## **Abstract**

Canterra is facing a growing challenge with 15% annual employee turnover, impacting productivity, increasing recruitment costs, and straining key relationships with clients and partners. This report analyzes the key factors driving attrition: age, income, job satisfaction, and tenure. Our findings highlight that younger employees, particularly those under 25, and dissatisfied workers are the most at risk of leaving the company. Moreover, income disparities contribute to employee turnover, as those who feel underpaid are more likely to seek opportunities elsewhere.

Based on these insights, we recommend specific actions to address the root causes of turnover:

* Enhance career development programs for younger employees to increase retention and foster future leaders.
* Improve employee job satisfaction through recognition, feedback, and work-life balance initiatives to reduce dissatisfaction and increase loyalty.
* Competitive pay and targeted benefits to ensure employees feel valued and reduce the impact of income disparities.
* Engagement Strategies for Long-Term Employees to prevent burnout and keep them motivated and invested in the company.

Through these measures, Canterra can decrease turnover by 15-20%, improve employee satisfaction, and establish a more stable, high-performing workforce. The strategy will also help save money on hiring and training, which is also beneficial for business success in the long term.

### **Introduction**

Workplace turnover is among the most costly and disruptive problems any organization can encounter. For Canterra, which has around 4,000 employees, a 15% annual attrition rate costs money – it drains resources, reduces productivity, and damages key knowledge. It does not only raise hiring and training costs, but also delays projects, damages partnerships with customers and partners, and undermines organizational workforces.

At Canterra, reducing attrition is no longer an option – it’s a business necessity. Turnover requires a firm to pinpoint its cause and make bold decisions to minimize the effects. By harnessing data and predictive models, Canterra can not only reduce the immediate costs of attrition but also lay the foundation for sustained success in a competitive job market.

This article employs a data-driven approach to pinpoint the hidden causes of turnover and define the most promising tactics for mitigating these risks. It’s the first and most important step to transforming attrition into an opportunity to improve retention, employee satisfaction, and business performance.

### **Notable Market Trends**

Turnover is a significant challenge for any company. The cost of replacing a worker is between 50% and 200% of the salary received by that employee per year, depending on the position (King, 2022). Researchers have uncovered several key factors behind employee turnover, emphasizing the need for proactive retention efforts.

* **Career Development Opportunities:** The employees appreciate job development. A Gallup poll in 2023 discovered that 87% of millennials value career growth as their number-one concern when choosing an employer (Hickman, 2020). Yet, only 29% of workers feel they have ample opportunities to develop within their current organization, which suggests a disconnect between what people expect and what organizations provide (SHRM, 2024).
* **Employee Engagement & Job Satisfaction:** Workplace engagement is directly associated with retention. Gallup found that highly engaged workers are 59% less likely to leave an organization than their disengaged peers (King, 2022). Yet 85% of all workers in the world don’t actually attend work, which means that turnover is highly correlated with low engagement (Montgomery, 2018). Low perception and underperformance are chief motivators of engagement.
* **Compensation & Benefits:** Compensation remains one of the most critical factors influencing employee retention. A recent WTW survey found that 56% of employees would leave their current job for better pay elsewhere, and 39% stated they would leave for better health insurance (Allcot, 2022). Salary disparities and lack of benefits can accelerate turnover in competitive job markets.
* **Flexible Work Arrangements:** Flexibility at the office is no longer an option - it is a requirement. JazzHR also finds that 67% of workers consider flexible work options a deciding factor when accepting employment (Sparks, 2023). Further, the absence of flexible work hours causes employees 2.5 times more likely to quit a job for a remote or hybrid-work position (Lovich & Sargeant, 2023).

A fiercely competitive job market poses a challenge to employers as much as it presents an opportunity. Workers have now expected far more from the Great Resignation and are seeking positions that will support their values, mental health, and life-style. When workers increasingly demand flexibility and an improved employee experience, companies must respond to ever-higher expectations for pay while also adapting to a hyper-competitive workforce. Businesses that aren’t up to speed on these demands are in danger of losing top talent to competitors that offer better working conditions and greater personal satisfaction.

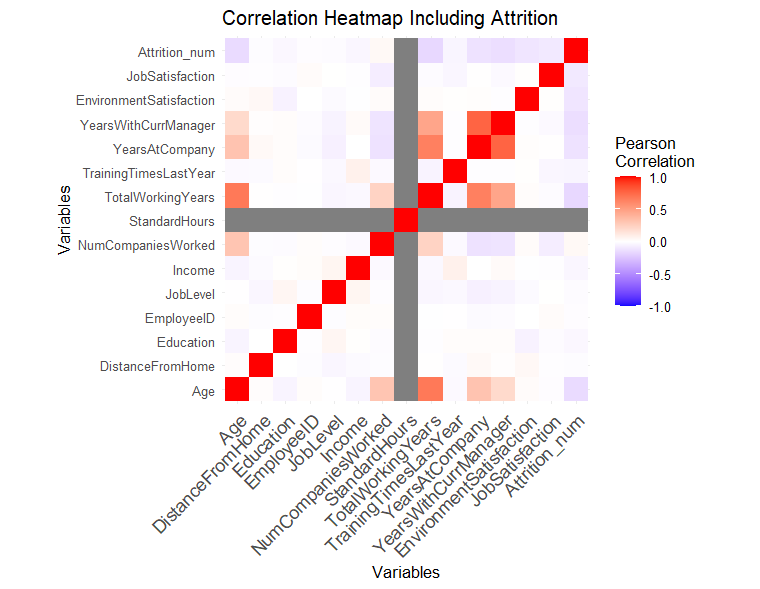
Furthermore, inflation, labor shortages, and fast-moving work habits remain the determinants of recruitment and retention. All of these make workforce management challenging, and thus companies have an additional obligation to modify their strategies in order to stay ahead of the competition. To attract and retain top talent while positioning themselves as employers of choice, businesses must cultivate an inclusive, adaptable, and purpose-driven culture that aligns with an economy increasingly focused on balance, meaningful interaction, and purpose.

Companies that take advantage of these concerns will both increase their ability to attract and retain talent and better position themselves for the long term in a changing economic environment. This report translates these general trends into Canterra and provides actionable guidance to help keep people, mitigate high turnover, and create a more committed, satisfied, and productive workforce.

### **Exploratory Data Analysis and Data Preparation**

The journey of preparing the dataset for analysis started with a review of the structure and context of variables. The dataset consisted of 4,410 observations with eight categorical and numerical variables and had "attrition" as the target variable. Before cleaning the data, descriptive statistics were calculated to learn more about Canterra employees. The average age of employees at Canterra was 36 years old, and most employees had spent about 7 years with the company. The Attrition variable was shown to be dramatically imbalanced, with 16.6% being labeled as "yes," indicating an employee has left, and 83.4% as "no," indicating the employee just stayed. At this step, this was critical since the correction of this imbalance should provide a balanced and reliable prediction.

