



HR TALKS

Attrition at IBM

What factors are influencing employee turnover at IBM?

Presented by : Vickie Yang

Attrition at IBM

PROJECT GOAL

This data analysis predicts employee attrition and studies the factors that contribute to turnover.

STAKEHOLDERS

The primary stakeholder of this analysis are HR and IBM Business Unit leads.

PROJECT USE CASE

Talent retention and succession planning.

Dataset Overview

DATA SHAPE

1470 data points, 35 features.

Target feature: 'Attrition' - 237 attrition employees.

DATAPOINT DETAILS

Features: income, personal information, job satisfaction, training time, department, job level, etc.

Departments: Sales', 'Research & Development', 'Human Resources'

DATA CLEANING

Converted to binary values - 'Attrition'

Removed irrelevant features, conducted multicollinearity analysis

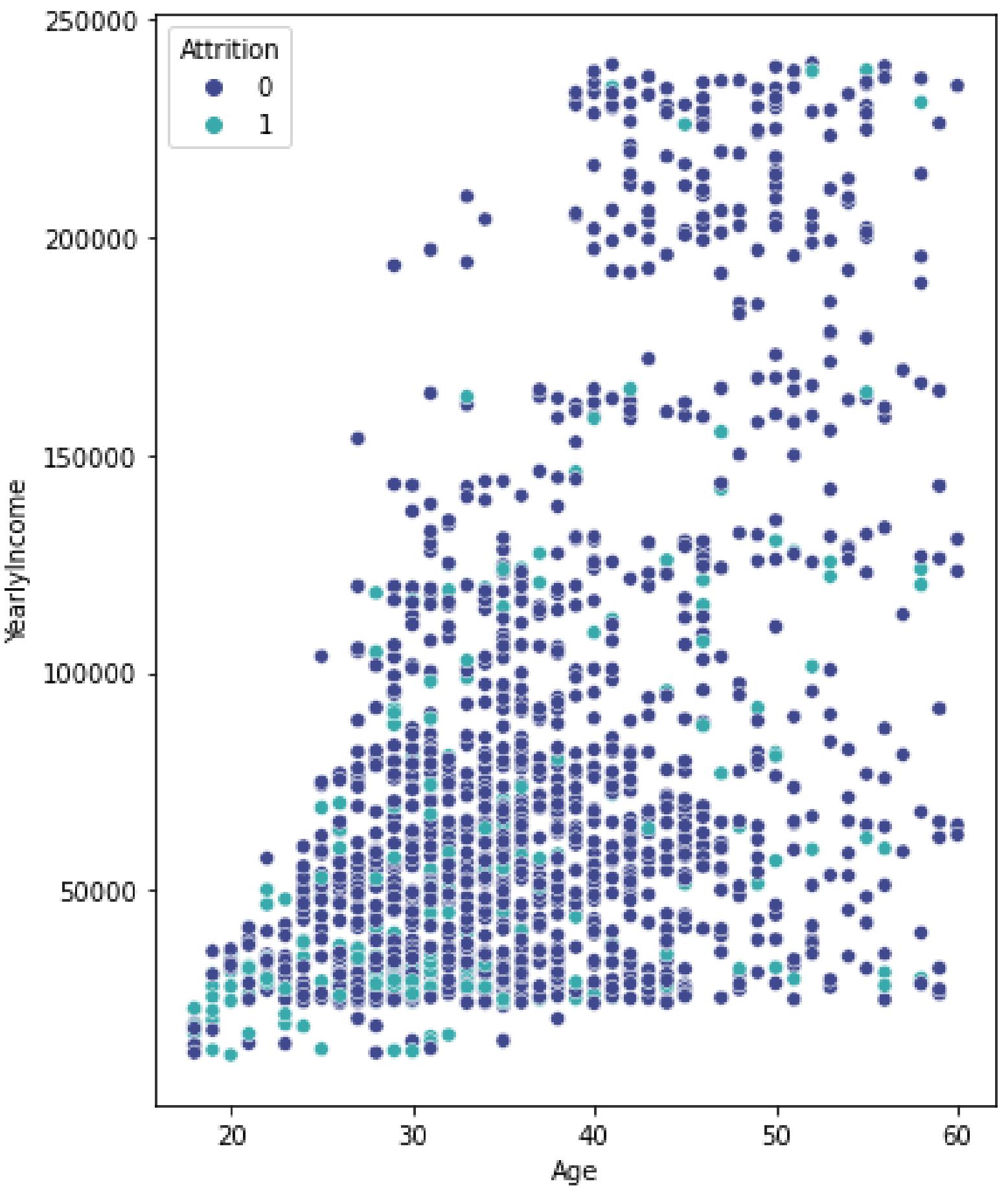
Attrition Overview by Age and Income

The majority of attrition employees are around and below 30 years old.

Most attrition employees are earning below ~50k.

