HR
ANALYTICS

MINI PROJECT ON HR ANALYTICS

SUBMITTED TO :- ABHISHEK TIWARI

SUBMITTED BY :-

SHREYASH GAUR (201500674)

DEVESH SRIVASTAV (201500217)

SHIVANSHU AGRAWAL (201500666)

SIDDHANT YADAV (201500689)

PRAKHAR VERMA (201500495)

About

HR analytics is revolutionizing the way human resources departments operate, leading to higher efficiency and better results overall. Human resources have been using analytics for years. However, the collection, processing and analysis of data has been largely manual, and given the nature of human resources dynamics and HR KPIs, the approach has been constraining HR. Therefore, it is surprising that HR departments woke up to the utility of machine learning so late in the game. Here is an opportunity to try predictive analytics in identifying the employees most likely to get promoted.

Aim

Company is facing a problem inidentifying the right people for promotion (only for manager position and below) and prepare them in time. The final promotions are only announced after the evaluation and this leads to delay in transition to their new roles.
Hence, company needs our help in identifying the eligible candidates at aparticular checkpoint so that they can expedite the entire promotion cycle.Company have provided multiple attributes around Employee's past and current performance along with demographics. Now the task is to predict whether a potential employee at checkpoint in the test set will be promoted or not after the evaluation process.

OBJECTIVE

Our client is a large MNC and they have 9 broad verticals across the organization. One of the problem your client is facing is around identifying the right people for promotion (only for manager position and below) and prepare them in time. Currently the process, they are following is:

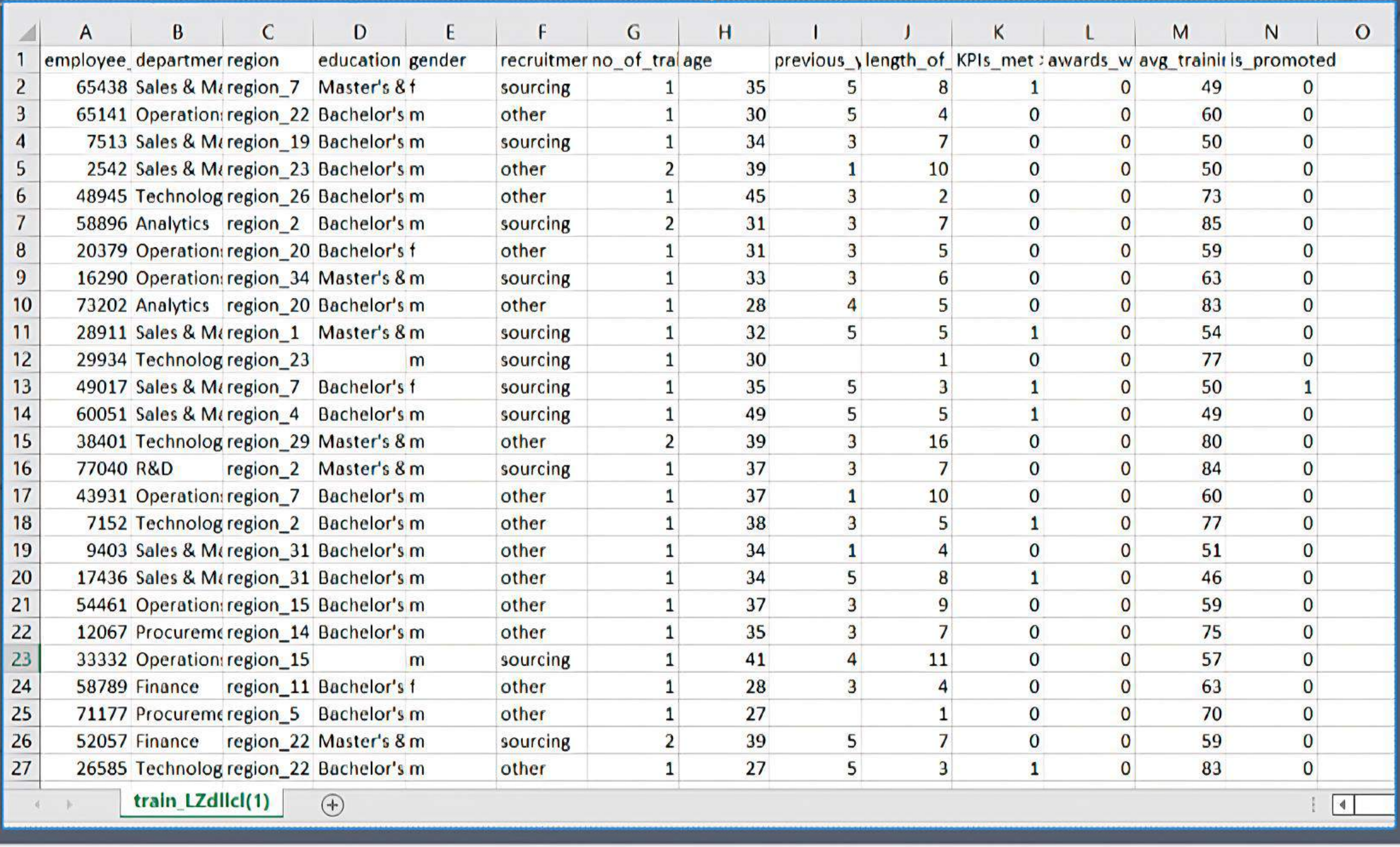
1. They first identify a set of employees based on recommendations/ past performance

