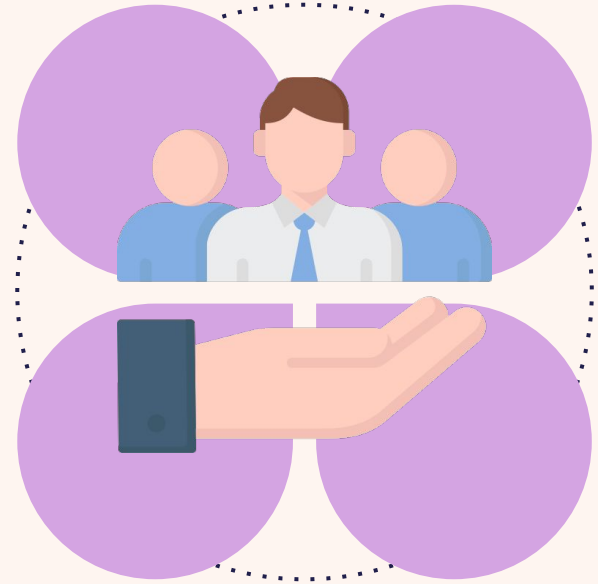


Employee Attrition & Retention Analysis: Insights for HR

Data Analysis Project
Conducted by Kamilia Firzanah (2024)





Introduction






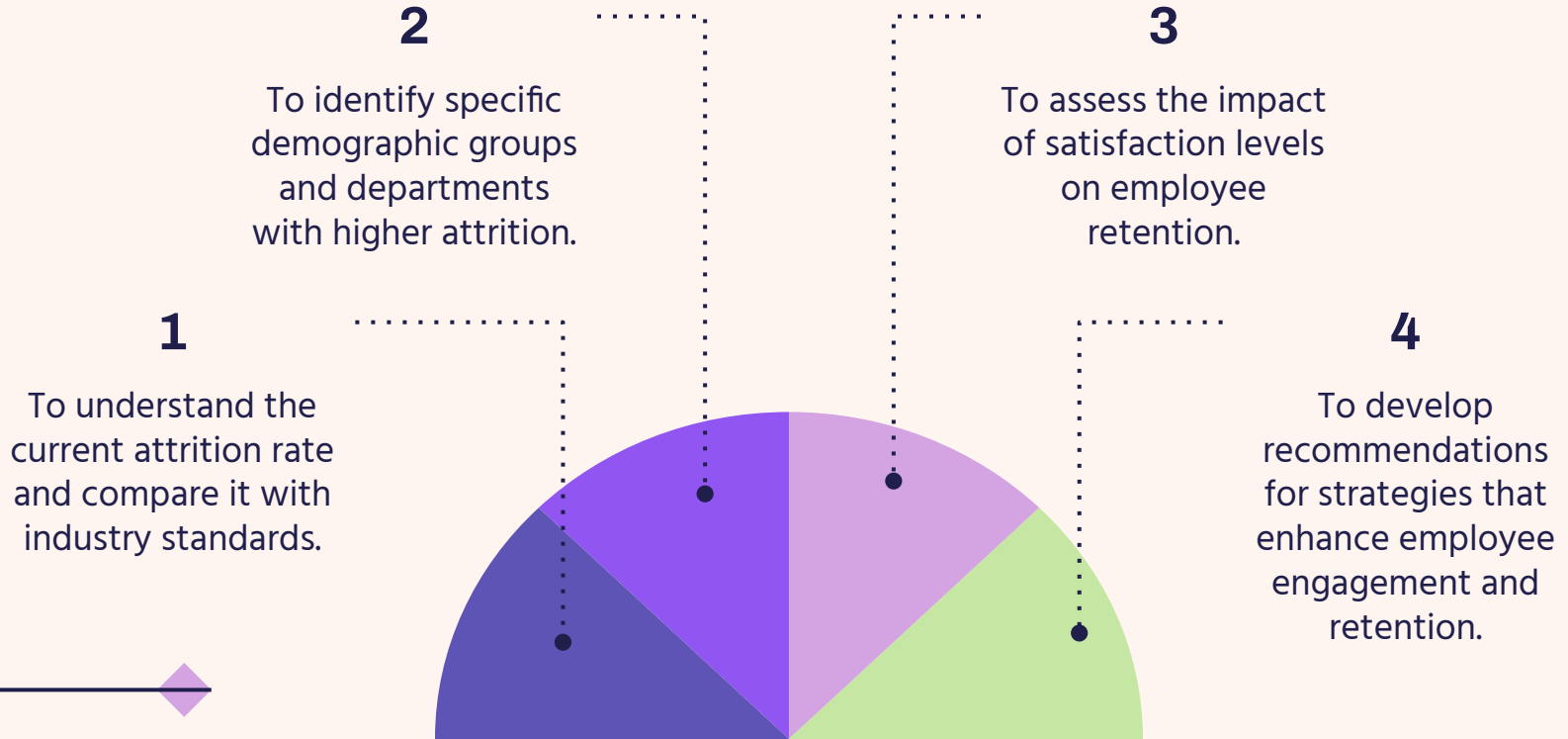
Introduction

Employee attrition has become a pressing challenge for many organizations, especially as talent markets grow more competitive and the costs of hiring and training new employees continue to rise.

This HR Employee Attrition Analysis project seeks to **uncover the main factors behind employee attrition** and provide HR with data-driven insights to **enhance retention strategies.**



Objectives



Data & Tools

The Kaggle logo, featuring the word "kaggle" in a blue, lowercase, sans-serif font, centered within a light green rounded square.

kaggle

The dataset used in this analysis is IBM HR Analytics Employee Attrition & Performance, sourced from [Kaggle](#). It's a fictional dataset created by IBM data scientists.



