# Corporate Employee Attrition Analysis

**APROJECT REPORT**

Submitted By

# TeamID:PNT2022TMID38728

**Team Leader: POOVARASAN S (420619106009) Team Member: PAVITHRAN V (420619106008)**

# Team Member: SURIYA C (420619106014)

# Team Member: KISHORE M (420619106005)

# Team Member: SARASWATHI M (420619106011)

In partial fulfillment for the award of the degree of

# BACHELOR OF ENGINEERING

IN

# ELECTRONICS AND COMMUNICATIONS ENGINEERING ANNAI TERESA COLLEGE OF ENGINEERING, THIRUNAVALUR

**ANNAUNIVERSITY::CHENNAI600025**

**TABLEOFCONTENTS**

# CHAPTER NO TITLE PAGENO

1. [INTRODUCTION 4](#_TOC_250023)
   1. [ProjectOverview **4**](#_TOC_250022)
   2. [Purpose **4**](#_TOC_250021)
2. LITERATURESURVEY 5
   1. [Existingproblem **5**](#_TOC_250020)
   2. [References **5**](#_TOC_250019)
   3. [ProblemStatementDefinition **5**](#_TOC_250018)
3. IDEATION&PROPOSEDSOLUTION
   1. [Empathy MapCanvas **6**](#_TOC_250017)
   2. [Ideation &Brainstorming **6**](#_TOC_250016)
   3. [Proposed Solution **8**](#_TOC_250015)
   4. [ProblemSolutionfit **9**](#_TOC_250014)
4. [REQUIREMENTANALYSIS](#_TOC_250013)
   1. [Functional requirement **10**](#_TOC_250012)
   2. [Non-Functional requirements **10**](#_TOC_250011)
5. [PROJECTDESIGN](#_TOC_250010)
   1. [DataFlowDiagrams **11**](#_TOC_250009)
   2. [Solution&Technical Architecture **11**](#_TOC_250008)
   3. [UserStories **12**](#_TOC_250007)
6. PROJECTPLANNING&SCHEDULING
   1. [SprintPlanning&Estimation **13**](#_TOC_250006)
   2. [SprintDeliverySchedule **13**](#_TOC_250005)
   3. ReportsfromJIRA
7. [CODING&SOLUTIONING](#_TOC_250004)
   1. Feature1 **13**
   2. Feature2 **14**
   3. DatabaseSchema
8. [TESTING](#_TOC_250003)
   1. [TestCases **15**](#_TOC_250002)
   2. [UserAcceptanceTesting **15**](#_TOC_250001)
9. [RESULTS 17](#_TOC_250000)
10. ADVANTAGES&DISADVANTAGES 18
11. CONCLUSION 18
12. FUTURESCOPE 19
13. APPENDIX 20

# 1.INTRODUCTION

# Projectoverview

Employee attrition has become a vital problem across the world.Itisoneofthecrucialissuesfacedbybusinessleaderswithincompanieswhere they lose the most talented employees. A good employee isalways an asset to the organization and their resignation can lead tovarious problems like financial losses, overall performance, and loss ofacquiredknowledge.Furthermore,hiringnewemployeesisfarexorbitant, taxing, and time-consuming in comparison to recruiting theexisting one. It is very time-consuming to recruit a new employee as ittakeshimmonthsfortraining,adjustingtotheculture,rules,andenvironment.Therefore,upcomingtrendsandtechnologyusingMachine Learning Algorithms must be exploited for the benefitofbusinessorganizations.Knowingthereasonbeforehandfortheemployee attrition,companies can mitigate this loss. Thisanalysisprovides a conclusive review of employee attrition from the data setIBMHRAnalyticsEmployeeAttritionPerformance.

# Purpose

* + 1. Hardik P. K. ( 2016) , researched on “a study on employeeattrition:withspecialreferencetoKeralaITIndustry”.Hisresearchexamined the relationship between organizational factors and attrition ofIT professional’s. The result can conclude that the organizational factorsplayed significant role in predicting the variance in turnover intention(attrition) of Kerala IT professionals. Therefore, the HR managers in ITorganizationsmaytakeintoconsiderationtheproblemswithorganizational

factorsoftheirworkerstoreducetheturnoverintentionoftheskilledemployees**.**

# LITERATURESURVEY

# ExistingProblem

The Existing system includes only few attributes for analysis andalso deals with qualitative observations and simple statistical analysis.Thequalitative observations deal with data and can be observed throughhuman senses.They do not involve measurements or number. Due to theincrease in IOT and connected device,we now have access to so much ofdataandalong withitanincreaseneeds tomanageand understanddata.

# References

* + 1. From Big Data to Deep Data to support people analytics foremployee attrition prediction, Nesrine Ben Yahia, HlelJihen, RicardoColomo-Palacio(2021)
    2. MachineLearningApproachforEmployeeAttritionAnalysis.Dr.

R. S. Kamath | Dr. S. S. Jamsandekar | Dr. P. G. Naik ,Published inInternational Journal of Trend in Scientific Research and Development(ijtsrd),(March2019)

* + 1. Investigation of early career teacher attrition(ECT) and theimpact of induction programs in Western Australia, Janine E.Wyatt,MichaelO’Neill (2021)

