

SULIT



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

FACULTY OF COMPUTING
UTM Johor Bahru

**UNIVERSITI TEKNOLOGI
MALAYSIA**

SUBJECT CODE : **MCSD1013**

SUBJECT NAME : **BUSINESS INTELLIGENCE AND ANALYTICS**

SECTION : **01**

TIME : **2 weeks**

DATE/DAY : **26/1/2023**

VENUES :

INSTRUCTIONS:

This alternative assessment is an **INDIVIDUAL** assessment consists of **THREE (3)** questions/phases:

- 1) Data preparation
- 2) Dashboard development
- 3) Reporting of analysis

Each student will be evaluated based on **ALL** the questions/phases given.

(Please attach this page as the front page of your report)

(Please Write Your Lecturer Name And Section In Your Answer Booklet)

Name	Chong Xian Jun
I/C No.	980714-01-7629
Year / Course	2022231/MCD1013
Section	01
Lecturer Name	Dr. Nor Erne Nazira binti Bazin

This question paper consists of 22 () printed pages excluding this page.

Alternative Assessment Report

Name : Chong Xian Jun

Matrix No.: MCS211047

1.0 Introduction

1.1 Business Problem Formulation

1.2 Research Questions

2.0 Data Preparation

2.1 Data Description

2.2 Data Preprocessing

3.0 Report of Analysis

3.1 Analytics on Feature Importance for Attrition in IBM

3.2 EDA on Attrition of Employees by Performance Rating

3.3 Career Prospect vs Attrition of Outstanding Employees

3.4 Work Life Balance vs Attrition of Outstanding Employees

3.5 Monitor Dashboard - Predicting Probability of Attrition for Each JobRole

4.0 Conclusion

5.0 Attachments

1.0 Introduction

Problem Statement

The tech industry is currently experiencing an alarming attrition rate of over 20% among their employees despite the competitive market. To tackle this problem, companies need to carefully analyze their HR department's employees attrition data in order to develop continuously optimized strategies for retaining top talent.

Risk

Companies risk losing their competitive advantage to rivals who are better at retaining top talent should they fail to implement optimized retention strategies based on data-driven insights.

Problem Background

In the technology industry, where the value of talent resources is increasingly important for organizations, companies are willing to invest heavily in order to attract and retain top talent. While there have been layoffs and hiring freezes in the mega-tech companies recently, the attrition rate in tech industry is predicted to remain as high as 20% for 2022-2023 according to HR experts of HCLTech.

This high level of attrition becomes a major risk to tech companies that will not be resolved in the near future. Hence, in this study, we make use of a set of fictional IBM employee data released by IBM (source: IBM HR Analytics Employee Attrition & Performance | Kaggle) to uncover the impact of different factors on the attrition rate of employees in IBM, extracting insights to help optimize their talent management. As talent optimization is a complex multivariate problem, an intuitive interface that allows the HR managers to look into all of the important features easily becomes gravely needed.

Stakeholder

HR managers in IBM

1.1 Research Questions

This project revolves around one single research question: “How to retain more talent in IBM?”

The content of the analysis is then broken down into several parts:

1. What are the most important features contributing to attrition?
 - Find feature importance of each feature on attrition rate
 - Visualize and analyze the few most important factors.
2. What are the strategies that can be used to optimize talent retention at different times?
3. How can predictive analytics be used to increase the managerial capabilities of the human resources department to better manage employee attrition and retain more talent?

2.0 Data Preparation

2.1 Data Description

In this study, we make use of a set of fictional IBM employee data released by IBM (source: [IBM HR Analytics Employee Attrition & Performance | Kaggle](#)) to uncover the impact of different factors on the attrition rate of employees in IBM, extracting insights to help optimize their talent management. The attributes of the imported data are described in detail in the table below.

No.	Attribute Name	Data type	Description	Example
1	Age	int	The age of employees.	18, 19, 20,...
2	Attrition	category	The state of attrition of employees.	'Yes', 'No'
3	BusinessTravel	category	Shows how often an employee goes on a business trip.	'Travel_Frequently', 'Travel_rarely', 'Non-Travel'
4	DailyRate	int	The average daily pay of an employee.	102, 103, 200,...
5	Department	category	The department where an employee works in.	'Human Resources', 'Research & Development', 'Sales'
6	DistanceFromHome	int	The distance from work to home. (assumed to be in km)	1, 2, 3...
7	Education	int	The education level of employee (1: 'Below College', 2: 'College', 3: 'Bachelor', 4: 'Master', 5: 'Doctor')	1, 2, 3, 4, 5
8	EducationField	category	The education field of an employee.	'Life Sciences', 'Medical', 'Technical Degree',...
10	EmployeeNumber	int	The employee ID.	1, 2, 4,...
11	EnvironmentSatisfaction	int	The level of employee's satisfaction on the environment. (1: 'Low', 2: 'Medium', 3: 'High', 4: 'Very High')	1, 2, 3, 4
12	Gender	category	Gender of employees.	'Male', 'Female'

