**Alberto Gell**

**Dat-430**

**Baseline Establishment**

The HR attrition dataset provides insight into the factors that contribute to employees leaving the company and how we can predict who is at risk. The dataset includes key features such as Attrition-Label, JobRole, MonthlyIncome, JobSatisfaction, and YearsAtCompany.

In the dataset, Attrition-Label indicates whether an employee left the company (Yes/No), and Attrition\_Binary represents this information in binary format (1 for Yes, 0 for No). Below is a screenshot that shows the baseline for the analysis, focusing on these key features:

A graph with red and blue bars

Description automatically generated

**Exploratory Data Analysis (EDA)**

I explored the relationship between various factors and the relationship to attrition. Below are the key insights gathered from this exploratory data analysis:

* **Attrition by Job Role**: The Sales Representative role has the highest attrition rate at 45.5%, while Research Scientist has the lowest at 27.7%.
* **Job Satisfaction and Attrition**: Employees with lower job satisfaction scores are more likely to leave, especially in Sales and Human Resources roles.
* **Monthly Income and Attrition**: Despite higher salaries, Sales Representatives still have higher attrition rates, suggesting that factors beyond income, such as job satisfaction and workload, contribute to turnover.

**Visualization:**

Using Plotly, I created a bar chart to visualize Attrition by Job Role.

A screenshot of a computer

Description automatically generated

**Job Satisfaction vs. Training and Performance Rating vs. Training**

In addition to analyzing attrition, I also investigated the effect of Training on JobSatisfaction and PerformanceRating using linear regression models.

**Job Satisfaction vs. Training**

* Intercept (Baseline Job Satisfaction): 2.21
* Coefficient (Impact of Training): 0.162

These outputs from the linear regression model show that for each additional unit of training, job satisfaction increases by 0.162 units. This highlights the critical role of training in improving employee satisfaction, which can directly impact retention.

**Performance Rating vs. Training**

* Intercept (Baseline Performance Rating): 2.49
* Coefficient (Impact of Training): 0.114

These outputs show that for every additional unit of training, performance rating increases by 0.114 units. This suggests that employees who receive more training perform better and are more likely to stay engaged with their work.

**Comparing factors between job roles with the highest attrition and lowest attrition**

I compared the Sales Representative role (highest attrition) and the Research Scientist role (lowest attrition) across key features such as MonthlyIncome, YearsAtCompany, JobSatisfaction, and PerformanceRating. Here are the key insights

A screenshot of a computer

Description automatically generated

** Monthly Income: Sales Representatives** earn a higher monthly income ($7,986) compared to **Research Scientists** ($6,858), yet they still exhibit a higher attrition rate. This suggests that income alone is not enough to retain employees in high-stress roles.

** Years at Company: Sales Representatives** tend to stay with the company longer (16.57 years) than **Research Scientists** (10.76 years). Despite their longer tenure, the high attrition rate for Sales Representatives could indicate burnout or dissatisfaction with job growth opportunities.

 **Job Satisfaction: Job Satisfaction** is lower for **Sales Representatives** (2.61) compared to **Research Scientists** (2.72), reinforcing the need for initiatives that address employee satisfaction in high-attrition roles.

** Performance Rating: Performance ratings** are slightly lower for **Sales Representatives** (2.70) compared to **Research Scientists** (2.95), which might contribute to higher attrition in the Sales role.

**Feature Selection and Engineering for Predictive Model**

Based on the EDA, I selected the following features for the predictive model:

* JobRole
* MonthlyIncome
* JobSatisfaction
* YearsAtCompany
* OverTime

To prepare the data for the predictive models, I imported the necessary python libraries and performed one-hot encoding for categorical variables JobRole and OverTime

A screenshot of a computer

Description automatically generated

**Predictive Models**

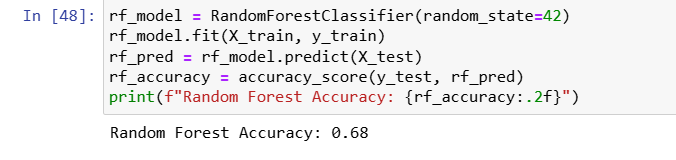
To predict employee attrition, I applied two models: Logistic Regression and Random Forest Classifier

**Logistic Regression**

**A screenshot of a computer code

Description automatically generated**

**Random Forest Classifier**

****

**Conclusion And Recommendations**

The primary objective of this analysis was to identify key factors driving employee attrition and to predict which employees are at the highest risk of leaving the company. After performing exploratory data analysis (EDA) and applying predictive models such as Logistic Regression and Random Forest, several important insights were uncovered.

