Aziz Bohra
Jerin Easo Thomas

Employee Attrition

Statement of goals

The aim of this study is to identify the factors that contribute to employee attrition among IBM staff. We are attempting to answer the question of **why employees quit IBM positions and what influenced them to take that decision**. The organization may utilize this information to make the required adjustments to lower employee turnover and retain talented individuals, which will ultimately improve the performance and reputation of the business. Other businesses in the same industry can enhance their talent retention strategy and increase employee satisfaction by having a better understanding of the elements that contribute to employee churn.

Description of your data

The IBM HR Analytics Employee Attrition & Performance dataset obtained from Kaggle is available at the provided link.
https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset
This dataset contains information about a company's employees, such as job roles, performance assessments, and job satisfaction levels. The dataset comprises **35 columns and 1,470 rows**. Some of the features that were investigated are listed in the table below:

Feature	**Type**	**Description**
Age	Numerical	Employee's age
Monthly Income	Numerical	Employee's monthly income
No of Companies	Numerical	No. of Companies the employee has worked till now
Years with Curr Manager	Numerical	No. of years worked with the Current Manager
Total Working Years	Numerical	Total number of years the employee has worked
Years at Company	Numerical	Number of years the employee has worked at the company
Education	Ordinal	Education level (Below College - PhD)
Performance Level	Ordinal	Performance Rating (Low - Outstanding)
Job Satisfaction	Ordinal	Employee's level of job satisfaction (Low - Very High)
Job Level	Ordinal	Designation (Entry Level - Executive Level)

Environment Satisfaction	Ordinal	Workplace Environment (Low - Very High)
Job Involvement	Ordinal	Employee's level of involvement in their job (Low - Very High)
Work Life Balance	Ordinal	Employee's level of satisfaction with their work-life balance (Bad - Best')
Attrition	Categorical	Whether the employee has left the company (Yes/No)
Gender	Categorical	Sex of the Employee (Male/Female)
Business Travel	Categorical	Frequency of business travel (Non-Travel, Travel_Rarely, Travel_Frequently)
Department	Categorical	HR/R&D/Sales
Education Field	Categorical	Type of Discipline
Job Role	Categorical	Employee's job role (Sales Executive, Research Scientist, Research Director, Human Resources, etc)
Marital Status	Categorical	Married/Unmarried

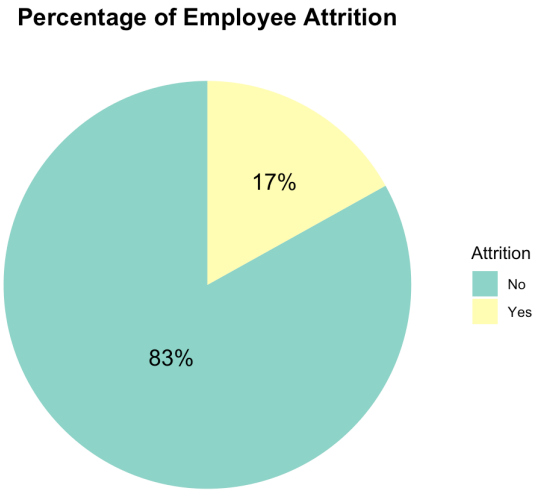

Fig 1: Pie Chart for percentage of employee attrition.

The distribution of the target variable ‘Attrition’ is depicted in Fig 1, which says that this dataset is imbalanced and we have to be cautious as this might create a bias in our predictive analysis.

Answering our questions:

Obtaining an answer to the research question that forms the foundation of this study may not be a straightforward process and may require more than simply examining the data. Our approach involved a systematic step-by-step process to approach the answers we sought, and we drew inferences based on the results we obtained. To avoid exploring dataset features aimlessly, we conducted research to identify the most relevant features and used concrete evidence to guide our analysis.

We observed that Job satisfaction is a significant factor that influences an employee's decision to leave a company. It is influenced by aspects like, their contentment with work environment, pay, work-life balance, and career advancement opportunities. Analyzing job satisfaction would provide us insights into the causes of employee departure and help organizations to identify and resolve issues, improve retention of valuable employees.

Based on the above we asked ourselves: *How important is Job Satisfaction with respect to Attrition?*

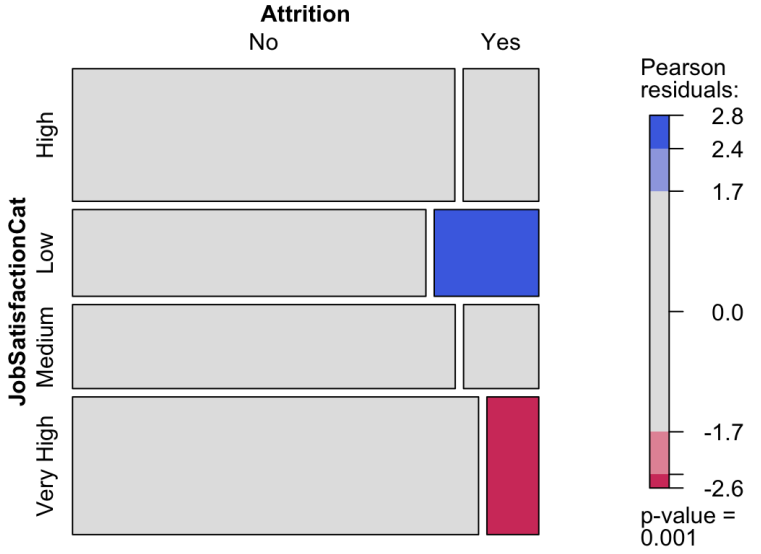

Fig 2.1: Mosaic Plot for Job Satisfaction and Attrition

In the beginning we plotted a mosaic plot to get an understanding of attrition based on an employee's job satisfaction level. We found that Low and Very High are the categories in which attrition has an influence. Moving forward, we aimed to delve deeper into identification of relationships between other variables and attrition, and thus performed the following visualizations with Monthly Income as a key factor.

Is there any relationship between Attrition and Monthly Income?

