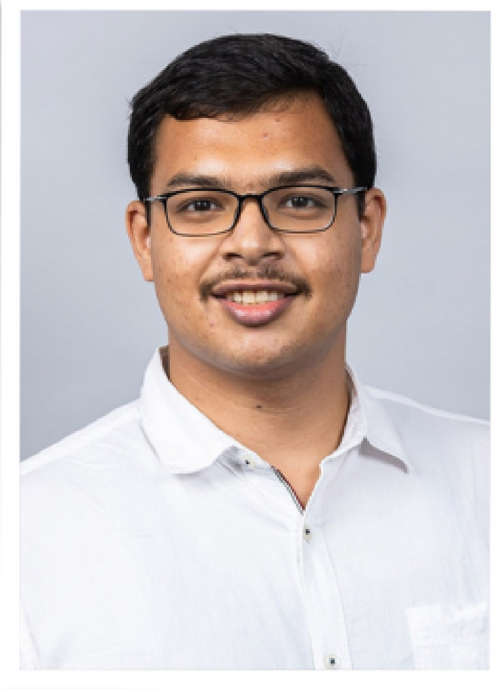
Datatrons
Group 2

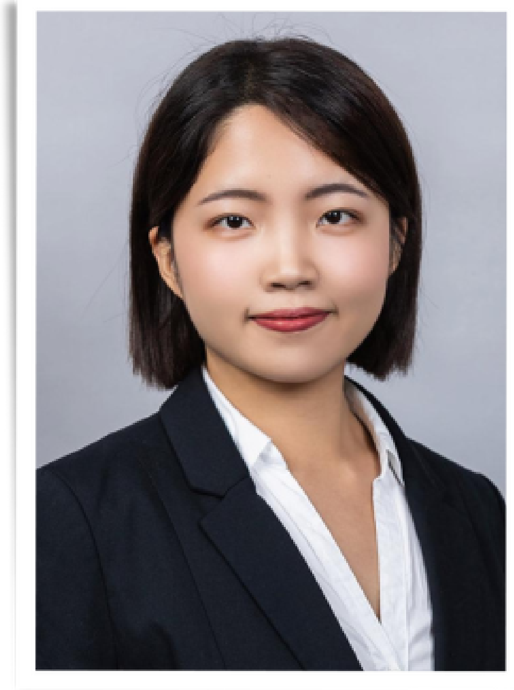
WHAT'S THE VARIABLE?

OUR TEAM

Diverse team with wealth of experience!



DEEP SINGH
DATA ENGINEER



JIA PEI
DATA ANALYST

MANASA
BUSINESS ANALYST

DEEPTHI
STATISTICIAN

VIDIT
QUANT ANALYST

NOUMAN
INTERN

AGENDA

- Introduction
- Dataset
- Pre-processing
- Analysis
- Conclusion

INTRODUCTION

WHAT’S THE VARIABLE?

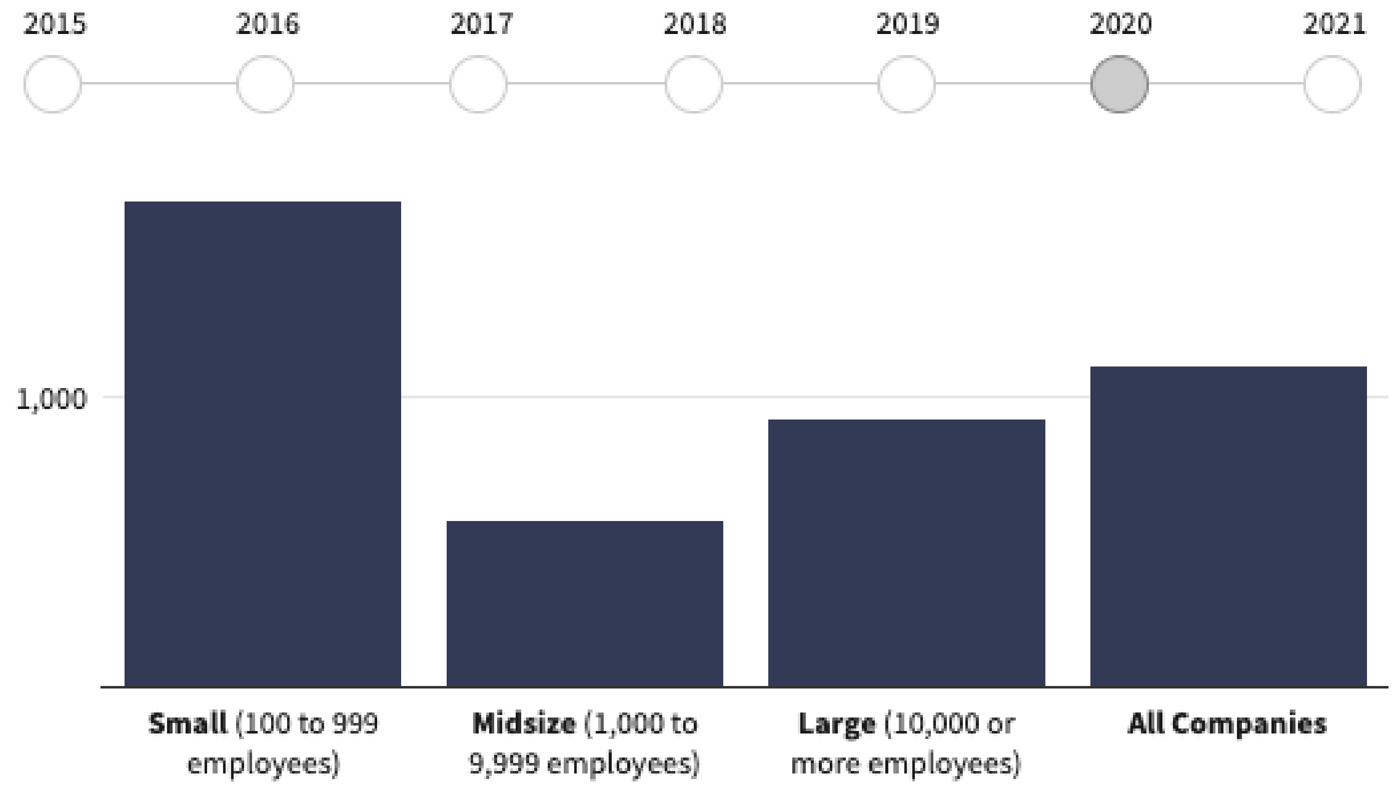
- Why is it important to know the variable?
- Cost of hiring a new employee is more than just their salary