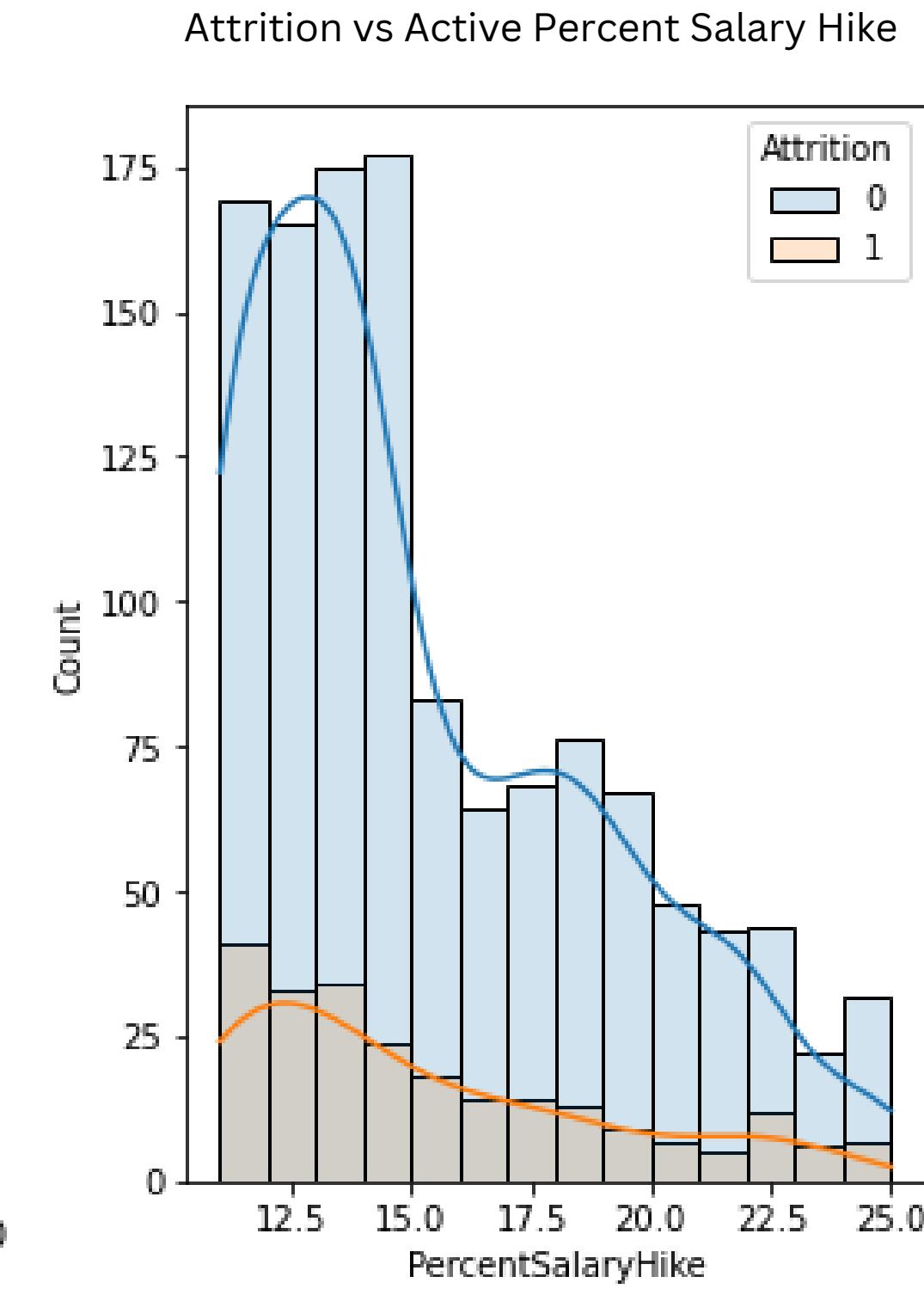
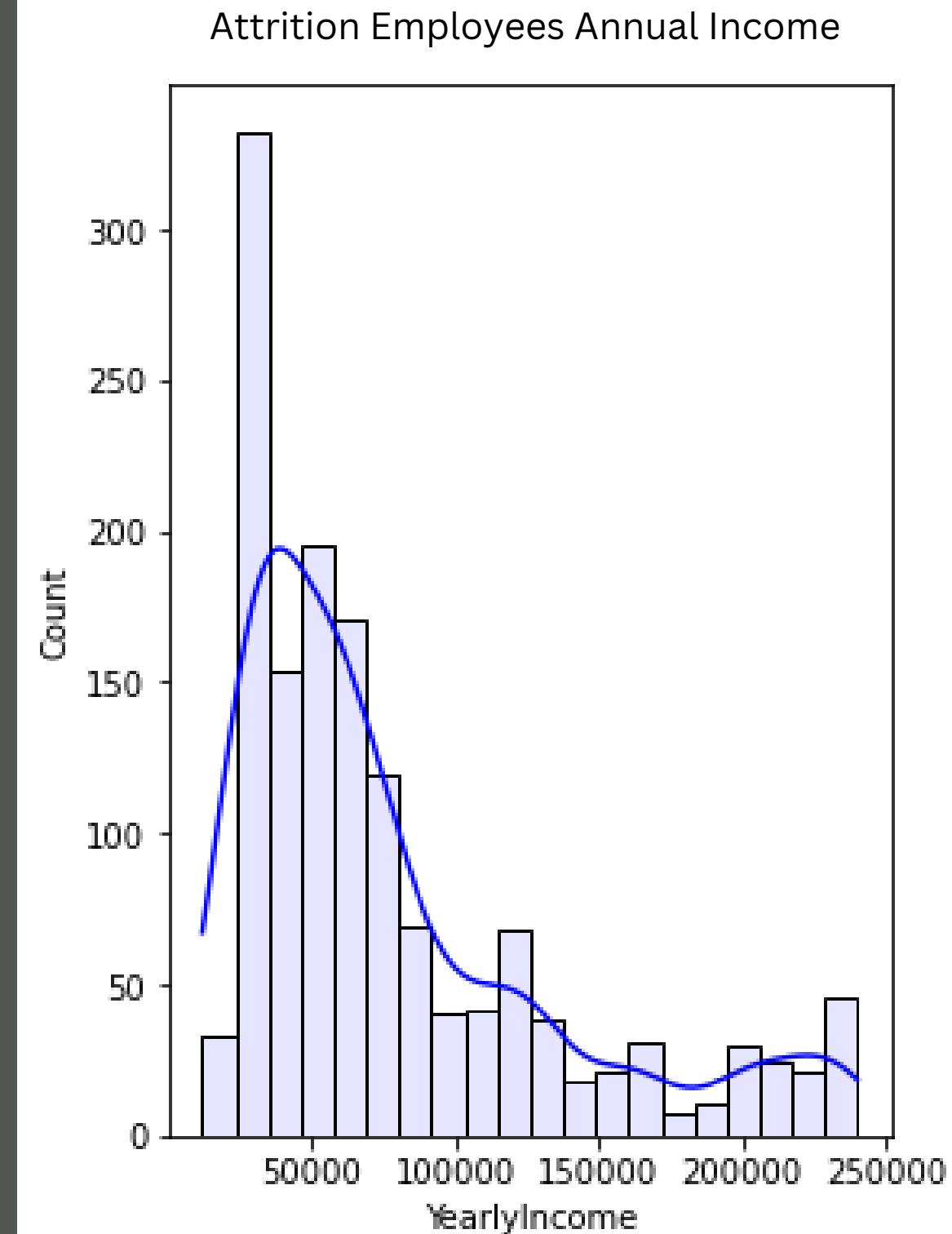
Target 0 = Active

