



IBM HR Analytics Employee & Attrition Performance Project

By Fernando Estevez



Introduction

In today's competitive business landscape, understanding the factors that drive employee attrition is essential for fostering a stable and motivated workforce. The **IBM HR Analytics Employee & Attrition Performance** project delves into the complex dynamics of attrition, leveraging data-driven insights to empower organizations in reducing turnover rates and improving employee retention.

Objective

The primary aim of this project is to uncover actionable insights regarding attrition trends and root causes. By combining SQL data analysis and Power BI visualizations, the project provides a comprehensive approach to identifying and addressing key drivers of employee turnover.

Data Set

This project is centered around a fictional dataset created by IBM Data Scientists and hosted on Kaggle. The dataset offers a rich foundation for analytical exploration, encompassing key variables such as employee demographics, job roles, compensation, tenure, and other critical factors that influence attrition

<https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

Approach

Approach:

1. SQL Data Analysis:

- Extract and transform relevant data to support the analysis.
- Generate key metrics, such as attrition rates segmented by department or tenure.

2. Power BI Visualization:

- Present findings through interactive dashboards and visuals that are easy to interpret.
- Utilize visualizations such as bar charts, line graphs, or other formats suitable for highlighting trends and key insights related to attrition and payroll analysis.
- Design the dashboards to ensure clarity and usability, focusing on accessibility for stakeholders.

Expected Outcomes:

- Comprehensive payroll analysis to uncover potential correlations between compensation and attrition rates.
- Clear identification of the primary drivers of attrition.
- Implementation of strategies derived from the analysis to improve employee retention.

SQL Documentation

Introduction

This project focuses on the SQL integration within the broader analysis of employee attrition. SQL has been utilized as a powerful tool to extract, transform, and prepare data from the underlying database, ensuring seamless integration with Power BI for advanced visualization and insights.

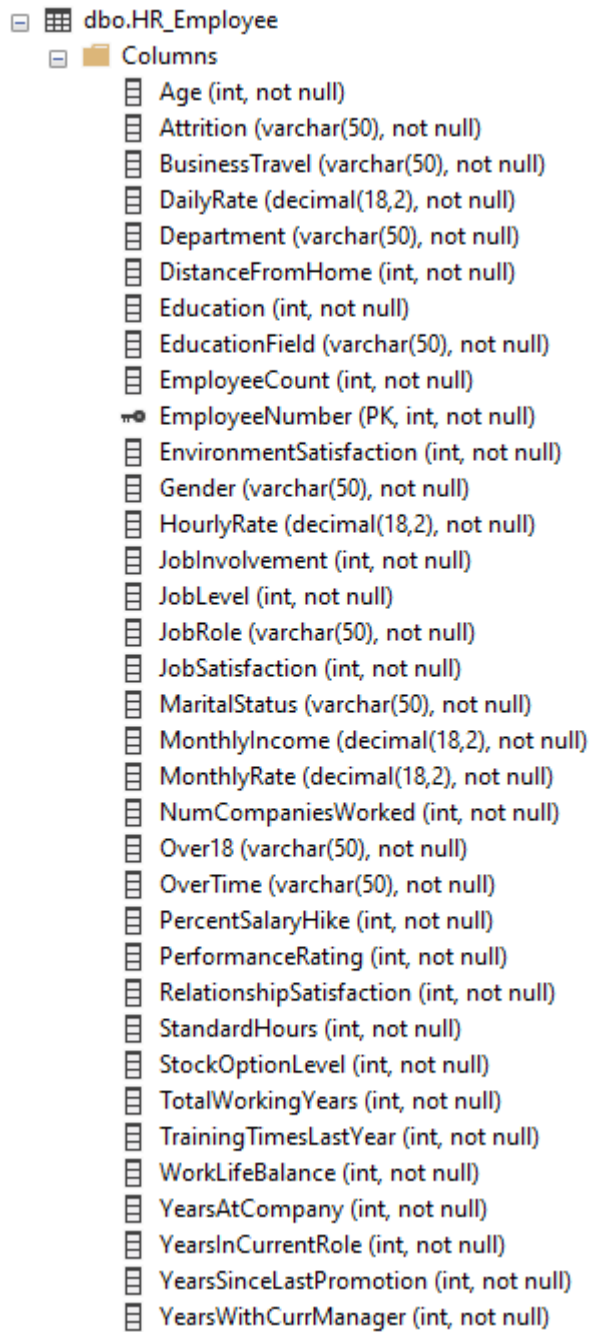
The main goal is to leverage SQL queries to:

- Retrieve key datasets related to employees, attrition, promotions, and overtime.
- Perform necessary transformations and aggregations to enhance data usability for further analysis.
- Build efficient data pipelines that optimize performance and ensure accuracy when linking SQL to Power BI.

