ANALYZING EMPLOYEE ATTRITION

Uncovering Drivers of Employee Turnover

OUTLINE

1. Executive Summary
2. Introduction
3. Data Exploration
4. Data Analysis
5. Key Insights
6. Future Direction
7. Appendix

Data Science Project by Dr. Elena Vieth

Contact: elenavieth@gmail.com
LinkedIn: www.linkedin.com/in/elenavieth

Executive Summary

The Cost of Attrition

Employee turnover is an unavoidable challenge for businesses, but when it happens too frequently, it becomes a costly issue. Studies estimate that **replacing an employee can cost anywhere between 90% to 200% of their annual salary**, factoring in hiring expenses, onboarding, training, and productivity losses. Beyond financial costs, high turnover disrupts team stability, lowers morale, and hinders long-term growth.

While traditional HR strategies focus on exit interviews or broad employee engagement surveys, these methods often come too late — when an employee has already decided to leave. The key question for any organization is:

Can we predict which employees are likely to leave and take proactive steps to retain them?

This report provides a **data-driven framework** to answer this question.

How This Analysis Helps Businesses

This study explores how companies can **use data analytics to anticipate employee attrition risks and design targeted retention strategies**. Using a fictional dataset from IBM HR Analytics, we analyze key drivers of turnover and demonstrate how businesses can apply predictive modeling to improve employee retention.

Although the dataset does not represent real employees, it mirrors common HR challenges that companies face. This case study serves as an example of how businesses can harness their internal HR data to prevent avoidable turnover and improve workforce stability.

Executive Summary

Key Findings: What Makes Employees Leave?

Through advanced analytics, we identified several **clear patterns in employee turnover:**

Employees working overtime are more likely to leave. Reviewing workload and ensuring work-life balance can mitigate this risk.

Lower-income employees are more likely to leave, indicating the need for competitive compensation.

Employees with stock options exhibit lower attrition. Expanding stock options or other financial incentives could enhance retention.

Shorter tenures with the current manager correlated with higher attrition, highlighting the importance of strong managerial relationships.

Business Impact: The Cost-Saving Potential of Retention

By proactively addressing overtime, compensation, managerial relationships, and financial incentives, organizations can significantly reduce preventable turnover. Projected savings for targeted retention strategies are:

- Reducing attrition by just **10-20%** could result in **millions in cost savings annually**, depending on company size.
- Improved employee retention leads to **higher productivity, stronger team morale, and increased institutional knowledge**—all of which contribute to long-term growth.

Rather than reacting to resignations, organizations can use data to act early, intervene proactively, and retain their top talent.

Executive Summary

Next Steps for HR Leaders & Business Decision-Makers

- **Leverage HR Data to Track Retention Risks:** Identify at-risk employees before they decide to leave.
- **Prioritize Employee Well-being & Workload Balance:** Address overtime and burnout risks with structured workload management.
- **Refine Compensation & Incentive Strategies:** Ensure salaries remain competitive and explore long-term financial incentives like stock options.
- **Invest in Leadership Development:** Train managers to foster stronger, trust-based relationships with their teams.
- **Continue Monitoring & Improving Attrition Models:** Adapt strategies as new employee sentiment data becomes available.

By adopting a data-driven approach to retention, companies can reduce turnover, improve workforce satisfaction, and achieve substantial cost savings over time.

Why This Matters

Employee attrition isn't just an HR problem—it's a business risk that impacts financial performance, productivity, and growth. This case study highlights how predictive analytics can transform HR decision-making, turning reactive policies into proactive retention strategies.

> ***Start the retention process when the person is still open to staying and not after they've already told you they're leaving.***
>
> Jeff Weiner, Executive Chairman at LinkedIn

INTRODUCTION

Introduction

The Problem: Employee Attrition Hurts Business Growth

Every company faces employee turnover, but when attrition rates become too high, they hurt profitability, productivity, and team morale.

- **Financial Impact:** Hiring a replacement for an employee who leaves can cost up to 200% of their annual salary—including recruitment, onboarding, and lost productivity.
- **Operational Disruptions:** Frequent resignations create team instability, lower morale, and lead to the loss of institutional knowledge.
- **Missed Growth Opportunities:** Instead of investing in innovation and workforce development, HR teams often get stuck in a cycle of reactive hiring.

Traditional HR methods—such as exit interviews, surveys, and general engagement initiatives—often come too late, once an employee has already decided to leave.

This report explores a data-driven approach that helps companies move from reactive to proactive workforce management.

How Can Data Help?

- Rather than waiting for employees to leave, HR teams can **use historical workforce data** to identify employees who are at risk of resigning.
- By analyzing factors such as overtime, salary, job satisfaction, and managerial relationships, businesses can develop **targeted retention strategies** before employees become disengaged.
- This approach allows HR professionals to **focus on preventing turnover rather than constantly hiring replacements**.

Predictive analytics allows HR teams to make better, faster decisions—helping businesses retain top talent and reduce unnecessary costs.

Sources:
https://www.shrm.org/topics-tools/news/employee-relations/attrition-definition-types-causes-mitigation-tips
Cascio, W.F. 2006. Managing Human Resources: Productivity, Quality of Work Life, Profits (7th ed.). Burr Ridge, IL: Irwin/McGraw-Hill.
Mitchell, T.R., Holtom, B.C., & Lee, T.W. 2001. How to keep your best employees: Developing an effective retention policy

Introduction

A Use Case: What This Study Demonstrates

To illustrate how data science can transform HR decision-making, this report examines employee attrition using the IBM HR Analytics Employee Attrition dataset.

- **Important Note:** This dataset is fictional and does not represent real employees. However, it was designed to simulate real-world workforce trends and is commonly used as a case study in HR analytics.
- **Why this matters:** Although the dataset is not sourced from a real company, it provides a realistic example of how businesses can apply predictive modeling to understand why employees leave and what can be done to improve retention.

The main objective of this project is to address several critical questions related to employee attrition, including:

1

What demographic and job-related factors drive attrition?
Understanding the key drivers behind employee turnover helps identify at-risk groups and informs targeted retention strategies.

2

Can we predict employee attrition using machine learning models?
By building and evaluating classification models, we aim to accurately forecast attrition and prioritize at-risk employees for intervention.

3

What actionable insights can be derived to mitigate attrition?
Leveraging advanced tools like SHAP analysis, this study highlights specific, data-driven recommendations to reduce turnover and enhance workforce stability.

Data Exploration

Data Exploration

What Data Tells Us About Employee Attrition

Before predicting employee turnover, it's important to **understand the workforce** itself. This study uses a fictional HR dataset that mimics real-world employee records, including:

- Demographics: Age, marital status, education.
- Job Details: Department, job role, salary, experience.
- Work Conditions: Overtime, travel, stock options.
- Employee Satisfaction: Relationship with manager, job satisfaction, work-life balance.

