

Employee Attrition Analysis Report

Name: Russell Jamieson

Date: August 17, 2024

Exploring the correlation between employee attrition and various factors such as age, salary, job satisfaction, etc.

Background

Attrition is the departure of employees from an organization for any particular reason. Reasons may range from retirement or resignation, voluntary attrition, to termination or death, involuntary attrition. Employee attrition is an inevitable part of any business. However, when attrition crosses a particular threshold, it becomes a cause for concern. For example, attrition amongst employees of a particular minority group can hurt the diversity of an organization. Attrition amongst senior leaders can lead to a significant gap in organizational leadership.

Objective

The main objective of this analysis is to identify key factors contributing to employee attrition.

Dataset

- The dataset used in this analysis was provided by [Kaggle.com](https://www.kaggle.com)
- The dataset contains 1470 rows of data.
- The data contains 35 columns namely:
 - Age
 - Attrition
 - BusinessTravel
 - DailyRate
 - Department
 - DistanceFromHome
 - Education
 - EducationField
 - EmployeeCount
 - EmployeeNumber
 - EnvironmentSatisfaction
 - Gender
 - HourlyRate
 - JobInvolvement
 - JobLevel
 - JobRole
 - JobSatisfaction
 - MaritalStatus

- MonthlyIncome
- MonthlyRate
- NumCompaniesWorked
- Over18
- OverTime
- PercentSalaryHike
- PerformanceRating
- RelationshipSatisfaction
- Standard Hours
- StockOptionLevel
- TotalWorkingYears
- TrainingTimesLastYear
- WorkLifeBalance
- YearsAtCompany
- YearsInCurrentRole
- YearsSinceLastPromotion
- YearsWithCurrManager

The following variables are considered factors for attrition: Education, EnvironmentSatisfaction, JobInvolvement, JobSatisfaction, PerformanceRating, RelationshipSatisfaction, WorkLifeBalance

Each factor has ranking system with the following definitions:

Education: 1: "Below College" (Secondary School/Trade School/No College Education) 2: "College" (Currently enrolled in a post-secondary institution) 3: "Bachelor" (Earned Bachelor's degree) 4: "Master" (Earned Master's degree) 5: "Doctor" (Earned Doctorate)

EnvironmentSatisfaction: 1: "Low" 2: "Medium" 3: "High" 4: "Very High"

JobInvolvement 1: "Low" 2: "Medium" 3: "High" 4: "Very High"

JobSatisfaction 1: "Low" 2: "Medium" 3: "High" 4: "Very High"

PerformanceRating 1: "Low" 2: "Medium" 3: "High" 4: "Very High"

RelationshipSatisfaction 1: "Low" 2: "Medium" 3: "High" 4: "Very High"

WorkLifeBalance 1: "Poor" 2: "Fair" 3: "Good" 4: "Great"

Data Preparation

We processed and cleaned the data through Excel, as our CSV file was easily accessible through it and modifications were made simpler.

We began with checking to ensuring that there existed no duplicated rows, and removing any if such exists. We then deleted any unnecessary columns for our analysis. This included removing the columns: Over18, DailyRate, HourlyRate, MonthlyRate, StandardHours, StockOptionLevel, EmployeeCount.

We created a new column, Salary, which calculated the yearly salary of each employee by multiplying the MonthlyIncome of each employee by 12.

We standardized our data ensuring every column was of the right format before beginning the analyzing process.

Exploratory Data Analysis

We performed our analysis in SQL using SQL Server. The updated CSV file was imported into our created database EmployeeAttrition using the Import/Export Wizard and the resulting table was named Attrition_Table.

Overall Attrition

To begin our data analysis, we calculated the overall attrition rate of the organization.

```
SELECT COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM AttritionTable;
```

Output:

Results		Messages	
	TotalEmployees	TotalAttrition	AttritionRate
1	1470	237	16.122448979591

The company's total attrition rate was 16.12%.

Attrition per Department

Next, we calculated the attrition rate across all departments.

```
SELECT Department,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY Department
ORDER BY AttritionRate DESC;
```

Output:

	Results	Messages			
	Department	TotalEmployees	TotalAttrition	AttritionRate	
1	Sales	446	92	20.627802690582	
2	Human Resources	63	12	19.047619047619	
3	Research & Development	961	133	13.839750260145	

The results showed that the Sales department had the highest rate of attrition with 20.62%.

