

Title: IBM HR Analytics Employee Attrition Modeling with Python Data Science

Author: Jon Ramangwana

Date: 3 December 2023

Table of Contents

1. Introduction.....	3
1.1 Background	3
1.2 Aim.....	3
1.3 Objectives.....	3
1.4 Research question.....	4
2. Method	4
3. Results.....	5
4. Analysis.....	10
5. Conclusion	11
6. Recommendations.....	11
7. References	11

1. Introduction

1.1 Background

Employee attrition has a lot of effects that are not desired by different companies because profit gets lost when employees leave. Cost of getting new employees, training, tax considerations etc. affect financial performance. IBM a big business in different countries employs many individuals. This makes retention of employees a huge concern. To look at causes of employees leaving their company IBM needed a prediction of reasons for this attrition.

A company that cannot retain employees can lose profit in a form of annual wages for each employee, social programs, mental health issues etc. Understanding factors affecting attrition can be important for any organization considering social and cooperate consequences (Skelton, 2018).

A huge challenge to deal with this can be a company's ability to keep its workers. When employees leave main cause of this would be lack of attractive work conditions. Looking at budget, company performance, cost-effectiveness, etc. there can be a challenge to correct these issues (Astivik et al, 2020).

Many companies know about different disadvantages of employee attrition. This cause them to try to retain a lot of employees by providing benefits at work. Information about why employees leave help Human Resources personnel to keep valuable employees. Information would need to get analyzed to get insights to make decisions for improving retention (Boomhower et al, 2018).

Quality of employees affects how a company grows. Artificial intelligence has an effect on employee management to improve quality when applied correctly. Analysis of factual data rather than assumptions can assist a company to make decisions that can exponentially grow business at the same time provide clients with needed services (Fallucchi et al, 2020).

Getting skills, developing skilled workers, retaining workers, planning for future activity etc. can be done effectively with Data Science. This also helps to understand company responsibilities to drive growth (Kaewnaknaew et al, 2022). This project aims to find factors that affect IBM employee attrition by building a logistic regression model to predict employees who might leave this company among other Data Science methods.

1.2 Aim

1. To model employee attrition for IBM.

1.3 Objectives

1. To find age distribution of employees.
2. To explore attrition by age.
3. To explore data for employees who left.
4. To find education distribution of employees.
5. To provide a bar chart for number of married and unmarried employees.
6. To build a logistic regression model to predict employees who are likely to attrite.

1.4 Research question

An American multinational company named IBM found in 170 countries that do computing, software, hardware, business need to know what affects attrition of employees. Indispensable employees may leave this company to cause profit loss hence this concern. A question that need an answer would be:

1. What makes employees leave IBM?

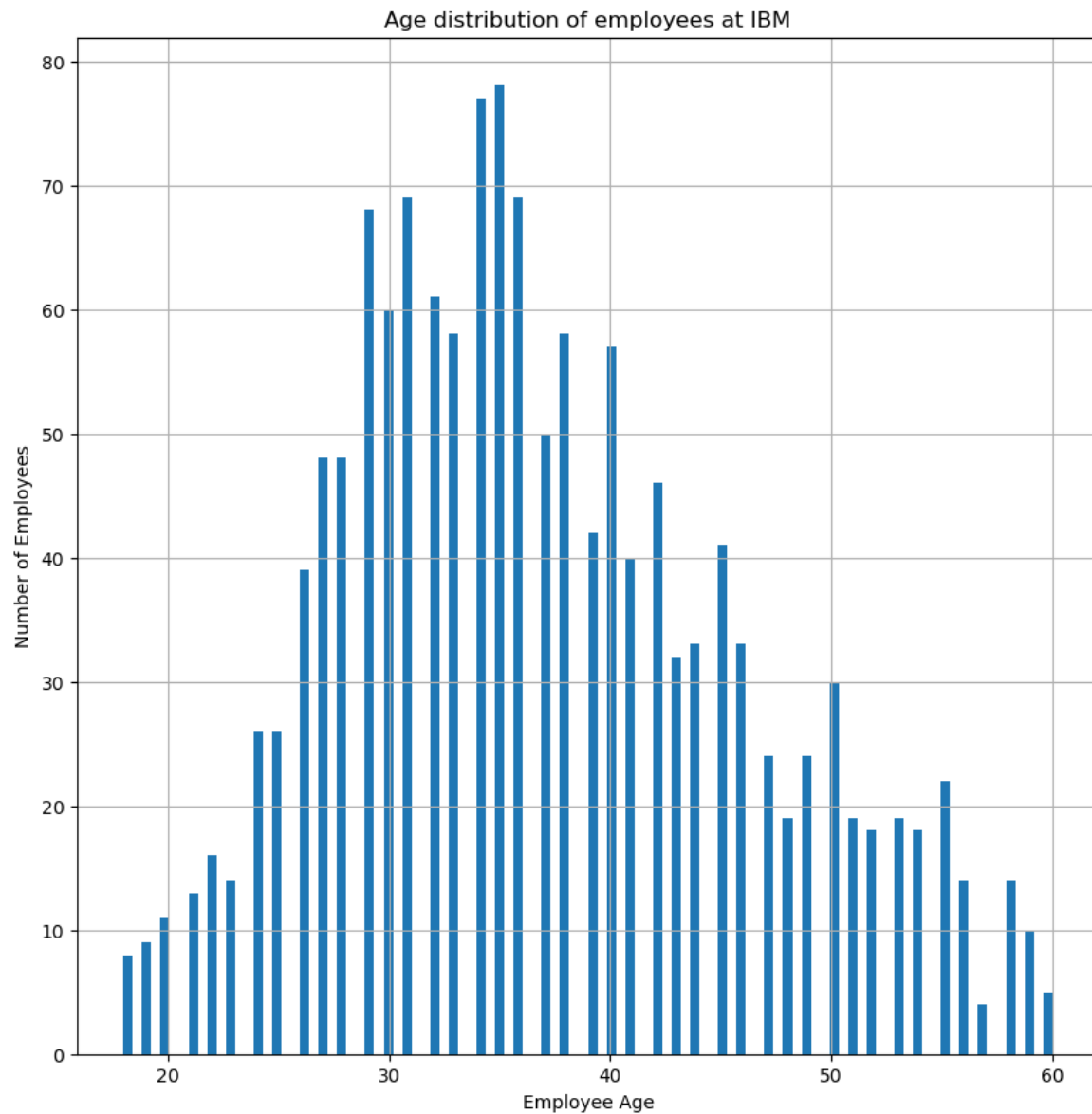
2. Method

IBM attrition dataset was downloaded. This dataset was imported to Jupyter notebook together with libraries such as matplotlib.pyplot, pandas, numpy, seaborn etc. Exploratory data analysis was done. Steps followed to do analysis include:

1. Finding employee age distribution.
2. Exploring attrition by age.
3. Exploring data for employees who left.
4. Finding distribution of employees by education field.
5. Provide a bar chart for employee marriage status.
6. Build a logistic regression model to predict which employees are likely to attrite.

