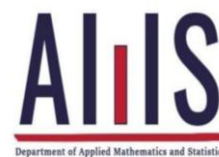Institute of Technology of Cambodia

Department of Applied Mathematics and Statistics

Report for Data Visualization

Topic: "Employee Attrition Analysis"

Group: I4-AMS-C7

Name of Students	ID
PHOEUN RAJAME	e20210057
PHOEURN KIMHOR	e20210823
SEK SOMUNINEATH	e20210929

Lecturer: Dr. PHAUK Sokkhey (Course & TP)

Academic Year 2024-2025

TABLE OF CONTENTS

1. Introduction

1.1.   What is employee attrition?

Employee attrition refers to the gradual reduction in a company's workforce due to employees leaving the organization voluntarily or involuntarily, such as through resignation, retirement, termination, or death. Unlike turnover, where positions are often filled promptly, attrition typically implies that the role is not immediately replaced. In the long run, unchecked attrition can lead to workforce shrinkage and loss of organizational capacity.

1.2.   Why does attrition matter?

Attrition is a significant issue for organizations because it affects both operational efficiency and financial performance. When valuable employees leave, they take with them critical knowledge, skills, and experience that are not easily replaced. High attrition rates can lead to a breakdown in team dynamics, reduced employee morale, increased workloads on remaining staff, and delays in project delivery. Moreover, a consistent pattern of employee exits can damage a company's employer brand, making it harder to attract and retain top talent.

2. Problem Statement

High employee turnover poses a critical and costly challenge for organizations and the U.S. economy. The continuous loss of talent results in substantial financial burdens, with replacement costs averaging one-third [1] to one-half of an employee's annual salary, and significantly higher for skilled and senior roles, potentially reaching up to two times the salary. Beyond these direct expenses, elevated attrition rates severely impact organizational profitability, leading to companies being approximately 23% less profitable than those with strong retention. At a national level, employee turnover is projected to cause an economic loss exceeding $400 billion annually by 2030 [2], potentially reaching $1 trillion when accounting for all hidden costs such as lost productivity, decreased morale, and operational disruptions. This pervasive issue undermines organizational stability, impedes growth, and

represents a significant drain on economic resources, necessitating effective strategies to mitigate its adverse effects.

3. Objective

The objective of this project is to analyze the IBM Employee Attrition dataset to uncover patterns and insights related to employee turnover. Specifically, the analysis aims to:

- Investigate how job roles, educational backgrounds (EducationField), and gender influence attrition rates.
- Identify key factors that contribute to employee resignation or retention.
- Provide data-driven insights that can help organizations develop strategies to improve employee satisfaction and reduce attrition.

4. Dataset Overview

The data utilized for this project is a fictional dataset meticulously created by IBM data scientists, publicly available on Kaggle. This dataset is designed to simulate realistic employee attributes and behaviors relevant to attrition analysis, comprising 1435 rows and 35 columns, thereby providing a rich and comprehensive set of features for detailed investigation.

Within this dataset, a variety of categorical variables are present, many of which are represented by numerical encodings corresponding to specific qualitative levels. For clarity and accurate interpretation, these mappings are crucial:

- 'Education' is categorized from 1 ('Below College') to 5 ('Doctor')
- 'EnvironmentSatisfaction' ranges from 1 ('Low') to 4 ('Very High')
- 'JobInvolvement' is similarly scaled from 1 ('Low') to 4 ('Very High')
- 'JobSatisfaction' also spans from 1 ('Low') to 4 ('Very High')
- 'PerformanceRating' is mapped from 1 ('Low') to 4 ('Outstanding')
- 'RelationshipSatisfaction' progresses from 1 ('Low') to 4 ('Very High')
- 'WorkLifeBalance' is categorized from 1 ('Bad') to 4 ('Best')

