ALLIANCE
UNIVERSITY
Alliance School of Business

EXECUTIVE POST GRADUATE DIPLOMA IN MANAGEMENT

Title: Analysis on Employee Attrition & Performance

PROJECT REPORT SUBMITTED TO ALLIANCE UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE COURSE:

Prepared by

NAME: RAKSHITHA .R

REGISTRATION NO.: 2022EPGD05ASB025

BATCH: MAY - 2022

SPECIALIZATION: BUSINESS INTELLIGENCE AND ANALYTICS

Under the guidance of

Dr. / Prof. NANDISH MANANGI

ALLIANCE
UNIVERSITY
Alliance School of Business

Executive Post Graduate Diploma in Management

Term – IV: Project Work

Student Declaration Form

This is to declare that the Report entitled **“Analysis on Employee Attrition & Performance”** has been made for the partial fulfilment of the Course: Project Work in Term – IV by me at **Alliance School of Business under the guidance** of **Dr./Prof Nandish Manangi**

I confirm that this Report truly represents my work undertaken as a part of my Project Work. This work is not a replication of work done previously by any other person. I also confirm that the contents of the Report and the views contained therein have been discussed and deliberated with the Faculty Guide.

Signature of the Student :

Name of the Student : **RAKSHITHA .R**

Registration No. : **2022EPGD05ASB025**

Specialization : **Business Intelligence and Analytics**

ALLIANCE
UNIVERSITY
Alliance School of Business

Executive Post Graduate Diploma in Management

Faculty Guide Certificate

This is to certify that **Ms. Rakshitha R** Registration No. **2022EPGD05ASB025** has completed the Report entitled **“Analysis on Employee Attrition & Performance”** under my guidance for the partial fulfillment of the Course: Project Work in Term – IV of the Executive Post Graduate Diploma in Management.

Signature of Faculty Guide

Dr./Prof Nandish Manangi

Acknowledgement

I would want to convey my heartfelt gratitude to **Dr./Prof. Nandish Manangi**, my mentor, for his invaluable advice and assistance in completing my project. He was there to assist me every step of the way, and his motivation is what enabled me to accomplish my task effectively. I would also like to thank all of the other supporting personnel who assisted me by supplying the equipment that was essential and vital, without which I would not have been able to perform efficiently on this project.

I would also want to thank the University of **Alliance School of Business** for accepting my project in my desired field of expertise. I'd also like to thank my friends and parents for their support and encouragement as I worked on this assignment.

Signature of the Student:

Name of the Student : **RAKSHITHA .R**

Registration No. : **2022EPGD05ASB025**

Specialization : **Business Intelligence and Analytics**

TABLE OF CONTENTS

Chapter 1

1.1 Introduction

Employee attrition Performance and are two crucial elements that can have a big impact on an organization's success. Employee attrition, or the pace at which individuals leave a company, can increase the expenses of hiring and training new employees as well as result in the loss of institutional knowledge and expertise, lowered morale among surviving employees, and a reduction in institutional knowledge and experience. The productivity, quality, and customer happiness of an organisation can all be impacted by staff performance.

Employee attrition and performance can be analysed to provide important information regarding an organization's strengths and deficiencies. Organisations may improve employee engagement and retention by figuring out the underlying reasons for employee churn and taking action to solve them. Organisations may enhance employee performance and overall success by recognising high-performing individuals and offering them chances for growth and development.

This analysis will examine the connection between employee attrition and performance and offer suggestions for how businesses can manage both successfully to foster a productive workplace that encourages employee satisfaction, engagement, and success as a whole.

Employees depart from the company for a variety of reasons. The reason may be a higher wage at another company, family mobility, a preference for new technology, a higher position, etc. This unhappiness is a personal phenomenon that no organisation can effectively manage. Attrition causes organisations to suffer losses and incur additional costs, though. Employers invest a lot of time, money, and effort in training and developing their staff members to improve the effectiveness of their work. In the event that one employee departs the company, a new employee must be hired. In order to find and train a new employee, the same time, effort, and resources must be expended. The majority of IT organisations are currently suffering greatly from high attrition rates, which have a negative impact on many productivity and quality-related issues. Thus, it

Maintaining positive relationships with employees could be one of the key retention strategies. Any workplace's employee-employer connection is crucial. Employees that have good and healthy employer-employee relationships feel valued, motivated, and supported.

Employee satisfaction increases the likelihood that they will put in extra effort and stay on the job for a long time.

However, attrition happens when an employee's expectations of their company or of their employment aren't met. Each person has unique professional and personal aspirations. Both of those have a specific connection to the company he works for.

So it's critical to align employee expectations with organisational expectations. When expectations or interests are not aligned, it leads to job discontent and employee underperformance. Attrition occurs as a result. Employee develops the belief that he is not a good fit for the position as soon as he realises that his professional ambitions do not align with the organization's aims. He begins to explore for employment opportunities outside of the company.

The role that is promised to the employee and the role that he ultimately fills can differ greatly. This new position does not assist the employee in developing his professional abilities, which demotivates him to remain with the group or organisation. Lack of learning chances also seems to be a factor in individuals leaving their jobs. Attrition may result from numerous requirements mismatches.

This project work also aims to understand the expectations of both employers and employees at various levels about numerous employment-related topics. The article also seeks to understand the causes of industry attrition.

1.2 Objective of the study

The Objective of this study on employee attrition and performance is to examine the relationship between these two variables and offer suggestions for how businesses can manage them well to foster a productive workplace that encourages employee satisfaction, engagement, and success in general. The study specifically seeks to:

1. List the elements that affect employee retention, such as commute distance, industry competition, retirement, job satisfaction, pay and benefits, workload and work-life balance, leadership and management, organisational culture, work relationships, and opportunities for growth and development.

2. Analyse the effects of employee attrition on a company's performance, taking into account the loss of institutional knowledge and experience, the morale of the surviving staff, and the higher expenses of hiring and training new personnel.

3. Analyse staff performance data to find high performers and areas for development. This analysis should include indicators like productivity, quality, and customer happiness.

4. Give advice on how companies can manage employee attrition and performance in an efficient manner, including tactics to deal with the reasons of attrition, offer chances for professional advancement, and foster employee engagement and satisfaction.

The general Objective of this study is to offer perceptions and suggestions that can assist organisations in developing a favourable work environment that encourages employee retention, engagement, and success.

1.3 Scope of the study

The Scope of this study on employee attrition and performance is to examine the causes of attrition as well as how it affects an organization's performance. To identify high-performing individuals and areas for improvement, the study will also analyse employee performance data, including indicators like productivity, quality, and customer happiness.

The study will also offer suggestions on how organisations can manage employee attrition and performance successfully, including tactics to deal with the underlying causes of attrition, offer chances for professional advancement, and foster employee engagement and satisfaction.

However, the goal of this study is not to offer a comprehensive analysis of any particular businesses or organisations. The suggestions made will be all-encompassing and suitable for a variety of organisations. Additionally, the study will not address other significant elements like workplace diversity and inclusion, worker health and wellness, or technological improvements that may have an impact on employee turnover and performance.

The overall goal of this study is to offer a general understanding of the connection between employee attrition and performance and to offer suggestions for how businesses can manage these variables to foster a productive workplace that encourages employee satisfaction, engagement, and success.

1.4 Cause of Employee Attrition

A number of reasons, such as the following, can contribute to employee attrition:

1. Job dissatisfaction: When workers feel underpaid, overworked, or unappreciated at their jobs, they may search for alternative employment.

2. Lack of professional development opportunities: Employees want to feel like their careers are progressing and advancing. They can leave their current organisation in quest of better possibilities if they believe there are no opportunities for growth and development there.

3. Poor management and leadership: Poor management and leadership can foster a hostile work atmosphere by making people feel unappreciated and devalued, which increases attrition.

4. Issues with work-life balance: Employees may leave a company if they are unable to balance their personal and professional lives, which can result in burnout and reduced job satisfaction.

5. Organisational culture: If the workplace environment is poisonous, staff members could feel uneasy and wish to leave.

6. Compensation and benefits: Workers may resign if they believe their pay is unfair or their desired benefits are not provided.

7. Long commuting times can significantly contribute to a reduction in staff retention.

8. Competition in the industry: Strong industry competition may result in more job opportunities for employees, making it simpler for them to quit their current employer.

9. Retirement: Workers may leave their jobs for personal or other reasons.

Identifying and addressing the root causes of employee attrition can help organizations increase employee retention and engagement.

1.5 Types of Employee Attrition

There are two main types of employee attrition: voluntary and involuntary:

1. Voluntary attrition: This occurs when employees voluntarily leave an organization, such as through resignation, retirement, or a decision to pursue other opportunities. Voluntary attrition can be caused by a variety of factors, including job dissatisfaction, lack of growth opportunities, poor leadership and management, work-life balance issues, organizational culture, compensation and benefits, commute distance, and industry competition.

2. Involuntary attrition: This occurs when an organization terminates an employee's employment, such as through layoffs or performance-related issues. Involuntary attrition can be caused by a variety of factors, including changes in business strategy or market conditions, budget cuts, restructuring, or poor performance.

Both types of attrition can have a significant impact on an organization, as they can result in a loss of institutional knowledge and experience, decreased morale among remaining employees, and increased costs associated with recruiting and training new employees.

It is important for organizations to understand the root causes of both voluntary and involuntary attrition and take steps to address them in order to create a positive work environment that promotes employee retention and engagement.

1.6 Problem Identification

The problem that this analysis on employee attrition and performance seeks to address is the negative impact that high employee attrition rates can have on an organization's performance. High attrition rates can lead to a loss of institutional knowledge, decreased productivity, increased costs associated with recruiting and training new employees, and decreased employee morale.

The study also aims to pinpoint the elements that affect employee performance and how they affect employee attrition. Organisations may improve staff retention and engagement by addressing the fundamental reasons of attrition. This will raise productivity, boost customer satisfaction, and improve organisational performance overall.

In order to create plans for enhancing overall employee performance and satisfaction, the analysis also aims to pinpoint high-performing people and areas for development. Organisations can increase organisational performance over time by identifying and keeping high-performing workers, who help to maintain a good work environment that promotes growth and development.

1.7 Calculations of Employee Attrition

Employee attrition can be calculated using the following formula:

Attrition rate = (Number of employees who left during a period / Average total number of employees during the period) x 100

For example, let's say that a company had 500 employees at the beginning of the year and 50 employees left during the year. At the end of the year, the company had 475 employees.

To calculate the attrition rate, we would first need to calculate the average number of employees during the period:

Average total number of employees = (Beginning number of employees + Ending number of employees) / 2

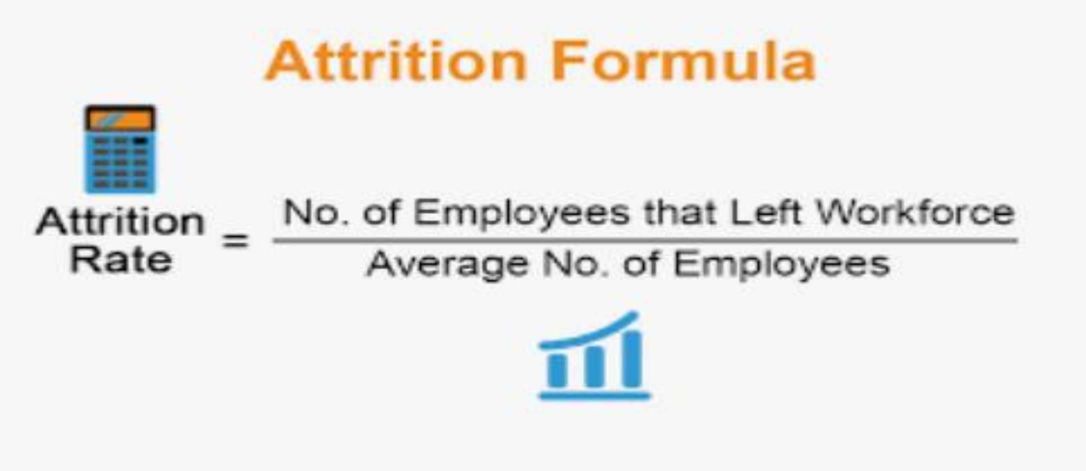
Average total number of employees = (500 + 475) / 2
Average total number of employees = 487.5

Next, we can calculate the attrition rate:

Attrition rate = (Number of employees who left during the year / Average total number of employees during the year) x 100

Attrition rate = (50 / 487.5) x 100
Attrition rate = 10.26%

Therefore, the attrition rate for this company during the year was 10.26%.



Attrition Formula

$$\text{Attrition Rate} = \frac{\text{No. of Employees that Left Workforce}}{\text{Average No. of Employees}}$$

Chapter 2

2.1 Methodology

Data collection is an essential part of exploratory data analysis. It refers to the process of finding and loading data into our system. Good, reliable data can be found on various public sites or bought from private organizations. Some reliable sites for data collection are Kaggle data base.

The study is conducted among working IT professionals of two different categories. This categorisation mainly was focused on experience level and role in the organisation. It was important to know the views of candidates who seek for the job for various reasons as well as the views of interviewers involved in the process of hiring the candidates. The research study involves reference of both primary and secondary data.

2.1.1 Data Collection

Type of data collected (primary and secondary data):

Data collection is the process of gathering the information which is needed for the research or the analysis. It can be acquired from primary or secondary sources, and data can be numerical or behavioural.

Primary data refers to the first-hand data collected by the researcher us.

Secondary data means the data collected by someone else earlier.

In this project the primary data has been collected from the website:

https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

2.1.2 Sampling Techniques

Simple Random Sampling: In this technique, a random sample of employees is selected from the entire population of employees in the organization. This technique is useful when the population is large and homogeneous.

2.1.3 Tools Used for Analysis

Some of the most common data tools used to create an EDA include:

Python: An interpreter, object-oriented programming language with dynamic semantics. Its high-level, built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components together. Python and EDA can be used together to identify missing values in a data set, which is important so you can decide how to handle missing values for machine learning.

Chapter 3.

Findings and Discussions:

3.1 Exploratory data analysis :

3.1.1 Introduction :

Exploratory Data Analysis is a process of examining or understanding the data and extracting insights or main characteristics of the data. EDA is generally classified into two methods, i.e. graphical analysis and non-graphical analysis.

It is essential to examine all variables in the dataset to:

- Catch mistakes
- Generate hypotheses
- See patterns in the data
- Extract important variables
- Detect outliers and anomalies
- Gain deep familiarity with the dataset
- Refine selection of features that will be used to build the machine learning models.

Exploratory data analysis tools

Specific statistical functions and techniques you can perform with EDA tools include:

- Clustering and dimension reduction techniques, which help create graphical displays of high-dimensional data containing many variables.
- Univariate visualization of each field in the raw dataset, with summary statistics.
- Bivariate visualizations and summary statistics that allow you to assess the relationship between each variable in the dataset and the target variable you're looking at.
- Multivariate visualizations, for mapping and understanding interactions between different fields in the data.
- K-means Clustering is a clustering method in unsupervised learning where data points are assigned into K groups, i.e. the number of clusters, based on the distance from each group's centroid. The data points closest to a particular centroid will be clustered under the same category. K-means Clustering is commonly used in market segmentation, pattern recognition, and image compression.

- Predictive models, such as linear regression, use statistics and data to predict outcomes.

