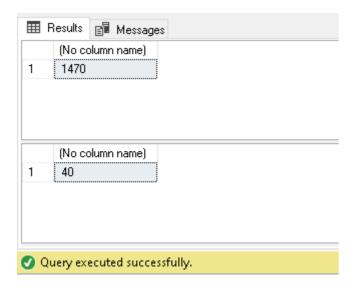# SQL Assignment
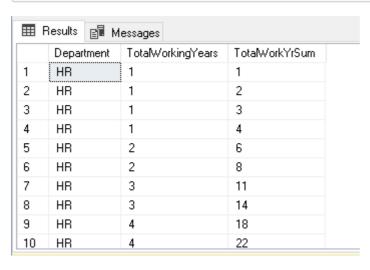
A. Return the shape of the table

```
select count(*) from emp;
select count(*) from INFORMATION_SCHEMA.COLUMNS where TABLE_NAME = 'emp';
```


B. Calculate the cumulative sum of total working years for each department

```
select
    Department, TotalWorkingYears, SUM(TotalWorkingYears)
    OVER(PARTITION BY Department ORDER BY TotalWorkingYears ROWS BETWEEN UNBOUNDED
PRECEDING
    AND CURRENT ROW) AS TotalWorkYrSum
FROM emp
    WHERE TotalWorkingYears > 0 ;
```


	Department	TotalWorkingYears	TotalWorkYrSum
1	HR	1	1
2	HR	1	2
3	HR	1	3
4	HR	1	4
5	HR	2	6
6	HR	2	8
7	HR	3	11
8	HR	3	14
9	HR	4	18
10	HR	4	22

C. Which gender have higher strength as workforce in each department.

```
SELECT Department, Gender, Gender_Count From(
SELECT
    Department, Gender, count(*) as Gender_Count,
    RANK() OVER(PARTITION BY Department
                ORDER BY COUNT(*) DESC) AS Gender_rank
    FROM emp
    GROUP BY Department, Gender) as _
    WHERE Gender_rank = 1;
```

	Department	Gender	Gender_Count
1	HR	Male	43
2	R&D	Male	582
3	Sales	Male	257

D. Create a new column AGE_BAND and Show Distribution of Employee's Age band group (Below 25, 25-34, 35-44, 45-55. ABOVE 55).
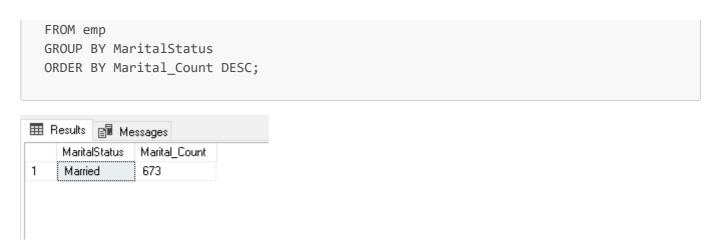
```
ALTER TABLE emp
ADD AGE_BAND INT;

UPDATE emp
SET AGE_BAND = (
    SELECT COUNT(*)
    FROM emp AS ed2
    WHERE ed2.CF_age_band = emp.CF_age_band)

select AGE_BAND from emp;
```

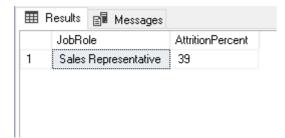|	AGE_BAND_COUNT
1	505
2	69
3	505
4	554
5	245
6	245

✓ Query executed successfully.

E. Compare all marital status of employee and find the most frequent marital status

```
SELECT
    TOP(1) MaritalStatus, Count(*) as Marital_Count
```

```
FROM emp
GROUP BY MaritalStatus
ORDER BY Marital_Count DESC;
```

	MaritalStatus	Marital_Count
1	Married	673

F. Show the Job Role with Highest Attrition Rate (Percentage)

```
SELECT TOP(1) JobRole,
TotalYes * 100/totalCount AS AttritionPercent
FROM (
    SELECT JobRole,
    COUNT(CASE
        WHEN Attrition = 'Yes' THEN 1
    END) AS TotalYes,
    COUNT(*) totalCount
    FROM emp
    GROUP BY JobRole
) _
ORDER BY AttritionPercent DESC;
```

