

Minor project on HR Analytics for a fast-growing organization called SimpleYard. SimpleYard is facing a significant challenge with a high employee attrition rate, which negatively impacts its growth and operational efficiency due to loss of skills, institutional knowledge, and increased hiring/training costs.

The case study seeks to answer several critical questions:

- Current workforce size and number of employees who have left
- Departments with the highest attrition rates
- Whether employees working on fewer than 3 projects are more likely to leave
- Correlation between the number of projects and time spent at the company for those who have left
- The influence of compensation levels on an employee's decision to leave

The goal of this HR Analytics case study is to perform exploratory data analysis (EDA) using employee records to:

- Analyze employee attrition patterns using statistical summaries and visualizations.
- Identify key factors associated with employees leaving, such as departmental distribution , number of projects , time spent at the company , salary levels , and promotion history.
- Quantify attrition rates and generate actionable insights.
- Support the HR team in developing data-driven employee retention strategies.

This analysis will involve data loading, preprocessing, visualization, and interpretation using tools like Pandas, NumPy, Matplotlib, and Seaborn.

The source is provided from the employee.csv file, which is attached below:



employee.csv

To analyze employee attrition patterns using statistical summaries and visualizations, we should start by calculating the overall attrition rate and then delve into how different features relate to employees leaving the company.

1. Calculate Overall Attrition Rate

2. Statistical Summary of Numerical Features by Attrition Status: We can use descriptive statistics ,mean, median, standard deviation, etc. to compare employees who left versus those who stayed across numerical features like

satisfactoryLevel, lastEvaluation, numberOfProjects, avgMonthlyHours, and timeSpent.company.

3. Statistical Summary of Categorical Features by Attrition Status: We will analyze the distribution of categorical features such as

dept, salary, workAccident, and promotionInLast5years for employees who left versus those who stayed to identify potential patterns.

4. Visualizations: We'll create various plots to visually represent these patterns, which could include:

- Bar plots for attrition rates by department and salary levels.
- Distribution plots for numerical features, distinguishing between employees who stayed and those who left

Source Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('employee.csv')
print("Initial Data Head")
print(df.head())
print("\nInitial Data Info")
print(df.info())

total_employees = df.shape[0]
employees_left = df[df['left'] == 1].shape[0]
print(f"\nWorkforce Size and Attrition Count")
print(f"Current workforce size: {total_employees}")
print(f"Number of employees who have left: {employees_left}")

attrition_rate = (employees_left / total_employees) * 100
print(f"\nOverall Attrition Rate")
print(f"Overall Attrition Rate: {attrition_rate:.2f}%")
```

```

print("\n Statistical Summary for Employees Who Stayed (left = 0) ")
print(df[df['left'] == 0].describe())

print("\n Statistical Summary for Employees Who Left (left = 1) ")
print(df[df['left'] == 1].describe())

print("\n Attrition by Department ")
dept_attrition = df.groupby('dept')['left'].value_counts(normalize=True).unstack().fillna(0)
dept_attrition['attrition_rate'] = dept_attrition[1] * 100
print(dept_attrition)

print("\n Attrition by Salary ")
salary_attrition =
df.groupby('salary')['left'].value_counts(normalize=True).unstack().fillna(0)
salary_attrition['attrition_rate'] = salary_attrition[1] * 100
print(salary_attrition)

print("\n Attrition by Work Accident ")
accident_attrition =
df.groupby('workAccident')['left'].value_counts(normalize=True).unstack().fillna(0)
accident_attrition['attrition_rate'] = accident_attrition[1] * 100
print(accident_attrition)

print("\n Attrition by Promotion in Last 5 Years ")
promotion_attrition =
df.groupby('promotionInLast5years')['left'].value_counts(normalize=True).unstack().fillna(0)
promotion_attrition['attrition_rate'] = promotion_attrition[1] * 100
print(promotion_attrition)

sns.set_style("whitegrid")

plt.figure(figsize=(6, 4))
sns.countplot(x='left', data=df, palette='viridis')
plt.title('Overall Employee Attrition (0: Stayed, 1: Left)')
plt.xlabel('Attrition Status')
plt.ylabel('Number of Employees')
plt.xticks([0, 1], ['Stayed', 'Left'])
plt.show()

plt.figure(figsize=(12, 6))
sns.barplot(x=dept_attrition.index, y=dept_attrition['attrition_rate'], palette='coolwarm')

```