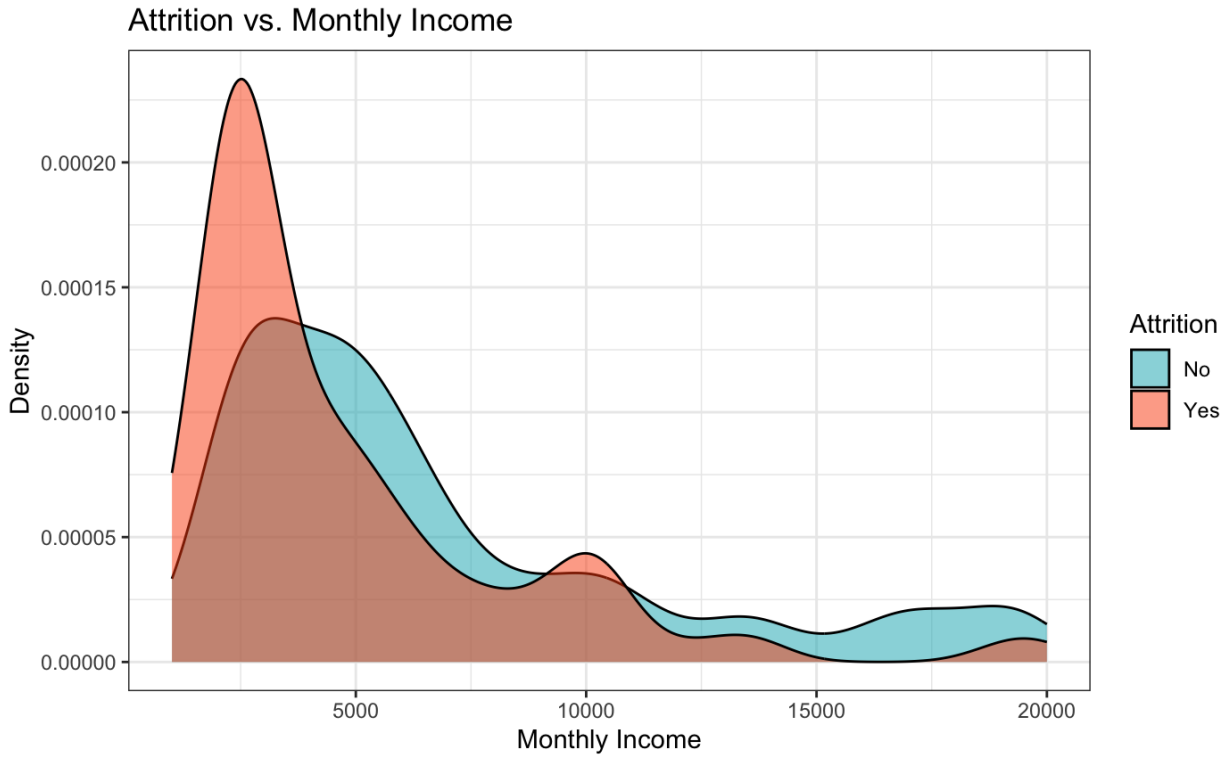

Fig 2.2 Line Graph (Income, Job Satisfaction, Attrition)

With Fig. 2.2 in place, we can clearly see that the majority of the data points are clustered below the monthly income of $10000, with the highest peak of attrition seen with monthly income less than $5000, clearly indicating that with less monthly income, there is a high chance an employee will leave the organization, whereas if monthly salary is greater than $15000, an employee leaving the organization is less likely. The second attrition peak may be seen at monthly income of $10,000, indicating that despite better pay, some employees quit the company for other reasons, which will be addressed in the following analysis.

Does the Department contribute any insights in the analysis of attrition?

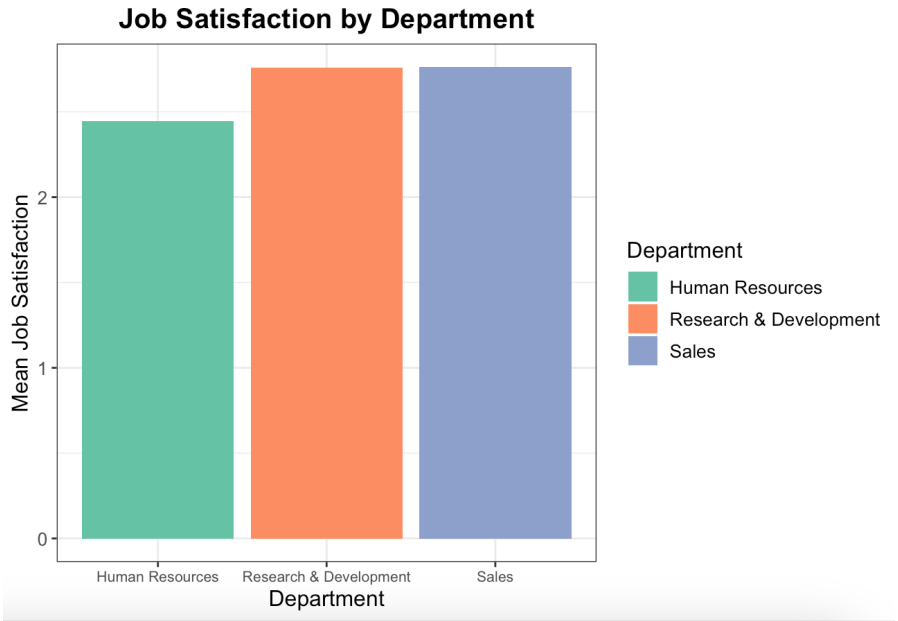

Fig 2.3: Histogram for Job Satisfaction by Department

On observing Fig 2.3 we can tell that the Research & Development and Sales department have almost the same average level of Job Satisfaction. Although this graph doesn't reveal anything as such, it is just developed as a context for the next plot.

