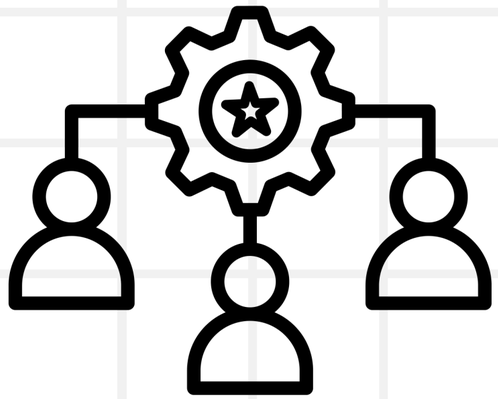
INST 737
HR ANALYTICS



Presentation by Harshitha Ramachandra

INTRODUCTION

The primary objective of this project is to leverage HR analytics using the R programming language to gain insights into employee attrition patterns and identify key factors influencing turnover.

Importance

- Understanding employee attrition is crucial for organizations to develop effective retention strategies, improve employee satisfaction and maintain workforce stability.
- By utilizing data-driven approaches, actionable insights are inferred that can inform HR decision-making and contribute to organizational success.

DATASET AND PROJECT STRUCTURE

Dataset Overview:

- The HR Analytics Dataset, sourced from Kaggle, contains comprehensive information on employee demographics, performance metrics, and attrition status.
- With approximately 15,000 employee observations and 10 features, the dataset provides a rich source of information for analysis.

Project Components:

- Data Collection: Obtaining the dataset from Kaggle and ensuring its well-organized structure.
- Data Cleaning: Handling missing values, outliers, and inconsistencies to prepare the data for analysis.
- Exploratory Data Analysis (EDA): Identifying underlying patterns and relationships within the dataset.
- Data Modeling: Building predictive models using machine learning algorithms to forecast employee attrition.
- Data Interpretation: Analyzing model results and deriving actionable insights for HR strategies.

EXPLORATORY DATA ANALYSIS (EDA)

Statistical Overview

- The dataset comprises approximately 15,000 employee observations and includes 10 features.
- The turnover rate within the company stands at about 24%, indicating a significant proportion of employees leave.
- On average, employee satisfaction levels are at 0.61, suggesting a moderate level of contentment within the workforce.

Attrition Analysis

Attrition (turnover) rate: 23.81%

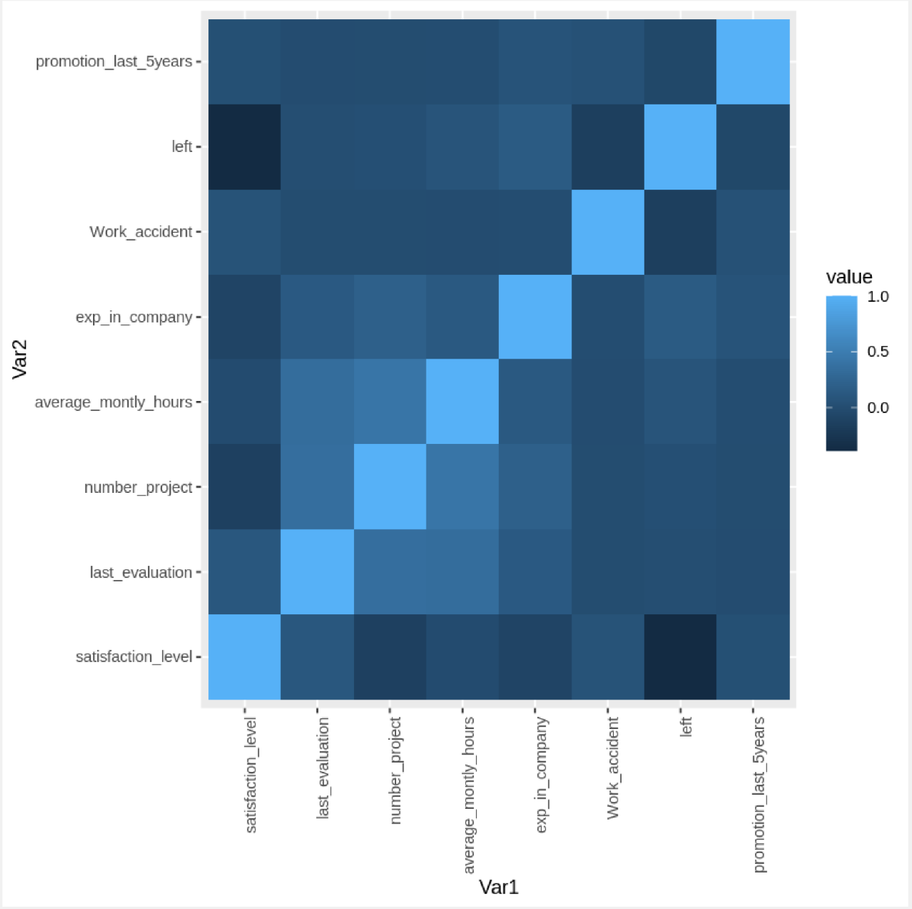
Distribution of turnover vs. non-turnover:

Turnover (1): 3,571 observations

Non-turnover (0): 11,428 observations

Category	Satisfaction level	Last evaluation	Number of projects	Average Monthly Hours	Experience in Company	Work Accident	Promotion in Last 5 Years
Turnover	0.44	0.72	3.86	207.42	3.88	0.05	0.005
Non-turnover	0.67	0.72	3.79	199.06	3.38	0.18	0.026

CORRELATION MATRIX & HEATMAP



Moderate Positive Correlations

- Project Count vs. Evaluation: 0.35
- Project Count vs. Average Monthly Hours: 0.42
- Average Monthly Hours vs. Evaluation: 0.34

Moderate Negative Correlation

- Satisfaction vs. Turnover: -0.39

- From the heatmap, there is a positive(+) correlation between projectCount, averageMonthlyHours, and evaluation. This could mean that the employees who spent more hours and did more projects were evaluated highly.
- For the negative(-) relationships, turnover and satisfaction are highly correlated. I'm assuming that people tend to leave a company more when they are less satisfied.

STATISTICAL ANALYSIS

One-Sample T-Test (Measuring Satisfaction Level)

- Hypothesis Testing:
  - Null Hypothesis (H0: pTS = pES): There is no difference in satisfaction level between employees who had turnover and the entire employee population.
  - Alternate Hypothesis (HA: pTS != pES): There is a difference in satisfaction level between employees who had turnover and the entire employee population.

Interpretation

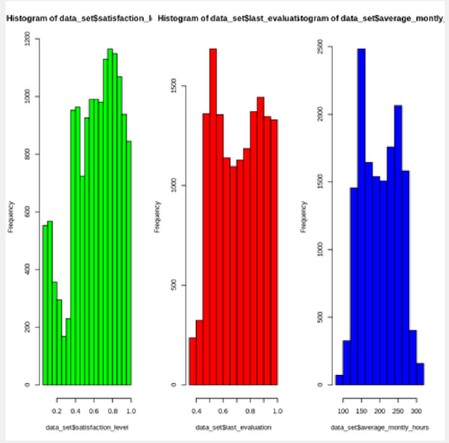
- The test result indicates a significant difference between the mean satisfaction of employees who had turnover and the entire employee population.
- The calculated t-statistic lies far outside the quantiles of the t-distribution corresponding to our confidence level and degrees of freedom.
- With a p-value of 9.01e-279 (much lower than the confidence level of 5%), we reject the null hypothesis.

T-Test Summary:

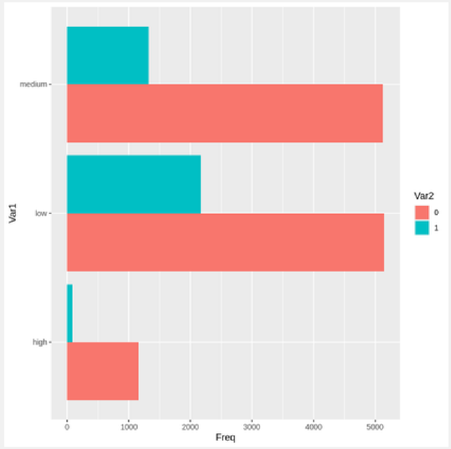
- T-Test = -39.109
- P-Value = 9.01e-279
- Reject Null Hypothesis