2.Selected employees go through the separate training and evaluation program for each vertical. These programs are based on the required skill of each vertical

3.At the end of the program, based on various factors such as training performance, KPI completion (only employees with KPIs completed greater than 60% are considered) etc., employee gets promotion

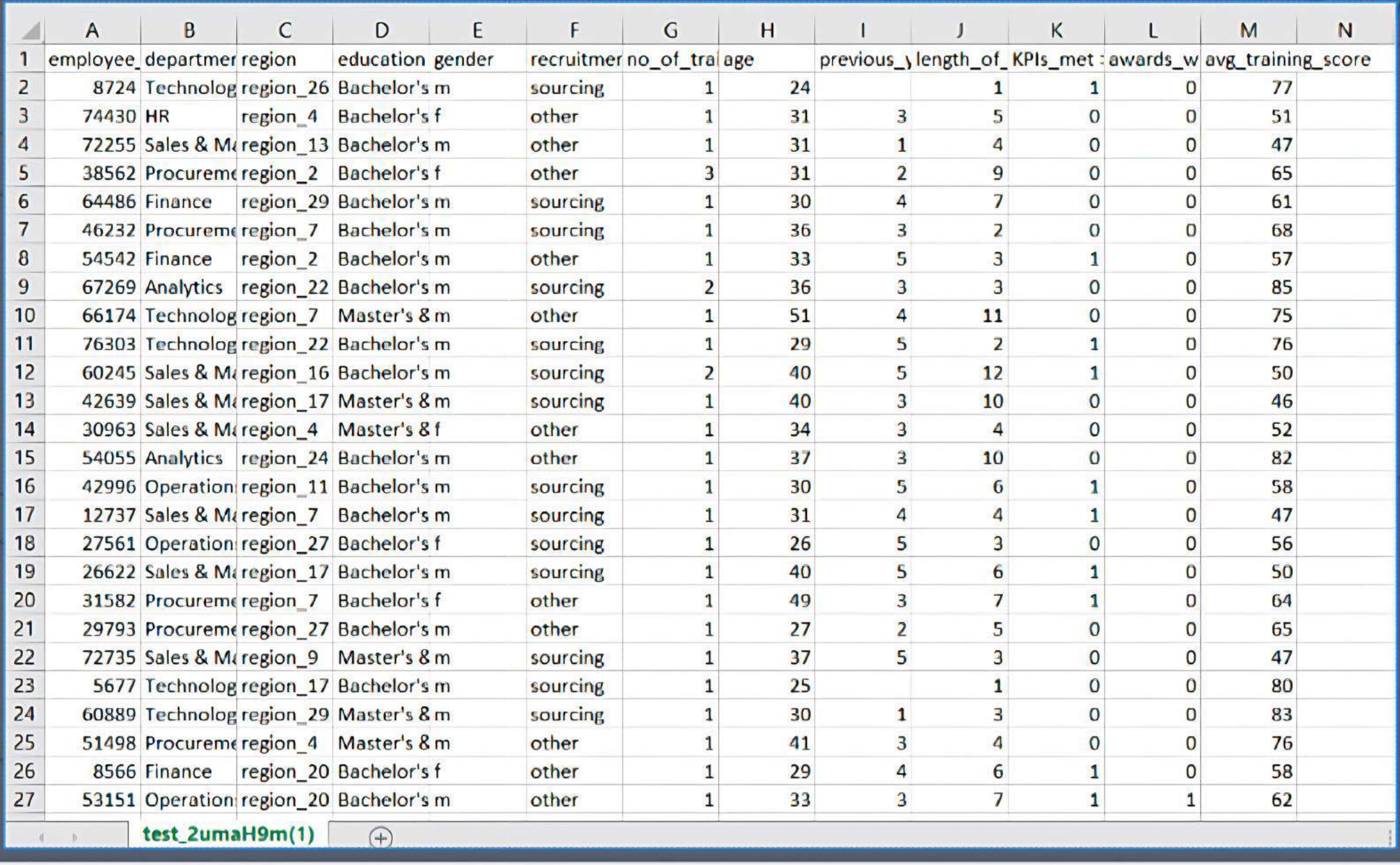
For above mentioned process, the final promotions are only announced after the evaluation and this leads to delay in transition to their new roles. Hence, company needs your help in identifying the eligible candidates at a particular checkpoint so that they can expedite the entire promotion cycle.