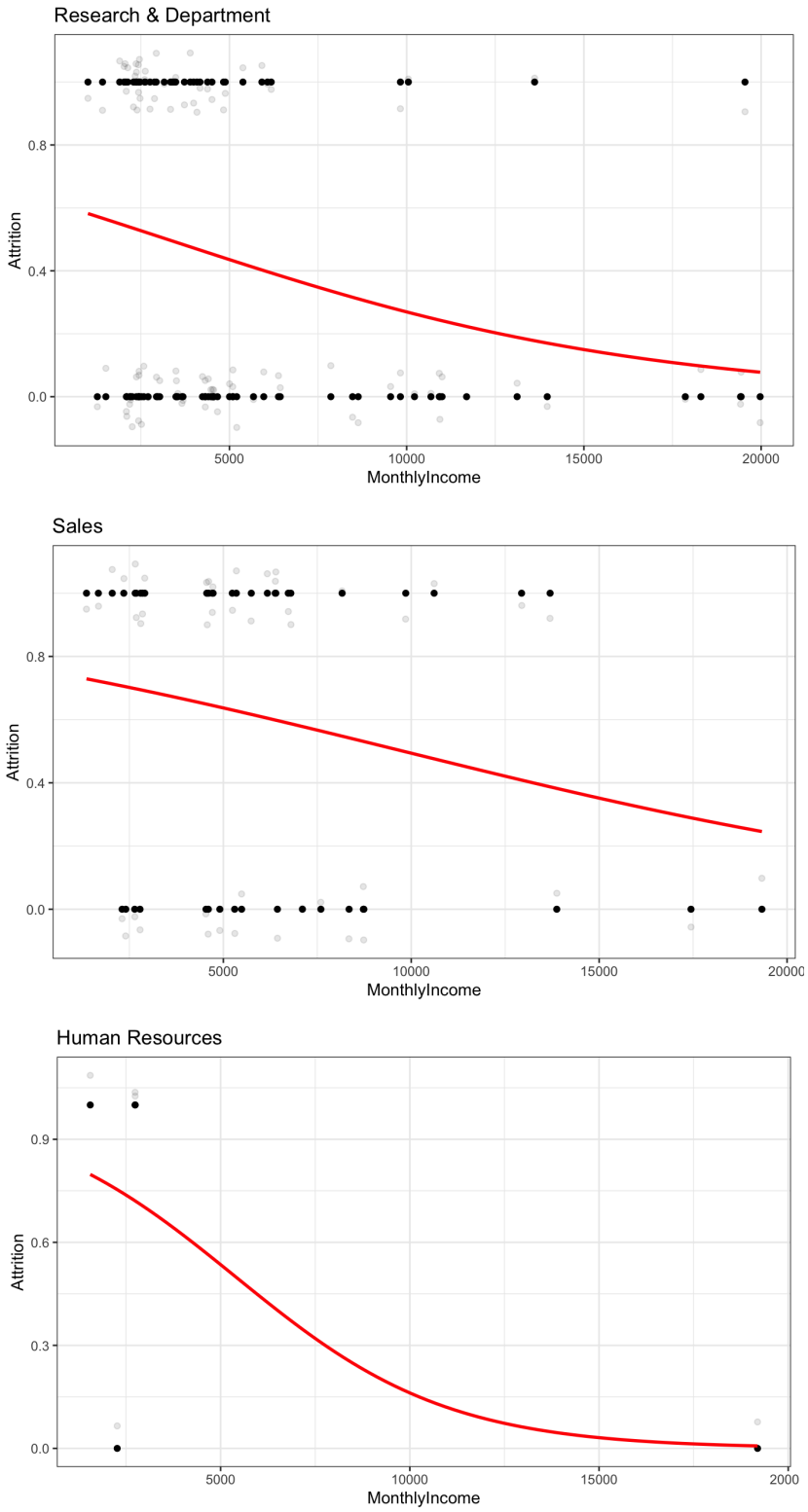

Fig 2.4 Probability of Attrition based on Monthly Income (Job Satisfaction=High)

From this Fig 2.4 we have numerous observations which will play a crucial role in our analysis. Firstly, it is clear that the Research & Development department has many data points and a clear indication that the dataset is imbalanced. Next, we observe that the probability of Attrition is high across all the departments when the Monthly Income is low. One thing which is interesting to look at is that the plot for the Sales Department is almost linear.

Now after exhaustive EDA on these variables we will switch gears to other variables ahead.

Are there any relation between Attrition, Marital Status and Overtime?

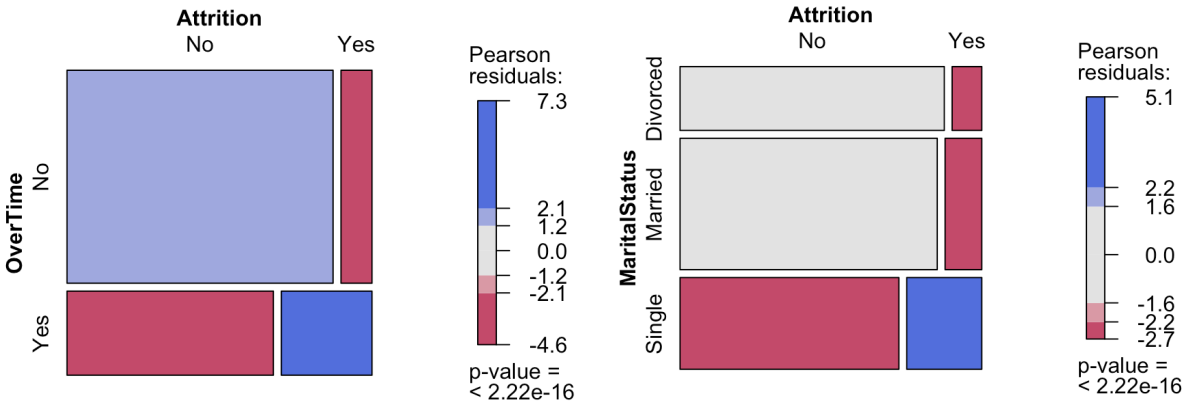

Fig 3: Mosaic Plot for OverTime, Marital Status and Attrition

Figure 3 depicts the mosaic plot for OverTime and Marital Status with Attrition. On the left hand side, we can see that the probability of Attrition is high when the employee is asked to do over-time which makes sense. Coming on the right hand side, we see that there is a high probability of Attrition when the employees are signal and this might be because these individuals will be ambitious and like to switch companies which result in their career growth. On the other hand, we can observe that Married employees are not really inclined to leave their jobs because of their families financial security.

Relation between Attrition vs Work Life Balance?

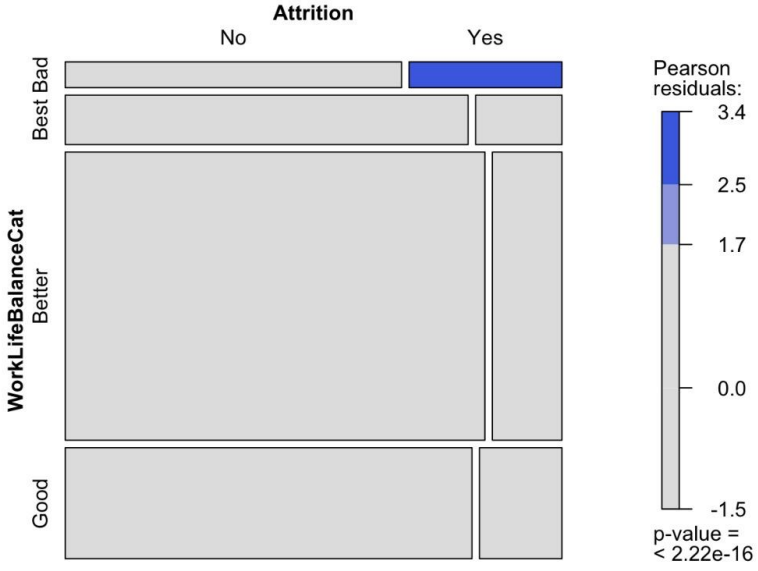

Fig 4: Mosaic plot for Work Life Balance vs Attrition

WorkLifeBalance was formerly an ordinal type with values ranging from 1 to 4, which was transformed to 1: 'Bad,' 2: 'Good,' 3: 'Better,' and 4: 'Best'. As seen in Fig 4, an employee with poor work-life balance is more likely to leave the organization, thus, one initiative that could be taken to prevent attrition would be to take various steps, such as reducing overtime and engaging in other outside activities, which could also increase Job Satisfaction.

