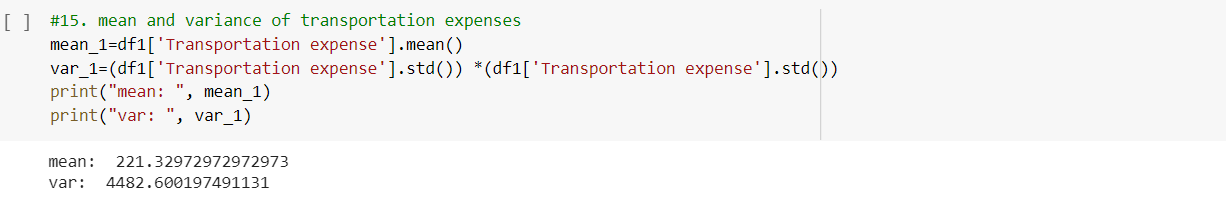




1. Evaluate the characteristics of the population distribution with mean and, deviation (a) Average age of highschoolers (b) Average age of graduates (c) Average age of postgraduates (2)



1. Compute the mean and the variance of "transportation expenses” for medical consultations (1).



1. Draw a histogram for age of Social drinkers. Histogram should be step-filled with bin of size 20. Find the skew of this distribution, and comment if it is positive or negative (5).

