



Exploring The Factors That Impact Employee Turnover

Name: M.M. Mohamed Sabath

Student Number: 017

Module: Final project of the advanced diploma

Course: Data Science (Full time)

Institute: National Institute of Business Management

Date of Submission: 30.04.2023

“A Project Report submitted for the partial fulfillment of the requirements of the Advanced Diploma in Supplies and Materials Management (Full Time) Programme”

Declaration

I hereby declare that the work presented in this project report was carried out independently.
by myself and have cited the work of others and given due reference diligently.

Name of the student – M.M. Mohamed Sabath

Signature -

Date – 2023.04.29

Supervisor's Certification

I certify that the above student carried out his/her project under my supervision and
guidance.

Name of the Supervisor – Mr. Thurairasa Balakumar

Signature –

Date –

Dedication

This research project is dedicated to my dear parents, who have been wonderful supporters of mine until the project was completed, as well as to my good friends, who have continuously given me their undivided attention and support while I worked to complete this study with genuine self-assurance.

Acknowledgement

Without the assistance and direction of numerous people, this project would not have been able to be carried out. I want to thank each and every one of them from the bottom of my heart.

I want to sincerely thank my supervisor, Mr. Thurairasa Balakumar, for providing me with this excellent chance and for overseeing and directing this project. I also want to thank you for giving us the information and tools we needed to complete this endeavor. We were able to finish this project thanks in large part to the wonderful cooperation and encouragement of my friends and family, for whom I would like to express my gratitude.

Abstract

The purpose of this study is to look into the connection between employee turnover in an organization and aspects connected to their jobs, personal fulfillment, and job satisfaction.

Findings from the study will help to

impact of specific job-related factors on employee attrition, the typical education level attained by staff members who leave the company, and the connection between an employee's opinion of work-life balance and attrition. This study can assist businesses in identifying the elements causing employee turnover and creating retention strategies by examining data on elements including employee compensation, job satisfaction, education, and work-related factors.

Table of Contents

Declaration	i
Dedication	ii
Acknowledgement	iii
Abstract	iv
Table of Contents	v
List of tables.....	vii
List of figures.....	vii
List of Abbreviations	viii
Chapter 1: Introduction	1
1.1 Background	1
1.2 Research problem.....	2
1.3 Research Questions	2
1.4 Objectives of the project	3
1.5 Scope of the research.....	3
1.6 Justification of the research.....	4
1.7 limitations.....	5
Chapter 2 – Literature Review	6
2.1 Introduction to the research theme	6
2.2 Theoretical explanation about the Key Words in the Topic.....	6

2.3 Findings by other researchers.....	7
2.4 The Research gap	9
2.5 Table for Variables, their definitions and sources.....	10
2.6 Chapter conclusion.....	13
Chapter 3: Methodology	14
3.1 Introduction	14
3.2 Population, sample and Sampling technique.....	14
3.3 Type of Data to be collected and data sources	14
3.4 Data collection tools and plan	15
3.5 Conceptual framework	15
3.6 Hypothesis	16
3.7 Operationalization Table	16
3.8 Methods of Data analysis	18
Chapter 4: Data Analysis	19
4.1 Data Analysis	19
4.2 Findings and Interpretation	32
Chapter 5: Discussion and Recommendations.....	41
5.1 Discussion	41
5.1 Recommendations.	43
5.3 Conclusions	44

Appendices.....	45
-----------------	----

List of tables

Table 2.5- Variables, their definitions and sources.....	11
Table 3.7- Operationalization table.....	19

List of figures

Figure 4.1.1: Checking the null values	22
Figure 4.1.2: Boxplot for checking the outliers	19
Figure 4.1.3 frequency table of employee attrition variable.....	20
Figure 4.1.4: Frequency table for education level of employees who left the organization	21
Figure 4.1.5: Pie chart for the education level of the employees who left the organization.	21
Figure 4.1.8: Bar chart of job satisfaction variable.....	23
Figure 4.1.9: Bar chart for employees environment satisfaction level	24
Figure 4.1.10: Bar chart for employees level of work-life balance	25
Figure 4.1.11: Coding and the output for performing first sub objective	26
Figure 4.1.12: Coding and the output for performing 2nd sub objective	27
Figure 4.1.13: Coding and the output for performing 3rd sub objective	28
Figure 4.1.14: Coding and the output for performing 3rd sub objective	29

Figure 4.1.15: Coding and the output for performing 2nd main objective	30
Figure 4.1.16: Coding and the output for performing 3rd main objective.....	31
Figure 4.2.1: Heatmap for the 1st sub objective output	33
Figure 4.2.2: Heatmap for the 2nd sub objective output.....	34
Figure 4.2.3: Heatmap for the 3rd sub objective output	36
Figure 4.2.5: Bar plot of the employee's education level	38
Figure 4.2.6: Heatmap for the 3rd main objective output.....	40

List of Abbreviations

ANOVA – Analysis of variance

Chapter 1: Introduction

Employer migration is a major worry for businesses in all sectors. Increased costs, decreased productivity, and a detrimental effect on an organization's overall performance can result from high personnel turnover rates. Therefore, each organization's success depends on its ability to identify the elements that influence employee turnover. This study intends to investigate how personal satisfaction and work-related characteristics affect employee turnover in a firm. Data on a variety of characteristics, including employee demographics, compensation, job satisfaction, education, and work-related aspects, will be used in the study. The study's goals include examining whether the chosen job-related factors have an impact on employee attrition, Is there a list of the highest educational level that former employees have attained?

1.1 Background

Employee churn is a major issue that many firms now confront. Among other detrimental effects, high employee turnover can increase costs, lower productivity, and lower morale. To build effective retention tactics and foster a more favorable work environment, employers must have a thorough understanding of the factors that influence employee turnover. The "HR_Analytics.csv" file offers insightful information about the characteristics that affect employee turnover, such as job-related variables, educational levels, work-life balance, and other significant variables.

This project uses the "HR_Analytics.csv" dataset to investigate the elements that influence employee turnover. The goals of the study include figuring out whether factors related to the job have an effect on employee attrition, identifying the average educational level attained by departing employees, and examining the connection between an employee's perception of their

work-life balance and attrition. The outcomes of this investigation may offer organizations helpful information about how to increase employee retention and foster a happier workplace.

1.2 Research problem

Finding the causes of employee turnover, a problem that many firms have, is the subject of the research problem for this project report. The productivity, morale, and costs of a business can all be significantly impacted by high staff turnover. Determining the root causes of employee attrition and creating methods to lower turnover rates are therefore crucial. Using the "HR_Analytics.csv" dataset, this research seeks to investigate the elements that influence employee turnover. The research problem entails determining whether job-related factors influence employee attrition, examining the most common educational level attained by employees who have left the company, and investigating the connection between an employee's perception of their work-life balance and attrition. This initiative can contribute significantly by solving these research issues.

1.3 Research Questions

1. Is there a significant impact of job satisfaction on employee attrition?
2. Does working overtime have a significant impact on employee attrition?
3. Is there a significant relationship between environmental satisfaction and employee attrition?
4. Does hourly rate have a significant impact on employee attrition?
5. What is the most common level of education achieved by employees who have left the organization?
6. Is there a significant relationship between an employee's perception of their work-life balance and their attrition rate?

1.4 Objectives of the project

- ❖ Identifying is there an impact on employee attrition from the selected job-related factors.
 - Sub objectives: Identifying is there from.
 1. Job satisfaction on attrition
 2. Working overtime with attrition
 3. Environment satisfaction on attrition
 4. Hourly rate on attrition
- ❖ Identify the most common level of education achieved by employees who have left the organization.
- ❖ Evaluate is there a relationship between an employee's perception of their work-life balance and their attrition.

1.5 Scope of the research

The goal of this study is to investigate the elements that affect employee turnover within a business. The study will cover a range of employee work-related factors, and personal satisfaction variables to understand their relationship with employee turnover. The study will be conducted in a specific organization according to the dataset source. The research will use a quantitative approach, analyzing the data collected from the organization's HR department. The amount of data that is available and the level of statistical power needed to answer the study questions will be used to establish the sample size. The project proposal's listed research questions are the only ones whose scope the study can cover. It's possible that other relevant subjects like performance, job satisfaction, and employee involvement weren't thoroughly covered. The overall goal of the study is to give insights into the elements that affect employee turnover in a particular firm and assist that organization in creating successful retention strategies in view of the results.

1.6 Justification of the research

The fact that staff turnover is a major problem for businesses in a variety of industries serves as the rationale for this study. The expenses associated with replacing staff members can be high and include both the direct costs associated with finding, hiring, and training new hires as well as the indirect costs associated with decreased output, low morale, and higher burden for the remaining staff. Employee turnover can also affect how successful a company is since it can diminish productivity, the caliber of work, and customer satisfaction.

Organizations can take action to address the root causes of turnover and put plans in place to lower employee turnover rates by understanding how employee work-related factors and personal happiness affect employee turnover. This study can offer helpful insights into the elements that affect employee attrition, and it can be helpful to form innovative solutions and strategies.

Additionally, firms can make improvements that can improve employee's productivity and their working place more comfortable for them and customer satisfaction by understanding the elements that affect employee satisfaction. Businesses can utilize the findings of this study to determine what needs to be altered to provide employees with a better work environment, which can increase employee retention rates and increase productivity.

1.7 limitations.

The variables included in the analysis could have some limitations. Even though the project's variables are all-inclusive, there might be additional significant elements that affect employee turnover that are not included in the statistics. Additionally, the analysis only uses quantitative data and ignores qualitative aspects like workplace culture, leadership style, and employee relationships that may affect employee turnover.

Last but not least, this project ignores outside variables like industry or economic trends that could affect employee turnover. my research effort may not be able to fully account for all of the elements that affect employee turnover because it is a complex problem.

Chapter 2 – Literature Review

2.1 Introduction to the research theme

Employee turnover in organizations and the factors that affect it are the focus of this project's research. Employee churn has been a recurring issue for many businesses throughout the world. High employee turnover can be costly and have a negative impact on an organization's productivity and effectiveness. Therefore, it is critical for businesses to understand the factors that contribute to employee turnover and to find solutions.

The existing research on employee turnover, including its causes, effects, and contributing variables, will be the primary emphasis of the literature evaluation for this project. The assessment will also look at the many approaches that businesses might take to cut down on staff churn and boost retention.

2.2 Theoretical explanation about the Key Words in the Topic

Employee Turnover: This key word states about employee leaving a company and being replaced by a new hiring is a means to employee turnover. An organization may incur high costs due to lost production and recruitment and training expenses if there is a high employee turnover rate.

Factors: Factors are the underlying causes or influences that have an impact on a specific result. Factors are the causes of employee departures from their occupations as well as those that may affect their choice to remain or go.

Impact: Impact describes the impact or influence that one item or person has on another. In the context of my study topic, I'm interested in examining the effects of numerous variables on

employee churn, including both organizational and individual variables (such job satisfaction and work-life balance).

The overall goal of my research is to comprehend the elements that cause employee turnover and how they affect organizational effectiveness. We can assist firms to better understand the causes of employee turnover and create strategies to lower it and increase retention by looking into these issues.

2.3 Findings by other researchers

- In the study by Boondarig Ronra and Prof. Manat Chaisawat (2020), they say there is a significant positive relationship between job satisfaction and organizational commitment among employees working in the Thai tourism industry. Salary, working environment, chances for professional progression, and employee perks are some of the variables that affect job satisfaction.
- In the study by Ahituv, A. and Lerman, R.I. (2010). They found that there is a relationship between employee attrition and the job satisfaction variable.
- In the study by Tulangow, M.J., Saerang, D.P.E. and Rumokoy, F.S. (2018). They say that work motivation and job satisfaction have a negative effect on turnover intention, while the work environment has a positive effect on turnover intention.

- In the study by Alamdar Hussain Khan and Muhammad Aleem Found that organizational culture has a significant positive impact on job satisfaction, and job satisfaction is negatively related to employee turnover intention.

- A study by Sousa-Poza, A. and Sousa-Poza, A.A. (2007) found that both job satisfaction and organizational commitment have a significant negative impact on turnover intention. The authors suggest that enhancing job satisfaction and organizational commitment can help reduce employee turnover intention in the restaurant industry.

- A study by Zimmerman, R.D. and Darnold, T.C. (2009) found that work-life balance is positively related to job satisfaction and that job satisfaction is negatively related to turnover intentions. The authors suggest that improving work-life balance can lead to higher job satisfaction, which in turn can help reduce employee turnover intentions.

- A study by Muhamad Ekhsan he says that work discipline and work motivation have a significant negative impact on employee turnover intention, while job satisfaction has a significant positive impact on work discipline and work motivation. The authors suggest that enhancing work discipline, work motivation, and job satisfaction can help reduce employee turnover intention.

- A study by Kaushalya, R.K.N. and Perera, G.D.N. (2018) found that job satisfaction and work-life balance are negatively related to turnover intention. The authors suggest that

enhancing job satisfaction and promoting work-life balance can help reduce employee turnover intention among academics in Sri Lankan universities.

- A study by Steel and Warner (2019) found that job satisfaction and organizational commitment were negatively related to turnover intention, while job stress was positively related to turnover intention. The study also found that the effects of job stress on turnover intention were stronger for employees with low job satisfaction and organizational commitment.
- According to a study by Griffeth et al. (2019), employee turnover is influenced by a combination of individual, organizational, and environmental factors. The study found that job satisfaction, organizational support, and job alternatives were the strongest predictors of turnover.

2.4 The Research gap

According to the literature study, there are still some knowledge gaps in the elements that influence employee turnover. Although several studies have identified several job-related and personal factors that affect turnover, further research is still required to fully understand the intricate interactions between these elements. More research is needed on the interactions between these elements' effects on turnover because earlier studies have mostly concentrated on the influence of individual factors. In addition, despite considerable study on the connections between work-life balance and employee turnover, there is still a lack of information regarding how this relationship

differs depending on what kind of business and job function. Overall, additional study is required to examine the intricate and ever-changing nature of the variables affecting employee turnover.

