

EMPLOYEE ATTRITION PREDICTION

PRESENTED BY

THANH VINH GIANG (ALVIN)

MASON SEIFADDINI

REKHA RENUKAPPA

UMA SELVARAJ



SHOULD I QUIT MY JOB? 16 SIGNS IT'S TIME TO SAY YES

- Your responsibilities have increased, but the pay hasn't.
- You don't get any feedback about your performance.
- There is a high turnover rate among employees.
- You have lost interest or passion for your job.
- Your knowledge and skills are under-utilised.
- You feel that there is no growth left for you.
- You're consistently stressed, negative, or unhappy at work.
- You're not learning.
- You have a terrible boss.
- You're experiencing abuse.
- Your workplace is a toxic environment.
- Your feedback and ideas are not heard.
- You dread going to work in the morning.
- Your health is being negatively impacted.
- Your company seems financially unstable.



WHAT IS HR ANALYTICS

HR analytics is the process of collecting and analyzing Human Resources (HR) data in order to improve an organization's workforce performance.

The process can also be referred to as talent analytics, people analytics, or even workforce analytics.

This method of data analysis takes data that is routinely collected by HR and correlates it to HR and organizational objectives. Doing so provides measured evidence of how HR initiatives are contributing to the organization's goals and strategies.



OBJECTIVE

Employee attrition is the rate at which employees leave a company. The goal of this analysis is to model employee attrition and determine the most dominant contributing factors that govern this turnover.

Through this kind of analysis, we can understand how many employees are likely to leave, while also determining which employees are at the highest risk and for what reasons.



Attrition Rate for HR



$$\text{Attrition rate} (\%) = \frac{\text{number of employee departures}}{\text{average number of employees}} \times 100$$

Factors impacting attrition

Internal factors



- Compensation
- Job satisfaction
- Learning & development

External factors



- Workforce demographics
- Industry shifts
- Economic conditions

• What are the statistics for employee attrition?

• The average employee turnover rate in Australian workplaces, for the 12 months to the end of October 2023, was 14%.

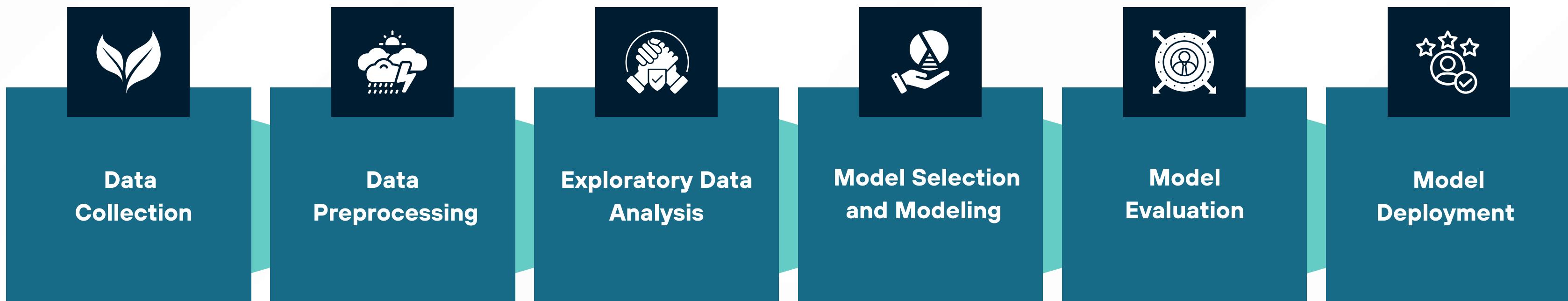




MOTIVATION

The project stems from the potential to improve employee satisfaction, reduce costs, enhance organizational performance, and create a positive workplace culture. It's an opportunity to use data and analytics to make meaningful changes that benefit both employees and the organization as a whole.

SYSTEM ARCHITECTURE DIAGRAM



The dataset on employee attrition from Kaggle contains 35 columns and 1,470 rows.

Key Features of the Dataset:

1. Demographics: Includes attributes like age, gender, marital status, and education level.
2. Job-related Information: Includes job role, job satisfaction, work-life balance, and job level.
3. Compensation: Includes salary information, stock options, and years at the company.
4. Performance Metrics: Performance rating and training times.
5. Attrition Label: Indicates whether an employee left the company (attrition) or stayed.

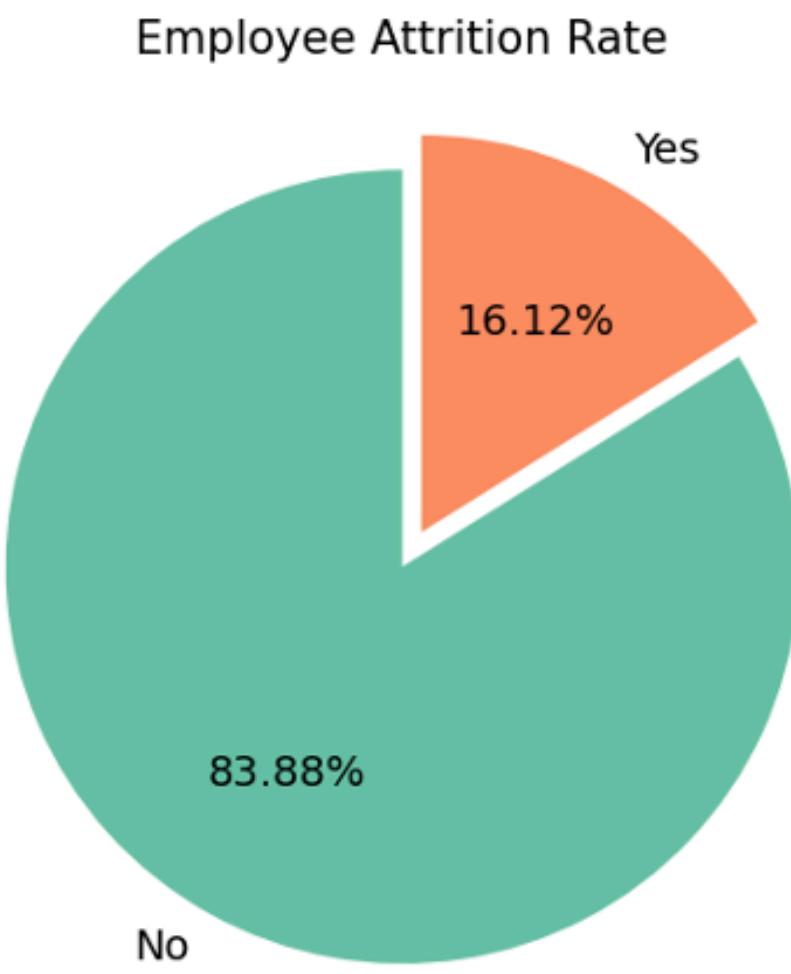
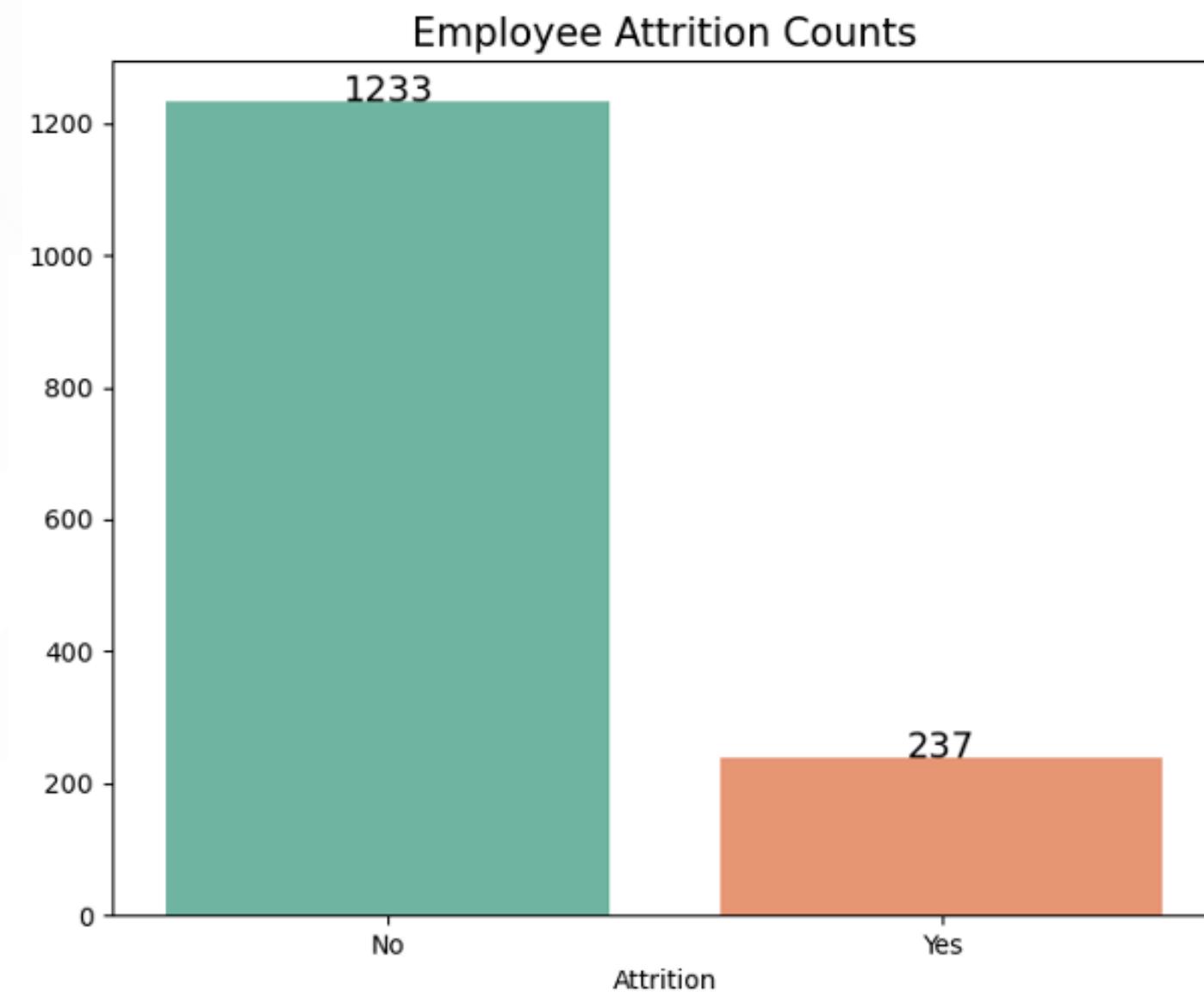
Potential Uses:

1. Attrition Prediction: The dataset is commonly used to build machine learning models that predict whether an employee is likely to leave based on various factors.
2. Feature Importance: Helps identify the most important factors contributing to employee attrition.
3. Employee Retention Strategies: Insights from the analysis can help HR departments design better retention strategies.



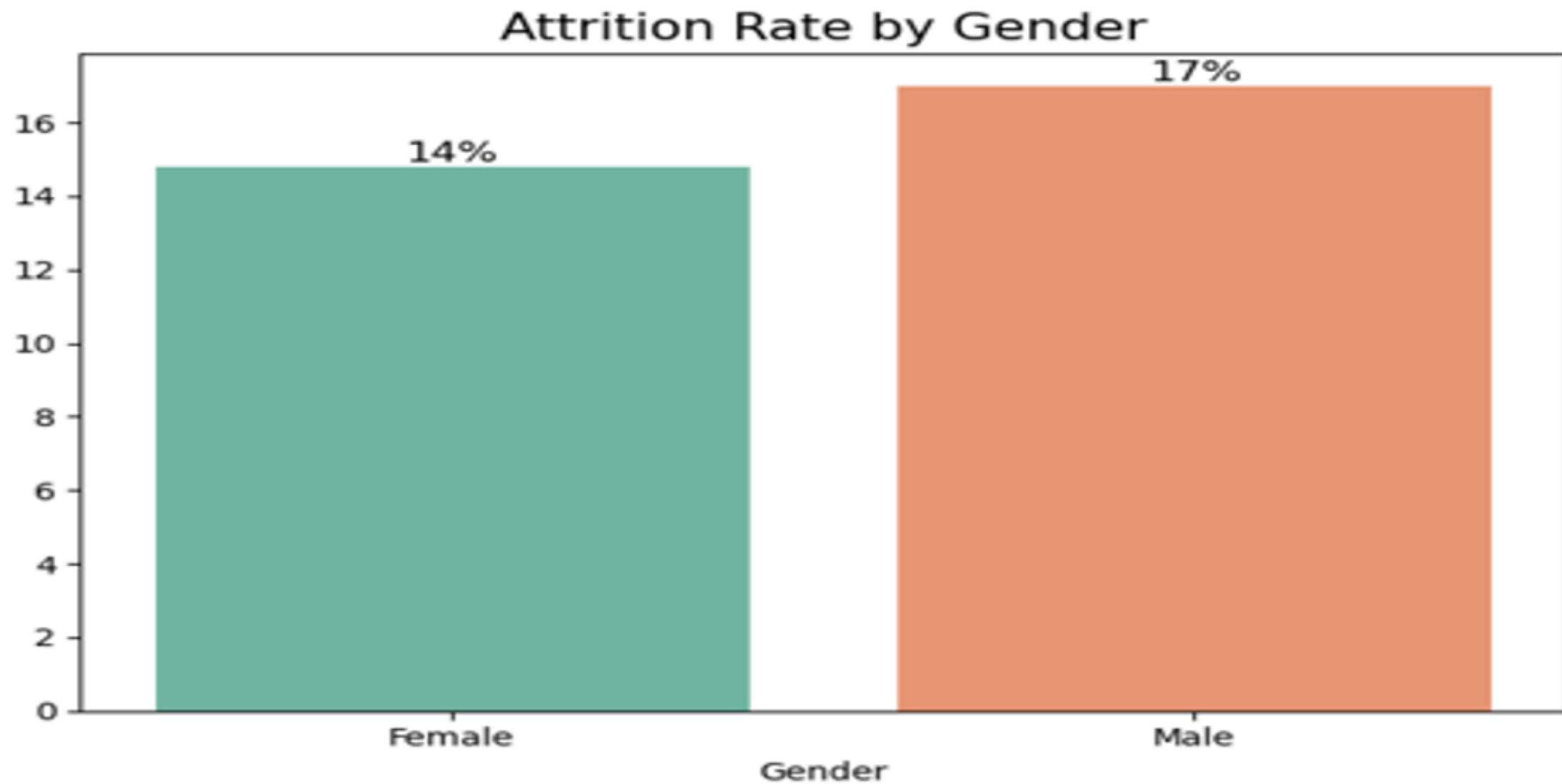
kaggle™





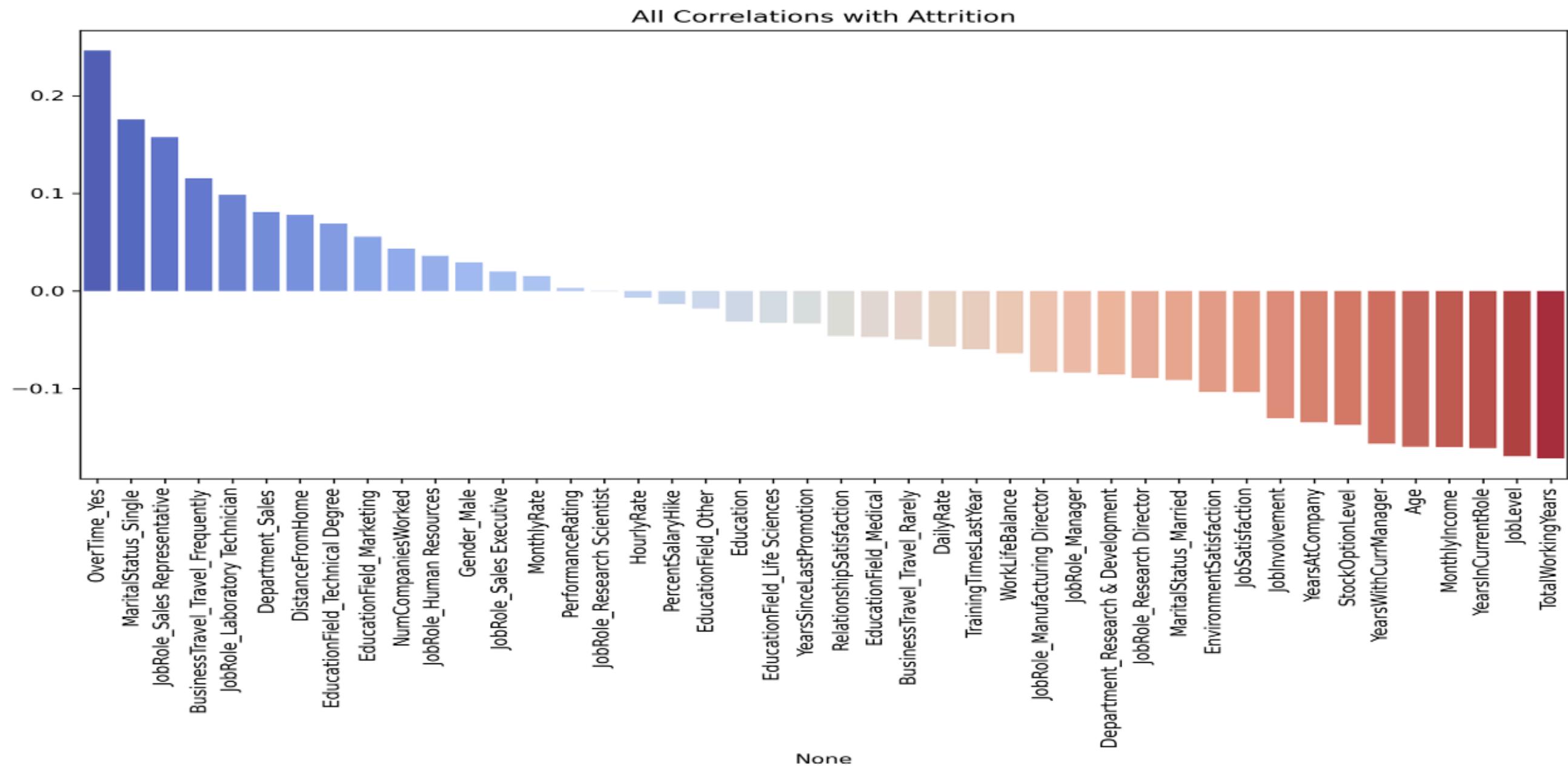
Inference:

1. The employee attrition rate of this organization is 16.12%.
