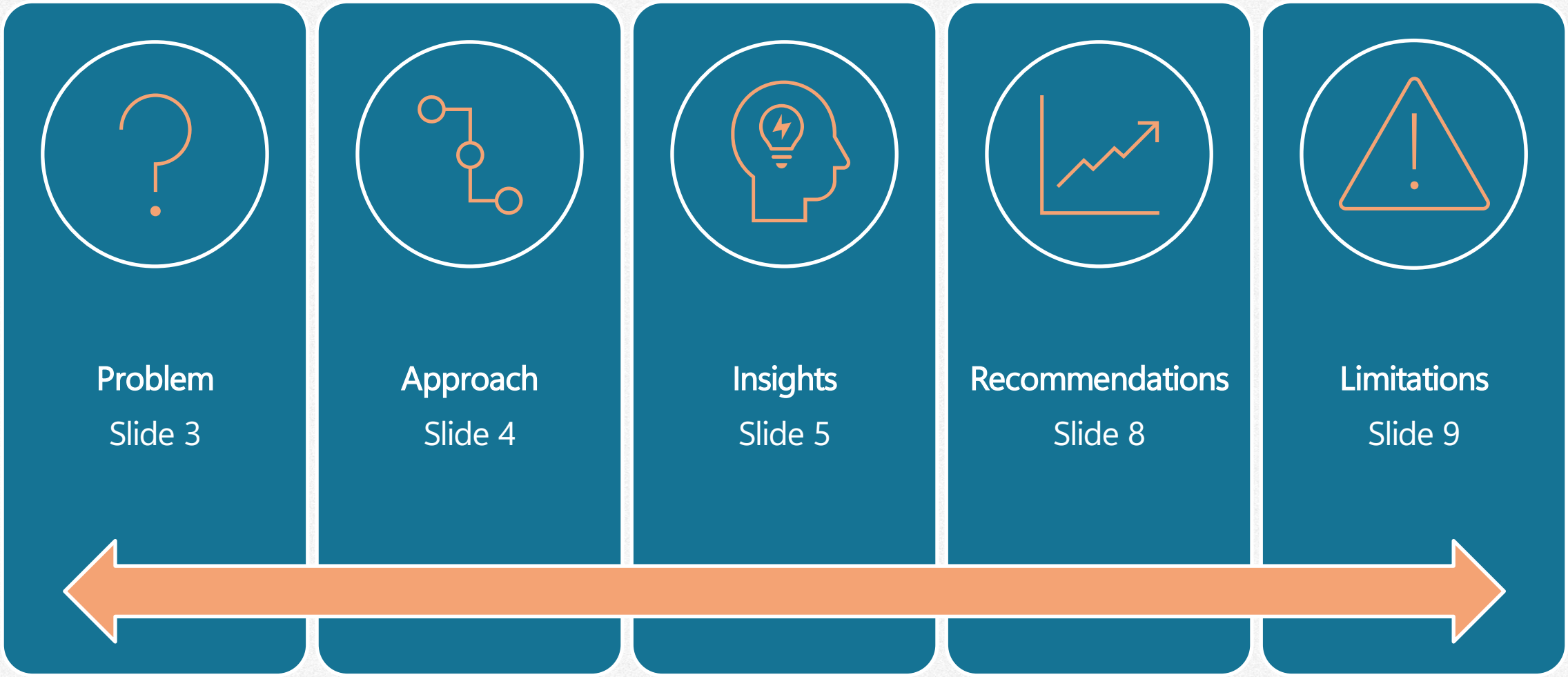
Acme Aroma Employee Turnover Project

Product of the Acme Aroma Data Science Department

Nick Wheatley

August 15, 2022

Table of Contents



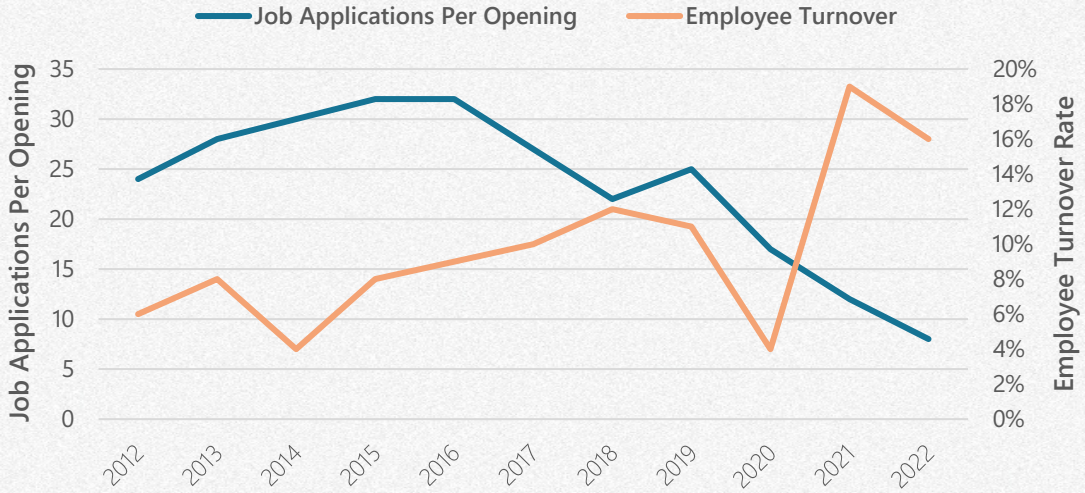
Acme Aroma Employee Turnover Project

Problem Statement: High Attrition Reduces Talent

Background

Acme Aroma has an attrition problem. The strong labor market is seeing more people leaving and fewer people applying for new roles. As shown in the visual below, in the last two years, Employer Turnover rates have doubled with only one-third the number of new applicants per job.

With the acquisition cost per new employee now at 3000 rupees, the attrition problem costs Acme Aroma **100-150% a full-time salary** in addition to decreased efficiency on the production line. This issue will perpetuate as strong talent leaves with a limited pool of new talent to replace them.



Root Cause Analysis

Review of attritted employees revealed that, compared to employees who stayed, **attritted employees experienced lower salary**, lower job level, lower salary hikes, and a slower rate of promotion among many other things. Root Cause analysis will permit us to understand which of these variables drive employee turnover the most.

> "Employee Turnover rates have doubled with only one-third the number of new applicants."

Predictive Insights

Better understanding of these variables further leads to an improved comprehension of employee sentiment, and the ability to better predict likely attrition in the future. HR will thus be able to make needed interventions to keep strong talent and decrease Acme Aroma's Employee Turnover rates.

With 2022 **employee attrition at 16%** and acquisition costs at an all-time high, the need for root cause analysis and predictive insight is strongly present. Their implementation could guide HR initiatives, generating significant savings and an improved employee experience at Acme Aroma .

