

Project Report

Corporate Employee Attrition Analysis

Submitted By

PNT2022TMID153278

Sruti K	-	SSNCE195002118
Maadhava Muralidharan B	-	SSNCE195002068
Vasuundhara K	-	SSNCE195002124
Sonam T	-	SSNCE195002115

Table of Content

1. INTRODUCTION

- a. Project Overview
- b. Purpose

2. LITERATURE SURVEY

- a. Existing problem
- b. References
- c. Problem Statement Definition

3. IDEATION & PROPOSED SOLUTION

- a. Empathy Map Canvas
- b. Ideation & Brainstorming
- c. Proposed Solution
- d. Problem Solution fit

4. REQUIREMENT ANALYSIS

- a. Functional requirement
- b. Non-Functional requirements

5. PROJECT DESIGN

- a. Data Flow Diagrams
- b. Solution & Technical Architecture
- c. User Stories

6. PROJECT PLANNING & SCHEDULING

- a. Sprint Planning & Estimation
- b. Sprint Delivery Schedule

- c. Reports from JIRA
- 7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**

- a. Feature 1
- b. Feature 2
- c. Database Schema (if Applicable)
- 8. **TESTING**

- a. Test Cases
- b. User Acceptance Testing
- 9. **RESULTS**

- a. Performance Metrics
- 10. **ADVANTAGES & DISADVANTAGES**
- 11. **CONCLUSION**
- 12. **FUTURE SCOPE**
- 13. **APPENDIX**

- Source Code
- GitHub & Project Demo Link

CHAPTER 1

INTRODUCTION

1.1 PROJECT OVERVIEW

The present study is bound to assess the causes of attrition and to find the expectation of employees to retain them in the organization. The primary data is collected through the structured questionnaire from 100 stakeholders, through stratified random sampling. It includes various levels of employees in the automobile industry. Chi-square, Anova, comparison and correlation are used for the analysis. The result shows that employees with experience are not promoted due to which they are dissatisfied. Employees receiving low income, no promotion or no career growth, developing stress may look for better opportunities. Employees except job security to retain themselves in the organization. Due to lack of growth opportunities they prefer to change job. So, the organization has to provide atleast required benefits and rewards to employees to make them retain and explore their career.

1.2 PURPOSE

Attrition is the shrinkage in employees number through retirement, voluntary/ involuntary resignation or death. It is also called as the total turnover. The employee turnover has become a biggest issue in many organizations.(jeen,2014). Attrition reflects the organization internal strength and weakness as well as the ability of the company. Attrition is considered as a twofold issue in many company, first that is related to cost and second retaining the experienced talent. Most of the organizations are battling with the increasing rate of attrition globally. (Kadam & Thakar, 2014). Application of Human resources has no alternative and they have continuous role to play in any organization. So, the major problem against human resources is attrition. As the rate of attrition increases day to day the organization production and productivity will be effected. (Chandrasekar, 2011). Employee attitude and candidate absconding are the major entanglement for every organization. Both these are bigger than attracting talent. In current knowledge impelled market place employees are the most vital human capital assets and attrition is the serious problem here. Attitude impacts organizations competing advantage. Employee attitude is a global issue. If the employees who are trained well and adapted to the organization leave the organization means it creates a vacuum which 116 MAN IN INDIA loses key skills,

knowledge and business relationship. Maximum effectiveness, growth and progress of the organization become lower. Every organization need to be clear in the concept that attracts and retains a potential employee and need to predict attrition early in the recruitment process to curtail significant loss of productivity among hiring managers, recruiters and consequent loss of money and revenue.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING PROBLEM

When considering attrition, many leaders tend to focus on the problem of high turnover – with good reason. Recruiting, hiring, onboarding, and training new employees costs businesses billions each year. Companies also suffer productivity losses – and lost profits – when there is a large amount of continuous churn in the workforce.

Top talent, in particular, can be very difficult and expensive to replace. The more talented the worker, the greater the consequences of attrition: Replacing an individual employee typically costs one-half to two times the worker's annual salary.

Financial considerations aside, businesses are better off when they can retain good employees and the organizational knowledge they possess.

But too little attrition can also be a problem. The right amount of attrition – with the right people turning over at the right time – is desirable.

The goal with employee attrition and retention is to strike the right balance of holding on to top talent while accepting that some level of attrition is healthy; employee attrition analytics enables organizations to find that balance.

2.2 REFERENCES

Rajesh Verma, Aanchal Aggarwal (2012) [1], articulated that the last decade saw an outbreak in the growth and development of the Indian Economy. It was accompanied by the revolution in all fronts and a radical change in the way life insurance business was done. . It is expected that Indian Insurance sector will be amongst the top 3 in the world by 2020. To reach up to this level the insurance companies have to take steps to reduce the increasing rate of attrition. As this study

says that the attrition rate increasing in the sales force in the Insurance sector, the author made focus on the reasons for attrition among the sales force and how to manage the attrition rate.

Asma farooque, Habibuddin (2015) [2], investigated several factors that are affecting high attrition rate in hotel industry. A study conducted by an industry chamber located in New Delhi reveals that the attrition rate in the hospitality industry in India is set to double to nearly 50 per cent by 2010, up from the earlier 25 per cent growing at an alarming rate of 10 per cent per annum. The result showed that the lack in salary, stress in job, compensation, better opportunities and stake issues were the major reasons behind the problem of attrition in the company.

Batty Dorance Jeen (2014) [3], mentioned that many organizations have the concern of employee turnover. It is highly destructive to both the organization as well as the employees. The research was conducted for the retail industry in Bangalore. Despite the incentives, motivational techniques and old practises of HRM there is still high attrition rate so study showcases ways to reduce the intentions of employee leaving the organization. This is performed by distributing questionnaires to retail outlets in Bangalore. The study concluded that turnover intention has influence on attrition factors such as QWL, career growth, working hours, personal/family reasons, and relation with internal co – worker, welfare, working condition, and salary.

Venkata Naga Manjula, Ruchita Ramani, Swati John (2013) [4], articulated that the objective of the study is to understand the growth of ITES sector in India. The study in the BPO Industry is to understand and gauge the attrition rate, its intensity and make a causal analysis, to design strategy to stabilise the sector by suggesting mitigating the attrition.

Shivani Mishra, Deepa Mishra (2013) [5], articulated that the study carried out is for shipping industry of Kutch, Gujarat. The purpose of the study is to analyse the turnover and the commitment to identify several domains of organization, human resource practices and other like employee characteristic and environmental factors, which may have a positive or negative impact on employees' intention to stay with an organization.

Vibha Gupta (2013) [6], stated that in recent years the turnover is high in BPO sectors either by absenteeism or employees absconding without any prior notice. Turnover rates for permanent Agents/Executives were 15.6% in 2009 and 35% in 2012. Department of Human Resources which also tracks attrition of temporary employees measured the turnover rate for temporary employees to be 77% in 2012. Therefore study is focused on recruitment and retention challenges that the IT/BPO industry currently faces and to examine ways to reduce high turnover rates among

first year Employees in the leading Domestic Call Center based in Indore.

Gayatri Negi (2013) [7], stated that the study explains that the attrition rate is inevitable but manageable cause its hard to provide permanent and promising workforce. This research shows the cause of attrition from various point of view and what are the effects for both the employer and employee. It also gives us a view on the positive side of the attrition and the role of leadership skills in controlling attrition.

