# IBM HR Analytics Employee Attrition & Performance

## Tools used in building the Project:

1. Excel
2. SQL Server
3. Power BI

## Objective of the Project:

1. **Understand Current Turnover Rates**: Gain a comprehensive understanding of the current employee turnover rate and analyse the demographic distribution of attrition by age, gender, education, department and job role.
2. **Identify Key Factors Influencing Turnover**: Examine the main factors contributing to employee attrition, including job satisfaction indicators (job involvement and work-life balance, salary factors (monthly income and salary hikes), and benefit factors (stock option levels), to uncover patterns and correlations that drive higher attrition rates.

## Dataset Overview:

The dataset includes the following columns:

- Age: The employee's age in years.
- Attrition: Indicates whether the employee has left the company ("Yes") or is still employed ("No").
- Business Travel: The frequency of business travel for the employee, typically categorized as "Non-Travel," "Travel Rarely," or "Travel Frequently."
- Daily Rate: The daily salary rate for the employee.
- Department: The department where the employee works, such as "Sales," "Research & Development," or "Human Resources."
- Distance From Home: The distance in miles from the employee's home to the workplace.
- Education: The education level of the employee, usually on a scale from 1 to 5, with 1 being the least education (e.g., high school) and 5 being the highest (e.g., doctorate).
- Education Field: The field of education in which the employee studied, such as "Life Sciences," "Medical," "Marketing," etc.
- Employee Count: The count of employees in the dataset. Usually a constant (e.g., 1) for each record.
- Employee Number: A unique identifier for each employee in the dataset.

- Environment Satisfaction: Employee satisfaction with the work environment, typically rated on a scale of 1 to 4, with 1 being the least satisfied and 4 the most satisfied.
- Gender: The gender of the employee, usually "Male" or "Female."
- Hourly Rate: The hourly wage of the employee.
- Job Involvement: A rating of the employee's involvement in their job, usually on a scale from 1 to 4, with 1 being least involved and 4 most involved.
- Job Level: The level or rank of the employee's position within the company, usually on a scale (e.g., from 1 to 5) representing progression in the hierarchy.
- Job Role: The specific role of the employee, such as "Sales Executive," "Research Scientist," or "Laboratory Technician."
- Job Satisfaction: A rating of the employee's satisfaction with their job, usually on a scale of 1 to 4, with 1 being least satisfied and 4 most satisfied.
- Marital Status: The marital status of the employee, e.g., "Single," "Married," or "Divorced."
- Monthly Income: The monthly salary of the employee.
- Monthly Rate: The monthly rate for the employee.
- Num Companies Worked: The number of companies the employee has worked for prior to the current one.
- Over18: Indicates if the employee is over 18 years old, often marked as "Y" for all entries.
- Over Time: Whether the employee works overtime ("Yes") or not ("No").
- Percent Salary Hike: The percentage increase in the employee's salary.
- Performance Rating: A performance rating of the employee, typically on a scale (e.g., 1 to 4), where 4 indicates the highest performance.
- Relationship Satisfaction: A rating of the employee's satisfaction with personal relationships, typically on a scale from 1 to 4.
- Standard Hours: The standard number of working hours, usually a constant value.
- Stock Option Level: The stock option level of the employee, usually ranging from 0 (no options) to 3 (higher levels).
- Total Working Years: The total years the employee has been working professionally.
- Training Times Last Year: The number of times the employee attended training sessions in the last year.
- Work Life Balance: A rating of the employee's work-life balance, typically on a scale of 1 to 4, with 1 being the least balanced and 4 the most balanced.
- Years At Company: The total number of years the employee has worked at the current company.
- Years In Current Role: The number of years the employee has been in their current role.
- Years Since Last Promotion: The number of years since the employee's last promotion.

- Years With Curr Manager: The number of years the employee has worked with their current manager.

