**Exploratory Data Analysis on Employee Attrition Dataset**

## **Introduction**

Employee turnover, commonly known as attrition, poses a persistent challenge for organizations across industries, exerting significant implications on productivity, organizational culture, and bottom-line performance. As employees depart from an organization, they take with them valuable knowledge, skills, and experience, disrupting workflows, team dynamics, and overall operational efficiency. Moreover, high attrition rates can erode morale, diminish employee engagement, and create a sense of instability within the workforce, ultimately impacting organizational culture and employee satisfaction levels.

In today's competitive business landscape, where talent retention is paramount for sustaining organizational success, understanding the underlying factors driving employee attrition is imperative. By identifying these factors and their impacts, organizations can develop targeted strategies to mitigate attrition risks, foster a more stable and engaged workforce, and ultimately enhance organizational performance.

This project embarks on a journey to explore the complexities of employee attrition through data analysis and visualization, utilizing the Employee Attrition dataset sourced from IBM. Our research objectives are twofold: firstly, to uncover insights into the trends and patterns of employee turnover, and secondly, to identify actionable strategies for mitigating attrition risks and enhancing retention efforts within organizations.

By delving into the rich repository of employee data encompassing demographic attributes, job-related factors, and indicators of job satisfaction, aim to shed light on the multifaceted dynamics of employee turnover. Through rigorous analysis and visualization techniques, seek to uncover hidden patterns, correlations, and drivers of attrition, providing stakeholders with the knowledge needed to devise effective retention strategies.

In summary, this project represents a concerted effort to bridge the gap between data analytics and organizational strategy, leveraging the power of data-driven decision-making to tackle one of the most pressing issues facing modern workplaces. By uncovering the hidden drivers of attrition and proposing evidence-based solutions, aim to catalyse positive change and usher in a new era of retention-focused organizational culture.

## **Literature Review**

Previous research in organizational psychology and human resource management has extensively explored the dynamics of employee turnover and retention strategies. Studies such as "An Effectiveness of Human Resource Management Practices on Employee Retention in Institute of Higher Learning: - A Regression Analysis" by Eric Ng Chee Hong et al., have investigated the impact of various human resource management practices on employee retention. The study highlights factors such as empowerment, training, appraisal systems, and compensation, emphasizing their significant influence on retention rates. It underscores the importance of addressing employees' needs for training, fair compensation, and recognition to enhance retention strategies, particularly in Asian cultures.

Similarly, research like "Boosting employee retention through CSR: A configurational analysis" by Leemen Lee and Li-Fei Chen has explored the relationship between corporate social responsibility (CSR) initiatives and employee job satisfaction and retention intention. The study utilizes fuzzy-set qualitative comparative analysis (fsQCA) to uncover complex relationships between CSR, employee needs, and organizational outcomes. It emphasizes the positive impact of CSR practices coupled with high fulfilment of employee ERG needs on job satisfaction and retention intention, highlighting the importance of considering environmental and consumer-focused CSR factors for optimal employee outcomes.

Additionally, analyses of specific datasets, such as the "IBM Employee Attrition Analysis" by Shenghuan Yang and Md Tariqul Islam, have employed various techniques including correlation matrix, Random Forest, K-means Clustering, and binary logistic regression to identify key determinants of attrition. The study identifies factors such as monthly income, age, job level, job satisfaction, and frequency of travel as significant predictors of attrition. It underscores the need for further qualitative analysis to understand diverse employee intentions and recommends strategies to improve job satisfaction and prioritize employee well-being to enhance retention rates.

Furthermore, studies like "Predicting Employee Attrition Using Machine Learning Techniques" by Francesca Fallucchi et al., have explored the application of machine learning techniques in predicting employee attrition. The study analyses a dataset provided by IBM analytics and identifies factors such as monthly income, age, overtime, and distance from home as significant predictors of attrition. It highlights the potential for further research to improve attrition classifiers and offers insights for HR practitioners to implement proactive measures to mitigate employee turnover risks.

## **Methodology**

In this section, we delve into the specific steps involved in the analysis process, providing detailed insights into the data preprocessing techniques utilized to ensure data quality and consistency.