DATASET(FOR TRAINIG)



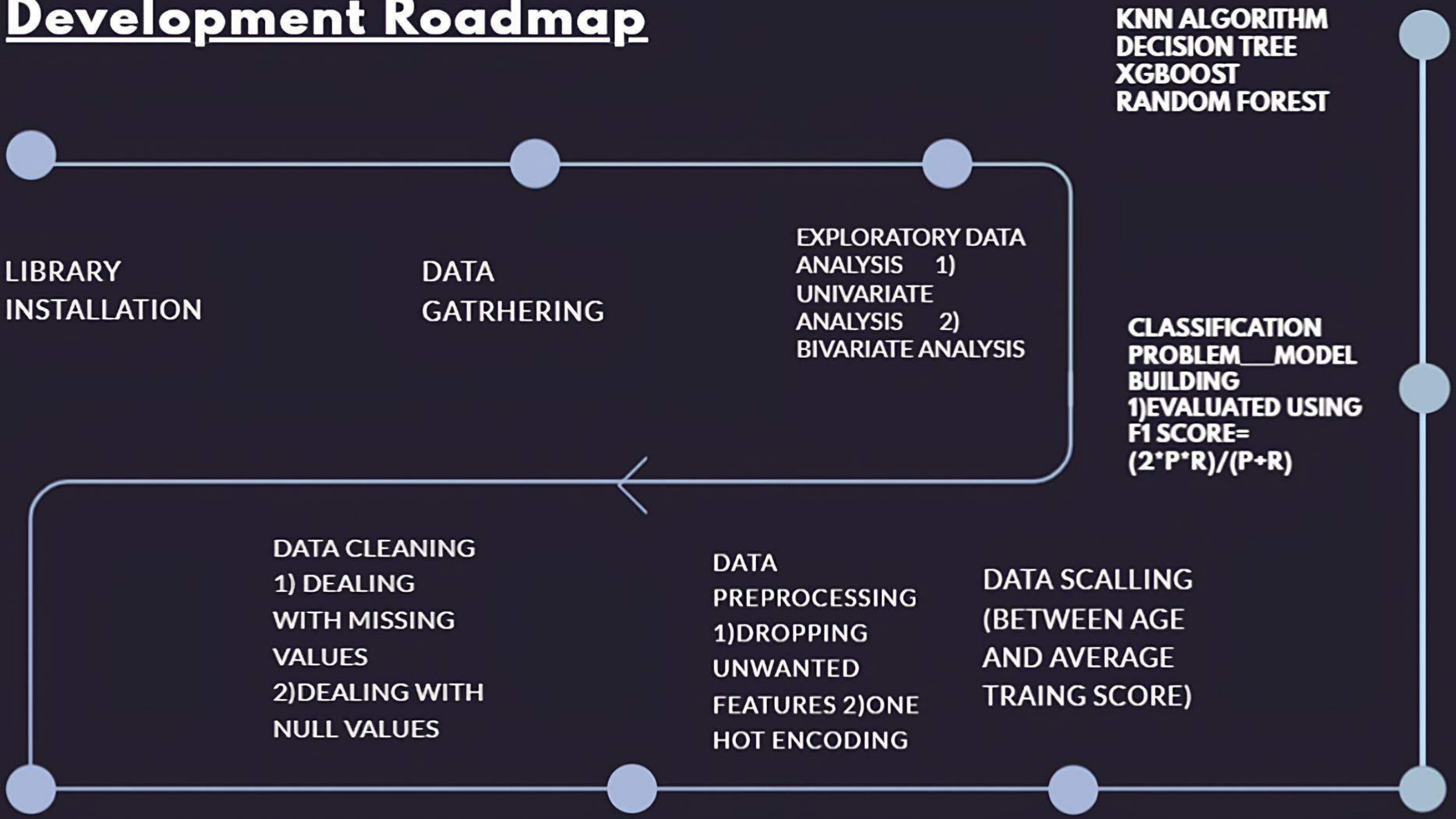
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	employee_	departmer	region	education	gender	recruitmer	no_of_tra	age	previous_y	length_of_	KPIs_met :	awards_w	avg_trainir	is_promoted	
2	65438	Sales & Ma	region_7	Master's &	f	sourcing	1	35	5	8	1	0	49	0	
3	65141	Operation:	region_22	Bachelor's	m	other	1	30	5	4	0	0	60	0	
4	7513	Sales & Ma	region_19	Bachelor's	m	sourcing	1	34	3	7	0	0	50	0	
5	2542	Sales & Ma	region_23	Bachelor's	m	other	2	39	1	10	0	0	50	0	
6	48945	Technolog	region_26	Bachelor's	m	other	1	45	3	2	0	0	73	0	
7	58896	Analytics	region_2	Bachelor's	m	sourcing	2	31	3	7	0	0	85	0	
8	20379	Operation:	region_20	Bachelor's	f	other	1	31	3	5	0	0	59	0	
9	16290	Operation:	region_34	Master's &	m	sourcing	1	33	3	6	0	0	63	0	
10	73202	Analytics	region_20	Bachelor's	m	other	1	28	4	5	0	0	83	0	
11	28911	Sales & Ma	region_1	Master's &	m	sourcing	1	32	5	5	1	0	54	0	
12	29934	Technolog	region_23		m	sourcing	1	30		1	0	0	77	0	
13	49017	Sales & Ma	region_7	Bachelor's	f	sourcing	1	35	5	3	1	0	50	1	
14	60051	Sales & Ma	region_4	Bachelor's	m	sourcing	1	49	5	5	1	0	49	0	
15	38401	Technolog	region_29	Master's &	m	other	2	39	3	16	0	0	80	0	
16	77040	R&D	region_2	Master's &	m	sourcing	1	37	3	7	0	0	84	0	
17	43931	Operation:	region_7	Bachelor's	m	other	1	37	1	10	0	0	60	0	
18	7152	Technolog	region_2	Bachelor's	m	other	1	38	3	5	1	0	77	0	
19	9403	Sales & Ma	region_31	Bachelor's	m	other	1	34	1	4	0	0	51	0	
20	17436	Sales & Ma	region_31	Bachelor's	m	other	1	34	5	8	1	0	46	0	
21	54461	Operation:	region_15	Bachelor's	m	other	1	37	3	9	0	0	59	0	
22	12067	Procureme	region_14	Bachelor's	m	other	1	35	3	7	0	0	75	0	
23	33332	Operation:	region_15		m	sourcing	1	41	4	11	0	0	57	0	
24	58789	Finance	region_11	Bachelor's	f	other	1	28	3	4	0	0	63	0	
25	71177	Procureme	region_5	Bachelor's	m	other	1	27		1	0	0	70	0	
26	52057	Finance	region_22	Master's &	m	sourcing	2	39	5	7	0	0	59	0	
27	26585	Technolog	region_22	Bachelor's	m	other	1	27	5	3	1	0	83	0	