13	HourlyRate	int	The average daily pay of an employee.	30, 94, 61,...
14	JobInvolvement	category	The level of job involvement of employee. (1: 'Low', 2: 'Medium', 3: 'High', 4: 'Very High')	1, 2, 3, 4
15	JobLevel	int	Job level indicated as integer. The higher the number, the higher the level.	1, 2, 3, 4, 5
16	JobRole	category	The role taken by an employee.	'Sales Executive', 'Research Scientist', 'Laboratory Technician', ...
17	JobSatisfaction	int	Indicator of job satisfaction of an employee. (1: 'Low', 2: 'Medium', 3: 'High', 4: 'Very High')	1, 2, 3, 4
18	MaritalStatus	category	Marital status of an employee.	'Single', 'Married', 'Divorced'
19	MonthlyIncome	int	The monthly fixed salary of an employee.	5993, 5130, 2090,...
20	MonthlyRate	int	The monthly rate of an employee, including variable allowances, bonus, incentive payments,...	19479, 24907, 2396,...
21	NumCompaniesWorked	int	The number of companies an employee worked at.	1, 2, 3,...
22	Over18	category	Whether the employee is above 18. ('Y': above 18, 'N': below 18)	'Y', 'N'
23	OverTime	category	Whether the employee works overtime.	'Yes', 'No'
24	PercentSalaryHike	int	Percentage increase in salary.	11, 23, 15,...
25	PerformanceRating	int	Performance rating (1: 'Low', 2: 'Good', 3: 'Excellent', 4: 'Outstanding')	1, 2, 3, 4
26	RelationshipSatisfaction	int	Level of satisfaction of an employee with relationships in the company. (1: 'Low', 2: 'Medium', 3: 'High', 4: 'Outstanding')	1, 2, 3, 4
28	StockOptionLevel	int	The stock option level obtained by an employee.	0, 1, 2, 3

29	TotalWorkingYears	int	Working experience of employee by the total number of working years.	10, 17, 6,...
30	TrainingTimesLastYear	int	Hours spent training last year.	3, 2, 5,...
31	WorkLifeBalance	int	Level of Work-life balance (1: 'Bad', 2: 'Good', 3: 'Better', 4: 'Best')	1, 2, 3, 4
32	YearsAtCompany	int	The number of years an employee works in IBM.	6, 10, 0,...
33	YearsInCurrentRole	int	The number of years of an employee working in their current role.	4, 7, 0,...
34	YearsSinceLastPromotion	int	The number of years of an employee since their last promotion.	0, 1, 3,...
35	YearsWithCurrManager	int	The number of years of an employee working with their current manager.	0, 5, 7,...

