

# **Corporate Employee Attrition Rate Analysis**

## **A PROJECT COMPONENT REPORT**

*Submitted by*

**Thoshinny B (Reg.No.412519104154)**

**Jayashree S(Reg.No.412519104050)**

**Iswarya G (Reg.No.412519104048)**

**Kaaviya Darshini P (Reg.No.412519104054)**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SRI SAIRAM ENGINEERING COLEGE**

**(An Autonomous Institution affiliated to Anna University Chennai)**

## **CONTENTS**

### **1. INTRODUCTION**

1. Project Overview
2. Purpose

### **2. LITERATURE SURVEY**

1. Existing problem
2. References
3. Problem Statement Definition

### **3. IDEATION & PROPOSED SOLUTION**

1. Empathy Map Canvas
2. Ideation & Brainstorming
3. Proposed Solution
4. Problem Solution fit

### **4. REQUIREMENT ANALYSIS**

1. Functional requirement
2. Non-Functional requirements

### **5. PROJECT DESIGN**

1. Data Flow Diagrams
2. Solution & Technical Architecture
3. User Stories

### **6. PROJECT PLANNING & SCHEDULING**

1. Sprint Planning & Estimation
2. Sprint Delivery Schedule
3. Reports from JIRA

### **7. CODING & SOLUTIONING (Explain the features added in the project along with code)**

1. Feature 1
2. Feature 2
3. Database Schema (if Applicable)

### **8. TESTING**

1. Test Cases
2. User Acceptance Testing

## **9. RESULTS**

### **1. Performance Metrics**

## **10. ADVANTAGES & DISADVANTAGES**

## **11. CONCLUSION**

## **12. FUTURE SCOPE**

## **13. APPENDIX**

Source Code

GitHub & Project Demo Link

## **1. INTRODUCTION**

### **1.1 PROJECT OVERVIEW:**

Around the world, employee churn has grown to be a serious issue. The loss of the best personnel is one of the major problems that company owners deal with in their organisations. A competent employee is always a benefit to the company, and when they leave, it can cause a number of issues, including financial losses, performance issues, and knowledge loss. In addition, compared to recruiting new personnel, hiring new workers is far more expensive, time-consuming, and labor-intensive. It takes a lot of time to hire a new employee because it takes him months to complete his training and get accustomed to the organization's mission, norms, and environment. Therefore, it is imperative that corporate organisations take advantage of emerging trends and technology that use machine learning algorithms. Companies can lessen this cost if they are aware of the cause of employee attrition in advance. This report offers a thorough examination of employee attrition using the IBM HR Analytics Employee Attrition Performance data set.

### **1.2 PURPOSE**

Hardik P. K. (2016) conducted research on "an employee attrition study: with particular reference to Kerala's IT industry." His study investigated the link between organisational variables and the attrition of IT professionals. We can draw the conclusion from the results that organisational characteristics were crucial in predicting the variation in turnover intention (attrition) among Kerala IT workers. In order to decrease the intention of competent personnel leaving their firms, HR managers in IT organisations may pay attention to the issues with organisational elements of their employees.

## **2. LITERATURE SURVEY**

### **2.1 EXISTING PROBLEM**

The current method only uses a small number of features for analysis and deals with both qualitative observations and straightforward statistical analysis. The qualitative observations are data-related and perceptible to humans. They are not quantified or numerical. We now have access to a tonne of data, which comes with a rise in the requirement to manage and comprehend data due to the growth of IOT and linked devices.

### **2.2 REFERENCES**

1. 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)
2. From Big Data to Deep Data to support people analytics for employee attrition prediction, Nesrine Ben Yahia, Hlel Jihen, Ricardo Colomo-Palacio( 2021)
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

# Empathy Map Canvas

Gain insight and understanding on solving customer problems.

1


Build empathy and keep your focus on the user by putting yourself in their shoes.



### 3.2 IDEATION AND BRAINSTORMING

#### Step-1: Team Gathering, Collaboration and Select the Problem Statement.

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

➔

#### 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

How might we predict the employee attrition considering all the factors influencing the company's growth??

Key rules of brainstorming

To run a smooth and productive session

- 🗣️ Stay in topic.
- 💡 Encourage wild ideas.
- ⏸️ Defer judgment.
- 👂 Listen to others.
- 🗣️ Go for volume.
- 👁️ If possible, be visual.

## Step-2: Brainstorm, Idea Listing

2

### Brainstorm

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

🕒 10 minutes

#### TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing

#### Thoshinny B

Examine each and every aspect that influences an employee to leave an organisation.

Get an assessment from professionals who evaluate the cause for the employee's departure visually to estimate the employee's departure.

Calculate the average number of employees who have departed the organisation recently.

Employees that work in a stressful environment may leave the organisation. Consequently, to forecast the business environment.

#### Iswarya G

Examine the likelihood of staff attrition in the future and calculate it.

Surveying people who are entering the company to gather their information so that we may analyse it and choose the ideal employee for the company.

employees' relationships with each other are shaky

Calculate by comparing the staff attrition rate for the current and previous years.

#### Jayashree S

to determine whether they offer pleasant working conditions

Take the employee rate from the prior year and get the average for the current year.

To forecast whether employees are dissatisfied with their pay

Generate by comparing the dropout rate for the present and previous years' employees.

#### Kaaviya Darshini P

HR aims to enliven the employee departments

to be aware that flexible working hours are offered to employees from the perspective of the company.

Using past information from former employees to identify the causes of attrition

To determine if the person lacks the necessary skill set for the job



## Step-3 : Grouping.

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 can break it up into smaller sub-groups.

 20 minutes

#### TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.



#### Evaluate using past records.

Calculate by comparing the staff addition rate for the current and previous years.

Take the employee rate from the prior year and get the average for the current year.

Using past information from former employees to identify the causes of attrition.

#### Evaluate by analysing the impact of external factors.

To determine whether they offer pleasant working conditions

Employees that work in a similar environment may have the opportunity to increase the business environment.

To determine if the person lacks the necessary skill set for the job

#### Presuming by consulting experts

Get an assessment from professionals who evaluate the data for the employee's departure visually to estimate the employee's departure.

#### Evaluate using surveys.

Surveying people who are entering the company to gather data information so that we may analyze it and choose the staffful employees for the company.

#### Evaluate using working hours and salaries:

To forecast whether employees are dissatisfied with their pay

To be aware that flexible working hours are offered to employees from the perspective of the company.

#### Evaluate using other factors:

HR aims to enhance the employee departments

Using past information from former employees to identify the causes of attrition.

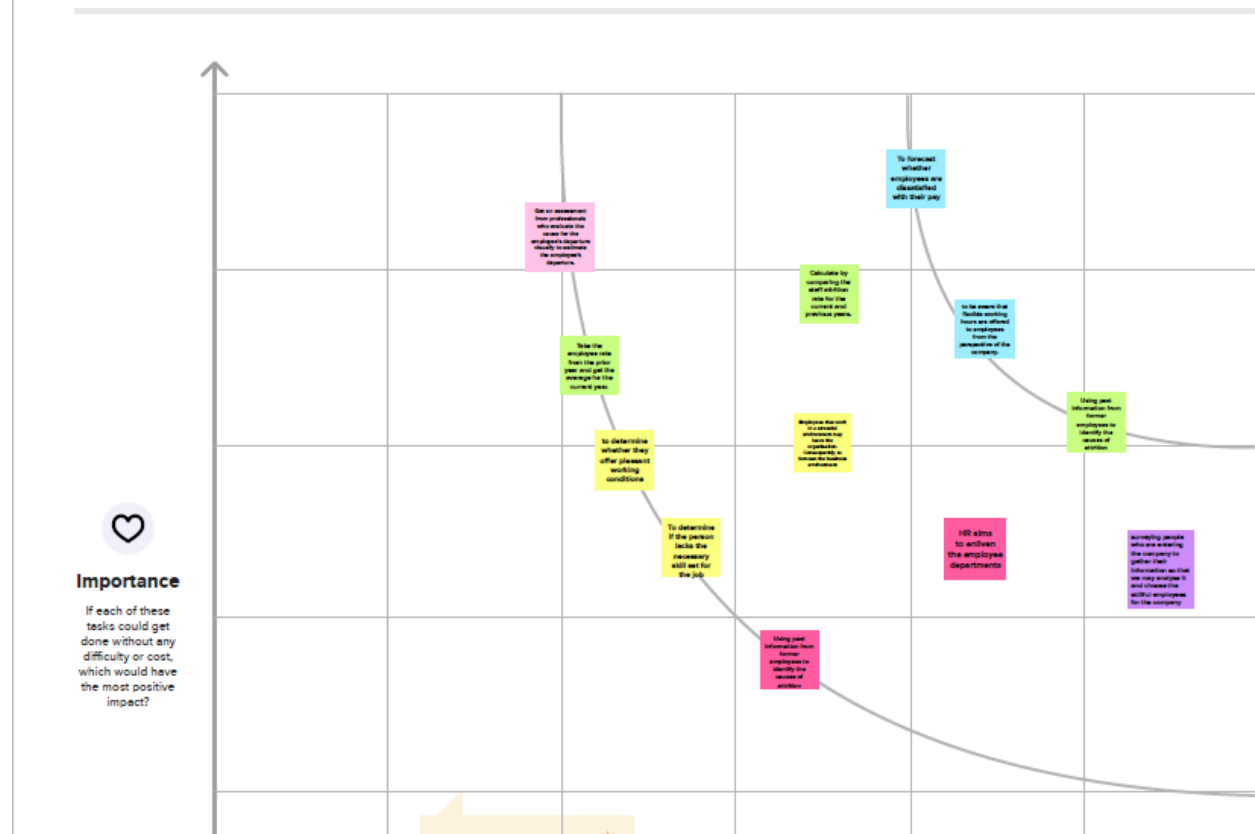
## Step-4 : Idea prioritization

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



## 3.3 PROPOSED SOLUTION

The current method only uses a small number of features for analysis and deals with both qualitative observations and straightforward statistical analysis. The qualitative observations are data-related and perceptible to humans. They are not quantified or numerical. We now have access to a tonne of data, which comes with a rise in the requirement to manage and comprehend data due to the growth of IOT and linked devices.

