HR Data Analysis – Who's Likely to Quit?

Employee retention is critical for organizational success. Understanding which employees are at risk of quitting enables proactive engagement, reducing turnover costs and preserving talent.

In this case study focuses on predicting which employees are most likely to quit by using historical employee data. With this information, the company can take action early to keep the employees from leaving.

Here will first understand before going to the topic, **what is employee attrition?**

Employee attrition refers to the natural reduction in workforce that occurs when employees leave an organization and are not immediately replaced. This includes voluntary departures (resignations, retirements) and involuntary separations (layoffs, terminations). Understanding and predicting attrition patterns has become crucial for modern organizations.

1.INTRODUCTION

- **Explain the importance of predicting employee attrition.**

  Predicting employee attrition means finding out which workers might leave the company soon. This is very important for companies because if they know who might quit, they can try to keep those employees by solving their problems. This helps the company save money (because hiring new people costs a lot) and keeps the work running smoothly. It also makes sure that important jobs don't stay empty for too long.

- **Who benefits? (HR managers, team leads, organizational leadership)**

  When employees stay at a company for a long time, everyone benefits. But some people benefit the most, like HR managers, team leads, and organizational leadership.

  - HR managers-They are incharge of hiring and keeping employees happy. If less people leave the company, HR doesn't have to spend so much time and money finding and training new workers .This saves the company money and they can plan better strategies, like improving work space and culture, giving them promotion based on their work, or hiking salary.
  - Team leads-If they work closely with employees every day, people stay longer, teams become stronger. So they don't have to train for new people. If the employee is

unhappy or overworked in their work, then team lead need to support and engage with them. Then this makes their job easier and helps the team succeed.

- Organizational leadership-Here a company management and director will get benefits too. When attrition rate is low, the company runs more smoothly. It saves money, keeps skilled workers, and helps the company to grow higher. Happy employees also lead to better performance and customer service. Finally they get a clear view of employee satisfaction and can make better business decisions.
- Finance Teams-Finance departments benefit from more accurate budget forecasting. They can better predict costs related to hiring, training, and temporary staffing needs, leading to more precise financial planning.

- **What business problems does this analysis address?**

When employees leave a company can cause serious challenges for any organization. In this analysis helps to identify and solve several important business problems.

- When employees quit unexpectedly, companies have to spend a lot of time and money to find and train new people. This increases hiring costs and slows down work and remains the project pending. By predicting which employees are likely to leave, companies can take steps to keep them, which helps reduce these costs, and they can save their time too.
- If losing skilled and experienced workers can affect the overall performance of a team or department. Important projects may be delayed, and the quality of work might decrease. This analysis helps managers take action early, so they can support at-risk employees and improve job satisfaction.
- High employee turnover can hurt a company's reputation. If people keep leaving, others may think it is not a good place to work. This can make it hard to attract new talent. By using data to understand why people leave, the company can create a better work environment.
- This analysis helps businesses reduce hiring costs, improve employee satisfaction, maintain strong teams, and protect their reputation. It supports better decision-making and creates a more stable and productive workplace.

Example: Predicting attrition helps HR prioritize retention strategies, reduce hiring costs, and maintain workforce stability.

In this case study, why the predicting attrition is more important in advance means, it helpful for HR team, team lead, and companies director and management to supports better decision-making and creates a more stable and productive workplace to keep well skilled employees in their company, which is benefits for them. When HR knows who might quit, they can take early steps to stop it from happening. For example, they can talk to those employees, and try understanding their problems, and try to make their work life better.

This helps keep good employees from leaving, which is called a retention strategy. It also saves money. Hiring new employees costs a lot—companies have to spend on job advertisements, interviews, training, and more. If fewer people leave, then companies don't have to spend so much on hiring and it have it reputation and so many people will come and join the company, it will have a good name and fame in the cooperate world. Lastly, when employees stay longer, the workplace becomes more stable and the company can grow high. Teams work better together, and the company can achieve its goals more easily. It's better for everyone when the workforce is steady and not always changing. So, predicting attrition helps HR make smart plans to keep employees happy, save money, and keep the company strong.

