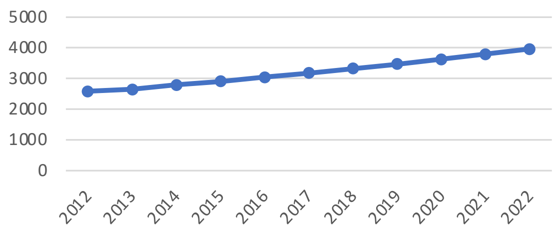
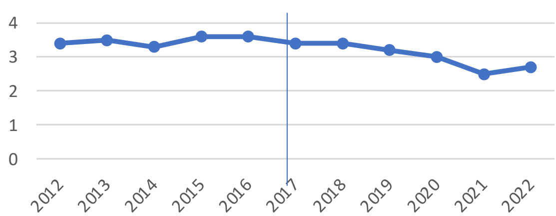
Learning How to Retain Talent at Acme Aromas

Stefan Bund, Investigator, bund@umich.edu

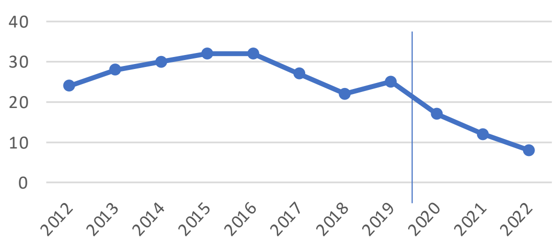
2012 –'22: Number of Employees



Job Satisfaction Declines



Job Applications Per Opening



While the Acme Workforce is Expanding…

Acme Aromas' workforce has grown 50% since 2012. In 2020, job satisfaction reached a new low, where it began falling precipitously. The internal culture at Acme is worth studying.

Acme has installed data science capabilities, enabling the situation to be understood. Factors inside the company, however, operate in tandem with conditions outside.

Worker Happiness is Receding…

By 2016, signs of external forces began to exert an influence, and declining job application rates began to occur.

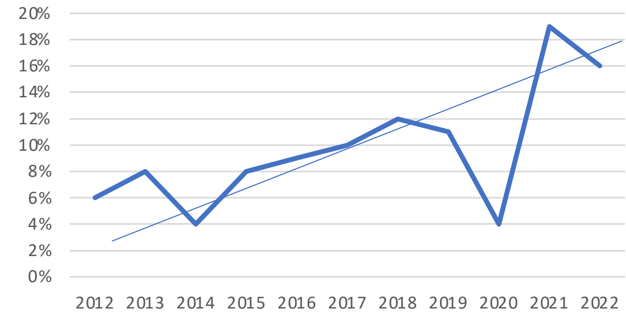
The monies required to attract talent since 2020 have skyrocketed, pointing to very high rates of employment in the Indian manufacturing sector. In 2020 it became necessary to spend up to 15,000R to acquire a new hire; this sum doubled in 2021.

And Applicants are Disappearing

The phenomenon at Acme is that a business is expanding, offering more opportunity to more employees.

Growth of business, however, has happened as employee happiness has declined. Also, as business volume has increased, the departure of employees has surged, while incoming new hiring has grown vastly more challenging.

Employee Turnover



How Can Acme Intervene, to Retain Talent?

The growth in Acme's human capital has potentially mirrored the Indian skilled manufacturing sector. With this trend continuing outside Acme's gates, there exists the opportunity to invest in the analytic tools needed to diagnose Acme's culture, within.

Acme has what is needed to diagnose workplace conditions which precede worker departure.

What is necessary, at this time, is to develop models to understand how and why Acme employees might leave an otherwise successful business.

A continued investment in root cause analysis will generate models, to improve employee retention, essential to continued growth, at Acme.

Our Approach
How We'll Analyze Attrition at Acme Aromas

Our Goals: What We Can Achieve, In the Study

Learn the Causes of Attrition
Leavers have left a trail of clues, before they left. We can study this cluster of clues and rank the most-likely signs of attrition, based on these cases.