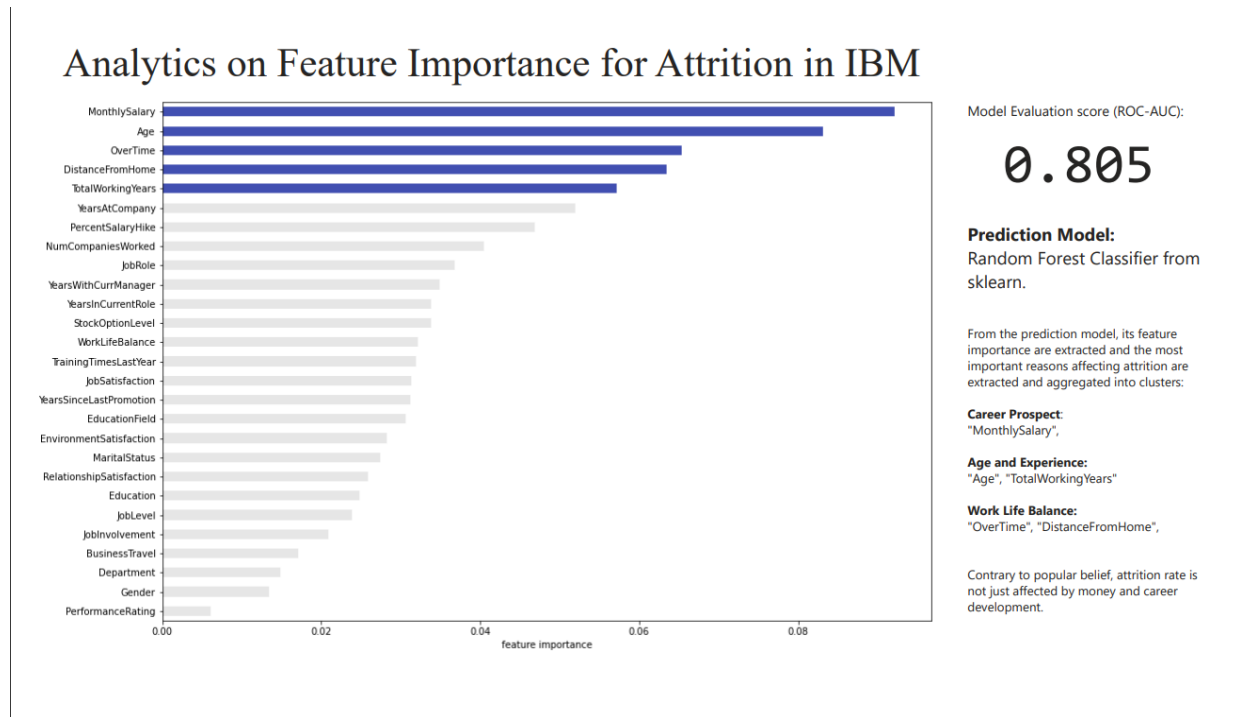
2.2 Data Preprocessing

From the data description above, there are several issues with the imported dataset that can be optimized for ease of further processing.

No.	Problem	Elaboration	Preprocessing Step
1	Existence of non-informative categories.	“EmployeeCount” and “StandardHours” are given constants, which do not provide any information for the attrition status of an employee.	Drop “EmployeeCount” and “StandardHours” columns.
2	Remove redundant columns.	'DailyRate', 'HourlyRate', 'MonthlyRate', 'Over18', 'StandardHours', 'EmployeeCount', 'EmployeeNumber' are the redundant columns that do not provide extra info to the dataset	Remove the stated redundant columns.
3	Inconsistent data format.	Several ordinal features are encoded as integers but few are not.	Tables with maps for ordinal features: “Education”, “EnvironmentSatisfaction”, “JobInvolvement”, “JobSatisfaction”, “PerformanceRating”, “RelationshipSatisfaction”, “WorkLifeBalance” are integrated to the dataset to decode the integers to their represented categories.
4	Suboptimal data type.	<p>True/False features are recorded as “Yes” or “No” text data type, which are less consistent and harder to work with.</p> <p>Ordinal features with integers as data type will be interpreted as numerical values in machine learning models.</p>	<p>Replace the values as True or False and change to a boolean True/False data type.</p> <p>Ordinal features with integers as data type: “JobLevel” and “StockOptionLevel” are changed into text.</p>

3.0 Report of Analysis

3.1 Analytics on Feature Importance for Attrition in IBM



As the dataset is high-dimensional, it is difficult to judge which features are important to include in the analysis of attrition in IBM. This is evaluated by conducting a preliminary predictive analytics and getting its feature importance. By doing so, we are able to extract the important features that might lead to attrition in IBM and guide the following analysis.

In this case, a random forest classifier from scikit-learn library is used to train the predictive model by using Python script in PowerBI. The accuracy of the model is measured by AUC-ROC with K-fold cross validation, of which a decent score of 0.805 is retrieved. It is thus decided that the random forest classifier model can be a preliminary guide for deeper analysis on the attrition in IBM.

The feature importance is visualized on the chart above, and the top 5 features are highlighted in blue, i.e. "MonthlySalary", "Age", "OverTime", "DistanceFromHome" and

“TotalWorkingYears”, which can be aggregated into 3 larger groups of features, that is career prospect, age and experience, work life balance. While money remains to be the most important factor in predicting attrition, other features play an important role to affect the attrition rate in IBM too.

In the upcoming analysis in the following dashboards, these few top features will be the main perspectives for the discussion.

3.2 EDA on Attrition of Employees by Performance Rating

EDA on Attrition of Employees by Performance Rating



Overall, IBM has a decently low attrition rate of near 16%. Looking into the top features extracted from the the model feature importance, we can observe several important patterns:

- **Higher salary leads to lower attrition rate.** Most attrition happens among employees with a monthly salary range in between 0 - 5000, and a relatively smaller amount of them in the salary range in between 5000 - 10000. At an even higher salary range, the attrition rate becomes significantly lower.
- **Increased overtime predicts higher rate of attrition.** When OverTime is recorded True, employee attrition is reported to be nearly 3 times higher (30.53%) than when OverTime is False (10.44%).
- **Increased age reduces the likelihood of attrition,** which reverses when the employee is older than 50 years old when they approach retirement age.

- **The distance from home to workplace has a weak positive correlation with the attrition rate.** Nonetheless, it is important to note that the correlation is not a strong one as employees who live further from the workplace are still frequently reported to have a lower rate of attrition.

