A Study on Employee Attrition Prediction and Analysis

Presentation Outline

1. Business Objectives
2. Data Source
3. Methodology
4. Evaluation/Results
5. Summary

Business Objectives:

Determine the factors that lead to attrition

Predict the likeliness of attrition

Reduce employee attrition

Data Source

Source: SAMPLE DATA: HR Employee Attrition and Performance

Statistics:

- **Rows – 1470 (excluding headers)**
- **Features – 33**
- **Primary Feature – Attrition (Yes/No)**

Methodology

Approach

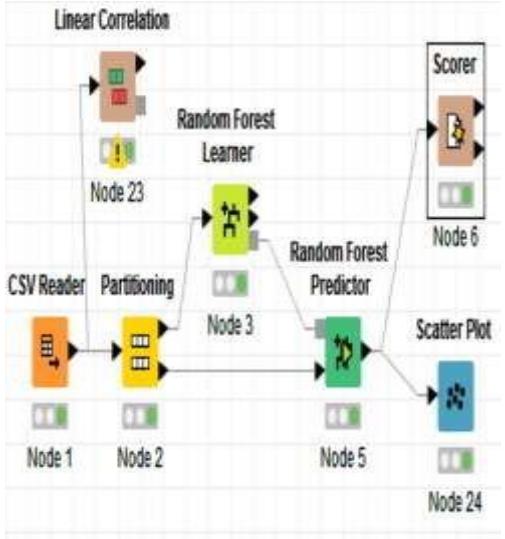
33 features extracted from raw data

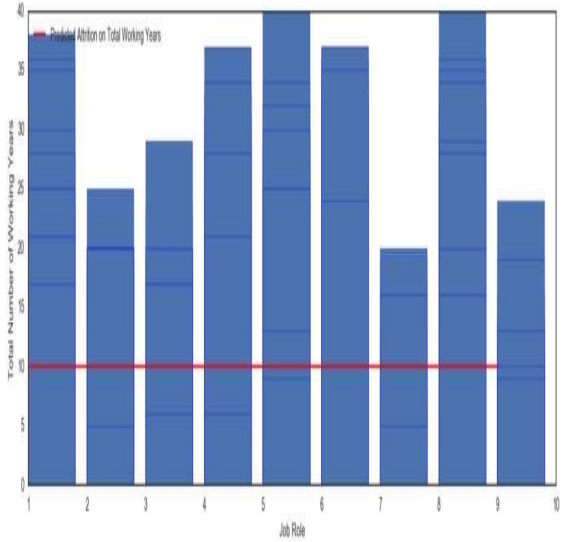
Build predictive model using multiple techniques

Optimal model identified through testing and evaluation

Identified top variables that explained attrition

Feature Name	Feature Description
Age	Age of Employee
Attrition	Did the employee left the company?
BusinessTravel	How often the employee travels for work?
DailyRate	Employee's daily rate
Department	Where the employee worked
DistanceFromHome	Distance the employee travels from home to office
Education	Employee's highest educational attainment
EducationField	The major the correspondent graduated with
EmployeeCount	Number of employee per field
EmployeeNumber	Employee ID number
EnvironmentSatisfaction	Is the candidate satisfied with their work environment
Gender	Gender of employee
HourlyRate	Employee's hourly rate
JobInvolvement	How involved the employee is with their work?
JobLevel	Job level of employee
JobRole	Employee's job role





Methodology

Machine Learning Algorithm:

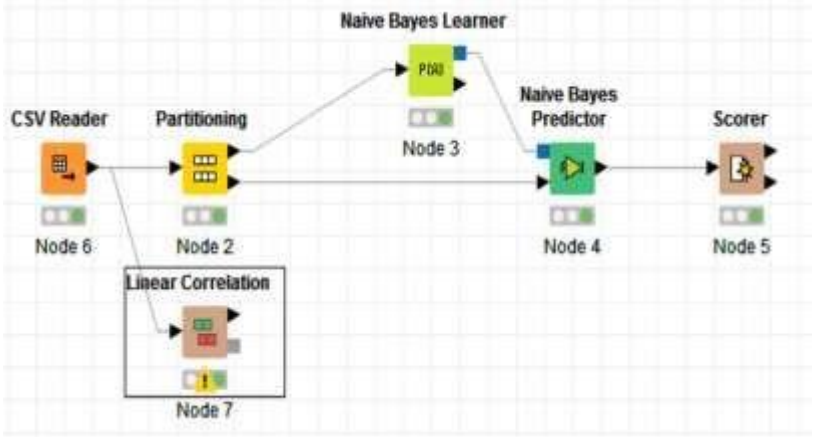
- Naïve Bayes
- Decision Tree
- Random Forest
- k-Means

Tools of Disposal:

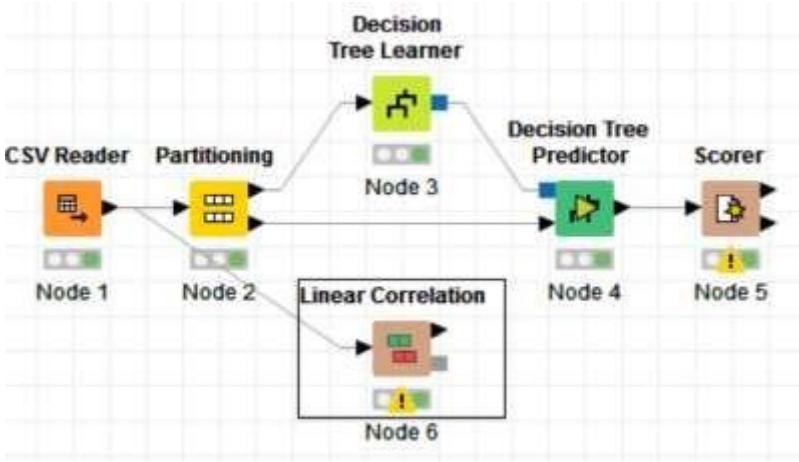
- KNIME Analytics Platform
- Jupyter Notebook (Python)
- Open Refine
- Tableau

