CORPORATE EMPLOYEE ATTRITION

TEAM ID : PNT2022TMID02953

TEAM MEMBERS

KESAVAN M

MANIKANDAN M

MONISH KUMAR R

NIRMAL RAJ S

Industry Mentor(s) Name SHANAWAZ ANWAR

Faculty Mentor(s) Name : KALAIARASAN T R

SRI KRISHNA COLLEGE OF ENGINEERING AND TECHNOLOGY
COIMBATORE

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 dataset ‘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.

2. LITERATURE SURVEY

2.1 Existing Problem

- Various researchers framed different conceptual ideas and models based on their researches. In 1982, Bill Mobley suggested a model that is known as the Traditional model of attrition. Mobley suggests that the decision to quit the job is due to major dissatisfaction in the current job. According to Mobley, an individual try to evaluate his/her current job and level of satisfaction/dissatisfaction. If they are dissatisfied, they may have the idea of quitting and assess the cost of quitting and success of searching for an alternative job. After identifying the alternative job, decision is taken whether to stay or quit the job.

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)

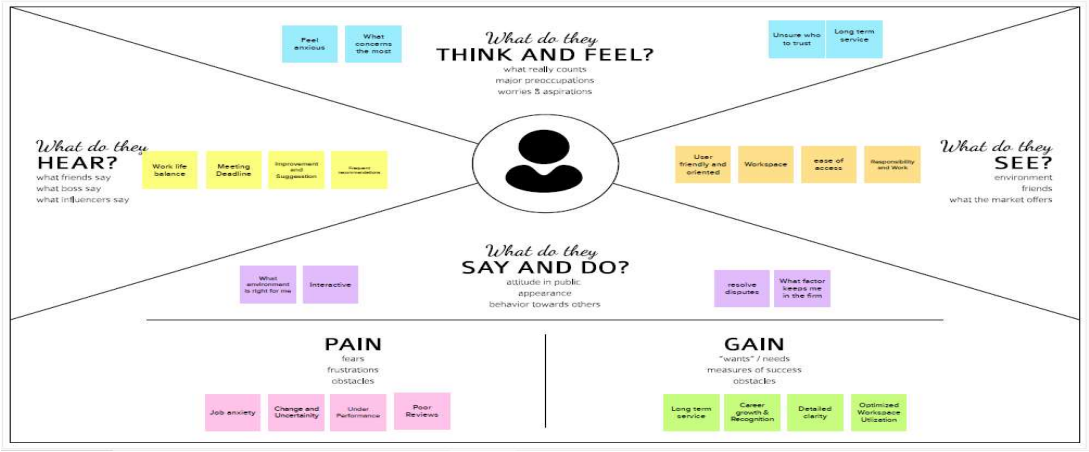
4.EMPLOYEE ATTRITION PREDICTION USING DEEP NEURAL NETWORK, Salah Al-Darraji, Dhafer G. Honi, Francesca Fallucchi, Ayad I. Abdulsada, Romeo Giuliano and Husam A. Abdulmalik,(3 November 2021)

2.3 Problem Statement Definition

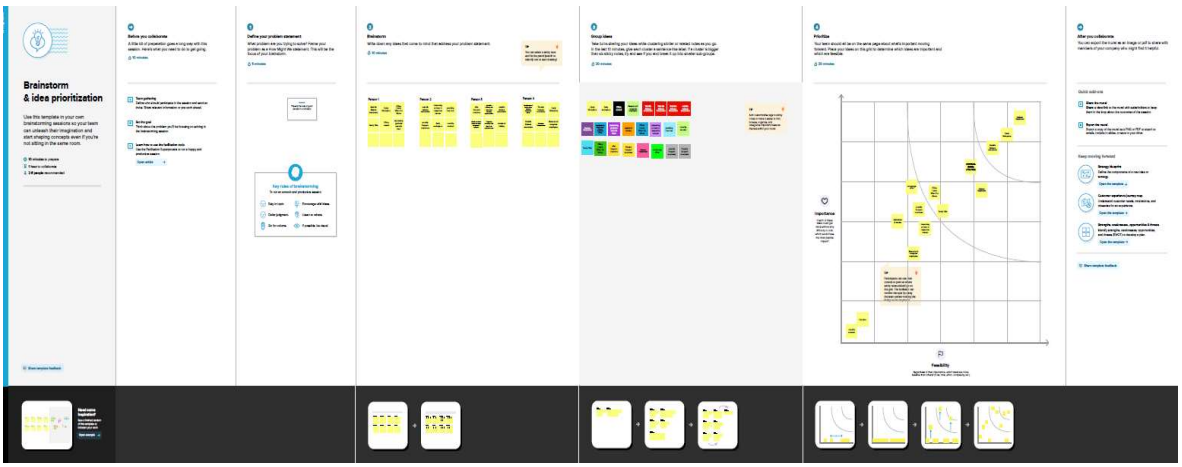
- To create a dashboard and perform analysis of employee attrition in corporates using IBM Cognos analytics platform.
- To reduce the employee attrition rate through data analytics, data visualization by analysing the major factors that causes attrition.

