

# **Corporate Employee Attrition Analysis**

## **A PROJECT REPORT**

Submitted By

**Team ID: PNT2022TMID38728**

**Team Leader: POOVARASAN S (420619106009)**

**Team Member: PAVITHRAN V (420619106008)**

**Team Member: SURIYA C (420619106014)**

**Team Member: KISHORE M (420619106005)**

**Team Member: SARASWATHI M (420619106011)**

In partial fulfillment for the award of the

degree of

**BACHELOR OF ENGINEERING**

**IN**

**ELECTRONICS AND COMMUNICATIONS ENGINEERING**

**ANNAI TERESA COLLEGE OF ENGINEERING, THIRUNAVAILUR**

**ANNA UNIVERSITY::CHENNAI 600025**

## TABLE OF CONTENTS

| <b>CHAPTER NO</b> | <b>TITLE</b>                            | <b>PAGE NO</b> |
|-------------------|---|----------------|
| <b>1</b>          | <b>INTRODUCTION</b>                     | <b>4</b>       |
|                   | Project Overview                        | 4              |
|                   | Purpose                                 | 4              |
| <b>2</b>          | <b>LITERATURE SURVEY</b>                | <b>5</b>       |
|                   | Existing problem                        | 5              |
|                   | References                              | 5              |
|                   | Problem Statement Definition            | 5              |
| <b>3</b>          | <b>IDEATION &amp; PROPOSED SOLUTION</b> |                |
|                   | Empathy Map Canvas                      | 6              |
|                   | Ideation & Brainstorming                | 6              |
|                   | Proposed Solution                       | 8              |
|                   | Problem Solution fit                    | 9              |
| <b>4</b>          | <b>REQUIREMENT ANALYSIS</b>             |                |
|                   | Functional requirement                  | 10             |
|                   | Non-Functional requirements             | 10             |
| <b>5</b>          | <b>PROJECT DESIGN</b>                   |                |
|                   | Data Flow Diagrams                      | 11             |
|                   | Solution & Technical Architecture       | 11             |
|                   | User Stories                            | 12             |

|           |  |           |
|-----------|--|-----------|
| <b>6</b>  | <b>PROJECT PLANNING &amp; SCHEDULING</b> |           |
|           | Sprint Planning & Estimation             | <b>13</b> |
|           | Sprint Delivery Schedule                 | <b>13</b> |
|           | Reports from JIRA                        |           |
| <b>7</b>  | <b>CODING &amp; SOLUTIONING</b>          |           |
|           | Feature 1                                | <b>13</b> |
|           | Feature 2                                | <b>14</b> |
|           | Database Schema                          |           |
| <b>8</b>  | <b>TESTING</b>                           |           |
|           | Test Cases                               | <b>15</b> |
|           | User Acceptance Testing                  | <b>15</b> |
| <b>9</b>  | <b>RESULTS</b>                           | <b>17</b> |
| <b>10</b> | <b>ADVANTAGES &amp; DISADVANTAGES</b>    | <b>18</b> |
| <b>11</b> | <b>CONCLUSION</b>                        | <b>18</b> |
| <b>12</b> | <b>FUTURE SCOPE</b>                      | <b>19</b> |
| <b>13</b> | <b>APPENDIX</b>                          | <b>20</b> |

## **1. INTRODUCTION**

### **Project overview**

Employee attrition has become a vital problem across the world. It is one of the crucial issues faced by business leaders within companies where they lose the most talented employees. A good employee is always an asset to the organization and their resignation can lead to various problems like financial losses, overall performance, and loss of acquired knowledge. Furthermore, hiring new employees is far exorbitant, taxing, and time-consuming in comparison to recruiting the existing one. It is very time-consuming to recruit a new employee as it takes him months for training, adjusting to the culture, rules, and environment. Therefore, upcoming trends and technology using Machine Learning Algorithms must be exploited for the benefit of business organizations. Knowing the reason beforehand for the employee attrition, companies can mitigate this loss. This analysis provides a conclusive review of employee attrition from the data set IBM HR Analytics Employee Attrition Performance.

### **Purpose**

[1] Hardik P. K. ( 2016) , researched on “a study on employee attrition: with special reference to Kerala IT Industry”. His research examined the relationship between organizational factors and attrition of IT professional’s. The result can conclude that the organizational factors played significant role in predicting the variance in turnover intention (attrition) of Kerala IT professionals. Therefore, the HR managers in IT organizations may take into consideration the problems with organizational

factors of their workers to reduce the turnover intention of the skilled employees.

## 1. LITERATURE SURVEY

### Existing Problem

The Existing system includes only few attributes for analysis and also deals with qualitative observations and simple statistical analysis. The qualitative observations deal with data and can be observed through human senses. They do not involve measurements or number. Due to the increase in IOT and connected device, we now have access to so much of data and along with it an increase needs to manage and understand data.

### References

1. From Big Data to Deep Data to support people analytics for employee attrition prediction, Nesrine Ben Yahia, Hlel Jihen, Ricardo Colomo-Palacio( 2021)
- 2.Machine Learning Approach for Employee Attrition Analysis.Dr. R. S. Kamath | Dr. S. S. Jamsandekar | Dr. P. G. Naik ,Published in International Journal of Trend in Scientific Research and Development (ijtsrd), (March 2019)
3. Investigation of early career teacher attrition(ECT) and the impact of induction programs in Western Australia, Janine E.Wyatt, MichaelO'Neill (2021)

