# IBM HR Analytics Project Report

Project Title: Understanding and Reducing Employee Attrition

Tools: A Comprehensive Analysis Using Excel and Power BI

Submitted by: Mumtaj

Date: [Submission Date]

**Executive Summary:**

This HR analytics project explores the key drivers of employee attrition using data from a corporate employee database. The goal was to uncover patterns and factors that influence why employees leave the organization, using techniques such as exploratory data analysis (EDA), correlation analysis, hypothesis testing, and linear regression.

The dataset included variables such as age, income, job level, distance from home, years at the company, and job satisfaction. EDA revealed that attrition is more common among younger employees, those with lower monthly income, fewer years at the company, and greater commuting distances.

Statistical tests, including t-tests, confirmed significant differences in income between employees who left and those who stayed. The regression analysis, while showing a modest R-squared value, identified income, age, and managerial relationship as important predictors.

Key findings suggest that compensation, work-life balance, and employee engagement are critical areas for intervention. Business recommendations include reviewing pay structures, offering flexible work arrangements, and investing in leadership training.

This project demonstrates the importance of data-driven HR strategies to improve retention and strengthen workforce stability.

**Introduction:**

An organisation depends a lot on its people. HR analytics is an essential tool for comprehending employee behaviour, reducing turnover and building performance. As we see more competition and demand for talent, organizations should use employee data to inform strategic HR decisions.  
  
This study shall concentrate on the analysis of IBM HR Analytics Employee Attrition & Performance dataset in Microsoft Excel. In the dataset some of the attributes of employees like their age, education, work experience, job satisfaction, work life balance, income prospect, attrition status are provided. Through this data we hope to learn what factors contribute to employee attrition and performance.  
  
The goal of this project is apply predictive models which can help in predicting which employees are more likely to terminate the employment and provide businesses with findings which can employed to take corrective action to prevent employee attrition and increase employee satisfaction, ultimately leading to better productivity and organizational success.

**Problem Statement:**

Employees are one of the most valuable assets of any organization. Companies invest a lot of time and money in hiring and developing talent. When employees—especially skilled or experienced ones—leave, it results in significant losses and disrupts overall productivity.

Attrition can happen for many reasons: low pay, lack of growth opportunities, poor leadership, absence of recognition, or limited work flexibility. As a result, many employees feel dissatisfied and start looking for better opportunities, causing the attrition rate to rise.

To reduce this, the HR team needs a way to identify which employees are at risk of leaving. By analyzing historical employee data, they can better understand the reasons behind attrition and take preventive steps—like improving work conditions or targeted hiring—to minimize the impact on the organization.

**Objective:**

The goal is to find out the most important factors which will influence the cause of time to quit or leave and het worst attribute of employee attrition. The aim is to analyse the patterns/ reasons of attrition in the employees and make relevant changes in the workplace in order to increase them employee retention. This study also aims to assist HR in data-based decision-making in order to avoid attrition and, thus, improve organizational performance.

**Hypotheses:**

1. Higher employee attrition leads to increased costs in recruitment, onboarding, and training of new employees.
2. Increased employee attrition negatively affects organizational productivity, performance, and profitability.
3. Employees with low satisfaction and limited growth opportunities are more likely to leave the organization.
4. Departments with poor management or high workload have higher attrition rates than others.

**About the Dataset:**

The dataset used in this project is titled “IBM HR Analytics Employee Attrition & Performance” and is publicly available on Kaggle. It contains detailed information about 1,470 employees from IBM, with the goal of predicting factors that influence employee attrition.

The dataset includes both categorical and numerical variables across different HR-related dimensions such as:

* Demographics: Age, Gender, Marital Status, Education
* Job-related: Job Role, Department, Job Level, Years at Company
* Performance-related: Job Involvement, Job Satisfaction, Performance Rating
* Compensation-related: Monthly Income, Hourly Rate, Stock Option Level
* Work environment: Environment Satisfaction, Work-Life Balance, Training Times
* Target Variable: Attrition (Yes/No)

There are a total of 35 columns (features) and 1,470 rows (employees), with no missing values, making it ideal for analysis in Excel without additional data imputation.