This detailed structure, coupled with these explicit categorical mappings, enables a thorough exploration of employee characteristics and their potential influence on attrition patterns.
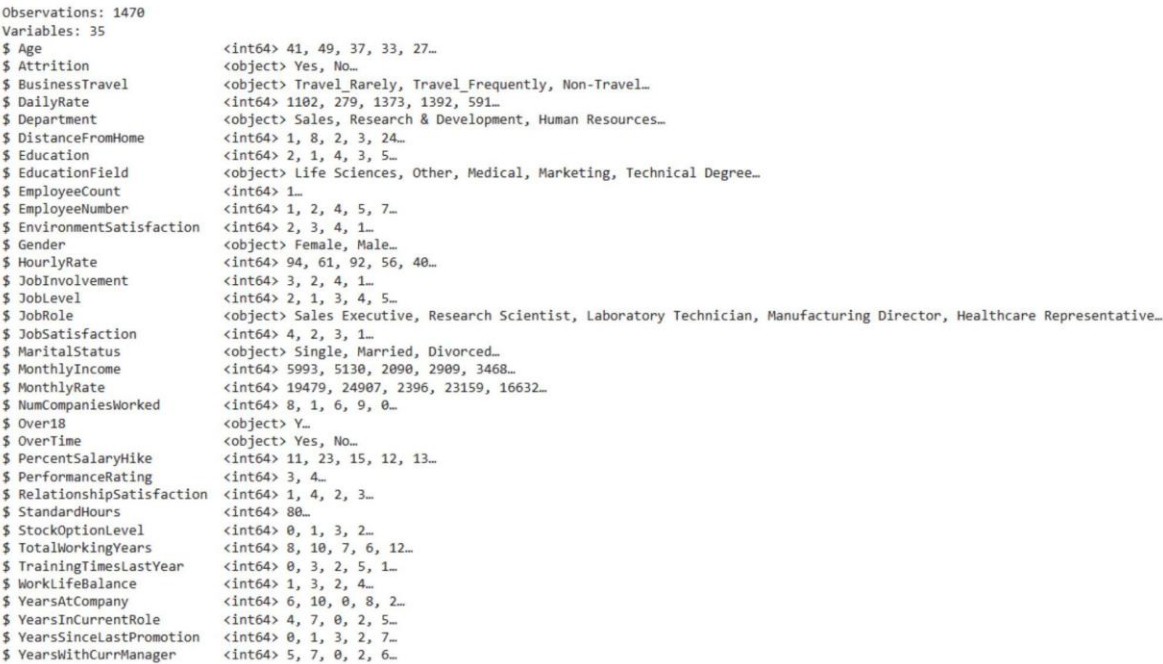
```
Observations: 1470
Variables: 35
$ Age                        <int64> 41, 49, 37, 33, 27…
$ Attrition                  <object> Yes, No…
$ BusinessTravel             <object> Travel_Rarely, Travel_Frequently, Non-Travel…
$ DailyRate                  <int64> 1102, 279, 1373, 1392, 591…
$ Department                 <object> Sales, Research & Development, Human Resources…
$ DistanceFromHome           <int64> 1, 8, 2, 3, 24…
$ Education                  <int64> 2, 1, 4, 3, 5…
$ EducationField             <object> Life Sciences, Other, Medical, Marketing, Technical Degree…
$ EmployeeCount              <int64> 1…
$ EmployeeNumber             <int64> 1, 2, 4, 5, 7…
$ EnvironmentSatisfaction    <int64> 2, 3, 4, 1…
$ Gender                     <object> Female, Male…
$ HourlyRate                 <int64> 94, 61, 92, 56, 40…
$ JobInvolvement             <int64> 3, 2, 4, 1…
$ JobLevel                   <int64> 2, 1, 3, 4, 5…
$ JobRole                    <object> Sales Executive, Research Scientist, Laboratory Technician, Manufacturing Director, Healthcare Representative…
$ JobSatisfaction            <int64> 4, 2, 3, 1…
$ MaritalStatus              <object> Single, Married, Divorced…
$ MonthlyIncome              <int64> 5993, 5130, 2090, 2909, 3468…
$ MonthlyRate                <int64> 19479, 24907, 2396, 23159, 16632…
$ NumCompaniesWorked         <int64> 8, 1, 6, 9, 0…
$ Over18                     <object> Y…
$ OverTime                   <object> Yes, No…
$ PercentSalaryHike          <int64> 11, 23, 15, 12, 13…
$ PerformanceRating          <int64> 3, 4…
$ RelationshipSatisfaction   <int64> 1, 4, 2, 3…
$ StandardHours              <int64> 80…
$ StockOptionLevel           <int64> 0, 1, 3, 2…
$ TotalWorkingYears          <int64> 8, 10, 7, 6, 12…
$ TrainingTimesLastYear      <int64> 0, 3, 2, 5, 1…
$ WorkLifeBalance            <int64> 1, 3, 2, 4…
$ YearsAtCompany             <int64> 6, 10, 0, 8, 2…
$ YearsInCurrentRole         <int64> 4, 7, 0, 2, 5…
$ YearsSinceLastPromotion    <int64> 0, 1, 3, 2, 7…
$ YearsWithCurrManager       <int64> 5, 7, 0, 2, 6…
```

