CITY
FURNITURE

EMPLOYEE TURNOVER

UF ISOM

CONTENT

01 OUR TEAM

02 EXECUTIVE SUMMARY

03 WORKFLOW

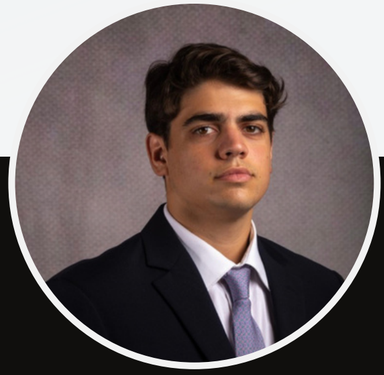
04 RECOMMENDATIONS

05 DATA HANDLING AND EDA

06 MODELING

07 Q&A

OUR TEAM



Manuel Ucha Rodriguez

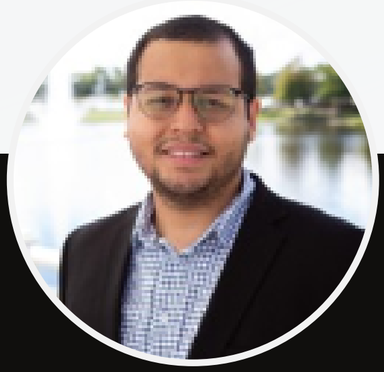
Team Lead/
Data Scientist

Chenxin Zhang

Data Engineer

Oliver Costantino

Modeler



Lewis Rincon Castano

Data Analyst

Jiayuan Liu

SME/Functionalist

EXECUTIVE SUMMARY

Business Objective

The student team was tasked with building a prediction model that generates the probability that an employee will leave. this is to help City Furniture with their high turnover problem

Outcome

We found out that the amount of work (shifts/hours) was the biggest factor affecting turnover. The less work the more likely the employees were to turnover

Recommendation

The student team will brainstorm for a recommendation to use the model and solve the sponsor's business probelm

WORKFLOW

The data team will perform EDA on the received dataset. With this we created a hypothesis and had a better knowledge of data when modeling

DATA ANALYSIS

Using the EDA and R the team performed 5 different models with the different imputed datasets to give the best chance at high accuracy

MODELING

With the model the data team brainstormed possible recommendations and possible solutions to lower the high turnover

DEPLOYMENT

HYPOTHESIS & PERSONA

Hypothesis: After an initial looking at the data, we hypothesized that low level employees that work minimal shifts would be most likely to turnover.

Persona of an employee likely to leave: A Service Technician that works 30 hours or less per week.