Model Building & Evaluation:

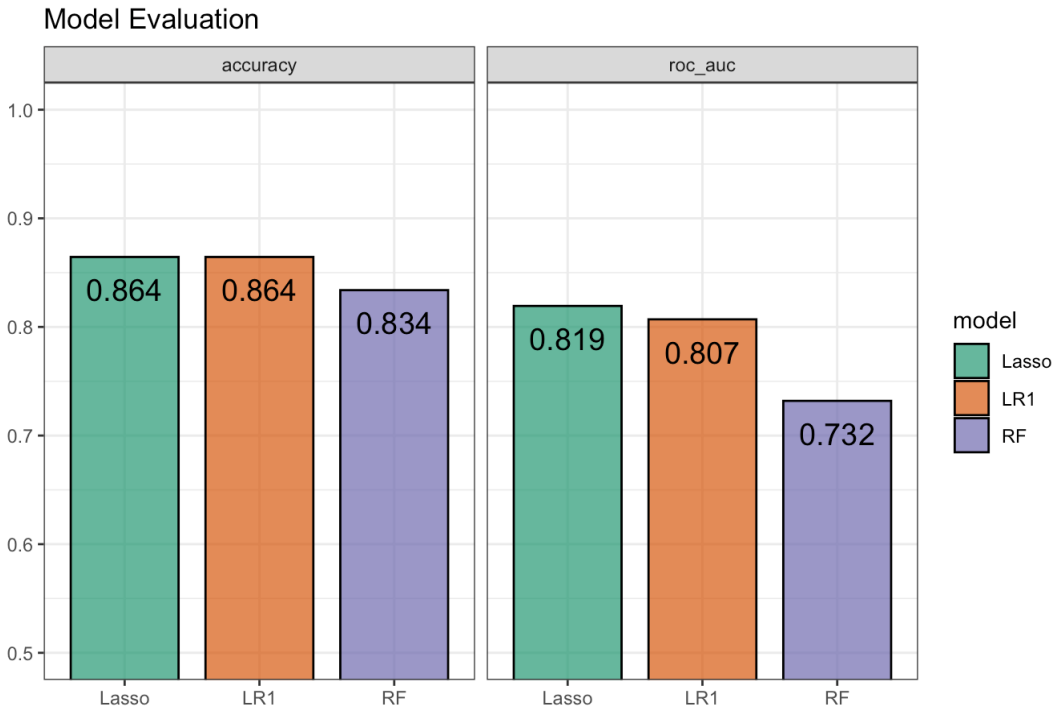

Fig 5.1: Model Evaluation

We moved on to the model construction phase once we had a decent idea of which features influence Attrition. Initial observations suggest that the Lasso Regression and Logistic Regression models have an accuracy of 86.4%, while the Random Forest model has an accuracy of 83.1%, as seen on the left side of Fig 5.1. So, how do we choose between Lasso Regression and Logistic Regression? To determine this, we calculated and plotted the ROC accuracy, which is shown on the right hand side of Fig 5.1, and reveals that the Lasso Regression has the best ROC accuracy (82%).

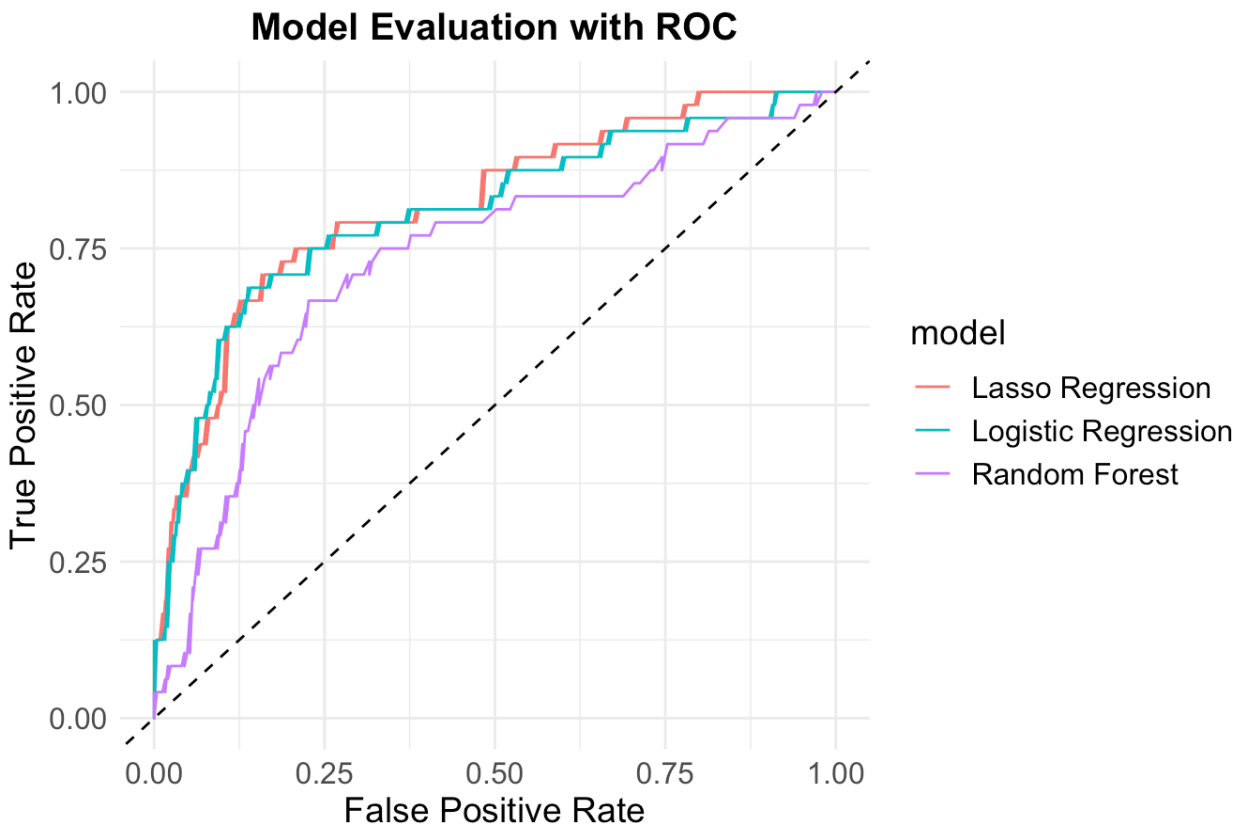

Fig 5.2: Model Evaluation with ROC

The ROC curves for all models are shown in Fig 5.2, and it can be seen that for the most part, Lasso Regression has the biggest area under the curve, corroborating our earlier conclusion that Lasso Regression performs the best of the three models.

How do we determine which features have the highest influence over Attrition?