2.5 Table for Variables, their definitions, and sources

Table 2.5- Variables, their definitions, and sources

Variable	Definition	Source
Employee turnover	The rate at which employees leave the organization, either voluntarily or involuntarily.	Kaggle Employee Attrition and Factors Dataset
Employee demographics	The characteristics of the employees, such as age, gender, education, and job level.	Kaggle Employee Attrition and Factors Dataset
Work-related factors	Factors related to an employee's job, such as workload, job satisfaction, overtime, and work-life balance.	Kaggle Employee Attrition and Factors Dataset
Personal satisfaction	satisfaction An employee's overall satisfaction with their life outside of work, including family, health, and other personal factors.	Kaggle Employee Attrition and Factors Dataset
Age	The age of the employee	Kaggle Employee Attrition and Factors Dataset
Attrition	Whether the employee has left the company (Yes/No)	Kaggle Employee Attrition and Factors Dataset
BusinessTravel	The frequency of travel for the employee (Non-Travel, Travel_Rarely, Travel_Frequently)	Kaggle Employee Attrition and Factors Dataset

DailyRate	The daily rate of pay for the employee	Kaggle Employee Attrition and Factors Dataset
Department	The department the employee works in (Sales, Research & Development, Human Resources)	Kaggle Employee Attrition and Factors Dataset
DistanceFromHome	The distance in miles between the employee's home and workplace	Kaggle Employee Attrition and Factors Dataset
Education	The level of education of the employee that he or she has achieved.	Kaggle Employee Attrition and Factors Dataset
EducationFeild	The field of education studied by the employee (Life Sciences, Medical, Marketing, Technical Degree, Human Resources, Other)	Kaggle Employee Attrition and Factors Dataset
EnvironmentSatisfaction	The level of satisfaction the employee has with their work environment (1 = Low, 2 = Medium, 3 = High, 4 = Very High)	Kaggle Employee Attrition and Factors Dataset
Gender	The gender of the employee (Male, Female)	Kaggle Employee Attrition and Factors Dataset
HourlyRate	The hourly rate of pay for the employee	Kaggle Employee Attrition and Factors Dataset
JobInvolvement	The level of involvement the employee has with their job (1 = Low, 2 = Medium, 3 = High, 4 = Very High)	Kaggle Employee Attrition and Factors Dataset
JobLevel	The level of the employee's job within the company (1 = Entry	Kaggle Employee Attrition and Factors Dataset

	level, 2 = Manager, 3 = Senior Manager)	
JobRole	The role of the employee within the company (Sales Executive, Research Scientist, Laboratory Technician, Manufacturing Director, Healthcare Representative, Manager, Sales Representative, Research Director, Human Resources)	Kaggle Employee Attrition and Factors Dataset
JobSatisfaction	The level of satisfaction the employee has with their job (1 = Low, 2 = Medium, 3 = High, 4 = Very High)	Kaggle Employee Attrition and Factors Dataset
MaritalStatus	The marital status of the employee (Single, Married, Divorced)	Kaggle Employee Attrition and Factors Dataset
MonthlyIncome	The monthly income of the employee	Kaggle Employee Attrition and Factors Dataset
MonthlyRate	The monthly rate of pay for the employee	Kaggle Employee Attrition and Factors Dataset
NumCompaniesWorked	The number of companies the employee has worked for	Kaggle Employee Attrition and Factors Dataset
Over18	Whether the employee is over 18 years of age (Yes/No)	Kaggle Employee Attrition and Factors Dataset
OverTime	Whether the employee works overtime (Yes/No)	Kaggle Employee Attrition and Factors Dataset
PercentSalaryHike	The percentage increase in salary for the employee compared to their previous salary	Kaggle Employee Attrition and Factors Dataset
PerformanceRating	The employee's performance rating (1 = Low, 2 = Good, 3 = Excellent, 4 = Outstanding)	Kaggle Employee Attrition and Factors Dataset

RelationshipSatisfaction	The level of satisfaction the employee has with their relationships at work (1 = Low, 2 = Medium, 3 = High, 4 = Very High)	Kaggle Employee Attrition and Factors Dataset
StandardHours	The standard number of working hours per day for the employee	Kaggle Employee Attrition and Factors Dataset
StackOptionLevel	The level of stock options awarded to the employee (0 = None, 1 = Low, 2 = Medium, 3 = High)	Kaggle Employee Attrition and Factors Dataset
TotalWorkingYears	The total number of years the employee has worked	Kaggle Employee Attrition and Factors Dataset
TrainingTimesLastYear	The number of times the employee was trained last year	Kaggle Employee Attrition and Factors Dataset

2.6 Chapter conclusion

According to the assessment of the literature, when we see the employee turnover, it is a complicated problem that is impacted by a variety of societal, organizational, and environmental factors. Some of the important elements that have been discovered as attrition predictors include job satisfaction, organizational support, options for careers, work-life balance, job expectations, and job resources. By investigating the effects of work-life balance on employee attrition, as well as the effects of other job-related characteristics and the most common educational level attained by departing employees, this research seeks to close this gap in the literature.

Chapter 3: Methodology

3.1 Introduction

Any research project must have a methodology chapter that details the overall structure and technique of the study. The approaches and methods that will be utilized to gather and evaluate data are thoroughly explained in this chapter. The data for this project will come from Kaggle, and the methodology's main goal will be to determine how employee work-related aspects and personal issues affect employee turnover in the workplace.

This study's methodology will make use of a quantitative research approach that entails the gathering and examination of numerical data. the major method used by the survey to obtain data. The survey is intended to gather information on the relevant elements, such as employee demographics, workplace-related issues, individual satisfaction, and turnover intentions. Looking at the Dataset enables us to observe that. To ascertain the link between these factors, the data will be evaluated using statistical techniques including regression analysis and the Chi-Squared test.

3.2 Population, sample, and Sampling technique

All the organization's employees would make up the study's population. This population's subset, the sample, would be examined. This dataset's originator is not known. The sample technique employed, however, is not mentioned in the dataset documentation, and the data is instead based on an employee satisfaction survey carried out in a sizable firm.

3.3 Type of Data to be collected and data sources.

The types of data that has been collected for this project are mainly quantitative data, which is numerical data that can be measured and analyzed statistically. The primary data source for this project is the Employee Attrition and Factors dataset that I got from Kaggle.

3.4 Data collection tools and plan

In here I have taken data which is freely available on Kaggle. Normally for the data collection, we can use tools such as Survey Questionnaires, Company Records, Literature Review. From this, for this project for the data collection plan, I may use Literature Review to collect data from other studies and reports on similar topics if I need more data.

3.5 Conceptual framework

- ✓ Employee turnover: Employee turnover, which is defined as the voluntary or involuntary separation of employees from the firm, is the study's primary outcome variable.
- ✓ Work-Related Factors: The study will also investigate the connections between job satisfaction, workload, overtime, work-life balance, and performance and employee turnover.
- ✓ Personal Satisfaction: The study will examine the impact of personal satisfaction on employee turnover, including factors such as job security, compensation, and benefits.
- ✓ Data Collection: The study will collect data from the Employee Attrition and Factors dataset, as well as from other relevant sources.
- ✓ Data analysis: It will be done using descriptive statistics, regression analysis, and the chi-squared test.
- ✓ Results: The analysis's conclusions, including any noteworthy correlations and the strength of the link between the variables and employee turnover, will be presented in the research.
- ✓ Conclusion: Based on the findings of the analysis, the study will come to conclusions and offer suggestions that may be used to increase staff retention and decrease turnover within the company.

3.6 Hypothesis

- ✓ “There is a significant relationship between job satisfaction and employee turnover.”
- ✓ “There is a significant relationship between working overtime and employee turnover.”
- ✓ “There is a significant relationship between environment satisfaction and employee turnover.”
- ✓ “There is a significant relationship between hourly rate and employee turnover.”
- ✓ “There is a significant relationship between employees’ work-life balance and employee turnover.”

3.7 Operationalization Table

Table 3.7 -Operationalization table

Variable	Definition	Indicators	Measures
Attrition	Whether or not the employee has left the company	"Attrition"	Binary variable: 1 = Yes, the employee has left; 0 = No, the employee is still with the company
Daily Rate	The daily rate of pay for the employee	"DailyRate"	Numeric
Education	The level of education of the employee	"Education"	Categorical variable: 1 = Below College, 2 = College, 3 = Bachelor, 4 = Master, 5 = Doctor

Variable	Definition	Indicators	Measures
Education Field	The field of education of the employee	"EducationField"	Categorical variable: 1 = Life Sciences, 2 = Medical, 3 = Marketing, 4 = Technical Degree, 5 = Human Resources, 6 = Other
Environment Satisfaction	The employee's satisfaction with their work environment	"EnvironmentSatisfaction"	Likert scale from 1 (low) to 4 (high)
Gender	The gender of the employee	"Gender"	Binary variable: 1 = Female, 2 = Male
Hourly Rate	The hourly rate of pay for the employee	"HourlyRate"	Numeric
Job Involvement	The level of involvement the employee has in their job	"JobInvolvement"	Likert scale from 1 (low) to 4 (high)
JobSatisfaction	The degree to which an employee is satisfied with their job	"JobSatisfaction"	Likert scale from 1 (low) to 4 (high)
MonthlyIncome	The monthly income of the employee	"MonthlyIncome"	

3.8 Methods of Data analysis

- **Descriptive Statistics:** The main variables in the dataset can be summed up and described using descriptive statistics. It can reveal information on the typical age of departing employees, the typical educational level attained by departing employees, and the typical hourly wage of departing employees.
- **Inferential Statistics:** This is to test the hypothesis and determine the link between variables using inferential statistics. The effects of work-related variables, such as job satisfaction, working overtime, workplace satisfaction, and hourly rate, on employee attrition can be determined using regression analysis. The most frequent educational level attained by departing employees can be determined using the chi-square test. The perception of a work-life balance by an employee and their attrition can be compared using correlation analysis.
- **Data visualization:** It is a technique that may be used to show data visually to spot patterns and trends. It can give information about the distribution of departing employees across various age brackets, educational levels, and hourly wages. Finding the connections between employee attrition, hourly pay, working overtime, and job satisfaction can also be helpful.

Chapter 4: Data Analysis

4.1 Data Analysis

In the analysis part, before presenting the data that has to be analyzed, it has to be checked whether it has any null values or not also for the Outliers.

```
In [6]: #Checking is there any null values
print(df1.isnull().sum())

Attrition          0
JobSatisfaction    0
OverTime           0
EnvironmentsSatisfaction 0
HourlyRate         0
Education          0
WorkLifeBalance    0
dtype: int64
```

Figure 4.1.1 Checking the null values.

According to that, In the above image it shows that the data and the variables that must be analyzed have no null values. So furthermore, when we check the outliers, I have only one numerical variable for that.

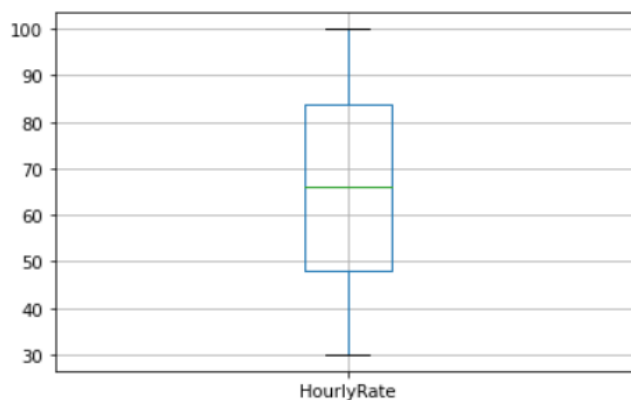


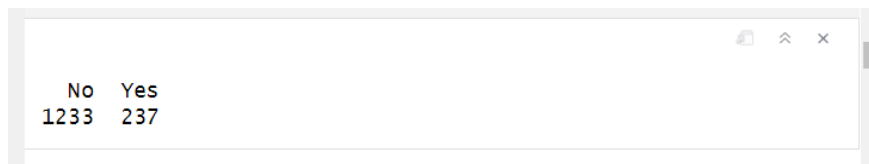
Figure 4.1.2 Boxplot for checking the outliers.

In here, the above image figure 2 shows that the variable doesn't have any outliers. Now can move to the presentation part.

The variables I have chosen for the analysis are.

- Employees attrition whether they left the job or not - Attrition.
- The employees' level of job satisfaction - JobSatisfaction
- Whether the employee work overtime or not - OverTime
- Employees level of satisfaction on the working environment - EnvironmentSatisfaction
- Hourly paying rate of employees - HourlyRate
- Education level of the employees - Education
- Level of employees' work-life balance - Work-LifeBalance

Employee Attrition



No	Yes
1233	237

Figure 4.1.3: frequency table of employee attrition variable

The above table clearly shows that there are 1233 'No' and 237 'Yes' which means 237 employees have left the organization. So, we can say that most of the employees are still working in the organization.

Education level of the employees who left the organization.

1	2	3	4	5
31	44	99	58	5

Figure 4.1.4: Frequency table for education level of employees who left the organization

The above figure 4.1.4 shows the count number of employees from each educational level separately from those who left the organization.

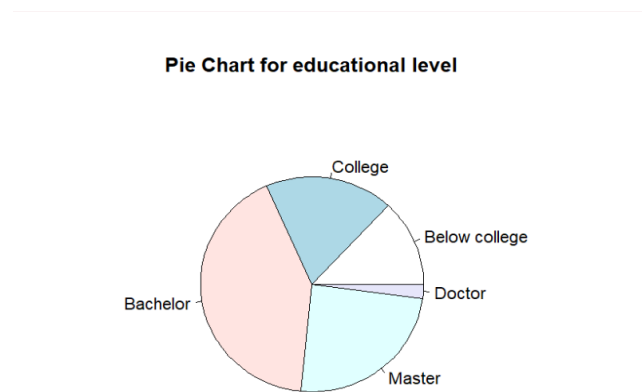
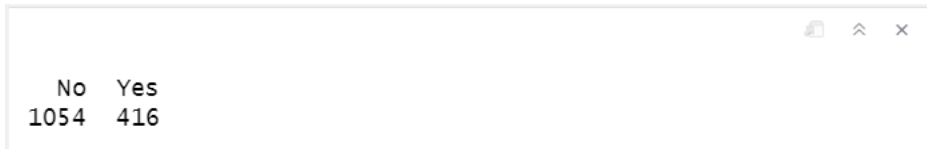


Figure 4.1.5: Pie chart for the education level of the employees who left the organization.

From the above figure 4.1.5, this pie chart clearly shows that most of the employees are having bachelor's degree at the same time the smaller number of people are having doctors.

Working overtime



No	Yes
1054	416

Figure 4.1.6: Frequency table for working overtime variable.

Figure 4.1.6 shows the status of the working overtime of the employees. There are only 416 employees who work overtime. From the above table we can clearly say that a greater number of employees do not work overtime, in other words, a smaller number of employees are working overtime in the organization.