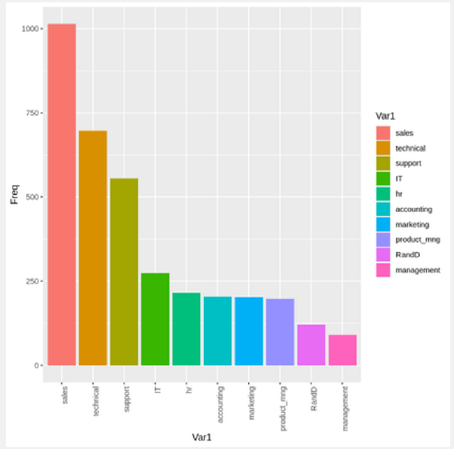
Based on the statistical analysis of a one-sample t-test, there is a significant difference in satisfaction level between employees who had turnover and the entire employee population.

DATA VISUALIZATIONS

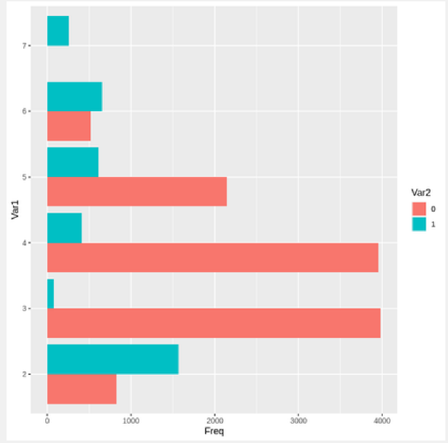


Employees with lower satisfaction levels were more likely to turnover, indicating a potential link between satisfaction and retention.



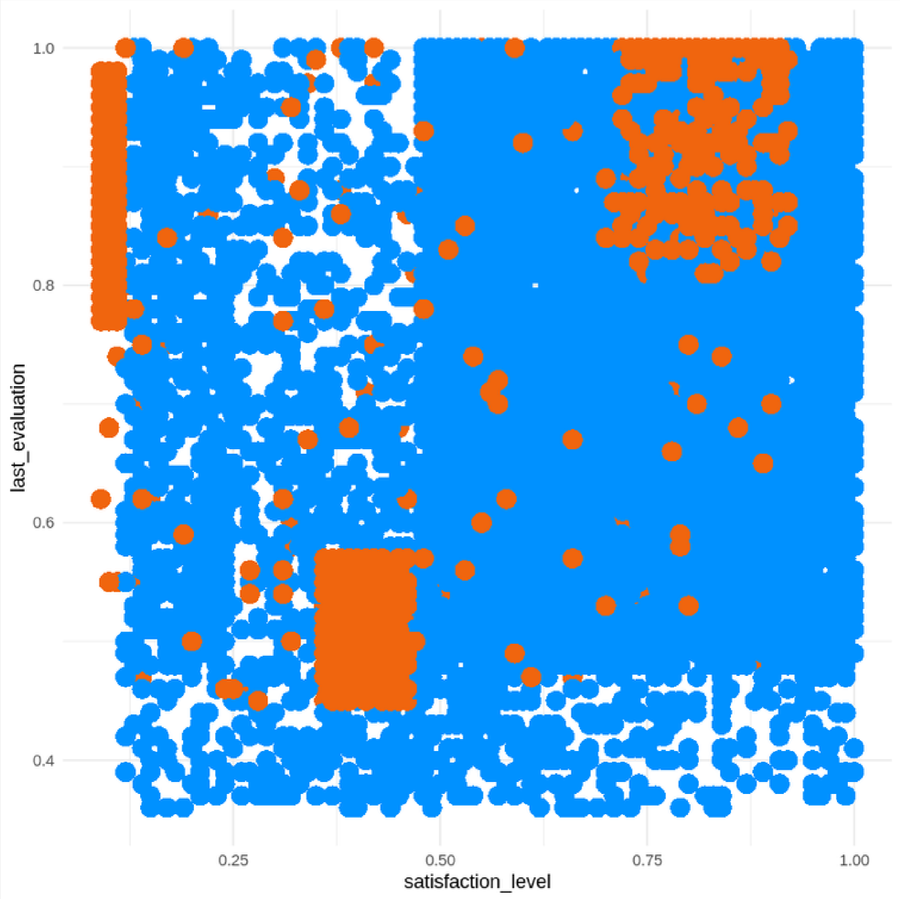


Turnover rates are higher among employees with low to average salaries, as well as those in sales, technical, and support departments.