2.DATA CLEANING AND PREPARATION

Data cleaning and preparation is the process of getting data ready for analysis. Before we can use data to find answers or make decisions, we need to make sure it is correct and organized. Sometimes we collect data's by mistake, or add by mistake. This may give us miss values, wrong numbers or repeated entries. So in the data cleaning we can fix this problems by correcting the error, filling the missing the values, or removing duplicates. After cleaning data, we will prepare a new data, this means putting it in the right format, give proper structure to it (like names to columns, organising the data) that makes easy to the work.

This step is very important because if the data is messy or wrong, the results of the analysis will also be wrong. Clean and well-prepared data helps us get accurate and useful results.

- **Check for missing or inconsistent data.**

  The first step is to identify if there are any missing values in the dataset. Missing data can lead to incorrect or biased results. Assessment of Missing Data Patterns, before deciding how to handle missing values, analysts must understand why data is missing. Missing data can be completely random, missing at random, or missing not at random. For example, if high-performing employees consistently skip satisfaction surveys, this creates a systematic bias that affects analysis results.

  Similarly, inconsistent data such as typos, incorrect formatting, or invalid entries should be corrected or removed. For example, department names might be recorded as "IT," "Information Technology," or "Tech" for the same department. Job titles may vary slightly, such as "Manager" versus "Mgr" or "Senior Developer" versus "Sr. Developer." These inconsistencies need standardization before analysis.

  Data Format Issues-Dates might be recorded in different formats, salary information may include or exclude currency symbols, and categorical variables might have different capitalization patterns. These formatting issues must be resolved to ensure accurate analysis.

Duplicate Records-Employee records might be duplicated due to system errors, employee rehiring, or data migration issues. Identifying and handling duplicates is essential for maintaining data integrity.

- **Handle categorical variables (e.g., Gender, Department).**

Here categorical variables means, columns that contain text values instead of numbers. "Gender" has values like 'Male' or 'Female'. "Department" has values like "Sales", "Research & Development", etc. Sometimes these values cannot be used directly in most machine learning models because ML models understand only numbers. So we need to convert them into a numerical format, this method is called "Encoding".

Also, we should explore the frequency of values in these columns to understand the distribution and spot any unusual or incorrect categories.

- **Explore distributions and outliers in numeric fields.**

Before analysing or modelling the data, it's important to understanding **Variable distributions**, here numeric variables in HR data often have unique distribution patterns. Salary data typically shows right-skewed distributions (e.g., if most employees have a very low or very high income), tenure might have bimodal distributions (new hires and long-term employees), and performance ratings often cluster around average values. Understanding these patterns (e.g., most employees who left may fall within a certain age or income range) helps in choosing appropriate analysis methods.

Identifying Outliers-Outliers (extreme values that may not be valid) in HR data can represent legitimate extreme cases or data entry errors. Values that is much higher or lower than the rest. For example, if most salaries are around ₹ 50,000 and one person earns ₹ 5,00,000 that's an outlier.

Why this is important in the data science because, outliers can affect the average. They can be errors or special cases. Understanding data distribution helps us choose the right model and pre-processing steps.

- **Create new features if needed (e.g., tenure buckets).**

Sometimes, creating new columns (features) from existing data can improve the analysis.

Creating Meaningful Categories Raw data often needs transformation into more meaningful categories for analysis. Continuous variables like age can be grouped into life stage categories (early career, mid-career, pre-retirement), and tenure can be categorized into retention risk periods (0-1 years, 1-3 years, 3-5 years, 5+ years).

Derived Variables New variables can be created by combining existing ones. For example, a work-life balance score might be derived from hours worked and overtime frequency. Promotion velocity can be calculated from tenure and position level changes. These derived variables often provide more insights than individual variables alone.

Time-Based Features Employee data has temporal aspects that can be captured through feature engineering. Seasonality in hiring, tenure milestones, and time since last promotion can all influence attrition likelihood. These time-based features help capture patterns that might not be apparent in static data.