DATASET (FOR TESTING)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	employee_	departmer	region	education	gender	recruitmer	no_of_tra	age	previous_	length_of_	KPIs_met	awards_w	avg_training_score	
2	8724	Technolog	region_26	Bachelor's	m	sourcing	1	24		1	1	0	77	
3	74430	HR	region_4	Bachelor's	f	other	1	31	3	5	0	0	51	
4	72255	Sales & M	region_13	Bachelor's	m	other	1	31	1	4	0	0	47	
5	38562	Procureme	region_2	Bachelor's	f	other	3	31	2	9	0	0	65	
6	64486	Finance	region_29	Bachelor's	m	sourcing	1	30	4	7	0	0	61	
7	46232	Procureme	region_7	Bachelor's	m	sourcing	1	36	3	2	0	0	68	
8	54542	Finance	region_2	Bachelor's	m	other	1	33	5	3	1	0	57	
9	67269	Analytics	region_22	Bachelor's	m	sourcing	2	36	3	3	0	0	85	
10	66174	Technolog	region_7	Master's &	m	other	1	51	4	11	0	0	75	
11	76303	Technolog	region_22	Bachelor's	m	sourcing	1	29	5	2	1	0	76	
12	60245	Sales & M	region_16	Bachelor's	m	sourcing	2	40	5	12	1	0	50	
13	42639	Sales & M	region_17	Master's &	m	sourcing	1	40	3	10	0	0	46	
14	30963	Sales & M	region_4	Master's &	f	other	1	34	3	4	0	0	52	
15	54055	Analytics	region_24	Bachelor's	m	other	1	37	3	10	0	0	82	
16	42996	Operation	region_11	Bachelor's	m	sourcing	1	30	5	6	1	0	58	
17	12737	Sales & M	region_7	Bachelor's	m	sourcing	1	31	4	4	1	0	47	
18	27561	Operation	region_27	Bachelor's	f	sourcing	1	26	5	3	0	0	56	
19	26622	Sales & M	region_17	Bachelor's	m	sourcing	1	40	5	6	1	0	50	
20	31582	Procureme	region_7	Bachelor's	f	other	1	49	3	7	1	0	64	
21	29793	Procureme	region_27	Bachelor's	m	other	1	27	2	5	0	0	65	
22	72735	Sales & M	region_9	Master's &	m	sourcing	1	37	5	3	0	0	47	
23	5677	Technolog	region_17	Bachelor's	m	sourcing	1	25		1	0	0	80	
24	60889	Technolog	region_29	Master's &	m	sourcing	1	30	1	3	0	0	83	
25	51498	Procureme	region_4	Master's &	m	other	1	41	3	4	0	0	76	
26	8566	Finance	region_20	Bachelor's	f	other	1	29	4	6	1	0	58	
27	53151	Operation	region_20	Bachelor's	m	other	1	33	3	7	1	1	62	