### 3.4 PROBLEM SOLUTION FIT

Project Title: Corporate Employee Attrition Analysis

Project Design Phase-I - Solution Fit

Team ID: PNT2022TMID03798

4.

Define CS, fit into CC	<b>1. CUSTOMER SEGMENT(S)</b> <ul style="list-style-type: none"> <li>HR of an organization who is responsible for employee's satisfaction in their job</li> <li>Talent Acquisition team</li> <li>Organization team</li> </ul>	<b>6. CUSTOMER CONSTRAINTS</b> <ul style="list-style-type: none"> <li>Communication between employees and organization</li> <li>Budget</li> <li>Datasets about employees</li> <li>Data analysts</li> <li>Time</li> <li>Risk</li> </ul>	<b>5. AVAILABLE SOLUTIONS</b> <ul style="list-style-type: none"> <li>Employee's compensation benefits</li> <li>Promotion offering to employees</li> <li>Data about Salary expectations of employees</li> <li>Feedback from employees</li> <li>Promote flexibility</li> </ul>	Evaluate AS, Affordability
List down all the problems	<b>2. JOBS-TO-BE-DONE / PROBLEMS</b> <ul style="list-style-type: none"> <li>Mentor need to be assigned to know and update employee's skills and personal issues if any.</li> <li>To admire their good performance and motivate them</li> <li>To assign the tasks based on their skill set</li> <li>Poor work life balance</li> </ul>	<b>9. PROBLEM ROOT CAUSE</b> <ul style="list-style-type: none"> <li>Manual analyzation of employee attrition takes more time</li> <li>Working in the same post for many years without promotion</li> <li>Low salary</li> <li>Lack of flexibility</li> <li>Lack of growth</li> <li>Poor workspace culture</li> </ul>	<b>7. BEHAVIOUR</b> <ul style="list-style-type: none"> <li>Getting feedback from employees regularly</li> <li>If they have any issues that can be solved, organization can find the best way to retain the best employee</li> </ul>	Evaluate for the solution

Identify strong TR & EM	<div>3. TRIGGERS</div> <div>TR</div> <ul style="list-style-type: none"><li>Employee who has not been promoted for years but watching their fellow mates developing in their work position.</li><li>Lack of employee development opportunities</li><li>Lack of motivation factors ie. Bonus, promotion, allowance</li><li>Lack of friendly workspace which other organizations provide.</li></ul>	<div>10. YOUR SOLUTION</div> <div>SU</div> <ul style="list-style-type: none"><li>Create a dashboard using monthly feedback from employees with either their identity known or unknown.</li><li>Data from the feedback needs to be constructed as a graph using data analytics to know which factor mostly affects the resignation of employees. The first preference has to be given for the major factor to construct solution.</li></ul>	<div>8. CHANNELS of BEHAVIOUR</div> <div>CH</div> <div>8.1 ONLINE</div> <ul style="list-style-type: none"><li>By using online feedback forms and algorithms, situation can be predicted and examined.</li><li>By using graphs and charts, information can be extracted</li></ul> <div>8.2 OFFLINE</div> <p>Employees physically communicate to share their issues and views</p>	Identify strong TP
	<div>4. EMOTIONS: BEFORE / AFTER</div> <div>EM</div> <div><div>Before</div><div>1. Stress</div><div>2. Dissatisfaction</div><div>3. Lack of smooth relationship with colleagues</div></div> <div><div>After</div><div>1. Motivation</div><div>2. Friendly relationship with colleagues and team lead</div><div>3. Satisfied and work with interest</div></div>			

## REQUIREMENT ANALYSIS

### 4.1 Functional requirement

Following are the functional requirements of the proposed solution.

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 Authentication	Using the details in the database, verify the user's attempted login.
FR-4	Retention Analysis	Analysis of employee attrition based on sentiment, the workplace, daily involvement, etc.
FR-5	Employee management	Verification and administration of the registered employee information.
FR-6	Progress management	Each employee's advancement inside the organisation is added.
FR-7	Input Credentials	After uploading dataset, graphical analysis of the attrition rate will be generated.
FR-8	Processing Methods	IBM Cognos Analytics Dashboard is used to find the attrition rate using prediction method.

### 4.2 Non-functional Requirements:

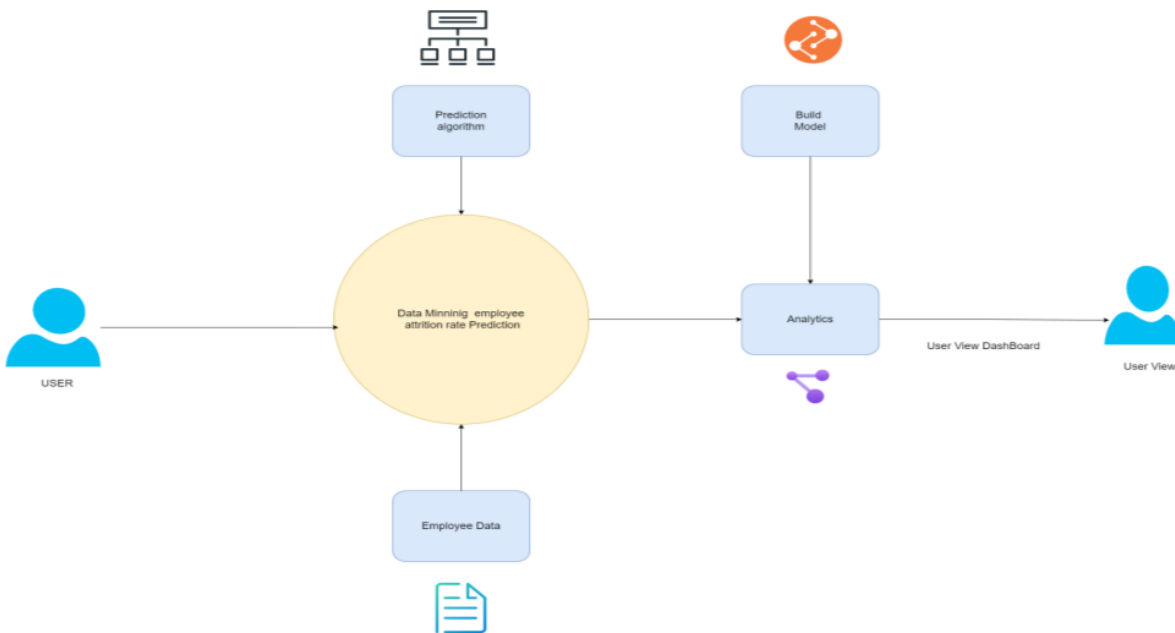
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	All users of this software will find it simple to use and require few procedures. All graphical user interface (GUI) languages must be user-friendly and understandable to non-technical users in 100% of cases.
NFR-2	Security	The system's data administrator is the only person who has the ability to modify the access rights for a certain system's information. High security standards must be in place for the user's data.
NFR-3	Reliability	Any failed changes must cause the database update procedure to roll back any linked updates. The only person who can change the dataset is the user. No one else will be allowed to do so.

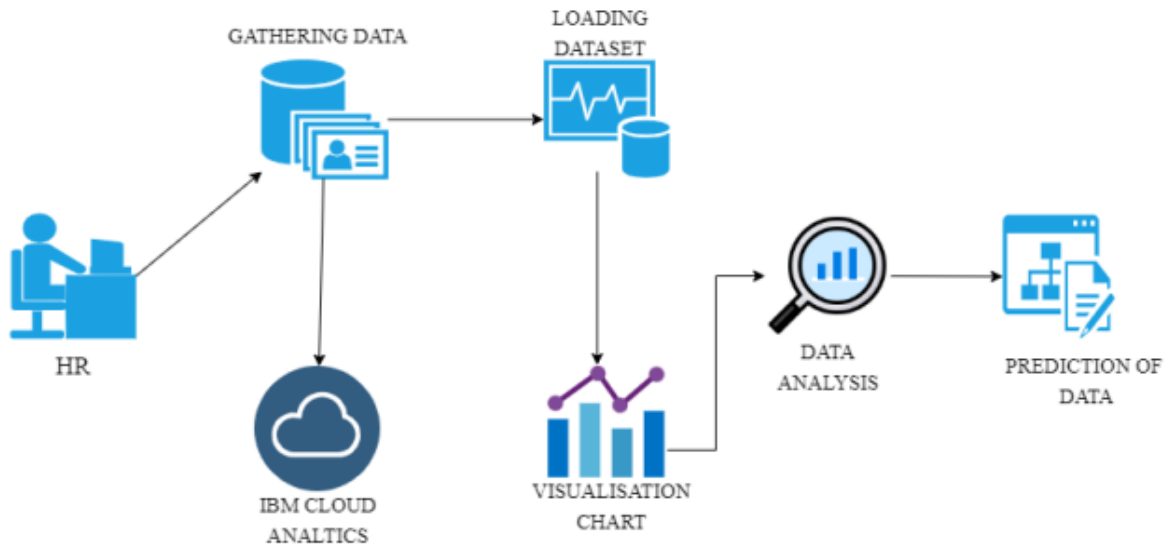
NFR-4	<b>Performance</b>	The system's efficiency must support the system's excellence.
NFR-5	<b>Availability</b>	The deployment of a new module shouldn't affect the accessibility of the home page, dashboard, or checkout pages, and it shouldn't take more than an hour. The rest of the pages that might encounter issues must present a notice with a countdown indicating when the system will be back up.
NFR-6	<b>Scalability</b>	The website's attendance cap must be expandable to accommodate 200,000 users at once. When the employee dataset from the companies is used for analysis, the dashboard is scalable for the businesses. The model effectively foretells the futuristic approach and suggests protective measures.