Target 1 = Attrition



Income Level for Attrition Employees

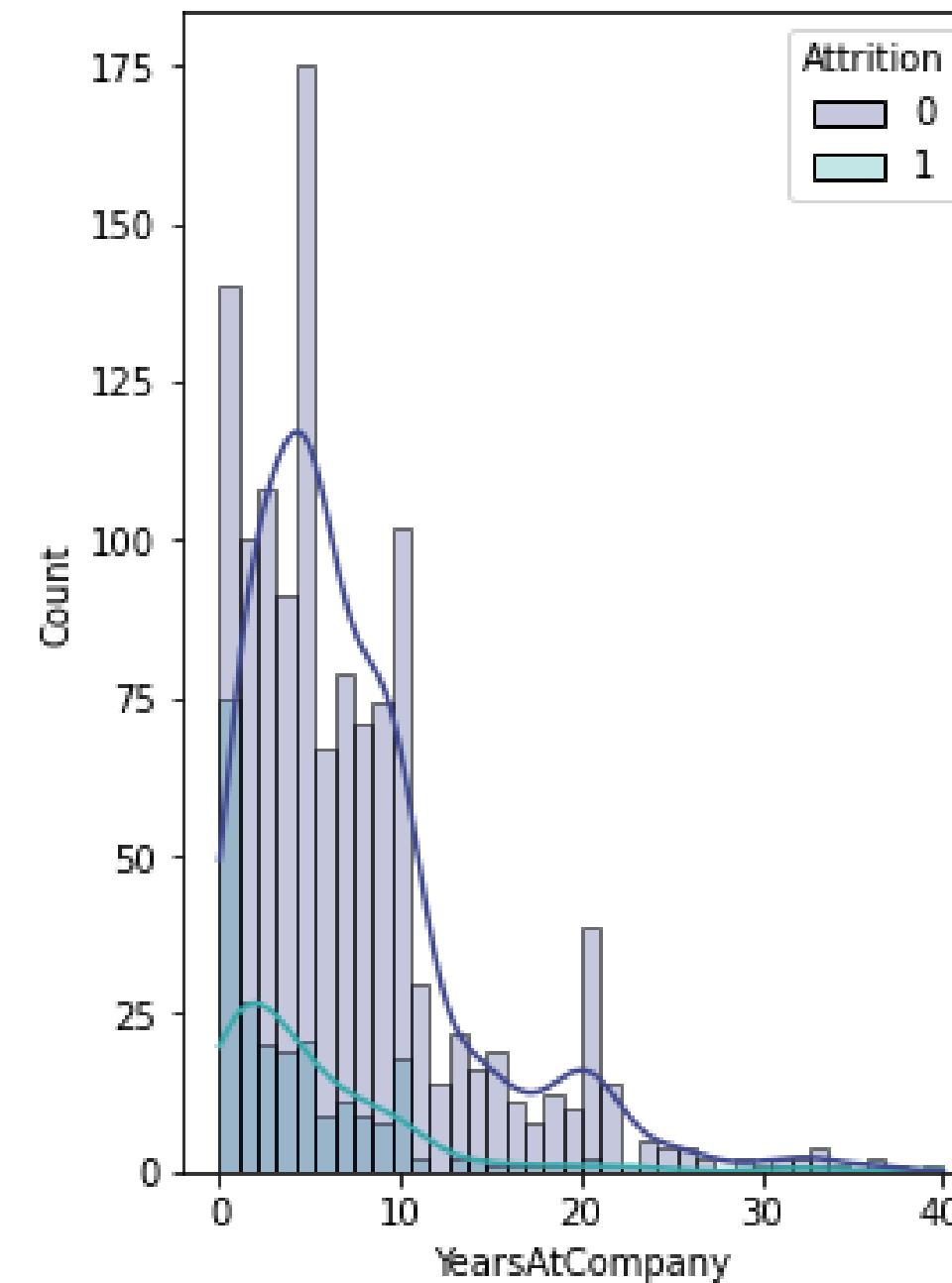
The majority of terminated employees are below ~50k and less than 14% of salary increase