RECOMENDATION

UNDERSTANDING MODEL RESULTS AND EMPLOYEES FLAGGED

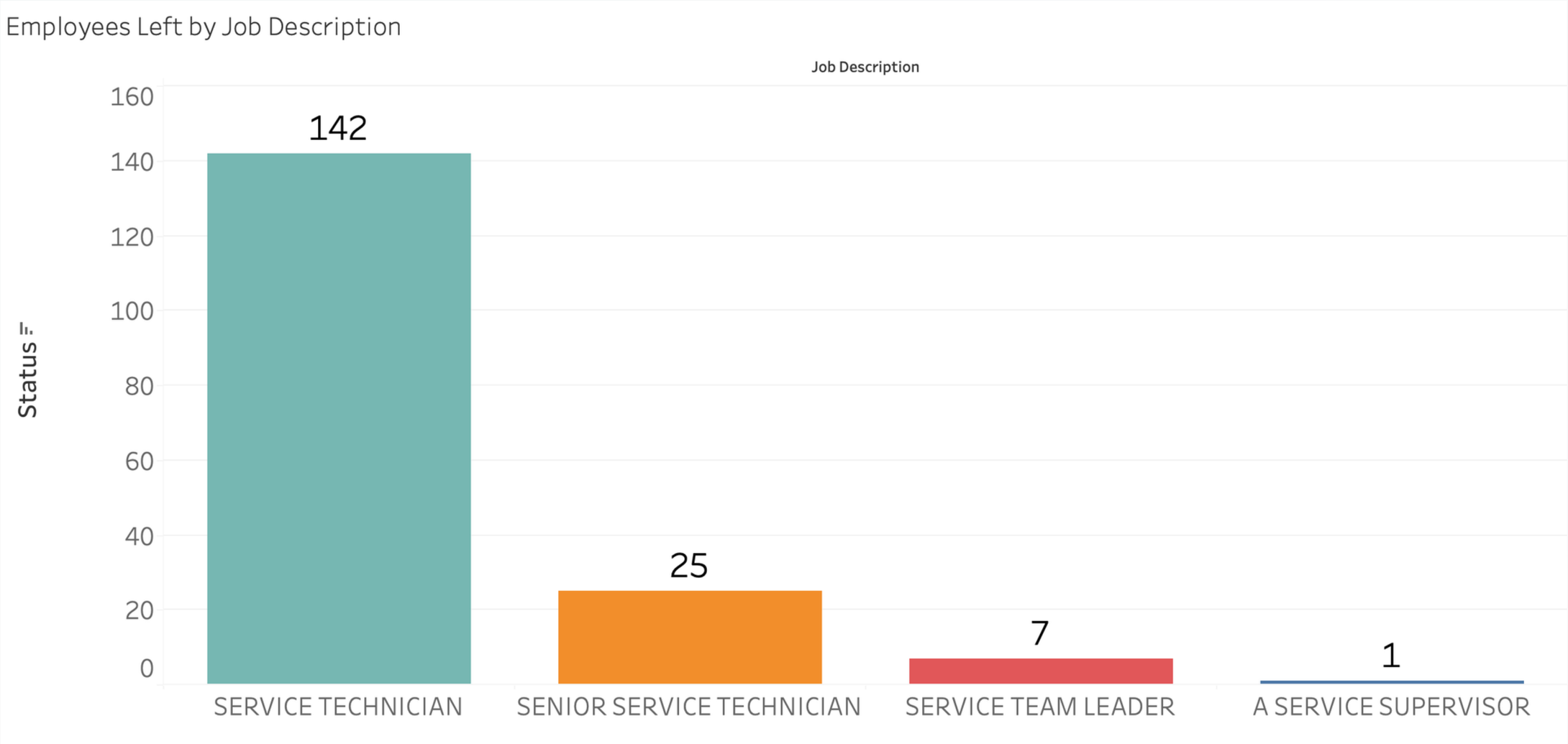
SUPERVISOR OPENS A LINE OF DIALOGUE

WORKERS CAN BE GIVEN MORE HOURS

EMPLOYEE STAY AND LOWER TURNOVER

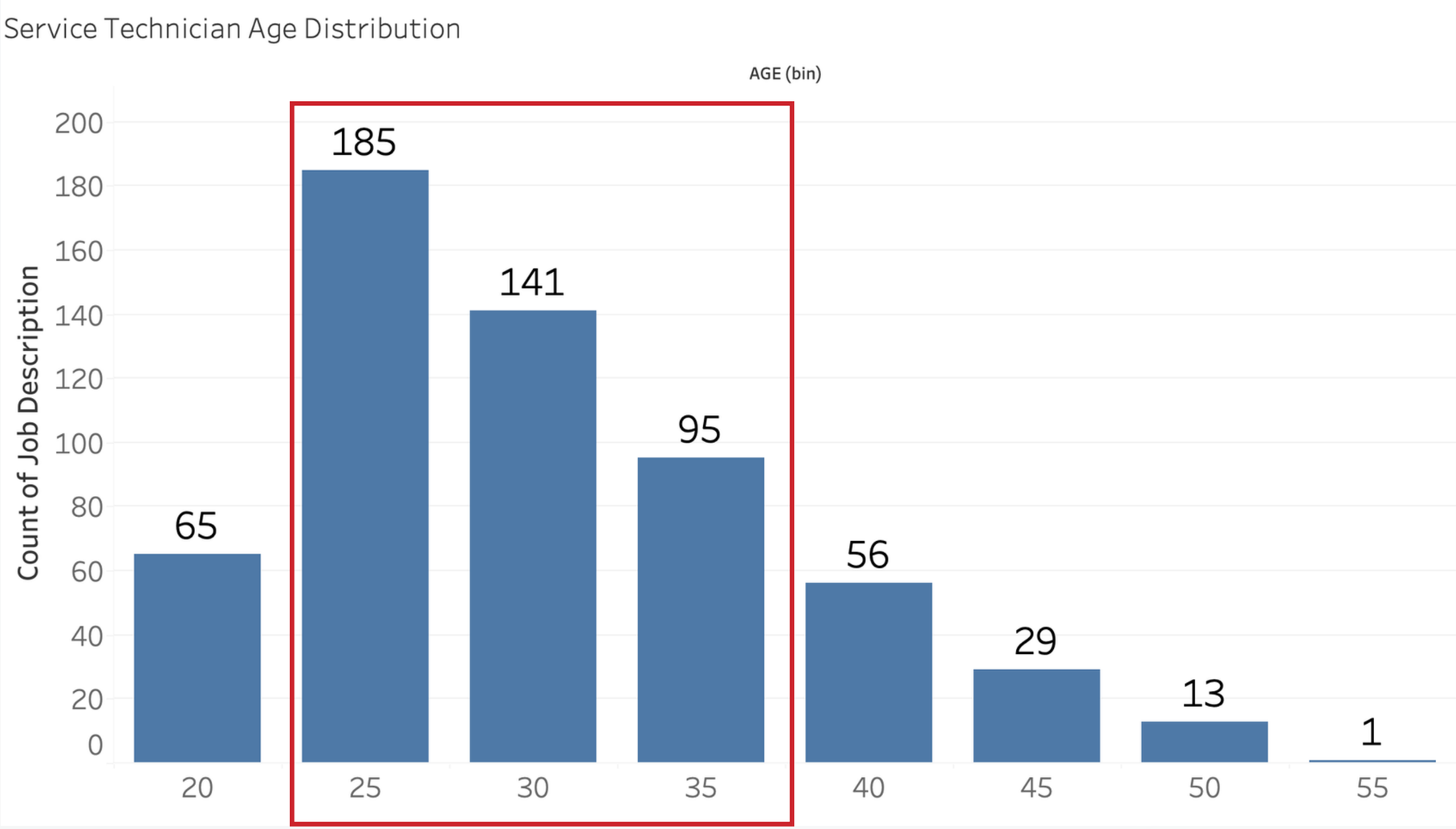
Give employees more hours if dialogue fails.

HOW TO PREVENT TURNOVER FOR SERVICE TECHNICIANS' ROLE.



More service technicians left compared to other roles

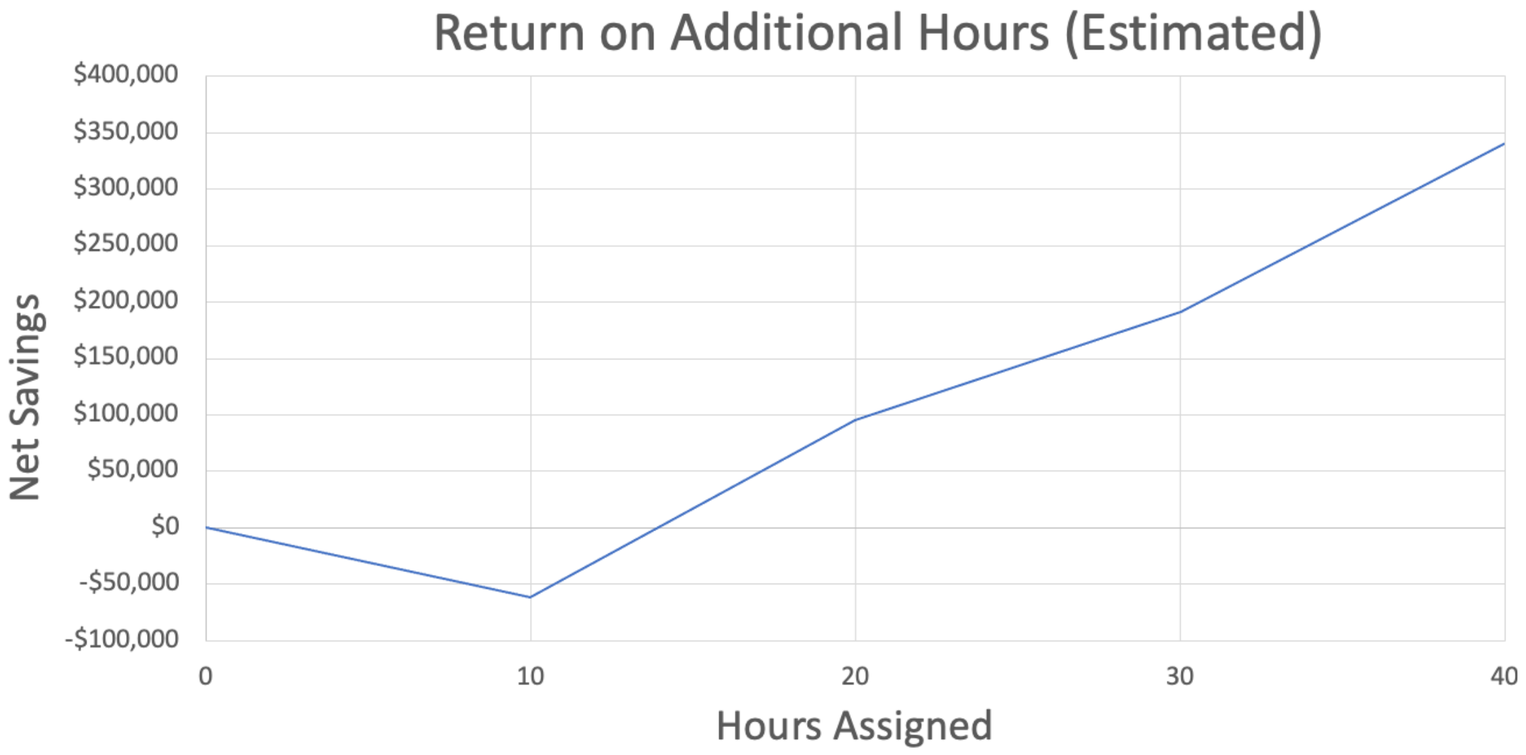
HOW TO PREVENT TURNOVER FOR SERVICE TECHNICIANS' ROLE.



Generation Z (born after 1997) values work flexibility and a balance between work and personal life, and prefers socially responsible companies with open communication and opportunities to share their ideas with management.