Predict Attrition
Before it happens using a logistical regression approach. We will locate the odds of attrition by locating the factors that contribute to attrition, most likely.

Propose a Course of Action
Once the major associated causes of attrition are found, we will recommend a course of action, where management will use a toolbox to remediate those causes.

Using Data From Real Cases at Acme

Get Relevant Data
Human Resource system data will be used in studying employees where attrition took place.

Use a Wide Set of Possible Factors
29 major factors could contribute to attrition. It's up to us to better understand how these factor have caused attrition in the past.

We'll possess a host of variables for thousands of employees, such as:

- *Age*
- *Time in position*
- *Time under current manager*
- *Time since last raise*

Test for Errors, in Our Predictions
Errors within our model are detected using measures of recall and accuracy. We will combine these measures in order to prevent:
Type 1 errors, where we incorrectly detect attrition
Type 2 errors, where we fail to detect that we correctly predicted attrition

Discovering Why Attrition Happens at Acme

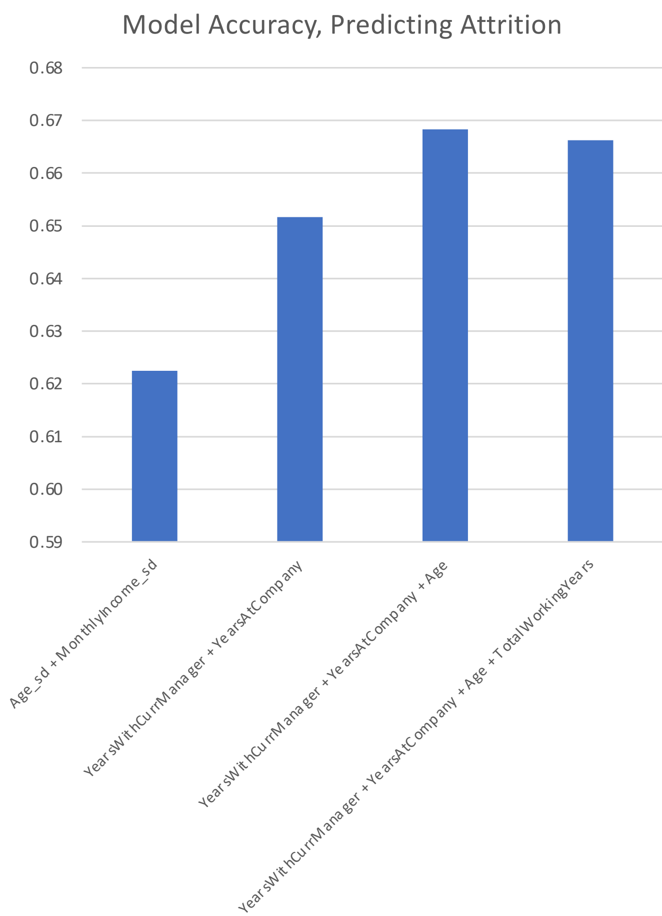
Explore the Data
We use correlation matrixes to explore data, revealing highly probable root causes.

Identify Causes
The odds of causal roots of attrition are converted to a probability, where we can identify the most likely causes. This logistic regression is a cornerstone to our approach, where we seek to categorize key factors leading to the problem. Once highly-likely causes are identified, we can identify attrition-likely employees with high probability. Using high probability metrics is the cornerstone to avoiding errors in prediction, such false positives.

Develop a Model
Key metrics are appearing, which suggest that attrition becomes more likely, as certain variables come into focus.

These measures of accuracy and recall are combined to form a confusion matrix, where we will compare the predictive power of different models. Additionally, the averaging of model recall and accuracy form an F1 metric, also useful in comparing models.

The Model for Attrition: Key Variables



Combinations of High Probability Causes Creates Better Models at Acme

When we combine the most likely sources of Attrition , we get increasingly high likelihoods of Attrition.