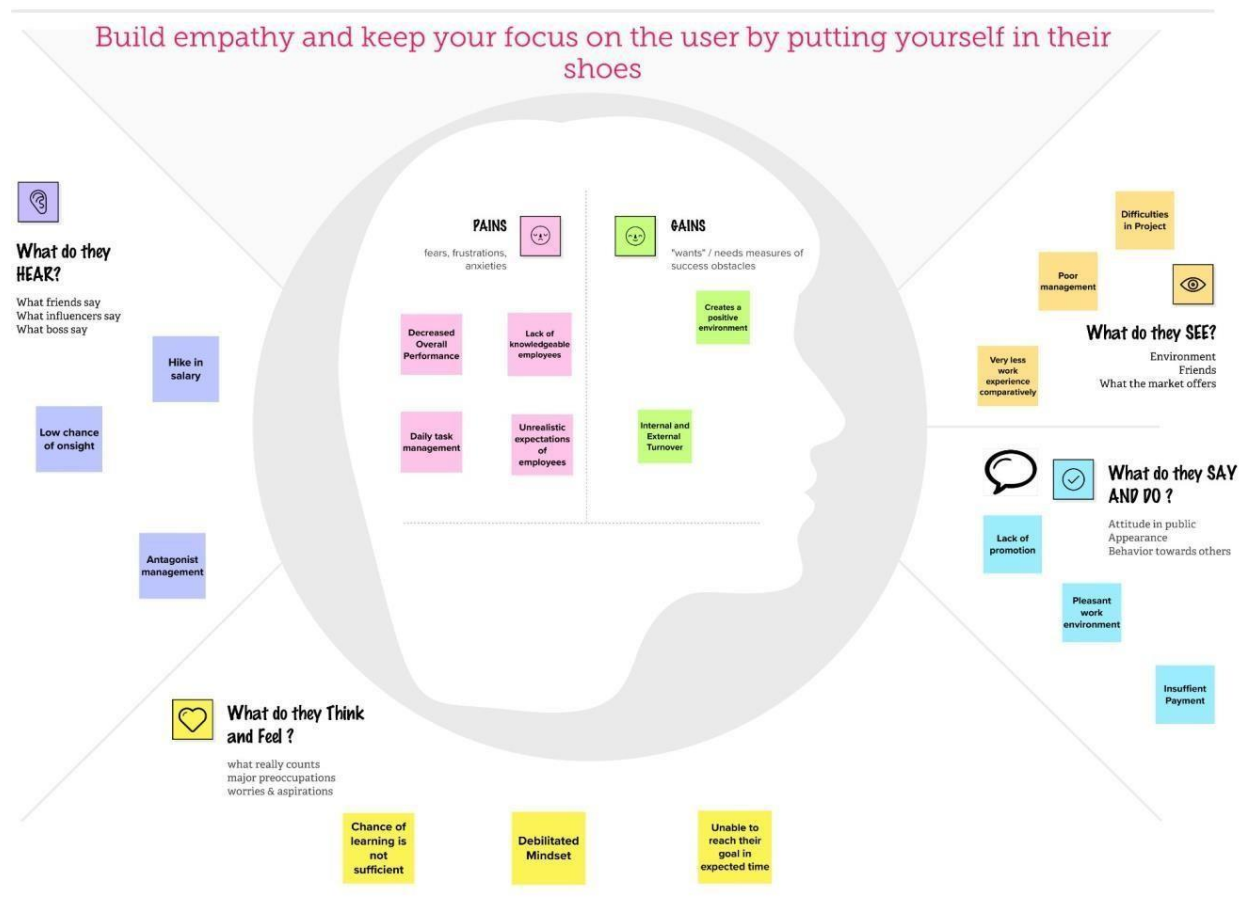
### Problem Statement Definition

- To create a dashboard and perform analysis of employee attrition in corporates using IBM Cognos analytics platform.
- To reduce the employee attrition rate through data analytics,

data visualization by analysing the major factors that causes attrition.


### 3. IDEATION AND PROPOSED SOLUTION

#### Empathy Map Canvas



Ideation & Brainstorming

Template



### Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

10 minutes to prepare

1 hour to collaborate

2-8 people recommended

Share template feedback

➔

**Before you collaborate**

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A

**Team gathering**

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B

**Set the goal**

Think about the problem you'll be focusing on solving in the brainstorming session.

C

**Learn how to use the facilitation tools**

Use the Facilitation Superpowers to run a happy and productive session.

Open article ➔

1

**Define your problem statement**

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

Employee Attrition is a major cost to an organization and predicting such attrition is the most important requirement of the human resources department in many organizations. In this problem our task is to predict the attrition rate of employee of an organization.

Key rules of brainstorming

To run an smooth and productive session

Stay in topic.

Defer judgment.

Go for volume.

Encourage wild ideas.

Listen to others.

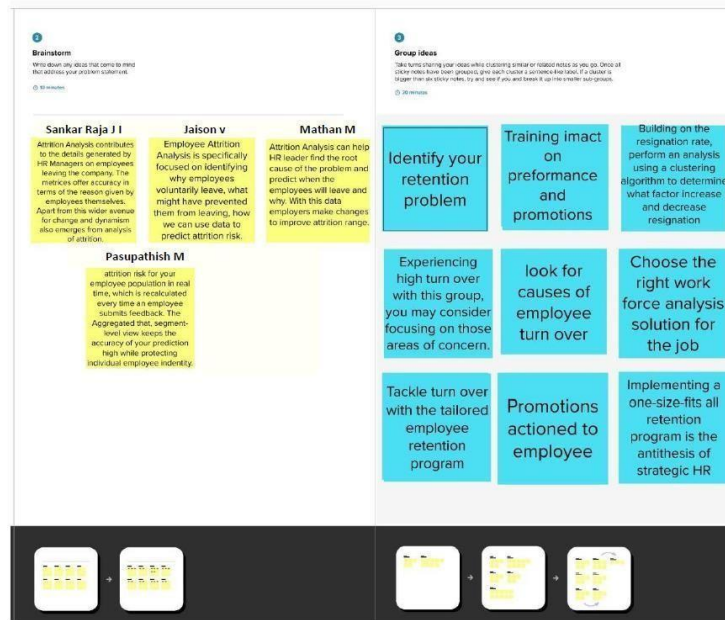
If possible, be visual.

Need some inspiration?

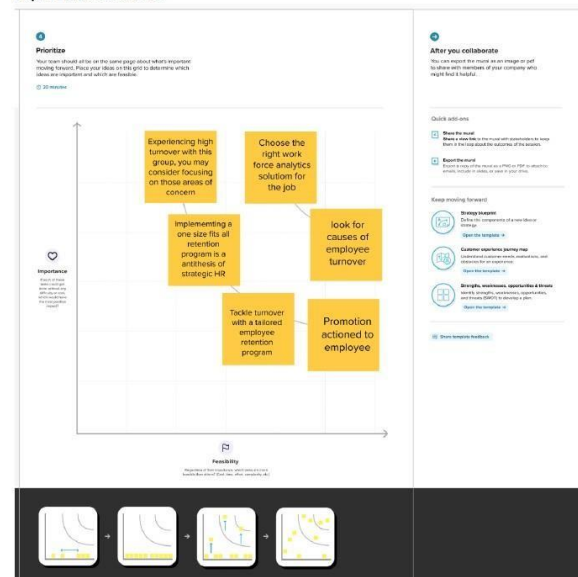
See a finished version of this template or kickstart your work.