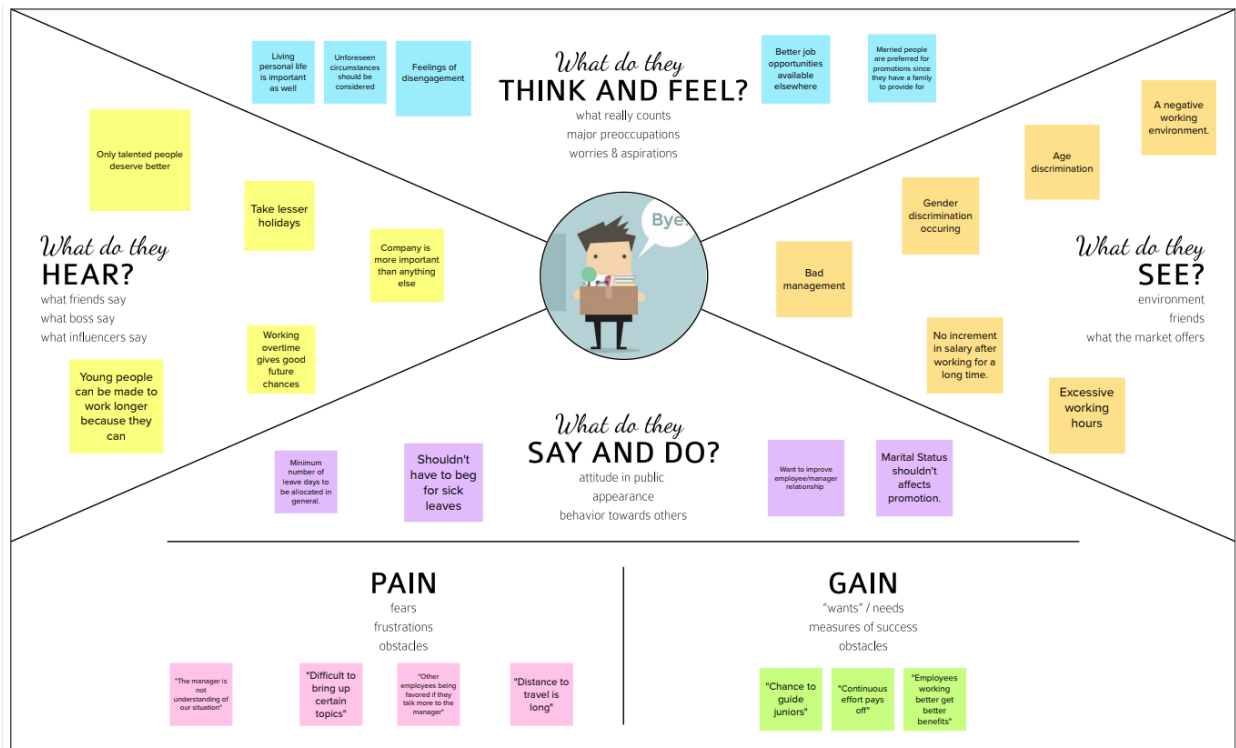
2.3 PROBLEM STATEMENT DEFINITION

Employee attrition is the gradual reduction in employee numbers. It is 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. Our aim is to determine which factors keep employees at the company and which prompt others to leave and predict the attrition rate of employees of an organization.

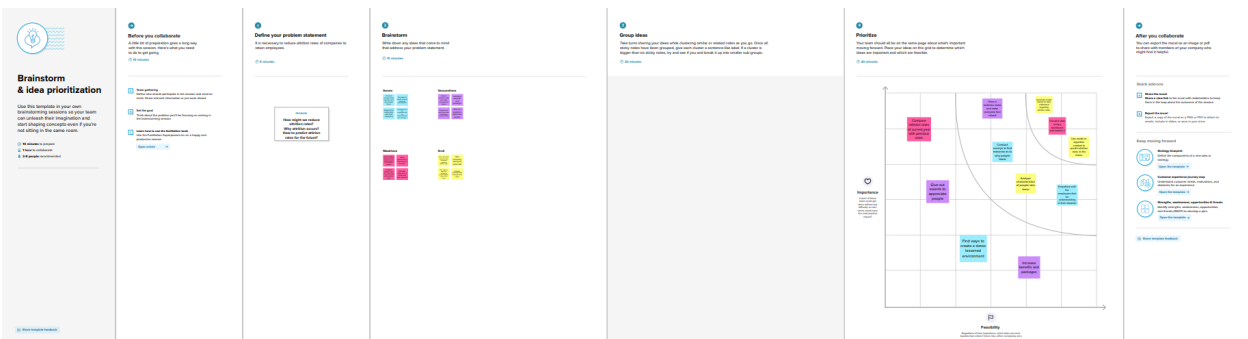
CHAPTER 3

IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING



3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	For any organization to be successful the important thing to do is to attract and retain people with good talent. But there are many factors due to which employees are prompted to leave. The objective is to identify and improve these factors so as to prevent loss of good people.
2.	Idea / Solution description	Collect data for attrition rate based on Business travel, Department, Job role, Marital status, Salary Hike etc. and analyse the obtained dataset to see what causes the attrition rates to increase and come up with ways to reduce it.
3.	Novelty / Uniqueness	So whenever the attrition rate comes to a peak during any time over the years the HR analyst who observes this notifies the company and solutions are proposed to bring it down.
4.	Social Impact / Customer Satisfaction	The company gets to retain good talent. It also helps to strengthen employee relationships and enhance team morale. Also the employees who remain start to feel that their job is fulfilling when they get better benefits. Everyone gains a sense of purpose.
5.	Business Model (Revenue Model)	The company gets to save their resources. They don't have to waste money, productivity and valuable time in finding new employees, hire them, on-board them, teach them, and keep them. Retaining experienced employees also helps them get quality work done.
6.	Scalability of the Solution	Efficiency of work done by the company is improved and the teams work better and quicker due to experience and better understanding of each other.

3.4 PROBLEM SOLUTION FIT

Define CS, fit into CC	1. CUSTOMER SEGMENT Who is your customer? The Company itself The HR	6. CUSTOMER CONSTRAINTS What constraints prevent your customers from taking action or limit their choices of solutions? I.e. spending power, budget, no cash, network connection, available devices. Analyzing attrition rates and trying to retain employees is assumed to be time consuming. The HR thinks it is better to keep recruiting newer employees to make up for the loss. No proper data is available to analyze form. Lack of interest to get things done	5. AVAILABLE SOLUTIONS Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have? i.e. pen and paper is an Recruitment of younger and newer employees. An advantage is that they can save money by giving smaller salaries to newbies.	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides. Need to analyze the rise of attrition rate of a company and visualize the findings to better understand the relationships formed and reasons as to why people leave. Visualizing findings as part of a dashboard to have different kinds of charts to observe from.	9. PROBLEM ROOT CAUSE What is the real reason that this problem exists? What is the back story behind the need to do this? Company not flexible to change. Too much work exerted and no proper growth is seen (that is there is no increase in salary after years of work or no promotion given). Other companies offer better plans and work life balance.	7. BEHAVIOUR What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and They try to get feedbacks from employees on regular intervals. Attain insights about their needs and expectations.	
Focus on J&P, tip into BE, understand RC	3. TRIGGERS What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news. High attrition rates that causes the company to lose valuable employees. Too much time and money being spent on recruiting new employees each time triggers them to rethink and come up with ways to retain the employees.	10. YOUR SOLUTION If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behavior. There is a high need to increase retention rates. So initially surveys are conducted each year where employees fill out the reasons they leave (Because of stress, growth plans, work culture, job satisfaction etc.) Next a dataset is built. This helps the HR to compare attrition rates each year. Visualization of data using dashboard helps us analysed data better. Next a model is generated based on the dataset available. This helps us to see changes in attrition rates in the current year and come up with ways to eliminate it	8. CHANNELS of BEHAVIOUR 8.1 ONLINE What kind of actions do customers take online? Extract online channels from #7 8.2 OFFLINE What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development. Online: A warm off boarding program is incorporated and proper description of jobs available is mentioned with growth plans. Offline: Meetups like dinner are done and trips are arranged to allow people to get to know each other and increase the team bond.	Focus on J&P, tip into BE, understand RC
	4. EMOTIONS: BEFORE / AFTER How do customers feel when they face a problem or a job and afterwards? i.e. lost, insecure -> confident, in control - use it in your communication strategy & design. Before: Company got frustrated having to keep recruiting again and again each time someone leaves. It is also stressful for the HR After: The company feels relieved now and is able to save resources.			

CHAPTER 4

REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registering into IBM via Gmail or LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Account creation for IBM Cognos.	Create an account for Cognos Analytics for Students through Gmail or LinkedIn.
FR-4	Analyse Dataset in IBM Cognos	Upload dataset, prepare it, explore it and create the final dashboard.
FR-5	Data Analyse using Python	Using algorithms and known ml models to analyse given dataset
FR-6	Final deliverable	Reach conclusions from dashboard and python data analysis code.