Acme Aroma Employee Turnover Project

Approach: Enhanced Regression Turnover Model

Model Objectives

Explain and quantify the relationship between employee turnover and identified contributing variables

Use the quantified relationship to predict the probability of employee attrition

Objectives

Acme Aroma may understand and predict Employee Turnover with an Enhanced Regression Turnover Model. The objectives of this model are 1) explain and quantify the relationship between employee turnover and identified contributing variables and 2) use the quantified relationship to predict the probability of employee attrition in the future. **This model will be used to determine the best HR initiative to reduce turnover**.

Data Background

The data for this project was procured from Pegasus, the Human Resources Information System. Much of the data preparation was already completed by the previous data scientist allowing a focus on modeling. Survey data was gathered from a conjoint study conducted by an outside firm.

The key response variable in the model is attrition – a "yes" or "no" variable indicating if an employee has left the company. Among others, the relationship to attrition of the following explanatory variables will be assessed in determining key predictors:

- Job Satisfaction
- Education Level
- Gender
- Age
- Years at Company
- Marital Status
- Salary Hikes
- Engagement Scores

Analytical Approach

The model will use regression to examine the association between employee turnover and explanatory variables and quantify the relationship allowing turnover prediction. **This approach will employ logistic regression to return a probability between 0 and 1 of attrition**; the closer to 1, the more likely the employee is to leave the company.