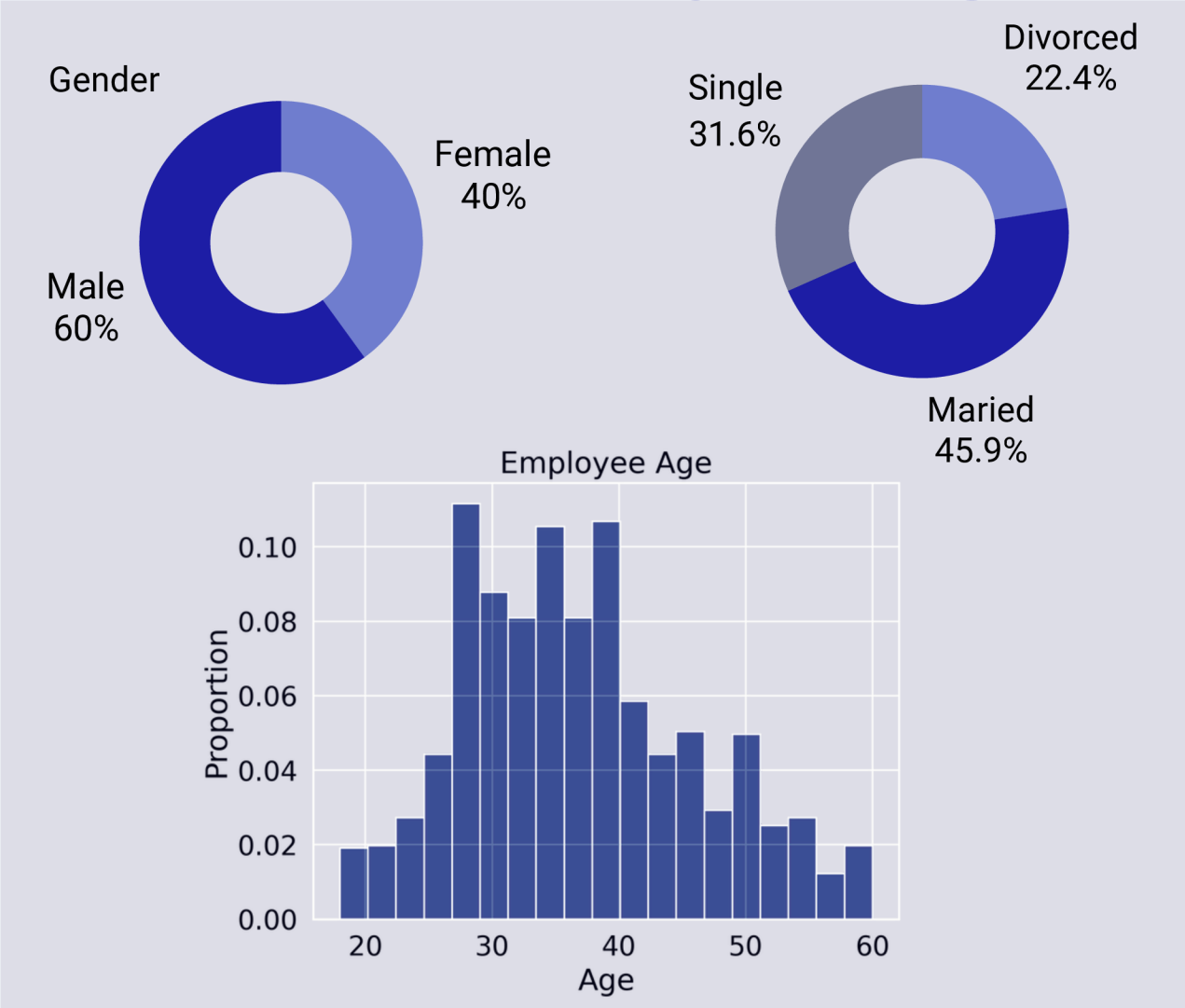
By subsequently analyzing the relationship between attrition these factors, we can **identify early warning signs of disengagement and attrition**.

Key Insights

- Predominantly **early to mind-career professionals**, with 1 to 5 years of company tenure
- Mostly **MINT backgrounds** and job roles
- Over 50% **earn ≤ $6,032** per month
- Most employees **do not work overtime** and report high job satisfaction and good work-life balance

Data Exploration

Employee Demographics

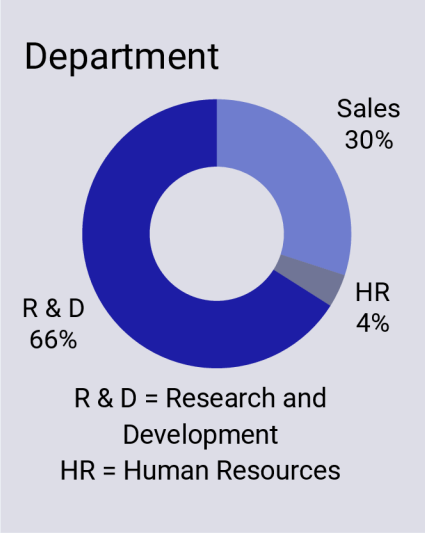


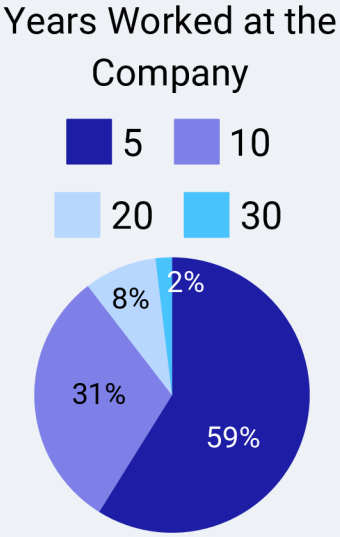
Exploration of **employee demographic data** revealed that gender is fairly balanced within the dataset. Most employees are either married or single and between 25 and 45 years of age.

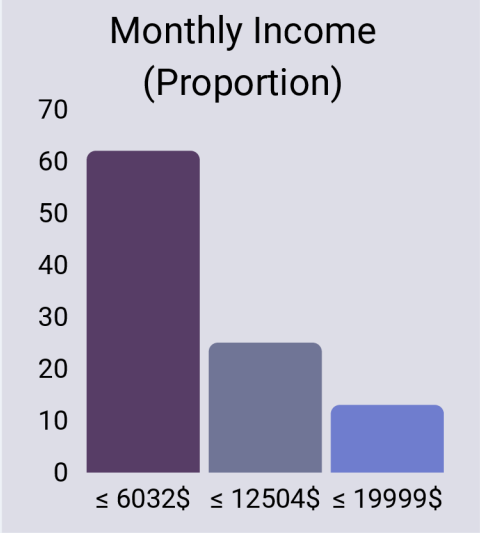
Understanding workforce demographics helps organizations tailor benefits, engagement initiatives, and career development programs.