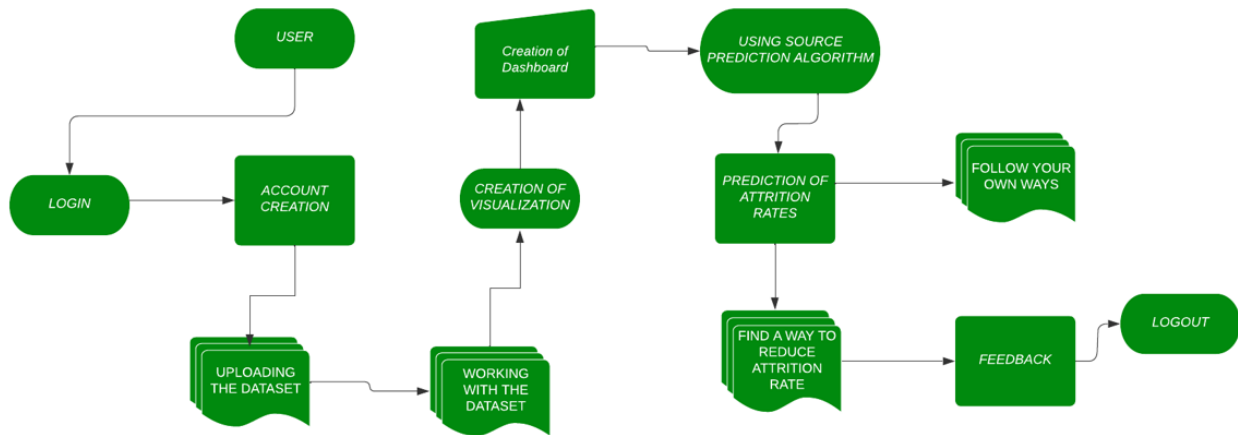
4.2 NON FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	HR should find it easy to see the dashboard and trends obtained from the data analysed using python.
NFR-2	Security	Permission to access should only be provided to admin and HR of company.
NFR-3	Reliability	Updates should be made to dataset according to change in employee's status like getting promoted ,relocating etc. Dataset should not be modifiable by anyone else without access control.
NFR-4	Performance	Dashboard should reflect changes of dataset.
NFR-5	Availability	Adding new functionality shouldn't affect old ones. Analysed data in the form of dashboard and trends should constantly be available.
NFR-6	Scalability	Dataset should accommodate large number of employees' data and model to predict trends should give realistic solutions.

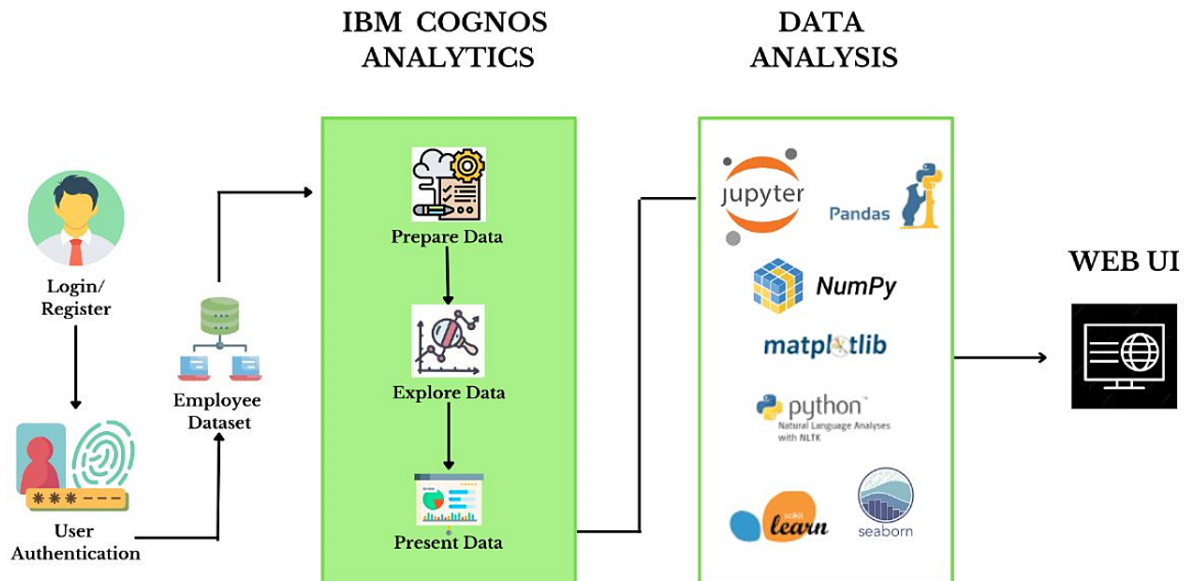
CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAM



5.2 SOLUTION & TECHNICAL ARCHITECTURE



5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance Criteria	Priority
Customer	Accessing the Application	USN-1	As a user, I should be able to access the application from anywhere and use on any devices	User can access the application using the browser on any device	High
	Uploading Dataset	USN-2	As a user, I should be able to upload the Dataset	User can upload the Dataset	High
	Viewing the Results	USN-3	As a user, I should be able to view the results	The result of the prediction is displayed	High
	Viewing Other Prediction	USN-4	As a user, I should be able to see other close predictions	The accuracy of other values must be displayed	High
	Usage Instruction	USN-5	As a user, I should have a usage instruction to know how to use the application	The usage instruction is displayed on the home page	Medium

CHAPTER 6

PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Collecting and preparing datasets	USN-1	As a user, I collect the required information about the corporate employee from the higher officials or from the office administration.	2	low	B. Maadhava Muralidharan, Sonam T
Sprint-1		USN-2	As a user, I can also get the employee details through the company database.		High	B. Maadhava Muralidharan, Sonam T
Sprint-1		USN-3	As a user, I segregate the data in a representable form which is used for the	1	high	B. Maadhava Muralidharan, Sonam T

			further steps.			
Sprint-2	Data visualization	USN-1	As a user, I analyze the data through visualization	2	medium	Vasuundhara , Sruti K
Sprint-2		USN-2	As a user, I analyze the data through dashboards		high	Vasuundhara , Sruti K
Sprint-2		USN-3	As a user, I analyze the data in the form of stories, graph, reports, etc.		low	Vasuundhara Sruti K
Sprint-3	Data analyzing	USN-1	As a user, I finally represent the results gained from the data analytics using python	2	high	B. Maadhava Muralidharan, Sonam T
Sprint-3		USN-2	Through python, I can calculate the attrition results		medium	B. Maadhava Muralidharan, Sonam T
Sprint-4	Reporting the results	USN-1	As a user, I can prepare reports from	1	medium	Vasuundhara, Sruti K

			the data analysis process			
Sprint-4		USN-2	From the reports, I can take necessary actions which results in employee attrition.		low	Vasuundhara, Sruti K

6.2 SPRINT DELIVERY SCHEDULE

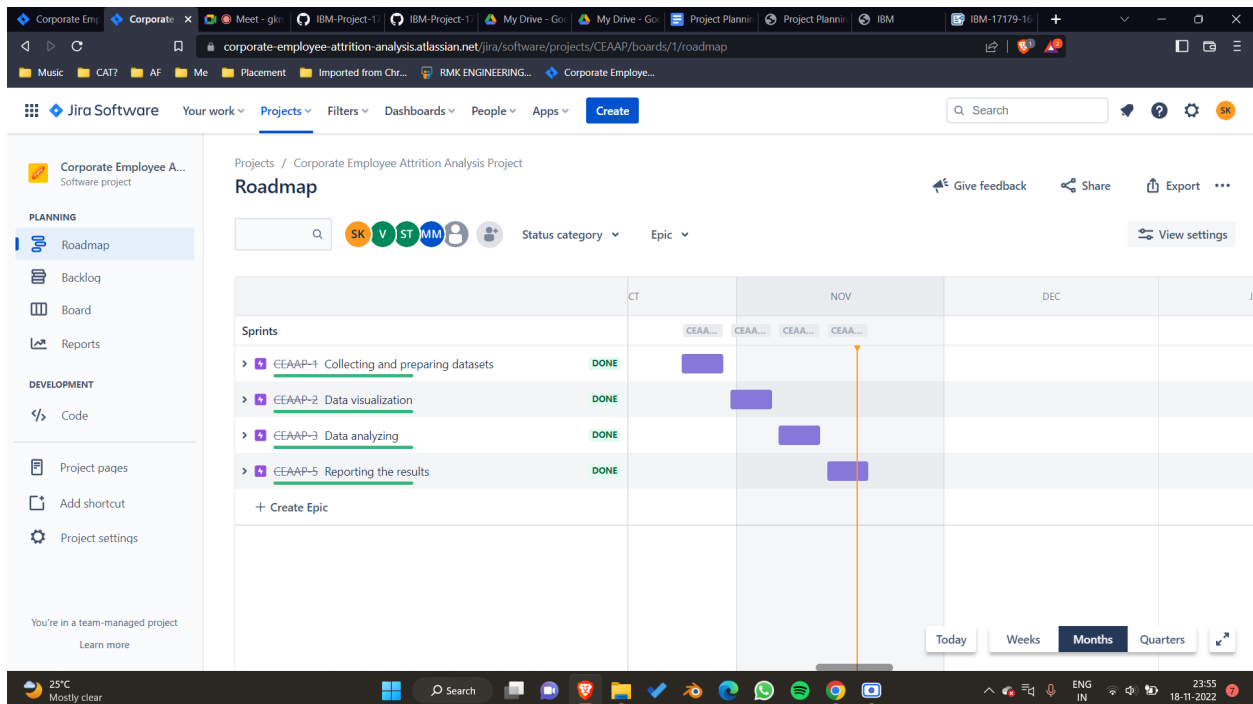
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	29 October 2022	05 November 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	05 November 2022	06 November 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	08 November 2022	09 November 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	11 November 2022	16 November 2022

