Corporate Employee Attrition Analytics

(A PROJECT COMPLETION REPORT)

Team ID:PNT2022TMID54360

Submitted by,

1.S Kirti Sri (Reg No.310619106062)
2.R Krithika(Reg No.310619106064)
3.K G Nikitha(Reg No.310619106088)
4.H Mohammed Razeen(Reg No.310619106081)

DEPARTMENT OF ELECTRONICS COMMUNICATIONENGINEERING

EASWARI ENGINEERING COLLEGE,RAMAPURAM CHENNAI

(An Autonomous Institution affiliated to Anna University Chennai)

1.INTRODUCTION

1.1 PROJECT OVERVIEW

The overview of this project is to predict the attrition rate for each employee, to find who's more likely to leave the organisation.It will help organization to find ways to prevent attrition or plan in advance the hiring of the new candidate.Attrition proves to be a costly and time-consuming problem for the organization and it also leads to the loss of probability.The scope of the project extends to companies to all industries.

1.2 PROJECT PURPOSE

The key to success in any organization is attracting and retaining top talent. As an HR analyst one of the key task is to determine which factors keep employees at the company and which prompt others to leave. Given in the data is a set of data points on the employees who are either currently working within the company or have resigned. The objective is to identify and improve these factors to prevent loss of good people.

2.LITERATURE SURVEY

2.1 Existing problem

In certain cases employee attrition depends more on employees' satisfaction level as compared to other attributes.In this context, employee attrition presents a critical problem and a big risk for organizations as it affects not only their productivity but also their planning continuity.Employee attrition presents a critical problem and a big risk for organizations as it affects not only their productivity but also their planning continuity. In this context, the salient contributions of this research are as follows.

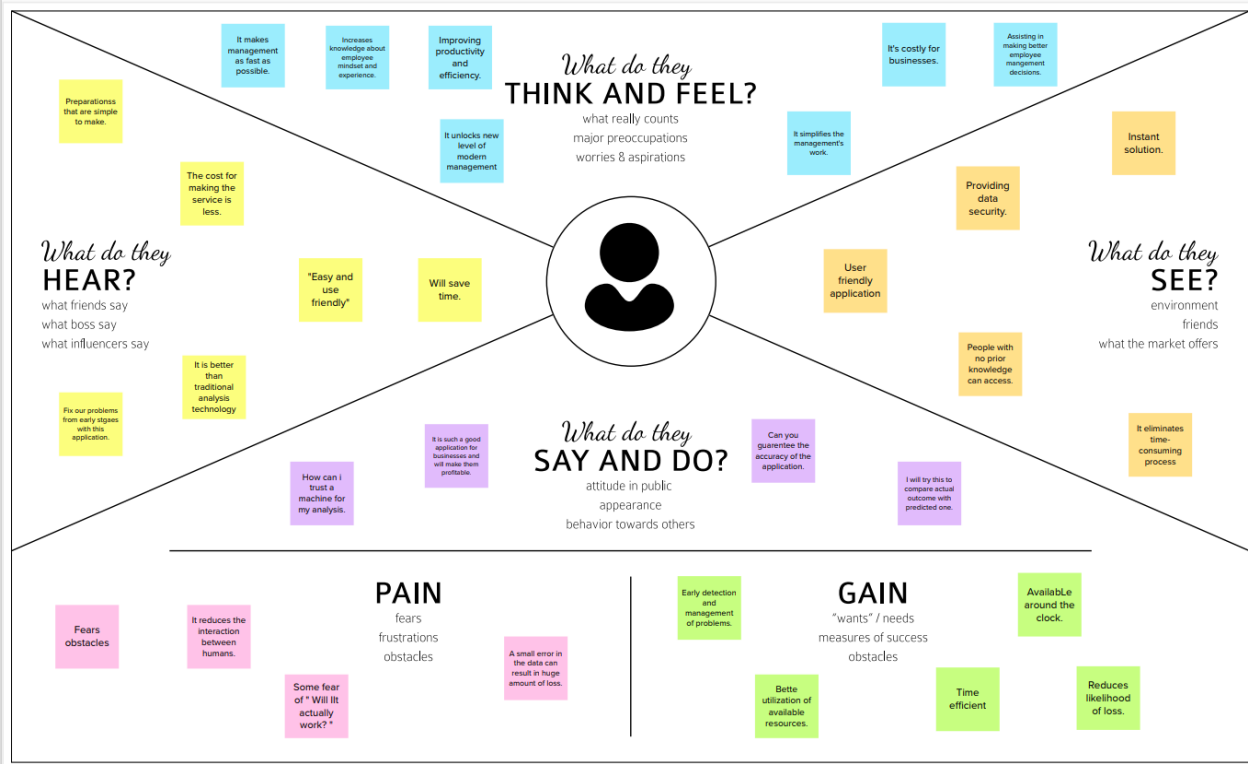
2.2Problem Statement Definition

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, the task is to predict the attrition rate of employees of an organization.