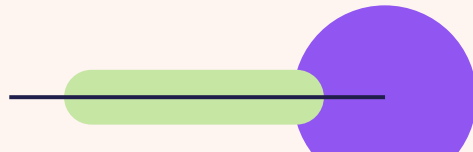
The dataset contains **1,470 rows and 35 columns**, including 17 numerical columns and 18 categorical columns.



The tools used in this analysis are **Python** (programming language) and **Google Colab** (IDE). Python libraries used are `opendatasets`, `pandas`, `matplotlib`, and `seaborn`.



Data Preparation



Data Preview

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4
...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	3
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	1
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	2
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	4
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	1

1470 rows × 35 columns

Column Names & Description

No.	Column Name	Description
1	Age	Employee age
2	Attrition	Did employee leave company or not?
3	Business Travel	Travel_Rarely, Travel_Frequently, or Non-Travel for work
4	DailyRate	How much an employee made in one day
5	Department	Department/Field employee worked
6	DistanceFromHome	Distance from home in minutes
7	Education	Years of education completed
8	EducationField	Employee's major/minor or academic focus
9	EmployeeCount	Employee count per row
10	EmployeeNumber	Primary key for employees
11	EnvironmentSatisfaction	Employee environment satisfaction on a scale from 1-4
12	Gender	Male or female
13	HourlyRate	Hourly rate of employee
14	JobInvolvement	Job involvement on a scale from 1-4
15	JobLevel	Job level on a scale from 1-4
16	JobRole	Job roles include: Nurse, Other, Therapist, Administrative, Admin
17	JobSatisfaction	Job satisfaction on a scale from 1-4

Column Names & Description

No.	Column Name	Description
18	MaritalStatus	Single, Married, Divorced
19	MonthlyIncome	Total monthly income of employee
20	MonthlyRate	Total monthly rate of employee
21	NumCompaniesWorked	Number of companies employee is employed at
22	Over18	Employee is over the age of 18: Y or N
23	OverTime	Employee did overtime: Yes or No
24	PercentSalaryHike	Percent of employee's salary raise
25	PerformanceRating	Performance of an employee on a scale from 1-4
26	RelationshipSatisfaction	Employee's personal relationship satisfaction on a scale from 1-4
27	StandardHours	Total hours an employee works in two weeks
28	StockOptionLevel	Level of stock options granted to an employee from 1-4
29	TotalWorkingYears	How many years employee worked
30	TrainingTimesLastYear	How many training events employee had last year
31	WorkLifeBalance	Employee's rating of work life balance on a scale from 1-4
32	YearsAtCompany	Years employee is at current company
33	YearsInCurrentRole	Years employee is in current position
34	YearsSinceLastPromotion	Years since employee's last promotion

Data Assessment & Cleaning

Check for column' names consistency

1. Age	18. MaritalStatus
2. Attrition	19. MonthlyIncome
3. BusinessTravel	20. MonthlyRate
4. DailyRate	21. NumCompaniesWorked
5. Department	22. Over18
6. DistanceFromHome	23. OverTime
7. Education	24. PercentSalaryHike
8. EducationField	25. PerformanceRating
9. EmployeeCount	26. RelationshipSatisfaction
10. EmployeeNumber	27. StandardHours
11. EnvironmentSatisfaction	28. StockOptionLevel
12. Gender	29. TotalWorkingYears
13. HourlyRate	30. TrainingTimesLastYear
14. JobInvolvement	31. WorkLifeBalance
15. JobLevel	32. YearsAtCompany
16. JobRole	33. YearsInCurrentRole
17. JobSatisfaction	34. YearsSinceLastPromotion
	35. YearsWithCurrManager

Column names

No inconsistency were found.

All column names are written in sentence case with no separator between the words.

Data Assessment & Cleaning

Detect duplicates

```
df.duplicated().sum()
```

```
0
```

Number of duplicates

No duplicates values were found.

Each record is unique.

Data Assessment & Cleaning

Detect missing values

Age	0.0	MaritalStatus	0.0
Attrition	0.0	MonthlyIncome	0.0
BusinessTravel	0.0	MonthlyRate	0.0
DailyRate	0.0	NumCompaniesWorked	0.0
Department	0.0	OverTime	0.0
DistanceFromHome	0.0	PercentSalaryHike	0.0
Education	0.0	PerformanceRating	0.0
EducationField	0.0	RelationshipSatisfaction	0.0
EmployeeNumber	0.0	StockOptionLevel	0.0
EnvironmentSatisfaction	0.0	TotalWorkingYears	0.0
Gender	0.0	TrainingTimesLastYear	0.0
HourlyRate	0.0	WorkLifeBalance	0.0
JobInvolvement	0.0	YearsAtCompany	0.0
JobLevel	0.0	YearsInCurrentRole	0.0
JobRole	0.0	YearsSinceLastPromotion	0.0
JobSatisfaction	0.0	YearsWithCurrManager	0.0

Percentage of missing values of each column

No missing values were found.

Each row and column contains values.

Data Assessment & Cleaning

Detect constant columns

	0
EmployeeCount	1
Over18	1
StandardHours	1

Constant columns and the unique value

Three (3) constant columns were found.

Constant column is a column that has one unique value only. In the data used, the columns are EmployeeCount, Over18, and StandardHours.