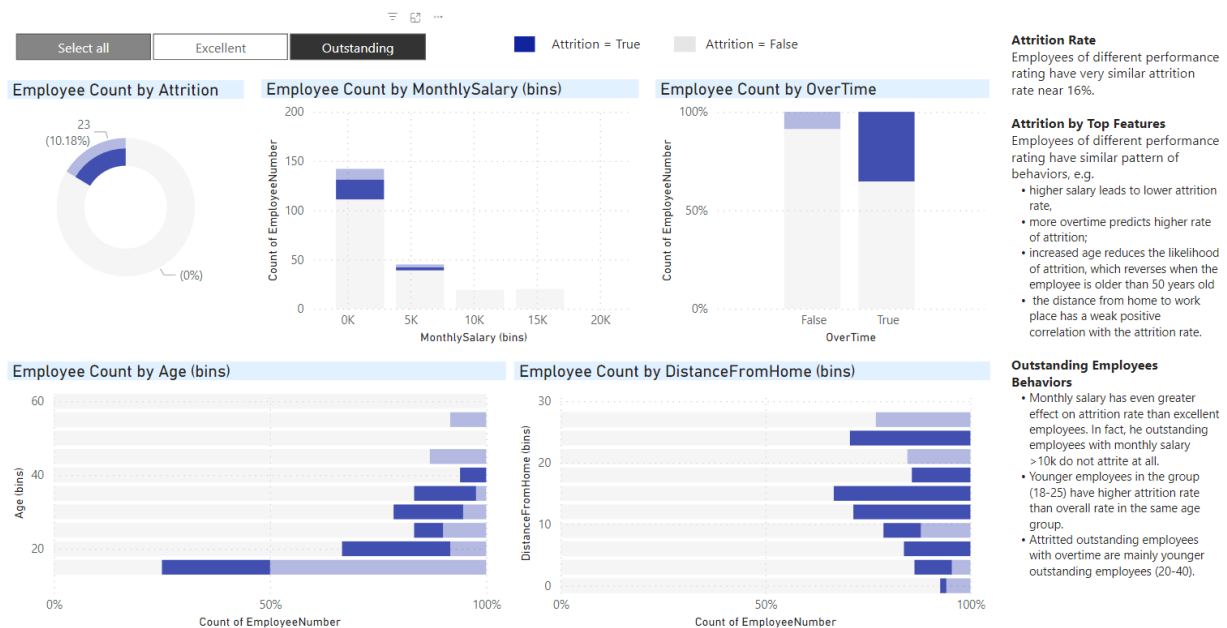
While these observations are consistent with our common understanding about workplace attrition, the charts provide more in depth quantified information on the impact of each feature on attrition rate in IBM. From the observations mentioned above, several suggestions can be extracted to reduce the rate of attrition in IBM:

- Increase the monthly salary for employees that IBM wishes to retain.
- Minimize the frequency of overtime.
- Adapt to the needs of the younger employees (20-35 years old) as they make up most of the attrition rate. This is especially important for the long term planning of the company as the attrition rate of the current older employees will grow as they age.

Nevertheless, these generic suggestions are suboptimal solutions for talent retention that are unable to optimize the company resources as they use up resources indifferently. In the age of knowledge, high performing employees have become one of the most important assets and the anchor for talent retention strategies in tech companies. Therefore, it becomes ever so important to design an environment that can facilitate the retention of high performing employees. In order to do so, we need to delve deep into the features that affect the attrition rate among them.

In the following section, outstanding employees are used to represent talents in IBM following the dataset schema.

EDA on Attrition of Employees by Performance Rating



In this dataset with 1407 samples, there are only 226 high performing employees that are labeled as ‘outstanding’. Out of the 226 outstanding employees, there is an attrition rate of 16.37% similar to the overall attrition rate of the full dataset. While these outstanding employees share similar characteristics with the other employees in terms of the impact of top features on attrition, there are several significant differences:

- **Career prospects.** Monthly salary has an even greater effect on attrition rate than excellent employees. In fact, there is no attrition among the outstanding employees with a monthly salary >10k. This shows that the outstanding employees take greater importance in their career prospects than the others.
- **Work life balance.** The effect of overtime on attrition rate is stronger amongst outstanding employees. When OverTime is recorded True, employee attrition is reported to be more than 4 times higher (35.38%) than when OverTime is False (8.57%). This is coupled with the fact that most attrition amongst outstanding employees with overtime are mainly younger employees (20-40), giving a clear message that the younger outstanding employees do not appreciate overtime. Moreover, the positive correlation between distance from home to workplace and the rate of attrition is more apparent

amongst the outstanding employees too. These observations combined to show that the outstanding employees, especially the younger ones, wish to have more time for themselves outside of working hours.

These two aspects are further elaborated in the following discussions to understand the reasons leading to attrition of outstanding employees.