Development Roadmap



LIBRARY INSTALLATION

DATA GATRHERING

EXPLORATORY DATA ANALYSIS 1) UNIVARIATE ANALYSIS 2) BIVARIATE ANALYSIS

DATA CLEANING 1) DEALING WITH MISSING VALUES 2)DEALING WITH NULL VALUES

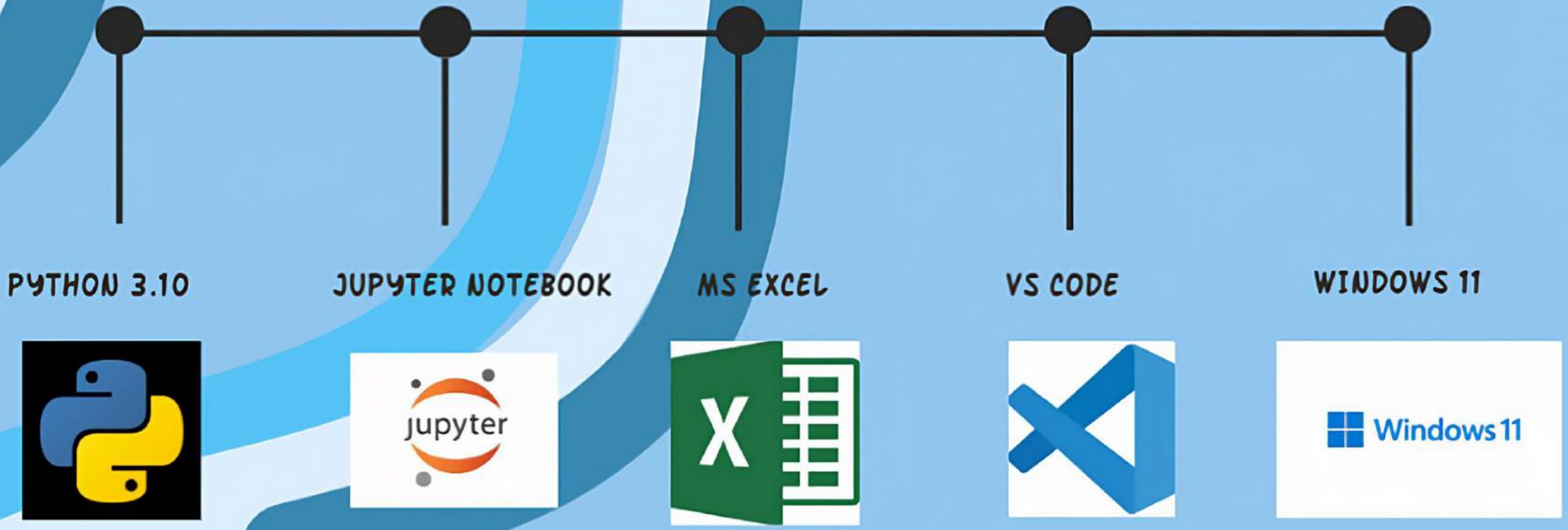
DATA PREPROCESSING 1)DROPPING UNWANTED FEATURES 2)ONE HOT ENCODING

DATA SCALLING (BETWEEN AGE AND AVERAGE TRAING SCORE)

CLASSIFICATION PROBLEM___MODEL BUILDING 1)EVALUATED USING F1 SCORE= (2*P*R)/(P+R)

KNN ALGORITHM
DECISION TREE
XGBOOST
RANDOM FOREST

Technologies Used



PYTHON 3.10

JUPYTER NOTEBOOK

MS EXCEL

VS CODE

WINDOWS 11

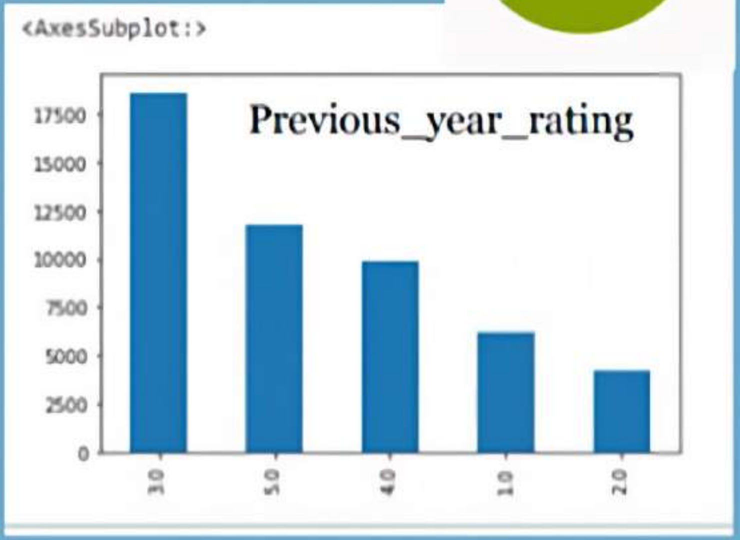
WORK-FLOW

Missing Value Treatment

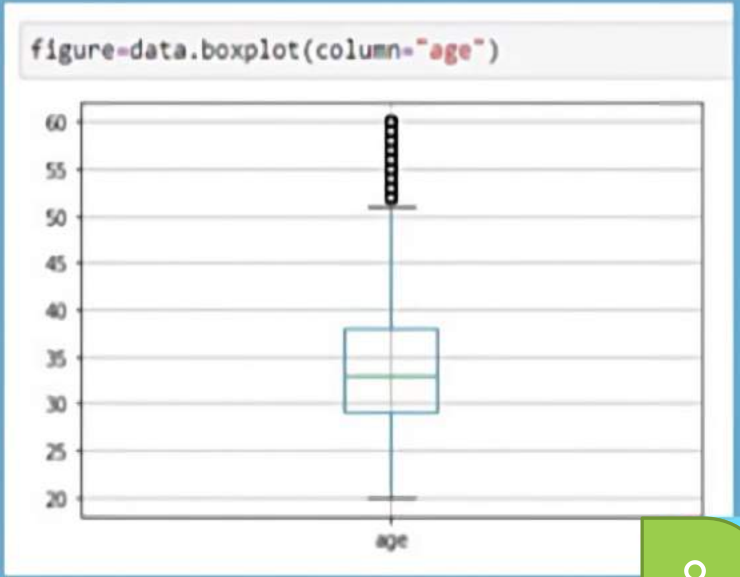
- Previous_year_rating has also missing values. By looking into distribution , mode value doesn't guarantee effectiveness of missing values.
- So Adding a separate col (after the mode value treatment)for weightage given to NAN values would be helpful.