## 5. PROJECT DESIGN

### 5.1 Data Flow Diagrams



## 5.2 Solution and Technical Architecture



## 5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a CEO, 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 an employee, I can register for the application by entering my mail, password, and confirming password	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Google	I can register & access the dashboard with Google Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook login	Medium	Sprint-1
Customer (CEO)	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
Customer (Employee)		USN-6	As a user, I can log into the application by entering email & password	I can access my account/dashboard	High	Sprint-3
CEO	Dashboard	USN-7	As a CEO, I can use the predict button to know which factor keeps the employee at the company and which prompts others to leave	I can view the visualizations	High	Sprint-4
EMPLOYEE		USN-8	As an employee of the organization, I can view, fill and submit the survey form that is displayed	I can see the acknowledgement message for submitting the survey	High	Sprint-4

## 6. PROJECT PLANNING

### 6.1 Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dashboard	USN-1	I provide the information about the employees who work for our company as a user for the attrition detail.	5	High	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-1		USN-2	I will examine the dataset as an Analyst and carry out exploratory data analysis in Cognos Analytics.	3	High	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-2	Report	USN-3	I want minimal, simpler visualisations that report certain events as a user.	2	Low	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-2		USN-4	I'll create a report using Cognos Analytics as an analyst.	3	Medium	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-3	Story	USN-5	As a user, I can only comprehend the Analysis in an animated dataset display.	3	Medium	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-3		USN-6	I utilise Cognos Analytics as an analyst to produce an animated presentation (story) using the dataset.	3	Medium	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-4	Predictive Analysis	USN-7	I want to use the dataset as a user to forecast the company's attrition rate.	5	Medium	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B
Sprint-4		USN-8	As an Analyst, I'll carry out Prediction Analysis using a variety of Python modules.	3	High	Iswarya.G, Jayashree.S, Kaaviya Darshini.P, Thoshinny.B

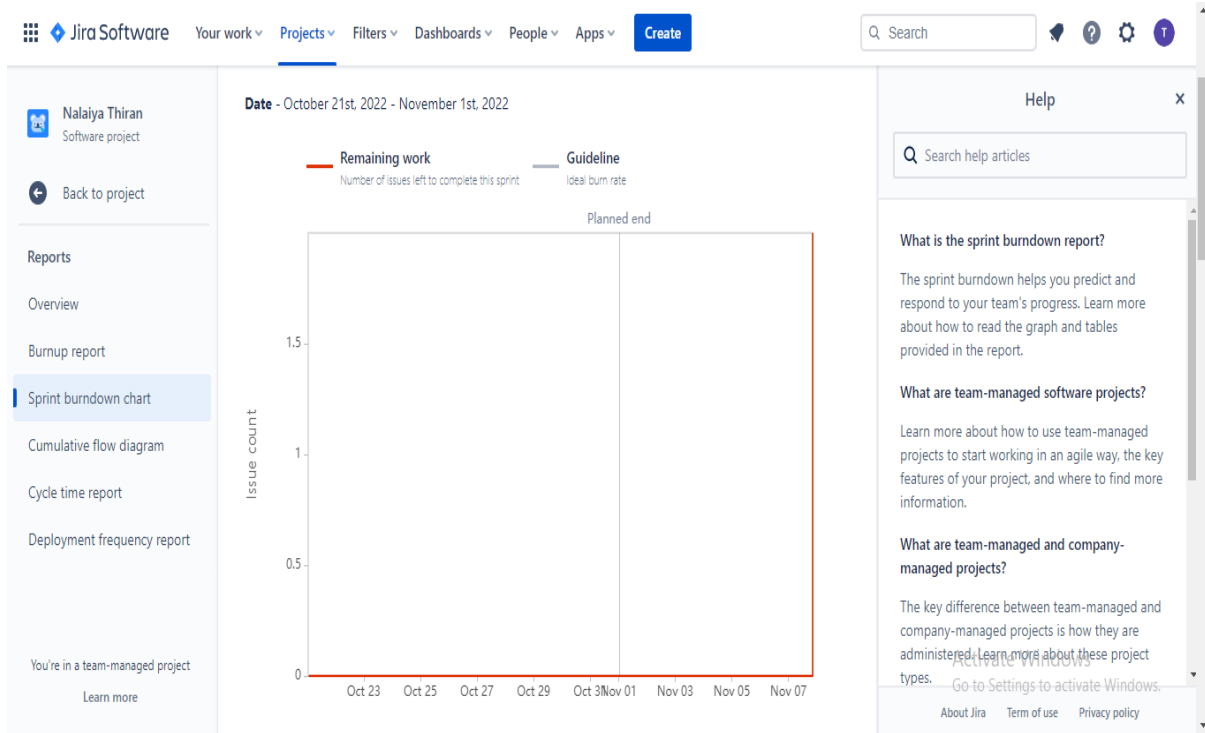
### 6.2 Sprint Delivery Schedule

Project Tracker, Velocity & Burndown Chart: (4 Marks)

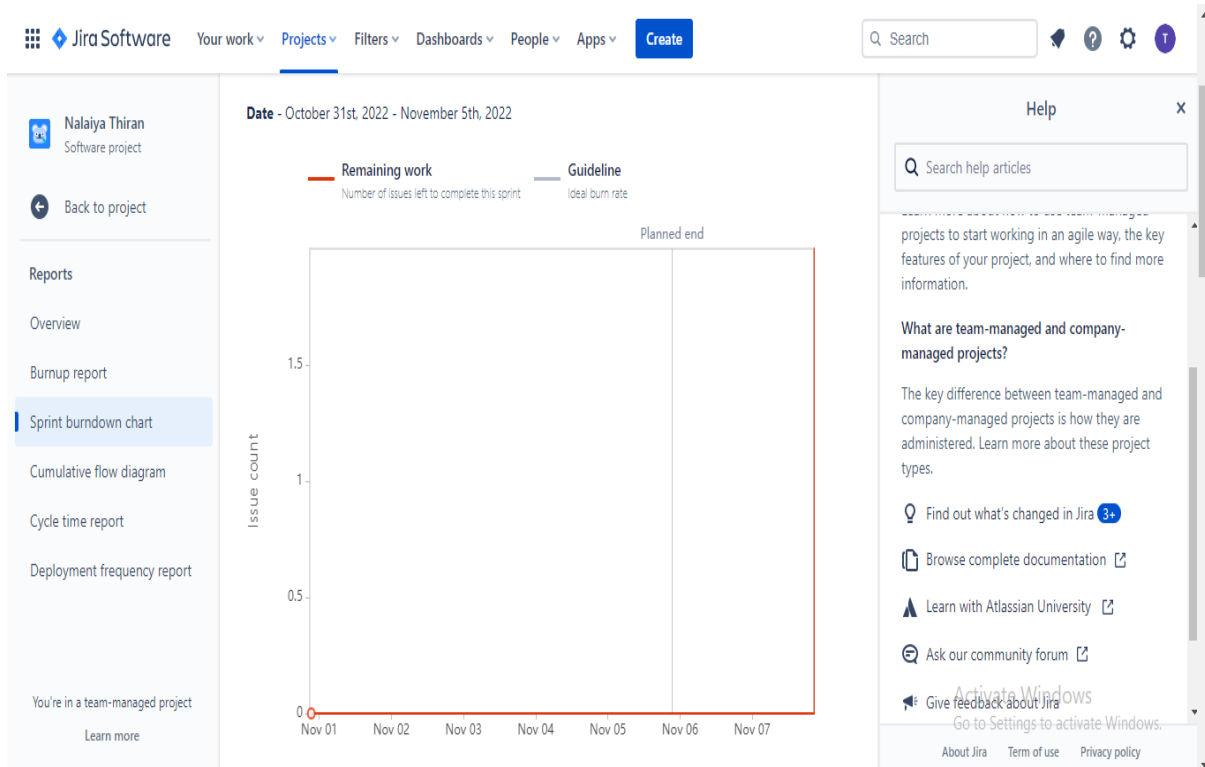
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	5	6 Days	27 Oct 2022	01 Nov 2022	5	01 Nov 2022
Sprint-2	5	6 Days	31 Oct 2022	05 Nov 2022	5	05 Nov 2022
Sprint-3	5	6 Days	07 Nov 2022	12 Nov 2022	5	12 Nov 2022
Sprint-4	5	6 Days	14 Nov 2022	19 Nov 2022	5	19 Nov 2022