Hourly Rate

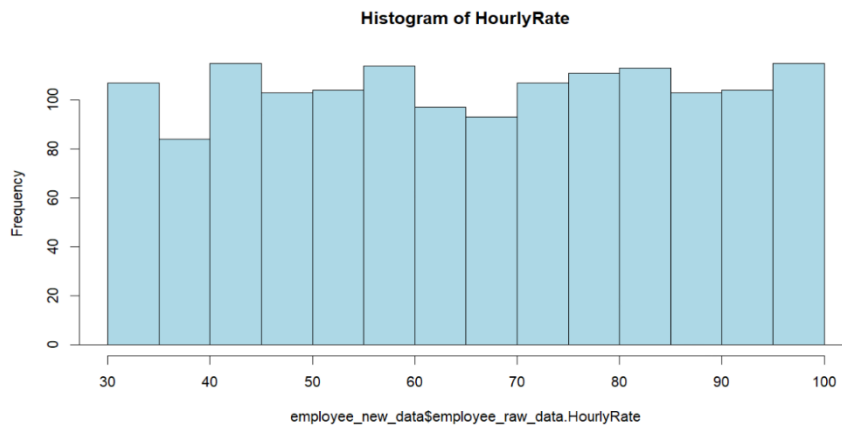


Figure 4.1.7: Histogram for employees' hourly rate

The above figure 4.1.7 shows the histogram of the employees' hourly pay rate. From this we can say that the employee salary range is somewhere around 80k – 120k range. Additionally, most of the employees' salaries are almost with a small difference gap.

Job Satisfaction

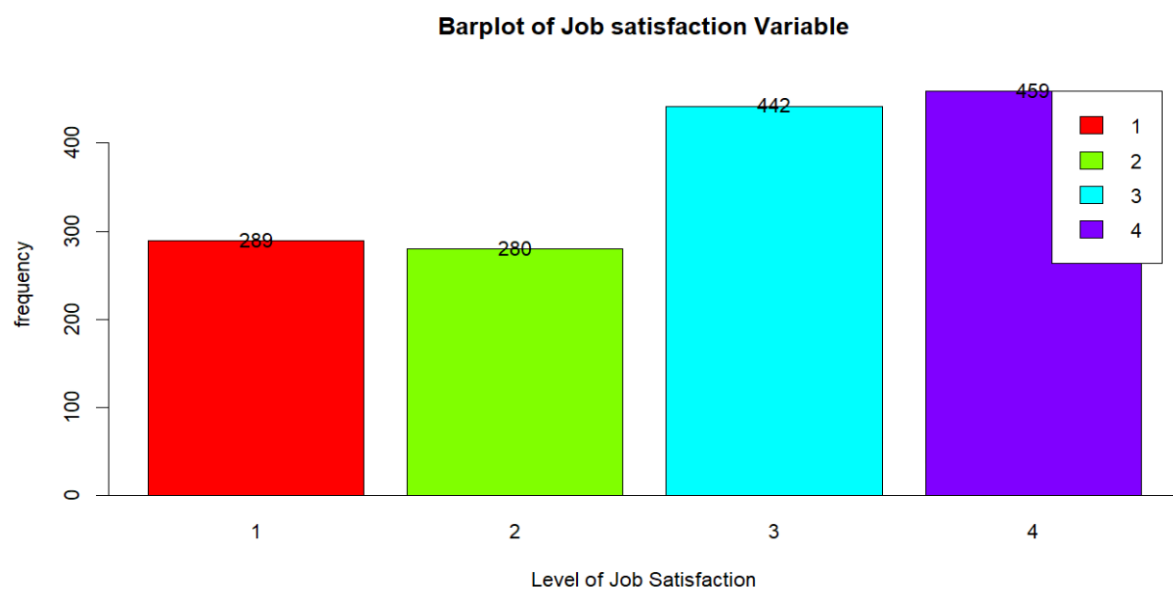


Figure 4.1.8: Bar chart of job satisfaction variable

The above figure 4.1.8 shows the job satisfaction level of employees. In here the 1 means the low satisfaction level and 4 means high satisfaction. The above bar chart shows the level of job satisfaction of the employees. This above chart clearly shows that most of the employees have higher level satisfaction and at the same time the half of that amount have low satisfaction level.

Environment Satisfaction

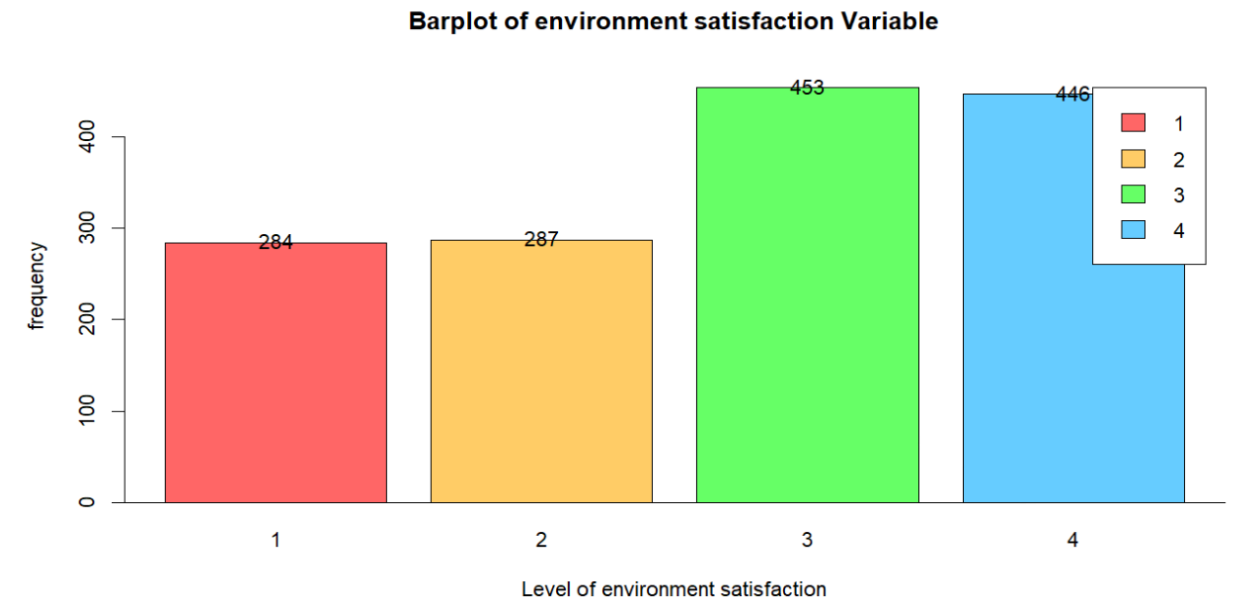


Figure4.1.9: Bar chart for employees' environment satisfaction level

The above figure 4.1.11 shows the environmental satisfaction level of employees. In here the 1 means the low satisfaction level and 4 means high satisfaction.

The above bar chart shows the level of environmental satisfaction of the employee. This above chart clearly shows that, the most of the employees have more than medium level of satisfaction.

Work-Life Balance

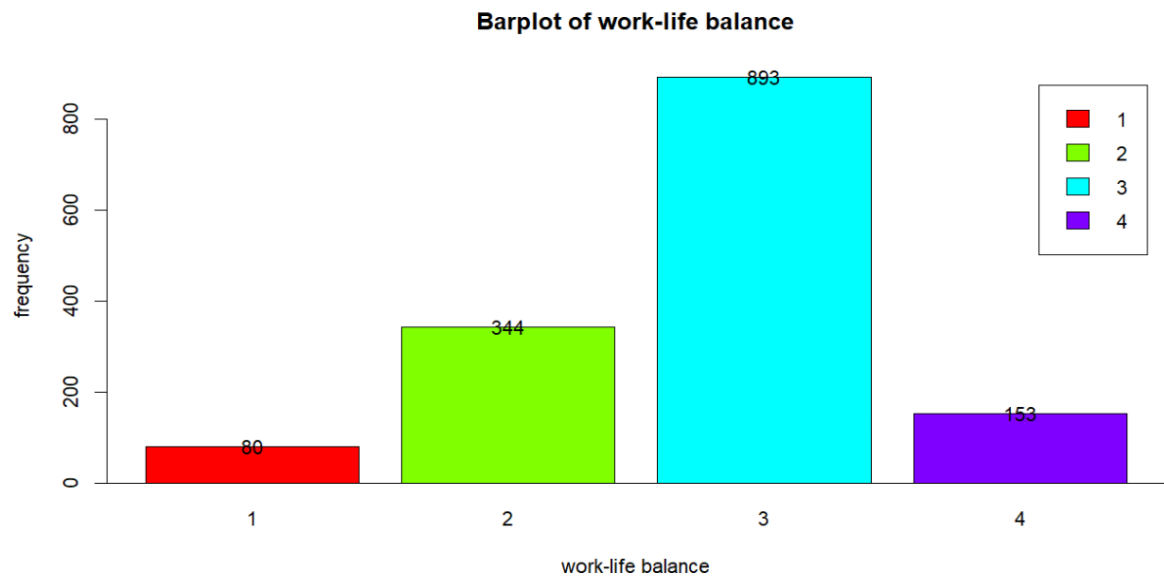


Figure 4.1.10 : Bar chart for employees level of work-life balance

The above figure 14 shows the level of work life balance of the employees. In here, 1 means very hard to manage and the 4 means can manage very easily. The above bar chart shows the level of work-life balance of the employees. So, from this we can say that the greatest number of the employees have a work- life balance in a good manner. Which means they can manage their work-life balance.

Analysis of the collected data

Main Objective 1:

Identifying is there an impact on employee attrition from the selected job-related factors.

Sub objective 1 - Identifying is there a relationship between job satisfaction and attrition.

```
#Here dependent variable is attrition and the independent variable is job satisfaction

#Null hypothesis (H0) -There are no relationship between job satisfaction and attrition.
#Alternative hypothesis (H1) - - There is a relationship between job satisfaction and attrition.

table <- table(employee_new_data$employee_raw_data.JobSatisfaction,employee_new_data$employee_raw_data.Attrition)
table

##
##      No Yes
## 1 223  66
## 2 234  46
## 3 369  73
## 4 407  52

test <- chisq.test(table)

test

##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 17.505, df = 3, p-value = 0.0005563
```

Figure 4.1.11: Coding and the output for performing first sub objective

Here, I have used a chi-squared test of independence to perform the first objective. So, for that we must create a frequency table first. After that we ran the chi-squared test to check whether is there any relationship between job satisfaction and employee attrition. Accordingly, I got the p value of 0.0005563.

Sub objective 2 - Identifying is there a relationship between working overtime and attrition

```
#Here dependent variable is attrition and the independent variable is working overtime

#Null hypothesis (H0) -There are no relationship between overtime and attrition.
#Alternative hypothesis (H1) - - There is a relationship between overtime and attrition.

table1 <- table(employee_new_data$employee_raw_data.OverTime,employee_new_data$employee_raw_data.Attrition)
table1

##
##      No Yes
## No  944 110
## Yes 289 127

test1 <- chisq.test(table1)

test1

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table1
## X-squared = 87.564, df = 1, p-value < 2.2e-16
```

Figure 4.1.12 : Coding and the output for performing 2nd sub objective

For this objective also I have used chi-squared test of independence as a statistical method to check whether there is any relationship between working overtime and employee attrition. For that I have followed the method that I have used in the previous objective. First, I created contingency table and then I ran the chi-square test. Accordingly, I got the p-value as 2.2e-16.

Sub objective 3 - Identifying is there a relationship between Environment satisfaction and attrition.

```
table2 <- table(employee_new_data$employee_raw_data.EnvironmentSatisfaction,employee_new_data$employee_raw_data.Attrition)
table2
```

```
##
##      No Yes
## 1 212  72
## 2 244  43
## 3 391  62
## 4 386  60
```

```
test2 <- chisq.test(table2)
test2
```

```
##
##  Pearson's Chi-squared test
##
## data:  table2
## X-squared = 22.504, df = 3, p-value = 5.123e-05
```

Figure 4.1.13: Coding and the output for performing 3rd sub objective.

To achieve this objective, I have also used chi-squared test of independence as a statistical analysis method. So here in this objective we must check whether there is any relationship between environmental satisfaction and employee attrition. So accordingly, we got 5.123e-05 as a p- value.

Sub objective 4 - Identifying is there a relationship between Hourly rate and attrition

```
# Fit logistic regression model
model <- glm(employee_new_data$employee_raw_data.Attrition ~ employee_new_data$employee_raw_data.HourlyRate,
data = employee_raw_data, family = "binomial")

# Summarize the model
summary(model)
```

Call:
glm(formula = employee_new_data\$employee_raw_data.Attrition ~
employee_new_data\$employee_raw_data.HourlyRate, family = "binomial",
data = employee_raw_data)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.5889155	0.2397695	-6.627	3.43e-11 ***
employee_new_data\$employee_raw_data.HourlyRate	-0.0009159	0.0034896	-0.262	0.793

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1298.6 on 1469 degrees of freedom
Residual deviance: 1298.5 on 1468 degrees of freedom
AIC: 1302.5

Number of Fisher Scoring iterations: 3

Figure 4.1.14: Coding and the output for performing 3rd sub objective.

For this objective I have used logistic regression as a statistical method in order to find whether is there any relationship between hourly rate and attrition. So here first I subset the variables and in addition to that I converted the attrition variable as a factor. Then I performed test in order to achieve the output. So here I got 0.793 as a p-value as an output.

Main Objective 2:

Identifying the most common level of education achieved by employees who have left the organization.

```
# Filter the dataframe to include only employees who have left the organization
left_df = df1[df1['Attrition'] == 'Yes']
```

```
# Count the number of occurrences of each education level and Sorting the result in descending order
```

```
education_counts = left_df['Education'].value_counts()
education_counts = education_counts.sort_values(ascending=False)
education_counts
```

```
3    99
4    58
2    44
1    31
5     5
Name: Education, dtype: int64
```

```
# Selecting the first element to obtain the most common level of education achieved by employees who have left the organization
most_common_education = education_counts.index[0]
print("The most common level of education achieved by employees who have left the organization is:", most_common_education)
```

```
The most common level of education achieved by employees who have left the organization is: 3
```

Figure 4.1.15: Coding and the output for performing 2nd main objective.

Here, this objective is to check what is the most common level of education achieved by the employees who left the organization. So, for that first I have extracted the employees who left the organization. In addition to that I have counted the number of occurrences of each educational level and sorted the output in descending order to see the most common level of the study level.

