

HR People Data Analysis with a focus on Attrition

Status **Completed** ▾

Timing Oct 1, 2024 to Oct 2, 2024

Owners Jessica Gutstein

Overview

Employee attrition rates are exceeding desired levels within the company. This project seeks to understand the root causes of employee attrition and identify data-driven solutions to improve employee retention and enhance overall organization performance. This project uses Excel, SQL (in BigQuery), Power BI, and Python (in Google Colab) to analyze attrition rates in a public HR dataset.

Project Objectives

- Define main variables causing attrition.
- Study the correlation between employee data variables.
- Create data visualizations to understand attrition causes.
- Provide business insights to improve employee retention.

Strategy

Approach

Analyze public HR dataset to extract and create insights using Excel, SQL, PowerBi, and Python.

Target audience

- Executive stakeholders
- HR Leadership
- HR Business Partners

Measurement

- Correlation matrix
- Regression analysis
- Statistical summary
- Data visualization plotting

Project questions

What is the correlation between the variables and attrition?

What is the variable with the highest correlation with attrition?

What other variables have the most impact on attrition?

What are the recommended actions to improve retention?

Data

The data was obtained from a public [dataset in Kaggle](#), and contains information regarding employee demographics, attrition, and satisfaction rankings.

Results

What is the correlation between the variables and attrition?

In order to analyze the data, it was necessary to clean the dataset and manipulate the current variables from text to numerical values. From there, a correlation matrix was created in Python and Excel to compare variables to attrition. Please refer to the Python notebook (HR_People_Data.ipynb) and Excel Analysis in the Github (HR People Data.xlsx).

The correlations can be visualized below:

	<i>Correlation with Attrition</i>		
OverTime#	0.24611799	Department#	-0.063990596
TotalWorkingYears	-0.1710632	WorkLifeBalance	-0.063939047
JobLevel	-0.1691048	TrainingTimesLastYear	-0.059477799
YearsInCurrentRole	-0.160545	DailyRate	-0.056651992
MonthlyIncome	-0.1598396	RelationshipSatisfaction	-0.045872279
Age	-0.159205	NumCompaniesWorked	0.043493739
YearsWithCurrManager	-0.1561993	YearsSinceLastPromotion	-0.033018775
StockOptionLevel	-0.1371449	Education	-0.03137282
YearsAtCompany	-0.1343922	Gender #	-0.029453253
JobInvolvement	-0.130015957	JobRole#	-0.027930476
JobSatisfaction	-0.103481126	MaritalStatus#	-0.023803636
EnvironmentSatisfaction	-0.103368978	MonthlyRate	0.015170213
DistanceFromHome	0.077923583	PercentSalaryHike	-0.013478202
EducationField#	0.075437215	EmployeeNumber	-0.010577243
		HourlyRate	-0.00684555

What is the variable with the highest correlation with attrition?

The results highlighted the most correlated variables being:

1. Overtime
2. Total Working Years (Negative Correlation)
3. Job Level (Negative Correlation)
4. Years in Current Role (Negative Correlation)
5. Monthly Income (Negative Correlation)
6. Age (Negative Correlation)
7. Years with Current Manager (Negative Correlation)
8. Stock Option Level (Negative Correlation)
9. Years at Company (Negative Correlation)

Overtime is the variable with the highest correlation.

What other variables have the most impact on attrition?

The other variables that have an impact on attrition are related to age, job level, and years of experience. There is a negative correlation in these variables, indicating that early career professionals are more prone to leaving the company within a short period of time.

Further analysis of the data using SQL and Excel shows that employees from the lowest Job Level had the highest attrition rate at 26.34%.

Results from Attrition_joblevel.sql			
JobLevel	total_employees	attrition	AttritionRate
1	543	143	26.34
2	534	52	9.74
3	218	32	14.68
4	106	5	4.72
5	69	5	7.25

Additionally, workers who leave the company have an average of 8 years of experience, as opposed to 12 years of experience of those who stay in the company, indicating that early professionals have a higher rate of attrition.

Results from Average_TotalWorkingYears.sql				
Attrition	average_TotalWorkingYears	max_TotalWorkingYears	min_TotalWorkingYears	count_TotalWorkingYears
FALSE	12	38	0	1233
TRUE	8	40	0	237

Moreover, looking at age and salary distributions, the employees with lower salaries were showing more prone to leaving the company compared to employees who stayed. For example, employees in their 30s who left were making an average of \$4,794 monthly compared to those who stayed at an average of \$5,733.