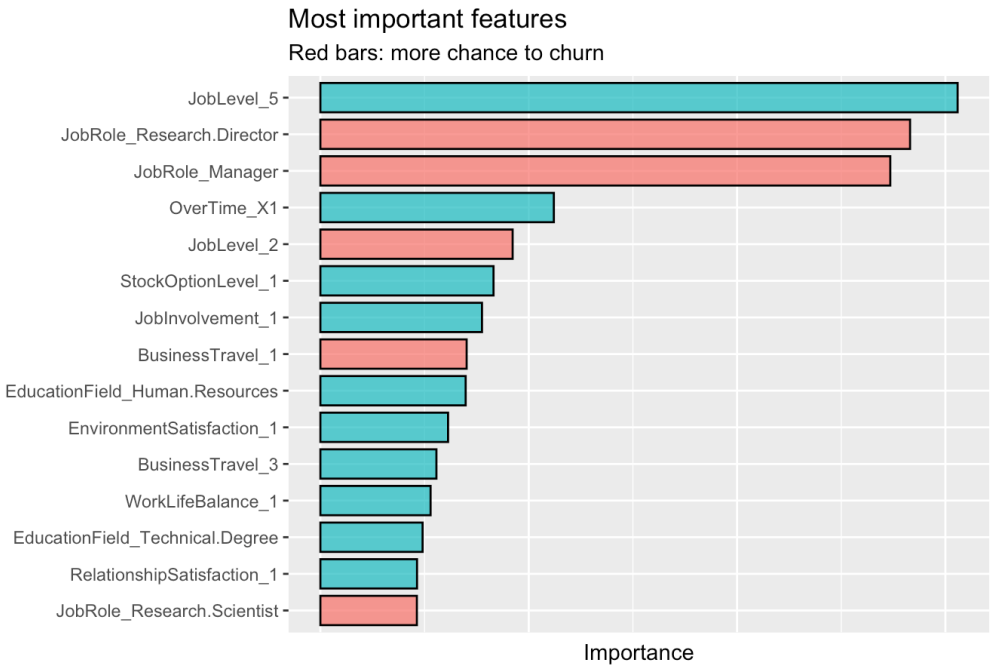

Fig 5.3: Lasso Regression Most Important Features

The top 15 features that have the largest influence on the Attrition obtained using Lasso Regression, which has the best accuracy, are shown in Fig 5.3. Consider an example where the feature job level and value 5 are extremely likely to be attained beyond the age of 45, which is typically the age of retirement; similarly, higher work time leads to higher attrition over time. On the other hand, high-paying professional roles such as research director and manager are less likely to leave the organization.

After exhaustive analysis and modeling, we finally think we have the answer to our prime question which was "***Which factors are associated with employee attrition at IBM?***". The answer is Job Level, Job Role, OverTime, StockOptionLevel, Job Involvement, Business Travel, Education Field, Environment Satisfaction and WorkLife Balance are the features which influence the target variable "Attrition" the most.

This study has major implications for firms looking to retain staff and lower turnover rates. Organizations can take proactive actions to address these challenges and enhance their work environment, remuneration, and career development possibilities by measuring job satisfaction and identifying the variables that contribute to employee leaving. This can lead to increased employee satisfaction and loyalty, as well as higher retention rates. The project's findings can also be used to generate targeted interventions and tactics to reduce attrition and improve employee retention, which can save firms money on recruiting, training, and productivity. This initiative also emphasizes the significance of constantly analyzing and monitoring employee job satisfaction in order to proactively detect and address issues before they lead to significant attrition rates.

Identification of work left to do/limitations

Although our research and models yield good results, we need more evidence before deploying this project in the real work to solve problems which we aimed for. We say this because there are few constraints to this:

1. Because the study is based on a single dataset, it may not be applicable to different sectors or organizations.
2. The survey excludes specific reasons why people leave their positions, such as personal or family concerns, which may or may not be related to job satisfaction or other workplace characteristics.
3. The study does not investigate the impact of external factors on employee attrition, such as the economy or industry trends.
4. The study does not assess the impact of organizational interventions or adjustments to promote work satisfaction and reduce attrition.

The limitations of this project highlight the need for future research that can address the gaps in our understanding of employee attrition. We have researched about the future scope of this application and have following points to propose as the future study:

- Longitudinal study: Tracking employee attrition and job satisfaction over time can provide a deeper understanding of the factors that influence employee turnover and job satisfaction.
- Qualitative research, like interviews or focus groups with current and past employees, can provide a more in-depth understanding of why individuals quit their employment and what aspects contribute to job satisfaction.
- Comparing attrition and job satisfaction rates across industries might provide insight into industry-specific issues that lead to employee turnover.
- Analysis of the impact of organizational interventions or changes to promote job satisfaction and reduce attrition might provide significant information on the effectiveness of such interventions.
- External factors analysis: Examining the impact of external factors on employee attrition and job satisfaction, such as the economy, industry trends, or competition, can provide a larger framework for understanding employee turnover.