```

plt.title('Attrition Rate by Department')
plt.xlabel('Department')
plt.ylabel('Attrition Rate (%)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()

plt.figure(figsize=(8, 5))
salary_order = ['low', 'medium', 'high']
sns.barplot(x=salary_attrition.index, y=salary_attrition['attrition_rate'],
order=salary_order, palette='plasma')
plt.title('Attrition Rate by Salary Level')
plt.xlabel('Salary Level')
plt.ylabel('Attrition Rate (%)')
plt.tight_layout()
plt.show()

plt.figure(figsize=(10, 6))
sns.scatterplot(x='numberOfProjects', y='timeSpent.company', data=leavers_df)
plt.title('Number of Projects vs. Time Spent at Company for Employees Who Left')
plt.xlabel('Number of Projects')
plt.ylabel('Time Spent at Company (Years)')
plt.grid(True)
plt.show()

```

Result: (Running Python file in Visual Studio Code)

```

[Running] python -u "c:\Deep_learning projects\HRAnalytics\miniproject_1.py"
Initial Data Head
|   satisfactoryLevel  lastEvaluation  ...   dept  salary
0           0.38           0.53  ...  sales    low
1           0.80           0.86  ...  sales  medium
2           0.11           0.88  ...  sales  medium
3           0.37           0.52  ...  sales    low
4           0.41           0.50  ...  sales    low

[5 rows x 10 columns]

```

```
Initial Data Info
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14999 entries, 0 to 14998
Data columns (total 10 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   satisfactoryLevel                     14999 non-null float64
1   lastEvaluation                       14999 non-null float64
2   numberOfProjects                     14999 non-null int64
3   avgMonthlyHours                     14999 non-null int64
4   timeSpent.company                   14999 non-null int64
5   workAccident                        14999 non-null int64
6   left                                14999 non-null int64
7   promotionInLast5years               14999 non-null int64
8   dept                                14999 non-null object
9   salary                              14999 non-null object
dtypes: float64(2), int64(6), object(2)
memory usage: 1.1+ MB
```

Workforce Size and Attrition Count
Current workforce size: 14999
Number of employees who have left: 3571

Overall Attrition Rate
Overall Attrition Rate: 23.81%