Main Objective 3:

Evaluate is there a relationship between an employee's perception of their work-life balance and their attrition.

```
table6 <- table(employee_new_data$employee_raw_data.WorkLifeBalance,employee_new_data$employee_raw_data.Attrition)
table6
```

```
##
##      No Yes
## 1  55  25
## 2 286  58
## 3 766 127
## 4 126  27
```

```
test6 <- chisq.test(table6)
test6
```

```
##
##  Pearson's Chi-squared test
##
## data:  table6
## X-squared = 16.325, df = 3, p-value = 0.0009726
```

Figure 4.1.16: Coding and the output for performing 3rd main objective.

This is the 3rd main and the final objective. In order to achieve this objective I have used the chi-squared test of independence as a statistical method to find whether is there any relationship between employees perception about their work life balance and the attrition variable. So here I have got 0.0009726 as a p-value.

4.2 Findings and Interpretation

In this part, we are going to discuss about the objective outputs and hypothesis. Accordingly, the first main objective is,

❖ Identifying is there an impact on employee attrition from the selected job-related factors.

Sub objective 1: Identifying is there an impact from working overtime on attrition.

For this objective I have used Chi-squared test of independent. In here the dependent and the independent variable are attrition and job satisfaction accordingly.

Hypothesis:

There is no relationship between job satisfaction and attrition. (H0)

There is a relationship between job satisfaction and attrition. (H1)

For this, the p-value I got is less than 0.05. So, we can reject null hypothesis(h0). Which means, there is a relationship between job satisfaction and attrition variables.

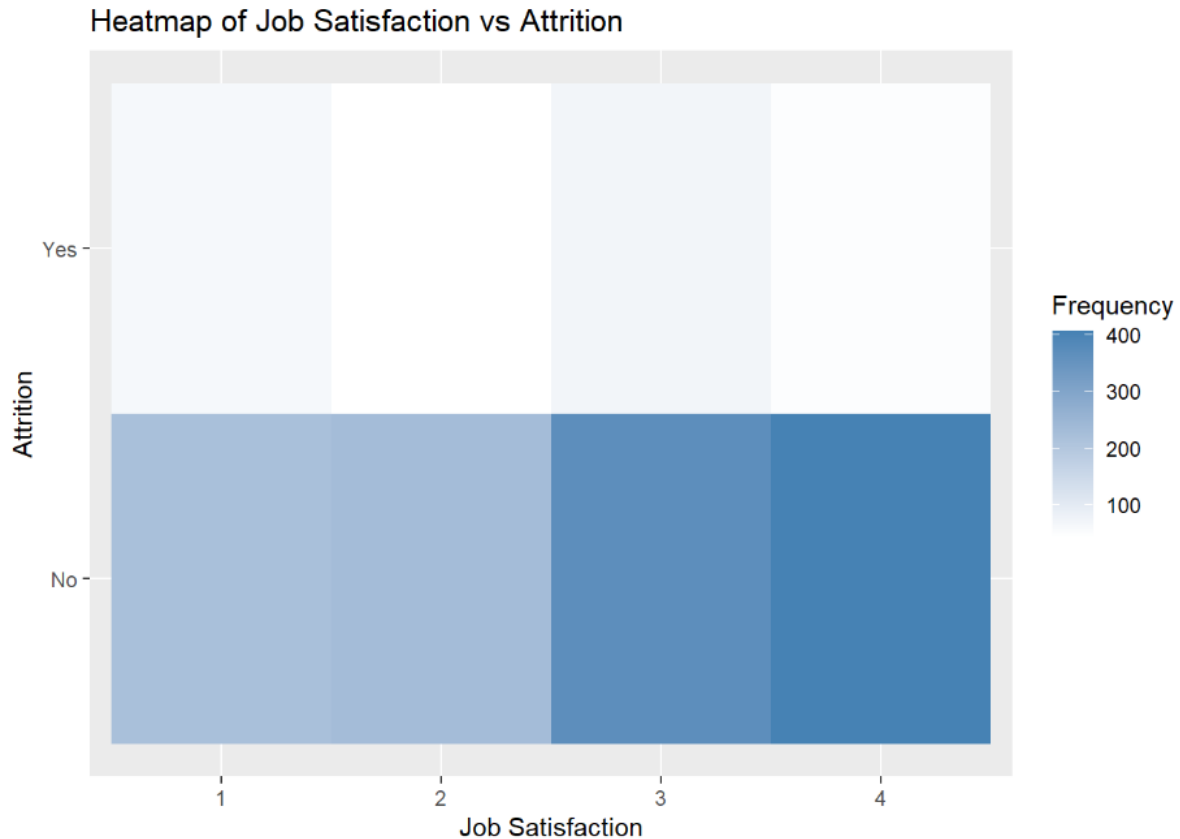


Figure 4.2.1: Heatmap for the 1st sub objective output

In here, Cells with higher frequencies will be represented by darker colors on the heatmap, while cells with lower frequencies will be lighter.

Here the cell corresponding to high Job Satisfaction and Attrition = No has a high frequency, it means that many employees with high job satisfaction did not leave the company. The cell corresponding to low Job Satisfaction and Attrition = no has a low frequency, it means that only fewer employees with low job satisfaction did not leave the company.

Sub objective 2: Identifying is there an impact from working overtime on attrition.

For this objective also I have used Chi-squared test of independent. In here the dependent and the independent variable are attrition and working overtime accordingly.

Hypothesis:

There is no relationship between working overtime and attrition. (H0)

There is a relationship between working overtime and attrition. (H1)

For this, the p-value I got is less than 0.05. So, we can reject null hypothesis. Which means, there is a relationship between working overtime and attrition variables.

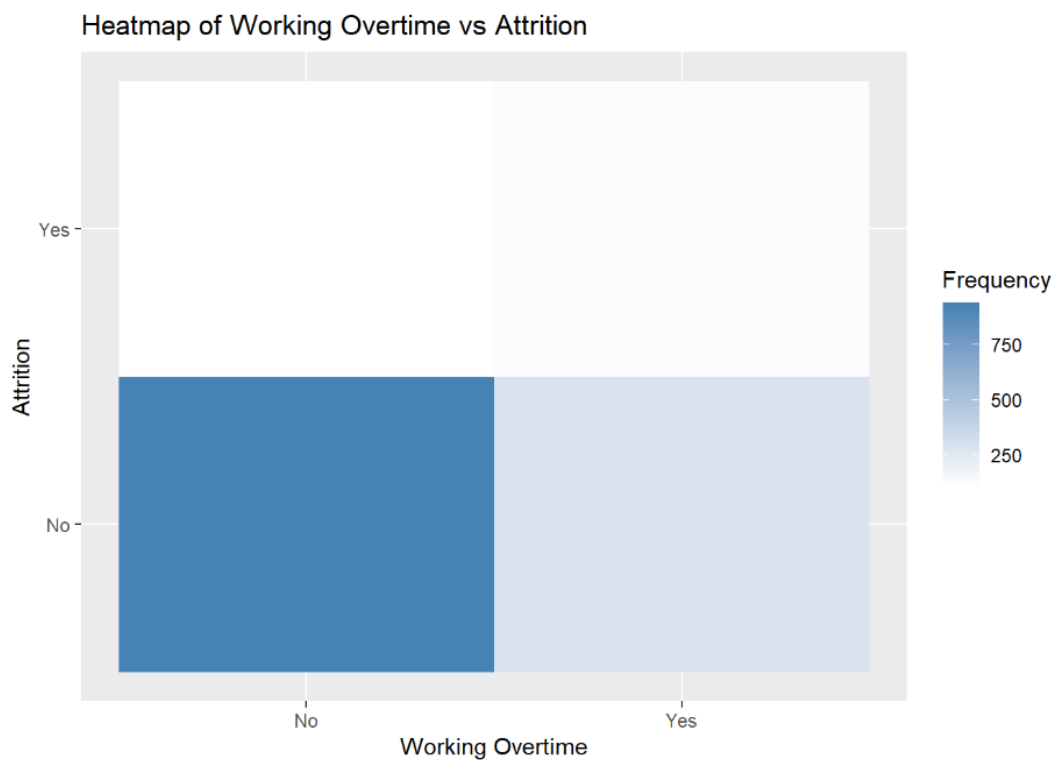


Figure 4.2.2: Heatmap for the 2nd sub objective output

A high frequency in a cell indicates that there are many occurrences of that specific combination of Working Overtime and Attrition levels. The cell corresponding to "Working Overtime = No" and "Attrition = No" has a high frequency, it means that many employees who did not work overtime don't have attrition which means they did not leave the company. In other word we can say that,

A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Working Overtime and Attrition levels. The cell corresponding to "Working Overtime = Yes" and "Attrition = No" has a low frequency, it means that there are only fewer employees who worked overtime did not leave the company.

Sub objective 3: Identifying is there an impact from environment satisfaction on attrition.

In order to achieve this objective also I have used Chi-squared test of independent. In here the dependent and the independent variable are attrition and environment satisfaction accordingly.

Hypothesis:

There is no relationship between environment satisfaction and attrition. (H0)

There is a relationship between environmental satisfaction and attrition. (H1)

For this, the p-value I got is less than 0.05. So, we can reject null hypothesis. Which means, there is a relationship between environmental satisfaction and attrition variables.

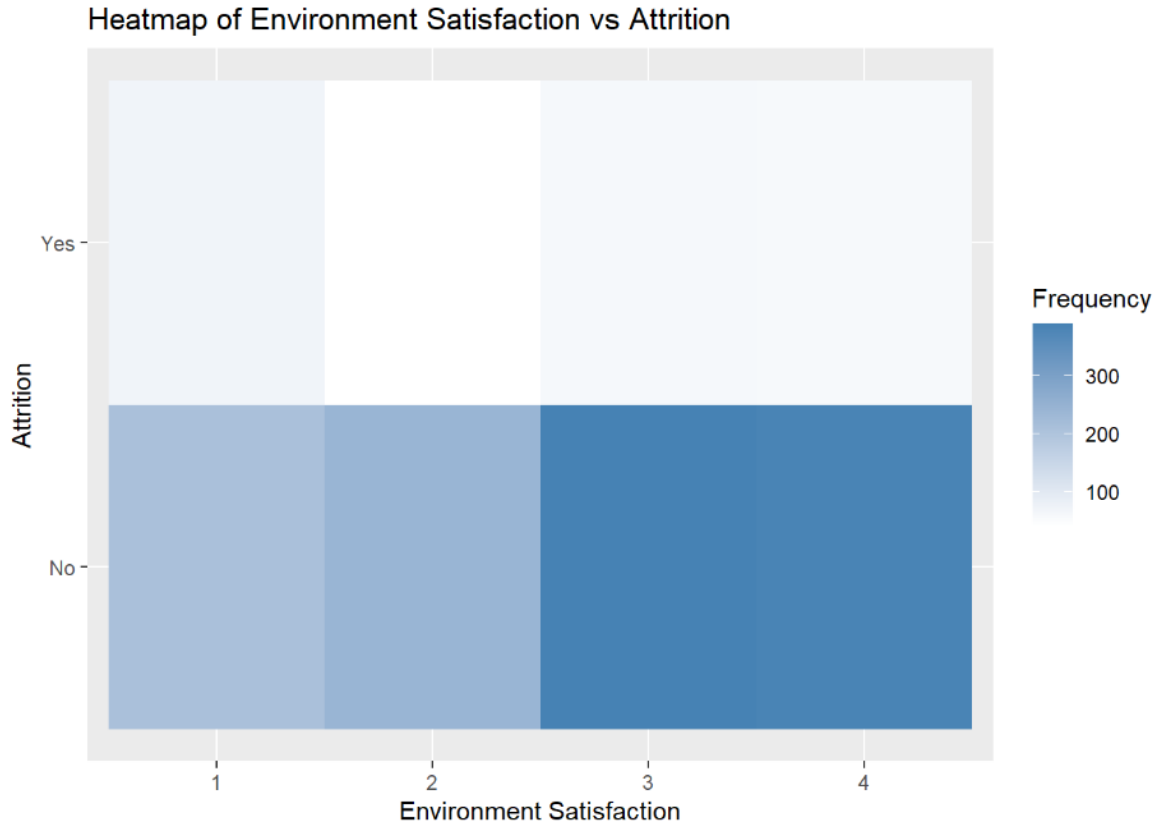


Figure 4.2.3: Heatmap for the 3rd sub objective output

Here a high frequency in a cell indicates that there are many occurrences of that specific combination of Environment Satisfaction level and Attrition status. The cell corresponding to "Environment Satisfaction = 3 and 4" and "Attrition = No" has a high frequency, it means that many employees with a level 3 and 4 environment satisfaction did not leave the company (They do not have attrition)

In other words, A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Environment Satisfaction level and Attrition status. The cell corresponding to "Environment Satisfaction = 2 and 1" and "Attrition = NO" has a low frequency, it means that fewer employees with a level 2 and 1 environment satisfaction are staying in the company which means so many left the company.

Sub objective 4: Identifying is there an impact from hourly rate on attrition.

For this objective, I have used Logistic Regression. In here, here dependent variable is attrition, and the independent variable is Hourly rate of pay variable.

Hypothesis:

There is no relationship between hourly rate and attrition. (H0)

There is a relationship between hourly rates and attrition. (H1)

According to the results of the Logistic regression test, the p-value (0.793) indicates that there is no correlation between hourly rate and attrition. However, I can still visualize this output using a boxplot to see the distribution of Hourly Rate across the two Attrition groups.

(Yes or no).