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



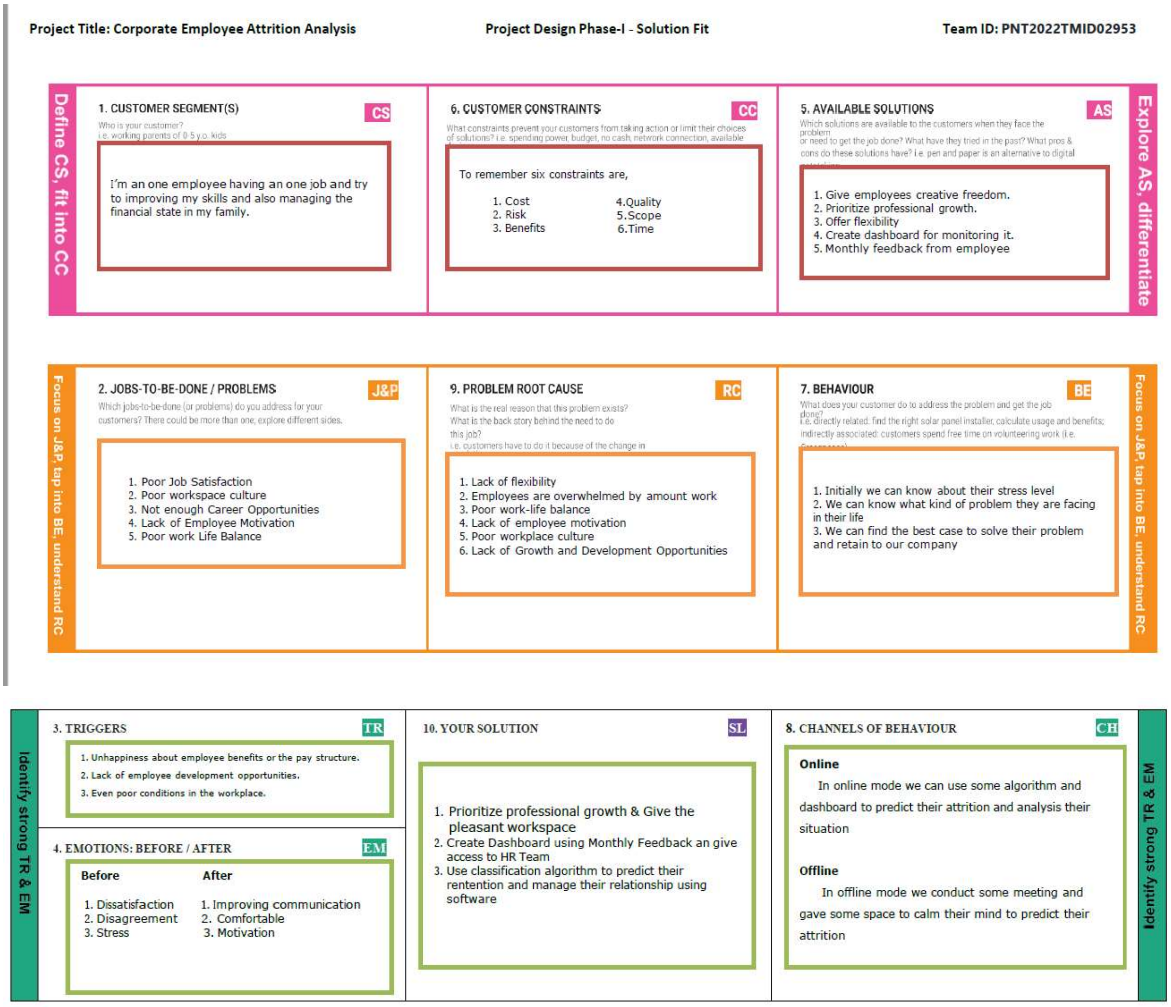
3.2 Ideation & Brainstorming



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 rapid growth in technology,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