2. According to experts in the field of Human Resources, says that the attrition rate 4% to 6% is normal in organization.
3. So we can say the attrition rate of the organization is at a dangerous level.
4. Therefore the organization should take measures to reduce the attrition rate.



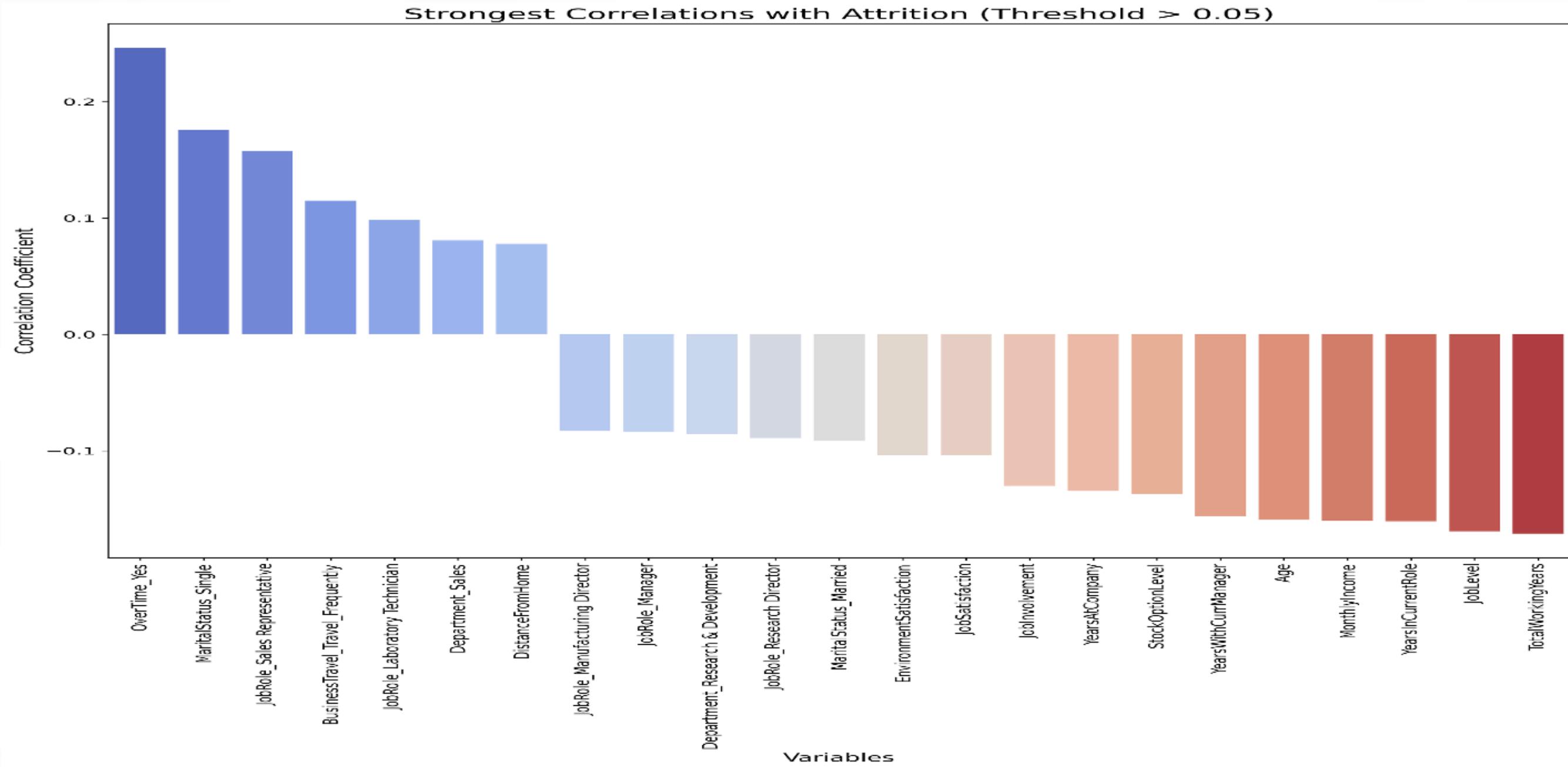
- The attrition rate among males is 17.01%, slightly higher than the overall rate.
- The attrition rate among females is 14.80%, which is lower than the overall rate and the male attrition rate.