Logistic regression will yield coefficients for each explanatory variable used to infer their impact on attrition (root cause analysis). These coefficients will then be used to predict future attrition outcomes.

Evaluation Methods

To assess the accuracy of the model, this project considers Type I and Type II errors in testing. A Type I error occurs if an employee is predicted to quit and they do not (false positive), and a Type II error occurs if an employee is not flagged as a flight-risk, but they leave the company (false negative).

These errors are considered in an F1-score and ROC Curve that assess model performance. F1-scores close to 1 indicate few Type I and Type 2 errors, and an ROC Curve area close to 1 indicates the same.

Acme Aroma Employee Turnover Project

Insights: Several Variables Correlate with Attrition

Model Description

Once meaningful explanatory variables (monthly income, job satisfaction, etc.) have been identified using exploratory data analysis (EDA), they are fed into the regression model for evaluation. Using logistic regression, the model compares these variables against attrition to determine their weight in driving employee turnover.

The output of the regression is a series of variable coefficients that explain the corresponding variable's relationship to attrition. These may also be used to predict the likelihood of turnover in the future. The optimal HR initiative to reduce attrition was selected by applying these coefficients to adjusted key metrics to explore their impact on expected attrition.

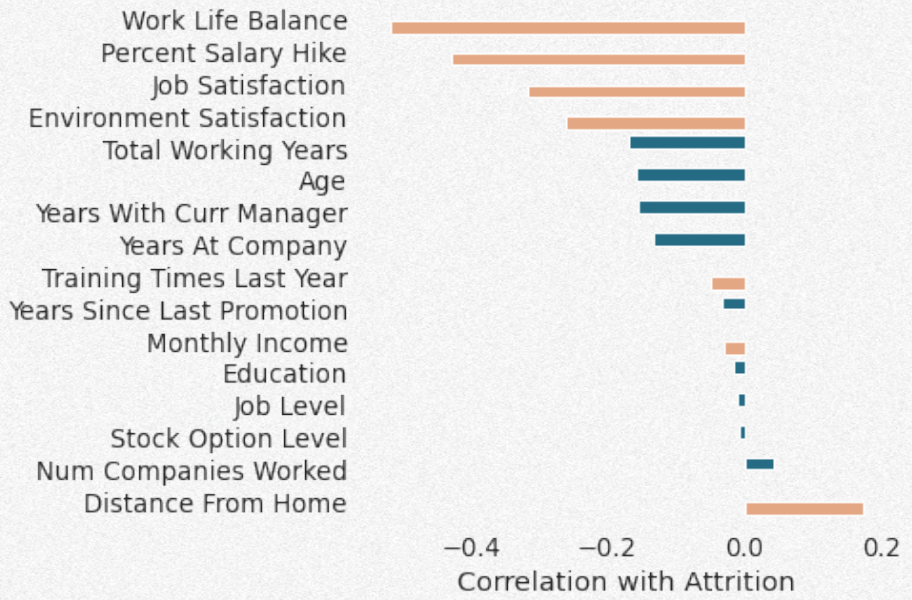
Exploratory Data Analysis

Initial data provided 24 independent variables to compare against attrition. While it was possible to use all variables in the model, doing so could introduce unnecessary bias, thus variables were selected through the process of exploratory data analysis.

In reviewing visual indicators such as histograms, bar plots, and correlation heat maps, five variables displayed a strong impact on attrition as detailed to the right. "Training Times Last Year" and "Monthly Income" were added to evaluate all possible HR initiatives. In the bar plot, a negative correlation suggests that as the variable increases, attrition decreases. A positive correlation indicates an increase in the variable produces an increase in attrition.