Data Assessment & Cleaning

Convert data types

Convert categorical columns to 'category'

object_cols

```
['Attrition',  
'BusinessTravel',  
'Department',  
'EducationField',  
'Gender',  
'JobRole',  
'MaritalStatus',  
'OverTime']
```

**Nominal columns
stored as 'object'**

ordinal_cols

```
['Education',  
'EnvironmentSatisfaction',  
'JobInvolvement',  
'JobLevel',  
'JobSatisfaction',  
'PerformanceRating',  
'RelationshipSatisfaction',  
'WorkLifeBalance',  
'StockOptionLevel']
```

**Ordinal columns
stored as 'int64'**

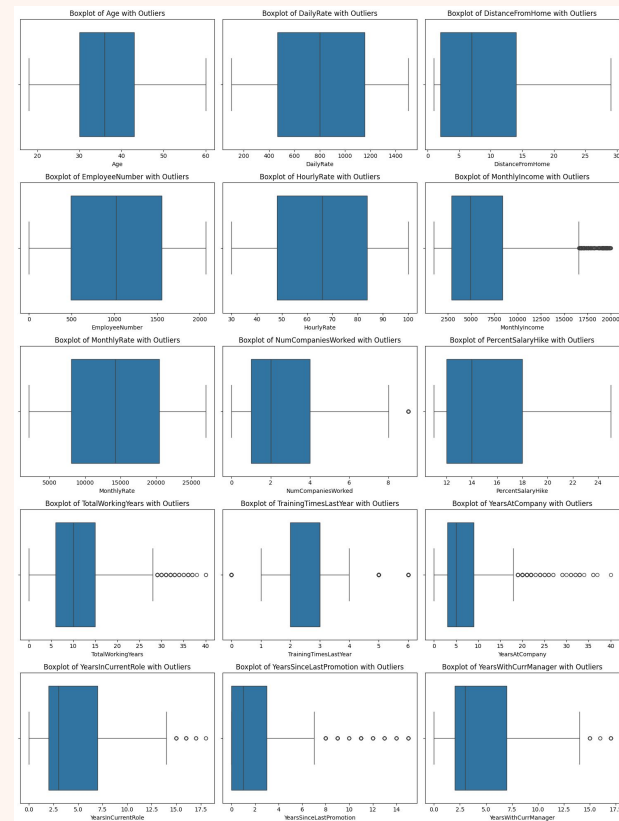
**Convert to
'category'**

```
convert_to_cat = object_cols + ordinal_cols  
for col in convert_to_cat:  
    df[col] = df[col].astype('category')
```

Data Assessment & Cleaning

Detect outliers

According to the box plots, there are 8 columns containing outliers. However, **I'm going to keep these outliers** because they're most likely **true outliers**, which are data points that significantly deviates from the general trend or pattern of the dataset, but is not due to errors in data collection or entry.



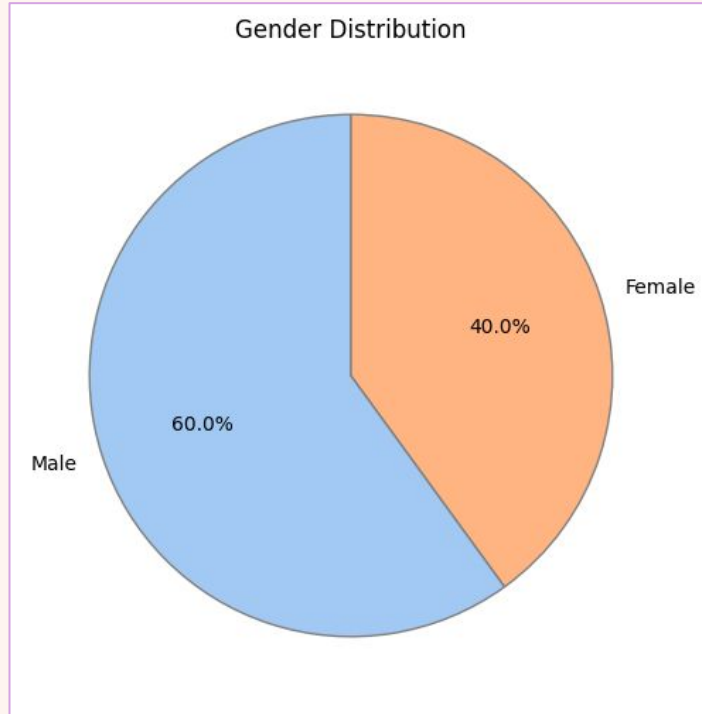


Analysis



Demographic

Univariate Analysis



Number of Employees & Gender Distribution

There were **1,470 employees**, including 882 males (60%) and 588 females (40%).

This is **close to a balanced gender distribution**, with a slightly higher proportion of male employees.