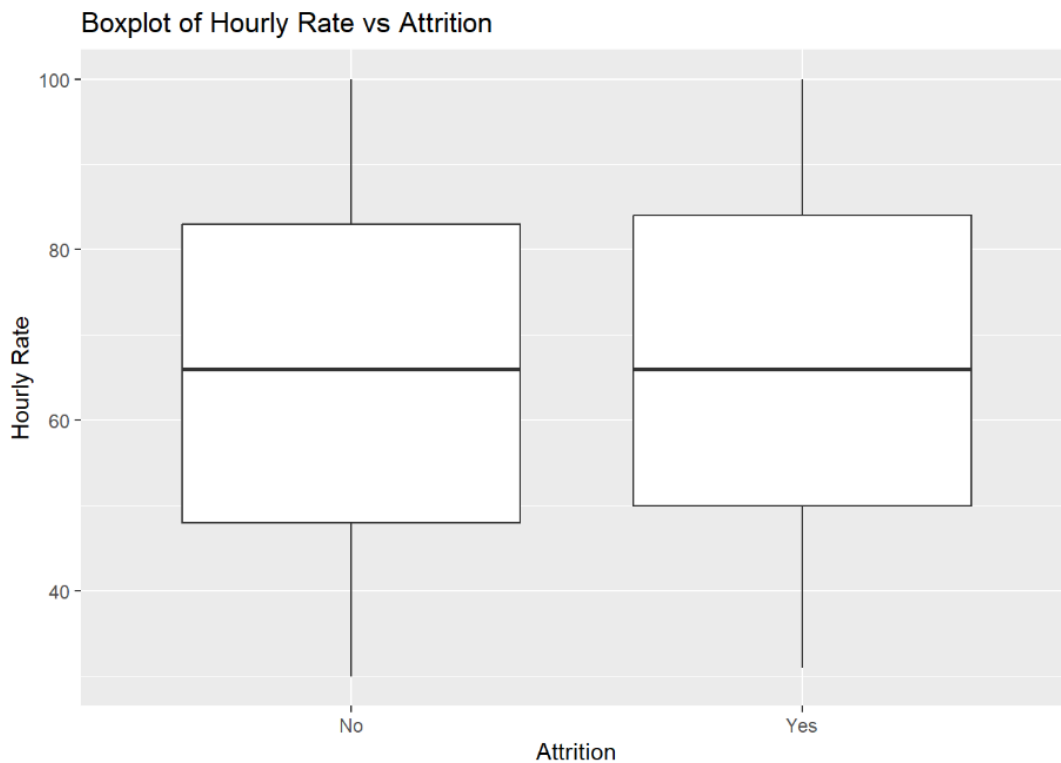


Figure 4.2.4: The boxplot of Hourly Rate vs Attrition

Since the Logistic regression suggests that there is no significant relationship between hourly rate and attrition variable, we do not have substantial differences in the boxplot.

Main Objective 2: Identifying the most common level of education achieved by employees who have left the organization.

Here, this objective is to check what is the most common level of education achieved by the employees who left the organization.

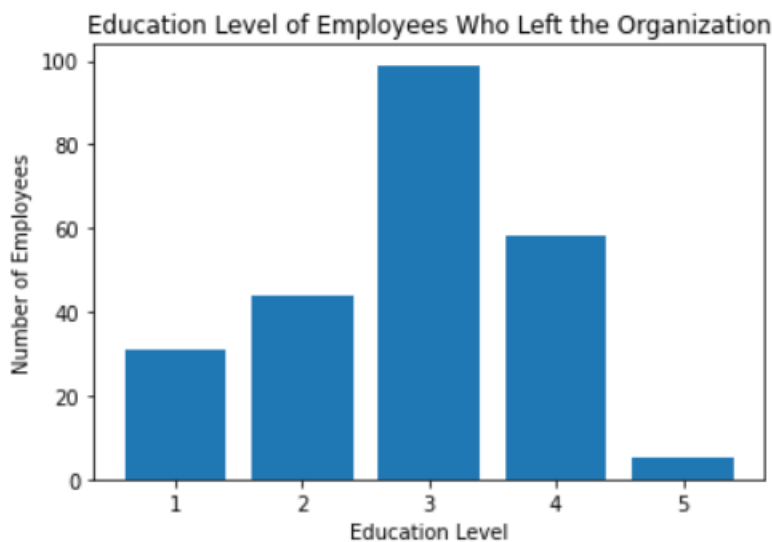


Figure 4.2.5: Bar plot of the employees' education level

So, from the sorted and the descending ordered table and from this above bar chart we can clearly see that the most common level of education level achieved the employees who left the organization is 3 which means bachelor's degree.

Main Objective 3: Evaluate is there a relationship between an employee’s perception of their work-life balance and their attrition.

Here in this objective, the dependent variable is attrition, and the independent variable is work-life balance variable.

Null hypothesis (H0) -There is no relationship between work-life balance variable and attrition.

Alternative hypothesis (H1) - - There is a relationship between work-life balance variable and attrition.

For this, the p-value I got is less than 0.05. So, we can reject null hypothesis. Which means, there is a relationship between employees’ perception about their work-life balance and attrition variables.

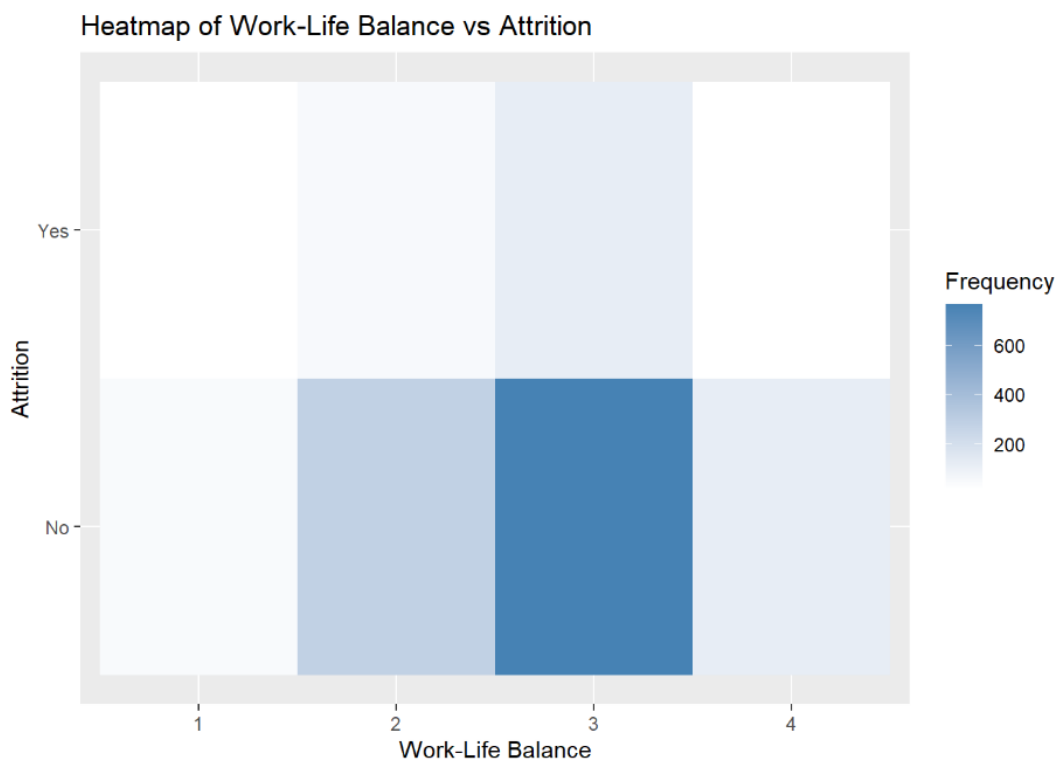


Figure 4.2.6: Heatmap for the 3rd main objective output

A high frequency in a cell indicates that there are many occurrences of that specific combination of Work-Life Balance level and Attrition status. The cell corresponding to "Work-Life Balance = 3" and "Attrition = No" has a high frequency, it means that many employees with a level 3 work-life balance did not leave the company (they did not have attrition).

A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Work-Life Balance level and Attrition status. For example, if the cell corresponding to "Work-Life Balance = 1" and "Attrition = Yes" has a low frequency, it means that fewer employees with a level 1 work-life balance left the company (experienced attrition).

Chapter 5: Discussion and Recommendations

5.1 Discussion

In this discussion part of the report, we are going to explore and outcome that we got from statistical analysis. The things I found through my statical analysis are.

1. There is a relationship between job satisfaction and attrition: So, we have statistical evidence that we can say there is a relationship between these 2 variables.
2. There is a relationship between working overtime and attrition: To prove this objective also I ran chi-squared test and now It proves that those both variables have a relationship.
3. There is a relationship between environmental satisfaction and attrition: To prove this also I used chi-squared test and now It has proper statistical evidence that says these 2 variables have a relationship.
4. There is no relationship between hourly rate of pay of the employees and the attrition: According to our dataset, I used ANOVA test (Analysis of variance) to check whether is there any relationship between these 2 variables or not. According to that output, there is no evidence to show that there is any relationship between those 2 variables.
5. The most common educational level achieved by the employees who left the organization is bachelor's degree.
6. There is a relationship between employees' perception about their work-life balance and attrition variables: To prove this objective also I ran chi-squared test and now It proves that those both variables have a relationship.

My study's findings, which are in line with those of Boondarig Ronra and Prof. Manat Chaisawat (2020), Ahituv and Lerman (2010), and Zimmerman and Darnold (2009), show a substantial correlation between job satisfaction and attrition. These studies also stress the significance of numerous aspects that affect job happiness, including pay, the work atmosphere, opportunities for career advancement, and incentives for employees. These conclusions are coming by my research, which also implies that businesses should focus on these issues in order to increase employee retention and job satisfaction.

According to Tulangow et al. (2018), work motivation and job satisfaction have a negative impact on turnover intention, which is consistent with my study's findings that there is a relationship between working overtime and attrition. Additionally, in line with Muhamad Ekhsan's findings, my research emphasizes the value of environmental pleasure and its connection to attrition.

My research found that there is no correlation between hourly rates of pay and attrition, which suggests that other factors at work, such work-life balance and job satisfaction, may be more crucial in influencing an employee's decision to leave a company. By focusing on enhancing these factors rather than just depending on monetary incentives, this insight may prove helpful for businesses trying to retain personnel.

My results on the most common educational level attained by departing employees (a bachelor's degree) can help businesses better recognize the kind of workers who are most likely to quit and create targeted retention plans in response.

5.1 Recommendations.

For this recommendation part must give recommendations based on my analysis and findings.

Some of my recommendations would be,

Raise job satisfaction: Companies should concentrate on raising different aspects of job satisfaction, including competitive compensation, chances for professional growth, and employee benefits. This can be accomplished by regularly surveying employees to determine areas that could be improved and then putting these areas the focus of focused projects.

Promote work-life balance: This is about motivating the employees to keep up a highly satisfied healthy work-life balance by providing them with flexible work hours, remote working opportunities, and plenty of vacation time to spend their time with their families. Implement guidelines that assist staff in effectively juggling their personal and professional obligations, such as facilitating access to childcare facilities, supporting mental health support, and setting up team-building exercises to foster a supportive workplace.

Improving the work environment: This is about Creating a welcoming and peaceful workplace that encourages teamwork, communication, and employee engagement to improve the working environment. Also we can provide staff with training and development opportunities to improve their skills.

Develop targeted retention strategies: Identify the employees who are more willing to quit (for example, those with a bachelor's degree) and build specific retention efforts for these groups. To assist people in advancing their careers, this can entail providing mentorship programs, professional development opportunities, and specialized training.

By executing these suggestions into practice, businesses may better retain talent and can be able to solve the problem of employee attrition, which will mostly result in a workforce that is more engaged and productive.

5.3 Conclusions

In conclusion, this study's findings point to several significant relationships between variables like job satisfaction, work-life balance, and attrition. To effectively lower staff turnover and increase retention, it is crucial to address these aspects, according to my findings and those from previous literature.

The study found a relationship between attrition and all of these factors, including job satisfaction, working overtime, and workplace satisfaction. I also discovered that there is no connection between attrition and hourly compensation.

Additionally, I found that there is a positive relationship between attrition and work-life balance variable and there are a higher attrition rate for those with bachelor's degrees. These findings gave me several suggestions that companies might adopt to improve their working environment, work-life balance, and job happiness to reduce staff turnover and increase employee retention. Among other things, these suggestions call for raising job satisfaction, encouraging work-life balance, strengthening the working environment, and creating specific retention plans.

Even though this research has offered insightful information, more research is still needed in some areas. Future studies should investigate other aspects of employee attrition, like business culture, employee engagement, and leadership philosophies. Additionally, longitudinal studies could be carried out to determine the best methods for lowering staff turnover across different industries and to look at the long-term consequences of employing retention tactics.

Appendices

Python

Codes

```
#Importing Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Load the dataset into a pandas dataframe
df = pd.read_csv('HR_Analytics.csv.csv')
```

df

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	1
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	8	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	10	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	7	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	8	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	6	
...
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	...	3	80	1	17	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	...	1	80	1	9	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	...	2	80	1	6	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	...	4	80	0	17	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	...	1	80	0	6	

1470 rows × 35 columns

```
#Extracting the variables
df1 = df[['Attrition', 'JobSatisfaction', 'OverTime', 'EnvironmentSatisfaction', 'HourlyRate', 'Education', 'WorkLifeBalance']]
df1
```

	Attrition	JobSatisfaction	OverTime	EnvironmentSatisfaction	HourlyRate	Education	WorkLifeBalance
0	Yes	4	Yes	2	94	2	1
1	No	2	No	3	61	1	3
2	Yes	3	Yes	4	92	2	3
3	No	3	Yes	4	56	4	3
4	No	2	No	1	40	1	3
...
1465	No	4	No	3	41	2	3
1466	No	1	No	4	42	1	3
1467	No	2	Yes	2	87	3	3
1468	No	2	No	4	63	3	2
1469	No	3	No	2	82	3	4

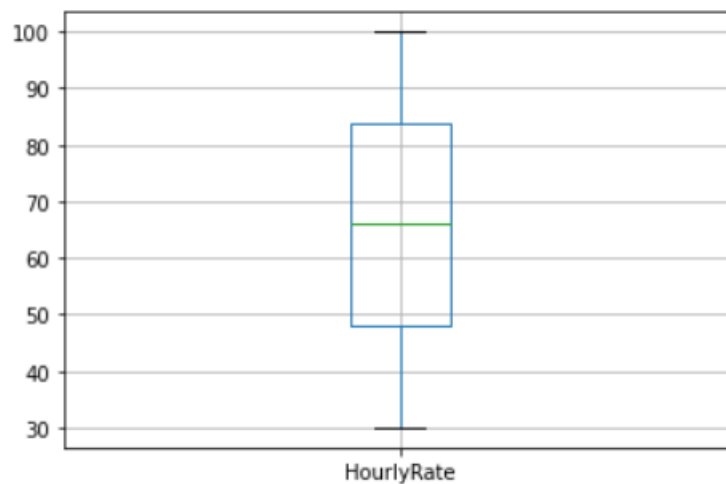
1470 rows × 7 columns

```
#Checking is there any null values
print(df1.isnull().sum())
```

```
Attrition          0
JobSatisfaction    0
OverTime           0
EnvironmentSatisfaction  0
HourlyRate         0
Education          0
WorkLifeBalance    0
dtype: int64
```

```
#Checking for outliers
df1.boxplot(column = ['HourlyRate'])
```