## 6.3 Reports from JIRA

### Sprint - 1

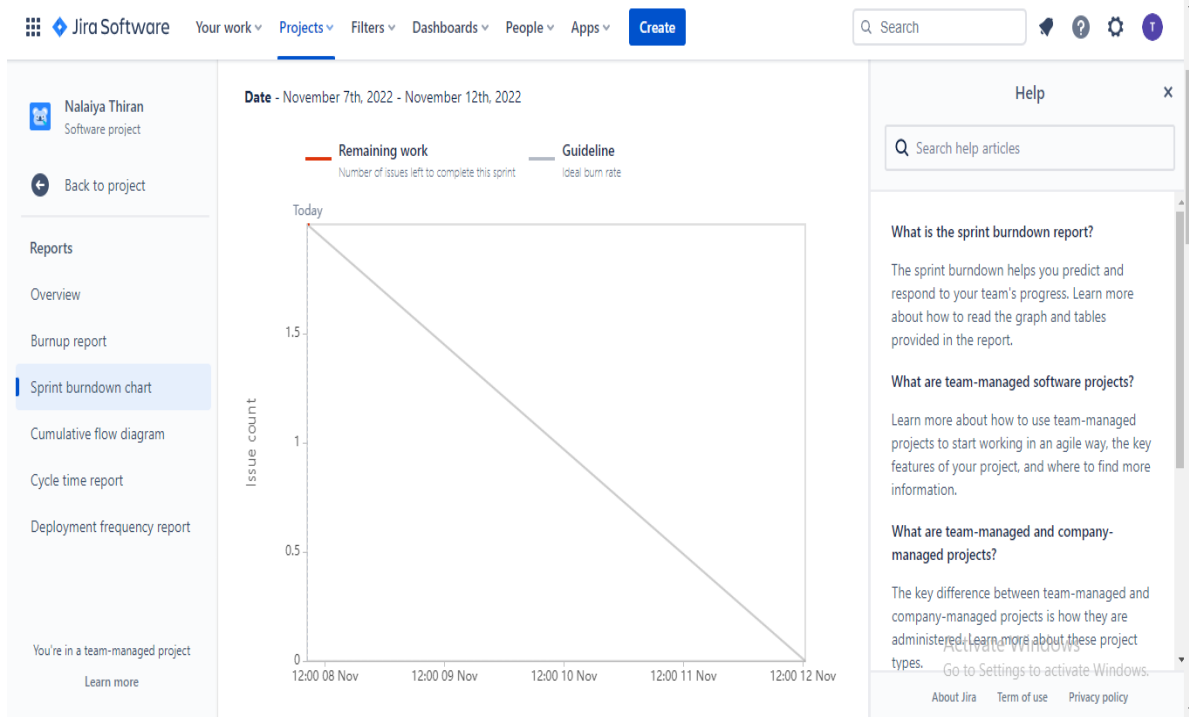


### Sprint - 2

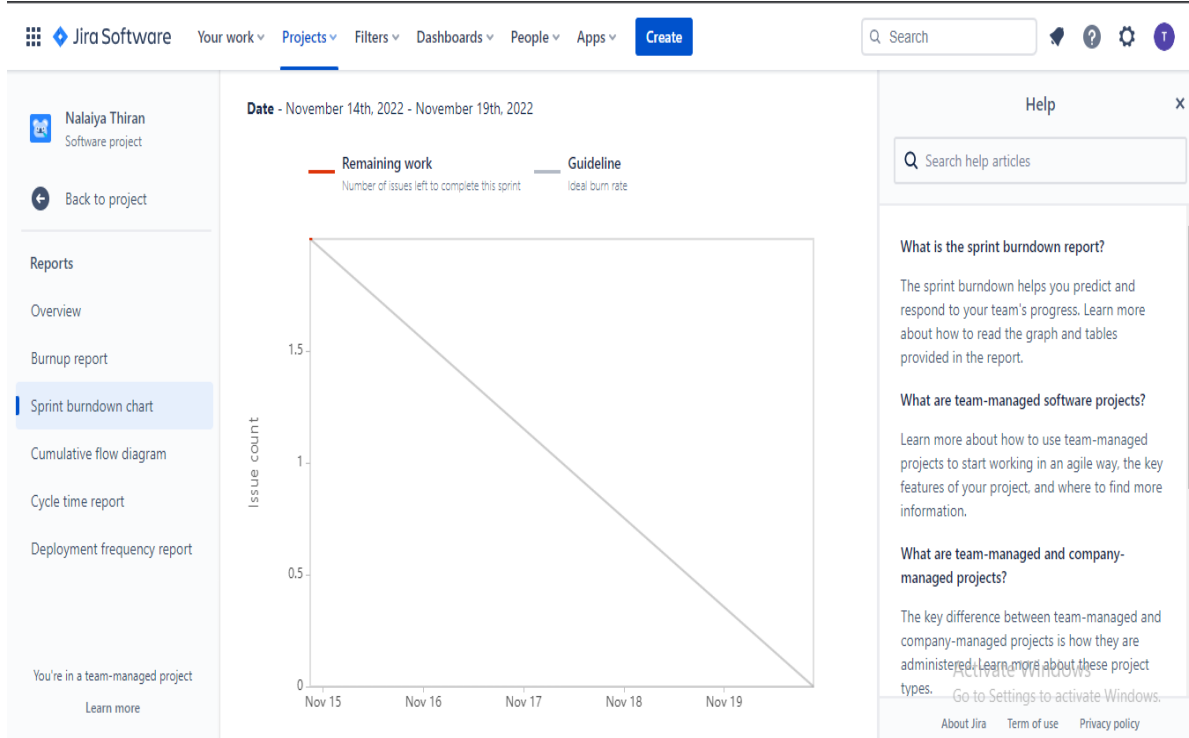




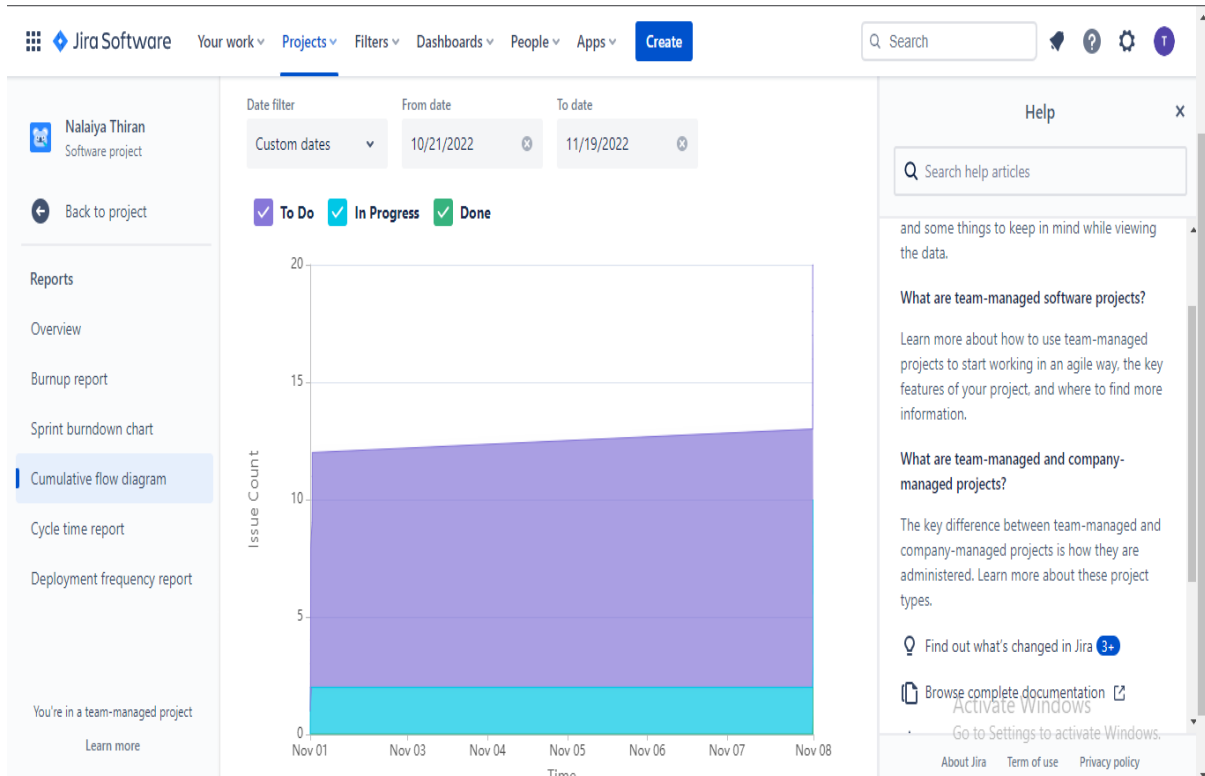
## Sprint - 3



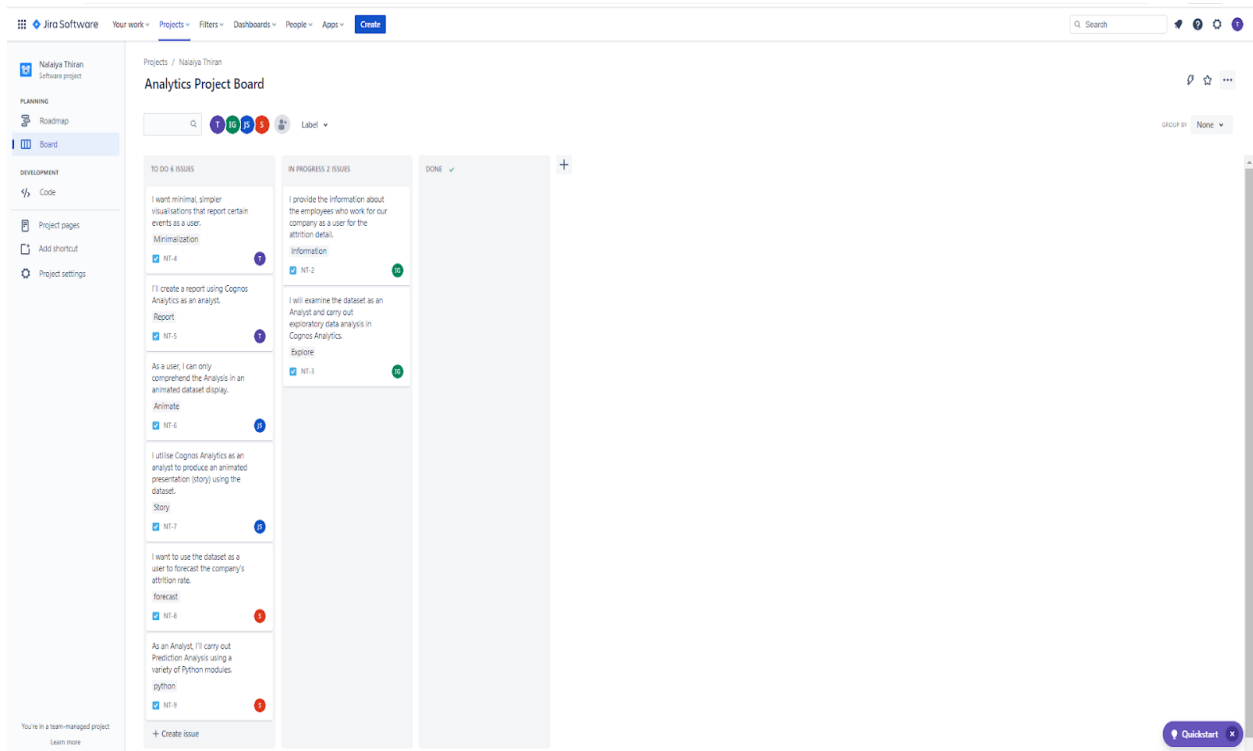
## Sprint - 4



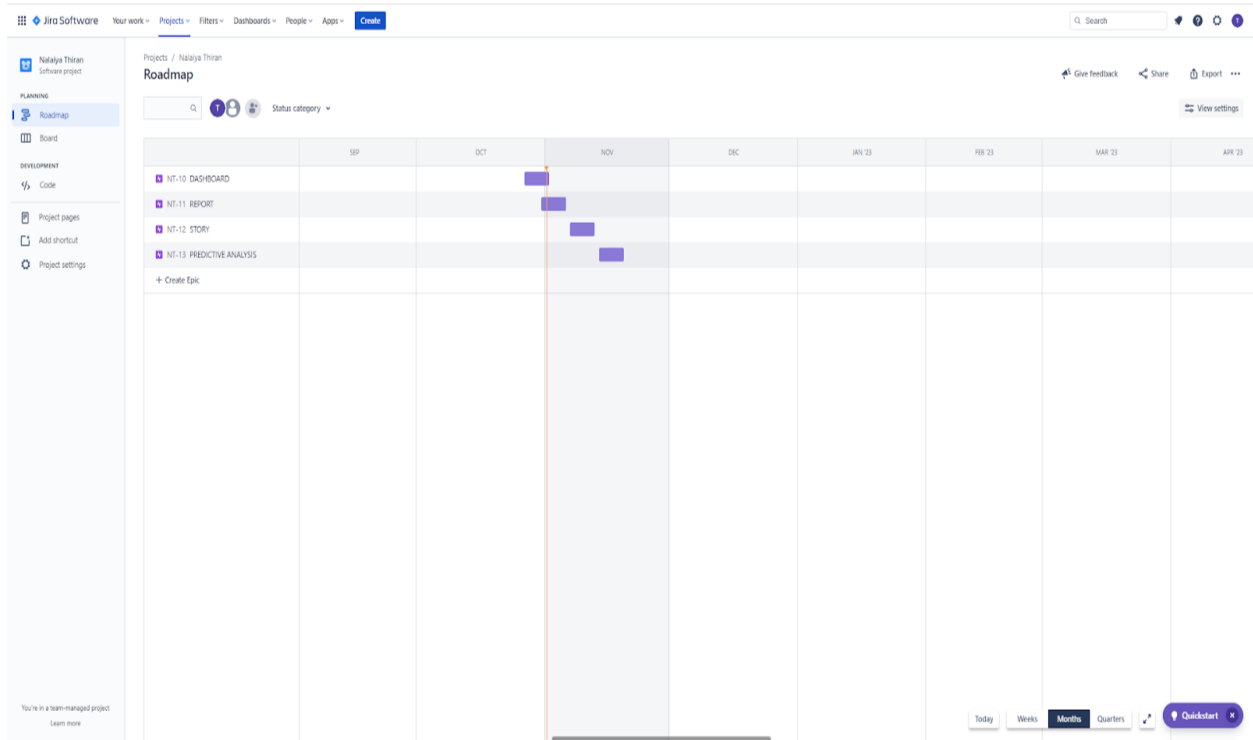
## Cumulative Flow Diagram :



## Board :



## Road map :



## 7. CODING AND SOLUTIONING

### #DATA MODULE

```
#importing required libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import glob
```

### #import MODEL SELECTION

```
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import GridSearchCV
```

### #import MODEL

```
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
```

```

from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier

#import MODEL SCORES
from sklearn.metrics import confusion_matrix , accuracy_score
,classification_report

```

### **#Load Data**

```

df1 = pd.read_csv("/content/employee_survey_data.csv")
df2 = pd.read_csv("/content/general_data.csv")
df3 = pd.read_csv("/content/manager_survey_data.csv")

```

### **#Data Preprocessing**

```

print("Employee survey data")
df1.head()

```

```

df1.tail()

```

```

print("General data")
df2.head()

```

```

df2.tail()

```

```

print("Manager survey data")
df3.head()

```

```

df3.tail()

```

### **Merging the datasets**

#### **Merging the Data from 1 and 2 based on the Common Column -'EmployeeID'**

```

df1_2 = df1.merge(df2,on ="EmployeeID", suffixes=('_left'))
/usr/local/lib/python3.7/dist-packages/pandas/core/frame.py:9203:
FutureWarning: Passing 'suffixes' as a <class 'str'>, is not supported and
may give unexpected results. Provide 'suffixes' as a tuple instead. In the
future a 'TypeError' will be raised.
    validate=validate,

```

```

#Merged data
df1_2.head()

```

Similarly Merging Dataframe 1,2 and 3 based on the Common Column 'EmployeeID'

```
df = df1_2.merge(df3,on ="EmployeeID", suffixes=('left'))
