TeamId : PNT2022TMID53479

Project Report

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

6.2 Sprint Delivery Schedule

6.3 Reports from JIRA

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

7.1 Feature 1

7.2 Feature 2

7.3 Database Schema (if Applicable)

8. TESTING

8.1 Test Cases

8.2 User Acceptance Testing

9. RESULTS

9.1 Performance Metrics

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

Source Code

GitHub & Project Demo Link

TeamId : PNT2022TMID53479

Date	17th November 2022
Team ID	PNT2022TMID53479
Project Name	Project - Corporate Employee Attrition Analytics
Team Members	Team Lead - Kesav S J Team Member 1 - Koushik K Team Member 2 -Deekshitha M Team Member 3 - Giridhar Prashanth

INTRODUCTION

The two primary components that contribute to the development and growth of the nation are corporate businesses and industries. Manpower, often known as the workforce, is a crucial component of any corporation. Performance and expansion of the business depend on how long the employees stay in their jobs. The fundamental difference between attrition and retention is that each has a different aim, but they are fundamentally related since one clears the way for the other. Global marketplaces are getting more competitive over time, which has altered workplace culture. The existence of the labor force, the emerging imbalance between the supply and demand of competent workers, and the growing importance of work-life balance have made it difficult for the company's HR and management to find the right candidate for the right role. The two faces that reflect the approach to determine business employment trends, general business growth, motivation, and growth are attrition and retention. Because losing a valuable employee has a negative impact on knowledge value, uneasy coworkers lost capital, and the organization's reputation, it is observed that globally competitive organizations spend a significant amount of interest, time, and money on employee attrition. This ultimately results in the failure of the business or organization.

1.1 PROJECT OVERVIEW

The organization's success depends on its ability to draw in and keep outstanding personnel. Identifying the factors that retain employees at the organisation and those that cause others to quit is a crucial responsibility for HR. A number of data points about the personnel who are either still employed by the firm or have left it are included in the data. To stop the company from losing talented employees, it is important to recognise and address these problems.

1.2 PURPOSE

a. To analyze the factors that causes the employee attrition through predictive analysis and to give suggestions by modelling techniques to reduce the cause of retention.

b. Visualization Charts are prepared to highlight the insights for the given dataset

c. Creating dashboard for the HR and managers for understanding the reasons for attrition and to take necessary measures in the organization.

2. LITERATURE SURVEY

Reduction in the number of employees in a company is referred to as employee attrition. Employee attrition has been a recognised problem for the corporate sector during the past twenty years. Employees depart from the company for a variety of reasons. Among the causes include the need for high pay, changes in technology or roles, obstacles in the workplace, etc. High attrition increases the cost of various company characteristics and functions. The overall cost to the employees is increased by recruitment, training, and development expenses.

2.1 EXISTING PROBLEM

Both the employer and the employee have recently lost faith in one another. The former believes that an employee can quit the company at any moment, whereas the latter believe that the former can dismiss the employee at any time. Regardless of who is at fault, a loss of workers is unavoidable. Attrition refers to

this labor loss, regardless of the cause. Attrition is a prevalent issue in any organization, regardless of the kind of business or organizational structure, which not only hinders output and results in high long-term expenses and a loss of goodwill to the organisation. Therefore, it becomes necessary to investigate this complex issue and find workable answers.

2.2 REFERENCES

1. **TITLE**: From Big Data to Deep Data to Support People Analytics for Employee Attrition Prediction.

YEAR: 2021

AUTHORS: Nesrine Ben Yahia; Jihen Hlel; Ricardo Colomo-Palacios

DESCRIPTION: In the era of data science and big data analytics, firms and their HR managers can reduce attrition by using people analytics, which transforms how businesses and their human resources (HR) managers find and keep talent. Staff attrition is a big problem for businesses in this situation since it affects both production and the continuity of planning. The main contributions that this study has made in this situation are listed below. We start by proposing a people's The analytics approach to employee attrition prediction shifts from a large data environment to a deep data one by focusing on data quality rather than quantity.

2. **TITLE**: Towards Understanding Employee Attrition using a Decision Tree Approach

YEAR: 2019

AUTHORS: Saadat M Alhashmi

DESCRIPTION: The severe issue of employee attrition has been the subject of research for several decades. This problem has been approached using a variety of methods, including psychological studies and exit interviews. The goal is to prevent or minimise employees leaving a company before hiring a replacement. Recently, researchers in the field of artificial intelligence have also addressed this problem due to the amount of data. With the aid of publicly available data and a decision tree approach, this study tackled the problem of staff attrition. The results of this work-in-progress study are encouraging, and subsequent work-studies will add more factors and test the model using data from a nearby supermarket.

3. **TITLE**: Employee Attrition System Using Tree Based Ensemble Method

YEAR: 2022

AUTHORS: Vimoli Mehta; Shrey Modi

DESCRIPTION: 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 declines, 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 long time to find a new employee because it takes him months to get trained and get used to the surroundings. Therefore, commercial organisations must take advantage of emerging trends and technology that uses machine learning algorithms. Companies can reduce this loss by knowing the cause of staff churn before it happens. Using the dataset "IBM HR Analytics Employee Attrition Performance"

and the tree-based Ensemble Machine Learning Model, this article offers a thorough analysis of employee attrition. The decision of an employee to quit the company is connected to a number of statistically important factors. To acquire the best outcomes from the currently available tree approaches, the study assesses the tree-based ensemble.

4. **TITLE**: Early Prediction of Employee Attrition using Data Mining Techniques

YEAR: 2019

AUTHORS: Sandeep Yadav; Aman Jain; Deepti Singh

DESCRIPTION: Take away our best 20 employees, and we [Microsoft] become a mediocre firm, according to a comment attributed to Bill Gates. Bill Gates' comment brought our attention to one of the main issues with employee churn in the workplace. Any firm must pay a hefty price for employee attrition (turnover), which could ultimately affect how efficiently it operates as a whole. According to CompData Surveys, total turnover climbed from 15.1 percent to 18.5 percent over the previous five years. Finding a qualified and experienced employee is a difficult endeavour for any firm, and replacing such workers is even more difficult. In addition to raising the major cost of human resources (HR), this has an effect on an

organization's market worth. Despite these realities, the literature that has contributed to numerous misunderstandings between HR and employees receives little attention. As a result, the purpose of this study is to present a methodology for predicting employee churn by applying classification algorithms to analyse the particular behaviours and qualities of the employee.

5. **TITLE**: Prediction of Employee Attrition Using data mining

YEAR: 2018

AUTHORS: R. Shiva Shankar; J. Rajanikanth; V.V. Sivaramaraju; K.V.S.S.R. Murthy

DESCRIPTION: Employee attrition has recently grown to be a significant issue in enterprises.

Employee attrition is a significant problem for firms, particularly when skilled, technical, and critical people leave for another company that offers greater opportunities. As a result, replacing a skilled person costs money. As a result, we examine the frequent causes of employee attrition using data on both present and historical employees. On the human resource data, we employed well-known classification algorithms, such as Decision tree, Logistic Regression, SVM, KNN, Random Forest, and Naive Bayes, in order to reduce employee attrition. To do this, we apply the feature selection approach to the data and analyse the outcomes to stop staff attrition. The ability to foresee employee turnover helps businesses expand economically by lowering the cost of their human resources

2.3 PROBLEM STATEMENT DEFINITION

Problem Statement

Employees are the most important part of an organization. Successful employees meet deadlines, make sales, and build the brand through positive customer interactions. Employee attrition is a major cost to an organization and predicting such attritions is the most important requirement of the Human Resources department in many organizations. In this problem, our task is to predict the attrition rate of employees in an organization. Among all employee-related problems, employee attrition is one of the key problems in today's scenario despite the changes in the external environment. Attrition is said to be a gradual

reduction in the number of employees through resignation, death, and retirement. A high attrition rate indicates that the employees have a lot of issues with the organization. Consequently, they'll only spread the bad word about the company. This will pose a huge risk to the company's reputation and make it difficult for the employer to find the right replacements. Every organization wants its valuable employees to be a part of its organization for a long period. Still, when many employees start leaving, it will be a