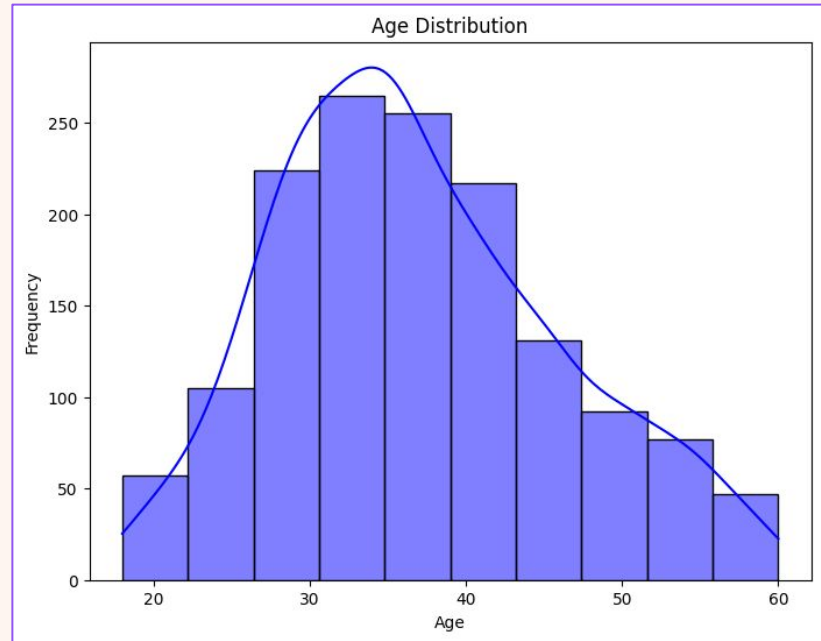
Demographic

Univariate Analysis

Age

Employees' ages range **from 18 to 60 years old, with an average age of 36.**

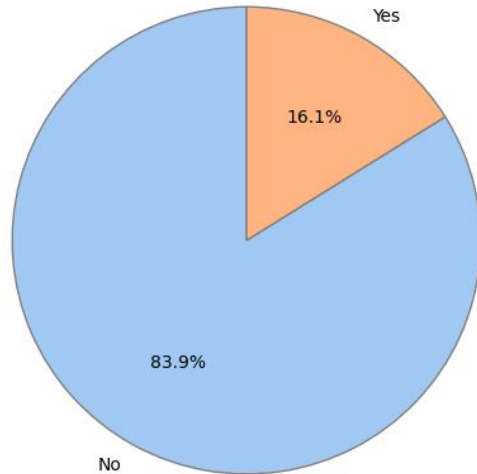
This indicates that **the company has a diverse workforce in terms of age**, including both early-career employees and those approaching retirement.



Demographic

Univariate Analysis

Attrition Distribution



Attrition Distribution & Rate

Out of 1,470 employees, **237 employees (16.12%) have left the company**, while 1,233 employees (83.9%) chose to stay.

The overall attrition rate is 16.12%.

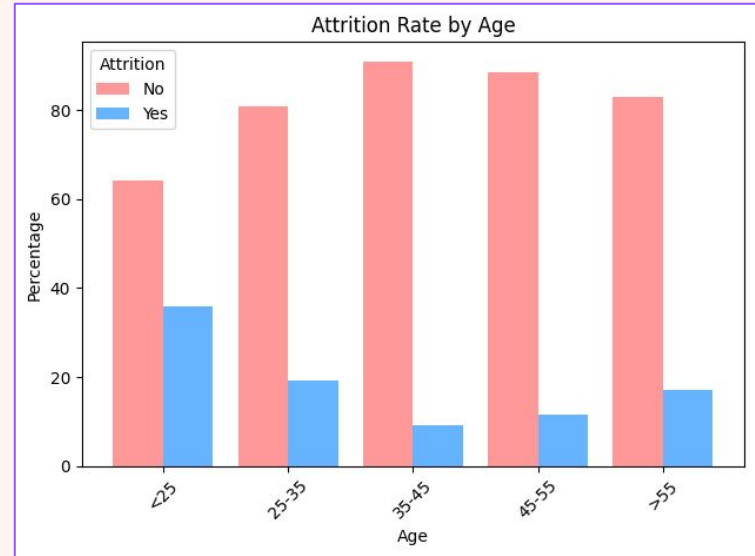
According to experts, healthy organizations typically maintain an attrition rate of 10% or less, suggesting that **this company's attrition rate may be a potential area for improvement.**

Bivariate Analysis

Attrition & Age

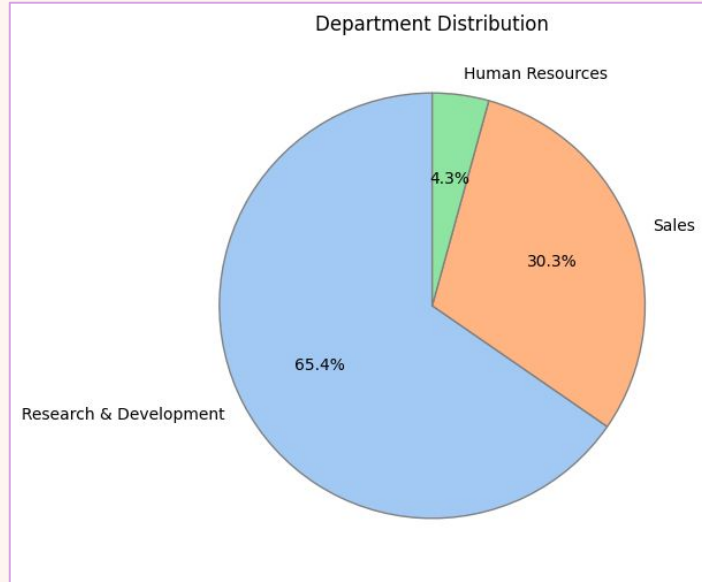
Employees under the age of 25 experienced the highest attrition rate.

This is most likely because this age group tends to be early in their careers, often seeking new experiences, career advancement, or educational opportunities, making them more open to job changes.