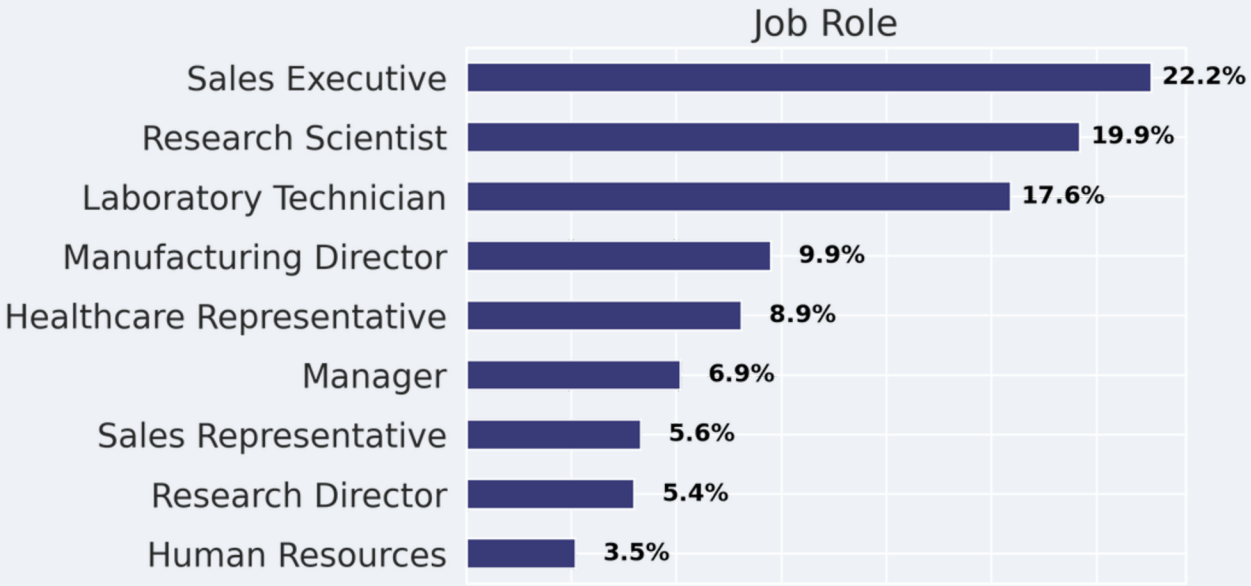
Data Exploration

Employee Profiles







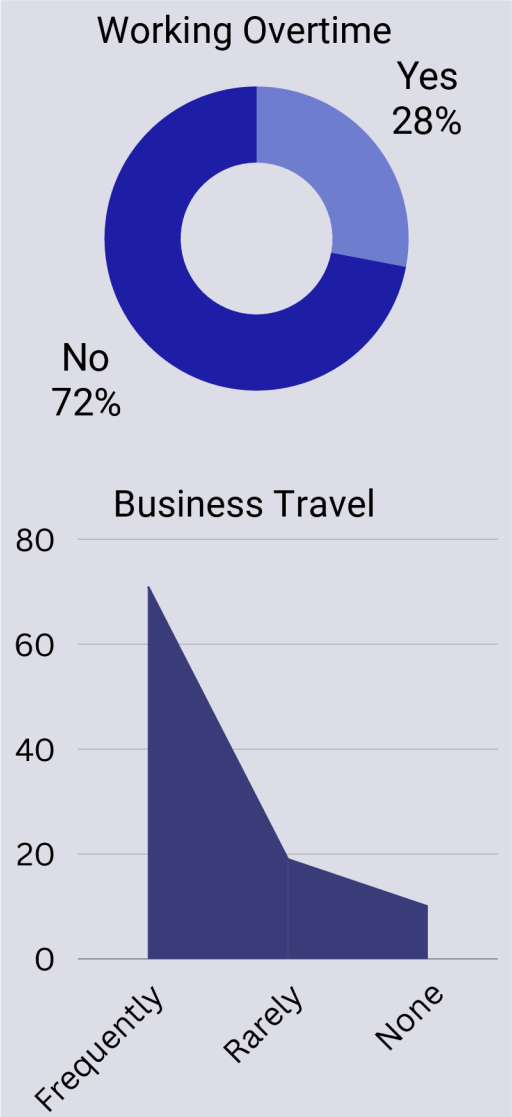


Most employees works in **Research & Development**. Frequent **job roles** are Sales Executive, Research Scientist & Laboratory Technician.

More than half of the employees **earn less than or equal to 6032$**. The majority **works 1-5 years** at the company.

Data Exploration

Job Demands & Employee Satisfaction



Most employees do not work overtime. A majority of employees travel for business, with only one 10th indicating no travel.

More than half have a high to very high job satisfaction and about 80% rate their work-life balance good or better.

Data Analysis

Data Analysis

What Predictive Modeling Reveals About Employee Attrition

The primary goal of this analysis was to **predict which employees are at risk** of leaving the company (attrition) so that proactive retention strategies can be developed. **Machine learning** was chosen for this, as it enables the model to learn patterns in historical data, such as how salary, job satisfaction, or overtime influence attrition. This way, the model cannot only be utilized to **uncover existing factors** influencing attrition in the workforce, but also **predict future** outcomes and identify at-risk employees.

By analyzing patterns in work history, job conditions, and employee satisfaction, we built a predictive model that helps identify employees at risk of resigning before they leave.

This approach allows HR teams to:

- Spot early warning signs of disengagement.
- Prioritize interventions for at-risk employees.
- Reduce unnecessary turnover costs.

How Well Does The Model Work?

Our model correctly identifies most employees who are at risk of leaving, allowing HR teams to **analyze the current key drivers** (e.g., overtime, low salary, poor job satisfaction) and take targeted short-term actions.

While predictive analytics improves decision-making, no model is perfect. Some employees predicted to leave may stay, and some interventions may not work. Improvement in the models ability to predict individual future departures accurately could enable HR to use it for **long-term forecasting**.

Data Analysis

Working overtime increases attrition likelihood, while not working overtime protects against leave.

Monthly income on the low and high end of the scale increases or protects against attrition, respectively.

Owning at least one company stock is a protective factor against attrition.

Employees with fewer **years under their current manager** are at higher attrition risk.

Younger employees show higher attrition likelihood. **Middle aged as well as senior employees** are similarly less likely to leave.

Having worked for less companies is a protective factor, while having **more previous employers** increases attrition likelihood.

Employees with low **relationship satisfaction** exhibit distinct attrition likelihood, while average to high satisfaction similarly reduces risk.

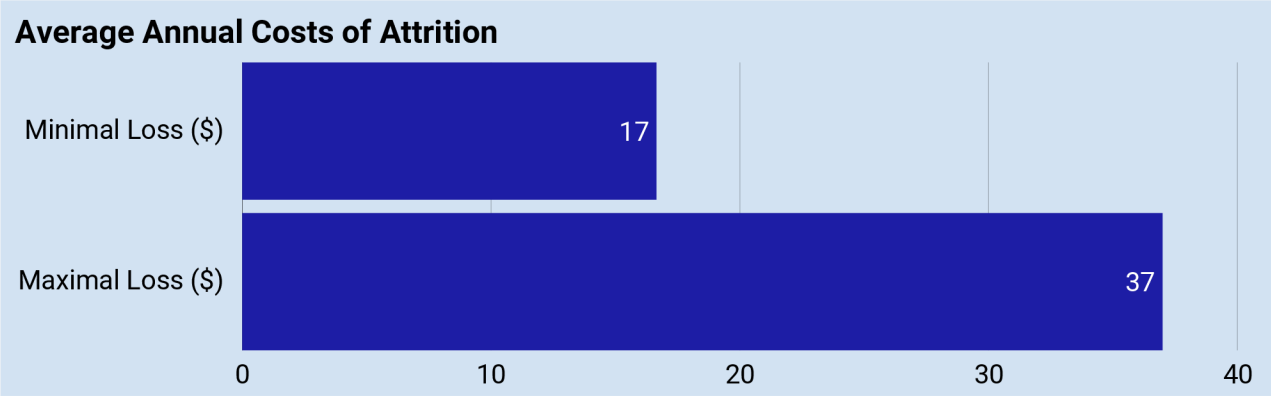
Employees with low **environment satisfaction** show increased attrition likelihood, while average to very high satisfaction reduces attrition similarly.