Figure 4.1: Dataset Structure and Variable Types

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction
22	No	Non-Travel	1123	Research & Development	16	2	Medical	1	22	4
39	Yes	Travel_Rarely	895	Sales	5	3	Technical Degree	1	42	4
35	No	Travel_Rarely	464	Research & Development	4	2	Other	1	53	3
27	No	Travel_Rarely	1240	Research & Development	2	4	Life Sciences	1	54	4
48	Yes	Travel_Rarely	626	Research & Development	1	2	Life Sciences	1	64	1
45	No	Travel_Rarely	1339	Research & Development	7	3	Life Sciences	1	86	2
36	Yes	Travel_Rarely	318	Research & Development	9	3	Medical	1	90	4
36	No	Travel_Rarely	132	Research & Development	6	3	Life Sciences	1	97	2
32	No	Travel_Rarely	827	Research & Development	1	1	Life Sciences	1	134	4
37	No	Non-Travel	1040	Research & Development	2	2	Life Sciences	1	139	3
34	No	Travel_Rarely	1031	Research & Development	6	4	Life Sciences	1	151	3
19	No	Travel_Rarely	1181	Research & Development	3	1	Medical	1	201	2
51	No	Travel_Rarely	1169	Research & Development	7	4	Medical	1	211	2
38	No	Travel_Rarely	1261	Research & Development	2	4	Life Sciences	1	271	4
38	Yes	Travel_Rarely	1180	Research & Development	29	1	Medical	1	282	2
41	No	Travel_Rarely	896	Sales	6	3	Life Sciences	1	298	4
59	No	Travel_Rarely	142	Research & Development	3	3	Life Sciences	1	309	3
41	No	Travel_Rarely	1411	Research & Development	19	2	Life Sciences	1	334	3
37	Yes	Travel_Frequently	504	Research & Development	10	3	Medical	1	342	1
51	No	Travel_Rarely	833	Research & Development	1	3	Life Sciences	1	353	3

Figure 4.2: Sample Employee Attrition Dataset

5. Data Visualization

In this section explores the root causes of attrition through visual analysis drawn from multiple interactive dashboards. The visual data helps isolate high-risk segments and contributing factors, providing an evidence-based foundation for targeted retention strategies.

5.1. Attrition Overview

Figure 5.1: Attrition Overview Dashboard

- **Overall Attrition Rate:** The company has an attrition rate of **16.12%**, with **237 employees** having left. This is a significant percentage and indicates a need for intervention.

- **Average Years at Company for Attrition:** Employees who leave ("Yes" attrition) have an average tenure of **5.13 years**, significantly lower than those who stay ("No" attrition) at **7.37 years**. This suggests a retention challenge within the first few years of employment.