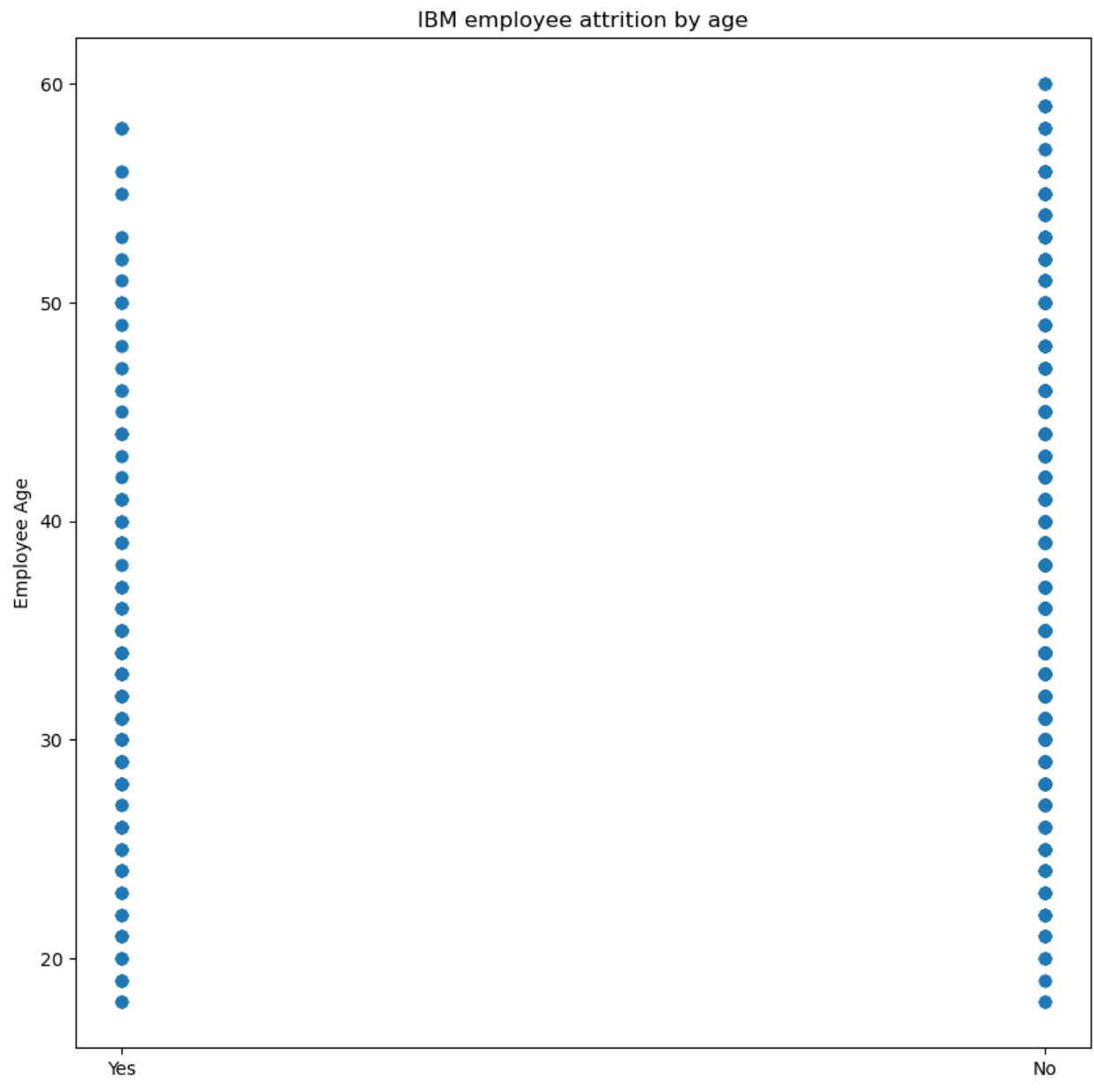
3. Results

Graph 1