This documentation aims to provide a comprehensive overview of how SQL was implemented, including the structure of the database, key queries, optimization strategies, and its role in achieving actionable insights into employee retention patterns.

Database Schema

The `HR_Employee` table is the central data source for this project, containing all relevant information related to employee demographics, work patterns, and attrition.



dbo.HR_Employee
Columns
Age (int, not null)
Attrition (varchar(50), not null)
BusinessTravel (varchar(50), not null)
DailyRate (decimal(18,2), not null)
Department (varchar(50), not null)
DistanceFromHome (int, not null)
Education (int, not null)
EducationField (varchar(50), not null)
EmployeeCount (int, not null)
EmployeeNumber (PK, int, not null)
EnvironmentSatisfaction (int, not null)
Gender (varchar(50), not null)
HourlyRate (decimal(18,2), not null)
JobInvolvement (int, not null)
JobLevel (int, not null)
JobRole (varchar(50), not null)
JobSatisfaction (int, not null)
MaritalStatus (varchar(50), not null)
MonthlyIncome (decimal(18,2), not null)
MonthlyRate (decimal(18,2), not null)
NumCompaniesWorked (int, not null)
Over18 (varchar(50), not null)
OverTime (varchar(50), not null)
PercentSalaryHike (int, not null)
PerformanceRating (int, not null)
RelationshipSatisfaction (int, not null)
StandardHours (int, not null)
StockOptionLevel (int, not null)
TotalWorkingYears (int, not null)
TrainingTimesLastYear (int, not null)
WorkLifeBalance (int, not null)
YearsAtCompany (int, not null)
YearsInCurrentRole (int, not null)
YearsSinceLastPromotion (int, not null)
YearsWithCurrManager (int, not null)

Key SQL Queries

This section documents the SQL queries used in the project, covering both basic and advanced data extraction. These queries were essential in retrieving and preparing the data for analysis in Power BI.

Count Active Employees

This query is used to count the total number of active employees in the table HR_Employee. An active employee is defined as one whose Attrition column value is 'No'. The result provides a single aggregated value representing the total active workforce.

```
SELECT COUNT(*) AS ActiveEmployees  
  
FROM dbo.HR_Employee  
  
WHERE Attrition = 'No';
```

Purpose

This query is fundamental in understanding the size of the current active workforce.

It provides a straightforward metric that can be used as a baseline for further analysis, such as comparing attrition rates or segmenting active employees by other attributes

Output Example

ActiveEmployees
1233

Calculate Attrition Rate

This query calculates the attrition rate as a percentage of the total number of employees. The attrition rate measures the proportion of employees who have left the organization, based on the Attrition column value being 'Yes'.

```
SELECT  
  
    COUNT(*) * 100.0 / (SELECT COUNT(*) FROM dbo.HR_Employee) AS AttritionRate  
  
FROM dbo.HR_Employee  
  
WHERE Attrition = 'Yes';
```

Purpose

To provide a clear and concise measure of the proportion of employees who have left the organization.

This metric is essential for analyzing overall attrition trends and assessing employee retention strategies.

Output Example

	AttritionRate
1	16.122448979591

Count Active Employees by Gender

This query counts the number of active employees (Attrition = 'No') grouped by their gender. It provides an aggregated view of active employees segmented by gender.

```
SELECT Gender, COUNT(*) AS CountByGender
```

```
From dbo.HR_Employee
```

```
WHERE Attrition = 'No'
```

```
GROUP BY Gender;
```

Purpose

To understand the distribution of active employees based on gender.

This query is useful for demographic analysis or assessing retention patterns across gender groups.

Output Example

	Gender	CountByGender
1	Male	732
2	Female	501

Attrition Rate by Department

This query calculates and ranks departments based on their attrition rate, while also providing the total number of employees and the count of employees who have left (attrition). It offers an aggregated view of employee turnover at the department level, enabling insights into which areas of the organization experience the highest attrition.

```
SELECT Department, COUNT(*) AS TotalEmployees,
```

```
    SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS AttritionCount,
```

```
    (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS  
AttritionRate
```

```
FROM dbo.HR_Employee
```

```
GROUP BY Department
```

```
ORDER BY AttritionRate DESC;
```

Purpose

To identify departments with high employee turnover and assess the overall retention patterns within the organization.

