Human Recourse Department

❖ Project Overview:

This project focuses on developing a comprehensive solution to streamline the hiring and retention processes within the Human Resources (HR) department. By leveraging data-driven insights, automation, and strategic tools, the goal is to reduce the time, capital, and skills required for effective talent acquisition and employee retention. The project aims to improve efficiency, lower recruitment costs, and support decision-making, ultimately enabling companies to focus more on core business activities rather than administrative tasks associated with HR functions.

❖ Problem Statement:

Hiring and retaining employees are critical yet highly complex tasks for Human Resource (HR) departments, demanding significant investment in terms of capital, time, and expertise. The challenge is particularly acute for small businesses, where owners often spend approximately 40% of their working hours on non-revenue-generating activities such as recruiting. Additionally, the cost of recruiting new employees is considerable, with companies typically spending 15%-20% of an employee's salary to onboard new talent. This project aims to address these challenges by streamlining and optimizing HR processes to reduce costs and save time, ultimately helping businesses focus more on growth and productivity.

❖ Project Objective:

The primary objective of this project is to create a predictive model that identifies employees who are likely to quit, along with understanding the underlying factors influencing their decisions. We will explore the reasons employees may want to leave, the motivations that encourage them to stay, and effective strategies to enhance their engagement and retention.

In addition, the project will analyze the costs associated with employee turnover, including direct expenses related to recruitment, onboarding, and training of new hires, as well as indirect costs such as lost productivity and decreased morale among remaining staff. By addressing these key areas, the project aims to provide actionable insights for HR departments to foster a more stable and motivated workforce, thereby reducing turnover rates and associated costs.

Human Recourse Department

❖ Techniques Used:

In this project, we will utilize Logistic Regression and Random Forest Classifier as the primary techniques for building our predictive model.

➢ Model Implementation

1. Logistic Regression Classifier
    - Use logistic regression for binary classification to predict employee attrition.
    - Evaluate performance metrics such as accuracy, precision, recall, and F1 score.
2. Random Forest Classifier
    - Implement a random forest classifier to leverage ensemble learning for improved prediction accuracy.
    - Evaluate feature importance to identify significant predictors of attrition.
3. Artificial Neural Networks (ANN)
    - Design and implement a neural network architecture suitable for the dataset.
    - Utilize activation functions, loss functions, and optimization techniques to enhance learning.

==========

- Understand how to leverage the power of data science to reduce employee's turnover and transform human resources department.
- Understand the theory behind logistic regression and random forest classifiers
- Train a logistic regression classifier and random forest classifier model using scikit learn
- Apply 'sigmoid' functions to obtain probability
- Load and manipulate dataset using pandas data frame
- Develop a function in python and apply it to pandas data frame Perform exploratory data analysis using matplotlib and Seaborn libraries
- Plot Kernel density estimate (KDE) plots, box plots, and count plots
- Convert categorical variables into dummy variables
- Divide dataset into training and testing using scikit learn
- Understand the theory and intuition behind artificial neural networks and how to apply them to perform classification tasks
- Evaluate classification models and present results using confusion matrix and classification reports
- Understand the difference between precision, recall, and F1-score.

==========

Human Recourse Department

Task 1: Understand the Problem Statement and Business Case

In today's competitive business landscape, managers face significant challenges in hiring and retaining top talent within their organizations. The complexities involved in these tasks require substantial capital, time, and specialized skills. Retaining high-performing employees is crucial, as they generate significant value for the company.

To effectively manage this challenge, organizations need to implement strategies that ensure top talent feels valued and motivated to stay. This includes offering competitive compensation packages, providing ongoing training and development opportunities, and implementing performance-based bonuses. By focusing on these areas, businesses can foster a supportive environment that encourages employee retention and maximizes the contributions of their most valuable assets—their employees.