Explanatory Variables

- Work Life Balance
- Percent Salary Hike
- Job Satisfaction
- Environment Satisfaction
- Distance From Home
- Training Times Last Year
- Monthly Income



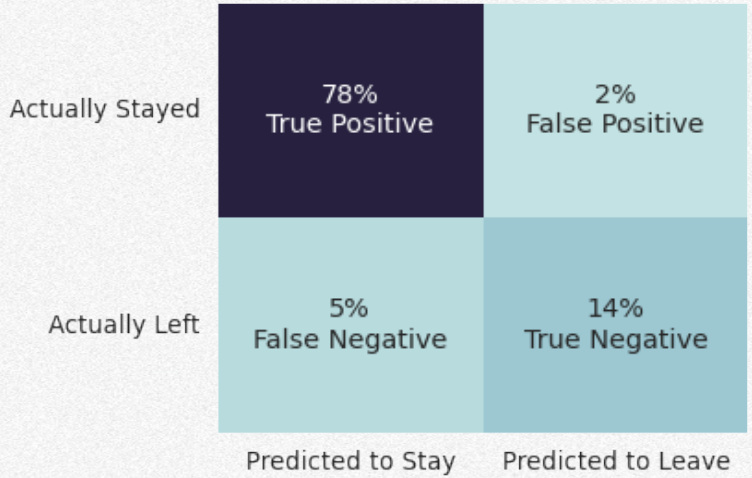
Acme Aroma Employee Turnover Project

Insights: Current Model Outperforms its Predecessor

Performance Overview

Accepting model results first requires trusting its accuracy. To measure this, several tests were applied each describing performance in a different way. In logistic regression, using the same data to train and test a model can lead to overfitting – a modeling error that overstates accuracy. To prevent this, the data was separated into training and testing datasets.

After training the model with the training dataset, the produced regression coefficients were applied to the test dataset to understand how predicted attrition fared against actual results. Accuracy was assessed by comparing employees who were correctly predicted to stay or leave (True Positive and True Negative) against those predicted to leave but stayed (False Positive, aka Type I error) and those predicted to stay but left (False Negative, aka Type II error). The results of the tests are shown in the visual below.



Precision, Recall, and F1-Score

Identifying True and False Positives and Negatives permits a better understanding of model accuracy. **Precision, Recall, and the F1-score are accuracy metrics that describe the model's ability to correctly assess likelihood of attrition.**

Each metric falls between 0 and 1 with a Precision score of 1 indicating no False Positives, and a Recall score of 1 indicating no False Negatives. F1-score combines these metrics as detailed below.

$$Precision = \frac{True\ Positive}{Employees\ Who\ Stayed}$$

$$Recall = \frac{True\ Positive}{Employees\ Predicted\ to\ Stay}$$

$$F1\text{-}score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The model's few False Negatives and False Positives produce Precision and Recall scores of 98% and 94% relatively. **Precision and Recall combine to result in a strong F1-score of 96%.**

Precision = 98%

Recall = 94%

F1-score = 96%

Acme Aroma Employee Turnover Project

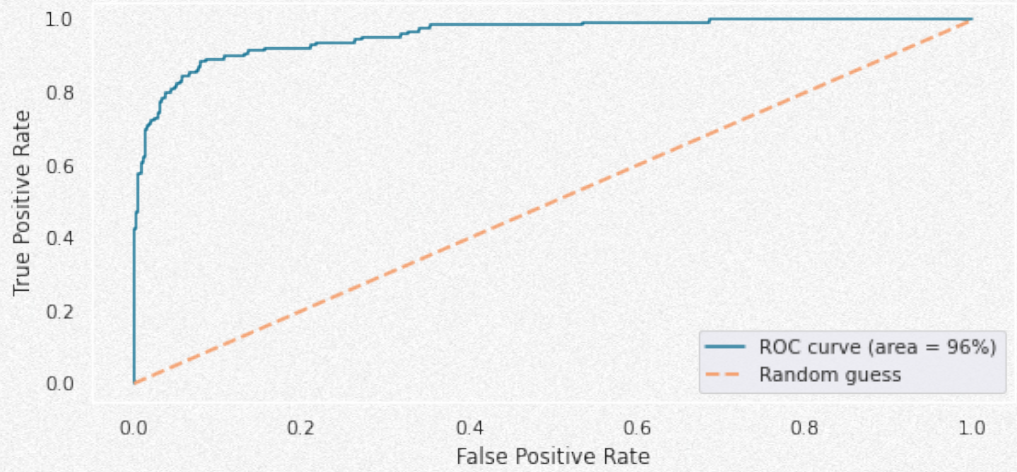
Insights: Satisfaction Indicators Drive Attrition