```
Statistical Summary for Employees Who Stayed (left = 0)

```

	satisfactoryLevel	lastEvaluation	...	left	promotionInLast5years
count	11428.000000	11428.000000	...	11428.0	11428.000000
mean	0.666810	0.715473	...	0.0	0.026251
std	0.217104	0.162005	...	0.0	0.159889
min	0.120000	0.360000	...	0.0	0.000000
25%	0.540000	0.580000	...	0.0	0.000000
50%	0.690000	0.710000	...	0.0	0.000000
75%	0.840000	0.850000	...	0.0	0.000000
max	1.000000	1.000000	...	0.0	1.000000

[8 rows x 8 columns]

Statistical Summary for Employees Who Left (left = 1)

	satisfactoryLevel	lastEvaluation	...	left	promotionInLast5years
count	3571.000000	3571.000000	...	3571.0	3571.000000
mean	0.440098	0.718113	...	1.0	0.005321
std	0.263933	0.197673	...	0.0	0.072759
min	0.090000	0.450000	...	1.0	0.000000
25%	0.130000	0.520000	...	1.0	0.000000
50%	0.410000	0.790000	...	1.0	0.000000
75%	0.730000	0.900000	...	1.0	0.000000
max	0.920000	1.000000	...	1.0	1.000000

[8 rows x 8 columns]

Attrition by Department

left	0	1	attrition_rate
dept			
IT	0.777506	0.222494	22.249389
RandD	0.846252	0.153748	15.374841
accounting	0.734029	0.265971	26.597132
hr	0.709066	0.290934	29.093369
management	0.855556	0.144444	14.444444
marketing	0.763403	0.236597	23.659674
product_mng	0.780488	0.219512	21.951220
sales	0.755072	0.244928	24.492754
support	0.751009	0.248991	24.899058
technical	0.743750	0.256250	25.625000

Attrition by Salary

left	0	1	attrition_rate
salary			
high	0.933711	0.066289	6.628941
low	0.703116	0.296884	29.688354
medium	0.795687	0.204313	20.431275

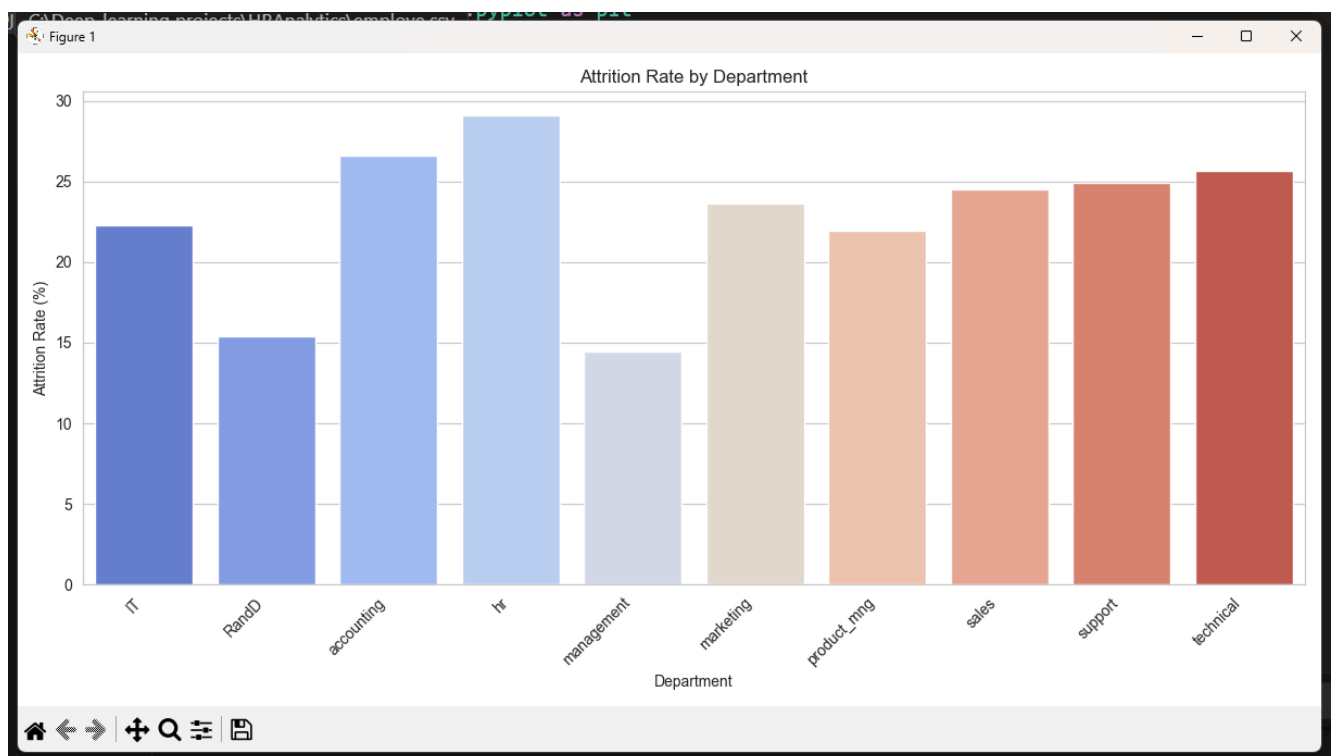
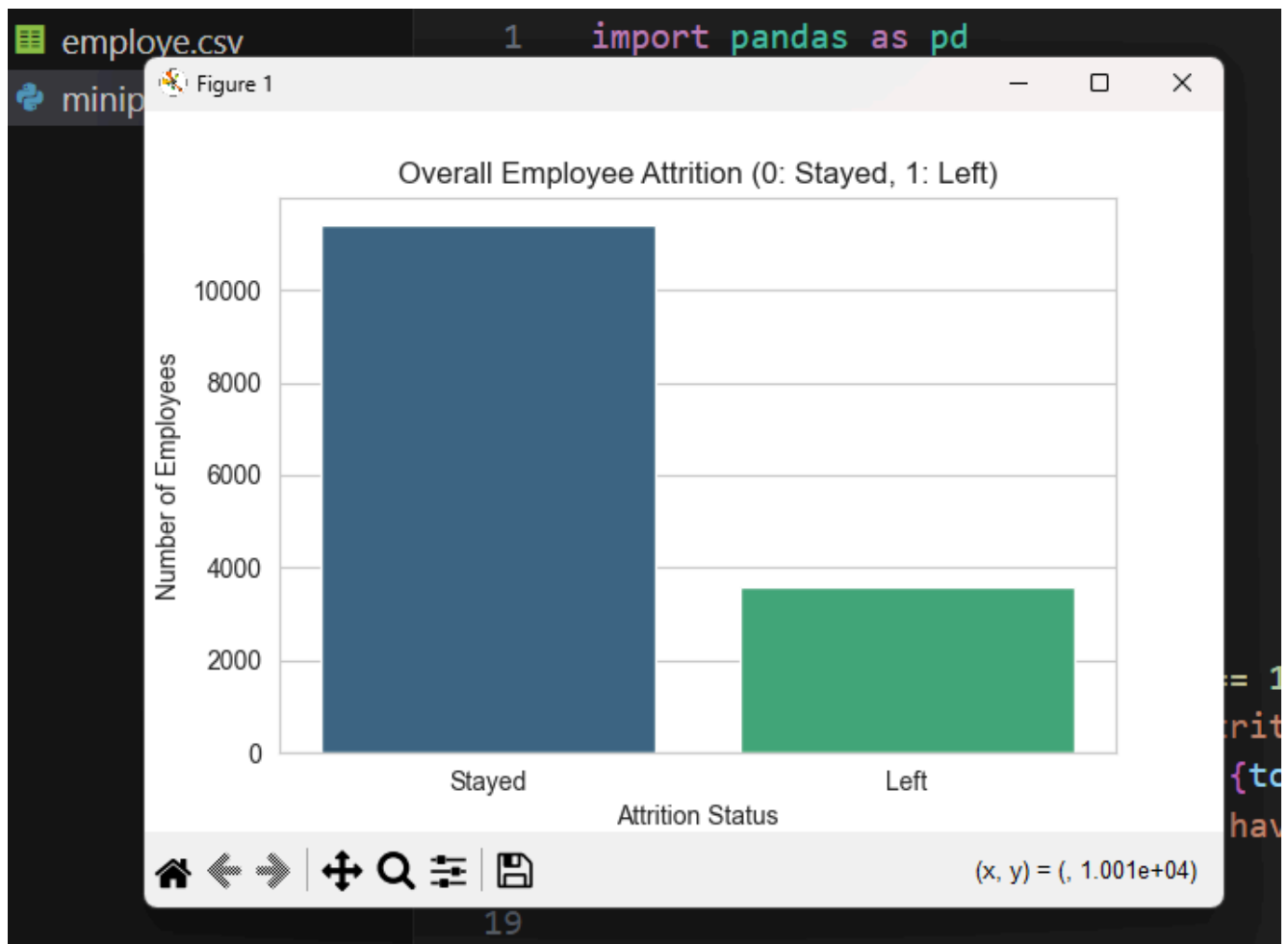
Attrition by Work Accident

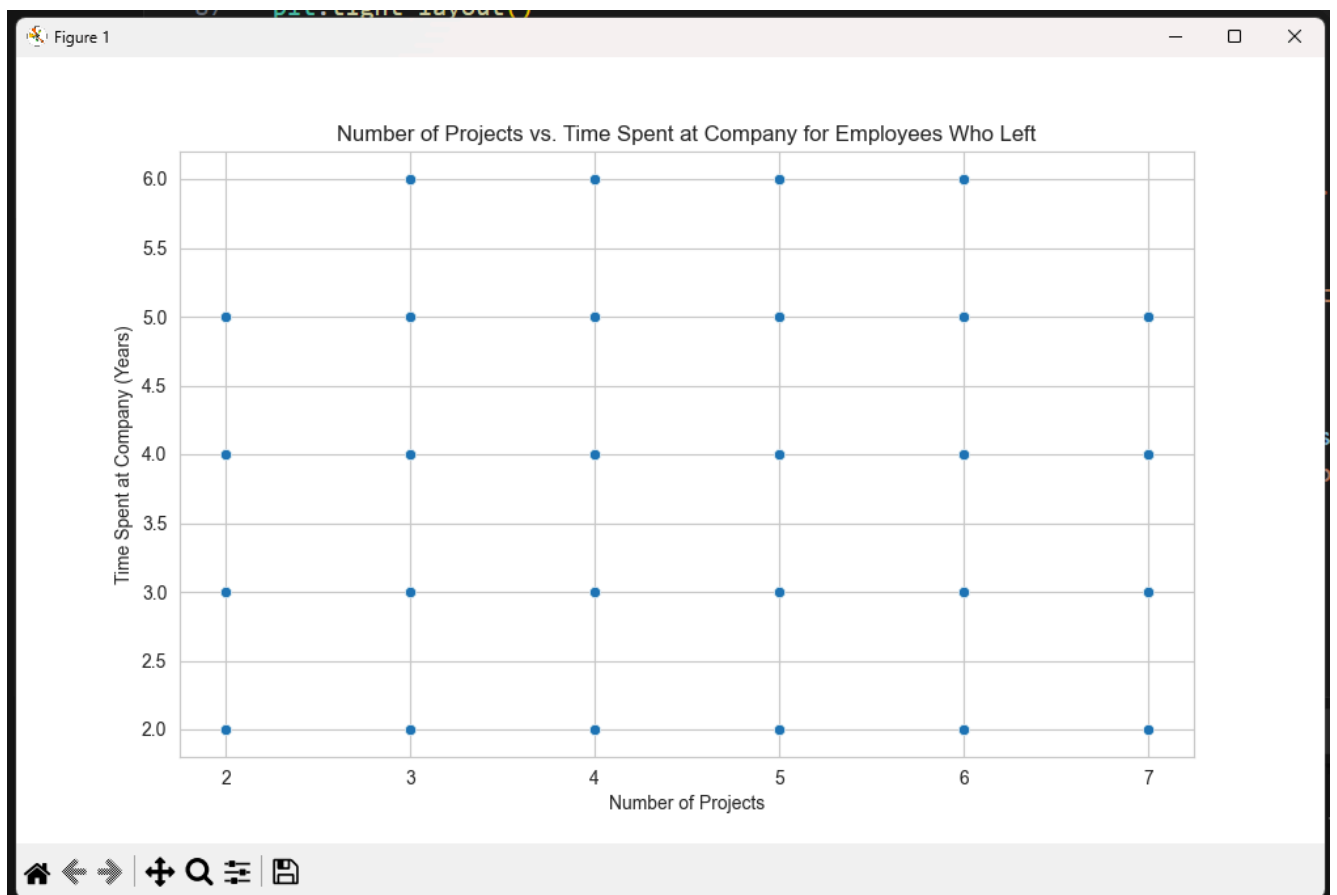
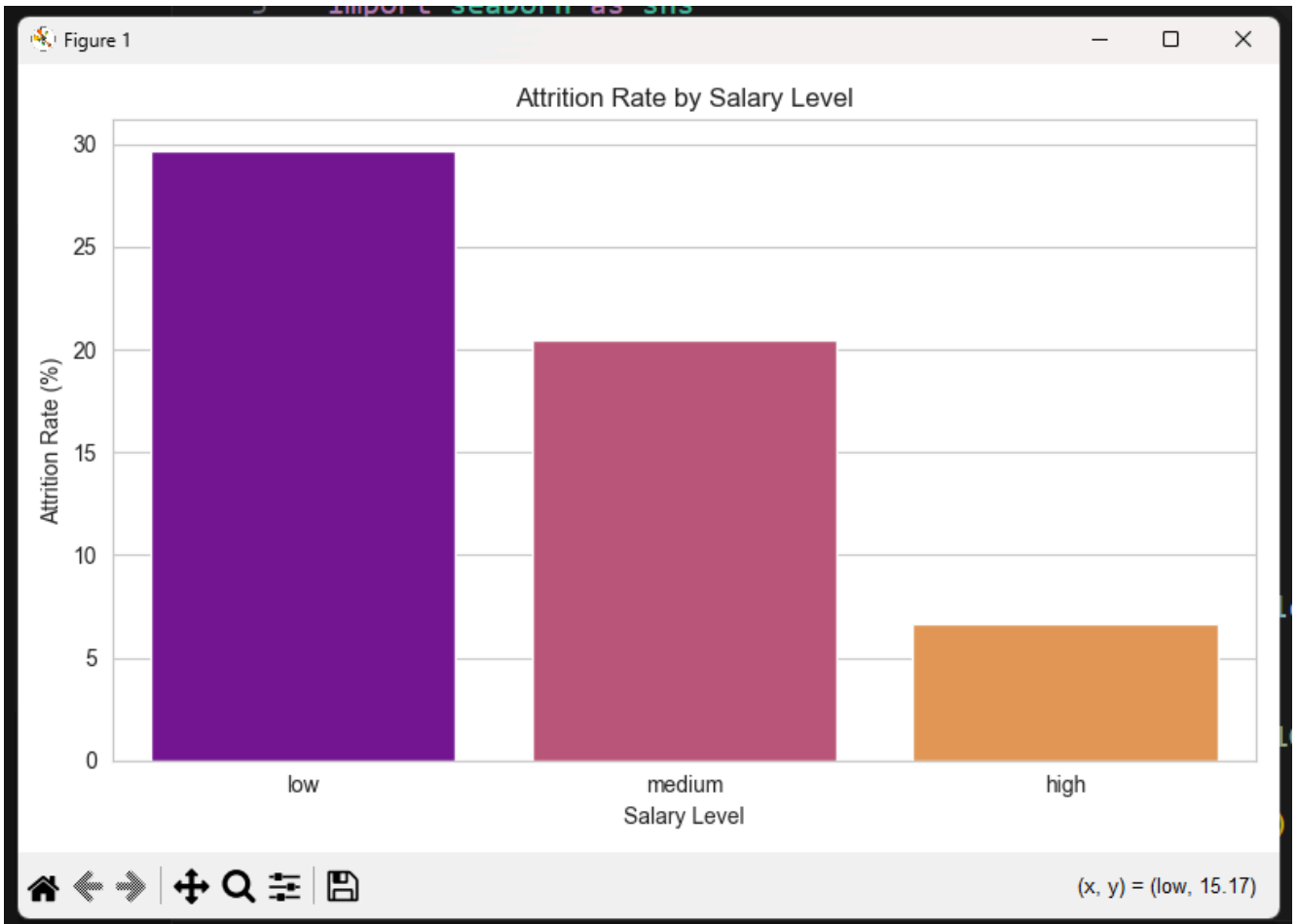
left	0	1	attrition_rate
workAccident			
0	0.734840	0.265160	26.515978
1	0.922084	0.077916	7.791609

Attrition by Promotion in Last 5 Years

left	0	1	attrition_rate
promotionInLast5years			
0	0.758038	0.241962	24.196185
1	0.940439	0.059561	5.956113

Visualization:





Overall Attrition Rate:

The overall attrition rate at SimpleYard is 23.81%. This means that nearly a quarter of the employees have left the organization.