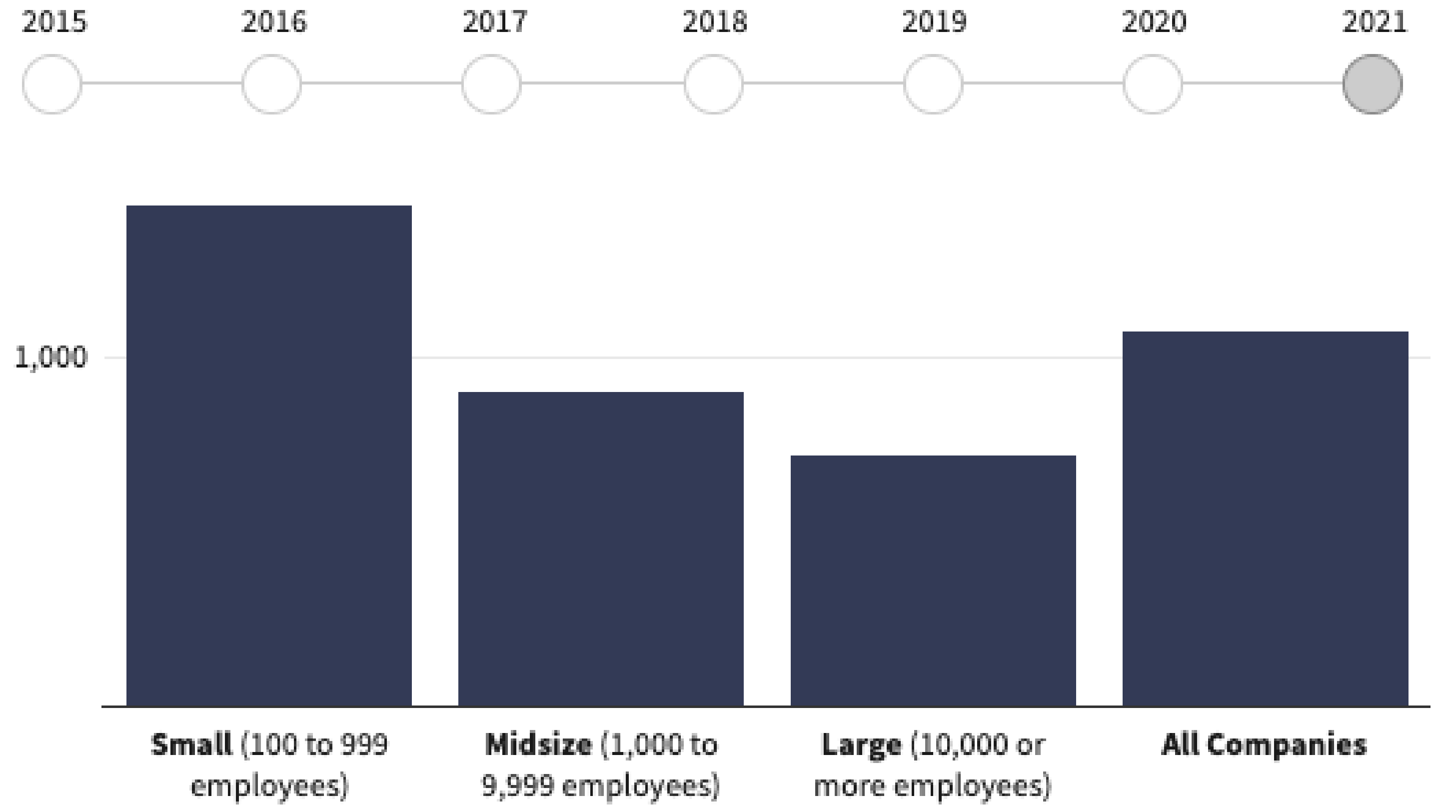
WHY IS IT IMPORTANT TO KNOW THE VARIABLE?

$1,000-$5,000 per hire

https://hiring.workopolis.com/article/how-to-calculate-cost-to-hire-and-why-its-important/

THE COST OF TRAINING





Reference: https://www.investopedia.com/financial-edge/0711/the-cost-of-hiring-a-new-employee.aspx

- Companies spent $92.3 billion in 2020-2021 on training according to a recent study by Training Magazine.
- Employees devoted an average of 64 hours to training.
- Small businesses pay significantly more to train each new employee

GOALS

- Promote better workspace setting
- Help employers to retain Their employees
- Focus on analyzing variables that lead towards attrition and work to decrease the same.

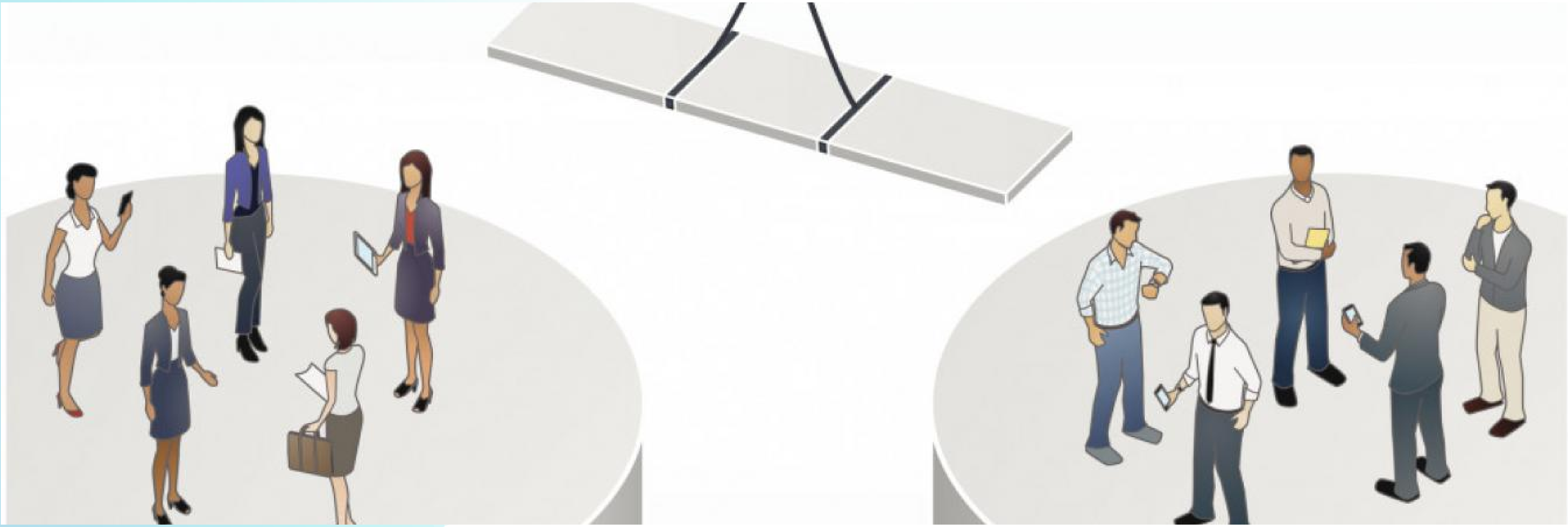
DATASET

- The dataset we've used is fabricated by the Data Scientists at IBM
- Obtained from kaggle.com
- The dataset is composed of 35 columns and 1471 rows, with a total of 51485 data points