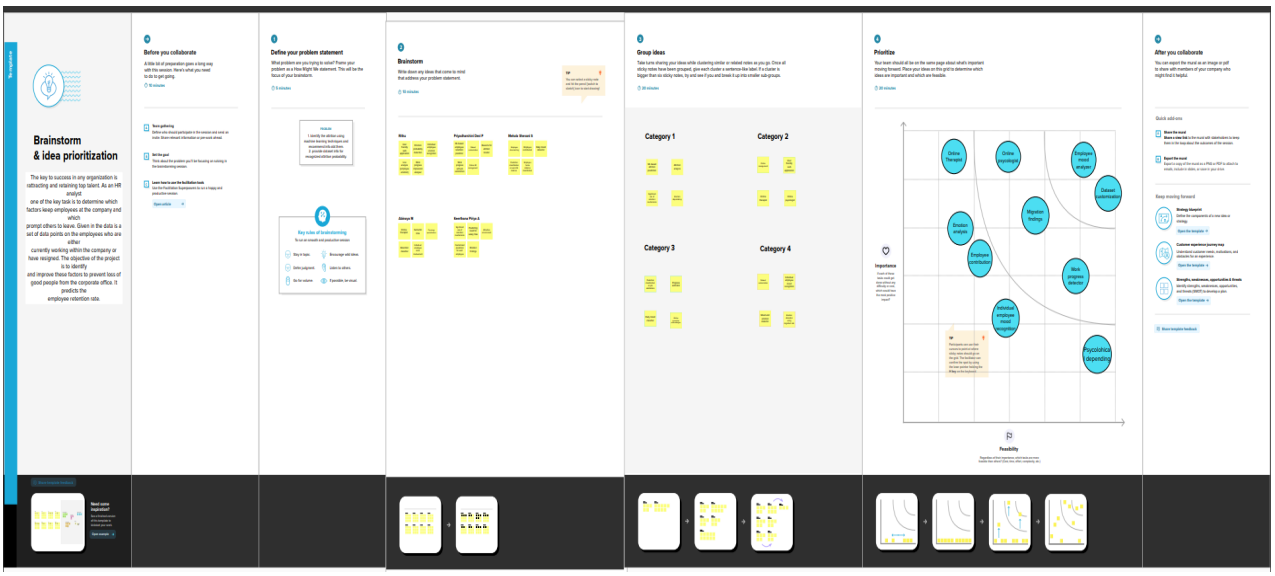
What does the platform affect?	Corporate HR managers.
What are the boundaries of the problem?	Management that faces the issues of employee attrition.
What is the issue?	In organizations, employees are leaving and it might be a concern for the organization. It is more cost-effective to keep the employees a company already has.
Why is it important to fix the problem?	It is required for the growth of the organization. It is important to retain the employees.
What solution to solve this issue?	An automated system is introduced to identify different reasons for employees leaving the organization by environment, job satisfaction, work-life balance.
What methodology used to solve the issue?	Machine learning techniques are used to predict the employee attrition in an organization.

3.IDEATION & PROPOSED SOLUTION

3.1Empathy Map Canvas



3.2 Ideation & Brainstorming



3.3 Proposed Solution

3.3 Problem Solution fit

Project Title: Corporate Employee Attrition Analysis | **Project Design Phase-I - Solution Fit Template** | **Team ID:** PNT2022TMIDxxxxxx

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

HRs of corporate companies

6. CUSTOMER CONSTRAINTS CC

- Spending power
- Budget constraints
- Resource constraints

5. AVAILABLE SOLUTIONS AS

A dashboard to maintain the various reasons for employees leave their jobs, and to provide tips to HRs on tackling different issues related to attrition.