Results from Attritionbyage&income.sql						
Attrition	Age	min_monthlyincome	25th_percentile	average_monthlyincome	75th_percentile	max_monthlyincome
Active	60s	5220	5405	10268	10883	19566
Non-Active	50s	2339	2683	8338.13	10650	19859
Active	50s	2066	5473	11379.35	17046	19999
Non-Active	40s	2018	2778	7045.65	9619	19545
Active	40s	2001	4385	8699.03	13496	19973
Active	30s	1129	3280	5733.66	6932	19431
Non-Active	30s	1081	2610	4794.27	6074	13610
Non-Active	20s	1009	2323	3186.48	3464	9854
Active	20s	1052	2532	4012.18	4741	16124
Non-Active	10s	1102	1569	1841.7	2121	2564
Active	10s	1051	1200	1772.14	2552	2994

The next focus area was Years in Current Role. Once again, employees who left the company stayed an average of 2.9 years, compared to employees who stayed at an average of 4.5 years.

Results by Average_yearsinrole.sql				
Attrition	max_yearsinrole	min_yearsinrole	avg_yearsinrole	count_yearsinrole
FALSE	18	0	4.48	1233
TRUE	15	0	2.9	237

Finally, employees who had received overtime had an attrition rate of 30% compared to employee who never received overtime at a attrition rate of 10%.

Results by Attrition_Overtime.sql		
OverTime	Attrition	Overtime_EmployeeCount
TRUE	TRUE	127
TRUE	FALSE	289
FALSE	FALSE	944
FALSE	TRUE	110

The following SQL queries were used to obtain the results above (refer to them in the Github project folder):

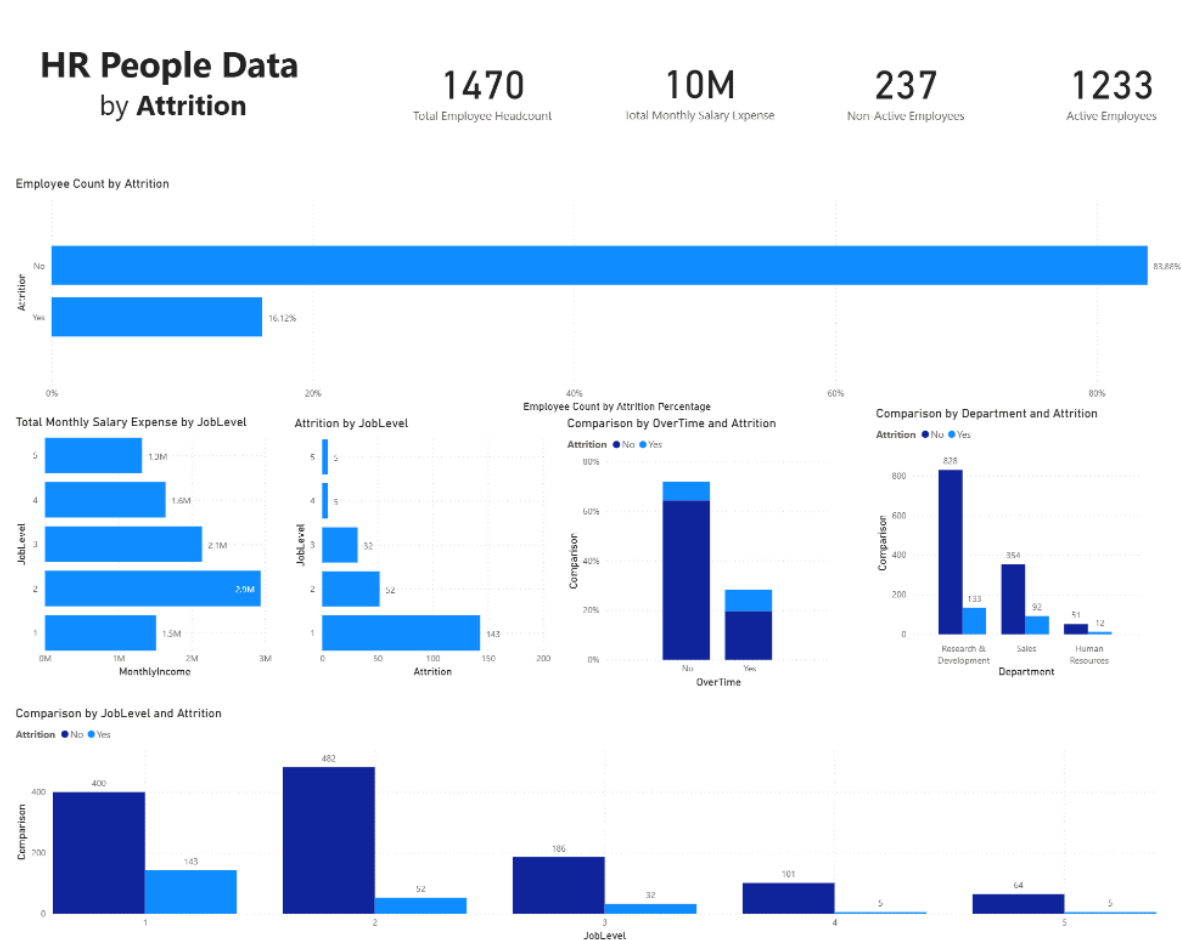
- Attrition_Overtime.sql
- Attrition_joblevel.sql
- Attritionbyage&income.sql
- Average_TotalWorkingYears.sql
- Average_yearsinrole.sql