Univariate Analysis

Department



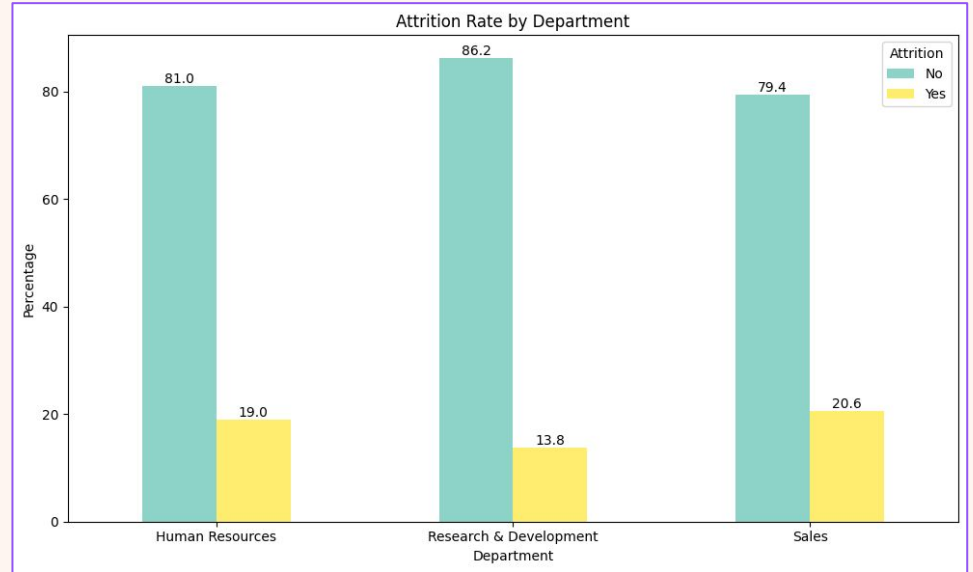
Employees in the R&D department dominate the company, making up 65% (961 employees) of the workforce, followed by Sales with 30.3% (446 employees) and HR with 4.3% (63 employees).

Bivariate Analysis

Attrition & Department

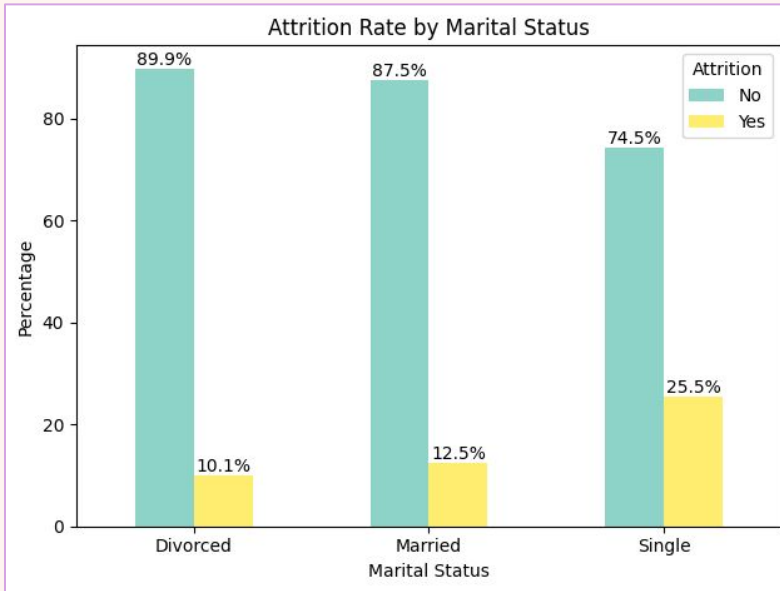
The R&D department has the lowest attrition rate at 13.84%. Since R&D employees make up the largest portion of the company, this suggests that they are generally content with their roles.

However, **attention is needed in the HR and Sales departments, as both have attrition rates of 19.05% and 20.63%, respectively, which are above the company average of 16.1%.**



Bivariate Analysis

Attrition & Marital Status



A larger proportion of **single employees (25.35%)** tend to leave the company compared to married and divorced employees. This could be attributed to factors like seeking new opportunities, exploring different career paths, or lifestyle choices.

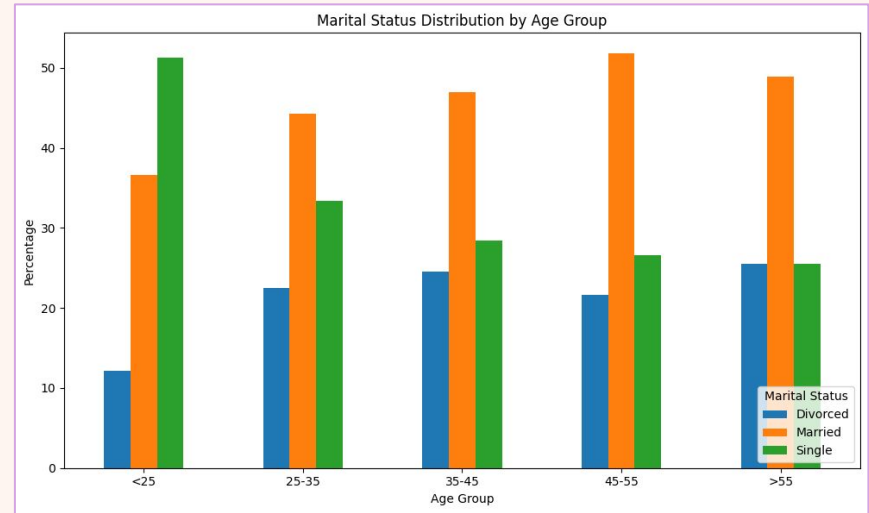
Multivariate Analysis

Attrition, Marital Status, and Age Group

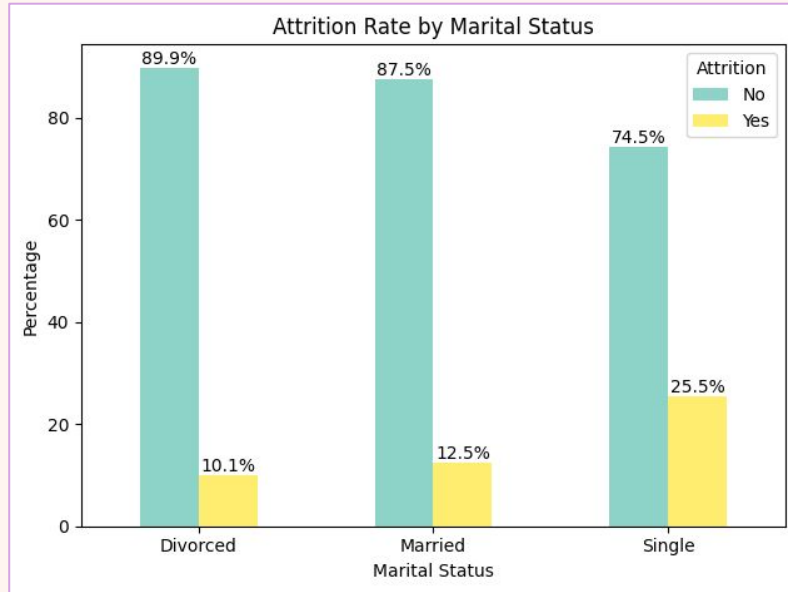
Previously, we gained insight that employees under the age of 25 have the highest attrition rate. Now, we want to analyze the relationship between age and marital status.