Types of exploratory data analysis

There are four primary types of EDA:

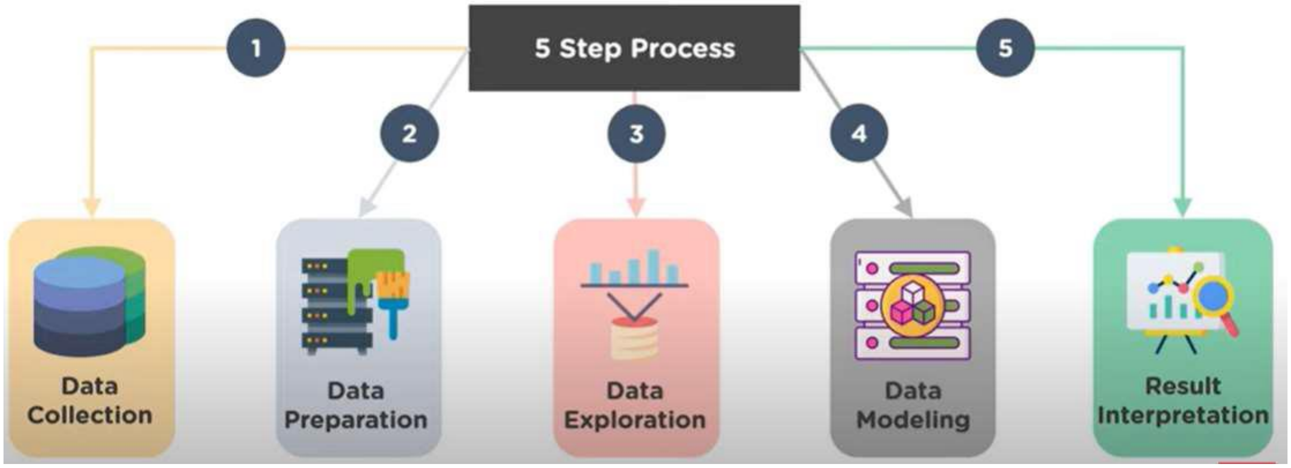
- Univariate non-graphical.
- Univariate graphical.
- Multivariate no graphical:
- Multivariate graphical:

Exploratory Data Analysis Tools

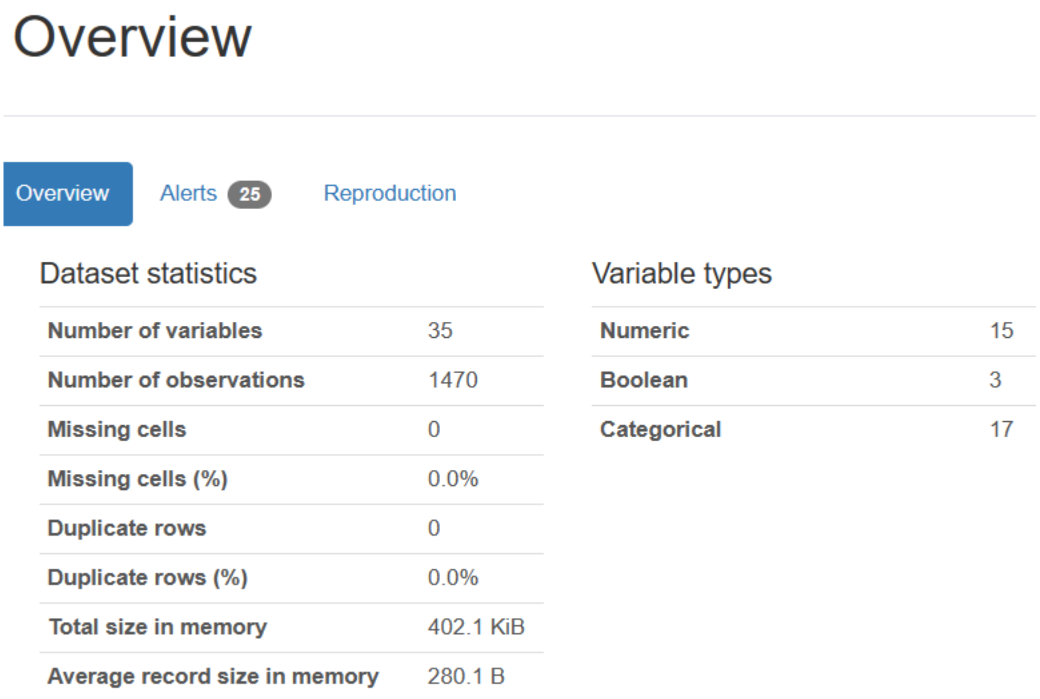
Some of the most common data science tools used to create an EDA include:

- **Python:** An interpreter, object-oriented programming language with dynamic semantics. Its high-level, built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as a scripting or glue language to connect existing components together. Python and EDA can be used together to identify missing values in a data set, which is important so you can decide how to handle missing values for machine learning.
- **R:** An open-source programming language and free software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians in data science in developing statistical observations and data analysis.

3.1.2 Steps Involved in Exploratory Data Analysis:



3.2 Data Overview



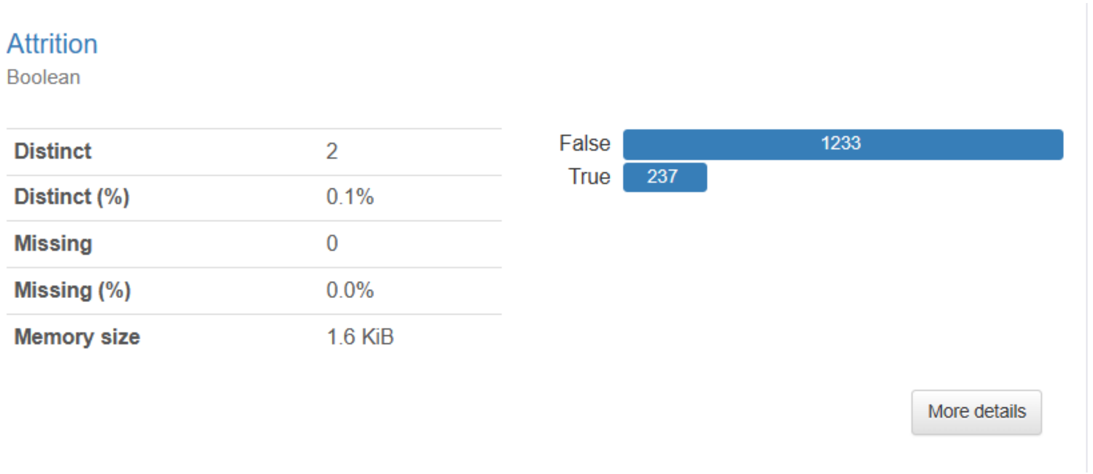
Overview

Dataset statistics

Number of variables	35
Number of observations	1470
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	402.1 KiB
Average record size in memory	280.1 B

Variable types

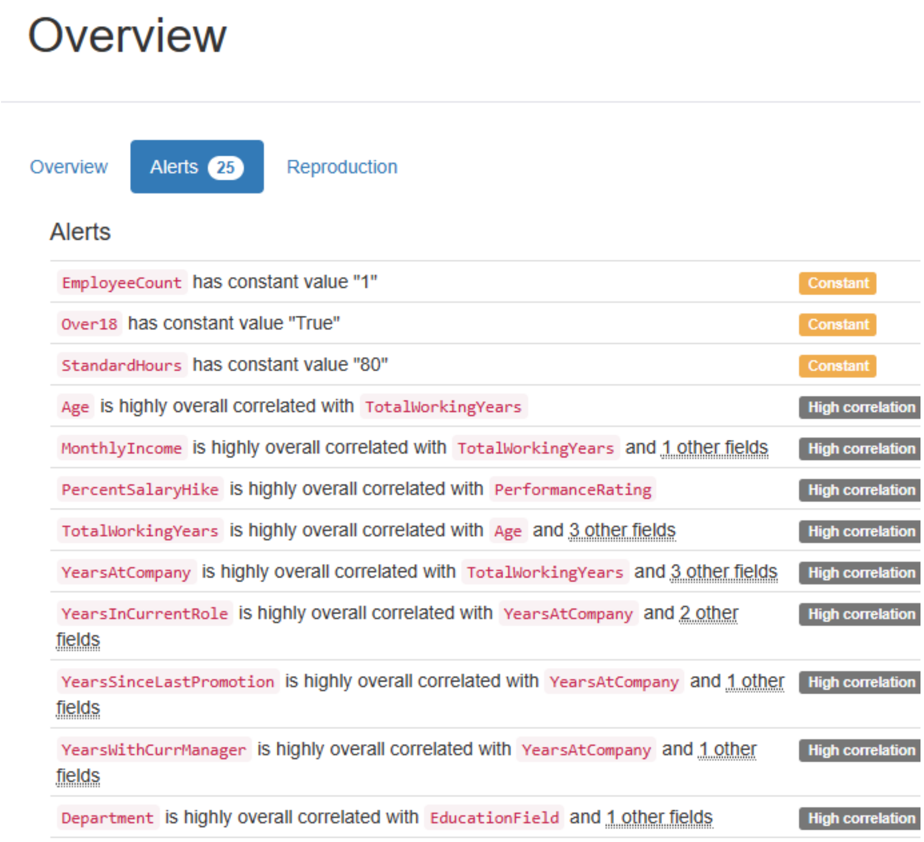
Numeric	15
Boolean	3
Categorical	17



Attrition

Boolean

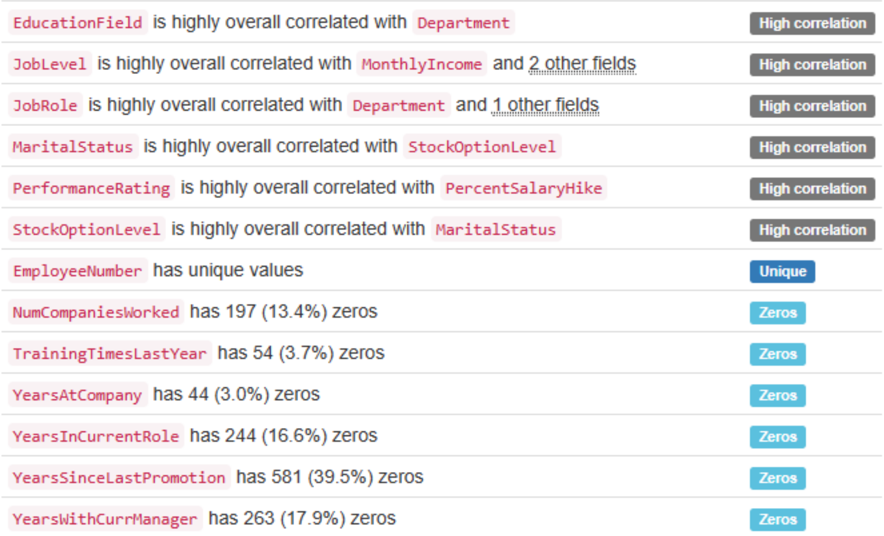
Distinct	2
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.6 KiB



Overview

Alerts

EmployeeCount has constant value "1"	Constant
Over18 has constant value "True"	Constant
StandardHours has constant value "80"	Constant
Age is highly overall correlated with TotalWorkingYears	High correlation
MonthlyIncome is highly overall correlated with TotalWorkingYears and 1 other fields	High correlation
PercentSalaryHike is highly overall correlated with PerformanceRating	High correlation
TotalWorkingYears is highly overall correlated with Age and 3 other fields	High correlation
YearsAtCompany is highly overall correlated with TotalWorkingYears and 3 other fields	High correlation
YearsInCurrentRole is highly overall correlated with YearsAtCompany and 2 other fields	High correlation
YearsSinceLastPromotion is highly overall correlated with YearsAtCompany and 1 other fields	High correlation
YearsWithCurrManager is highly overall correlated with YearsAtCompany and 1 other fields	High correlation
Department is highly overall correlated with EducationField and 1 other fields	High correlation



EducationField is highly overall correlated with Department	High correlation
JobLevel is highly overall correlated with MonthlyIncome and 2 other fields	High correlation
JobRole is highly overall correlated with Department and 1 other fields	High correlation
MaritalStatus is highly overall correlated with StockOptionLevel	High correlation
PerformanceRating is highly overall correlated with PercentSalaryHike	High correlation
StockOptionLevel is highly overall correlated with MaritalStatus	High correlation
EmployeeNumber has unique values	Unique
NumCompaniesWorked has 197 (13.4%) zeros	Zeros
TrainingTimesLastYear has 54 (3.7%) zeros	Zeros
YearsAtCompany has 44 (3.0%) zeros	Zeros
YearsInCurrentRole has 244 (16.6%) zeros	Zeros
YearsSinceLastPromotion has 581 (39.5%) zeros	Zeros
YearsWithCurrManager has 263 (17.9%) zeros	Zeros

3.3 Importing Data :

Introduction

IBM created this dataset to allow analysts to uncover the reason for employee attrition.

We will look into the attrition rate, as the average company's employee attrition rate is below 20%. For any company with a higher than 20% attrition rate, there is a concern for the company's turnover rate. A company that wants to have a "good" attrition rate should be aiming at 10% attrition rate. As employees requires time, energy, and money to train, the lower the attrition rate, the better it is to have a sustainable company culture.

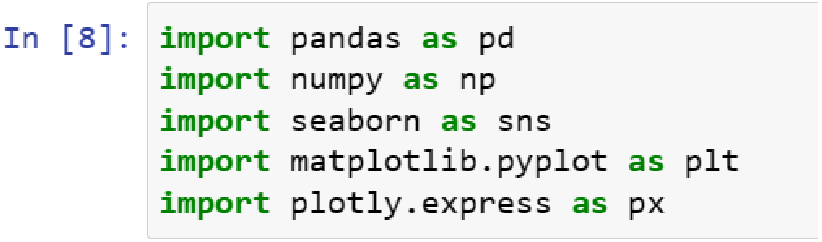
Before diving into the dataset and start analyzing, I have listed out three questions that I will try to find in my analysis and based on the finding to generate recommendations to the management.

Business Question

1. What is the attrition rate? Is it high or low?
2. What are the factors affecting the attrition rate?
3. What recommendation can be provided to the management based on the analytics?