Human Recourse Department

Data Set

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7
...	...	...	...	...	...	...	...	...	...	...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061

EnvironmentSatisfaction	Gender	HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked
2	Female	94	3	2	Sales Executive	4	Single	5993	19479	8
3	Male	61	2	2	Research Scientist	2	Married	5130	24907	1
4	Male	92	2	1	Laboratory Technician	3	Single	2090	2396	6
4	Female	56	3	1	Research Scientist	3	Married	2909	23159	1
1	Male	40	3	1	Laboratory Technician	2	Married	3468	16632	9
...	...	...	...	...	...	...	...	...	...	...
3	Male	41	4	2	Laboratory Technician	4	Married	2571	12290	4

Over18	OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear
Y	Yes	11	3	1	80	0	8	0
Y	No	23	4	4	80	1	10	3
Y	Yes	15	3	2	80	0	7	3
Y	Yes	11	3	3	80	0	8	3
Y	No	12	3	4	80	1	6	3
...	...	...	...	...	...	...	...	...
Y	No	17	3	3	80	1	17	3

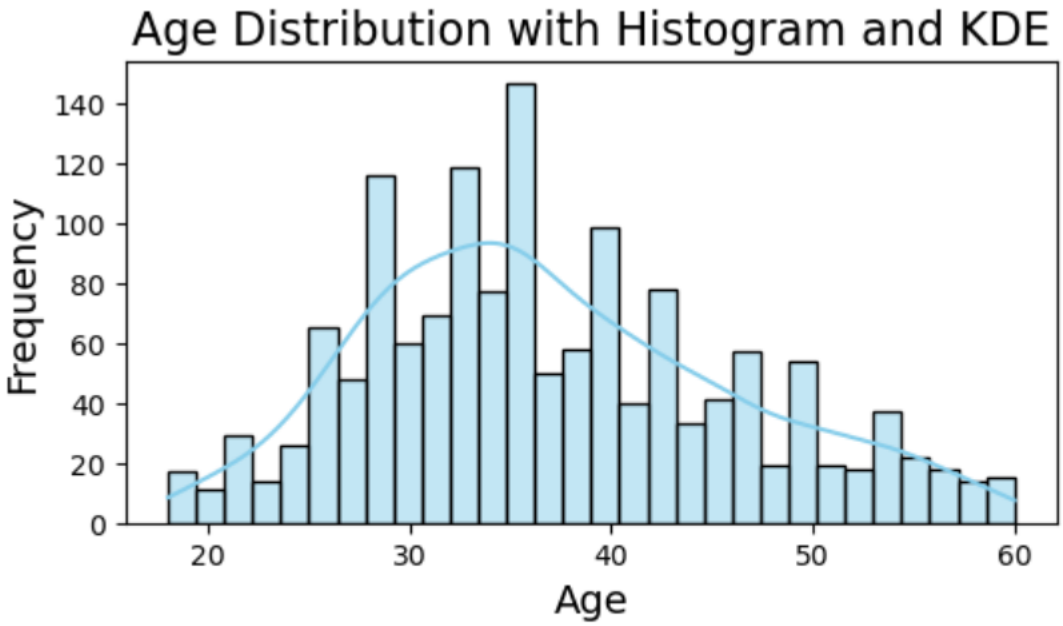
TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
0	1	6	4	0	5
3	3	10	7	1	7
3	3	0	0	0	0
3	3	8	7	3	0
3	3	2	2	2	2
...	...	...	...	...	...
3	3	5	2	0	3

Human Recourse Department

❖ 'Age' Column

The 'Age' column in the dataset provides insights into the age distribution of employees:

1. Mean Age: Approximately 36.9 years, indicating that the average employee is in their mid-30s.
2. Minimum Age: 18 years, representing the youngest employee in the dataset.
3. Maximum Age: 60 years, indicating the oldest employee.
4. Age Distribution: The majority of employees fall within the age range of 25 to 38 years, suggesting a workforce primarily composed of younger to mid-career professionals. This demographic may influence retention strategies, as this age group often seeks career development opportunities and work-life balance.



❖ Employee Attrition

The analysis of employee attrition provides key insights into workforce stability:

- Total Employees: 1,470
- Employees Who Left the Company: 237
  - Percentage of Employees Who Left: Approximately 16.12%
- Employees Who Stayed: 1,233
  - Percentage of Employees Who Stayed: Approximately 83.88%

Human Recourse Department

❖ Comparison of Key Metrics Between Employees Who Stayed and Left

This analysis compares the mean and standard deviation of various attributes between employees who stayed with the company and those who left. The findings highlight significant differences in several key areas:

1. Age:
    - Mean Age of Employees Who Stayed: Higher than those who left, suggesting that older employees may feel more stable or engaged in their roles.
2. Daily Rate:
    - Mean Daily Rate of Employees Who Stayed: Higher compared to those who left, indicating that better compensation may contribute to retention.
3. Distance from Home:
    - Distance of Employees Who Stayed: Generally closer to home, which could positively influence their job satisfaction and likelihood to stay.
4. Environment Satisfaction:
    - Satisfaction Levels of Employees Who Stayed: Higher overall, indicating a more positive work environment which is likely to encourage employee retention.
5. Job Satisfaction:
    - Job Satisfaction of Employees Who Stayed: Generally higher than that of those who left, suggesting that job roles and responsibilities contribute significantly to their decision to remain with the company.
6. Stock Option Level:
    - Stock Option Level of Employees Who Stayed: Tends to be higher, which may provide additional motivation to stay, as it aligns their interests with the company's long-term performance.