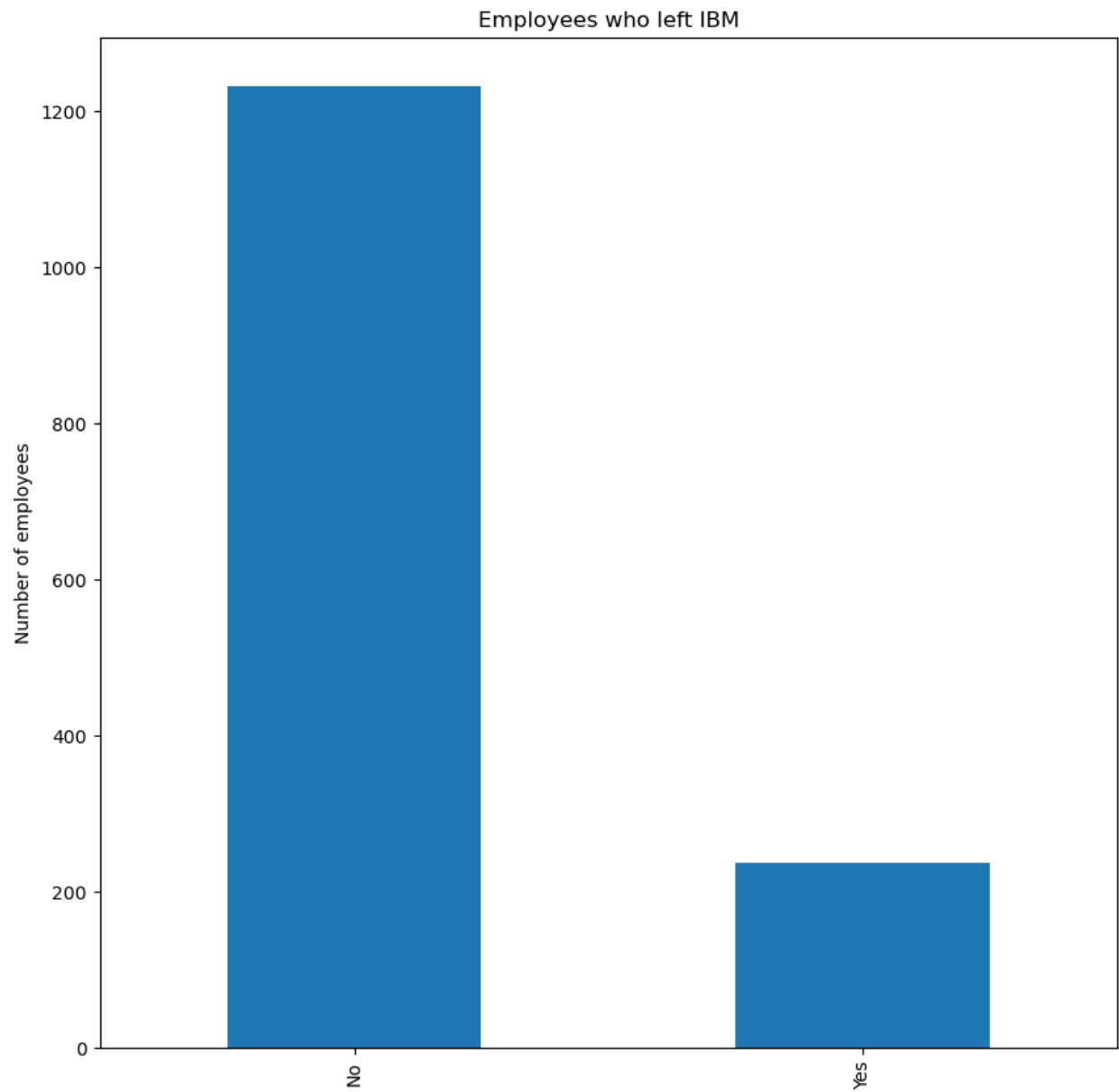
Histogram above shows age distribution of IBM employees.