### **Data Collection:**

The dataset utilized in this analysis was sourced from data.world, specifically the IBM company employee dataset. This dataset encompasses a wide array of employee demographic attributes, job-related factors, and indicators of job satisfaction, providing a rich source of information for analysing employee attrition patterns and trends.

### **Data Preprocessing:**

Prior to analysis, the dataset underwent meticulous preprocessing to address any inconsistencies or missing values that could potentially impact the integrity of the analysis. Several techniques were employed:

* **Handling Missing Values**: Missing values within the dataset were identified and addressed through a combination of imputation and removal strategies, depending on the nature and extent of missingness. For numerical variables, missing values were imputed using appropriate measures such as mean, median, or mode, while for categorical variables, missing values were either imputed with the most frequent category or treated as a separate category, depending on the context.
* **Data Transformation**: In addition to handling missing values, certain data transformations were performed to enhance the interpretability and analysis of the dataset. This included creating new derived variables or calculation fields to capture relevant insights not explicitly present in the original dataset. For example, calculation fields were created to categorize employees into different age groups, tenure at the company, or commute distance categories, enabling more granular analysis and visualization.

### **Tools and Technologies:**

Tableau, a powerful data visualization tool, was leveraged for creating interactive dashboards to visualize and analyse the employee attrition data. Tableau's intuitive interface and robust features facilitated the creation of dynamic and insightful visualizations, allowing for deeper exploration of the data.

### **Exploratory Data Analysis (EDA):**

Exploratory Data Analysis (EDA) played a pivotal role in gaining insights into the characteristics and patterns of the dataset. Summary statistics were calculated for numerical variables, while various visualization techniques such as bar charts, donut charts, and lollipop charts were employed to explore relationships and trends among different variables. EDA provided a foundational understanding of the distribution of data and identified key factors influencing employee attrition.

### **Dashboard Design and Implementation:**

Interactive dashboards were meticulously designed and implemented in Tableau to facilitate exploratory analysis and storytelling. The dashboards were thoughtfully organized to present key insights and trends in employee attrition, allowing users to interactively explore the data and uncover meaningful patterns. Design elements such as colour coding, filters, and tooltips were strategically utilized to enhance usability and visual appeal.

### **Storytelling and Analysis:**

The methodology also encompassed storytelling and analysis, where key findings and insights derived from the data were presented in a narrative format. The analysis focused on identifying patterns, trends, and relationships within the data, with a particular emphasis on factors influencing employee attrition. The storytelling component provided context and interpretation to the visualizations, helping to convey the significance of the findings to stakeholders and decision-makers.

## **Visualization Interpretation**

In this section, we aim to provide clear and concise interpretations of the visualizations presented throughout the Exploratory Data Analysis (EDA) and Dashboard Design sections. By elucidating the key insights, trends, and patterns observed in the data, endeavour to establish a robust connection between the visualizations and the research objectives and problem statement.

### **Exploratory Data Analysis (EDA):**

During the EDA phase, various visualizations were employed to uncover insights into employee attrition trends and factors influencing attrition rates. Notably, the analysis revealed several compelling findings:

* **Attrition by Gender**: The visualization showcasing attrition rates by gender unveiled a nuanced picture, with males exhibiting a slightly higher attrition rate compared to females. This insight prompts further exploration into potential gender-specific factors contributing to attrition within the organization.
* **Attrition by Age Group**: The exploration of attrition rates across different age groups uncovered a notable trend, with younger employees, particularly those in the 25-34 age bracket, exhibiting higher attrition rates. This finding underscores the importance of targeted retention strategies tailored to address the needs and preferences of younger employees.
* **Attrition by Department**: Analysis of attrition rates across different departments highlighted significant disparities, with certain departments experiencing higher attrition rates compared to others. For instance, the Sales department exhibited a relatively higher attrition rate, indicating potential areas for improvement in retention strategies within this department.