Most Service Technicians are Gen Z. They prioritize work flexibility & dialogue

ROI ANALYSIS FOR SERVICE TECHNICIANS

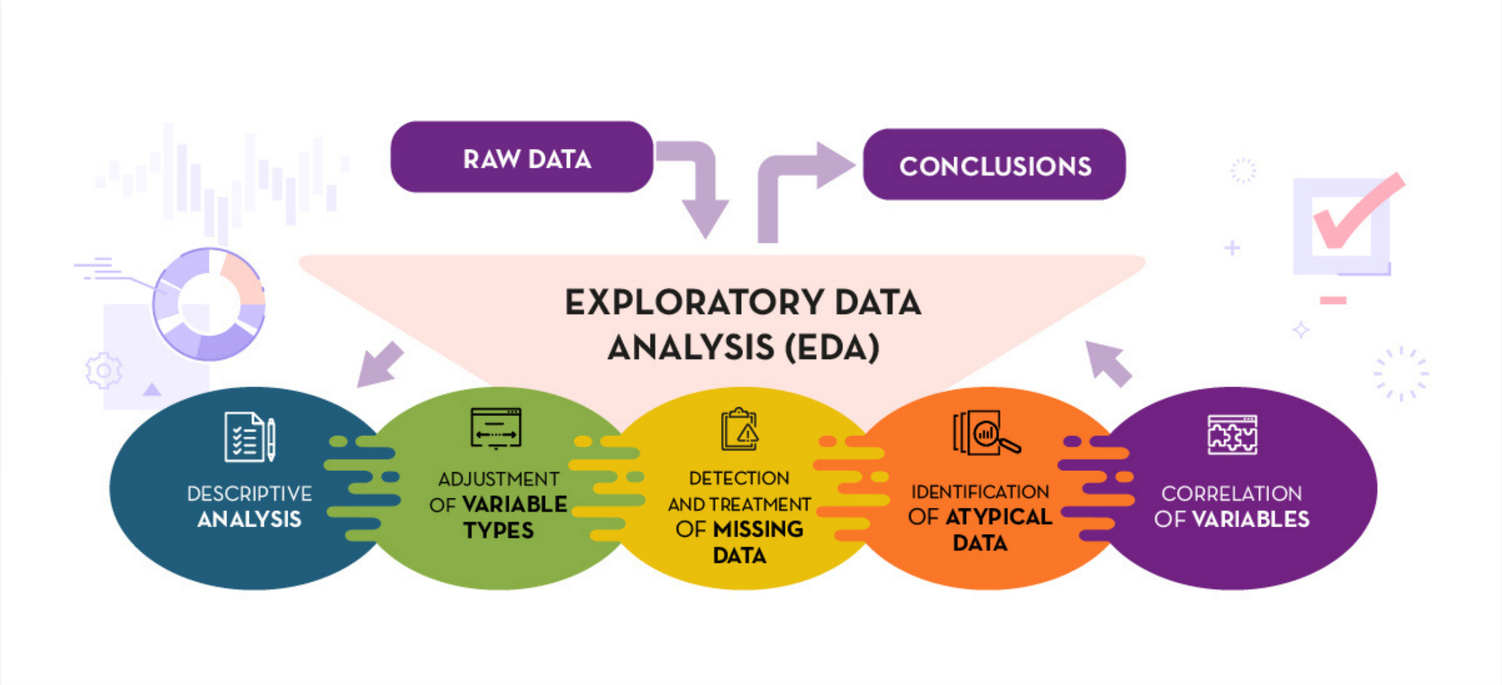


Assumptions :

- 142 Leavers
- Wage : $20/hr
- Cost to train new hire: $7,500
- Probabilities are calculated using the hours worked probability plot
- All values are estimated

IN-DEPTH PROJECT

DATA HANDLING & EDA



We use exploratory data analysis(EDA) to evaluate the data we received. This method could detect missing values, identify outliers and explore correlations between variables.

EDA help us ensure the results we produce could give City Furniture a deeper understanding of their current employee turnover situation, also support City Furniture's future decision.

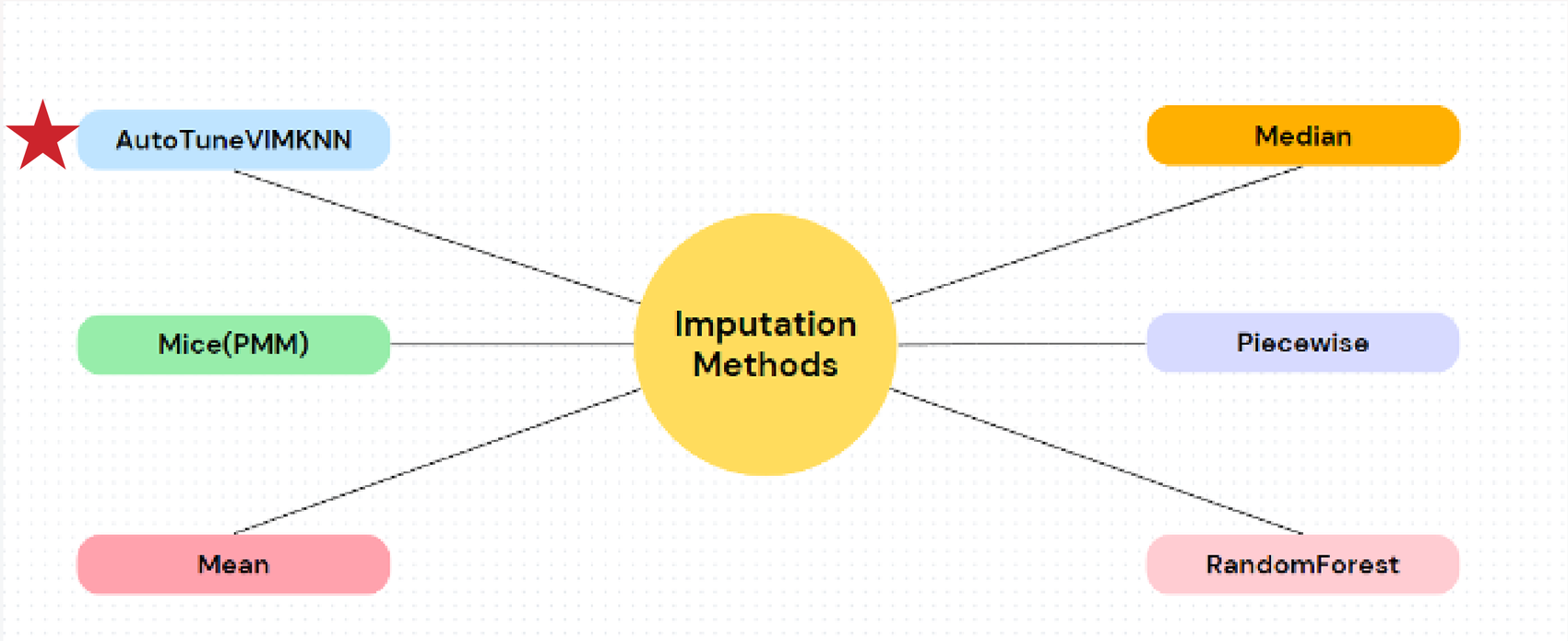
DATA HANDLING & EDA
MISSING VALUES

Total number of cells:
26820

Missing cells:
6424

24%

IMPUTATION



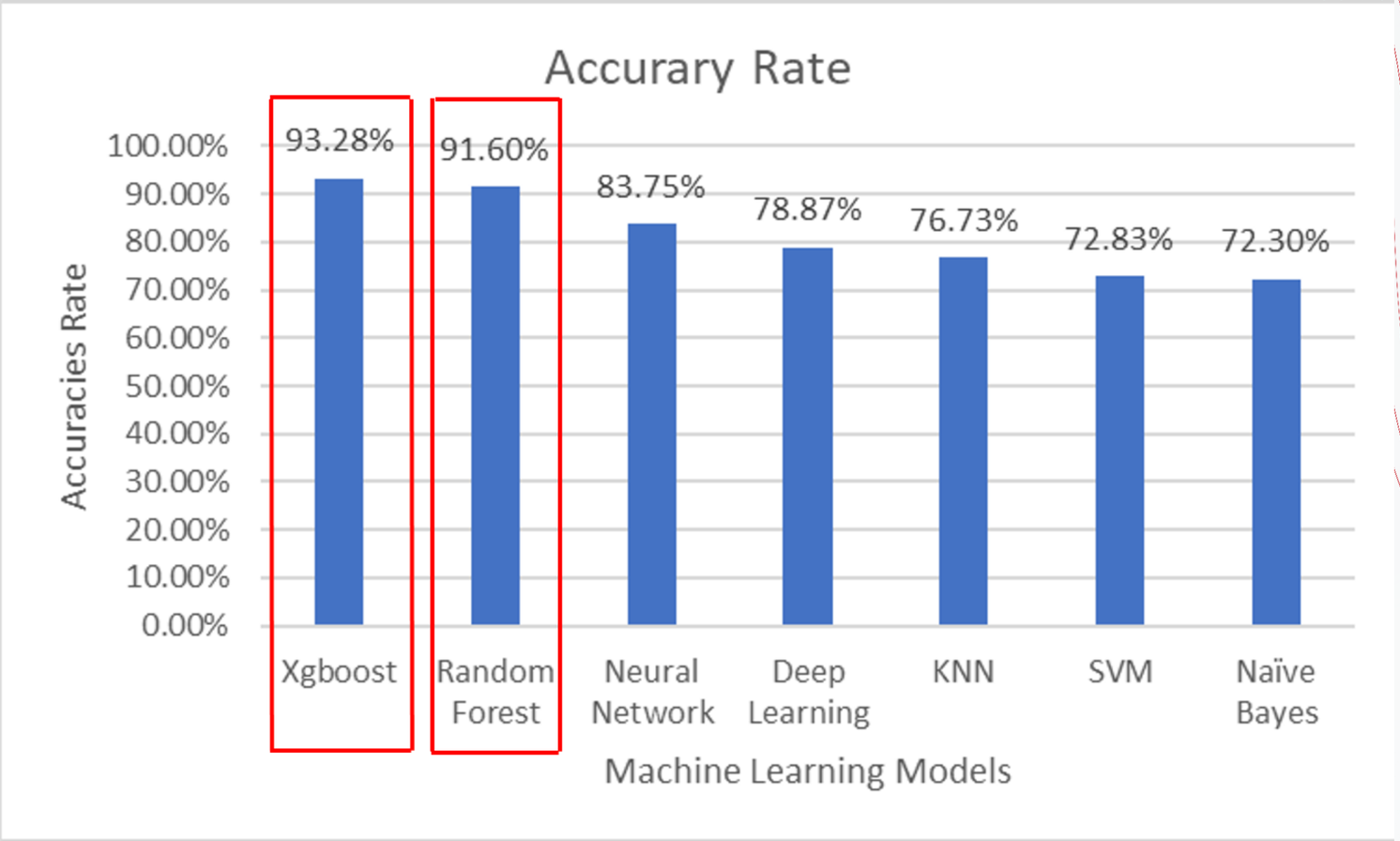
For every model we built, we applied all six imputed datasets.