According to this graph, **the majority of single employees belong to the <25 age group.** This aligns with the general trend that younger people are less likely to be married or divorced.

From this, we can conclude that **employees who are single and under the age of 25 tend to leave the company more.** This could be attributed to factors like career exploration, fewer commitments, or a desire for change.



Bivariate Analysis



Attrition & Marital Status

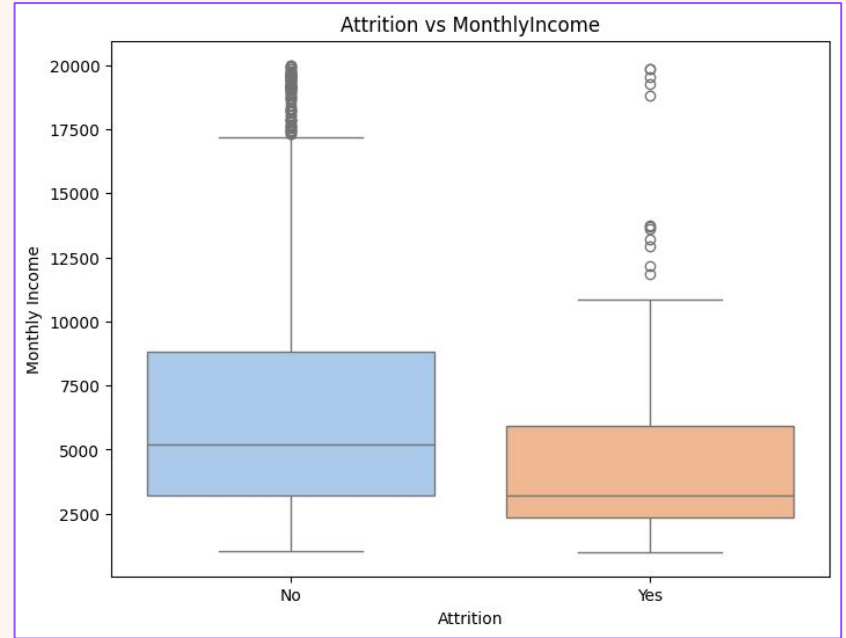
A larger proportion of **single employees (25.35%)** tend to leave the company compared to married and divorced employees. This could be attributed to factors like seeking new opportunities, exploring different career paths, or lifestyle choices.

Bivariate Analysis

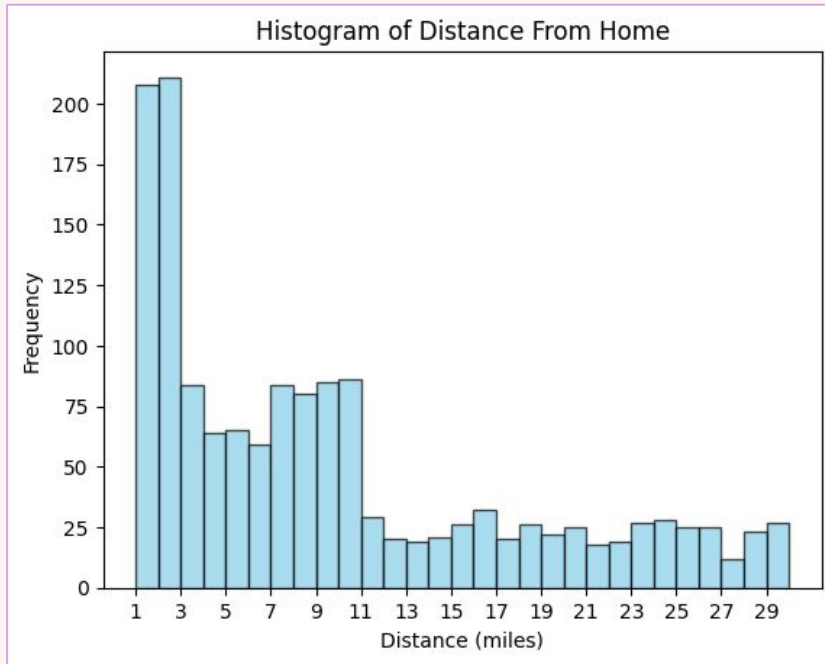
Attrition and Monthly Income

The boxplot suggests a potential relationship between monthly income and attrition. Employees who stayed tend to have a higher median monthly income compared to those who left.

This suggests that **employees with higher income might be more satisfied with their compensation and less likely to leave the company.**



Univariate Analysis



Distance From Home

The majority of employees live close to the workplace, with the highest frequency observed for distances between 1 and 3 miles.

