

Employee Database Analysis

a. Return the shape of the table

```
SELECT
  (SELECT COUNT(*) FROM EmployeeData) AS row_bo,
  (SELECT COUNT(*) FROM INFORMATION_SCHEMA.COLUMNS WHERE TABLE_NAME =
    'EmployeeData') AS no_columns;
```

	row_bo	no_columns
1	1470	40

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
  Cumulative_sum_year
FROM EmployeeData;
```

	Department	TotalWorkingYears	Cumulative_sum_year
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 Has Higher Strength as Workforce in Each Department

```
WITH GenderCounts AS (
  SELECT
    Department,
    Gender,
    COUNT(*) as counts,
    SUM(CASE WHEN Gender = 'Male' THEN 1 ELSE 0 END) AS Males,
    SUM(CASE WHEN Gender = 'Female' THEN 1 ELSE 0 END) AS Females
  FROM
```

```
EmployeeData
GROUP BY
    Department,
    Gender
)
SELECT
    Department,
    Gender AS ProminentGender,
    counts,
    RANK() OVER (PARTITION BY Department ORDER BY counts DESC) AS Gender_Rank
FROM GenderCounts;
```

	Department	ProminentGender	counts	Gender_Rank
1	HR	Male	43	1
2	HR	Female	20	2
3	R&D	Male	582	1
4	R&D	Female	379	2
5	Sales	Male	257	1
6	Sales	Female	189	2

d. Create a New Column **AGE_BAND** and Show Distribution of Employee's Age Band Group

```
ALTER TABLE EmployeeData
ADD Age_Band NVARCHAR(50);
```

```
UPDATE EmployeeData
SET Age_Band = CASE
    WHEN Age < 25 THEN 'Below 25'
    WHEN Age BETWEEN 25 AND 34 THEN '25-34'
    WHEN Age BETWEEN 35 AND 44 THEN '35-44'
    WHEN Age BETWEEN 45 AND 55 THEN '45-55'
    ELSE 'Above 55'
END;

SELECT
    Age_Band,
    COUNT(*) AS Count
FROM EmployeeData
GROUP BY Age_Band;
```

	Age_Band	Count
1	45-55	267
2	35-44	505
3	Above 55	47
4	Below 25	97
5	25-34	554

e. Compare All Marital Status of Employees and Find the Most Frequent Marital Status

```

SELECT
    MaritalStatus,
    Count,
    RANK() OVER (ORDER BY Count DESC) AS Freq_Rank
FROM (
    SELECT
        MaritalStatus,
        COUNT(*) AS Count
    FROM EmployeeData
    GROUP BY MaritalStatus
) AS _

```

	MaritalStatus	Count	Freq_Rank
1	Married	673	1
2	Single	470	2
3	Divorced	327	3

- Most frequent marital status is "Married"

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

```

WITH AttritionRate AS (
    SELECT
        JobRole,
        (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
Attrition_Percentage
    FROM EmployeeData
    GROUP BY JobRole
)
SELECT TOP 1
    JobRole,
    Attrition_Percentage
FROM AttritionRate
ORDER BY Attrition_Percentage DESC;

```

	JobRole	Attrition_Percentage
1	Sales Representative	39.759036144578

- Sales Representetatives have the highest attrition rate

g. Show Distribution of Employee's Promotion, Find the Maximum Chances of Employee Getting Promoted

```
SELECT
    Department,
    YearsSinceLastPromotion,
    COUNT(*) AS EmployeeCount
FROM EmployeeData
WHERE YearsSinceLastPromotion = 0
GROUP BY Department, YearsSinceLastPromotion
ORDER BY Department, YearsSinceLastPromotion;
```

	Department	YearsSinceLastPromotion	EmployeeCount
1	HR	0	24
2	R&D	0	379
3	Sales	0	178

Insight: From above data we can see that most employees from R&D Department got a most recent promotion.

```
SELECT
    Department,
    JobRole,
    YearsSinceLastPromotion,
    COUNT(*) AS EmployeeCount
FROM EmployeeData
GROUP BY Department ,JobRole, YearsSinceLastPromotion
ORDER BY EmployeeCount DESC;
```

	Department	JobRole	YearsSinceLastPromotion	EmployeeCount
1	R&D	Laboratory Technician	0	126
2	Sales	Sales Executive	0	122
3	R&D	Research Scientist	0	121
4	R&D	Research Scientist	1	82
5	Sales	Sales Executive	1	75
6	R&D	Laboratory Technician	1	64
7	R&D	Manufacturing Director	0	56
8	R&D	Research Scientist	2	45
9	Sales	Sales Representative	0	44
10	R&D	Manufacturing Director	1	42

From further analyzing the data, we can observe that large number of Laboratory Technicians from R&D, Sales Executives from Sales and Research Scientists from R&D Department got recent promotions.