AutoTuneVIMKNN is the best imputation method for our models

MODELING

Types of Models Used:

- Xgboost
- Random Forest
- Neural Network
- Deep Learning
- K Nearest Neighbor
- Support Vector Machine
- Naïve Bayes



XGbooost and Random Forest produced the highest accuracy

WHY RANDOM

- We have opted to choose the Random forest model with KNN v2 imputed dataset to be the final model going forward.
- It has a **high accuracy of 91.6%**.
- It has the **highest true negatives of 57 (leave-leave)** and the **lowest false positives with 17**.
- Therefore, in spite of XGboost having a higher accuracy of 93.3%, we felt the random forest model better **fit our business problem.**

91.6%

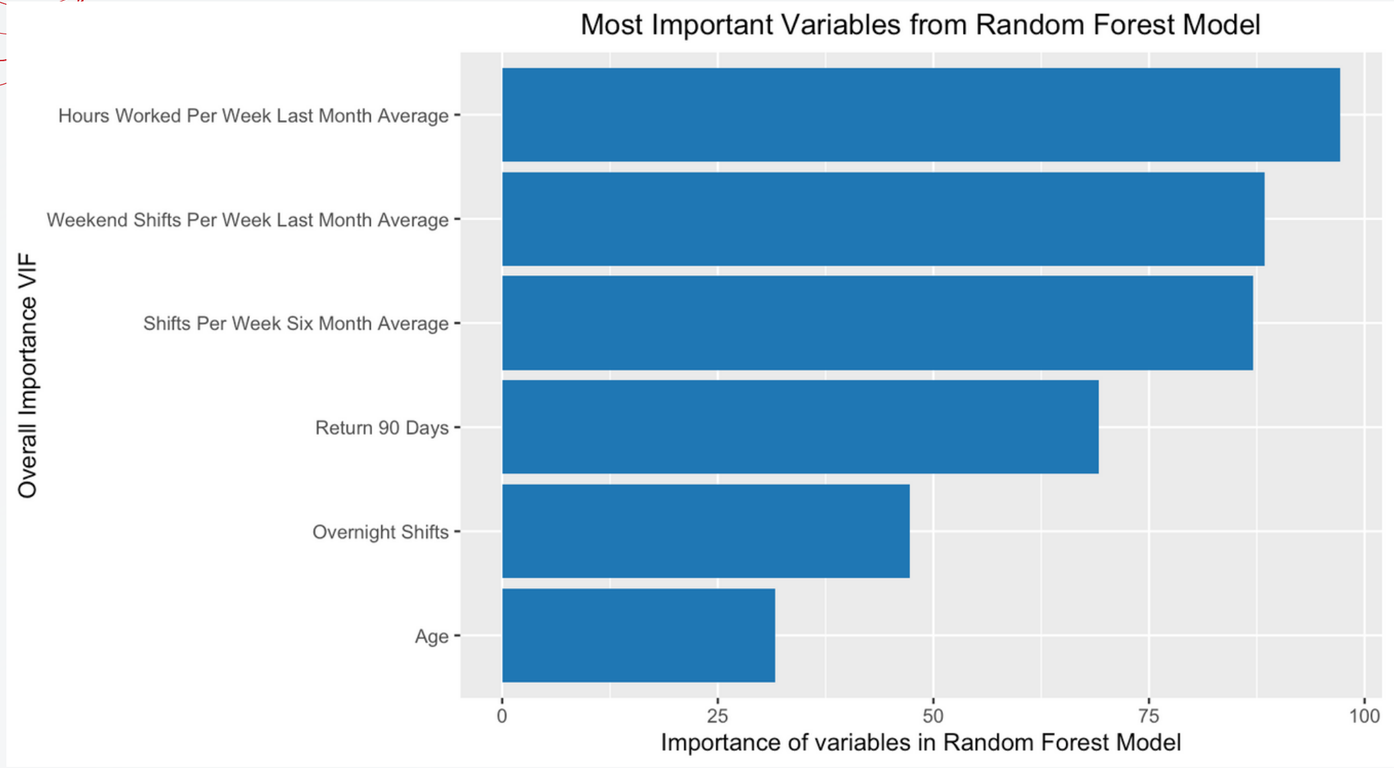
Accuracy Rate

57 (leave-leave)

17

(Stay-Leave)