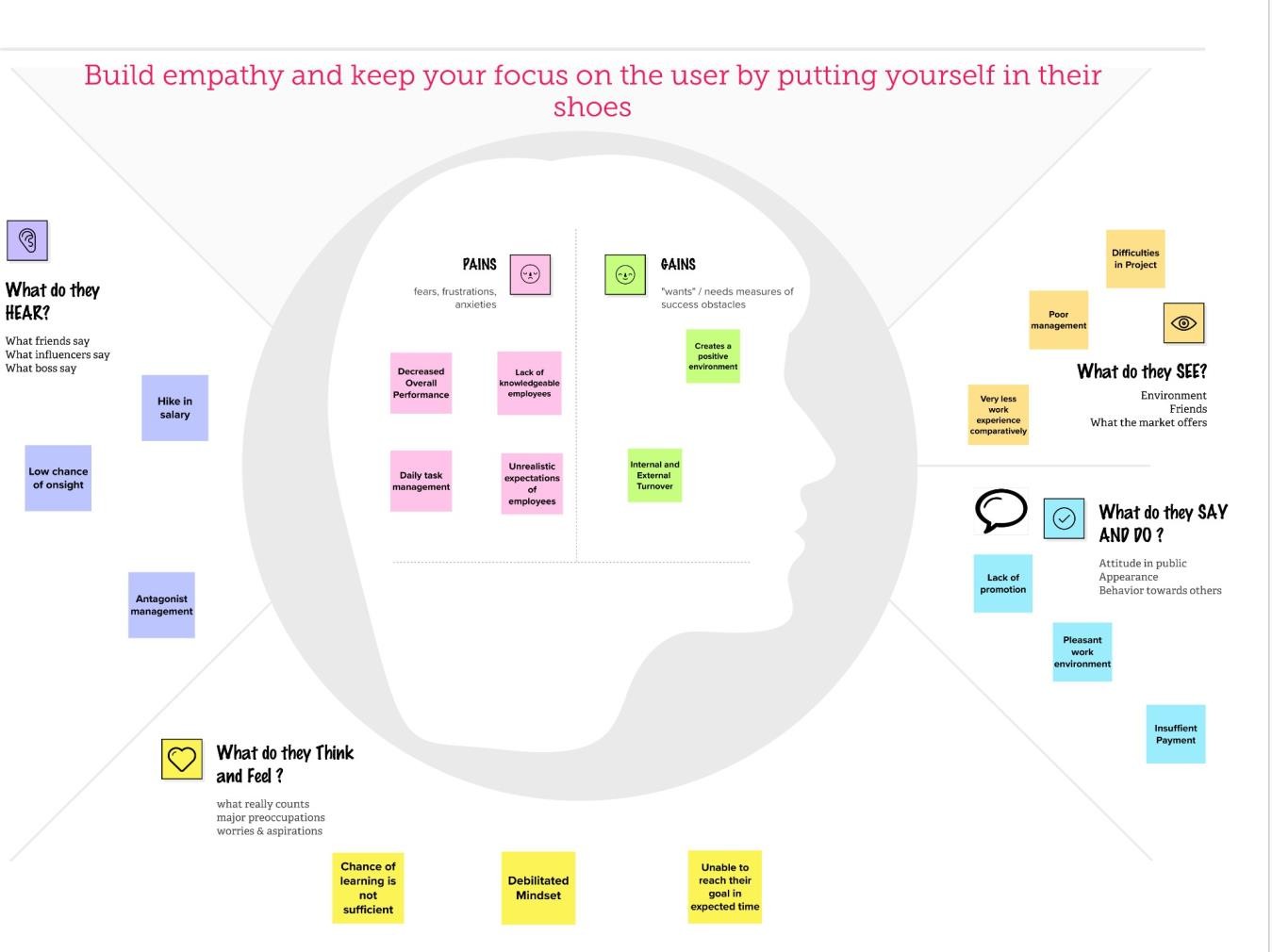
# ProblemStatementDefinition

* TocreateadashboardandperformanalysisofemployeeattritionincorporatesusingIBMCognosanalyticsplatform.
* Toreducetheemployeeattritionratethroughdataanalytics,

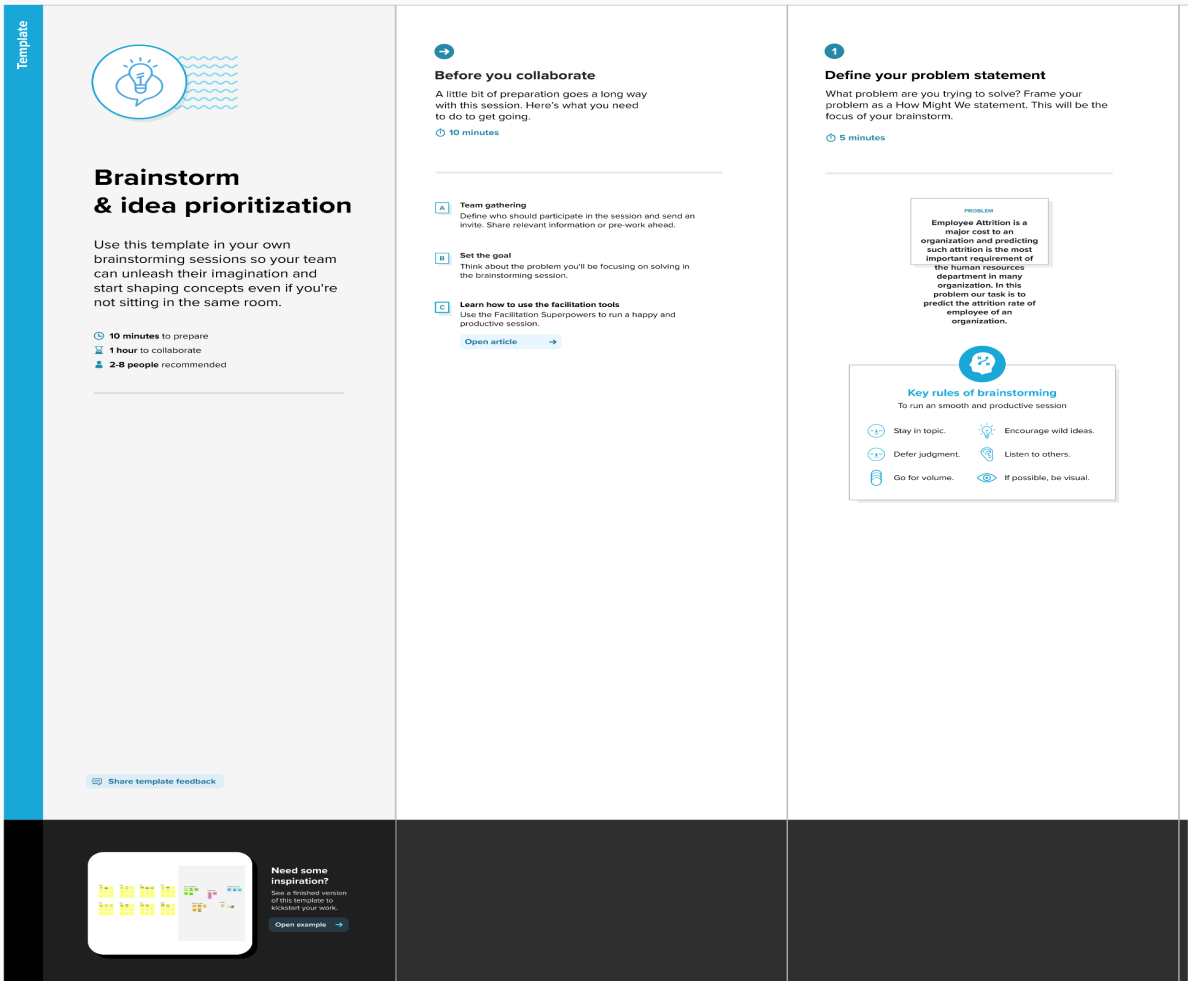
datavisualizationbyanalysingthemajorfactorsthatcausesattrition.

