Attrition Investigation Report

Predictive Modeling & Recommendations for Bain and Company

Problem Statement & Objectives

- **Problem:** Bain and Company is facing an unprecedented amount of employee attrition/churn rates.
- **Objectives:**
  - Analyze company data to gain insights into potential causes.
  - Use data to build a model to accurately predict attrition.
  - Make recommendations on what steps can be taken to lower attrition.

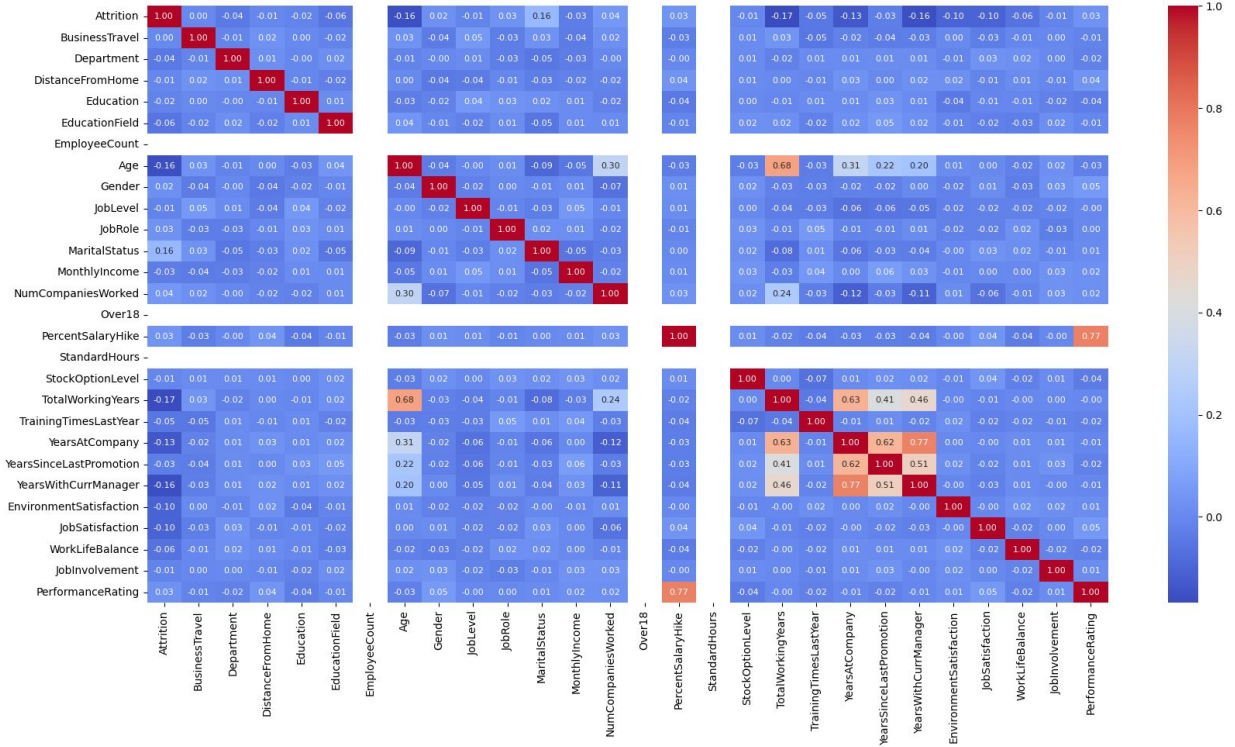
Methodology - Data Analysis

- Combine the general employee data with the surveys to make one dataset to work with.
- Examine relationship between each data field with attrition, and with each other data field.
- Use that insight to decide what fields we want our prediction model to learn from.

Data Analysis Insights

Correlation Matrix

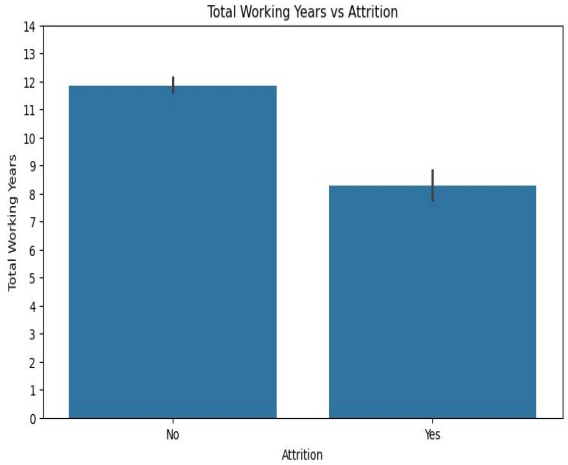
- Employee count, over 18, and standard hours are fields that have no correlation, we can omit them from model training
- Fields describing seniority (age, years at company) are highly correlated with each other , so we can limit to using one to simplify training
- Total working years and monthly income have the strongest inverse correlation to attrition