The best model combined the attributes which were most likely associated with Attrition. As we add more highly likely causes for Attrition, the general accuracy increases.

Generally, the causes of Attrition increase as:
- **They work under the same manager**
- **Years at the company increases**
- **The employee's age increases**
- **Total time at Acme increases**

In essence, the stagnation of an employee's career predicts attrition.

dim1	dim2	corr
YearsWithCurrManager	YearsAtCompany	0.769212
Age	TotalWorkingYears	0.680419
TotalWorkingYears	YearsAtCompany	0.626876
YearsAtCompany	YearsSinceLastPromotion	0.618409
YearsWithCurrManager	YearsSinceLastPromotion	0.510224

Problem	Approach	Insight 2	Recommend	Limitations

Model Performance: Measures of Predictive Capacity

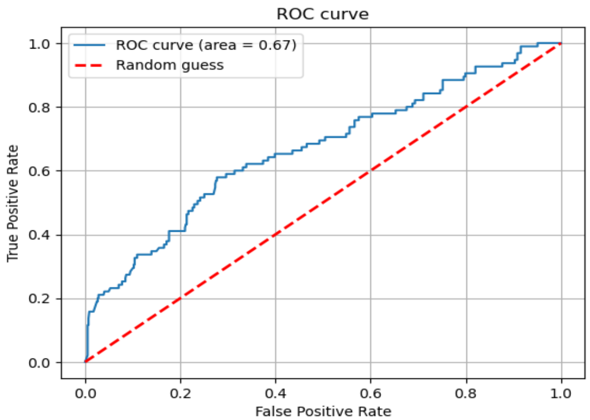
Measuring How Frequently the Model Misses: Measuring the Area Under the Curve

In comparing measures of our model's predictive capabilities, we measured how well it predicted actual cases of attrition, versus its errors. These errors are typically called Type I and Type II errors, and they relate to our ability to correctly identify attrition cases.

In the case of the model, we correctly identify cases of Attrition 67% of the time. This updated model offers an improvement of 5 percentage points over the prior model, which arrived at predictions at a 62% rate.

The ROC statistic measures the ratio of True Positives to False Positives, within the general body of Acme employees, where attrition took place.

Thus, given other models we tried, this model arrives at the best predictive capability.



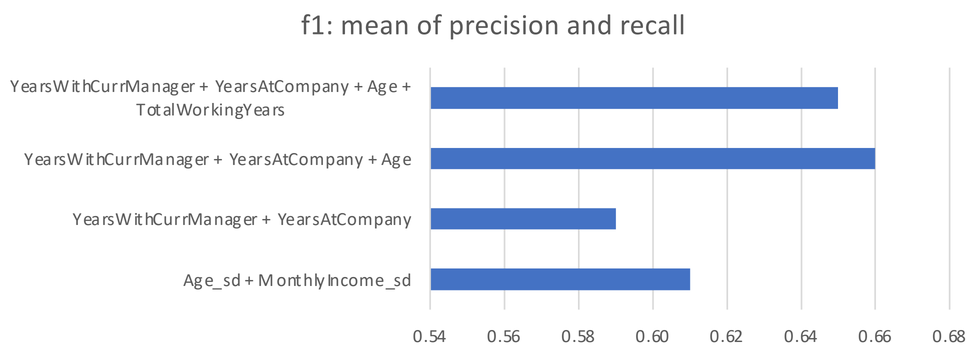
Choosing Models Based on the Number of Correct Predictions

When considering a model, the accuracy of the model reflects how correctly our model detects actual instances of attrition.

Models also possess recall, which measure the model's ability to detect actual causes of attrition out of all predictions, including misses.

Our model was assessed using an F1 score, which averages the recall and accuracy. This model was chosen over others due to its relatively high accuracy and recall rate, while maintaining the highest possible Area Under the Curve (AUC) score.