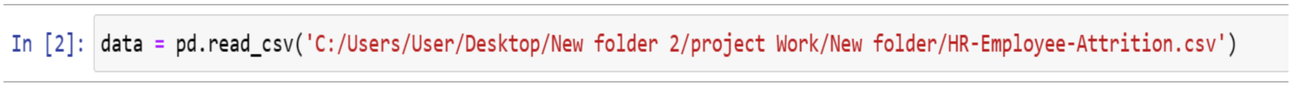
We are using Jupyter Notebook and Python programming for analysis of the above data. First step is to

Import the python libraries



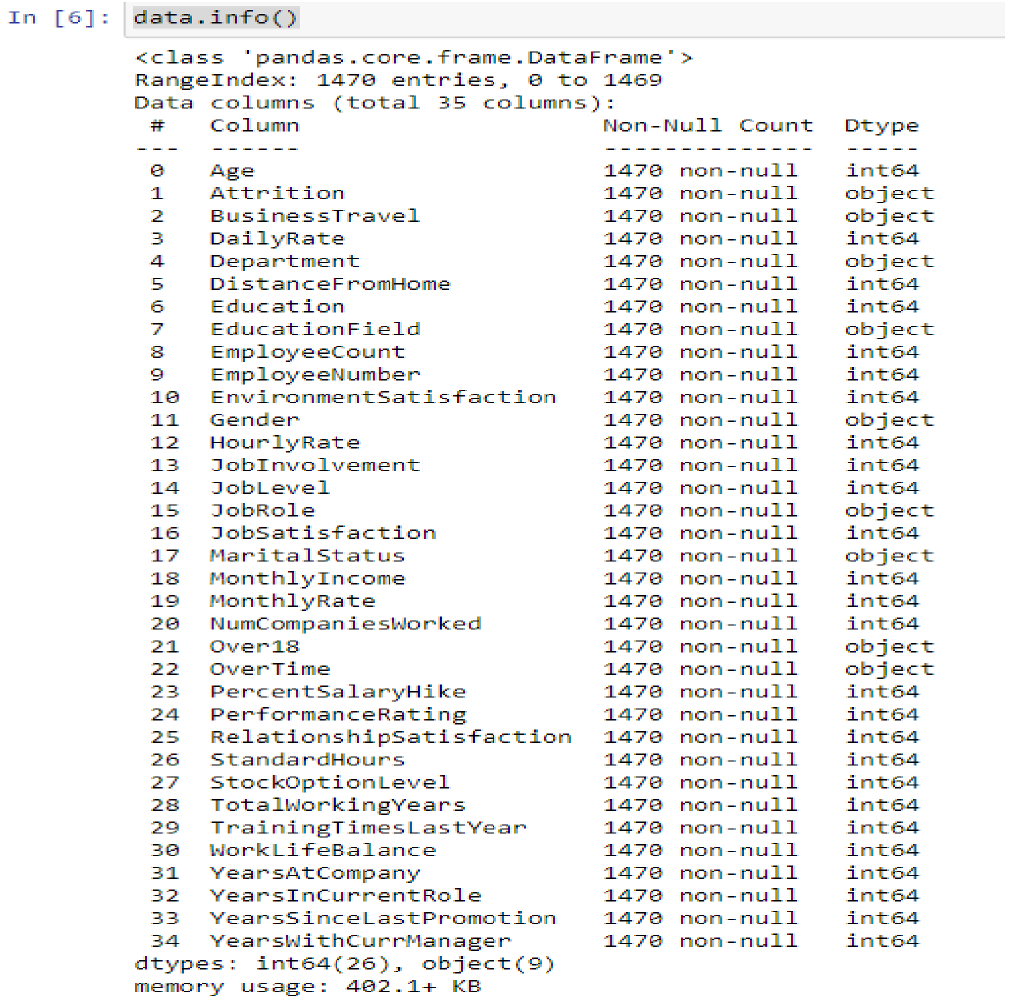
```
In [8]: import pandas as pd
        import numpy as np
        import seaborn as sns
        import matplotlib.pyplot as plt
        import plotly.express as px
```

Next step is to import and read the dataset into Jupyter Notebook.



```
In [2]: data = pd.read_csv('C:/Users/User/Desktop/New folder 2/project Work/New folder/HR-Employee-Attrition.csv')
```

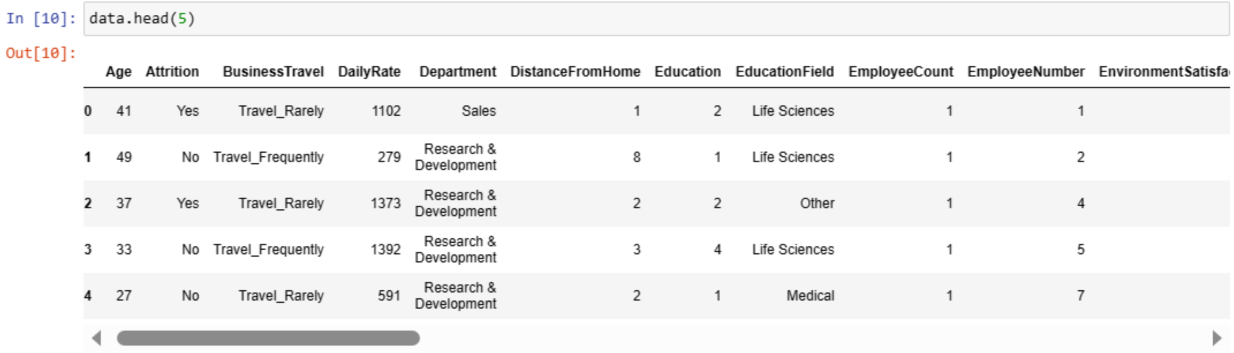
Using the data.info() to check the info of the data.



```
In [6]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1470 entries, 0 to 1469
Data columns (total 35 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Age                       1470 non-null   int64
 1   Attrition                 1470 non-null   object
 2   BusinessTravel            1470 non-null   object
 3   DailyRate                 1470 non-null   int64
 4   Department                1470 non-null   object
 5   DistanceFromHome          1470 non-null   int64
 6   Education                 1470 non-null   int64
 7   EducationField            1470 non-null   object
 8   EmployeeCount             1470 non-null   int64
 9   EmployeeNumber            1470 non-null   int64
 10  EnvironmentSatisfaction   1470 non-null   int64
 11  Gender                    1470 non-null   object
 12  HourlyRate                1470 non-null   int64
 13  JobInvolvement            1470 non-null   int64
 14  JobLevel                  1470 non-null   int64
 15  JobRole                   1470 non-null   object
 16  JobSatisfaction           1470 non-null   int64
 17  MaritalStatus             1470 non-null   object
 18  MonthlyIncome             1470 non-null   int64
 19  MonthlyRate               1470 non-null   int64
 20  NumCompaniesWorked        1470 non-null   int64
 21  Over18                    1470 non-null   object
 22  OverTime                  1470 non-null   object
 23  PercentSalaryHike         1470 non-null   int64
 24  PerformanceRating         1470 non-null   int64
 25  RelationshipSatisfaction  1470 non-null   int64
 26  StandardHours             1470 non-null   int64
 27  StockOptionLevel          1470 non-null   int64
 28  TotalWorkingYears         1470 non-null   int64
 29  TrainingTimesLastYear     1470 non-null   int64
 30  WorkLifeBalance           1470 non-null   int64
 31  YearsAtCompany            1470 non-null   int64
 32  YearsInCurrentRole        1470 non-null   int64
 33  YearsSinceLastPromotion   1470 non-null   int64
 34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
memory usage: 402.1+ KB
```

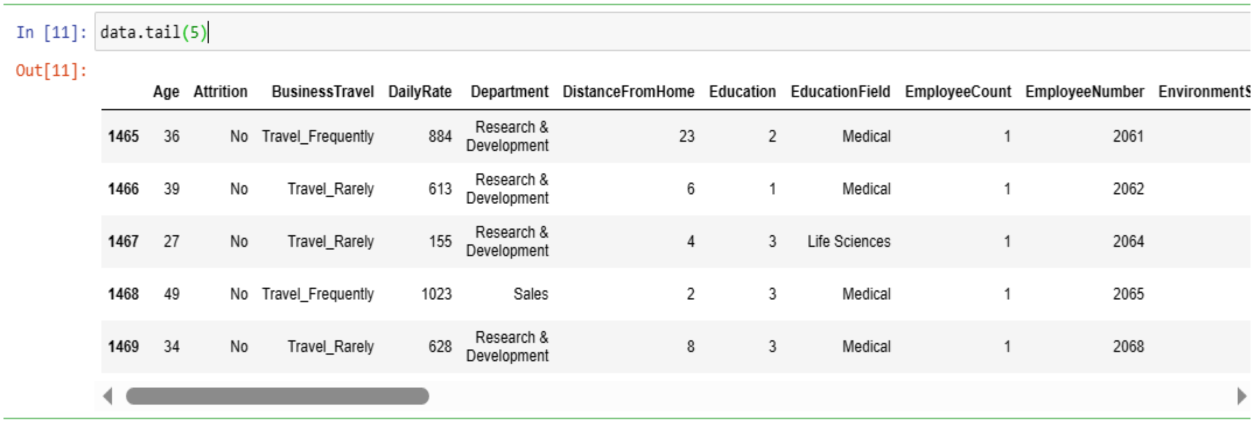
Other useful method is info that shows a summary of the dataset, like number of observations, columns, variable type and the total memory usage.
The dataset have 1,470 entire , 35 columns and with no null values.
The data types of the variables are divided in 2 6 integer and 9 object.

Using the data.head() , and data.tail(5) command we can check the first 5 rows and last 5 rows respectively of the dataset.



```
In [10]: data.head(5)
Out[10]:
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfa
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	



```
In [11]: data.tail(5)
```

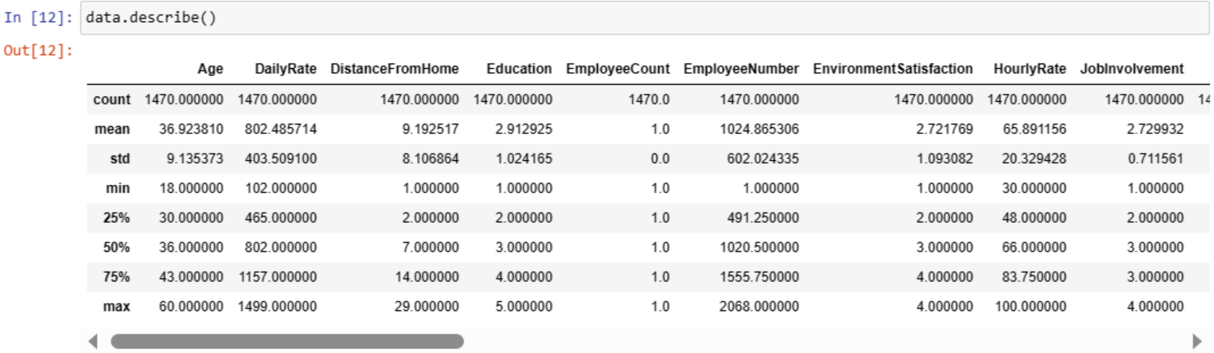
Out[11]:

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentS
1465	36	No	Travel_Frequently	884	Research & Development	23	2	Medical	1	2061	
1466	39	No	Travel_Rarely	613	Research & Development	6	1	Medical	1	2062	
1467	27	No	Travel_Rarely	155	Research & Development	4	3	Life Sciences	1	2064	
1468	49	No	Travel_Frequently	1023	Sales	2	3	Medical	1	2065	
1469	34	No	Travel_Rarely	628	Research & Development	8	3	Medical	1	2068	

3.4 Descriptive Analytics:

The descriptive Analytics is used to simplify and summarize the mainly characteristics of the dataset.

The Pandas method describe generates a descriptive statistics that summarize the central tendency, dispersion and shape of the dataset.



```
In [12]: data.describe()
```

Out[12]:

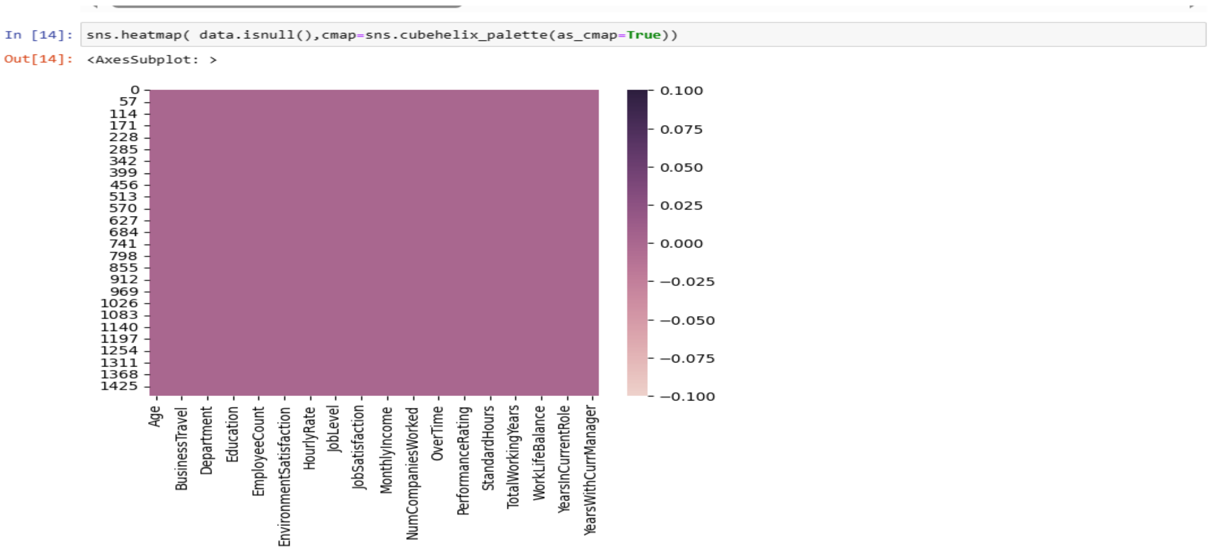
	Age	DailyRate	DistanceFromHome	Education	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	HourlyRate	JobInvolvement	
count	1470.000000	1470.000000	1470.000000	1470.000000	1470.0	1470.000000	1470.000000	1470.000000	1470.000000	14
mean	36.923810	802.485714	9.192517	2.912925	1.0	1024.865306	2.721769	65.891156	2.729932	
std	9.135373	403.509100	8.106864	1.024165	0.0	602.024335	1.093082	20.329428	0.711561	
min	18.000000	102.000000	1.000000	1.000000	1.0	1.000000	1.000000	30.000000	1.000000	
25%	30.000000	465.000000	2.000000	2.000000	1.0	491.250000	2.000000	48.000000	2.000000	
50%	36.000000	802.000000	7.000000	3.000000	1.0	1020.500000	3.000000	66.000000	3.000000	
75%	43.000000	1157.000000	14.000000	4.000000	1.0	1555.750000	4.000000	83.750000	3.000000	
max	60.000000	1499.000000	29.000000	5.000000	1.0	2068.000000	4.000000	100.000000	4.000000	

3.5 Data Cleaning

Data cleaning is the process of eliminating unneeded variables and values from your dataset as well as removing any anomalies. Such abnormalities might distort the data unduly, affecting the outcomes. The following procedures may be taken to clean data:

a. Remove missing values, outliers, and extraneous rows/columns.
b. We are re-indexing and reformatting our data.

It's now time to tidy up the dataset. You must first determine the number of missing values in each column as well as the proportion of missing values to which they contribute.