KNIME Machine Learning Workflow

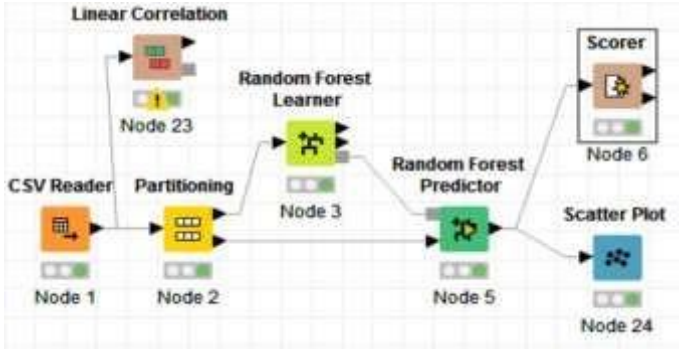
Naïve Bayes:



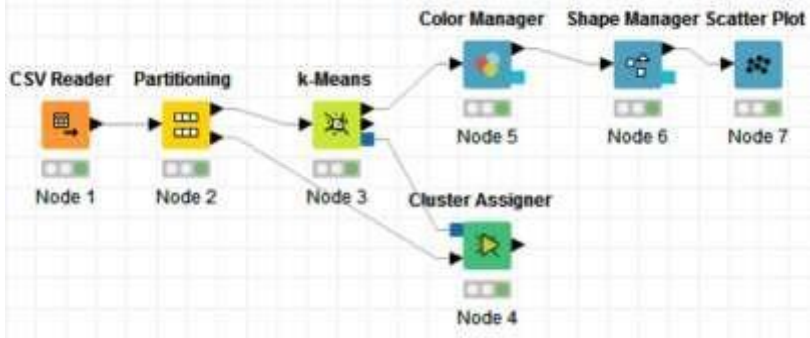
Decision Tree:



Random Forest:



K-Means:

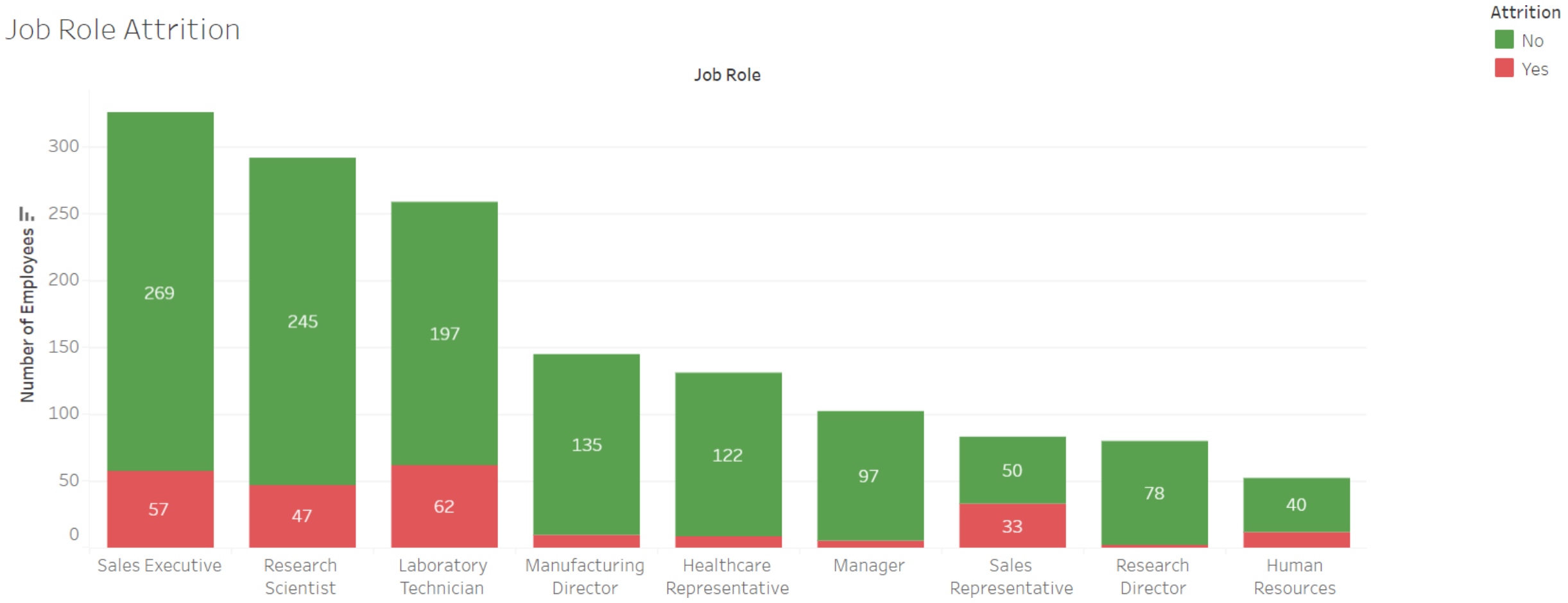


Evaluation/Results

16%

out of 1470 employees have expressed leaving the company

Sales Executive roles made up the largest segment of the database. However, Laboratory Technician roles delivered a highest rate of employees who wanted to leave the company.



Accuracy Percentage

Machine Learning Algorithm	Accuracy Percentage Score	Error Percentage
Naïve Bayes	81.4%	18.6%
Decision Tree	77.0%	23.0%
Random Forest	86.4%	13.6%

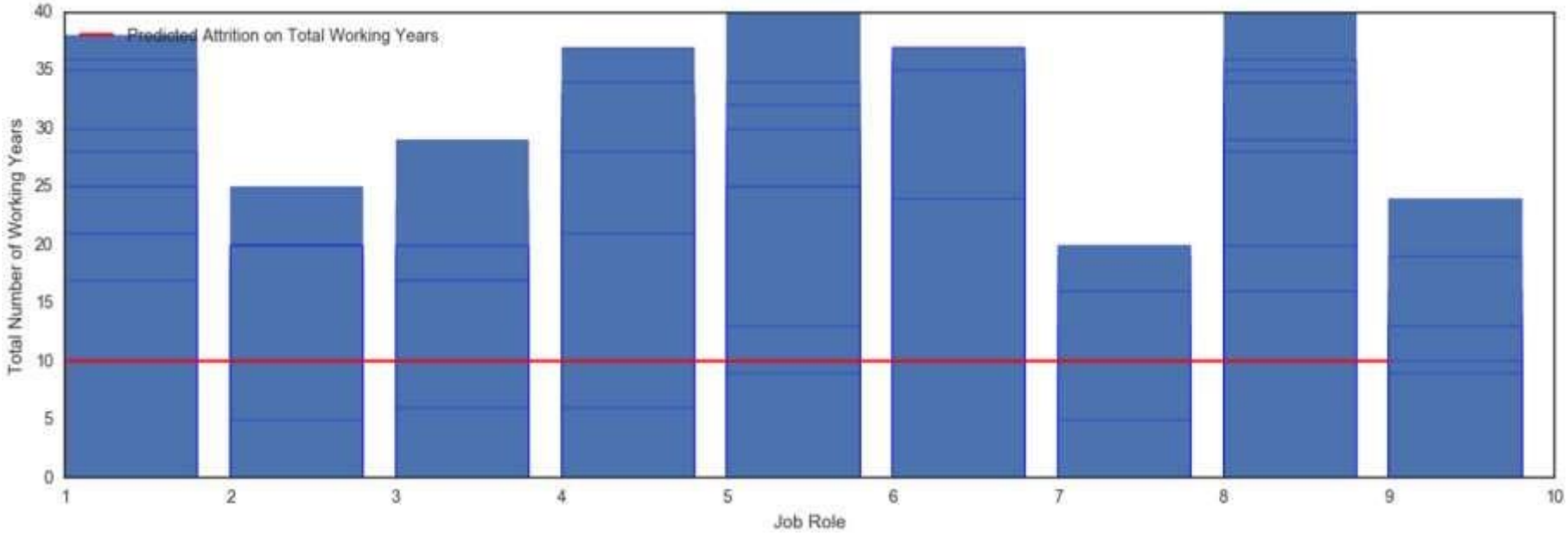
Summary of the top factors identified as impacting attrition by using Random Forest algorithm