Attrition per Job Role

We then analyzed which job roles were more prone to attrition.

```
SELECT JobRole,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY JobRole
ORDER BY AttritionRate DESC;
```

Output:

	Results	Messages			
	JobRole	TotalEmployees	TotalAttrition	AttritionRate	
1	Sales Representative	83	33	39.759036144578	
2	Laboratory Technician	259	62	23.938223938223	
3	Human Resources	52	12	23.076923076923	
4	Sales Executive	326	57	17.484662576687	
5	Research Scientist	292	47	16.095890410958	
6	Manufacturing Director	145	10	6.896551724137	
7	Healthcare Representative	131	9	6.870229007633	
8	Manager	102	5	4.901960784313	
9	Research Director	80	2	2.500000000000	

We can observe that the role of Sales Representatives had the highest attrition rate with a staggering **39.76%**. This is a significant **increase of 15.82%** from the next role, Laboratory Technician, with an attrition rate of **23.94%**. The lowest attrition rate recorded was **2.5%** from Research Directors with only 2 directors leaving the company.

Attrition by Salary

We wished to see whether salary correlated to attrition rates. With our employees earning a wide range for their salary, we needed a way to categorize each employee's earnings. We created a column **CompanyAverage** and assigned to it values based on the employee's salary.

If an employee was making below the average company salary, we assigned the value *Below Average* in the **CompanyAverage** column. If they were making above the average company salary, the value *Above Average* was assigned to them. Otherwise, they were assigned the value *Average*.

```
ALTER TABLE Attrition_Table
ADD CompanyAverage varchar(50);

UPDATE Attrition_Table
SET CompanyAverage = CASE WHEN Salary > (SELECT AVG(Salary) FROM Attrition_Table)
THEN 'Above Average'
WHEN Salary < (SELECT AVG(Salary) FROM Attrition_Table) THEN 'Below Average'
ELSE 'Average'
END;
```

We found that all employees fell into either one of two categories: *Above Average* or *Below Average*.

Calculating the attrition rate per salary using the following code:

```
SELECT CompanyAverage,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY CompanyAverage
ORDER BY AttritionRate DESC;
```

Output:

	CompanyAverage	TotalEmployees	TotalAttrition	AttritionRate
1	Below Average	977	185	18.935516888433
2	Above Average	493	52	10.547667342799

From the results, we can see that employees who earned less than company average were **8.39%** more likely to leave than employees who made more.

Correlation between attrition with age

Another factor we considered with employee attrition rates is the age of the employees. With our diverse team ranging in age from 18 to 60, we decided to recategorize our employees into three groups: Young Adults for those ages 18 to 30, Adults for those ages 31 to 44, Seniors for 45 and up.

Similar to our Salary grouping, we created a column **AgeGroup** and assigned the values *Young Adult*, *Adult*, and *Senior* based on the employee's age.

```
ALTER TABLE Attrition_Table
ADD AgeGroup varchar(50);

UPDATE Attrition_Table
SET AgeGroup = CASE WHEN Age BETWEEN 18 AND 30 THEN 'Young Adult'
WHEN Age BETWEEN 31 AND 44 THEN 'Adult'
ELSE 'Senior'
END;
```

Calculating the attrition rate per salary using the following code:

```
SELECT AgeGroup,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY AgeGroup
ORDER BY AttritionRate DESC;
```

Output:

Results		Messages		
	AgeGroup	TotalEmployees	TotalAttrition	AttritionRate
1	Young Adult	386	100	25.906735751295
2	Adult	811	103	12.700369913686
3	Senior	273	34	12.454212454212

We see that Young Adults are twice as likely to leave the company than any other age group. This may call into question our inclusivity to younger employees in the workplace.