ALL CORRELATIONS WITH ATTRITION



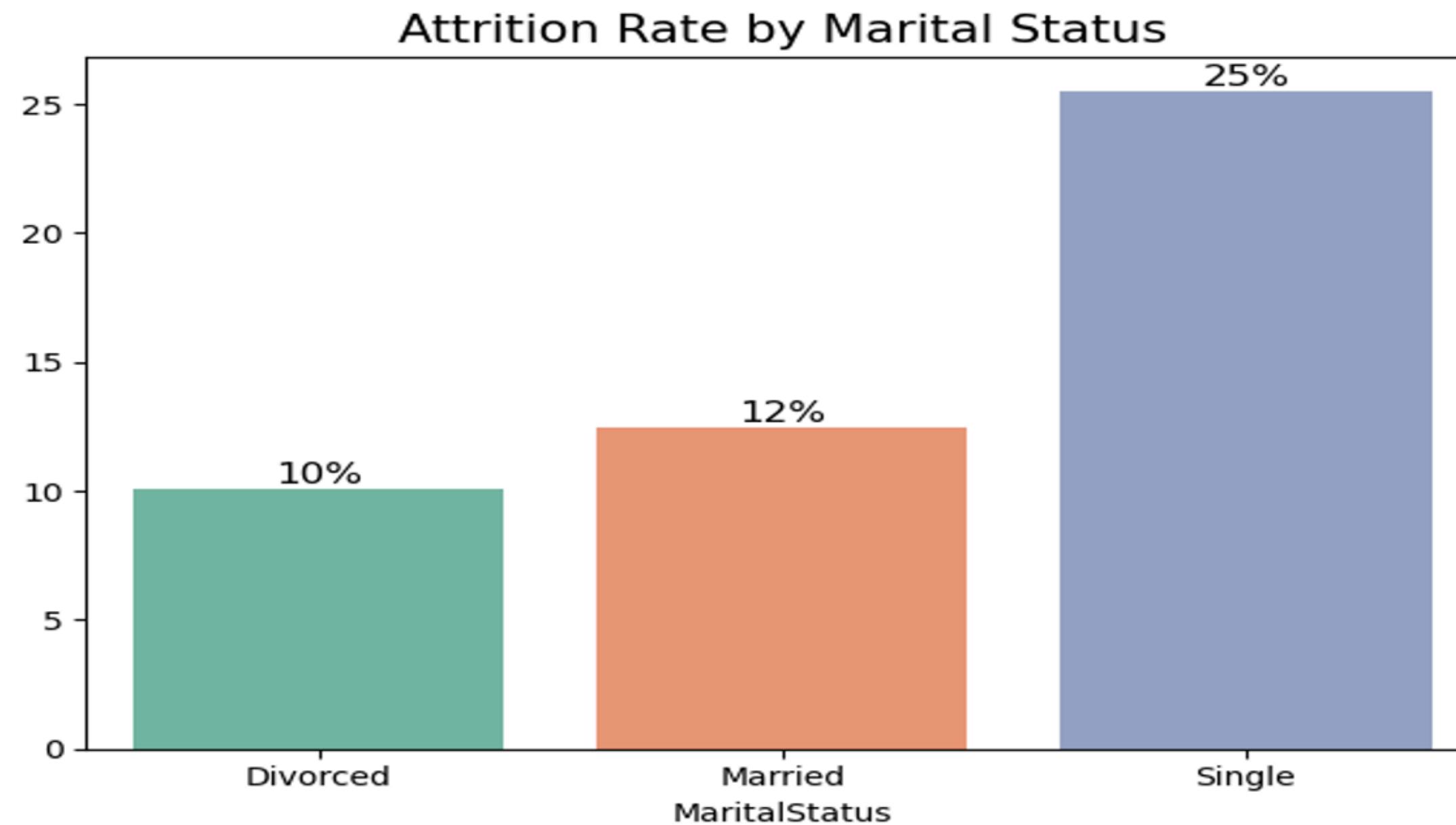
Correlation analysis shows the influence of various factors on attrition, with variables like overtime and job role being significantly correlated.

STRONGEST CORRELATIONS WITH ATTRITION (THRESHOLD > 0.05)



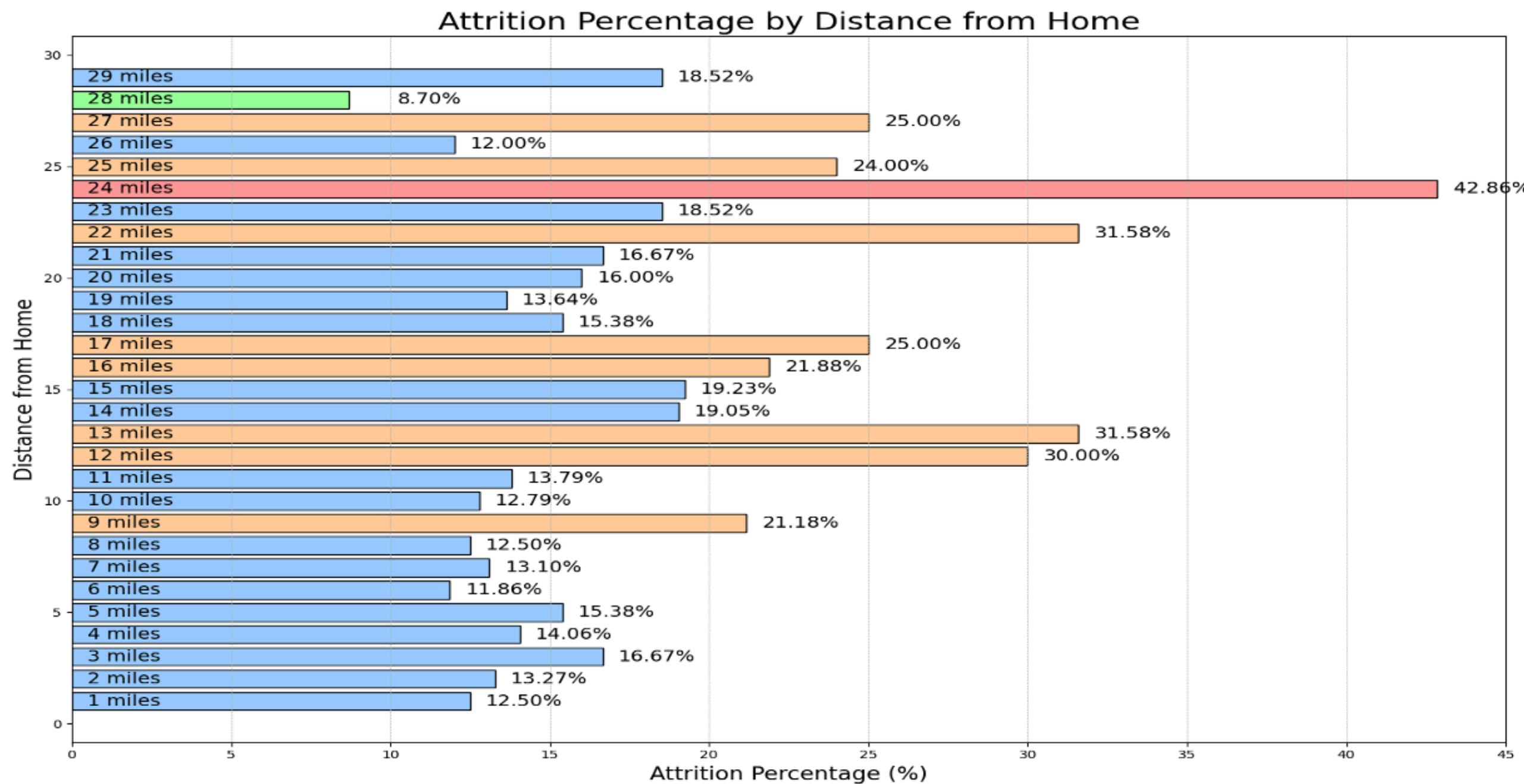
Filtering for stronger correlations reveals key factors like overtime, marital status, and specific job roles impacting attrition.