**Key Findings:**

**Job Satisfaction and Performance vs. Training**

A linear regression model was developed to predict job satisfaction and performance ratings based on training data. The results showed that:

* For each additional unit of training, job satisfaction increases by 0.162 units. This indicates that employee development and learning opportunities play a critical role in boosting overall satisfaction levels.
* For each additional unit of training, performance rating increases by 0.114 units. This highlights that employees who undergo more training not only feel more satisfied but also perform better.

**Attrition by Job Role**:

* Sales Representatives have the highest attrition rate at 45.5%. Despite having one of the highest average monthly incomes ($7,986), their attrition rate is significantly higher than other roles. This suggests that income alone does not prevent turnover.
* Research Scientists, on the other hand, have the lowest attrition rate at 27.7%, even with lower average monthly incomes ($6,858). Their higher JobSatisfaction (2.72 vs. 2.61 for Sales Representatives) and PerformanceRating (2.95 vs. 2.70) seem to play a larger role in retaining them.

**Job Satisfaction and Overtime**:

* Lower JobSatisfaction is a significant predictor of attrition. Employees who report lower satisfaction are more likely to leave, particularly in roles such as Sales and HumanResources.
* Overtime was another critical factor. Employees who work overtime are more likely to leave the company, potentially due to burnout and stress.

**Predictive Models**:

* The Random Forest Classifier performed better than the Logistic Regression model, with an accuracy of 68% compared to 64%. While these accuracy rates are moderate, they indicate that the model can reasonably predict which employees are at risk of leaving.
* The Random Forest model's ability to capture non-linear relationships between features such as JobSatisfaction, MonthlyIncome, and Overtime suggests that a more complex combination of factors, rather than a single variable, drives employee attrition.

The analysis demonstrates that **JobSatisfaction, Training, OverTime**, and **Workload** are all key contributors to employee attrition. While **MonthlyIncome** is important, it does not play as significant a role as factors like **JobSatisfaction** and the amount of **Training** employees receive. Even employees with higher salaries, such as **Sales Representatives**, have higher attrition rates due to lower job satisfaction and higher overtime demands. Also, the results from the training impact analysis show that **training programs** directly improve both **job satisfaction** and **performance**. This means that by offering more professional development opportunities, HR can not only reduce turnover but also enhance overall employee productivity.

**Recommendations**

**Enhance Job Satisfaction Through Training:**

* Targeted Training Programs: Since training positively impacts job satisfaction and performance, HR should prioritize training programs tailored to specific roles, especially those with high attrition like Sales Representatives and Human Resources. The data shows that an increase in training leads to better satisfaction, which is crucial for retaining employees in high-stress roles.
* Career Development Opportunities: Establish clear development and promotion paths for employees, coupled with personalized training plans. This will keep employees engaged and satisfied with their career progression.

**Manage Workload and Overtime**:

* Work-Life Balance Initiatives: As identified, OverTime plays a significant role in attrition. Implement flexible work schedules, limit mandatory overtime, and promote a culture of work-life balance to mitigate burnout and reduce turnover.
* Monitor Workload Distribution: Regularly assess the workload of employees, especially in high-attrition roles like Sales Representatives. Redistribute tasks where necessary and consider hiring additional staff to alleviate pressure on existing employees.

**Performance Recognition Programs:**

* Performance-Based Incentives: Implement performance recognition programs that reward employees who show improvements in performance after completing training. This will not only motivate employees to participate in training but will also boost JobSatisfaction and overall PerformanceRatings.