Conclusion

The comparison reveals that employees who stayed generally have more favorable conditions regarding age, compensation, proximity to work, and job satisfaction. These factors can inform HR strategies aimed at improving retention by enhancing employee satisfaction and addressing potential concerns among those who may be considering leaving the company.

Human Recourse Department

❖ Employee Retention Analysis: Factors Influencing Longevity in the Company

• Strong Positive Correlations

1. Total Working Years & Age: Correlation of 0.680 suggests that older employees tend to have more years of total working experience.
2. Job Level & Total Working Years: Correlation of 0.782 indicates that higher job levels are associated with longer total working years.
3. Monthly Income & Job Level: Strong correlation of 0.950 indicates that as job levels increase, monthly income also increases significantly.
4. Monthly Income & Total Working Years: Correlation of 0.773 shows a strong link between longer total working years and higher monthly income.

Strong Negative Correlations

1. Attrition & Total Working Years: Correlation of -0.171 indicates that employees with fewer total working years are more likely to leave the company.
2. Attrition & Job Level: Correlation of -0.169 suggests that lower job levels are associated with higher attrition rates.
3. Percent Salary Hike & Job Level: Correlation of -0.034 indicates that a lower salary hike may be linked to lower job levels, though the correlation is weak.

Moderate Correlations

1. Age & Education: Correlation of 0.208 suggests a slight positive relationship, indicating that older employees may have higher education levels.
2. Job Satisfaction & Years At Company: Correlation of 0.311 indicates that longer tenure at the company is associated with higher job satisfaction.
3. Environment Satisfaction & Job Satisfaction: Correlation of 0.070 shows a weak positive relationship, suggesting that employees satisfied with their environment may also be satisfied with their job.

Key Insights

- Age and Experience: Older employees tend to have more experience, which correlates with higher job levels and income.
- Attrition Factors: Attrition is negatively correlated with both job level and total working years, suggesting that increasing job levels and tenure could potentially reduce employee turnover.
- Job Satisfaction: Job satisfaction appears to improve with years at the company, indicating that retaining employees may enhance their satisfaction levels.