Graph 2



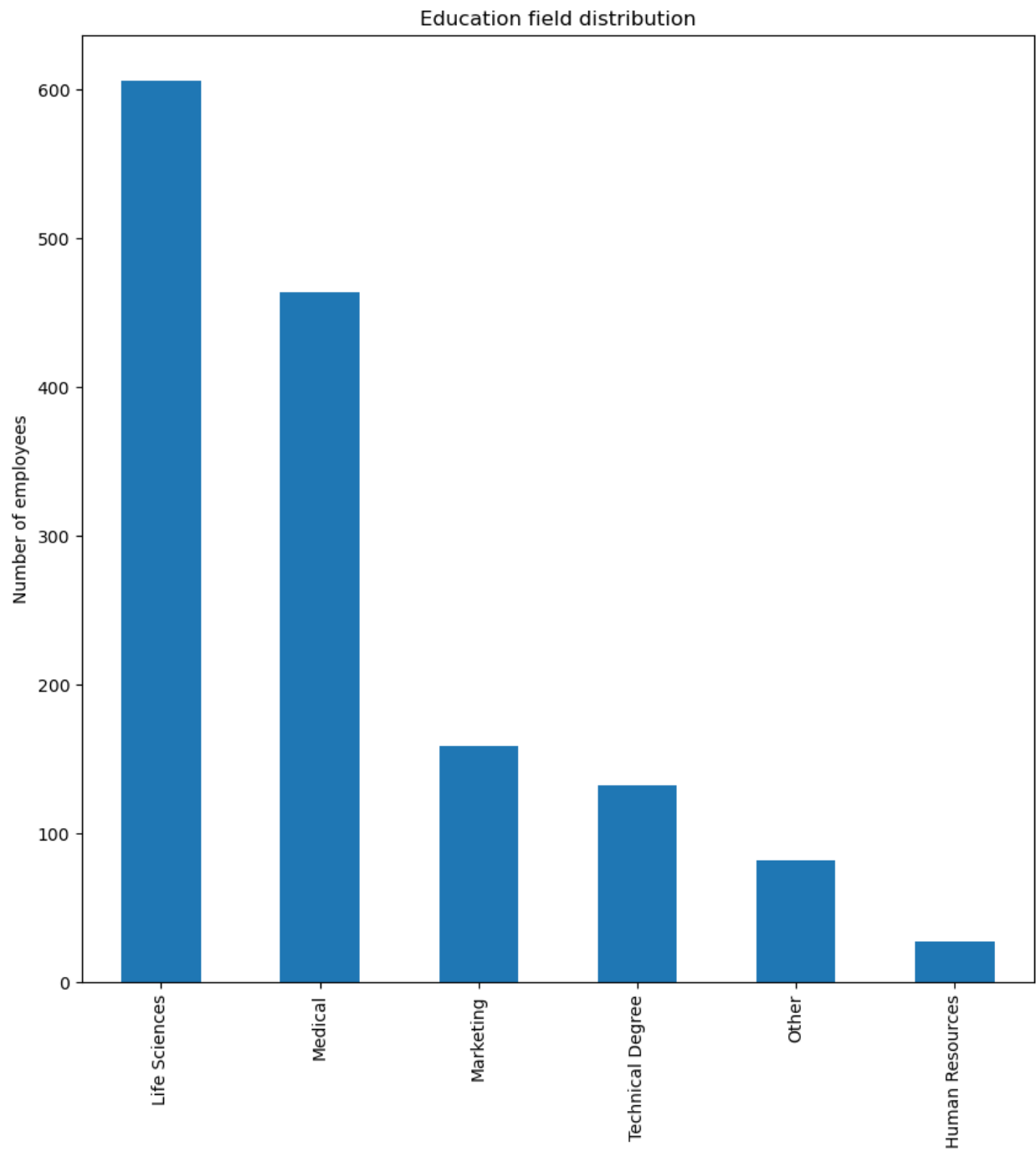
Scatter plot above shows IBM employee attrition by age.