- **Attrition by Department:**

  o **Research & Development** has the highest absolute number of attritions. While it also has the largest employee base, the sheer volume is noteworthy.

  o **Sales** has a moderate level of attrition.

  o **Human Resources** has the lowest attrition, both in absolute numbers and proportionally.

- **Attrition by Education:**

  o Education levels 3 and 4 (likely Bachelor's and Master's/Ph.D.) show the highest absolute numbers of attrition.

- **Attrition by Job Role:**

o **Sales Executive** and **Research Scientist** roles show the highest absolute numbers of attrition.

o **Laboratory Technician**, **Manufacturing Director**, and **Healthcare Representative** also contribute significantly.

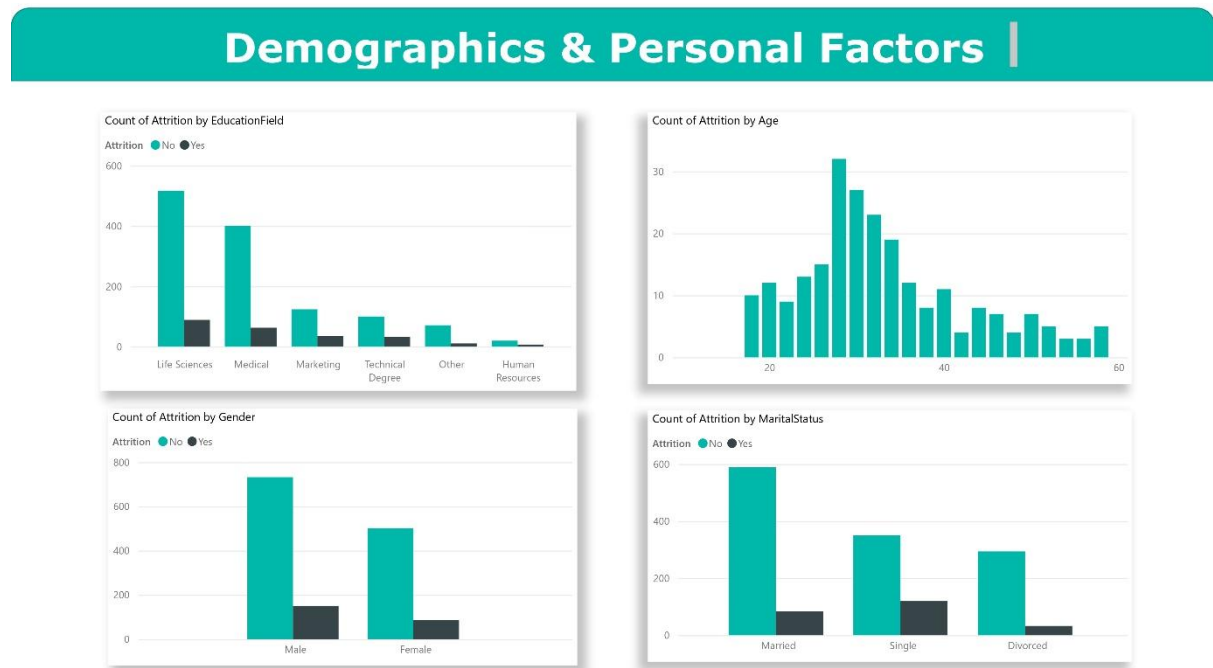## 5.1. Demographics & Personal Factors

Figure 5.2: Demographics & Personal Factors Dashboard

Attrition by Education Field:

- **Life Sciences** and **Medical** fields show the highest absolute numbers of attrition.

Attrition by Age:

- The highest concentration of attrition occurs in the **early to mid-30s** (roughly 30-35 years old). Attrition generally decreases with age.

Attrition by Gender:

- **Males** account for a higher absolute number of attritions than Females. (Note: To understand if this is a *rate* issue, we'd need to know the male/female distribution in the total workforce).