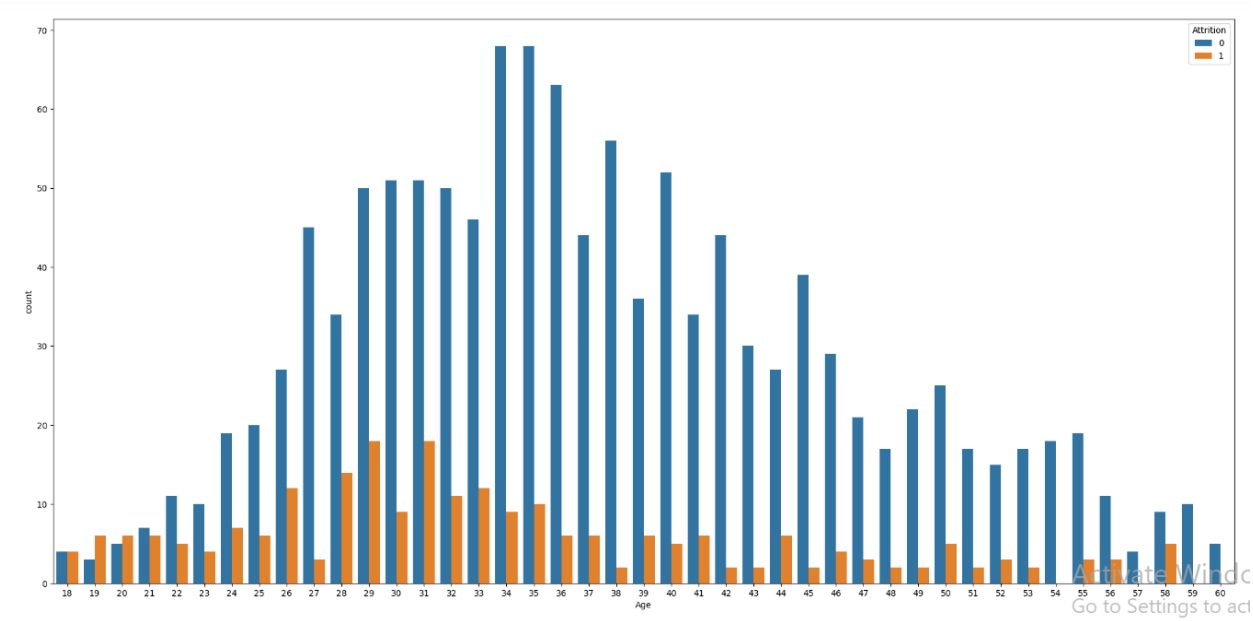
Human Recourse Department

Recommendations

1. Retention Strategies: Focus on employee engagement programs that enhance job satisfaction for long-tenured employees.
2. Career Development: Provide clear pathways for career advancement to reduce attrition, particularly for employees at lower job levels.
3. Salary Review: Regularly assess salary hikes in relation to job levels to ensure competitive compensation, which may reduce turnover.

These insights could guide HR strategies to improve employee retention and satisfaction within the organization.

❖ Key Points on Age and Attrition Data:



1. Age Distribution:
   - The dataset includes employees aged between 18 and 60.
   - The highest employee counts are observed in the age group of 29 (50 employees) and 30 (51 employees), indicating a concentration of staff in their late twenties to early thirties.
2. Attrition Rates:
   - Overall Attrition:
     - Employees aged 18 to 60 exhibit varied attrition rates, with some age groups having significantly higher counts of employees who left (attrition = 1).

- Peak Attrition Age:
  - The 19-year-olds have a notable attrition rate of 6, compared to 3 who stayed, highlighting a 66.7% turnover.
  - The age group of 27 has the highest count of employees (45) who stayed with the company, but only 3 left, resulting in a low attrition rate of 6.3%.

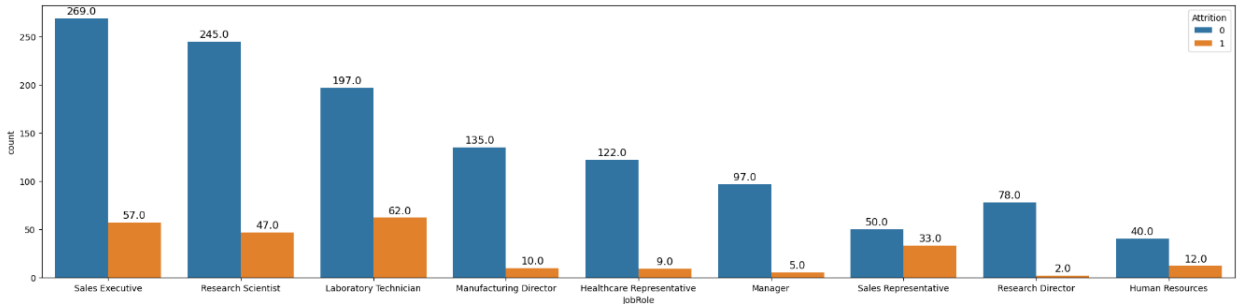
3. Age Group Insights:
   - Young Employees (18-25):
     - The attrition counts for younger employees show a mixed trend, with high attrition at ages 18 and 19 but lower counts at ages 20 to 25.
   - Mid-Aged Employees (26-35):
     - This group has a mix of attrition, with counts of employees staying generally higher than those who left, especially for ages 26, 29, and 30.
   - Older Employees (36-60):
     - Higher stability is noted in older employees, particularly between ages 34 and 36, where more than 60 employees remained in the company, despite lower overall numbers.
4. Specific Age Insights:
   - Age 34 and 35 have the highest counts of employees who stayed (68 each) with relatively low attrition (9 and 10 respectively).
   - Employees aged 40 and above show a trend of higher stability, with attrition counts remaining low compared to younger age groups.
5. General Trends:
   - There is a general trend of declining attrition rates as employees age, suggesting that older employees tend to stay longer in the company.
   - Ages 36 to 40 show particularly low attrition rates, indicating potential factors of job satisfaction or loyalty.
6. Potential Areas for Improvement:
   - The company may need to investigate why younger employees (especially those around 18-24) exhibit higher attrition rates and implement strategies to enhance retention among this demographic.
7. Recommendations for Retention Strategies:
   - Implement mentorship programs that pair younger employees with more experienced staff.
   - Conduct exit interviews to gather insights into why younger employees are leaving.
   - Consider tailored engagement strategies for younger age groups to increase job satisfaction and reduce turnover.