Analysis of Attrition Patterns

1. Numerical Features

By comparing the descriptive statistics and distribution plots for employees who stayed versus those who left, several patterns emerge:

- Satisfaction Level (**satisfactoryLevel**):
 - Employees who left generally had significantly lower satisfaction levels (mean of ~0.44) compared to those who stayed (mean of ~0.67).
 - The distribution plot clearly shows two peaks for employees who left: one group with very low satisfaction (below 0.2) and another group with relatively high satisfaction (around 0.7-0.9), but still lower than the peak for employees who stayed. This suggests both extremely dissatisfied employees and those who might be highly competent but are leaving for other reasons.
- Last Evaluation (**lastEvaluation**):
 - The average **lastEvaluation** scores are very similar for both groups (around 0.71-0.72).
 - However, the distribution plot reveals that employees who left tend to have either very low evaluation scores (below 0.5) or very high scores (above 0.8), with a noticeable dip in the middle range. This could indicate that both underperformers and high-performers are prone to leaving.
- Number of Projects (**numberOfProjects**):
 - Employees who left worked on a slightly higher average number of projects (~3.86) compared to those who stayed (~3.79).
 - The distribution plot for leavers shows distinct peaks at 2 projects, 4 projects, and 6-7 projects. This suggests that employees with very few projects (possibly underutilized) and those with a very high number of projects (possibly overworked) are more likely to leave.
- Average Monthly Hours (**avgMonthlyHours**):
 - On average, employees who left worked slightly more monthly hours (~207) than those who stayed (~199).
 - The distribution plot shows three distinct groups among leavers: those working low hours (around 130-160), those working average hours (around 220-250), and those working very high hours (above 260). This

indicates that both underwork and overwork could be contributing factors.