Employees left more often when handling projects 2, 6, and 7, while those staying usually managed 3, 4, and 5 projects. Notably, everyone with 7 projects left, suggesting a link between project count and turnover.

DATA VISUALIZATIONS



A compelling observation is the presence of distinct clusters representing different employee profiles based on satisfaction and evaluation scores, potentially indicating underlying reasons for turnover, such as overwork or underperformance.

DATA MODELING - LOGISTIC REGRESSION

Objective: Predict employee attrition using Logistic Regression.
Methodology: Utilizes logistic regression to predict binary outcomes (0 or 1) based on various predictors.
Advantages: Interpretable model allowing understanding of factors influencing attrition and predictive capability for identifying probabilities of employee turnover.

Model Implementation:
Data Splitting: 75% Training Set, 25% Test Set.
Model Training: Iterative process of manipulating variables to identify significant predictors.
Packages Used: gmodels, Hmisc, caTools, ROCR.

Model Summary:
Coefficients: Factors such as satisfaction level, last evaluation, average monthly hours, salary (low, medium), role, and number of projects.
Model Evaluation: Accuracy (75.71%), Precision (69.83%), Recall (50.27%).

Interpretation: Significant predictors include satisfaction level, last evaluation, average monthly hours, salary, role and number of projects.

DATA MODELING - RANDOM FOREST

Random Forest model outperformed Logistic Regression.

Methodology:

Utilized Random Forest algorithm.

Trained with variables: satisfaction level, last evaluation, average monthly hours, salary, role, number of projects.

Parameters: 500 trees, 3 variables randomly sampled at each split.

Model Summary

Error Rate: 2.45%

Confusion Matrix:

True Negative (0's): 8465

False Positive (0's classified as 1's): 131

False Negative (1's classified as 0's): 145

True Positive (1's): 2508

Model Evaluation

Performance Metrics

Accuracy: 97.84%

Precision: 95.44%

Recall: 95.75%

Variable Importance

Key Predictors

Satisfaction Level: 93.56

Last Evaluation: 36.67

Average Monthly Hours: 79.98

Salary, Role, Number of Projects: Varying importance levels.

Interpretation:

Random Forest model provides robust predictive capability, enhancing insights for HR strategies to mitigate attrition and improve employee retention efforts.

MODEL INTERPRETATION

LOGICTIC REGRESSION MODEL

Lower satisfaction correlates with higher turnover.
Extreme evaluation scores increase turnover risk.
Overworked or underworked employees are more likely to leave.
Low and medium salaries relate to higher turnover.
Specific roles like HR, IT, and R&D impact turnover.
More projects lead to increased turnover risk.

Model Performance
Accuracy: 75.71%
Precision: 69.83%
Recall: 50.27%

RANDOM FOREST MODEL

Satisfaction Level: Most influential predictor, indicating its strong impact on turnover.
Last Evaluation: Significant, but less influential compared to satisfaction level.
Average Monthly Hours: Moderately important, suggesting workload affects turnover.
Salary, Role, Number of Projects: Varying importance levels, contributing to overall predictive capability.

Model Performance
Accuracy: 97.84%
Precision: 95.44%
Recall: 95.75%

CONCLUSION

- Workload Balance: Employees leave when underworked (less than 150hr/month) or overworked (more than 250hr/month).
- Evaluation Impact: Extreme evaluation scores (high or low) indicate higher turnover risk.
- Salary Influence: Low to medium salaries dominate employee turnover.
- Project Allocation: Employees with 2, 6, or 7 projects are at risk of leaving.
- Satisfaction Priority: Employee satisfaction is key to reducing turnover.
- Tenure Awareness: Employees with 4 or 5 yearsAtCompany need retention focus.

Actionable Strategies

- Implement workload balancing measures.
- Revise performance evaluation processes.
- Review and adjust salary structures.
- Optimize project allocation strategies.
- Prioritize initiatives to enhance employee satisfaction.
- Develop retention strategies tailored to employees with 4 or 5 years at Company.

REFERENCES

REFERENCES 1

Jacksonchou. (2019, June 19). HR Analytics. Kaggle. https://www.kaggle.com/code/jacksonchou/hr-analytics

REFERENCES 2

Wibjorn. (n.d.-b). Microsoft Learn: Build skills that open doors in your career. https://learn.microsoft.com/

THANK YOU