Open example ➔

7



### Step-3: Idea Prioritization



## Proposed Solution

The Existing system includes only few attributes for analysis and also deals with qualitative observations and simple statistical analysis. The qualitative observations deal with data and can be observed through human senses. They do not involve measurements or number. Due to the increase in IOT and connected device, we now have access to so much of data and along with it an increase needs to manage and understand data.



## Problem Solution fit

Project Title: Corporate Employee Attrition Analytics

Project Design Phase-I – Problem Solution Fit

Team ID: PNT2022TMID34476

|                        |   |   |  |                           |
|------------------------|---|---|--|---------------------------|
| Define CS, fit into CC | <b>1. CUSTOMER SEGMENT(S)</b> <span>CS</span> <ul style="list-style-type: none"> <li>➤ The customer of this project will be the HR professionals, the administration or the person with the higher power authority who are responsible for their lower-level employees.</li> <li>➤ The customer uses the employee data</li> </ul>   | <b>6. CUSTOMER CONSTRAINTS</b> <span>CC</span> <ul style="list-style-type: none"> <li>➤ The constraints which the which can't be used customer would face may be the lack of skilled employee or the for analysis amount of surplus employee would bring the issue in decision</li> <li>➤ Lack making in taking the appropriate results of communication.</li> <li>➤ Unstructured data</li> </ul> | <b>5. AVAILABLE SOLUTIONS</b> <span>AS</span> <ul style="list-style-type: none"> <li>➤ Initially the performance of the employee is observed manually by the higher officials.</li> <li>➤ But this may lead to imbalance in treating all employees as same.</li> <li>➤ But the analysis will be completely digital so that there may not occur any favouritism.</li> </ul> | Explore AS, differentiate |
|                        | <b>2. JOBS-TO-BE-DONE / PROBLEMS</b> <span>J&amp;P</span> <ul style="list-style-type: none"> <li>➤ Initially data has to be collected and formatted in a proper way.</li> <li>➤ A deep analysis of the employee data should be done in order to gain the results.</li> <li>➤ The problem which may arise here is sometimes the data may be an invalid or incorrect data which affects the results.</li> </ul> | <b>9. PROBLEM ROOT CAUSE</b> <span>RC</span> <ul style="list-style-type: none"> <li>➤ To identify the potential employees.</li> <li>➤ To find the reason of employee attrition</li> <li>➤ To improve the organization profit by retaining good talents.</li> <li>➤ To consider every employee performance.</li> </ul>   | <b>7. BEHAVIOUR</b> <span>BE</span> <ul style="list-style-type: none"> <li>➤ Directly related with the higher authorities.</li> <li>➤ Indirectly related with the knowledge of the employees.</li> </ul>   |                           |

|                         |   |   |   |
|-------------------------|---|---|---|
| Identify strong TR & EM | <b>3. TRIGGERS</b> <span>TR</span> <ul style="list-style-type: none"> <li>➤ With the analysis, the employee will be more aware of his responsibilities being done.</li> <li>➤ It encourages good employees to step forward in their career and it serves as a warning for those employees who are not being responsible in their work.</li> </ul> | <b>10. YOUR SOLUTION</b> <span>SU</span> <ul style="list-style-type: none"> <li>➤ The solution would be the attrition analytics which gains the useful results which may be beneficial both to the employees as well as to the organization.</li> </ul> | <b>8. CHANNELS OF BEHAVIOUR</b> <span>CH</span> <ul style="list-style-type: none"> <li>➤ The customers can perform visualization using different graphs, can draw many useful insights from it.</li> <li>➤ Using the results which was collected, the action may be taken offline. Preparing datasets can be done offline.</li> </ul> |
|                         | <b>4. EMOTIONS: BEFORE / AFTER</b> <span>EM</span> <ul style="list-style-type: none"> <li>➤ The good employees will be encouraged and the irresponsible ones will be noticed.</li> </ul>  |   |   |

## 4. REQUIREMENT ANALYSIS

### Functional requirement

| FR No. | Functional Requirement (Epic) | Sub Requirement (Story / Sub-Task)   |
|--------|-------------------------------|--|
| FR-1   | User Registration             | Registration through Form<br>Registration through Gmail<br>Registration through LinkedIn                         |
| FR-2   | User Confirmation             | Confirmation via Email<br>Confirmation via OTP   |
| FR-3   | User Feedback                 | Feedback through Form<br>Feedback through Gmail<br>Feedback through Instagram polls<br>Feedback through LinkedIn |
| FR-4   | User Rating                   | Rating via Mail<br>Rating through Message  |
| FR-5   | Employee Management           | Validating and managing the employee details   |
| FR-6   | Attrition Analytics           | Analysing and finding out the major reason for the attrition of employees using dataset                          |

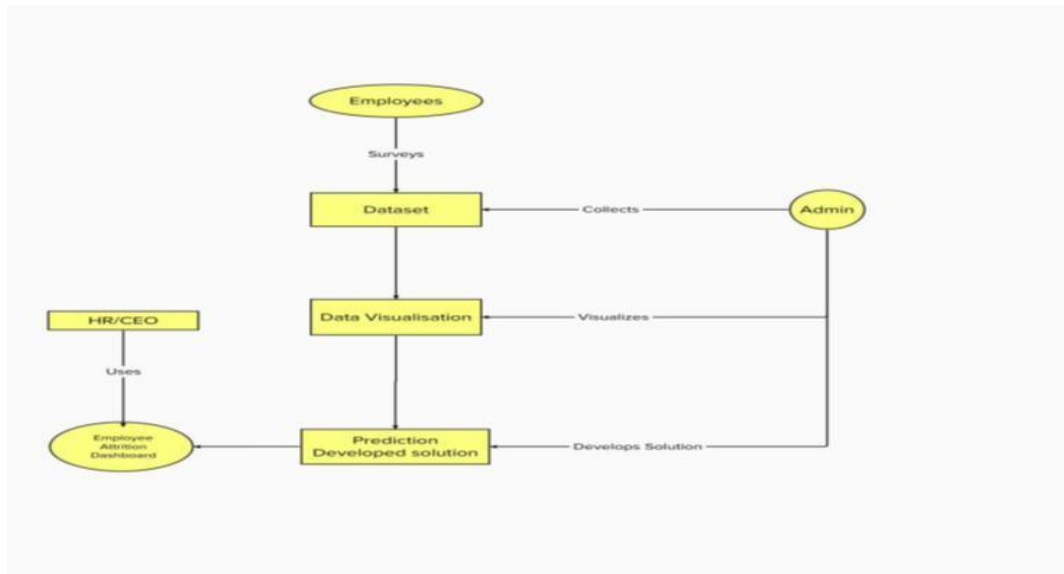
### Non-Functional requirements

Following are the non-functional requirements of the proposed solution.