This rich set of attributes allows for comprehensive HR analytics and insight into key factors affecting employee turnover and performance.

**Data Cleaning and Preprocessing:**

After importing the dataset into Excel, the next step was to clean and prepare the data for analysis. Data cleaning is essential to remove inconsistencies, errors, and irrelevant information that may affect the accuracy of the analysis.

The process of data cleaning saw a few fundamental steps undertaken to ensure that the dataset was correct, consistent, and ready for analysis. One was checking for duplicate rows through Excel's "Remove Duplicates" option, but none was found. Missing values were identified using "Go To Special → Blanks"; in contrast, the dataset was full, with no null or blank entries. In a bid for consistency, text fields like Gender and Department were changed with respect to formatting, and all numeric fields were checked to fit the correct format of numbers. A few columns were renamed—for instance, TotalWorkingYears was changed to Experience, and MonthlyIncome was changed to Income—for better reading and easier use.

Furthermore, derived columns were created to add analytical strength to the dataset. An example could be an Age Group column, with IF statements stating that individuals between 18 and 25 are labeled young; those between 26 and 35 as mid-age; those between 36 and 45 as experienced; and those above 46 years as senior. Outliers were detected in important numerical columns like MonthlyIncome, TotalWorkingYears, and DistanceFromHome using the Interquartile Range (IQR) method, with an upper limit of Q3 + 1.5 × IQR. The analysis discovered 25 monthly income outliers, 7 distance from home outliers, and 5 total working years outliers. Instead of removing outliers, which might have caused loss of the dataset, the value of the outlier was capped to the upper bound using an IF formula in Excel: =IF(cell\_value > UpperBound, UpperBound, cell\_value). This helped preserve the data while taking care of its extreme values.

**Exploratory Data Analysis (EDA):**

In this step, different employee attributes were examined for patterns and trends related to attrition. Charts, with PivotTables and conditional formatting in Excel, were used to visualize certain factors-likewise job satisfaction, income, experience, age, and department-to determine their influence on an employee leaving the organization. Each factor was analyzed individually by bar charts, pivot tables, and filters to get a deeper understanding of the trends.

**1.Job Satisfaction vs. Attrition**

Employees with low job satisfaction (levels 1 and 2) had higher attrition rates, indicating dissatisfaction is a strong driver for quitting. Those with higher satisfaction (levels 3 and 4) were more likely to stay.

***Hypothesis :*** Low Job Satisfaction Leads to Higher Attrition

Why this matters: Job satisfaction is directly linked to employee morale and engagement.

What we found:

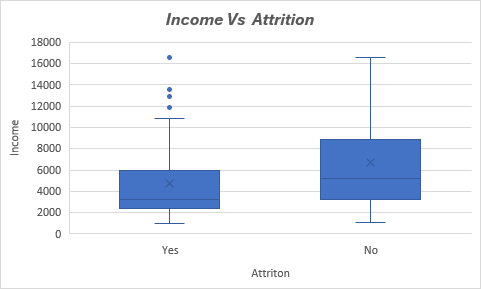
* + Attrition was highest in job satisfaction levels 1 and 2.
  + Levels 3 and 4 had significantly lower attrition.

Conclusion: Employees with low job satisfaction are much more likely to leave.  
🔹 Hypothesis Accepted

***Interpretation***: Improving job satisfaction—through recognition, better leadership, or career development—can reduce employee turnover.

**2.Monthly Income vs. Attrition**

Employees with lower monthly income had a higher attrition rate, while those earning more were less likely to leave.



The box plot shows that employees who left the company generally had lower monthly incomes than those who stayed. Both the median and mean income are lower in the attrition group, highlighting compensation as a key factor in retention. Additionally, more outliers in the lower-income group suggest that even a few well-paid employees left, though this was less common.