### **Dashboard Design and Implementation**

The dashboard design and implementation phase involved creating interactive visualizations in Tableau to explore the Employee Attrition dataset and uncover insights into employee turnover trends and factors influencing attrition. Four distinct dashboards were developed, each focusing on different aspects of the dataset:

#### **Dashboard 1: Employee Attrition Overview**

This dashboard provides a comprehensive overview of employee attrition within the organization. Key performance indicators (KPIs) such as total employees, current employees, attrition rate, and total attrition are displayed prominently. Additionally, the dashboard features multiple worksheets showcasing attrition trends across various demographic and job-related factors, including gender, department, job role, age group, education level, and daily commute distance.

A close-up of a graph

Description automatically generated

The dashboard begins by highlighting the following KPIs: Total Employees (1470), Current Employees (1233), Attrition Rate (16.12%), and Total Attrition (237). These KPIs provide a snapshot of the workforce composition and attrition dynamics within the organization.

The dashboard features several visualizations to delve deeper into attrition patterns:

* **Attrition by Gender**: A lollipop chart illustrates the attrition rate by gender, revealing that 87 females and 150 males have experienced attrition.
* **Attrition by Department**: A donut chart showcases attrition rates across different departments. Attrition rates vary, with Human Resources experiencing 6 cases in both males and females (2.53%), Research & Development seeing 90 cases in males (37.97%) and 43 in females (18.14%), and Sales with 54 cases in males (22.78%) and 38 in females (16.03%).
* **Business Travel & Attrition**: A bar chart displays attrition rates based on business travel frequency. Employees who travel rarely have the highest attrition (150 cases), followed by around 70 cases for those who travel frequently, and approximately 10 for non-travellers.
* **Attrition by Job Role**: A bar chart reveals attrition rates across different job roles. Laboratory Technicians exhibit the highest attrition rate, followed by Sales Executives, while Research Directors have the lowest attrition rate.
* **Attrition by Age Group**: A dual-axis bar chart illustrates attrition rates across different age groups. The 25-34 age group shows the highest attrition rate, followed by the 35-44 age group, with the lowest attrition observed in the group over 55.
* **Attrition by Education**: A square chart displays attrition rates based on education level, indicating that employees with Bachelor's degrees have the highest attrition rate, followed by those with Master's degrees, while those with Doctorates have the lowest attrition rate.
* **Daily Commute Distance and Attrition**: A bar chart depicts attrition rates based on daily commute distance. The 1-7 miles range exhibits the highest number of attrition cases, while the 26-30 miles range has the least.

#### **Dashboard 2: Detailed Attrition Analysis**

This dashboard offers a comprehensive analysis of employee attrition, focusing on various demographic and job-related factors. Each visualization is designed as a color-coded bar chart, providing a clear representation of attrition rates across different categories.

A screenshot of a computer screen

Description automatically generated

* **Employee Statistics**: A single bar chart indicating key metrics: Total Employees (1470), Attritioned (237, representing a 16.12% attrition rate), and Current Employees (1233).
* **Employee Attrition and Gender:** It presents grouped bar charts, breaking down attrition rates by gender. Of the total 1470 employees, 882 are male and 588 are female. The attrition rate among males stands at 17%, with 150 employees affected, while for females, it is 15%, with 87 employees affected.
* **Attrition Rate by Job Satisfaction**: Visualization provides insights into attrition rates based on job satisfaction ratings. With a total of 1470 employees, ratings 1, 2, 3, and 4 exhibit attrition rates of 23%, 16%, 17%, and 11% respectively.
* **Employee and Attrition Rate by Marital Status**: This chart depicts attrition rates across marital statuses. Divorced employees (327) experience a 10% attrition rate, while for married employees (673) it is 12%, and for single employees (470) it is 26%.
* **Employee and Attrition by Department**: Visualization displays attrition rates across different departments. Human Resources (63 employees) witnesses a 19% attrition rate, Research & Development (961 employees) has a 14% attrition rate, and Sales (446 employees) shows a 21% attrition rate.
* **Employee and Attrition by Overtime**: Bar chart highlights attrition rates based on overtime. Among 1054 employees who work overtime, the attrition rate is 10%, whereas among 416 employees who do not work overtime, the attrition rate jumps to 31%.
* **Employee and Attrition by Years at Company**: In this visualization, attrition rates are analysed based on tenure. Employees with less than 5 years (580 employees) experience a 24% attrition rate, while those with 5-10 years (524 employees) have an 11% attrition rate, and so on.
* **Employee and Attrition by Job Role**: This bar chart breaks down attrition rates by job roles. Sales Reps (83 employees) exhibit the highest attrition rate at 40%, while Research Directors (80 employees) have the lowest at 2%.