Shorter commutes reduce attrition likelihood, while **longer commutes** increase it, with the former having a greater protective impact.

Medium to very high **job satisfaction** reduces attrition, while being dissatisfied increases it.

Key Insights

Key Insights



The average annual costs of attrition are estimated to range from $16.6 million (minimal estimate) to $37.0 million (maximal estimate), based on average employee income. Estimated maximal losses include direct replacement costs, lost productivity, and onboarding expenses, highlighting the urgent need for targeted retention strategies.

Turning Data Insights into Actionable HR Strategies

The predictive model identified key factors driving employee attrition. These findings can be translated into strategies that HR leaders and executives can implement to improve retention and workforce stability.

By addressing these risk factors, companies can:

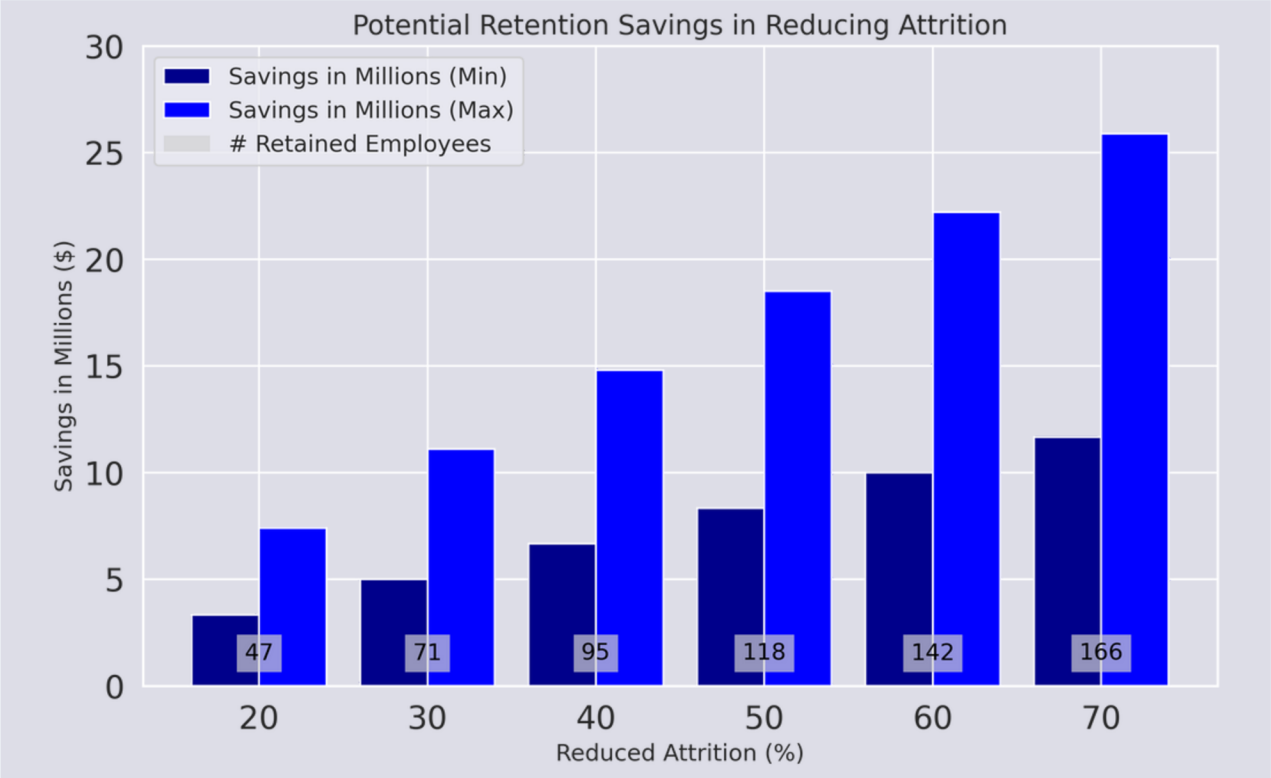
- Reduce unnecessary turnover and increase employee engagement.
- Save on hiring and training costs associated with high attrition.
- Strengthen workplace culture, productivity, and team stability.

Key Insights

Estimated Financial Impact of Reducing Attrition

By implementing data-driven retention strategies, a company can:

- Reduce attrition effectively, resulting in **millions in cost savings annually**.
- **Improve employee engagement and productivity**, leading to better team performance.
- Build a stronger, more stable workforce that drives l**ong-term business success**.



Estimated Financial Impact of Reducing Attrition

This plot illustrates the financial savings achievable by reducing employee attrition through targeted retention strategies. The analysis shows potential minimal and and maximal savings for reductions in attrition rates from 20% to 70%, with estimates based on the average employee income. These figures highlight the significant financial impact of retention strategies, which address direct replacement costs, lost productivity, and onboarding expenses.

Key Insights

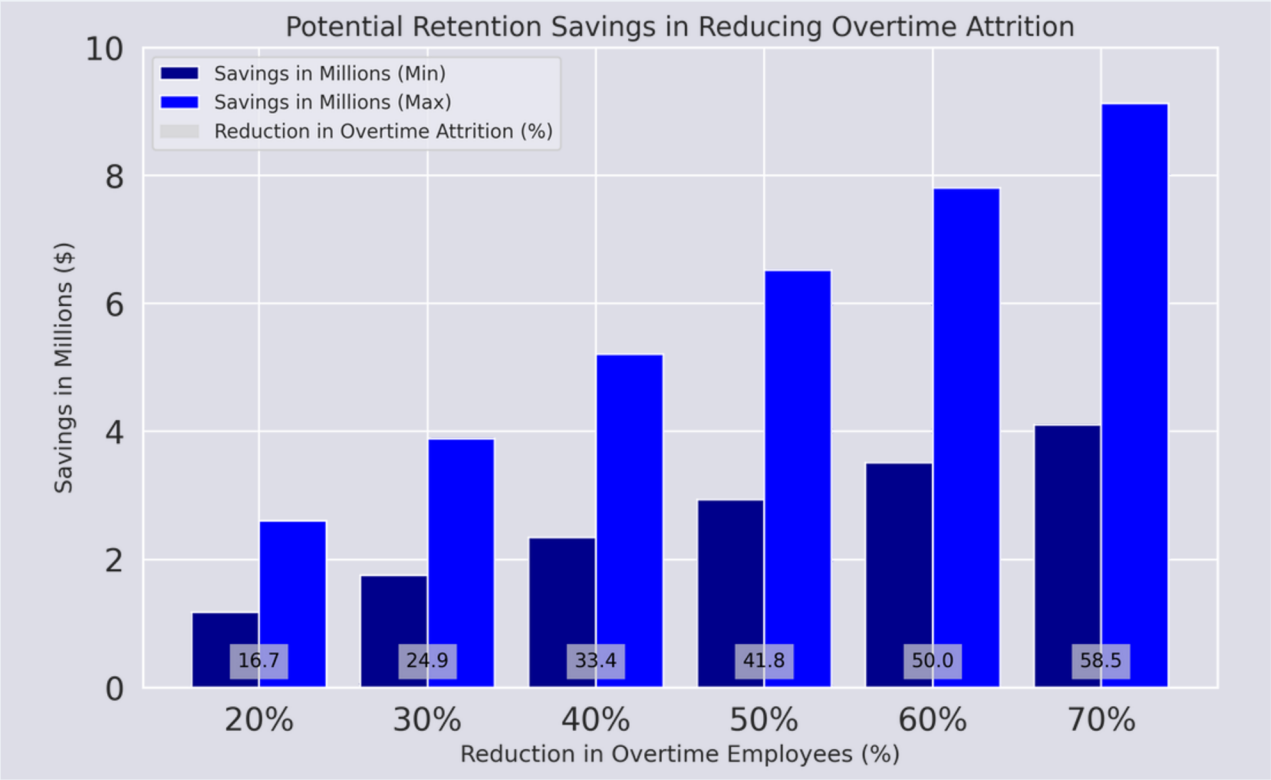
Targeting the Main Driver of Attrition: Working Overtime

Employees who regularly work overtime are significantly more likely to leave.
Recommended Strategy:

- Monitor employee workload and redistribute tasks where needed.
- Implement overtime policies that ensure sustainable work hours.
- Encourage flexible schedules to promote better work-life balance.