/usr/local/lib/python3.7/dist-packages/pandas/core/frame.py:9203:
FutureWarning: Passing 'suffixes' as a <class 'str'>, is not supported and
may give unexpected results. Provide 'suffixes' as a tuple instead. In the
future a 'TypeError' will be raised.
    validate=validate,
```

```
df.head()
```

```
df.info()
```

```
df.describe()
```

```
df.shape
(4410, 29)
```

## **Data Cleaning**

### **Checking the null and improper values**

```
df.isnull().any()
```

```
df.isnull().sum()
```

Replacing the NULL values based on the AVERAGE obtained from DESCRIPTION

```
df['EnvironmentSatisfaction'] = df['EnvironmentSatisfaction'].replace(np.nan,
2.723603)
df['JobSatisfaction'] = df['JobSatisfaction'].replace(np.nan, 2.728246)
df['WorkLifeBalance'] = df['WorkLifeBalance'].replace(np.nan, 2.761436)
df['NumCompaniesWorked'] = df['NumCompaniesWorked'].replace(np.nan, 2.694830)
df['TotalWorkingYears'] = df['TotalWorkingYears'].replace(np.nan, 11.279936)
```

### **Checking NULL values after Replacing it**

```
df.isnull().sum()
```

### **Describing the Mathematical and statistical Functions of All the Columns**

```
pd.set_option('display.max_columns', 30)
```

```
df.describe(include='all')
```

The attrition respect to the Employee is counted

```
attrition_count = pd.DataFrame(df["Attrition"].value_counts())
print(attrition_count)
```

### Chart for Employee Attrition

```
df.head()
```

```
plt.pie(attrition_count['Attrition'], labels= ['No','Yes'] , explode =
(0.1,0))
```

### Dropping the unwanted column

```
df.drop('EmployeeCount', axis=1, inplace=True)
df.shape
(4410, 28)
```

The Attrition column has Yes and No values. Its converted into Two columns by replacing 1/0

```
df.head()
```

```
attrition_split = pd.get_dummies(df['Attrition'],columns=['name'])
attrition_split.head()
```

Concatinating The attrition Split data with Original Data Frame

```
df = pd.concat([df, attrition_split],axis=1)
df.head()
```