***Hypothesis :*** Lower Monthly Income Increases Attrition

* Why this matters: Compensation is one of the top reasons employees change jobs.
* What we found:
  + Employees who left had a much lower average monthly income compared to those who stayed.
* ✅ Conclusion: Lower-paid employees are more prone to quitting.  
  🔹 Hypothesis Accepted

***Interpretation:*** Compensation is a major factor in employee satisfaction and retention. Competitive salaries could reduce attrition.

**3.Age vs. Attrition**

The analysis shows that younger employees (<25) have the highest attrition rate (~35%), followed by those aged 26–35 (~19%). In contrast, employees aged 36+ show significantly lower attrition (8–12%).

This indicates that younger employees are more likely to leave, possibly for better opportunities or growth. The company should focus on retention strategies for younger age groups, such as career development, training, and flexible policies

**4.Experience (Total Working Years) vs. Attrition**

Most employees who left had less than 3 years of total working experience. The attrition rate declines sharply as experience increases.

The chart reveals that employees with less than 3 years of experience are more likely to leave the organization, with attrition being significantly higher in this group compared to experienced employees.

***Hypothesis*** : New Employees (<3 Years) Are More Likely to Quit

* Why this matters: Early-stage attrition wastes hiring and training costs.
* What we found:
  + A very high percentage of attrition came from employees with <3 years of tenure.
* ✅ Conclusion: New hires are more vulnerable to attrition.  
  🔹 Hypothesis Accepted

***Interpretation***: New or early-career employees are at a higher risk of leaving. More engagement and support for freshers could improve retention.

**5.Department vs. Attrition**

The Sales department experienced the highest attrition, followed by Research & Development.

The chart shows that employee attrition varies significantly across departments. The Sales department has the highest attrition, followed by Human Resources, while Research & Development retains most of its employees. This indicates a need for targeted retention strategies, especially in high-turnover areas like Sales.

***Hypothesis*** : Sales Department Has Higher Attrition

* Why this matters: Different departments may face unique pressures (e.g., targets).
* What we found:
  + The Sales department had the highest attrition rate.
  + R&D and HR had relatively lower attrition.
* Conclusion: Department plays a key role in attrition risk.  
  🔹 Hypothesis Accepted

***Interpretation:*** The **Sales department experiences the highest attrition**, possibly due to job pressure, performance-based stress, or lack of incentives. **Research & Development retains employees better**, which could be due to more stable roles, better work conditions, or career growth paths. **Human Resources falls in the middle**.

**6.Job Role vs. Attrition**

Sales Executives and Laboratory Technicians had the highest attrition rates. Roles like Managers showed the lowest.

The analysis of attrition by job role reveals that Sales Executives, Laboratory Technicians, and Research Scientists are the most likely to leave. These roles may need targeted interventions such as improved career paths, training, or incentives. In contrast, Human Resources and managerial roles experience lower attrition, indicating more job stability in these positions.

**Interpretation:** Lower-level or operational roles face more attrition. Providing better support or growth paths in these roles could help.

**7. Gender vs. Attrition**

Slightly higher attrition among female employees, though the gap was not very significant.

The attrition distribution by gender indicates that males account for a slightly higher number of exits compared to females. While the difference is not extreme, it may be worth exploring whether job roles, expectations, or work conditions differ by gender and contribute to this trend.

***Interpretation****:* May indicate a need to explore gender-specific challenges such as work-life balance, growth opportunities, or safety concerns.

**8. Education Field vs. Attrition**

Employees with education in Life Sciences and Medical fields showed higher attrition.

The analysis indicates that employees from Life Sciences and Medical backgrounds contribute the most to attrition. This could reflect role dissatisfaction, limited growth opportunities, or external market demand. Targeted retention strategies or better job fit analysis might be beneficial for these education groups.

***Interpretation****:* The organization may need to look into role alignment, expectations, or compensation in these education backgrounds.