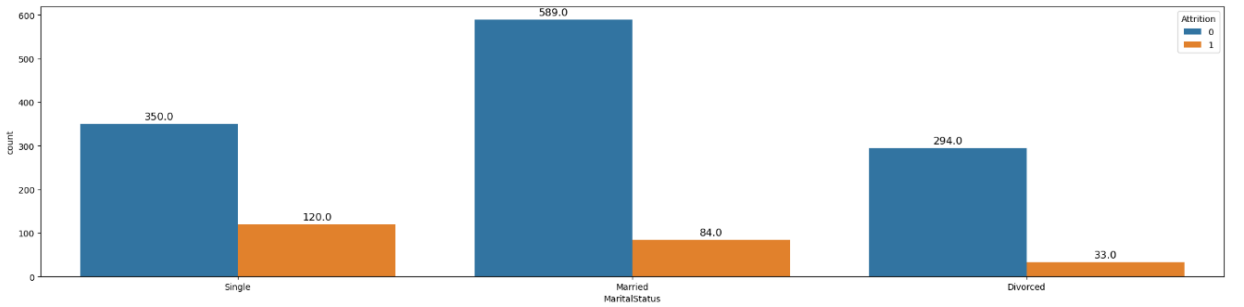
Human Recourse Department

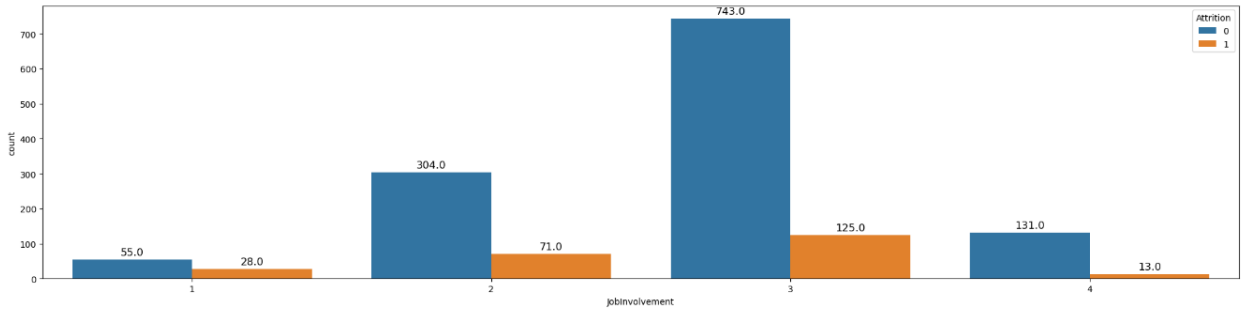
Conclusion:

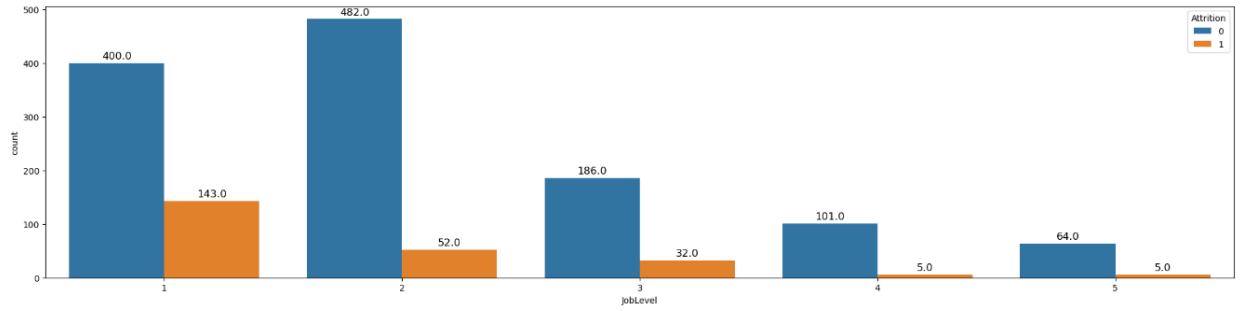
This data highlights significant trends in employee attrition based on age, providing insights that can inform retention strategies and help improve employee satisfaction across various age groups.