	JobRole	AttritionPercent
1	Sales Representative	39

G. Show distribution of Employee's Promotion, Find the maximum chances of employee getting promoted.

```
SELECT *
FROM emp

SELECT
JobRole,PerformanceRating,YearsInCurrentRole,YearsAtCompany,YearsSinceLastPromotio
n,
    JobInvolvement,TrainingTimesLastYear
FROM emp
GROUP BY Attrition,JobRole,PerformanceRating,YearsSinceLastPromotion,
    YearsInCurrentRole,YearsAtCompany,JobInvolvement,TrainingTimesLastYear
ORDER BY YearsAtCompany
```
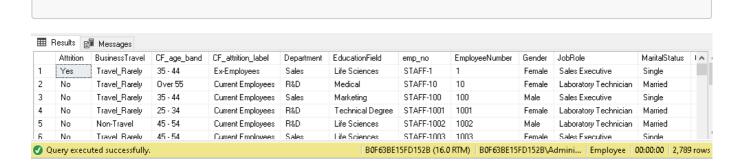
	Attrition	BusinessTravel	CF_age_band	CF_attrition_label	Department	EducationField	emp_no	EmployeeNumber	Gender	JobRole	MaritalStatus	
1	Yes	Travel_Rarely	35 - 44	Ex-Employees	Sales	Life Sciences	STAFF-1	1	Female	Sales Executive	Single	
2	No	Travel_Rarely	Over 55	Current Employees	R&D	Medical	STAFF-10	10	Female	Laboratory Technician	Married	
3	No	Travel_Rarely	35 - 44	Current Employees	Sales	Marketing	STAFF-100	100	Male	Sales Executive	Single	
4	No	Travel_Rarely	25 - 34	Current Employees	R&D	Technical Degree	STAFF-1001	1001	Female	Laboratory Technician	Married	
5	No	Non-Travel	45 - 54	Current Employees	R&D	Life Sciences	STAFF-1002	1002	Male	Laboratory Technician	Married	
6	No	Travel_Rarely	45 - 54	Current Employees	Sales	Life Sciences	STAFF-1003	1003	Female	Sales Executive	Single	

Query executed successfully.   B0F63BE15FD152B (16.0 RTM) | B0F63BE15FD152B\Admini... | Employee | 00:00:00 | 2,789 rows

I. Find the rank of employees within each department based on their monthly income

```
SELECT
    emp_no, Department, MonthlyIncome,
    DENSE_RANK() OVER(PARTITION BY Department ORDER BY MonthlyIncome DESC)
    AS Rank
FROM emp;
```

	emp_no	Department	MonthlyIncome	Rank	
1	STAFF-1338	HR	19717	1	
2	STAFF-1625	HR	19658	2	
3	STAFF-1973	HR	19636	3	
4	STAFF-734	HR	19189	4	
5	STAFF-731	HR	19141	5	
6	STAFF-140	HR	18844	6	

Query executed successfully.   B0F63BE15FD152B (16.0 RTM) | B0F63BE15FD152B\Admini... | Employee | 00:00:00 | 1,470 rows

J. Calculate the running total of 'Total Working Years' for each employee within each department and age band.

```
SELECT Department,CF_AGE_BAND,TotalWorkingYears,
SUM(TotalWorkingYears) OVER(PARTITION BY Department,CF_AGE_BAND
ORDER BY TotalWorkingYears ROWS BETWEEN UNBOUNDED PRECEDING
AND CURRENT ROW) AS TotalWorkYrSum
FROM emp
WHERE TotalWorkingYears > 0
```

	Department	CF_AGE_BAND	TotalWorkingYears	TotalWorkYrSum	
1	HR	25 - 34	1	1	
2	HR	25 - 34	1	2	
3	HR	25 - 34	2	4	
4	HR	25 - 34	2	6	
5	HR	25 - 34	3	9	
6	HR	25 - 34	4	13	

Query executed successfully.   B0F63BE15FD152B (16.0 RTM) | B0F63BE15FD152B\Admini... | Employee | 00:00:00 | 1,459 rows

K. For each employee who left, calculate the number of years they worked before leaving and compare it with the average years worked by employees in the same department.

```
SELECT emp_no, dept.Department, YearsAtCompany, avgyr_dept
FROM emp LEFT JOIN
(
    SELECT Department, AVG(YearsAtCompany) as avgyr_dept
    FROM emp
    GROUP BY Department
) as dept
ON dept.Department = emp.Department
```

	emp_no	Department	YearsAtCompany	avgyr_dept
1	STAFF-1	Sales	6	7
2	STAFF-10	R&D	1	6
3	STAFF-100	Sales	15	7
4	STAFF-1001	R&D	2	6
5	STAFF-1002	R&D	9	6
6	STAFF-1003	Sales	9	7