Outliers Treatment

- Checking outliers in variables of the dataset related to Business scenarios.
- Age col found to be have outliers.
- Treating Outliers for Continuous var-As age is normally distributed we can eliminate or restrict the age distribution to **3 standard deviation (std)** or we can perform **IQR(if variable is skewed)**
- Few Nominal categorical variable have outliers like e.g. requriment_channel,Awards_won but may these outliers could be valuable info for the business.



```
<AxesSubplot:>
```

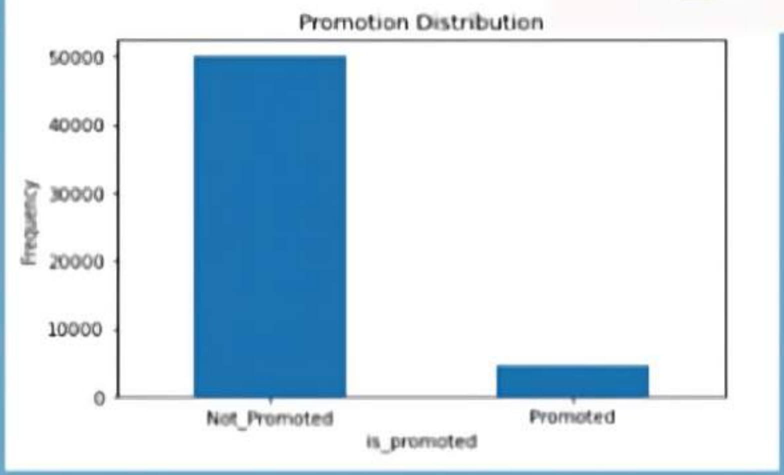


```
figure=data.boxplot(column="age")
```

WORK-FLOW

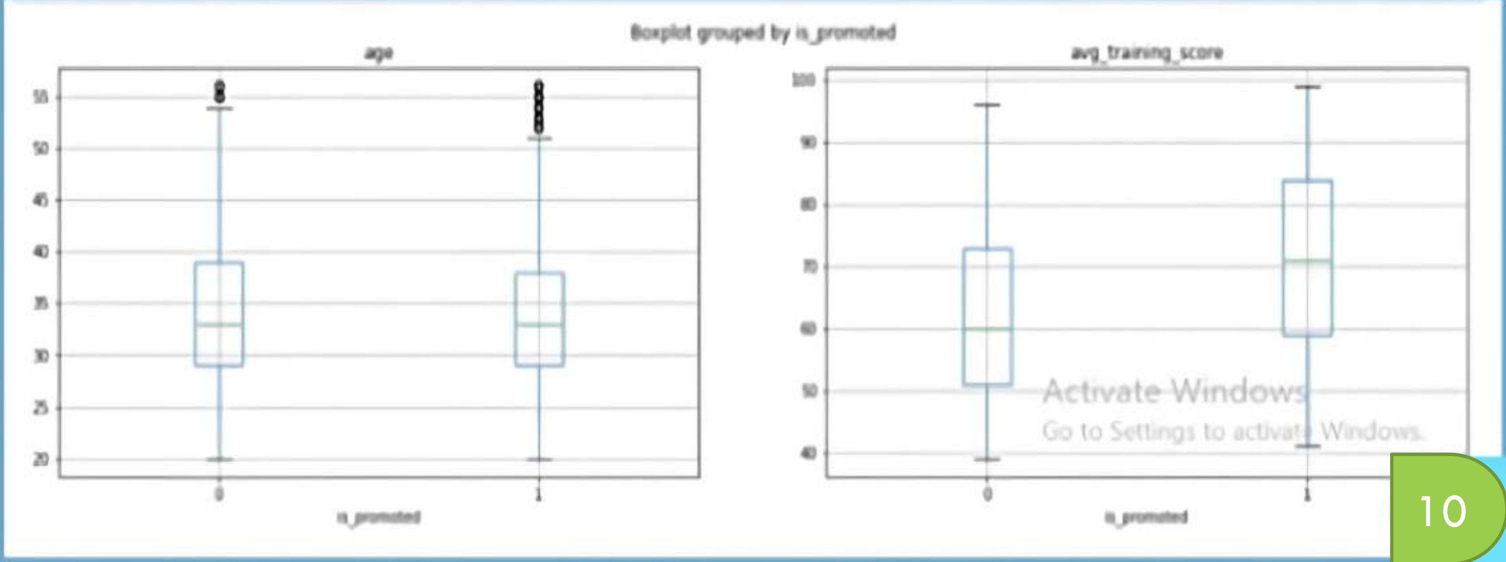
Exploratory Data Analysis

- Distribution of Target Variable is Highly Imbalanced ,So there is chance of ML model getting biased .
- In Order to avoid this we need to perform sampling algorithm.Here i have used Over sampling Method(Avoid using Under Sampling because risk of less Information)