ATTRITION RATE BY MARITAL STATUS



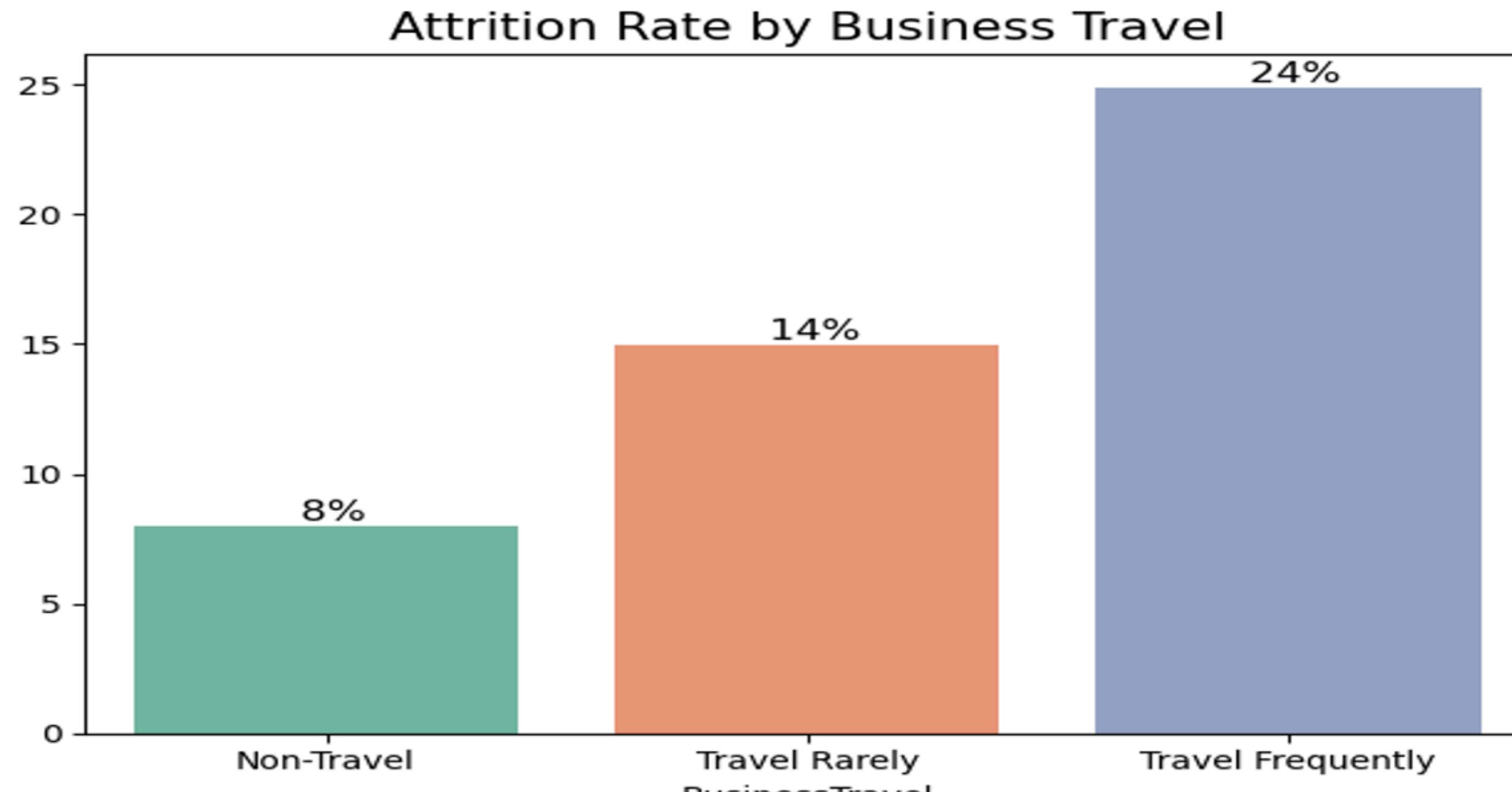
Single employees have the highest attrition rate, suggesting marital status plays a role in employee retention.

ATTRITION RATE BY DISTANCE FROM HOME



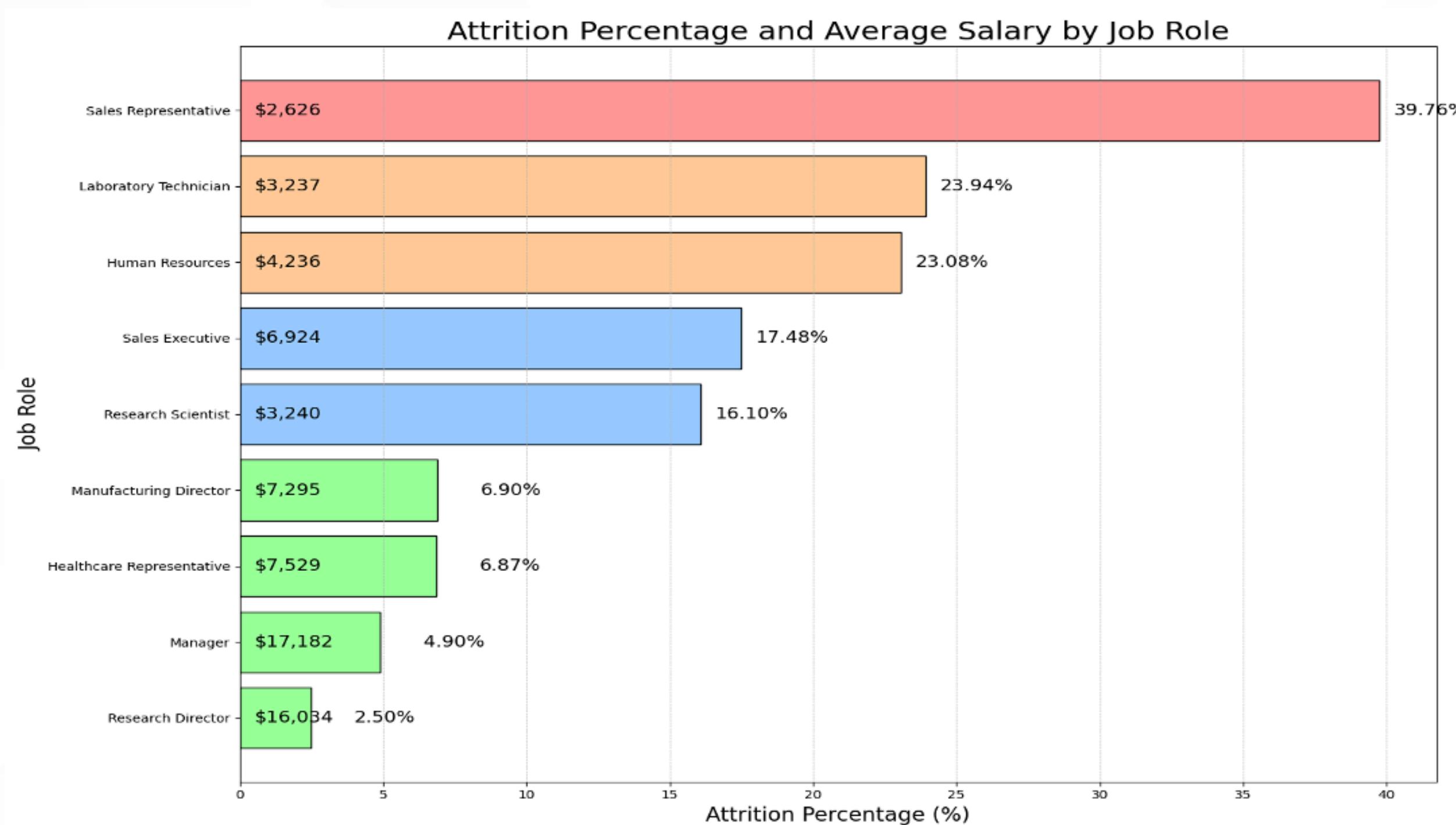
Distance from home shows mixed impact on attrition, with some living far having low attrition rates, indicating other factors might override distance.

