Corporate Employee Attrition Analysis

A PROJECT REPORT

Submitted By

Team ID : PNT2022TMID34476

Team Leader : SANKAR RAJA J I (961819104076)

Team Member : JAISON V (961819104038)

Team Member : PASUPATHISH M (961819104061)

Team Member : MATHAN M (961819104055)

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

PONJESLY COLLEGE OF ENGINEERING, NAGERCOIL

ANNA UNIVERSITY::CHENNAI 600025

TABLE OF CONTENTS

1. INTRODUCTION

1.1 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.

1.2 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

2.1 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.

2.2 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)

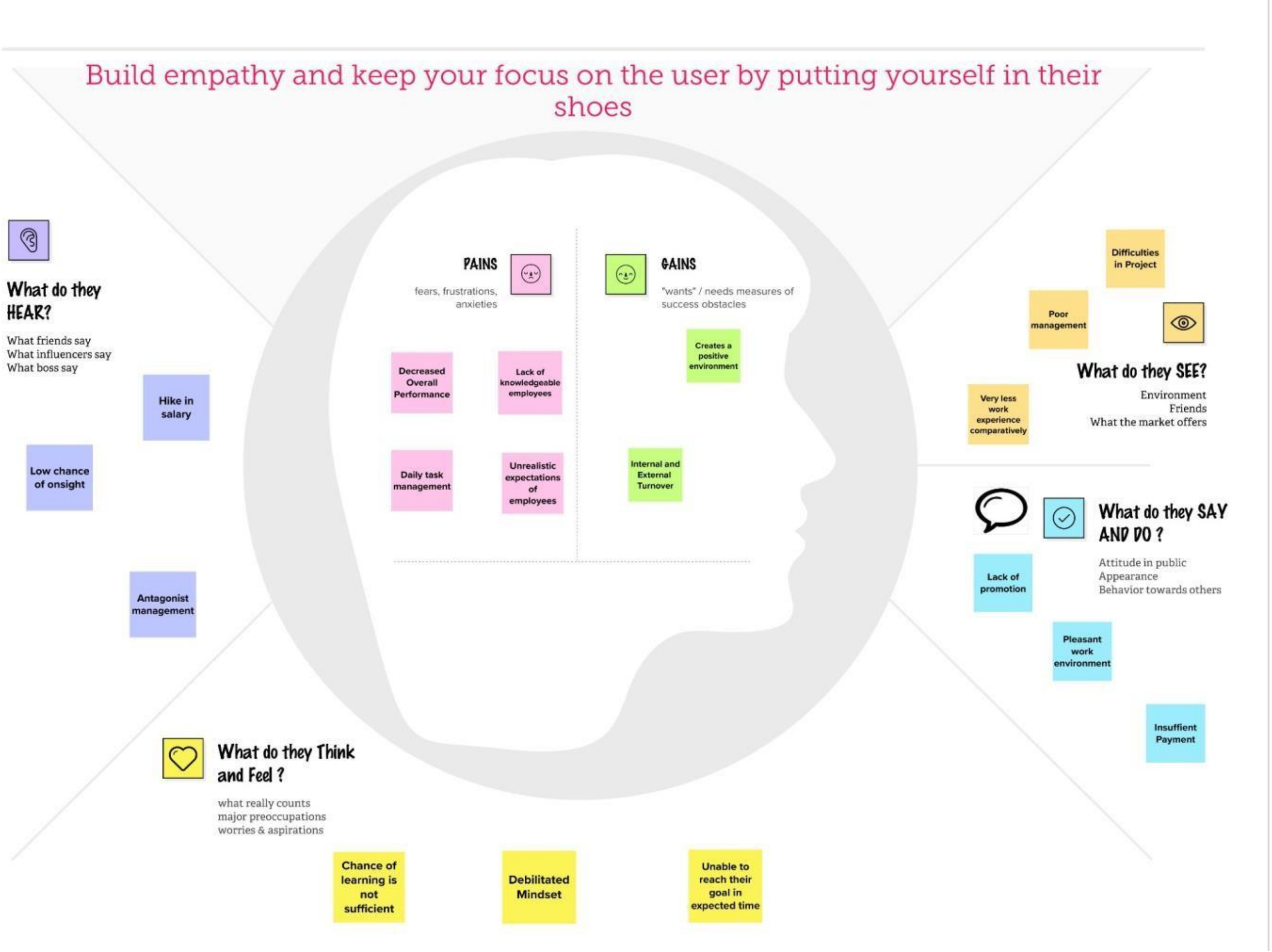
2.3 Problem Statement Definition

- 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,

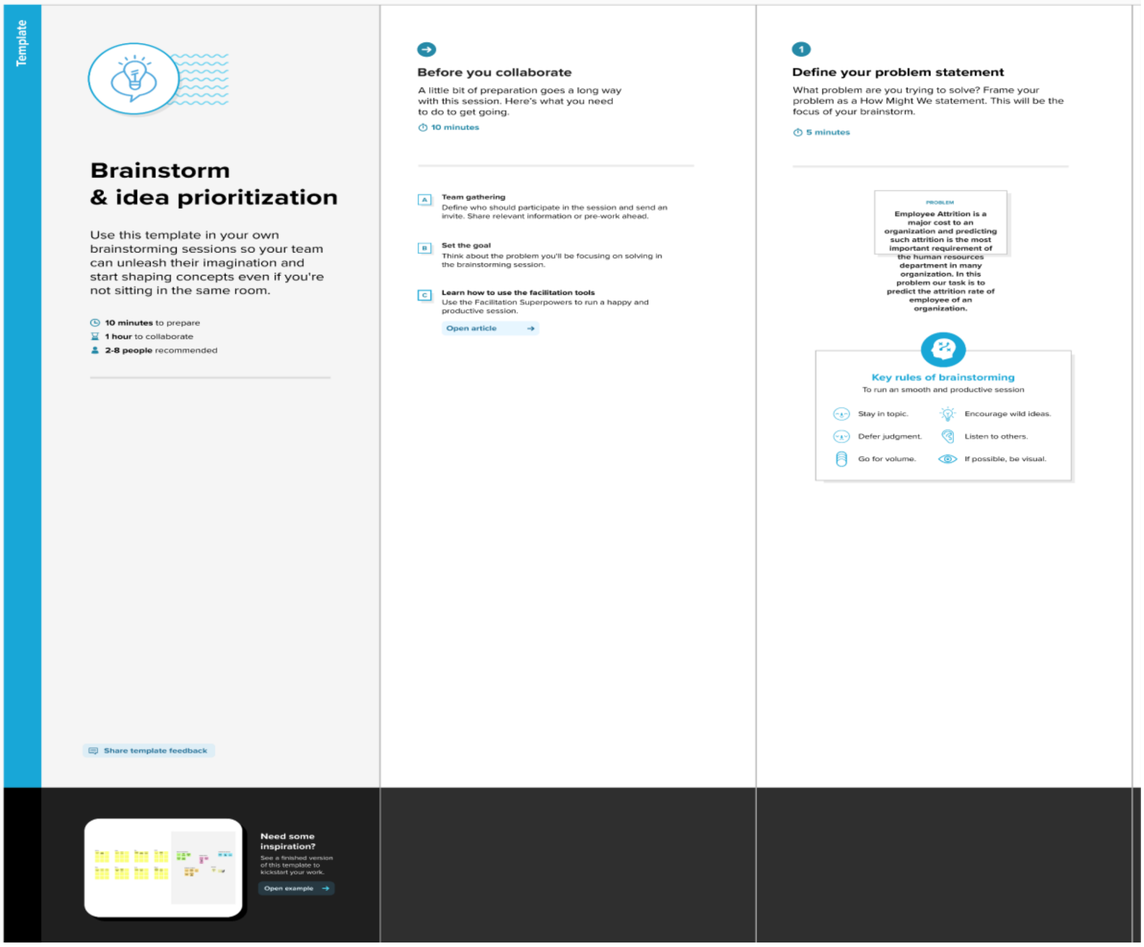
data visualization by analysing the major factors that causes attrition.

