**Employee Data**

Submitted by: Dipshikha Purkayastha

Contact: 9101732847

Email: dipshikhapurkayastha@gmail.com

**Objective:** Transform current employee data from a columnar format into a historical, row-based versioning format suitable for database storage.

**EXCEL FILE**

Steps used to convert the input data into a structured format representing historical records of employee compensation, engagement, and performance reviews.

* **Employee Code:** The employee code is given in the input file.
* **Manager Employee Code:** This information is given in the input file. This data represent which employee is working under which manager.
* **Last Compensation:** This column represents the last compensation information of any employee in the company.
  + If the employee is new joining than the last compensation will be 0.
* **Compensation:** This column represents the compensation information given to any employee.
  + Assumptions:
  + Every employee joins the company with a compensation of 20,000.
  + Every year the revised and if the employee rating is less than 9.5 the compensation is changed to 10,000.
* **Last Pay rise date:** This column represents the date the last pay was revised. Every year on 1st January the pay is revised.
* **Variable Pay:** No information is given related to this. Variable Pay is an additional amount given to an employee for their performance.
* **Tenure in Org.:** This column represents the years an employee is present in the company.
  + Assumptions:
  + Years an employee is present in the company based on the effective date and end date.
  + If the employee continues with the company, then the last entry for that particular employee represents the year difference between the day of joining and present date.
* **Engagement Score**: This column represents the number of engagements of an employee.
* **Performance Rating:** This column represents the rating of employees.
  + Every year in the month of June rating is announced.
  + For new employee who joins before June of any year the row is represented as “NOT RATED”.
  + For employees whose rating is not present in the database then it is represented as ‘NA ‘. (In this case employee 1)
* **Effective Date:** This column represents the effective date of any task related to the employee.
  + Joining date
  + Engagement tasks 1 &2
  + Rating review date
  + Date form when new Compensation is given
  + Date of Exist
* **End Date:** This column represents the date up to which the last was affected. In case the employee continues with the company then the last end date for any employee will be 01-01-2100.

**SQL**

Objective: A data set is given which is a response to a recently run survey in a company on different themes or "Drivers".

Analysing the dataset to give meaningful information to HR leaders.

* Query 1 provides the average score group by drivers.
* Query 2 provides all the scores by department and driver combination.
* Query 3 provides all scores by location and driver combination
* Query 4 provides all scores by gender and driver
* Query 5 provides all scores by age group. I use the Case function to create an age group.
* Query 6 provides all scores by managers. I use CTE to filter the condition of maintaining anonymity.

**GoogleSheet**

The Google sheet contains the pivot table and heatmap for the data where red represents the low values, green represents high value and rest are value between low and high value.