Univariate and Bi Variate Analysis

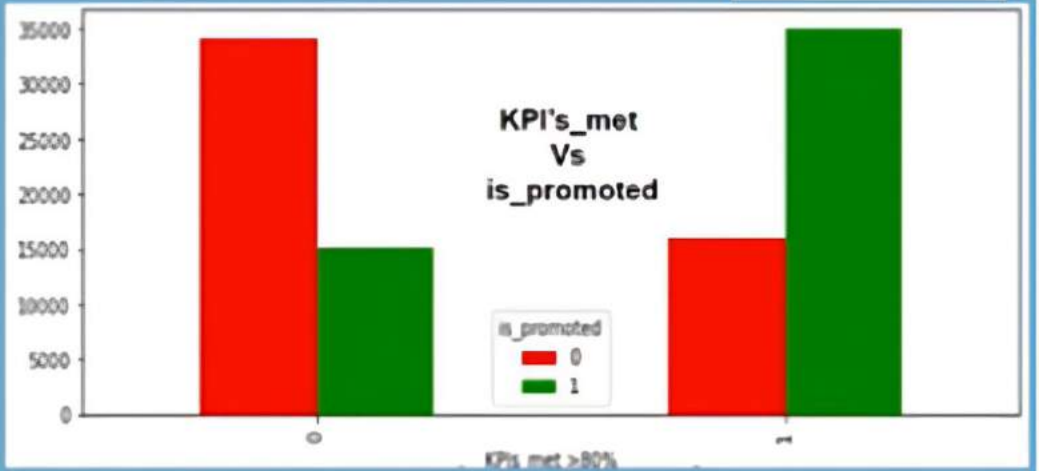
- We can see between Age and Avg_Training_Score Continuous variable Avg_Training_Score fluctuates more w.r.t promotion.
- So that we can say Avg_Training_Score variable is more influential to Target Var than Age.



WORK-FLOW

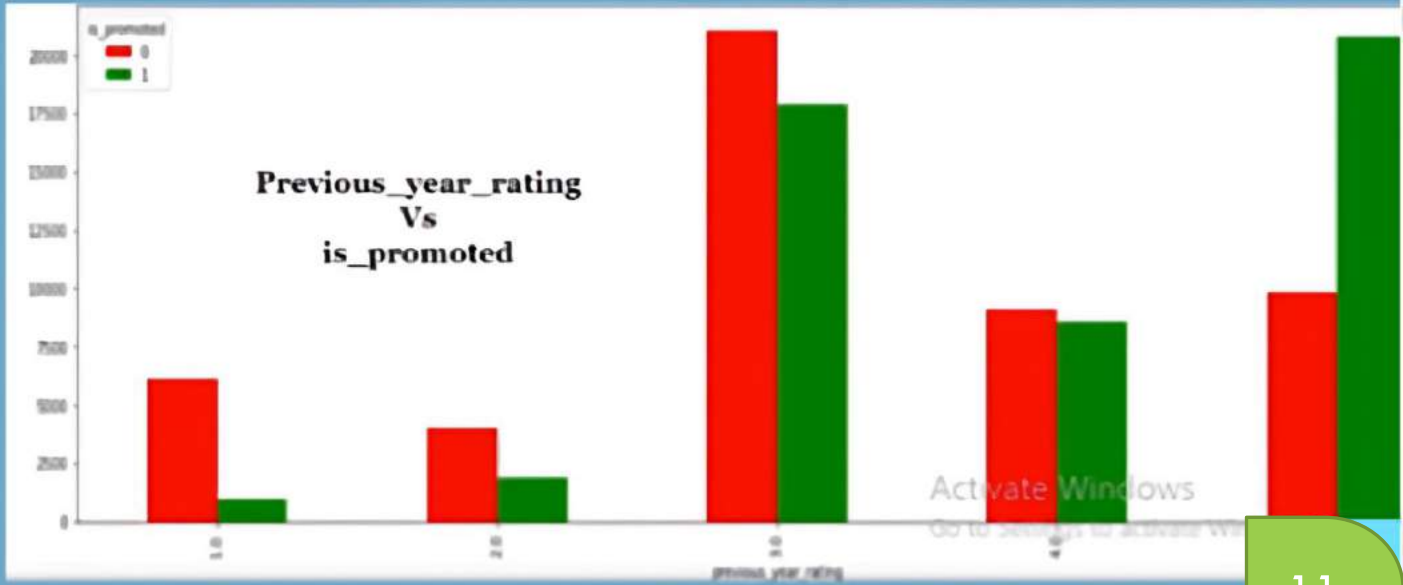
Univariate and Bi Variate Analysis

- From Categorical variable we can visualise those employees who has KPIs Metrics >80 has more likely to get promoted.



KPI's_met
Vs
is_promoted

- Those employees who had 3 and 5 as Previous_year_rating were comparatively got promoted most.
- But we can see those who got 4 and 5 chances of getting promotion is high as the ratio of not getting promotion is low as compared to 1, 2 & 3 .