The top 3 factors that were identified are as follows:

1. **Job Role**
2. **Total Working Years**
3. **Overtime**

Row ID	D score
JobRole	1.321
TotalWorkingYears	1.19
OverTime	1.068
MonthlyIncome	0.817
YearsAtCompany	0.782
WorkLifeBalance	0.778
MonthlyRate	0.567
Age	0.534
YearsWithCurrManager	0.533
YearsInCurrentRole	0.455
DailyRate	0.384
StockOptionLevel	0.369
MaritalStatus	0.308
JobLevel	0.304
HourlyRate	0.282
YearsSinceLastPromotion	0.238
Department	0.206
JobInvolvement	0.188
Education	0.182
PercentSalaryHike	0.171
EducationField	0.17
NumCompaniesWorked	0.133
JobSatisfaction	0.125
DistanceFromHome	0.094
BusinessTravel	0.062
EnvironmentSatisfaction	0.057
TrainingTimesLastYear	0.03
Gender	0
PerformanceRating	0
RelationshipSatisfaction	0

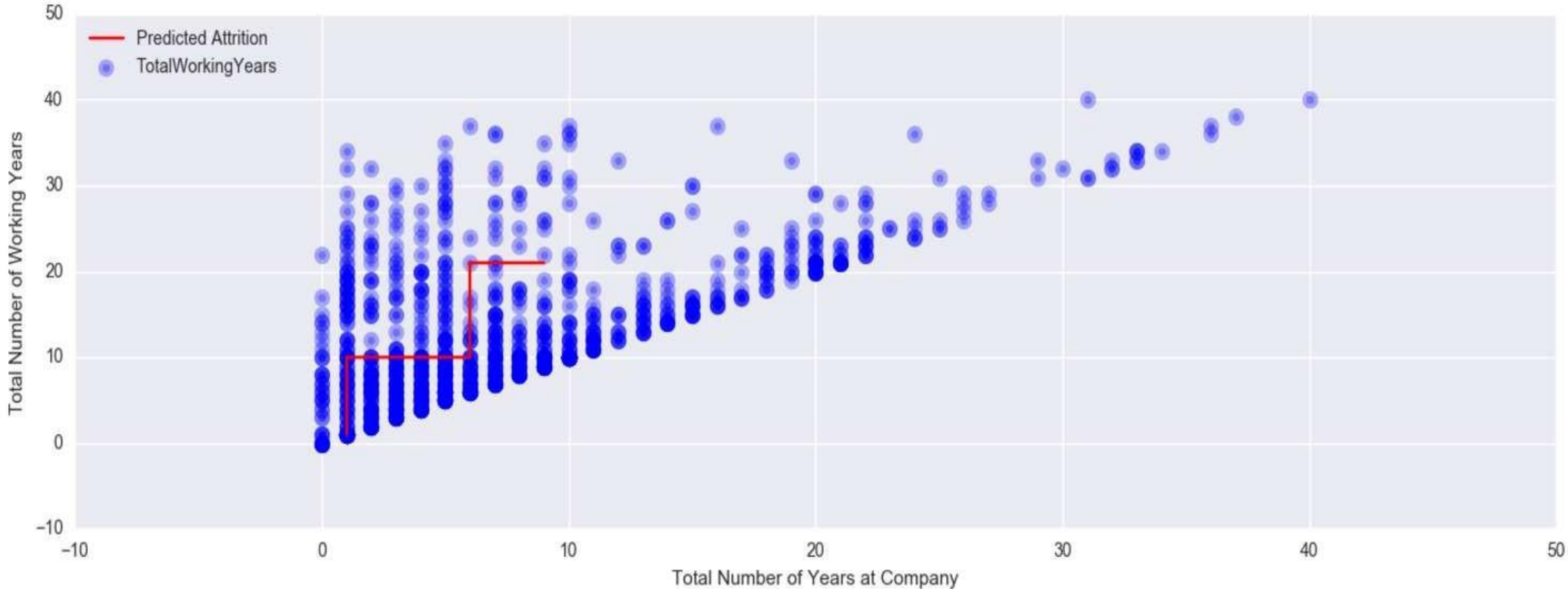
Employees with a total of 10 working years are most likely to leave by 14% for all roles



