# 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**

# CHAPTER NO TITLE PAGE NO

1. [INTRODUCTION 4](#_TOC_250023)
   1. [Project Overview **4**](#_TOC_250022)
   2. [Purpose **4**](#_TOC_250021)
2. LITERATURE SURVEY 5
   1. [Existing problem **5**](#_TOC_250020)
   2. [References **5**](#_TOC_250019)
   3. [Problem Statement Definition **5**](#_TOC_250018)
3. IDEATION & PROPOSED SOLUTION
   1. [Empathy Map Canvas **6**](#_TOC_250017)
   2. [Ideation & Brainstorming **6**](#_TOC_250016)
   3. [Proposed Solution **8**](#_TOC_250015)
   4. [Problem Solution fit **9**](#_TOC_250014)
4. [REQUIREMENT ANALYSIS](#_TOC_250013)
   1. [Functional requirement **10**](#_TOC_250012)
   2. [Non-Functional requirements **10**](#_TOC_250011)
5. [PROJECT DESIGN](#_TOC_250010)
   1. [Data Flow Diagrams **11**](#_TOC_250009)
   2. [Solution & Technical Architecture **11**](#_TOC_250008)
   3. [User Stories **12**](#_TOC_250007)
6. PROJECT PLANNING & SCHEDULING
   1. [Sprint Planning & Estimation **13**](#_TOC_250006)
   2. [Sprint Delivery Schedule **13**](#_TOC_250005)
   3. Reports from JIRA
7. [CODING & SOLUTIONING](#_TOC_250004)
   1. Feature 1 **13**
   2. Feature 2 **14**
   3. Database Schema
8. [TESTING](#_TOC_250003)
   1. [Test Cases **15**](#_TOC_250002)
   2. [User Acceptance Testing **15**](#_TOC_250001)
9. [RESULTS 17](#_TOC_250000)
10. ADVANTAGES & DISADVANTAGES 18
11. CONCLUSION 18
12. FUTURE SCOPE 19
13. APPENDIX 20

# 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**.**

# 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)

* + 1. Investigation of early career teacher attrition(ECT) and the impact of induction programs in Western Australia, Janine E.Wyatt, MichaelO’Neill (2021)