❖ Comparison between different features where employees stay and leave the job.



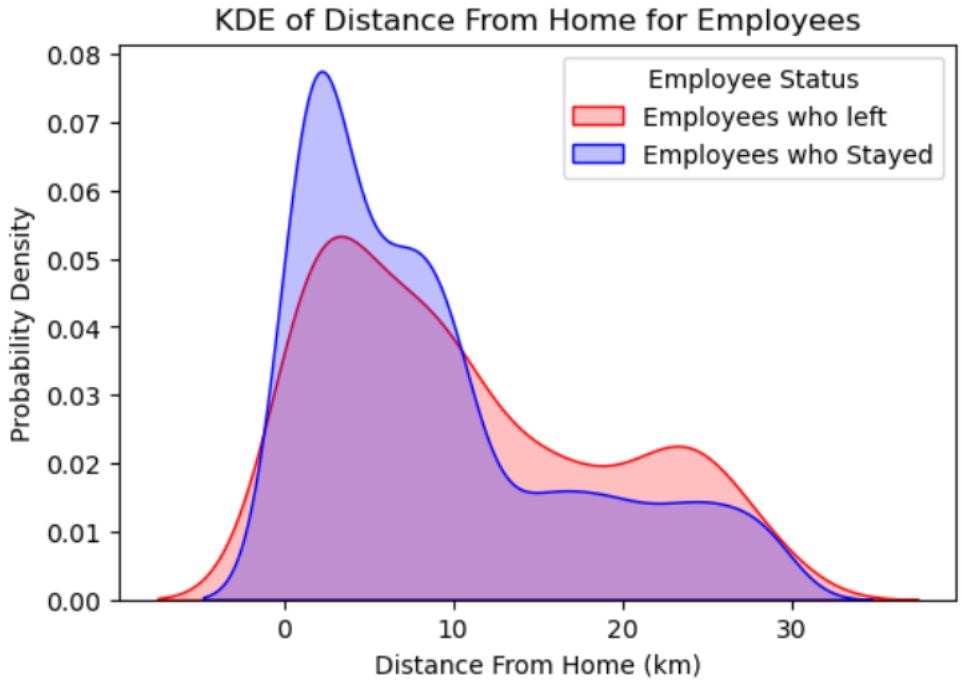






Human Recourse Department

❖ Distance Range and Attrition and stay Data for Employees:



➢ Summary of Distance Range and Attrition Data:

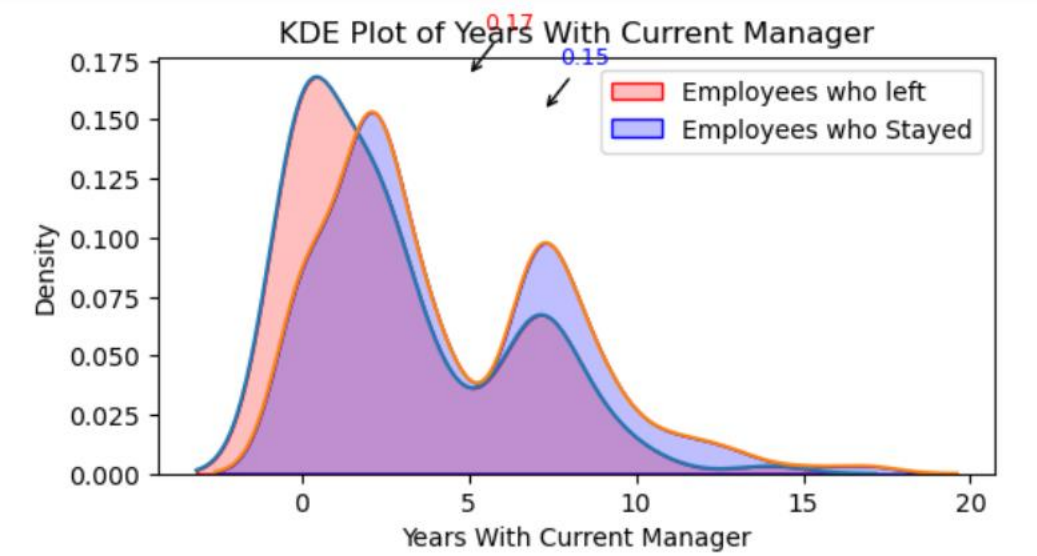
1. Distance Overview:
   - Employees who left the company are categorized by commuting distances: 1-5 km (26), 5-10 km (28), 10-20 km (14), 20-30 km (9), and 30 km+ (10).
2. High Attrition in Closer Ranges:
   - The 1-10 km distance group has the highest attrition (54 employees), while those in the 10-30 km ranges show lower counts.
3. Trend:
   - Attrition decreases with increased distance from the company, indicating employees living farther away may have higher job satisfaction or stability.
4. Recommendations:
   - Consider flexible work options, commuting support, and engagement initiatives to retain employees in closer distance ranges.
5. Conclusion:
   - Addressing the needs of employees living within 1-10 km could improve retention rates.