Velocity:

Imagine we have a 6-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \text{sprint duration/velocity} = 20/6 = \sim 3$$

6.3 REPORTS FOR JIRA



CHAPTER 7

CODING & SOLUTIONIN

▼ Importing required packages:

```
import numpy as np
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
%matplotlib inline
warnings.filterwarnings('ignore')
```

▼ Mounting the dataset from drive:

```
import google.colab #to import the google colab
from google.colab import drive #import the drive from google colab
drive.mount('/content/gdrive') #mounting the drive with the colab
```

```
Mounted at /content/gdrive
```

```
employee =pd.read_csv('/content/gdrive/My Drive/Employee-Attrition.csv')
#employee =pd.read_csv( '/content/Employee-Attrition.csv' )
```

▼ Displaying the dataset:

```
employee.info()
pd.set_option('display.max_columns', 500)
```

```
employee
```

```
employee.head()
```

```
employee.tail()
```

Finding number of people who left the company and who stayed.

```
leave = employee['Attrition'][employee['Attrition']=='Yes']
```

```
leave = len(leave)
stays = employee['Attrition'][employee['Attrition']=='No']
stays = len(stays)
labels = ['People who Leave', 'People who Stay']
Attrition = [leave, stays]
explode = [0, 0.1]
palette_color = sns.color_palette('Accent')
#People who leave the company
```

```
fig, ax = plt.subplots(figsize=(13, 8))
sns.set_context('paper')
plt.style.use('seaborn')
```

```
plt.pie(Attrition, labels=labels, colors=palette_color,
        explode=explode, autopct='%0f%%')
```

```
plt.title('Attrition in Company ', fontsize=18)
```

```
plt.show()
```

```
#Replacing 'Attrition', 'OverTime', 'Over18' columns with integer before performing any vi
employee['Attrition'] = employee['Attrition'].apply(lambda x:1 if x == 'Yes' else 0)
employee['OverTime'] = employee['OverTime'].apply(lambda x:1 if x == 'Y' else 0)
employee['Over18'] = employee['Over18'].apply(lambda x:1 if x == 'Yes' else 0)
```

```
employee.head()
```

```
employee.isnull().sum()
```

▼ Data Exploration and insight gathering from the dataset:

1. Histogram distribution of all parameters of an employee:

```
employee.hist(bins = 30, figsize = (20,20), color = 'b')

# Dropping 'EmployeeCount' , 'Standardhours' and 'Over18' as they do not change from one em
# Dropping 'EmployeeNumber' since it is not needed.
employee.drop(['EmployeeCount', 'StandardHours', 'Over18', 'EmployeeNumber'], axis = 1, inplace=True)

employee.head()
```

2. Finding Correlations between all row values when each other using heatmap:

```
correlations = employee.corr()
fig, ax = plt.subplots(figsize = (20,20))
sns.heatmap(correlations, annot = True)
```

Conclusion:

Age is strongly correlated with total working years :

Value obtained: 0.68

The people who worked for longer don't leave easily.

Monthly income is strongly correlated with Job level:

Value obtained: 0.95

Better designations promote people to stay.

Performance rate is correlated with Percent Salary Hike:

Value obtained: 0.77

People like to work better when they are given better incentives

3. Analysing age of all employees of the company using bargraph:

```
plt.figure(figsize = [25,12])
sns.countplot(x = 'Age', hue = 'Attrition', data = employee)
```

Conclusion:

▼ We observe that people between age range 26-35 tend to leave the company

4. How Unproductive Workplace affects attribution:

```
data=pd.read_csv('/content/gdrive/My Drive/Employee-Attrition.csv')
#data =pd.read_csv( '/content/Employee-Attrition.csv' )

fig, axs = plt.subplots(ncols=2, nrows=2,figsize=(13, 9))
plt.suptitle('Are Attrition Caused by Unproductive Workplace ?', fontsize= 18)
sns.set_style('darkgrid')

ax1 = sns.histplot(x='EnvironmentSatisfaction', hue = 'Attrition', data=data, ax=axs[0, 0])
ax2 = sns.histplot(x='RelationshipSatisfaction', hue = 'Attrition', data=data, ax=axs[0, 1])
ax3 = sns.histplot(x='WorkLifeBalance', hue = 'Attrition', data=data, ax=axs[1, 0], palette=
ax4 = sns.histplot(x='JobSatisfaction', hue = 'Attrition', data=data, ax=axs[1, 1], palette=

for ax in axs.flatten():
    plt.sca(ax)
    plt.xticks(rotation = 30)

plt.show()
```

Conclusion:

Environment Satisfaction, Relationship Satisfaction, Job Satisfaction: Low attrition between employee with lowest and highest environment satisfaction

Work Life Balance : High attrition for employee with high work life balance

5. Identifying how benefits affecting employee retention trend in company

```
fig, axs = plt.subplots(ncols=2, nrows=2, figsize=(13, 8))
plt.suptitle('Are Attrition Caused by the Lack of Benefits ?', fontsize= 18)
sns.set_style('darkgrid')

ax1 = sns.histplot(x='HourlyRate', hue = 'Attrition', data=data, ax=axs[0, 0], palette='Ac
ax2 = sns.histplot(x='StockOptionLevel', hue = 'Attrition', data=data, ax=axs[0, 1], palet
ax3 = sns.histplot(x='YearsSinceLastPromotion', hue = 'Attrition', data=data, ax=axs[1, 0]

for ax in axs.flatten():
    plt.sca(ax)
    plt.xticks(rotation = 30)

plt.show()
```

Conclusion:

Hourly Rate : Attrition not changing much due to worker's hourly rate

Stock Option Leve & Year Since Last Promotionl : Causes high attrition

6. How SocioEconomics Characteristics of employee affect attrition:

```
plt.figure(figsize = [20,20])

plt.subplot(611)
sns.countplot(x = 'JobRole', hue = 'Attrition', data = employee)

plt.subplot(612)
sns.countplot(x = 'MaritalStatus', hue = 'Attrition', data = employee)

plt.subplot(613)
sns.countplot(x = 'JobInvolvement', hue = 'Attrition', data = employee)

plt.subplot(614)
```

```
sns.countplot(x = 'JobLevel', hue = 'Attrition', data = employee)
plt.subplot(615)

sns.countplot(x='Gender', hue = 'Attrition',data=employee)
plt.subplot(616)

sns.countplot(x='Education', hue = 'Attrition',data=employee)
#Note Attrition : 1-Leave, 0-Stay
```

Conclusion:

Single employees tend to leave compared to married and divorced

Sales Representatives tend to leave compared to any other job

Less involved employees tend to leave the company

Less experienced (low job level) tend to leave the company

Attrition was relatively high on worker with education level 3

7.Visualizing the Probability Density of a continuous variable using KDE (Kernel Density Estimate)

KDE describes the probability density at different values in a continuous variable

```
left = employee[employee['Attrition'] == 1 ]
stay = employee[employee['Attrition'] == 0 ]
```

7.1 Analysing how distance from home causes employees to leave or stay:

```
plt.figure(figsize = [15,10])
sns.kdeplot(left['DistanceFromHome'], label = 'Employee who left', shade = True, color = 'red')
sns.kdeplot(stay['DistanceFromHome'], label = 'Employee who stay', shade = True, color = 'blue')

plt.xlabel('Distance from Home')
```