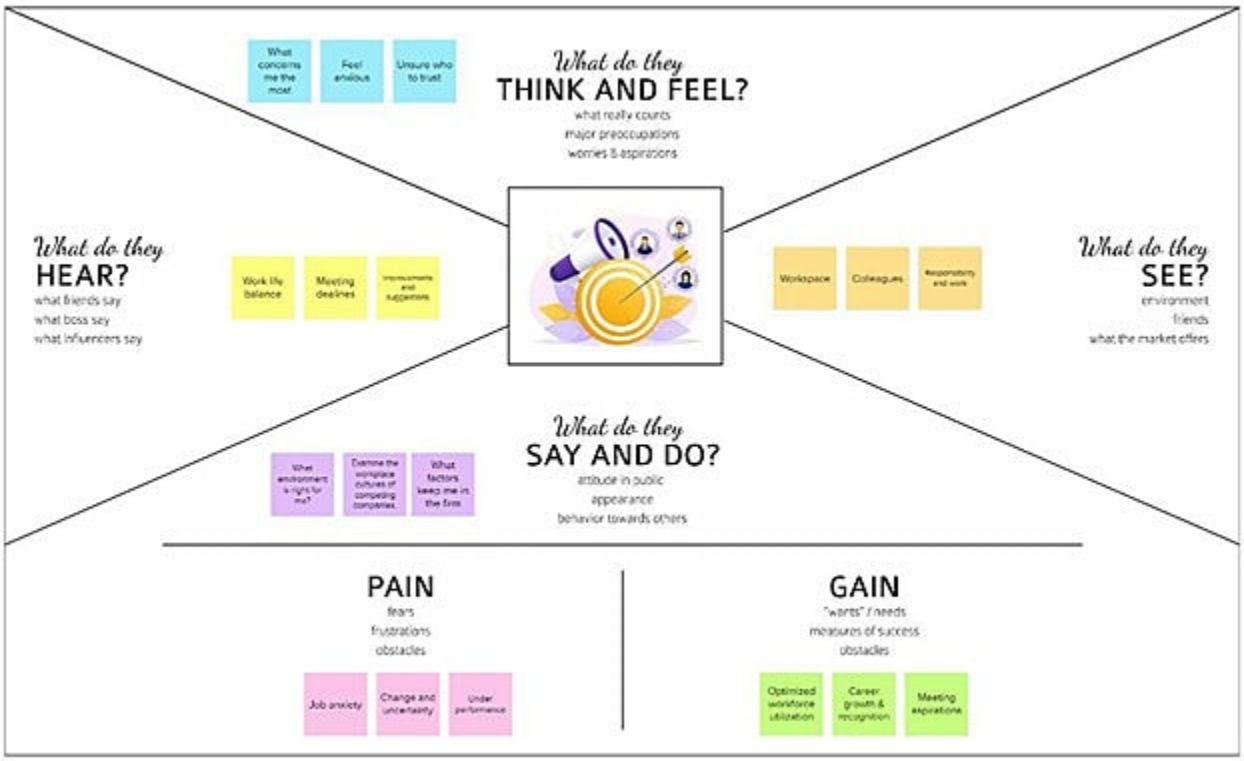
concern for the organization. The key to success for any organization is attracting and retaining top talent. One of the key tasks is to determine which factors keep employees at the company and which prompt others to leave. It's more cost-effective to keep the employees a company already has. • A company needs to maintain a pleasant working atmosphere to make their employees stay in that company for a longer period. To reduce the cost of attrition, organizations need to ensure that employees' aspirations are met.

Business Model/Impact

• Organizations can use this tool to manage the team. • Reduction in Hiring Cost

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



TeamId : PNT2022TMID53479

3.2 IDEATION AND BRAINSTORMING

1)Collection of Ideas

Deekshitha M	• Study about Employee Attrition rates. • Collection of Data • Finding out Real world causes for Attrition • Incorporating impactful factors for attrition such as inflation • Emotional Factors of Employees to be considered • Time, Work Patterns and Changing lifestyle influencing employee attrition • Information Extraction from Employees' Statements • testing the reliability of data • Formulate solutions for lowering attrition
Koushik K	• Data Collection • Checking the Credibility of the Data • Deciding the Algorithm to be used for the Analysis • Inferring the reason for attrition manually • Performing Analytics using various methods • Inferring the insights • Comparing the Results got with the previous results • Influencing Factors are segregated and re-checked • Deriving outcomes and preventive measures to lower attrition
Giridhar Prasanth	• Understanding what makes employees unhappy • segregating the available data • Data collection of employee's emotions • analyzing with past survey results • find out the solutions • choosing best algorithm for analysis • Using data to predict attrition risks • find the root cause of the problem and predict when employee leave • Building a custom employee retention model

Kesav S J	• Collection of Data from various sources • Finding out the results of past surveys • Testing the correctness of the Data • Incorporating Past survey results with our available data • Joining more datasets and forming a single reliable dataset • Providing insights based on various conditions • Representing the Data in form of Dashboards • Finding out the impact percent of each factor • Providing solutions for retaining employees based on the analysis

2) Grouping of ideas

TeamId : PNT2022TMID53479

1)Data collection

- Collection of Data from various sources
- Segregating the available data
- testing the realiability of data
- Testing the correctness of the Data
- Checking the Credibility of the Data
- Incorporating Past survey results with our available data
- Joining more datasets and forming a single reliable dataset
- Testing the correctness of the Data

2)Manual Insights

- Inferring the reason for attrition manually
- Finding out Real world causes for Attrition
- Understanding what makes employees unhappy

3)Analytics

- Performing Analytics using various methods
- choosing best algorithm for analysis
- Finding out the results of past surveys
- Finding out the impact percent of each factor
- Comparing the Results got with the previous results

4)Factors

- Incorporating impactful factors for attrition such as infation
- Emotional Factors of Employees to be considered
- Finding out Real world causes for Attrition
- Time, Work Patterns and Changing lifestyle infuencing employee attrition
- Understanding what makes employees unhappy

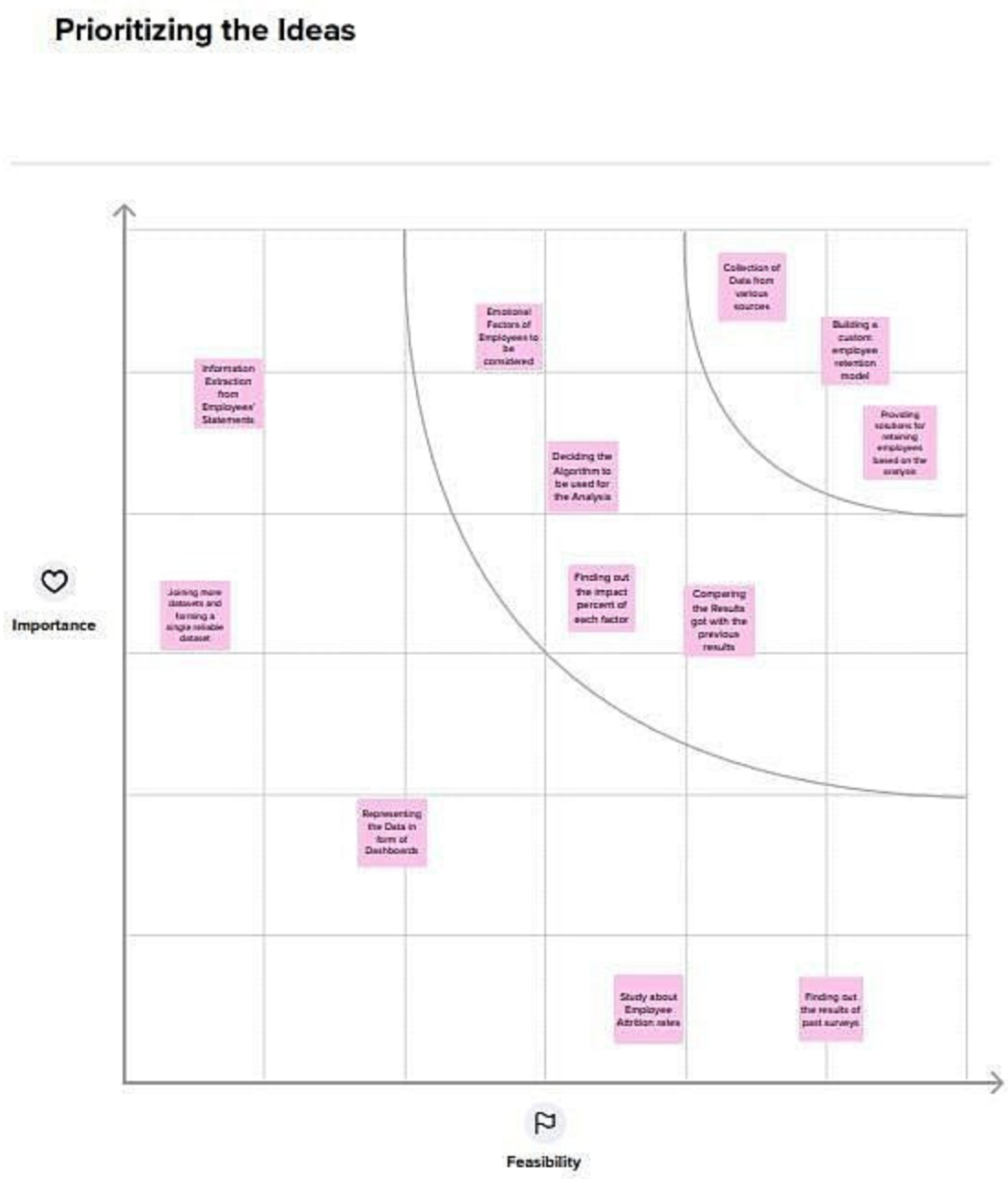
- Infuencing Factors are segregated and re-checked