Human Recourse Department

➢ Summary of Distance Range and Retention Data:

1. **Distance Overview:**
   - Employees who stayed with the company are categorized by commuting distances: 1-5 km (182), 5-10 km (183), 10-20 km (70), 20-30 km (55), and 30 km+ (55).
2. **High Retention in Closer Ranges:**
   - The 1-10 km distance group has the highest retention (365 employees), indicating a strong correlation between proximity and employee retention.
3. **Decreased Retention with Distance:**
   - Retention significantly drops for employees living 10 km or more away, with counts decreasing as the distance increases.
4. **Implications:**
   - Employees living within 1-10 km are more likely to remain with the company, suggesting that location may play a crucial role in job satisfaction and loyalty.
5. **Conclusion:**
   - To further enhance retention, the company may consider incentives or support for employees commuting from longer distances.

❖ Employee Retention Analysis: Insights by Years of Service



❖ Summary of Tenure and Employee Turnover:

1. Tenure Overview:
   - Employees who left the company are categorized by their tenure:
     - 0-1 years: 85
     - 1-2 years: 11
     - 2-3 years: 50

- 3-4 years: 19
- 4-5 years: 11
- 5-6 years: 4
- 6-7 years: 4
- 7-8 years: 31
- 8-9 years: 10
- 9-10 years: 6

2. High Turnover in Early Years:
   - The 0-1 years tenure group shows the highest turnover (85 employees), indicating significant attrition among newer employees.
3. Mid-Tenure Drop-Off:
   - The 2-3 years range also sees notable attrition (50 employees), suggesting potential dissatisfaction or career shifts around this tenure mark.
4. Low Turnover in Long Tenure:
   - Employees with more than 5 years of tenure exhibit very low turnover, with counts declining significantly beyond the 3-4 years mark.
5. Implications:
   - Focus on improving employee engagement and support during the 0-3 years period may help reduce turnover rates.
   - Understanding the reasons behind departures in early and mid-tenure could lead to effective retention strategies.
6. Conclusion:
   - To enhance retention, targeted initiatives should be implemented for employees in their initial years, especially focusing on those who have been with the company for less than three years.

❖ Summary of Employee Retention by Tenure who stayed:

1. Tenure Overview:
   - Employees who stayed with the company are categorized by their tenure:
     - 0-1 years: 178
     - 1-2 years: 65
     - 2-3 years: 294
     - 3-4 years: 123
     - 4-5 years: 87
     - 5-6 years: 27
     - 6-7 years: 25
     - 7-8 years: 185
     - 8-9 years: 97
     - 9-10 years: 58

2. Strong Retention in Mid-Tenure:
    - The 2-3 years tenure group has the highest retention (294 employees), indicating a significant number of employees choose to remain after their initial period.
3. Declining Retention with Increasing Tenure:
    - Retention rates generally decline as tenure increases after the 2-3 years period, particularly in the 5-6 years and 6-7 years ranges.
4. Consistent Retention Among Long-Term Employees:
    - Employees in the 7-8 years range also exhibit strong retention (185 employees), suggesting that those who stay beyond the initial years often remain committed to the company.
5. Implications:
    - The data suggests that fostering a supportive work environment during the 0-3 years period can contribute to higher retention rates.
    - Consideration should be given to strategies that maintain engagement among longer-tenured employees, particularly those in the 5-7 years range.
6. Conclusion:
    - To improve overall retention, the focus should be on enhancing employee satisfaction and support during their early years while also addressing engagement for employees with longer tenures.