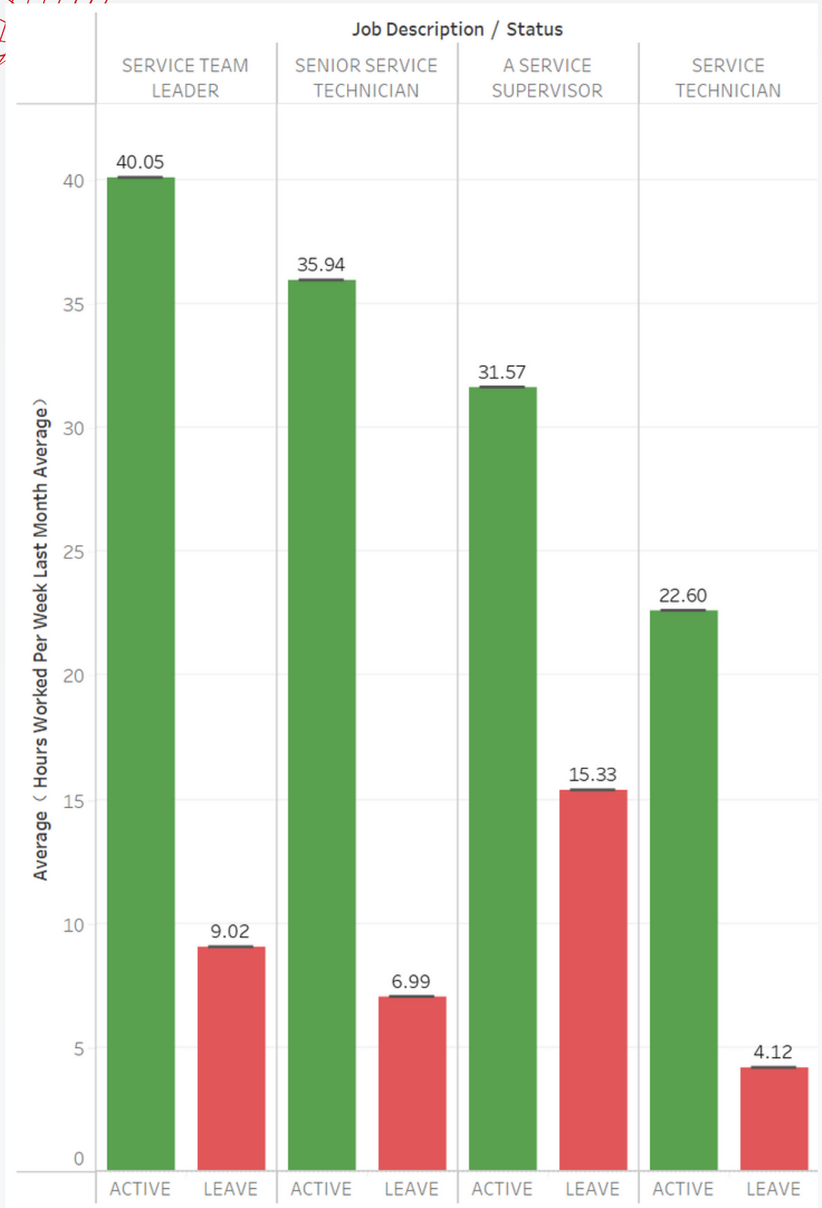
OUTCOME VIF



- This chart shows us the variables that were most important to our Random Forest Model. From this chart we can see that the hours and shifts worked plays a significant role in the outcome of our model.

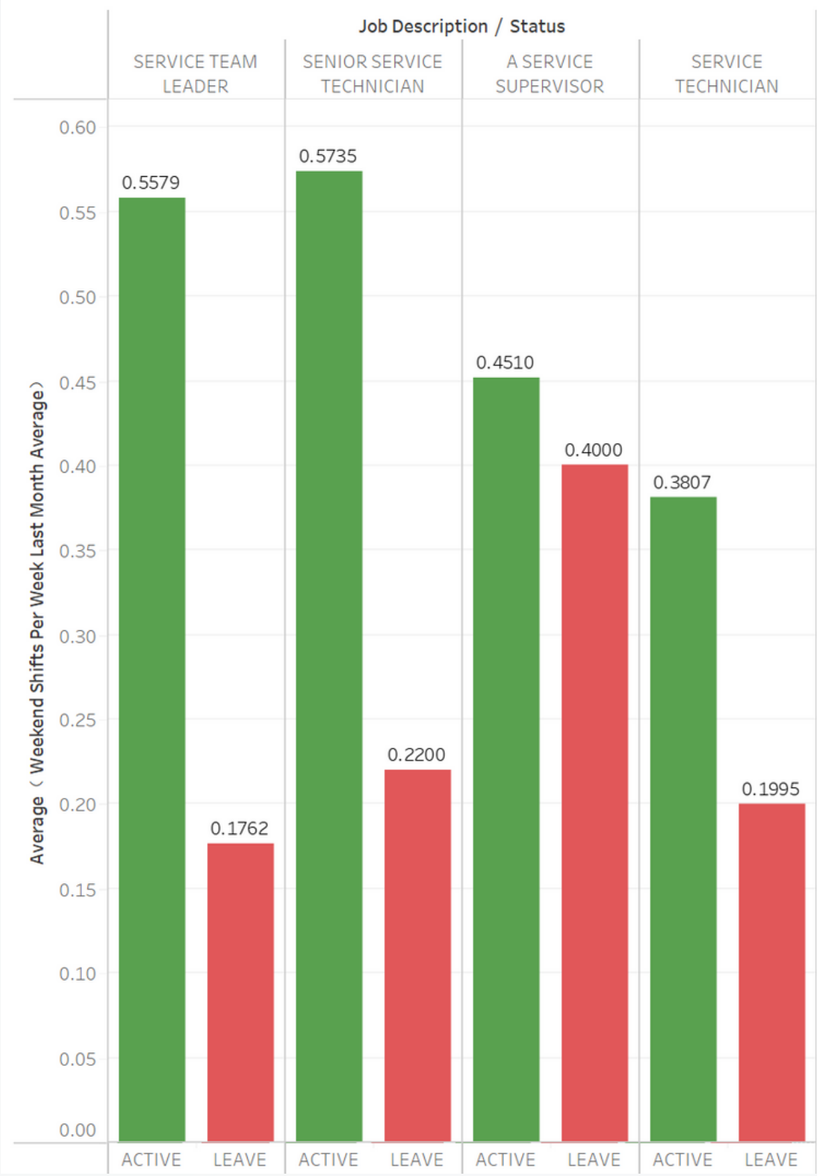
Hours worked per week was the most important variable

OUTCOME HOURS



- The less hours they worked the more likely the employees are going to leave.
- Also we can find that almost all the active workers have about twice more average working hours than the leave workers.

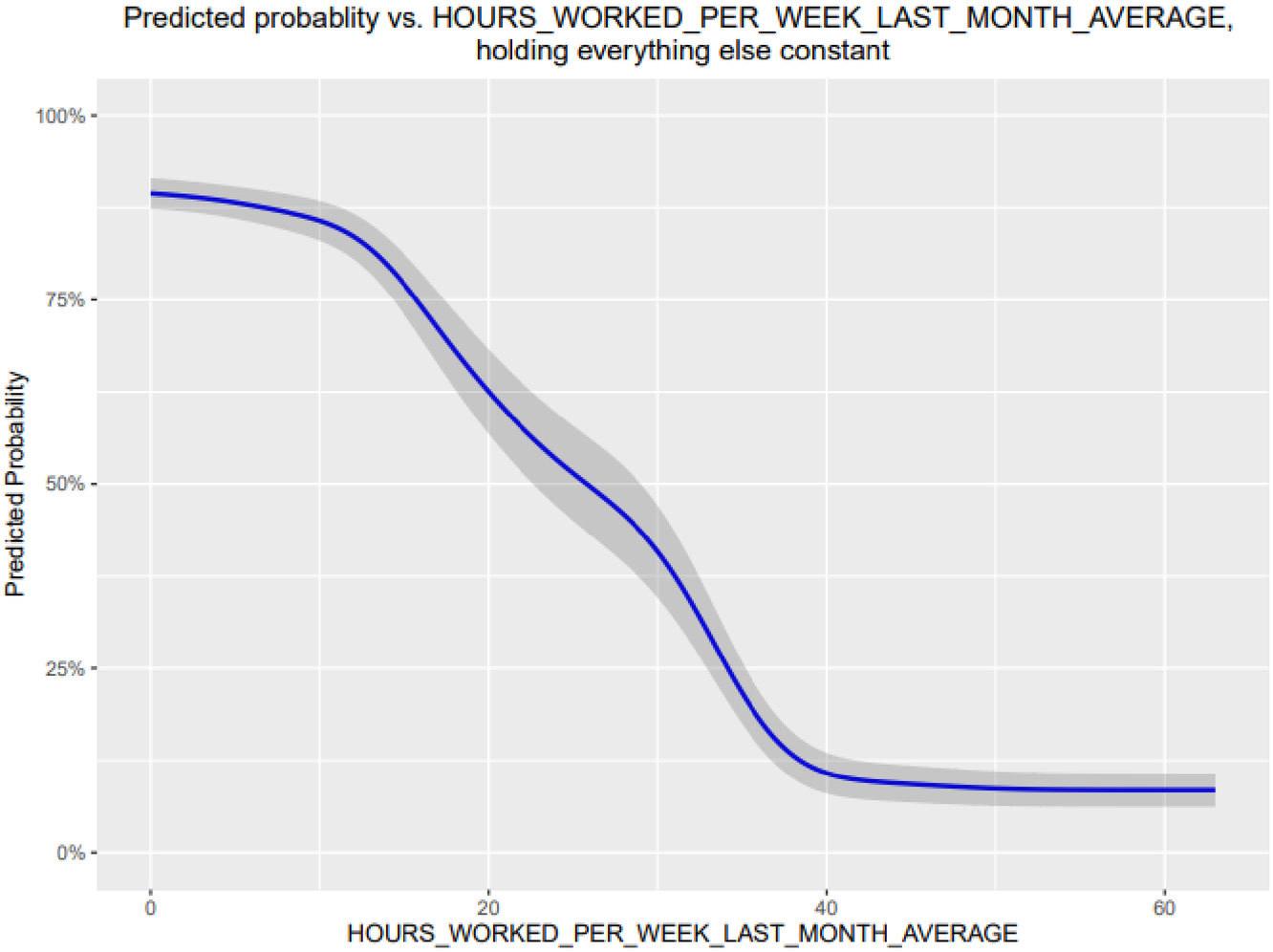
OUTCOME HOURS



- The less shifts they worked the more likely the employees are going to leave.
- Also we can find that almost all the active workers, except for service supervisor, have about twice more average working shifts than the leave workers.