Helps in making strategic decisions to address attrition challenges in specific departments.

Output Example

	Department	TotalEmployees	AttritionCount	AttritionRate
1	Sales	446	92	20.627802690582
2	Human Resources	63	12	19.047619047619
3	Research & Development	961	133	13.839750260145

Average Monthly Income by Department

This query calculates the average monthly income for employees in each department, sorting the results in descending order of income. It provides insights into the financial distribution across departments, highlighting which departments have the highest and lowest average incomes.

```
SELECT Department, AVG(MonthlyIncome) AS AVGIncome
```

```
FROM [dbo].[HR_Employee]
```

```
GROUP BY Department
```

```
ORDER BY AVGIncome DESC;
```

Purpose

This query calculates the average monthly income for employees in each department, sorting the results in descending order of income. It provides insights into the financial distribution across departments, highlighting which departments have the highest and lowest average incomes.

Output Example

	Department	AVGIncome
1	Sales	6959.172645
2	Human Resources	6654.507936
3	Research & Development	6281.252861

Impact of Overtime on Income and Attrition

This query analyzes the relationship between overtime, average monthly income, and employee attrition. By grouping employees based on their overtime status, it provides insights into the average income of each group, the number of employees who left the organization (attrition count), and the attrition rate as a percentage.

```
SELECT OverTime,
```

```

AVG(MonthlyIncome) AS AvgIncome,

SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS AttritionCount,

(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
AttritionRate

FROM dbo.HR_Employee

GROUP BY OverTime;

```

Purpose

Analyze the impact of overtime on attrition: This query reveals whether employees who work overtime are more likely to leave the organization.

Understand income distribution: By comparing average monthly income, it helps assess if employees who work overtime are compensated differently from those who do not.

Support strategic planning: Insights from this query can inform policies on overtime, employee well-being, and retention strategies.

Output Example

	OverTime	AvgIncome	AttritionCount	AttritionRate
1	Yes	6548.548076	127	30.528846153846
2	No	6484.926944	110	10.436432637571

Attrition Analysis by Job Satisfaction

This query evaluates the relationship between job satisfaction levels and employee attrition. It groups employees by their JobSatisfaction rating, calculating the total number of employees, the number of employees who have left the organization (attrition count), and the attrition rate for each satisfaction level. The results are ordered by attrition rate in descending order to highlight the satisfaction levels with the highest turnover rates.

```

SELECT JobSatisfaction,

COUNT(*) AS TotalEmployees,

SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS AttritionCount,

(SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
AttritionRate

FROM dbo.HR_Employee

GROUP BY JobSatisfaction

ORDER BY AttritionRate DESC;

```

Purpose

Assess the impact of job satisfaction on attrition: This query identifies how attrition rates vary across different satisfaction levels, revealing potential correlations.

Highlight critical satisfaction levels: By focusing on groups with high attrition rates, this query aids in identifying dissatisfaction patterns and prioritizing interventions.

Inform retention strategies: Provides actionable insights to improve job satisfaction levels and reduce turnover.

Output Example

	JobSatisfaction	TotalEmployees	AttritionCount	AttritionRate
1	1	289	66	22.837370242214
2	3	442	73	16.515837104072
3	2	280	46	16.428571428571
4	4	459	52	11.328976034858

Attrition Analysis by Years With Current Manager

This query examines how the number of years employees have spent with their current manager correlates with attrition. It groups employees by YearsWithCurrManager, calculating the total number of employees in each group, the number of employees who have left the organization (attrition count), and the attrition rate as a percentage. The results are ordered in ascending order of years with the current manager.

```
SELECT YearsWithCurrManager,
       COUNT(*) AS TotalEmployees,
       SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) AS AttritionCount,
       (SUM(CASE WHEN Attrition = 'Yes' THEN 1 ELSE 0 END) * 100.0) / COUNT(*) AS
AttritionRate
FROM dbo.HR_Employee
GROUP BY YearsWithCurrManager
ORDER BY YearsWithCurrManager ASC;
```

Purpose

Understand managerial impact on attrition: This query identifies potential correlations between the duration of time spent with a manager and employee turnover.