Attrition by Marital Status:

- **Single** employees have a significantly higher absolute number of attritions compared to Married or Divorced individuals, relative to their population size. This is a very strong indicator.
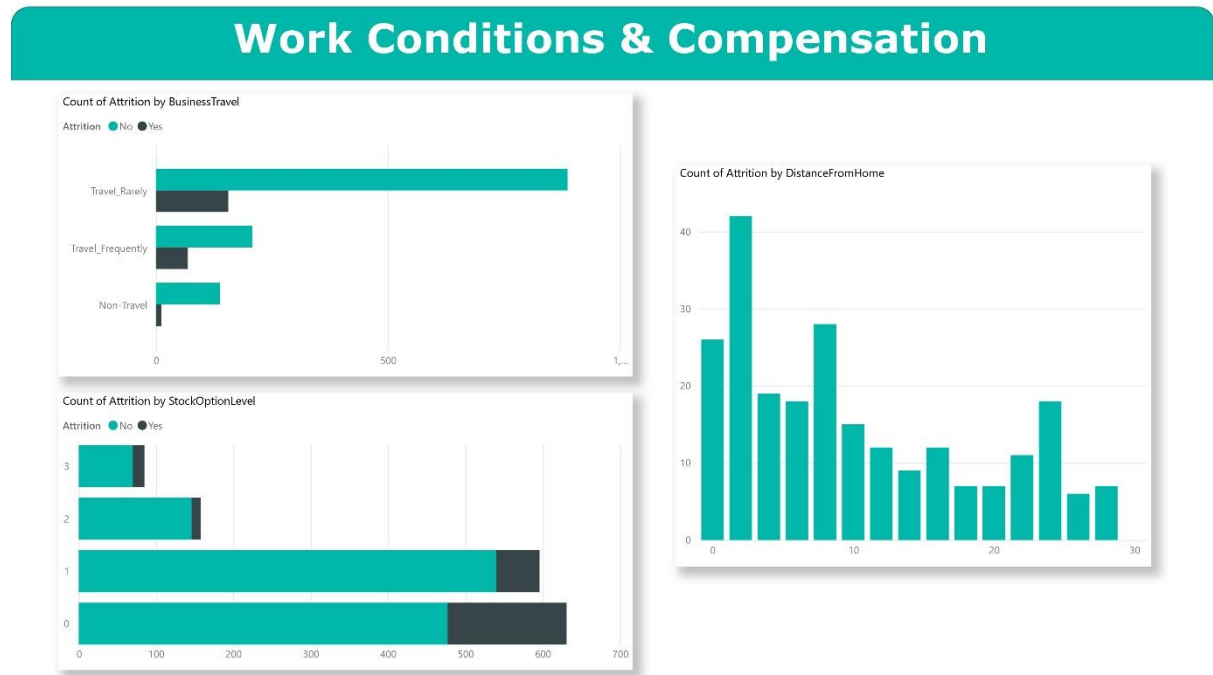
5.2.   Work Conditions & Compensation

Figure 5.3: Work Conditions & Compensation Dashboard

Attrition by Business Travel:

- Employees who **"Travel Rarely"** show the highest absolute number of attritions.

- **"Travel Frequently"** and **"Non-Travel"** roles have significantly lower absolute attrition.

Attrition by Distance From Home:

- A high number of attritions occur for employees living very **close to home (0-5 miles)**, and also a notable peak around **10 miles**, and then another smaller peak around **25 miles**.

Attrition by Stock Option Level:

- Employees with **Stock Option Level 0** (no stock options) have the highest absolute number of attritions, and a higher proportion of attrition compared to those with stock options.