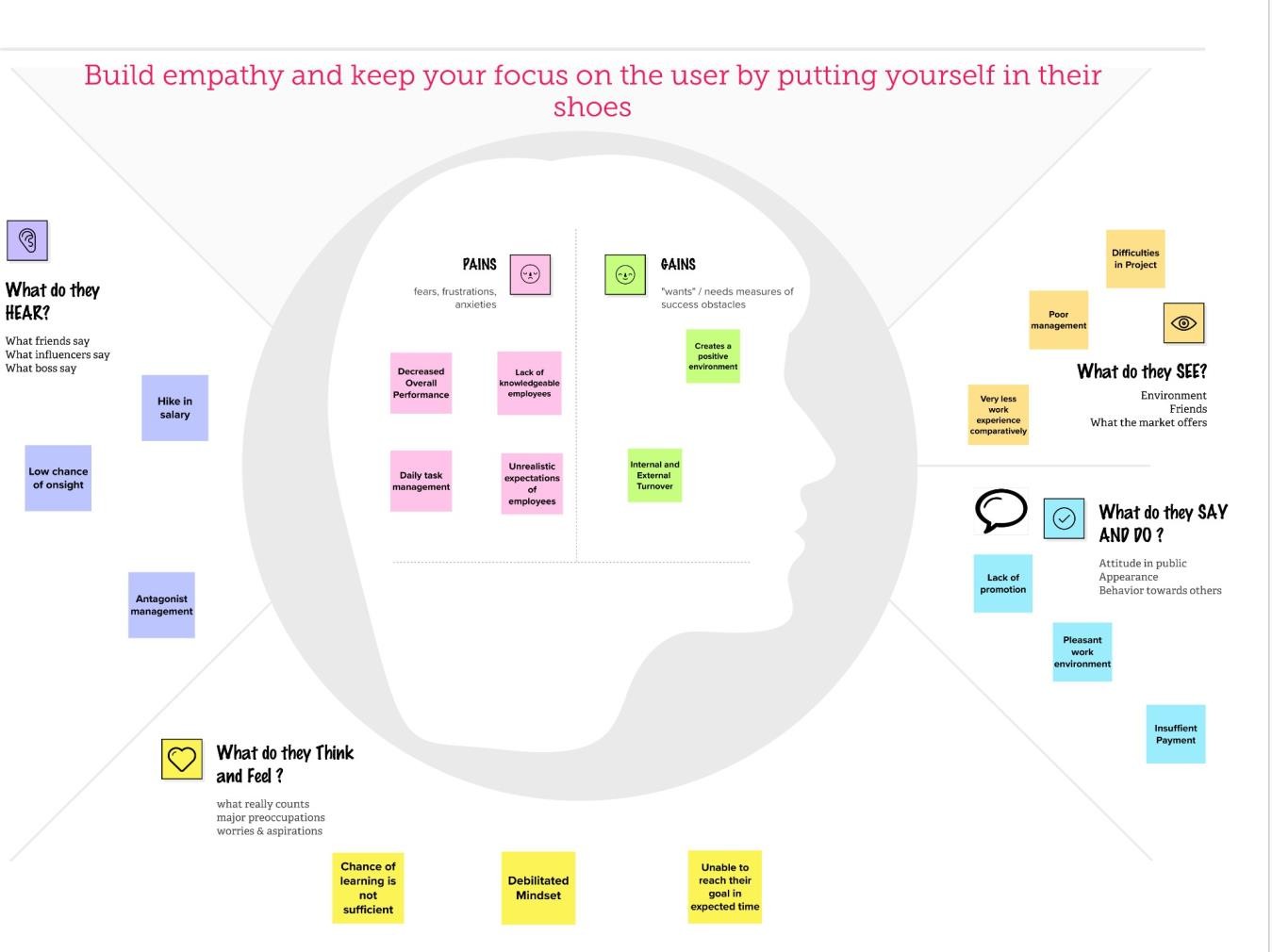
# Problem Statement Definition

* 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,

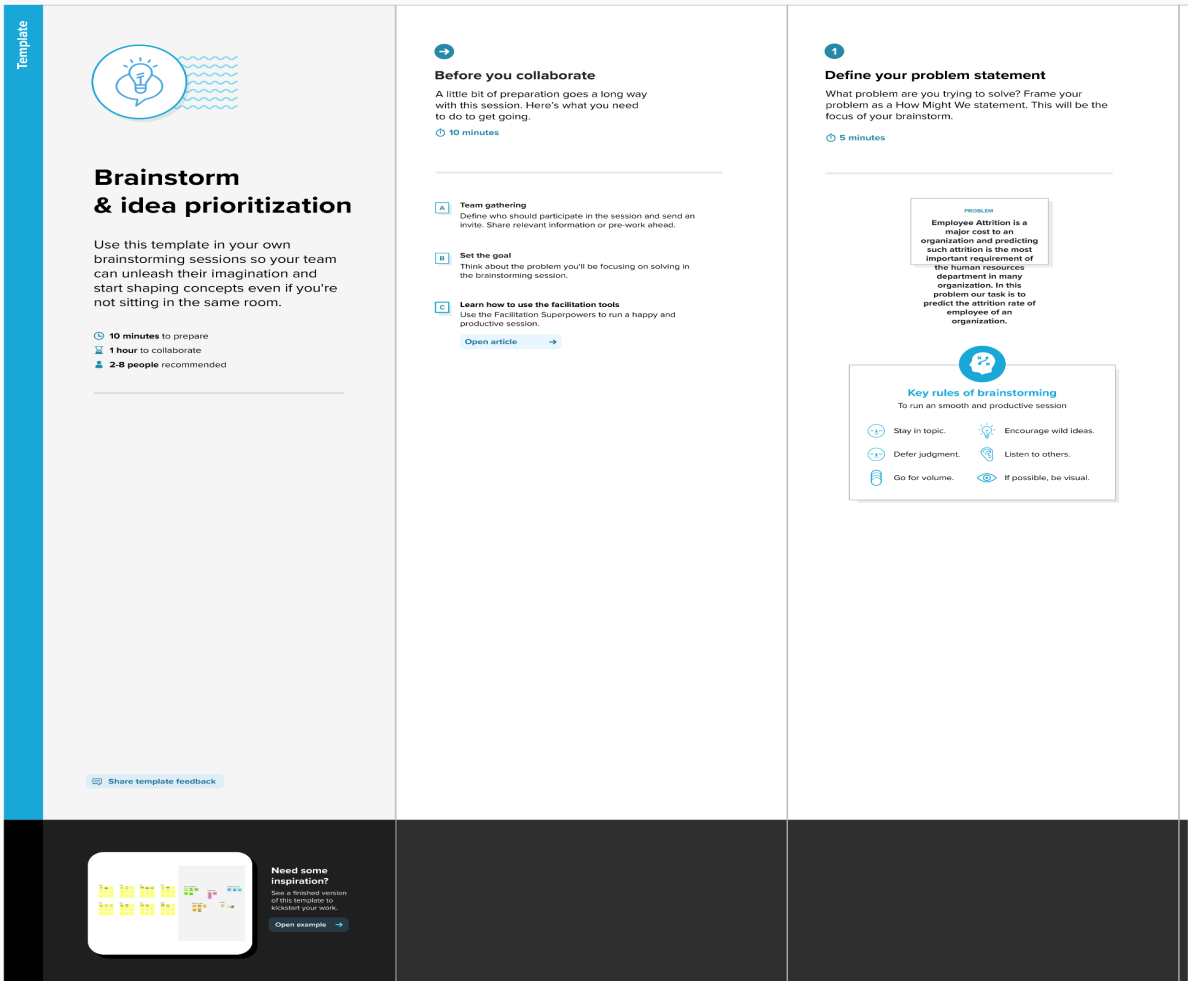
data visualization by analysing the major factors that causes attrition.