**Statistical Analysis of Attrition Factors:**

**Regression Analysis to Predict Attrition**

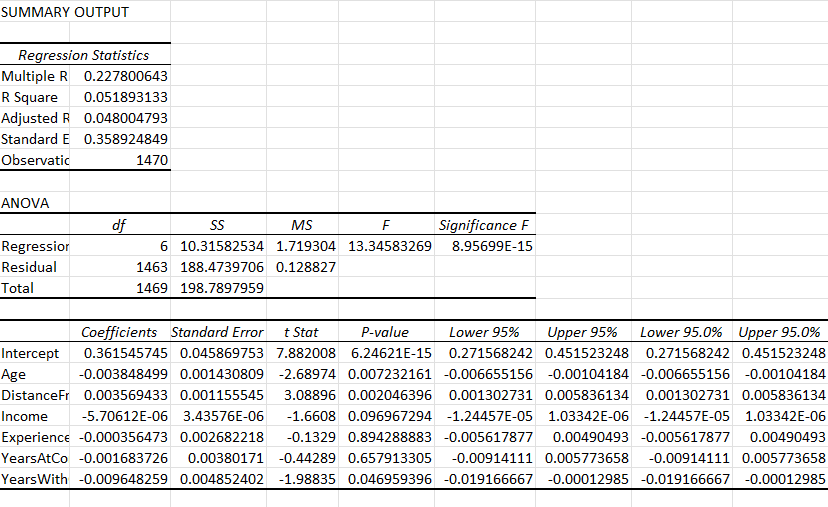
**Objective:**

The objective of this analysis is to understand how various employee-related factors such as Age, Distance from Home, Monthly Income, Total Working Years, Years at Company, and Years with Current Manager impact the likelihood of employee attrition. By building a linear regression model, we aim to identify which variables significantly influence attrition and provide actionable insights to help HR reduce employee turnover.

**Methodology:**

We performed a Linear Regression using Excel’s Data Analysis Toolpak. The target variable, Attrition, was converted to binary format (Yes = 1, No = 0). The following independent variables were selected for the model:

* Age
* Distance From Home
* Monthly Income
* Total Working Years (Experience)
* Years at Company
* Years with Current Manager



**Summary Output:**

* Multiple R (Correlation Coefficient): 0.228
* R Square: 0.052  
  This indicates that approximately 5.2% of the variation in attrition is explained by the model.
* Significant Variables (p-value < 0.05):
  + Age (p = 0.0072): Younger employees are more likely to leave.
  + Distance from Home (p = 0.0020): Employees living farther from the office tend to leave more.
  + Years with Current Manager (p = 0.047): Less time with the same manager is associated with higher attrition.
* Non-Significant Variables:
  + Monthly Income (p = 0.097)
  + Total Working Years (p = 0.894)
  + Years at Company (p = 0.658)

**Interpretation:**

The regression model suggests that younger employees, those living farther away, and those with shorter durations under their current manager are more likely to leave the organization. These insights can be valuable for designing targeted retention strategies. For example, improving manager-employee relationships, supporting younger employees, and offering flexible work options for those with long commutes can help reduce attrition.

Although the R-squared value is low (0.052), this is common in HR analytics, where employee behaviour is influenced by many unmeasured personal and organizational factors.

**Conclusion:**

While this regression model does not predict attrition with high accuracy, it provides useful directional insights. HR teams should focus on improving engagement with younger staff, offering flexibility to employees with long commutes, and fostering stable relationships between managers and their teams. These efforts can contribute to reducing attrition and improving organizational stability.