4. REQUIREMENT ANALYSIS

4.1 Functional requirement

collect Dataset	Data from different sources are collected in order to get optimized result

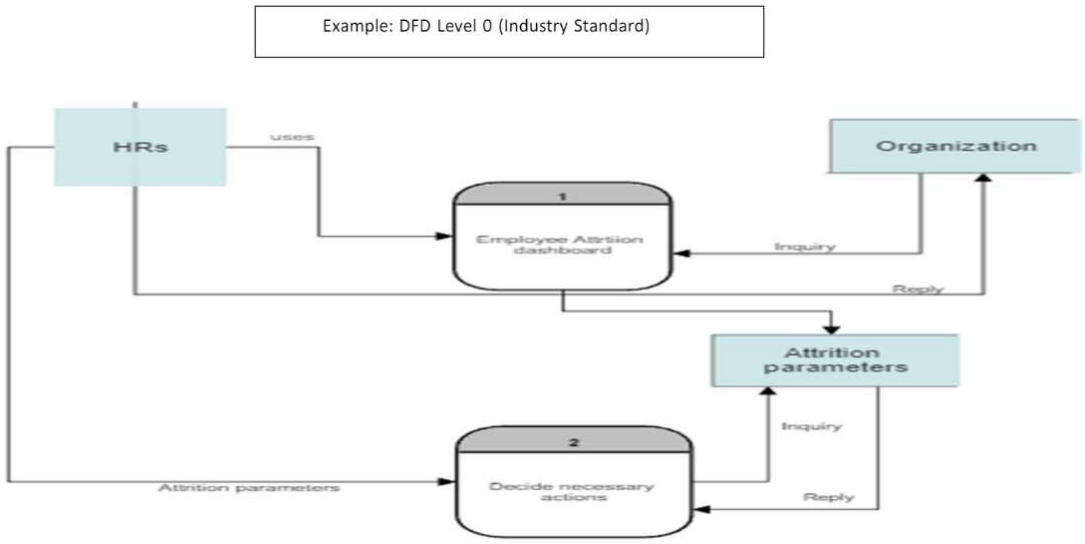
Data cleaning	When combining data from multiple sources there are duplicated data and hence we clean the data 1st
Data modelling	Identify the relationship between various parameters.
Prediction and analysis	The length of stay is predicted with the Machine learning algorithm
Functional Requirement (Epic)	**Sub Requirement (Story / Sub-Task)**