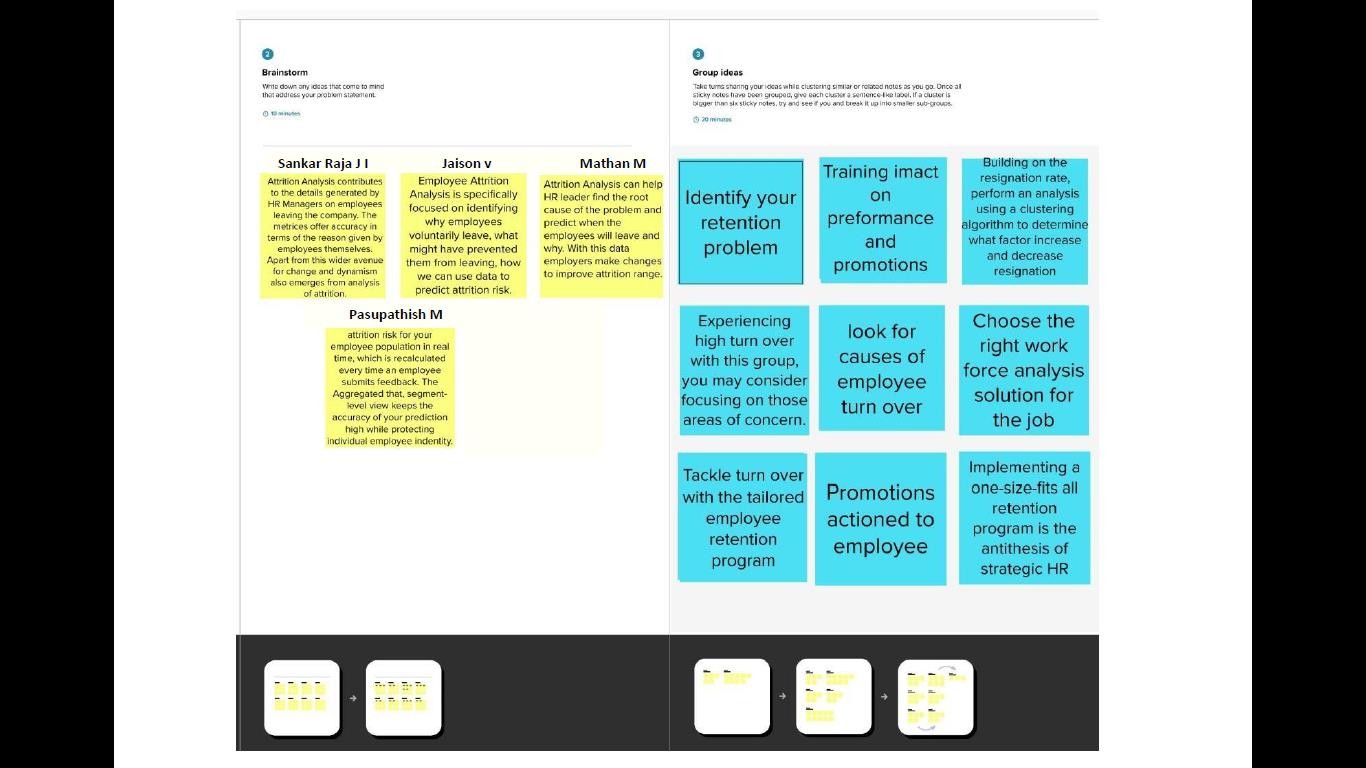
# IDEATION AND PROPOSED SOLUTION