7.2 Analysing how total working years causes employees to leave or stay:

```
plt.figure(figsize = [15,10])
sns.kdeplot(left['TotalWorkingYears'], label = 'Employee who left', shade = True, color = 'red')
sns.kdeplot(stay['TotalWorkingYears'], label = 'Employee who stay', shade = True, color = 'blue')

plt.xlabel('Total Working Years')
```

Conclusion:People stayed

7.3 Analysing how years with current manager causes employees to leave or stay:

```
plt.figure(figsize = [15,10])
sns.kdeplot(left['YearsWithCurrManager'], label = 'Employee who left', shade = True, color
sns.kdeplot(stay['YearsWithCurrManager'], label = 'Employee who stay', shade = True, color

plt.xlabel('Years With Current Manager')
```

Conclusion:Attrition happens only when employee had worked with a manager for 2-3 years

8.Job Roles vs Monthly income:

```
plt.figure(figsize = [10,15])
sns.barplot(x = employee['MonthlyIncome'] , y = employee['JobRole'])
```

▼ Training and Testing Dataset

```
employee.head(2)
```

```
X_cat = employee[['BusinessTravel','Department','EducationField','Gender','JobRole','Marit
X_cat
```

```
from sklearn.preprocessing import OneHotEncoder
onehotencoder = OneHotEncoder()
X_cat = onehotencoder.fit_transform(X_cat).toarray()
X_cat
```

```
X_cat.shape
```

```
X_cat = pd.DataFrame(X_cat)
```

```
# note that we dropped the target 'Attrition'
X_numerical = employee[['Age','DailyRate','DistanceFromHome','Education','EnvironmentSatis
```

```
X_all = pd.concat([X_cat, X_numerical], axis = 1)
X_all.head()
```

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
X = scaler.fit_transform(X_all)
X
```

```
y = employee['Attrition']  
y
```

▼ Train and Evaluate Logistic Regression Classifier

```
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)
```

```
X_train.shape
```

```
X_test.shape
```

```
from sklearn.linear_model import LogisticRegression  
from sklearn.metrics import accuracy_score
```

```
model = LogisticRegression()  
model.fit(X_train, y_train)
```

```
LogisticRegression()
```

```
y_pred = model.predict(X_test)
```

```
# Testing Set Performance  
y_pred
```

```
from sklearn.metrics import confusion_matrix, classification_report  
print('Accuracy {} %'.format( 100 * accuracy_score(y_pred, y_test)))
```

```
Accuracy 83.69565217391305 %
```

```
cm = confusion_matrix(y_pred, y_test)  
sns.heatmap(cm, annot = True)
```

```
print(classification_report(y_test, y_pred))
```

Conclusion: Data analysis using Logistic Regression gives good precision for given dataset. So insights obtained are quite precise.

▼ Train and Evaluate A Random Forest Classifier

```
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier()
model.fit(X_train, y_train)

RandomForestClassifier()