4.2 Non-Functional requirements

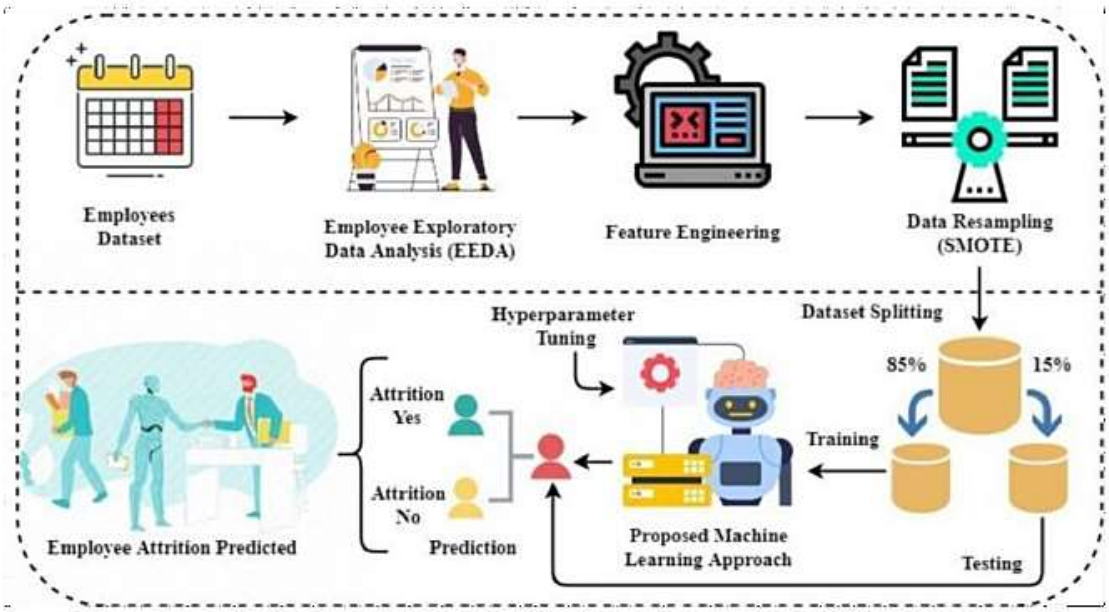
Non-Functional Requirement	Description
Usability	User can view and visualise the data through the interactive dashboard and predict the length of stay of patients with machine learning algorithm
Security	IBM Cognos provides better security. The dataset uploaded to the dashboard cannot be downloaded or accessed by external sources
Reliability	The dashboard and the prediction is very reliable and provide prediction with more accuracy
Performance	The length of stay of patients is predicted with more accuracy
Availability	The predicted length of stay and the visualization will be available in cognos analysis
Scalability	The software is scalable and extendable. Because it allow multiple user to handle the data at the same time

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
Employees	Registration	USN-1	The employees can register to be a part of the organization	I can access my account / dashboard	High	Sprint-1
		USN-2	As an employee, I will receive confirmation email	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As an employee, I can register for the application through G mail	I can get a verification link through email	Medium	Sprint-1
	Login	USN-4	As a employee, I can log into the application by entering email & password	I can enter the application	High	Sprint-2
	About	USN-5	I can view the Dashboard, Story and Report for attrition rates and determining the factors leading to them	I can get an idea about the project	Low	Sprint-2
	Launch	USN-7	As a HR, I can upload various analyzed parameters from the computer through link given in the PDF	I can choose any employee ('s all parameters) from my device	High	Sprint-2
	Link	USN-8	As a HR, I can review an employee's performance and offer appraisals biannually or Quarterly	I can view the employee's parameters on the dashboard along with the attrition rate.	High	Sprint-3
		USN-9	I can also upload csv format of employee retention parameters from cloud.	I can view the employee's parameters on the dashboard along with the attrition rate.	Medium	Sprint-3

6. PROJECT PLANNING

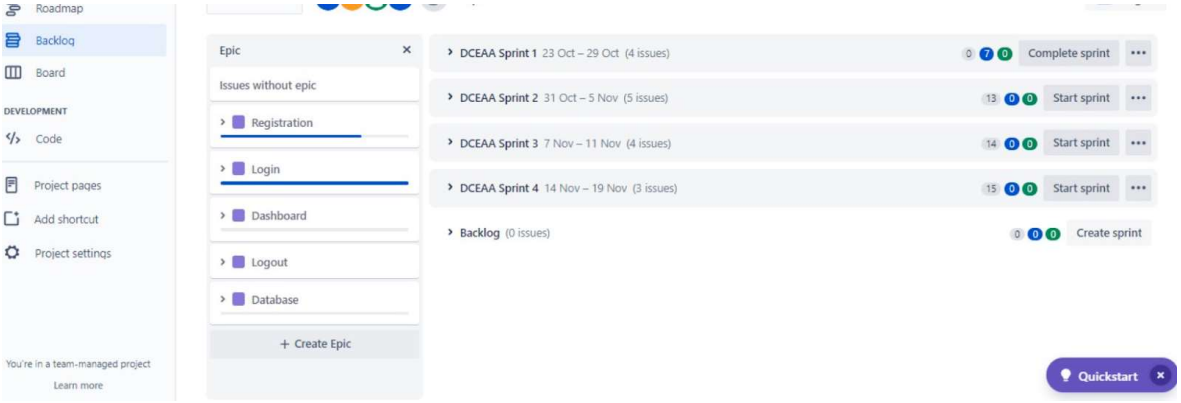
6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I should be able to register in the application.	3	Medium	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S
Sprint-1	Authentication	USN-2	The registered user should be authenticated and verified and logged in.	2	Low	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S
Sprint-2	Dataset upload and creating dashboards.	USN-3	As a user, I should be able to upload thedataset and do exploratory analysis andexplore patterns.	2	Medium	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S
Sprint-2		USN-4	I present the data using analytical tools and present the data using charts and graphs.	3	Medium	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S
Sprint-3	Model creation and testing	USN-5	I split the data into test and train dataand create the model.	5	High	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S
Sprint-4	Model Output	USN-6	The model is used to predict the attritionrate.	5	High	MONISH KUMAR R,KESAVAN M,MANIKANDAN M,NIRMAL RAJ S