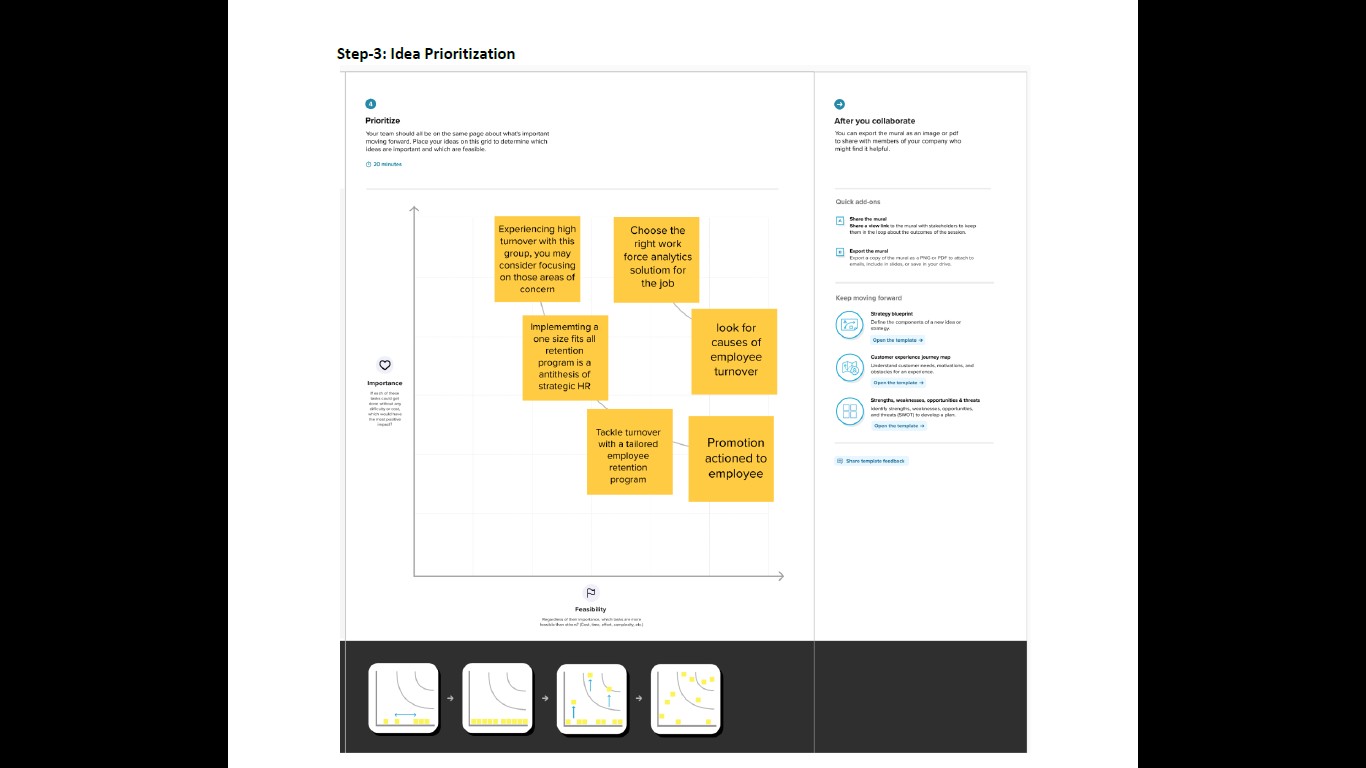
# Empathy Map Canvas



# Ideation & Brainstorming