Questions to explore:

During data exploration, several important questions should be examined to identify potential causes of attrition:

- **Which departments show higher attrition rates?**

  This helps understand if specific areas within the organization are losing more employees than others, possibly due to workload, management, or job roles.

- **Does overtime or low job satisfaction correlate with attrition?**

  Employees who frequently work overtime or report low job satisfaction may feel burned out or unhappy, making them more likely to leave.

- **Is there a relationship between salary and attrition?**

  Analysing whether employees with lower income are more likely to resign can help design better compensation strategies.

This comprehensive approach to data cleaning and preparation ensures that the employee attrition analysis is built on a solid foundation, leading to more accurate insights and better business decisions.

3.EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis is the process of examining and understanding data before building predictive models. In employee attrition studies, EDA helps discover patterns, relationships, and trends that explain why employees leave the organization. This analysis provides the foundation for making informed business decisions and developing effective retention strategies. It helps us understand the

data better and provides meaningful insights that guide decision-making and model building.

In this section, we focus on understanding how different features are related to employee attrition. We use various visual tools such as bar charts, boxplots, and heat maps to analyse the data.

Purpose of EDA in Attrition Analysis,

Understanding the Data Story, EDA helps tell the story hidden in employee data. It reveals which factors most strongly influence employee decisions to leave, identifies unexpected patterns, and provides insights that might not be obvious from looking at raw numbers alone.

Identifying Key Risk Factors, through systematic analysis, EDA identifies the most important factors that contribute to employee attrition. These insights help prioritize retention efforts and allocate resources to areas with the highest impact.

Validating Business Assumptions, Organizations often have assumptions about why employees leave. EDA helps confirm or challenge these assumptions with actual data, leading to more effective retention strategies based on evidence rather than intuition.

- ## Visualize attrition rates by department, job role, and gender.

  Understanding **who is leaving** and **from where** is one of the key goals of EDA. To do this, we explore how attrition varies across different departments, job roles, and gender categories.

  - **By Department:** Attrition rates for each department, making it easy to identify which departments have the highest and lowest turnover. This visualization helps HR teams focus their retention efforts on problem areas. We compare the attrition rate in departments like Sales, Research & Development, and Human Resources. This helps identify if any specific department is experiencing more employee exits.
  - **By Job Roles:** Some job roles may have higher stress levels, limited growth opportunities, or higher workloads. Visualizing attrition by job role helps spot such trends.
  - **By Gender:** Exploring gender-wise attrition gives insights into whether male or female employees are more likely to leave. This can help in identifying any gender-based workplace challenges or gaps.
  - **Bar Charts for Categorical Analysis:** Bar charts are the most effective way to compare attrition rates across different categories. They provide clear visual comparisons that are easy to understand and interpret.
  - **Box Plot for Numerical Variable Analysis:** Box plots show the distribution of numeric variables and help identify differences between employees who stay and those who leave.
  - **Salary Distribution:** Box plots comparing salary ranges between current employees and those who left can reveal compensation-related attrition patterns. If departing

employees cluster in certain salary ranges, this suggests targeted compensation issues.

- **Age Distribution:** Age-based attrition patterns often show distinct trends. Younger employees might leave for career growth opportunities, while older employees might leave for retirement or work-life balance reasons.
- **Tenure Analysis**: Box plots of tenure data reveal how long employees typically stay before leaving. This information helps identify critical retention periods when interventions are most needed.
- **Heatmaps for Correlation Analysis:** Heatmaps provide a visual representation of how different factors relate to each other and to attrition. They use colour coding to show the strength of relationships between variables.
- **Feature Correlation:** Heatmaps show which employee characteristics are most strongly correlated with attrition. Darker colours typically indicate stronger relationships, helping identify the most important predictive factors.
- **Department vs. Role Patterns:** Two-dimensional heatmaps can show attrition patterns across department and role combinations, revealing specific job categories with highest risk.