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas


Build empathy and keep your focus on the user by putting yourself in their shoes
What do they HEAR?
What friends say
What influencers say
What boss say
Hike in salary
Low chance of onsight
Antagonist management
PAINS
fears, frustrations, anxieties
Decreased Overall Performance
Lack of knowledgeable employees
Daily task management
Unrealistic expectations of employees
GAINS
"wants" / needs measures of success obstacles
Creates a positive environment
Internal and External Turnover
Difficulties in Project
Poor management
What do they SEE?
Environment
Friends
What the market offers
Very less work experience comparatively
What do they SAY AND DO ?
Attitude in public
Appearance
Behavior towards others
Lack of promotion
Pleasant work environment
Insuffient Payment
What do they Think and Feel ?
what really counts
major preoccupations
worries & aspirations
Chance of learning is not sufficient
Debilitated Mindset
Unable to reach their goal in expected time

3.2 Ideation & Brainstorming



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

Need some inspiration?

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

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.

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

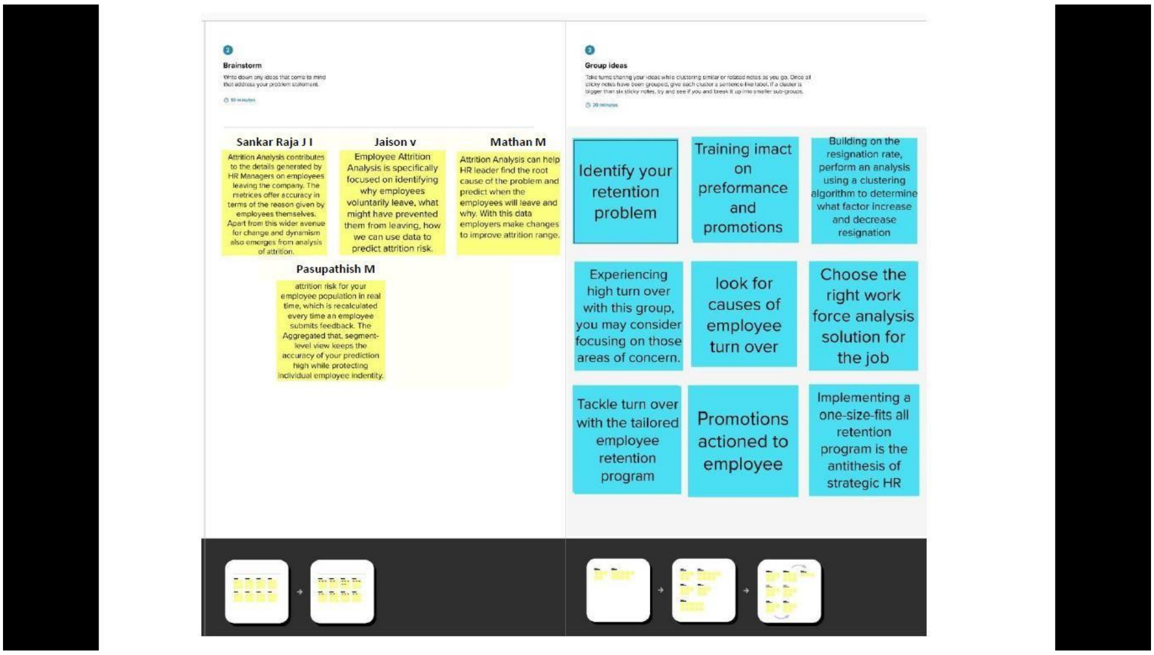
PROBLEM

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 organization. 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.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.



2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

Sankar Raja J I

Attrition Analysis contributes to the details generated by HR Managers on employees leaving the company. The metrices offer accuracy in terms of the reason given by employees themselves. Apart from this wider avenue for change and dynamism also emerges from analysis of attrition.

Jaison v

Employee Attrition Analysis is specifically focused on identifying why employees voluntarily leave, what might have prevented them from leaving, how we can use data to predict attrition risk.

Mathan M

Attrition Analysis can help HR leader find the root cause of the problem and predict when the employees will leave and why. With this data employers make changes to improve attrition range.

Pasupathish M

attrition risk for your employee population in real time, which is recalculated every time an employee submits feedback. The Aggregated that, segment-level view keeps the accuracy of your prediction high while protecting individual employee indentity.

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

Identify your retention problem

Training imact on preformance and promotions

Building on the resignation rate, perform an analysis using a clustering algorithm to determine what factor increase and decrease resignation

Experiencing high turn over with this group, you may consider focusing on those areas of concern.