Advanced Analysis

We will now take a deeper glance at employee attrition by observing combining multiple factors to see their total effect on attrition

Job Role and Overtime

Firstly, we will observe the correlation between employee job roles and whether they work overtime, and see its impact on attrition.

```
SELECT JobRole, OverTime,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY JobRole, OverTime
ORDER BY AttritionRate DESC;
```

Output:

Results		Messages			
	JobRole	OverTime	TotalEmployees	TotalAttrition	AttritionRate
1	Sales Representative	1	24	16	66.66666666666666
2	Laboratory Technician	1	62	31	50.00000000000000
3	Human Resources	1	13	5	38.461538461538
4	Research Scientist	1	97	33	34.020618556701
5	Sales Executive	1	94	31	32.978723404255
6	Sales Representative	0	59	17	28.813559322033
7	Human Resources	0	39	7	17.948717948717
8	Laboratory Technician	0	197	31	15.736040609137
9	Manager	1	27	4	14.814814814814
10	Sales Executive	0	232	26	11.206896551724
11	Manufacturing Direc...	1	39	4	10.256410256410
12	Healthcare Represe...	0	94	7	7.446808510638
13	Research Scientist	0	195	14	7.179487179487
14	Manufacturing Direc...	0	106	6	5.660377358490
15	Healthcare Represe...	1	37	2	5.405405405405
16	Research Director	1	23	1	4.347826086956
17	Research Director	0	57	1	1.754385964912
18	Manager	0	75	1	1.3333333333333

Although there is no apparent correlation, we would like to note that the 5 highest attrition rates are for job roles that worked overtime.

(OverTime = 1 represents an employee who works overtime and OverTime = 0 represents one who does not)

We may need to amend our policies on overtime and be more lenient on deadlines for tasks given to our employees.

Job and Environment Satisfaction

Next, we will see how job and environment satisfaction correlates to overall attrition.

```
SELECT JobSatisfaction, EnvironmentSatisfaction,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
GROUP BY JobSatisfaction, EnvironmentSatisfaction
ORDER BY AttritionRate DESC;
```

Output:

	JobSatisfaction	EnvironmentSatisfaction	TotalEmployees	TotalAttrition	AttritionRate
1	1	1	53	20	37.735849056603
2	2	1	47	16	34.042553191489
3	1	2	55	14	25.454545454545
4	3	1	100	22	22.000000000000
5	1	4	87	16	18.390804597701
6	1	3	94	16	17.021276595744
7	3	4	130	22	16.923076923076
8	4	1	84	14	16.666666666666
9	3	2	84	13	15.476190476190
10	2	3	87	12	13.793103448275
11	2	2	59	8	13.559322033898
12	4	3	144	18	12.500000000000
13	3	3	128	16	12.500000000000
14	2	4	87	10	11.494252873563
15	4	2	89	8	8.988764044943
16	4	4	142	12	8.450704225352

The results show that the higher either the job and environment satisfaction was, the lower the attrition rate. The highest attrition rate corresponded to the lowest job and environment satisfaction and the lowest attrition rate corresponded to the highest job and environment satisfaction. This may be a cause for concern for our overall job environment.

Attrition by Promotion

Finally, we look at the correlation between attrition and promotion.

```
SELECT YearsSinceLastPromotion,
COUNT(EmployeeNumber) AS TotalEmployees,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) AS TotalAttrition,
COUNT(CASE WHEN Attrition = 1 THEN 1 END) * 100.0 / COUNT(EmployeeNumber) AS
AttritionRate
FROM Attrition_Table
```