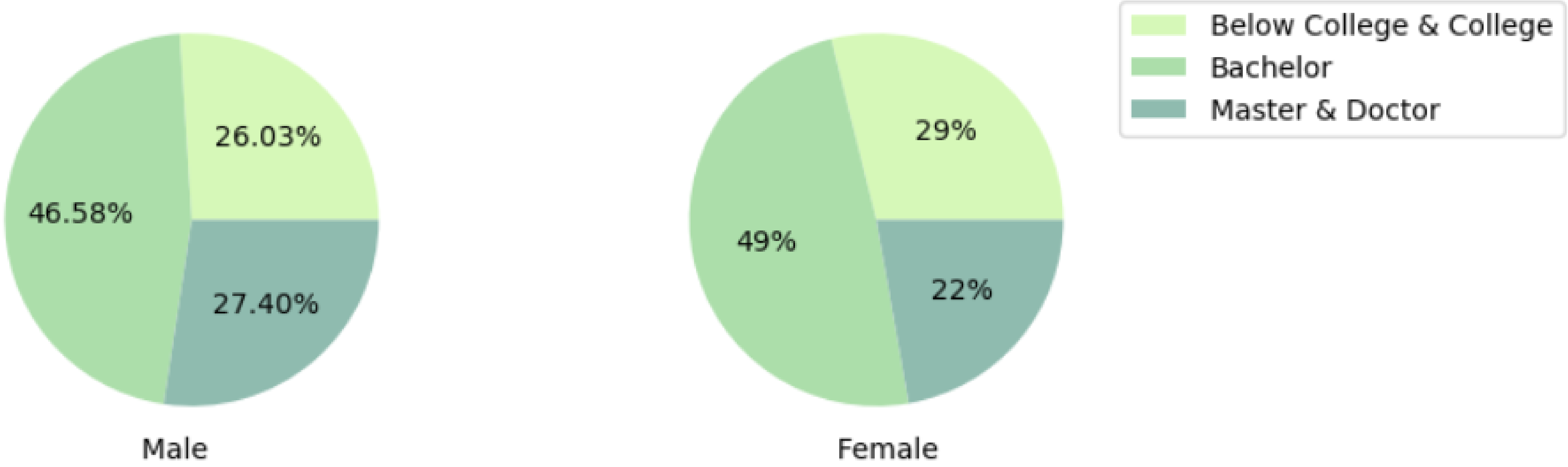
PRE-PROCESSING

ANALYSIS 1

Analyzing how hourly wages, marital status, and work-life balance affect attrition with respect to gender.



WHAT IS THE EDUCATIONAL LEVEL OF THE MALE AND FEMALE EMPLOYEES WHO HAVE LEFT THE COMPANY AND WERE PAID LESS THAN THE MEAN WAGES?



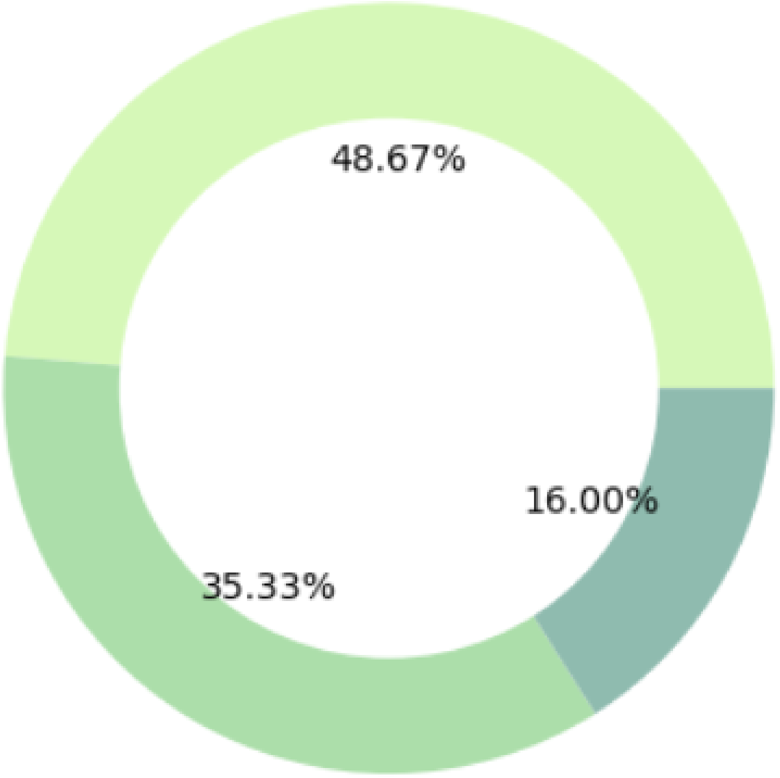
The majority of Employees who attrited had a bachelor's degree. Possible reasons?

- pursuing higher education
- Switching to better jobs early in their careers

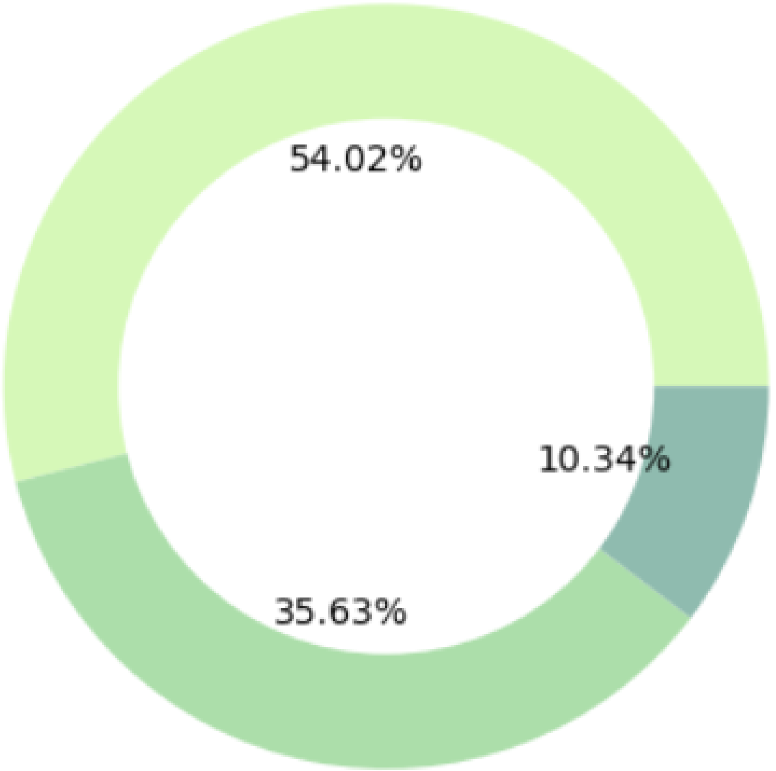
Employees with a master's or doctorate degree may have attrited due to:

- being underpaid and overqualified

WHAT IS THE MARITAL STATUS OF THE EMPLOYEES WHO LEAVE THE COMPANY?