APPENDIX

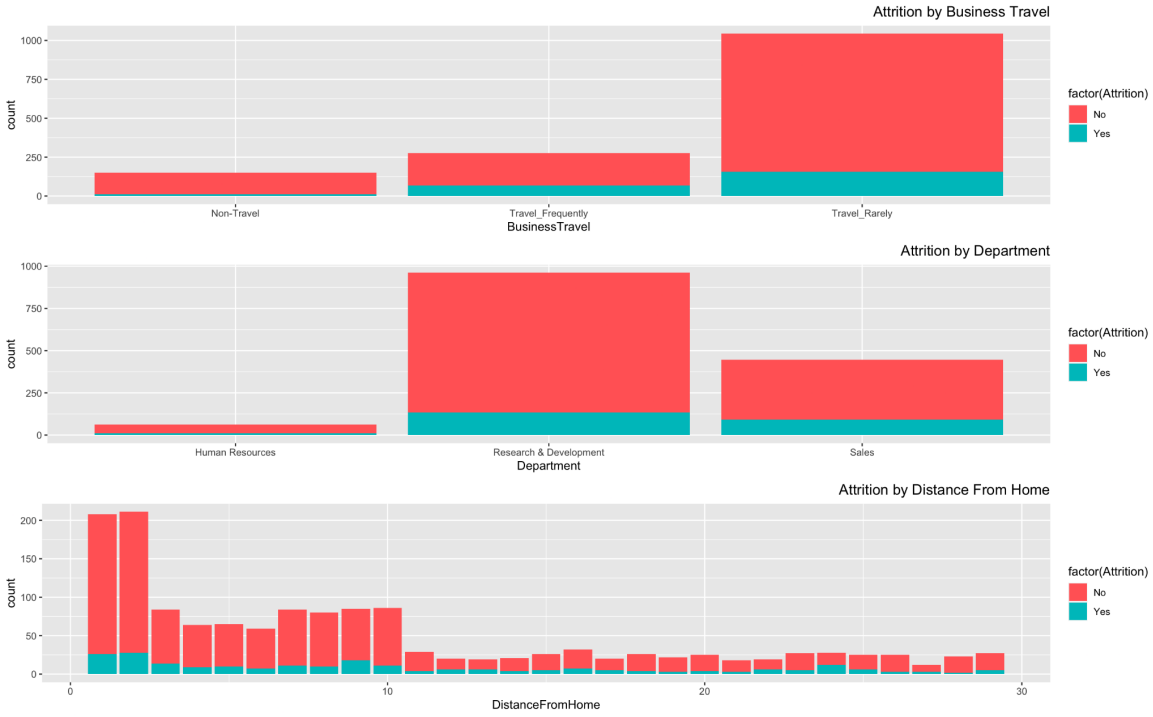

Fig A: Attrition vs Business Travel, Department, Distance From Home

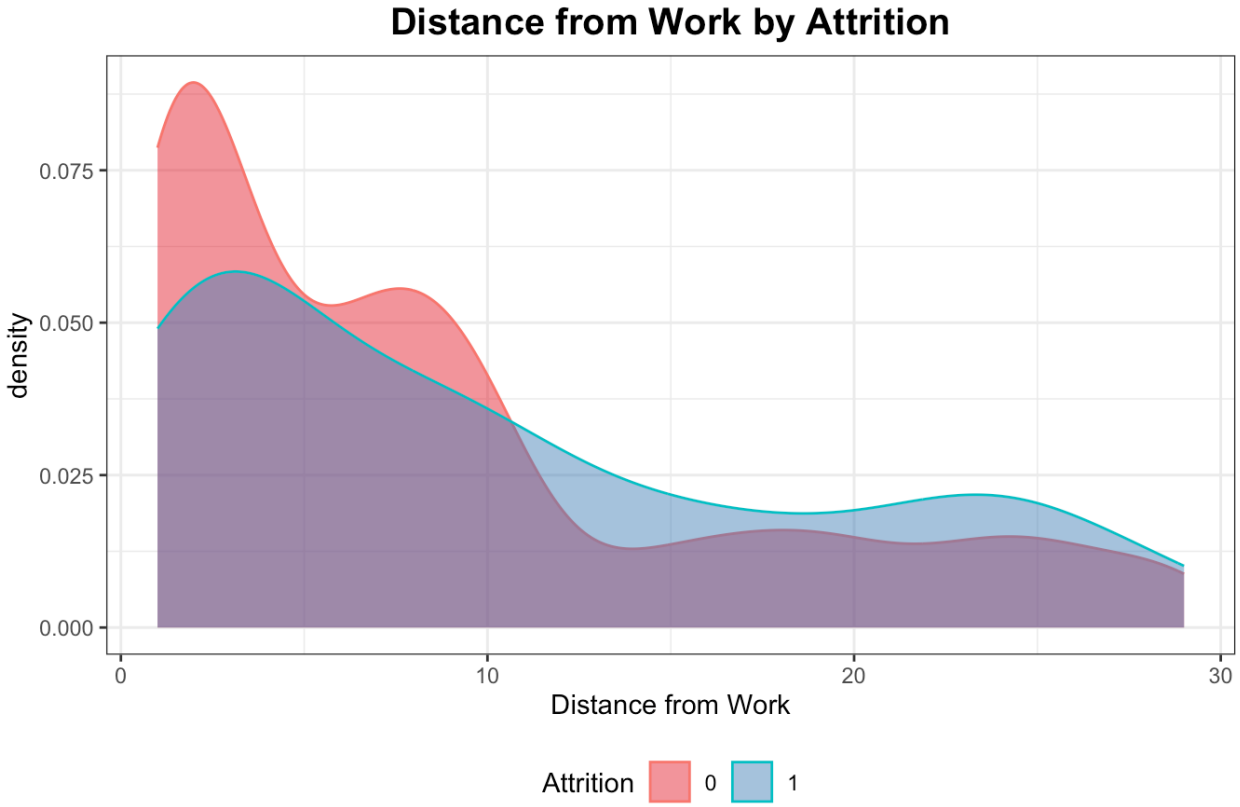

Fig B: Distance from Work by Attrition