Either Yes or No is required So dropping out No column

```
df = df.drop(['Attrition','No'], axis=1)
df.head()
```

### Visualization

```
plt.figure(figsize=(20,10))
sns.heatmap(df.corr(),cmap="YlGnBu")
```

```

sns.barplot(x='JobLevel', y='Yes', data=df )

sns.barplot(x='BusinessTravel', y='Yes', data=df)

sns.barplot(x='WorkLifeBalance', y='Yes', data=df)

sns.barplot(x='MaritalStatus', y='Yes', data=df)

sns.barplot(x='JobInvolvement', y='Yes', data=df)

```

### Data Wrangling

```

from sklearn.preprocessing import LabelEncoder
for column in df.columns:
    if df[column].dtype==np.number:
        continue
    else:
        df[column]=LabelEncoder().fit_transform(df[column])

```

### Model Selection

```

#Cross Validation
k = KFold(n_splits = 5)

#Logistic Regression
lr_model = LogisticRegression()
lr_score = cross_val_score(lr_model , x_train , y_train ,cv = k ,scoring =
'neg_mean_squared_error')
lr_score.mean()

#Random Forest Classifier
rf_model = RandomForestClassifier()
rf_score = cross_val_score(rf_model , x_train , y_train ,cv = k ,scoring =
'neg_mean_squared_error')
rf_score.mean()

#Support Vector machine Classifier
svm_model = SVC()
svm_score = cross_val_score(svm_model , x_train , y_train ,cv = k ,scoring =
'neg_mean_squared_error')

```

```

svm_score.mean()

#Decison Tree Classifier
dt_model = DecisionTreeClassifier()
dt_score = cross_val_score(dt_model , x_train , y_train ,cv = k ,scoring =
'neg_mean_squared_error')
dt_score.mean()

plt.figure(figsize = (14 , 6))
plt.subplot(1,2,1)
x = ['Logistic Regression','Random Forest' , 'Support Vector' , 'Decision
Tree']
y = [lr_score.mean() , rf_score.mean() ,svm_score.mean() , dt_score.mean()]
plt.title('Neg Mean square error for Models')
sns.barplot(y,x,palette="viridis")

plt.subplot(1,2,2)
plt.plot(x ,y,marker = 'o' ,color = 'b',mfc = 'g' ,ms =8 )
plt.title('Neg Mean square error')

model_params ={
    'RandomForestClassifier':
    {
        'model':RandomForestClassifier(),
        'param':
        {
            'n_estimators':[10 ,50 ,100,130],
            'criterion':['gini' , 'entropy'],
            'max_depth':range(4,8,1),
            'max_features':['auto' , 'log2']
        }
    },
    'SVC':
    {
        'model':SVC(),
        'param':
        {
            'C':[1,20],
            'gamma':[1,0.1],
            'kernel':['rbf']
        }
    }
}

```



```

    }
}

scores =[]
for model_name , mp in model_params.items():
    model_sel = GridSearchCV(estimator= mp['model'] ,param_grid= mp['param']
,cv = 4 ,return_train_score=False)
    model_sel.fit(x_train,y_train)

    scores.append({
        'model':model_name,
        'best_score':model_sel.best_score_,
        'best_params':model_sel.best_params_
    })

scores
[{'model': 'RandomForestClassifier',
  'best_score': 0.8694532368299026,
  'best_params': {'criterion': 'gini',
  'max_depth': 7,
  'max_features': 'auto',
  'n_estimators': 130}},
 {'model': 'SVC',
  'best_score': 0.8331716262440946,
  'best_params': {'C': 1, 'gamma': 1, 'kernel': 'rbf'}}]
```

### Random Forest ML Model

#### To predict the Employee Attrition

This encodes the Label values into Numerical data

```

from sklearn.preprocessing import LabelEncoder
for column in df.columns:
    if df[column].dtype==np.number:
        continue
    else:
        df[column]=LabelEncoder().fit_transform(df[column])

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=10, criterion='entropy',
```

```
random_state=0)
```

X and y values are seperated

```
x = df.drop(['Yes'], axis=1)
```

```
y = df['Yes']
```

The training and testing dataset is splitted

```
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size=0.3,
```

```
random_state=0)
```

Testing X and Y dataset after splitting

```
#checking x train
```

```
x_train.head()
```

```
#checking y train
```

```
y_train.tail()
```

### **Fitting the DATASET in RANDOM FOREST Model**

```
rf.fit(x_train, y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=10, random_state=0)
```

```
#Checking the scores of Training Data
```

```
rf.score(x_train,y_train)
```

```
#Predicting the y test data in Randomforest Model
```

```
pred = rf.predict(x_test)
```

```
#Importing Accuracy score package to calculate the score of Prediction
```

```
from sklearn.metrics import accuracy_score
```

```
accuracy_score(y_test,pred)
```

```
0.9750566893424036
```

### **ML Model Accuracy**

The ML Model predicts the Testing Data with high Accuracy(97.50%)

This is the Best fit model for given data

## 8. TESTING

### 8.1 Test Cases

#### 1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

### 8.2 User Acceptance Testing

#### 2. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved.

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

#### 3. Test Case Analysis

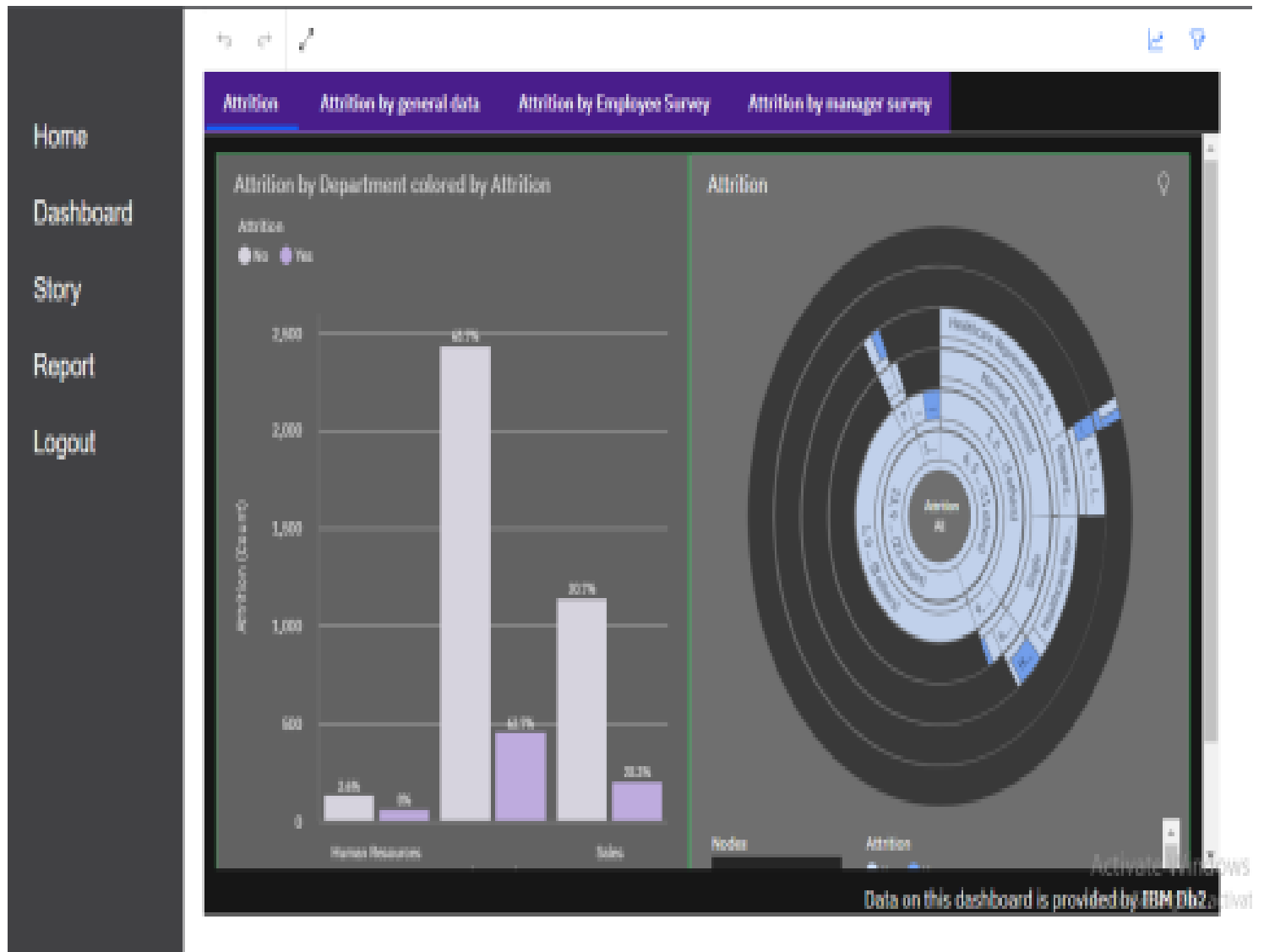
This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Register Page	6	0	0	6
Login Page	6	0	0	6
Main Page	5	0	0	5
Dashboard	4	0	0	4
Report	4	0	0	4
Story	4	0	0	4
Logout	2	0	0	2