MALE



FEMALE

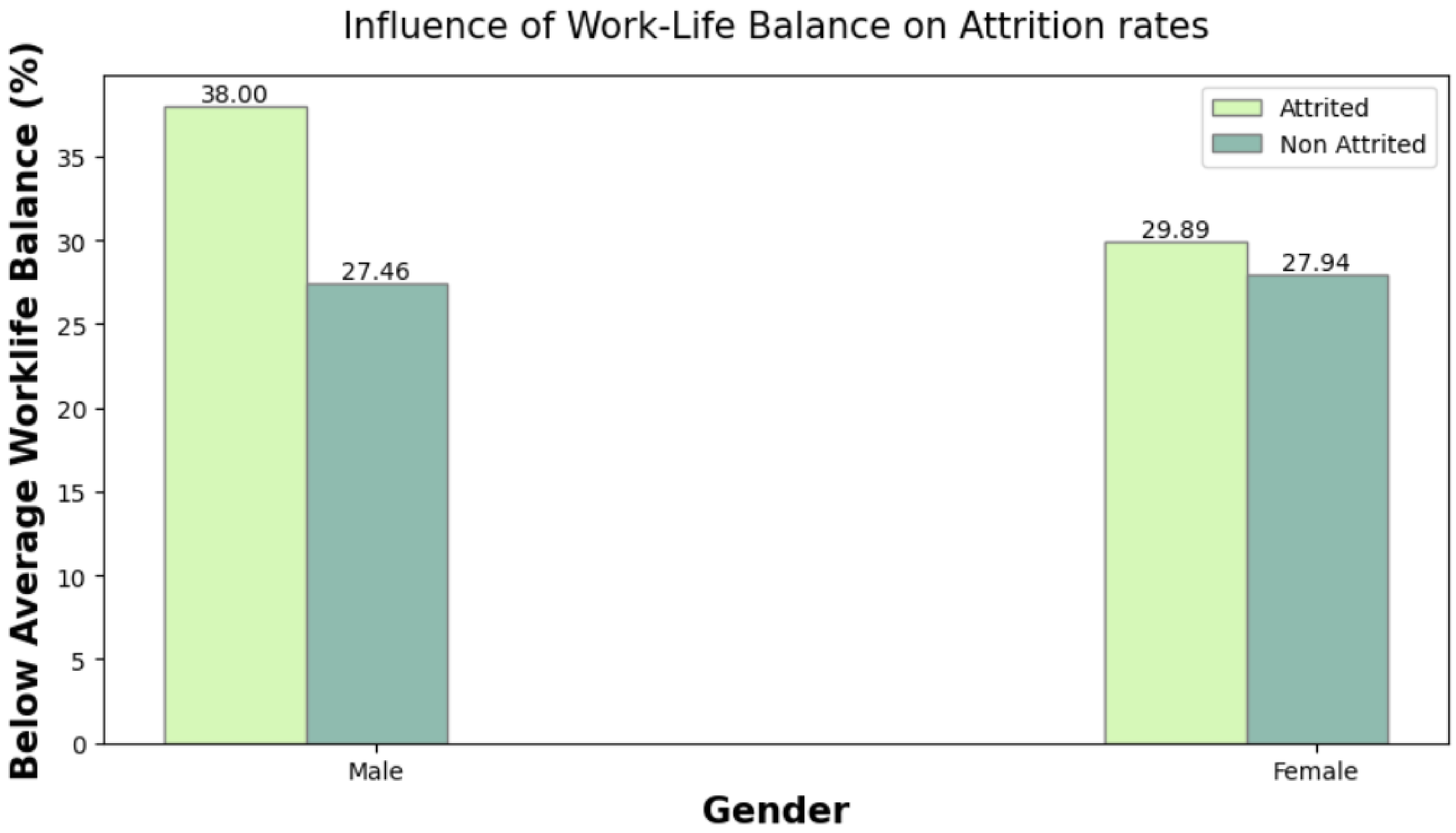
SINGLE

- Less constrained by family obligations.
- Flexible to relocate

MARRIED

- Unlike our initial hypothesis that married females are more prone towards attrition, our data proved otherwise.