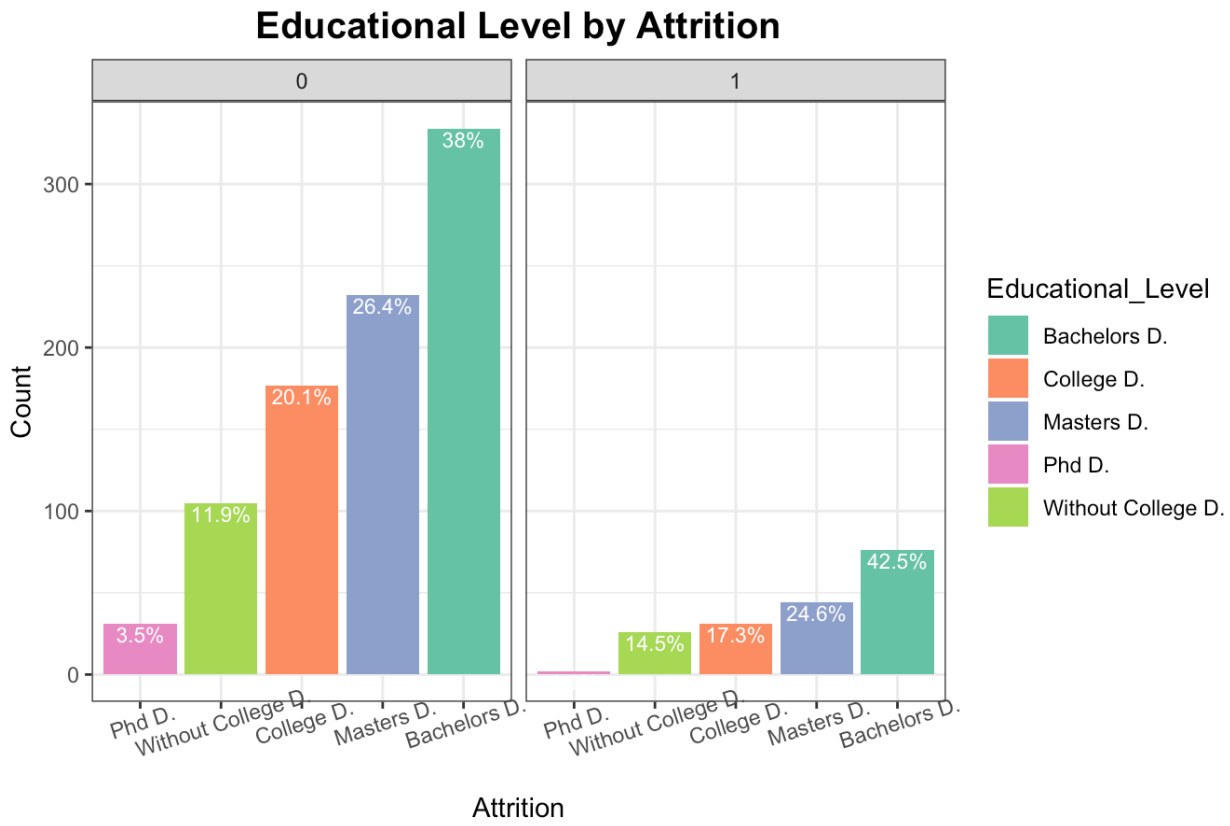

Fig C: Education Level by Attrition

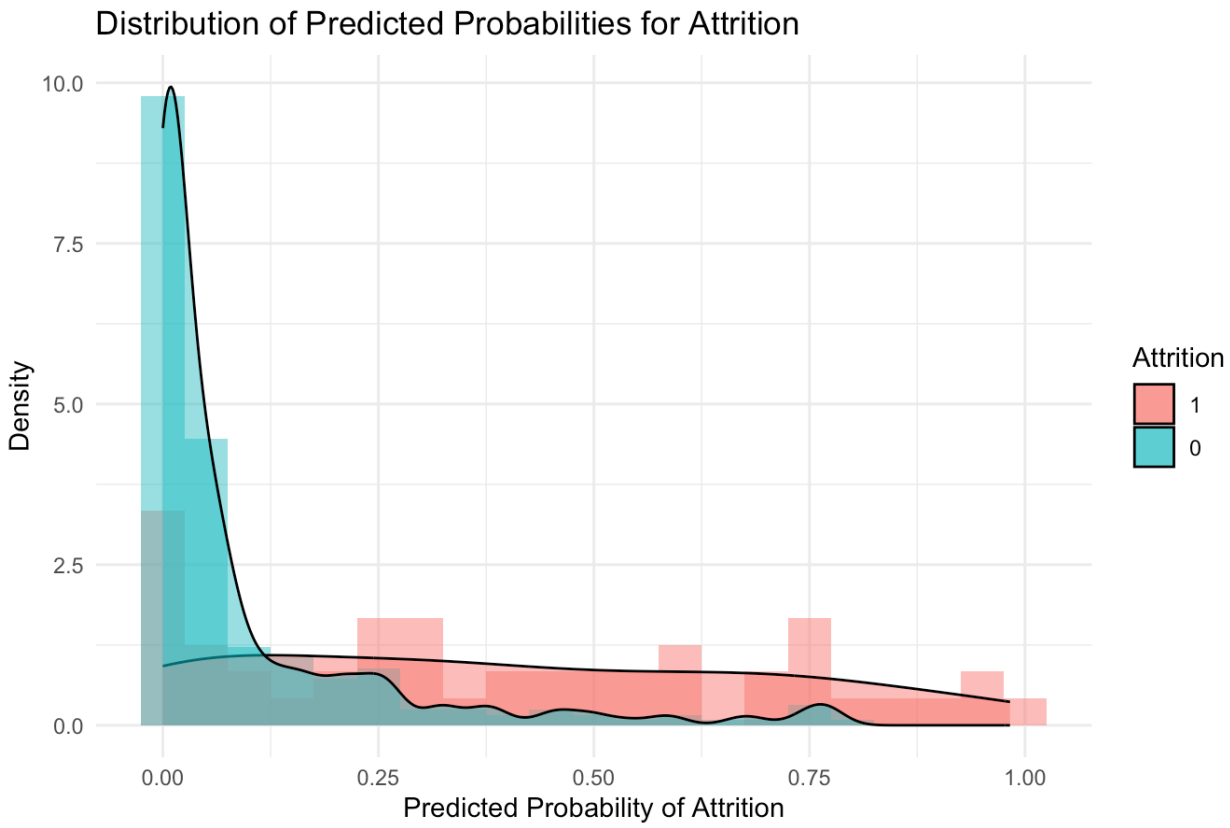

Fig D: Predicted Probabilities for Attrition

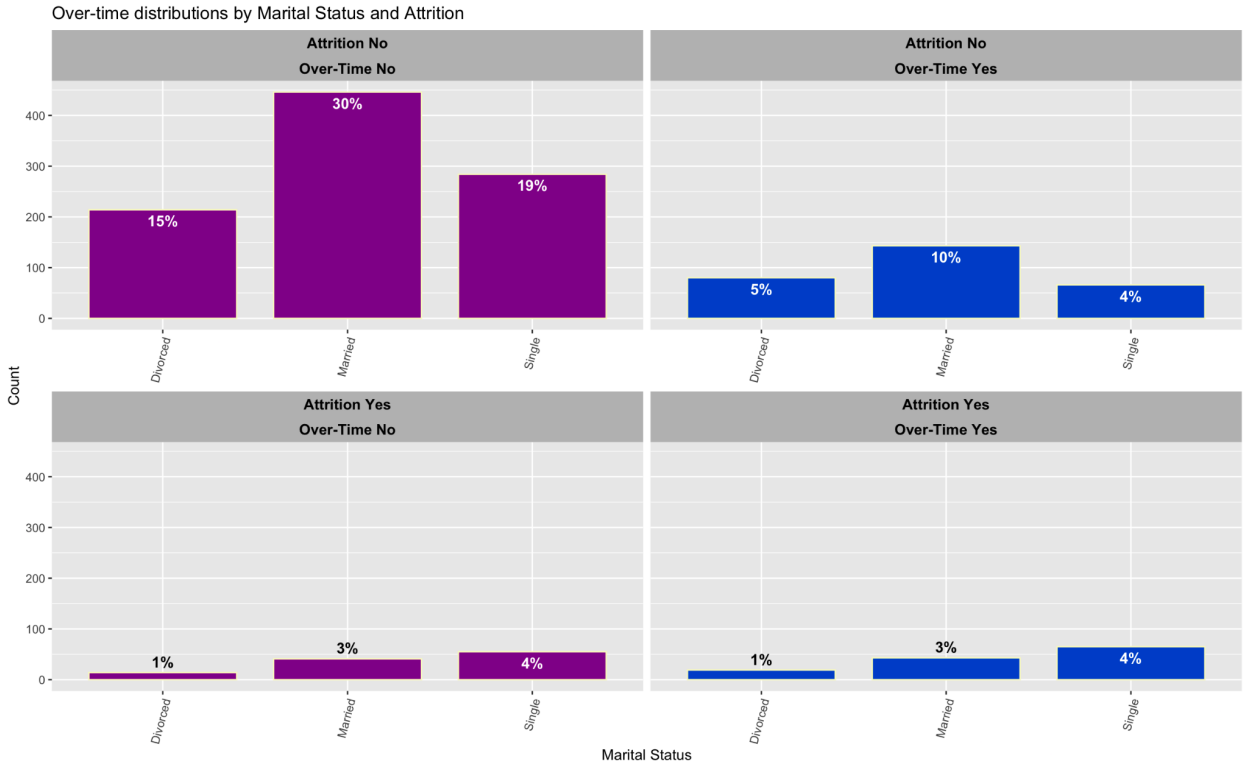