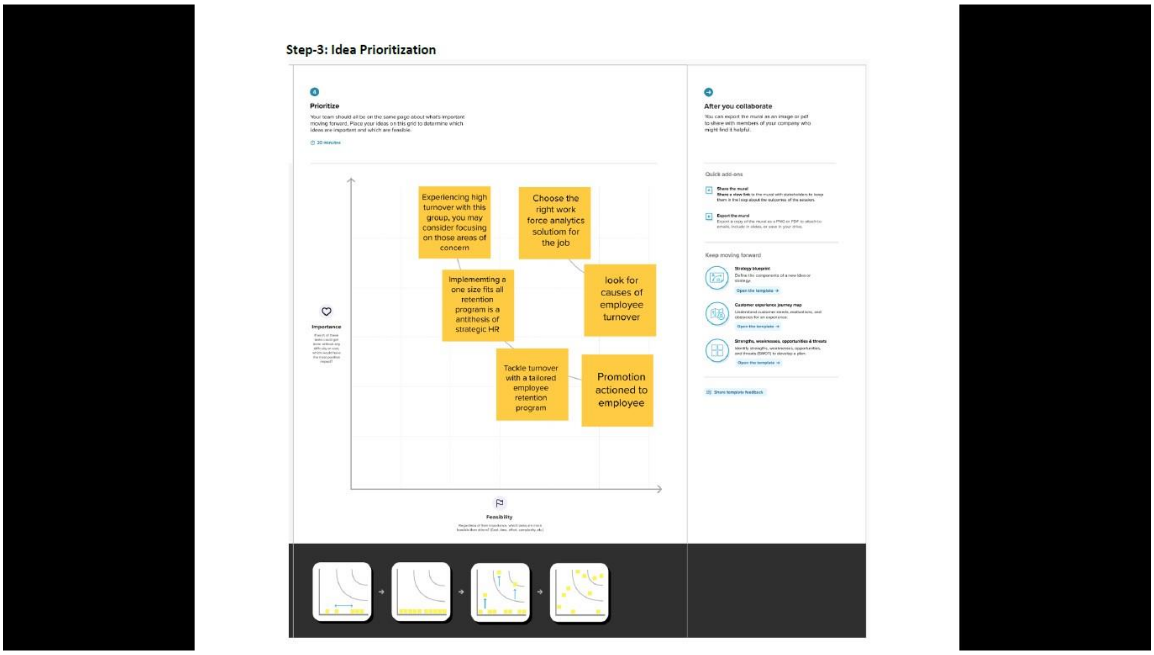
look for causes of employee turn over

Choose the right work force analysis solution for the job

Tackle turn over with the tailored employee retention program

Promotions actioned to employee

Implementing a one-size-fits all retention program is the antithesis of strategic HR



Step-3: Idea Prioritization

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

Experiencing high turnover with this group, you may consider focusing on those areas of concern

Choose the right work force analytics solutiom for the job

Implememting a one size fits all retention program is a antithesis of strategic HR

look for causes of employee turnover

Tackle turnover with a tailored employee retention program

Promotion actioned to employee

After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template →

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

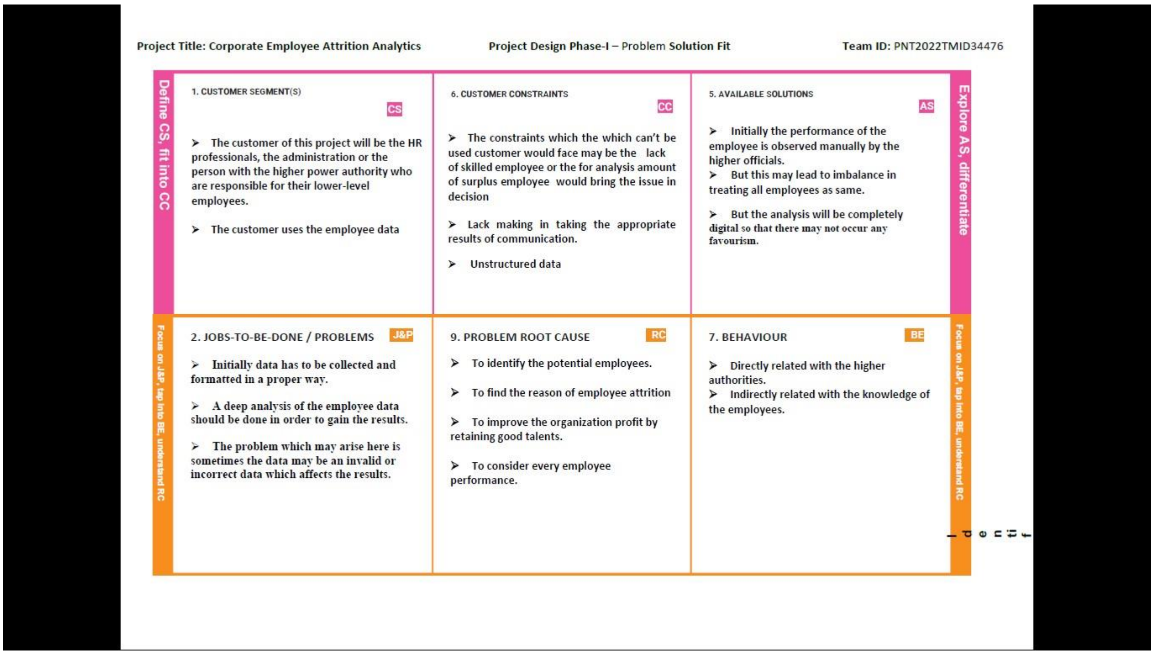
Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

Share template feedback

3.3 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.

3.4 Problem Solution fit



Project Title: Corporate Employee Attrition Analytics

Project Design Phase-I – Problem Solution Fit

Team ID: PNT2022TMID34476

1. CUSTOMER SEGMENT(S) CS

- 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.
- The customer uses the employee data

6. CUSTOMER CONSTRAINTS CC

- 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
- Lack making in taking the appropriate results of communication.
- Unstructured data

5. AVAILABLE SOLUTIONS AS

- Initially the performance of the employee is observed manually by the higher officials.
- But this may lead to imbalance in treating all employees as same.
- But the analysis will be completely digital so that there may not occur any favourism.

2. JOBS-TO-BE-DONE / PROBLEMS J&P

- Initially data has to be collected and formatted in a proper way.
- A deep analysis of the employee data should be done in order to gain the results.
- The problem which may arise here is sometimes the data may be an invalid or incorrect data which affects the results.

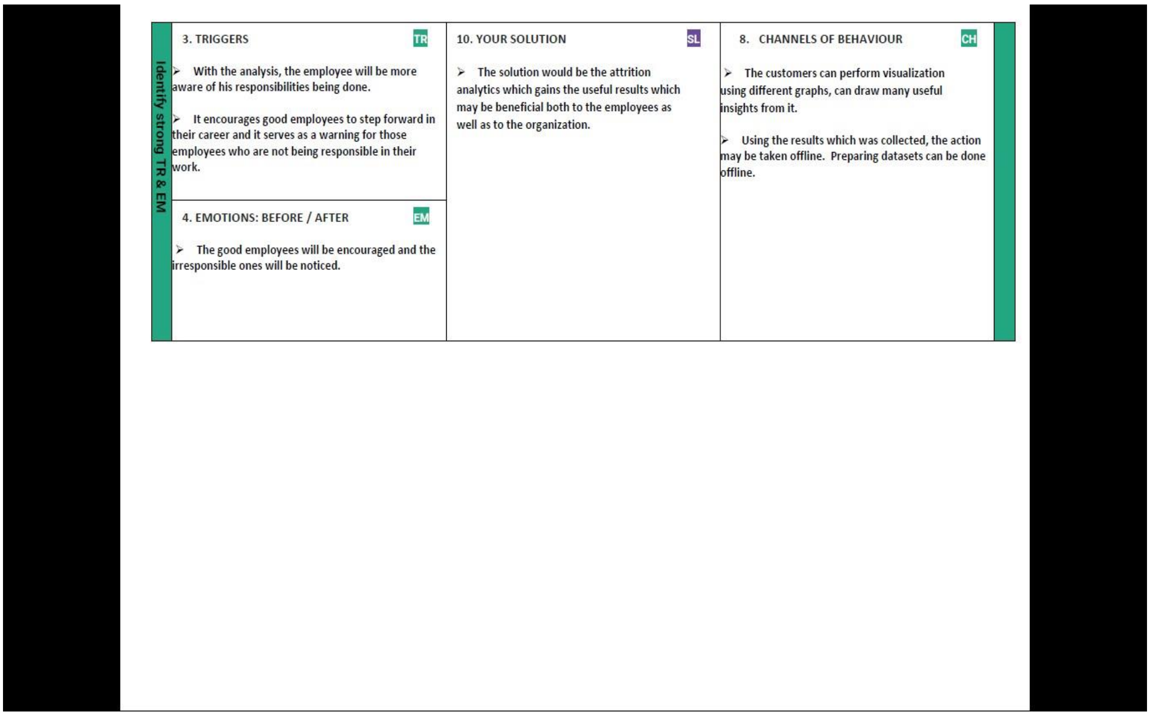
9. PROBLEM ROOT CAUSE RC

- To identify the potential employees.
- To find the reason of employee attrition
- To improve the organization profit by retaining good talents.
- To consider every employee performance.