- Employees with **Stock Option Level 1** also have significant attrition, but the "No Attrition" group is much larger.
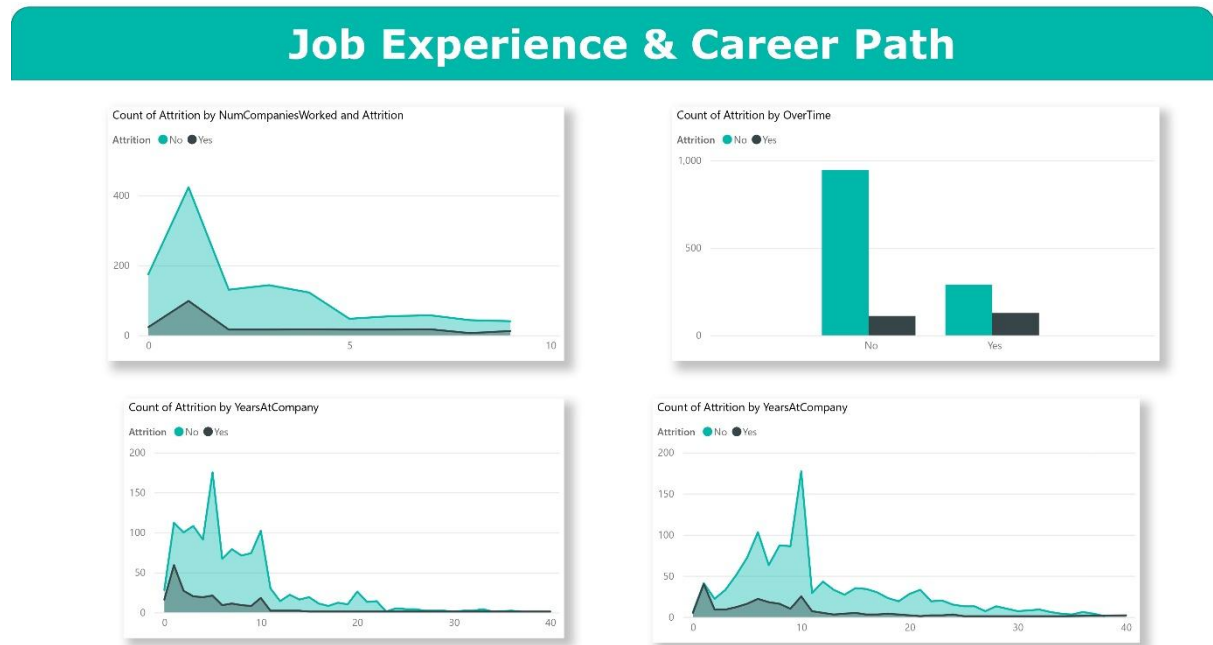
5.3. Job Experience & Career Path

Figure 5.4: Job Experience & Career Path Dashboard

Attrition by Number of Companies Worked:

- Employees who have worked for **1 or 2 companies** previously show the highest attrition. As the number of previous companies increases, attrition generally decreases.

Attrition by Overtime:

- Employees who **work Overtime** have a disproportionately higher attrition rate compared to those who do not. The number of attritions for those working overtime is substantial.

Attrition by Years at Company:

- The highest attrition occurs within the **first 1-3 years** and then again around the **7-10 year** mark. This reinforces the "Average Years at Company" insight from the overview.

Attrition by Total Working Years:

- Attrition is highest for employees with **fewer total working years** (e.g., 0-5 years) and peaks again around **10-15 years**.
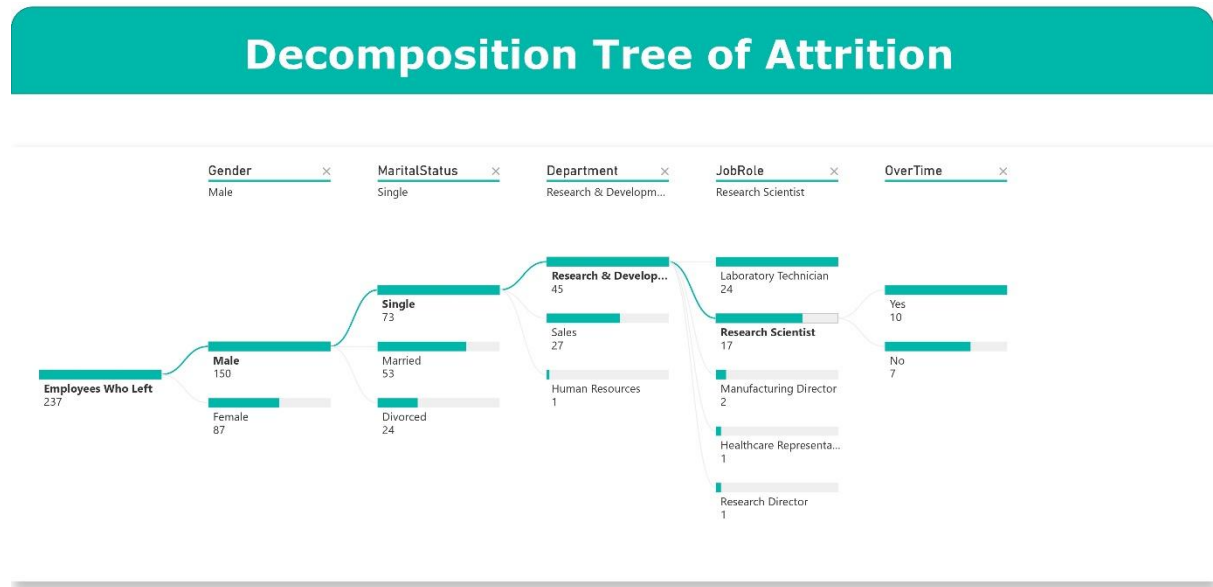
5.4. Decomposition Tree of Attrition

Figure 5.5: Decomposition Tree of Attrition

Dominant Attrition Path: The tree clearly highlights the most significant attrition pathway:

- Male (150 leavers) -> Single (73 leavers) -> Research & Development (45 leavers) -> Research Scientist (17 leavers) / Laboratory Technician (24 leavers) -> Overtime (Yes, 10 leavers for Research Scientists, No 7 leavers for Research Scientist)

This path identifies a highly specific segment: **Single Male Research Scientists and Laboratory Technicians in R&D who work overtime.** This is a critical target group.

6. Recommendations to Reduce Employee Attrition

Based on a comprehensive analysis of employee attrition dashboards, several actionable insights have emerged. The following recommendations target the most critical patterns and groups identified through data, with the goal of reducing overall attrition and improving organizational retention.

6.1. Early Career Retention & Onboarding Initiatives

Employees with 1–3 years of tenure are at a significantly higher risk of leaving, especially those in their early to mid-30s. These employees often face unclear career trajectories, lack of support, and limited professional development.

Recommendations:

- Implement **structured onboarding programs** that extend into the first 12–24 months, including mentor assignments and regular check-ins.

- Develop **fast-track career development plans** for new hires to clarify promotion and skill-growth paths.

- Promote **internal mobility** by advertising open positions across departments to allow employees to explore different career interests.

6.2. Support for High-Risk Demographic Groups

The data highlights that single male employees, particularly in Research & Development roles such as Research Scientists and Laboratory Technicians, experience disproportionately high attrition.

Recommendations:

- Introduce **peer-support networks** and social engagement programs targeting single or early-career employees to foster a greater sense of belonging.

- Offer **flexible work policies** and well-being support that cater to the specific challenges faced by these groups, such as isolation or burnout.

- Provide **role rotation opportunities** within R&D to keep work dynamic and intellectually engaging.

6.3. Overtime Management and Workload Distribution

A strong link exists between overtime work and attrition. Employees working extended hours, especially in technical roles, show significantly higher turnover.

Recommendations:

- Establish **workload monitoring systems** and flag individuals consistently logging high overtime.

- Limit overtime hours through proactive **resource planning and workload balancing** among teams.

- Compensate overtime work fairly through **additional pay, time off, or performance-based rewards**.

6.4. Enhancement of Compensation Packages with Equity

Employees without access to stock options exhibit a higher likelihood of leaving. This reflects both financial dissatisfaction and perceived lack of ownership in the company's success.