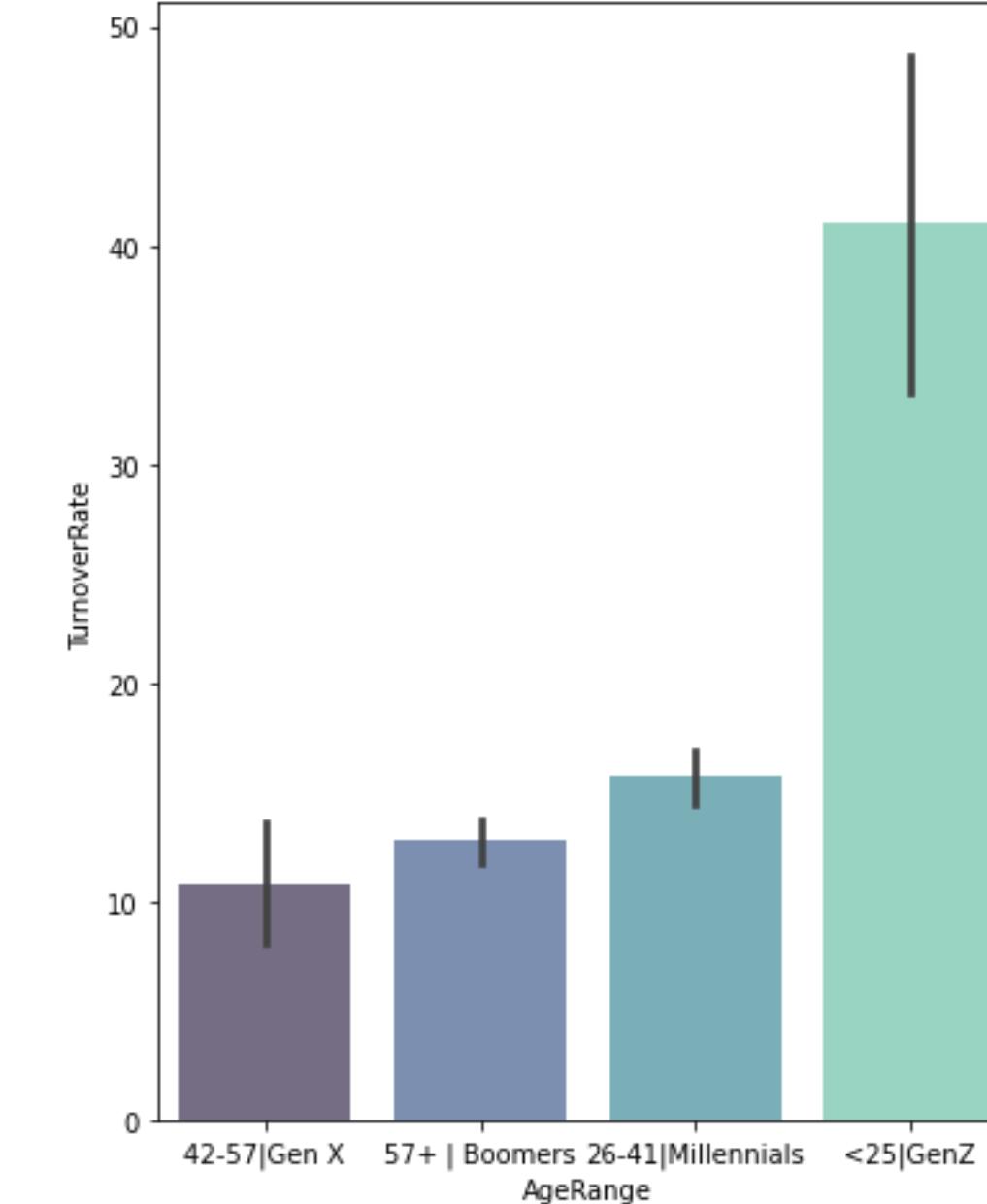
ATTRITION I

Years at Company, Age Distribution

Attrition vs Active Employees Yrs of Service



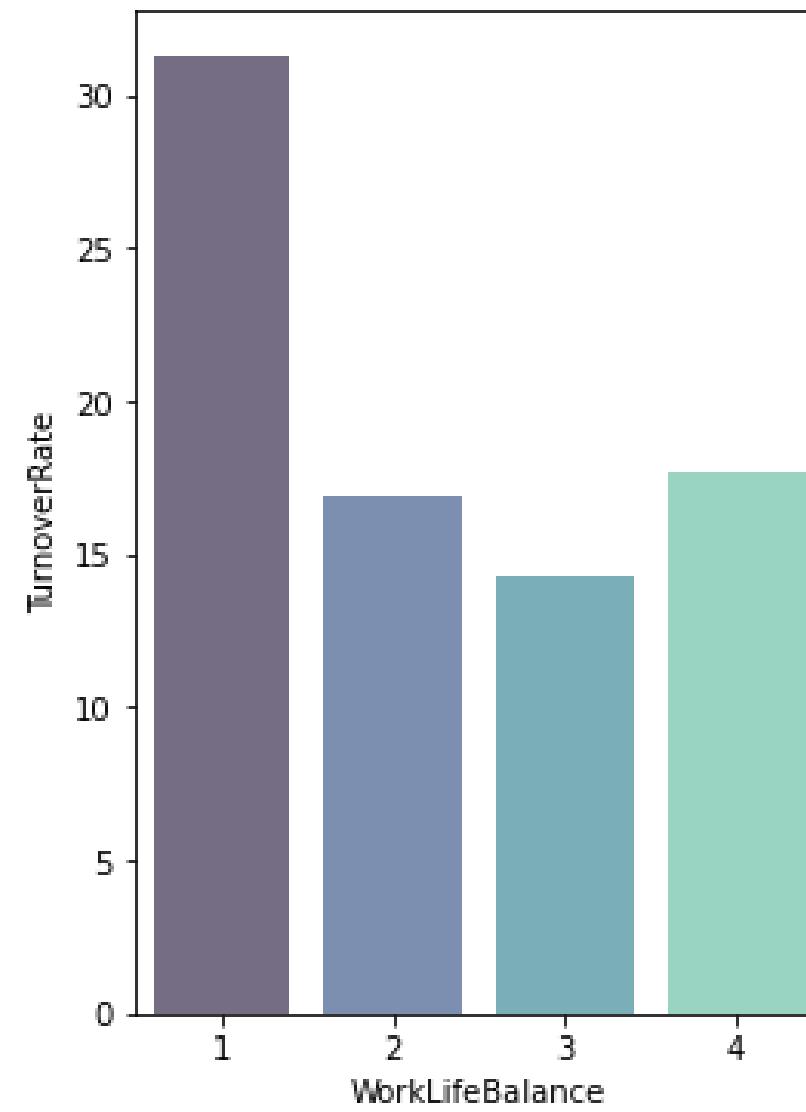
Turnover Rate by Generation Groups