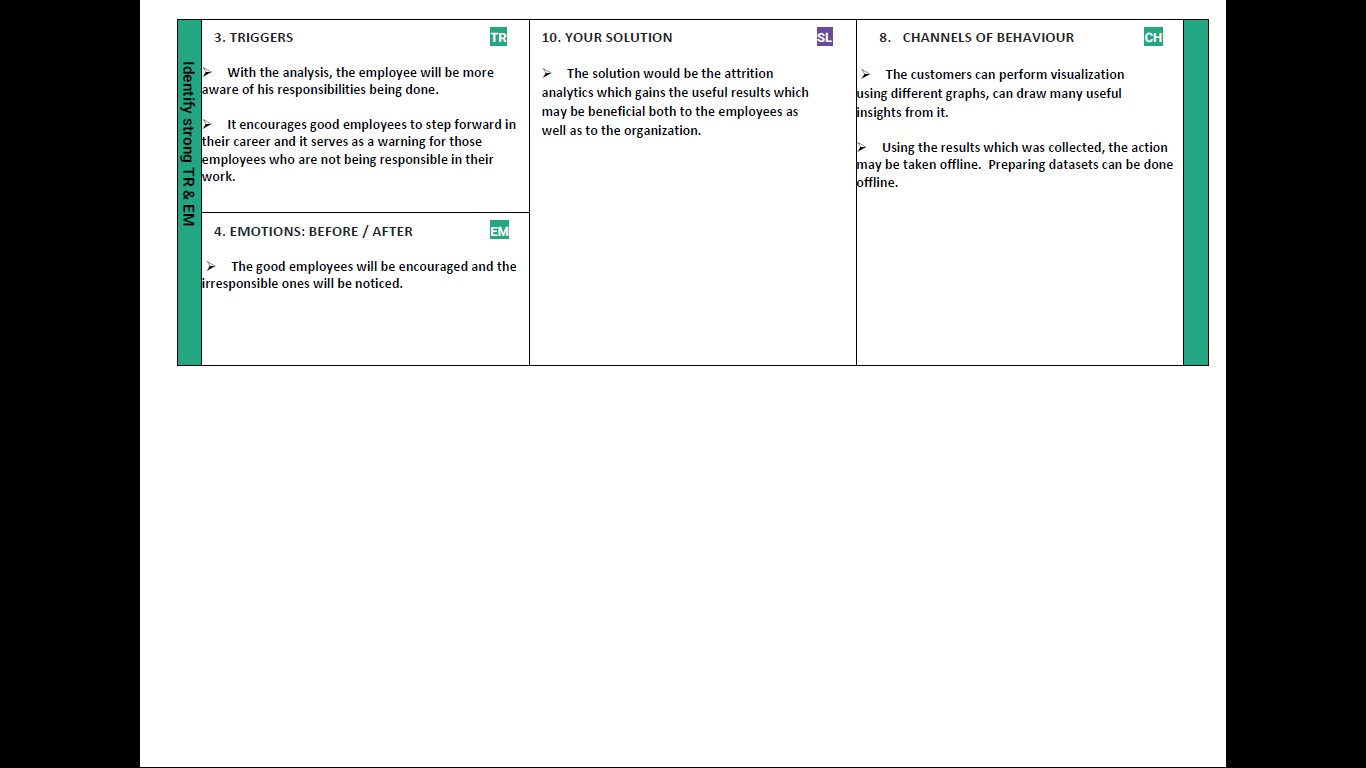
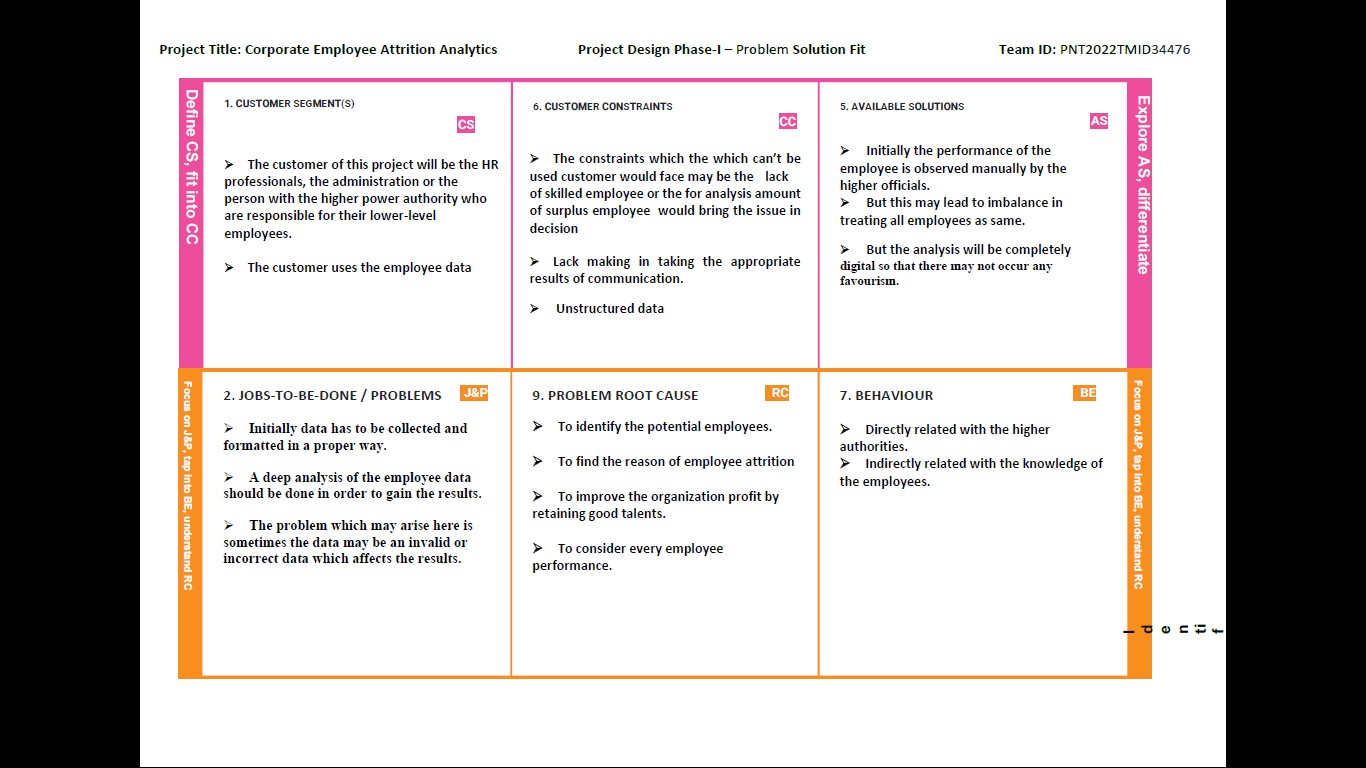