#### **Dashboard 3: Employee Distribution Analysis**

Employee Distribution Analysis dashboard provides valuable insights into the distribution of employees across various demographic and job-related categories within the organization. Through a series of visually informative bar charts, this dashboard offers a comprehensive overview of the workforce composition.

* **Employee Distribution by Gender**: This visualisation revealing a total workforce of 1470 employees, with 882 males and 588 females. This breakdown provides a clear understanding of the gender distribution within the organization.
* **Employee Distribution by Marital Status**: showcases the marital status distribution among employees. The bar chart illustrates that 673 employees are married, 470 are single, and 327 are divorced, offering insights into the marital status diversity within the workforce.
* **Employee Distribution by Age Group**: highlighting the distribution of employees across different age brackets. Notably, most employees fall within the 25-34 age group, followed by the 35-44 age group, demonstrating the age diversity present in the organization.

A screenshot of a graph

Description automatically generated

* **Employee Distribution by Education:** provides insights into the educational background of employees. With 572 employees holding bachelor’s degrees, 398 with Master's degrees, and 282 with College degrees, the organization exhibits a diverse range of educational qualifications among its workforce.
* **Employee Distribution by Education Field**: sheds light on the distribution of employees across various fields of study. The visualization indicates that most employees are from Life Sciences and Medical backgrounds, with smaller proportions from Marketing, Technical, and Human Resources fields.
* **Employee Distribution by Department**: showcases the distribution of employees across different departments within the organization. The largest department is Research and Development, followed by Sales and Human Resources, providing a clear overview of the departmental composition.

#### **Dashboard 4: Average Monthly Salary Distribution**

Average Monthly Salary Distribution dashboard provides insights into the distribution of average monthly salaries across various demographic and job-related categories within the organization. Through a series of bar charts, this dashboard offers a comprehensive overview of salary distribution trends.

* **Average Salary Distribution by Gender**: On average, females earn $6,687 per month, while males earn $6,381 per month. This comparison provides insights into gender-based salary discrepancies within the organization.
* **Average Salary Distribution by Age Group**: Highlights the average monthly salaries across different age brackets. Employees aged 45-55 have the highest average monthly salary of $10,449, while those under 25 have the lowest average monthly salary of $2,659.
* **Average Salary Distribution by Education**: Sheds light on the relationship between educational attainment and average monthly salary. Employees with Doctorates have the highest average monthly salary of $8,278, followed by those with master’s degrees and bachelor’s degrees.

A screenshot of a graph

Description automatically generated

* **Average Salary Distribution by Education Field**: Showcases the average monthly salaries across various fields of study. Employees in Marketing have the highest average monthly salary of $7,349, while those in technical fields have the lowest average monthly salary of $5,758.
* **Average Monthly Salary Distribution by Department:** provides insights into salary disparities across different departments within the organization. The Sales department has the highest average monthly salary of $6,959, followed by Human Resources and Research & Development.
* **Average Monthly Salary Distribution by Job Role**: Illustrates the average monthly salaries across different job roles. Managers have the highest average monthly salary of $17,182, while Sales Representatives have the lowest average monthly salary of $2,626.

## **Storytelling and Analysis**

In this section, we embark on a narrative journey through the data, unravelling the stories woven within the numbers and visualizations. Through a comprehensive analysis of key insights derived from the exploration of employee attrition, distribution, and salary trends, we aim to uncover actionable insights that can inform strategic decision-making and drive positive change within the organization.