5)Insights

- Information Extraction from Employees' Statements
- Formulate solutions for lowering attrition
- Inferring the insights
- Representing the Data in form of Dashboards
- Building a custom employee retention model
- Comparing the Results got with the previous results
- Deriving outcomes and preventive measures to lower attrition
- Using data to predict attrition risks
- find the root cause of the problem and predict when employee leave
- Providing solutions for retaining employees based on the analysis
- Providing insights based on various conditions

TeamId : PNT2022TMID53479

3) Prioritization



TeamId : PNT2022TMID53479

3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Losing productive people would directly affect the growth of any organization. Given the data of employees working or resigned, the task is to analyse the data and find out the factors which lead the employees to leave the organization. This will help in retaining the employees and reduce the attrition rates.
2.	Idea / Solution description	Based on the results of the analysis of employee attrition, improving on the factors that lead the employees to leave the organization, maintaining good relationship with the employees and promoting personal career growth would have a positive impact on the retention of employees.
3.	Novelty / Uniqueness	Analysing the given data along with external survey results obtained from employees directly. This will help in improving the accuracy of the results.

4.	Social Impact / Customer Satisfaction	Reduction in the loss of valuable employees could be achieved. The Software directly benefits the customer by providing insights on the specific factors which need to be improved. The above factors subsequently lead to the growth of the company as well as customer satisfaction.
5.	Business Model (Revenue Model)	We plan to implement this application using a subscription-based model. Based on the number of employees, the subscription plans may differ.
6.	Scalability of the Solution	This software will be scalable for any organization as it runs only on the particular company's employee dataset. Implementing this software with the help of cloud service providers helps in increasing the scalability.

TeamId : PNT2022TMID53479

3.4 PROBLEM SOLUTION FIT

<table>
<tr><td>CUSTOMER SEGMENT(S)

1. HR
2. Talent Acquisition Team
3. Head Hunter Organization</td><td>6. CUSTOMER CONSTRAINTS
Multitudinous factors that are diffcult to take into consideration for manual analysis.</td><td>5. AVAILABLE SOLUTIONS
Predict whether an employee will stay in the organization for a period of time.</td></tr>
<tr><td>2. JOBS-TO-BE-DONE / PROBLEMS
Develop solution to identify factors resposible for employees to leave an organization</td><td>9. PROBLEM ROOT CAUSE
1. Unsatisfactory work life balance
2. Low pay.
3. Toxic working environment
4. No scope for growth.</td><td>7. BEHAVIOUR
Re-negotiate salary and promotion.</td></tr>
<tr><td>3. TRIGGERS
Talented work force leaving the organization to work for the competitors.</td><td rowspan="2">10. SOLUTION
Use historic data of employee, previous employer and survey to find the factors resposible.</td><td rowspan="2">8. CHANNELS of BEHAVIOUR

Offline - Resigning</td></tr>
<tr><td>4. EMOTIONS: BEFORE / AFTER
In Control.</td></tr>
</table>

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

TeamId : PNT2022TMID53479

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	User needs to give their email and password.
FR-2	Login Page	Login with username and password.
FR-3	List of Analytics	The webpage consist of many analytical dashboards.
FR-4	User Dashboard	Take the data given by user and interactive dashboard can be created.
FR-5	Analysis and Estimation	Analyze the corporate employee attrition from the data and estimate corporate using the Data Driven Approach i.e., Cognos Analytics with Watson.

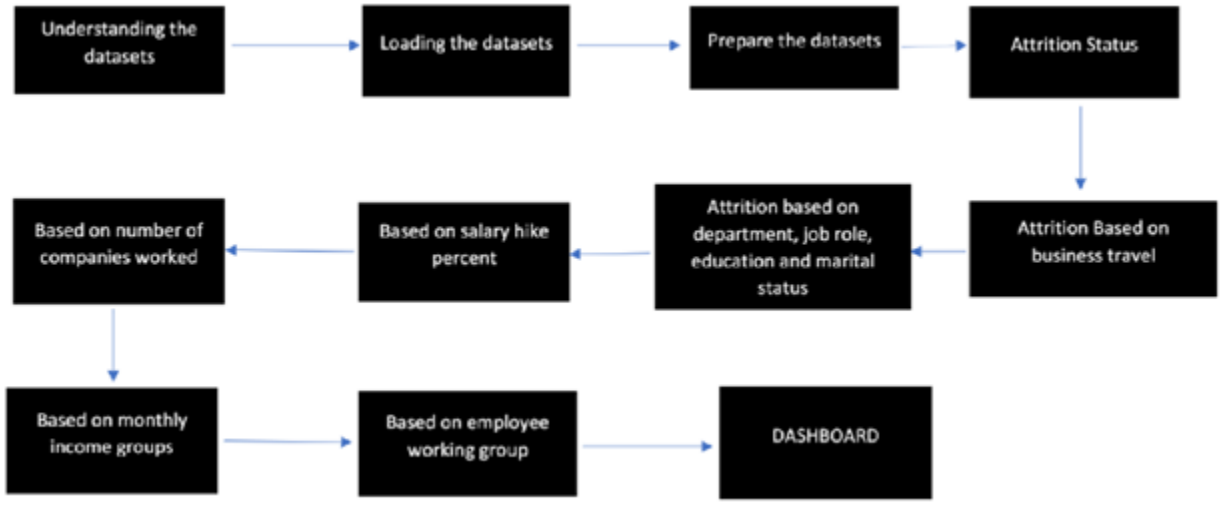
4.2 NON FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	All the data which is needed will be displayed in one which is easily understandable and will be useful for user to enhance the corporate employee attrition with higher accuracy and also, they can get the In sights of employees.
NFR-2	**Security**	Only recognized users can access the resource.
NFR-3	**Reliability**	A new Visualization and dashboard that is added or erased it won't affect other dashboards.

NFR-4	**Performance**	Data analytics helps in executing the existing algorithms faster with large data sets. Therefore, it will be helpful to people and Business persons to gain profit in their business.
NFR-5	**Availability**	By using the technique of data analysis, resource allocation can be made to make resources available at any time to achieve high profit.
NRF-6	**Scalability**	The data stored can be viewed and retrieved at any time and any where.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

