Data Wrangling

The data set consists of 4 csv files.
1. Employee Survey - Contains the feedback provided by the employee
2. General Data - Contains general data about the employee such as Dept, distance from home, gender, age etc
3. In Time - Contains the login time of the employee
4. Out Time - Contains the logout time of the employee
5. Manager Survey - Contains the feedback given by the manager about the employee

Employee Survey:
1. Employees have given feedback on three parameters (Environment Satisfaction, Job Satisfaction and WorkLifeBalance). Ratings are in the range of 1-5. There are certain rows where one of the three parameters is an "NA". In such cases, the "NA" is replaced with the average of the other two parameters.

Manager Survey:
1. No Null Values.

In Time and Out Time:
1. First column is unnamed in both In and Out csv files. Comparing the values in the first column with the values in other csv files, we can conclude that the first column is "Employee ID".
2. Non Working days (weekends, public holidays) contain NA which can be removed as it is a valid NA.
3. If the In time and Out time for a single employee is "NA" on the same date, we can conclude the employee was on a PL/SL. This is also a valid NA and can be removed.
4. In scenarios where one of the entries is missing (either In Time or Out Time) for an employee we can replace such "NA" values with the mean value of the in and out times of the employee.