**Employee Data Assessment Part**

The data provided is the data of more than 400 employees & this data is about the courses they are opting for from the selection of 4 options (Anti Bribery Fundamentals, Competition Law, Cyber Security Essentials, Employee Harassment). All these employees have been assigned to a different Business Divisions which are Corporate , Engineering, Environmental, Manufacturing, Mining & Transportation, & in these Business Divisions workers are working in different work locations with a work location ID Number(#) & the details given by the employees are their E-mail addresses along with this the data gives a brief knowledge about when these employees were hired with the exact date of joining, the most important information regarding this course is the completion status of the each particular course that the employees enrolled into and it is sub-divided into 3 categories which are Completed , In-Progress & Not Started which goes for the 3 types if the employees have completed the course they opted for , if their course is an on-going state or they didn’t even start the course they opted for.

**Problems with the Data: -**

* While going through the data it can be surveyed that the employees who work in the mining business division have no employees who have neither completed their course nor have started working on it yet.
* At least 8 data Rows are similar in the employee’s name section and has been repeated and the person named Bridget Salazar has been repeated 8 times but with different Work Id Location Number (#) but with the same Business Division.
* Similarly, the employee’s name with Jim Sellers in the employee’s name section has been repeated 10 times but the only difference is that the Business Division is different this time and the Work Location Id Number is same.
* The location work Id number provided to all the employee is uneven which means the employees are working in different areas are uneven that is in some locations there are workers up to 29 but in some locations, there are like 1 employee working in 1 location.
* There are multiple similar work location Id number (#) provided for different employee.
* Most of the people are in the Not Started state for their course which means most of them didn’t start their course yet & that is around 46.65% of the total number of employees which are more than 400.
* There is no date for the course start provided.
* Regardless of when the employees who completed the course the completion year for all of them is in the year 2022 only.
* The locations specifics are not exact & the classification according to the location ID number is not really directly achievable.
* There should have been a direct list for the names of the locations so that the classification & the arrangement could have been easier at the end for a proper analysis.

**Solutions & suggestions to fix the data: -**

* Employees working in the mining business division should either be motivated in order to get the course started or should be allowed with some free time to let them complete the course that they opted for.
* Repetition of data should be taken care of while entering the data as none of the data is unique & it is necessary that at least one of the fields provided in the data set should be distinct for proper classification.
* Employees should be uniformly divided to different locations so that the work load should be equal and should not cause any other employee a hectic experience throughout their work hour.
* Work location Id Number (#) with more than one employee should be brought to a halt.
* More & more employees who haven’t started their course should be encouraged more & should be given proper time or decrease work burden in order for them to let start & finish the course.
* A date should be provided for all the employees to start their course in order to finish it by the deadline.
* The employees which are in the ”In progress” state for their courses should be encouraged more to complete their course on time and shouldn’t let them be similar the employees from the Completed Status who completed their courses in the very last year.
* Locations should have been provided for distinct and more proper classification.
* The location ID number should have been more accurate for better approach towards the data .