y_pred = model.predict(X_test)
y_pred

# Testing Set Performance
cm = confusion_matrix(y_pred, y_test)
sns.heatmap(cm, annot = True)

print(classification_report(y_test, y_pred))
```

Conclusion: Data analysis using Random Forest Classifier also gives good precision. Hence reconfirmed that insights obtained are quite precise.

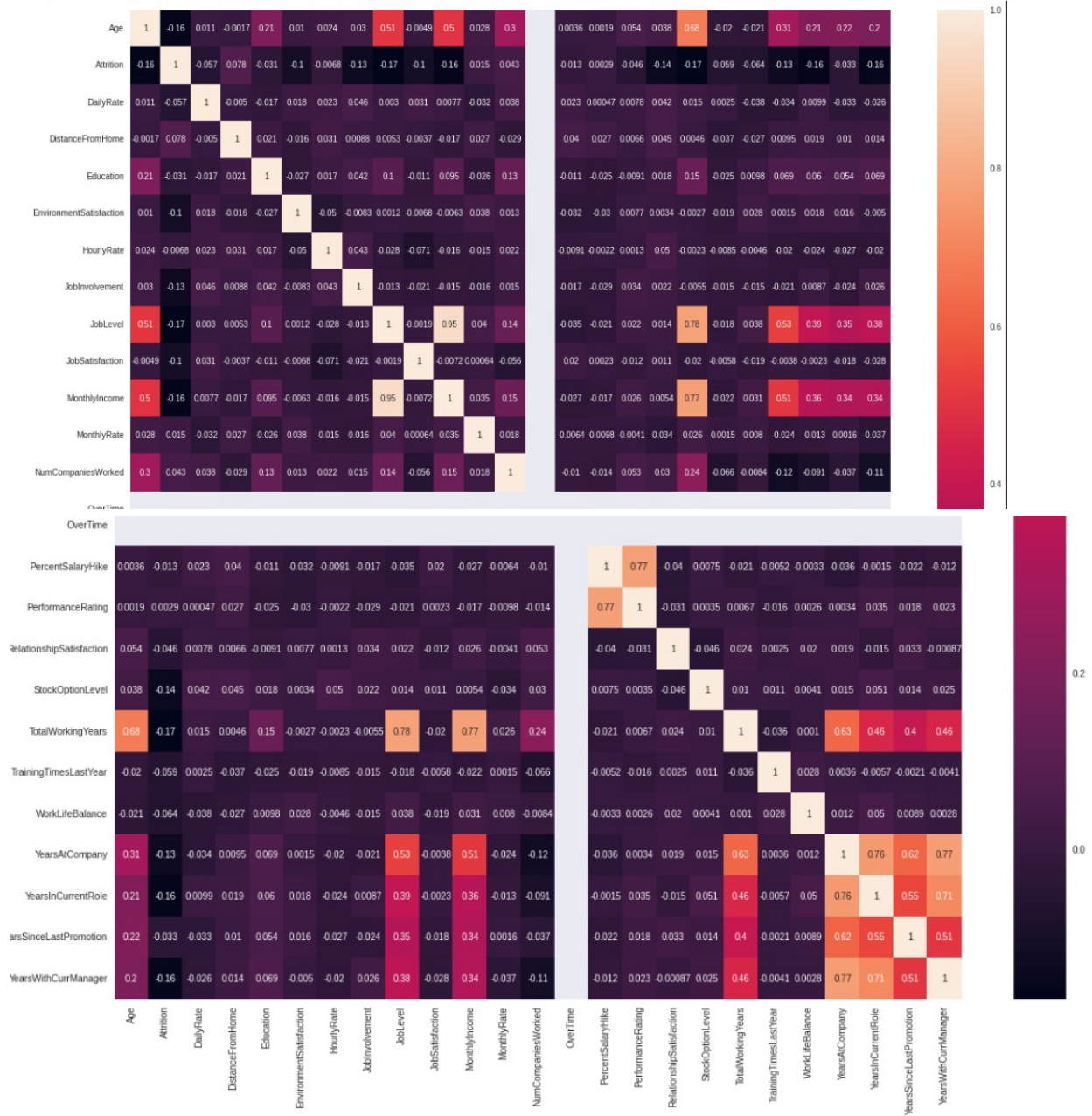
CHAPTER 9

RESULTS

9.1 PERFORMANCE METRICS

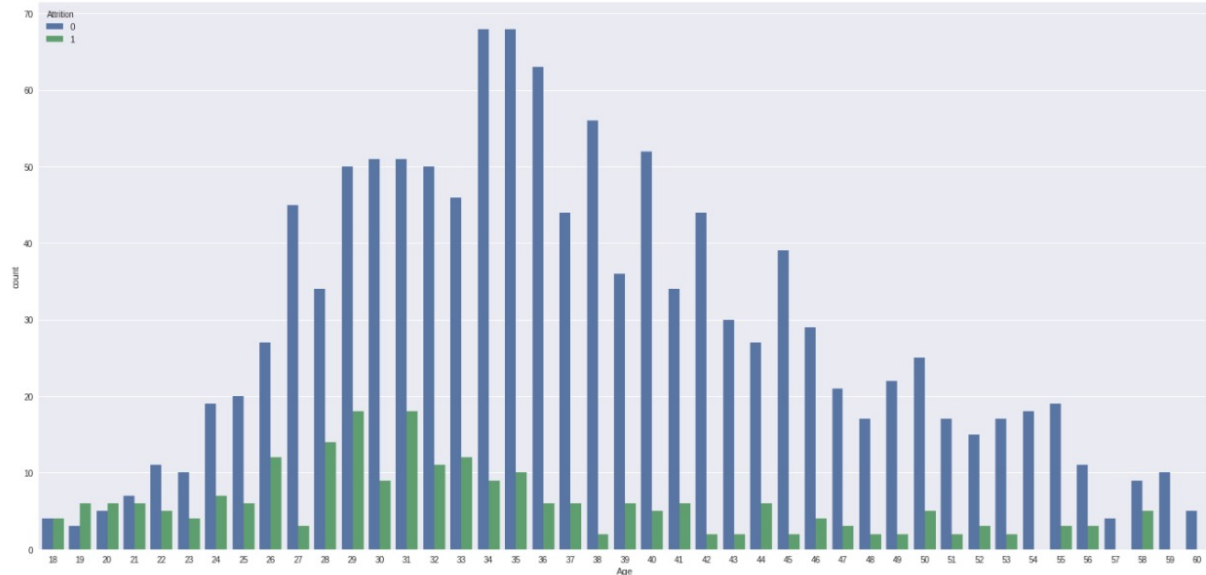
9.1.1 FINDING CORELATION USING HEAT MAP

<matplotlib.axes._subplots.AxesSubplot at 0x7ffb57332250>



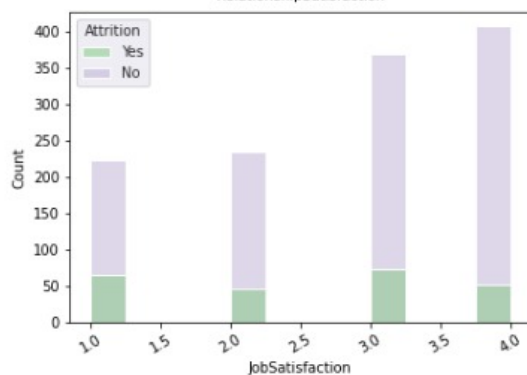
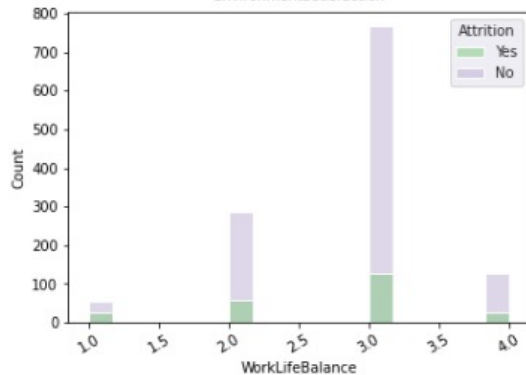
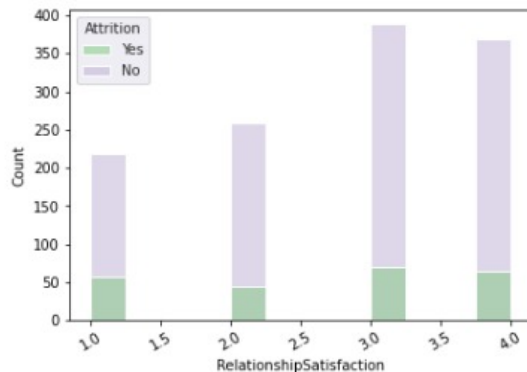
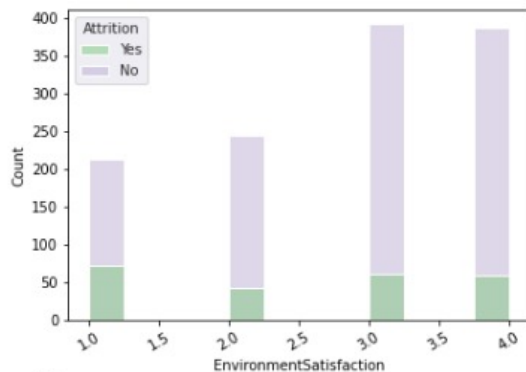
9.1.2 ANALYSING AGE OF EMPLOYEES