```
SELECT
    JobInvolvement,
    PerformanceRating,
    YearsSinceLastPromotion,
    COUNT(*) AS EmployeeCount
FROM EmployeeData
GROUP BY
    PerformanceRating,
    JobInvolvement,
    YearsSinceLastPromotion
ORDER BY YearsSinceLastPromotion, EmployeeCount DESC;
```

	JobInvolvement	PerformanceRating	YearsSinceLastPromotion	EmployeeCount
1	3	3	0	281
2	2	3	0	132
3	4	3	0	52
4	3	4	0	50
5	2	4	0	24
6	1	3	0	20
7	1	4	0	11
8	4	4	0	11
9	3	3	1	184
10	2	3	1	69

From Employees who got recent promotions, we can observe their Jobinvolment and Perfomrance Rating is 3.

h. Show the Cumulative Sum of Total Working Years for Each Department

- duplicate of b

```
SELECT
    Department,
    TotalWorkingYears,
    SUM(TotalWorkingYears) OVER (PARTITION BY Department ORDER BY
    TotalWorkingYears ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS
    Cumulative_sum_year
FROM EmployeeData;
```

	Department	TotalWorkingYears	Cumulative_sum_year
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

i. Find the Rank of Employees Within Each Department Based on Their Monthly Income

```
SELECT
    EmployeeNumber,
    Department,
    MonthlyIncome,
    RANK() OVER (PARTITION BY Department ORDER BY MonthlyIncome DESC) AS
Income_Rank
FROM EmployeeData;
```

	EmployeeNumber	Department	MonthlyIncome	Income_Rank
1	1338	HR	19717	1
2	1625	HR	19658	2
3	1973	HR	19636	3
4	734	HR	19189	4
5	731	HR	19141	5
6	140	HR	18844	6
7	644	HR	18200	7
8	148	HR	17328	8
9	1408	HR	16799	9
10	1550	HR	16437	10

Employee 1338 have highest income, followed by 1625

j. Calculate the Running Total of 'Total Working Years' for Each Employee Within Each Department and Age Band

```
SELECT
    EmployeeNumber,
    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 Running_Total_Working_Years
FROM EmployeeData;
```

	EmployeeNumber	Department	CF_age_band	TotalWorkingYears	Running_Total_Working_Years
1	1865	HR	25 - 34	1	1
2	1944	HR	25 - 34	1	2
3	1467	HR	25 - 34	2	4
4	1842	HR	25 - 34	2	6
5	1858	HR	25 - 34	3	9
6	847	HR	25 - 34	4	13
7	1844	HR	25 - 34	4	17
8	1207	HR	25 - 34	5	22
9	1987	HR	25 - 34	6	28
10	869	HR	25 - 34	6	34

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

```

WITH YearsWorked AS (
    SELECT
        EmployeeNumber,
        Department,
        TotalWorkingYears AS Years_Worked_Before_Leaving
    FROM EmployeeData
    WHERE Attrition = 'Yes'
),
AvgYears AS (
    SELECT
        Department,
        AVG(Years_Worked_Before_Leaving) AS Avg_Years_Worked
    FROM YearsWorked
    GROUP BY Department
)
SELECT
    Y.EmployeeNumber,
    Y.Department,
    Y.Years_Worked_Before_Leaving,
    A.Avg_Years_Worked,
    A.Avg_Years_Worked - Y.Years_Worked_Before_Leaving AS Diff
FROM YearsWorked Y, AvgYears A
WHERE Y.Department = A.Department
ORDER BY Department, Diff;

```

	EmployeeNumber	Department	Years_Worked_Before_Leaving	Avg_Years_Worked	Diff
1	1098	HR	24	6	-18
2	590	HR	11	6	-5
3	608	HR	8	6	-2
4	133	HR	7	6	-1
5	1747	HR	6	6	0
6	1818	HR	6	6	0
7	1844	HR	4	6	2
8	1467	HR	2	6	4
9	1842	HR	2	6	4
10	1714	HR	1	6	5

l. Rank the Departments by the Average Monthly Income of Employees Who Have Left