- **Analyse correlations between features and attrition.**

  To understand which features (factors) are most related to employee attrition, we analyse how each one behaves when compared to attrition status.

  For Example:

  - Does low job satisfaction lead to more attrition?
  - Is there a pattern between salary and attrition?
  - Do employees who frequently work overtime tend to leave more?
  - Does work-life balance have an impact on attrition?

We calculate **correlation values** and visualize them to see which variables are strongly linked to employees leaving. A positive or negative correlation helps in identifying important predictors.

- **Use bar charts, boxplots, and heatmaps for insights.**

- **Bar Charts:** Used to compare counts or percentages of attrition across categories like departments, job roles, and genders. Bar charts help identify where attrition is happening most often.
- **Box Plots:** Used to compare numeric values (like Monthly Income, Years at Company, Age) between employees who stayed and those who left. They show the spread and differences in values and highlight outliers.
- **Heatmaps:** Used to display correlation values between all features in a grid format. A heatmap shows which features are closely related to each other and to attrition, using colour shades for easier understanding.

These insights provide valuable input for the next stage — **building predictive models** and **creating retention strategies**.

4.KEY FINDINGS AND RECOMMENDATIONS

After analysing the data and exploring different patterns related to employee attrition, several important insights were discovered. These findings help in identifying the key risk factors that lead to employees leaving the organization. Based on these insights, actionable recommendations are provided for HR teams and organizational leadership.

Not all analytical findings are equally important for business decision-making. Key findings should be prioritized based on their potential impact on the organization, the strength of evidence supporting them, and the feasibility of implementing changes based on these insights.

High-Impact Findings: These discoveries affect large numbers of employees or have significant financial implications. For example, if analysis reveals that 60% of departing employees cite poor management as their primary reason for leaving, this finding affects the entire organization's leadership development strategy.

Statistically Significant Patterns: Findings should be based on statistically significant differences, not random variations. When reporting that certain employee groups have higher attrition rates, include confidence levels and sample sizes to establish credibility.

Actionable Insights: The most valuable findings are those that lead to specific business actions. Abstract insights about general employee satisfaction are less useful than specific findings about particular policies or practices that drive attrition.

- **Present top 3 insights related to attrition risks.**

Finding 1: Job Satisfaction and Overtime Impact

Discovery: Employees reporting low job satisfaction who also work regular overtime are 2.3 times more likely to leave the organization within 12 months compared to satisfied employees with normal working hours.

Finding 2: Departmental Attrition Disparities

Discovery: Sales and Information Technology departments show attrition rates 35% and 28% higher than the organizational average, respectively, with distinct underlying causes in each department.

Finding 3: Tenure-Based Attrition Patterns

Discovery: Employees in their second year of employment (12-24 months tenure) show the highest attrition risk, with 38% of all departures occurring during this period.

- **Example insights:**
  - "Employees with low job satisfaction and overtime are twice as likely to quit."

    From the analysis, it was found that employees who report low levels of job satisfaction and also work overtime regularly are at a **much higher risk of leaving** the organization. This suggests that dissatisfaction and overwork are strong indicators of potential attrition.
  - "Departments A and B have attrition rates 30% higher than average."

    The attrition rate in the **Sales** and **Human Resources** departments is significantly higher than in other departments. These two departments show a **30% greater employee turnover** than the overall company average, indicating that special attention is needed in these areas.

- **Recommendations for HR:**

Understanding HR Recommendations in Case Studies.

Based on the analysis and key findings from the employee attrition dataset, several areas of improvement have been identified. These recommendations are intended to help HR departments and organizational leadership reduce employee turnover, improve job satisfaction, and build a more stable and motivated workforce.

Recommendation1: Improve Work-Life Balance Initiatives for Overtime Employees.

The analysis clearly showed that employees who regularly work overtime and have low job satisfaction are much more likely to leave the company.