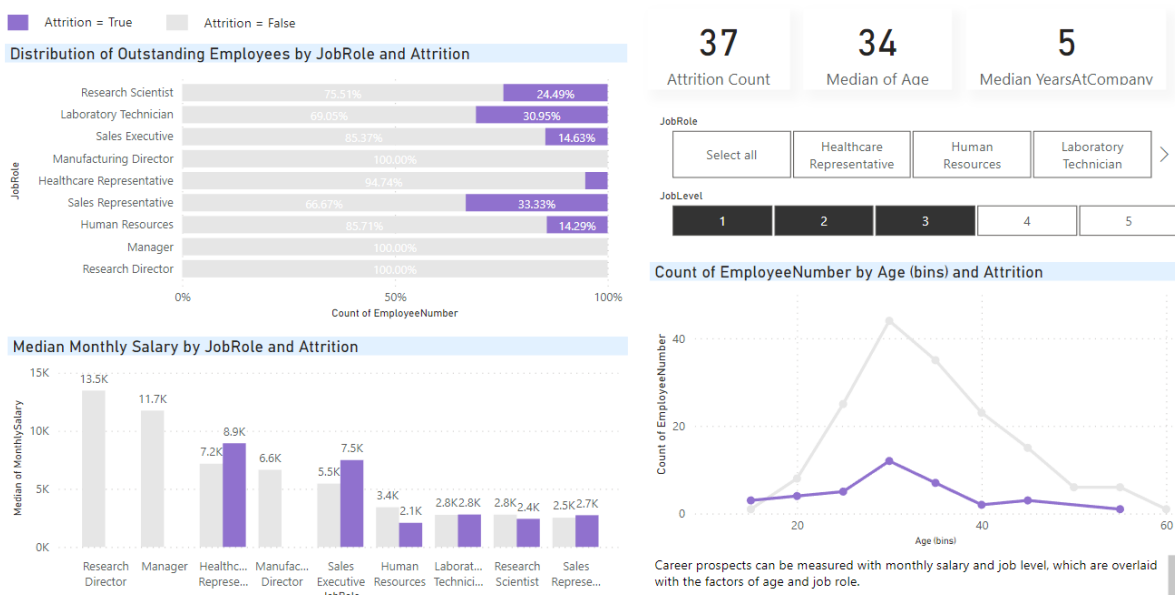
3.3 Career Prospect vs Attrition of Outstanding Employees

From a high level standpoint, career prospects can be measured from two dimensions: payment and job level. These features are then overlaid with two other important features, i.e. age and job roles to provide a comprehensive analysis on the effect of career prospects on attrition of outstanding employees. This dashboard tries to visualize the discussion via several charts:

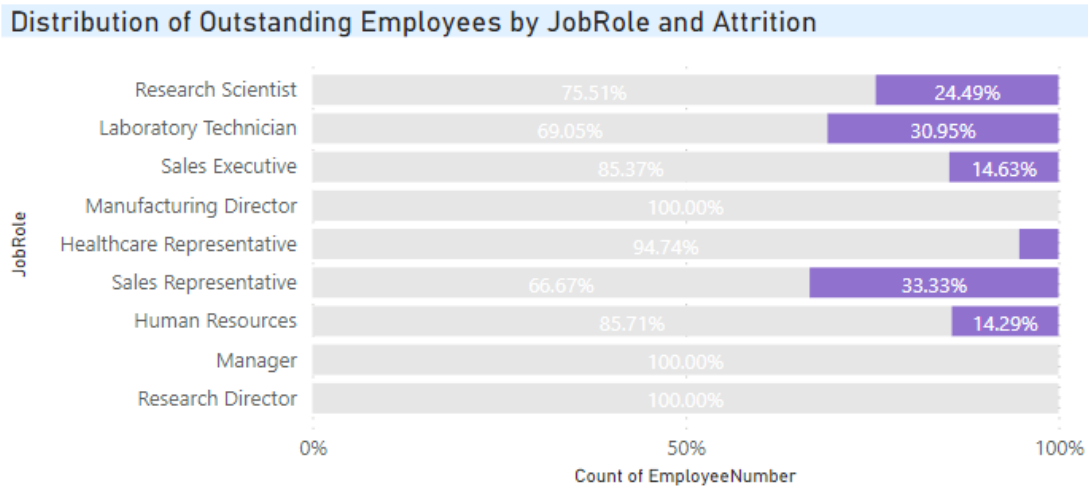
- The distribution of outstanding employees by job role and attrition.
- The median monthly salary by job role and attrition.
- The median monthly salary by age and attrition.

Job levels are used as a slicer to segregate the data into smaller groups and facilitate deeper discussion on a fine scale.