❖ Creating Testing and Training Dataset

- Split the dataset into training and testing sets to evaluate model performance.
- Use stratified sampling to maintain the distribution of the target variable (Attrition: 0 and 1).

➢ Data Cleaning

- Handle missing values by either imputing or removing them based on their impact on the dataset.
- Convert categorical variables into numerical format using OneHotEncoder for effective model training.
- Scale numerical features using MinMaxScaler to normalize data within a specified range (0 to 1).

➢ Target Variable

- Define Attrition as the target variable (0 for stayed, 1 for left).

➢ Model Implementation

1. Logistic Regression Classifier
    - Use logistic regression for binary classification to predict employee attrition.
    - Evaluate performance metrics such as accuracy, precision, recall, and F1 score.

Human Recourse Department

2. Artificial Neural Networks (ANN)
    - Design and implement a neural network architecture suitable for the dataset.
    - Utilize activation functions, loss functions, and optimization techniques to enhance learning.
3. Random Forest Classifier
    - Implement a random forest classifier to leverage ensemble learning for improved prediction accuracy.
    - Evaluate feature importance to identify significant predictors of attrition.

➢ Conclusion

- Compare and contrast the performance of the different classifiers to determine the most effective model for predicting employee attrition.
- Document findings and model performance metrics for reporting.

❖ Classification Performance Metrics | Logistic Regression Prediction

- Precision
    - Class 0 (Stayed): 0.89 (high precision indicates few false positives)
    - Class 1 (Left): 0.76 (moderate precision; some false positives present)
- Recall
    - Class 0 (Stayed): 0.97 (high recall indicates few false negatives)
    - Class 1 (Left): 0.41 (low recall indicates many false negatives)
- F1-Score
    - Class 0 (Stayed): 0.93 (excellent balance between precision and recall)
    - Class 1 (Left): 0.54 (lower score reflects challenges in correctly identifying class 1)
- Support
    - Class 0 (Stayed): 305 (number of actual instances)
    - Class 1 (Left): 63 (number of actual instances)
- Overall Accuracy:
    - Accuracy: 0.88 (88% of all predictions are correct)
- Macro Average:
    - Precision: 0.83
    - Recall: 0.69
    - F1-Score: 0.73 (average performance across classes without taking class imbalance into account)
- Weighted Average:
    - Precision: 0.87 (takes class imbalance into account)
    - Recall: 0.88
    - F1-Score: 0.86 (indicates the model's performance is better when considering class sizes)

Human Recourse Department

The Logistic Regression model for predicting employee attrition yielded an accuracy of 88%.

- Precision: 0.89 for employees who stayed (0) and 0.76 for those who left (1).
- Recall: 0.97 for class 0, indicating strong identification of staying employees, but only 0.41 for class 1, suggesting difficulty in recognizing those who left.
- F1-Score: 0.93 for class 0 and 0.54 for class 1, showing a better balance for staying employees.

Overall, while the model is effective in predicting those who stay, it struggles with identifying leavers, indicating a need for further improvement in this area.

❖ Random Forest Classifier Prediction

The Random Forest Classifier for predicting employee attrition achieved an overall accuracy of 85%.

- Precision: 0.85 for employees who stayed (0) and 0.73 for those who left (1).
- Recall: 0.99 for class 0, indicating excellent identification of staying employees, but only 0.17 for class 1, highlighting significant difficulty in recognizing those who left.
- F1-Score: 0.91 for class 0 and 0.28 for class 1, reflecting a strong performance for predicting stayers but poor performance for leavers.

In summary, while the model effectively identifies employees who stay, its ability to predict those who leave is limited, indicating a need for enhancement in capturing attrition cases.

❖ TensorFlow Model Prediction

The TensorFlow model for predicting employee attrition achieved an overall accuracy of 83%.

- Precision: 0.89 for employees who stayed (0) and 0.51 for those who left (1), indicating strong accuracy in predicting stayers but moderate accuracy for leavers.
- Recall: 0.91 for class 0, showing effective identification of employees who stayed, but only 0.44 for class 1, suggesting difficulty in capturing those who left.
- F1-Score: 0.90 for class 0 and 0.47 for class 1, reflecting solid performance in predicting stayers but weaker performance for leavers.

Overall, the model demonstrates reliable prediction capabilities for employees who stay but shows a need for improvement in predicting attrition.

Human Recourse Department