<matplotlib.axes._subplots.AxesSubplot at 0x7ffb51ca4390>



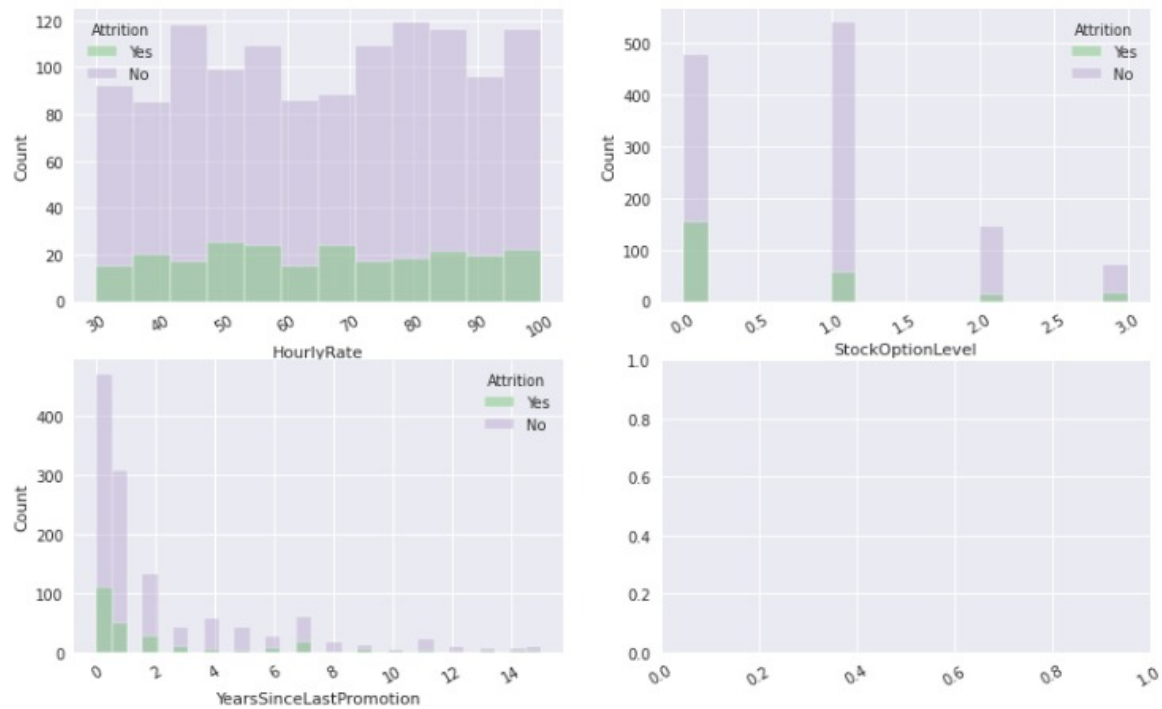
9.1.3 Attrition Caused By Unproductive Workplace

Are Attrition Caused by Unproductive Workplace ?



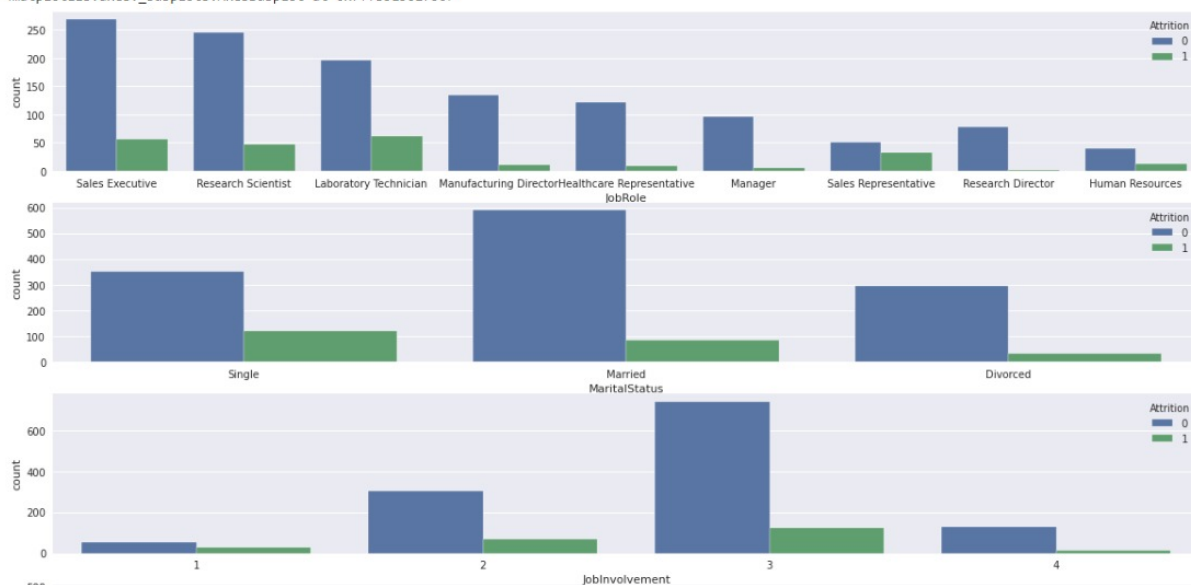
9.1.4 Attrition Caused By Lack of Benfits

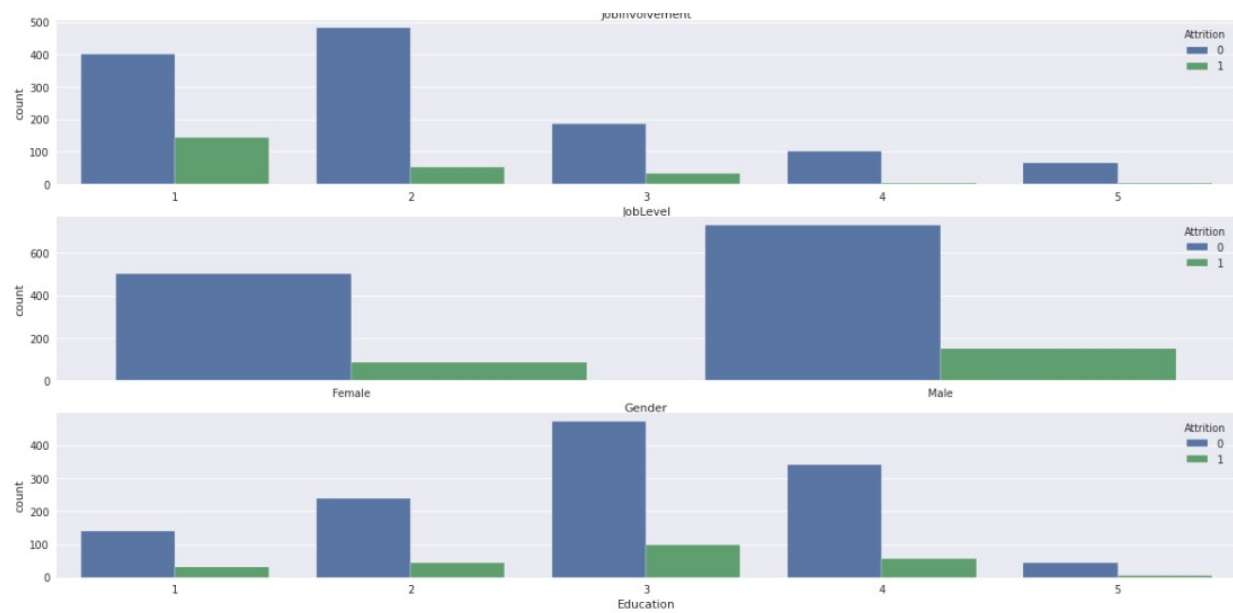
Are Attrition Caused by the Lack of Benefits ?



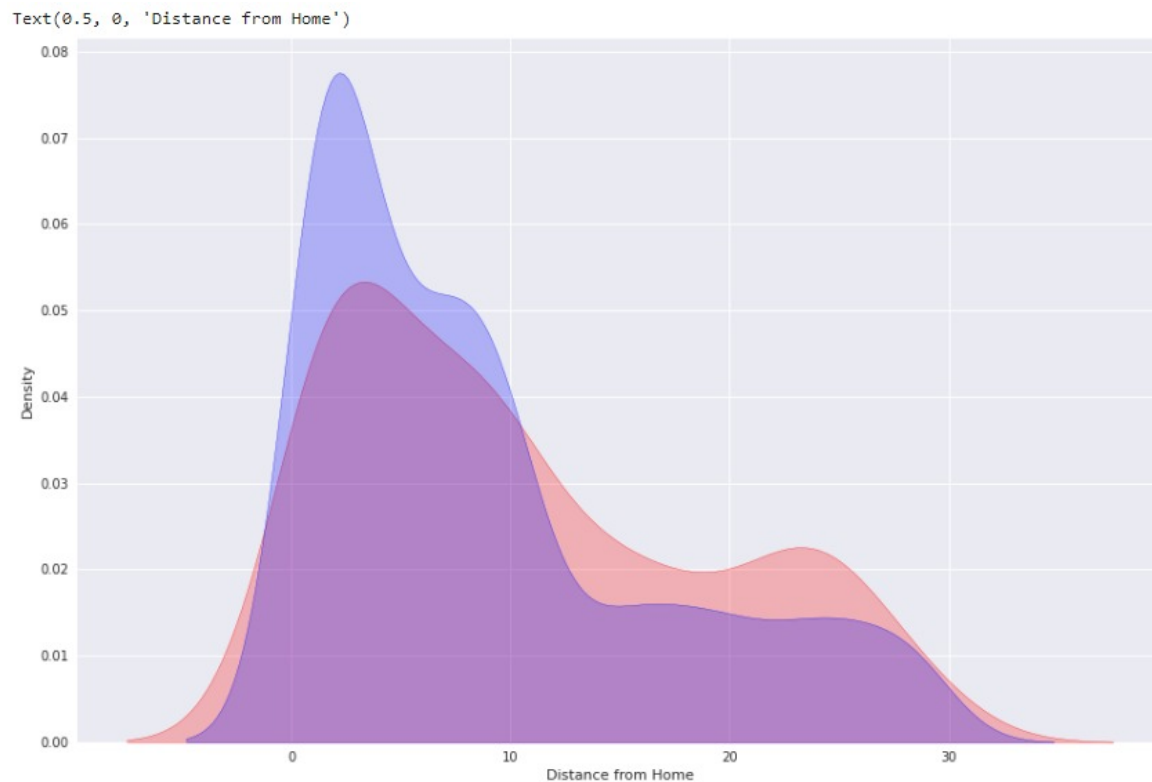
9.1.5 How SocioEconomics Characteristics of employee affect attrition:

<matplotlib.axes._subplots.AxesSubplot at 0x7ffb51362750>



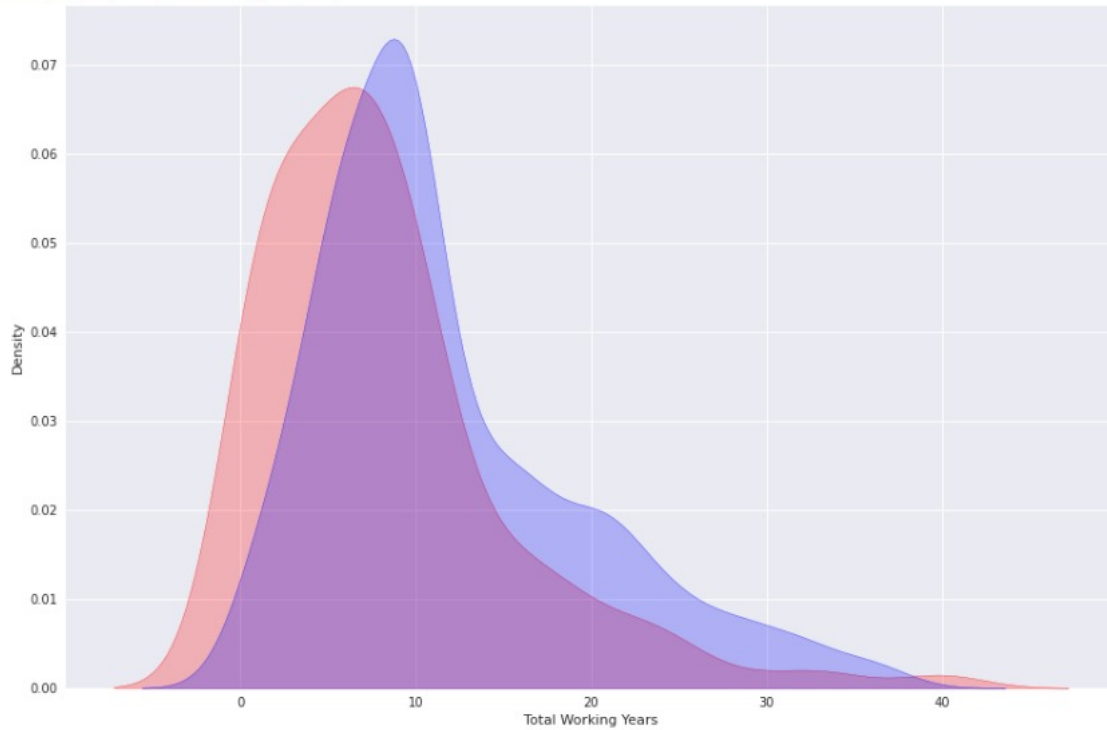


9.1.6 Kdeplot used to analyse how distance from home causes affects attrition



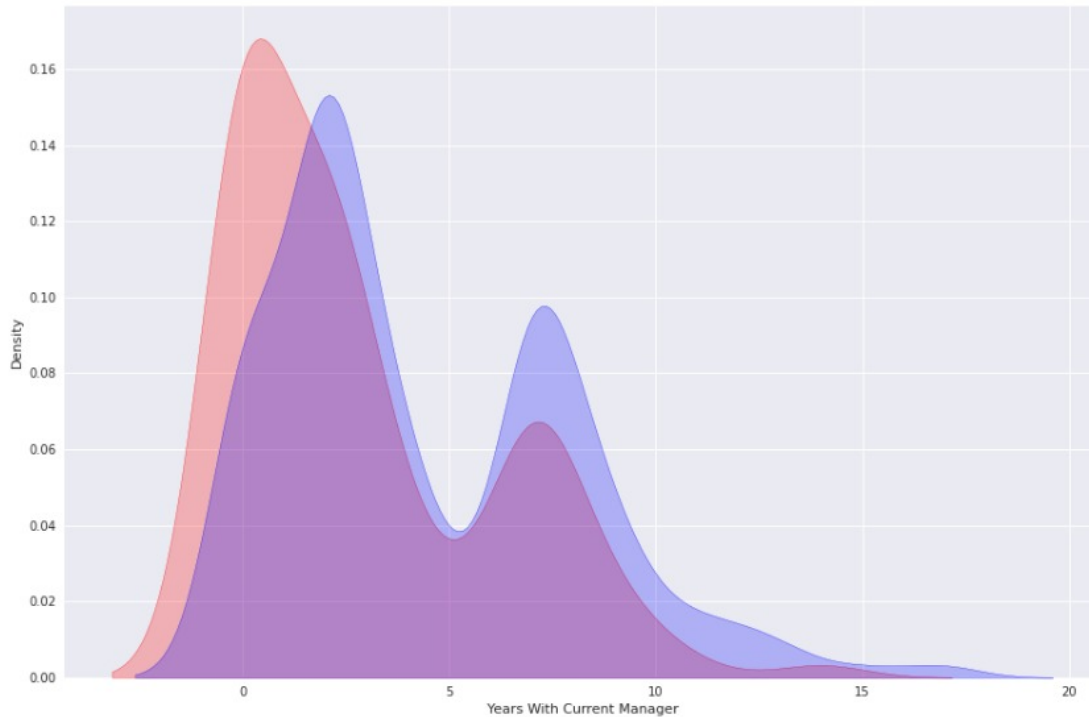
9.1.7 Kdeplot used to analyse how Total working hours affects attrition