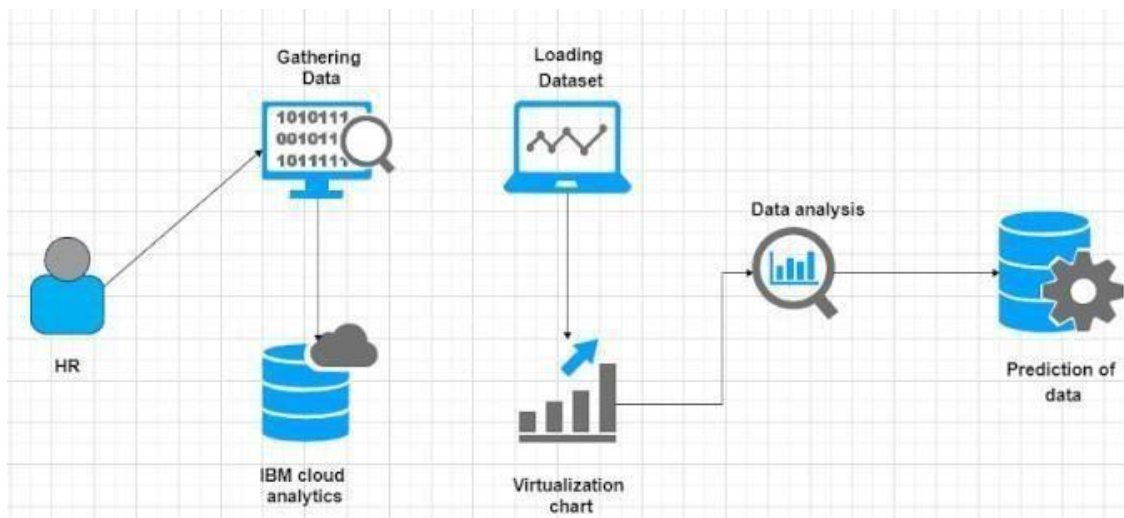
| FR No. | Non-Functional Requirement | Description   |
|--------|----------------------------|---|
| NFR-1  | <b>Usability</b>           | This Data Visualization shall be easy to use for all users with minimal instructions. 100% of the languages on the graphical user interface (GUI) shall be intuitive and understandable by non-technical users. |
| NFR-2  | <b>Security</b>            | The employee data is kept secure and their identity is hidden for the organization.   |
| NFR-3  | <b>Reliability</b>         | The Link shall be operable in all conditions. The system must be less prone to errors   |
| NFR-4  | <b>Performance</b>         | This software is portable and inter-operable. It works smoothly without generating errors. It also provides a faster response   |
| NFR-5  | <b>Portability</b>         | The link shall be portable to all operating platforms. Therefore, this link should not depend on the different operating systems.   |
| NFR-6  | <b>Scalability</b>         | Our solution is scalable for large and small datasets. It provides an efficient solution despite the size of the dataset.   |