ATTRITION RATE BY BUSINESS TRAVEL



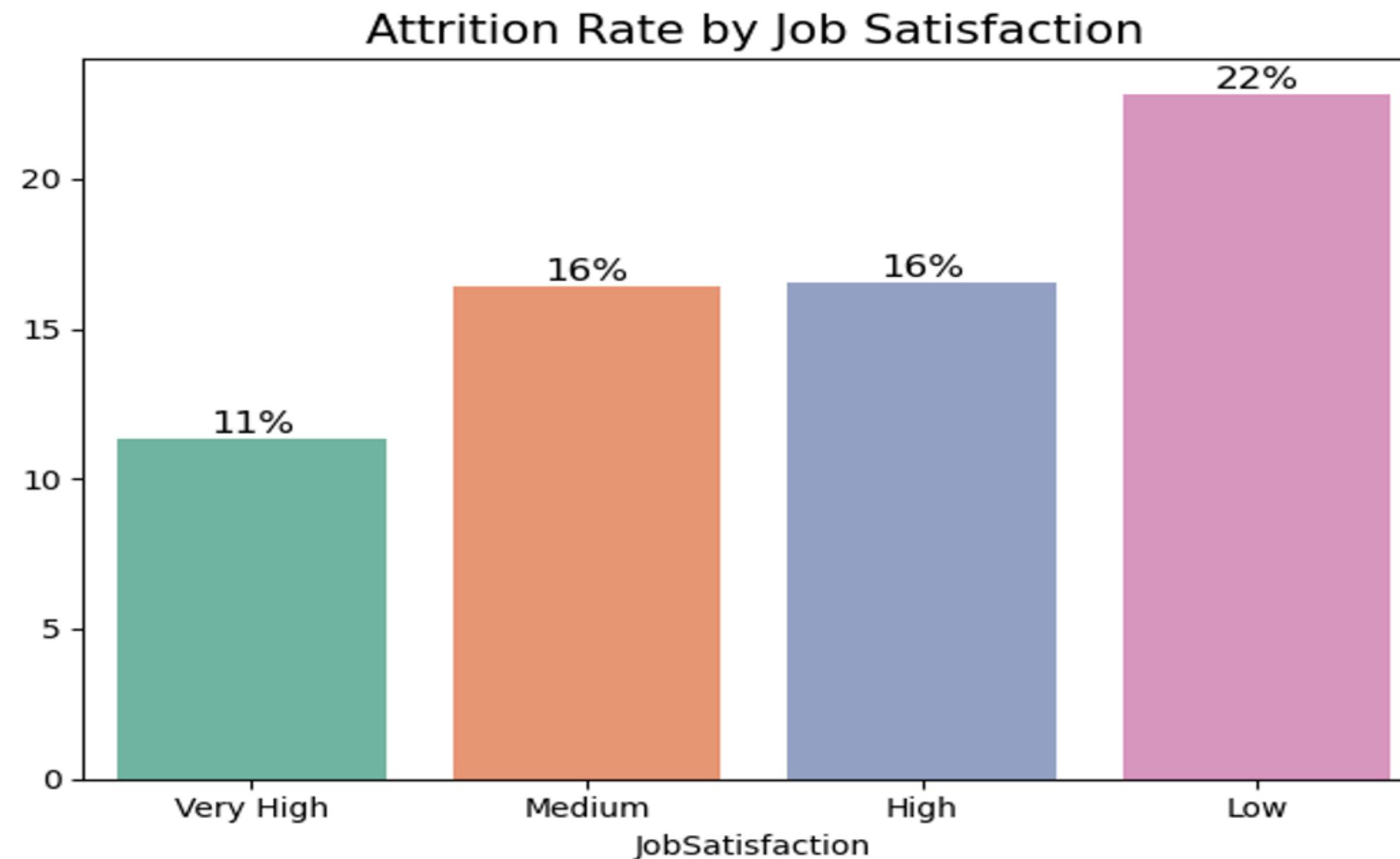
Frequent business travel is linked to higher attrition, emphasizing the strain of frequent travel on employees.

ATTRITION RATE AND AVERAGE SALARY BY JOB ROLE



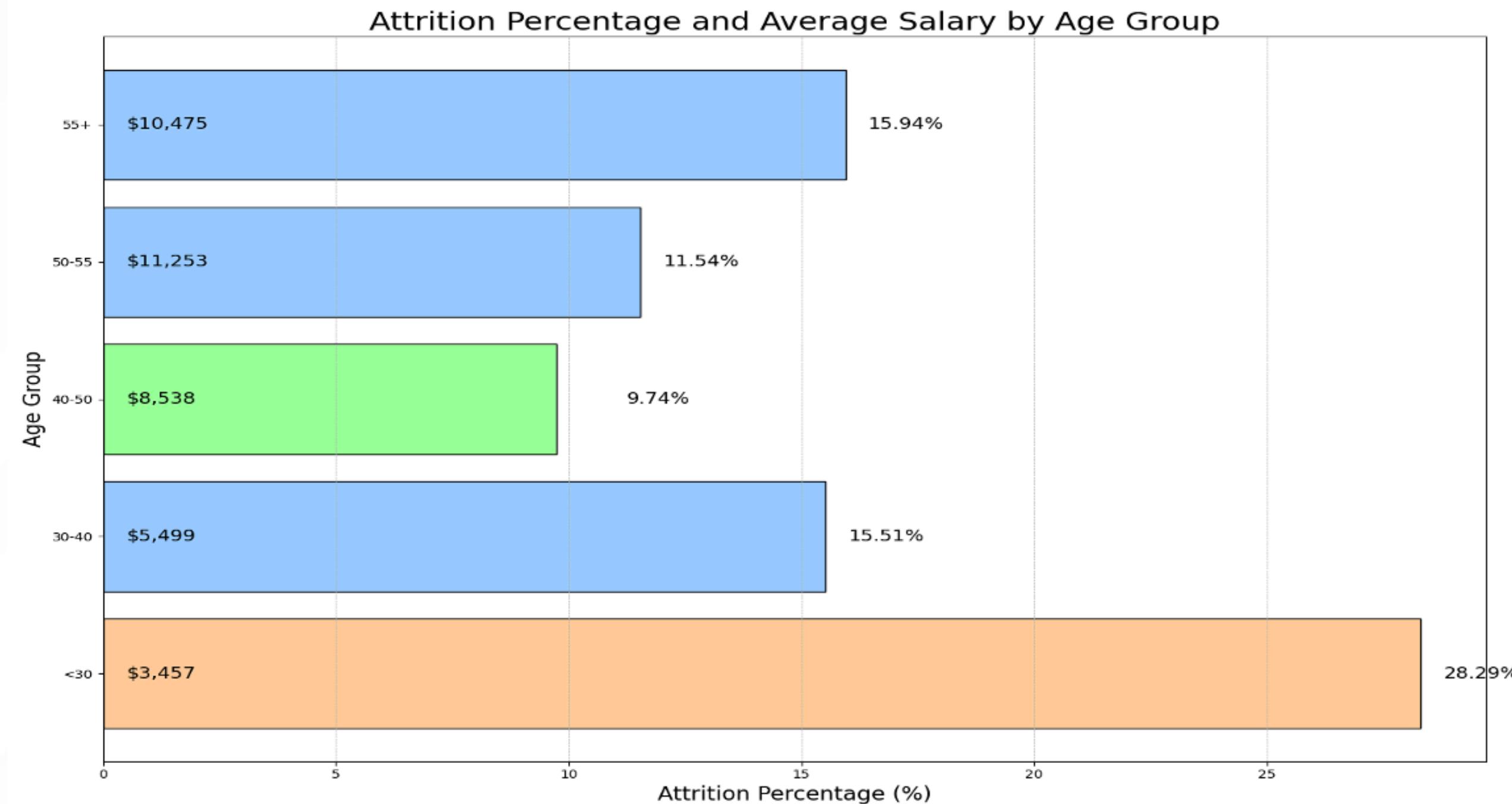
Sales roles, particularly Sales Representatives, have the highest attrition, possibly due to lower average salaries.