# Proposed Solution

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



# REQUIREMENT ANALYSIS

# 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 |

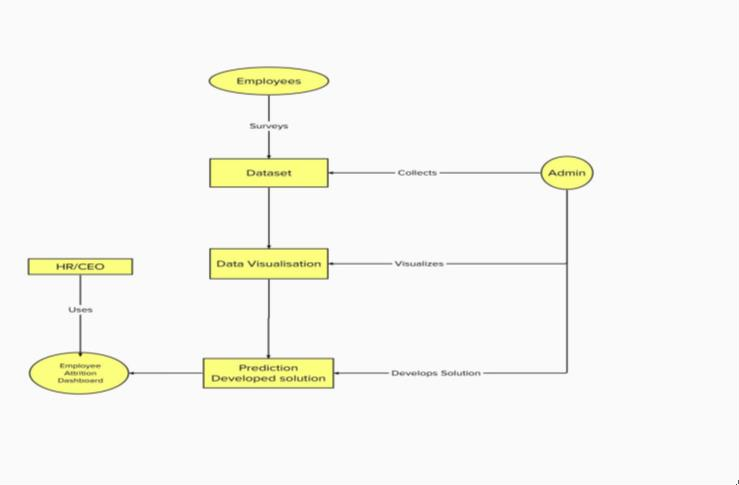
# Non-Functional requirements

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

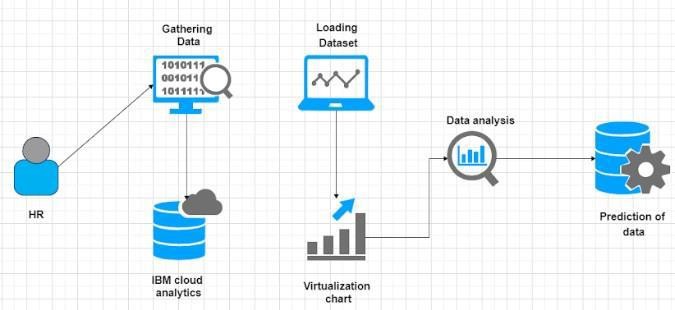
|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | **Usability** | This Data Visualization shall be easy to use for all users with minimal instructions. 100% of the languages on the graphical user interface (GUI) shall be intuitive and understandable 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. |