The Problem behind This Recommendation

When employees work overtime regularly, they experience stress, fatigue, and reduced personal time. This combination leads to job dissatisfaction and increases the likelihood of employees leaving the company. The analysis shows that employees working overtime with low job satisfaction are twice as likely to quit compared to other employees.

To address this: **Introduce flexible work policies** such as hybrid or remote options. **Encourage work-hour limits** to prevent burnout. **Promote the use of paid time off and wellness breaks**. **Conduct regular check-ins** with teams that are known to work extra hours.

Why This Recommendation Matters?

When employees feel overworked and underappreciated, their performance and mental health are affected. A healthy work-life balance can improve overall job satisfaction, engagement, and retention.

Recommendation2: Target Retention Programs in High Risk Departments.

Departments like Sales and Human Resources were found to have significantly higher attrition rates compared to other departments.

The Problem behind This Recommendation

The analysis shows that certain departments have attrition rates 30% higher than the company average. This means these departments have specific problems that make employees want to leave. Generic company-wide retention programs don't work because different departments have different issues.

To reduce turnover in these areas: **Conduct department-specific surveys** to understand employee concerns. **Offer mentoring and training opportunities** to support employee growth. **Recognize high-performing employees** and reward them accordingly. **Assign HR business partners** to monitor team satisfaction and provide timely support.

Why This Recommendation Matters?

Some departments may have unique challenges such as high workloads, customer pressure, or limited growth. Focused retention strategies in these departments can prevent mass exits and maintain stability.

Recomendation3. Offer salary reviews where income correlates with attrition risk.

The data showed a strong relationship between lower income and higher attrition risk. Employees with lower monthly income are more likely to look for better-paying opportunities elsewhere.

The Problem Behind This Recommendation

The analysis shows that employees with below-market salaries are more likely to leave the company. When employees feel underpaid, they become dissatisfied and start looking for better opportunities elsewhere. This problem is especially serious for high-performing employees who can easily find new jobs.

To address this: **Implement regular salary reviews** for employees, especially in the lower income group. **Provide performance-based bonuses** and incentives. **Ensure fair and transparent compensation policies** across all roles and departments. **Include financial well-being programs** such as budgeting workshops or financial planning support.

Why This Recommendation Matters?

Employees who feel underpaid are more likely to leave, even if other aspects of their job are satisfactory. Fair and competitive pay helps improve retention, motivation, and trust in the organization.

Implementing these HR strategies can significantly reduce employee attrition and improve workplace satisfaction. By focusing on employee well-being, fair compensation, and department-specific challenges, the company can build a more committed, productive, and long-lasting workforce.

5. CONCLUSION

Employee attrition is a major challenge for many organizations, leading to high recruitment costs, loss of skilled talent, and disruptions in team performance. This case study explored the patterns and risk factors behind employee attrition using data analysis and predictive techniques.

- **Summarize key takeaways.**
  - **Overtime and low job satisfaction** are strong indicators of attrition risk. Employees experiencing both are more likely to leave the company.
  - Departments like **Sales** and **Human Resources** show **higher attrition rates**, indicating the need for department-specific retention strategies.
  - Employees with **lower monthly income** are at greater risk of quitting, highlighting the importance of fair and competitive compensation.

These findings provide valuable insights into why employees leave and which groups are most at risk.

- **Explain how predictive insights can support strategic HR decisions to reduce turnover.**
  - By using data to **predict which employees are more likely to resign**, HR teams can take **proactive steps** instead of reacting after the employee has left. Predictive analytics helps in:
  - **Identifying high-risk employees early**, giving time to engage and support them.
  - **Designing targeted retention plans**, such as improving work-life balance, reviewing compensation, or offering career growth opportunities.
  - **Making data-driven decisions**, which are more accurate and reliable than assumptions.

With the help of predictive insights, organizations can **reduce turnover**, **improve employee satisfaction**, and **build a stronger, more stable workforce**.

Understanding and predicting employee attrition is not only about reducing numbers; it's about building a healthy workplace where employees feel valued, supported, and motivated to stay. Strategic use of data empowers HR to create positive changes that benefit both the employees and the organization.