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: HOURS WORKED PER WEEK LAST MONTH AVERAGE

- Employees with less than 30 hours of work are likely to quit between 25% and over 75%.
- Employees over 40 hours will likely quit around 15% or less.

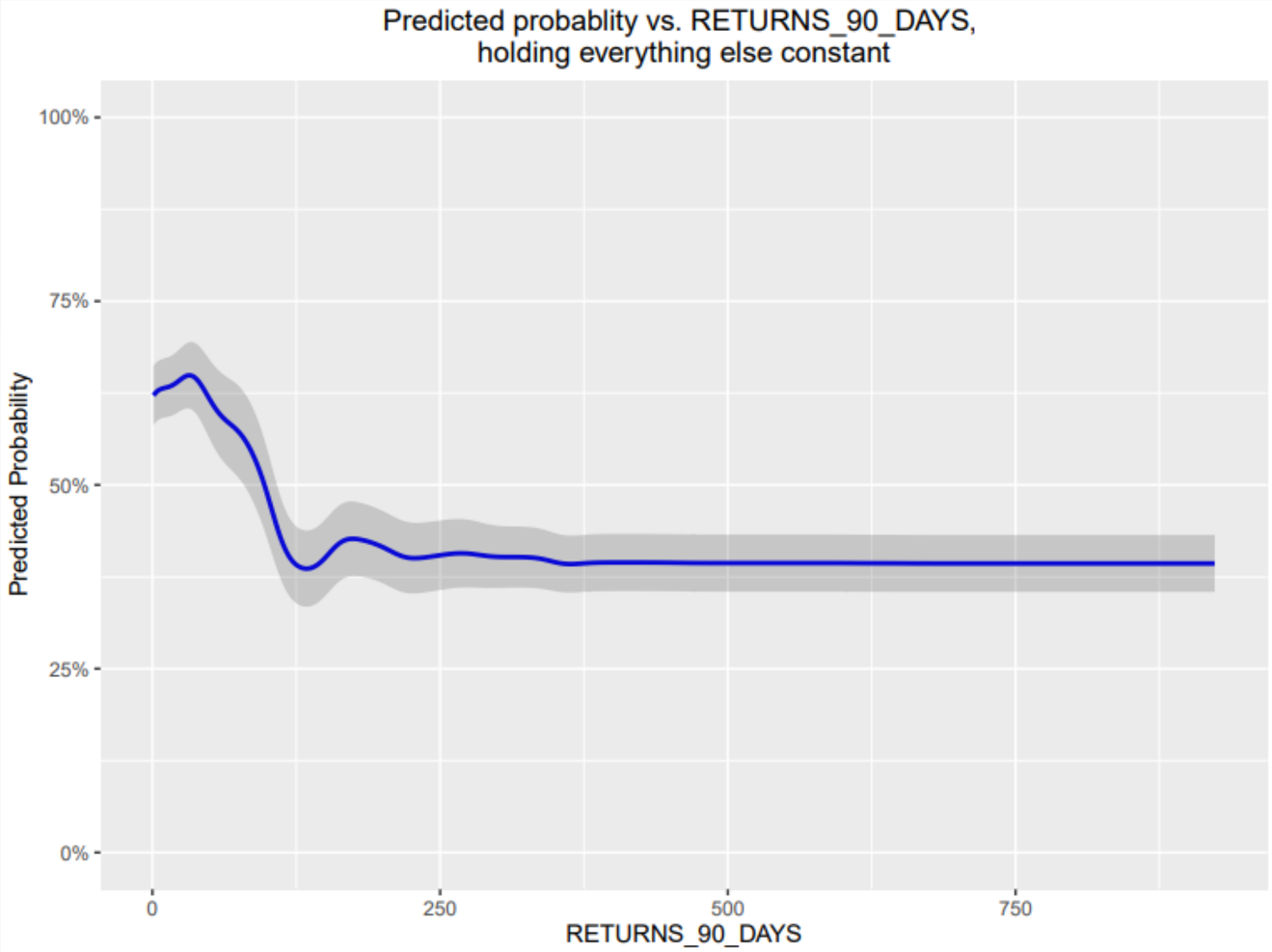


The probability of leaving the company reduced with a higher amount of working hours

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: RETURN 90 DAYS

- Employees with less than 250 returns are likely to quit between 40% and 65%.

- Employees with more than 400 returns have a constant probability of nearly 40%.

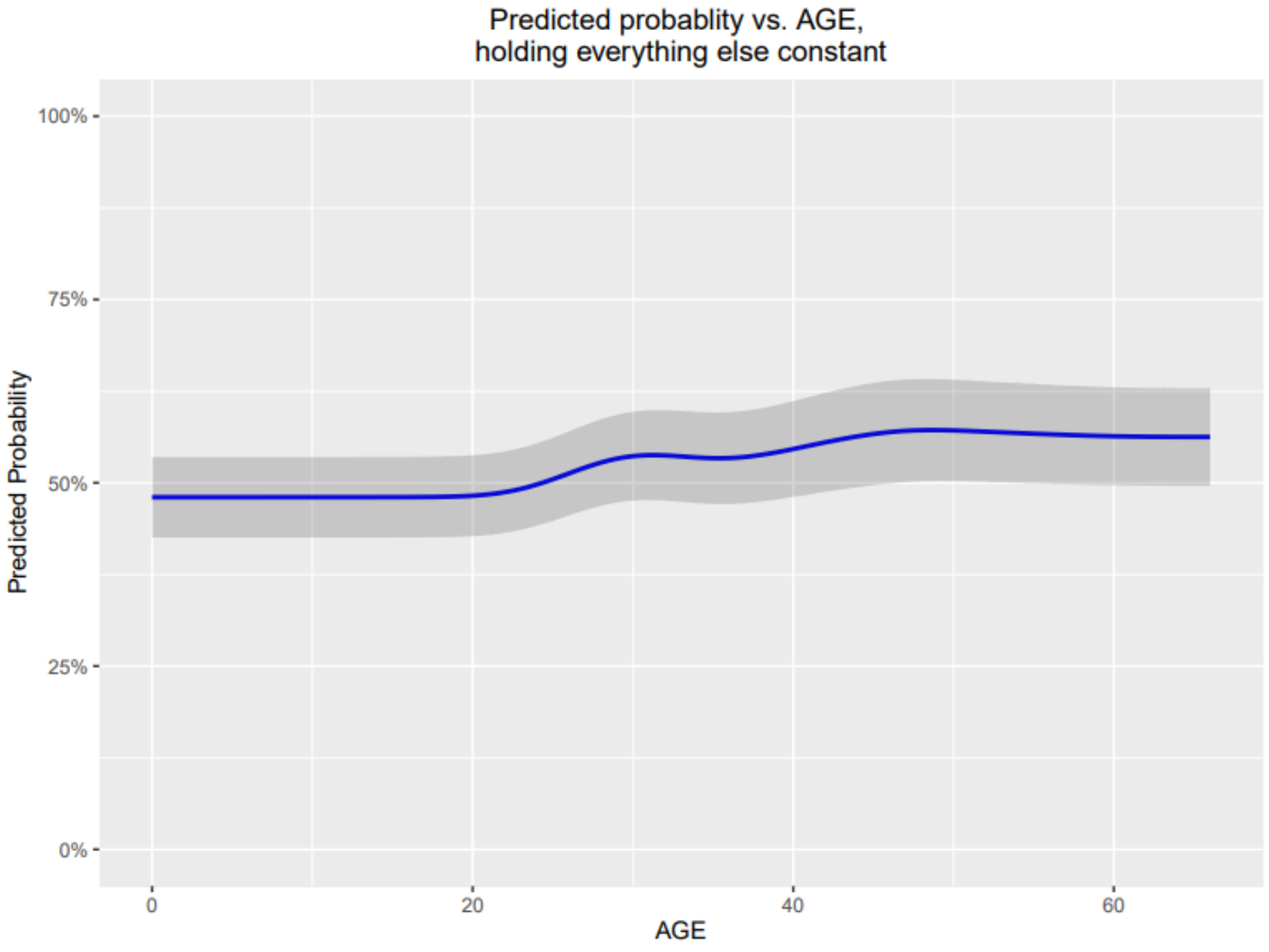


Employees with less than returns are more likely to quit.