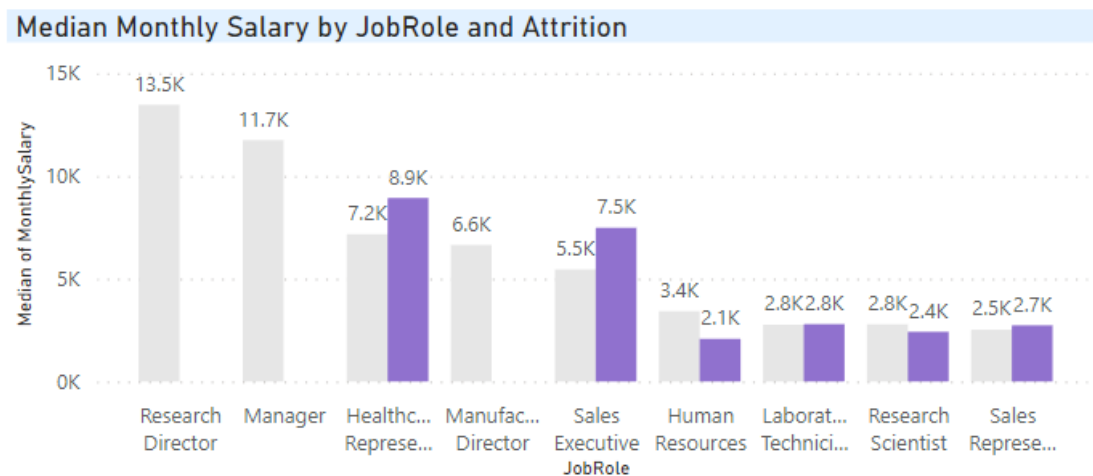
Career Prospect vs Attrition of Outstanding Employees



Out of the 226 outstanding employees in the dataset, there are 37 attritions, 27 of which are in job level 1, 6 from job level 2 and 4 from job level 3. In general, the attrited employees are most likely employees who are in the career development stage of life between 25-40 years old and served several years in the company.



Looking at the distribution of outstanding employees by job role and attrition, we can notice that the job roles that have the highest likelihood of attrition amongst the outstanding employees are sales representatives (33.33%), laboratory Technician (30.05%) and research scientist (24.49%) respectively.



The observation makes perfect sense as sales representatives, research scientists and laboratory technicians have the lowest median monthly salary compared with other job roles. The vast gap in monthly salary between different job roles is apparent, making it extremely hard to retain employees in the low paying job roles with higher monthly salary. In fact, the observations made from the two figures above shows that the higher the median monthly salary of a job role, the lower the likelihood of attrition.

Suggestion:

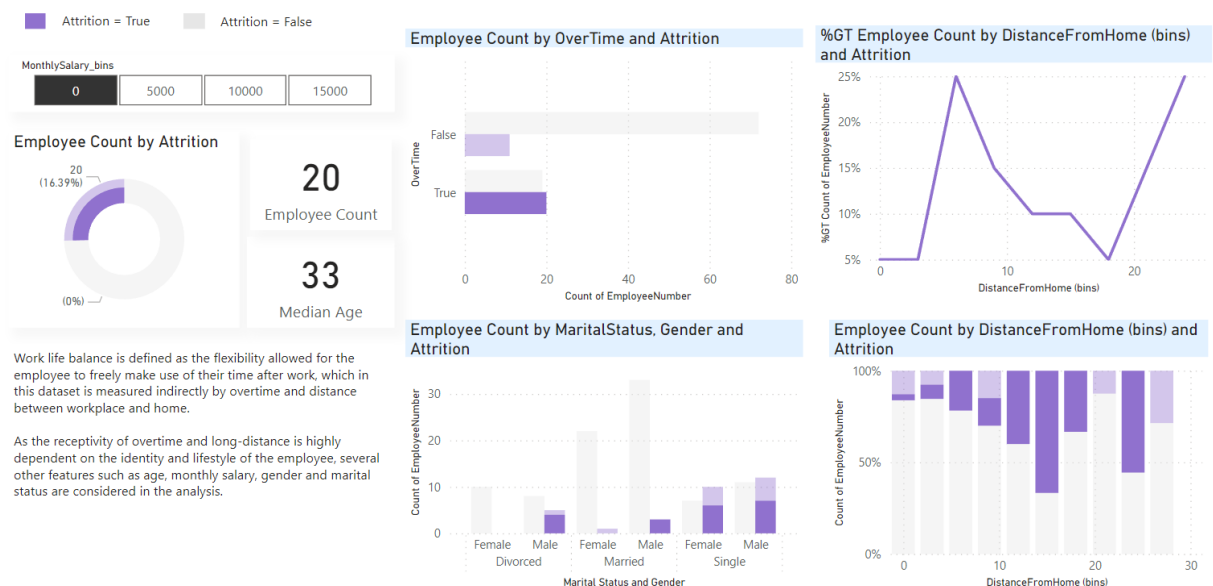
Thus, it is not just about increasing the monthly salary to retain outstanding employees. This suggests that IBM should conduct systemic change to increase its capacity for flexibility and upward mobility for the employees with lower paying jobs to reduce the loss of high performing employees. There are multiple ways that can be done to achieve this, e.g.