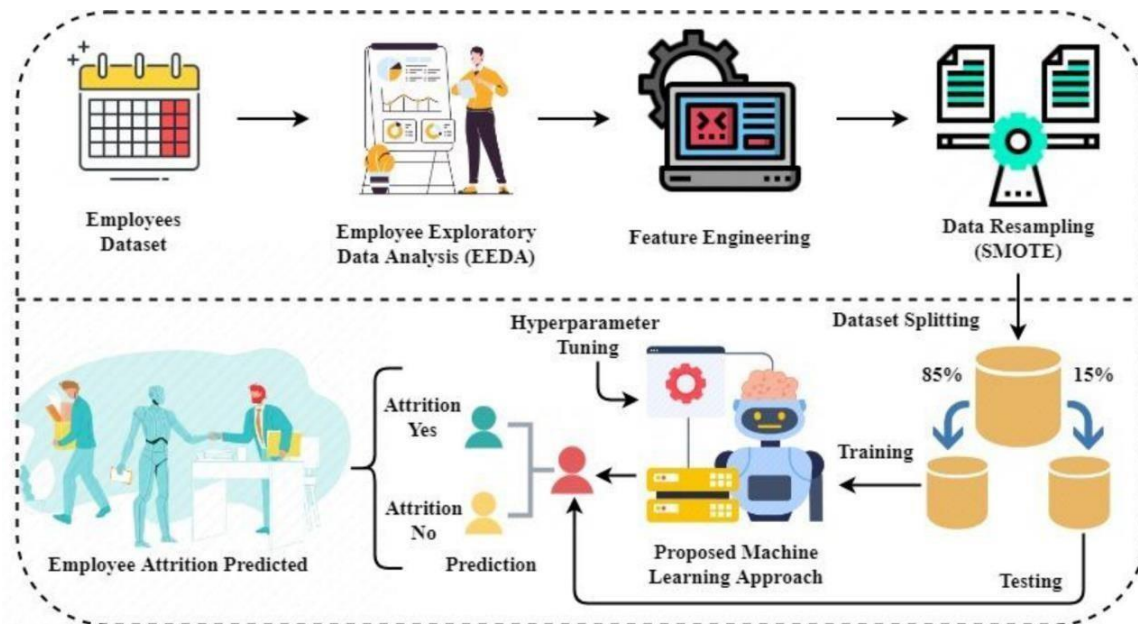
## 5. PROJECT DESIGN

### Data Flow Diagrams



### Solution & Technical Architecture





## User Stories

| User Type               | Functional Requirement (Epic) | User Story Number | User Story / Task   | Acceptance criteria  | Priority | Release  |
|-------------------------|-------------------------------|-------------------|---|--|----------|----------|
| Customer (Web user)     | Registration                  | USN-1             | As a user, I can register for the application by entering my email, password, and confirming my password. | I can access my account / dashboard  | High     | Sprint-1 |
|                         |                               | USN-2             | As a user, I will receive confirmation email once I have registered for the application                   | I can receive confirmation email & click confirm                                       | High     | Sprint-1 |
|                         |                               | USN-3             | As a user, I can register for the application through Facebook  | I can register & access the dashboard with Facebook Login                              | Low      | Sprint-2 |
|                         |                               | USN-4             | As a user, I can register for the application through Gmail   | I can register & access the dashboard with Gmail Login                                 | Medium   | Sprint-1 |
|                         | Login                         | USN-5             | As a user, I can log into the application by entering email & password                                    | I can access my account / dashboard  | High     | Sprint-1 |
|                         | Dashboard                     | USN-6             | Uploading the Dataset   | I can be able to upload my dataset   | High     | Sprint-2 |
|                         |                               | USN-7             | Working With Dataset  | I can be able to access my dashboard   | High     | Sprint-2 |
|                         |                               | USN-8             | Visualization   | I can be able to view the visual attrition rate of my dataset                          | High     | Sprint-3 |
|                         |                               | USN-9             | Working with Dashboard  | I can be able to view the various views of the attrition rate                          | High     | Sprint-3 |
| Customer Care Executive |                               | USN-10            | Asking Help / Feedback  | I can be able to ask help if I can face any issues or problems while using the webpage | Medium   | Sprint-4 |
| Administrator           |                               | USN-11            | Managing the Database   | I can assure that my data is in secure state   | High     | Sprint-4 |
|                         |                               | USN-12            | Managing the over all process   | I can assure that my data and process is going good                                    | High     | Sprint-4 |

## 6. PROJECT PLANNING

### Sprint Planning & Estimation

### Sprint Delivery Schedule

| Sprint   | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as on Planned End Date) | Sprint Release Date (Actual) |
|----------|--------------------|----------|-------------------|---------------------------|---|------------------------------|
| Sprint-1 | 20                 | 6 Days   | 24 Oct 2022       | 29 Oct 2022               | 29 October 2022                                 | 05 November 2022             |
| Sprint-2 | 20                 | 6 Days   | 31 Oct 2022       | 05 Nov 2022               | 05 November 2022                                | 06 November 2022             |
| Sprint-3 | 20                 | 6 Days   | 07 Nov 2022       | 12 Nov 2022               | 08 November 2022                                | 09 November 2022             |
| Sprint-4 | 20                 | 6 Days   | 14 Nov 2022       | 19 Nov 2022               | 11 November 2022                                | 16 November 2022             |

## 7. CODING & SOLUTIONING

**F  
e  
a  
t  
u  
r  
e**

**1**

**#  
G  
E  
N  
E  
R  
A  
L**

**i  
m  
p  
o  
r  
t**

**p  
a  
n  
d  
a**

s

a

s

p

d

i

m

p

o

r

t

n

u

m

p

y

a

s

n

p

```
import seaborn as sns
import matplotlib.pyplot as plt
path = '/content/general_data.csv'
df=pd.read_csv(path)
```

```

df
df.shape
df.info()
df.select_dtypes('int64','float64').columns
cat_cols = df.select_dtypes('object').columns
cat_cols
df.describe().T
df
for cat in cat_cols:
    print(cat,'->',df[cat].unique())
    print()
print("All columns Unique values count")
for col in df:
    print(col, len(df[col].unique()), sep=': ')
plt.figure(figsize=(14,5))
plt.subplot(1,2,1)
sns.countplot(df['Attrition'],color='b',hue=df['Gender'])
plt.title('Attrition by Gender')
plt.subplot(1,2,2)
plt.pie(df['Attrition'].value_counts(),colors=['r','c'],explode=[0,0.1],autopct=
'%0.2f',labels=['No','Yes'])
plt.title('Attrition')
#HANDLING CATEGORICAL OUTPUT VARIABLE
df['Attrition'].replace({'Yes':1,'No':0},inplace=True)
df['Attrition'].head()
plt.figure(figsize=(20,8))
sns.boxplot(x='JobRole',y='MonthlyIncome',data=df,hue='Attrition',color='red')
col=['YearsInCurrentRole','YearsSinceLastPromotion','YearsWithCurrManager',
'YearsAtCompany']
plt.figure(figsize=(10,10))
for i,c in enumerate(col):
    plt.subplot(2,2,i+1)
    sns.distplot(df[c],color='b')

```

**F  
e  
a  
t  
u  
r  
e**

**2**

**#  
G  
E  
N**



E  
R  
A  
L

i  
m  
p  
o  
r  
t

p  
a  
n  
d  
a  
s

a  
s

p  
d

i  
m  
p  
o  
r  
t

n  
u  
m  
p  
y

a  
s

n  
p

```
import seaborn as sns
import matplotlib.pyplot as plt
#FEATURE ENGINEERING
from sklearn.preprocessing import LabelEncoder
from imblearn.over_sampling import SMOTE
path = '/content/general_data.csv'
df =pd.read_csv(path)
df
df.shape
```

```
df.info()  
df.select_dtypes('int64', 'float64').columns
```

```

cat_cols = df.select_dtypes('object').columns
cat_cols
df.describe().T
df
for cat in cat_cols:
    print(cat, '-> ', df[cat].unique())
    print()
print("All columns Unique values count")
for col in df:
    print(col, len(df[col].unique()), sep=': ')
plt.figure(figsize=(14,5))
plt.subplot(1,2,1)
sns.countplot(df['Attrition'], color='b', hue=df['Gender'])
plt.title('Attrition by Gender')
plt.subplot(1,2,2)
plt.pie(df['Attrition'].value_counts(), colors=['r','c'], explode=[0,0.1], autopct=
'%0.2f', labels=['No', 'Yes'])
plt.title('Attrition')
#HANDLING CATEGORICAL OUTPUT VARIABLE
df['Attrition'].replace({'Yes':1, 'No':0}, inplace=True)
df['Attrition'].head()
df.drop(columns=no_use, axis=1, inplace=True)
df.columns
df['Gender'].replace({'Male':1, 'Female':0}, inplace=True)
df['OverTime'].replace({'Yes':1, 'No':0}, inplace=True)
(df.Attrition.value_counts()/1470)*100
smote = SMOTE(sampling_strategy='minority')
x, y = smote.fit_resample(x, y)
print(x.shape, y.shape)
#now balanced
y.value_counts()
sns.countplot(y, palette='viridis')
plt.title('Now Class is Balanced')

```

## 8. TESTING

### Test Cases

### User Acceptance Testing

#### 1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issue of corporate employee attrition at the time of the release.