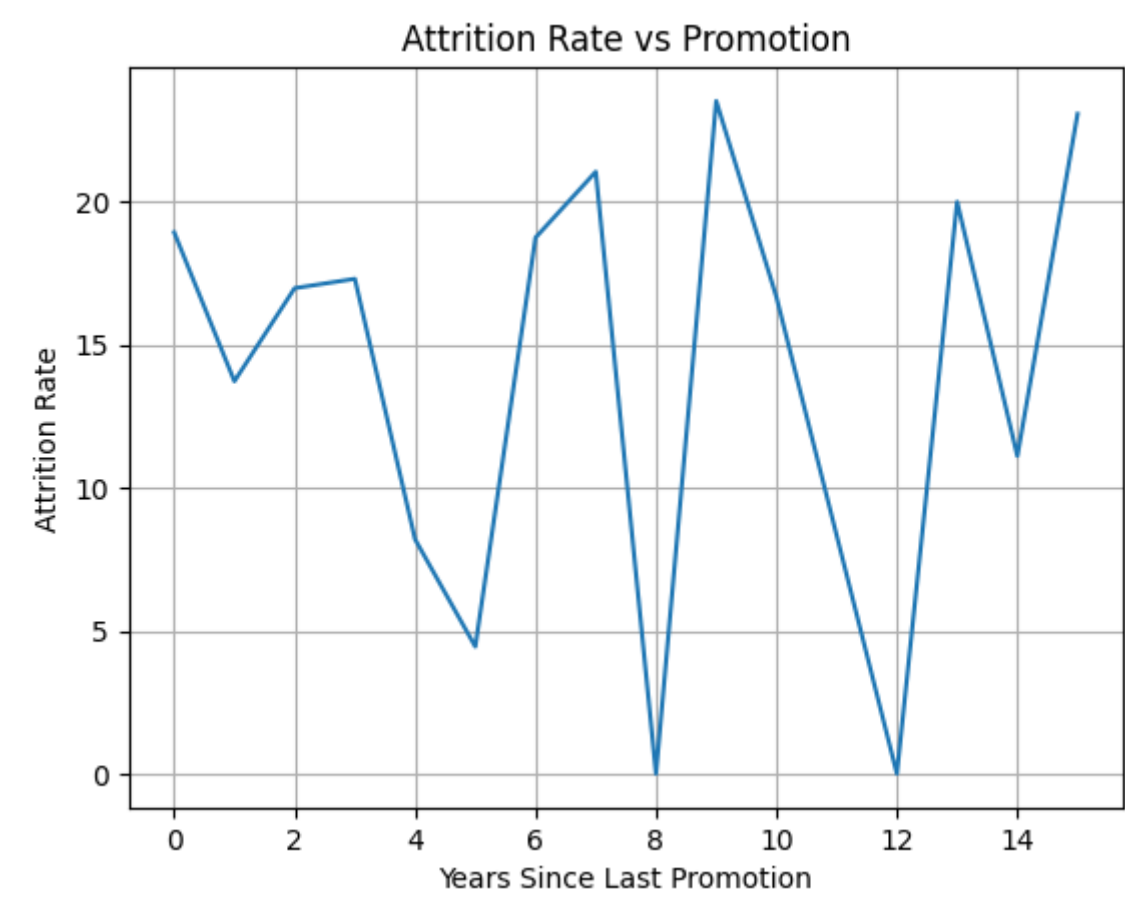


```
GROUP BY YearsSinceLastPromotion
ORDER BY YearsSinceLastPromotion;
```

Output:

	YearsSinceLastPromotion	TotalEmployees	TotalAttrition	AttritionRate
1	0	581	110	18.932874354561
2	1	357	49	13.725490196078
3	2	159	27	16.981132075471
4	3	52	9	17.307692307692
5	4	61	5	8.196721311475
6	5	45	2	4.4444444444444
7	6	32	6	18.750000000000
8	7	76	16	21.052631578947
9	8	18	0	0.000000000000
10	9	17	4	23.529411764705
11	10	6	1	16.666666666666
12	11	24	2	8.3333333333333
13	12	10	0	0.000000000000
14	13	10	2	20.000000000000
15	14	9	1	11.111111111111
16	15	13	3	23.076923076923

If we plot a line graph of Attrition Rate vs Years Since Last Promotion, it is evident that there is not correlation between the two.



Summary

In summary, we found that young adults, aged 18 - 30, were twice as likely to leave than those of any other age bracket. We also found that employees who work overtime were more likely to leave than ones who do not. Finally, employees who were dissatisfied with their job and the environment showed increase signs of leaving the organization.

Recommendations

- Implement better career growth opportunities/engagement programs in order to retain younger employees.
- Update office equipment, systems and spaces and foster a more positive and fun work environment to improve job and environment satisfaction.
- Implement flexible scheduling, analyze workload distribution and/or adjust staffing levels to decrease the amount overtime employees.

References

1. [Gartner](#)
2. [Spiceworks](#)