- Time Spent at Company ([timeSpent.company](#)):
 - Employees who left spent a slightly higher average time at the company (~3.88 years) compared to those who stayed (~3.38 years).
 - The distribution plot highlights significant peaks for leavers at 3, 4, 5, and 6 years. This suggests critical attrition points occur around the 3 to 6-year mark, particularly at 4-5 years.

2. Categorical Features

- Department ([dept](#)):
 - The HR department has the highest attrition rate at approximately 29.09%, followed closely by Accounting (26.60%) and Technical (25.63%).
 - Management (14.44%) and RandD (15.37%) departments show the lowest attrition rates.
 - The bar plot for attrition by department visually confirms these differences.
- Salary ([salary](#)):
 - Employees with low salaries have the highest attrition rate at 29.69%.
 - Medium salary employees show an attrition rate of 20.43%.
 - Employees with high salaries have a significantly lower attrition rate of only 6.63%.
 - The bar plot for attrition by salary level clearly illustrates this inverse relationship: lower salary correlates with higher attrition.
- Work Accident ([workAccident](#)):
 - Employees who had a work accident are much less likely to leave, with an attrition rate of 7.79%, compared to those who did not have a work accident (26.52%).
- Promotion in Last 5 Years ([promotionInLast5years](#)):
 - A significant disparity exists: employees who did not receive a promotion in the last 5 years have an attrition rate of 24.20%.
 - In contrast, employees who received a promotion in the last 5 years have a remarkably low attrition rate of 5.96%.
 - This strongly suggests that lack of promotion is a major contributor to employee churn.

Summary of Attrition Patterns and Key Factors

The analysis reveals several critical factors associated with employee attrition:

- Low Satisfaction and Extreme Workload: Employees are more likely to leave if they have very low satisfaction levels or if they are either significantly underworked (few projects, low hours) or significantly overworked (many projects, high hours).