Next, data cleaning and transformation processes were conducted to ensure the dataset's integrity. Missing values were identified and removed, while outliers were systematically examined and addressed to prevent potential biases that could impact the model's performance. Some variables were reformatted: Total Working Years, Number of Companies Worked, Environmental Satisfaction, and Job Satisfaction were transformed from character formats into numerical variables for better usability in the analysis. Similarly, the Attrition variable was converted to a binary format where "no" and "yes" were represented by 1 and 2, respectively, to make the input simple for the models. To ensure the data was appropriately prepared for model training and evaluation, the data was split into training and testing sets using a stratified sampling approach. The data was divided into 70% for training and 30% for testing.

A heatmap was used to assess variable correlations to uncover potential relationships and refine the analysis. This visualization showed some areas of multicollinearity that deserved further investigation. From this, the relationships between the key variables—Income, Business Travel, and Age—and the target variable, Attrition, were explored. The first relationship explored was the income distribution by Attrition; the Attritions (Appendix A) show that employees who stayed have a wider income range, with a higher median income than those who left. When looking at the relationship between attrition by business travel, employees who travel frequently exhibit higher attrition rates than those who travel rarely or not (Appendix B). Employees who traveled the most were 24.7% more likely to leave, which is significantly higher compared to other groups that were on Non-Travel (8.1%) or Travel Rarely (15.0%). The last relationship assessed was between age and attrition; the boxplot in Appendix C showed that younger individuals tended to leave the company at higher rates. This analysis painted a more nuanced picture of how these factors influence employee turnover. The lessons learned from this exploration guarantee that the model is based on a well-prepared dataset and a clear understanding of the factors increasing employee attrition at Canterra.

1. **Model Development**

Four models were developed after cleaning the data: logistic regression, decision tree, bagged tree, and random forest. These models were selected for their unique strengths in handling complex datasets and generating actionable insights.

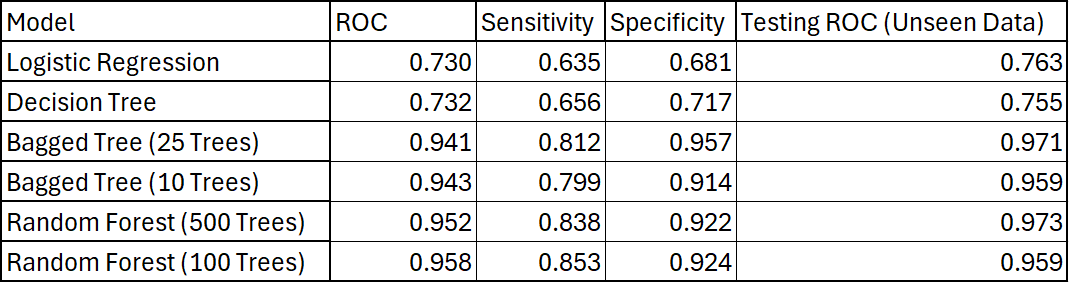
* **Logistic Regression** is ideal for analyzing binary variables like attrition, predicting the probability of outcomes as 'yes' or 'no.'
* **Decision Trees** provide interpretability by visually illustrating factors influencing employee attrition in a hierarchical structure that is easy to understand.
* **Bagged Trees** enhance model robustness and reduce overfitting by combining multiple decision trees.
* **Random Forest** builds on Bagged Trees by introducing random feature selection at each split, which improves prediction accuracy.

These four models complement each other in the analysis by offering a balanced output of interpretability and predictive power. Allowing easy identification in solving Canterra’s management’s retention issue while also maintaining a high degree of model performance.

* **Logistic Regression model:** The data underwent a 70/30 train/test split for the Logistic Regression model, and the class imbalance was addressed by down-sampling the negative class. Variables such as ‘EmployeeID’, ‘StandardHours’, ‘Job Level’, 'Distance From Home’, and ‘Gender’ were excluded as they did not contribute to model performance, and outliers were removed to ensure cleaner data.
* **Decision Tree:** The Decision Tree model was developed for its simplicity and interpretability in identifying key decision points. It used the Gini index as the splitting criterion to separate Attrition at each node. ‘EmployeeID’ was excluded for its lack of predictive value, and missing (NA) values were removed. Trained using the “rpart” method, the model's complexity parameter (cp) was tuned through grid search (0.001–0.1) to prevent overfitting. A 10-fold cross-validation ensured a balance between bias and variance for robust performance.
* **Bagged Tree Model:** The Bagged Tree model was developed to improve prediction stability and accuracy by aggregating results from multiple decision trees through bootstrap aggregation (bagging), reducing variance and overfitting. Two versions were created: one with 25 trees and a reduced version with 10 trees. Trained using the tree-bag method, the model followed similar preprocessing steps as the Decision Tree model. excluding ‘EmployeeID’ and removing NA values. A 10-fold cross-validation was applied, and the Receiver Operating Characteristic Area Under the Curve (ROC AUC) metric was used to evaluate predictive performance.
* **Random Forest Model:** The Random Forest model was developed to enhance predictive performance and address potential overfitting issues that arise in single decision tree models. Random Forest combines the predictions of multiple decision trees, where each tree is trained on a random subset of the data, and features are selected randomly at each split. This approach introduces diversity among the trees, which helps in reducing variance and improving the model's generalizability.

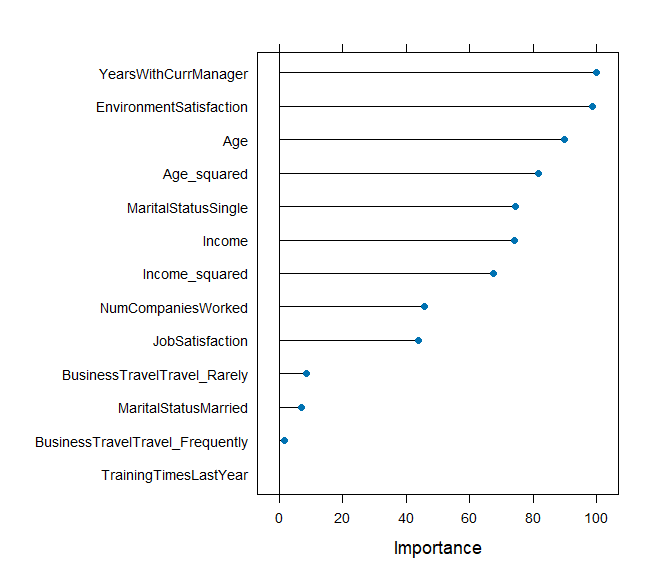
The four models provided a comprehensive framework for addressing Canterra’s management's retention challenges by balancing interpretability and performance, offering actionable insights into employee attrition, and supporting informed decision-making for long-term solutions.

