Capstone Project
Background and Objectives

HR Analytics: To develop a scientific method of understanding the

factors that lead to Attrition in an organization

Background

1.Client
A healthcare R&D organization

2.Background
The company 1470 employees . Over the last year , The HR team has collected and consolidated information through various sources (like survey , appraisal report, salary structure etc.) about each emplc

The company wants to understand what parameters impact the attrition in the organization and how can they make the work place a more conducive environment

Background

3. Objective

– To analyze attrition and correlate with other factors

– To develop employee attrition model

– To analyze employee sentiments

– Optional- Develop web application using R Shiny

DATA SCIENCE INSTITUTE

Background

4. Data

The data used for the same is the employee data as on 31st Dec '2022

The data available is the following :
1. Employee Attrition
2. Employee Demographics
3. Employee Income
4. Employee Job details
5. Exit Interview Responses

TA SCIENCE
INSTITUTE

Data : Employee Attrition

This data is mainly to understand whether the Employee has resigned or not anytime in the year 2022

EmpId	Attrition	Department
100001	Yes	Sales
100002	No	Research & Development
100003	Yes	Research & Development
100004	No	Research & Development
100005	No	Research & Development
100006	No	Research & Development
100007	No	Research & Development
100008	No	Research & Development
100009	No	Research & Development
100010	No	Research & Development
100011	No	Research & Development
100012	No	Research & Development
100013	No	Research & Development

Columns	Description	Type	Measurement	Possible values
EmpID	Employee ID	numeric	-	-
Attrition	whether he has resigned anytime in the year '19	Factor	-	Yes, No
Department	Department	Factor	-	Sales , HR , R&D

DATA SCIENCE
INSTITUTE

Data : Employee Demographics

Content

This data is mainly about the demographic factors of the employee

EmpId	Age	DistanceFromHome	Education	EducationField	Gender	MaritalStatus
100001	41	1	2	Life Sciences	Female	Single
100002	49	8	1	Life Sciences	Male	Married
100003	37	2	2	Other	Male	Single
100004	33	3	4	Life Sciences	Female	Married
100005	27	2	1	Medical	Male	Married

Columns	Description	Type	Measurement	Possible values
EmpID	Employee ID	numeric	-	-
Age	Age of the employee	numeric	-	-
DistanceFromHome	Distance from home in kms	numeric	-	-
Education	Score based on level of education	ordinal	1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'	1,2,3,4,5
Educationfield	Field of Education	factor	-	-
Gender	Gender	factor	-	Male, Female
MaritalStatus	Marital Status	factor	-	Single , Divorced , Married

DATA SCIENCE INSTITUTE

Data : Employee Income

This data is mainly to understand his income components

EmpId	StandardHours	HourlyRate	MonthlyIncome	PercentSalaryHike	StockOptionLevel
100001	200	30	5993	11	0
100002	200	26	5130	23	1
100003	200	10	2090	15	0
100004	200	15	2909	11	0

Columns	Description	Type	Measurement	Possible values
EmpID	Employee ID	numeric	-	-
StandardHours	Minimum no of hours to be spent at work in a month	Numeric	Since all are full time employees , 200 is the std hours across	200
HourlyRate	How much does the employee make per hour	integer	-	-
MonthlyIncome	Monthly Income	Numeric	-	-
PercentSalaryHike	Field of Education	Factor	-	-
StockOption Level	No of stocks as a part of compensation benefits	Factor	0- none 1- <20 2-21-40 3->40	0,1,2,3,

DATA SCIENCE
INSTITUTE

Data : Employee Job Details

This data is mainly to understand his job details .