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

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

To build a supervised machine learning model based on regression algorithms for forecasting the reason for employee attrition based on the following parameters.

- Work progress
- Employee satisfaction
- Subsidiaries
- Distance between offices & homes

9. PROBLEM ROOT CAUSE RC

- Manually analyzing reasons for employee attrition is a time - consuming process.
- Due to this, before identifying the cause the employee would have left the job and valuable human resource is lost.
- Lots of human effort is required.

7. BEHAVIOUR BE

- The model behaves so as to reduce the human efforts required to consolidate and confess the various reasons for employee attrition by manually analyzing them.
- It also brings the most important causes for employee attrition to spotlight for the main customers the HRs of the corporates.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

4.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
FR-2	User Confirmation	Confirmation via Email
FR-3	User Authentication	Authenticate the user's attempt to login using the database
FR-4	Retention analysis	Employee attrition analysis by sentiment, work environment, daily contribution etc.
FR-5	Employee management	Validating and managing the registered employee details.
FR-6	Progress management	Add the progress of each employee to the company.
FR-7	Predict button	The predict route is used for prediction and it contains all the codes which are used for predicting our results. Firstly, inside launch function we are having the following things: ● Getting our input and storing it. ● Select the necessary attributes for the prediction. ● Creating model. ● Predicting our results . ● Showcase the results with the help of dashboard. ● Finally run the application.

4.2 Non-Functional requirements

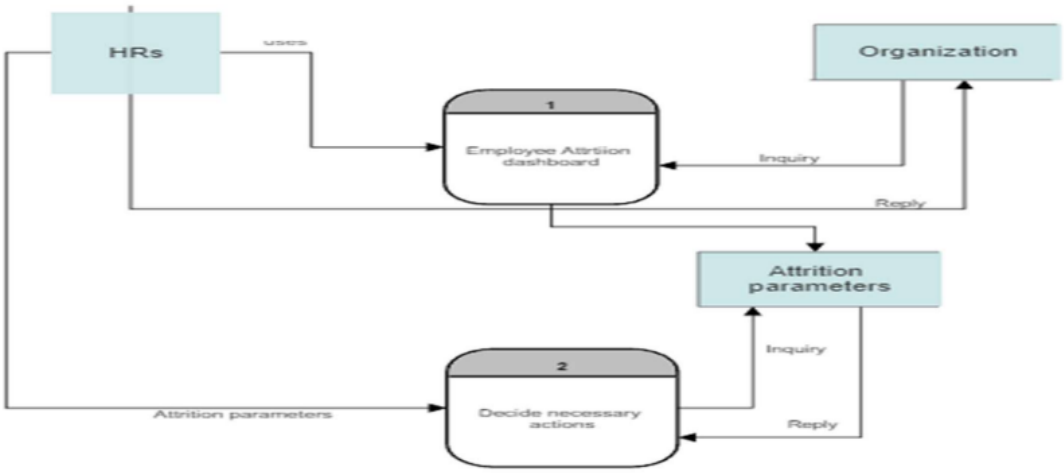
Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional	Requirement Description
NFR-1	Usability NFR-2	This software shall be easy to use for all users with minimal instructions. 100% of the languages on the graphical user interface (GUI) shall be intuitive and understandable by non-technical users.
NFR-2	Security	The user of the system should be provided the surety that their account details are secure. The system will provide security against cross site request forgery.
NFR-3	Reliability	The software shall be operable in all conditions. The system must be less prone to errors.
NFR-4	Performance	The performance of the system must assist the system's quality.
NFR-5	Extensibility	The software shall be extensible to support future developments and add-ons.
NFR-6	Portability	The software shall be 100% portable to all operating platforms. Therefore, this software should not depend on the different operating systems.
NFR-7	Scalability	The system must be able to handle an increase in workload without performance degradation