Recommendations:

- Expand eligibility for **stock options** to include non-executive and mid-level staff, particularly in high-attrition departments.

- Offer **performance-linked equity incentives** for key contributors with 1–3 years of service.

- Clearly communicate the **long-term financial value of equity** to improve perceived compensation competitiveness.

6.5. Reassess Work Travel and Exposure Opportunities

Interestingly, those who rarely travel for work exhibit higher attrition, suggesting a lack of variety or engagement.

Recommendations:

- Provide occasional **travel or external collaboration opportunities** even for "non-travel" roles, such as site visits, conferences, or training events.

- Rotate employees through **cross-functional projects or assignments** that increase exposure to broader organizational operations.

- Recognize and reward contributions beyond routine work to enhance engagement for employees in more static roles.

6.6. Predictive Retention Strategy with Targeted Intervention

The decomposition tree and other insights suggest a need for a data-driven approach to preemptively identify and support high-risk employees.

Recommendations:

- Develop an **internal attrition risk model** using available HR data to flag employees at higher risk based on tenure, overtime, demographic, and satisfaction metrics.

- For flagged individuals or groups, initiate **customized retention plans**, including conversations with managers, personalized development plans, and benefits reviews.

- Track the **effectiveness of these interventions** over time and refine the model to improve accuracy and outcomes.

By focusing on these key areas, the organization can address both systemic and group-specific drivers of attrition. A blend of structural support, targeted benefits, and personalized attention will be essential to retain top talent and foster long-term employee engagement.

7. Conclusion

This project embarked on a comprehensive analysis of the IBM Employee Attrition dataset with the primary objective of uncovering patterns and identifying key factors driving employee turnover. Through a detailed data visualization approach, we successfully investigated the influence of various attributes, including job roles, educational backgrounds, gender, work conditions, compensation, and career paths, on attrition rates.

Our analysis revealed several critical insights:

- **Early career employees (1-3 years tenure)**, particularly those in their early to mid-30s, exhibit a significantly higher risk of attrition, often due to unclear career trajectories and limited development opportunities.

- **Specific demographic segments**, notably single male employees in Research & Development roles (Research Scientists and Laboratory Technicians), are disproportionately affected by attrition.

- A strong correlation was identified between **overtime work** and increased turnover, particularly in technical roles.

- The absence of **stock options** significantly contributes to an employee's likelihood of leaving, highlighting a perceived lack of ownership and financial dissatisfaction.

- Counter-intuitively, employees with **minimal work travel** also show higher attrition, suggesting a need for broader exposure and engagement opportunities.

- The **decomposition tree** provided a powerful visual summary, pinpointing a highly vulnerable segment: single male Research Scientists and Laboratory Technicians in R&D who frequently work overtime.

The actionable recommendations derived from these insights, spanning enhanced onboarding, targeted support for high-risk groups, improved overtime management, expanded equity compensation, diversified exposure opportunities, and a predictive retention strategy which offer a robust framework for intervention. By focusing on these key areas, the organization can move towards a data-driven approach to talent management.

Ultimately, addressing these systemic and group-specific drivers of attrition is paramount. Implementing a blend of structural support, tailored benefits, and personalized attention will not only lead to a reduction in employee turnover but also foster greater employee satisfaction, engagement, and a more stable, productive workforce in the long term. Continuous monitoring and refinement of these strategies will be essential to ensure sustained success in retaining top talent and building a resilient organizational culture.

REFERENCES

[1] C. Pappas, "Employee Retention Statistics You Should Know In 2025," eLearning
    Industry. Accessed: Jun. 22, 2025. [Online]. Available:
    https://elearningindustry.com/employee-retention-statistics-you-should-know
[2] "38+ Employee Turnover Stats (2024-2027)," Exploding Topics. Accessed: Jun. 22,
    2025. [Online]. Available: https://explodingtopics.com/blog/employee-turnover-statistics