Graph 3



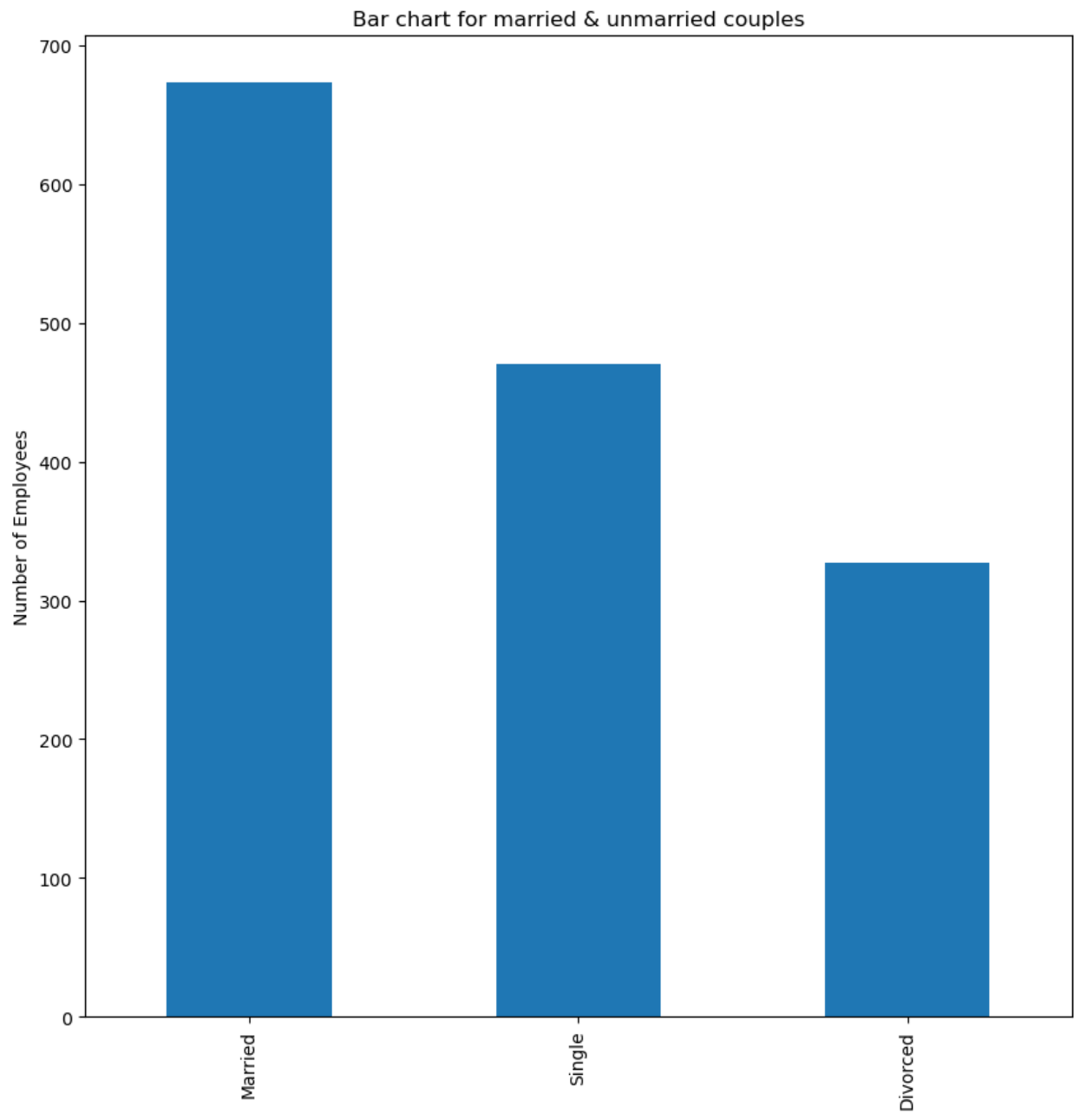
Bar graph above shows employees who left IBM.

Graph 4



Bar graph above shows IBM employee distribution by education.

Graph 5



Bar graph above shows IBM employee marriage status distribution.

Logistic Regression Model

1. Model score = 0.8401360544217688
2. Model precision = 0.8435374149659864
3. Roc- Auc = 0.646207162110127
4. Probability of attrition = 2.659×10^{-4} - 9.997×10^{-1} to 4 s.f

5. Confusion matrix:

[[371	0]	maps to	[[True positive	False positive]
[69	1]]		[False negative	True negative]

4. Analysis

Each result was observed to get insights from IBM employee attrition dataset. Graph 1 shows age distribution of IBM employees with number of employees on y-axis, employee age on x-axis. Old employees were at the age of 60, minimum age was below 20 years old. Age ranges with few employees were 10 – 24 & 50 – 60. Many employees were between 25 & 45. Most employees were at the age of 35. Age range of 10 – 24 had few people likely because at this age most people stay with parents while they go to school therefore there are less employed individuals. There are few employees between 50 & 60 because it's close to retiring age people would be getting old. Range of 25 to 45 has high number of employees with peak at the age of 35. A reason for this might be that people are active at this age. A lot of bread winners are at this age range.