1. **Model Performance**



Four models were evaluated, with two similar tuned versions also being tested to assess their predictive performance. The models were compared using key metrics: ROC, sensitivity, specificity, and testing ROC on unseen data. ROC measures the model's ability to distinguish between positive and negative outcomes, with values ranging from 0.5 (random guessing) to 1 (perfect classification). The distinction between ROC and testing ROC lies in the latter's ability to assess how well the model generalizes to new, unseen data. A high-testing ROC indicates minimal overfitting and strong predictive power. Sensitivity and specificity evaluate the model's accuracy in identifying true positives and negatives, respectively. Higher values in these metrics indicate better performance. Ultimately, the model with the highest ROC, sensitivity, and specificity is considered the most effective at predicting employee attrition.

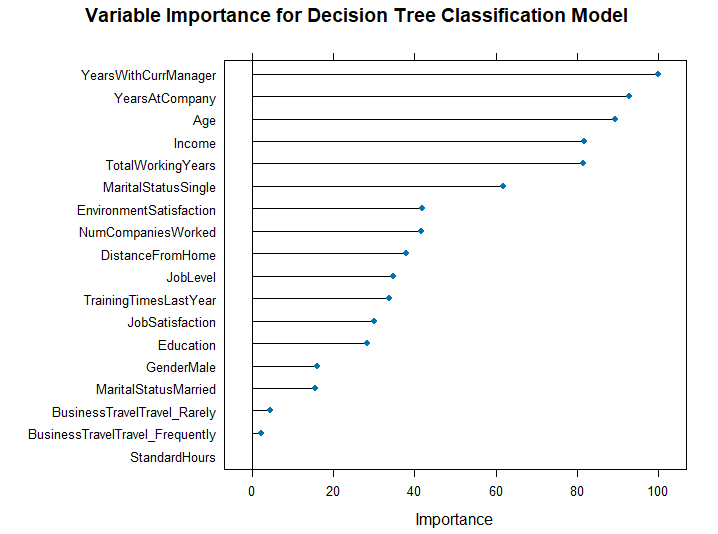
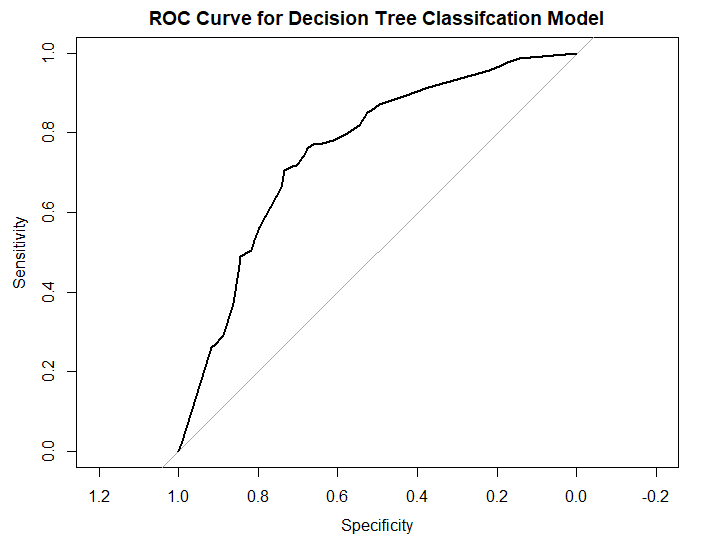
**Logistic Regression Model**

The final logistic regression model had an in-sample ROC AUC of 0.7295, a sensitivity of 0.635, and a specificity of 0.681. The out-of-sample model on the test set produced a ROC AUC of 0.763 (Appendix D). The model’s sensitivity indicates that 63.5% of the classifications were true positives for employees who were classified under attrition and left the company. The model’s specificity score indicates that 68.1% of the classifications were true negatives for employees who did not leave the company. For variable importance, years with current manager, environment satisfaction, age, Marital status was single, and income was the top 5 variables that contributed to the predictive performance of the model for attrition (Appendix K). Age, environment satisfaction, job satisfaction, and income all significantly negatively affected attrition while business travel, marital status, and the number of companies worked all had significantly positively affected attrition (Appendix L). These results suggest that Canterra should focus on improving job and environment satisfaction for employees. Targeted retention efforts should be directed at employees with higher risks of attrition who travel frequently, single employees, and employees who have a work history with a large number of companies. 

**Decision Tree Model**

After completing the training phase, the model identified the best cp value of 0.001, which yielded a ROC AUC score of 0.732, a sensitivity of 0.656 ,and specificity of 0.717. When using the model on unseen data the area under the curve was 0.755. The sensitivity (true positive rate) measures the model’s ability to correctly identify employees who will leave, meaning (on average) the model correctly identifies 65.6% of employees who will leave the company. The specificity (true negative rate) is the ability of the model to correctly identify employees who will stay, meaning the model (on average) correctly identifies 71.7% of employees who are likely to stay.

This suggests that the model has a moderate predictive capability. Additionally, variable importance was assessed which resulted in the most critical predictors being Total Working Years, Age and Income . Meaning that these are the top 3 variables that affect employee Attrition. These features were also analyzed through a scatter plot.

Overall, the decision tree model has a moderate ability to successfully predict attrition and underscores the importance of tenure, experience, and compensation when it comes to employee retention.

**Bagged Tree Model**

The default Bagged Tree achieved an average ROC AUC of 0.941, average sensitivity being 0.812. and average specificity of 0.957. The smaller ten bootstrap replications achieved an average ROC AUC of 0.943 with a sensitivity of 0.799 and specificity ranging from 0.914.

For the default Bagged Tree model, the variable importance analysis highlighted age, income, and total working years as the most critical predictors of employee attrition (Appendix E). While for the smaller variation the analysis highlighted income as the most important variable followed by age and total working years (Appendix F). Although the top three variables are the same, the order of importance differs.

Finally, all models were tested with unseen data in order to validate their performance. The 25-tree-bagged model had an ROC AUC of 0.971 (Appendix I). The 10-tree bagged model had an ROC AUC of 0.959 (Appendix J), while the single-tree decision tree had an ROC AUC of 0.755. The results indicate that both the Bagged Tree and reduced Bagged Tree version were very strong models in terms of predictive analysis, but also there’s significant improvement in predictive analysis using the Bagged Tree method.