```
WITH DepartmentIncome AS (
    SELECT
        Department,
        AVG(MonthlyIncome) AS Avg_Monthly_Income
    FROM EmployeeData
    WHERE Attrition = 'Yes'
    GROUP BY Department
)
SELECT
    Department,
    Avg_Monthly_Income,
    RANK() OVER (ORDER BY Avg_Monthly_Income DESC) AS Department_Rank
FROM DepartmentIncome;
```

	Department	Avg_Monthly_Income	Department_Rank
1	Sales	5908	1
2	R&D	4108	2
3	HR	3715	3

- Sales have the employees who left with highest average monthly income followed by R&D.
- In HR Department, average income of employees who left are comparatively low.

m. Find If There Is Any Relation Between Attrition Rate and Marital Status of Employee

```
SELECT
    MaritalStatus,
    (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
Attrition_Percentage
FROM EmployeeData
GROUP BY MaritalStatus;
```


	MaritalStatus	Attrition_Percentage
1	Single	25.531914893617
2	Divorced	10.091743119266
3	Married	12.481426448736

- Insight:

Attrition rate is higher in Singles.

Attrition rate is lower in Married and Divorced.

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

```
SELECT TOP 1
    Department,
    (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
Attrition_Percentage
FROM EmployeeData
GROUP BY Department
ORDER BY Attrition_Percentage DESC;
```

Results		Messages
	Department	Attrition_Percentage
1	Sales	20.627802690582

Sales department have highest attrition rate.

o. Calculate the Moving Average of Monthly Income Over the Past 3 Employees for Each Job Role

```
WITH LaggedData AS (
    SELECT
        JobRole,
        EmployeeNumber,
        MonthlyIncome,
        LAG(MonthlyIncome, 1) OVER (PARTITION BY JobRole ORDER BY EmployeeNumber)
AS PrevIncome1,
        LAG(MonthlyIncome, 2) OVER (PARTITION BY JobRole ORDER BY EmployeeNumber)
AS PrevIncome2
    FROM
        EmployeeData
)
SELECT
    JobRole,
    EmployeeNumber,
    PrevIncome1,
    PrevIncome2,
    MonthlyIncome,
```

```

    (COALESCE(MonthlyIncome, 0) + COALESCE(PrevIncome1, 0) + COALESCE(PrevIncome2,
0)) /
    (CASE
        WHEN PrevIncome2 IS NOT NULL THEN 3
        WHEN PrevIncome1 IS NOT NULL THEN 2
        ELSE 1
    END) AS MovingAverageMonthlyIncome
FROM
    LaggedData;

```

	JobRole	EmployeeNumber	PrevIncome1	PrevIncome2	MonthlyIncome	MovingAverageMonthlyIncome
1	Healthcare Representative	13	NULL	NULL	5237	5237
2	Healthcare Representative	36	5237	NULL	10248	7742
3	Healthcare Representative	40	10248	5237	6465	7316
4	Healthcare Representative	70	6465	10248	9884	8865
5	Healthcare Representative	83	9884	6465	10096	8815
6	Healthcare Representative	117	10096	9884	4152	8044
7	Healthcare Representative	119	4152	10096	13503	9250
8	Healthcare Representative	124	13503	4152	10673	9442
9	Healthcare Representative	139	10673	13503	5163	9779
10	Healthcare Representative	145	5163	10673	7484	7773

p. Identify Employees with Outliers in Monthly Income Within Each Job Role

```

WITH IncomeStats AS (
    SELECT
        JobRole,
        EmployeeNumber,
        MonthlyIncome,
        PERCENTILE_CONT(0.25) WITHIN GROUP (ORDER BY MonthlyIncome) OVER
(PARTITION BY JobRole) AS Q1,
        PERCENTILE_CONT(0.75) WITHIN GROUP (ORDER BY MonthlyIncome) OVER
(PARTITION BY JobRole) AS Q3
    FROM EmployeeData
)
SELECT
    EmployeeNumber,
    JobRole,
    MonthlyIncome,
    CASE
        WHEN MonthlyIncome < Q1 - (Q3 - Q1) * 1.5 THEN 'Low'
        WHEN MonthlyIncome > Q3 + (Q3 - Q1) * 1.5 THEN 'High'
        ELSE 'Not an Outlier'
    END AS OutlierType
FROM IncomeStats
WHERE MonthlyIncome < Q1 - (Q3 - Q1) * 1.5
    OR MonthlyIncome > Q3 + (Q3 - Q1) * 1.5;

```

	EmployeeNumber	JobRole	MonthlyIncome	OutlierType
1	1985	Laboratory Technician	6323	High
2	1737	Laboratory Technician	6472	High
3	1315	Laboratory Technician	6674	High
4	1132	Laboratory Technician	6782	High
5	944	Laboratory Technician	7403	High
6	1516	Manager	11244	Low
7	613	Manager	11557	Low
8	153	Manager	11631	Low
9	376	Manager	11849	Low
10	428	Manager	11878	Low

q. Gender Distribution Within Each Job Role, Show Each Job Role with Its Gender Domination