```
In [14]: sns.heatmap( data.isnull(),cmap=sns.cubehelix_palette(as_cmap=True))
Out[14]: <AxesSubplot: >
```

The heatmap visualisation of missing (null) values in the supplied pandas Data Frame 'data' is shown above.

The heatmap is created using the Seaborn library's sns.heatmap () method. The first input to this method is a Boolean DataFrame constructed by data.isnull() that includes True values wherever there are null values in the original DataFrame 'data'.

The Heatmap shows the correlation between different features. The Key features that are highly correlated with attrition, ordered from highest positive correlation to highest negative correlation:

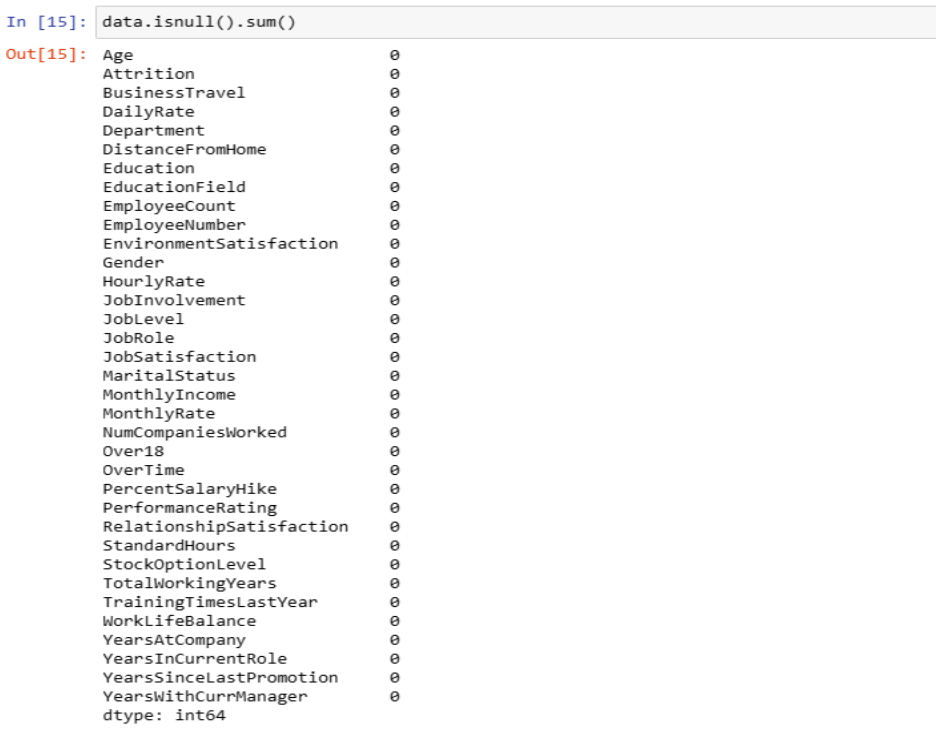
Features that might influence people to leave the company include:

- Job Level: Employees with lower job levels
- Monthly Income: Employees with lower monthly income
- Age: Younger employees
- Total Working Years: Employees with fewer years of experience
- Job Involvement: Employees with lower job involvement
- Stock Option Level: Employees with a lower stock option level
- Distance From Home: Employees who live further away from work

Features that might influence people to stay with the company include:

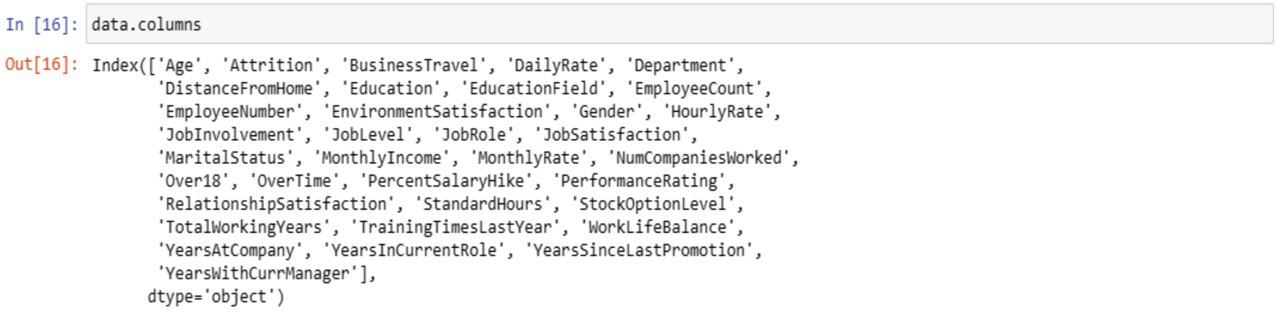
- Job Satisfaction: Employees with higher job satisfaction
- Environment Satisfaction: Employees with higher environment satisfaction
- Work Life Balance: Employees with a better work life balance
- Job Involvement: Employees with higher job involvement
- Job Role: Certain job roles, such as Manager and Research Director

Note: It's important to note that correlation does not necessarily imply causation, and further analysis would be needed to establish the causal relationships between these features and attrition. Nonetheless, the correlation heatmap provides a useful starting point for identifying potentially important features for predicting attrition.



```
In [15]: data.isnull().sum()

Out[15]: Age                         0
         Attrition                   0
         BusinessTravel              0
         DailyRate                   0
         Department                  0
         DistanceFromHome            0
         Education                   0
         EducationField              0
         EmployeeCount               0
         EmployeeNumber              0
         EnvironmentSatisfaction     0
         Gender                      0
         HourlyRate                  0
         JobInvolvement              0
         JobLevel                    0
         JobRole                     0
         JobSatisfaction             0
         MaritalStatus               0
         MonthlyIncome               0
         MonthlyRate                 0
         NumCompaniesWorked          0
         Over18                      0
         OverTime                    0
         PercentSalaryHike           0
         PerformanceRating           0
         RelationshipSatisfaction    0
         StandardHours               0
         StockOptionLevel            0
         TotalWorkingYears           0
         TrainingTimesLastYear       0
         WorkLifeBalance             0
         YearsAtCompany              0
         YearsInCurrentRole          0
         YearsSinceLastPromotion     0
         YearsWithCurrManager        0
         dtype: int64
```

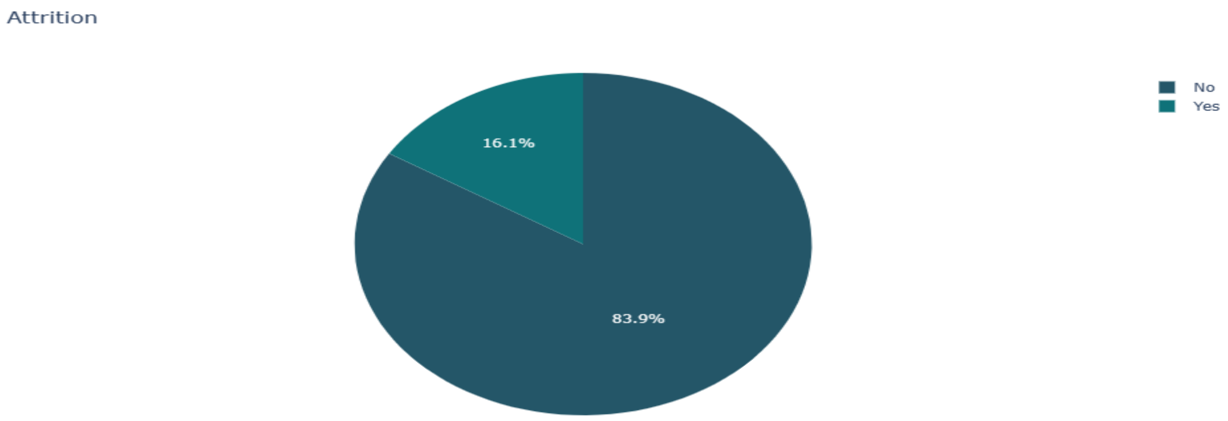
As you can see there is no NUL Value



```
In [16]: data.columns

Out[16]: Index(['Age', 'Attrition', 'BusinessTravel', 'DailyRate', 'Department',
                'DistanceFromHome', 'Education', 'EducationField', 'EmployeeCount',
                'EmployeeNumber', 'EnvironmentSatisfaction', 'Gender', 'HourlyRate',
                'JobInvolvement', 'JobLevel', 'JobRole', 'JobSatisfaction',
                'MaritalStatus', 'MonthlyIncome', 'MonthlyRate', 'NumCompaniesWorked',
                'Over18', 'OverTime', 'PercentSalaryHike', 'PerformanceRating',
                'RelationshipSatisfaction', 'StandardHours', 'StockOptionLevel',
                'TotalWorkingYears', 'TrainingTimesLastYear', 'WorkLifeBalance',
                'YearsAtCompany', 'YearsInCurrentRole', 'YearsSinceLastPromotion',
                'YearsWithCurrManager'],
               dtype='object')
```

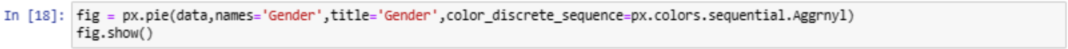
Above are the columns available in the Data set.

3.6 Data visualization

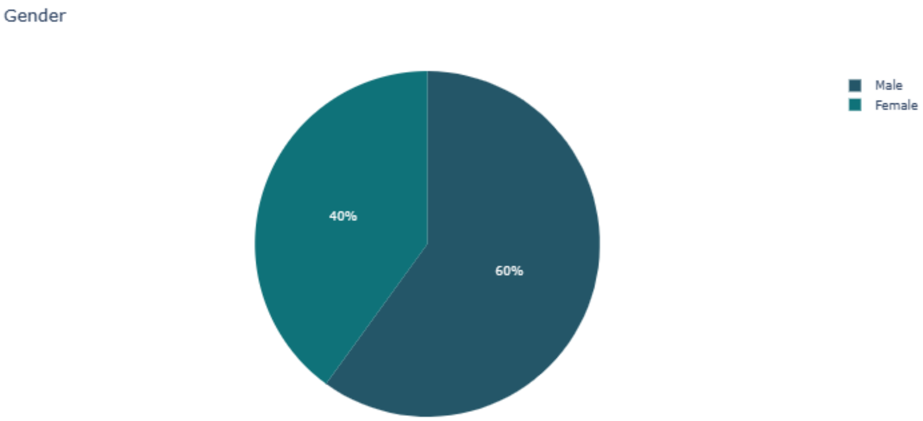


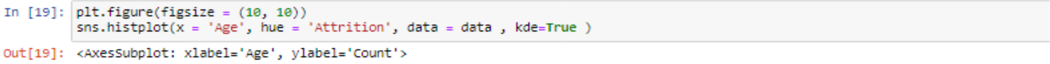
We may learn from the Pie conversation that just 16.1% of the staff had departed the firm.

However, when we look at the percentage of males in the firm, we see that it is larger than the percentage of females.

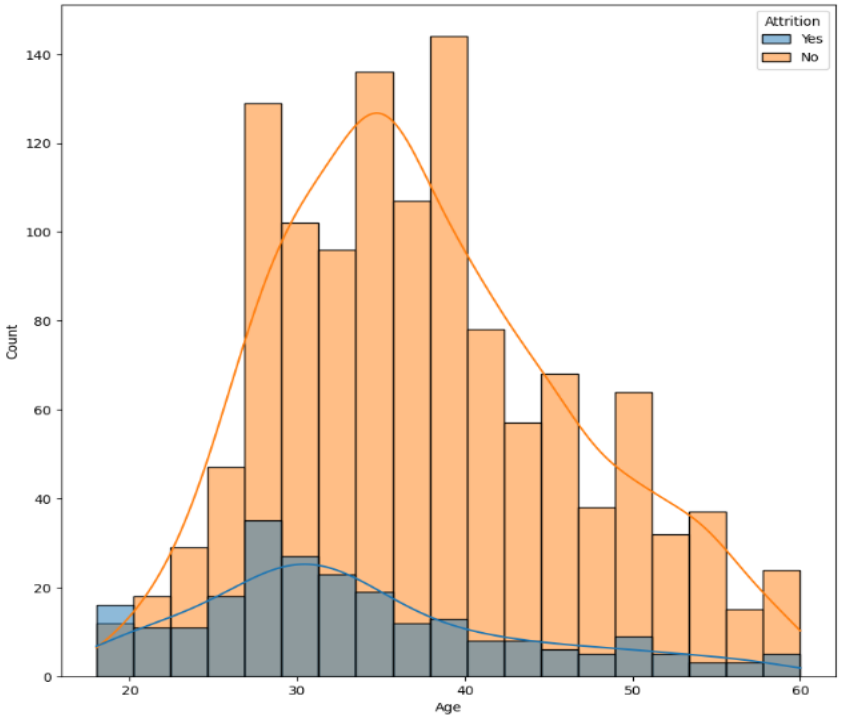


```
In [18]: fig = px.pie(data,names='Gender',title='Gender',color_discrete_sequence=px.colors.sequential.Aggrnyl)
         fig.show()
```





```
In [19]: plt.figure(figsize = (10, 10))
         sns.histplot(x = 'Age', hue = 'Attrition', data = data , kde=True )

Out[19]: <AxesSubplot: xlabel='Age', ylabel='Count'>
```

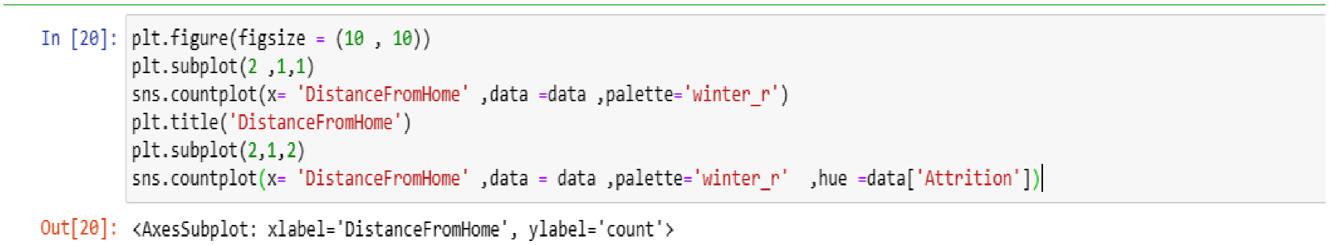


According to the above histogram plot, the bulk of employees are between the ages of 28 and 36, while the age range is 18 to 60. The Company appears to be heavily reliant on individuals under the age of 40.

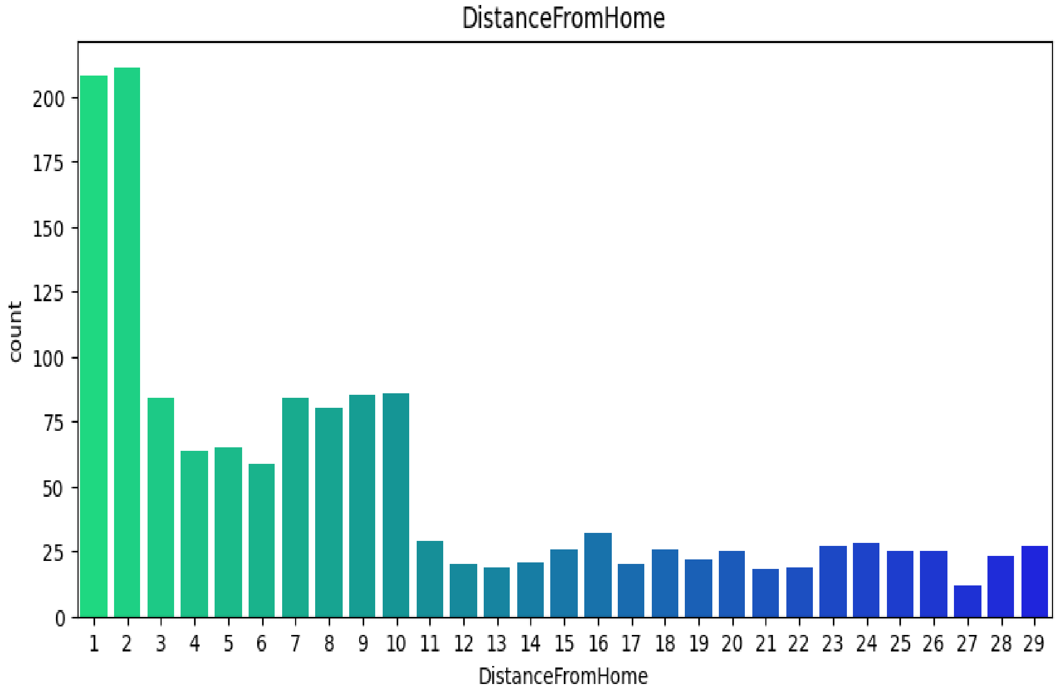
In the first subplot, a countplot is created using sns.countplot () with 'DistanceFromHome' on the x-axis and the count of observations on the y-axis. The 'palette' parameter is used to set the color scheme for the countplot to 'winter_r', which is a reversed version of the default Seaborn 'winter' color palette. The title of the subplot is set to 'DistanceFromHome'.

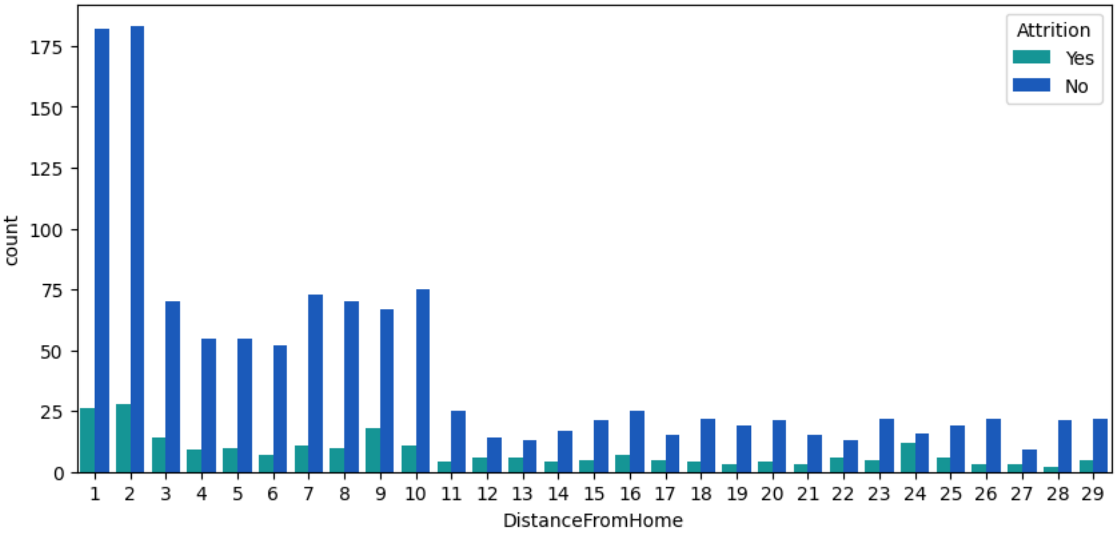
In the second subplot, another countplot is created using sns.countplot (), but this time with the addition of the 'hue' parameter. The 'hue' parameter is set to 'Attrition', which means that the countplot will show the distribution of the 'DistanceFromHome' column for each value of the 'Attrition' column. This creates two separate bars for each 'DistanceFromHome' value, one for employees who have 'Attrition' and another for those who do not have 'Attrition'

By creating two countplots side by side, we can quickly compare the distribution of the 'DistanceFromHome' column across all employees in the first subplot and the distribution of the same column for employees with and without 'Attrition' in the second subplot. This can help us identify any patterns or trends in the data that may be related to attrition.



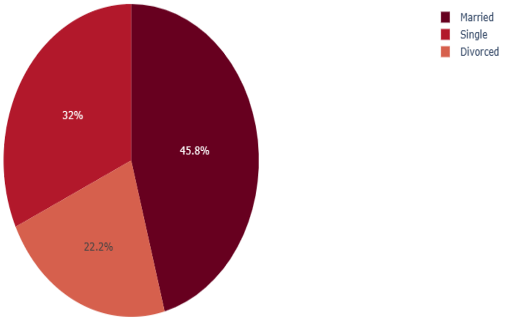
```
In [20]: plt.figure(figsize = (10 , 10))
         plt.subplot(2 ,1,1)
         sns.countplot(x= 'DistanceFromHome' ,data =data ,palette='winter_r')
         plt.title('DistanceFromHome')
         plt.subplot(2,1,2)
         sns.countplot(x= 'DistanceFromHome' ,data = data ,palette='winter_r'  ,hue =data['Attrition'])