#### 2. Defect Analysis

| Resolution     | Severity 1 | Severity 2 | Severity 3 | Severity 4 | Subtotal |
|----------------|------------|------------|------------|------------|----------|
| By Design      | 3          | 2          | 0          | 0          | 5        |
| Duplicate      | 4          | 0          | 2          | 0          | 6        |
| External       | 3          | 2          | 0          | 0          | 5        |
| Fixed          | 1          | 0          | 1          | 0          | 2        |
| Not Reproduced | 0          | 3          | 3          | 0          | 6        |
| Skipped        | 0          | 0          | 3          | 2          | 5        |
| Won't Fix      | 0          | 0          | 1          | 0          | 1        |
| Totals         | 11         | 7          | 10         | 2          | 30       |

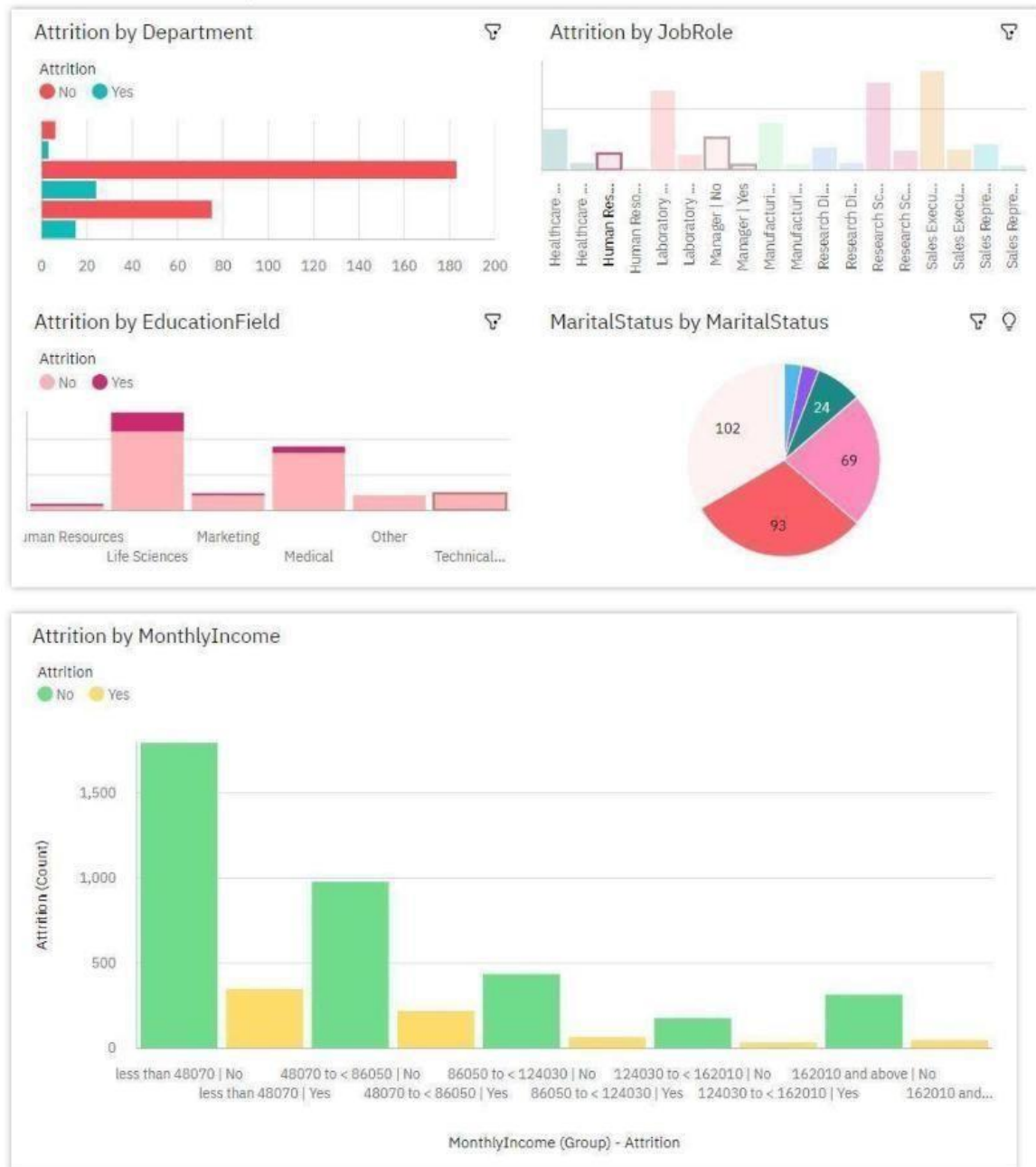
### 3. Test Case Analysis

|                     |   |   |   |   |
|---------------------|---|---|---|---|
| Database            | 2 | 0 | 0 | 2 |
| Dashboard           | 1 | 0 | 0 | 1 |
| Visualize the data  | 8 | 0 | 0 | 8 |
| Logistic Regression | 4 | 0 | 0 | 4 |

| Section                    | Total Cases | Not Tested | Fail | Pass |
|----------------------------|-------------|------------|------|------|
| Login Page                 | 1           | 0          | 0    | 1    |
| Employee Attrition Details | 1           | 0          | 0    | 1    |

## 9. RESULTS

### 9.1 Performance Metrics



## **9. ADVANTAGES & DISADVANTAGES**

### **9.1 Advantages**

**Data Collection :** The study is conducted among working IT professionals of two different categories. This categorization mainly was focused on experience level and role in the organization. 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.