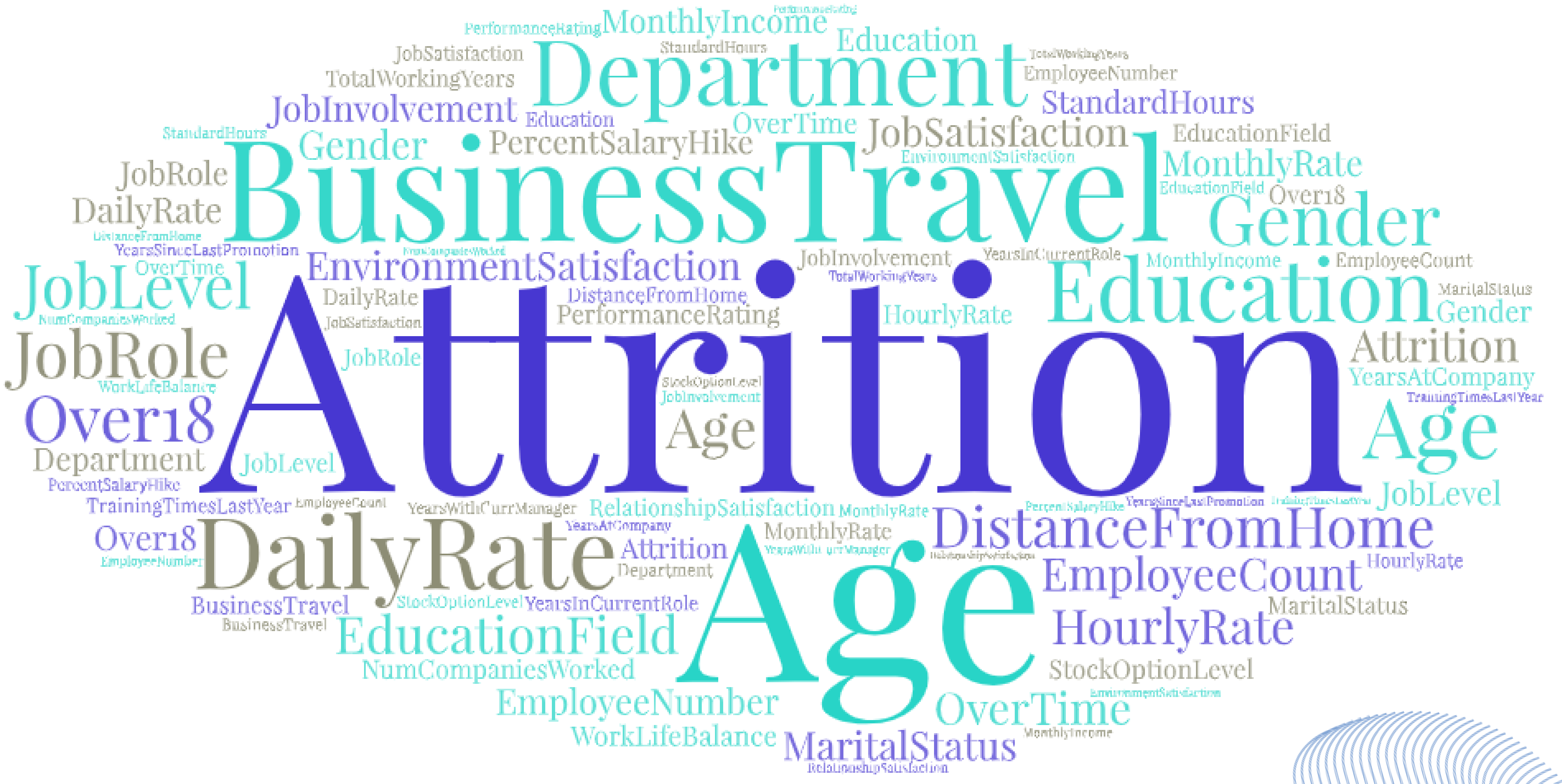
THE EFFECT OF WORK-LIFE BALANCE ON ATTRITION FOR EACH GENDER



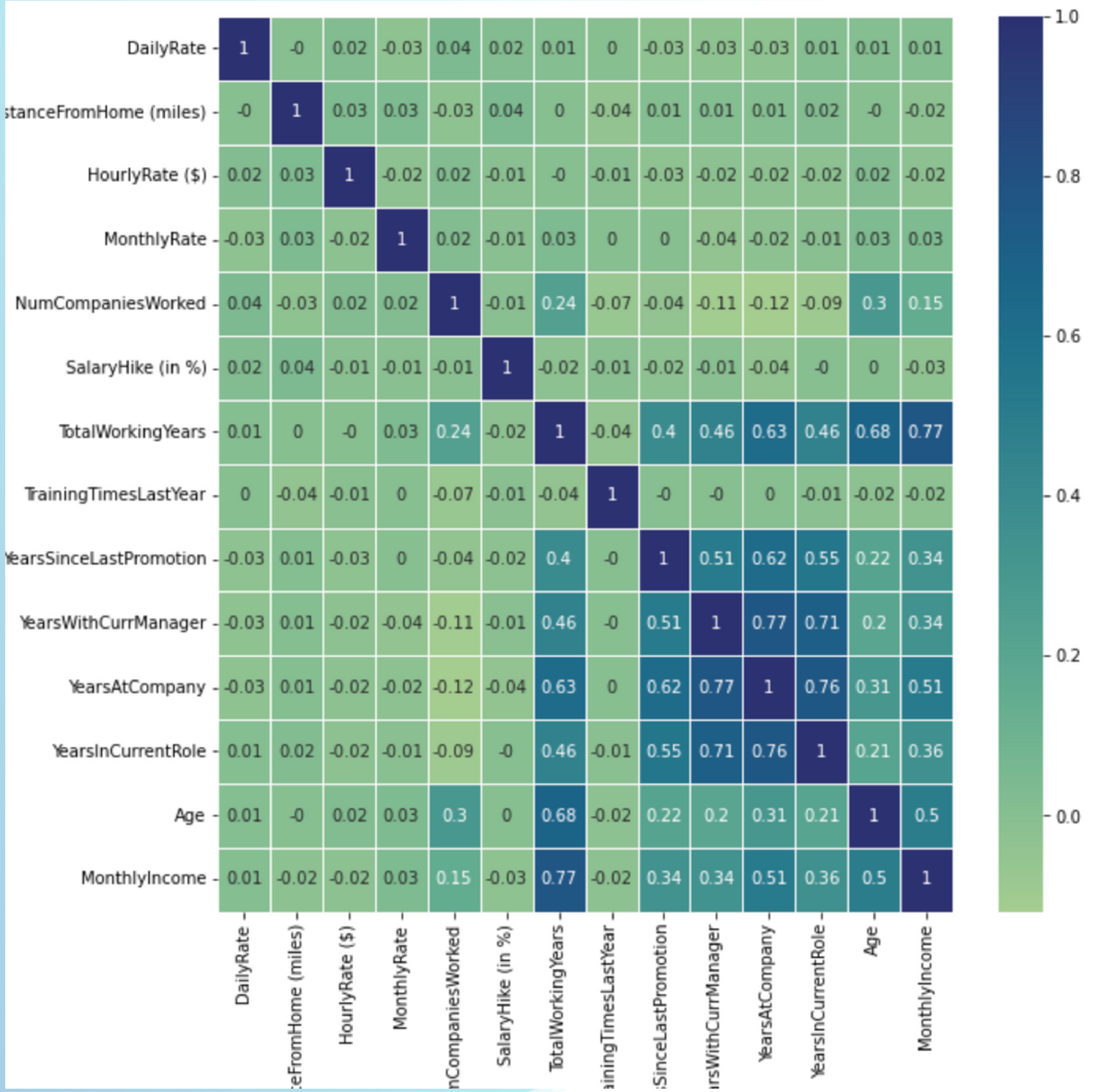
Males are more likely than females to quit their jobs when they have a poor work-life balance.

ANALYSIS 2

- Detect the 10 most significant variables that affect the attrition rate.
- Predict the likeliness of attrition for a new employee.