<AxesSubplot:>



```
# Main objective 2 - Identifying the most common level of education achieved by employees who have Left the organization.
#For that first we have to Calculate the mode of variable "Education" to check where the employee attrition equals to "Yes"
# Filter the dataframe to include only employees who have Left the organization
left_df = df1[df1['Attrition'] == 'Yes']
```

```
# Count the number of occurrences of each education Level and Sorting the result in descending order
education_counts = left_df['Education'].value_counts()
education_counts = education_counts.sort_values(ascending=False)
education_counts
```

```
3    99
4    58
2    44
1    31
5     5
Name: Education, dtype: int64
```

```
# Selecting the first element to obtain the most common level of education achieved by employees who have Left the organization
most_common_education = education_counts.index[0]
print("The most common level of education achieved by employees who have left the organization is:", most_common_education)
```

The most common level of education achieved by employees who have left the organization is: 3

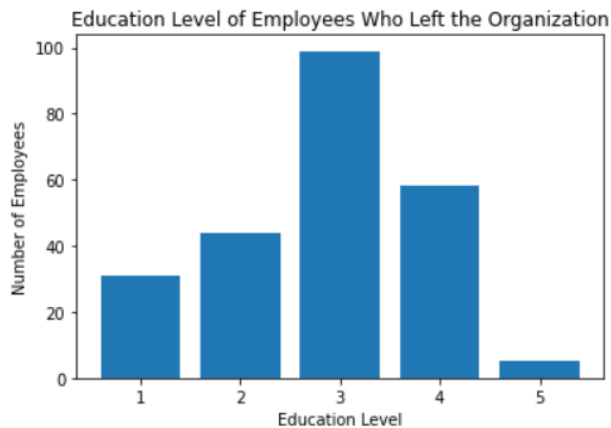
```
import matplotlib.pyplot as plt

# Data
education_levels = education_counts.index
count = education_counts.values

# Creating the bar chart
plt.bar(education_levels, count)

plt.title('Education Level of Employees Who Left the Organization')
plt.xlabel('Education Level')
plt.ylabel('Number of Employees')
plt.xticks(education_levels)

# Show the bar chart
plt.show()
```



So from here we can say that the most common level of education achieved by employees who have left the organization is: 3

R codes

#Reading the data set

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
employee_raw_data = read.csv("HR_Analytics.csv")
View(employee_raw_data)
```


#Extracting the using variables

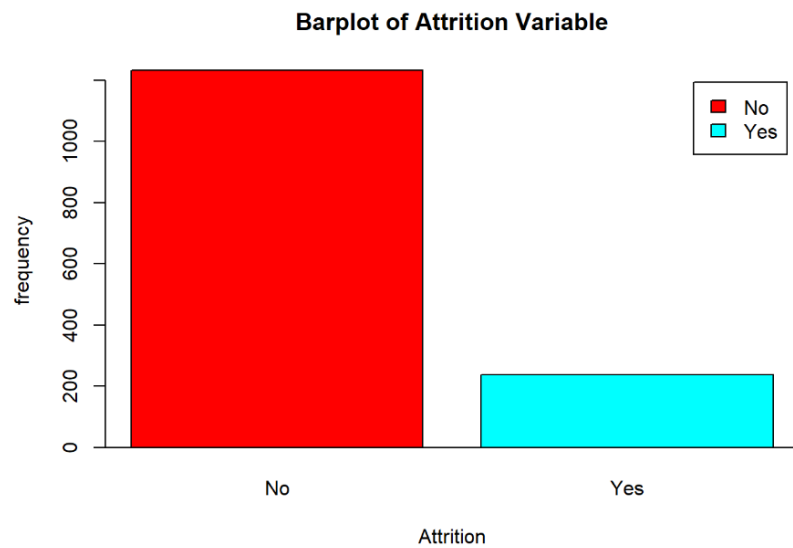
```
employee_new_data <- data.frame(employee_raw_data$JobSatisfaction,employee_raw_data$Attrition,employee_raw_data$O  
verTime,employee_raw_data$EnvironmentSatisfaction,employee_raw_data$HourlyRate,employee_raw_data$Education,employ  
ee_raw_data$WorkLifeBalance)  
  
View(employee_new_data)
```

#Visualization the variables seperately

```
#Contingency table for attrition  
table1 <- table(employee_new_data$employee_raw_data.Attrition)  
table1
```

```
##  
##   No   Yes  
## 1233  237
```

```
# Plot the bar chart  
barplot(table1,beside = F,  
        col=rainbow(2),  
        legend=rownames(table1),  
        main="Barplot of Attrition Variable",  
        xlab="Attrition",  
        ylab="frequency")  
abline(h=0)
```



Filter the dataframe to include only employees who have left the organization

```
left_df <- subset(employee_new_data, employee_new_data$employee_raw_data.Attrition == "Yes")
```

Counting the number of occurrences of each education level of the employees who left the organization.

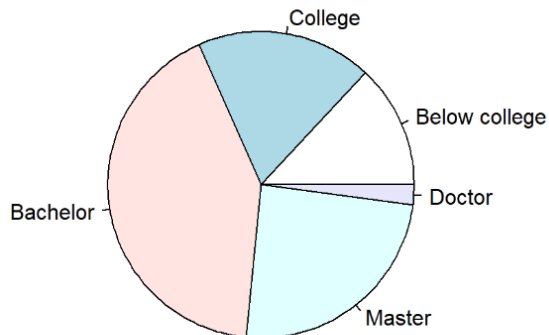
```
edu_counts = table(left_df$employee_raw_data.Education)
edu_counts
```

```
##
##  1  2  3  4  5
## 31 44 99 58  5
```

#Plotting pie chart to identify the level of education of the employees who left the organization

```
catego <- c("Below college", "College", "Bachelor", "Master", "Doctor")
val <- c(31, 44, 99, 58, 5)
pie(val, labels = catego, main = "Pie Chart for educational level")
```

Pie Chart for educational level

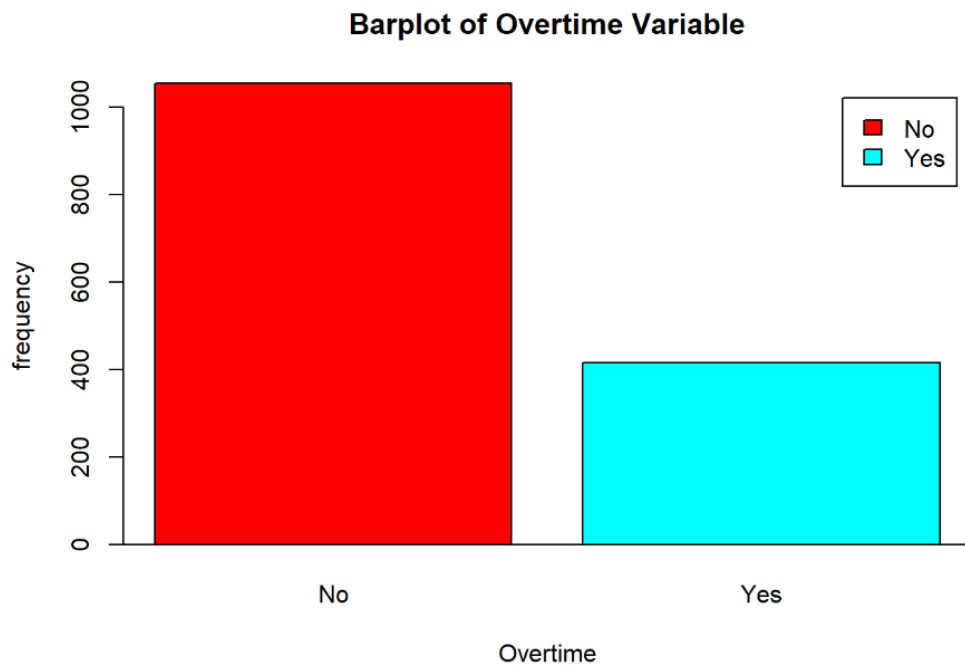


```
#Contingency table for Overtime
table2 <- table(employee_new_data$employee_raw_data.Overtime)
table2
```

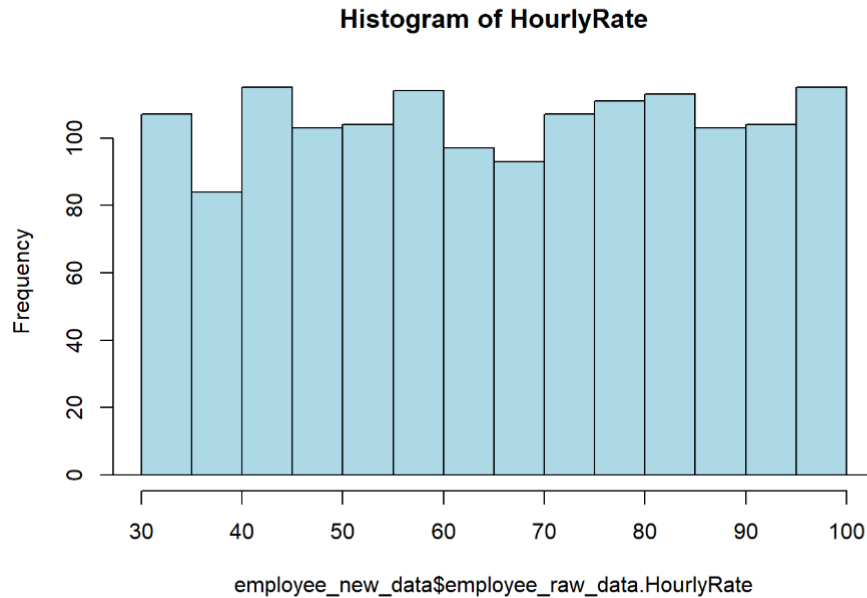
```
##
##   No  Yes
## 1054  416
```

#Creating Bar plot for Over time variable

```
# Plot the bar chart
barplot(table2,beside = F,
        col=rainbow(2),
        legend=rownames(table2),
        main="Barplot of Overtime Variable",
        xlab="Overtime",
        ylab="frequency")
abline(h=0)
```



```
#Create histogram for HourlyRate
hist(employee_new_data$employee_raw_data.HourlyRate,col="lightblue",
      main = "Histogram of HourlyRate")
abline(h=0)
```



```
#Contingency table for job satisfaction
table3 <- table(employee_new_data$employee_raw_data.JobSatisfaction)
table3
```

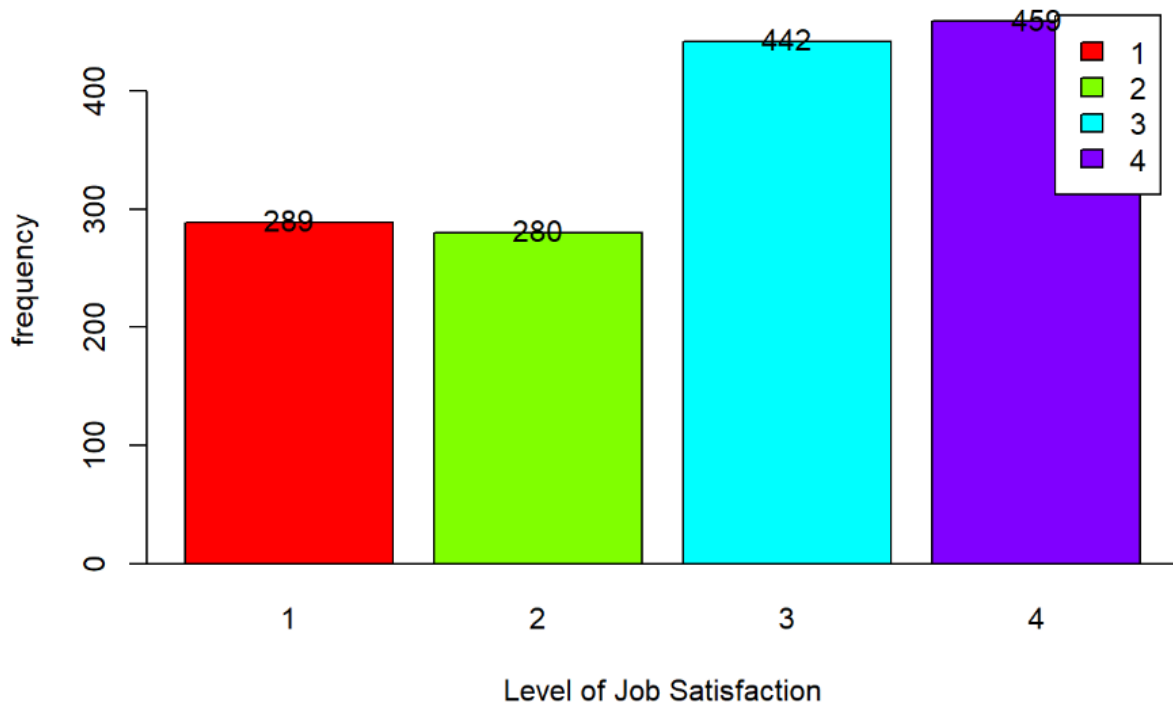
```
##
##  1  2  3  4
## 289 280 442 459
```