7. BEHAVIOUR BE

- Directly related with the higher authorities.
- Indirectly related with the knowledge of the employees.

Define CS, fit into CC
Explore AS, differentiate
Focus on J&P, tap into BE, understand RC
Focus on J&P, tap into BE, understand RC
Identif



3. TRIGGERS TR

- With the analysis, the employee will be more aware of his responsibilities being done.
- 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.

4. EMOTIONS: BEFORE / AFTER EM

- The good employees will be encouraged and the irresponsible ones will be noticed.

10. YOUR SOLUTION SL

- 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.

8. CHANNELS OF BEHAVIOUR CH

- The customers can perform visualization using different graphs, can draw many useful insights from it.
- Using the results which was collected, the action may be taken offline. Preparing datasets can be done offline.

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Feedback	Feedback through Form Feedback through Gmail Feedback through Instagram polls Feedback through LinkedIn
FR-4	User Rating	Rating via Mail 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

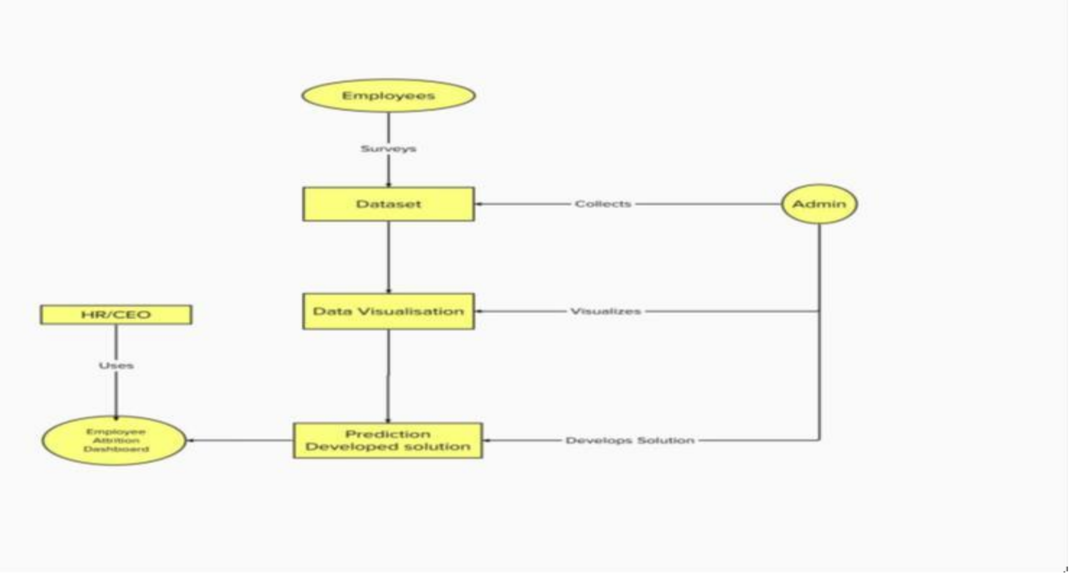
4.2 Non-Functional requirements

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

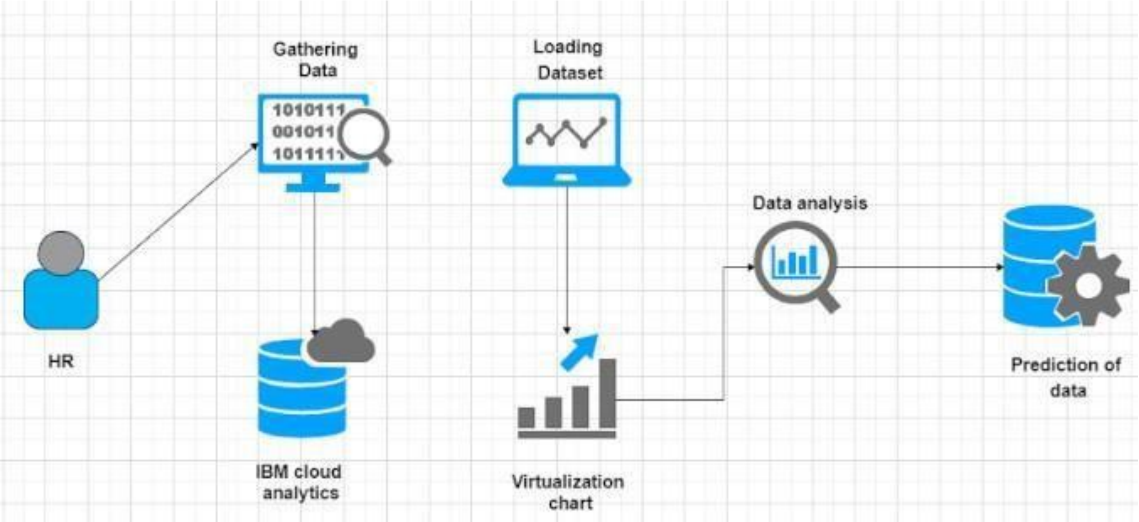
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This Data Visualization shall be easy to use for all users with minimal instructions. 100% of the languages on the graphical user interface (GUI) shall be intuitive and understandable by non-technical users.
NFR-2	**Security**	The employee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