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: AGE

- The probability of leaving the company varies between 47% and 55% between ages 20 and 40.

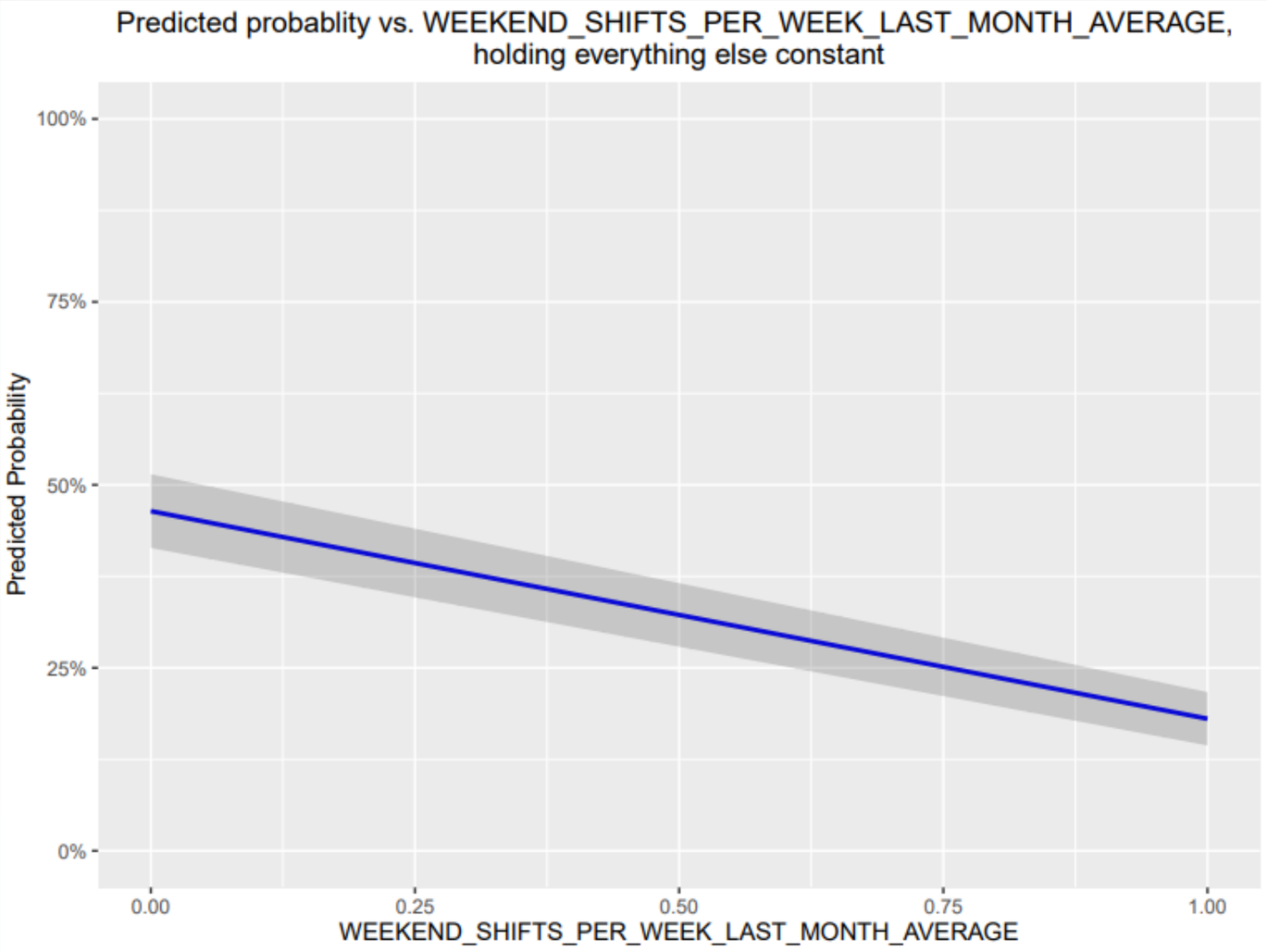
- Employees older than 40 years are slightly more likely to leave the company than younger workers; their probability rate is between 55% AND 60%



Employees older than 40 are more likely to leave the company than younger employees

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: WEEKEND SHIFTS PER WEEK LAST MONTH AVERAGE

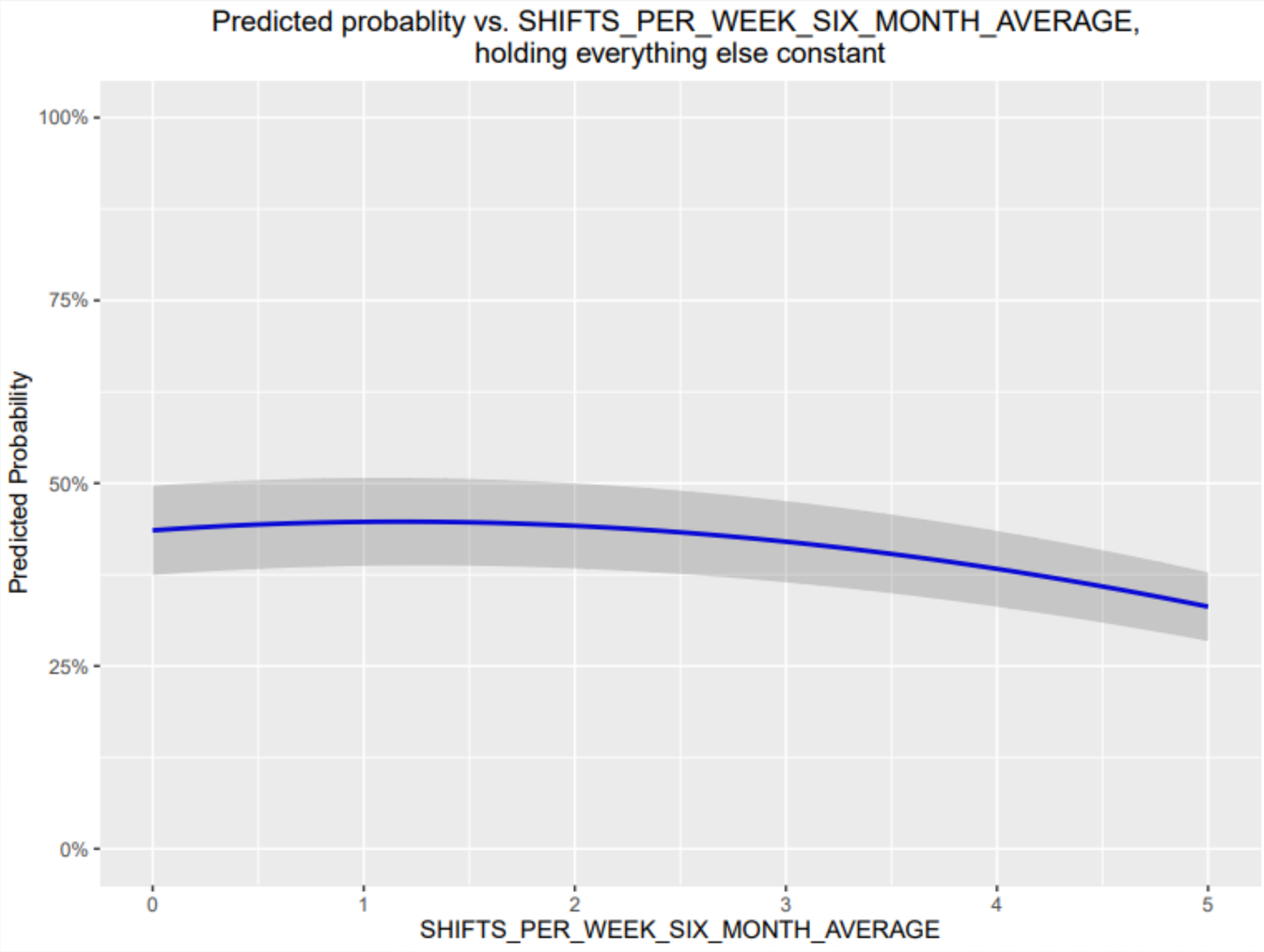
- Employees with less than a one-weekend shift average have a probability of quitting between 20% and 45%