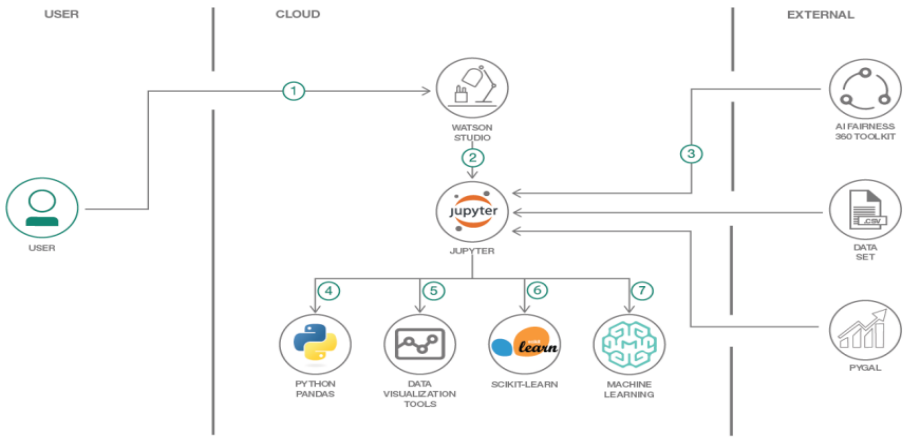
5.PROJECT DESIGN

5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 Technical Architecture



5.3 User Stories

Use the below template to list all the user stories for the product.

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 organisation by entering in form and getting authenticated via email.	I can access my account / dashboard	High	Sprint-1
		USN-2	As an employee, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As an employee, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-1
		USN-4	As an employee, I can register for the application through Gmail	I can get a verification link through email	Medium	Sprint-1
	Login	USN-5	As a employee, I can log into the application by entering email & password	I can enter the application	High	Sprint-2
	About	USN-6	I can click on the "About" to get the idea on employee attrition rate prediction using parameters like work environment, sentiments of employee, daily employee engagement and work progress.	I can get an idea about the project	Low	Sprint-2
	Launch	USN-7	As a HR, I can upload various analysed parameters from the computer.	I can choose any employee ('s all parameters) from my device	High	Sprint-2
	Predict	USN-8	As a HR, I can perform prediction using predict button	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	Medium	Sprint-3

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

TITLE	DESCRIPTION	DATE
Literature Survey & Information	Gathering Literature survey on the selected project & gathering information by referring technical papers,research publications etc.	28 SEPTEMBER 2022
Prepare Empathy Map	Prepare empathy map canva to capture the user pains & gains, prepare list of problem statements	24 SEPTEMBER 2022
Ideation	List the idea by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	25 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document which includes the novelty, feasibility of idea, business model, social impact, scalability of solution etc.	23 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	30 SEPTEMBER 2022
Solution Architecture	Prepare solution - architecture document.	28 SEPTEMBER 2022
Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	20 OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	8 OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	9 OCTOBER 2022
Technology Architecture	Prepare the technology architecture diagram.	10 OCTOBER 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	22 OCTOBER 2022

6.2 Sprint Delivery Schedule

Sprint	Duration	Sprint Start Date	Sprint End Date(Planned)	Release Date(Actual)
Sprint-1	6 Days	24 Oct 2022	29 Oct 2022	29 Oct 2022
Sprint-2	6 Days	31 Oct 2022	05 Nov 2022	05 Nov 2022
Sprint-3	6 Days	07 Nov 2022	12 Nov 2022	12 Nov 2022
Sprint-4	6 Days	14 Nov 2022	14 Nov 2022	19 Nov 2022

7.CODING & SOLUTIONING

To enhance with the coding we added many features and methodologies to bring out an errorless results.

The following features have been added in our project

Business Understanding:

Before solving the problem in the Business domain it needs to be understood properly. Business understanding forms a concrete base, which further leads to easy resolution of queries. We should have the clarity of what is the exact problem we are going to solve.

Analytic Understanding:

Based on the above business understanding one should decide the analytical approach to follow. The approaches can be of 4 types: Descriptive approach (current status and information provided), Diagnostic approach (A.K.A statistical analysis, what is happening and why it is happening), Predictive approach (it forecasts on the trends or future events probability) and Prescriptive approach (how the problem should be solved actually).

Data Requirements:

The above chosen analytical method indicates the necessary data content, formats and sources to be gathered. During the process of data requirements, one should find the answers for questions like ‘what’, ‘where’, ‘when’, ‘why’, ‘how’ & ‘who’.

Data Collection:

Data collected can be obtained in any random format. So, according to the approach chosen and the output to be obtained, the data collected should be validated. Thus, if required one can gather more data or discard the irrelevant data.

Data Understanding:

Data understanding answers the question “Is the data collected representative of the problem to be solved?”. Descriptive statistics calculates the measures applied over data to access the content and quality of matter. This step may lead to

reverting the back to the previous step for correction.