```
WITH GenderCount AS
(
SELECT
    JobRole,
    SUM(CASE WHEN Gender = 'Male' THEN 1 ELSE 0 END) AS Male_Count,
    SUM(CASE WHEN Gender = 'Female' THEN 1 ELSE 0 END) AS Female_Count
FROM EmployeeData
GROUP BY JobRole
)
SELECT
    *,
    CASE
        WHEN Male_count > Female_Count THEN 'Male'
        WHEN Female_Count > Male_count THEN 'Female'
        ELSE 'Equal'
    END AS Dominant_Gender
FROM
    GenderCount;
```

	JobRole	Male_Count	Female_Count	Dominant_Gender
1	Sales Representative	45	38	Male_Domination
2	Manager	55	47	Male_Domination
3	Healthcare Representative	80	51	Male_Domination
4	Laboratory Technician	174	85	Male_Domination
5	Sales Executive	194	132	Male_Domination
6	Manufacturing Director	73	72	Male_Domination
7	Human Resources	36	16	Male_Domination
8	Research Director	47	33	Male_Domination
9	Research Scientist	178	114	Male_Domination

Insight every department in male dominated.

r. Percent Rank of Employees Based on Training Times Last Year

```
SELECT
    EmployeeNumber,
    TrainingTimesLastYear,
    PERCENT_RANK() OVER (ORDER BY TrainingTimesLastYear) * 100 AS PercentRank
FROM EmployeeData;
```

	EmployeeNumber	TrainingTimesLastYear	PercentRank
1	1	0	0
2	56	0	0
3	58	0	0
4	90	0	0
5	125	0	0
6	178	0	0
7	246	0	0
8	275	0	0
9	353	0	0

s. Divide Employees into 5 Groups Based on Training Times Last Year

```
SELECT
    EmployeeNumber,
    TrainingTimesLastYear,
    NTILE(5) OVER (ORDER BY TrainingTimesLastYear) AS Training_Group
FROM EmployeeData;
```

	EmployeeNumber	TrainingTimesLastYear	Training_Group
1	1	0	1
2	56	0	1
3	58	0	1
4	90	0	1
5	125	0	1
6	178	0	1
7	246	0	1
8	275	0	1
9	353	0	1
10	361	0	1

t. Categorize Employees Based on Training Times Last Year as - Frequent Trainee, Moderate Trainee, Infrequent Trainee

```
SELECT
    EmployeeNumber,
    TrainingTimesLastYear,
    CASE
        WHEN TrainingTimesLastYear <= 2 THEN 'Infrequent Trainee'
        WHEN TrainingTimesLastYear <= 4 THEN 'Moderate Trainee'
        ELSE 'Frequent Trainee'
    END
```

```
END AS TraineeCategory
FROM EmployeeData
ORDER BY TrainingTimesLastYear DESC;
```

	EmployeeNumber	TrainingTimesLastYear	TraineeCategory
1	30	6	Frequent Trainee
2	42	6	Frequent Trainee
3	54	6	Frequent Trainee
4	153	6	Frequent Trainee
5	154	6	Frequent Trainee
6	217	6	Frequent Trainee
7	258	6	Frequent Trainee
8	307	6	Frequent Trainee
9	309	6	Frequent Trainee
10	381	6	Frequent Trainee

u. Categorize Employees as 'High', 'Medium', or 'Low' Performers Based on Their Performance Rating

```
SELECT
    EmployeeNumber,
    PerformanceRating,
    CASE
        WHEN PerformanceRating >= 4 THEN 'High'
        WHEN PerformanceRating = 3 THEN 'Medium'
        ELSE 'Low'
    END AS PerformanceCategory
FROM EmployeeData
ORDER BY PerformanceRating DESC;
```

	EmployeeNumber	PerformanceRating	PerformanceCategory
1	10	4	High
2	11	4	High
3	12	4	High
4	2	4	High
5	27	4	High
6	33	4	High
7	60	4	High
8	61	4	High
9	62	4	High
10	68	4	High

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