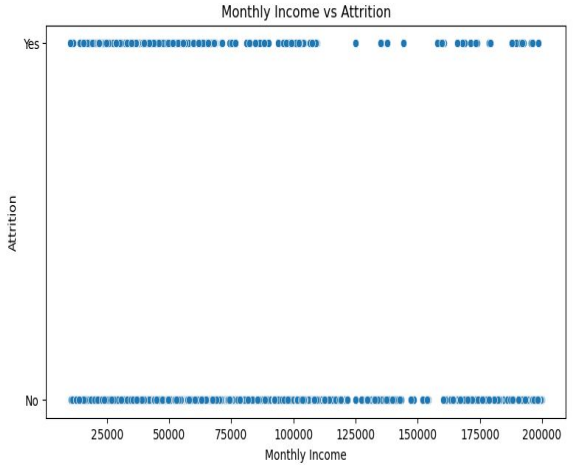


Data Analysis Insights

Highly Correlated Fields

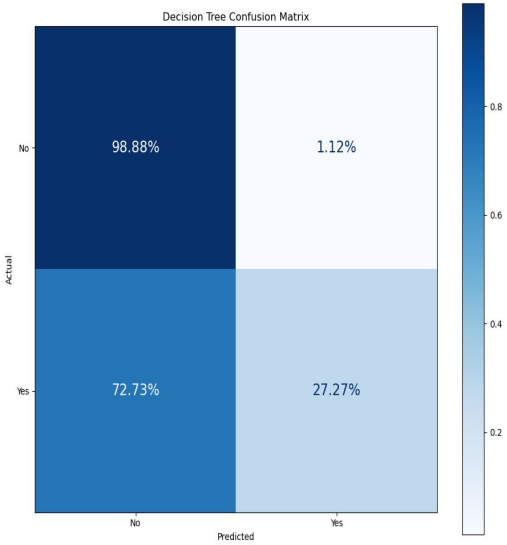
- Total working years and monthly income had the highest correlation with attrition (This is disregarding the other fields related to age and seniority)
- As employees age, the bar graph shows less attrition. Similarly, the scatter plot shows attrition thinning out with higher salaries
- This gives us a hint that they may be strong influencers to our models
- We took the liberty of also removing marital status from training as to avoid discriminatory biases in our model

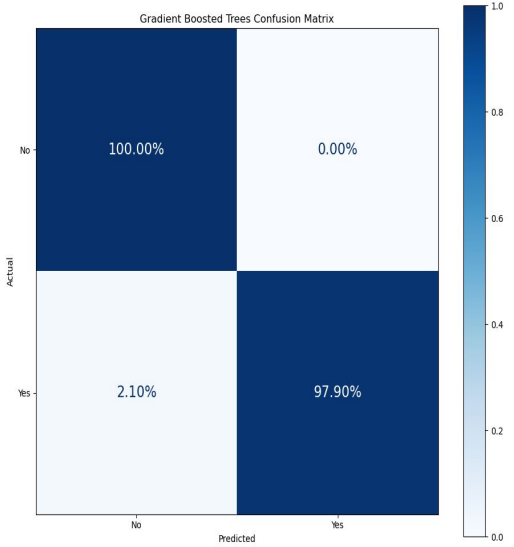




Methodology - Predictive Modeling

- For our predictive model we evaluated between a decision tree, and gradient boosted trees(a decision tree ensemble).
- Decision trees are very useful for interpreting how the model makes predictions, which gives us insights into what impacts attrition the most.
- We will use these metrics to evaluate model performance:
  - **Accuracy:** The model's overall ability to predict whether an employee left the company.
  - **Precision:** The model's ability to avoid false positives while predicting attrition.
  - **Recall:** The model's ability to avoid false negatives, or to correctly predict all instances of employees leaving.
- While accuracy and precision are important, we believe a high recall score is most useful to Bain & Co.