Query executed successfully.   B0F63BE15FD152B (16.0 RTM) | B0F63BE15FD152B\Admini... | Employee | 00:00:00 | 1,470 rows

L. Rank the departments by the average monthly income of employees who have left.

```
SELECT Department,AvgMonthlyIncome,
RANK() OVER(ORDER BY AvgMonthlyIncome DESC)
AS Income_Rank
FROM (
    SELECT Department,avg(MonthlyIncome) AvgMonthlyIncome
    FROM emp
    WHERE Attrition = 'Yes'
    GROUP BY Department
) AS _
```

	Department	AvgMonthlyIncome	Income_Rank
1	Sales	5908	1
2	R&D	4108	2
3	HR	3715	3

Query executed successfully.   B0F63BE15FD152B (16.0 RTM) | B0F63BE15FD152B\Admini... | Employee | 00:00:00 | 3 rows

M. Find the if there is any relation between Attrition Rate and Marital Status of Employee.

```
SELECT
    Attrition,MaritalStatus, COUNT(*) as MaritalCount
FROM emp
GROUP BY
    MaritalStatus, Attrition
ORDER BY
    MaritalCount DESC;
```

N. Show the Department with Highest Attrition Rate (Percentage)

```
SELECT TOP(1) Department,
(COUNT(CASE
    WHEN Attrition = 'Yes' THEN 1
END) * 100 ) / COUNT(*) AS DYesPercent
FROM emp
GROUP BY Department
ORDER BY DYesPercent DESC
```


O. Calculate the moving average of monthly income over the past 3 employees for each job role.

```
SELECT emp_no,MonthlyIncome,
AVG(MonthlyIncome)
OVER(ORDER BY MonthlyIncome ROWS BETWEEN 2 PRECEDING AND CURRENT ROW)
AS MovingAvgIncome
FROM emp
```


P. Identify employees with outliers in monthly income within each job role. [Condition :Monthly_Income < Q1 - (Q3 - Q1) * 1.5 OR Monthly_Income > Q3 + (Q3 - Q1)]

```
WITH EmpCTE AS (
    SELECT JobRole, MonthlyIncome,
        PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY MonthlyIncome) OVER() AS
```

```
Q1,
        PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY MonthlyIncome) OVER() AS
Q2,
        PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY MonthlyIncome) OVER() AS
Q3
    FROM emp
)
SELECT JobRole, MonthlyIncome
FROM EmpCTE
WHERE MonthlyIncome < Q1 - (Q3 - Q1) * 1.5 OR MonthlyIncome > (Q3 + (Q3 - Q1));
```

	JobRole	MonthlyIncome
1	Sales Executive	13872
2	Healthcare Representative	13964
3	Healthcare Representative	13966
4	Manufacturing Director	13973
5	Manager	14026
6	Manager	14118

Query executed successfully. B0F63BE15FD152B (16.0 RTM) B0F63BE15FD152B\Admini... Employee 00:00:00 146 rows

Q. Gender distribution within each job role, show each job role with its gender domination.
[Male_Domination or Female_Domination]

```
SELECT Gender,JobRole
FROM (
    SELECT Gender,JobRole,
    RANK() OVER(PARTITION BY JobRole ORDER BY COUNT(*) DESC)
    AS Rank
    FROM emp
    GROUP BY Gender,JobRole
) AS _
WHERE Rank = 1
```

	Gender	JobRole
1	Male	Healthcare Representative
2	Male	Human Resources
3	Male	Laboratory Technician
4	Male	Manager
5	Male	Manufacturing Director
6	Male	Research Director

Query executed successfully. Assignment1_empDB.sql - B0F63BE15FD152B.Employee TM) B0F63BE15FD152B\Admini... Employee 00:00:00 9 rows

R. Percent rank of employees based on training times last year

```
SELECT emp_no,TrainingTimesLastYear,
PERCENT_RANK() OVER(ORDER BY TrainingTimesLastYear)
AS TrainingPercentage
FROM emp
```