In conclusion, the visual summaries, including the ROC Curves and comparison between different variables of importance plots provide a comprehensive overview of the model’s performance and highlights how the Random Forest model does the best at predicting attrition. By prioritizing factors such as “Income”, “YearsAtCompany”, “Age”, and other significant factors seen on Appendix E organizations can effectively retain employees.

**Random Foresting**

The initial model demonstrated strong predictive power, achieving a ROC AUC of 0.952, sensitivity being 0.838, specificity being 0.922. OOB estimate of error rate is 8.49%, which means that the model has an error rate of 8.49% for unseen data. The confusion matrix revealed that among instances labeled “No”, 428 were correctly classified while 61 were misclassified as “Yes” (class error rate of 12.47%). In addition, among instances labeled “Yes” 467 were classified correctly while 22 were misclassified as “No” (class error of 4.5%).

The adjusted model achieved a ROC AUC of 0.958, with sensitivity ranging from 0.853 and specificity being 0.924. The OOB error of the tuned model is 9.71%, significantly higher than the original model. The classification errors were also higher with “No” having a class error rate of 14.929% and “Yes” having an error of 4.499%. The OOB error rate and classification errors rose compared to the initial model. The important variables for the initial model were age, income, then total working years (Appendix G). For the tuned model it was income, total working years, then age (Appendix H).

When testing all models with unseen data, the initial model outperformed the tuned model, with the area under the curve being 0.972, compared to the smaller 100 tree model, with the area under the curve being 0.959. Based on the results, the initial Random Forest model is a stronger predictor of employee attrition due to its superior performance on unseen data, lower OOB error rate, and better classification accuracy.

1. **Interpretation of Results**

The output of the four models demonstrates different levels of effectiveness when analyzing and predicting employee attrition. While each model has its own strengths and weaknesses, the output of each model and the way it analyzes the data differs, which results in different outputs.

The Logistic Regression model demonstrated its limited ability to predict employee attrition with an ROC of 0.730 and testing ROC of 0.763. The sensitivity of 63.5% indicates that the model struggles to identify employees that leave and its specificity of 68.1% also would suggest it isn’t suitable for identifying employees that stay. These metrics suggest that while the simplicity of the model does provide a good general understanding of factors that might influence attrition, the predictive capabilities are insufficient compared to other models.

The Decision Tree model offers slightly better interpretability with similar ROC values of 0.732 and testing ROC of 0.755, while having much higher sensitivity and specificity of 65.5% and 71.7% respectively. The hierarchical structure of the Decision Tree highlighted variables such as Years with Current Manager, Total Working Years, and Age that affected attrition the most, with StandardHours, business travel, and marital status affecting attrition the least. Although the model improved compared to the Logistic Regression model, the other two models offer a significant leap in predictive power

The Bagged Tree model, through aggregation, reduced the variance and overfitting of the Decision Tree model. With ROCS of 0.941 and 0.943 for the 25-tree and 10-tree versions, respectively, the two models achieved testing ROCs of 0.971 and 0.959. The sensitivity and specificity also increased to around the 80s for sensitivity and 90s for specificity. Interestingly, the variable importance analysis indicated that Age, Income, and total working years were the most important predictors for attrition in both the 25-tree and 10-tree model–although in different orders of their significance. The least impactful variables were similar to the Decision tree, that being: StandardHours, business travel, and marital status. The consistency of the VIP analysis also serves to strengthen the confidence in the reliability of the models.

Finally, the Random Forest model achieved the highest ROC of 0.972 due to its ability to combine both random feature selection with ensemble learning. This allowed the model to also achieve the highest sensitivity of 83.8% and specificity of 92.9%. Although the tuned model with 100 trees was also highly effective, it showed a slight reduction in ROC, which suggests that the initial version was a stronger model. Like the Bagged Tree model, the most and least effective VIP predictors were the same across both variations of the Random Forest model (Age, Income, Total Working Years & StandardHours, business travel, marital status) but ordered differently in importance.

In conclusion, the analysis suggests that the Random Forest model with 500 trees is best at predicting employee attrition when applied to unseen data. To address attrition, Canterra would need to focus on sectors such as Age, Income, and Total Working Years if the organization wants to identify high-risk employees or develop retention strategies that focus on improving compensation, tenure, and work satisfaction for younger employees.

1. **Limitations and Assumptions**

The following are some limitations that must be considered when using the Random Forest model with 500 trees:

* **Hyperparameters:** The model's performance is susceptible to hyperparameters, including the number of features selected (mtry) at each split and the total number of trees. These have to be tuned carefully for optimum performance.
* **Overfitting:** While Random Forest reduces overfitting concerning a single decision tree, the most complex models can overfit in the presence of noisy or redundant features.
* **Class Imbalance:** Class imbalance, as reflected by higher error rates for the "No" class, remains a challenge in the data, with “No” having a higher error rate of 12.47%, compared to 4.5% for the "Yes" class.
* **Misclassifications:** The OOB error rate of 8.49% does show room for improvement, especially when one considers real-world applications where misclassifications can have grave consequences.

Some key assumptions underpin the effectiveness of the Random Forest model:

* **Independent**: It assumes that observations are independent, which may not hold if there are correlations or dependencies within the data.
* **Predictive Power:** The model requires that input features are relevant and have predictive power; irrelevant or noisy features can dilute effectiveness.
* **Balanced:** Balanced representation of both classes (“Yes” and “No” for attrition) is crucial, as imbalanced data can skew predictions even after balancing efforts.
* **Multicollinearity:** Random Forest is robust regarding multicollinearity, a high degree of correlation among features reduces interpretability and biases variable importance.
* **Large Amount of Data:** Sufficiently large data is required to train the ensemble effectively, as smaller datasets risk redundancy or overfitting.
* **No Bias:** Last but not least, if all the features hold the same importance at the outset, no bias is naturally drawn toward their impact on the target variable.

The previously said limitations and assumptions cannot be left unaddressed to keep the model reliable and true to Canterra’s objectives.