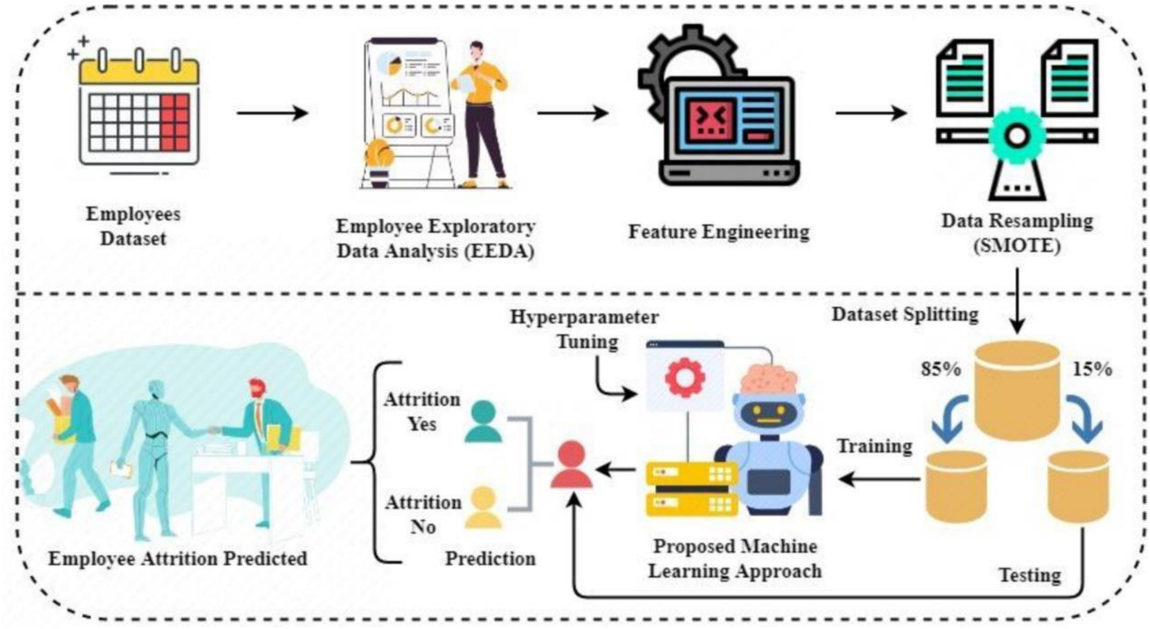
5. PROJECT DESIGN

5.1 Data Flow Diagrams

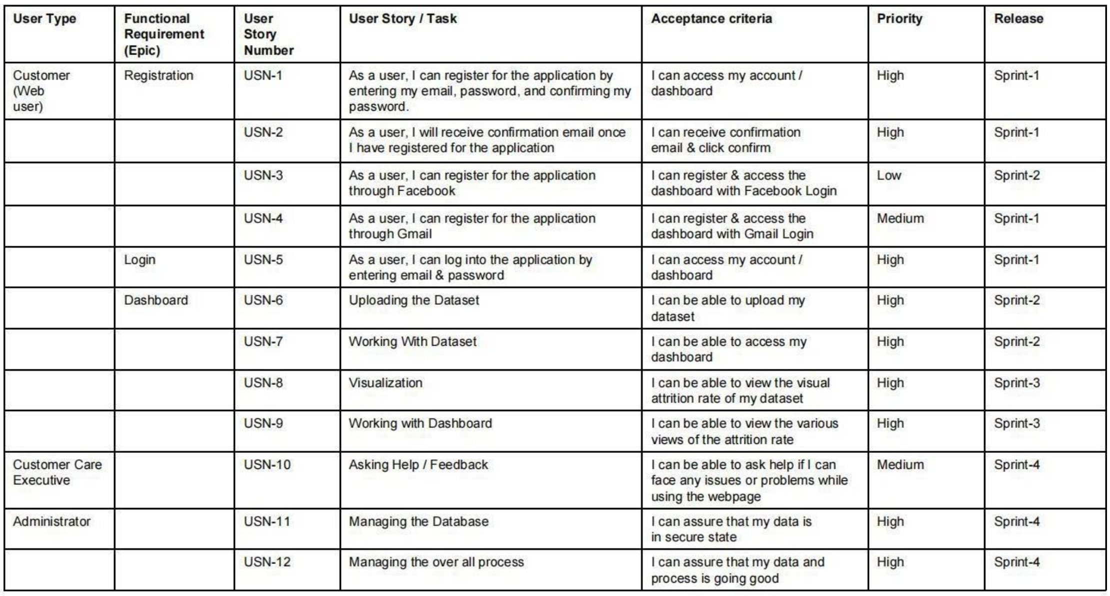


5.2 Solution & Technical Architecture





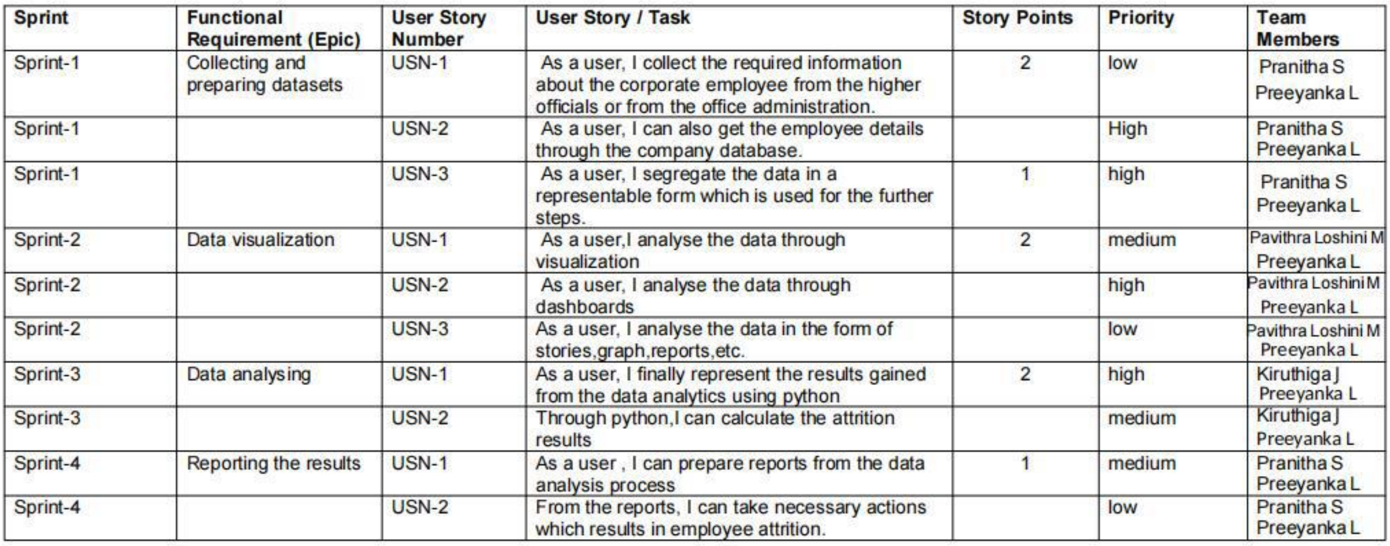
5.3 User Stories



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

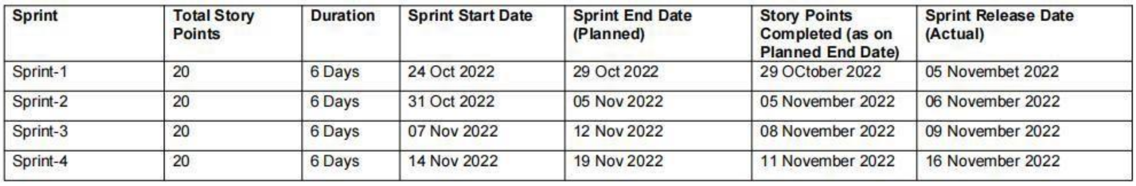
6. PROJECT PLANNING

6.1 Sprint Planning & Estimation



Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Collecting and preparing datasets	USN-1	As a user, I collect the required information about the corporate employee from the higher officials or from the office administration.	2	low	Pranitha S Preeyanka L
Sprint-1		USN-2	As a user, I can also get the employee details through the company database.		High	Pranitha S Preeyanka L
Sprint-1		USN-3	As a user, I segregate the data in a representable form which is used for the further steps.	1	high	Pranitha S Preeyanka L
Sprint-2	Data visualization	USN-1	As a user,I analyse the data through visualization	2	medium	Pavithra Loshini M Preeyanka L
Sprint-2		USN-2	As a user, I analyse the data through dashboards		high	Pavithra Loshini M Preeyanka L
Sprint-2		USN-3	As a user, I analyse the data in the form of stories,graph,reports,etc.		low	Pavithra Loshini M Preeyanka L
Sprint-3	Data analysing	USN-1	As a user, I finally represent the results gained from the data analytics using python	2	high	Kiruthiga J Preeyanka L
Sprint-3		USN-2	Through python,I can calculate the attrition results		medium	Kiruthiga J Preeyanka L
Sprint-4	Reporting the results	USN-1	As a user , I can prepare reports from the data analysis process	1	medium	Pranitha S Preeyanka L