Business Impact:

Reducing overtime can improve retention rates and increase employee satisfaction, leading to higher productivity and lower absenteeism.



This plot illustrates the financial savings achievable by **reducing overtime-related employee attrition through targeted retention strategies**. The analysis shows potential minimal and maximal savings for reductions in employees working overtime from 20% to 70%, with estimates based on the average employee income. These figures highlight the significant financial benefits of addressing overtime, which was identified as a mayor contributor of employee attrition.

Key Insights

Further Retention Strategies Based on Data Insights

Competitive Compensation to Retain Top Talent
Employees earning below $6,000 per month are at higher risk of leaving.

- Conduct salary benchmarking to stay competitive in the job market.
- Offer performance-based raises and retention bonuses.
- Introduce non-monetary benefits like professional development funding.

Expand Stock & Equity Incentives
Employees with stock options have lower attrition rates than those without.

- Expand stock-based compensation beyond executives.
- Offer equity vesting plans that reward long-term tenure.
- Educate employees on the long-term value of stock options.

Strengthen Leadership & Managerial Stability
Employees who haven't spent long under their current manager are more likely to leave.

- Implement leadership training to improve managerial effectiveness.
- Encourage mentorship between employees and senior leaders.
- Conduct "stay interviews" to assess concerns before employees leave.

Improve Job Satisfaction & Employee Engagement
Employees with low job satisfaction ratings are much more likely to leave.

- Conduct regular engagement surveys to track employee sentiment.
- Improve career development programs and internal mobility options.
- Recognize and reward high performers with promotions and incentives.

Consider Hybrid & Flexible Work Models
Employees with long commutes show slightly higher attrition rates.

- Offer remote work or hybrid work options where possible.
- Provide commuter benefits (subsidized transport, flexible start times).
- Ensure equal career growth for remote and on-site employees.

Future Directions

Future Directions

How Businesses Can Continue Leveraging Predictive Analytics

This study demonstrates that data-driven insights can significantly **enhance employee retention strategies**. By identifying key drivers of attrition, HR leaders can implement **targeted interventions** before employees decide to leave.

To ensure long-term success, businesses must continue refining their workforce strategies using **real-time data monitoring, feedback loops, and continuous optimization**.

Build an Ongoing Employee Retention Monitoring System

Instead of treating attrition analysis as a one-time study, companies should:

- Track key employee risk indicators (e.g., overtime, job satisfaction).
- Use dashboards to visualize workforce trends and predict risks.

Early detection of disengagement allows HR teams to act before employees decide to leave, reducing unnecessary turnover.

Expand Data Collection for a More Comprehensive View

While this analysis provides strong insights, businesses can enhance predictive accuracy by incorporating:

- Employee engagement surveys regarding job satisfaction and career growth concerns.
- Exit interview data to identify additional patterns in voluntary turnover.
- Performance and training metrics assessing how career development affects retention.

Test & Measure the Impact of Retention Strategies

Testing different approaches ensures HR efforts are cost-effective and data-driven rather than one-size-fits-all.

- Pilot new retention programs (e.g., mentorship initiatives)
- Measure effectiveness (turnover decreased/engagement improved?).
- Refine strategies based on data feedback.

Appendix

Data Exploration

Dataset Overview

Columm	Datatype (Scale)
Age	Continuous (Years)
Attrition	Binary (Yes|No)
Business Travel	Categorical (Never|Rarely|Frequently)
Department	Categorical (Departments)
Distance From Home	Continuous (Miles)
Education	Categorical (Below College to Doctor)
Education Field	Categorical (fields)
Environment Satisfaction	Categorical (Low to Very High)
Gender	Binary (Male|Female)
Job Involvement	Categorical (Low to Very High)
Job Level	Categorical (1 to 5)
Job Role	Categorical (Roles)
Job Satisfaction	Categorical (Low to Very High)
Marital Status	Categorical (Single|Married|Divorced)
Monthly Income	Continuous (Dollars)
Number Companies Worked At	Discrete (Count)
Overtime	Binary (Yes|No)
Percent Salary Hike	Continuous (Percentage)
Performance Rating	Categorical (Low to Outstanding)
Relationship Satisfaction	Categorical (Low to Very High)
Stock Option Level	Categorical (0 to 4)
Total Working Years	Continuous (Years)
Training Time Last Year	Discrete (Count)
Work Life Balance	Categorical (Bad to Best)
Years at Company	Continuous (Years)
Years in Current Role	Continuous (Years)
Years Since Last Promotion	Continuous (Years)
Years With Current Manager	Continuous (Years)

Data Analysis

Objectives of the Analysis

The primary goal of this analysis was to predict which employees are at risk of leaving the company (attrition) so that proactive retention strategies can be developed. ***Machine learning was chosen for this, as it enables the model to learn patterns in historical data, such as how salary, job satisfaction, or overtime influence attrition.*** This way, the model cannot only be utilized to uncover existing factors influencing attrition in the workforce, but also predict future outcomes and identify at-risk employees.

Several models that are suited for classification problems (i.e., predicting the probability of an employee leaving), such as Logistic Regression, Random Forest, and Gradient Boosting, were utilized. Subsequently, evaluation metrics were used to identify the model that performs best for predicting employee attrition. ***Model evaluation techniques focused on 1) how well each model identifies employees who are leaving the company and 2) how well the model van distinguish between the two classes*** (non-attrition versus attrition).

After evaluating these models using recall and precision, Gradient Boosting emerged as the most reliable option, helping us identify at-risk employees with high accuracy and take proactive measures to reduce attrition. ***To gain meaningful results, the most impactful features were investigated by analyzing SHAP (shapely additive explanations) values.*** SHAP assigns each feature a contribution value that quantifies its impact on the model's output for individual predictions. By ranking features based on their average impact across all predictions, decisions about which features should be prioritized can be made, providing actionable insights into the key drivers of employee attrition.