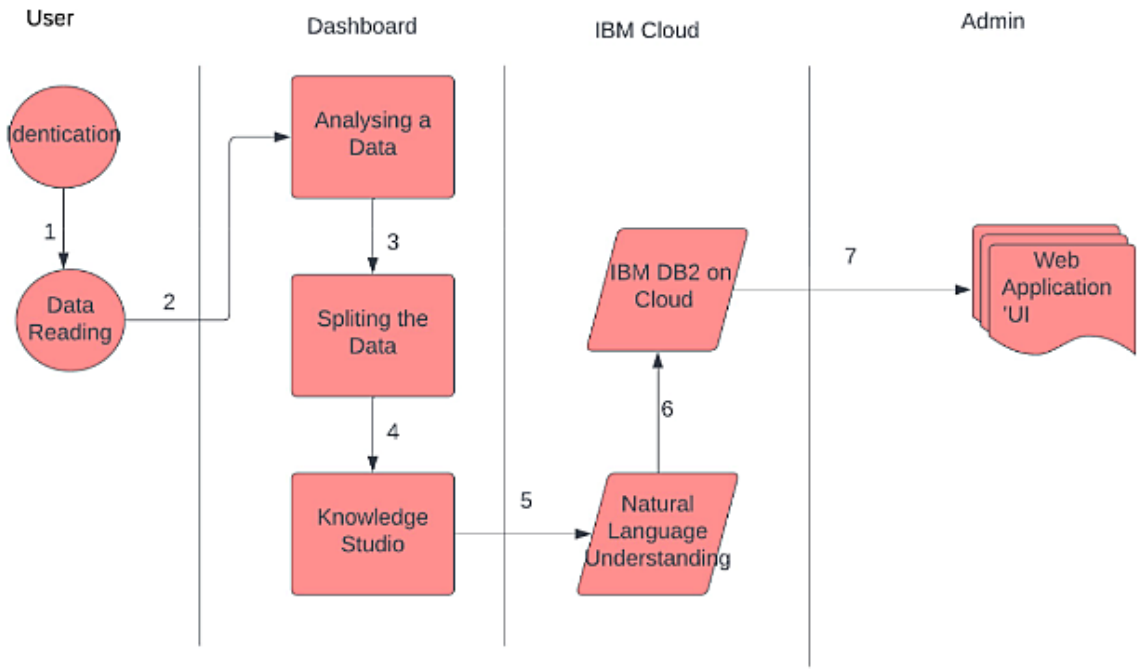


TeamId : PNT2022TMID53479

5.2 SOLUTION & TECHNICAL ARCHITECTURE

1. The Process involves cleaning the input dataset first.
2. Datasets are from various sources including the given one, plus the survey results obtained from employees
3. Data preprocessing is then done to remove all unnecessary or unstructured data and also to make it structured ▪ After Pre-processing, using a machine learning algorithm (Supervised learning), we are classifying the common factors leading to attrition.
4. Also, the prediction of future attrition rates is projected with the available data ▪ Finally, the output is displayed to the user

Solution Architecture



TeamId : PNT2022TMID53479

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Number	User Story / Task Story	Acceptance criteria	Priority	Release
Customer (CEO)	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
Customer (Employee)		USN-2	As an employee, I can register for the application by entering my mail, password,and confirming password.	I can access my account/dashboard	High	Sprint-1
		USN-3	As a user, I can register for the application	I can register & dashboard with login	Medium	Sprint-2
Customer (CEO)	Login	USN-4	As a user, I can log into the application by entering email & password	I can access my account/ dashboard	High	Sprint-3
Customer (Employee)		USN-5	As a user, I can log into the application by entering email and password.	I can access my account /dashboard	High	Sprint-3
CEO	Dashboard	USN-6	As a CEO, I can use the	I can view the	HIgh	

			predict button to know which factor keeps the employee at the company and which prompts others to leave.	visual chart.		Sprint-4
Employee		USN-7	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 & SCHEDULING6

6.1 SPRINT PLANNING & ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dashboard	USN-1	As a user, I give the details of the employees working in our organization for the attrition detail.	5	High	Kesav S J Koushik K, Deekshitha M, Giridhar Prashanth
Sprint-1		USN-2	As an Analyst, I will check the dataset and perform exploratory data analysis in Cognos Analytics	3	High	Kesav S J Koushik K, Deekshitha M, Giridhar Prashanth
Sprint-2	Report	USN-3	As a user, I want Simpler limited number of	2	Low	Kesav S J Koushik K,

			visualizations that report a particular event			Deekshitha M, Giridhar Prashanth
Sprint-2		USN-4	As an Analyst, I will use Cognos Analytics to generate a report	3	Medium	Kesav S J Koushik K, Deekshitha M, Giridhar Prashanth
Sprint-3	Story	USN-5	As a user, I can only understand the Analysis in animated presentation of	3	Medium	Kesav S J Koushik K, Deekshitha M, Giridhar Prashanth
Sprint-3		USN-6	As an Analyst, I use Cognos Analytics to create an animated presentation (Story) of the dataset	3	Medium	Kesav S J Koushik K Deekshitha M Giridhar Prashanth
Sprint-4	Predictive Analysis	USN-7	As a user, I want to predict the attrition rate of the company from the dataset	5	Medium	Kesav S J Koushik K Deekshitha M Giridhar Prashanth
Sprint-4		USN-8	As an Analyst, I will perform Prediction Analysis by utilizing various libraries in python	3	High	Kesav S J Koushik K Deekshitha M Giridhar Prashanth

TeamId : PNT2022TMID53479

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	12 Nov 2022
Sprint-4	5	6 Days	14 Nov 2022	19 Nov 2022	5	19 Nov 2022

We have a 6-day sprint duration, and the velocity of the team is 5 (points per sprint). To calculate the team's average velocity (AV) per iteration unit (story points per day)

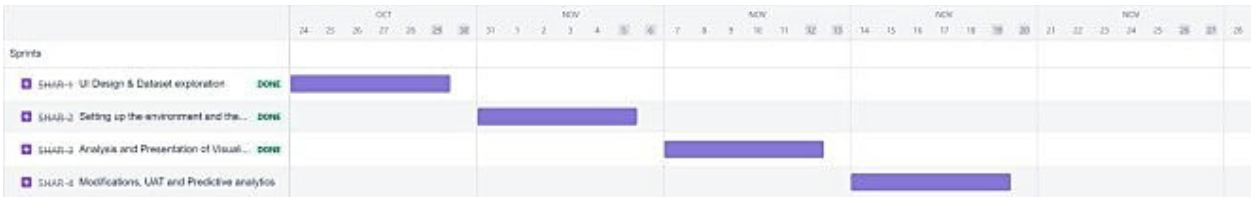
AV = SPRINT DURATION/VELOCITY

= 6/5

= 1.2

6.3 REPORTS FROM JIRA

TeamId : PNT2022TMID53479



7. CODING & SOLUTION

7.1 - FEATURE 1

In this project , we have done visualization by considering several criterias like

- ➤ age
- ➤ gender
- ➤ department
- ➤ business travel
- ➤ number of companies worked
- ➤ monthly income, etc.

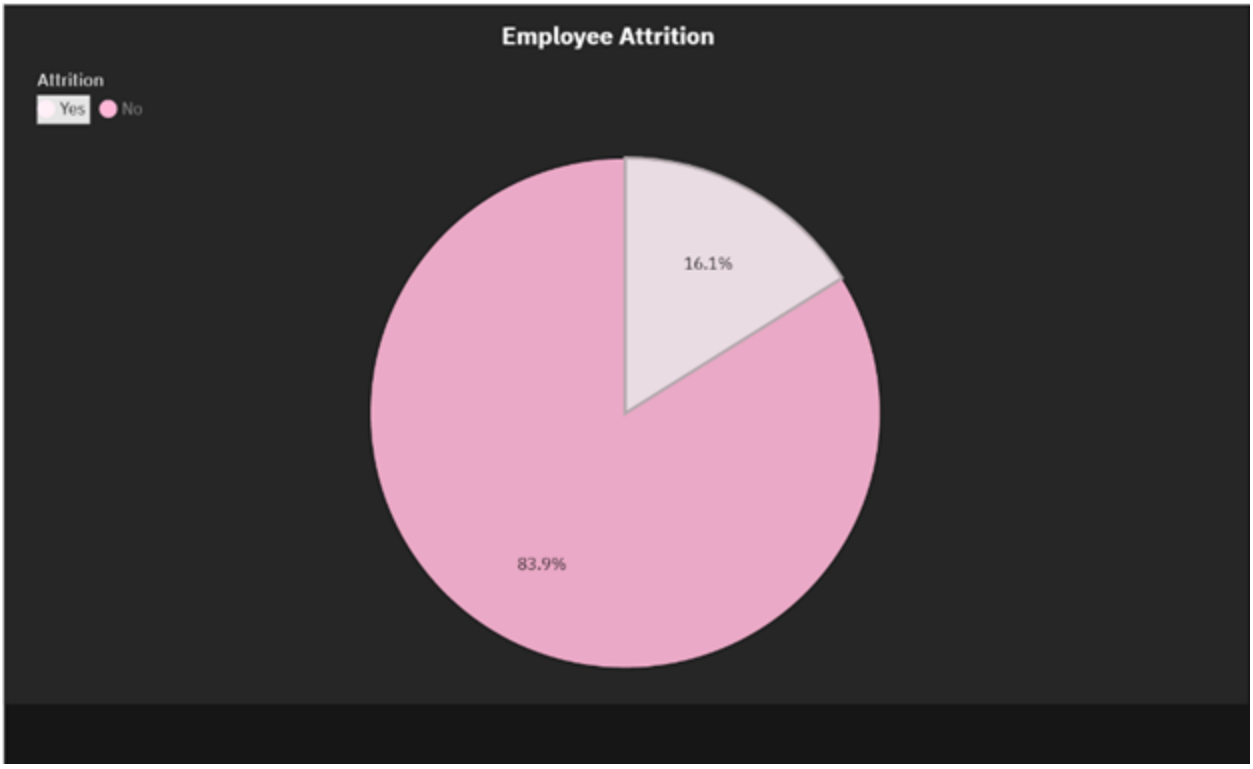
Considering all this during the visualization process makes it more accurate to exactly identify the root caue for the attrition of the employees.