Out[20]: <AxesSubplot: xlabel='DistanceFromHome', ylabel='count'>
```

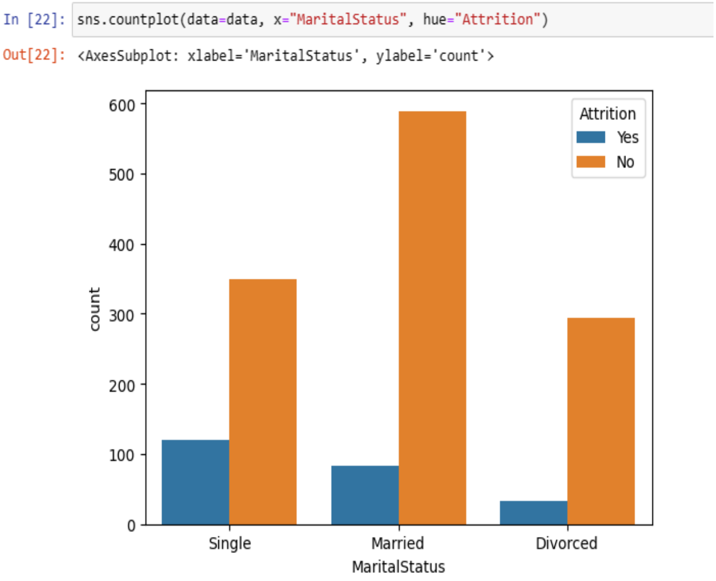




The sns.countplot () function from the Seaborn library is used to create the plot. The 'x' parameter is set to 'Marital Status' to specify that we want to count the number of observations for each value in the 'Marital Status' column. The 'hue' parameter is set to 'Attrition' to indicate that we want to separate the counts by the value in the 'Attrition' column, which can have two possible values ('Yes' and 'No').

The resulting countplot will have one bar for each unique value in the 'Marital Status' column, and each bar will be split into two colors corresponding to the 'Attrition' column. This allows us to compare the proportion of employees who have and have not experienced attrition within each category of marital status.



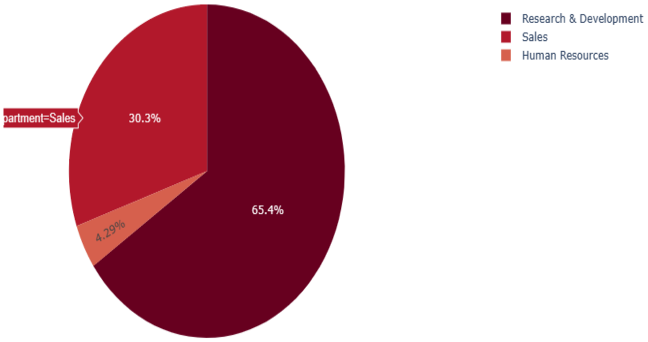


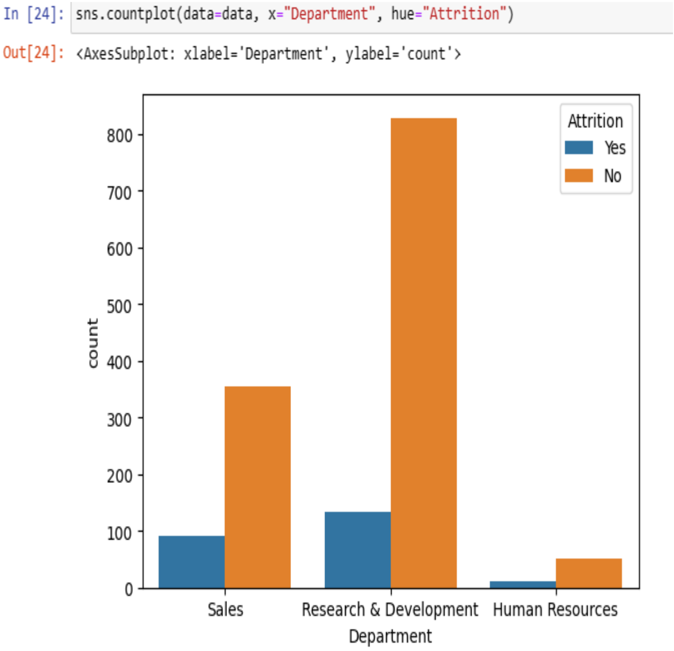
```
In [22]: sns.countplot(data=data, x="MaritalStatus", hue="Attrition")
Out[22]: <AxesSubplot: xlabel='MaritalStatus', ylabel='count'>
```

The sns.countplot () function from the Seaborn library is used to create the plot. The 'x' parameter is set to 'Department' to specify that we want to count the number of observations for each value in the 'Department' column. The 'hue' parameter is set to 'Attrition' to indicate that we want to separate the counts by the value in the 'Attrition' column, which can have two possible values ('Yes' and 'No').

The resulting countplot will have one bar for each unique value in the 'Department' column, and each bar will be split into two colors corresponding to the 'Attrition' column. This allows us to compare the proportion of employees who have and have not experienced attrition within each department.

We can observe that Research and Development Department followed by sales are more likely to depart than other positions



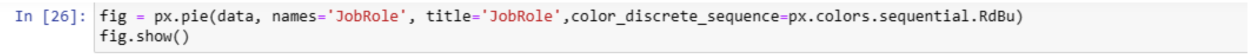


```
In [24]: sns.countplot(data=data, x="Department", hue="Attrition")

Out[24]: <AxesSubplot: xlabel='Department', ylabel='count'>
```

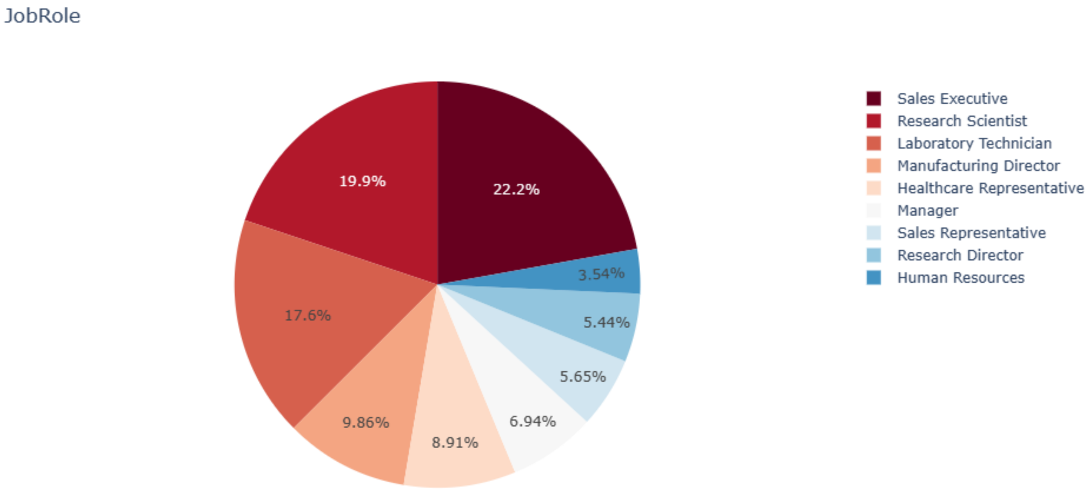
The sns.countplot () function from the Seaborn library is used to create the plot. The 'y' parameter is set to 'JobRole' to specify that we want to count the number of observations for each value in the 'JobRole' column and display the bars vertically. The 'hue' parameter is set to 'Attrition' to indicate that we want to separate the counts by the value in the 'Attrition' column, which can have two possible values ('Yes' and 'No').

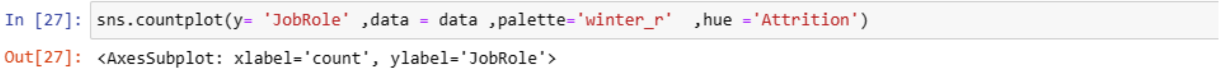
The resulting countplot will have one bar for each unique value in the 'JobRole' column, and each bar will be split into two colors corresponding to the 'Attrition' column. This allows us to compare the proportion of employees who have and have not experienced attrition within each job role.

We can observe that’ sales executive’,’ sales representative', and 'lab technician' are more likely to depart than other positions.

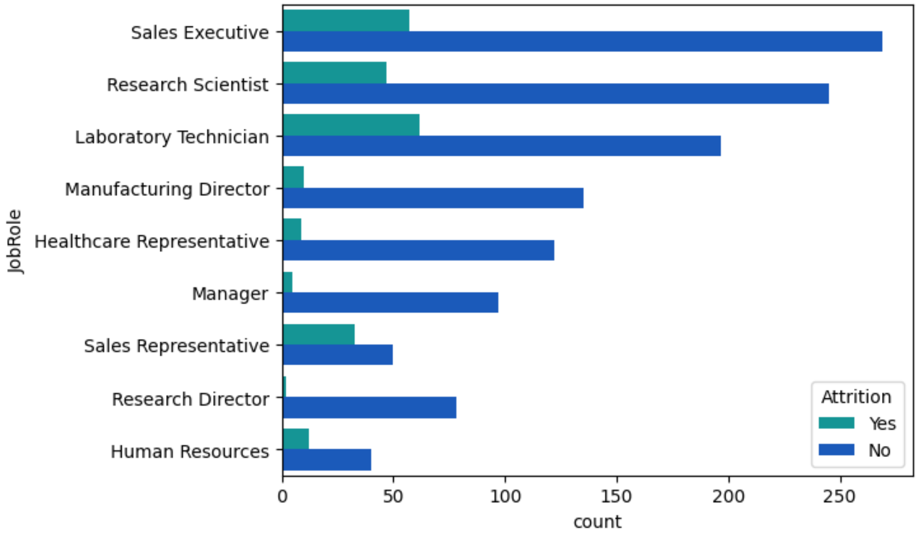


```
In [26]: fig = px.pie(data, names='JobRole', title='JobRole',color_discrete_sequence=px.colors.sequential.RdBu)
         fig.show()
```



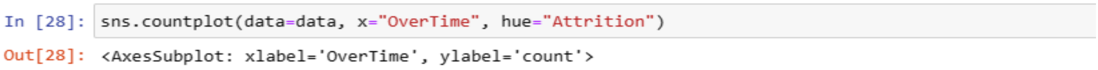


```
In [27]: sns.countplot(y= 'JobRole' ,data = data ,palette='winter_r'  ,hue ='Attrition')
Out[27]: <AxesSubplot: xlabel='count', ylabel='JobRole'>
```

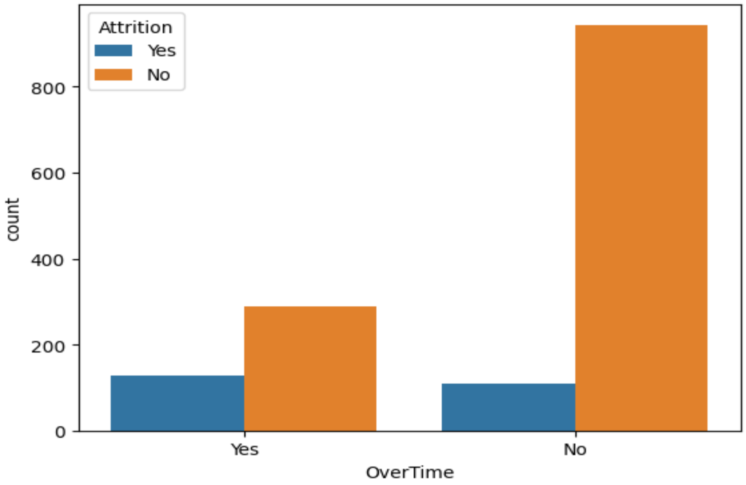


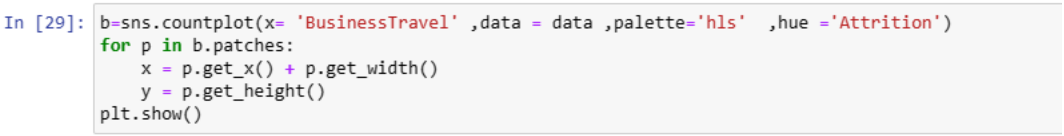
This code creates a countplot that shows the distribution of the 'Business Travel' column in the given pandas DataFrame 'data', broken down by the 'Attrition' column. The sns.countplot () function from the Seaborn library is used to create the plot. The 'x' parameter is set to 'Business Travel' to specify that we want to count the number of observations for each value in the 'Business Travel' column. The 'hue' parameter is set to 'Attrition' to indicate that we want to separate the counts by the value in the 'Attrition' column, which can have two possible values ('Yes' and 'No').

The resulting countplot will have one bar for each unique value in the 'Business Travel' column, and each bar will be split into two colors corresponding to the 'Attrition' column. This allows us to compare the proportion of employees who have and have not experienced attrition within each category of business travel.

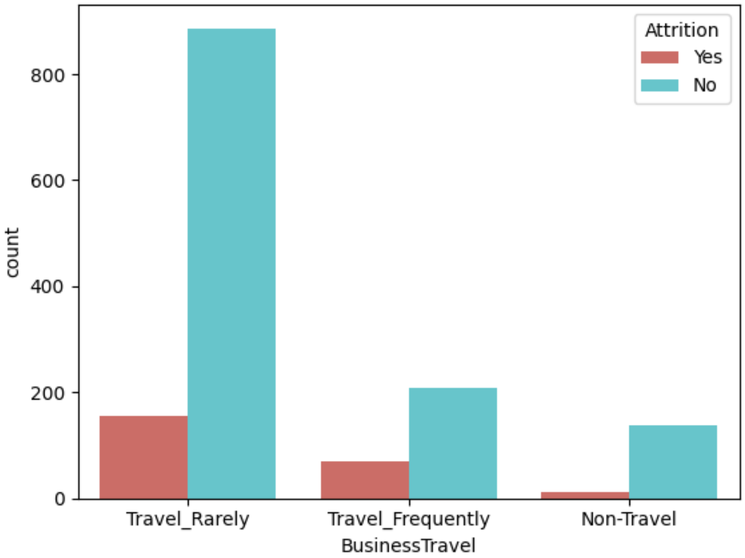