Data Preparation:

Let's understand this by connecting this concept with two analogies. One is to wash freshly picked vegetables and second is only taking the wanted items to eat in the plate during the buffet. Washing of vegetables indicates the removal of dirt i.e. unwanted materials from the data. Here noise removal is done. Taking only eatable items in the plate is, if we don't need specific data then we should not consider it for further process. This whole process includes transformation, normalization etc.

Modelling:

Modelling decides whether the data prepared for processing is appropriate or requires more finishing and seasoning. This phase focuses on the building of predictive/descriptive models.

Evaluation:

Model evaluation is done during model development. It checks for the quality of the model to be assessed and also if it meets the business requirements. It undergoes diagnostic measure phase (the model works as intended and where are modifications required) and statistical significance testing phase (ensures about proper data handling and interpretation).

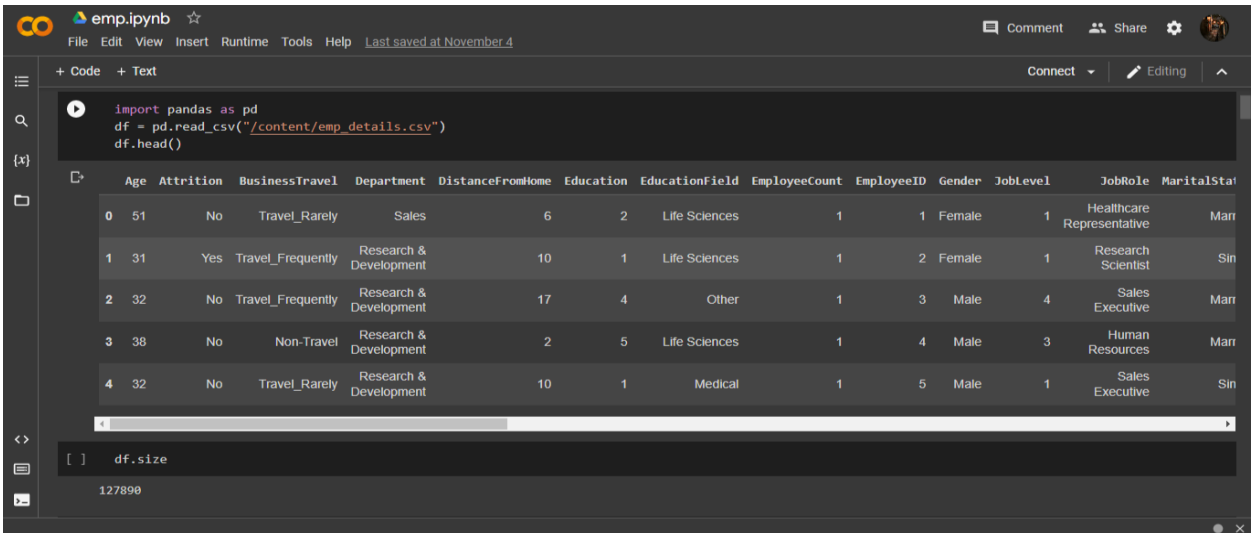
Deployment:

As the model is effectively evaluated it is made ready for deployment in the business market. Deployment phase checks how much the model can withstand in the external environment and perform superiorly as compared to others.