## Steps involved in building the Project:

1. Study the dataset visually in Excel.

2. Normalise the dataset into different tables with Employee Number as the primary key.

3. Import all the tables along with the main dataset in SQL Server Management Studio

4. Verify the data imported in the HR Analytics Database

   **SQl Query:**

   ```
   SELECT * FROM employee_fact

   SELECT * FROM education

   SELECT * FROM experience

   SELECT * FROM job

   SELECT * FROM work

   SELECT * FROM ratings

   SELECT * FROM compensation

   SELECT * FROM employee

   SELECT * FROM IBM_HR_Dataset        --Main Dataset
   ```

5. Clean the Data and prepare for export to Power BI
   - **CHECKING FOR NULL VALUES IN EACH OF THE COLUMNS**
     SQL QUERY:
     ```
     SELECT
       SUM(CASE WHEN Age IS NULL THEN 1 ELSE 0 END) AS Null_Age,
       SUM(CASE WHEN Attrition IS NULL THEN 1 ELSE 0 END) AS
     Null_Attrition,
       SUM(CASE WHEN BusinessTravel IS NULL THEN 1 ELSE 0 END) AS
     Null_BusinessTravel,
     ```

```
  SUM(CASE WHEN DailyRate IS NULL THEN 1 ELSE 0 END) AS
Null_DailyRate,
  SUM(CASE WHEN Department IS NULL THEN 1 ELSE 0 END) AS
Null_Department,
  SUM(CASE WHEN DistanceFromHome IS NULL THEN 1 ELSE 0 END) AS
Null_DistanceFromHome,
  SUM(CASE WHEN Education IS NULL THEN 1 ELSE 0 END) AS
Null_Education,
  SUM(CASE WHEN EducationField IS NULL THEN 1 ELSE 0 END) AS
Null_EducationField,
  SUM(CASE WHEN EmployeeCount IS NULL THEN 1 ELSE 0 END) AS
Null_EmployeeCount,
  SUM(CASE WHEN EmployeeNumber IS NULL THEN 1 ELSE 0 END) AS
Null_EmployeeNumber,
  SUM(CASE WHEN EnvironmentSatisfaction IS NULL THEN 1 ELSE 0
END) AS Null_EnvironmentSatisfaction,
  SUM(CASE WHEN Gender IS NULL THEN 1 ELSE 0 END) AS
Null_Gender,
  SUM(CASE WHEN HourlyRate IS NULL THEN 1 ELSE 0 END) AS
Null_HourlyRate,
  SUM(CASE WHEN JobInvolvement IS NULL THEN 1 ELSE 0 END) AS
Null_JobInvolvement,
  SUM(CASE WHEN JobLevel IS NULL THEN 1 ELSE 0 END) AS
Null_JobLevel,
  SUM(CASE WHEN JobRole IS NULL THEN 1 ELSE 0 END) AS
Null_JobRole,
  SUM(CASE WHEN JobSatisfaction IS NULL THEN 1 ELSE 0 END) AS
Null_JobSatisfaction,
  SUM(CASE WHEN MaritalStatus IS NULL THEN 1 ELSE 0 END) AS
Null_MaritalStatus,
  SUM(CASE WHEN MonthlyIncome IS NULL THEN 1 ELSE 0 END) AS
Null_MonthlyIncome,
  SUM(CASE WHEN MonthlyRate IS NULL THEN 1 ELSE 0 END) AS
Null_MonthlyRate,
  SUM(CASE WHEN NumCompaniesWorked IS NULL THEN 1 ELSE 0
END) AS Null_NumCompaniesWorked,
  SUM(CASE WHEN Over18 IS NULL THEN 1 ELSE 0 END) AS
Null_Over18,
  SUM(CASE WHEN OverTime IS NULL THEN 1 ELSE 0 END) AS
Null_OverTime,
  SUM(CASE WHEN PercentSalaryHike IS NULL THEN 1 ELSE 0 END) AS
Null_PercentSalaryHike,
  SUM(CASE WHEN PerformanceRating IS NULL THEN 1 ELSE 0 END) AS
Null_PerformanceRating,
```

```
  SUM(CASE WHEN RelationshipSatisfaction IS NULL THEN 1 ELSE 0
END) AS Null_RelationshipSatisfaction,
  SUM(CASE WHEN StandardHours IS NULL THEN 1 ELSE 0 END) AS
Null_StandardHours,
  SUM(CASE WHEN StockOptionLevel IS NULL THEN 1 ELSE 0 END) AS
Null_StockOptionLevel,
  SUM(CASE WHEN TotalWorkingYears IS NULL THEN 1 ELSE 0 END) AS
Null_TotalWorkingYears,
  SUM(CASE WHEN TrainingTimesLastYear IS NULL THEN 1 ELSE 0 END)
AS Null_TrainingTimesLastYear,
  SUM(CASE WHEN WorkLifeBalance IS NULL THEN 1 ELSE 0 END) AS
Null_WorkLifeBalance,
  SUM(CASE WHEN YearsAtCompany IS NULL THEN 1 ELSE 0 END) AS
Null_YearsAtCompany,
  SUM(CASE WHEN YearsInCurrentRole IS NULL THEN 1 ELSE 0 END) AS
Null_YearsInCurrentRole,
  SUM(CASE WHEN YearsSinceLastPromotion IS NULL THEN 1 ELSE 0
END) AS Null_YearsSinceLastPromotion,
  SUM(CASE WHEN YearsWithCurrManager IS NULL THEN 1 ELSE 0 END)
AS Null_YearsWithCurrManager
FROM IBM_HR_Dataset;
```