Columns	Description	Type	Measurement	Possible values
EmpID	Employee ID	numeric	-	-
BusinesTravel	Does the employee need to travel for work : Never , Rarely or Frequently	factor	-	Non-Travel, Travel_Rarely, Travel_Frequently
JobLevel	Level of Job.	Ordinal	1 =lowest ;5-highest	1,2,3,4,5
JobRole	Job Role	Factor	-	-
Jobsatisfaction		Factor	1=Low; 2=Medium; 3=High; 4=Very High	1,2,3,4
NumCompaniesworked	No of Companies worked previously	integer	-	-
PerformanceRating		Factor	1=Low; 2=Good ;3=Excellent; 4=Outstanding	1,2,3,4
Overtime	If the person has worked over 250 hrs in the month, OT= Yes	Factor	-	Yes,No
TotalWorkingYears	Total years of Experience	integer	-	-
Trainings	No of Training programs attended in	integer	-	-
WorkLifeBalance	Rating based on a survey conducted for work life balance	Ordinal	1=Bad; 2=Good; 3=Better; 4=Best	1,2,3,4
YearsAtCompany	No of year spent in the current organzation	integer	-	-
YearsatCurrentRole	No of years in the current role	Integer	-	-
YearssinceLastPromotion	No of years since the last promotion	Integer	-	-
YearsWithCurrManager	No of years with current manager	Integer	-	-

EmpId	Busin...	...me
100001	Trav	
100002	Trav	
100003	Trav	
100004	Trav	
100005	Trav	
100006	Trav	
100007	Trav	
100008	Trav	
100009	Trav	
100010	Trav	
100011	Trav	
100012	Trav	

DATA SCIENCE INSTITUTE

Data : Exit interview responses

Content

The response captured for the question "What is the main reason for your resignation ? You can be completely honest as this response is anonymous
Sample Size : 50

Job good . Remuneration is horrible

I have had no job satisfaction in the past 3 years

The salary Structure is something i am not happy with . I pay too much in taxes despite being compensated as per market Standards and dont end up saving anything. Also my variable comp

My work life balance was something i needed to check. I was working every weekend and was taking a toll on my health

No training , no upgradation , no change in the last 1 and half year i have been here . this is the first time in my career where i feel i havent gained any knowledge

Wanted a role change . Was not considered for the internal opening for the lateral shift . Hence decided to move

I wish my boss understood me more and took my ideas into consideration

I feel this job doesnt give me any flexibility . Too much travel and i never get a work from home despite the same being a part of the organazation's Structure

I didnt get promoted despite being promised . I dont think that is fair

Better opportunity is the reason i am moving out . Else all was good but i got an opportunity i couldnt resist

The Organzation needs to rethink their rewards , Recognition and appreciation policies . These benefi...

| Description	Type
Consumer Responses	text

DATA SCIENCE INSTITUTE

Next steps

Data management

- Compile all 4 data files based on Employee ID
- Data cleaning , Handling missing values and completing Basic Data checks
- Check if any variables needed to be feature coded i.e made into groups or want to be left as continuous variables e.g Age

## Descriptive Statistics & Data visualization
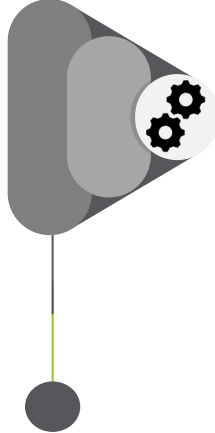
- Understanding the data better like checking attrition rate dept wise , is salary playing a role in attrition etc.
- How can this data be presented better visually ?
- Once again post Data visualization check if any variable needs to be feature coded

Predictive modelling

- Answer the objectives a) To understand factors to pre-empt which employees who are likely to leave in '20 b) To identify the parameters that lead to early churn (within 18 months of joining)
- Using different Predictive model techniques to find Significant variables
- Ensure you follow all steps like Train and test data , checking for Multicollinearity
- Check if any other ML technique fits better

Text mining
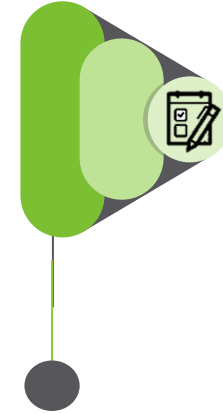
- Understand the employee responses using various Text mining Techniques
- Answer the objective:To understand the main reasons for quitting the organization

10

DATA SCIENCE INSTITUTE