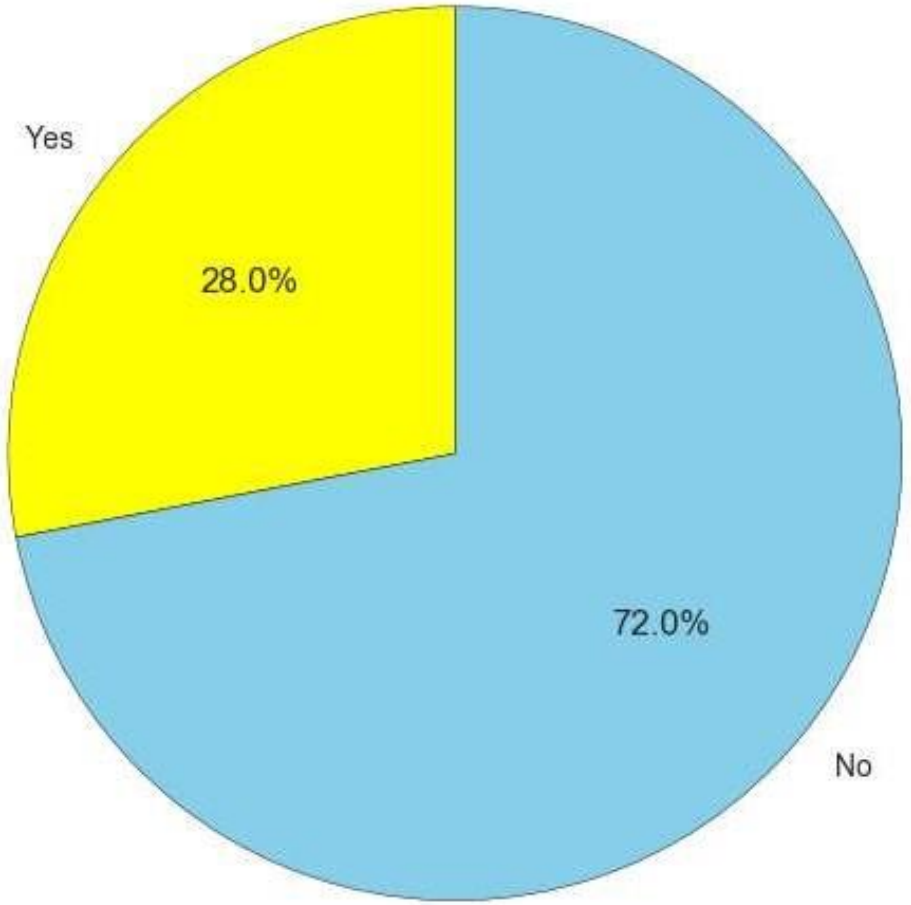
Legend:

1– Sales Executive
2– Research Scientist
3 – Laboratory Technician
4– Manufacturing Director
5– Healthcare Representative
6 – Manager
7– Sales Representative
8– Research Director
9 – Human Resources

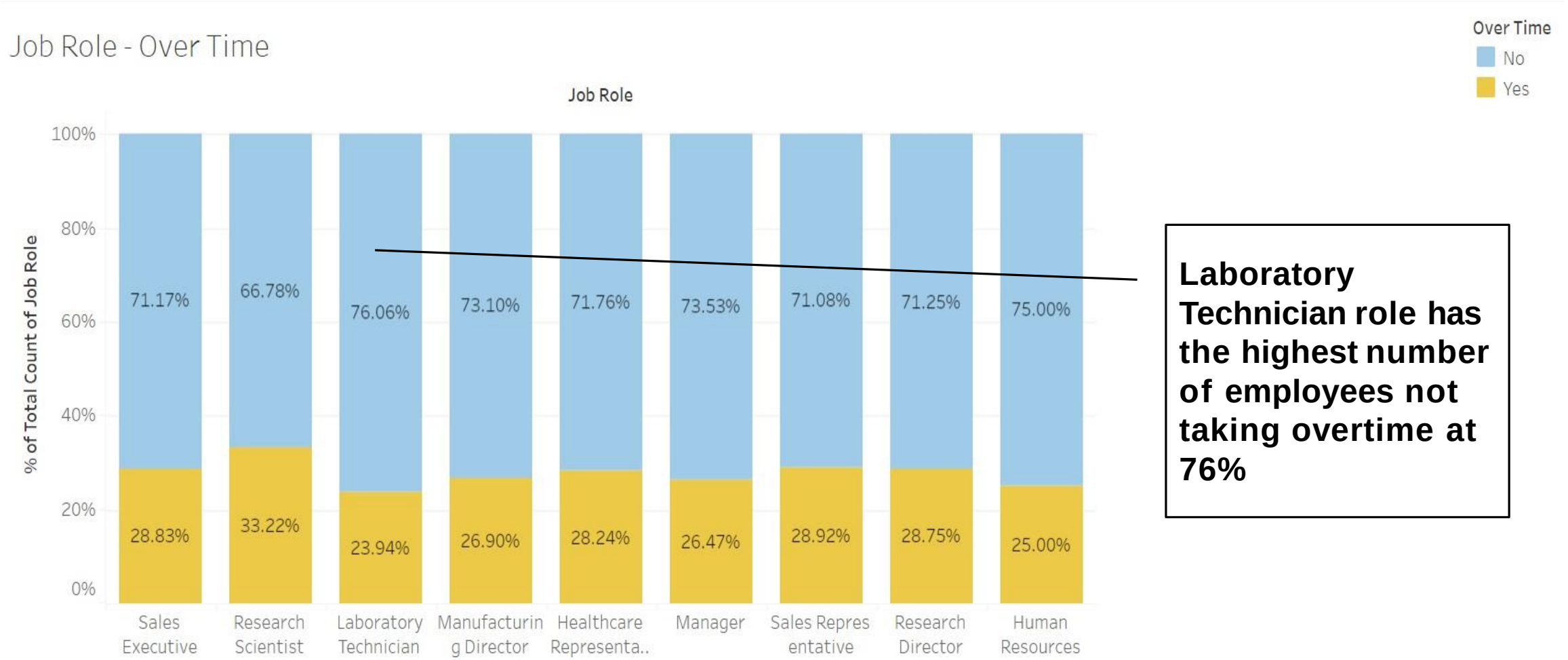
Employees who have 20 total working years are most likely to leave within 6 to 9 years of their current company company