Graph 2 shows scatter plot for IBM employee attrition by age. There are two separable classes on this plot first plot labeled “Yes” on x-axis for employees who left, second plot labeled “No” for employees who stayed at this company. “No” plot had each age value represented while “Yes” plot has missing age values above 50. When people get above 50 they might leave to take a break. Old age would be a reason for this.

Graph 3 shows a bar graph for employees who left IBM. Bar labeled “No” on x-axis represents workers who stayed. IBM retained over 1200 workers compared to over 200 workers who had to attrite. Employees who left translate into approximately 14% of sum of employees.

Graph 4 shows a bar graph for employee distribution by education. Life sciences education carry a huge number of employees compared to other education fields. Life sciences had 600 employees followed by Medical field with 450. There was a big decrease in number of employees from 450 to 150 in department of Marketing then a little below 150 for Technical degree. Other Education fields had a total below 100 while Human resources education field had less than 50 employees. These numbers depend on IBM, company needs. They would be expected to get more employees from an education field that can provide them with relevant skills for their business.

Graph 5 shows a bar chart for married & unmarried employees. There are more than 650 married employees as shown by first bar. Single employees are above 450 while divorced employees comprise of over 300 employees, this was a minimum. Many employees were married indicating stability at work. Another reason for many married employees at IBM can be that maximum number of employees are between 25 & 45 as shown on graph 1. At this age many people would be married.

Logistic regression model had a score of 84.0% & precision of 84.3% showing a decent model. Roc – Auc percentage was 64.6% a value that represents decent separability of classes also showing a decent model. Confusion matrix had a true positive of 371, a false positive of 0, a false negative of 69, a true negative of 1. False values were low compared to true values. Probability of attrition

rition for random employees entered to model was between 2.659×10^{-4} & 9.997×10^{-1} . Chance of these employees leaving the company was low compared to chance of employees staying at IBM.

5. Conclusion

Main cause of employee attrition shown by this specific data exploration was old age. IBM was retaining most of its employees with low attrition.

6. Recommendations

This company was keeping many employees meaning their HR department was taking correct steps to retain employees. Further steps can be followed to reduce number of people who leave such as:

1. Improve benefits.
2. Improve working conditions to retain valuable old employees.
3. Maintain favorable conditions provided to employees.

7. References

1. Astvik, W., Welander, J., Larsson, R., (2020), Reasons for staying: Longitudinal study of work conditions predicting social workers' willingness to stay in their organization, The British Journal of Social Work, 50(5), 1382-1400, <https://doi.org/10.1093/bjsw/bcz103>
2. Fallucchi, F., Coladangelo, M., Romeo, G., De Luca, E.W., (2020). Predicting employee attrition using machine learning techniques, Employee Attrition: What Makes an Employee Quit?, Computers, 9, 86, <https://doi.org/10.3390/computers9040086>
3. Frye, A.L., Boomhower, C., Smith, M., Vitovsky, L., Fabricant, S.M. (2018). Employee Attrition: What Makes an Employee Quit?, SMU Data Science Review, 1(1), 9, <https://scholar.smu.edu/datasciencereview/vol1/iss1/9>
4. Kaewnaknaew, C., Siripipatthanakul, S., Phayaphrom, B., Limma, P., (2022). Modelling of talent management on construction companies' performance: A model of business analytics in Bangkok, International Journal of Behavioral Analytics, 2(1), 1-17, <https://ssrn.com/abstract=4018709>
5. Skelton, A.R., Nattress, D., Dwyer, R.J., (2020), Predicting manufacturing employee intentions, Journal of Economics, Finance and Administrative Science, 25(49), 101, <https://doi.org/10.1108/JEFAS-07-2018-0069>