What are the recommended actions to improve retention?

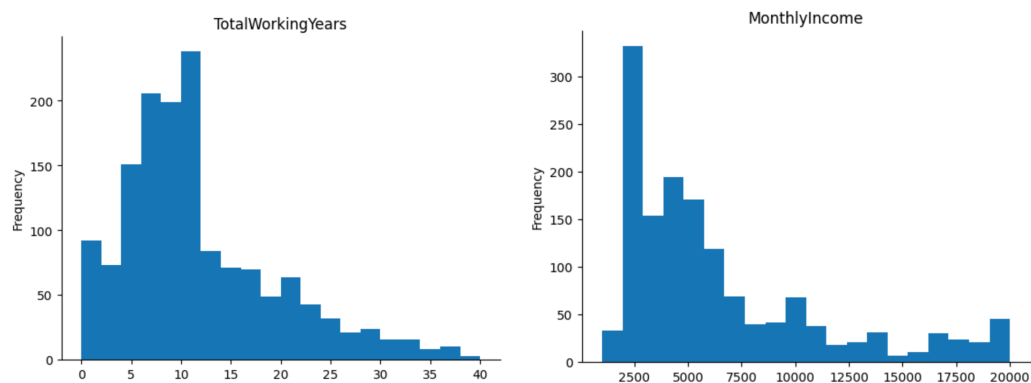
Based on the analysis, the recommended actions to improve retention would be to hone the focus on early career development, overtime limit, and employee incentives (such as training development, opportunities for growth, and higher stock/bonuses options based on job performance) who make the lower range of monthly income compared to average and above.

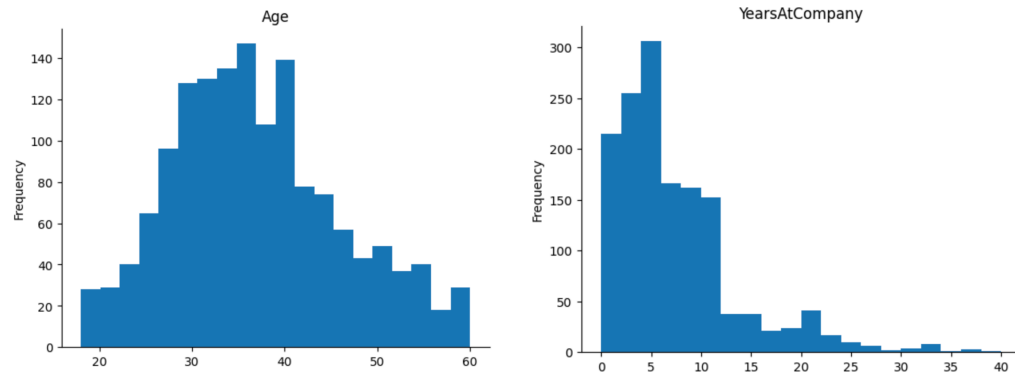
Data Visualization

The way to showcase the insights found and provide daily updates was creating a [PowerBi dashboard](#) on current attrition at the company. This way key stakeholders could have real time updates and watch to see if areas of focus are improving with the recommendations given.