```
In [28]: sns.countplot(data=data, x="OverTime", hue="Attrition")
Out[28]: <AxesSubplot: xlabel='OverTime', ylabel='count'>
```





```
In [29]: b=sns.countplot(x= 'BusinessTravel' ,data = data ,palette='hls'  ,hue ='Attrition')
         for p in b.patches:
             x = p.get_x() + p.get_width()
             y = p.get_height()
         plt.show()
```



This code generates a Kernel Density Estimate (KDE) plot using the Seaborn library, comparing the distribution of the 'YearsWithCurrManager' column for employees who left the company ('Yes' in the ‘Attrition' column) and those who stayed ('No' in the 'Attrition' column) in the given pandas DataFrame 'data'.

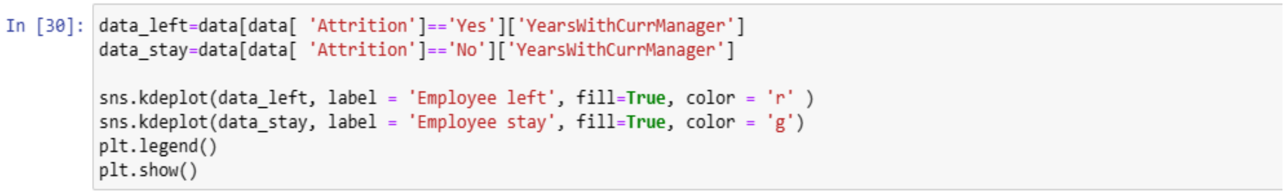
First, the code generates two new pandas Series: 'data left,' which contains just the values from the 'YearsWithCurrManager' column for workers who leave the firm, and 'data stay,' which contains only the values from the 'YearsWithCurrManager' column for employees who stayed.

The plot is then generated using the sns.kdeplot () method. This function generates a KDE plot, which is a method for estimating the probability density function of a random variable. In this situation, the random variable is the 'YearsWithCurrManager' column.

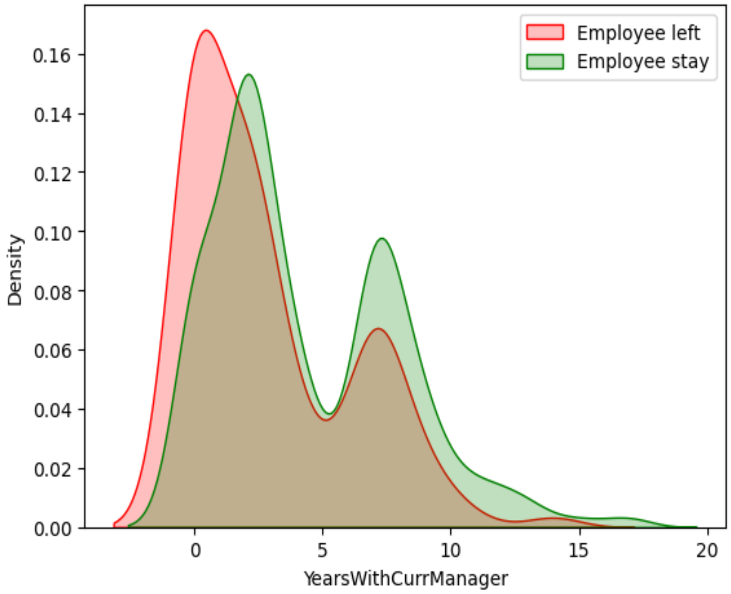
To provide the values for the KDE plot, the 'data_left' and 'data_stay' Series are used as inputs to the function. To define the labels for the two KDE curves, the 'label' argument is set to 'Employee left' and 'Employee remain', respectively. To fill the region under the curves, the 'fill' option is set to 'True'. To specify the colours for the two curves, the 'colour' parameter is set to 'r' and 'g'.

Finally, plt.legend () and plt.show () are used to add a legend to the plot and display it.

We may examine the distributions of the 'YearsWithCurrManager' column for workers who departed and those who stayed using this graphic. It can assist us in identifying any patterns or discrepancies between the two groups.



```
In [30]: data_left=data[data[ 'Attrition']=='Yes']['YearsWithCurrManager']
         data_stay=data[data[ 'Attrition']=='No']['YearsWithCurrManager']

         sns.kdeplot(data_left, label = 'Employee left', fill=True, color = 'r' )
         sns.kdeplot(data_stay, label = 'Employee stay', fill=True, color = 'g')
         plt.legend()
         plt.show()
```



The education field Life Sciences had the highest percentage of attrition, followed by Medical and Marketing.

The numerical values in several columns of the pandas DataFrame 'data' with their corresponding string labels to make the data more human-readable and interpretable. Here is what each line of cod does:

1.data['Education'].replace([1,2,3,4,5],["BelowCollege","College","Bachelor","Master","Doctor"],inplace=True) replaces the values 1, 2, 3, 4, and 5 in the 'Education' column with the corresponding string labels "Below College", "College", "Bachelor", "Master", and "Doctor", respectively. The inplace=True parameter ensures that the changes are made to the original DataFrame.

2. data['EnvironmentSatisfaction'].replace([1,2,3,4],["Low","Medium","High","Very High"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'EnvironmentSatisfaction' column with the corresponding string labels "Low", "Medium", "High", and "Very High", respectively.

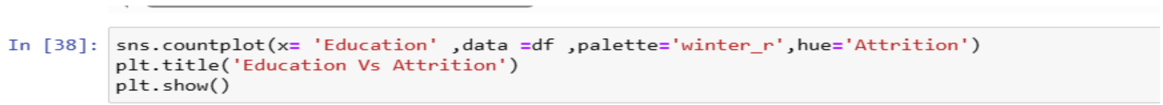
3. data['JobInvolvement'].replace([1,2,3,4],["Low","Medium","High","Very High"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'JobInvolvement' column with the corresponding string labels "Low", "Medium", "High", and "Very High", respectively.

4. data['JobSatisfaction'].replace([1,2,3,4],["Low","Medium","High","Very High"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'JobSatisfaction' column with the corresponding string labels "Low", "Medium", "High", and "Very High", respectively.

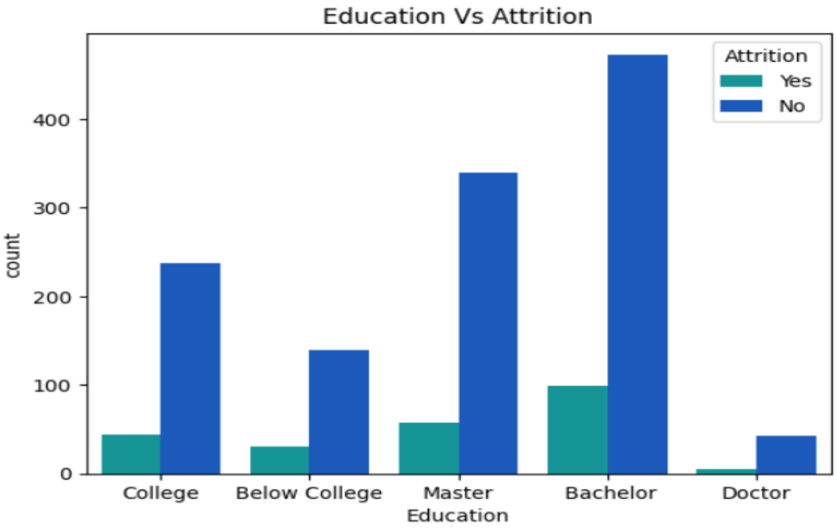
5.data['PerformanceRating'].replace([1,2,3,4],["Low","Good","Excellent","Outstanding"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'PerformanceRating' column with the corresponding string labels "Low", "Good", "Excellent", and "Outstanding", respectively.

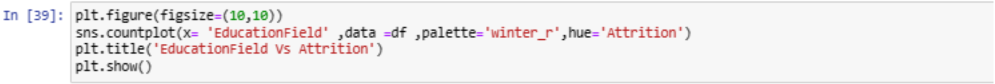
6. data['RelationshipSatisfaction'].replace([1,2,3,4],["Low","Medium","High","Very High"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'RelationshipSatisfaction' column with the corresponding string labels "Low", "Medium", "High", and "Very High", respectively.

7. data['WorkLifeBalance'].replace([1,2,3,4],["Bad","Good","Better","Best"],inplace=True) replaces the values 1, 2, 3, and 4 in the 'WorkLifeBalance' column with the corresponding string labels "Bad", "Good", "Better", and "Best", respectively.

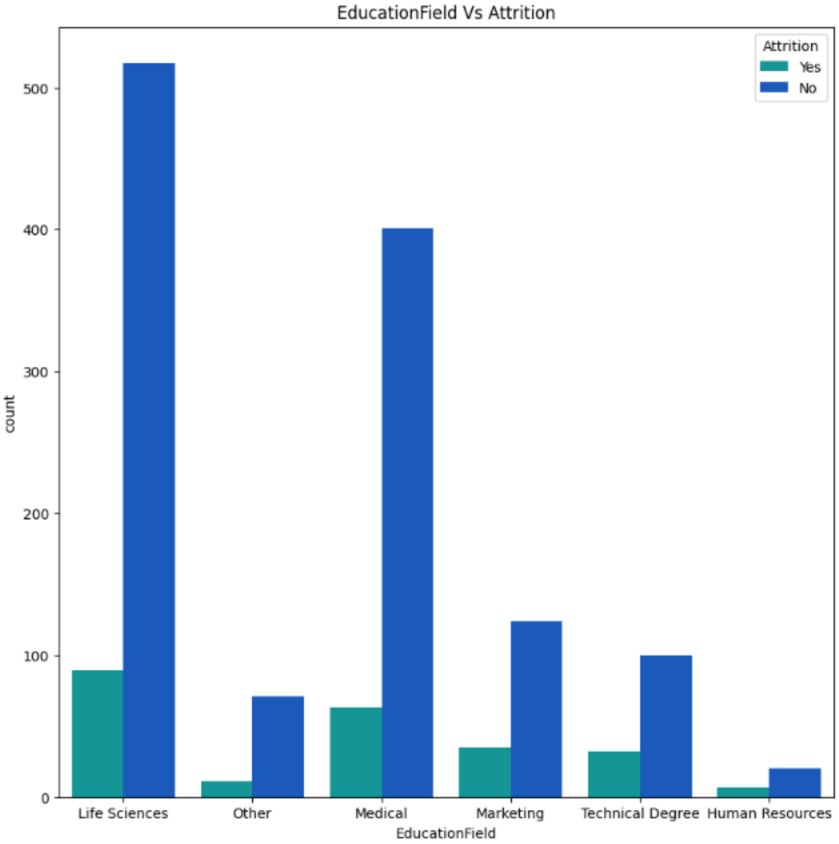


```
In [38]: sns.countplot(x= 'Education' ,data =df ,palette='winter_r',hue='Attrition')
         plt.title('Education Vs Attrition')
         plt.show()
```

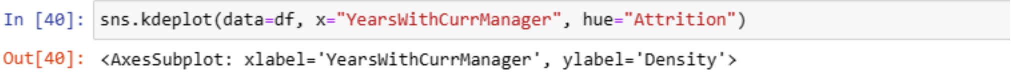




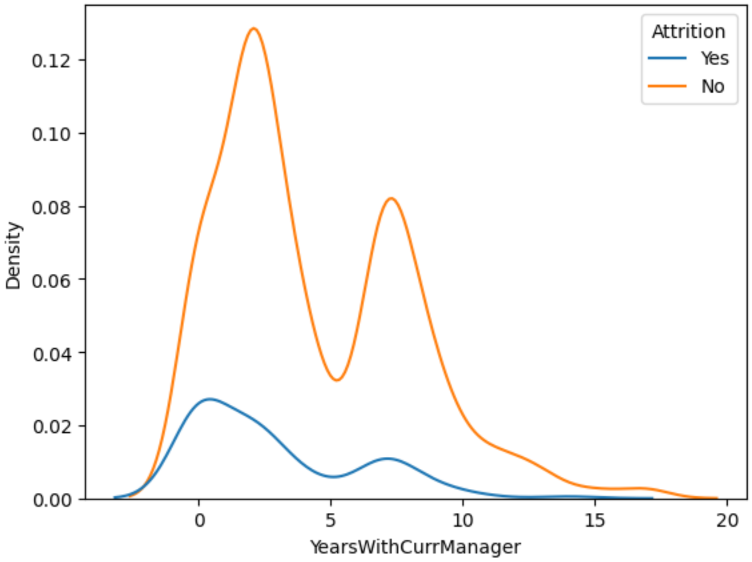
```
In [39]: plt.figure(figsize=(10,10))
         sns.countplot(x= 'EducationField' ,data =df ,palette='winter_r',hue='Attrition')
         plt.title('EducationField Vs Attrition')
         plt.show()
```



we can see more employees tend to leave with less than 2 years with the current managers

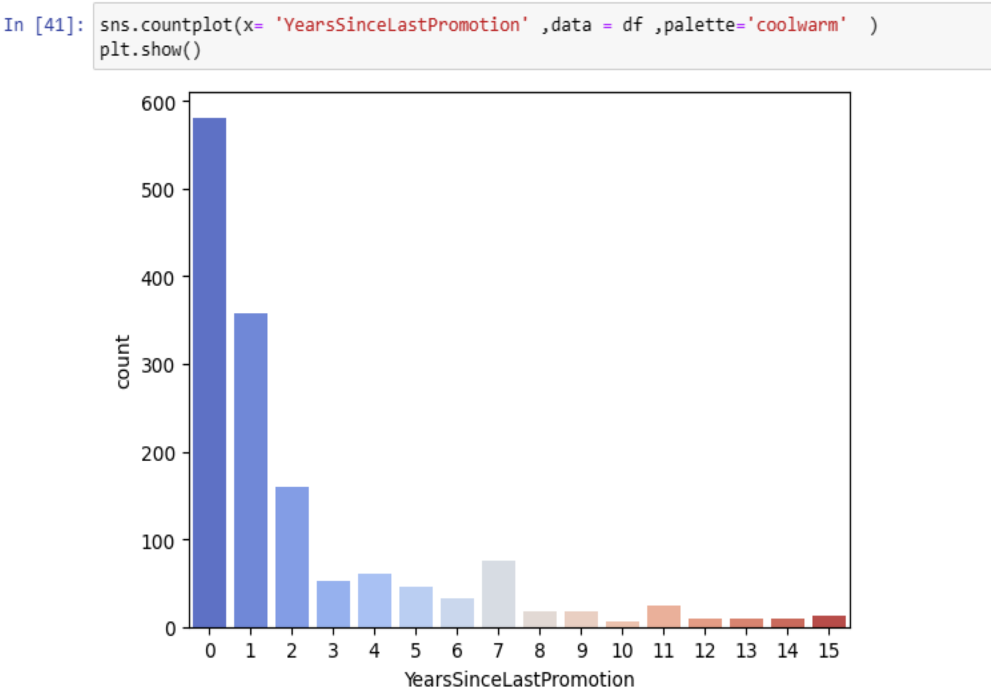