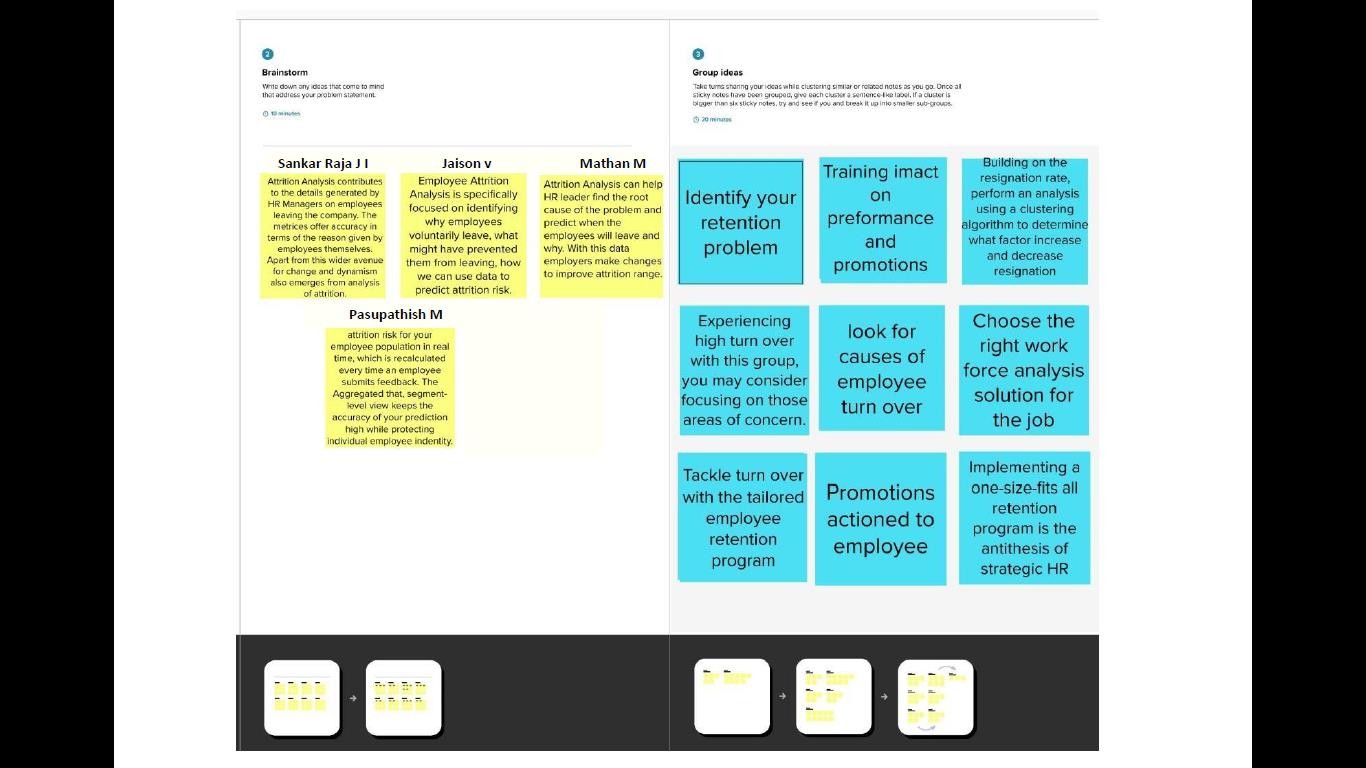
# IDEATIONANDPROPOSEDSOLUTION

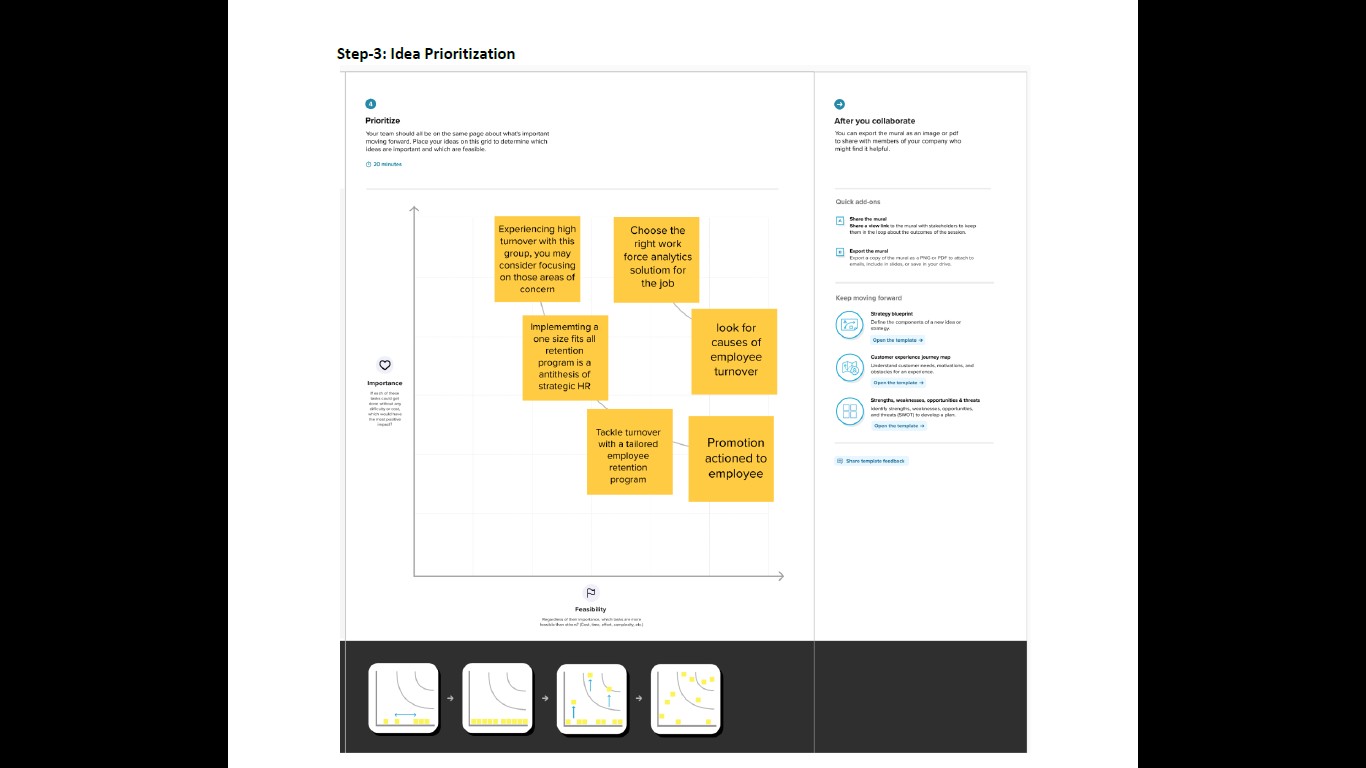
# EmpathyMapCanvas



# Ideation&Brainstorming





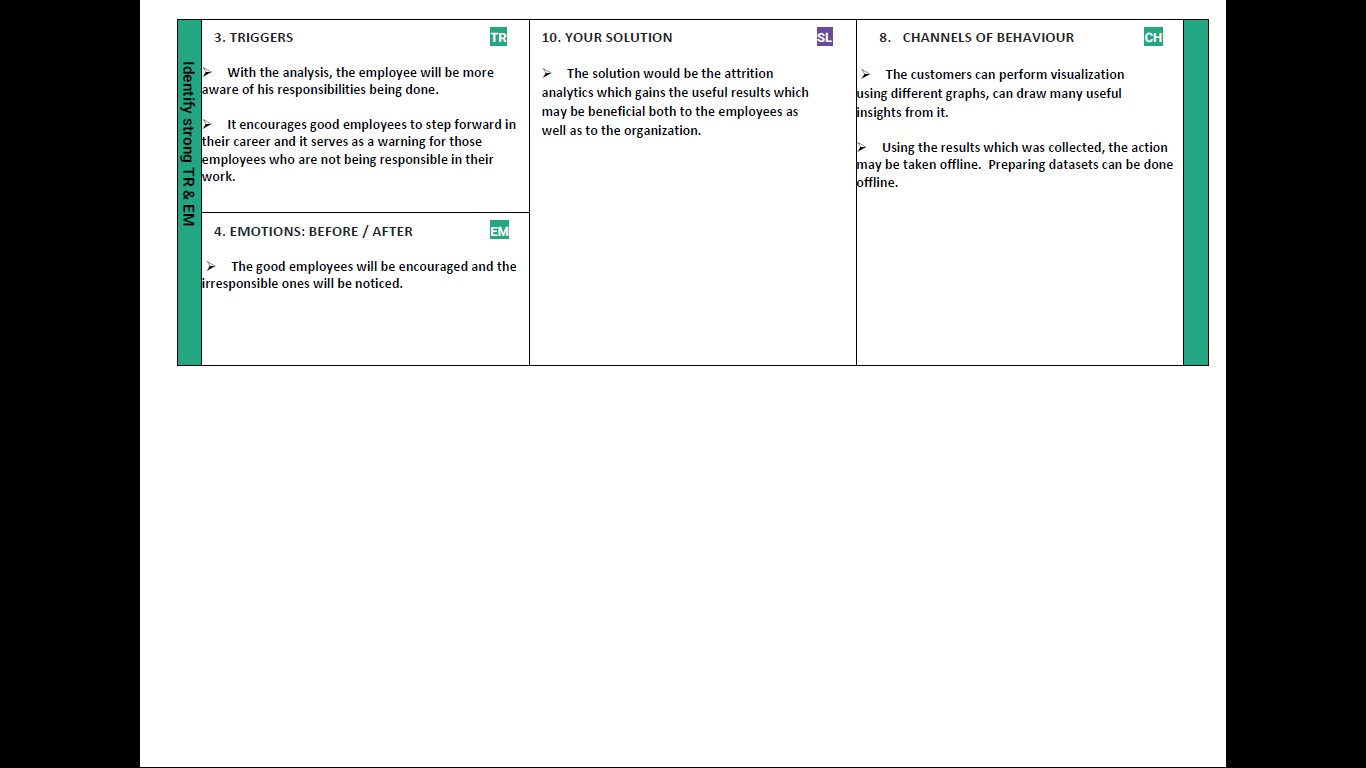
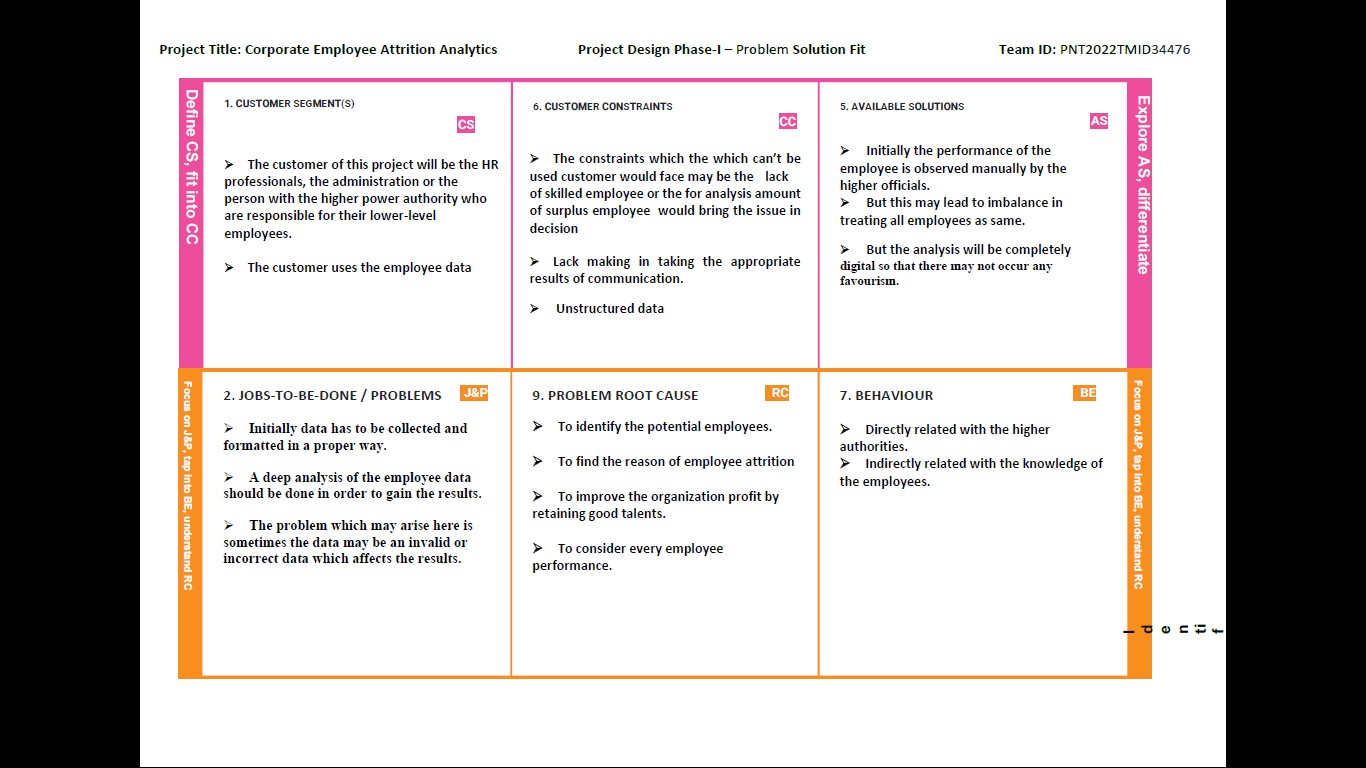


# ProposedSolution

TheExistingsystemincludesonlyfewattributesforanalysis

and also deals with qualitative observations and simple statistical analysis.The qualitative observations deal with data and can be observed throughhuman senses.They do not involve measurements or number. Due to theincrease in IOT and connected device,we now have access to so much ofdataand along with itanincreaseneedsto manageand understanddata.