### **Attrition Trends and Drivers**

Our analysis of employee attrition revealed compelling insights into the underlying trends and drivers shaping attrition rates within the organization. Notably, we observed a diverse array of factors influencing attrition, ranging from demographic characteristics to job-related variables. Key findings include:

* **Gender Disparities**: While males exhibited a slightly higher attrition rate compared to females, further investigation is warranted to understand the underlying reasons driving these disparities. Gender-specific factors such as career advancement opportunities, work-life balance, and organization
* al culture may play pivotal roles in shaping attrition trends.
* **Age Dynamics**: Younger employees, particularly those in the 25-34 age bracket, emerged as a vulnerable demographic group with higher attrition rates. This finding underscores the importance of tailored retention strategies to address the unique needs and aspirations of younger talent within the organization.
* **Departmental Variances**: Analysis of attrition rates across different departments revealed significant disparities, with certain departments experiencing higher attrition rates than others. Factors such as job satisfaction, career advancement prospects, and work environment dynamics may contribute to these departmental variances, necessitating targeted interventions to improve retention within specific areas.

### **Workforce Distribution Insights**

The exploration of employee distribution across various dimensions offered valuable insights into the composition of the organization's workforce. Key observations include:

* **Gender Representation**: While the workforce exhibited a relatively balanced gender distribution, further efforts may be needed to ensure gender equity across all levels of the organization. Addressing gender disparities in leadership roles and promoting diversity and inclusion initiatives can foster a more inclusive workplace culture.
* **Age Diversity**: The workforce demonstrated age diversity, with employees spanning across different age brackets. Embracing age diversity not only enriches the organizational culture but also fosters knowledge sharing, mentorship opportunities, and cross-generational collaboration.

### **Salary Disparities and Factors Influencing Pay**

Analysis of average monthly salary distribution shed light on disparities in compensation across various demographic and job-related categories. Key insights include:

* **Gender Pay Gap**: While males and females exhibited comparable average monthly salaries, further examination is warranted to identify and address any underlying gender-based pay disparities. Implementing equitable compensation practices and conducting regular pay equity audits can help mitigate gender pay gaps within the organization.
* **Education and Job Role Influence**: Employees with higher educational qualifications and those occupying senior positions tend to command higher average monthly salaries. Recognizing and rewarding employees based on their skills, experience, and contributions can foster a culture of fairness and meritocracy within the organization.

### **Strategic Implications and Recommendations**

The insights gleaned from our analysis have significant strategic implications for the organization, paving the way for targeted interventions and initiatives aimed at enhancing employee satisfaction, retention, and organizational performance. Key recommendations include:

* **Tailored Retention Strategies**: Develop tailored retention strategies targeting vulnerable demographic groups such as younger employees and those in high-attrition departments. These strategies should focus on addressing key drivers of attrition, such as career development opportunities, work-life balance, and recognition programs.
* **Diversity and Inclusion Initiatives**: Implement diversity and inclusion initiatives to foster a more inclusive workplace culture, where employees from diverse backgrounds feel valued, respected, and empowered to contribute their best. Promoting diversity in leadership roles and providing unconscious bias training can help create a more equitable and inclusive work environment.
* **Equitable Compensation Practices**: Review and revise compensation practices to ensure fairness and equity across all demographic and job-related categories. Conduct regular pay equity audits, address any gender-based pay disparities, and establish transparent salary structures to promote fairness and transparency.

## **Conclusion**

In conclusion, our exploration of the Employee Attrition dataset has provided valuable insights that can inform strategic decision-making and drive positive change within the organization. By leveraging the power of data analytics, we have uncovered actionable insights into employee attrition, distribution, and salary trends, offering a deeper understanding of the factors shaping workforce dynamics.

Our analysis has underscored the importance of addressing attrition challenges through targeted interventions and initiatives aimed at enhancing employee satisfaction and retention. Key findings highlight the need for tailored retention strategies, diversity and inclusion initiatives, and equitable compensation practices to create a supportive and inclusive work environment where employees feel valued, engaged, and empowered.