- Design pathways that allow low-level employees to move up in their careers.
- Provide more in-job training to prepare low level employees for transition upwards into higher paying jobs in the company.

3.4 Work Life Balance vs Attrition of Outstanding Employees

Other than career prospects, work life balance is another aspect that can be another huge part of the equation to retain outstanding employees. In the context of this discussion, work life balance is defined as the flexibility allowed for the employee to freely make use of their time after work, which in this dataset is measured indirectly by overtime and distance between workplace and home as they can use up a lot of time outside of standard working hours considerably. As the receptivity of overtime and long-distance is highly dependent on the identity and lifestyle of the employee, several other features such as age, gender and marital status are considered in the analysis. As we have known the monthly salary being the most important predictor of attrition, the dataset is sliced based on monthly salary for further analysis.

Work Life Balance vs Attrition of Outstanding Employees

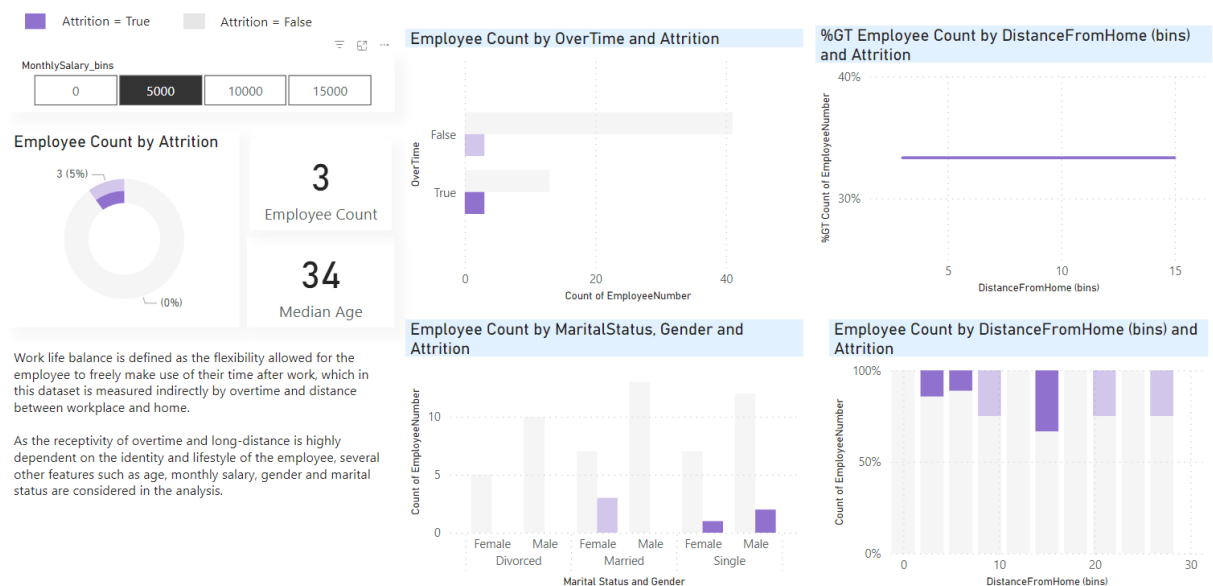


At monthly salary in the range of 0-5000, there are 31 attrited employees, making up 25.41% of the population. For the employees in this group who attrited, 20 out of 31 of them (64.52%) reported OverTime as True, showing a strong correlation between overtime and attrition. The groups of attrited employees with the highest proportion of overtime are married and divorced

male, followed by single male and female. As the distance between home and workplace increases, the likelihood of attrition increases when OverTime is true.

The observations above imply that most employees in this group who are attrited take importance in work life balance as both overtime and increased distance from home increase their rate of attrition. To these employees, it is important to respect their personal time and minimize overtime as much as possible to reduce their rate of attrition.

Work Life Balance vs Attrition of Outstanding Employees



At a monthly salary in the range of 5000-10000, there are only 6 attrited employees (25.41%) out of the population of 60. Out of the 6 of them, only 3 of them (50%) reported OverTime as True. At the same time, the distance between their homes and workplace have very little or no impact on the rate of attrition.

The observations show that work life balance is less important for this group of employees as both overtime and increased distance from home do not correlate with their rate of attrition. It can be explained by several plausible reasons:

- The outstanding employees at higher salary ranges are more career oriented.
- Higher salary can negate the effect of the lack of work life balance on attrition rate.

There is no attrition among outstanding employees with a salary above 10000.

