**EDA Project Presentation :**

**Name Of Candidate :-** Meeniga Nagendra

**Dataset Name :-** HR dataset ("HR\_DATA\_Excel.xlsx")

**Description :-**

An HR dataset typically includes various attributes related to employees in an organization. These attributes can cover personal information, job details, performance metrics, and other relevant data. In Dataset It have 44 columns and 1470 rows.

**Columns Info :-**

0 Attrition :- Indicates whether an employee has left the company (Yes/No).

1 Business\_Travel :- Describes the frequency of business travel.

2 CF\_age\_band :- Age band or group of employees.

3 CF\_attrition\_label :- Another indicator for employee attrition.

4 Department :- The department in which the employee works.

5 Education\_Field :- Stream of Education

6 emp\_no :- Employee Number

7 Employee\_Number :- employee number

8 Gender :- Gender of employee

9 Job\_Role :- job role of an employee

10 Marital\_Status :- employee is married or not

11 Over\_Time :- employee works over time

12 Over18 :- Indicates whether the employee is over 18 years old (Yes/No)

13 Training\_Times\_Last\_Year :- Number of training sessions the employee attended last year

14 -2 :-

15 0 :-

16 Age :- Age of the employee

17 CF\_attrition\_count :- Count related to employee attrition.

18 CF\_attrition\_counts :- Another count related to employee attrition

19 CF\_attrition\_rate :- Attrition rate, potentially calculated based on counts.

20 CF\_current\_Employee :- Employee working or note in that company

21 Daily\_Rate ;- Daily rate of pay for the employee.

22 Distance\_From\_Home :- Distance form home to office in KM's

23 Education :- Qualification of employee (masked)

24 Employee\_Count :- Employees count

25 Environment\_Satisfaction :- Satisfaction level regarding the work environment.

26 Hourly\_Rate :- Hourly rate of pay for the employee.

27 Job\_Involvement :- Level of involvement in the job.

28 Job\_Level :- level of Job

29 Job\_Satisfaction :- if employee is satisfied?

30 Monthly\_Income :- Monthly income of the employee.

31 Monthly\_Rate :- Monthly pay rate.

32 Num\_Companies\_Worked :- Number of companies the employee has worked for.

33 Percent\_Salary\_Hike :- Percentage increase in salary.

34 Performance\_Rating :- Rating of the employee's performance.

35 Relationship\_Satisfaction ;- Satisfaction level regarding work relationships.

36 Standard\_Hours :- per week standard working hours

37 Stock\_Option\_Level :- company stock option level or Level of stock options provided to the employee.

38 Total\_Working\_Years :- Total number of years the employee has worked.

39 Work\_Life\_Balance :- Balance between work and personal life.

40 Years\_At\_Company :- Employee total years at current company

41 Years\_In\_Current\_Role ;- employee total years in current role

42 Years\_Since\_Last\_Promotion :- Number of years since the last promotion.

43 Years\_With\_Curr\_Manager :- Number of years the employee has been with the current manager.

**Information About Columns :-**

0 Attrition 1470 non-null object

1 Business\_Travel 1470 non-null object

2 CF\_age\_band 1470 non-null object

3 CF\_attrition\_label 1470 non-null object

4 Department 1470 non-null object

5 Education\_Field 1470 non-null object

6 emp\_no 1470 non-null object

7 Employee\_Number 1470 non-null int64

8 Gender 1470 non-null object

9 Job\_Role 1470 non-null object

10 Marital\_Status 1470 non-null object

11 Over\_Time 1470 non-null object

12 Over18 1470 non-null object

13 Training\_Times\_Last\_Year 1470 non-null int64

14 -2 1470 non-null int64

15 0 1470 non-null int64

16 Age 1470 non-null int64

17 CF\_attrition\_count 1470 non-null int64

18 CF\_attrition\_counts 237 non-null float64

19 CF\_attrition\_rate 1470 non-null int64

20 CF\_current\_Employee 1470 non-null int64

21 Daily\_Rate 1470 non-null int64

22 Distance\_From\_Home 1470 non-null int64

23 Education 1470 non-null object

24 Employee\_Count 1470 non-null int64

25 Environment\_Satisfaction 1470 non-null int64

26 Hourly\_Rate 1470 non-null int64

27 Job\_Involvement 1470 non-null int64

28 Job\_Level 1470 non-null int64

29 Job\_Satisfaction 1470 non-null int64

30 Monthly\_Income 1470 non-null int64

31 Monthly\_Rate 1470 non-null int64

32 Num\_Companies\_Worked 1470 non-null int64

33 Percent\_Salary\_Hike 1470 non-null int64

34 Performance\_Rating 1470 non-null int64

35 Relationship\_Satisfaction 1470 non-null int64

36 Standard\_Hours 1470 non-null int64

37 Stock\_Option\_Level 1470 non-null int64

38 Total\_Working\_Years 1470 non-null int64

39 Work\_Life\_Balance 1470 non-null int64

40 Years\_At\_Company 1470 non-null int64

41 Years\_In\_Current\_Role 1470 non-null int64

42 Years\_Since\_Last\_Promotion 1470 non-null int64

43 Years\_With\_Curr\_Manager 1470 non-null int64

dtypes: float64(1), int64(30), object(13)

**Analysis Qustions :-**

### 1 .Take The Department and gender column find the male and female employees with department wise.

### 2.Print the male and female how many they are married, unmarried and devorced.

### 3.if they are married they are satisfied with their job or not.

### 4.if they are married since how many years they are work in same company.

### 5.if they are working regarding good environment or high salary.

**Model Fitting :-**

**1.Feature Selection :-** In My Feature selection I have taken three models

1.Decision Tree Classifier

2.Recursive Feature Elimination (RFE)

3.Feature Selection

Based on my Feature selection I found 4 un-necessary colunms out of 33 columns. I dropped the 4 un-necessary

1. Employee\_Number 2.Gender 3.Education 4. Performance\_Rating

**2.Standardaze Data :-**

**3.Create Model :-** I have taken 7 model in my Model creation

1.Logistic Regression

2.KNeighbors Classifier

3.Random Forest Classifier

4.Support Vector Machine (SVM)

5.Bagging Classifier

6.Ada Boost Classifier

7.XG Boost Classifier

In all my models Random Forest Classifier is given more accuracy and I created a pipe line based on my highest accuracy model Random Forest Classifier.

**4.Create Pipe line :-**

**5.User Test :-**

**6.Save Model :-** I have saved my model as “HR\_Data\_Set\_final\_model.sav”.

**7.Create Python file :-** I have created a python file through using syder notebook.

File name :- “hrdata.py”.

**All File Names :-**

Dataset Name  **:-** HR dataset ("HR\_DATA\_Excel.xlsx")

Python Notebook File name :- Nagendra\_EDA\_Project\_HR\_Dataset.ipynb

Model File Name :- HR\_Data\_Set\_final\_model.sav

Python file Name :- hrdata.py