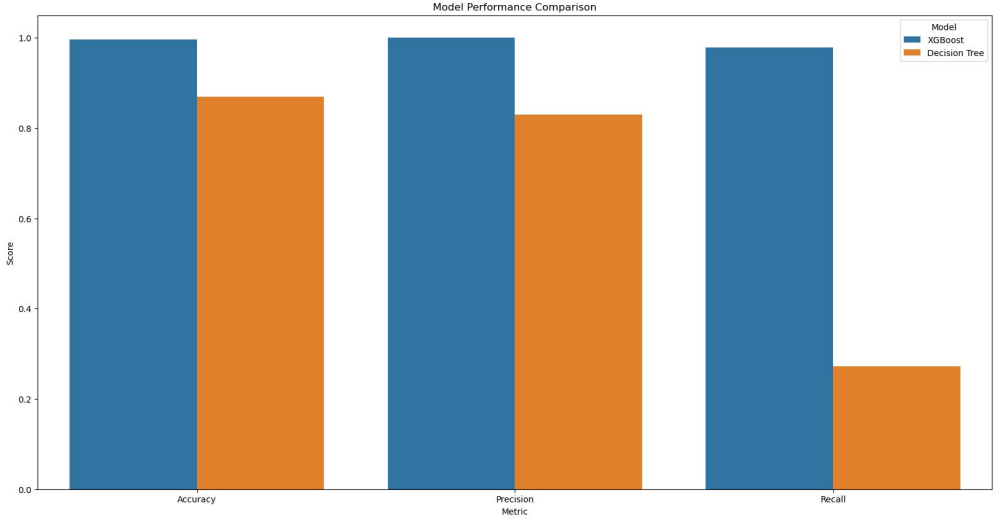


Confusion Matrix

- Both models did exceptionally well at predicting non-attrition
- Both models were also exceptional at not falsely predicting attrition
- The decision tree falsely predicted non-attrition at a very high rate, while the gradient boosted model rarely did
- The decision tree failed to predict attrition at a very high rate, the gradient boosted model rarely failed to do so

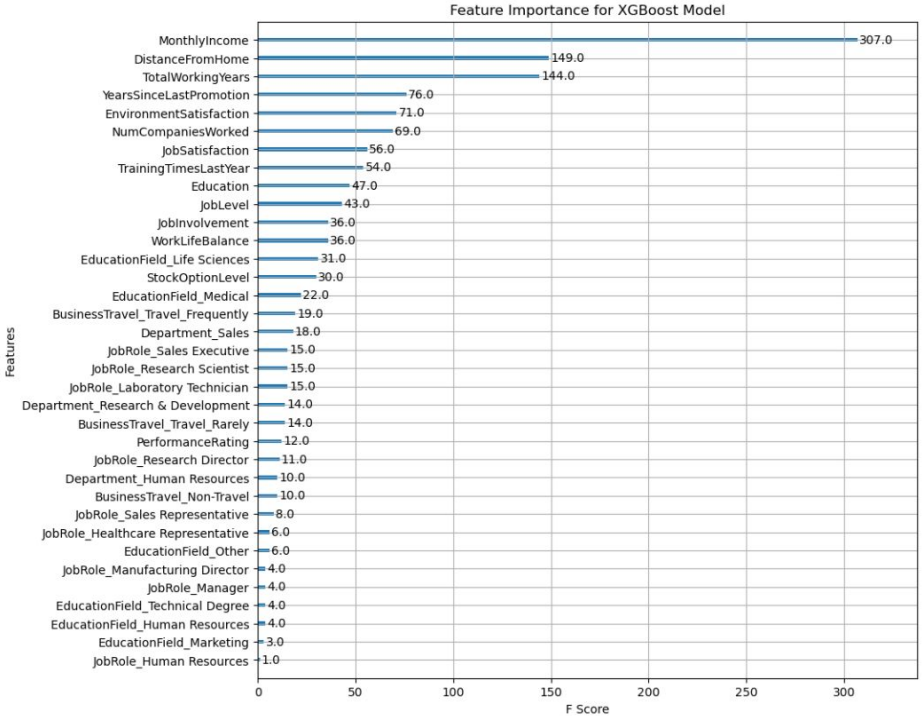
Model Performance

- For our metrics of accuracy, precision and recall, the gradient boosted model outperformed the single decision tree model every time.
- The decision to use the gradient boosted model for our insights was straightforward



Predictive Model Insights

- The model found monthly income to be the most important piece of information for predicting attrition.
- The model also found distance from home and total working years to both be extremely important at a near equal level



Conclusions & Recommendations

Practical Approaches to Address High Attrition

- Distance From Home
  - Remote and/or hybrid work options
  - Reimbursement for gas/public transportation/child care (lessen commute burden)
- Total Working Years
  - Mentorship program pairing newer employees with seniors
  - Incentives/rewards for company loyalty
- Using The Predictive Model
  - Model can be used to give managers insight into colleagues that might need more of their attention