We thus addressed issues of errors, accuracy and recall within one model, settling on variables which offered the most comprehensive survey of employees, and why they would leave the firm.



Problem | Approach | **Insight 3** | Recommend | Limitations

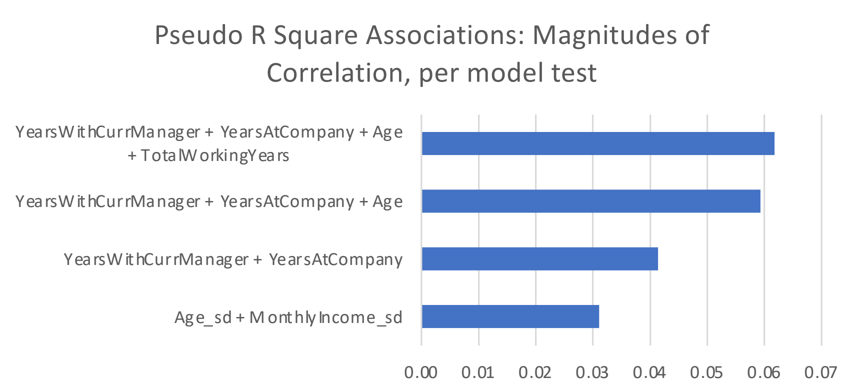
Why Attrition Takes Place At Acme: a Root Association Analysis

Identifying, and Testing Highly Predictive Variables

After preparing data for the analysis process, the variables with the highest correlation, or linear relationship, to attrition were identified. **Correlation translates into the probability that these variables contribute to attrition**.

Accepting these as qualifiers for a highly inclusive model to explain the factors associated with attrition, they were factored into various models.

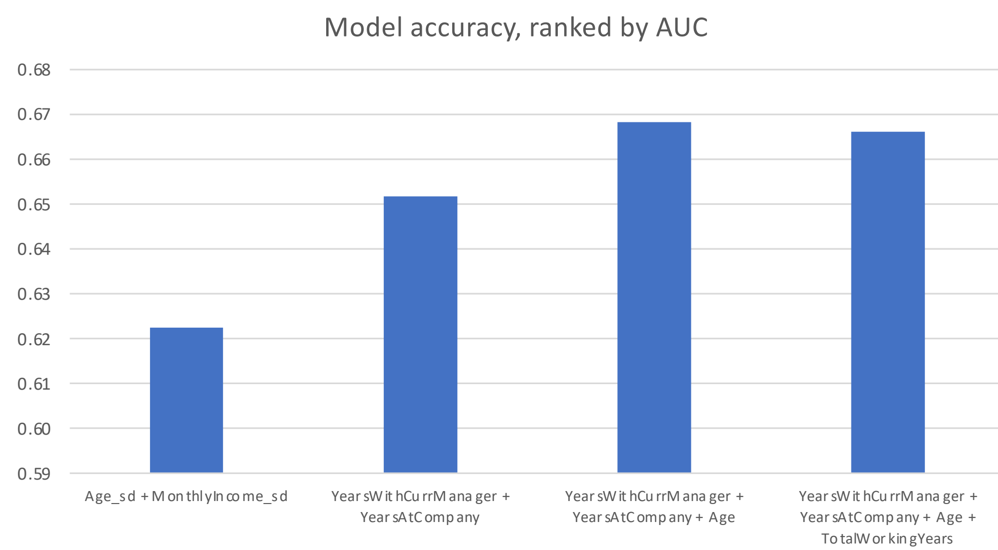
We tried various combinations of these highly correlated variables and determined their odds of correctly predicting attrition.



After ranking these highly predictive models, we identified cases where the model delivered real cases where attrition was likely. These actual cases helped to define where combinations of variables contributed to matches, within the entire employee body.

The variables with the greatest predictive capacity expressed a magnitude of likelihood, given their odds ratios.