**Result:** No Null Values were found

- **CHECKING FOR DUPLICATE ROWS**

```
SQL QUERY:
WITH CTE_Duplicates AS (
  SELECT
    *,
    ROW_NUMBER() OVER (
      PARTITION BY Age, Attrition, BusinessTravel, DailyRate, Department,
DistanceFromHome, Education, EducationField,
          EmployeeCount, EmployeeNumber,
EnvironmentSatisfaction, Gender, HourlyRate, JobInvolvement, JobLevel,
          JobRole, JobSatisfaction, MaritalStatus, MonthlyIncome,
MonthlyRate, NumCompaniesWorked, Over18,
          OverTime, PercentSalaryHike, PerformanceRating,
RelationshipSatisfaction, StandardHours,
          StockOptionLevel, TotalWorkingYears, TrainingTimesLastYear,
WorkLifeBalance, YearsAtCompany,
          YearsInCurrentRole, YearsSinceLastPromotion,
YearsWithCurrManager
      ORDER BY EmployeeNumber -- Specify a unique column for order
    ) AS RowNum
```

```
  FROM
    IBM_HR_Dataset
)
SELECT *
FROM CTE_Duplicates
WHERE RowNum > 1;
```

**Result:** No Null Values were found

- **UPDATING THE TABLES WITH THE PROPER VALUES**

SQL QUERY:

```
UPDATE education
SET education = CASE
  WHEN education = 1 THEN 'Below College'
  WHEN education = 2 THEN 'College'
  WHEN education = 3 THEN 'Bachelor'
  WHEN education = 4 THEN 'Master'
  WHEN education = 5 THEN 'Doctor'
END
WHERE education IN (1, 2, 3, 4, 5);

UPDATE job
SET JobInvolvement = CASE
  WHEN JobInvolvement = 1 THEN 'Low'
  WHEN JobInvolvement = 2 THEN 'Medium'
  WHEN JobInvolvement = 3 THEN 'High'
  WHEN JobInvolvement = 4 THEN 'Very High'
END
WHERE JobInvolvement IN (1, 2, 3, 4);

UPDATE ratings
SET EnvironmentSatisfaction = CASE
  WHEN EnvironmentSatisfaction = 1 THEN 'Low'
```

```
    WHEN EnvironmentSatisfaction = 2 THEN 'Medium'
    WHEN EnvironmentSatisfaction = 3 THEN 'High'
    WHEN EnvironmentSatisfaction = 4 THEN 'Very High'
END
WHERE EnvironmentSatisfaction IN (1, 2, 3, 4);

UPDATE ratings
SET RelationshipSatisfaction = CASE
    WHEN RelationshipSatisfaction = 1 THEN 'Low'
    WHEN RelationshipSatisfaction = 2 THEN 'Medium'
    WHEN RelationshipSatisfaction = 3 THEN 'High'
    WHEN RelationshipSatisfaction = 4 THEN 'Very High'
END
WHERE RelationshipSatisfaction IN (1, 2, 3, 4);

UPDATE ratings
SET WorkLifeBalance = CASE
    WHEN WorkLifeBalance = 1 THEN 'Bad'
    WHEN WorkLifeBalance = 2 THEN 'Good'
    WHEN WorkLifeBalance = 3 THEN 'Better'
    WHEN WorkLifeBalance = 4 THEN 'Best'
END
WHERE WorkLifeBalance IN (1, 2, 3, 4);

UPDATE ratings
SET JobSatisfaction = CASE
    WHEN JobSatisfaction = 1 THEN 'Low'
    WHEN JobSatisfaction = 2 THEN 'Medium'
    WHEN JobSatisfaction = 3 THEN 'High'
```

```
    WHEN JobSatisfaction = 4 THEN 'Very High'
END
WHERE JobSatisfaction IN (1, 2, 3, 4);

UPDATE ratings
SET PerformanceRating = CASE
    WHEN PerformanceRating = 1 THEN 'Low'
    WHEN PerformanceRating = 2 THEN 'Good'
    WHEN PerformanceRating = 3 THEN 'Excellent'
    WHEN PerformanceRating = 4 THEN 'Outstanding'
END
WHERE PerformanceRating IN (1, 2, 3, 4);
```