Fig E: Over-time distribution by Marital Status and Attrition

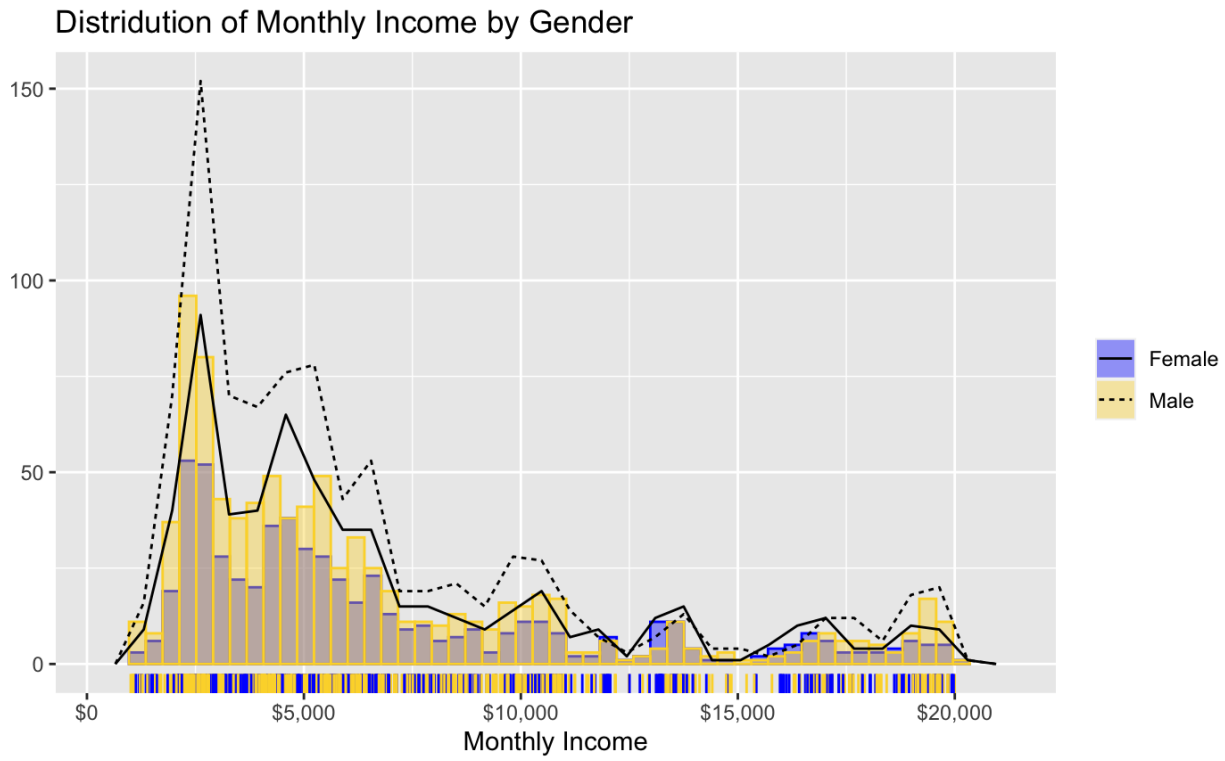

Fig F: Distribution of Monthly Income by Gender

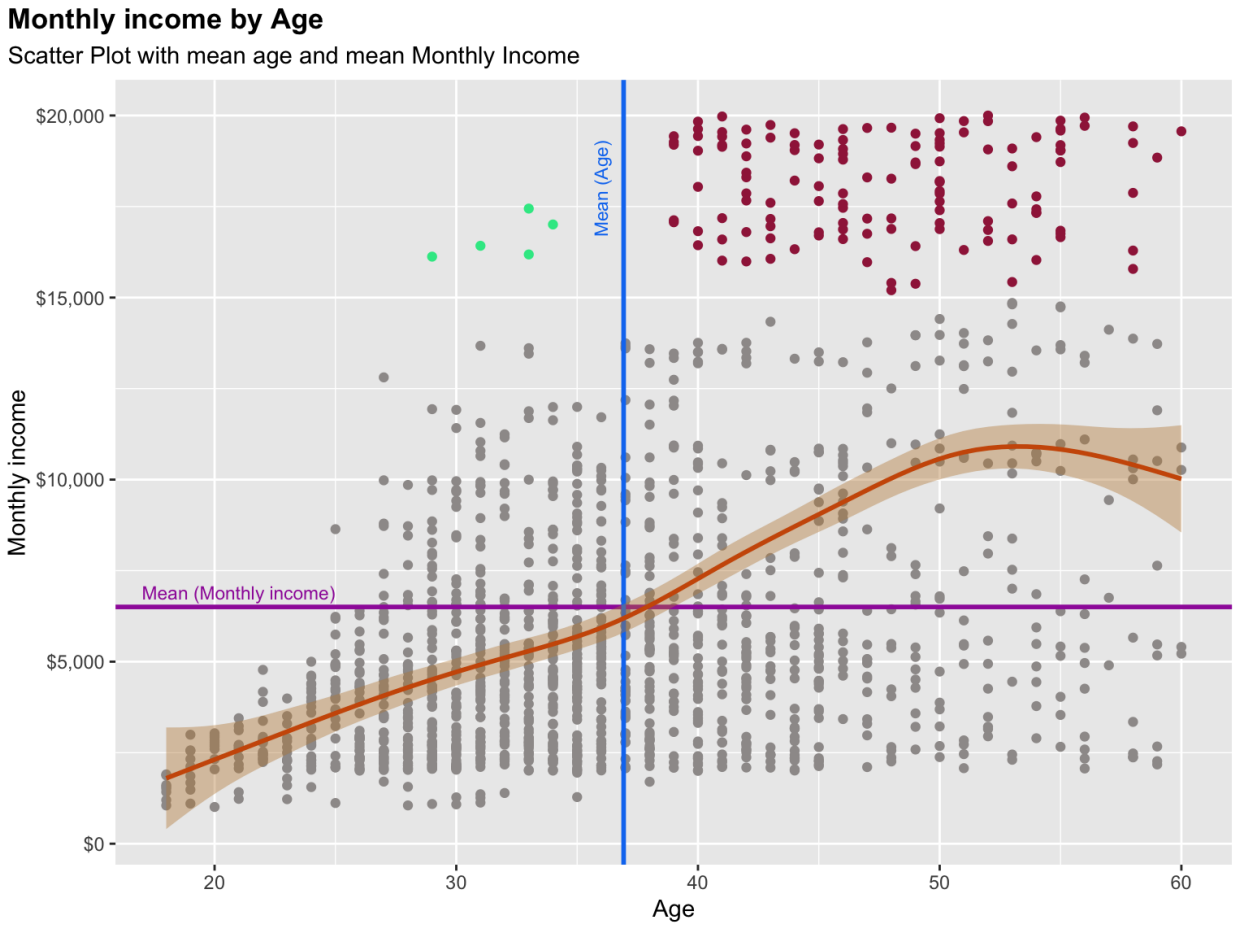

Fig G: Monthly Income by Age