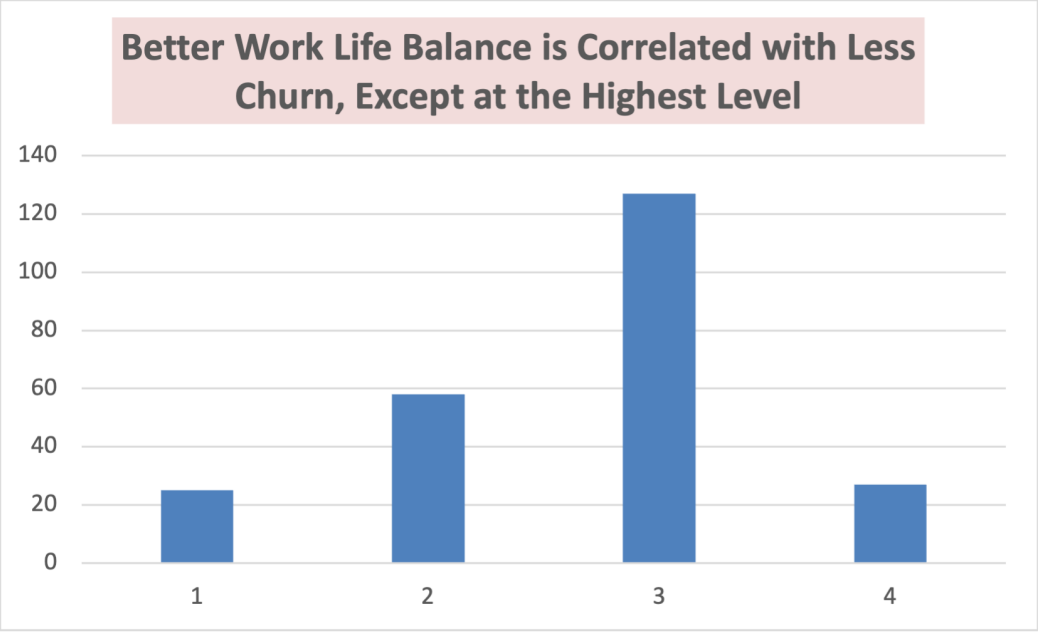
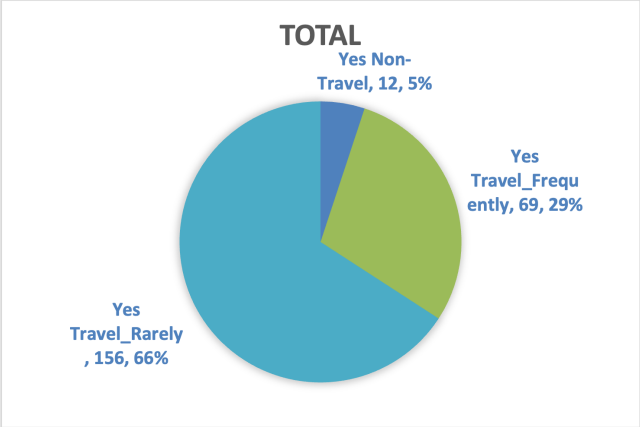
Analysis of IBM Employee Attrition

The goal of this project is to understand what factors contribute to an employee's decision to leave IBM, how those factors relate to each other, and if we can use that information to make a reasonable prediction (both via an initial analysis of our data and by using a logistic regression model) on whether or not an employee is likely to leave the organization. The goal of this discovery is to givre IBM insights on which employees to allocate incentives to in order to reduce employee attrition.

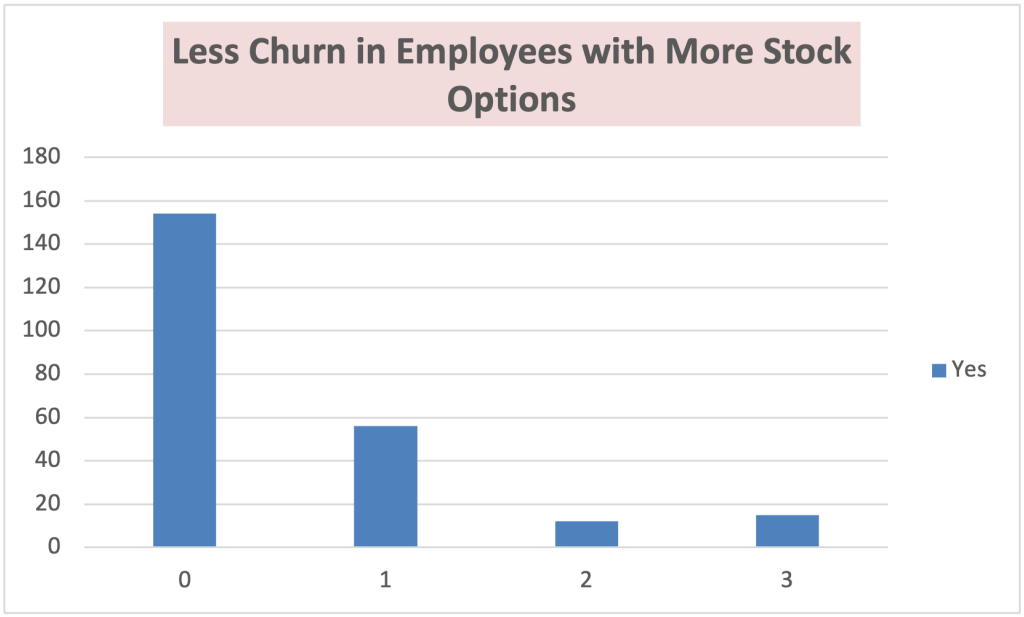
Below is a chart depicting the relationship between work-life balance and attrition. The scale from 1-4 is defined as a ratio of work hours to non-work hours. As that ratio gets larger (i.e, more working hours, less non-working hours), a larger amount of employees are shown to leave the company (churn). We can see that there's a discrepancy at the largest scale level because attrition substantially decreases there. Further analysis is needed to understand the impact of confounding factors.