**Correlation Matrix of Numerical Variables**

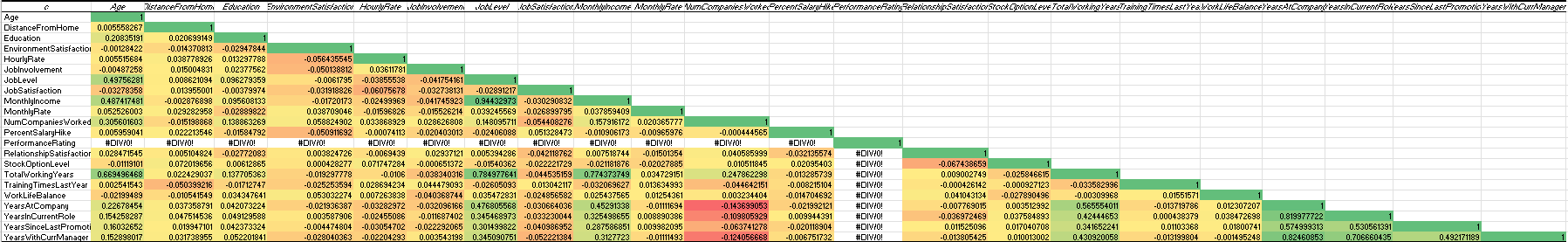
**Objective:**

To identify the strength and direction of relationships between key numerical variables and help understand which variables are potentially related to employee attrition or may cause multicollinearity in regression.

**Methodology:**

We calculated the Pearson correlation coefficients between all major numerical variables. The correlation values range between:

* +1: Perfect positive correlation
* 0: No correlation
* –1: Perfect negative correlation

****

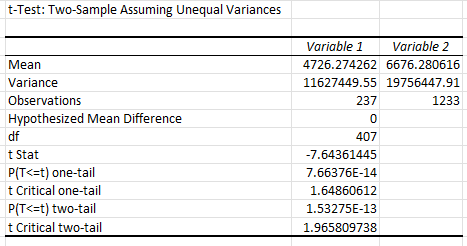
* Job Level and Monthly Income have a strong positive correlation (r = 0.944), confirming that senior employees earn more.
* Total Working Years and Monthly Income are also strongly correlated (r = 0.774), indicating experience leads to higher pay.
* Years With Current Manager has high correlation with Years At Company (r = 0.825) and Years In Current Role (r = 0.706), showing long-term tenure trends.
* Weak or no correlation exists between Distance From Home and other variables, suggesting commute distance doesn’t affect tenure or satisfaction significantly.
* Performance Rating showed data issues (#DIV/0!) and was excluded from interpretation.

**Conclusion:**  
Correlation analysis provided insight into which factors are interrelated, such as tenure, income, and experience, helping identify stable versus variable dimensions in employee profiles.

**t-Test – Monthly Income vs. Attrition:**

**Objective:**

To determine if there is a statistically significant difference in the monthly income of employees who left the organization compared to those who stayed.



To verify whether lower salaries contribute to employee attrition, a t-test assuming unequal variances was performed using Excel’s Data Analysis Toolpak. The test compared monthly income of employees who left the company (Attrition = Yes) against those who stayed.

* Mean Income (Attrition = Yes): ₹4726.27
* Mean Income (Attrition = No): ₹6676.28
* p-value (two-tailed): 1.53 × 10⁻¹³
* t Stat: -7.64

The p-value is well below the standard significance level (0.05), indicating that the difference in average monthly income is statistically significant.

**Conclusion:**

Employees who earn lower monthly incomes are more likely to leave the organization. This supports the hypothesis that “Lower Monthly Income Increases Attrition.”

**Key Findings:**

The exploratory data analysis and statistical testing provided several important insights into employee attrition. First, monthly income for employees who attrited was way less than for those who stayed. This was found through the two-sample t-test, which did confirm this difference was statistically significant (p-value < 0.001). Also, those employees living afar from the office were suggested to have a disposition to leave, evidenced by correlation and regression analysis. Then younger employees seemed to leave more commonly-a negative and statistically significant association existed between age and attrition. Another big issue concerning attrition was working with the direct manager, whereby employees with a longer tenure working under the same manager would be less likely to quit.

The analysis also indicated that employees with fewer years of total experience and shorter duration of company service were more prone to attrition. On the other hand, variables falling under satisfaction, such as job satisfaction and environment satisfaction, had weak relationships with attrition and were not significant within the regression model. Besides, the predictive linear regression model for attrition had a low value for R² (around 0.05), which shows that it actually explains only a small share of the variance in attrition. In view of this, it can be inferred that while a few variables surely affect attrition, linear regression may not be the fitting model to employ to yield predictions for an outcome such as attrition, being binary. On that premise, logistic regression may serve better for a future examination.