Employees with fewer weekend shifts average are more likely to quit their jobs

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: SHIFTS PER WEEK SIX MONTH AVERAGE

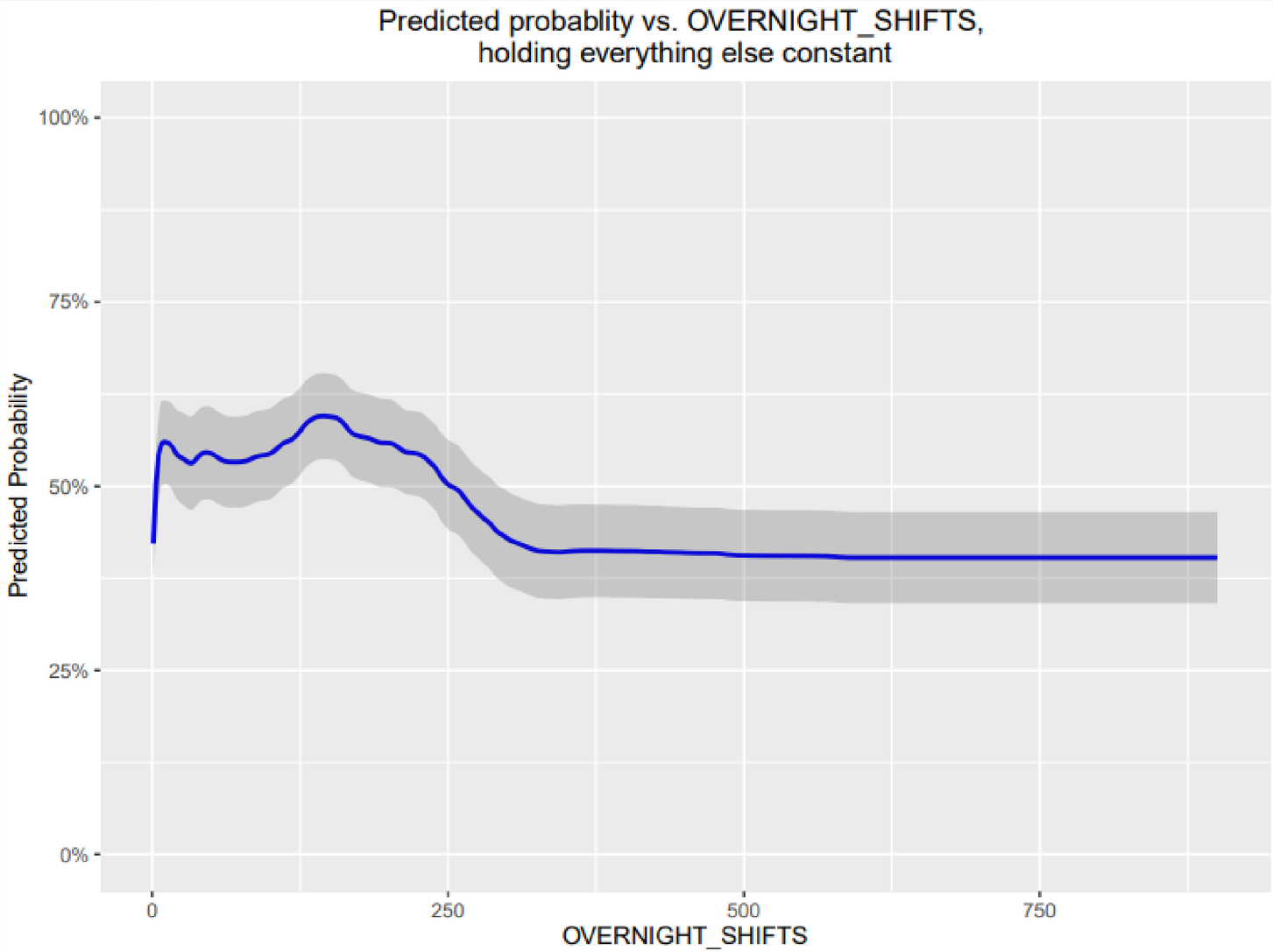
- Employees with less than four shifts per week on average have a probability of quitting between 32% and 45%



Employees with fewer shifts per weekend are more likely to quit.

PROBABILITY PLOTS FROM THE TOP SIX VARIABLES OF IMPORTANCE: OVERNIGHT SHIFTS

- Employees with 250 or fewer overnight shifts have a probability of leaving the company between 50% and 60%.
- This probability is reduced to nearly 40% when the employees have more than 250 overnight shifts.



Employees with more overnight shifts are less likely to leave the company

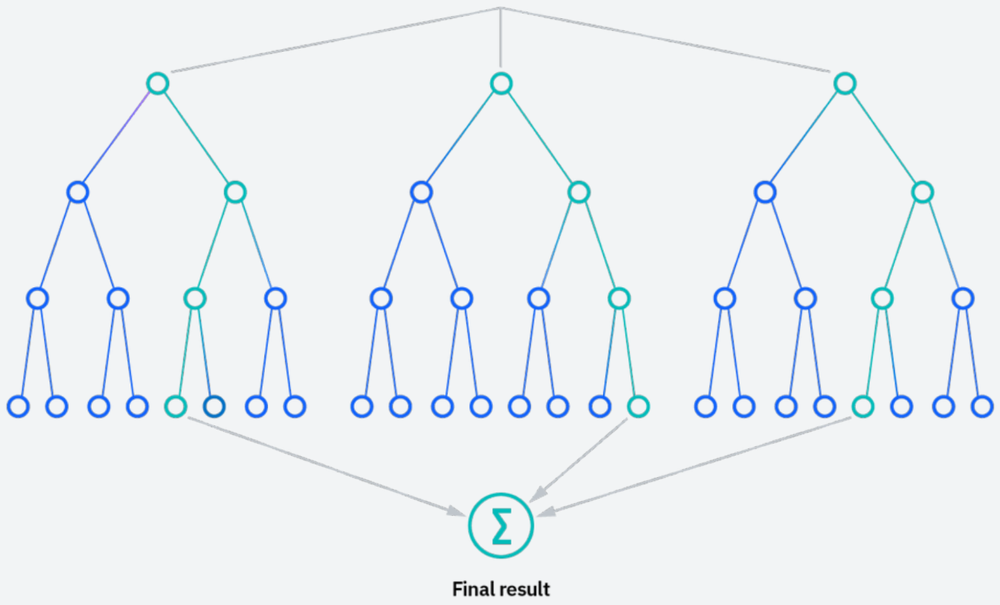
CONCLUSION

QUESTIONS & ANSWERS

ACKNOWLEDGEMENTS

WHAT IS RANDOM FOREST

- Random forest is the most simple and widely used algorithm. Used for both classification and regression. It is an ensemble of randomized decision trees. Each decision tree gives a vote for the prediction of target variable. Random forest reduces the variance of the individual decision trees by randomly selecting trees and then either average them or picking the class that gets the most vote.



“There should be some actual values in the feature variables of the dataset, which will give the classifier a better chance to predict accurate results, rather than provide an estimation. Missing values should be handled from training the model. The predictions from each tree must have very low correlations.”
(https://www.simplilearn.com/tutorials/data-science-tutorial/random-forest-in-r)