72% of employees do not take over time, in contrast to 28% of those who did

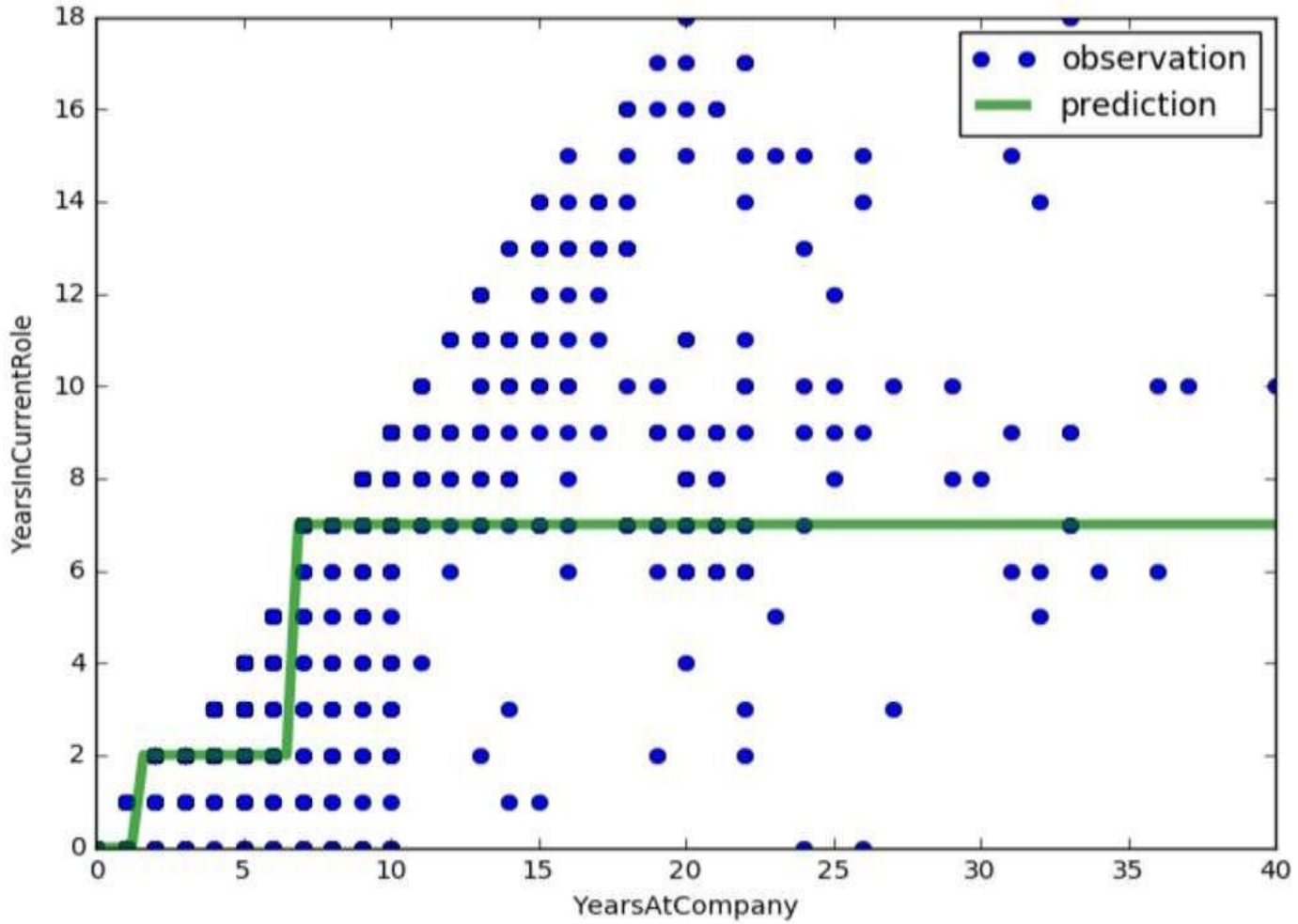


Research Scientist role has the most number of employees taking overtime at 33%

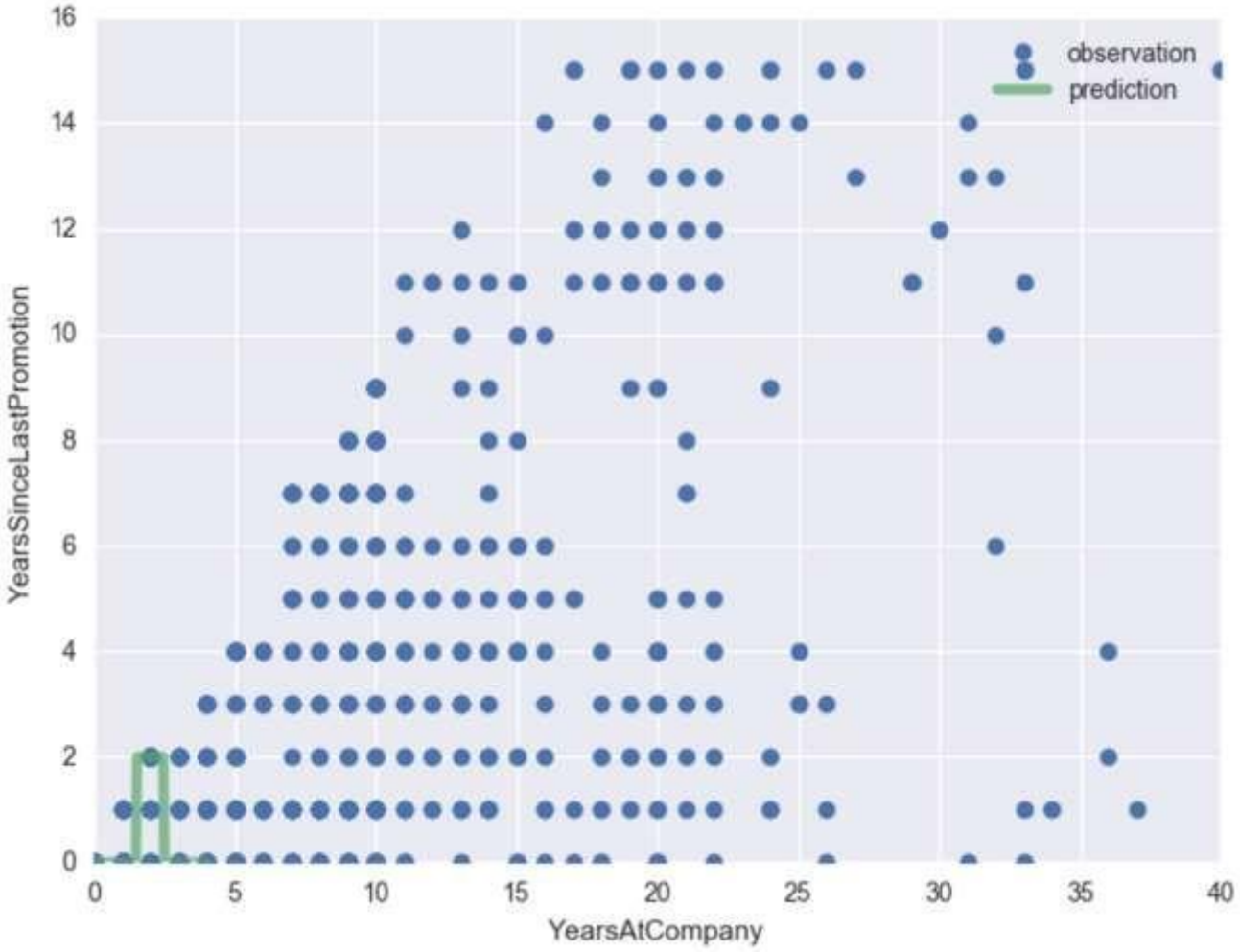


Laboratory Technician role has the highest number of employees not taking overtime at 76%

Employees are most likely to leave the company in 7 years in their current role by 15%.