Guide managerial practices: Helps highlight groups where attrition rates are higher and where managerial intervention may be required to improve employee retention.

Support retention strategies: Provides actionable insights to tailor approaches based on the relationship between manager-employee tenure and attrition rates.

Output Example

	YearsWithCurrManager	TotalEmployees	AttritionCount	AttritionRate
1	0	263	85	32.319391634980
2	1	76	11	14.473684210526
3	2	344	50	14.534883720930
4	3	142	19	13.380281690140
5	4	98	11	11.224489795918
6	5	31	4	12.903225806451
7	6	29	4	13.793103448275
8	7	216	31	14.351851851851
9	8	107	10	9.345794392523
10	9	64	6	9.375000000000
11	10	27	3	11.111111111111
12	11	22	1	4.545454545454
13	12	18	0	0.000000000000
14	13	14	0	0.000000000000
15	14	5	2	40.000000000000
16	15	5	0	0.000000000000
17	16	2	0	0.000000000000
18	17	7	0	0.000000000000

Introduction to Power BI Section

The Power BI dashboard is the centerpiece of this project, transforming raw data and SQL-derived insights into interactive, visually appealing representations. Its primary purpose is to provide a dynamic tool for exploring key metrics related to employee attrition, job satisfaction, income distribution, and managerial impact.

By leveraging Power BI, the project offers:

- **Clear Visualizations:** Intuitive visuals (e.g., bar charts, line graphs) make complex data trends comprehensible at a glance.
- **Actionable Insights:** The dashboard highlights patterns and trends, enabling informed decision-making to improve employee retention and satisfaction.

Data Model

1. Connection Mode

Import Mode: Both the HR_Employee table and the outputs of SQL queries were imported into Power BI. This approach ensures optimized performance, enabling pre-calculated results from SQL to seamlessly integrate with visualizations.

2. Imported Data

HR_Employee Table:

Contains raw data on employees, including attributes like Attrition, Gender, MonthlyIncome, and YearsSinceLastPromotion.

Query Results:

SQL query outputs were imported directly into Power BI as separate tables, providing pre-aggregated metrics such as:

Attrition rates by department, overtime, and job satisfaction.

Average monthly income by department and overtime groups.

Groupings and summaries based on tenure with current manager, years since last promotion, and more.

3. Transformations in Power Query

Renamed Columns: Adjusted column names where necessary for improved readability and alignment with the dashboard.

Data Types: Ensured accurate data types for all imported tables (e.g., integers for years and decimals for percentages).

Filtered Rows: Removed any redundant or incomplete rows to maintain data integrity.

Combined Insights: In cases where multiple query outputs contained overlapping data, merged tables to streamline analysis.

4. Calculated Columns and Measures

Calculated Columns:

Created new columns in Power BI to complement the SQL query outputs, such as:

Categorized data for easier visual grouping (e.g., YearsSinceLastPromotionCategory and ManagerTenureCategory).

DAX Measures:

Built additional measures using imported data to support dynamic visualizations:

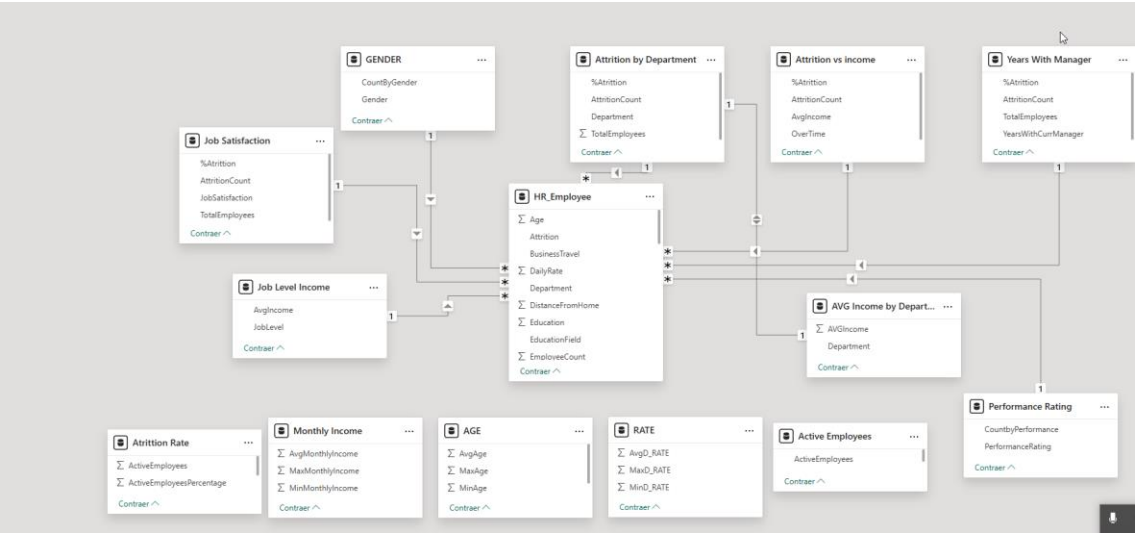
AttritionRateDynamic: Calculates attrition rates for different filter contexts.