**Primary Data** Primary data is collected through a field survey with the help of a structured self-administrated Questionnaire. The survey consisted of close ended questions by the means of convenience sampling. The scaling technique installed in the questionnaire is 5-point rating scale. Total 120 respondent were IT professionals belonging to the organizations from Nagpur, Pune and Mumbai cities in Maharashtra.

**Secondary Data** Secondary data is collected by referring to the Journals, research papers and published data in the form of books and newspapers.

### **Type of Research :**

The research paper adopted the descriptive research design methodology. **Sample Design, Sample Size and Sampling Method** The sample selected for the study is an Indian Information Technology Industry. The nature of the sample is restricted to working professionals in Information Technology sector and is collected through the convenience sampling technique. The sample size was 120 respondents.

## **9. CONCLUSION**

Employees as well as organizations must be clear with their expectations regarding the job profile. Any sort of mismatch leads to discrepancy and employees may fail to perform at their job. This eventually leads to attrition. 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.

## **10. FUTURE SCOPE**

Research findings suggest that attrition reasons in IT organizations primarily revolve around professional growth and challenges in the organization. Although economic factors happen to be the most influential factor, professionals may settle for second best criteria of their preference that is career growth and supportive work policies in the organization. On the other hand, candidates who aspire to have a better job than the one in hand are more interested in securing the next job. Young talent wants to work on latest technology and functional domain. IT professionals who are young career makers are less influenced by Brand name or geographical area. Most of the IT professionals look for challenging role and position in the organization. Candidates as well as senior professionals believe that challenging work motivate them to maintain the interest in the work life. Employees as well as organizations must be clear with their expectations regarding the job profile. Any sort of mismatch leads to discrepancy and employees may fail to

perform at their job. This eventually leads to attrition. 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. Further this research can make more detailed conclusions over “mapping of candidates’ expectations with organizations’ requirement” by collecting the data focusing on all the steps of recruitment and selection process.

## **11. APPENDIX**

### **12.1 Source Code**



```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns

```

#### DATASET 1

```
df1=pd.read_csv('/content/drive/MyDrive/attrition/employee_attrition_train.csv')
```

```

from google.colab import drive
drive.mount('/content/drive')

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

```
df1
```

|      | Age  | Attrition | BusinessTravel    | DailyRate | Department             | DistanceFromHome | Education | EducationField | EmployeeCount | EmployeeNumber | ... | RelationshipSatisfaction | Sta |
|------|------|-----------|-------------------|-----------|------------------------|------------------|-----------|----------------|---------------|----------------|-----|--------------------------|-----|
| 0    | 50.0 | No        | Travel Rarely     | 1126.0    | Research & Development | 1.0              | 2         | Medical        | 1             | 997            | ... | 3                        |     |
| 1    | 36.0 | No        | Travel Rarely     | 216.0     | Research & Development | 6.0              | 2         | Medical        | 1             | 178            | ... | 4                        |     |
| 2    | 21.0 | Yes       | Travel Rarely     | 337.0     | Sales                  | 7.0              | 1         | Marketing      | 1             | 1780           | ... | 2                        |     |
| 3    | 50.0 | No        | Travel Frequently | 1246.0    | Human Resources        | NaN              | 3         | Medical        | 1             | 644            | ... | 3                        |     |
| 4    | 52.0 | No        | Travel Rarely     | 994.0     | Research & Development | 7.0              | 4         | Life Sciences  | 1             | 1118           | ... | 4                        |     |
| ...  | ...  | ...       | ...               | ...       | ...                    | ...              | ...       | ...            | ...           | ...            | ... | ...                      | ... |
| 1024 | NaN  | No        | Travel Rarely     | 750.0     | Research & Development | 28.0             | 3         | Life Sciences  | 1             | 1596           | ... | 4                        |     |
| 1025 | 41.0 | No        | Travel Rarely     | 447.0     | Research & Development | NaN              | 3         | Life Sciences  | 1             | 1814           | ... | 1                        |     |
| 1026 | 22.0 | Yes       | Travel Frequently | 1256.0    | Research & Development | NaN              | 4         | Life Sciences  | 1             | 1203           | ... | 2                        |     |
| 1027 | 29.0 | No        | Travel Rarely     | 1378.0    | Research & Development | 13.0             | 2         | Other          | 1             | 2053           | ... | 1                        |     |
| 1028 | 50.0 | No        | Travel Rarely     | 264.0     | Sales                  | 9.0              | 3         | Marketing      | 1             | 1591           | ... | 3                        |     |

1029 rows × 35 columns

```
In [ ]: df1.columns
```

```
Out[ ]: 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')
```

```
In [ ]: df1.dtypes
```

```
Out[ ]: Age                float64
Attrition                object
BusinessTravel            object
DailyRate                float64
Department                object
DistanceFromHome          float64
Education                 int64
EducationField            object
EmployeeCount             int64
EmployeeNumber            int64
EnvironmentSatisfaction    int64
Gender                    object
HourlyRate                int64
JobInvolvement            int64
JobLevel                  int64
JobRole                   object
JobSatisfaction            int64
MaritalStatus             object
MonthlyIncome             int64
MonthlyRate               int64
NumCompaniesWorked        int64
Over18                    object
OverTime                  object
PercentSalaryHike         int64
PerformanceRating         int64
RelationshipSatisfaction   int64
StandardHours             int64
StockOptionLevel          int64
TotalWorkingYears         int64
TrainingTimesLastYear     int64
WorkLifeBalance           int64
YearsAtCompany            int64
YearsInCurrentRole        int64
YearsSinceLastPromotion   int64
YearsWithCurrManager      int64
dtype: object
```