ATTRITION RATE BY JOB SATISFACTION



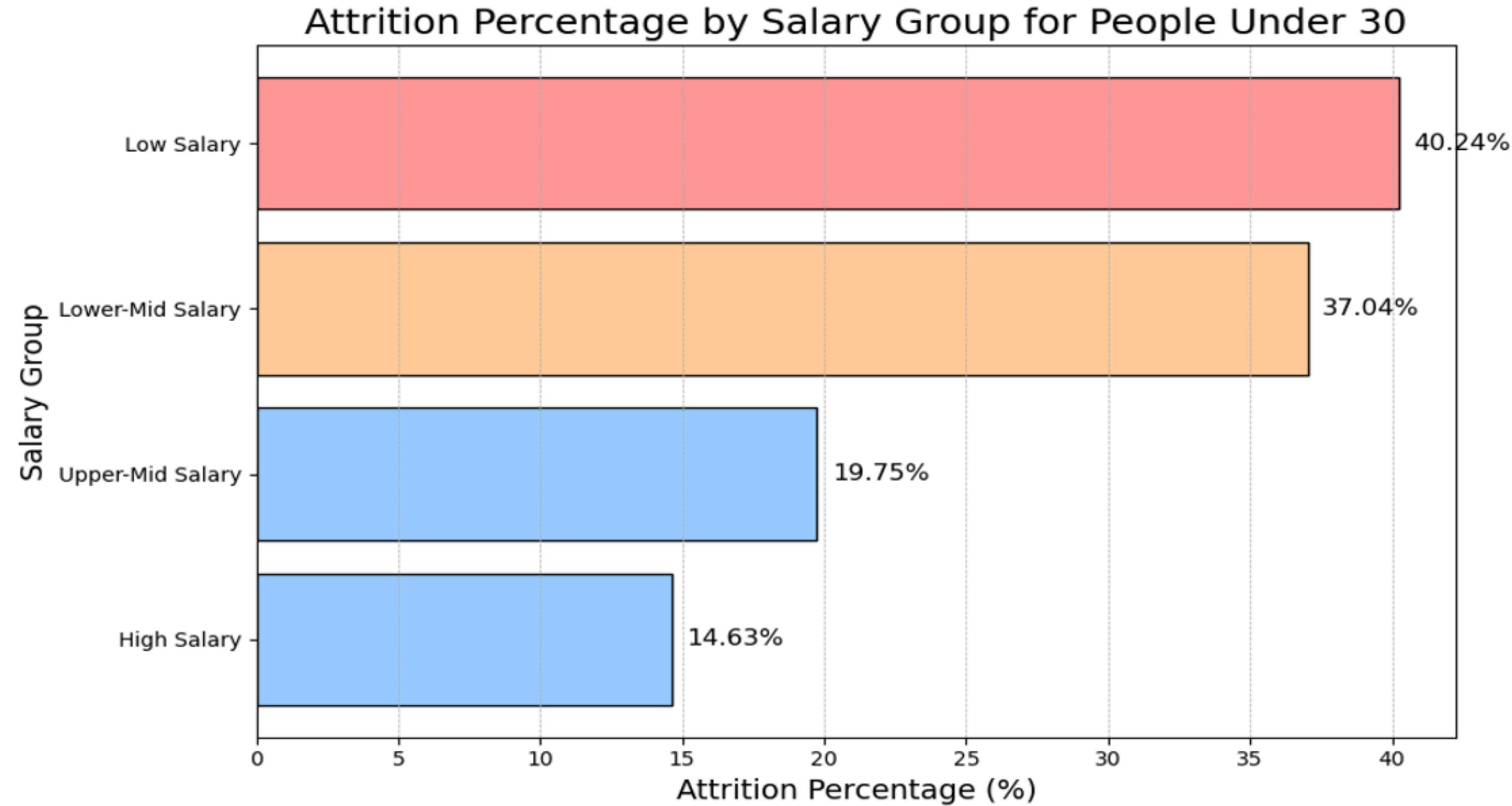
Low job satisfaction strongly correlates with higher attrition, underscoring the importance of employee engagement.

ATTRITION RATE AND AVERAGE SALARY BY AGE GROUP



Younger employees (<30) show the highest attrition, often linked to lower salaries and career growth opportunities.

ATTRITION RATE BY SALARY GROUP FOR PEOPLE UNDER 30



Lower salary groups under 30 show significantly higher attrition, emphasizing the need for competitive compensation.

MACHINE LEARNING

1. Preprocessing:

- Encode categorical variables.
- Balance the dataset using SMOTE
- Split the data into features (X) and target (y).
- Perform a train-test split.

2. Model Training:

- Train a Random Forest classifier.
- Perform hyperparameter tuning using GridSearchCV.

3. Evaluate the Model:

- Assess the model's performance using metrics like accuracy, precision, recall, and F1-score.

4. Feature Importances:

- Extract and visualize the feature importances from the trained model.

PREPROCESSING

Encode

- Encode categorical variables.

Balance

- 20% Attrition Yes
- 80% Attrition No
- Create synthetic examples of minority class using SMOTE

Split

- Split the data into features (X) and target (y).

Perform

- Perform a train-test split.
70% train
30 test

MODEL TRAINING

Model Choice:

Random Forest classifier

- An ensemble learning algorithm.
- Constructing multiple decision trees
- Handle both numerical and categorical features.
- Provides insights into feature importances

Hyperparameter Tuning:

GridSearchCV includes:

- number of trees (`n_estimators`)
- maximum tree depth (`max_depth`),
- minimum samples required to split a node (`min_samples_split`),
- minimum samples required at a leaf node (`min_samples_leaf`).

The best hyperparameters:

- `n_estimators` = 100,
- `max_depth` = 30,
- `min_samples_split` = 2,
- `min_samples_leaf` = 1.