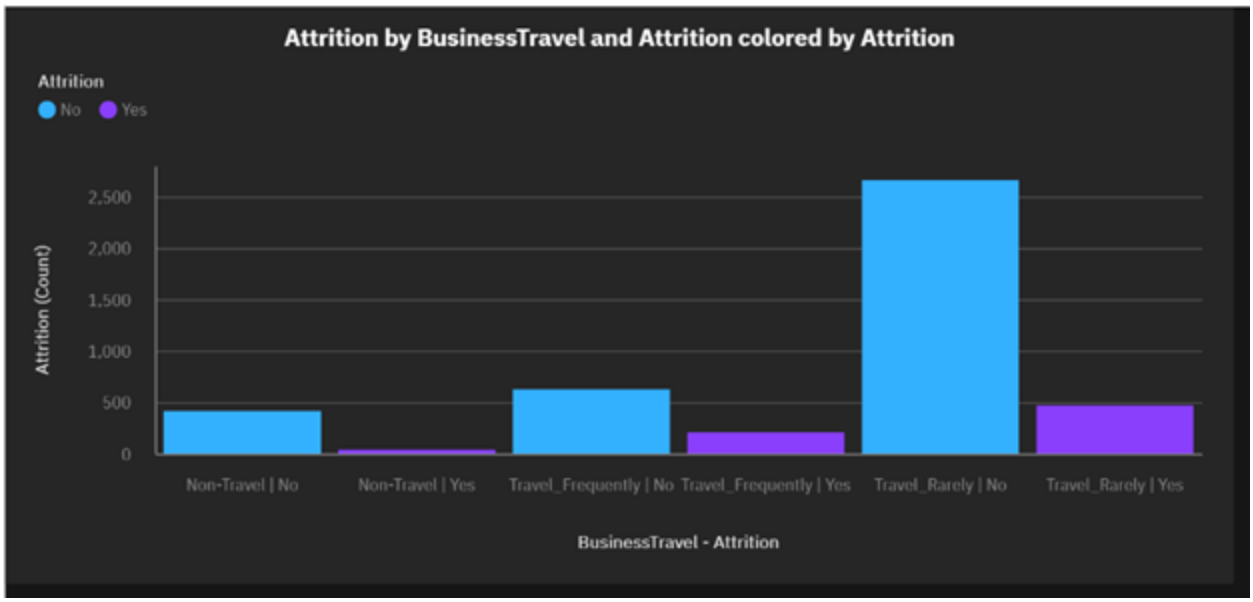
7.2 FEATURE 2:

The dataset is also understood by various factors to consider the missing or unnecessary values in it. Python is used inorder to make the process quite easy and visualisation is also performed using python.Random forest classifier is used in training and testing the datasets which yields almost 97% of accuracy