1. **Ethical and Regulatory Considerations**

When assessing Canterra’s employee attrition, the use of personally identifiable employee information must adhere to ethics and regulatory guidelines to ensure equality, confidentiality, and accountability. This includes adherence to data protection laws that govern the acquisition, handling, and utilization of employee data, as well as equitable predictive modeling. The following are some of the important regulations and risks involved in using your employees’ personal data:

1. **Data Privacy and Confidentiality:** The collection and analysis of personal employee data (age, income, job satisfaction, etc) must adhere to data protection regulations such as the General Data Protection Regulation (GDPR) of the European Union, which focuses on transparency, employee consent, and the ability to withhold consent (European Union, 2016). There are also regulations in the United States, such as the California Consumer Privacy Act (CCPA), that provide employees with privacy rights and ensure data collection is open and transparent (State of California Department of Justice, 2018). Employers are required to abide by these laws to avoid heavy penalties in the form of legal and financial fines for improper data processing or misuse.
2. **Bias and Fairness in Predictive Modeling:** Predictive models should not discriminate due to age, gender, or other non-recognizable characteristics. Title VII of the Civil Rights Act of 1964 and the Age Discrimination in Employment Act (ADEA) shield workers from discrimination on the job (US EEOC, 1967)(US EEOC, 1964). Unfairly biased models, including those that disproportionately benefit younger workers, could result in lawsuits under Equal Employment Opportunity Commission (EEOC) guidelines (US EEOC, 1964). We need periodic audits of predictive models for consistency to avoid accidental bias and ensure we align with anti-discrimination rules.
3. **Informed Consent and Transparency:** Workers should be aware of the use of their data and will be given clear consent for data processing. The GDPR mandates that employees be informed about the purpose of data collection and can revoke consent at any point (European Union, 2016). In the United States, the Fair Credit Reporting Act (FCRA) promotes openness regarding the use of personal information for employment purposes (Federal Trade Commission, 2013). Employers should ensure that employees understand and consent to how their data will be used, both to stay in line with the law and to secure their trust.

The ethics and compliance requirements regarding the use of individual employee data in Canterra highlight the need for transparency, equity, and compliance with laws such as GDPR, CCPA, and anti-discrimination legislation. Predictive modeling can provide valuable information on employee retention, but only when done in a manner that preserves privacy, does not bias, and fosters trust. By following data protection regulations, regularly reviewing models for fairness, facilitating informed consent, and fostering engagement with employees, Canterra can leverage predictive analytics effectively and responsibly to reduce attrition, employee satisfaction, and foster a positive workplace culture. This kind of morality will not only reduce the risk of litigation but also create a positive and inclusive culture where workers feel appreciated and involved.

1. **Recommendations and Conclusion**

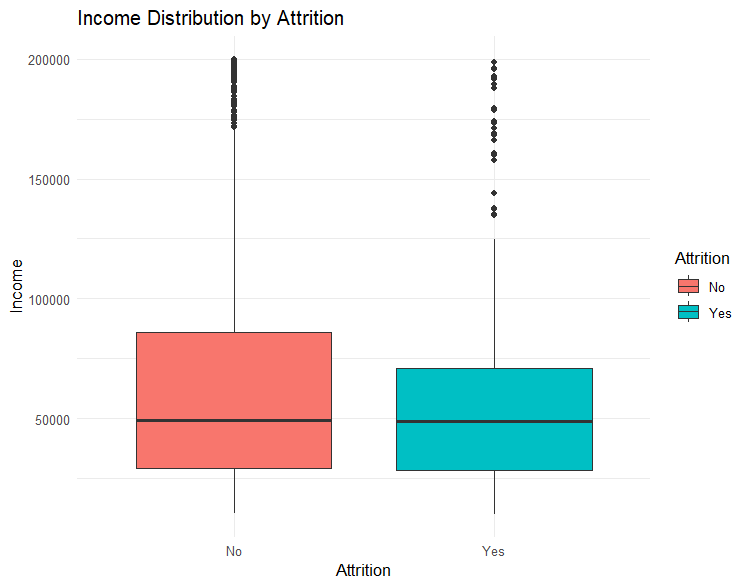
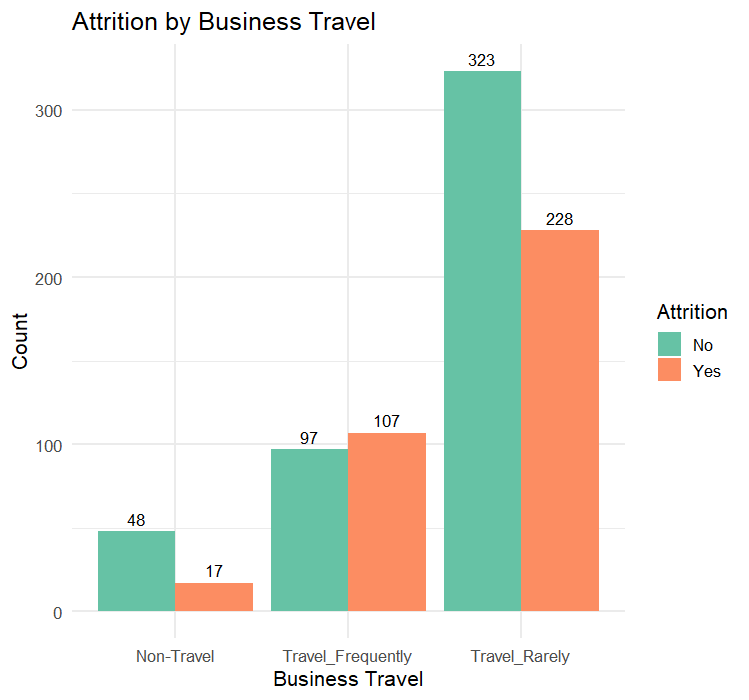
The analysis of employee attrition at Canterra, especially using the Random Forest model, has uncovered some crucial insights into what drives turnover. The model pointed to a few key factors—age, income, and total working years—that strongly influence whether employees stay or leave. With these findings in mind, we recommend several strategies to help reduce turnover, boost employee satisfaction, and build a more stable, loyal workforce.