```
In [40]: sns.kdeplot(data=df, x="YearsWithCurrManager", hue="Attrition")
Out[40]: <AxesSubplot: xlabel='YearsWithCurrManager', ylabel='Density'>
```

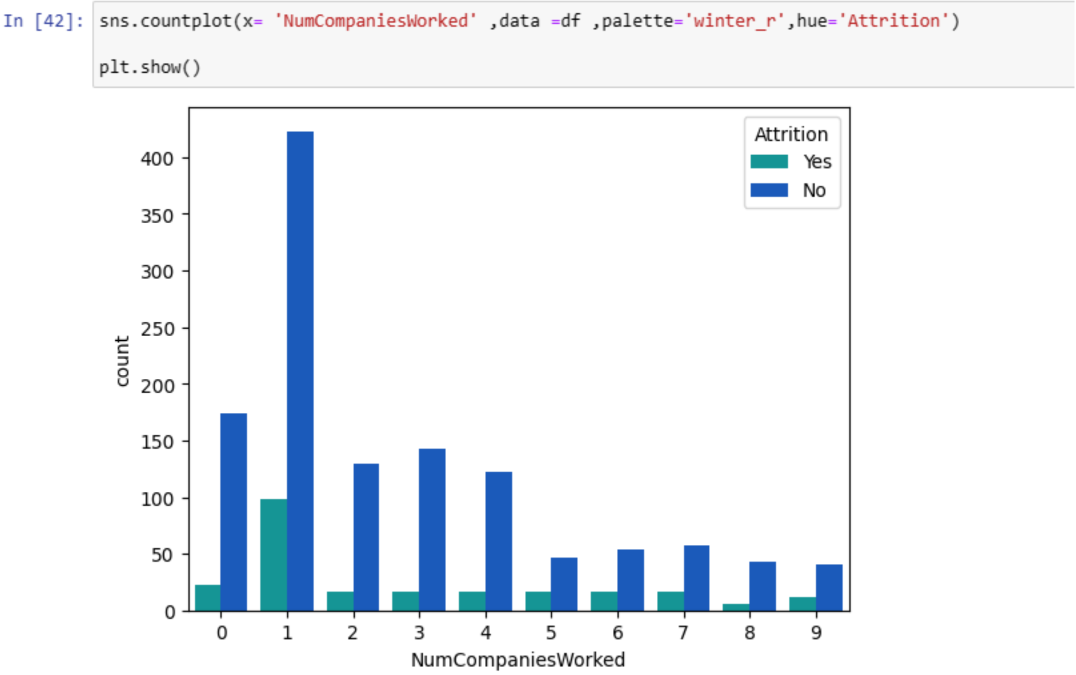


A count plot of the column 'NumCompaniesWorked' in the pandas DataFrame 'df' supplied. The plot is coloured based on the value of the 'Attrition' column, which has two options: 'Yes' or 'No'. The count plot is a form of bar plot that uses bars to display the number of observations in each category. This plot allows us to examine the frequency of 'NumCompaniesWorked' for both employee groups ('Attrition' = 'Yes' and 'Attrition' = 'No') and discover any variations. The 'palette' argument specifies the plot's colour palette. It is set to 'winter_r' in this scenario. The 'hue' argument introduces a new category variable into the display. It is set to 'Attrition' in this example, allowing us to view the number of observations for both groups.

Employees with 1 year experience has the highest percentage to leave



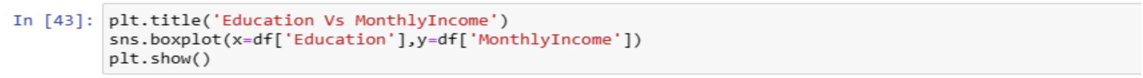
```
In [41]: sns.countplot(x= 'YearsSinceLastPromotion' ,data = df ,palette='coolwarm'  )
         plt.show()
```



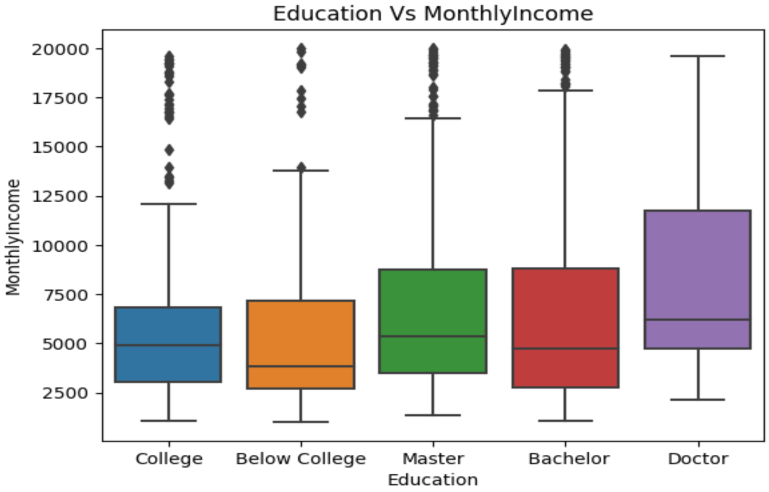
```
In [42]: sns.countplot(x= 'NumCompaniesWorked' ,data =df ,palette='winter_r',hue='Attrition')

         plt.show()
```

This code generates a box plot of the 'MonthlyIncome' column, categorised by the 'Education' column, in the provided pandas DataFrame 'df'. The box plot is a graphical depiction of the data distribution based on five summary statistics (minimum, first quartile, median, third quartile, and maximum). It lets us to see the data's range and distribution, as well as detect any outliers. The 'plt.title()' method changes the plot's title to 'Education vs MonthlyIncome'. The box plot is created using the'sns.boxplot()' method. The 'x' option defines the data grouping column, which is 'Education' in this example. The 'y' option defines the box plot column, which is 'MonthlyIncome' in this example. The command 'plt.show()'

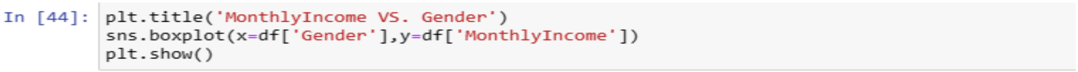
As the level of education increases , average monthly income increases



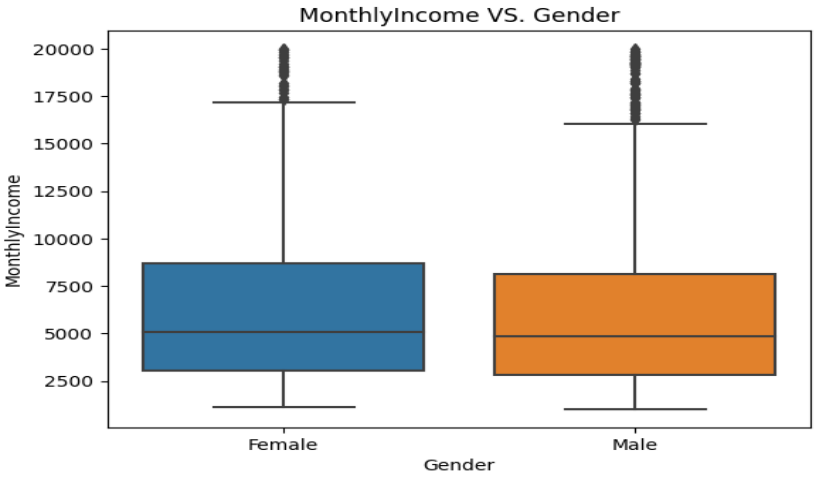
```
In [43]: plt.title('Education Vs MonthlyIncome')
         sns.boxplot(x=df['Education'],y=df['MonthlyIncome'])
         plt.show()
```



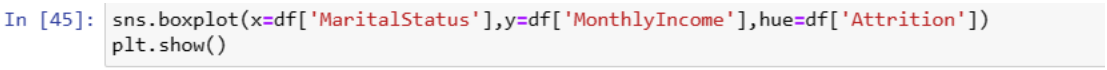
This code generates a box plot of the 'MonthlyIncome' column, grouped by the 'Gender' column, in the provided pandas DataFrame 'df'. The box plot is a graphical depiction of the data distribution based on five summary statistics (minimum, first quartile, median, third quartile, and maximum). It lets us to see the data's range and distribution, as well as detect any outliers. The 'plt.title()' method changes the plot's title to 'MonthlyIncome VS. Gender'. The box plot is created using the'sns.boxplot()' method. The 'x' option defines the data grouping column, which is 'Gender' in this example. The 'y' option defines the box plot column, which is 'MonthlyIncome' in this example. The command 'plt.show()'



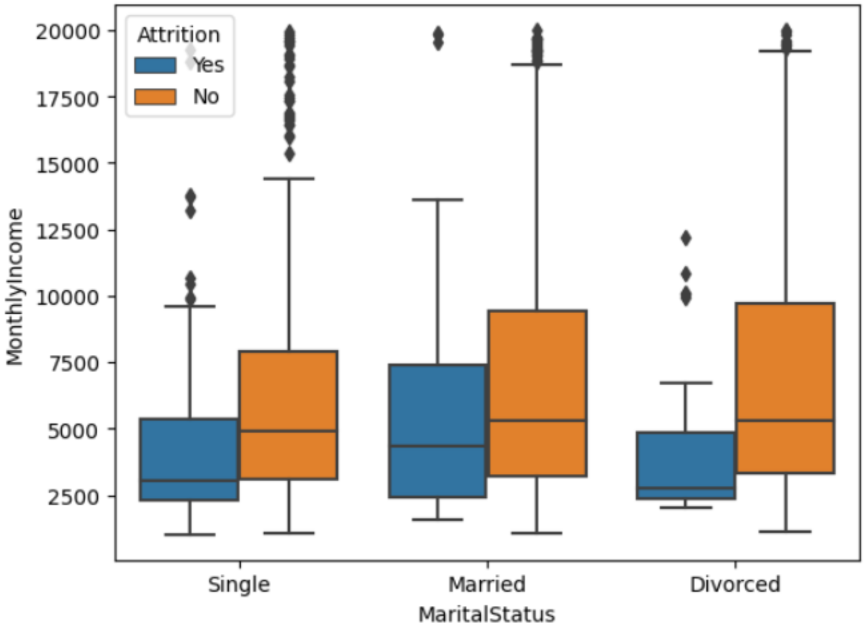
```
In [44]: plt.title('MonthlyIncome VS. Gender')
         sns.boxplot(x=df['Gender'],y=df['MonthlyIncome'])
         plt.show()
```



This code creates a box plot of the 'MonthlyIncome' column in the given pandas DataFrame 'df', grouped by the 'MaritalStatus' column and colored by the 'Attrition' column. The box plot is a graphical representation of the distribution of the data based on five summary statistics (minimum, first quartile, median, third quartile, and maximum). It allows us to see the range and distribution of the data, as well as identify any outliers. The 'sns.boxplot()' function is used to create the box plot. The 'x' parameter specifies the column used for grouping the data, which is 'MaritalStatus' in this case. The 'y' parameter specifies the column used for the box plot, which is 'MonthlyIncome' in this case. The 'hue' parameter specifies the column used for color-coding the boxes based on the 'Attrition' column. The 'plt.show()' command is used to display the plot in a separate window or within the Jupyter notebook.