```
SELECT
    EmployeeNumber,
    WorkLifeBalance,
```

```
CASE
    WHEN WorkLifeBalance = 1 THEN 'Poor'
    WHEN WorkLifeBalance = 2 THEN 'Fair'
    WHEN WorkLifeBalance = 3 THEN 'Good'
    ELSE 'Excellent'
END AS Work_Life_Balance_Category
FROM EmployeeData
ORDER BY WorkLifeBalance DESC;
```

	EmployeeNumber	WorkLifeBalance	Work_Life_Balance_Category
1	40	4	Excellent
2	42	4	Excellent
3	51	4	Excellent
4	77	4	Excellent
5	101	4	Excellent
6	102	4	Excellent
7	132	4	Excellent
8	129	4	Excellent
9	139	4	Excellent
10	160	4	Excellent

w. Group Employees into 3 Groups Based on Their Stock Option Level Using the [NTILE] Function

```
SELECT
    EmployeeNumber,
    StockOptionLevel,
    NTILE(3) OVER (ORDER BY StockOptionLevel) AS StockOption_Group
FROM EmployeeData;
```

	EmployeeNumber	StockOptionLevel	StockOption_Group
1	1	0	1
2	4	0	1
3	5	0	1
4	8	0	1
5	12	0	1
6	15	0	1
7	19	0	1
8	27	0	1
9	28	0	1
10	30	0	1

x. Find Key Reasons for Attrition in Company

```
SELECT
    Attrition,
    BusinessTravel,
    Department,
```

```

    MaritalStatus,
    JobRole,
    AVG(MonthlyIncome) AS AvgIncome,
    COUNT(*) AS Count
FROM EmployeeData
WHERE Attrition = 'Yes'
GROUP BY Attrition, BusinessTravel, Department, MaritalStatus, JobRole
ORDER BY Count DESC;

```

	Attrition	BusinessTravel	Department	MaritalStatus	JobRole	AvgIncome	Count
1	Yes	Travel_Rarely	Sales	Single	Sales Executive	7518	21
2	Yes	Travel_Rarely	R&D	Single	Laboratory Technician	2679	20
3	Yes	Travel_Rarely	R&D	Single	Research Scientist	2535	19
4	Yes	Travel_Rarely	R&D	Married	Laboratory Technician	2694	15
5	Yes	Travel_Rarely	Sales	Married	Sales Executive	7986	14
6	Yes	Travel_Rarely	Sales	Single	Sales Representative	2516	11
7	Yes	Travel_Frequently	Sales	Single	Sales Representative	1892	10
8	Yes	Travel_Frequently	R&D	Single	Laboratory Technician	3493	9
9	Yes	Travel_Rarely	R&D	Divorced	Laboratory Technician	3032	9
10	Yes	Travel_Rarely	R&D	Married	Research Scientist	3441	9

- insights:

From the first table we can observe that the most of the employees left are not married and are from Sales or R&D Department.

```

WITH WorkingYears AS
(
SELECT
    *,
    CASE
        WHEN YearsAtCompany <= 5 THEN '0-5'
        WHEN YearsAtCompany <= 10 THEN '5-10'
        WHEN YearsAtCompany <= 15 THEN '10-15'
        WHEN YearsAtCompany <= 20 THEN '10-20'
        ELSE '0-5'
    END AS WorkingYearBand
FROM
    EmployeeData
)
SELECT
    Attrition,
    WorkingYearBand,
    COUNT(*) AS Total_Left_Employees,
    AVG(Age) AS Avg_Age,
    AVG(MonthlyIncome) AS Avg_MonthlyIncome,
    AVG(WorkLifeBalance) AS Avg_WorkLifeBalance,
    AVG(JobSatisfaction) AS Avg_JobSatisfaction,
    AVG(EnvironmentSatisfaction) AS Avg_EnvironmentSatisfaction,
    AVG(JobInvolvement) AS Avg_JobInvolvement,
    AVG(PerformanceRating) AS Avg_PerformanceRating

```

```
FROM WorkingYears
WHERE Attrition = 'Yes'
GROUP BY Attrition, WorkingYearBand
ORDER BY Total_Left_Employees, WorkingYearBand DESC;
```

	Attrition	WorkingYearBand	Total_Left_Employees	Avg_Age	Avg_MonthlyIncome
1	Yes	10-20	5	42	9882
2	Yes	10-15	7	33	7477
3	Yes	5-10	55	36	5608
4	Yes	0-5	170	32	4260

- Insight:

It is clear that most of the employees who left the company are worked between 0 to 10 year in the company and they also have a lower average monthly income.