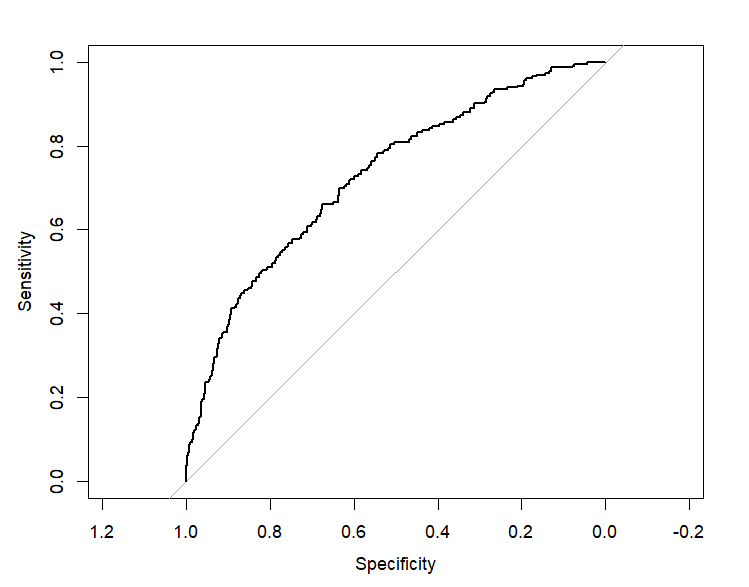
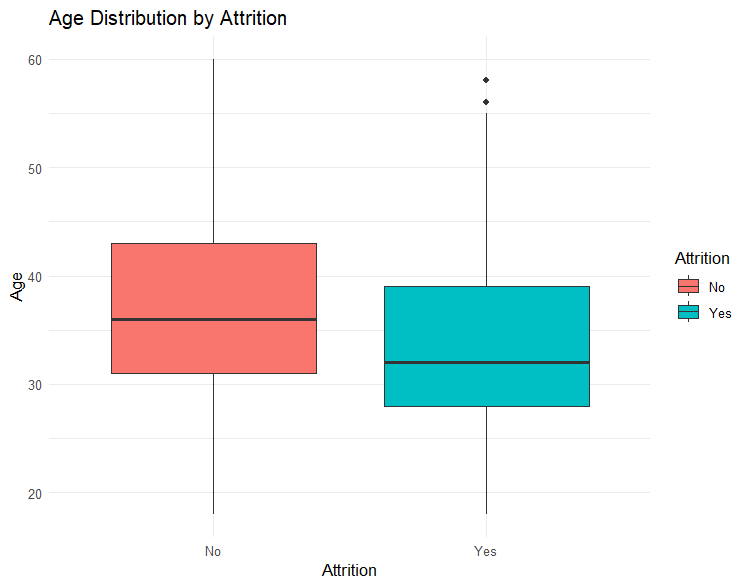
1. **Enhance Career Development Programs for Younger Employees to Increase Retention:** The model highlighted that younger employees, particularly those under 25, are at a higher risk of leaving. To keep them engaged, Canterra should roll out targeted career development programs like mentorship, leadership training, and clear growth paths. This will not only help retain these employees but also build a future pipeline of leaders, reducing turnover by up to 10% and ensuring the company’s long-term growth.
2. **Improve Employee Job Satisfaction Through Recognition, Feedback, and Work-Life Balance Initiatives**: Job satisfaction was a major driver of attrition, with dissatisfied employees being 55% more likely to leave. Employees who don’t feel recognized or get enough feedback are at an even higher risk. To tackle this, Canterra can create robust recognition programs, establish regular feedback channels, and offer more flexible work options like remote work and flexible hours. This will address dissatisfaction directly, improve morale, and hopefully cut turnover by 8-10%.
3. **Competitive Pay and Targeted Benefits Boost Employee Retention:** Compensation also played a big role in employee retention. The model showed that income disparities, particularly among lower-income employees, were a key factor in turnover. Canterra should regularly benchmark salaries to make sure they’re competitive, introduce performance-based bonuses, and offer targeted benefits like wellness programs and financial support. These measures will help employees feel valued and more satisfied, reducing attrition.
4. **Implement Strategies for Long-Term Employees to Foster Continued Engagement and Prevent Burnout:** While employees with longer tenures generally stay longer, there’s still a risk of burnout for those who’ve been with Canterra for years. The model showed that these employees may feel stagnated, which can lead to disengagement. To keep them motivated, Canterra should offer leadership opportunities, skill rotations, and high-impact projects. This will ensure that long-term employees remain challenged, recognized, and invested in the company’s success.

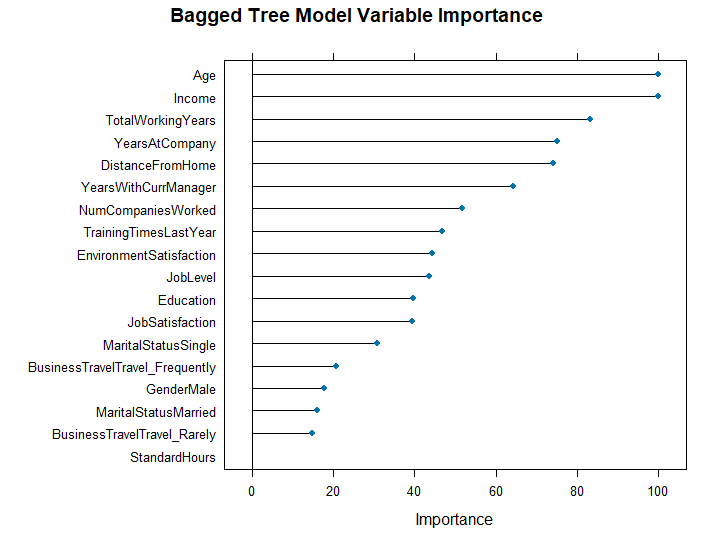
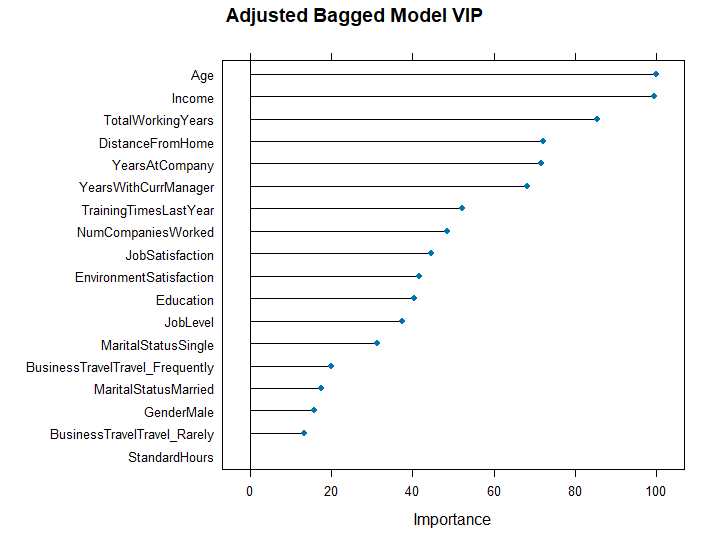
The analysis has given us a clear picture of what’s driving turnover at Canterra—things like age, income, job satisfaction, and how long employees have been with the company. With 15% attrition each year, it’s becoming a real challenge that’s costing the company time, money, and stability. By focusing on things like career development for younger employees, improving job satisfaction with more recognition and flexibility, and offering competitive pay, Canterra can make a big difference in reducing turnover and building a more committed and happy workforce.