ROC/AUC

ROC is another tool for evaluating model performance that also considers Type I and Type II errors. It is a graphical technique for displaying model accuracy that plots the model's "True Positive Rate" (aka Recall) against its "False Positive Rate". The resulting curve should look like a steep initial slope that plateaus as "True Positive Rate" approaches 1.0. The lower the slope, the less accurate the model

The total value for the area under the curve (AUC) represents total model accuracy. A perfect model with no Type I or Type II errors would have an AUC of 1.0.

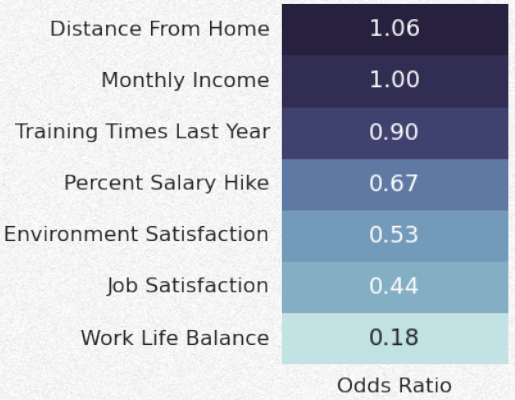
Applying the trained model to the test dataset yielded an AUC of 96% with the curve below. Combined with the F1-score, **performance measurements suggest that the model will accurately predict employee attrition 96% of the time**.



Root Cause Analysis

To quantify each variable's impact on attrition, the model's regression coefficients are converted to odds ratios. These ratios may be interpreted as the change in attrition for every 1 unit change in the explanatory variable. Odds close to 1 convey little impact and odds less than 1 represent an inverse relationship with attrition. Thus, as Work Life Balance increases 1 unit, risk of attrition decreases 82% (1.0 - 0.18).

Work Life Balance's impact on attrition appears strong. However, each variable is measured in different units. Monthly Income has a max of 199,990 Rs where Work Life Balance maxes at 5. To understand the impact of each variable, the needed change to generate a 10% reduction in attrition was measured to the right. The top three variables deliver the greatest change to attrition per unit of adjustment, with Work Life Balance driving the greatest impact.



Adjustment to Reduce Attrition 10%

Work Life Balance	**+0.12 Units**
Job Satisfaction	**+0.18 Units**
Environment Satisfaction	**+0.21 Units**
Percent Salary Hike	+30%
Training Times Last Year	+1 Training
Distance From Home	-1.57 Km
Monthly Income	+33,978 Rs

Acme Aroma Employee Turnover Project

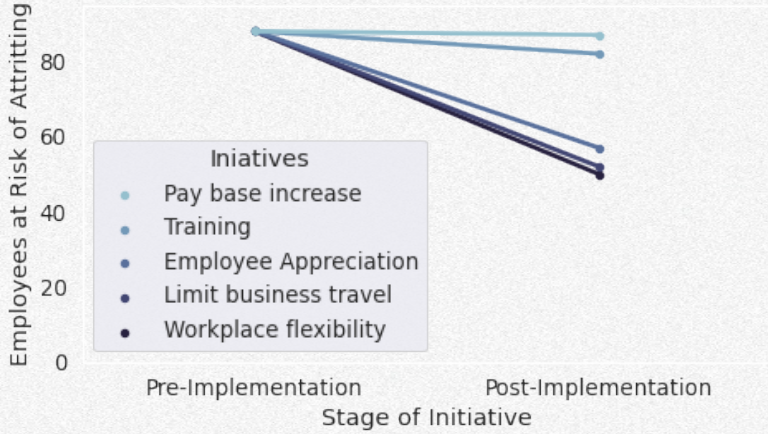
Recommendation: Pursue Employee Appreciation