Data Analysis

Model Evaluation

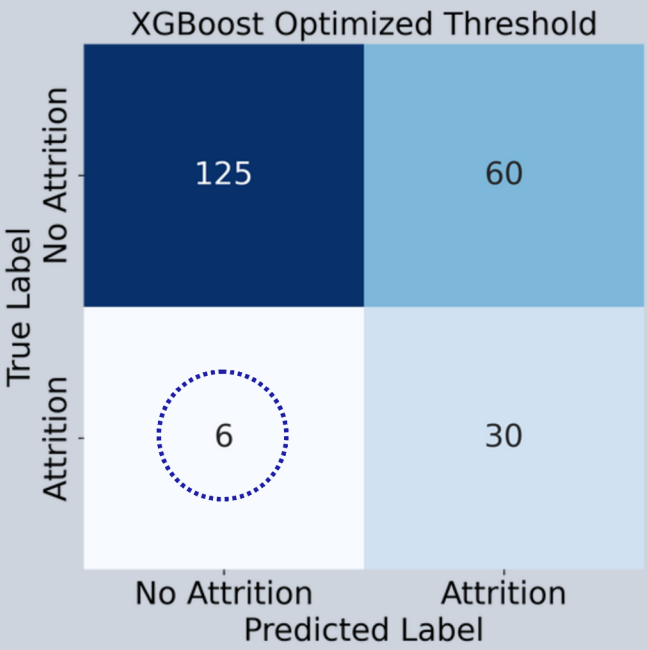
XGBoost			
	Precision	Recall	F1 Score
Non-Attrition	0.95	0.68	0.79
Attrition	0.33	0.83	0.48

The model is excellent at identifying employees who will stay (95% non-attrition precision). Among all employees who actually stay, the model correctly identifies 68% of them (non-attrition recall). The thus model misses 32% of stayers, but since this group is not a financial risk (they won't leave), this is less critical than missing actual attrition cases.

Among all employees who actually leave, the model correctly identifies most (83% recall for attrition). By prioritizing recall for attrition, the model ensures most employees at risk of leaving are detected. However, some employees predicted to leave may actually stay (33% precision for attrition), which means that some interventions may be unnecessary, but this trade-off is intentional to minimize missed attrition cases.

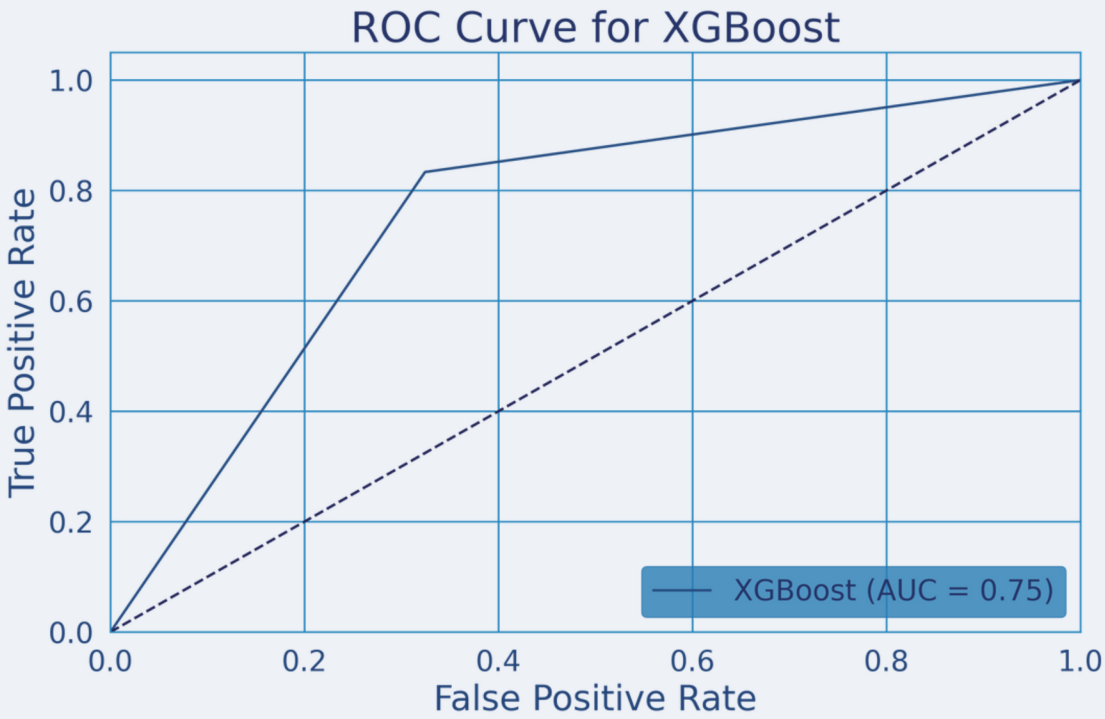
Missing an at-risk employee is more costly because it means loosing the opportunity to intervene and potentially retain that employee.

Accordingly, models minimizing false negatives (i.e., the model fails to identify an attrition case) were prioritized, as indicated by a high recall score.



Data Analysis

Overview of Model Performance



While minimizing false negatives was a priority, it is crucial that a model consistently performs well across all scenarios, meaning that it is able to correctly distinguish between attrition and non-attrition cases. We therefore used additional metrics that ensure the model is robust and reliable, providing confidence in its predictions.

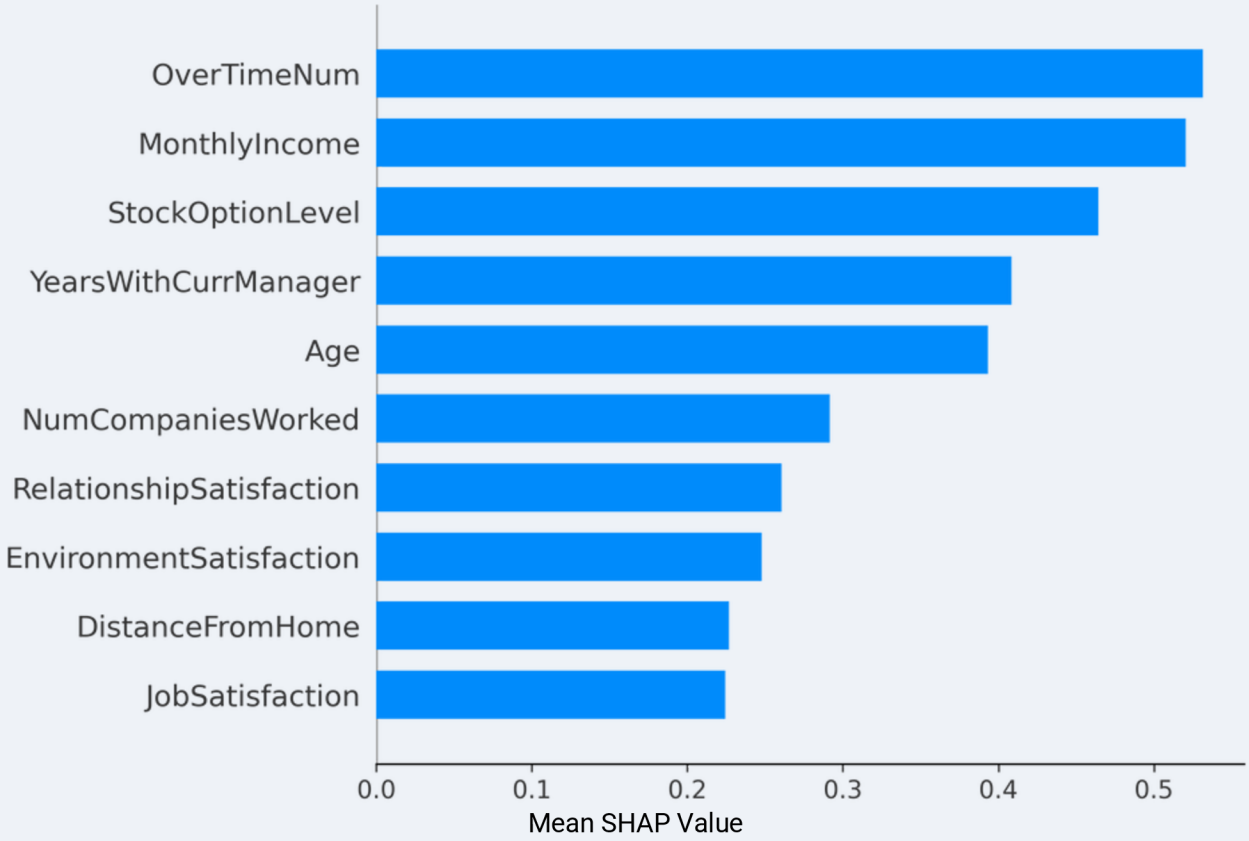
The model has good overall discrimination power, as displayed by the ROC-AUC Curve, meaning it effectively differentiates between employees who will stay vs. those who will leave.