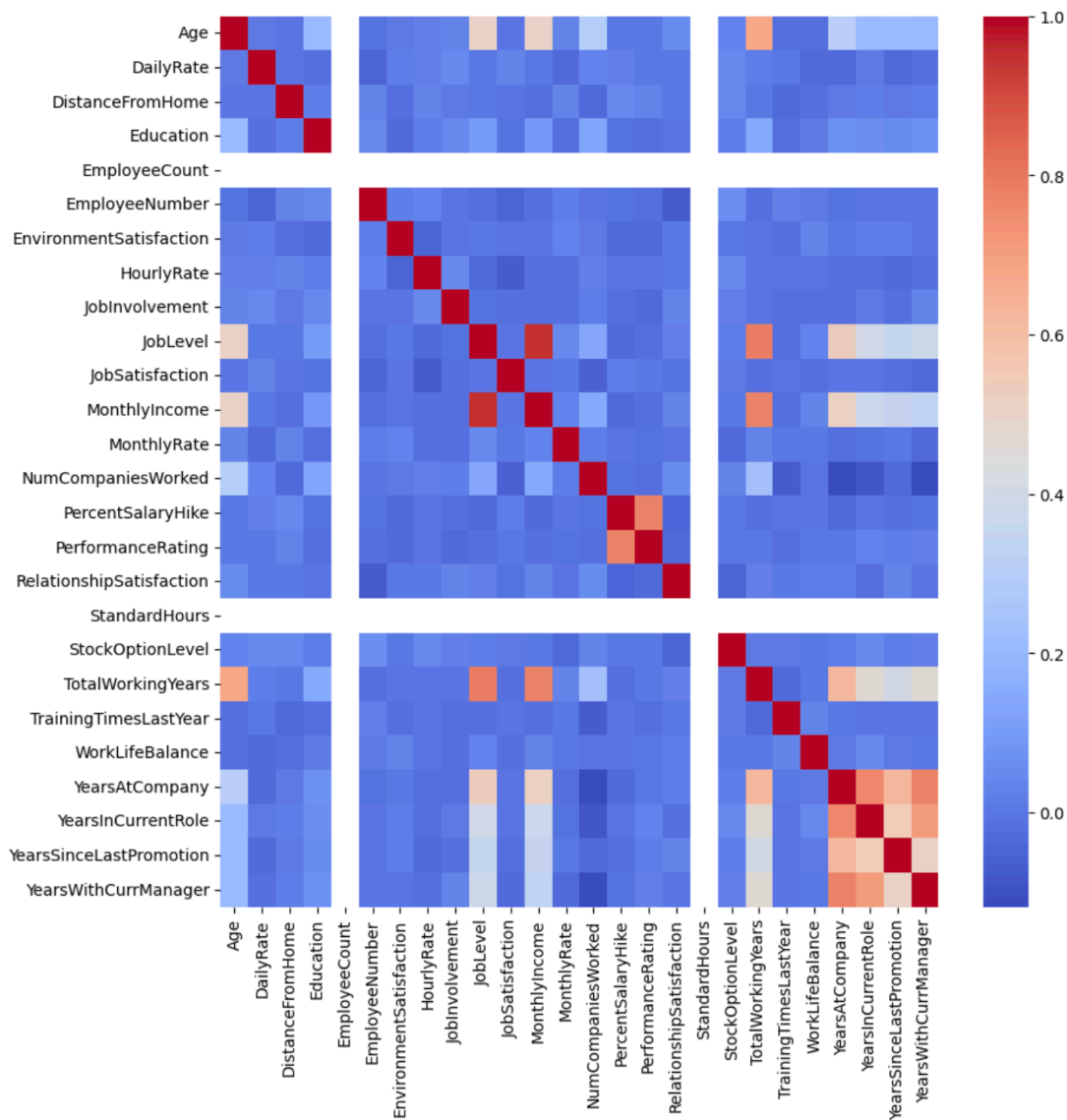


Moreover, Python was used to visualize the data further. The Matplotlib and Pandas libraries were used to describe the statistics of the data and create visualizations, including a correlation heatmap and histograms. For example, the following histograms show the distribution of employees across Total Working Years, Monthly Income, Years at Company, and Age.





The following heatmap was created using the Pandas and Seaborn libraries in Python, and shows the correlation between all the variables in the dataset.



Conclusion

As indicated in the findings, this organization should prioritize enhancing retention among early-career professionals through a series of incentives designed to mitigate the likelihood of their departure, such as training development, opportunities for growth, and higher stock/bonuses options based on job performance. Furthermore, the company should concentrate on limiting overtime, as this factor exhibits the strongest correlation with attrition.

Resources

- Excel Analysis: [HR People Data](#)
- SQL Queries:
 - [Attrition_Overtime.sql](#)
 - [Attrition_joblevel.sql](#)
 - [Attritionbyage&income.sql](#)
 - [Average_TotalWorkingYears.sql](#)
 - [Average_yearsinrole.sql](#)
- Power BI Dashboard: [People Dashboard](#) and [pdf version](#)
- Python Colab: [HR People Data Notebook](#)