- Time at Company: A critical period for attrition appears to be between 3 to 6 years of tenure.
- Compensation: Employees with lower salaries are substantially more likely to leave.
- Lack of Career Progression: Not receiving a promotion in the last five years dramatically increases the likelihood of an employee leaving the company.
- Departmental Differences: Some departments, particularly HR, Accounting, and Technical, experience higher attrition rates.
- Work Accidents: Paradoxically, employees who had a work accident are less likely to leave, possibly due to support mechanisms in place.

To support SimpleYard's HR team in developing data-driven employee retention strategies, the following actions are recommended, leveraging the insights from the analysis:

1. Re-evaluate and Adjust Compensation Structures:

- Strategy: Conduct a comprehensive review of salary bands, particularly for roles typically associated with "low" salary levels, and benchmark them against industry standards. Implement adjustments to ensure competitive compensation, especially for critical roles or departments experiencing high attrition.
- Data-Driven Rationale: Employees with low salaries have the highest attrition rate at 29.69%, significantly higher than medium (20.43%) and high (6.63%) salary groups. Compensation levels directly influence an employee's decision to leave.

2. Enhance Career Development and Promotion Opportunities:

- Strategy: Establish clear career progression paths and promotion criteria within the organization. Ensure regular performance reviews are linked to development plans and provide consistent, transparent opportunities for advancement. Focus on identifying and fast-tracking high-potential employees.
- Data-Driven Rationale: The attrition rate for employees who did not receive a promotion in the last 5 years is 24.20%, drastically higher than the 5.96% attrition for those who were promoted. Lack of promotion history is a strong predictor of attrition.

3. Optimize Workload and Project Management:

- Strategy: Implement systems to monitor and balance employee workloads and project assignments. For employees with
- fewer than 3 projects, investigate potential underutilization and offer new challenges or cross-functional opportunities. For those with a
- high number of projects (6-7) or excessive average monthly hours (above 260), assess burnout risk and redistribute tasks or provide additional resources.
- Data-Driven Rationale: Employees working on fewer than 3 projects are more likely to leave. Both very low and very high
- `numberOfProjects` and `avgMonthlyHours` correlate with higher attrition. The analysis showed distinct peaks for leavers at both ends of the project and hours spectrums.

4. Strengthen Employee Engagement and Address Satisfaction Gaps:

- Strategy: Conduct targeted "stay interviews" with current employees, especially those showing signs of low satisfaction, to understand their concerns and actively work towards solutions. Implement feedback mechanisms and act on employee suggestions to improve overall workplace satisfaction.
- Data-Driven Rationale: Employees who left had significantly lower `satisfactoryLevel` on average (~0.44) compared to those who stayed (~0.67). Analyzing employee attrition patterns using statistical summaries and visualizations can help identify key factors.

5. Implement Targeted Department-Specific Retention Programs:

- Strategy: Develop tailored retention initiatives for departments with the highest attrition rates, specifically HR (29.09%), Accounting (26.60%), and Technical (25.63%). This could involve department-specific training, leadership development, team-building activities, or addressing unique challenges within those teams.
- Data-Driven Rationale: Identifying which departments are experiencing the highest rates of attrition is a critical question. Departmental distribution is a key factor associated with employees leaving the company.

6. Proactive Retention at Key Tenure Milestones:

- Strategy: Design programs to proactively engage employees around their 3, 4, 5, and 6-year anniversaries. This could include career planning discussions, mentorship opportunities, or special recognition to acknowledge their contributions and reaffirm their value to the company.
- Data-Driven Rationale: The number of projects correlates with time spent at the company, particularly for those who have left. The analysis showed significant peaks in attrition at the 3, 4, 5, and 6-year marks in [timeSpent.company](#).

7. Leverage Positive Workplace Experiences:

- Strategy: Analyze the factors contributing to the lower attrition rate among employees who experienced a work accident (7.79% vs. 26.52% for those without). If this indicates strong company support or empathy, identify these positive elements and explore how to extend similar feelings of support and security across the broader organization.
- Data-Driven Rationale: Work accident is a factor associated with employees leaving the company. The data showed a significantly lower attrition rate for employees who had a work accident.

Summary:

This HR Analytics project for SimpleYard addresses a critical challenge: a high employee attrition rate impacting growth and efficiency. The primary objective is to leverage data analytics to identify the key factors contributing to employees leaving the organization.

Through exploratory data analysis of employee records, using statistical summaries and visualizations, the project identified significant patterns. Key factors associated with attrition include:

- Low compensation levels
- Lack of promotion opportunities
- Extreme workloads (both under- and over-utilization)
- Specific departmental distributions (e.g., HR, Accounting, Technical having higher attrition)
- Critical tenure periods (e.g., around 3-6 years at the company)
- Low employee satisfaction levels

By quantifying these attrition rates and analyzing these patterns, the project provides actionable insights to support SimpleYard's HR team in developing targeted, data-driven retention strategies aimed at reducing future employee churn and fostering a more stable workforce.