Sprint-4		USN-2	From the reports, I can take necessary actions which results in employee attrition.		low	Pranitha S Preeyanka L

6.2 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 Novembet 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

7.1 Feature 1

```
#GENERAL
import pandas as pd
import numpy as np
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 =
'%.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')
```

7.2 Feature 2

```
#GENERAL
import pandas as pd
import numpy as np
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 =
'%.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

8.1 Test Cases

8.2 User Acceptance Testing

1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issue of corporateemployee 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

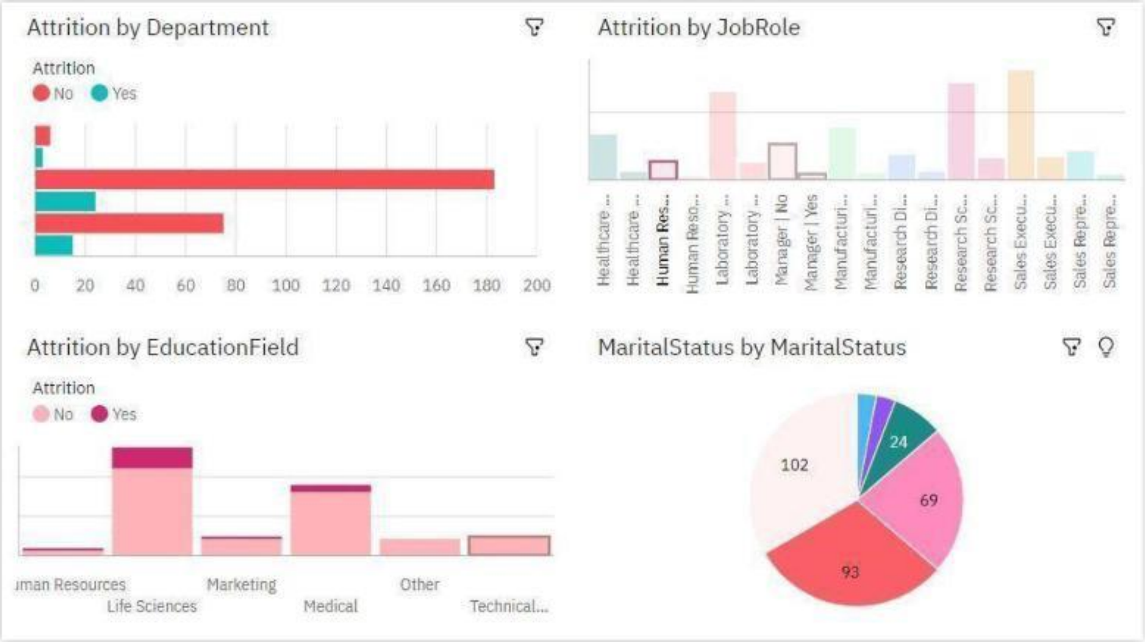
3. Test Case Analysis

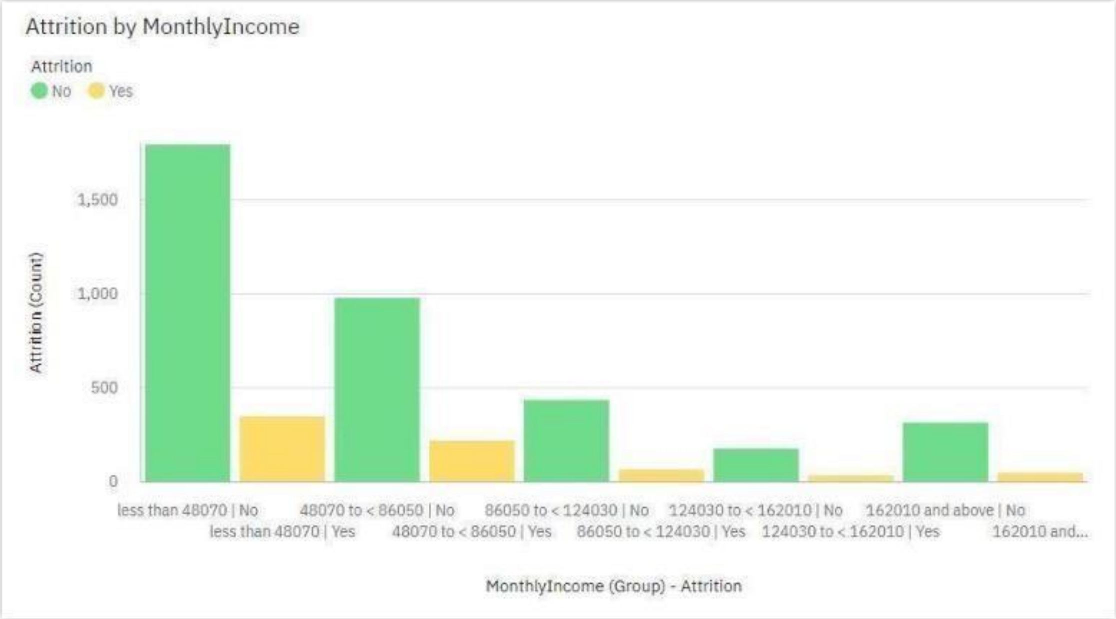
Database	2	0	0	2
Dashboard	1	0	0	1
Visualize the data	8	0	0	8
Logistic Regression	4	0	0	4

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

9.1Advantages

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

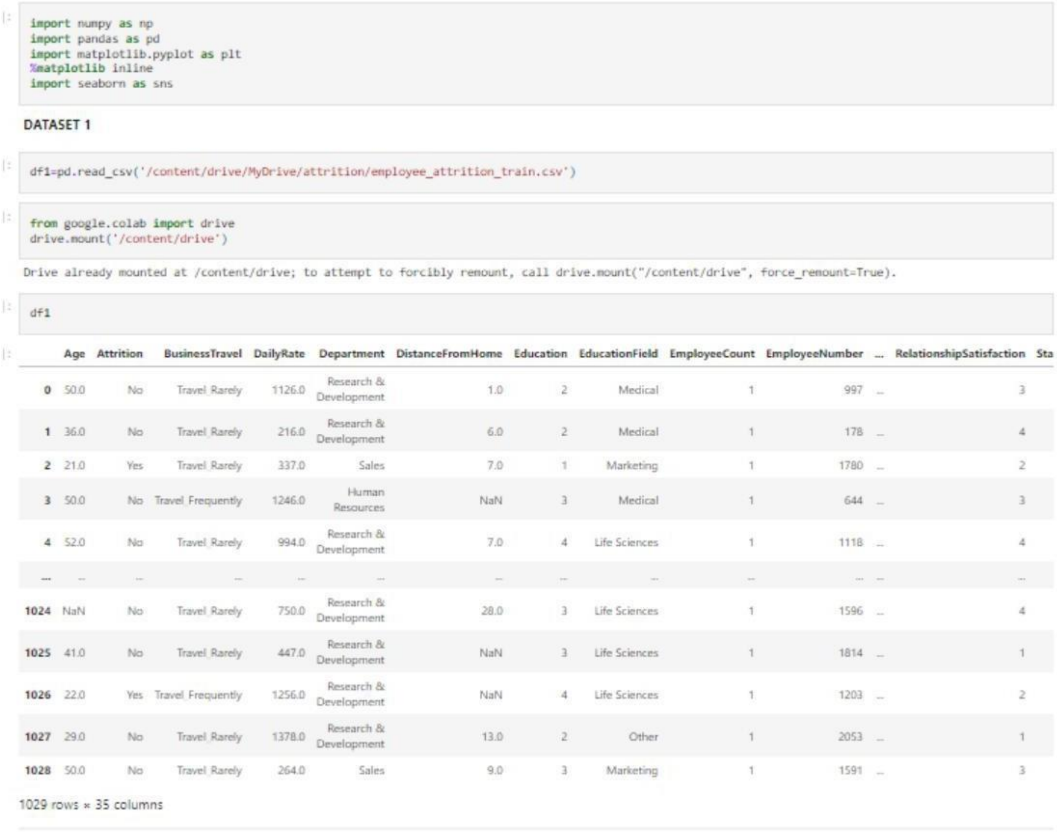
Employees as well as organizations must be clear with their expectations regarding the job profile. Any sort of mismatch leads to discrepancy and employees may fail to perform at theirjob. This eventually leads to attrition. Organizations should state the requirements and expectations unambiguously. 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 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')
```