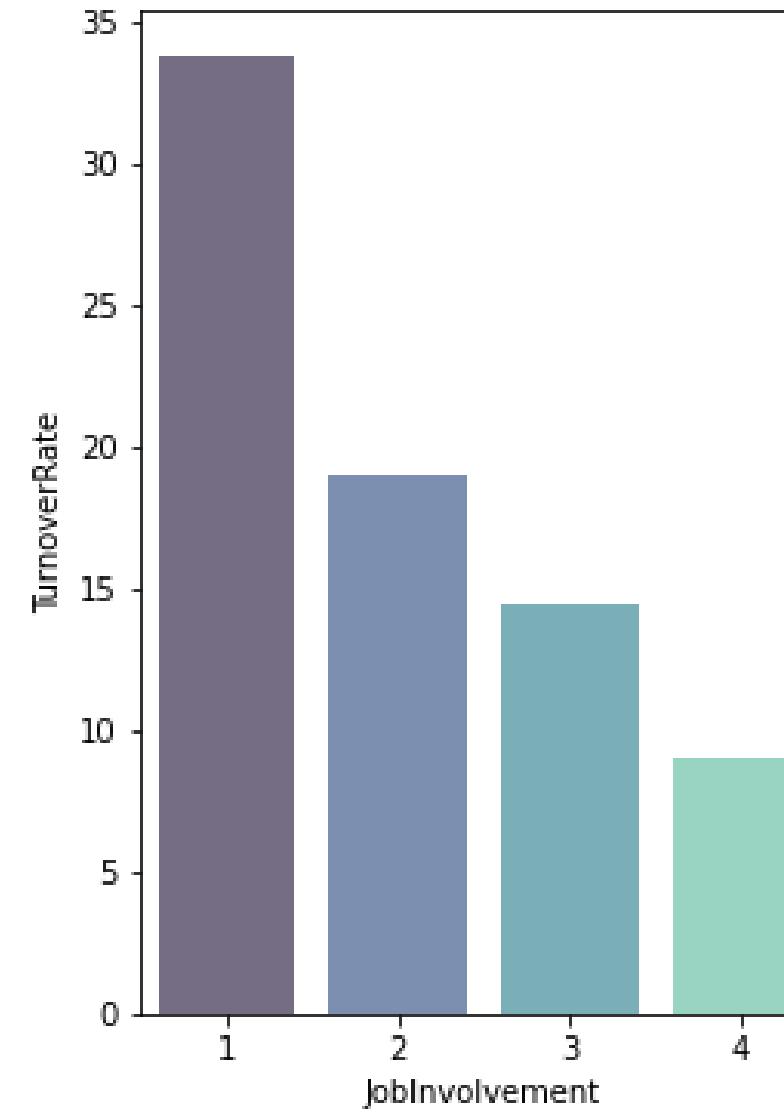
- Employees who are with IBM less than 5 years are most likely to leave the company
- Younger age groups (Millennials and Gen Z) have higher turnover rates