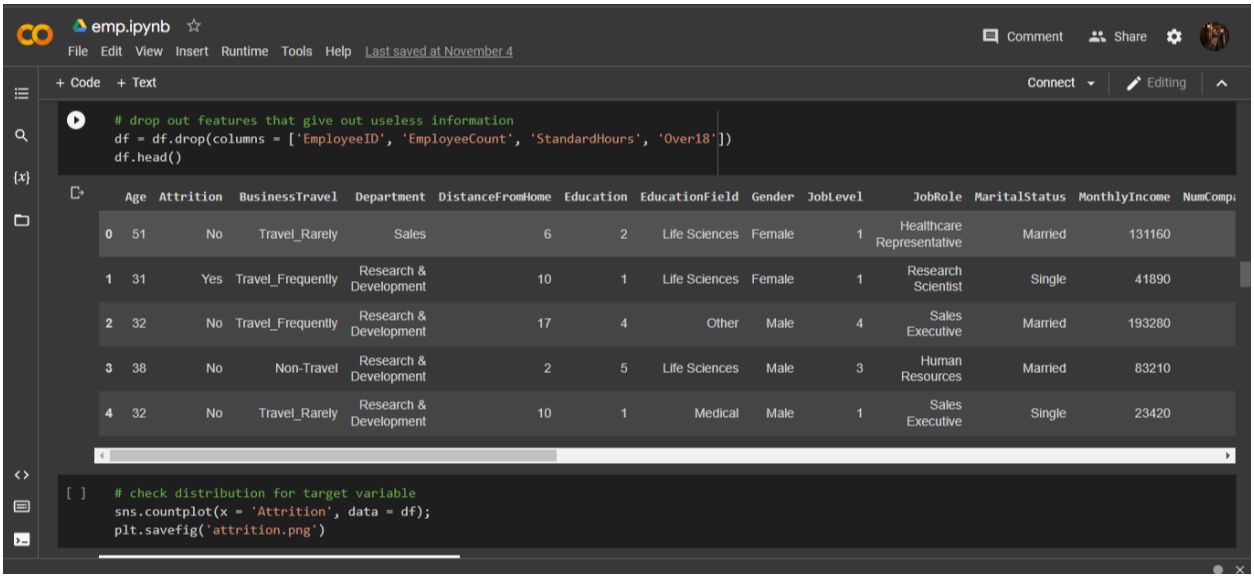
Feedback:

Feedback is the necessary purpose which helps in refining the model and accessing its performance and impact. Steps involved in feedback define the review process, track the record, measure effectiveness and review with refining.

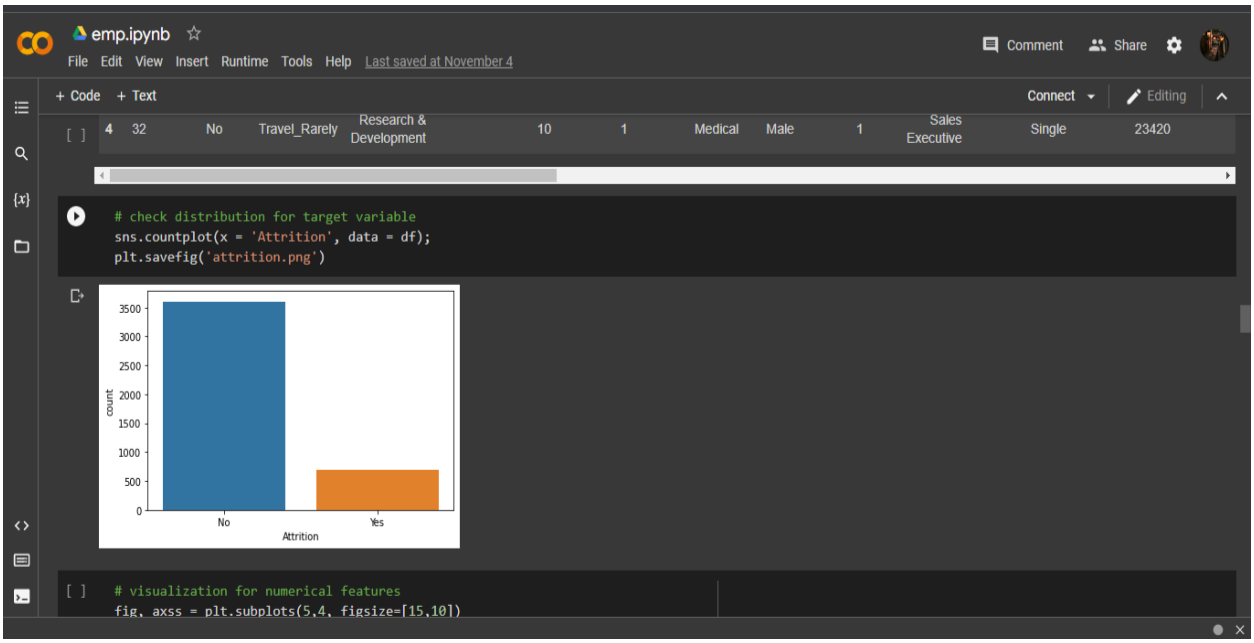
This figure shows the dataframe (which isn't cleaned and sanitized as it has lots of null values)