Furthermore, our examination of workforce distribution has emphasized the significance of embracing diversity and fostering an inclusive workplace culture. By promoting age diversity, gender equity, and educational inclusivity, the organization can harness the full potential of its workforce and drive innovation, collaboration, and organizational excellence.

As we look to the future, it is essential for the organization to prioritize ongoing monitoring and analysis of workforce dynamics to adapt to changing trends and challenges effectively. By embracing data-driven decision-making and leveraging advanced analytics tools and techniques, the organization can gain deeper insights into employee behaviour and preferences, enabling proactive interventions and strategic planning.

In summary, analysis has provided valuable insights that can empower the organization to optimize workforce management strategies, enhance employee satisfaction and retention, and ultimately, thrive in today's dynamic and competitive business landscape. Through a commitment to data-driven approaches and a focus on creating a culture of inclusion and empowerment, the organization can position itself as a leader in employee satisfaction and retention, driving sustainable growth and success in the years to come.

## **Appendix**

### **Multivariate data exploration visualization**

In the appendix section, we present additional insights derived from multivariate charts, providing a more nuanced understanding of workforce demographics and compensation structures within the organization.

**Employee statistics by gender and education**

A screenshot of a computer

Description automatically generated

The employee statistics by education field and gender reveal notable patterns across different educational backgrounds. In the life sciences and medical fields, we observe a higher representation of male employees compared to females, with 366 and 274 males, respectively, compared to 240 and 190 females. Similarly, in technical and marketing fields, male representation is higher, with 80 males compared to 52 females in technical roles and 90 males compared to 69 females in marketing roles. However, in human resources, the gender disparity is more pronounced, with only 19 male employees compared to 8 females. These findings underscore the need for gender diversity initiatives, particularly in fields where male dominance is prevalent.

**Average salary by Job role and gender**

A graph with blue and white bars

Description automatically generated

The analysis of average salary by job role and gender unveils intriguing insights into compensation differentials across various positions. Among managers, while the average salary is slightly higher for females ($17,409) compared to males ($16,915), the reverse trend is observed among research directors, where males earn an average salary of $15,144 compared to $16,658 for females. Interestingly, in roles such as healthcare representatives and manufacturing directors, female employees tend to earn marginally higher salaries than their male counterparts. However, in sales executive and human resource roles, male employees earn higher average salaries than females. These variations highlight the complex interplay between gender, job roles, and compensation, suggesting the need for equitable pay practices and gender-neutral salary structures to ensure fairness and transparency.

**Average salary by education and gender**

A blue and white bar chart

Description automatically generated

The examination of average salary by education and gender offers insights into the relationship between educational attainment and compensation levels. While male employees with doctorates command the highest average salary ($9,241), females with doctorates earn substantially lower ($7,463) on average. Similarly, among employees with bachelor's degrees, males earn an average salary of $6,811, while females earn $6,313 on average. Interestingly, females with master's degrees ($6,804) earn slightly higher average salaries compared to males ($6,878). These findings underscore the importance of addressing gender-based pay disparities and promoting educational inclusivity to ensure equitable compensation practices within the organization.

## **References**

1. Employee Attrition Dataset: (<https://data.world/aaizemberg/hr-employee-attrition>)
2. [An Effectiveness of Human Resource Management Practices on Employee Retention in Institute of Higher learning: - A Regression Analysis](https://www.researchgate.net/profile/Charles-Ramendran/publication/267725979_An_Effectiveness_of_Human_Resource_Management_Practices_on_Employee_Retention_in_Institute_of_Higher_learning_-A_Regression_Analysis/links/54d0dacf0cf298d6566937a0/An-Effectiveness-of-Human-Resource-Management-Practices-on-Employee-Retention-in-Institute-of-Higher-learning-A-Regression-Analysis.pdf)
3. [Boosting employee retention through CSR: A configurational analysis](https://onlinelibrary.wiley.com/doi/full/10.1002/csr.1511?saml_referrer)
4. [IBM Employee Attrition Analysis](https://arxiv.org/pdf/2012.01286.pdf)
5. [Predicting Employee Attrition Using Machine Learning Techniques](https://www.mdpi.com/2073-431X/9/4/86)