- **CATEGORIZING THE VALUES OF SOME COLUMNS:**

SQL QUERY:

```
ALTER TABLE employee
ADD Age_Group VARCHAR(50);

UPDATE employee
SET Age_Group = CASE
    WHEN Age BETWEEN 18 AND 30 THEN '18-30'
    WHEN Age BETWEEN 31 AND 40 THEN '31-40'
    WHEN Age BETWEEN 41 AND 50 THEN '41-50'
    WHEN Age BETWEEN 51 AND 60 THEN '51-60'
    ELSE 'Unknown'
END;

ALTER TABLE experience
ADD years_at_company VARCHAR(50);

UPDATE experience
SET years_at_company = CASE
    WHEN YearsAtCompany BETWEEN 0 AND 1 THEN '0-1'
    WHEN YearsAtCompany BETWEEN 2 AND 5 THEN '2-5'
    WHEN YearsAtCompany BETWEEN 6 AND 10 THEN '6-10'
    WHEN YearsAtCompany BETWEEN 11 AND 20 THEN '11-20'
    WHEN YearsAtCompany BETWEEN 21 AND 30 THEN '21-30'
```

```
  WHEN YearsAtCompany BETWEEN 31 AND 40 THEN '31-40'
  ELSE 'Out of Range'
END;

ALTER TABLE experience
ADD years_in_current_role VARCHAR(50);

UPDATE experience
SET years_in_current_role = CASE
  WHEN YearsInCurrentRole BETWEEN 0 AND 1 THEN '0-1'
  WHEN YearsInCurrentRole BETWEEN 2 AND 5 THEN '2-5'
  WHEN YearsInCurrentRole BETWEEN 6 AND 10 THEN '6-10'
  WHEN YearsInCurrentRole BETWEEN 11 AND 15 THEN '11-15'
  WHEN YearsInCurrentRole >= 16 THEN '15+'
  ELSE 'Out of Range'
END;

ALTER TABLE experience
ADD years_since_last_promotion VARCHAR(50);

UPDATE experience
SET years_since_last_promotion = CASE
  WHEN YearsSinceLastPromotion BETWEEN 0 AND 1 THEN '0-1 Years'
  WHEN YearsSinceLastPromotion BETWEEN 2 AND 5 THEN '2-5 Years'
  WHEN YearsSinceLastPromotion BETWEEN 6 AND 10 THEN '6-10 Years'
  WHEN YearsSinceLastPromotion BETWEEN 11 AND 15 THEN '11-15
Years'
  ELSE 'Out of Range'
END;

ALTER TABLE experience
ADD years_with_current_manager VARCHAR(50);

UPDATE experience
SET years_with_current_manager = CASE
  WHEN YearsWithCurrManager BETWEEN 0 AND 1 THEN '0-1'
  WHEN YearsWithCurrManager BETWEEN 2 AND 5 THEN '2-5'
  WHEN YearsWithCurrManager BETWEEN 6 AND 10 THEN '6-10'
  WHEN YearsWithCurrManager BETWEEN 11 AND 15 THEN '11-15'
  WHEN YearsWithCurrManager >= 16 THEN '15+'
  ELSE 'Out of Range'
END;

ALTER TABLE employee_fact
```

```
ADD income_slab VARCHAR(50);

UPDATE employee_fact
SET income_slab = CASE
  WHEN MonthlyIncome BETWEEN 1000 AND 3000 THEN '1k-3k'
  WHEN MonthlyIncome BETWEEN 3001 AND 5000 THEN '3k-5k'
  WHEN MonthlyIncome BETWEEN 5001 AND 8000 THEN '5k-8k'
  WHEN MonthlyIncome BETWEEN 8001 AND 12000 THEN '8k-12k'
  WHEN MonthlyIncome BETWEEN 12001 AND 15000 THEN '12k-15k'
  WHEN MonthlyIncome BETWEEN 15001 AND 20000 THEN '15k-20k'
  ELSE 'Out of Range'
END;

ALTER TABLE employee_fact
ADD Salary_hike_percentage VARCHAR(50);

UPDATE employee_fact
SET Salary_hike_percentage = CASE
  WHEN PercentSalaryHike BETWEEN 10 AND 12 THEN '10-12%'
  WHEN PercentSalaryHike BETWEEN 13 AND 16 THEN '13-16%'
  WHEN PercentSalaryHike BETWEEN 17 AND 20 THEN '17-20%'
  WHEN PercentSalaryHike BETWEEN 21 AND 25 THEN '21-25%'
  ELSE 'Out of Range'
END;
```