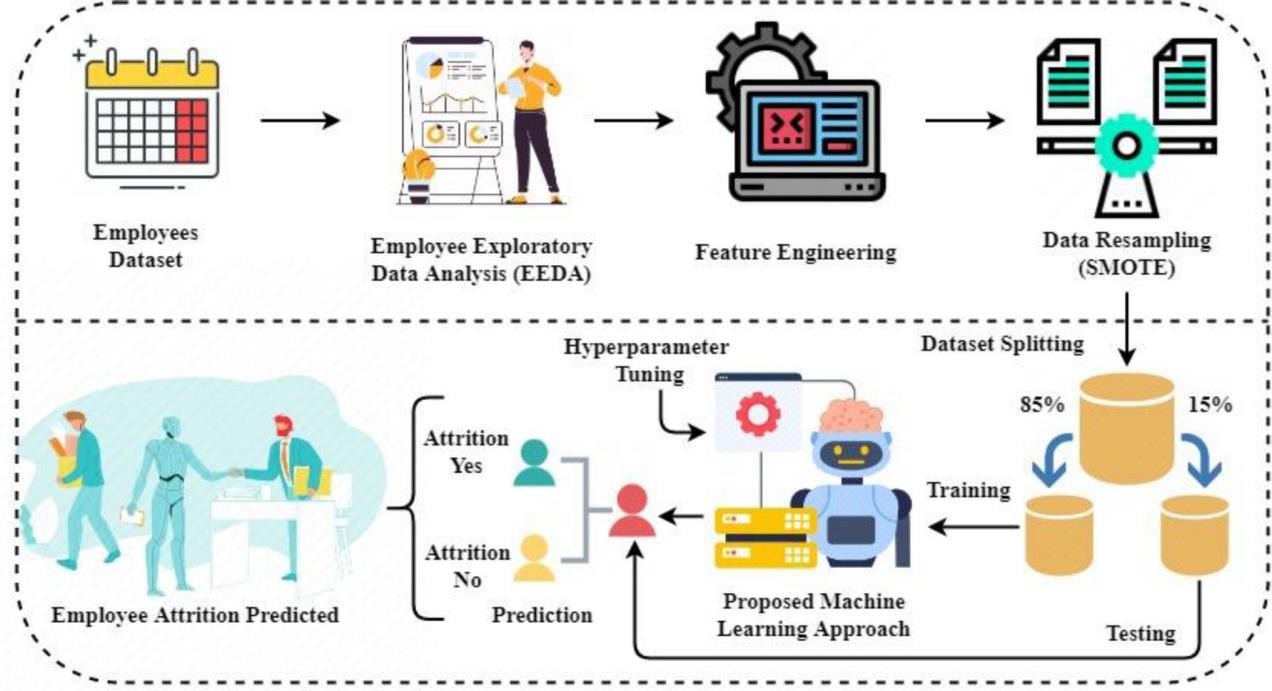
# PROJECT DESIGN

# Data Flow Diagrams

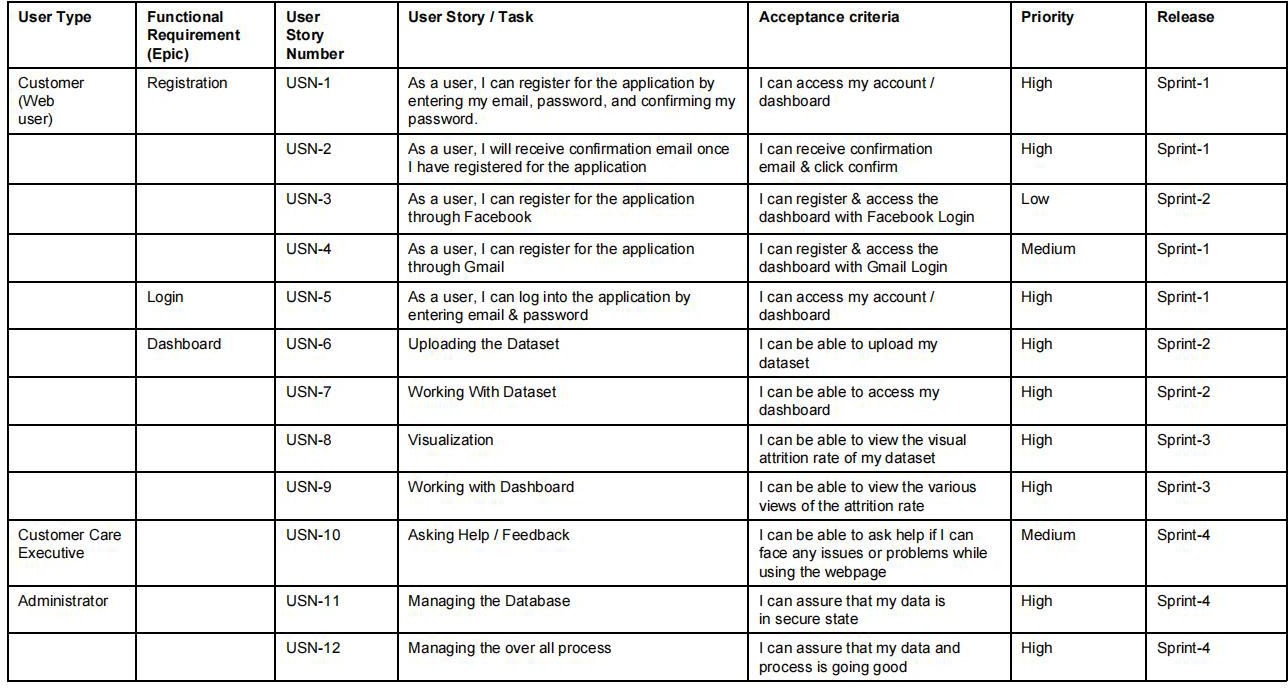


# Solution & Technical Architecture





# User Stories



# PROJECT PLANNING

# Sprint Planning & Estimation

****

# Sprint Delivery Schedule



# CODING & SOLUTIONING

* 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')

* 1. **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')

# 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 corporateemployee attrition at the time of the release.

* + 1. **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 |

* + 1. **Test Case Analysis**

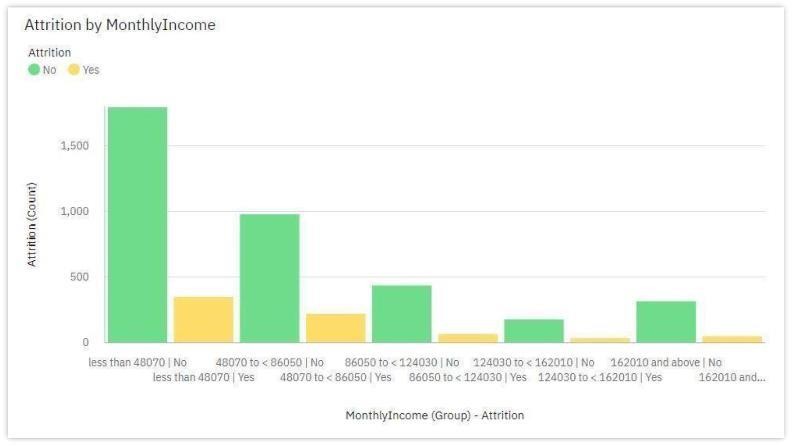
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Database | 2 | 0 | 0 | 2 |
| Dashboard | 1 | 0 | 0 | 1 |
| Visualize the data | 8 | 0 | 0 | 8 |
| Logistic Regression | 4 | 0 | 0 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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.