# ProblemSolutionfit



# REQUIREMENTANALYSIS

# Functionalrequirement

|  |  |  |
| --- | --- | --- |
| **FRNo.** | **FunctionalRequirement(Epic)** | **SubRequirement(Story/Sub-Task)** |
| FR-1 | UserRegistration | Registration through FormRegistrationthroughGmail  RegistrationthroughLinkedIN |
| FR-2 | UserConfirmation | ConfirmationviaEmailConfirmationviaOTP |
| FR-3 | UserFeedback | Feedback throughFormFeedback through GmailFeedbackthroughInstagrampolls  FeedbackthroughLinkedIn |
| FR-4 | UserRating | RatingviaMail  RatingthroughMessage |
| FR-5 | EmployeeManagement | Validatingandmanagingtheemployeedetails |
| FR-6 | AttritionAnalytics | Analysingandfindingoutthemajorreasonfortheattritionofemployeesusingdataset |

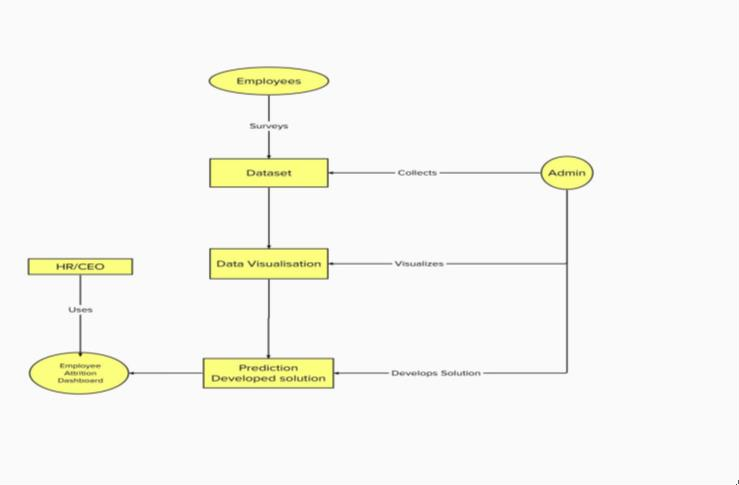
# Non-Functionalrequirements

Followingarethenon-functionalrequirementsoftheproposedsolution.

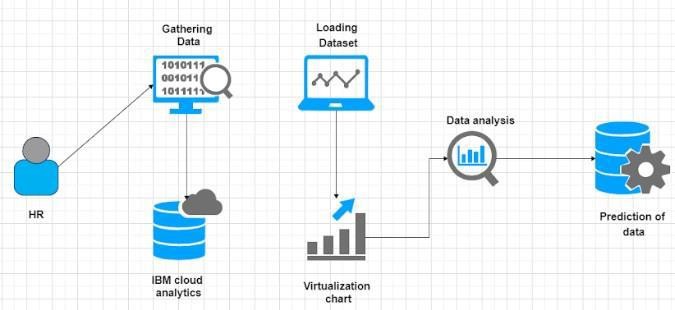
|  |  |  |
| --- | --- | --- |
| **FRNo.** | **Non-FunctionalRequirement** | **Description** |
| NFR-1 | **Usability** | This Data Visualization shall be easy to use for allusers with minimal instructions. 100% of thelanguagesonthegraphicaluserinterface(GUI)shallbe intuitive andunderstandablebynon-technical  users. |
| NFR-2 | **Security** | Theemployeedataiskeptsecureandtheiridentityishiddenfortheorganization. |
| NFR-3 | **Reliability** | TheLinkshallbeoperableinallconditions.Thesystemmustbelesspronetoerrors |
| NFR-4 | **Performance** | Thissoftwareisportableandinter-operable.Itworkssmoothlywithoutgeneratingerrors.Italso provides  afasterresponse |
| NFR-5 | **Portability** | Thelinkshallbeportabletoalloperatingplatforms.  Therefore,thislinkshouldnotdependonthedifferentoperatingsystems. |
| NFR-6 | **Scalability** | Oursolutionisscalableforlargeandsmalldatasets.Itprovides anefficientsolutiondespite the sizeof  thedataset. |