Key Insights

- **Understanding Current Attrition Patterns:** HR can analyze the key drivers (e.g., overtime, low salary, poor job satisfaction) and take targeted short-term actions.
- **Long-Term Forecasting:** However, since only 1 in 3 employees predicted to leave actually leaves, the model struggles with predicting individual future departures accurately.

Data Analysis

Feature Importance



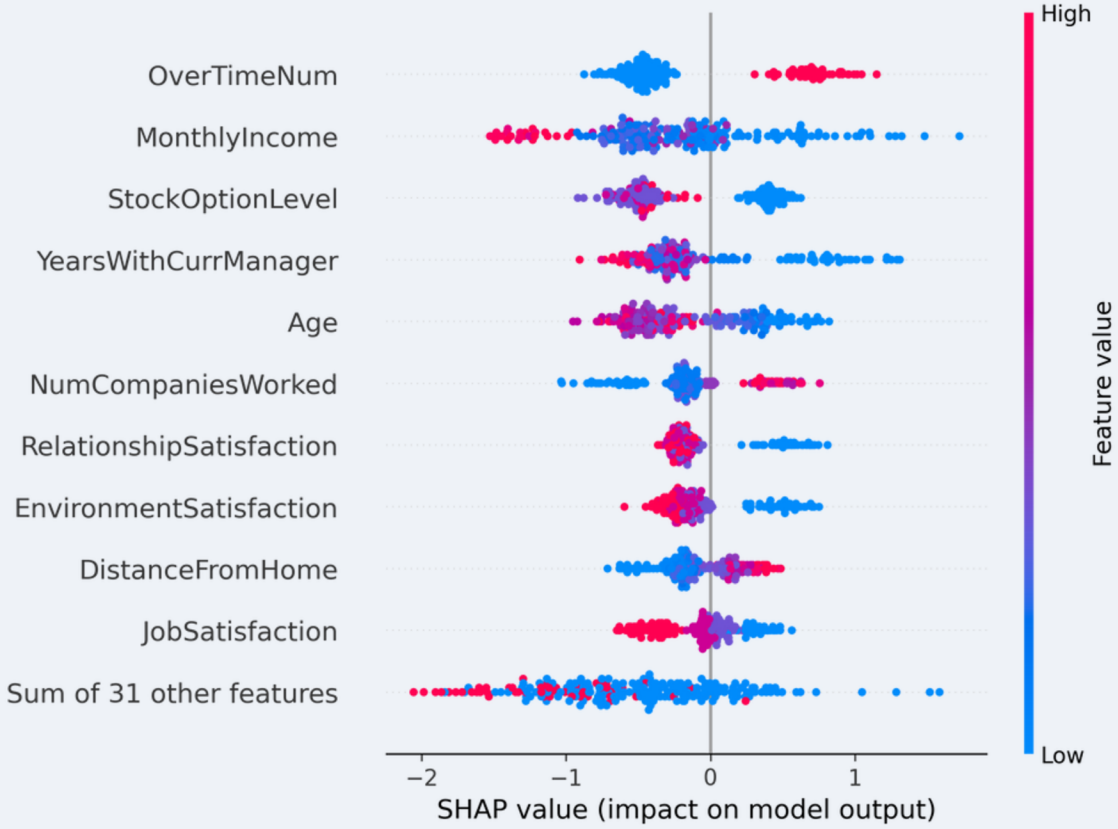
The SHAP summary bar plot illustrates the global importance of the top 10 features in the model's predictions. Each bar represents the mean absolute SHAP value for a given feature, indicating its average contribution to the model’s output across all predictions. Features with taller bars exert a greater influence on the predictions, meaning they are more critical in determining the model's decisions.

Key Insights:

- Overtime has the highest mean SHAP value, showing it is the most impactful feature in predicting attrition, followed closely by monthly income and stock option level.
- Features such as job satisfaction and distance from home, while still important, contribute less significantly to the model's predictions compared to the top-ranking features.

Data Analysis

Feature Importance



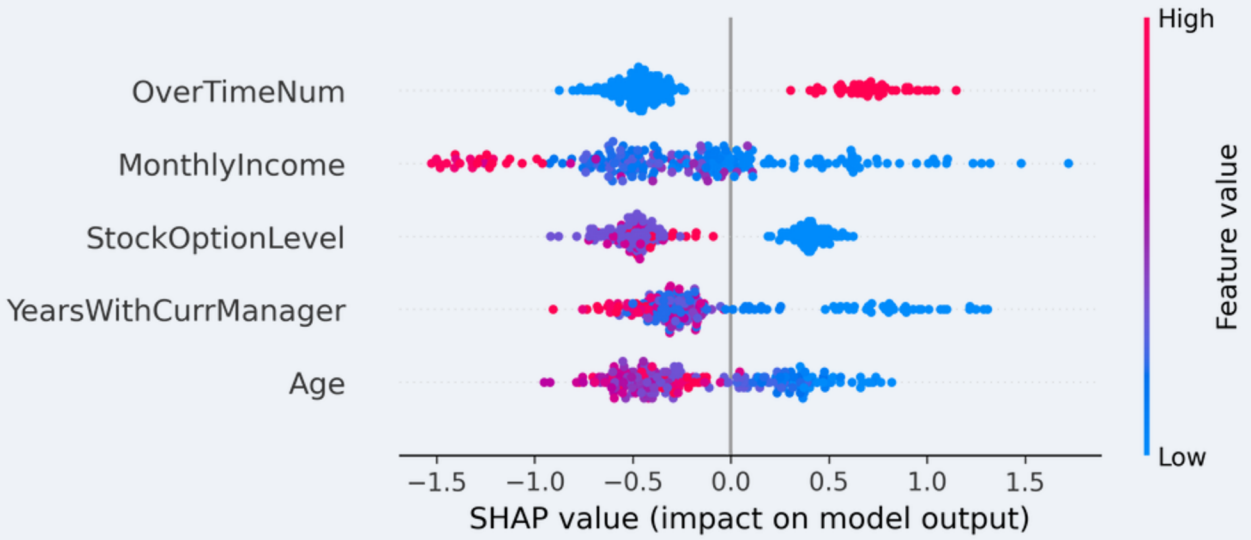
The SHAP beeswarm plot provides a visualization of how the top 10 features contribute to model predictions. The features are ranked by their impact on the model's predictions from high to low. SHAP values displayed on the x-axis quantify the contribution of a feature to the model's prediction for a specific observation, with negative values indicating that the feature decreases the likelihood of attrition, while positive values indicate that the feature increases the likelihood of attrition. Feature Values are signified by a color scale, from low feature values in blue color to high feature values displayed in red.