The most likely root causes to attrition were tested for the odds that they contributed to attrition. The odds that each were associated were tested



Taking Action to Solve Attrition: Recommendations for Acme

Combatting the Root Causes of Attrition: Stagnation

In our testing, the most plausible causes of attrition were identified. These variables paint a picture of an employee who has received training but has ceased to move forward. This pattern of career stagnation appears most frequently, in relation to attrition.

Thus, we should start the conversation regarding career stagnation, and shift the focus toward how to remediate that employee experience. If stagnation is the primary cause for attrition, how can Acme act?

A plausible intervention to counteract that experience is to offer training. More income, more appreciation or other actions may not possess the same counteractive impact. Also, these employees likely need coaching in order to realize their potential or gain more value to the firm.

```
IMPACT MODEL 1: ADDRESS ATTRITION CAUSES WITHIN THE TOP 75TH PERCENTILE
---------------------------------------------------------------------
Years With Current Manager        TOTAL COUNT      PCT
                                  813              18.435374149659864
Years At Company                  TOTAL COUNT      PCT
                                  1344             30.476190476190478
Age                               TOTAL COUNT      PCT
                                  1041             23.605442176870746
Total Working Years               TOTAL COUNT      PCT
                                  1066             24.17233560090703
---------------------------------------------------------------------
Acme Employee Size, Q4 2022       4410
# highest risk of attrition       24  %
size of impact, population        1058
cost of remediation, Q1 2023      $ 7,935,000.00
```

Solving Stagnation with Recommended Interventions

Given the model features, a search was done to identify employees who fit the mold for career stagnation. We determined the quartile value for each variable, then identified employees who occupied the top 25% of each variable. Thus, the Acme employees who possessed the highest scores within these four variables were identified as opportune candidates for intervention.

The percentage of employees who both fit the model and fell within the top percentiles of the distribution met thresholds for intervention.

The picture of how Acme can solve the attrition crisis is falling into place. We now know:

- Who is likely to leave Acme
- How they can be helped

Approximately 24% of Acme employees fit the model, are at risk for attrition, and would likely benefit from an intervention, such as Training.

The size of the group meeting the model parameters is 1,058.

If the training unit is priced at 7,500 IR, a total cost of IR 7,935 M is recommended.

Limitations to Our Data Science: How we manage uncertainty in our Study

Prediction Problem 1: Does the Past Resemble the Future?

We assume that the causes of attrition in the future resemble those in the past. Changing conditions in the economy can present new reasons for employees to seek employment elsewhere.

We can only discover the reasons why employees left in the past. Thus, if new economic conditions take place outside the firm, new causes of attrition may appear.

Additionally, the breadth of the employee data is important. In cases where data is not present, we will need to set data to minimums, thus limiting our ability to model data well.

Prediction Problem 2: False Positives

Our ability to predict future event hinges on our ability to recognize the leading reasons why attrition takes place.

Factoring in false positives and false negatives is part of the process. We avoid incorrectly identifying false cases of likely attrition by not allowing certain variables to imbalance out model.

At this time our analysis depends on the ratio of Type 1 and Type 2 errors, in a process known as Reporter Operator Curve, or Area Under the Curve.

False and True positives are compared as a statistic, in gauging the best model for the analysis.

Prediction Problem 3: Comparing Models

In order to arrive at the best possible model for prediction, various combinations of variables are used. These variables are chosen based on their correlation to attrition which is a linear relationship. Since other linear methods could be used later, different correlated variables, as causes of attrition, may be identified later. Thus, our analysis of the causal variables of attrition are best effort, given the current HRIS data. The model should be rerun later, when more data is available, and potentially with a different linear method, to discover causal variables.

Given the variables which corelate to attrition, we construct models with the greatest odds of predicting attrition. One limit to the study is the fact that we must choose one model, among many, and compare these models given the Area Under the Curve method (AUC/ROC). The AUC technique must use included variables in the same order, in order to model the probabilities of each variable appropriately.