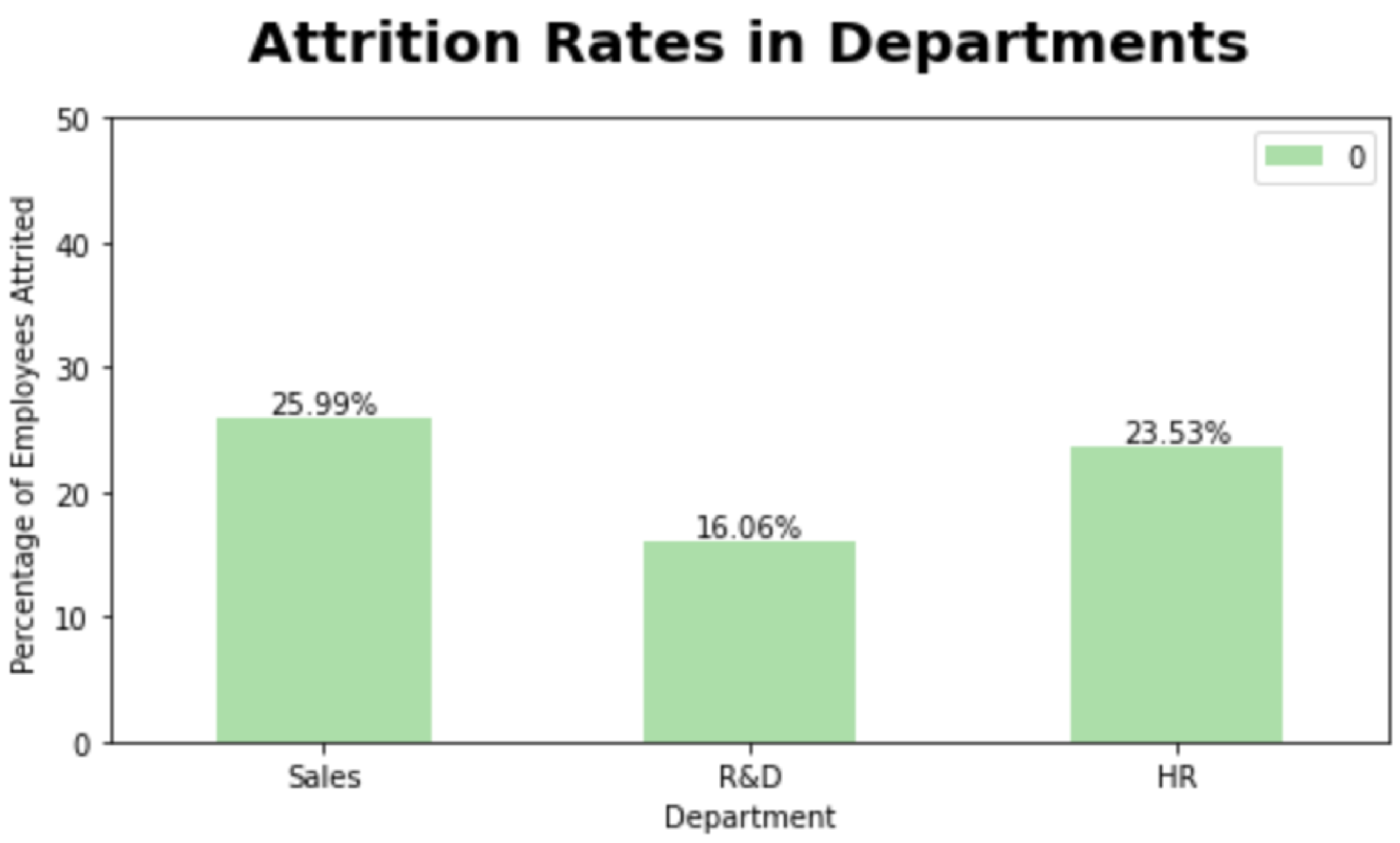


MULTICOLLINEARITY



- Purpose: Detect any specific combination of factors that result in high attrition rate
- Before running the model: Detect multicollinearity
- correlation significantly above .5:
  - 'Years with Current Manager' and 'Years at Company'
  - 'Age' and 'Total Working Years'
  - 'Monthly Income' and 'DailyRate'
- drop the variables - YearsAtCompany, YearsInCurrentRole, Age, MonthlyIncome

FURTHER REMOVE VARIABLES



- The variation in attrition rate does not change significantly on basis of the department
- How to find variables that contribute to the model more?
  - Akaike Information Criterion (AIC) score
- A lower AIC score indicates superior goodness-of-fit and a lesser tendency to over-fit.

OPTIMIZING VARIABLES

Why choose 10?

- based on our judgment

- Step 1:

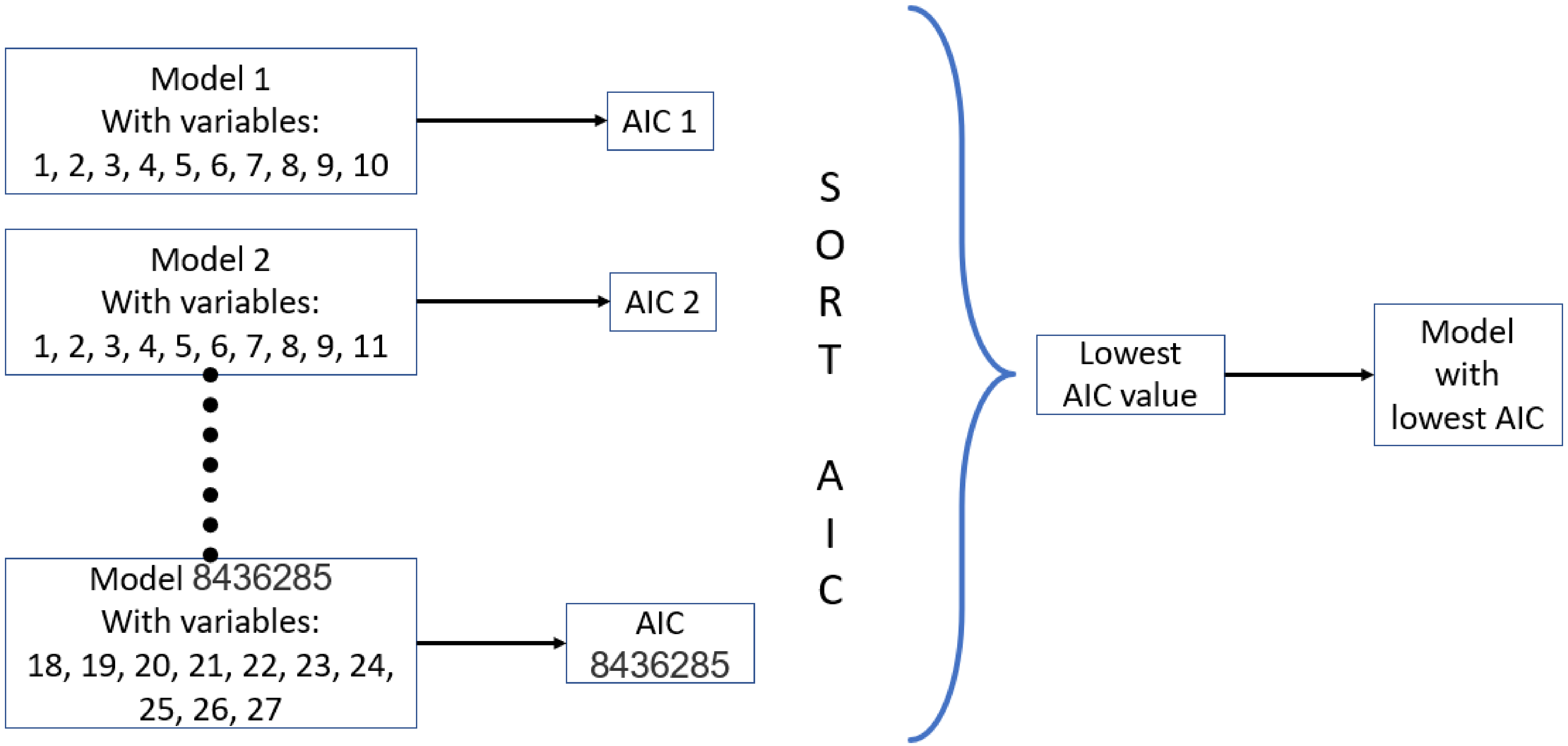
  Build a logistic regression model using 10 variables. Get AIC value.