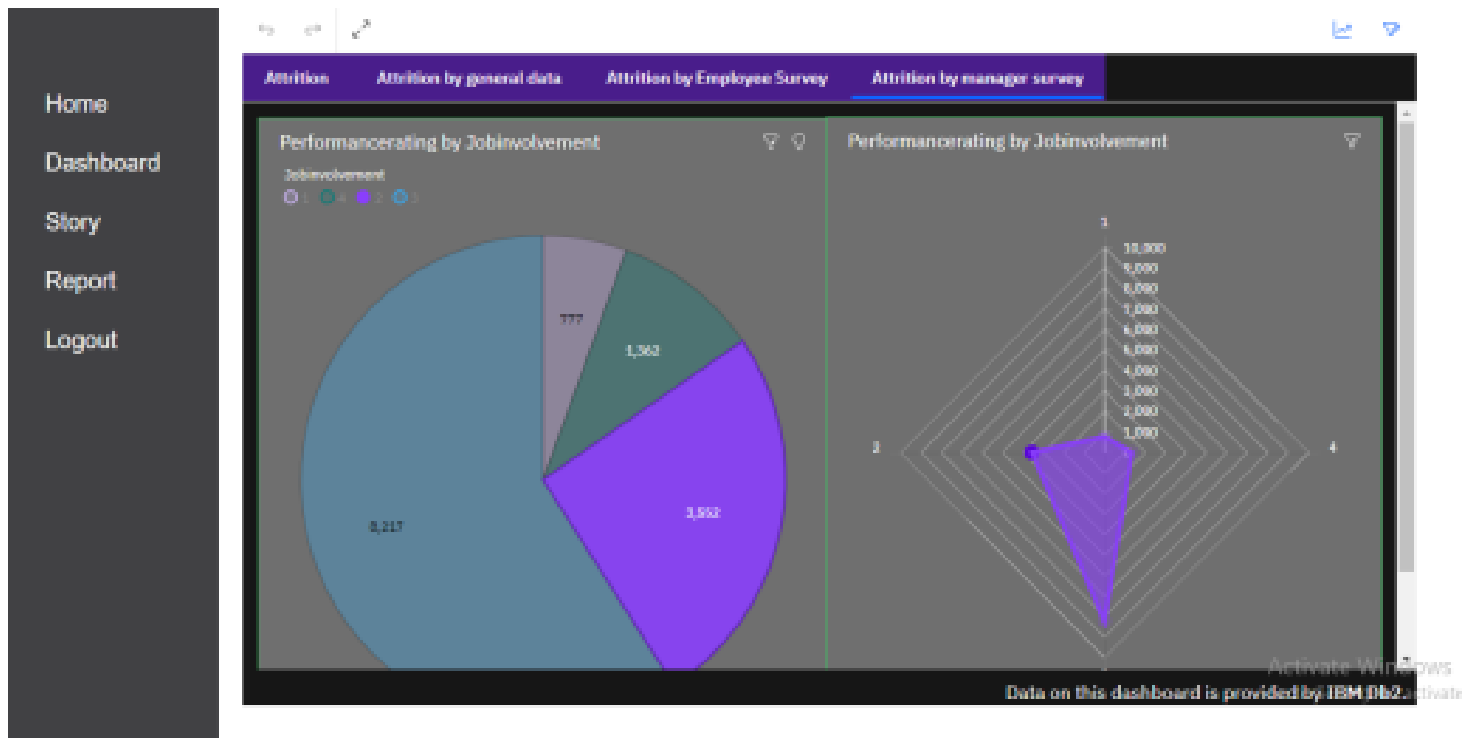
## 9. RESULTS

### 9.1 Performance Metrics

#### Attrition by Department:



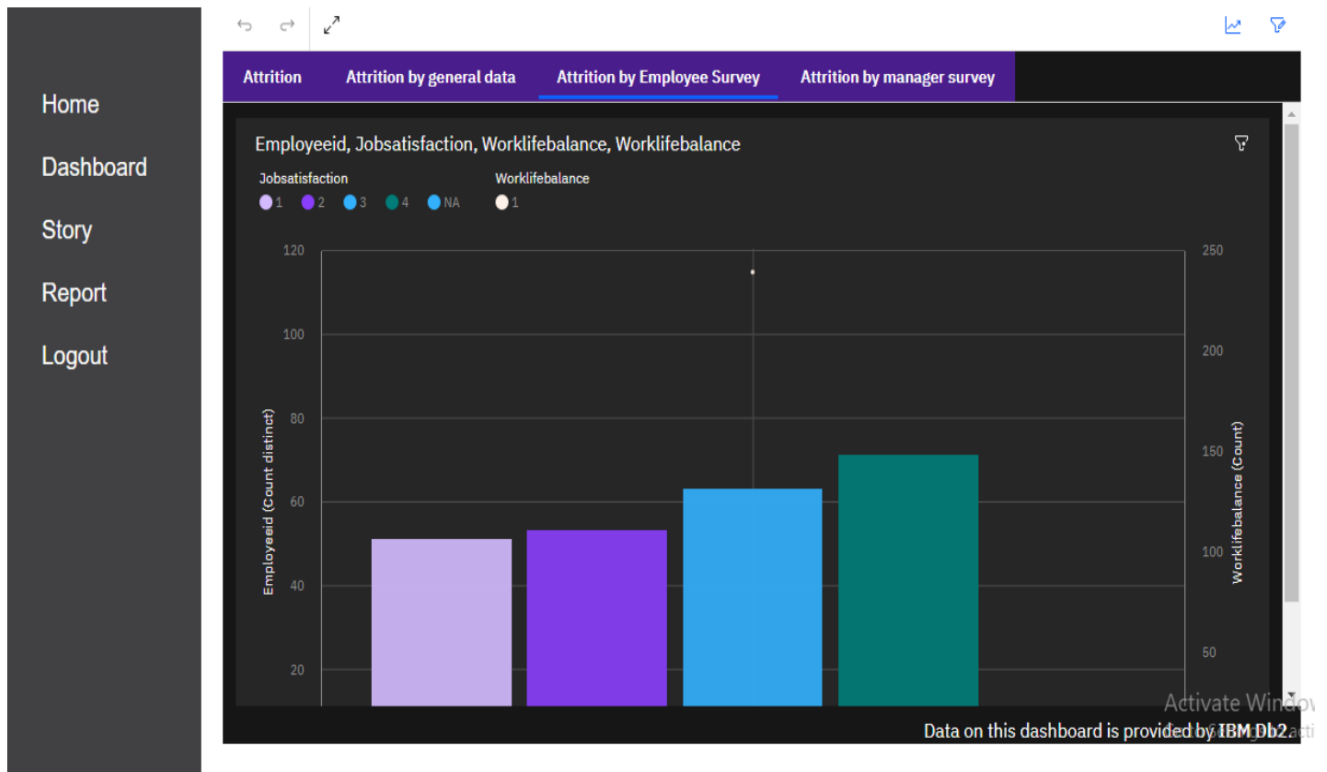
### Attrition by manager survey:



### Attrition and percent salary hike by employee count and education field:



## Attrition by employee survey

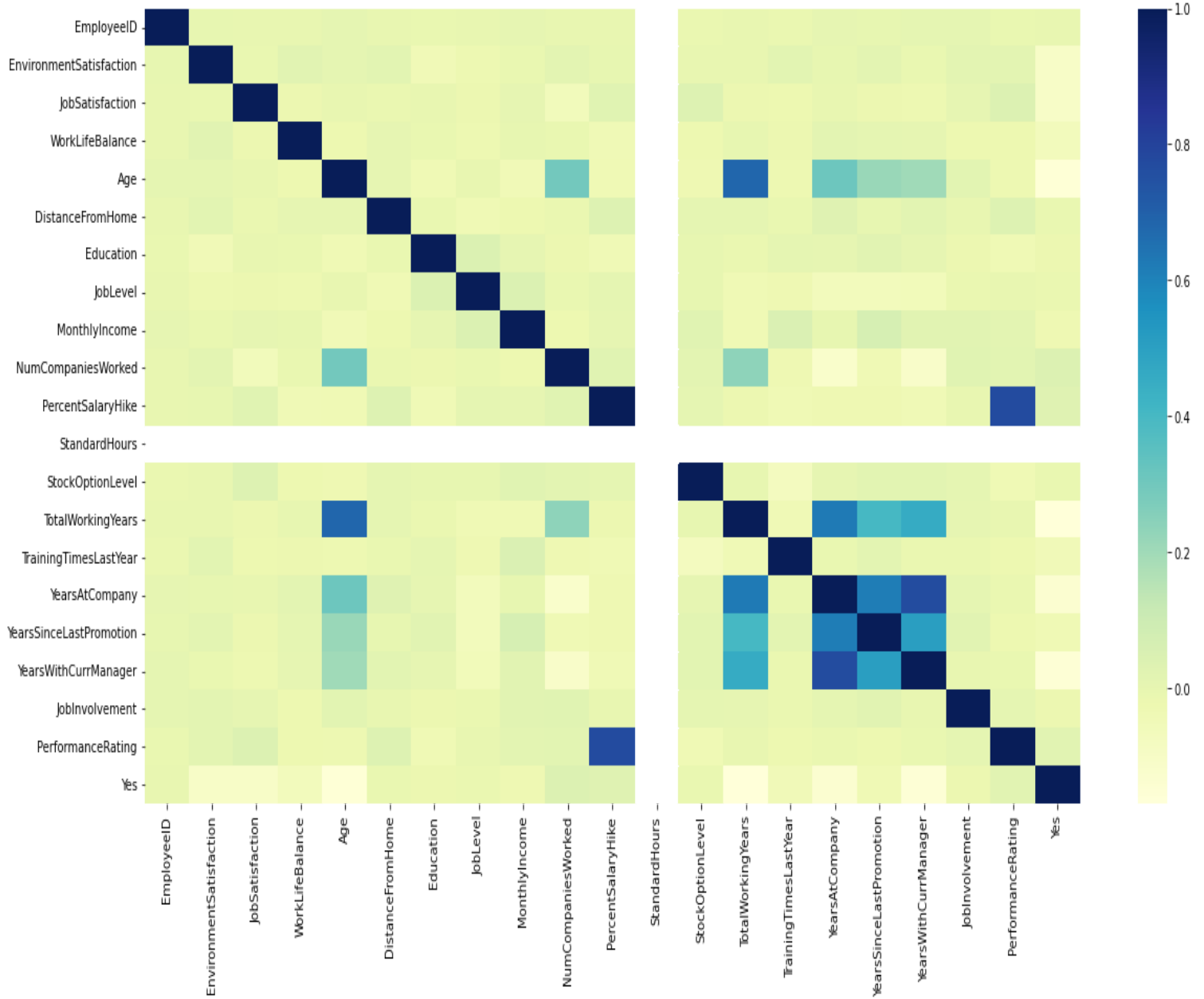


## REPORT



## 9.2 Performance Analysis

## Correlation matrix



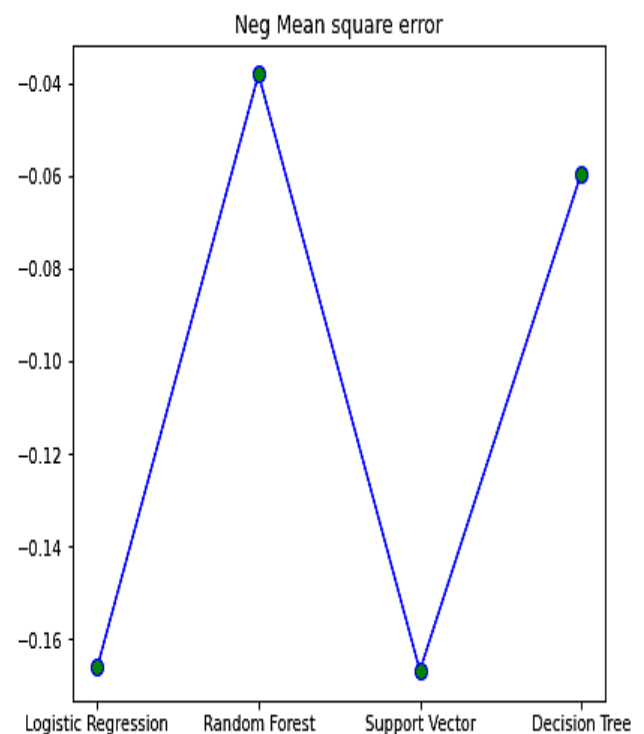
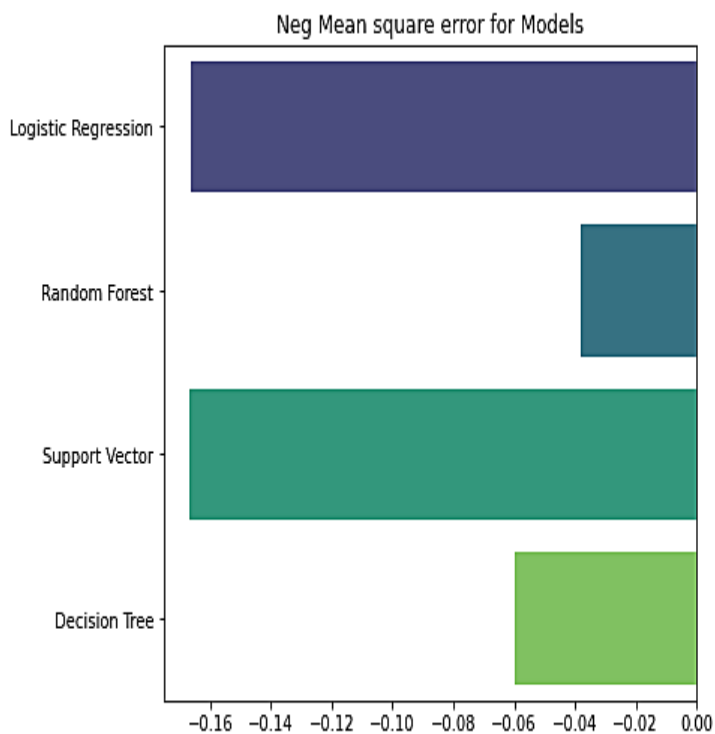
**Model Selection :**

```

✓ [198] plt.figure(figsize = (14 , 6))
0s
      plt.subplot(1,2,1)
      x = ['Logistic Regression', 'Random Forest', 'Support Vector', 'Decision Tree']
      y = [lr_score.mean(), rf_score.mean(), svm_score.mean(), dt_score.mean()]
      plt.title('Neg Mean square error for Models')
      sns.barplot(y,x,palette="viridis")

      plt.subplot(1,2,2)
      plt.plot(x ,y,marker = 'o' ,color = 'b',mfc = 'g' ,ms =8 )
      plt.title('Neg Mean square error')

```





```

[199] model_params = {
    'RandomForestClassifier':
    {
        'model': RandomForestClassifier(),
        'param':
        {
            'n_estimators': [10, 50, 100, 130],
            'criterion': ['gini', 'entropy'],
            'max_depth': range(4, 8, 1),
            'max_features': ['auto', 'log2']
        }
    },
    'SVC':
    {
        'model': SVC(),
        'param':
        {
            'C': [1, 20],
            'gamma': [1, 0.1],
            'kernel': ['rbf']
        }
    }
}

scores = []
for model_name, mp in model_params.items():
    model_sel = GridSearchCV(estimator= mp['model'], param_grid= mp['param'], cv= 4, return_train_score=False)
    model_sel.fit(x_train, y_train)

    scores.append({
        'model': model_name,
        'best_score': model_sel.best_score_,
        'best_params': model_sel.best_params_
    })

```

## MODEL OPTIMAL PARAMETER SELECTION USING GRID SEARCH CV

We obtained less less -ve mena sq error for SVC and random forest (NEARLY SAME).

Building optimal parameter model with both of them

```

scores
[[{'model': 'RandomForestClassifier',
  'best_score': 0.8681562199686834,
  'best_params': {'criterion': 'entropy',
                  'max_depth': 7,
                  'max_features': 'auto',
                  'n_estimators': 10}},
 {'model': 'SVC',
  'best_score': 0.8331716262440946,
  'best_params': {'C': 1, 'gamma': 1, 'kernel': 'rbf'}}]

```

**Selected Model** -> Random Forest Classifier model

## **10. Advantages and Disadvantages**

### **10.1. Advantages**

#### **Data Collection :**

The survey is being conducted among two types of working IT professionals. This categorization was primarily based on experience level and role in the organisation. It was critical to understand the perspectives of both job seekers and interviewers involved in the hiring process. The research study makes use of both primary and secondary data. Primary Data, A structured self-administered Questionnaire is used to collect primary data in a field survey. By using convenience sampling, the survey consisted of closed-ended questions. The 5-point rating scale is the scaling technique used in the questionnaire. A total of 120 respondents were IT professionals from organisations in the Maharashtra cities of Nagpur, Pune, and Mumbai. Secondary Data Secondary data is gathered by consulting journals, research papers, and published data in the form of books and newspapers.

#### **Type of research :**

The descriptive research design methodology was used for the research paper. Sample Design, Sample Size, and Sampling Method The study's sample is drawn from the Indian information technology industry. The sample is limited to working professionals in the Information Technology sector and was collected using the convenience sampling technique. The sample size was 120 respondents.

## **11. Conclusion**

Employees and organisations must be clear about their job expectations. Any mismatch leads to inconsistency, and employees may fail to perform their duties. This eventually leads to attrition. Organizations should state their requirements and expectations clearly. This assists candidates in deciding whether or not to accept the job. This eventually prevents further conflicts in employment terms.

## **12. Future Scope**

According to research findings, the primary causes of attrition in IT organisations are professional growth and organisational challenges. Although economic factors are the most influential, professionals may settle for the second best criteria of their preference, which is career growth and supportive work policies in the organisation. Candidates who want a better job than the one they have are more interested in getting the next one. Young talent wants to work on cutting-edge technology and functional domains. IT professionals in their early careers are less influenced by brand name or geographical area. Most IT professionals seek challenging roles and positions within organisations.

## 13. Appendix

### 13.1. Source Code

[https://github.com/IBM-EPBL/IBM-Project-25171-1659954502/blob/main/Final%20Deliverables/Final Code ML model Employee Attrition.ipynb](https://github.com/IBM-EPBL/IBM-Project-25171-1659954502/blob/main/Final%20Deliverables/Final%20Code%20ML%20model%20Employee%20Attrition.ipynb)

### 13.2 Github Repository Link

<https://github.com/IBM-EPBL/IBM-Project-25171-1659954502>

### 13.3 Demo video link

[https://drive.google.com/file/d/1XrilYZSwx6B5cXs-9wL3y5a75HYpRrQ0/view?usp=share link](https://drive.google.com/file/d/1XrilYZSwx6B5cXs-9wL3y5a75HYpRrQ0/view?usp=share_link)