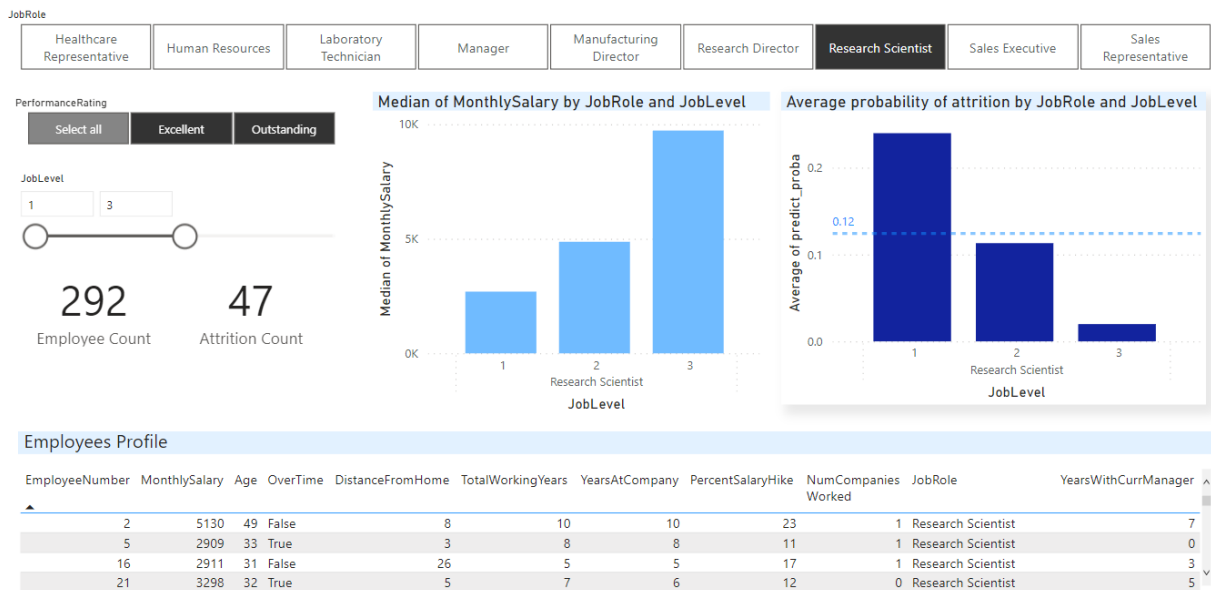
Suggestion:

As the company does not have unlimited monetary resources to spend, it is important that it makes use of strategies other than increasing salary to retain outstanding employees. To do so, it is important to customize retention strategies based on different target groups.

- For employees with a salary range between 0-5000, work life balance is a strong contributing factor to the prediction of attrition. The company can accommodate the needs of these employees by minimizing overtime and offering partial allowances to stay nearer to the workplace.
- For employees with a salary range between 5000-10000, features affecting work life balance do not predict attrition. In this group of employees, the company should focus on improving their career prospects.

3.5 Monitor Dashboard - Predicting Probability of Attrition for Each Job Role

Predicting Probability of Attrition for Each Jobrole



For human resource (HR) management in office, the HR department needs to know the attrition rate of each job. However, in practice, there is no way to calculate the rate of attrition for existing employees as every single existing employee will have attrition recorded as False.

Therefore, it is important that the human resource department has a means to predict the probability of attrition in the office. In this dashboard, a set of test data is used to represent the existing employees and the trained random forest classifier model is used to predict the attrition status and probability of each employee.

Suggestion:

By using various slicers in the dashboard: JobRole, PerformanceRating and JobLevel, the human resource managers are able to quickly estimate the probability of employee attrition for a concerned position, allowing for a significant increase in managerial capability.

4.0 Conclusion

In conclusion, the analysis with the IBM HE Analytics dataset has provided multiple ways that the HR managers to consider in their long term planning for talent retention in IBM.

The analytics on feature importance provide guidance for the directions that HR can use to analyze the attrition of outstanding employees in IBM. Upon exploratory data analysis, one can see that career prospects and work life balance are the two main reasons leading to attrition of outstanding employees. Further analysis on these two groups of features suggested that IBM should work on two main strategies:

1. increase its capacity for flexibility and upward mobility for the employees with lower paying jobs
2. accommodate the need of work life balance of the employees (especially those with lower monthly salary), e.g. by minimizing overtime and offering partial allowances to stay nearer to the workplace.

While the two suggested strategies cannot be done with only the HR department, the HR managers should work with other relevant stakeholders and integrate the two strategies mentioned above in the long-term talent retention planning in IBM.

Last but not least, a monitoring dashboard is built to allow human resource managers to quickly estimate the predicted probability of employee attrition, thereby increasing their managerial capability.

5.0 Attachments

1. Data:
 - a. CategoricalKeyMapping.xlsx
 - b. HR-Employee-Attrition.csv
 - c. Test_data.csv
2. AlternativeAssessment_ChongXianJun.pbix
3. pred_model.pkf
4. BIA-alternative_assessment.pdf