Assignment 3 - Employee Attrition Analysis

Problem Statement: You have been tasked with analysing employee attrition, employee performance, salary patterns and job roles within the company. You need to determine the following

Dataset Dictionary:

Column Name	Description	Data Type
Attrition	Whether the employee has left the company ('Yes' or 'No')	String
Business Travel	Frequency of travel for business purposes ('Travel_Rarely', 'Travel_Frequently')	String
CF_age band	Age range of the employee	String
CF_attrition label	Classification of employees ('Ex-Employees', 'Current Employees')	String
Department	Department where the employee works (e.g., 'Sales', 'R&D')	String
Education Field	Field of education (e.g., 'Life Sciences', 'Medical')	String
emp no	Employee identifier	String
Employee Number	Unique number assigned to each employee	Integer
Gender	Gender of the employee ('Male', 'Female')	String
Job Role	Job role of the employee (e.g., 'Sales Executive', 'Research Scientist')	String
Marital Status	Marital status of the employee ('Single', 'Married', 'Divorced')	String
Over Time	Whether the employee works overtime ('Yes' or 'No')	String
Over18	Whether the employee is over 18 years old ('Yes')	String
Training Times Last Year	Number of training sessions attended last year	Integer
Age	Age of the employee	Integer
CF_current Employee	Classification indicating if the employee is a current employee ('Yes', 'No')	String
Daily Rate	Daily pay rate of the employee	Integer
Distance From Home	Distance from home to work	Integer
Education	Level of education (1 to 5)	Integer
Employee Count	Count of employees in the dataset (typically '1' for each row)	Integer
Environment Satisfaction	Satisfaction with the work environment (1 to 4)	Integer
Hourly Rate	Hourly pay rate of the employee	Integer
Job Involvement	Level of job involvement (1 to 4)	Integer
Job Level	Level of the job (1 to 5)	Integer
Job Satisfaction	Satisfaction with the job (1 to 4)	Integer
Monthly Income	Monthly pay of the employee	Integer
Monthly Rate	Monthly rate of pay	Integer
Num Companies Worked	Number of companies the employee has worked for	Integer

Percent Salary Hike	Percentage increase in salary	Integer
Performance Rating	Performance rating (1 to 4)	Integer
Relationship Satisfaction	Satisfaction with relationships at work (1 to 4)	Integer
Standard Hours	Standard working hours (typically '80')	Integer
Stock Option Level	Level of stock options (0 to 3)	Integer
Total Working Years	Total number of years the employee has worked	Integer
Work Life Balance	Work-life balance rating (1 to 4)	Integer
Years At Company	Number of years the employee has been at the company	Integer
Years In Current Role	Number of years in the current role	Integer
Years Since Last Promotion	Number of years since the last promotion	Integer
Years With Curr Manager	Number of years working with the current manager	Integer

Create database HREmployeeDB.

a) Return the shape of the table
b) Calculate the cumulative sum of total working years for each department
c) Which gender have higher strength as workforce in each department
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).
e) Compare all marital status of employee and find the most frequent marital status
f) Show the Job Role with Highest Attrition Rate (Percentage)
g) Show distribution of Employee's Promotion, Find the maximum chances of employee getting promoted.
h) Show the cumulative sum of total working years for each department.
i) Find the rank of employees within each department based on their monthly income
j) Calculate the running total of 'Total Working Years' for each employee within each department and age band.
k) Foreach 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.
l) Rank the departments by the average monthly income of employees who have left.
m) Find the if there is any relation between Attrition Rate and Marital Status of Employee.
n) Show the Department with Highest Attrition Rate (Percentage)
o) Calculate the moving average of monthly income over the past 3 employees for each job role.
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)]
q) Gender distribution within each job role, show each job role with its gender domination. [Male_Domination or Female_Domination]
r) Percent rank of employees based on training times last year
s) Divide employees into 5 groups based on training times last year [Use NTILE ()]
t) Categorize employees based on training times last year as - Frequent Trainee, Moderate Trainee, Infrequent Trainee.

u) Categorize employees as 'High', 'Medium', or 'Low' performers based on their performance rating, using a CASE WHEN statement.
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.
w) Group employees into 3 groups based on their stock option level using the [NTILE] function.
x) **Find key reasons for Attrition in Company**