# 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.

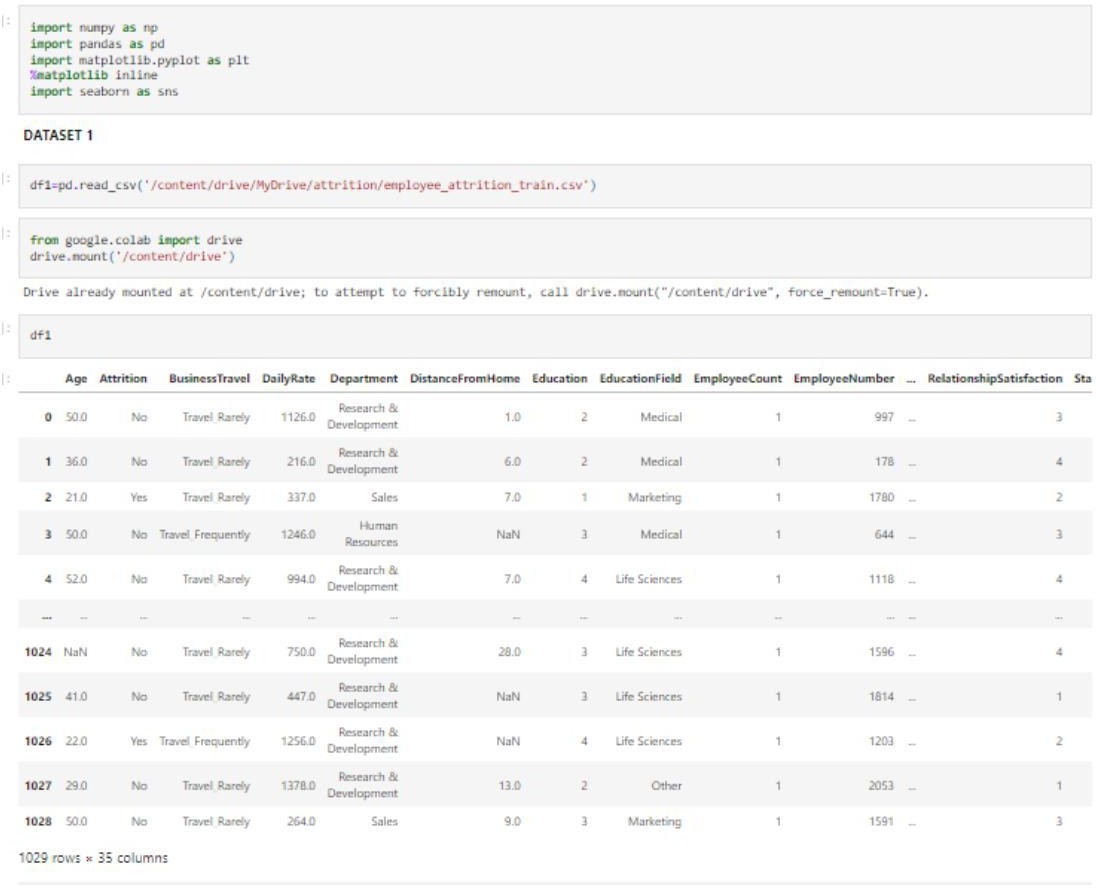
# FUTURE SCOPE

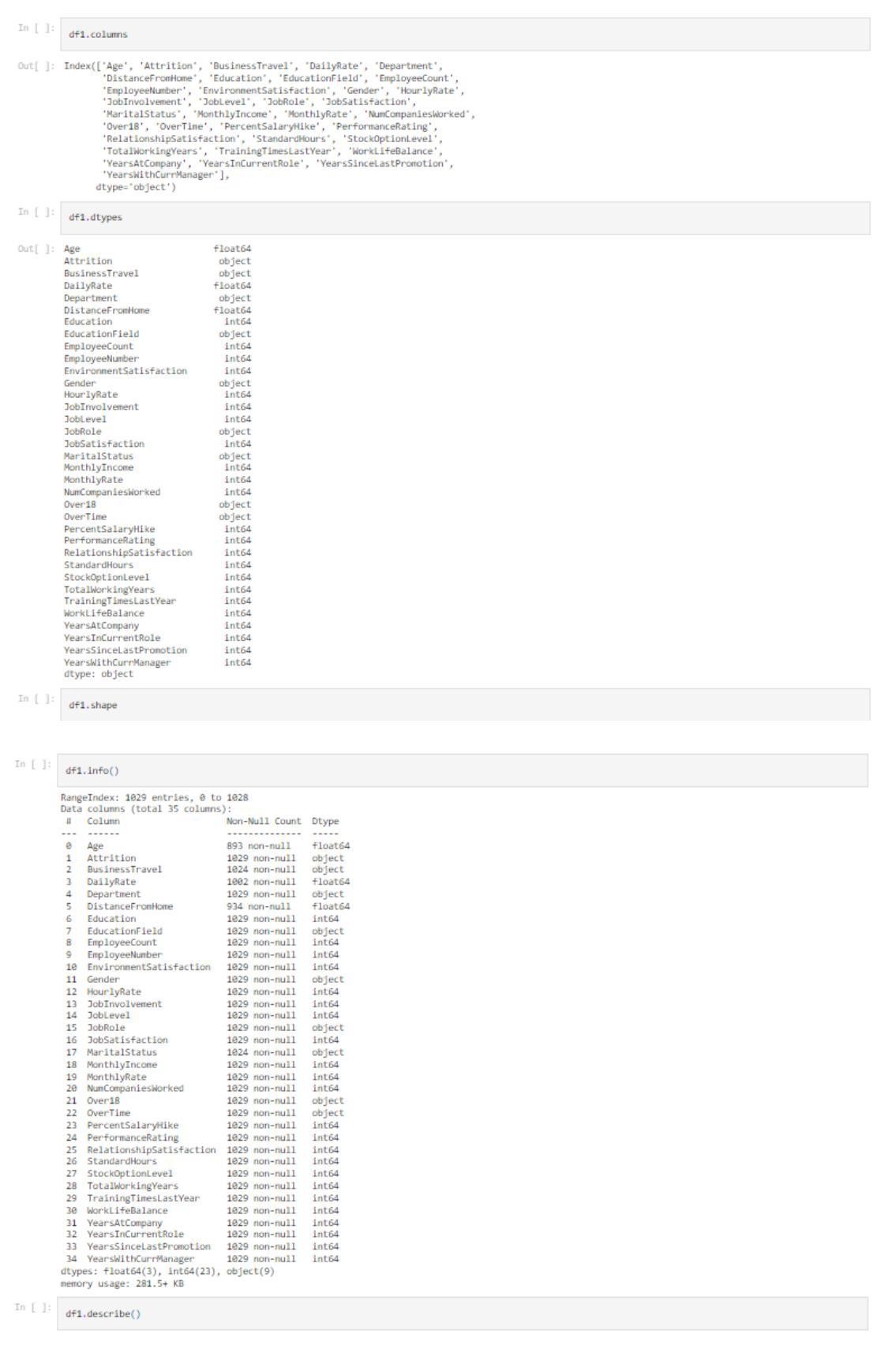
Research findings suggest that attrition reasons in IT organizations primarily revolve around professional growth and challenges in the organization. Although economic factors happen to the most influential factor, professionals may settle for second best criteria of their preference that is career growth 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.

# APPENDIX

**12.1 Source Code**





22