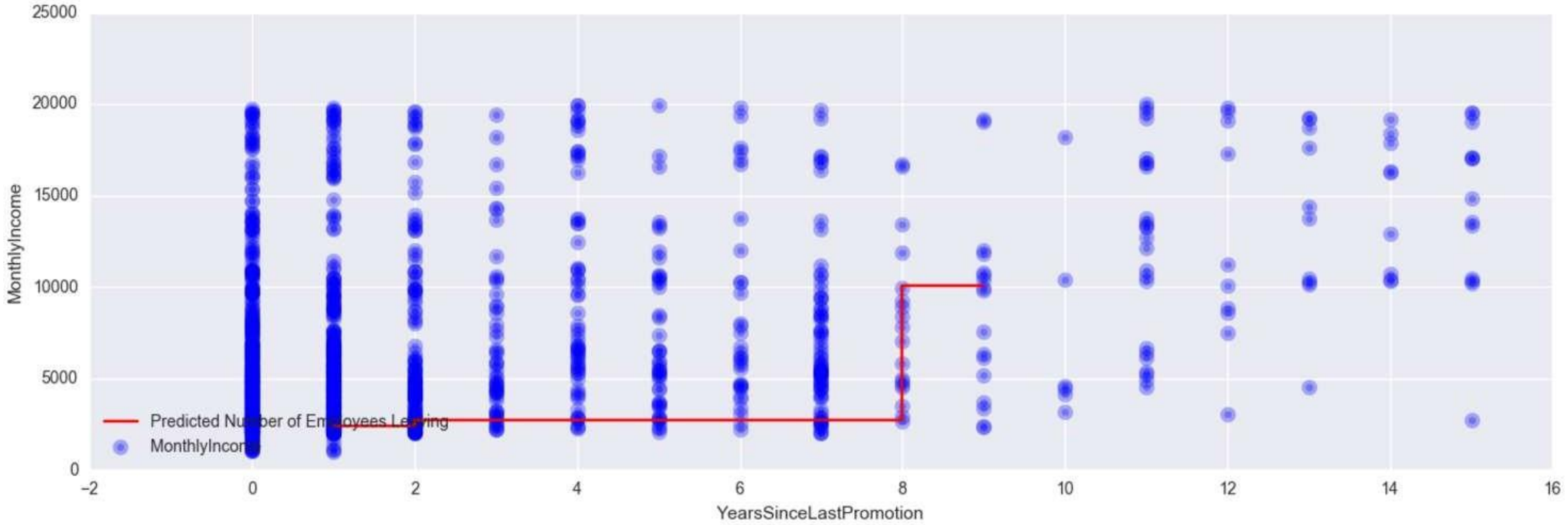


Employees who were last promoted 2 years ago, are most likely to leave the company by 7%.

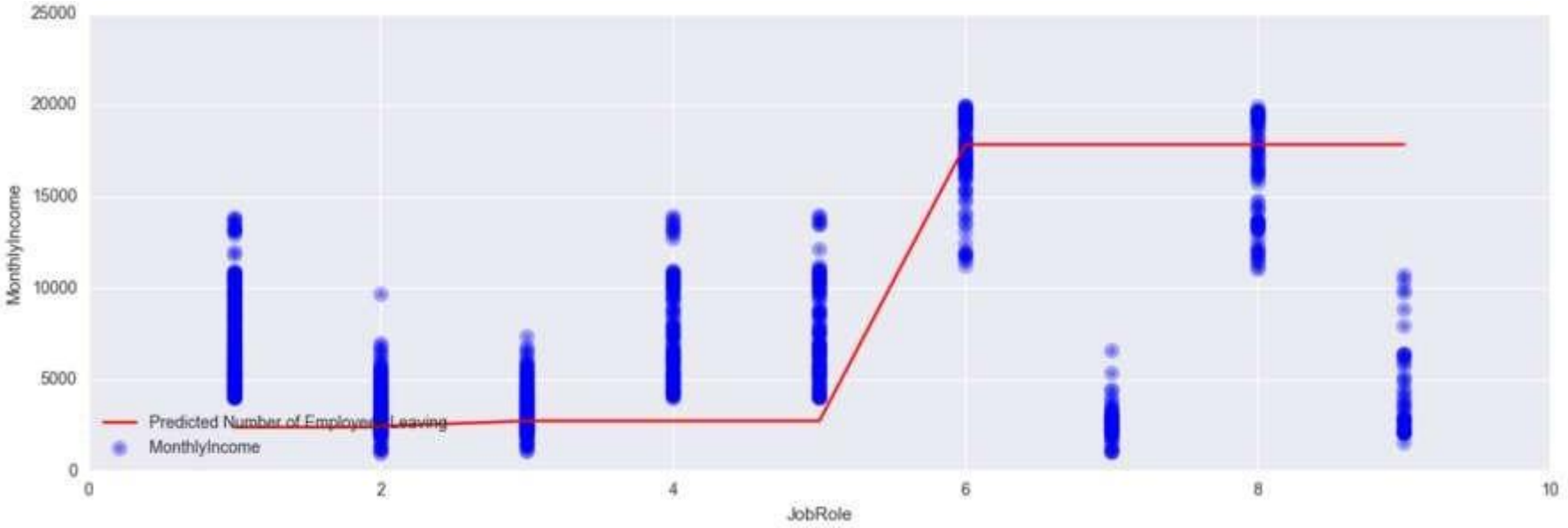


Research Scientist (25%) employees are ranked first. Laboratory Technician and Sales Executive are tied at 21% in second place.