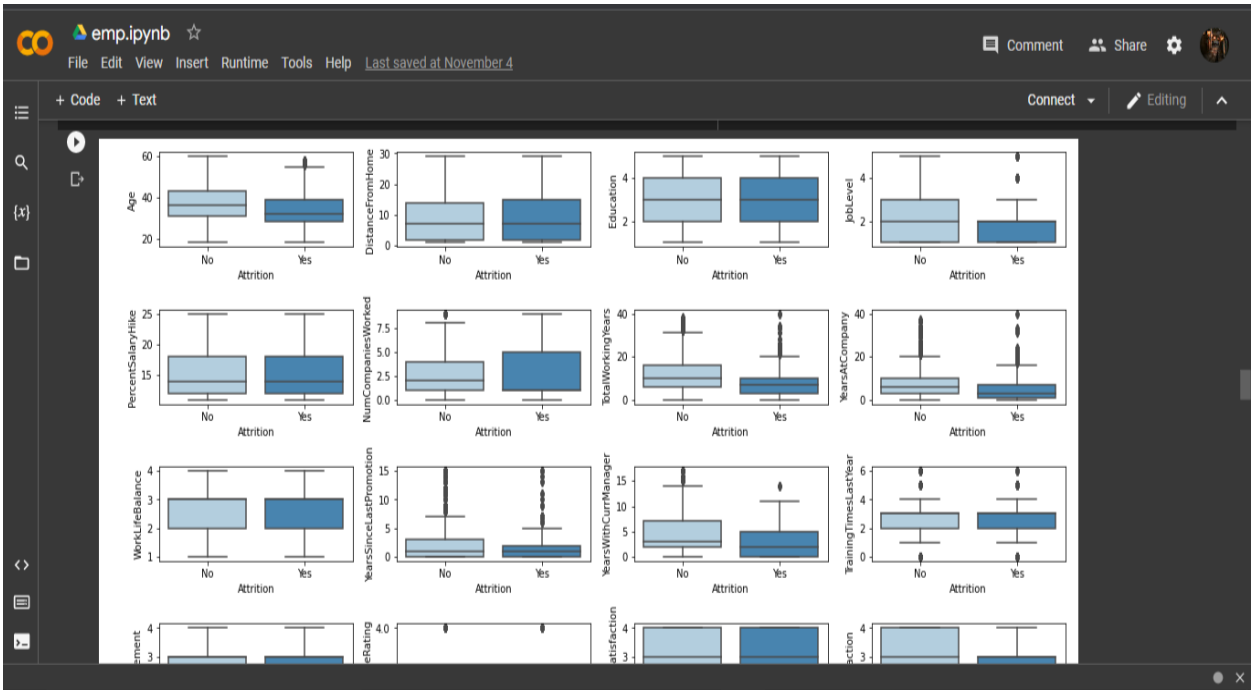
The below figure shows the dataset after removing unnecessary columns and the rows containing missing values and reordering the same.



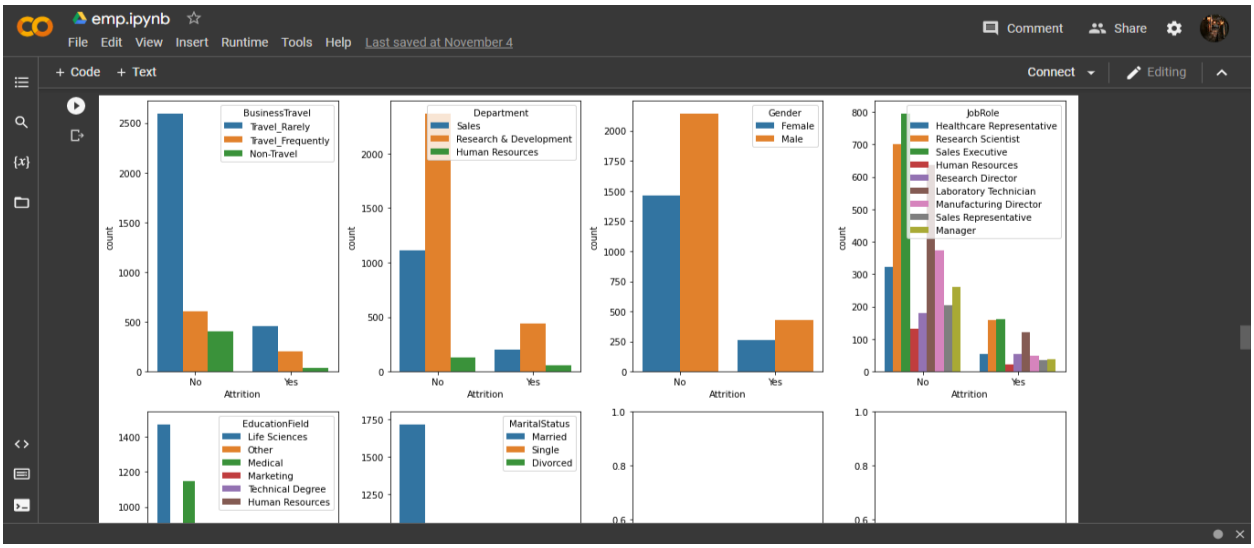
This figure shows the distribution of target variables