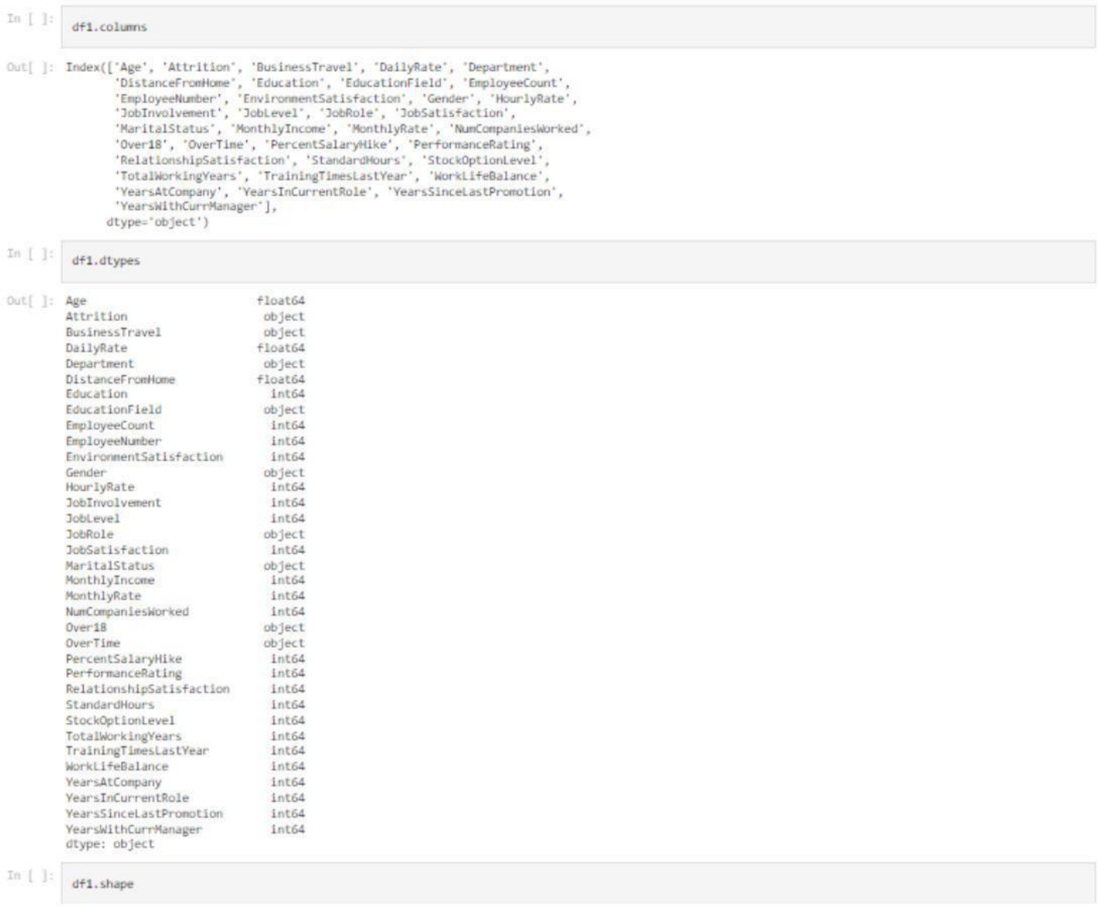
```
from google.colab import drive
drive.mount('/content/drive')
```