```
# Plotting the bar chart for Job satisfaction variable
barplot_data2 <- barplot(table3, beside = T,
                        col = rainbow(4),
                        legend = rownames(table3),
                        main = "Barplot of Job satisfaction Variable",
                        xlab = "Level of Job Satisfaction",
                        ylab = "frequency",
                        ylim = c(0, max(table3) + 20)) # Extend the y-axis range

abline(h = 0)

text(x = barplot_data2, y = table3 + 2, label = table3, cex = 1, col = "black")
```

Barplot of Job satisfaction Variable



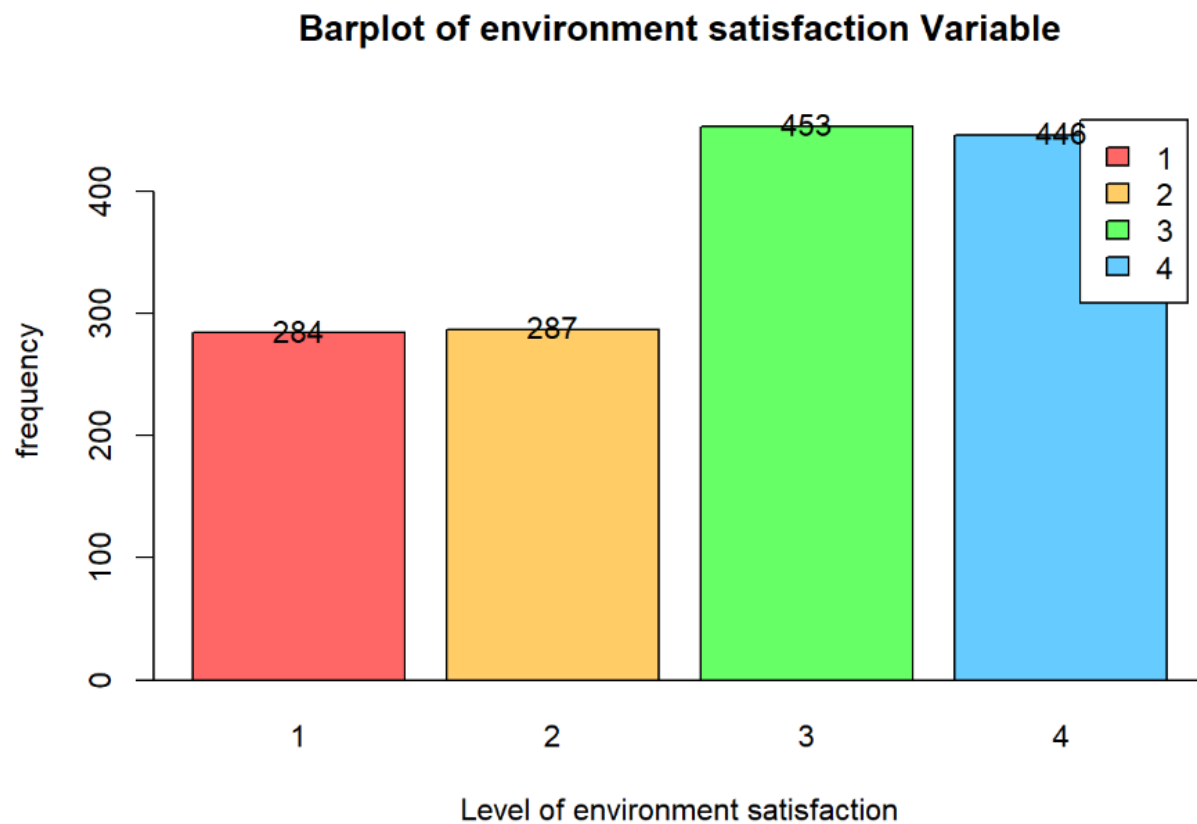
```
#Contingency table for environment satisfaction
table4 <- table(employee_new_data$employee_raw_data.EnvironmentSatisfaction)
table4
```

```
##
##  1  2  3  4
## 284 287 453 446
```

```
# Plotting the bar chart for environment satisfaction variable
barplot_data3 <- barplot(table4, beside = F,
  col = c("#FF6666", "#FFCC66", "#66FF66", "#66CCFF"),
  legend = rownames(table4),
  main = "Barplot of environment satisfaction Variable",
  xlab = "Level of environment satisfaction",
  ylab = "frequency",
  ylim = c(0, max(table4) + 20)) # Extend the y-axis range

abline(h = 0)

text(x = barplot_data3, y = table4 + 2, label = table4, cex = 1, col = "black")
```



```
#Contingency table for work-life balance
table5 <- table(employee_new_data$employee_raw_data.WorkLifeBalance)
table5
```

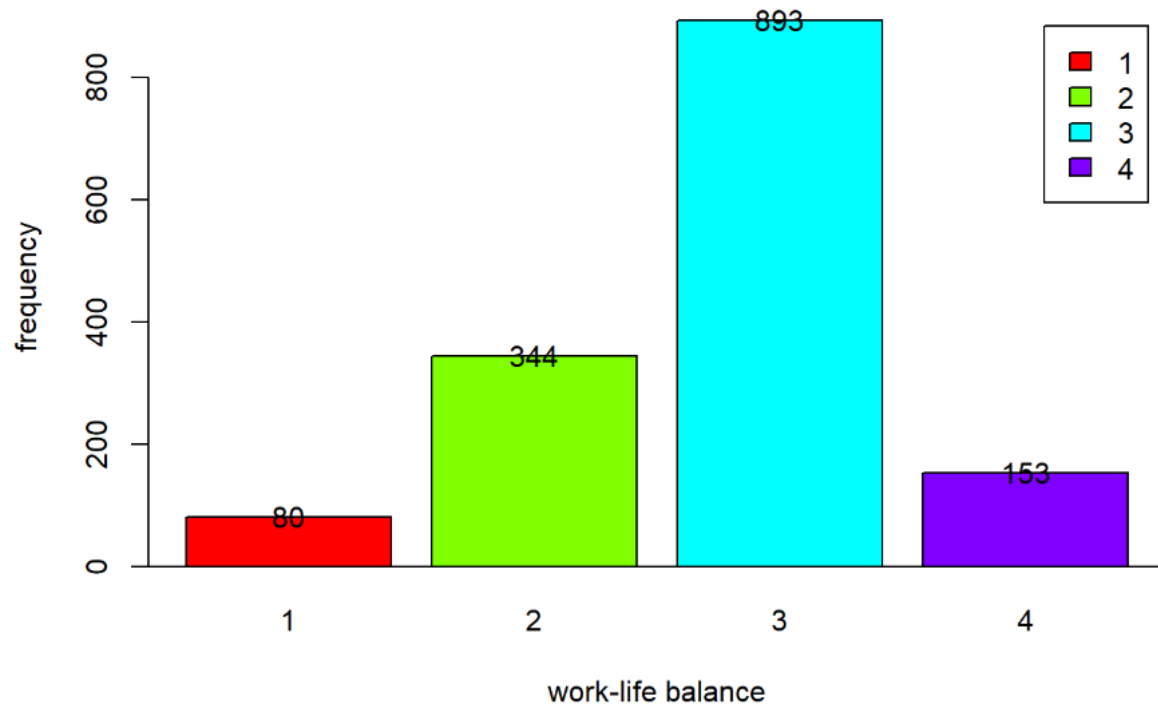
```
##
##  1  2  3  4
## 80 344 893 153
```

```
# Plotting the bar chart for work-life balance
barplot_data <- barplot(table5, beside = F,
  col = rainbow(4),
  legend = names(table5),
  main = "Barplot of work-life balance",
  xlab = "work-life balance",
  ylab = "frequency",
  ylim = c(0, max(table5) + 20)) # Extend the y-axis range

abline(h = 0)

# Adding data labels to the bar plot
text(x = barplot_data, y = table5 + 2, label = table5, cex = 1, col = "black")
```

Barplot of work-life balance



#Main Objectives

#Main Objective1 - Identifying is there an impact on employee attrition from the selected job-related factors.

#Sub objective 1 - Identifying is there a relationship between job satisfaction and attrition.

```
#Here dependent variable is attrition and the independent variable is job satisfaction

#Null hypothesis (H0) -There are no relationship between job satisfaction and attrition.
#Alternative hypothesis (H1) - - There is a relationship between job satisfaction and attrition.

table <- table(employee_new_data$employee_raw_data.JobSatisfaction,employee_new_data$employee_raw_data.Attrition)
table
```

```
##
##      No Yes
## 1 223  66
## 2 234  46
## 3 369  73
## 4 407  52
```

```
test <- chisq.test(table)

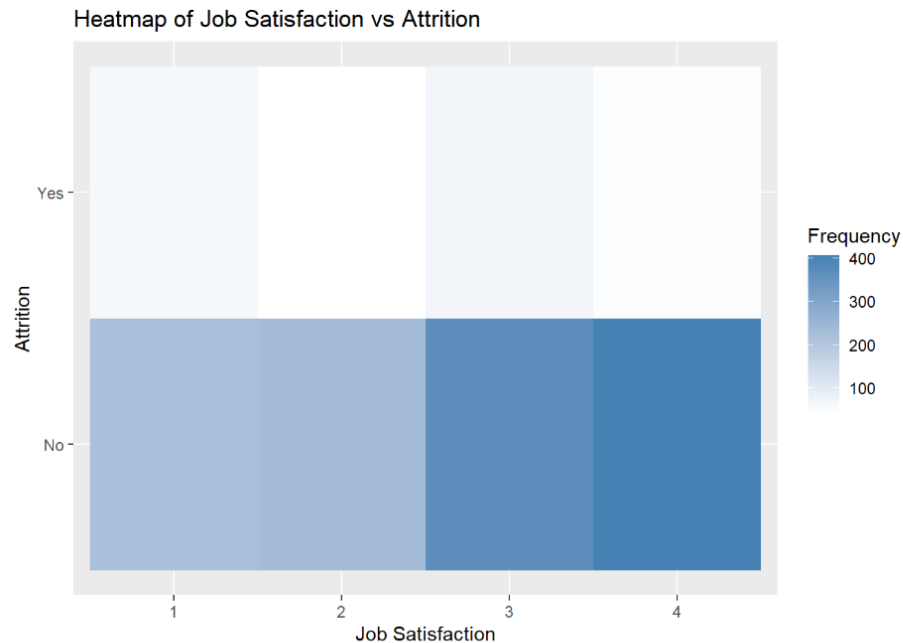
test
```

```
##
##  Pearson's Chi-squared test
##
## data:  table
## X-squared = 17.505, df = 3, p-value = 0.0005563
```

```
#For this, p-value is less than 0.05. So we can reject null hypothesis. Which means,There is a relationship between job satisfaction and attrition variables.
```

#Visualizing the output using a heatmap.

```
ggplot(as.data.frame(table), aes(x = Var1, y = Var2, fill = Freq)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "steelblue") +
  labs(x = "Job Satisfaction", y = "Attrition", fill = "Frequency") +
  ggtitle("Heatmap of Job Satisfaction vs Attrition")
```

#In here, Cells with higher frequencies will be represented by darker colors on the heatmap, while cells with lower frequencies will be lighter.

#Here the cell corresponding to high Job Satisfaction and Attrition = No has a high frequency, it means that many employees with high job satisfaction did not leave the company. The cell corresponding to low Job Satisfaction and Attrition = no has a low frequency, it means that only fewer employees with low job satisfaction did not leave the company.

#Sub objective 2 - Identifying is there a relationship between working overtime and attrition

#Here dependent variable is attrition and the independent variable is working overtime

#Null hypothesis (H0) - There are no relationship between overtime and attrition.

#Alternative hypothesis (H1) - - There is a relationship between overtime and attrition.

```
table1 <- table(employee_new_data$employee_raw_data.Overtime,employee_new_data$employee_raw_data.Attrition)
table1
```

```
##
##      No Yes
## No  944 110
## Yes 289 127
```

```
test1 <- chisq.test(table1)

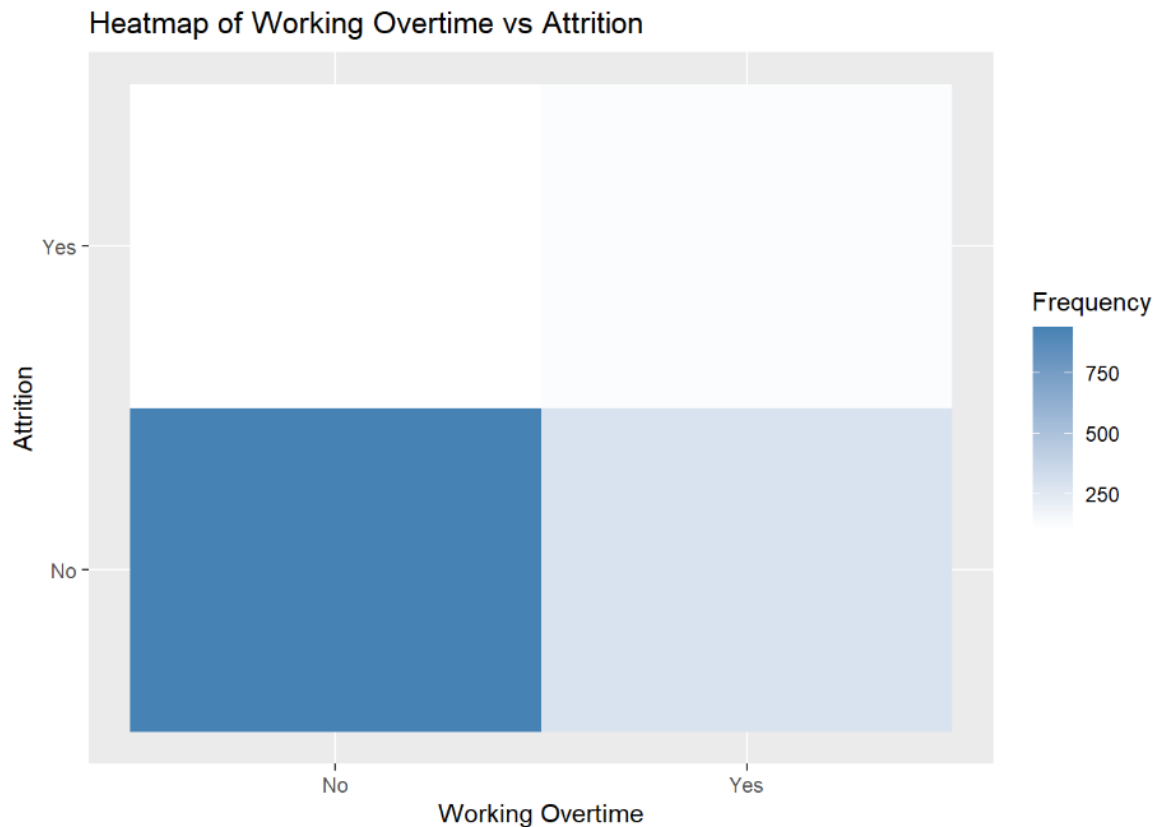
test1
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table1
## X-squared = 87.564, df = 1, p-value < 2.2e-16
```

#For this, p-value is less than 0.05. So we can reject null hypothesis. Which means, There is a relationship between working overtime and attrition variables.

#Visualizing the output using a heatmap.