The visualisations of the numerical variables is



Similarly of the non numerical variables



And then, we train the model and test for unknown data to predict the attrition rate. And the screenshot is attached herewith.

```
# Logistic Regression
start_time = time.time()
train_pred_log, acc_log, acc_cv_log = fit_ml_algo(LogisticRegression(), X_train,y_train, 10)
log_time = (time.time() - start_time)
print("Accuracy: %s" % acc_log)
print("Accuracy CV 10-Fold: %s" % acc_cv_log)
print("Running Time: %s" % datetime.timedelta(seconds=log_time))
```

```
Accuracy: 89.8
Accuracy CV 10-Fold: 88.63
Running Time: 0:00:01.753627
```

```
# SVC
start_time = time.time()
train_pred_svc, acc_svc, acc_cv_svc = fit_ml_algo(SVC(),X_train,y_train,10)
svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_svc)
print("Running Time: %s" % datetime.timedelta(seconds=svc_time))
```

```
Accuracy: 88.53
Accuracy CV 10-Fold: 85.91
Running Time: 0:00:00.497278
```

```
# Linear SVC
start_time = time.time()
train_pred_svc, acc_linear_svc, acc_cv_linear_svc = fit_ml_algo(LinearSVC(),X_train, y_train,10)
linear_svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_linear_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_linear_svc)
print("Running Time: %s" % datetime.timedelta(seconds=linear_svc_time))
```

```
Accuracy: 89.89
Accuracy CV 10-Fold: 88.73
Running Time: 0:00:01.055932
```

8.RESULTS

The proposed work utilizes the deep learning technique along with some preprocessing steps to improve the prediction of employee attrition. Extensive experiments have been conducted to show the practical value of our work. The prediction accuracy using the original dataset is about 91%, whereas it is about 94% using a synthetic dataset.

9.ADVANTAGES AND DISADVANTAGES

9.1 ADVANTAGES:

- This paper proposed the deep learning technique along with some preprocessing
- steps to improve the prediction of employee attrition.
- To get realistic results, we derived a balanced version from the original.
- Several factors are analyzed to reveal employee attrition intercorrelation and to demonstrate the dominant ones.

9.2 DISADVANTAGES

- It requires very large amount data in order to perform better than other techniques.

10.FUTURE SCOPES

The Attrition Prediction model estimates the attrition risk for your employee populations in real-time, which is recalculated every time an employee submits feedback. The aggregated, segment-level view keeps the accuracy of your predictions high while protecting individual employee identity.

11.CONCLUSION

On the whole, this project was a useful experience.We have gained new knowledge and skills we achieved several of my learning goals.We got insight into professional practice. We learned the different facets of working .We experienced that self exploration, as in many organisations, is an important factor for the progress of projects.Related to our study we learned more about employee attrition rate prediction and the various approaches and algorithms to achieve the same. There is still a lot to discover and to improve. The methods used at the moment are still not standardized and a consistent method is in development. Furthermore we have experienced that it is of importance of each strategy and how other one is better than the current algorithm and in which application. we found that the internship is not one sided, but it is a way of sharing knowledge, ideas and opinions and implementing the same to get results.The internship was also good to find out what our strengths and weaknesses are.This helped me to define what skills and knowledge.We believe that our time spent in learning and surfing regarding various algorithms and the mathematics behind was well worth it and contributed to finding an acceptable solution to build a model and predict the employee's attrition rate. Two main things that we've learned the importance of time-management skills and self-motivation.At last this project has gave us new insights and motivation to pursue a career in machine learning domain.

12.Link to code and executable file

https://github.com/IBM-EPBL/IBM-Project-42611-1660669649.git