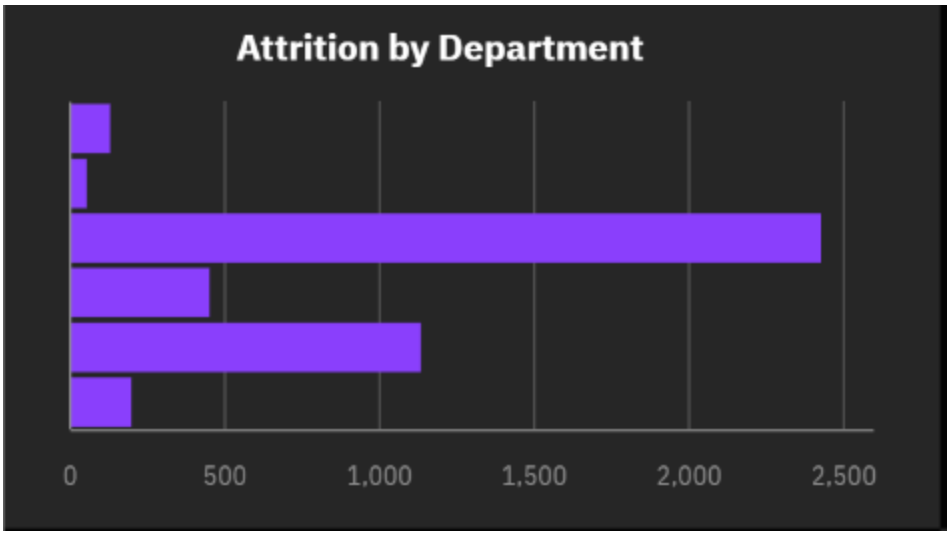
Interactions and Correlations

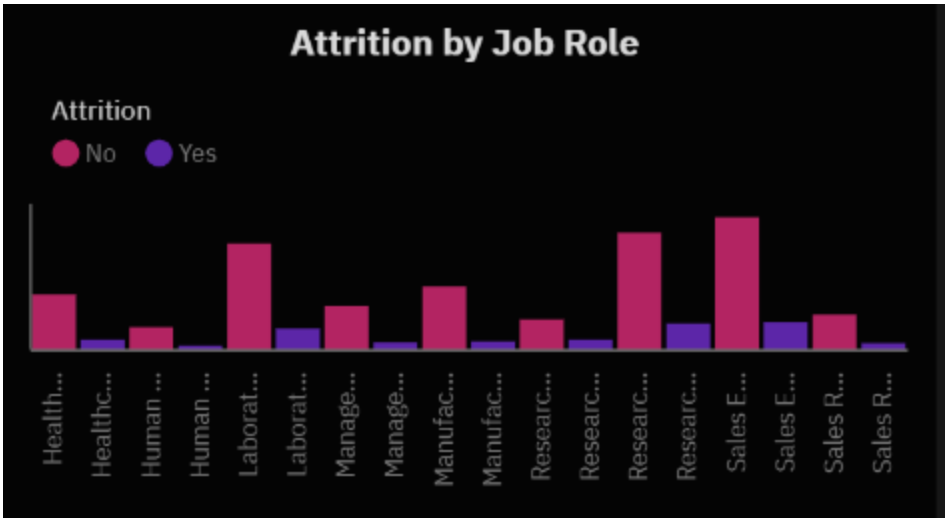
TeamId : PNT2022TMID53479

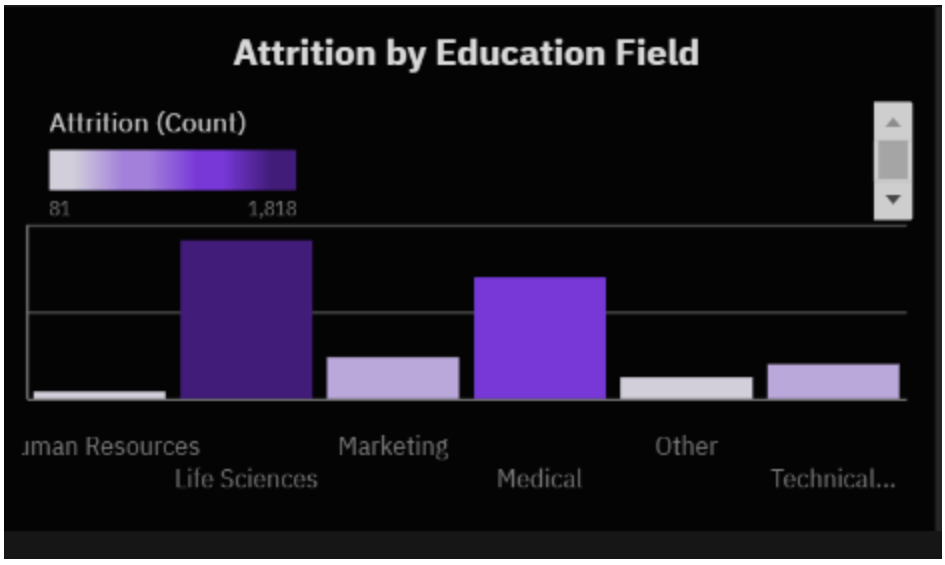




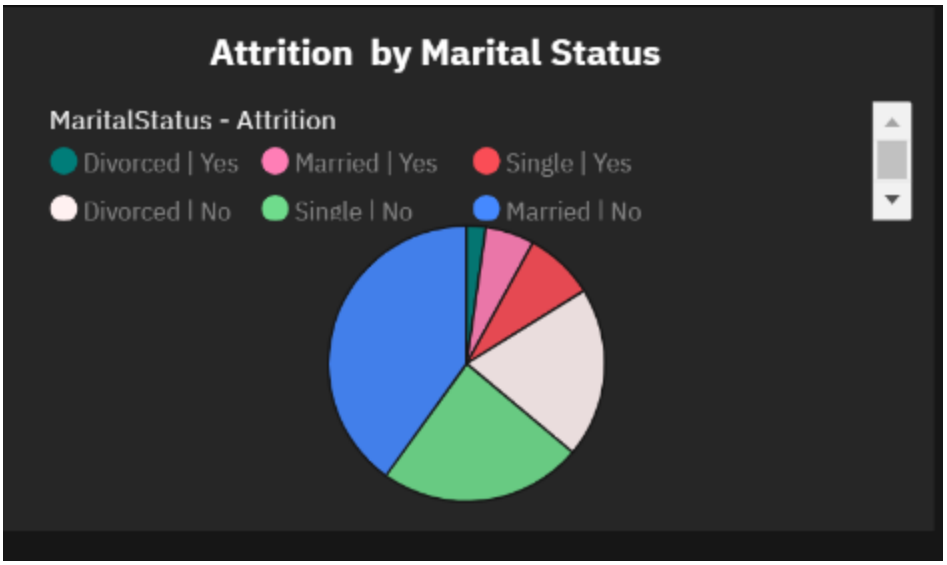
TeamId : PNT2022TMID53479

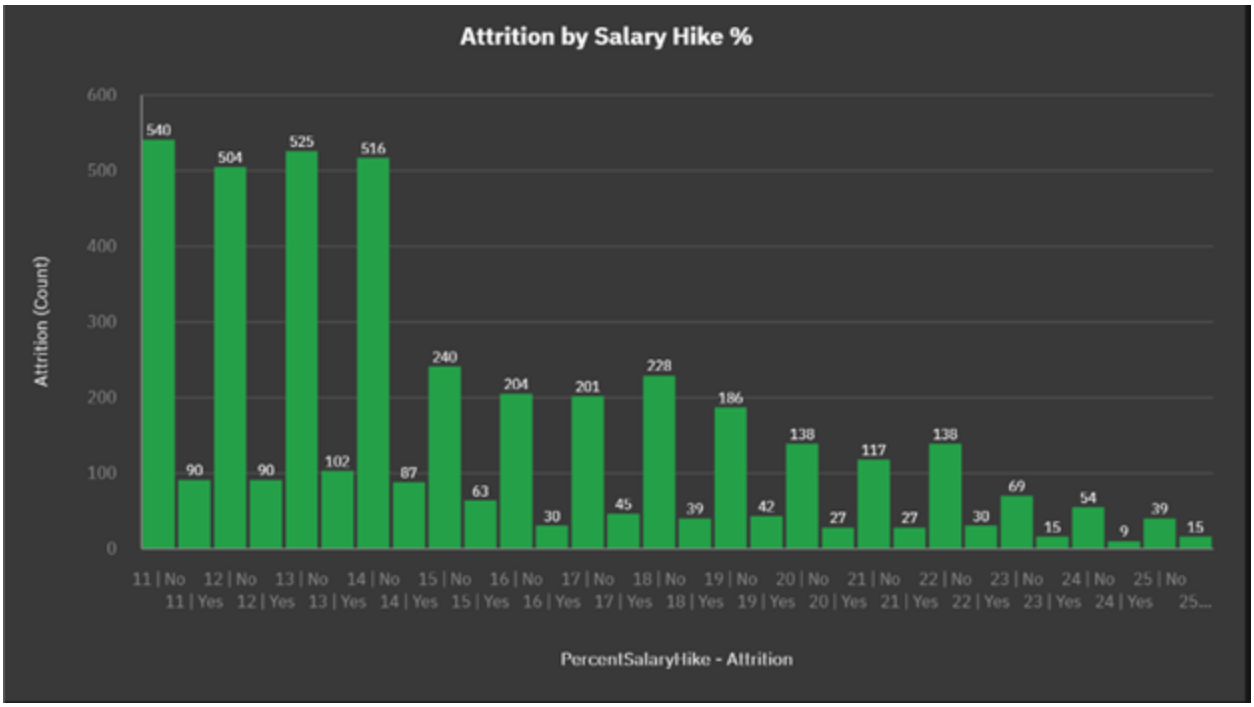




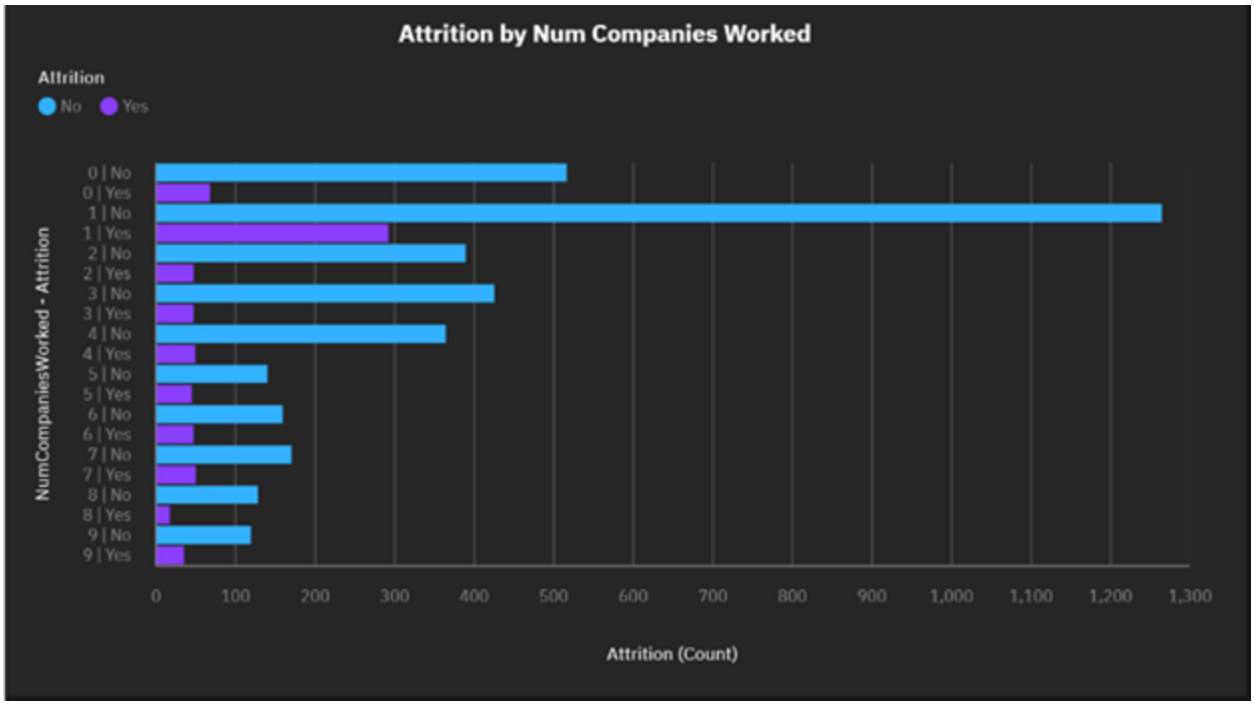


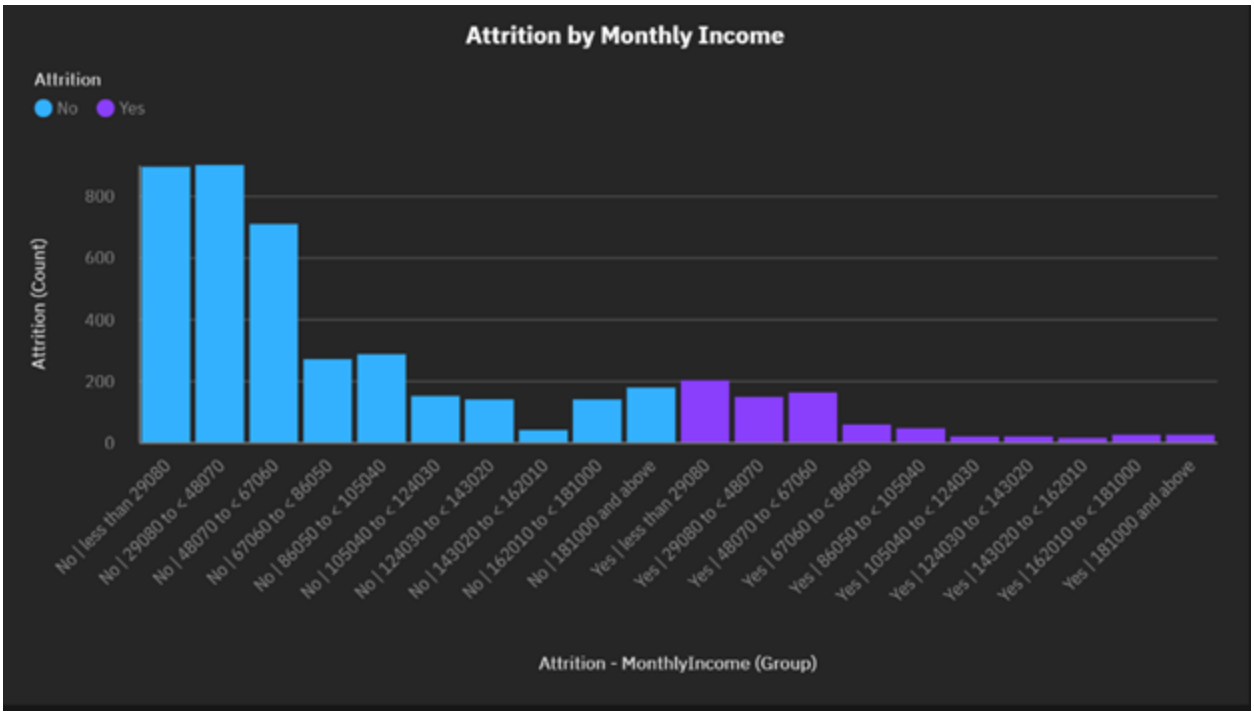
TeamId : PNT2022TMID53479



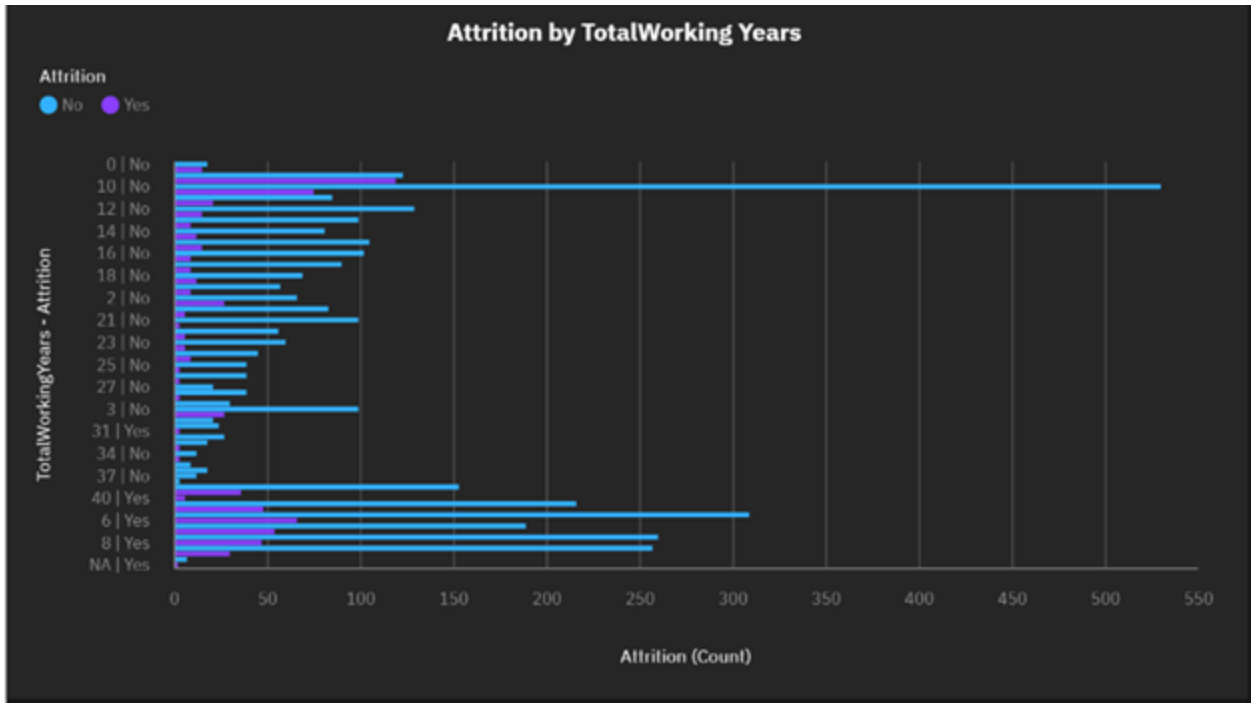


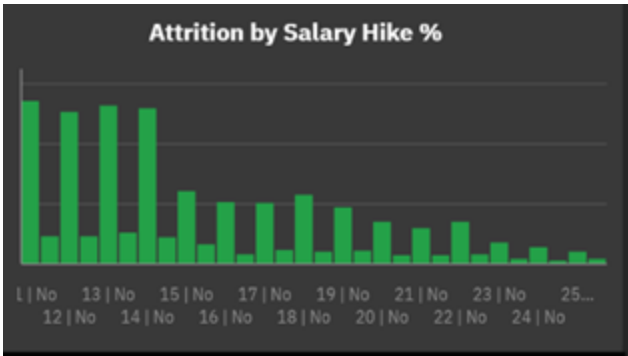
TeamId : PNT2022TMID53479

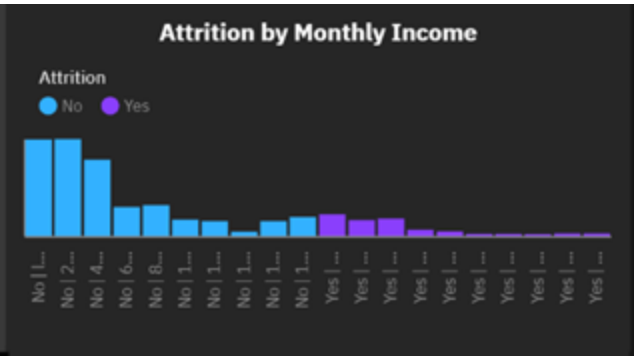


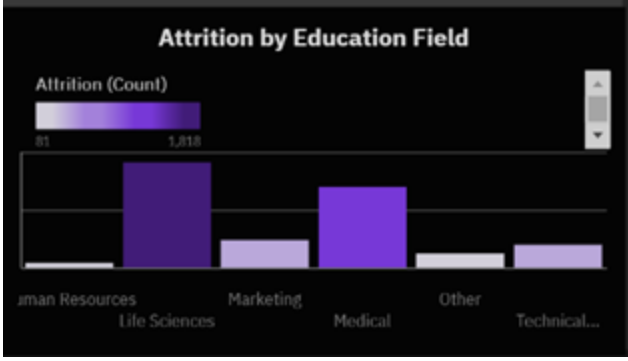


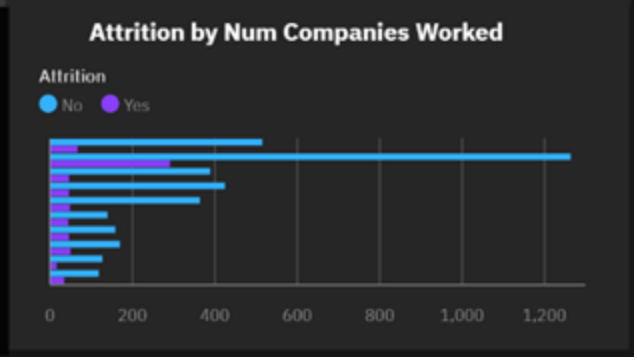
TeamId : PNT2022TMID53479











8. TESTING

8.1 TEST CASES

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status
CSV File upload	Functional	General_data.csv	Upload Successful/unsuccesful	Frontend Layout and button for uploading the CSV File	1.To check and prepare the data 2.To write python codes for uploading the csv file	https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	Uploads and Reads successfully	Working as expected	Pass
IBM Cognos Dashboard Embedment	Functional	General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	Verifying whether the dashboard of cognos analytics shows up in the web application	Cleaning and preparation of data	1.To test each data for test split 2.We need to write python code for each test split	https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	The should be passed as package	Working as expected	pass
Interactions	Functional	General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	Checking whether the interaction graph between any two components we select is working as expected	Cleaning and preparation of data	1.To prepare and clean the data and 2.To write python codes for each parameters	https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	The should be passed as package	Working as expected	pass
Correlations	Functional	General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	Correlations between all the available fields in 5 different methods	Cleaning and preparation of data	1.To prepare the data 2. To write python code for each parameter	https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	The parameter required for modelling can be identified and result is seen as heat plot	Working as expected	pass
EDA	Functional	General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	Development of Model	Cleaning and preparation of data	1.To find the parameter required for modelling 2. To write python code	https://www.kaggle.com/datasets/vjchoudhary7/hr-analytics-case-study General_data.csv, Employee_Survey_Data.csv, Manager_Survey_data.csv	The case is passed and the result is seen as bar graph	Working as expected	pass