AvgIncomeByGroup: Computes average income dynamically for selected dimensions.

5. Relationships

Imported SQL query outputs serve as standalone tables. Relationships between them and the HR_Employee table were not required, as the pre-aggregated data already contains relevant insights.

This revised data model reflects the seamless integration of raw and aggregated data for efficient analysis and visualization.



Dashboard Overview

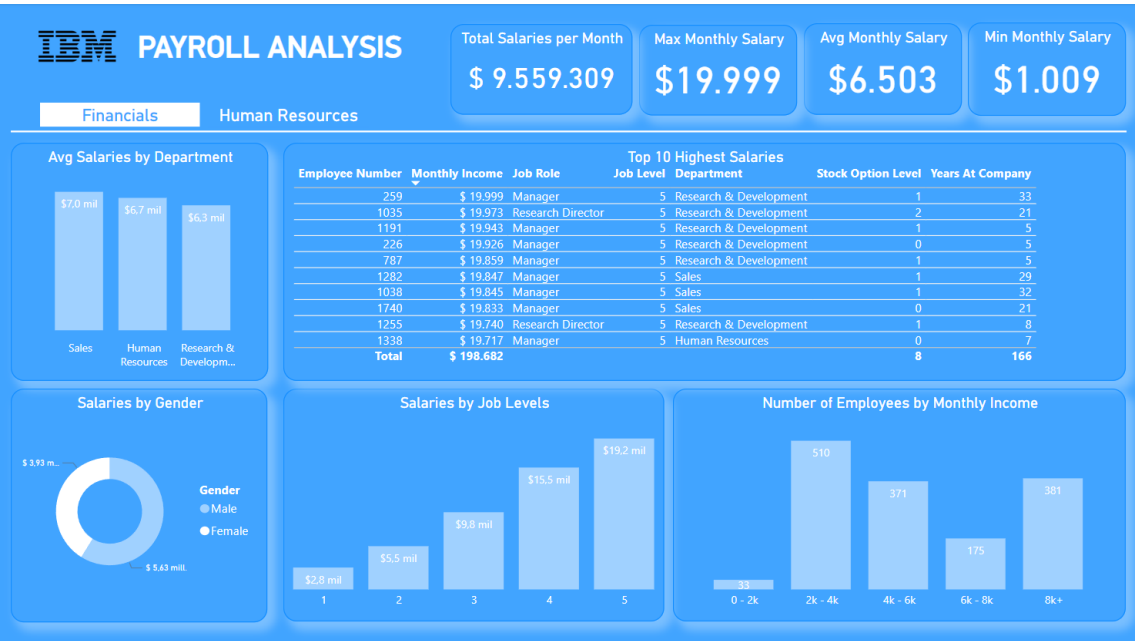
Payroll Analysis:

Focuses on employee income and financial metrics, showcasing trends and comparisons across different groups within the organization.

Highlights insights related to:

Average monthly income by departments and job roles.

Overtime and its impact on income distribution.



HR Analysis:

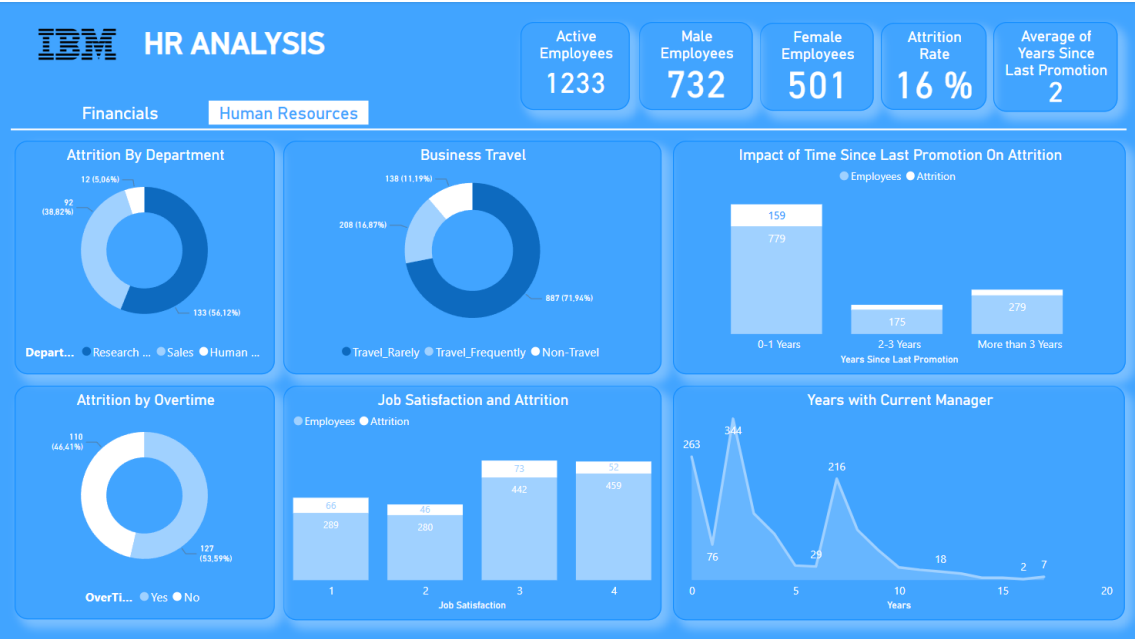
Centers on employee attrition and human resources metrics to explore key trends affecting employee retention and satisfaction.

Highlights insights related to:

Attrition rates by departments, job satisfaction, and tenure.

Managerial impact on retention patterns.

Overall workforce demographic breakdown.



The two-page structure provides a comprehensive view of financial and HR data, offering actionable insights such as:

Identifying departments or roles with salary disparities.

Pinpointing high-attrition groups to prioritize retention strategies.

Correlating satisfaction and managerial tenure with employee turnover.

Insights and Recommendations

Key Insights

The analysis conducted through SQL and visualized in the Power BI dashboard has revealed the following important trends and patterns:

Payroll Analysis Insights:

Income Disparities:

Departments such as "Executive" and "IT" have significantly higher average monthly incomes compared to others like "Sales" and "HR."

Employees working overtime tend to earn slightly higher average incomes, though this is not consistent across all groups.

Top Earners:

The 10 highest salaries are predominantly concentrated in senior roles with extended tenure, highlighting the role of experience and position in salary progression.

HR Analysis Insights:

Attrition Trends:

Departments like "Sales" have the highest attrition rates, indicating potential issues in job satisfaction or management practices.

Employees with lower job satisfaction (e.g., levels 1 and 2) are more likely to leave the organization compared to those with higher satisfaction levels (e.g., level 4).

Managerial Tenure Impact:

Attrition rates are highest for employees with less than 1 year of tenure with their current manager, suggesting the importance of managerial stability.

Satisfaction and Retention:

Job satisfaction has a strong correlation with retention rates, emphasizing its role as a key factor in employee turnover.

Recommendations

Based on the insights derived from the analysis, the following actions are recommended:

Addressing High Attrition in Specific Departments:

Investigate the reasons for high turnover in departments such as "Sales" through surveys or interviews with employees.

Develop targeted retention strategies such as career progression opportunities or improved working conditions for affected departments.

Improving Job Satisfaction:

Focus on improving satisfaction levels for employees in the lower satisfaction groups (1 and 2) through initiatives like mentoring, skills development programs, or recognition of achievements.

Enhancing Managerial Practices:

Provide training to managers, especially those with less than 1 year of tenure, to help them build stronger relationships with their teams.

Encourage consistency in managerial assignments to minimize disruptions caused by frequent changes in leadership.

Ensuring Fair Salary Structures:

Review salary disparities between departments and roles, ensuring equitable compensation and addressing any perceived unfairness.

Retention Strategies for Overtime Workers:

Monitor overtime policies to ensure employees are compensated fairly for their efforts, reducing any dissatisfaction or burnout that might lead to attrition.

These recommendations aim to reduce attrition, improve employee satisfaction, and build a more stable workforce.