**Employee Retention and Churn Analysis: Identifying Key Drivers of Tenure**

**Summary**: Analyze employee data to identify the factors most strongly associated with long tenure, using correlation analysis to reveal key drivers of retention and churn.

Understanding patterns in employee tenure can help a company develop a stable workforce and increase their productivity while minimizing costs spent on training and onboarding.

**Data Overview**:

I used a dataset that was generated by chatgpt that I also manually edited for my employee data in this project. The dataset included the hiring and salary information of 45 employees spanning from 2018-2024.

Basic data cleaning and formatting tasks were performed, including:

Some of the key columns were their department, hiring date, last working

* Handling blank cells
* Trimming spaces around text
* Formatting salaries as currency with comma separators
* Using data validation to ensure dates were realistic (hiring dates before start dates)
* Applying conditional formatting to highlight active employees in green (where the last working day was blank)

**Calculated metrics**:

-Tenure – the difference between their start and end working dates

-Retention Rate – the percent of employees at the start that are still employed

-Churn – the percent of employees that stopped working at the company

-Average onboarding, performance score, salary

-Correlation coefficient – indicates strength of relationship between two variables

- Tenure prediction formula:

**Analysis Methods:**

I calculated the correlation coefficient for the tenure and the variables I thought would have the greatest impact on tenure which included the onboarding duration, training period, salary, and performance score. The correlation coefficients between tenure and salary and performance score were the highest, with values of .75 and .5 respectively. I then used linear regression to quantify the effect that changes in an employee’s salary and performance score have on their tenure with the company. The tenure was the dependent variable and salary/performance score were the independent variables. The R squared value showed .57, which indicated 57% of the variance in an employee’s tenure could be explained by their salary and performance score. The coefficients for the salary and performance scores were 0.000005 and 0.016. This means that a 1 unit increase in salary would increase an employee’s tenure by 0.000005 years (4.3 hours) and a 1 unit increase in performance score would increase an employee’s tenure by 0.016 years (5.8 days). Of all the variables in my dataset, I found that performance score had the biggest impact on how long an employee worked at the company.

Next, I used the What If analysis tool in Excel to predict how much an employee’s tenure could be lengthened by varying their salary and performance score. I used the coefficients and intercept from the regression output in the formula to calculate an employee’s hypothetical tenure in years: **-**1.14 + 0.000005\*Salary + 0.016\*Performance\_Score. I created a table with increasing values for the performance score and salary that were greater than/equal to what the employees with the longest tenure had. The results were not what I expected, because it predicted a tenure of only .975 years with a salary and performance score that were higher than what was observed in the dataset. I think the predicted values for the tenure were inaccurate due to the large scale of the salary value compared to the other values in the data set. The salary was in hundred thousands, while the performance score values were less than 100.

**Top 3 Employees with the Longest Tenure:**

Sorting the dataset by tenure revealed:

* All three worked for **over 6 years**
* All were in the **IT department**
* Average salary: **$112,000**
* Average performance score: **82**
* Average onboarding: **43 days**

This suggests that employees in IT with salaries over $100,000 and high performance scores were most the likely to remain at the company.

**Tables and Visualizations**

Regression Calculation output

A close-up of a document

AI-generated content may be incorrect.

Scenario Analysis: Impact of Performance and Salary on Employee Tenure

A green and black text

AI-generated content may be incorrect.

Regression Scatter PlotsA graph of performance scores

AI-generated content may be incorrect.

Retention Overview: Employees Remaining vs. Employees DepartedA screenshot of a computer

AI-generated content may be incorrect.

Employee Metrics by Department: Compensation, Performance, Training, and StatusA screenshot of a calendar

AI-generated content may be incorrect.Annual Retention vs. Churn SummaryA screenshot of a computer

AI-generated content may be incorrect.

Key Characteristics of Top 3 Employees with Longest Tenure

A screenshot of a report

AI-generated content may be incorrect.

Multi-Line Chart: Trends in Retention Rates, Performance Scores, and Salary

A graph with red and blue lines

AI-generated content may be incorrect.

Employees by Department: Active and Departed

A screenshot of a graph

AI-generated content may be incorrect.

**Key Insights**:

- Declines in retention rate from 2018–2023 were accompanied by decreases in performance scores and salaries.

-The IT department had the highest average salary and the longest onboarding period.

-The IT department also experienced the highest number of departures but included employees with the longest tenure.

-Salary values should be scaled before using them in regression alongside performance scores to avoid inaccurate tenure predictions.

-Performance scores had a stronger impact on tenure than salary (by 0.25). This suggests that positive feedback and recognition from management may play a bigger role than salary in an employee’s decision to stay with a company.

**Conclusion**:

While the insights that can be derived from this study are somewhat limited due to the small and fictitious dataset, the analysis suggests that both increased salaries and positive managerial feedback may influence employee tenure. Salary alone does not appear to be the strongest driver of retention. Further research with a larger, real-world dataset could provide more definitive conclusions and better quantify the relative impact of compensation and management support on employee longevity