Employees who were last promoted 8 years ago, are most likely to leave the company with an average monthly income of $10,000



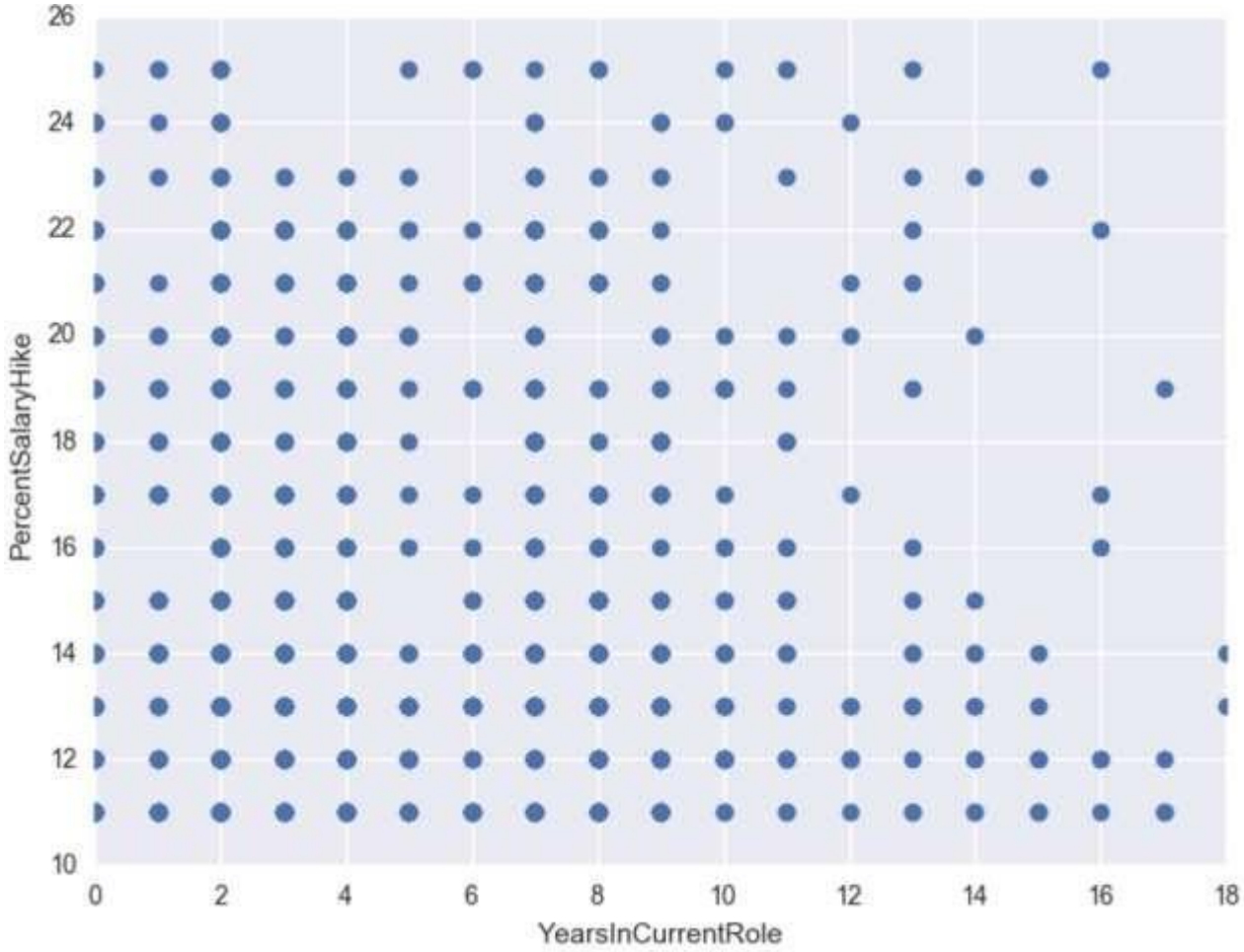
Manager and Research Director roles with an average monthly income of $18,000 are most likely to leave the company



Legend:

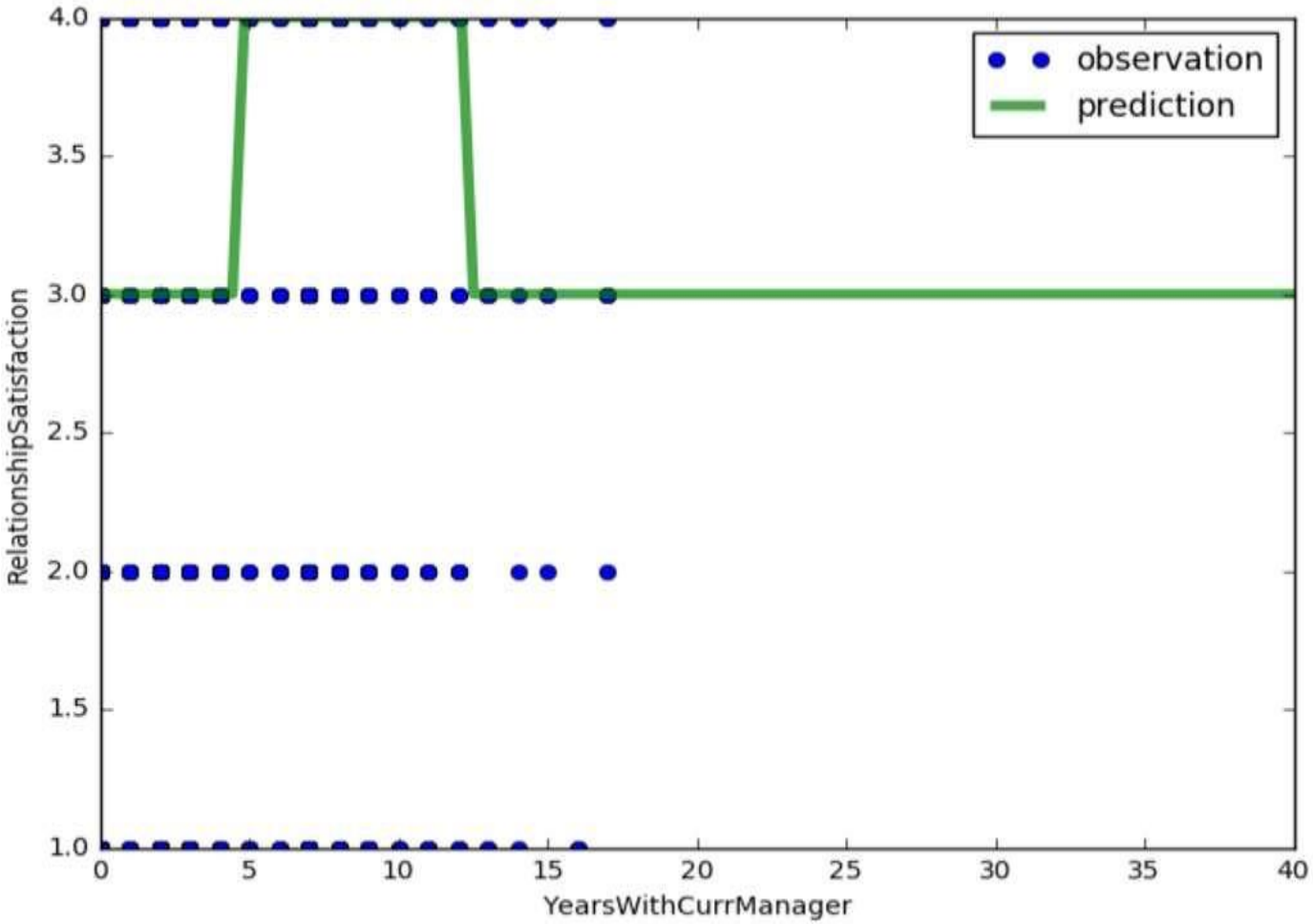
1– Sales Executive
2– Research Scientist
3 – Laboratory Technician
4– Manufacturing Director
5– Healthcare Representative
6 – Manager
7– Sales Representative
8– Research Director
9 – Human Resources