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

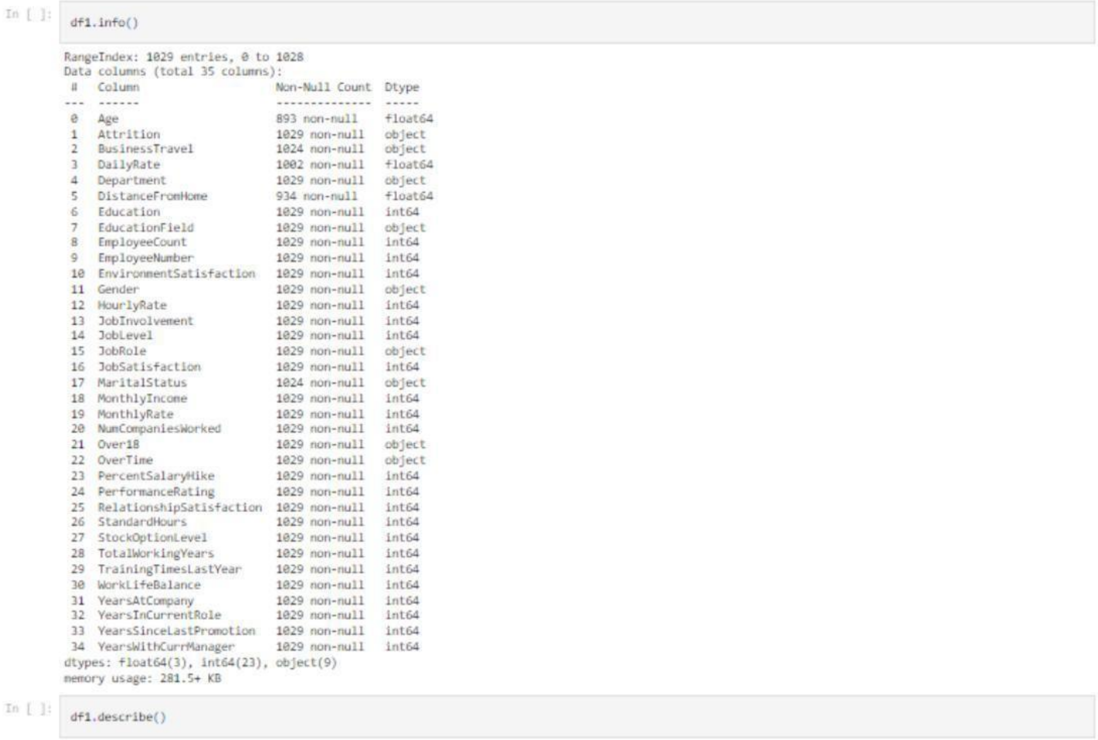
```
df1
```

	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	

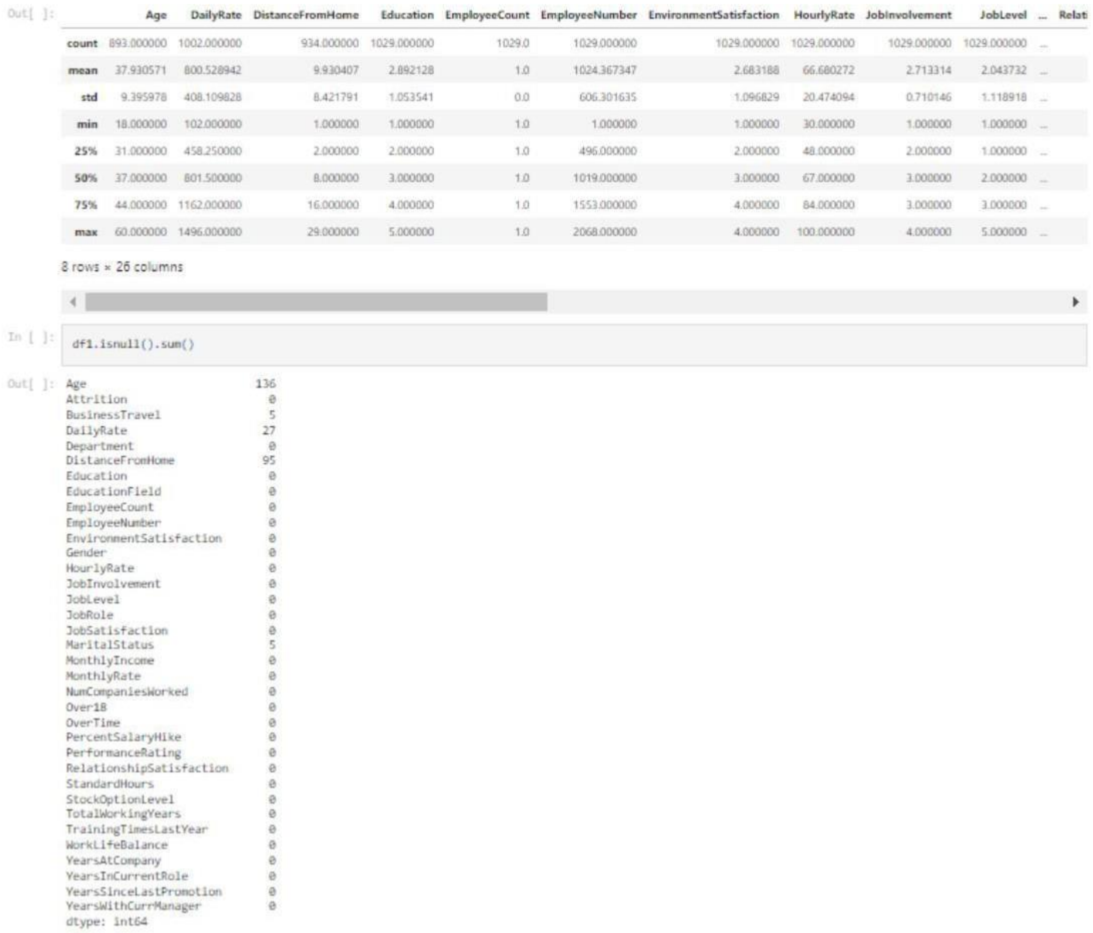
1029 rows × 35 columns



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

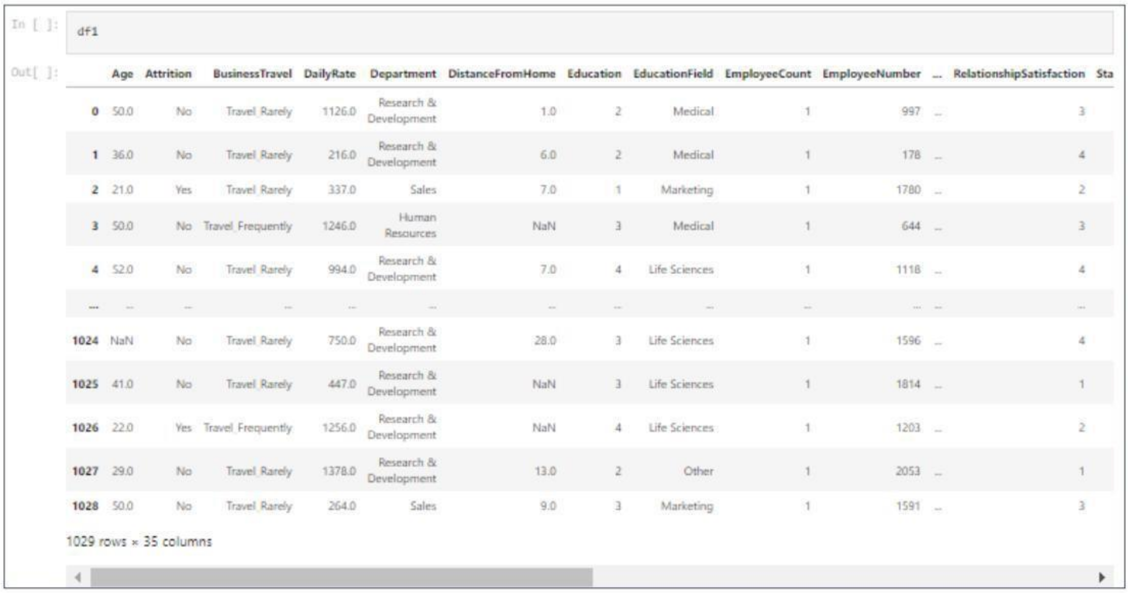
8 rows × 26 columns

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 × 35 columns