The predictive power of a feature can be visually inspected by how tightly associated feature's values are with specific SHAP value ranges:

- If a features high values consistently correspond to positive SHAP values, while low values correspond to negative values, this distinct spread suggest that this is a highly predictive feature.
- A smooth transition in the color gradient (e.g., blue to pink/red) across the SHAP value range shows that the feature captures an underlying pattern in the data.
- Features with a wider spread of SHAP values exert more influence on the model's predictions, implying higher predictive power.

Data Analysis

Feature Importance



Interpretation of Top 10 Features

Overtime

Working overtime (red dots) increases the likelihood of attrition (positive SHAP values) while not working overtime decreases it. Feature values relate to distinct SHAP values, making the predictive power of overtime valuable.

Monthly Income

Lower income (blue dots) is associated with higher attrition (positive SHAP values), higher income reduces attrition. While low and medium income values sharing similar SHAP values, high income seems to be a distinctive protective factor.

Stock Option Level

Owning no stock options distinctively increases attrition likelihood (blue dots), owning one or more stocks results in similar decreases in attrition likelihood.

Years With Current Manager

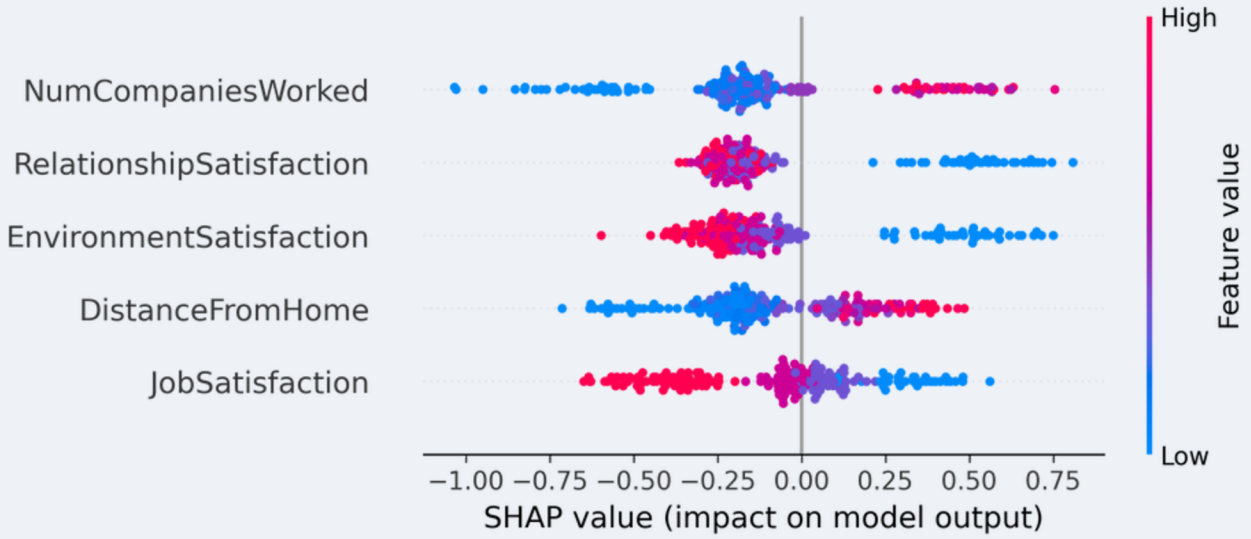
Employees with fewer years under their current manager (blue dots) are distinctively more likely to leave. Longer tenures with their manager reduce attrition likelihood, however, the relationship is not linear.

Age

Younger employees (blue dots) exhibit higher attrition likelihood. Middle aged and older employees are similarly less likely to leave.

Data Analysis

Feature Importance



Interpretation of Top 10 Features

Number of Companies Worked for

Having worked for more companies (red and lilac dots) increases, while having worked for less lowers attrition likelihood. Low values appear to be more distinct in their predictions compared to medium and high values.

Relationship Satisfaction

Employees with low relationship satisfaction (blue dots) increase attrition likelihood. Employees with average to high relationship satisfaction (lilac to red dots) have a similarly reduced risk of attrition.

Environment Satisfaction

Employees with low environment satisfaction (blue dots) are increase attrition likelihood, while average to very high satisfaction reduces attrition likelihood similarly.

Distance From Home

Shorter commutes (blue dots) tend to reduce attrition likelihood, while longer commutes increase it. Extreme values (short and long) are more predictive compared to medium distances.

Job Satisfaction

Medium to very high job satisfaction (lilac to red dots) reduces attrition, while being dissatisfied increases it. The spread is distinct even an medium levels.

Retention Strategies

Retention Strategies

High-Risk Scenarios (Attrition Driver)		
Factor	**Impact**	**Retention Strategy**
Overtime	Working overtime leads to attrition	Develop workload management strategies
Monthly Income	Low-income employees at high risk	Introduce competitive salaries & bonus structures
Stock Options	No stock options increase attrition	Introduce equity incentives
Relationship Satisfaction	Poor workplace relationships	Implement structured team-building & coaching
Years with Current Manager	Short tenure under manager	Provide leadership training, stay interviews
Job Satisfaction	Dissatisfied employees leave	Investigate in career growth & engagement initiatives
Environment Satisfaction	Poor work environment increases turnover	Improve office conditions and culture
Number of Companies Worked For	Frequent job changers more likely to leave	Offer personalized career development plans
Age	Younger employees more likely to leave	Offer career fast-tracking & mentorships
Distance from Home	Long commutes increase attrition	Provide remote working opportunities & commuter benefits

What makes employees leave?

Attrition is Multi-Factorial: A combination of high workload, low compensation, poor workplace and managerial relationships drives attrition.
Tailored Strategies Work Best: One-size-fits-all retention strategies don't work. Segmenting employees by risk factors (e.g., age, income, commute) allows for personalized solutions.

Retention Strategies

Retention Strategies

High-Retention Scenarios (Retention Driver)		
Factor	**Impact**	**Retention Strategy**
Overtime	Employees who do not work overtime	Reward teams that maintain productivity within regular hours
Monthly Income	Higher-income employees	Offer clear progression paths & performance-based raises
Stock Options	Employees with at least one stock option	Educate employees on how stock options build wealth.
Relationship Satisfaction	Employees with medium to high relationship satisfaction	Implement structured team bonding activities.
Years with Current Manager	Employees with more years with same manager	Support long-term manager-employee relationships
Job Satisfaction	Employees with high to very high job satisfaction ratings	Offer promotions & learning opportunities
Environment Satisfaction	Employees with average to high environment satisfaction ratings	Conduct surveys & act on workplace feedback.
Number of Companies Worked For	Employees who worked for fewer companies	Reward long company tenures and highlight them as success stories.
Age	Middle aged and senior employees	Offer long-term incentives like pensions & leadership training.
Distance from Home	Employees with short commutes	Provide office location flexibility or hybrid work.

What makes employees stay?

Retention is Multi-Factorial: Employees stay when they experience fair workload and compensation, a positive workplace culture and strong leadership.
Sustained Engagement Works Best: Employees who feel continuously valued, connected to their teams and leaders are less likely to leave.