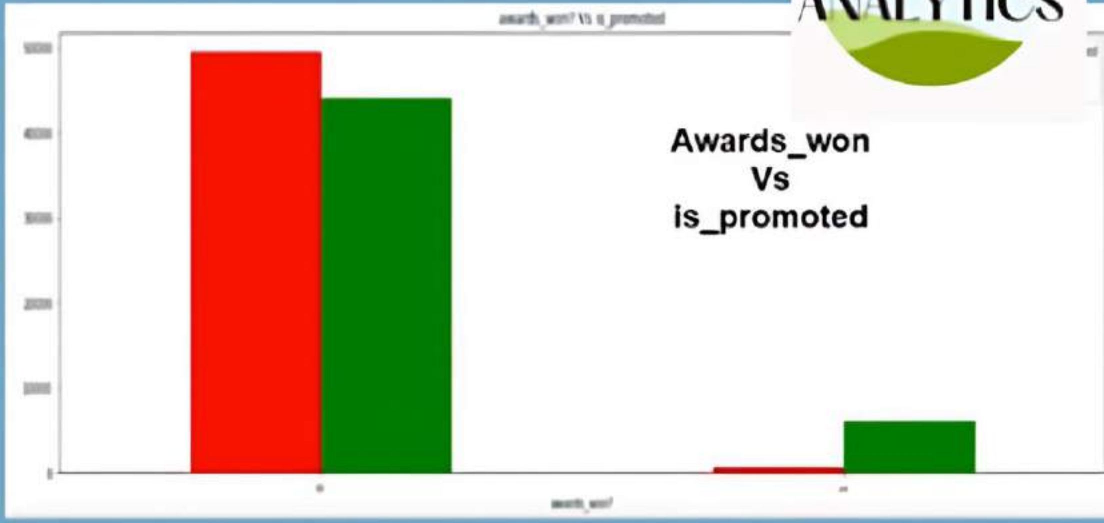


Previous_year_rating
Vs
is_promoted

WORK-FLOW

Univariate and Bi Variate Analysis

- The employees who didn't get Awards the chances of of not getting promoted is high as compared to employees who got Awards.



Measurement of Strength of Bi-variate Analysis

- Hypothesis Testing
    - ANOVA TESTING
    - CHI-Sqr
- ANOVA-Done between continuous and Target Variable.
- Chi_Sqr-Done between Categorical and Target Variable.
- All Categorical & Continuous in our DataSet have an impact as P value < 0.05 for all cols i.e. not enough evidence to accept the Null Hypothesis.

Hypothesis

Null (Ho) :There is not relationship between variable.

Alternative(Ha):There is some relationship between variables.

WORK-FLOW

HR ANALYTICS

NOTE

★ drop_first=TRU

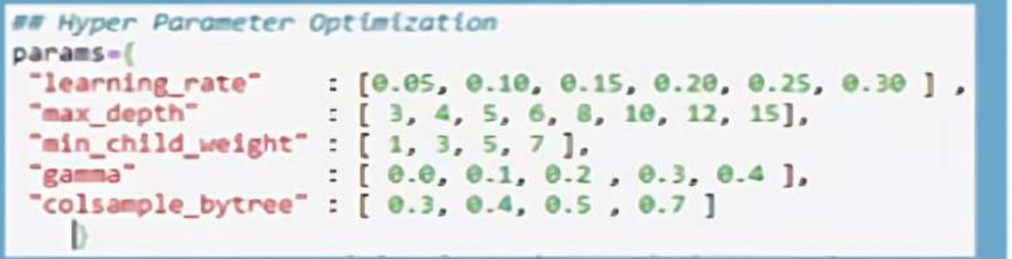
✖ Dummy Variable Trap ✖

Encoding Techniques

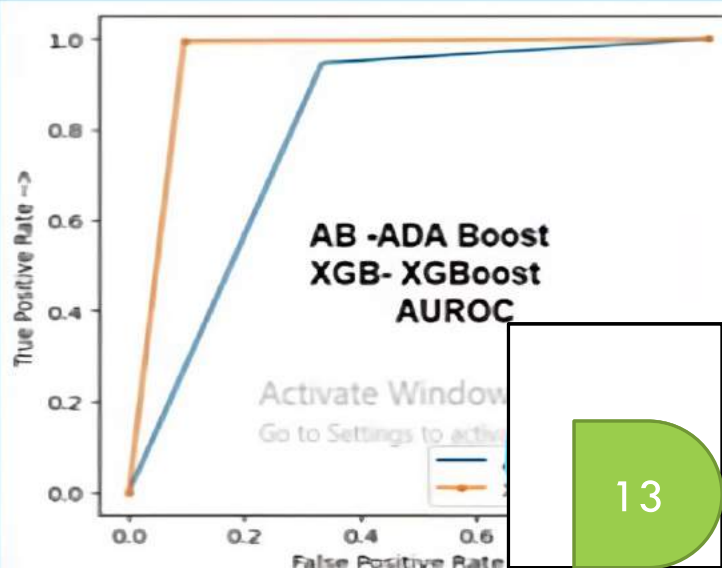
- Department : Mean Encoding & assigning the Rank
- Education : Ordinal Label Encoding
- Region : Mapping each no of repetitions to respective region.
- Recruiment_channel & Gender : One Hot Encoding (While one hot encoding always drop first column to avoid dimensional complexity)

Machine Learning Model Building

- Developing ML Model using Several Algorithm i.e. Logistic Regression,DecisionTreeClassifier,Random Forest,AdaBoost & XGBoost.
- Choosing Hyperparameter tuned XGBoost model which giving best Accuracy and ROC Value.
  - ***F1 Score*** : 94%
  - ***ROC-AUC*** : 95%



```
## Hyper Parameter Optimization
params={
 "learning_rate"    : [0.05, 0.10, 0.15, 0.20, 0.25, 0.30 ] ,
 "max_depth"        : [ 3, 4, 5, 6, 8, 10, 12, 15],
 "min_child_weight" : [ 1, 3, 5, 7 ],
 "gamma"            : [ 0.0, 0.1, 0.2 , 0.3, 0.4 ],
 "colsample_bytree" : [ 0.3, 0.4, 0.5 , 0.7 ]
    }
```



THANK YOU