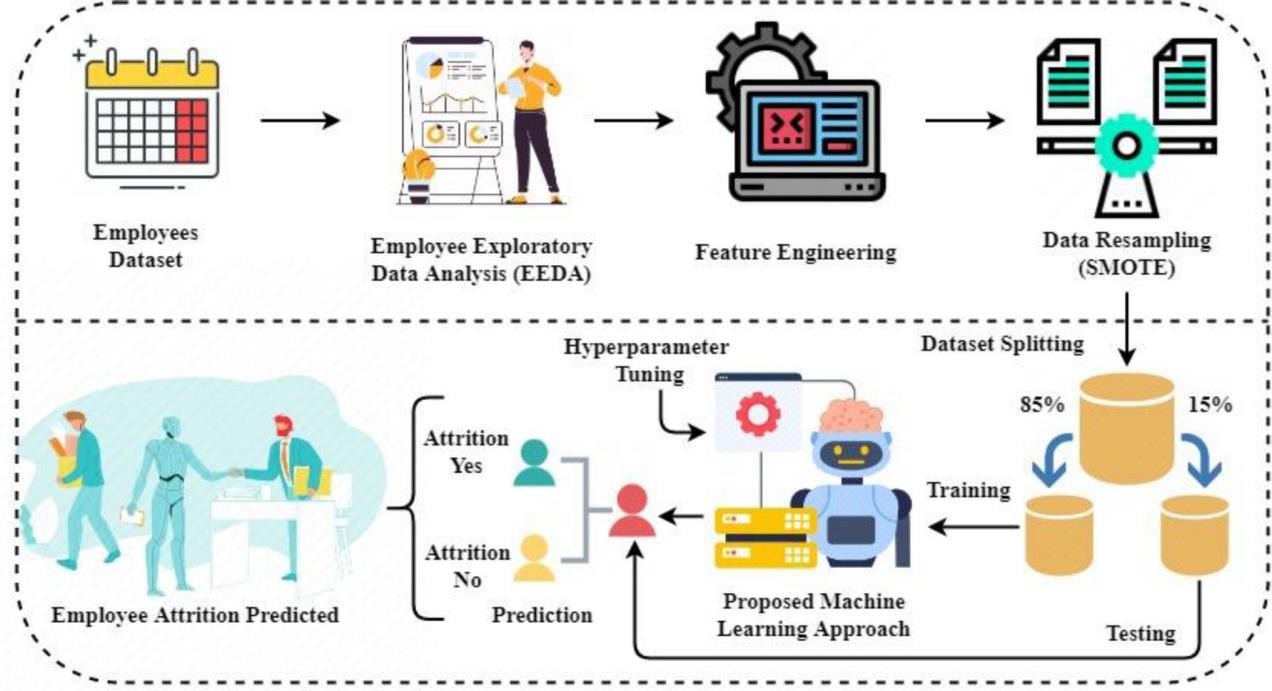
# PROJECTDESIGN

# DataFlowDiagrams

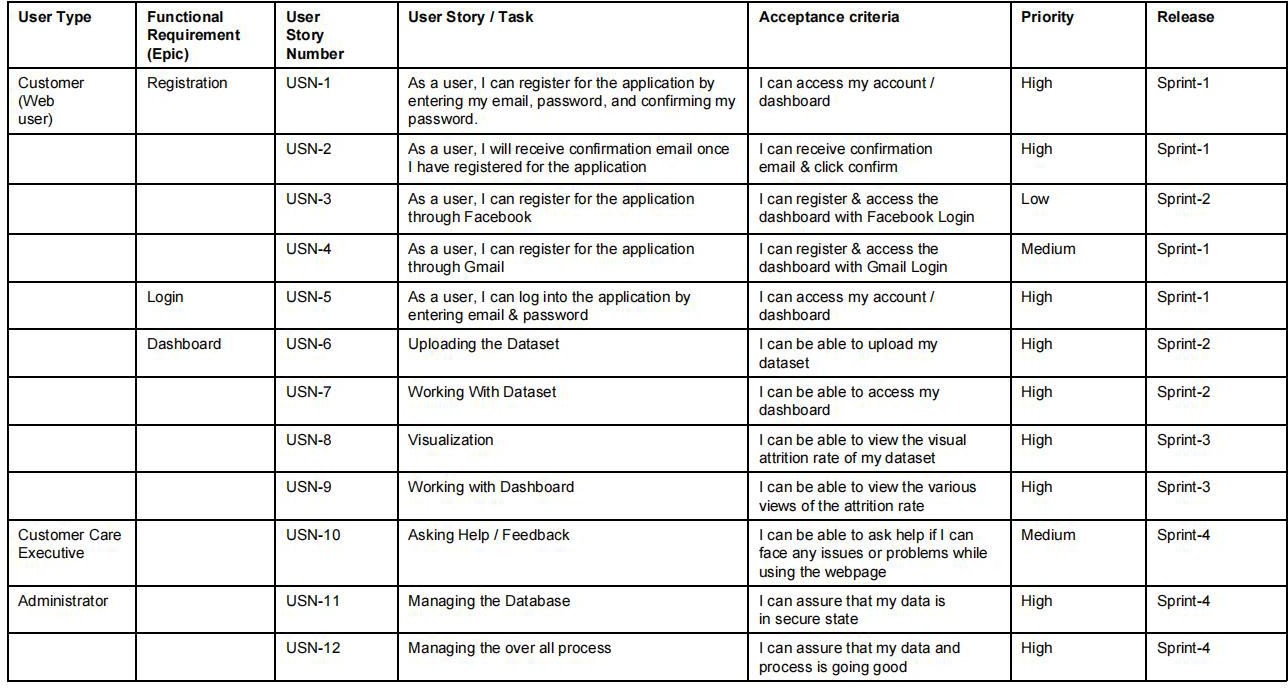


# Solution&TechnicalArchitecture





# UserStories



# PROJECTPLANNING

# SprintPlanning&Estimation

****

# SprintDeliverySchedule



# CODING&SOLUTIONING

* 1. **Feature 1**#GENERALimport pandas as pdimportnumpyasnp

importseabornassns

import matplotlib.pyplot as pltpath='/content/general\_data.csv'df=pd.read\_csv(path)

df

df**.**shapedf**.**info()

df**.**select\_dtypes('int64' ,'float64')**.**columnscat\_cols**=** df**.**select\_dtypes('object')**.**columnscat\_cols

df**.**describe()**.**Tdf

**for**cat**in**cat\_cols:

print(cat,'->',df[cat]**.**unique())print()

print("AllcolumnsUniquevaluescount")

**for**col**in**df:

print(col, len(df[col]**.**unique()), sep**=**': ')plt**.**figure(figsize**=**(14,5))plt**.**subplot(1,2,1)

sns**.**countplot(df['Attrition'] ,color **=**'b' ,hue **=**df['Gender'])plt**.**title('Attritionby Gender')

plt**.**subplot(1,2,2)

plt**.**pie(df['Attrition']**.**value\_counts(),colors**=**['r' ,'c'],explode**=**[0,0.1],autopct**=**

'%.2f' ,labels**=**['No' ,'Yes'])plt**.**title('Attrition')

*#HANDLINGCATEGORICALOUTPUTVARIABLE*

df['Attrition']**.**replace({'Yes':1 ,'No':0} ,inplace**= True**)df['Attrition']**.**head()

plt**.**figure(figsize**=**(20 ,8))

sns**.**boxplot(x**=**'JobRole', y **=**'MonthlyIncome' ,data **=** df,hue**=**'Attrition',color**=**'red')

col**=**['YearsInCurrentRole','YearsSinceLastPromotion','YearsWithCurrManager'

,'YearsAtCompany']plt**.**figure(figsize**=** (10 ,10))**for**i,c**in** enumerate(col):

plt**.**subplot(2 ,2,i**+**1)sns**.**distplot(df[c],color**=**'b')

* 1. **Feature 2**#GENERALimport pandas as pdimportnumpyasnp

importseabornassns

importmatplotlib.pyplotasplt#FEATUREENGINEERING

fromsklearn.preprocessingimportLabelEncoderfrom imblearn.over\_sampling import SMOTEpath='/content/general\_data.csv'

df=pd.read\_csv(path)df

df.shapedf.info()

df.select\_dtypes('int64','float64').columns

cat\_cols = df.select\_dtypes('object').columnscat\_cols

df.describe().Tdf

forcatincat\_cols:

print(cat,'->',df[cat].unique())print()

print("All columns Unique values count")forcol in df:

print(col, len(df[col].unique()), sep=': ')plt.figure(figsize =(14,5))plt.subplot(1,2,1)

sns.countplot(df['Attrition'] ,color ='b' ,hue =df['Gender'])plt.title('Attritionby Gender')

plt.subplot(1,2,2)

plt.pie(df['Attrition'].value\_counts() ,colors =['r' ,'c'] ,explode =[0,0.1],autopct ='%.2f',labels =['No' ,'Yes'])

plt.title('Attrition')

*#HANDLINGCATEGORICALOUTPUTVARIABLE*

df['Attrition'].replace({'Yes':1 ,'No':0} ,inplace = True)df['Attrition'].head()

df.drop(columns=no\_use,axis=1,inplace=True)df.columns

df['Gender'].replace({'Male':1,'Female':0},inplace=True)

df['OverTime'].replace({'Yes':1 ,'No':0} ,inplace = True)(df.Attrition.value\_counts()/1470)\*100

smote = SMOTE(sampling\_strategy='minority')x,y =smote.fit\_resample(x ,y)

print(x.shape,y.shape)*#now balanced*y.value\_counts()

sns.countplot(y ,palette='viridis')plt.title('NowClassisBalanced')

# TESTING

# TestCases

# UserAcceptanceTesting

* + 1. **PurposeofDocument**

The purpose of this document is to briefly explain the test coverage and open issueofcorporateemployee attritionatthetimeoftherelease.