- Step 2:

  Repeat Step 1 for 8.4M models.

- Step 3:

  Select the model with the lowest AIC value.



OPTIMIZING VARIABLES

Optimized Variables:

DistanceFromHome (miles)

EnvironmentSatisfaction

JobInvolvement

JobSatisfaction

MaritalStatus

NumCompaniesWorked

OverTime

PerformanceRating

TotalWorkingYears

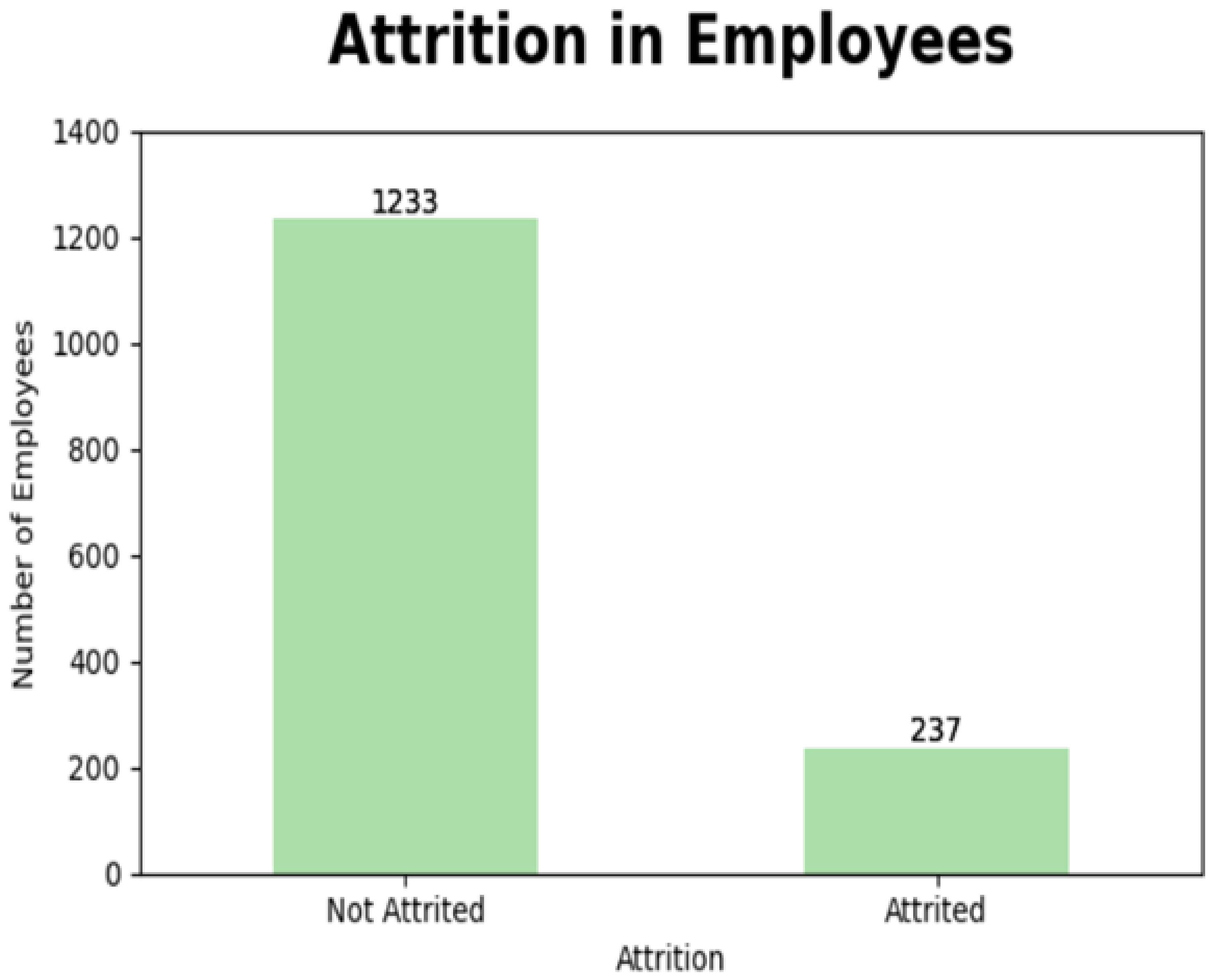
YearsSinceLastPromotion

Model Number	var1	var2	var3	var4	var5	var6	var7	var8	var9	var10	AIC values
7183910	3	7	10	13	14	16	17	19	22	25	1040.635018
5568007	2	3	7	10	13	14	16	17	22	25	1040.855640
5573019	2	3	7	11	13	14	16	17	25	26	1040.955168
5579234	2	3	7	13	14	16	17	22	25	26	1042.240874
8330012	7	10	13	14	16	17	19	22	25	26	1042.454901
2217510	0	3	7	10	13	14	16	17	19	22	1043.017188
7183948	3	7	10	13	14	16	17	22	25	26	1043.487727
5568705	2	3	7	10	13	16	17	21	22	25	1043.591947
5564725	2	3	7	10	11	14	16	17	25	26	1043.919037
5563945	2	3	7	10	11	13	14	17	25	26	1043.925429

Table of AIC values and the variables

TOWARDS PREDICTION

- Unbalanced data for the target variable: "Attrition."
- Less Data for a class translates to poor Prediction Accuracy for that class.
- We used SMOTE, which is an oversampling technique to solve this problem
- This solved the problem of unbalanced prediction accuracies.