Turnover Rate | Worklifebalance, JobInvolvement, JobSatisfaction

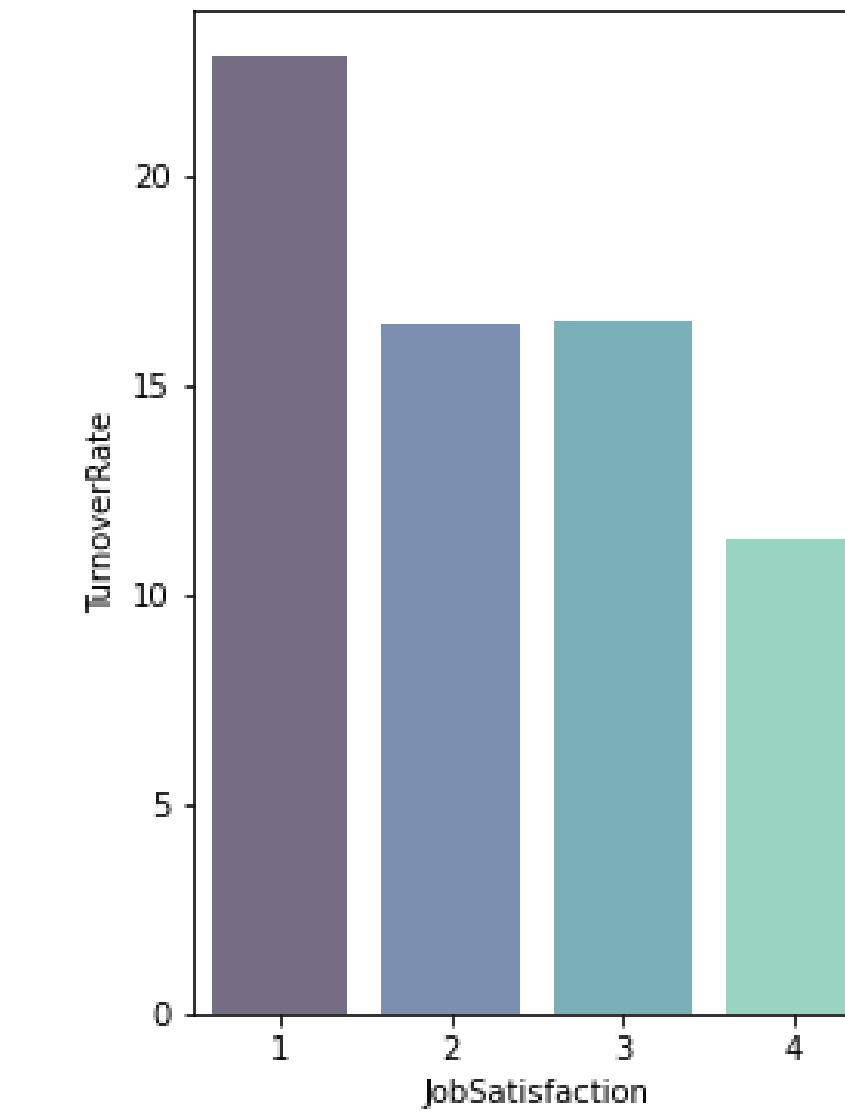
Attrition is heavily impacted by Work Life Balance, Job Involvement, Job Satisfaction



Sales (37.5%) has the highest turnover in the employees who rated "1" for work life balance



Sales has the highest turnover in the employees who rated "1" for Job Involvement



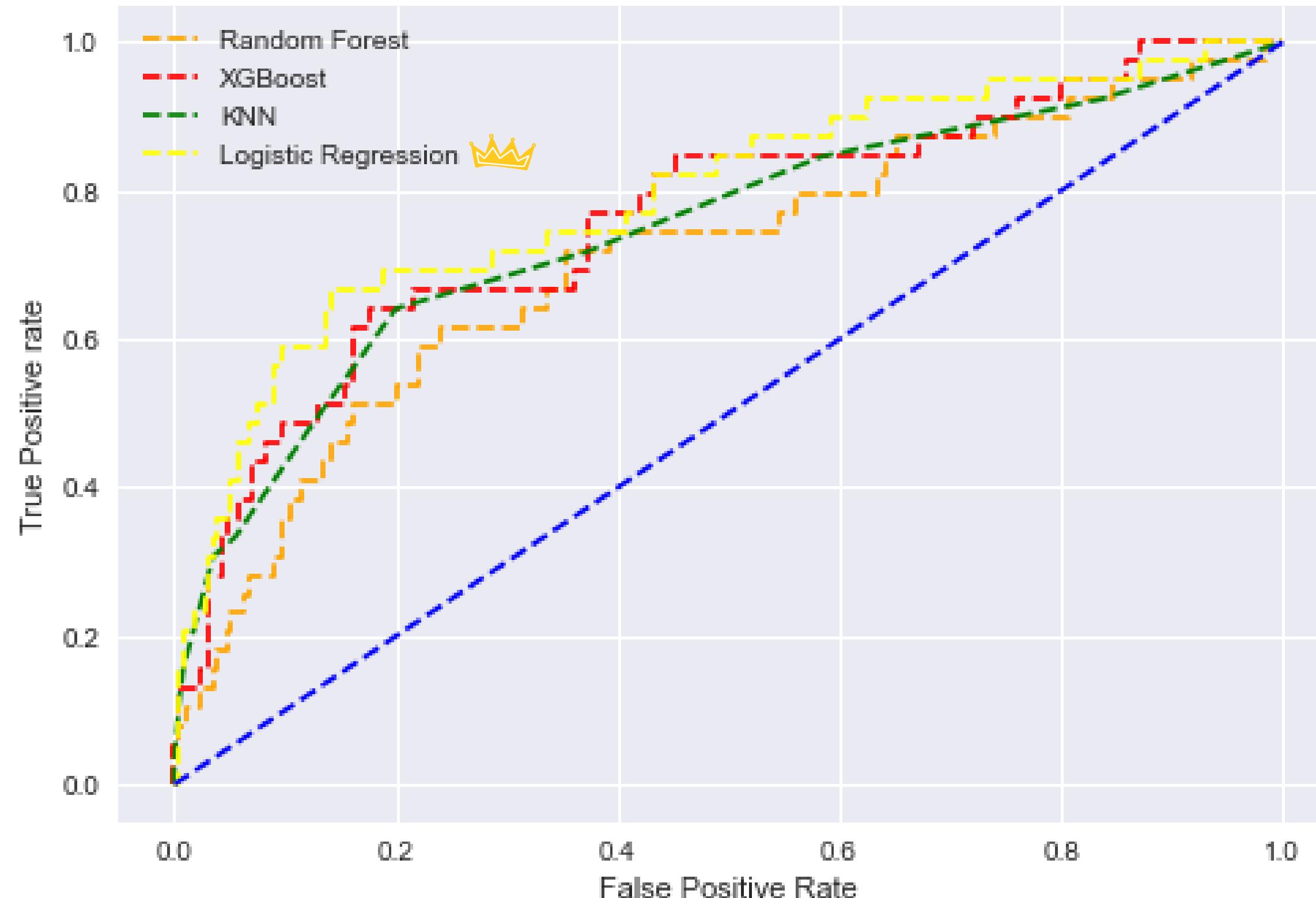
Human Resources (45%) has the highest turnover in the employees who rated "1" for Job Satisfaction