```
ggplot(as.data.frame(table1), aes(x = Var1, y = Var2, fill = Freq)) +  
  geom_tile() +  
  scale_fill_gradient(low = "white", high = "steelblue") +  
  labs(x = "Working Overtime", y = "Attrition", fill = "Frequency") +  
  ggtitle("Heatmap of Working Overtime vs Attrition")
```



#A high frequency in a cell indicates that there are many occurrences of that specific combination of Working Overtime and Attrition levels. The cell corresponding to "Working Overtime = No" and "Attrition = No" has a high frequency, it means that many employees who did not work overtime don't have attrition which means they did not leave the company. In other word we can say that,

#A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Working Overtime and Attrition levels. The cell corresponding to "Working Overtime = Yes" and "Attrition = No" has a low frequency, it means that there are only fewer employees who worked overtime did not leave the company

#Sub objective 3 - Identifying is there a relationship between Environment satisfaction and attrition

```
#Here dependent variable is attrition and the independent variable is environment satisfaction variable

#Null hypothesis (H0) -There are no relationship between environment satisfaction and attrition.
#Alternative hypothesis (H1) - - There is a relationship between environment satisfaction and attrition.

table2 <- table(employee_new_data$employee_raw_data.EnvironmentSatisfaction,employee_new_data$employee_raw_data.A
ttrition)
table2
```

```
##
##      No Yes
## 1 212  72
## 2 244  43
## 3 391  62
## 4 386  60
```

```
test2 <- chisq.test(table2)

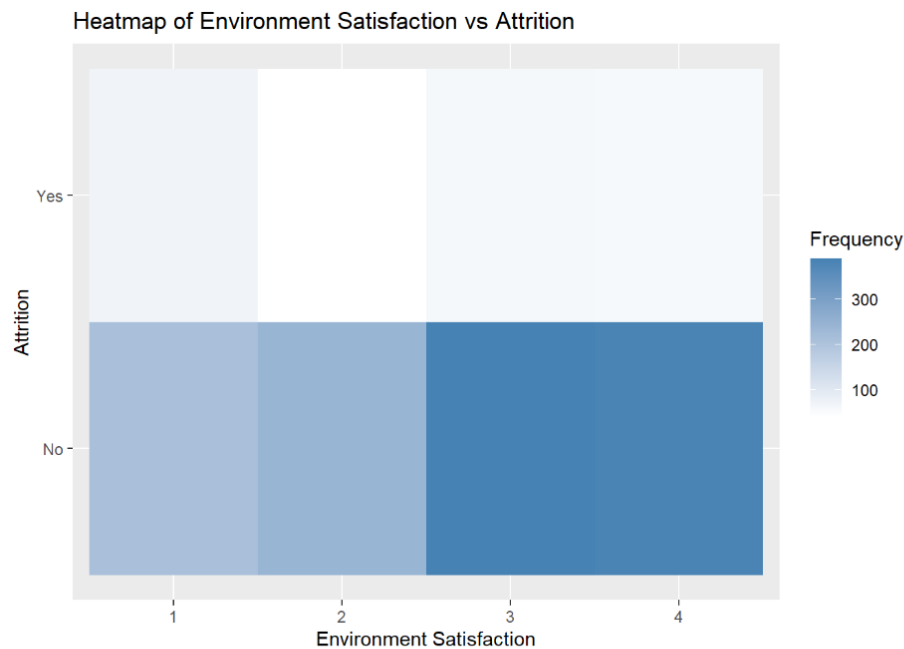
test2
```

```
##
##  Pearson's Chi-squared test
##
## data:  table2
## X-squared = 22.504, df = 3, p-value = 5.123e-05
```

```
#For this, p-value is less than 0.05. So we can reject null hypothesis. Which means,There is a relationship betwe
en environment satisfaction and attrition variables.
```

#Visualizing the output using heatmap

```
ggplot(as.data.frame(table2), aes(x = Var1, y = Var2, fill = Freq)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "steelblue") +
  labs(x = "Environment Satisfaction", y = "Attrition", fill = "Frequency") +
  ggtitle("Heatmap of Environment Satisfaction vs Attrition")
```



#A high frequency in a cell indicates that there are many occurrences of that specific combination of Environment Satisfaction level and Attrition status. The cell corresponding to "Environment Satisfaction = 3 and 4" and "Attrition = No" has a high frequency, it means that many employees with a level 3 and 4 environment satisfaction did not leave the company(They do not have attrition)

#Inother word, A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Environment Satisfaction level and Attrition status. The cell corresponding to "Environment Satisfaction = 2 and 1" and "Attrition = NO" has a low frequency, it means that fewer employees with a level 2 and 1 environment satisfaction are staying in the company which means so many left the company.

#Sub objective 4 - Identifying is there a relationship between Hourly rate and attrition

```
#Here dependent variable is attrition and the independent variable is Hourly rate of pay variable

#Null hypothesis (H0) -There are no relationship between Hourly rate and attrition.
#Alternative hypothesis (H1) - - There is a relationship between Hourly rate of and attrition.

# Subsetting the data to include only the relevant columns
employee_new_data1 <- subset(employee_new_data, select = c("employee_raw_data.Attrition", "employee_raw_data.HourlyRate"))

# Converting the Attrition variable to a factor
employee_new_data1$Attrition <- as.factor(employee_new_data1$employee_raw_data.Attrition)

# Performing the ANOVA test
anova_hr <- aov(employee_raw_data.HourlyRate ~ employee_raw_data.Attrition, data = employee_new_data1)

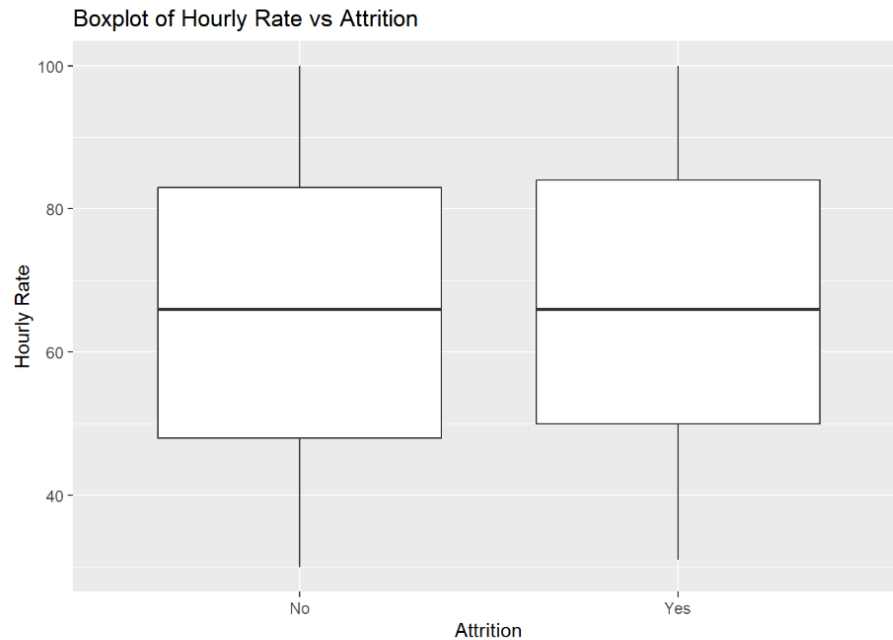
# Printing the results
summary(anova_hr)
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## employee_raw_data.Attrition  1      28    28.5   0.069   0.793
## Residuals              1468 607088   413.5
```

#Based on the output of the ANOVA test, we have a non-significant p-value (0.793), which suggests that there is no relationship between Hourly Rate and Attrition.

#Visualizing the output using boxplot

```
ggplot(employee_new_data1, aes(x = employee_raw_data.Attrition, y = employee_raw_data.HourlyRate)) +  
  geom_boxplot() +  
  labs(x = "Attrition", y = "Hourly Rate") +  
  ggtitle("Boxplot of Hourly Rate vs Attrition")
```



#Since the ANOVA test suggests no significant relationship, we do not have substantial differences in the boxplot.

#Main objective 3 - Evaluate is there a relationship between an employee's perception of their work-life balance and their attrition.

```
#Here dependent variable is attrition and the independent variable is work-life balance variable  
  
#Null hypothesis (H0) -There are no relationship between work-life balance variable and attrition.  
#Alternative hypothesis (H1) - - There is a relationship between work-life balance variable and attrition.  
  
table6 <- table(employee_new_data$employee_raw_data.WorkLifeBalance,employee_new_data$employee_raw_data.Attrition)  
table6
```

```
##  
##      No Yes  
## 1   55  25  
## 2  286  58  
## 3  766 127  
## 4  126  27
```

```
test6 <- chisq.test(table6)
```

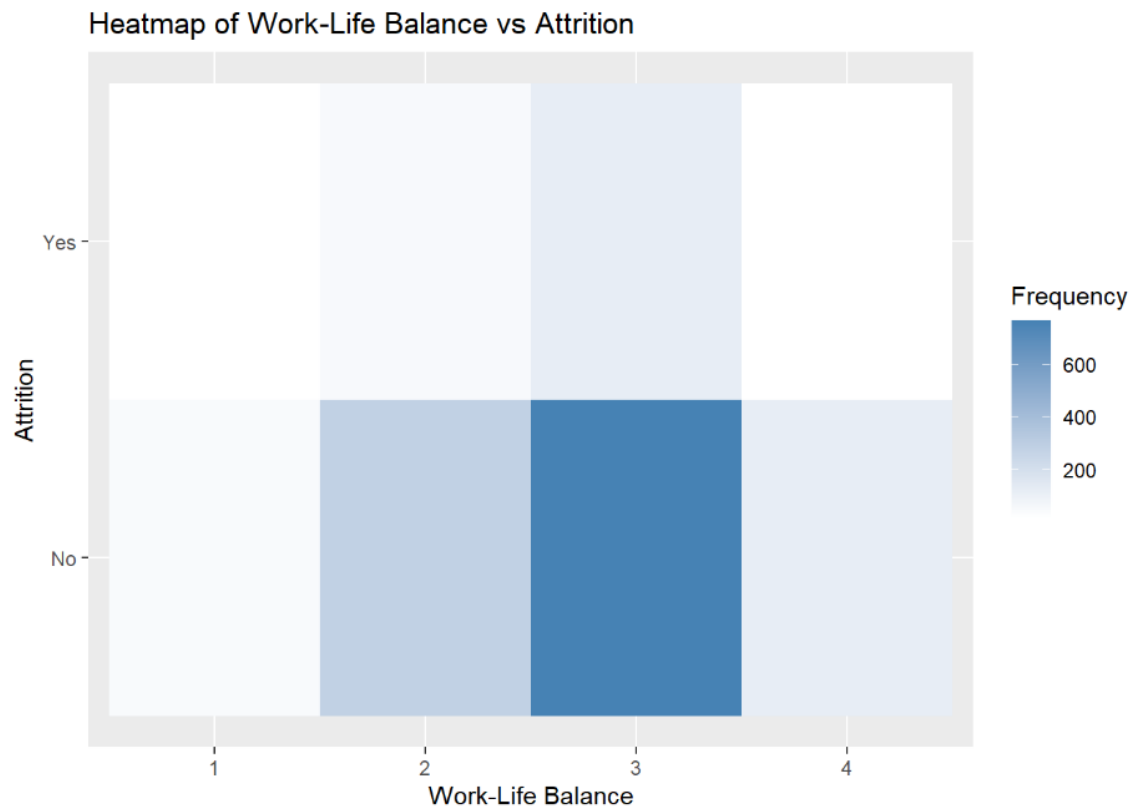
```
test6
```

```
##  
## Pearson's Chi-squared test  
##  
## data: table6  
## X-squared = 16.325, df = 3, p-value = 0.0009726
```

#For this, p-value is less than 0.05. So we can reject null hypothesis. Which means, There is a relationship between employees perception about their work-life balance and attrition variables.

#Visualizing the output using heatmap

```
ggplot(as.data.frame(table6), aes(x = Var1, y = Var2, fill = Freq)) +  
  geom_tile() +  
  scale_fill_gradient(low = "white", high = "steelblue") +  
  labs(x = "Work-Life Balance", y = "Attrition", fill = "Frequency") +  
  ggtitle("Heatmap of Work-Life Balance vs Attrition")
```



#A high frequency in a cell indicates that there are many occurrences of that specific combination of Work-Life Balance level and Attrition status. The cell corresponding to "Work-Life Balance = 3" and "Attrition = No" has a high frequency, it means that many employees with a level 3 work-life balance did not leave the company (they did not have attrition).

#A low frequency in a cell indicates that there are fewer occurrences of that specific combination of Work-Life Balance level and Attrition status. For example, if the cell corresponding to "Work-Life Balance = 1" and "Attrition = Yes" has a low frequency, it means that fewer employees with a level 1 work-life balance left the company (experienced attrition).

List of References

Data set –

The Devastator. (2020). Employee Attrition and Factors. Kaggle. Retrieved March 16, 2023, from <https://www.kaggle.com/datasets/thedevastator/employee-attrition-and-factors>

1. Ronra, B. and Chaisawat, A. (n.d.). *Factors Affecting Employee Turnover and Job Satisfaction: A Case Study of Amari Hotels and Resorts*. [online] Available at: http://tourismlibrary.tat.or.th/medias/PKPSU0007/PKPSU0007_fulltext.pdf.
2. Ahituv, A. and Lerman, R.I. (2010). Job turnover, wage rates, and marital stability: How are they related? *Review of Economics of the Household*, 9(2), pp.221–249. doi:<https://doi.org/10.1007/s11150-010-9101-6>.
3. Tulangow, M.J., Saerang, D.P.E. and Rumokoy, F.S. (2018). THE EFFECT OF JOB STRESS, WORK ENVIRONMENT AND WORKLOAD ON EMPLOYEE TURNOVER INTENTION (CASE STUDY AT PT. WIKA REALTY MANADO). *Jurnal EMBA : Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, [online] 6(1). doi:<https://doi.org/10.35794/emba.6.1.2018.19270>.
4. Khan, A. and Aleem, M. (n.d.). Scientific Papers. [online] Available at: <https://www.jois.eu/files/11.pdf>.

5. Sousa-Poza, A. and Sousa-Poza, A.A. (2007). The effect of job satisfaction on labor turnover by gender: An analysis for Switzerland. *The Journal of Socio-Economics*, 36(6), pp.895–913. doi:<https://doi.org/10.1016/j.socec.2007.01.022>.
6. Zimmerman, R.D. and Darnold, T.C. (2009). The impact of job performance on employee turnover intentions and the voluntary turnover process. *Personnel Review*, 38(2), pp.142–158. doi:<https://doi.org/10.1108/00483480910931316>.
7. Ekhsan, M. (2019). THE INFLUENCE JOB SATISFACTION AND ORGANIZATIONAL COMMITMENT ON EMPLOYEE TURNOVER INTENTION. *Journal of Business, Management, & Accounting*, [online] 1(1). Available at: <https://e-journal.stie-kusumanegara.ac.id/index.php/jobma/article/view/74>.
8. Steel, P. & Warner, J., 2019. Job satisfaction, organizational commitment, and turnover intention: A meta-analysis. *Journal of Applied Psychology*, 104(4), pp. 354-372.
9. Griffeth, R.W., Hom, P.W. & Gaertner, S., 2019. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 45(6), pp. 2439-2468.
10. Lee, Y.-H., Kim, S. & Yoon, J.H., 2018. The effects of work-life balance, job autonomy, and social support on job satisfaction and turnover intention among white-collar workers. *International Journal of Occupational Safety and Ergonomics*, 24(3), pp. 371-37