6. Import the Tables (excluding the main dataset) to Power BI.

7. Build the following measures required for the project.

DAX Queries of the measures:

1. Distance from Home = AVERAGE('work'[DistanceFromHome])
2. Monthly Income = AVERAGE(employee_fact[MonthlyIncome])
3. Number of Companies Worked = AVERAGE(experience[NumCompaniesWorked])
4. Salary Hike = AVERAGE(employee_fact[PercentSalaryHike])/100
5. Total Attrition = COUNTROWS(FILTER(employee_fact, employee_fact[Attrition] = "Yes"))
6. Total employees = COUNTROWS(employee)
7. Total Working Years = AVERAGE(experience[TotalWorkingYears])
8. Training Times last year = AVERAGE(experience[TrainingTimesLastYear])
9. Turnover Rate = DIVIDE(

```
    COUNTROWS(FILTER(employee_fact, employee_fact[Attrition] = "Yes")),
    COUNTROWS(employee_fact),
    0)
```

10. Years at Company = AVERAGE(experience[YearsAtCompany])
11. Years in current role = AVERAGE(experience[YearsInCurrentRole])
12. Years Since Promotion = AVERAGE(experience[YearsSinceLastPromotion])
13. Years With Current Manager =AVERAGE(experience[YearsWithCurrManager])

8. Build the visuals.

## Pages of the Project:

1. **Attrition Overview**: Visualizes demographic and departmental attrition trends to identify high-risk employee segments.

   **Insights:**

   **Gender:** Male employees have a higher attrition rate (63.29%) compared to females.

   **Marital Status:** Single employees (50.63%) exhibit higher attrition than married or divorced ones, indicating potential dissatisfaction among unmarried employees.

   **Job Roles:** Laboratory Technicians and Sales Executives experience the highest attrition, highlighting a potential need to investigate these roles' work conditions or challenges.

   **Education:** Most attrition occurs among employees with a Bachelor's degree, suggesting they may seek higher education or better opportunities.

   **Education Field:** The high attrition rate among employees with a background in Life Sciences reflects potential dissatisfaction or challenges specific to their roles or opportunities within the organization.

   **Age Group:** The majority of attrition is concentrated in the 18-30 age group, possibly reflecting career exploration or dissatisfaction among younger employees.

   **Departments:** Research & Development and Sales departments have the highest turnover rates, calling for targeted interventions in these areas.

2. **Salary & Satisfaction**: Evaluates the impact of job satisfaction, compensation, and work-life balance on employee turnover.

   **Insights:**

**Business Travel**: Employees who travel rarely show the highest attrition (65.82%), implying potential stress related to infrequent but disruptive travel.
**Income Category**: Employees earning between 1k-3k monthly experience the highest attrition, indicating dissatisfaction with low wages.
**Job Satisfaction**: Employees with high job satisfaction leave more frequently, reflecting potential lack of growth while it is second highest of low job satisfaction indicating disengagement or workplace dissatisfaction.
**Work-Life Balance:** High attrition for "better" and "good" work-life balance suggests that while balance is adequate, employees may leave for reasons like career growth, pay, or external opportunities.
**Environment Satisfaction**: Employees with low satisfaction in their work environment exhibit the highest turnover, pointing to workplace improvements as a key retention strategy.
**Stock Option Level**: High attrition at stock option level 0 indicates employees may feel undervalued or lack financial incentives to stay, prompting them to seek better benefits elsewhere.
**Salary Hike**: Moderate salary hikes (13%-16%) are associated with higher attrition, suggesting salary increases alone may not suffice to retain employees.

3. **Career Progression**: Assesses the role of career growth opportunities and tenure in driving attrition rates.

### Insights:

**Years at Company:** Most attrition occurs within the first five years, with a significant drop after that, highlighting onboarding and early engagement as critical phases.
**Years Since Last Promotion:** Employees who have not been promoted within 1-5 years are more likely to leave, underlining the importance of timely career advancement opportunities.
**Years with Current Manager:** Short tenures with a current manager (0-1 year) have the highest attrition, indicating potential gaps in leadership or alignment.
**Current Role Duration:** Employees in roles for 0-5 years leave at a higher rate, suggesting a need for role variety and progression.