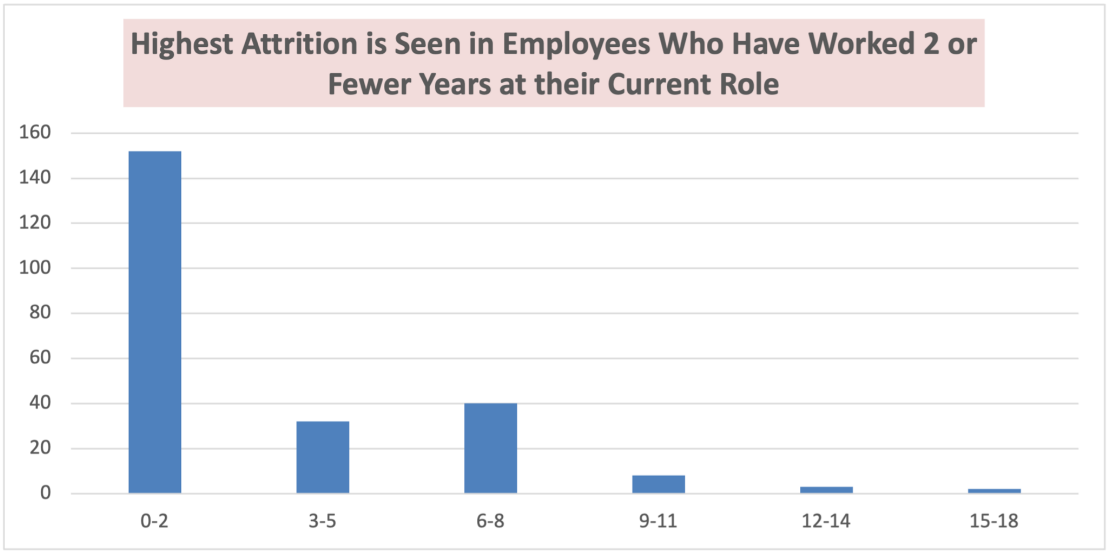
Next, we have a pie chart on the relationship between business travel and attrition. The lowest attrition occurs with employees who do not have business travel as a part of their role's responsibilities. Alternatively, the most attrition occurs when employees are required to travel for work, albeit "rarely". Finally the group in the middle is made of employees who travel frequently for work. My initial hypothesis is that business travel influences attrition most when it occurs inconsistently in an employee's schedule. Perhaps the employees who travel frequently or not at all are used to a consistent schedule, whereas those who travel rarely for work may not have that privilege.



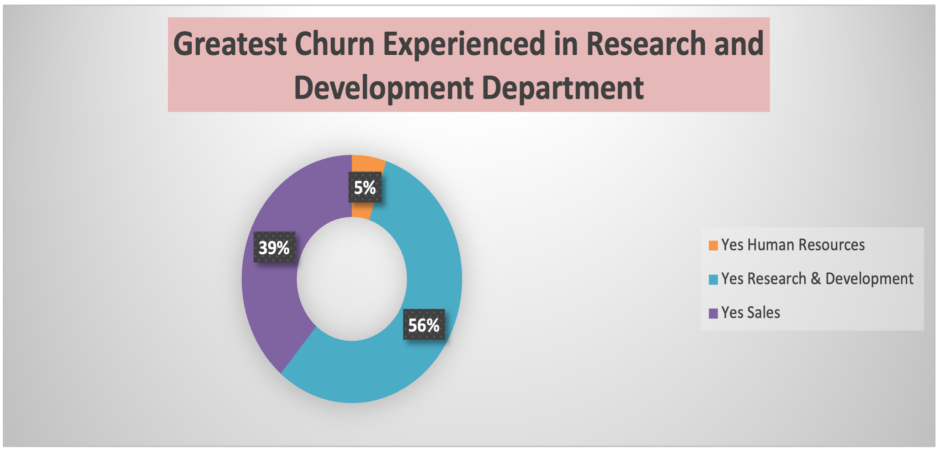
This next bar graph shows a relatively inverse relationship between employee stock options and attractions. We see that as the number of stock options increases, the number of churned employees decreases.



In the next bar graph, we see the relationship between the number of years an employee served in a job role and attrition. Here we see a somewhat inverse relationship, where the count of employees to churn is highest under the 3-year mark, and lowest under the 15-18 year mark.



Finally, we have a pie chart on the relationship between churn and department. We see the greatest percentage of churn in the Research and Development department, followed by Sales, then Human Resources.



As we can see from the above charts, many of our features present ambiguous results because of the influence of confounding factors. Further inspection of the data also suggests that some factors influence others' relationship with attrition. For that reason, I'd like to build a classification model to demystify the correlation of each factor with attrition, as well as to give us a way to quantify our prediction on which employees will leave the company.