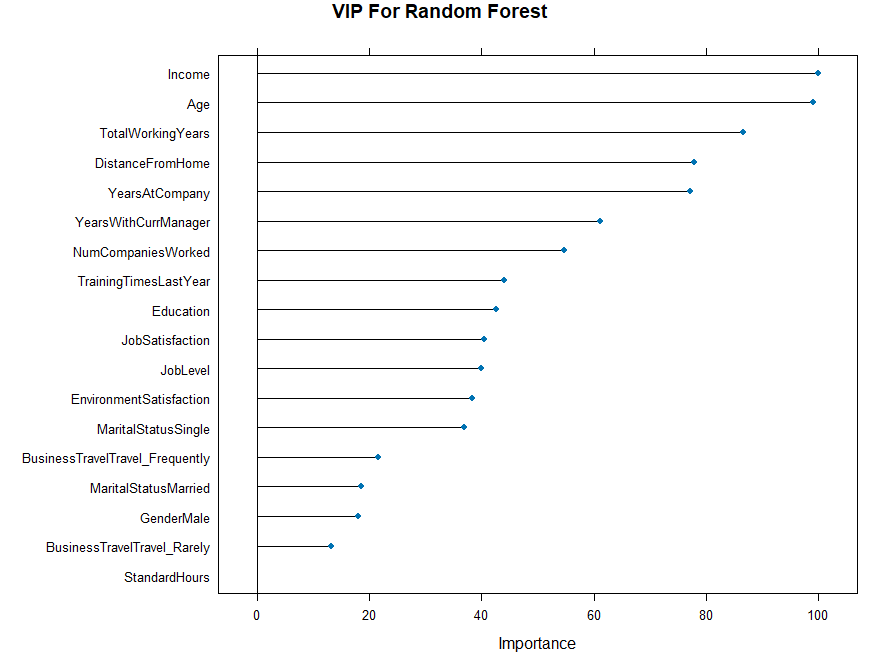
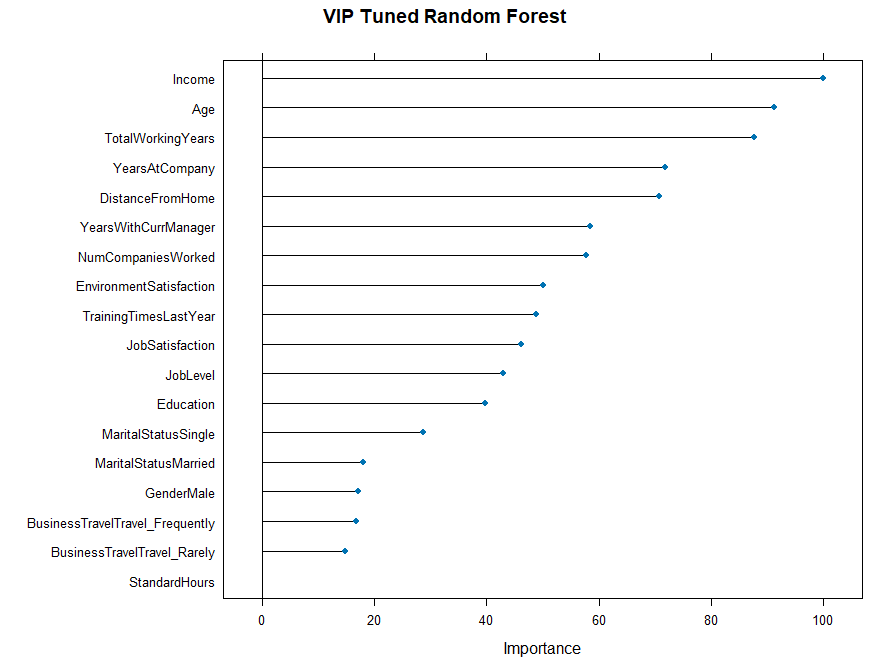
Taking action on these insights, while keeping track of how things are progressing, will not only help Canterra keep costs down but also position the company as a place employees want to stay. This approach will lead to better productivity, stronger relationships with clients, and a much more stable future for Canterra in an increasingly competitive job market.

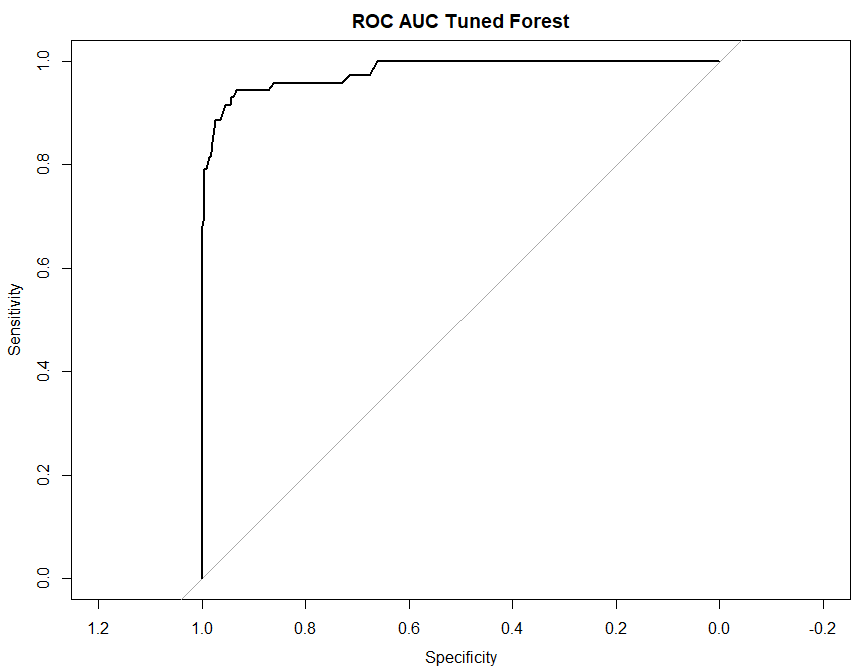
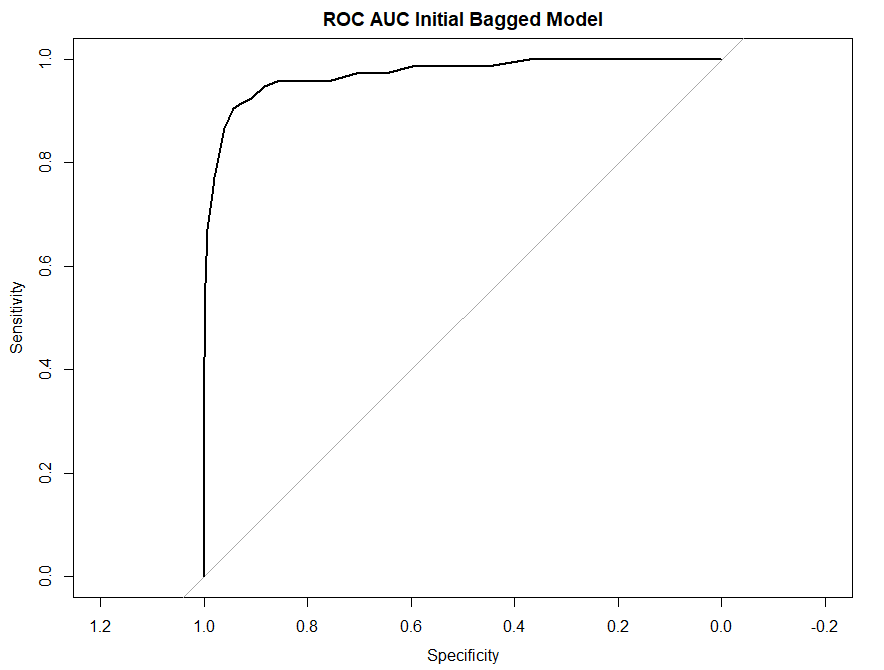
1. **Appendix**