There is a moderate negative correlation between employee's years in their current role and salary percent hike.



The longer the employee stays in their current role, their salary percent hike decreases over time

The higher the relationship status with their current manager, the most likely that employees will leave their current job



Legend:

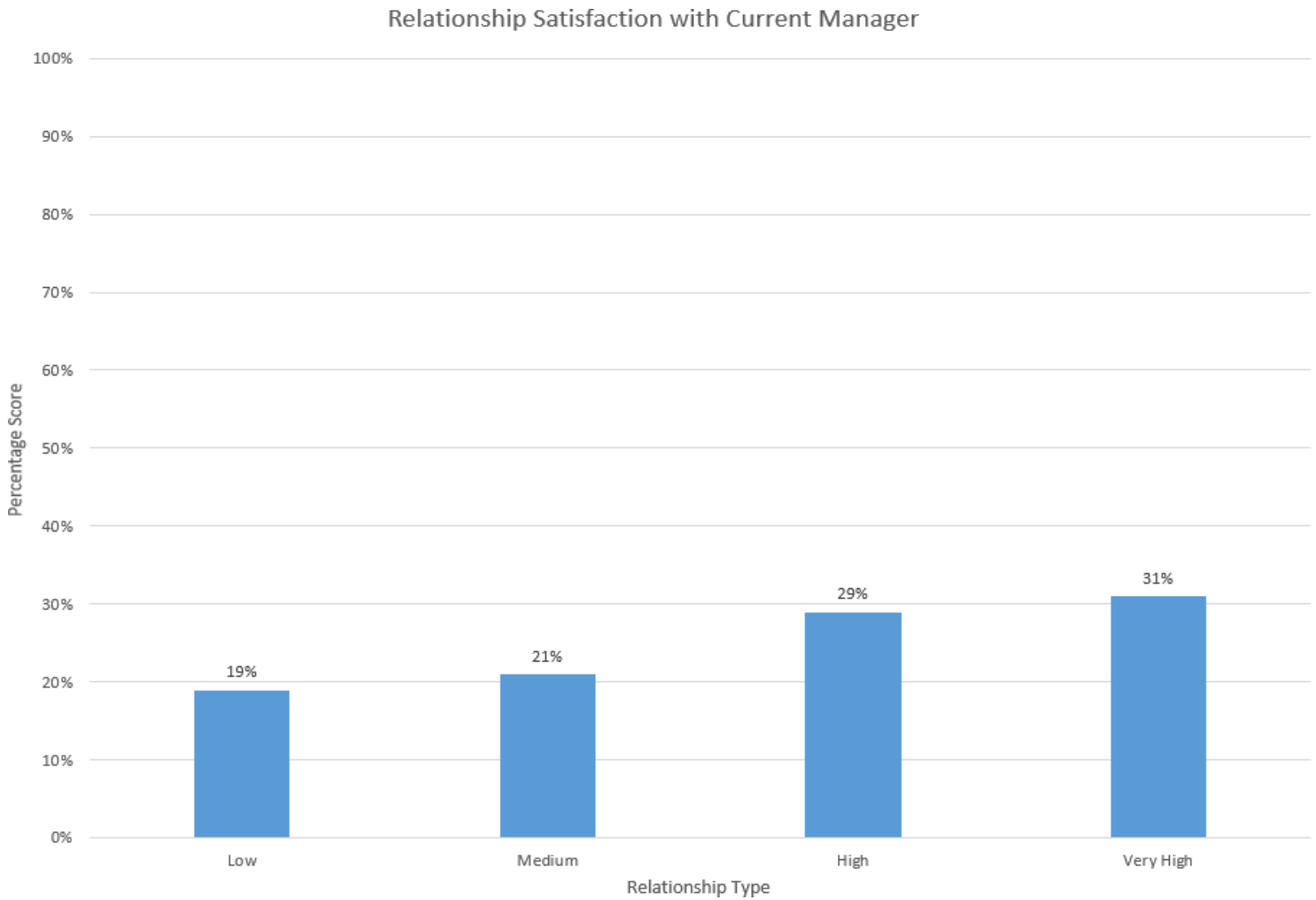
4.0 – Very High

3.0 – High

2.0 – Medium

1.0 – Low

Employees who have a ‘High Relationship Satisfaction score’ with their current managers are most likely to leave by 31%, followed by those with Very High Satisfaction score with 29%



Other things to consider :

30% of employees are not satisfied at their working environment

60% have scored ‘Better’ for work life balance, compared to 17% for ‘Best’

Summary

Insights

- Three factors that impacted the attrition as predicted by Random Forest:
  - Job Role
  - Total Number of Working Years
  - Overtime
- Employees who have 20 total working years are most likely to leave within 6 to 9 years of their current company
- 28% of employees work overtime. Majority of the employees are Research Scientists.
- Employees are most likely to leave the company in 7 years in their current role by 15%.
- The longer the employee stays in their current role, their salary percent hike decreases over time
- Employees may leave the company despite their positive relationship with their current managers.

Recommendations:

- Investigate the possible root causes of over time
- Promote work life balance within the company and the affected departments
- Check the market rate of rival companies

Improvements:

- More time in studying and testing the data
- Explore other programming languages and software like R

Q & A