* + 1. **DefectAnalysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Resolution** | **Severity1** | **Severity2** | **Severity3** | **Severity4** | **Subtotal** |
| ByDesign | 3 | 2 | 0 | 0 | 5 |
| Duplicate | 4 | 0 | 2 | 0 | 6 |
| External | 3 | 2 | 0 | 0 | 5 |
| Fixed | 1 | 0 | 1 | 0 | 2 |
| Not Reproduced | 0 | 3 | 3 | 0 | 6 |
| Skipped | 0 | 0 | 3 | 2 | 5 |
| Won't Fix | 0 | 0 | 1 | 0 | 1 |
| Totals | 11 | 7 | 10 | 2 | 30 |

* + 1. **TestCaseAnalysis**

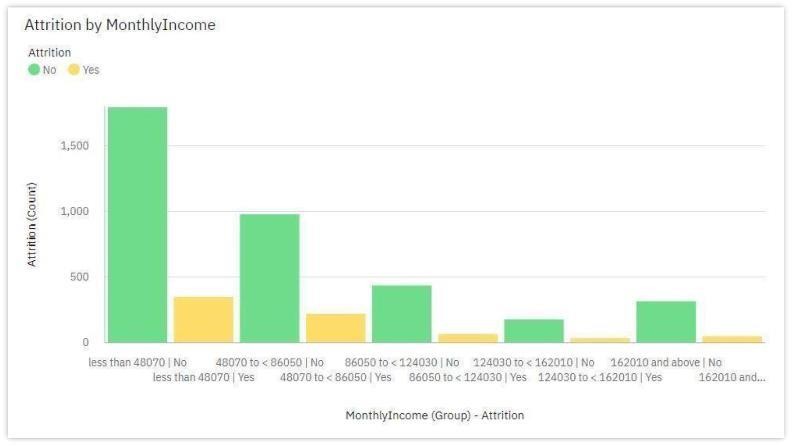
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Database | 2 | 0 | 0 | 2 |
| Dashboard | 1 | 0 | 0 | 1 |
| Visualize thedata | 8 | 0 | 0 | 8 |
| LogisticRegression | 4 | 0 | 0 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Section** | **TotalCases** | **NotTested** | **Fail** | **Pass** |
| LoginPage | 1 | 0 | 0 | 1 |
| EmployeeAttritionDetails | 1 | 0 | 0 | 1 |

# 9.RESULTS

**9.1PerformanceMetrics**

****



# 9. ADVANTAGES&DISADVANTAGES

9.1Advantages

Data Collection : The study is conducted among working ITprofessionalsoftwodifferentcategories.Thiscategorizationmainlywasfocusedonexperiencelevelandroleintheorganization.Itwasimportanttoknowtheviewsofcandidateswhoseekforthejobforvariousreasonsaswellastheviewsofinterviewersinvolvedintheprocessofhiringthecandidates.Theresearchstudyinvolvesreferenceofbothprimaryandsecondarydata.PrimaryDataPrimarydataiscollectedthroughafieldsurveywiththehelpofastructuredself-administratedQuestionnaire.Thesurveyconsistedofcloseendedquestionsbythemeansofconveniencesampling.Thescalingtechniqueinstalledinthequestionnaireis5-pointratingscale.Total120respondentwereITprofessionalsbelongingto theorganizationsfromNagpur,PuneandMumbaicitiesinMaharashtra.SecondaryDataSecondarydataiscollectedbyreferringtotheJournals,researchpapersandpublisheddataintheformofbooks andnewspapers.

Typeof Research:

Theresearchpaperadoptedthedescriptiveresearchdesignmethodology.SampleDesign,SampleSizeandSamplingMethodThesampleselected for the study is an Indian Information Technology Industry. Thenature of the sample is restricted to working professionals in InformationTechnologysectorandiscollectedthroughtheconveniencesamplingtechnique.The samplesize was120 respondents.

# CONCLUSION

Employeesaswellasorganizationsmustbeclearwiththeirexpectationsregardingthejobprofile.Anysortofmismatchleadstodiscrepancy and employees may fail to perform at theirjob. This eventuallyleadstoattrition.Organizationsshouldstatetherequirementsandexpectationsunambiguously.Thishelpscandidatesdecideupontoacceptthejobpositionornot.Thiseventuallyavoidsfurtherconflictsintheemploymentterms.

