1- Viewing raw data

SELECT \* FROM Employee;

SELECT \* FROM PerformanceRating;

2-Searching for missing VALUES

SELECT employeeid , COUNT(\*) AS missing\_count

FROM PerformanceRating

WHERE employeeid IS NULL;

3-Searching for duplicates in Employee table

SELECT \*,

COUNT(\*) OVER (PARTITION BY EmployeeID, FirstName, LastName, Age, Salary, Department, YearsAtCompany) AS DuplicateCount

FROM Employee\_cleaned

ORDER BY DuplicateCount DESC;

4-Searching for duplicates in PerformanceRating table

SELECT \*,

COUNT(\*) OVER (PARTITION BY employeeid , reviewdate , environmentsatisfaction , jobsatisfaction , relationshipsatisfaction , trainingopportunitieswithinyear , trainingopportunitiestaken , worklifebalance , selfrating , managerrating ) AS DuplicateCount

FROM PerformanceRating\_cleaned

ORDER BY DuplicateCount DESC;

5-Searching for outliers & invalid data

SELECT \*

FROM Employee

WHERE salary < 0 OR salary > 2000000

OR age < 18 OR age > 100

OR yearsatcompany > Age - 18;

SELECT \*

FROM PerformanceRating

WHERE EnvironmentSatisfaction < 0 OR EnvironmentSatisfaction > 5

OR JobSatisfaction < 0 OR JobSatisfaction > 5

OR WorkLifeBalance < 0 OR WorkLifeBalance > 5;

6- Data Integrity Check: Making sure all employees in PerformanceRating exist in Employee

SELECT pr.EmployeeID

FROM PerformanceRating pr

LEFT JOIN Employee e ON pr.EmployeeID = e.EmployeeID

WHERE e.EmployeeID IS NULL;