8.2 USER ACCEPTANCE TESTING

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	1	2	1	0	3
Duplicate	1	0	0	0	1
External	2	0	0	1	3
Fixed	7	2	3	0	12
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	1	0	0	1
Totals	11	5	6	2	23

TeamId : PNT2022TMID53479

Test Case Analysis

Section	Total Cases	Not Tested	Fail	Pass
CSV File upload	2	0	0	2
IBM Cognos Dashboard embedment	5	2	0	3
Interaction charts	4	0	0	4
Correlations	1	0	0	1
EDA	1	0	0	1

9. RESULTS

9.1 PERFORMANCE METRICS:

1. Attrition status by age :

➤ visualization performed by column chart

➤ Age by status = 92%

2. Employee count by department :

Visualization performed by bar chart

Employee count by department wise

➤ Human resource = 17%

➤ R&D = 89%

➤ Sales = 60%

3. Attrition based on business travel :

Visualization performed by waterfall chart percentage by business travel

➤ Non-travel =25%

➤ Travel frequently =75%

➤ Travel rarely =35%

➤ Sum=100%

4. Attrition based on department ,job role ,education &marital status:

Visualization performed by line & column chart percentage by

Department wise

i. Human resource =15%

ii. R&D =85%

iii. Sales =45%

iv. Education =69%
v. Job role =100%
vi. Marital status
➤ Male =80%
➤ Female =20%
5. Attrition based on salary hike percentage :
➤ Visualization performed by pie chart
➤ Salary hike percentage (overall) =95%(based on department wise)
6. Based on No.of companies worked:
Visualization performed by stocked column chart
No. of companies worked based on attrition
i. Human resource =15%
ii. R&D =65%
iii. Sales =35%
7. Visualization based on monthly income groups :
Visualization performed by scatterplot chart
➤ Monthly income percentage = 100%
8. Prediction based on employee working groups :
Visualization performed by network chart employee working groups
➤ percentile =75%
DASHBOARDS:
1. Attrition based on department by age department (visualization performed by bar chart)
➤ Human resource =17%
➤ R&D =89%
➤ SALES =60%
➤ OVERALL =91%
2. Analysis based on job involvement in daily rate :
Visualization performed by heat plot chart job involvement
➤ percentage =99%
3. Based on attrition :
➤ Visualization performed by scatterplot
➤ Attrition percentage =66%
➤ Business travel =99%
4. Calculating the employee performance :
Visualization performed by column chart
➤ No.of companies worked =59%
➤ Performance rating =84%

10. ADVANTAGES & DISADVANTAGES

Advantages:

1.Higher manpower cost

2.Stronger employee relationships

3.Setting a culture right

4.High performance

5.Improve employee satisfaction

6.Increased productivity

7.Increased Revenue

8.Morale improvement

Disadvantages:

1.Lack of knowledgeable people

2.Decreased overall performance

3.Poor work life balance

4.Create a negative image

5.Huge risk on company reputation

11.CONCLUSION

The following suggestion are given based on the analysis and modeling result:

CURRENT EMPLOYEES:

•Work life balance should be improved

•Work environment should be improved

•The manager of an employee should not be changed very often

•Employees should be provided relevant training regularly, especially for its younger employees

FUTURE EMPLOYEES (CHANGES IN HIRING PROCESS):

The company should follow either one of the strategies given below –

•Hire older people with decent work experience

•Hire young people and train them appropriately

12. FUTURE SCOPE

The future scope of the research is that these analysis and modeling helps in forecasting the cause of employee disengagement, enables HR managers develop long-term strategies to reduce attrition, Competitive measures to enhance company brand image, Develops and shapes drills that benefit both the management and the employees. The scope of this research can be extended to many numbers of samples and to other working fields other than corporations.

13. APPENDIX

Nowadays, employee attrition has become a serious issue regarding a company's competitive advantage. It's very expensive to find, hire and train new talents. It's more cost-effective to keep the employees a company already has. A company needs to maintain a pleasant working atmosphere to make their employees stay in that company for a longer period. A few years back it was done manually but it is an era of machine learning and data analytics. Now, a company's HR department uses some data analytics

tool to identify which areas to be modified to make most of its employees stay.

Source code:

```
# -*- coding: utf-8 -*-
"""corporate employee attrition analytics.ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/15UEJzaJpCsT7FtqqrsY8xUU_PpdObUUK
"""

#Description: This program predicts employee attrition

#Import the libraries
import numpy as np
import pandas as pd
import seaborn as sns

# load the data
from google.colab import files
uploaded = files.upload()

#Store the data int a dataframe
df = pd.read_csv('WA_Fn-UseC_-HR-Employee-Attrition.csv')

#Print the first 100 rows
df.head(100)



#Get the rows and columns
df.shape

#Get the column data types
df.dtypes
```

```
# Get a count of the empty values of each column
df.isna().sum()

#Check for any missign or null vlues in the data
df.isnull().values.any()

# View some statistics
df.describe()

#Get  a count of the number of  employee  that stayed and left the company
df['Attrition'].value_counts()

#Visualize the number of employees that stayed and left the company
sns.countplot(df['Attrition'])

#Checking the accuracy
print((1233-237)/1233)

#Show the number of employees that left and stayed by age
import matplotlib.pyplot as plt
plt.subplots(figsize=(12,4))
sns.countplot(x='Age',hue='Attrition',data=df,palette='colorblind')

#Print all of the datatypes and their unique values
for column in df.columns:
  if df[column].dtype == object:
    print(str(column) + ' : ' + str(df[column].unique()))
    print(df[column].value_counts())
    print('_____')

#Removing some unnecessary columns
df=df.drop('Over18',axis=1)
df=df.drop('EmployeeNumber',axis=1)
df=df.drop('StandardHours',axis=1)
df=df.drop('EmployeeCount',axis=1)
```

```
#Get the corelation
df.corr()

#Visualize the corelation
plt.figure(figsize=(14,14))
sns.heatmap(df.corr(),annot=True,fmt= '.0%')

#Transform the data
#Transform non-numerical into numerical columns
from sklearn.preprocessing import LabelEncoder
for column in df.columns:
  if df[column].dtype == np.number:
    continue
  df[column] = LabelEncoder().fit_transform(df[column])

#Create a new column
df['Age_Years'] = df['Age']

#Drop the age column
df = df.drop('Age',axis=1)

#Show the data frame
df

#Split the data
X = df.iloc[:,1:df.shape[1]].values
Y = df.iloc[:,0].values

#Split the data into 75% training and 25% testing
from sklearn.model_selection import train_test_split
X_train, X_test,Y_train, Y_test =train_test_split(X,Y, test_size = 0.25,random_state=0)

# Use the Random forest classifier
from sklearn.ensemble import RandomForestClassifier
forest = RandomForestClassifier(n_estimators=10,criterion='entropy',random_state=0)
forest.fit(X_train,Y_train)
```

```
#Get the accuracy on the training dataset
forest.score(X_train,Y_train)

#Show the confusion matrix and accuracy score for the model on the test data
from sklearn.metrics import confusion_matrix

cm= confusion_matrix(Y_test,forest.predict(X_test))
TN=cm[0][0]
TP=cm[1][1]
FN=cm[1][0]
FP=cm[0][1]

print(cm)
print('Model testing Accuracy = {}'.format((TP+TN) / (TP+TN+FN+FP)))
```

Github - https://github.com/IBM-EPBL/IBM-Project-17384-1659662477

Demo-
https://drive.google.com/file/d/1xBJv9yXhm8IVU2fZ92jCn4fGdD40_oDW/view?usp=share_link