Model Testing



Logistic Regression		Decision Tree		K Nearest Neighbour		XG Boosting		Random Forest	
0.89	Accuracy	0.78	Accuracy	0.88	Accuracy	0.85	Accuracy	0.85	
0.64	Precision	0.23	Precision	0.57	Precision	0.43	Precision	0.36	
0.46	Recall	0.28	Recall	0.30	Recall	0.38	Recall	0.17	
0.53	F1	0.25	F1	0.40	F1	0.40	F1	0.24	

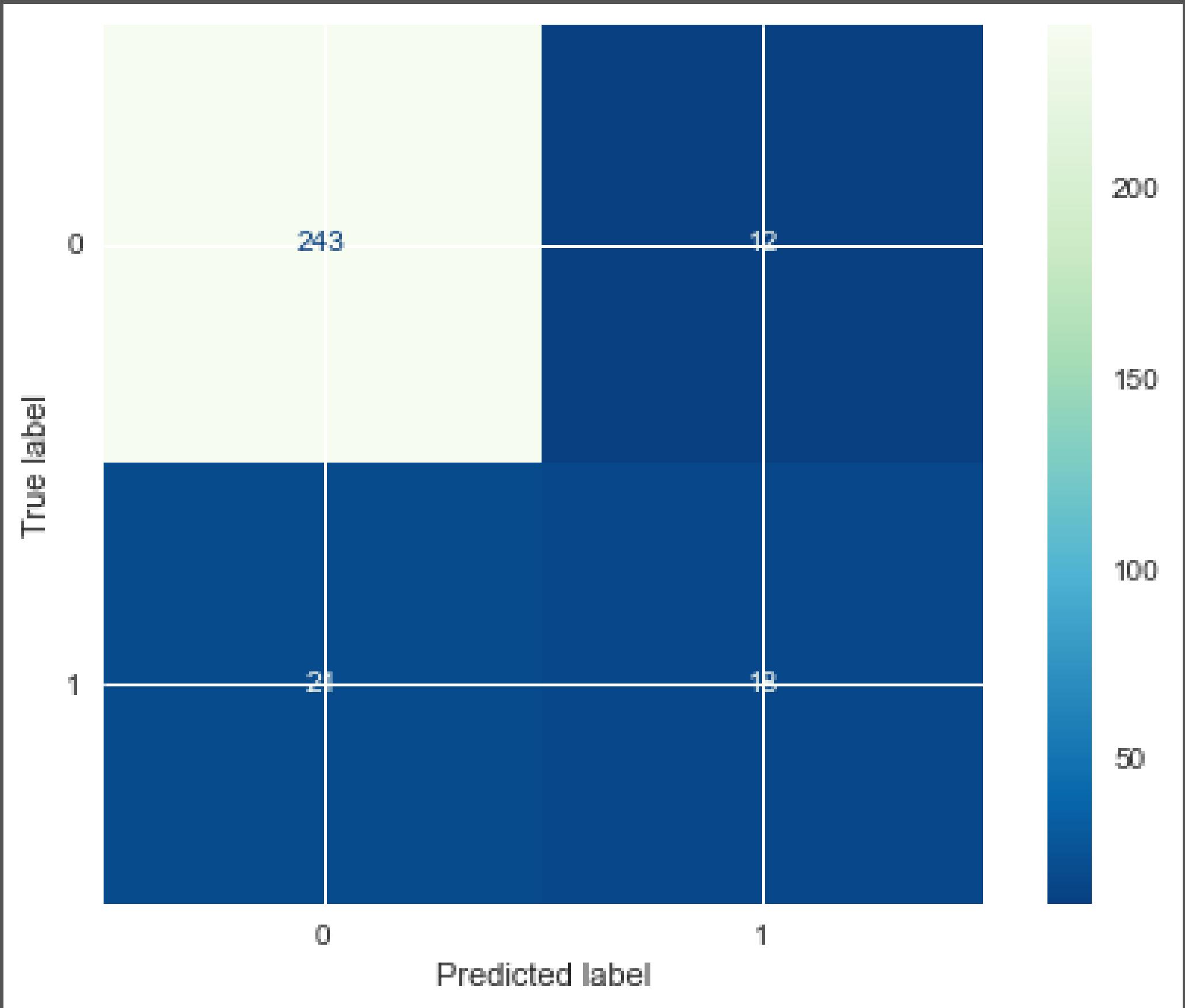
ROC Curve



Prediction Model - Logistic Regression

1 = Attrition
0 = Active

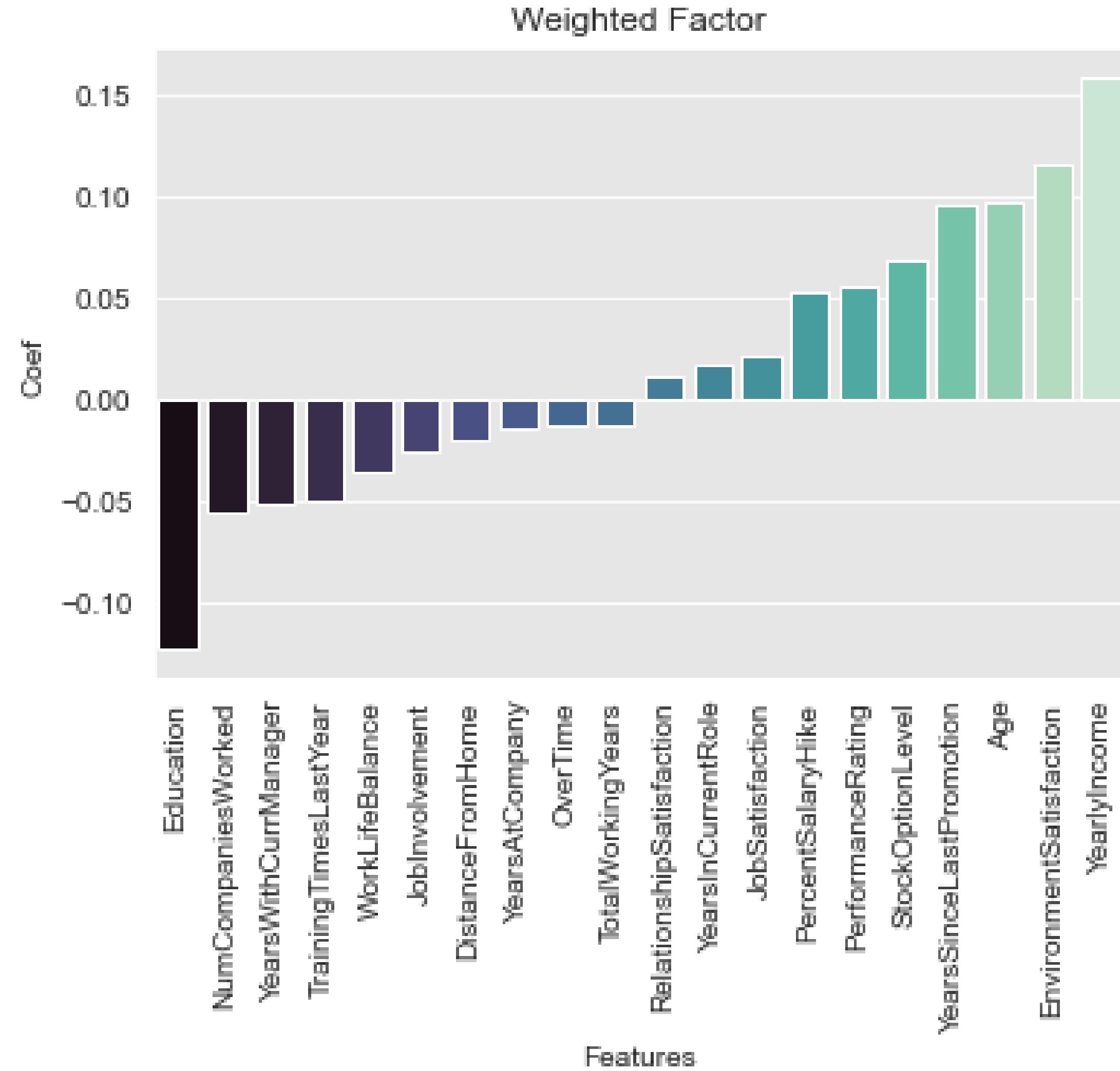
True Attrition (positive): 18
False Attrition (positive): 24
True Active (negative): 243
False Active (negative): 12



Feature Weights

For Attrition:
Yearly Income
EnvironmentSatisfaction
Age

For Active Employees :
Education
Numcompaniesworked
YearsWithCurrentMgr



What's Next

- Obtain more data to predict trend in attrition (time series, attrition reason, etc.)
- Obtain external features (i.e. unemployment rate, industry data, etc.)
- Continue tuning the models to achieve better results



Thank you!



tq20018@gmail.com

www.linkedin.com/in/yangvickie



t36yang