**Business Recommendations:**

Based on the detailed exploratory data analysis, statistical tests, and regression results, the following business recommendations are proposed to reduce employee attrition and improve workforce stability:

The analysis reveals that employees with lower monthly income are more likely to leave the organization. To address this, it is recommended that the company review and adjust its compensation policies, especially for lower-level positions, to remain competitive and equitable in the job market.

Additionally, younger employees and those with fewer years at the company show a higher tendency to leave. To retain such talent, the company should invest in career development programs, mentorship, and employee engagement initiatives tailored to early-career professionals.

Furthermore, a noticeable pattern was observed where employees living farther from the workplace are more likely to resign. This suggests the need to explore flexible work options such as remote work, hybrid models, or even location-based transfers to reduce commuting stress.

The analysis also suggests that a strong relationship with the immediate manager may contribute to employee retention. Therefore, it is advisable to enhance leadership development programs and focus on improving managerial communication, trust, and feedback mechanisms.

While the regression model's predictive power is limited, it still highlights key factors influencing attrition. The company should consider implementing regular employee satisfaction surveys and exit interviews to capture qualitative insights not available in numeric data. These proactive measures will enable the organization to better understand employee needs, address pain points, and ultimately reduce turnover.

**Conclusion**

In conclusion, this HR analytics project aimed to explore and understand the key factors influencing employee attrition using statistical and regression analysis. The analysis of various demographic, job-related, and performance-related attributes revealed that monthly income, age, years at the company, and distance from home have a significant impact on an employee’s decision to leave. The statistical tests, including correlation analysis and t-tests, confirmed meaningful differences between employees who left and those who stayed. Although the regression model had a modest explanatory power, it provided valuable insights into potential predictors of attrition. These findings highlight the importance of data-driven decision-making in human resource management. By acting on these insights, organizations can design better retention strategies, improve employee satisfaction, and ultimately reduce the cost and disruption associated with high employee turnover.

**POWER BI DASHBOARD:**

**📝 HR Insights Dashboard Summary**

This HR Insights Dashboard provides a comprehensive overview of employee attrition across various dimensions within the organization. Here are the key highlights:

**🔢 Overall Statistics**

* **Total Employees:** 961
* **Total Attrition Cases:** 133
* **Attrition Rate:** 13.84%
* **Average Age of Employees:** 37 years
* **Average Tenure:** 7 years
* **Average Monthly Income:** 6K

**🧑‍🎓 Attrition by Education Field**

* Most attrition comes from employees with a **Life Sciences** background (44.36%).
* **Technical and Medical** fields follow closely, indicating potential training or engagement gaps in those areas.

**📊 Attrition by Age Group**

* The highest attrition is in the **25–34 age group**, which may suggest job-hopping, dissatisfaction, or lack of growth opportunities.
* Attrition significantly drops after age 35.

**💼 Attrition by Job Role**

* **Research Scientists** show the highest attrition, followed by **Manufacturing Directors** and **Healthcare Representatives**.
* Lower attrition is observed among **Managers** and **Research Directors**.

**💰 Attrition by Monthly Income**

* Most attrition occurs in the **Below 3K** salary slab, suggesting compensation is a key factor.
* Attrition drops as income increases.

**🕒 Attrition Over Tenure**

* A spike in attrition appears in the **first few years**, implying early-stage disengagement or unmet expectations.
* Long-term employees show more stability.

**⚖️ Work-Life Balance vs Job Satisfaction**

* Employees with **lower job satisfaction (1 or 2)** and **poor work-life balance** show the highest attrition.
* Higher satisfaction levels correlate with reduced attrition.

**👥 Gender-wise Attrition**

* **Male** employees represent the majority of attrition cases, though a notable proportion of **female** attrition is also evident.

**📌 Conclusions & Recommendations**

1. **Focus retention strategies** on younger employees (25–34) and new hires.
2. **Improve compensation structures**, especially in the lower income bands.
3. **Address dissatisfaction** in high-attrition job roles like Research Scientist and Manufacturing Director.
4. **Enhance onboarding and early engagement programs** to reduce attrition within the first few years.
5. **Tailor support and development** programs for specific education fields and job roles.