Divide employees into 5 groups based on training times last year [Use NTILE ()]

```
SELECT emp_no,TrainingTimesLastYear,
NTILE(5) OVER(ORDER BY TrainingTimesLastYear)
AS Batch
FROM emp
```


T. Categorize employees based on training times last year as - Frequent Trainee, Moderate Trainee, Infrequent Trainee

```
WITH empCTE AS (
    SELECT emp_no, TrainingTimesLastYear,
          CASE
                WHEN TrainingTimesLastYear > 4 THEN 'Frequent Trainee'
                WHEN TrainingTimesLastYear > 2 THEN 'Moderate Trainee'
                ELSE 'Infrequent Trainee'
          END AS 'Training Frequency'
    FROM emp
)
SELECT emp_no, TrainingTimesLastYear, 'Training Frequency'
FROM empCTE
ORDER BY TrainingTimesLastYear DESC;
```


U. Categorize employees as 'High', 'Medium', or 'Low' performers based on their performance rating, using a CASE WHEN statement.

```
SELECT emp_no,PerformanceRating,
CASE
    WHEN PerformanceRating > 3 THEN 'High Performer'
    WHEN PerformanceRating > 1 THEN 'Medium Performer'
    ELSE 'Low Performer'
END AS 'Performance'
FROM emp
ORDER BY PerformanceRating DESC
```

	emp_no	PerformanceRating	Performance
1	STAFF-1010	4	High Performer
2	STAFF-1035	4	High Performer
3	STAFF-1056	4	High Performer
4	STAFF-10	4	High Performer
5	STAFF-103	4	High Performer
6	STAFF-1080	4	High Performer

Query executed successfully.   Assignment1 empDB.sql - B0F63BE15FD152B.Employee   B0F63BE15FD152B\Admini...   Employee   00:00:00   1,470 rows

V. Use a CASE WHEN statement to categorize employees into 'Poor', 'Fair', 'Good', or 'Excellent' ,work-life balance based on their work-life balance score.

```
WITH empCTE AS (
    SELECT emp_no, WorkLifeBalance,
            CASE
                WHEN WorkLifeBalance > 3 THEN 'Excellent WorkLifeBalance'
                WHEN WorkLifeBalance > 1 THEN 'Fair WorkLifeBalance'
                ELSE 'Poor WorkLifeBalance'
            END AS WorkLifeBalance_Ranking
    FROM emp
)
SELECT emp_no, WorkLifeBalance, WorkLifeBalance_Ranking
FROM empCTE
ORDER BY WorkLifeBalance DESC;
```

	emp_no	WorkLifeBalance	WorkLifeBalance_Ranking
1	STAFF-102	4	Excellent WorkLifeBalance
2	STAFF-1029	4	Excellent WorkLifeBalance
3	STAFF-1045	4	Excellent WorkLifeBalance
4	STAFF-1074	4	Excellent WorkLifeBalance
5	STAFF-101	4	Excellent WorkLifeBalance
6	STAFF-1081	4	Excellent WorkLifeBalance

Query executed successfully.   B0F63BE15FD152B (16.0 RTM)   B0F63BE15FD152B\Admini...   Employee   00:00:00   1,470 rows

W. Group employees into 3 groups based on their stock option level using the [NTILE] function.

```
SELECT StockOptionLevel,
NTILE(3) OVER(ORDER BY StockOptionLevel DESC)
AS 'Stock RANK'
FROM emp
```

X. Find key reasons for Attrition in Company

```
SELECT JobRole,Department,
AVG(YearsAtCompany) Company_year_Count,
AVG(YearsSinceLastPromotion) year_since_promotion_Count,
AVG(WorkLifeBalance) worklife_avg,
AVG(PercentSalaryHike) hike_percent,
AVG(MonthlyIncome) income_avg,
AVG(EnvironmentSatisfaction) env_satisfaction,
COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END) Attrition_rate,
COUNT(CASE WHEN Attrition = 'Yes' THEN 1 END)*100/ COUNT(*) Attrition_percent

FROM emp
GROUP BY JobRole,Department
ORDER BY Attrition_percent DESC
```


Insight :

Based on the analysis, the key reasons for the high Attrition rates in the company seem to be a combination of factors, including the specific job roles and departments (particularly Sales and R&D), low average tenures and limited career progression opportunities, suboptimal work-life balance and work environment satisfaction, as well as relatively low compensation levels in terms of salary hikes and monthly income.