Model Recommendation

With 96% accuracy ratings from ROC/AUC and F1-Score performance measures, and with strong indication of explanatory variable driving attrition, the data science department recommends this model for immediate implementation. While it is possible to further enhance this model in the future, with Acme Aroma's current attrition problem, the enhanced regression model as it currently exists can produce immediate financial benefit.

Initiative Performance

Each explanatory variable drives attrition to some degree. The model showed a decrease in attrition by implementing any one of HR's initiatives (see Table 1 In the Appendix). Flagging employees with greater than 50% probability of attrition as high risk of leaving, the model compared the change in number of high-risk individuals before and after the implementation of each initiative. The results are detailed in the visual to the right.



Introducing **workplace flexibility, limiting business travel, and increasing employee appreciation all show potential for a significant impact on employee turnover.** Pay base increase and employee training show less likelihood of driving positive impact. With an estimated 3000 rupees required to train new employees, deploying these initiatives represent the following possible savings:

Workplace Flexibility (+38 employees): 114,000 Rs
Limiting Business Travel (+36 employees): 108,000 Rs
Employee Appreciation (+31 employees): 93,000 Rs

However, of those flagged as likely to leave, a quarter are unable to work from home due to their roles (Lab Technicians and Manufacturing Directors) and 82% never or rarely travel. While working from home and reducing business travel may stimulate improved turnover rates in the future, they will not fix the problem today.

Initiative Recommendation

With added workplace flexibility and limited business travel offering a reduced benefit for the current employee base, **introducing an employee appreciation initiative is our recommended path going forward**. Not only will Acme Aroma experience reduced expenses in recruitment and onboarding costs, but it will also avoid decreased productivity from inexperienced staff.

Acme Aroma Employee Turnover Project

Limitations: The Model has its Caveats

Limitations in the Data

Despite the strong performance measures and root cause analysis, this model has its limitations. As is often the case with survey data, the information collected had its share of missing values. To properly train the model, missing values were replaced with their field's average. With a large enough dataset and few missing values, this should not lead to any significant shift in patterns, but it's possible some directional data was not considered affecting predictions.

Wary Assumptions

The accuracy of the model is predicated on two key assumptions:

1. A positive change in one variable does not negatively impact another variable
2. There is little to no bias in the survey responses

It is possible that increasing Job Satisfaction through employee appreciation initiatives will have unforeseen consequences on other variables that drive attrition.

It is also possible that bias exists in the survey data. While it is unlikely that respondents had incentive to falsify answers with the survey coming from an external firm, surveyees can blindly respond by selecting random answers to quickly end the survey. Doing so would negatively impact model performance.

Model Caveats

As seen with travel and work-from-home initiatives previously, certain employee cohorts would likely respond to initiatives differently than others. Model output should not be accepted unless explored thoroughly.

This is a possible reason for 5% of the test model predictions resulting in Type II errors. As the employee base grows, it may be beneficial to run training and testing data on individual cohorts, such as splitting by job role and travel frequency.

Appendix

Table 1: Initiative Options and Impact

Initiative	**Description**	**Effect**
Pay base increase	Increase base pay for all employees	Increase *MonthlyIncome* by 7500
Training	Provide additional professional development	Moves *TrainingTimesLastYear* up 0.5 on average.
Workplace flexibility	Allow work-from-home	*EnvironmentSatisfaction* increases by 0.75
Limit business travel	Reduce the amount of required business travel	*WorkLifeBalance* increases by 0.3
Employee Appreciation	Various employee appreciation initiatives	*JobSatisfaction* increases by 0.5