```
Text(0.5, 0, 'Total Working Years')
```



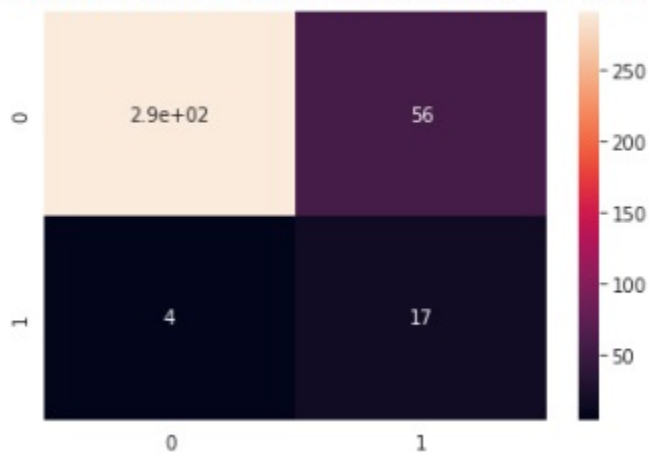
9.1.8 Kdeplot - How Years with current Manager affects attrition

```
Text(0.5, 0, 'Years With Current Manager')
```



9.1.9 Heat map for logistic Regression

<matplotlib.axes._subplots.AxesSubplot at 0x7fdb1cfa3290>

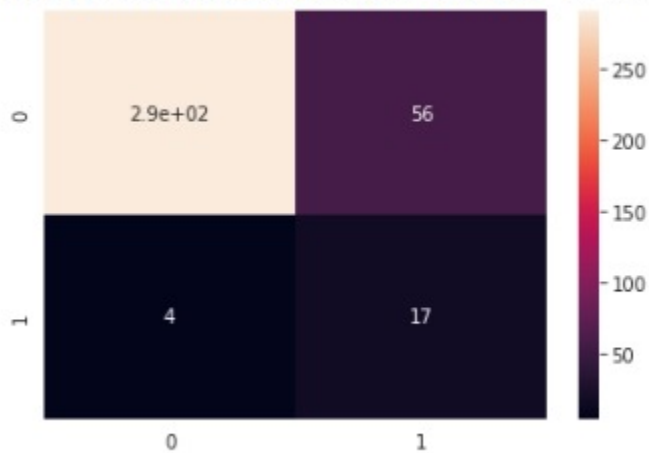


9.1.10 Classification report -logistic regression

	precision	recall	f1-score	support
0	0.84	0.99	0.91	295
1	0.81	0.23	0.36	73
accuracy			0.84	368
macro avg	0.82	0.61	0.63	368
weighted avg	0.83	0.84	0.80	368

9.1.11 Heat map-Random forest Classifier

<matplotlib.axes._subplots.AxesSubplot at 0x7fdb1cf3aa50>



9.1.12 Classification Report-Random Forest Classifier

	precision	recall	f1-score	support
0	0.83	1.00	0.91	300
1	0.89	0.12	0.21	68
accuracy			0.83	368
macro avg	0.86	0.56	0.56	368
weighted avg	0.84	0.83	0.78	368

CHAPTER 10

ADVANTAGES & DISADVANTAGES

ADVANTAGES

- Reduces manual work
- More accurate than average human
- Capable of handling a lot of data
- Can be used anywhere from any device

DISADVANTAGES

- Cannot handle complex data
- All the data must be in digital format
- Requires a high performance server for faster predictions
- Prone to occasional errors

CHAPTER 11

CONCLUSION

The major contributing factors and insights we have derived through data driven analysis of attrition rates are:

- Stock Option Level & Year Since Last Promotion
- Work Life Balance
- Single employees tend to leave compared to married and divorced
- Sales Representatives tend to leave compared to any other job
- Less involved employees tend to leave the company
- Less experienced (low job level) tend to leave the company
- Attrition was relatively high on worker with education level 3
- Distance from home also affects attrition

CHAPTER 12

FUTURE SCOPE

This project is far from complete and there is a lot of room for improvement. Some of the improvements that can be made to this project are as follows:

- Add support to enable easy gathering and creation of data set by incorporating surveys and other tools.
- Add support for Preprocessing of Dataset.
- Improve model that has Increased Accuracy.

This project has endless potential and can always be enhanced to become better. Implementing this concept in the real world will benefit several industries and reduce the workload on many workers, enhancing overall work efficiency.

CHAPTER 13

APPENDIX

Project Demo Link: [Youtube link for project demo](https://youtu.be/134MJrxMJOY)
<https://youtu.be/134MJrxMJOY>

Git Link: <https://github.com/IBM-EPBL/IBM-Project-17179-1659629778>

Source Link: https://github.com/IBM-EPBL/IBM-Project-17179-1659629778/blob/main/Final%20Deliverables/Corporate_Employee_Attrition_Rate.ipynb