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	5	6 Days	24 Oct 2022	29 Oct 2022	5	29 Oct 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	12Nov2022
Sprint-4	5	6 Days	14 Nov 2022	19 Nov 2022	5	19Nov2022

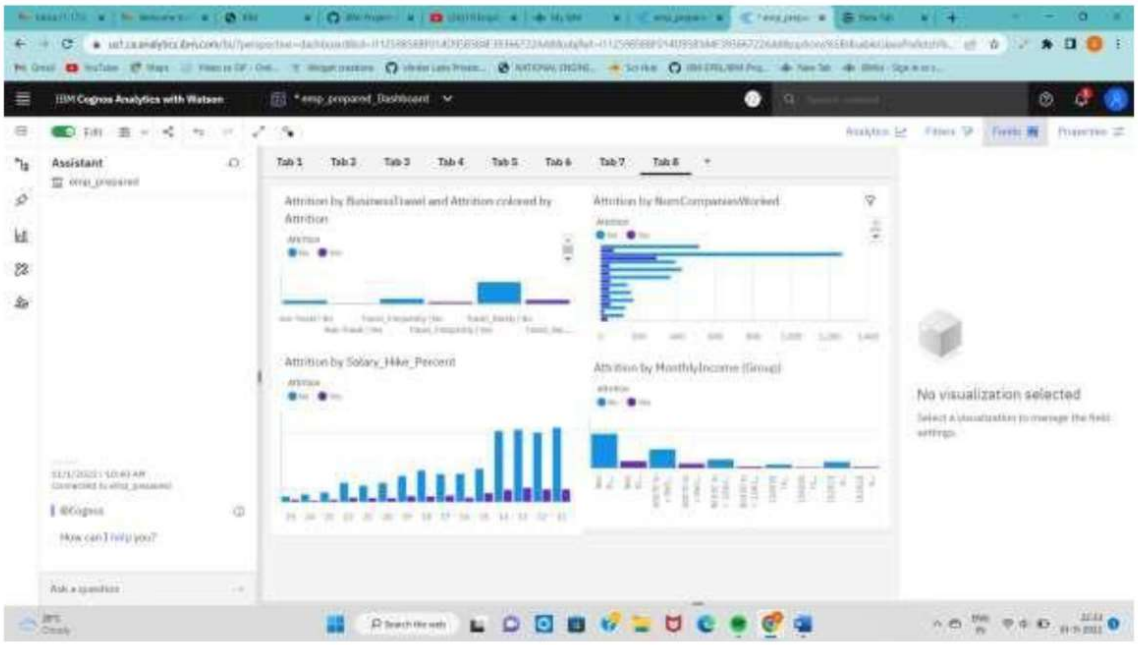
6.3 Reports from JIRA



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

7.1 Feature 1

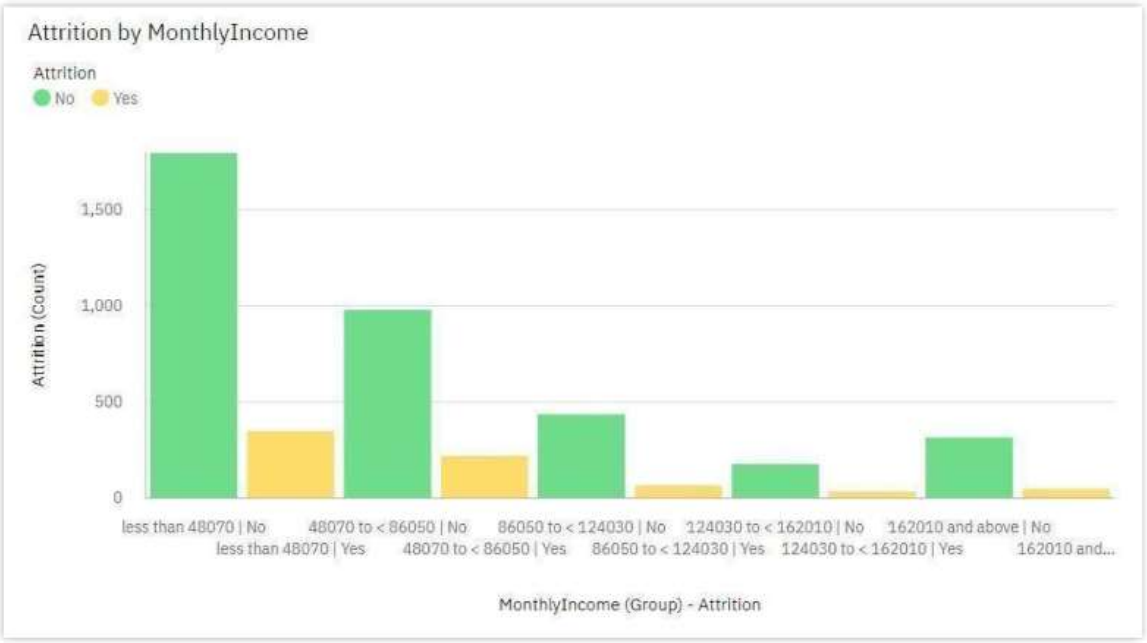
Dashboard



8. RESULTS

8.1 Performance Metrics


Attrition by Department
Attrition
No
Yes
0
20
40
60
80
100
120
140
160
180
200
Attrition by JobRole
Healthcare ...
Healthcare ...
Human Res...
Human Reso...
Laboratory ...
Laboratory ...
Manager | No
Manager | Yes
Manufacturi...
Manufacturi...
Research Di...
Research Di...
Research Sc...
Research Sc...
Sales Execu...
Sales Execu...
Sales Repre...
Sales Repre...
Attrition by EducationField
Attrition
No
Yes
iman Resources
Life Sciences
Marketing
Medical
Other
Technical...
MaritalStatus by MaritalStatus
102
24
69
93


Attrition by MonthlyIncome
Attrition
No
Yes
Attrition (Count)
1,500
1,000
500
0
less than 48070 | No
less than 48070 | Yes
48070 to < 86050 | No
48070 to < 86050 | Yes
86050 to < 124030 | No
86050 to < 124030 | Yes
124030 to < 162010 | No
124030 to < 162010 | Yes
162010 and above | No
162010 and...
MonthlyIncome (Group) - Attrition

9. ADVANTAGES & DISADVANTAGES

Advantages

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

The research study involves reference of both primary and secondary data. 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 organisations 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.

Limitations and Disclaimer

The research outcome is purely based on the experience, opinion, and the understanding level of the respondents. There is a scope of difference in results if the organisations under the survey are varied in geographical location. There may be limitations to generalize the findings of the survey completely

10. CONCLUSION

Employees as well as organizations must be clear with their expectations regarding the job profile. Any sort of mismatch leads to discrepancy and employees may fail to perform at their job. This eventually leads to attrition. Organizations should state the requirements and expectations unambiguously. This helps candidates decide upon to accept the job position or not. This eventually avoids further conflicts in the employment terms.

11. 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.

12. APPENDIX

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['NumCompaniesWorked']=df1['NumCompaniesWorked'].fillna(df1['NumCompaniesWorked'].mean())
```

```
In [ ]: df1['TotalWorkingYears']=df1['TotalWorkingYears'].fillna(df1['TotalWorkingYears'].mean())
```

```
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	1116	...	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

GITHUB LINK: https://github.com/IBM-EPBL/IBM-Project-24574-1659944796

DEMOLINK:https://drive.google.com/file/d/1gQ-G6hybB3OFP2EumLf2xESPsiU0MRlT/view?usp=drivesdk