```
In [45]: sns.boxplot(x=df['MaritalStatus'],y=df['MonthlyIncome'],hue=df['Attrition'])
         plt.show()
```

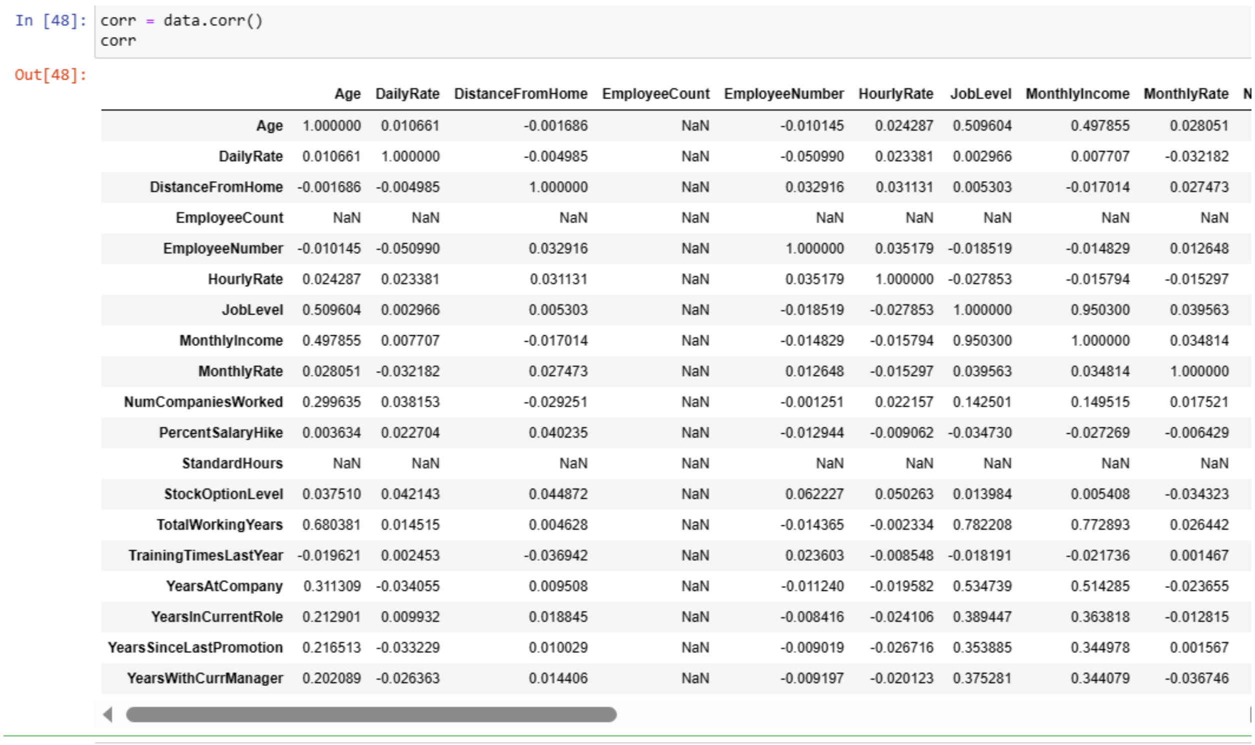


3.7 Correlation Analysis

The correlation is a very useful statitiscal analysis that describes the degree of relationship between two variables.

Let´s see the table below and the heat map to see what relationship are in the data.
Interpretation:

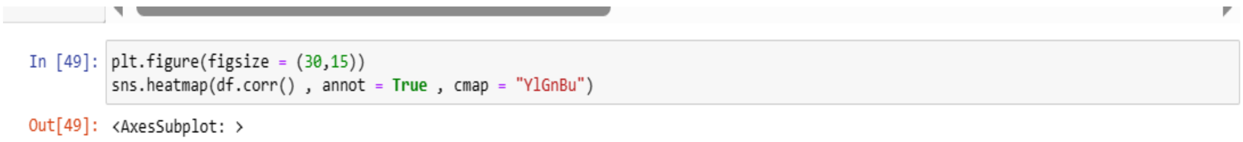
- Negative correlation of (-0.001) between distance from home to and the employees that left the company.
- The higher the total working years the higher the monthly income of an employee.
- The higher the percent salary hike the higher the performance rating.
- The higher the years with current manager the higher the years since last promotion.
- The higher the age the higher the monthly income.



In [48]:
```
corr = data.corr()
corr
```

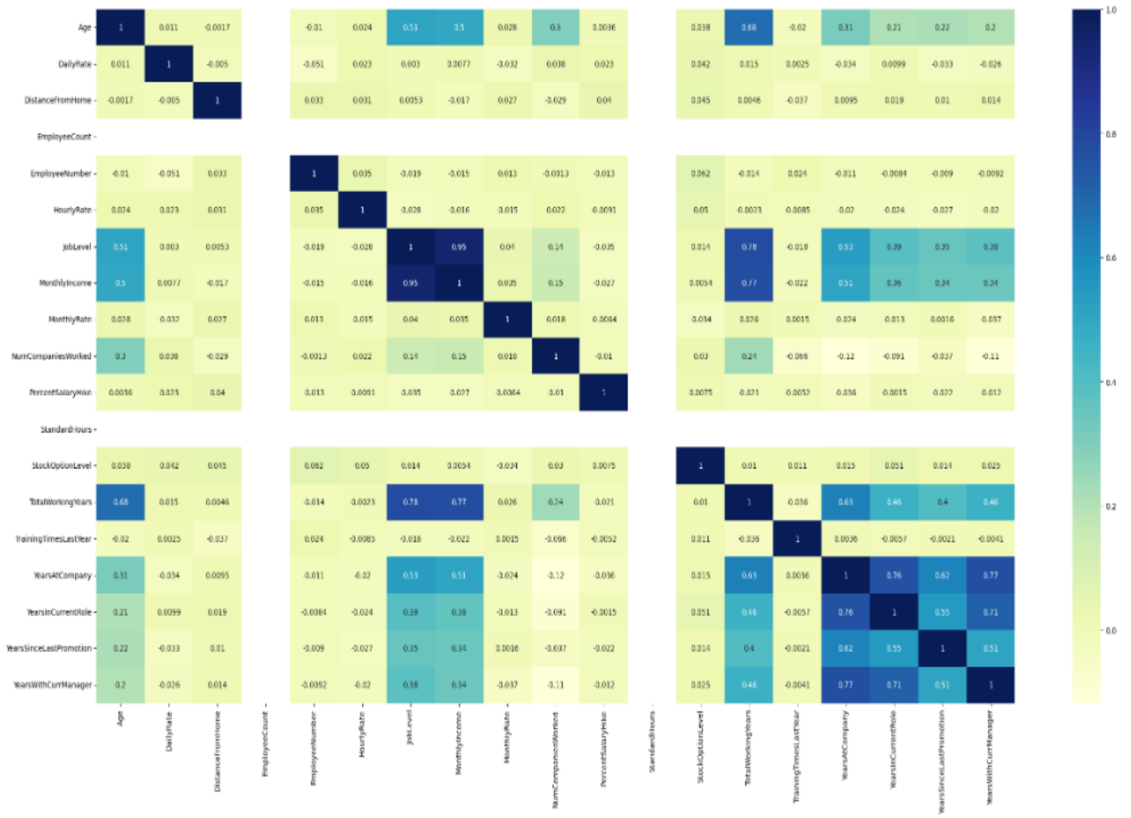
Out[48]:

	Age	DailyRate	DistanceFromHome	EmployeeCount	EmployeeNumber	HourlyRate	JobLevel	MonthlyIncome	MonthlyRate	N
Age	1.000000	0.010661	-0.001686	NaN	-0.010145	0.024287	0.509604	0.497855	0.028051	
DailyRate	0.010661	1.000000	-0.004985	NaN	-0.050990	0.023381	0.002966	0.007707	-0.032182	
DistanceFromHome	-0.001686	-0.004985	1.000000	NaN	0.032916	0.031131	0.005303	-0.017014	0.027473	
EmployeeCount	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
EmployeeNumber	-0.010145	-0.050990	0.032916	NaN	1.000000	0.035179	-0.018519	-0.014829	0.012648	
HourlyRate	0.024287	0.023381	0.031131	NaN	0.035179	1.000000	-0.027853	-0.015794	-0.015297	
JobLevel	0.509604	0.002966	0.005303	NaN	-0.018519	-0.027853	1.000000	0.950300	0.039563	
MonthlyIncome	0.497855	0.007707	-0.017014	NaN	-0.014829	-0.015794	0.950300	1.000000	0.034814	
MonthlyRate	0.028051	-0.032182	0.027473	NaN	0.012648	-0.015297	0.039563	0.034814	1.000000	
NumCompaniesWorked	0.299635	0.038153	-0.029251	NaN	-0.001251	0.022157	0.142501	0.149515	0.017521	
PercentSalaryHike	0.003634	0.022704	0.040235	NaN	-0.012944	-0.009062	-0.034730	-0.027269	-0.006429	
StandardHours	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
StockOptionLevel	0.037510	0.042143	0.044872	NaN	0.062227	0.050263	0.013984	0.005408	-0.034323	
TotalWorkingYears	0.680381	0.014515	0.004628	NaN	-0.014365	-0.002334	0.782208	0.772893	0.026442	
TrainingTimesLastYear	-0.019621	0.002453	-0.036942	NaN	0.023603	-0.008548	-0.018191	-0.021736	0.001467	
YearsAtCompany	0.311309	-0.034055	0.009508	NaN	-0.011240	-0.019582	0.534739	0.514285	-0.023655	
YearsInCurrentRole	0.212901	0.009932	0.018845	NaN	-0.008416	-0.024106	0.389447	0.363818	-0.012815	
YearsSinceLastPromotion	0.216513	-0.033229	0.010029	NaN	-0.009019	-0.026716	0.353885	0.344978	0.001567	
YearsWithCurrManager	0.202089	-0.026363	0.014406	NaN	-0.009197	-0.020123	0.375281	0.344079	-0.036746	

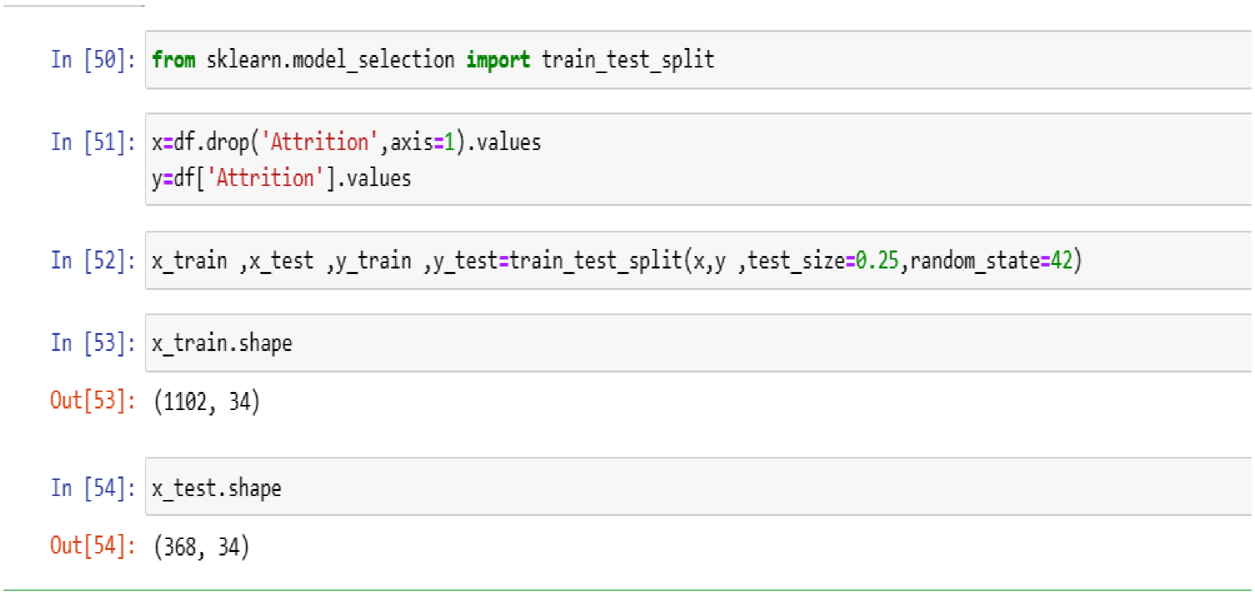


In [49]:
```
plt.figure(figsize = (30,15))
sns.heatmap(df.corr() , annot = True , cmap = "YlGnBu")
```

Out[49]: <AxesSubplot: >



3.9 Train Test Split



```
In [50]: from sklearn.model_selection import train_test_split

In [51]: x=df.drop('Attrition',axis=1).values
         y=df['Attrition'].values

In [52]: x_train ,x_test ,y_train ,y_test=train_test_split(x,y ,test_size=0.25,random_state=42)

In [53]: x_train.shape

Out[53]: (1102, 34)

In [54]: x_test.shape

Out[54]: (368, 34)
```

This code snippet is from the scikit-learn library in Python, and it is used to split the dataset into training and testing sets for a machine learning model.

Here's what each line of the code does: mfrom sklearn.model_selection import train_test_split: This imports the train_test_split function from the model_selection module of scikit-learn.

x = df.drop('Attrition', axis=1).values: This creates a new variable x that contains all the input features or independent variables in the dataset df, except for the column labeled Attrition.

y = df['Attrition'].values: This creates a new variable y that contains the target variable or dependent variable Attrition from the dataset df.

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=42): This splits the dataset into training and testing sets. The test_size parameter specifies the proportion of the dataset to be used for testing, in this case, 25%. The random_state parameter sets the random seed for reproducibility. The function returns four variables, x_train and y_train containing the training set, and x_test and y_test containing the testing set.

After splitting the dataset into training and testing sets, machine learning models can be trained on the training set and evaluated on the testing set to estimate their performance on new, unseen data.

Chapter 4

Conclusion

According to the data, training and overtime are the two biggest variables influencing employee attrition.

- **Overtime :** is necessary, it has been noticed that individuals with more seniority and experience are more likely to leave the organisation. This might be explained by their additional responsibilities or a deeper level of familiarity with the job. Therefore, it is advised that workload and staffing be considered in order to lessen the detrimental effects of overtime on employee attrition.

- **Lack of training:** Higher attrition rates are correlated with less training each year. However, a reduction in attrition rates is not always the result of greater training frequency. To maintain a low attrition rate, it is essential to ascertain the ideal amount of training needed by each department.

Employees tend to favour high-quality training delivered sporadically throughout the year as opposed to low-quality training delivered more regularly, so it is crucial to carefully choose the training material. This is especially true for positions like manager and representative in the healthcare industry.

A similar pattern is shown for all other job roles, with the highest turnover rates occurring when there is no training offered annually. However, it should be emphasised that each position has certain training requirements, and either too much or too little training might lead to higher turnover rates.

The analysis of the data revealed no evidence of a connection between attrition and either work satisfaction or job roles.

Numerous factors can affect employee attrition, or the pace at which individuals depart an organisation. The following are some typical factors that may have an impact on employee attrition:

1. Job satisfaction: Employees may be more likely to leave the company if they are not satisfied with their position, the workplace, or organisational culture.

2. Compensation and benefits: Employees have the right to quit if they believe their pay is unfair or their perks are insufficient.

3. Workload and work-life balance: Employees may be more prone to quit if they feel overworked or unable to strike a balance between their personal and professional lives.

4. Opportunities for development: Employees may be more likely to go elsewhere for employment if they believe there are no opportunities for promotion inside the company.

5. Management and leadership: Unhealthy management or leadership techniques can produce a toxic workplace that may drive away personnel.

6. Organisational culture: Employees may be more likely to quit if they don't feel their values or mission are shared by the organisation.

7. connections at work: A lack of positive working connections with coworkers or bosses can lead to employee discontent and attrition.

8. Employees may quit if their commute is excessively long or challenging.

9. Competition in the industry: If there is a shortage of talented individuals due to intense competition, it may be simpler for employees to locate alternative employment possibilities.

10. Retirement: Older workers may quit because they are retiring or prefer to work more slowly.

Analysing the link between staff attrition and performance can reveal important information about the advantages and disadvantages of a company.

Organizations should state the requirements and expectations unambiguously. This helps candidates decide upon to accept the job position or not. This eventually avoids further conflicts in the employment terms.

High employee churn rates can have a detrimental effect on an organization's performance since they can result in the loss of institutional knowledge and experience, lowered morale among surviving staff members, and higher recruitment and training expenses for new hires. In order to address the core causes of employee attrition, it is crucial to identify them.

The analysis of employee performance data, on the other hand, can reveal information about how well people are performing and whether there are any areas where they could do better. Organisations may boost employee engagement and retention by recognising high-performing employees and offering them chances for advancement.

Recommendation

I suggest changing two components of the company's present procedures to bring the attrition rate down to less than 10%. Although these suggestions are based on high correlations with the underlying issue, it should be remembered that correlation does not always imply causation.

1. Overtime : When it comes to using overtime, it should only be done as a last resort.

2. Tasks & Staffing It is advised to look into this area first because the dataset lacks data on department staffing or workload. Overtime is frequently caused by work surges or a lack of staff.

Employees could be willing to put in a short stint of extra if there is an unexpected spike in workload. However, if overtime requirements become common, workers may start to think about changing jobs or organisations. As a result, the business needs to evaluate the burden in each department to decide how to distribute the work. The business should then evaluate the workforce levels for each department while taking the simplified workload into account. If understaffing is the primary factor causing overtime, it could be required to hire new personnel.

Effective training is essential for every organization's success, yet year-round training delivery is not feasible. As a result, businesses should plan training according to the unique requirements of each department. This is crucial because every role calls for a particular kind of assistance.

Training Programmes by Department : Training materials must be thoroughly examined and modified to reflect the current demand of the workforce. The workload assessment used to save overtime can also be utilised when organising training sessions to comprehend employees' daily tasks and choose the most appropriate training opportunities.

Additionally, the number of sessions should be determined by the unique requirements of each department as training is conducted in tiers. In addition to receiving training on their job duties, management should also learn how to effectively train their team members. Managers should be trained, and human resources should assist managers in redesigning training programmes.

Here are some recommendations based on a general viewpoint for an organisation undertaking a study on staff attrition and performance:

1. Determine the elements that affect employee attrition: Analyse the reasons why employees are leaving the organisation in great detail. This could entail gathering information on, among other things, leadership, remuneration, benefits, and employee satisfaction.

2. Create methods to handle staff attrition: After identifying the underlying factors, create approaches to dealing with them. For instance, if low pay is a problem, think about raising salaries or providing workers with more benefits.

3. Analyse the statistics on employee performance: Data on employee performance should be gathered, including information on parameters like quality, productivity, and customer satisfaction. Identify high-performing employees and opportunities for development using this data.

4. Offer possibilities for development and growth: Identify high-performing personnel and offer chances for development and advancement within the company. Programmes for leadership development, mentoring, and training might be included in this.

5. Track results and make necessary strategy changes: Monitor personnel attrition rates and performance information often to ascertain the efficacy of the techniques being used. Make necessary adjustments to strategies to make sure they are benefiting the organisation.

Analysing employee performance and attrition can reveal important information about the advantages and disadvantages of a company. Organisations can foster a productive work environment that fosters employee satisfaction, engagement, and overall success by adopting a strategic approach to tackling employee attrition and encouraging staff performance.

References and Annexure.

1. Class lectures of the alliance university
2. https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset
3. Python Exploratory Data Analysis:
   https://www.datacamp.com/community/tutorials/exploratory-data-analysis-python