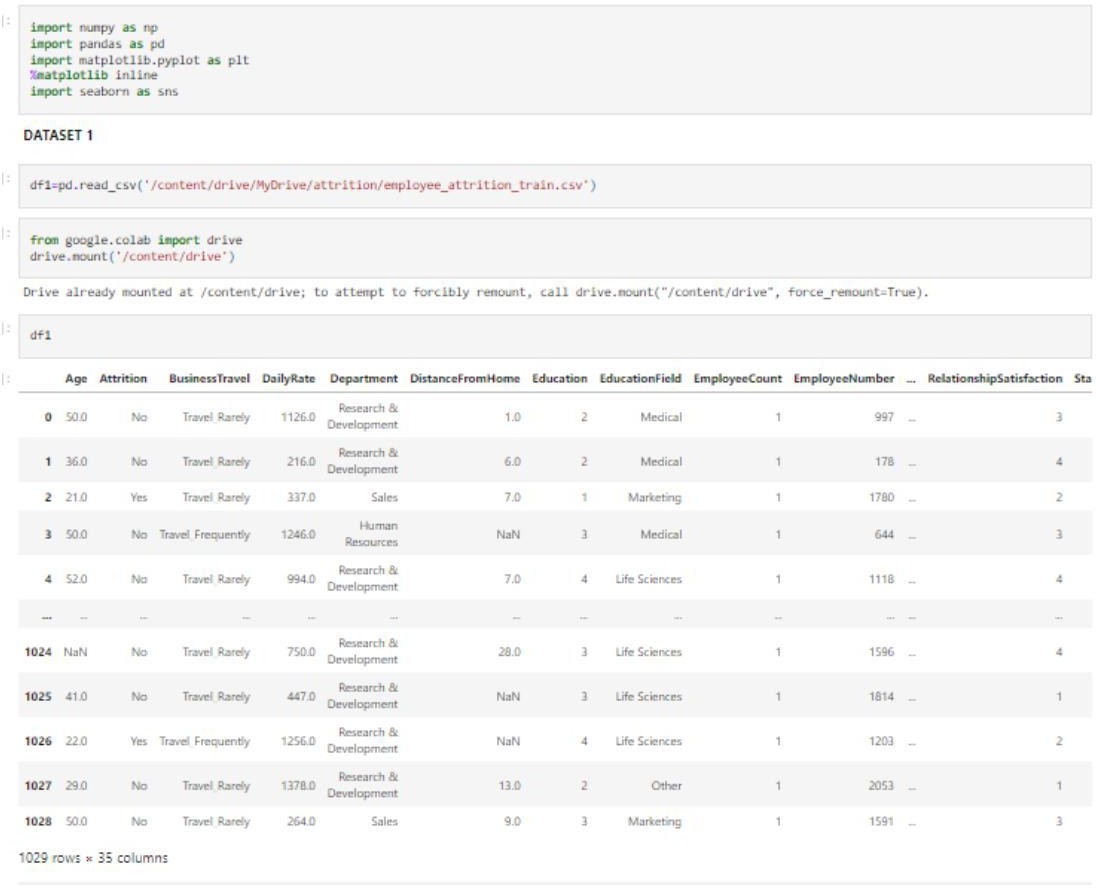
# FUTURESCOPE

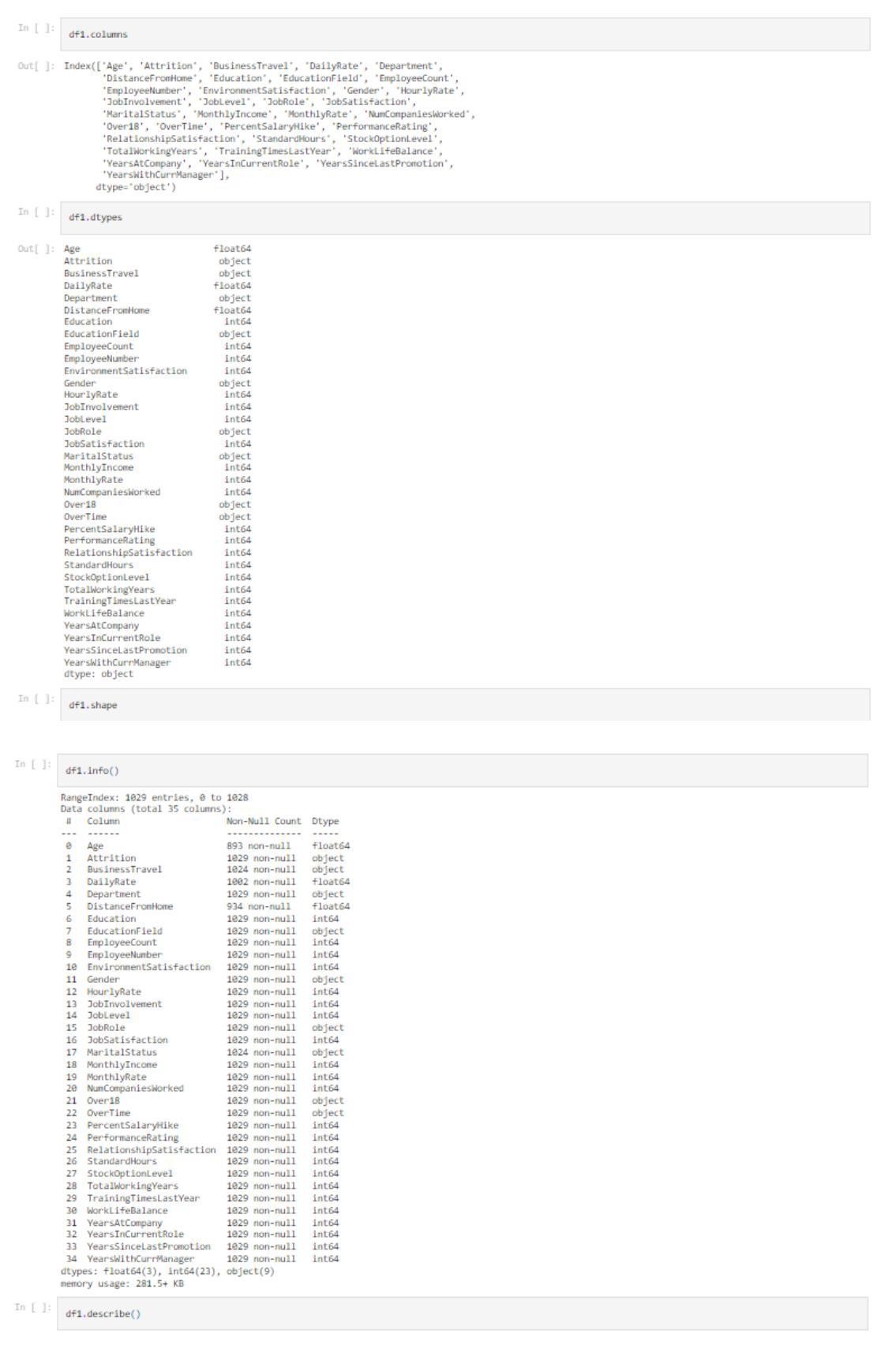
Research findings suggest that attrition reasons in IT organizationsprimarilyrevolvearoundprofessionalgrowthandchallengesintheorganization.Althougheconomicfactorshappentothemostinfluentialfactor,professionals may settle for second best criteria of their preference that iscareer growth and supportive work policies in the organization. On the otherhand,candidateswhoaspiretohaveabetterjobthantheoneinhandaremoreinterested in securing the next job. Young talent wants to work on latesttechnology and functional domain. IT professionals who are young careermakers are less influenced by Brand name or geographical area. Most of theIT professionals look for challenging role and position in the organization.Candidates as well as senior professionals believe that challenging workmotivate them to maintain the interest in the work life. Employees as well asorganizations must be clear with their expectations regarding the job profile.Anysort ofmismatchleadstodiscrepancyand employees may failto

perform at their job. This eventually leads to attrition. Organizations shouldstatetherequirementsandexpectationsunambiguously.Thishelpscandidatesdecide upon to accept the job position or not. This eventually avoids furtherconflicts in the employment terms. Further this research can make moredetailedconclusionsover“mappingofcandidates’expectationswithorganizations’requirement”bycollectingthedatafocusingonallthestepsofrecruitment andselectionprocess.

# APPENDIX

**12.1 SourceCode**





22