```
In [ ]: df1.shape
```

```
In [ ]: df1.info()
```

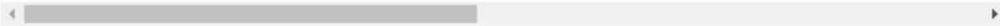
```
RangeIndex: 1029 entries, 0 to 1028
Data columns (total 35 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Age                    893 non-null    float64
 1   Attrition              1029 non-null   object
 2   BusinessTravel         1024 non-null   object
 3   DailyRate              1002 non-null   float64
 4   Department             1029 non-null   object
 5   DistanceFromHome       934 non-null    float64
 6   Education              1029 non-null   int64
 7   EducationField         1029 non-null   object
 8   EmployeeCount          1029 non-null   int64
 9   EmployeeNumber         1029 non-null   int64
10   EnvironmentSatisfaction 1029 non-null   int64
11   Gender                 1029 non-null   object
12   HourlyRate             1029 non-null   int64
13   JobInvolvement         1029 non-null   int64
14   JobLevel               1029 non-null   int64
15   JobRole                1029 non-null   object
16   JobSatisfaction        1029 non-null   int64
17   MaritalStatus          1024 non-null   object
18   MonthlyIncome          1029 non-null   int64
19   MonthlyRate            1029 non-null   int64
20   NumCompaniesWorked     1029 non-null   int64
21   Over18                 1029 non-null   object
22   OverTime               1029 non-null   object
23   PercentSalaryHike      1029 non-null   int64
24   PerformanceRating       1029 non-null   int64
25   RelationshipSatisfaction 1029 non-null   int64
26   StandardHours          1029 non-null   int64
27   StockOptionLevel       1029 non-null   int64
28   TotalWorkingYears      1029 non-null   int64
29   TrainingTimesLastYear  1029 non-null   int64
30   WorkLifeBalance        1029 non-null   int64
31   YearsAtCompany         1029 non-null   int64
32   YearsInCurrentRole     1029 non-null   int64
33   YearsSinceLastPromotion 1029 non-null   int64
34   YearsWithCurrManager   1029 non-null   int64
dtypes: float64(3), int64(23), object(9)
memory usage: 281.5+ KB
```

```
In [ ]: df1.describe()
```

Out[ ]:

|       | Age        | DailyRate   | DistanceFromHome | Education   | EmployeeCount | EmployeeNumber | EnvironmentSatisfaction | HourlyRate  | JobInvolvement | JobLevel    | ... | Relat |
|-------|------------|-------------|------------------|-------------|---------------|----------------|-------------------------|-------------|----------------|-------------|-----|-------|
| count | 893.000000 | 1002.000000 | 934.000000       | 1029.000000 | 1029.0        | 1029.000000    | 1029.000000             | 1029.000000 | 1029.000000    | 1029.000000 | ... | Relat |
| mean  | 37.930571  | 800.528942  | 9.930407         | 2.892128    | 1.0           | 1024.367347    | 2.683188                | 66.680272   | 2.713314       | 2.043732    | ... | Relat |
| std   | 9.395978   | 408.109828  | 8.421791         | 1.053541    | 0.0           | 606.301635     | 1.096829                | 20.474094   | 0.710146       | 1.118918    | ... | Relat |
| min   | 18.000000  | 102.000000  | 1.000000         | 1.000000    | 1.0           | 1.000000       | 1.000000                | 30.000000   | 1.000000       | 1.000000    | ... | Relat |
| 25%   | 31.000000  | 458.250000  | 2.000000         | 2.000000    | 1.0           | 496.000000     | 2.000000                | 48.000000   | 2.000000       | 1.000000    | ... | Relat |
| 50%   | 37.000000  | 801.500000  | 8.000000         | 3.000000    | 1.0           | 1019.000000    | 3.000000                | 67.000000   | 3.000000       | 2.000000    | ... | Relat |
| 75%   | 44.000000  | 1162.000000 | 16.000000        | 4.000000    | 1.0           | 1553.000000    | 4.000000                | 84.000000   | 3.000000       | 3.000000    | ... | Relat |
| max   | 60.000000  | 1496.000000 | 29.000000        | 5.000000    | 1.0           | 2068.000000    | 4.000000                | 100.000000  | 4.000000       | 5.000000    | ... | Relat |

8 rows × 26 columns



In [ ]:

```
df1.isnull().sum()
```

Out[ ]:

```
Age      136
Attrition      0
BusinessTravel      5
DailyRate      27
Department      0
DistanceFromHome    95
Education      0
EducationField      0
EmployeeCount      0
EmployeeNumber      0
EnvironmentSatisfaction  0
Gender      0
HourlyRate      0
JobInvolvement      0
JobLevel      0
JobRole      0
JobSatisfaction      0
MaritalStatus      5
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
```

In [ ]:

df1

Out[ ]:

|      | Age  | Attrition | BusinessTravel    | DailyRate | Department             | DistanceFromHome | Education | EducationField | EmployeeCount | EmployeeNumber | ... | RelationshipSatisfaction | Sta |
|------|------|-----------|-------------------|-----------|------------------------|------------------|-----------|----------------|---------------|----------------|-----|--------------------------|-----|
| 0    | 50.0 | No        | Travel Rarely     | 1126.0    | Research & Development | 1.0              | 2         | Medical        | 1             | 997            | ... | 3                        |     |
| 1    | 36.0 | No        | Travel Rarely     | 216.0     | Research & Development | 6.0              | 2         | Medical        | 1             | 178            | ... | 4                        |     |
| 2    | 21.0 | Yes       | Travel Rarely     | 337.0     | Sales                  | 7.0              | 1         | Marketing      | 1             | 1780           | ... | 2                        |     |
| 3    | 50.0 | No        | Travel Frequently | 1246.0    | Human Resources        | NaN              | 3         | Medical        | 1             | 644            | ... | 3                        |     |
| 4    | 52.0 | No        | Travel Rarely     | 994.0     | Research & Development | 7.0              | 4         | Life Sciences  | 1             | 1118           | ... | 4                        |     |
| ...  | ...  | ...       | ...               | ...       | ...                    | ...              | ...       | ...            | ...           | ...            | ... | ...                      | ... |
| 1024 | NaN  | No        | Travel Rarely     | 750.0     | Research & Development | 28.0             | 3         | Life Sciences  | 1             | 1596           | ... | 4                        |     |
| 1025 | 41.0 | No        | Travel Rarely     | 447.0     | Research & Development | NaN              | 3         | Life Sciences  | 1             | 1814           | ... | 1                        |     |
| 1026 | 22.0 | Yes       | Travel Frequently | 1256.0    | Research & Development | NaN              | 4         | Life Sciences  | 1             | 1203           | ... | 2                        |     |
| 1027 | 29.0 | No        | Travel Rarely     | 1378.0    | Research & Development | 13.0             | 2         | Other          | 1             | 2053           | ... | 1                        |     |
| 1028 | 50.0 | No        | Travel Rarely     | 264.0     | Sales                  | 9.0              | 3         | Marketing      | 1             | 1591           | ... | 3                        |     |

1029 rows x 35 columns

4