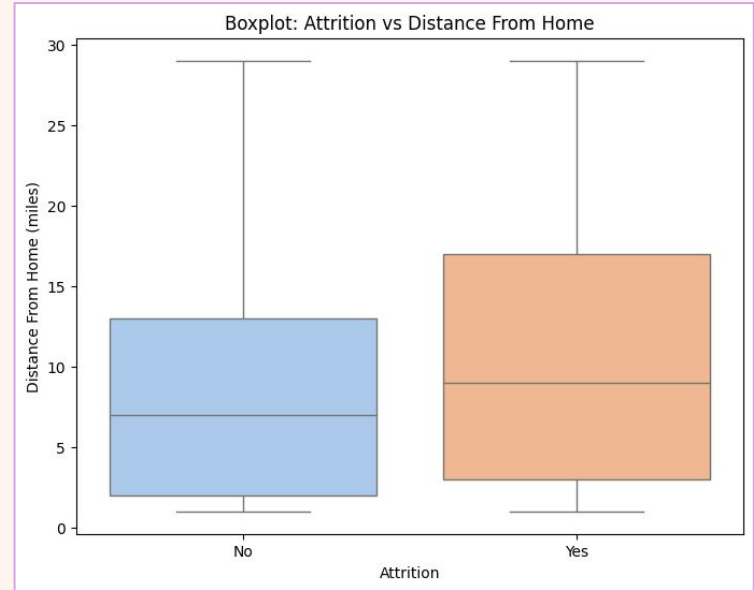
As the distance increases, the frequency generally decreases, indicating that **fewer employees commute from longer distances**. This suggests a concentration of employees living near the workplace.

Bivariate Analysis

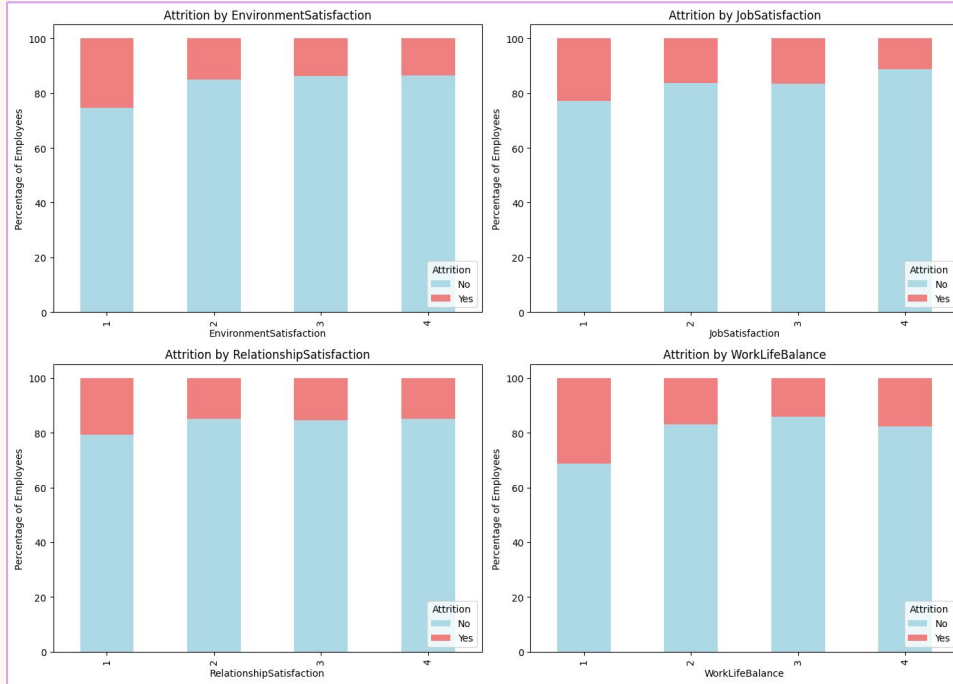
Attrition and Distance from Home

The boxplot shows that **the relationship between attrition and distance from home is not very strong.**

While there seems to be a slight increase in attrition rates for employees living farther away, the difference is not very significant. This suggests that **distance from home is not a major factor influencing employee attrition in this case.**



Univariate Analysis



Attrition and Satisfaction Level

Employees' satisfaction with their environment, job, relationships, and work-life balance are all strongly linked to attrition rates.

Those who are more satisfied with their work environment, job, relationships with colleagues and supervisors, and ability to balance work and personal life tend to have lower attrition rates.

Companies that focus on enhancing these aspects can create a more positive environment and improve employee retention.



Insights & Recommendations



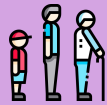
Insights



Demographics

- The company has 1470 employees, with a gender distribution of 60% male and 40% female, which is close to balanced.
- The overall **attrition rate is 16.12%**, above the ideal threshold of 10%, **suggesting an area for improvement.**
- The workforce is diverse, ranging in age from 18 to 60 years, with an average age of 36 years.

Insights



Attrition & Age

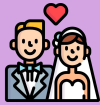
Employees **under 25 have the highest attrition rate**, likely due to career exploration and job changes at early career stages.



Attrition & Department

- R&D has the largest workforce (65%) and the lowest attrition rate (13.84%).
- **HR and Sales departments have attrition rates above the company average** (19.05% and 20.63%, respectively), indicating potential retention issues.

Insights



Attrition & Marital Status

Single employees, especially those under 25, are more likely to leave the company, possibly due to seeking career advancement or lifestyle changes.



Attrition & Monthly Income

Employees who stay tend to have a higher median monthly income, suggesting that **higher compensation correlates with lower attrition.**

Insights



Attrition & Distance from Home

While most employees live close to the workplace, distance from home does not strongly influence attrition rates, indicating that **commuting distance is not a major factor in attrition.**



Attrition & Satisfaction Levels

Satisfaction with the work environment, job, relationships, and work-life balance is strongly linked to attrition. **Employees with higher satisfaction are less likely to leave.**

Recommendations

for HR & Employee Retention

Improve Retention in HR and Sales Departments

HR and Sales have higher-than-average attrition rates. Investigate the reasons behind this, such as management issues, or implement targeted retention strategies in these departments.