MODEL EVALUATION

Before

Classification Report Before Hyperparameters Tuning

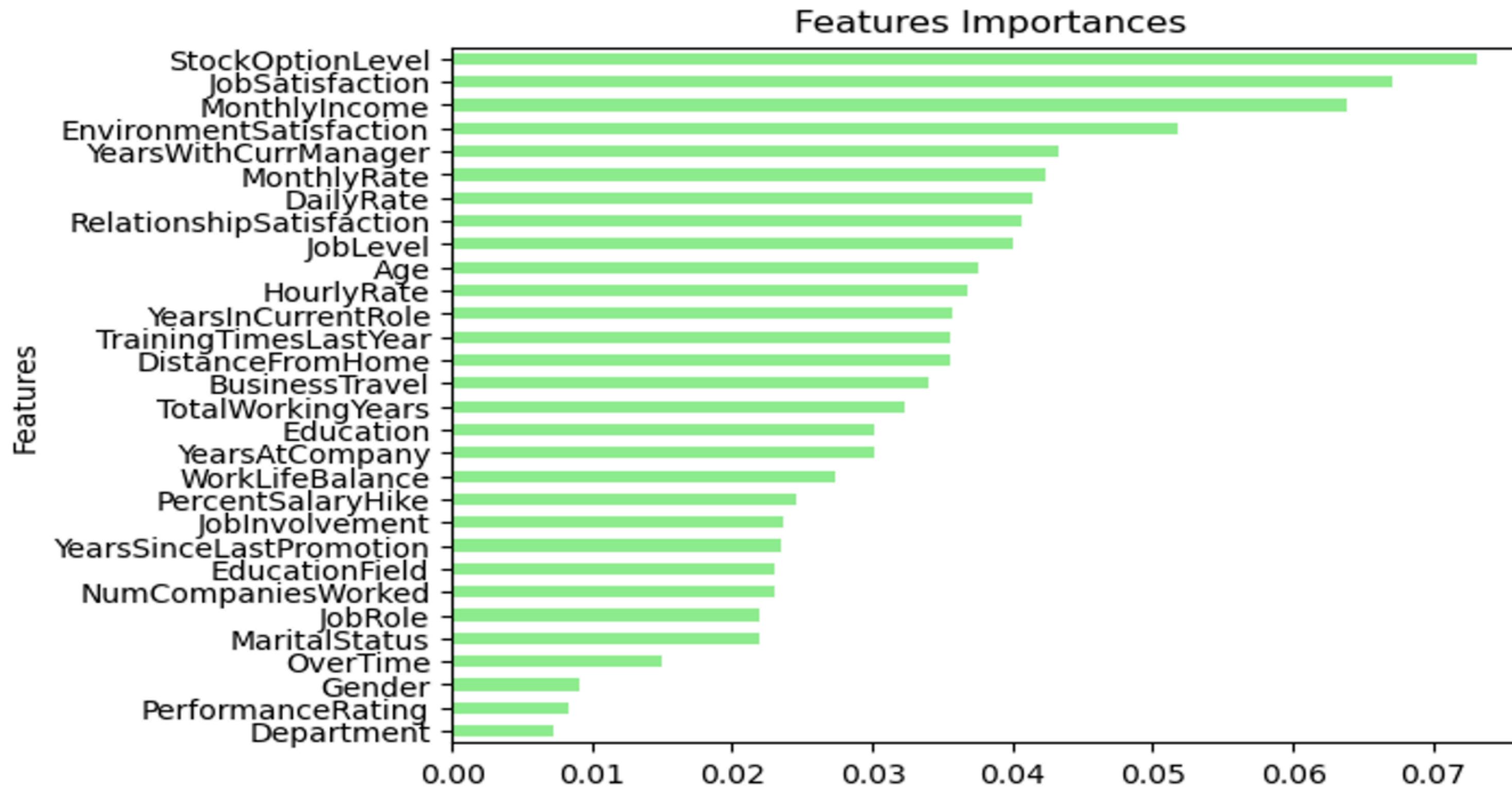
	precision	recall	f1-score	support
No Attrition	0.86	0.92	0.89	297
Attrition	0.92	0.86	0.89	320
accuracy			0.89	617
macro avg	0.89	0.89	0.89	617
weighted avg	0.89	0.89	0.89	617

After

Classification Report After Hyperparameters Tuning

	precision	recall	f1-score	support
No Attrition	0.88	0.94	0.91	297
Attrition	0.94	0.88	0.91	320
accuracy			0.91	617
macro avg	0.91	0.91	0.91	617
weighted avg	0.91	0.91	0.91	617

FEATURES IMPORTANCES





MODEL DEPLOYMENT

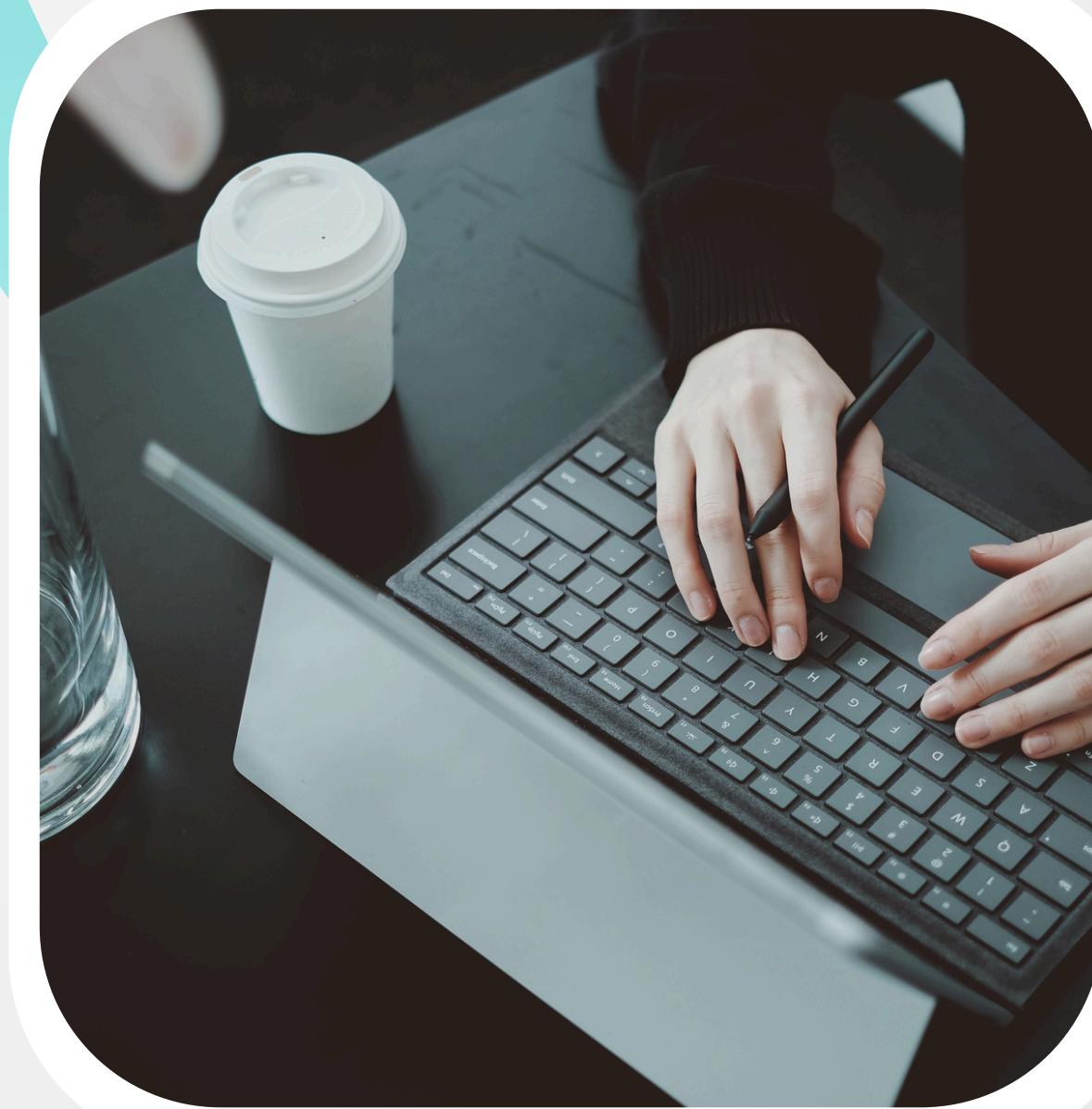


SUMMARY OF FINDINGS

- The Random Forest has a high accuracy score at 91%
- The Random Forest model has identified key factors contributing to employee attrition, with Stock Option Level, Salary and Job Satisfaction being the most influential.
- Practical Implications:
 - HR departments should consider focusing on competitive salary packages and ensuring high levels of job satisfaction to reduce attrition. Younger employees might benefit from targeted retention programs.
- Next Steps:
 - To further improve the model, exploring other models like Gradient Boosting Machines or tuning more hyperparameters could yield better results.

CONCLUSION

- In conclusion, we embarked on a comprehensive analysis of Analytics Attrition Dataset, from data loading to model evaluation.
- The results and visualizations generated throughout the process provide valuable information for decision-makers and professionals seeking to understand and mitigate employee attrition within the organization.
- The optimized model is now more reliable and capable of providing actionable insights for HR departments to identify employees at risk of leaving. By implementing such a model, organizations can proactively address attrition risks, thereby reducing turnover and retaining key talent.
- This project showcases the power of data analysis and machine learning in addressing real-world business challenges.



THANK YOU FOR ATTENTION