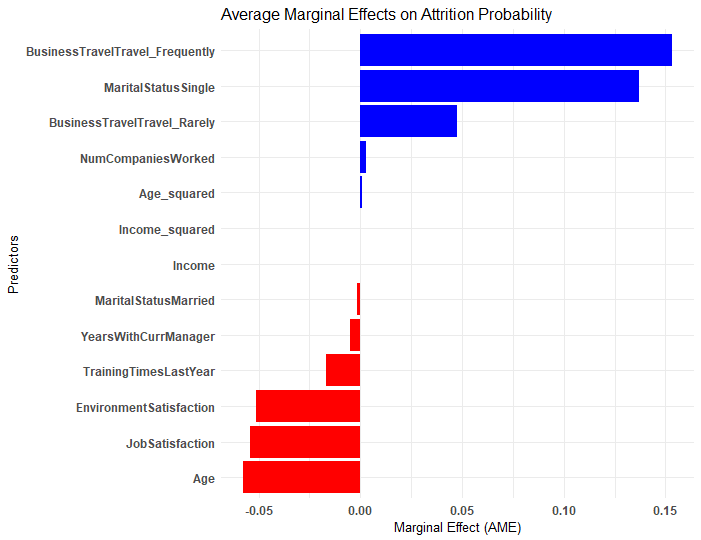
Appendix A. Income Distribution by Attrition Appendix B. Business Travel by —--------------------------------------------------------------------------------- -Attrition

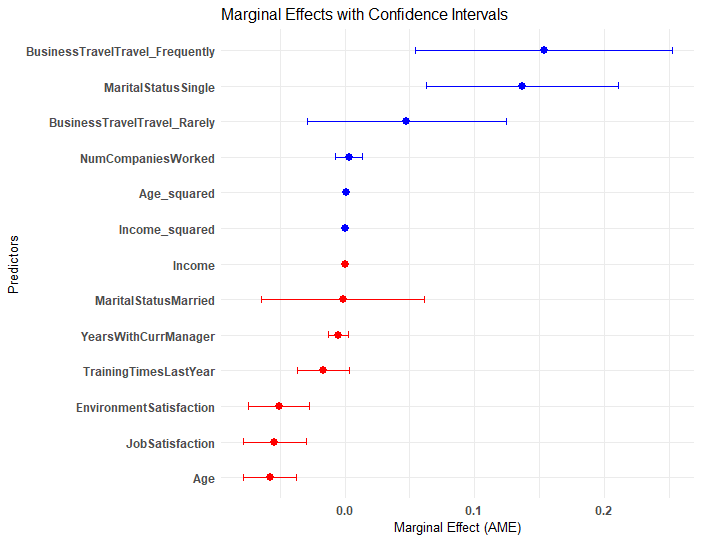
Appendix C. Age Distribution by Attrition Appendix D. Logistic Regression Curve ROC —----------------------------------------------------------------------------------------Curve —----------------------------------------------------------------------------------- 

Appendix E. Bagged Tree Model Variable Appendix F. Adjusted Bagged Model VIP —-------------------------------------------------------------------------------Importance—-----------------------------------------------------------------------------

Appendix G.VIP Random Forest Appendix H. VIP Tuned Random Forest

Appendix I. ROC AUC Bagged Model Appendix J: ROC Forest Tuned 

Appendix K: Avg. Marginal Effects Appendix L: Marginal Effects 



------------- ------

**Bibliography**

Allcot, D. (2022, April 20). Half of US Employees Would Leave Their Job for This One Reason, But Will Stay for These Benefits | GOBankingRates. GOBankingRates; GOBankingRates. https://www.gobankingrates.com/money/jobs/half-employees-would-leave-job-for-this-reason/?

European Union. (2016, April 17). Regulation - 2016/679 - EN - gdpr - EUR-Lex. European Union. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Federal Trade Commission. (2013, July 19). Fair Credit Reporting Act | Federal Trade Commission. Federal Trade Commission. https://www.ftc.gov/legal-library/browse/statutes/fair-credit-reporting-act

Hickman, A. (2020, January 29). What “Meaningful Feedback” Means to Millennials. Gallup.Com; Gallup. https://www.gallup.com/workplace/284081/meaningful-feedback-means-millennials.aspx#:~:text=Gallup%20research%20shows%20that%2087,every%20aspect%20of%20your%20workplace.

King, S. (2022). Why Do Employees Quit? Growth Force; GrowthForce. https://www.growthforce.com/blog/why-do-employees-quit?

Lovich, D., & Sargeant, R. (2023, August 15). Making Flexible Working Models Work | BCG. BCG Global; BCG Global. https://www.bcg.com/publications/2023/flexible-working-models?

Montgomery, A. (2018, August 1). 85% of Global Employees Are Disengaged: How to Build Employee Engagement - Intelligent Management -. Intelligent Management. https://intelligentmanagement.ws/85-global-employees-disengaged-how-to-build-employee-engagement/?

SHRM. (2024). The State of Global Workplace Culture. SHRM. https://www.shrm.org/content/dam/en/shrm/research/the-state-of-global-workplace-culture-2024.pdf

Sparks, S. (2023, April 5). 10 New Job Seeker Stats on Workplace Flexibility | GoCo.io. GoCo.Io; GoCo. https://www.goco.io/blog/workplace-flexibility-statistics?

State of California Department of Justice. (2018, October 15). California Consumer Privacy Act (CCPA) | State of California - Department of Justice - Office of the Attorney General. State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

US EEOC. (1964). Title VII of the Civil Rights Act of 1964. U.S. Equal Employment Opportunity Commission. https://www.eeoc.gov/statutes/title-vii-civil-rights-act-1964

US EEOC. (1967). Age Discrimination in Employment Act of 1967. US Equal Employment Opportunity Commission. https://www.eeoc.gov/statutes/age-discrimination-employment-act-1967

--------------------------------------------------