Focus on Reducing Attrition Among Younger Employees

Given that employees under 25 have the highest attrition rates, HR should focus on offering career development opportunities, mentorship, and clear growth paths for younger employees to encourage retention.

Enhance Job Satisfaction and Work Environment

Since satisfaction levels are strongly linked to attrition, HR should prioritize improving employees' satisfaction with their work environment, job role, relationships, and work-life balance. Flexible work arrangements, recognition programs, and a supportive company culture can help reduce attrition.



Recommendations

for HR & Employee Retention

Support Single Employees Under 25

The higher attrition rate among single employees, particularly those under 25, may indicate a desire for career exploration. Offering flexible work arrangements, career coaching, or rotation programs could help retain this group.

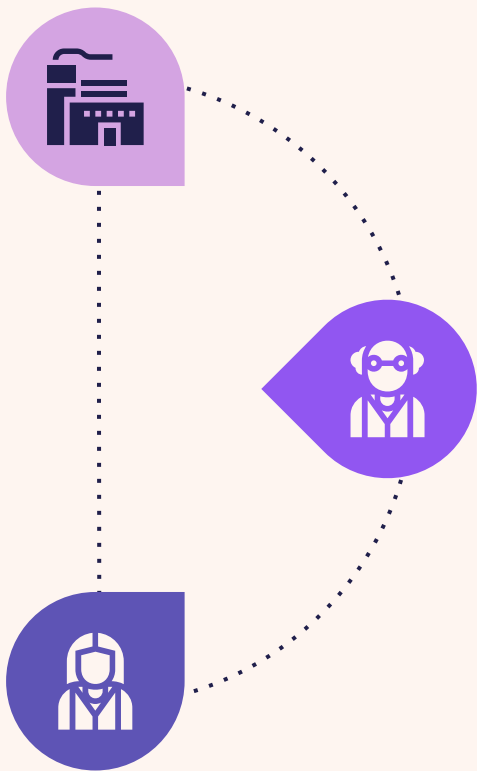
Salary & Compensation

Employees who stay tend to have higher incomes, suggesting that compensation plays a key role in retention. HR should review salary packages to ensure they are competitive and aligned with employee expectations.

Leverage Proximity to Workplace

While commuting distance does not seem to be a major factor in attrition, HR could still consider offering incentives for employees with longer commutes, such as travel allowances, remote work options, or flexible hours to improve overall job satisfaction.





Thank You!

If there's anything you want to discuss, feel free to contact me through:

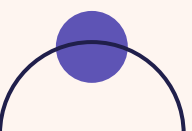


kamilia.firzanah@gmail.com



www.linkedin.com/in/kamilia-firzanah

If you want to access the code behind this project, click on: [Google Colab](#)

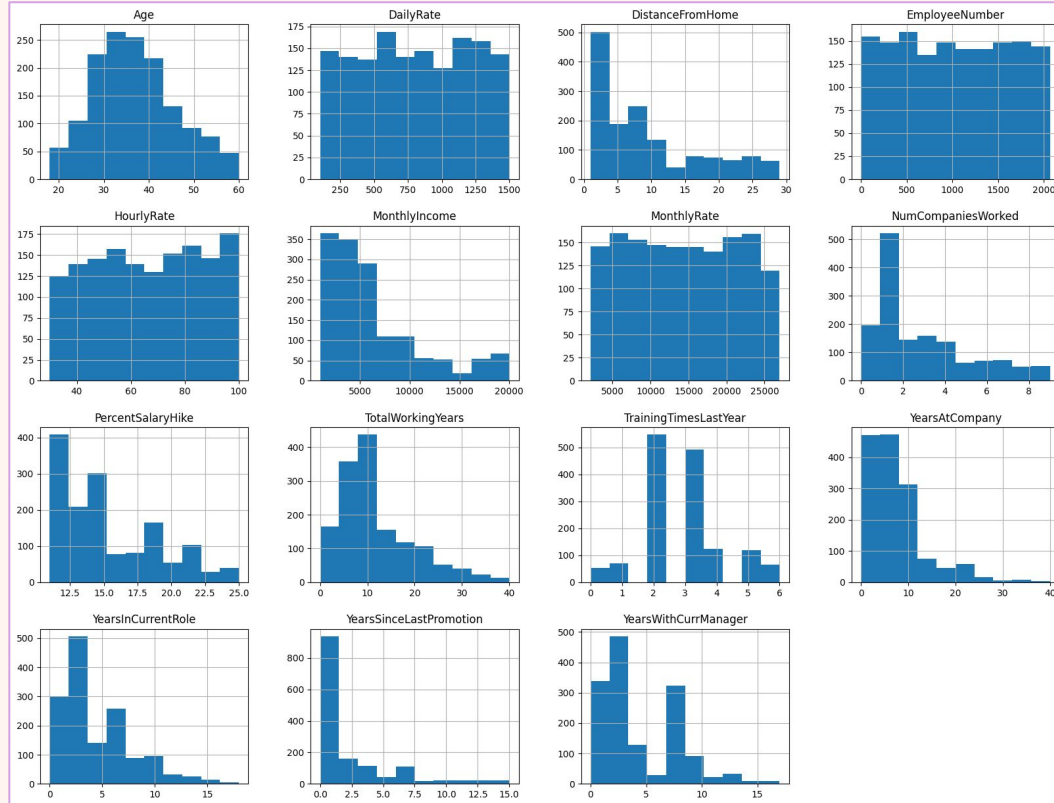




Additional



Histogram of Numerical Features



Correlation Matrix of Numerical Features