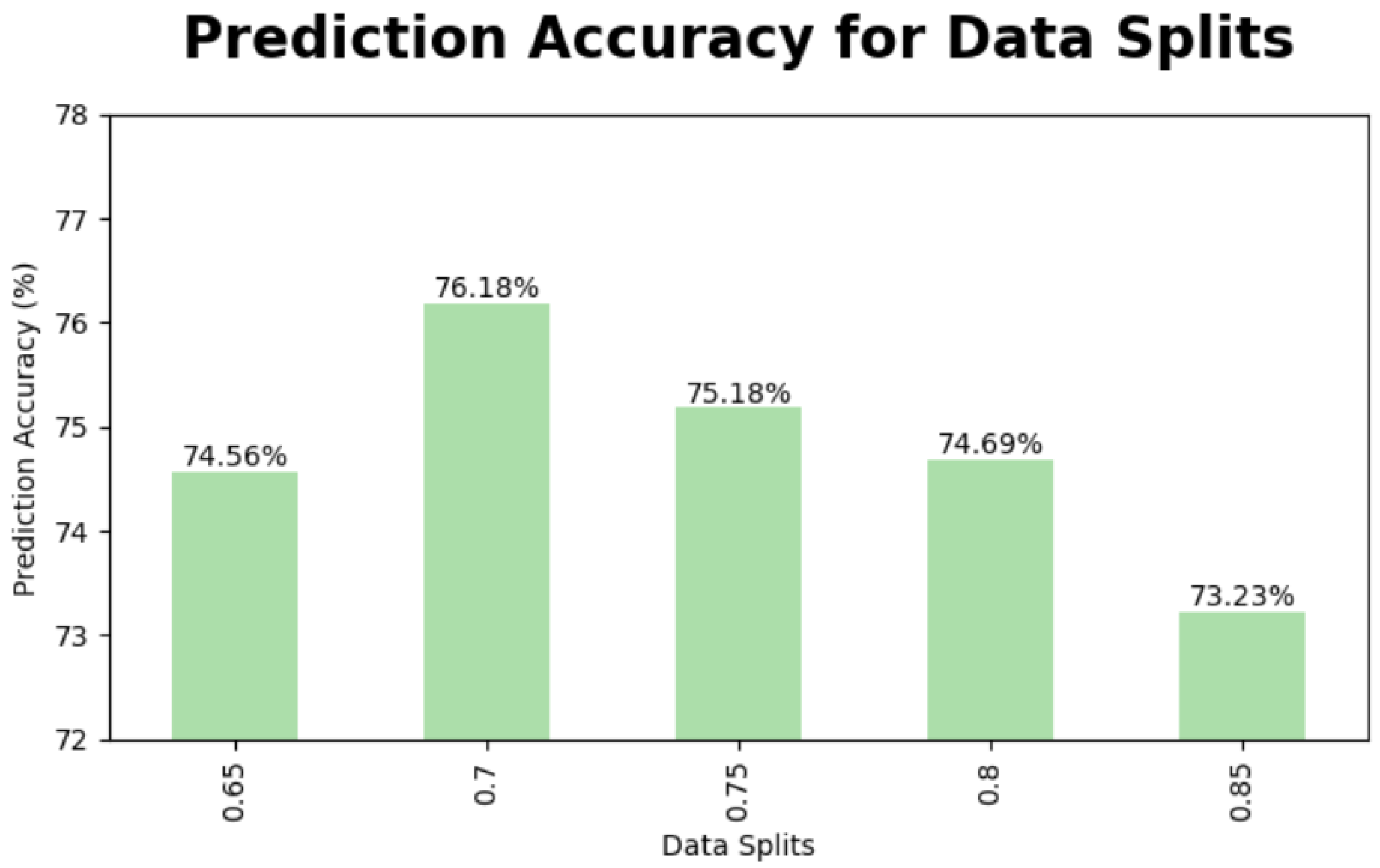


Comparison of Prediction Accuracies

	Attrited Employees	Non-Attrited Employees
Before SMOTE	0.31	0.98
SMOTE	0.76	0.73

OPTIMIZING DATA SPLIT

- Earlier we used cutoff of .8
- Achieved accuracy of 74.7%
- We tried to optimize for different cutoffs
- Cut off of .7 seems the most optimal
- Increasing accuracy from 74.7% to 76.2%

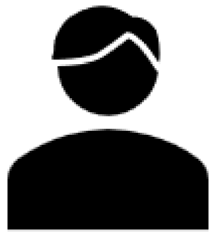


PREDICTING ATTRITION FOR A NEW EMPLOYEE

- We have our most critical factors
- Finalized our cutoff score for

  training dataset
- Model can be applied for prediction

Employee 1

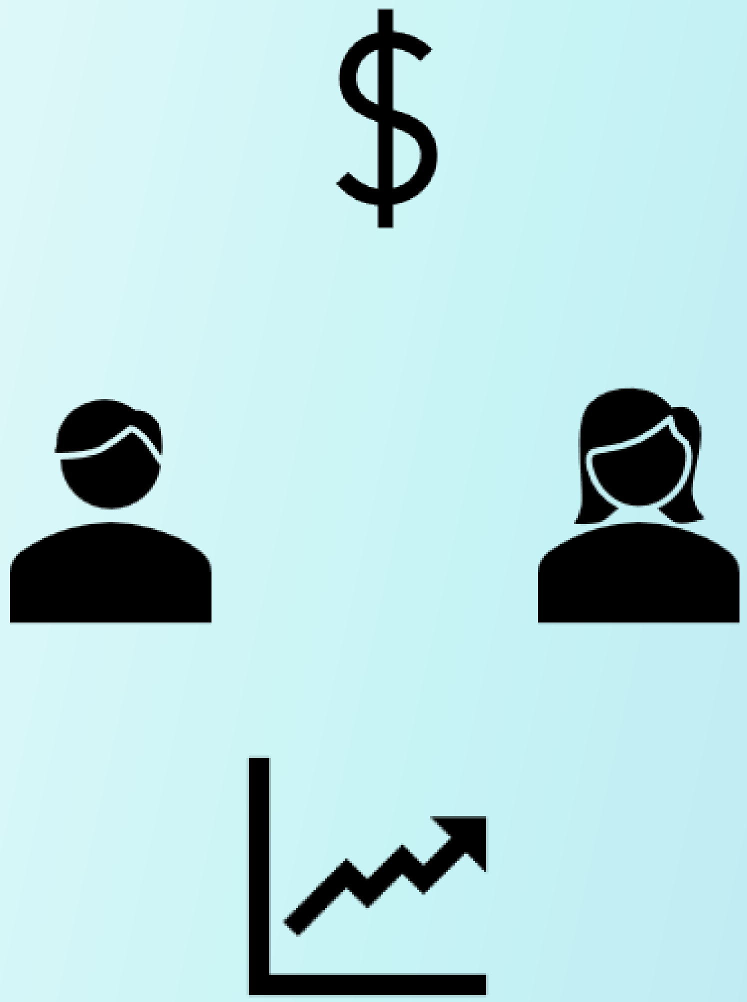


Employee 2

- Randomly selected 2 employees
- Treated them as new to the company
- Use the regression model to classify them

CONCLUSION

- Companies are spending trillions of dollars to overcome the challenge of attrition
- Although, same factors impact both gender, but for males, it's more pronounced
- Identified 10 most critical factors
- Built a prediction model



THANK YOU!