



PYTHON PROJECT - HR ANALYTICS



By Group Number 5:

Aishwarya Vishwanath
Pooja Raman
Anirudh Rao
Tawi Mankotia
Vishal Doshi

Introduction

- **HR-Force** is end-to-end talent acquisition solutions organization
- Experience in technology consulting and talent acquisition areas
- Actively practicing in the existing ultra-demanding business environment



Selection Stage:

Process involves multiple rounds of interview for the candidates whose resumes have been screened and cleared.

Offer Letter Stage:

The offer roll-out or rejection is guided by this score. Offer letter is released and accepted by the candidate.

Onboarding Stage:

Reneging is a major problem. This occurs when candidates who have been extended a job offer do not take up the company on their offer.

Dataset Description

Source : Harvard Business Review



Variable Name	Description	Type
DOJ extended	Binary variable identifying whether candidate asked for date of joining extension (Yes/No)	Input
Duration to accept the offer	Number of days taken by the candidate to accept the offer (continuous variable)	Input
Notice period	Notice period to be served in the parting company before candidate can join this company (continuous variable)	Input
Offered band	Band offered to the candidate based on experience and performance in interview rounds (categorical variable labelled E0 to E6)	Input
Percentage hike (CTC) expected	Percentage hike expected by the candidate (continuous variable)	Input
Percentage hike offered (CTC)	Percentage hike offered by the company (continuous variable)	Input
Joining bonus	Binary variable indicating if joining bonus was given or not (Yes/No)	Input
Gender	Gender of the candidate (Male/Female)	Input
Candidate source	Source from which resume of the candidate was obtained (categorical variables with categories: Employee referral/Agency/Direct)	Input
REX (in years)	Relevant years of experience of the candidate for the position offered (continuous variable)	Input
LOB	Line of business for which offer was rolled out (categorical variable)	Input
DOB	Date of birth of the candidate	Input
Joining location	Company location for which offer was rolled out for candidate to join (categorical variable)	Input
Candidate relocation status	Binary variable indicating whether candidate has to relocate from one city to another city for joining (Yes/No)	Input
HR status	Final joining status of candidate (Joined/Not-Joined)	Output

Research Question

- Identify the key drivers that influence the candidate to join or not join the company.
- Devise a predictive algorithm to calculate the probability of acceptance of an offer and joining the company after offer acceptance stage.



Identify
Strategic HR
Concern



Research
Design



Data
Management

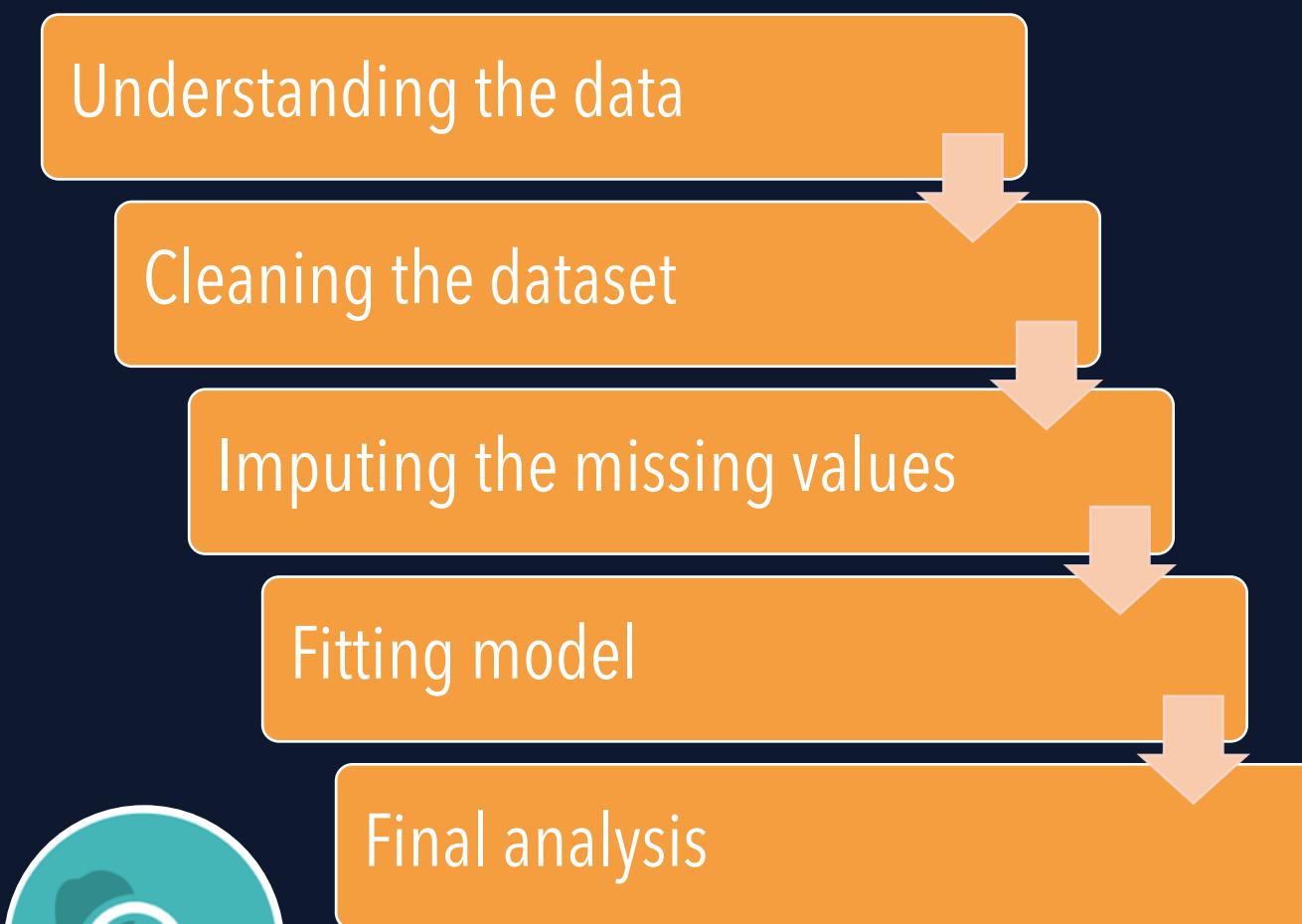


Data
Analysis



Data
Interpretation
and
Communication

Methodology



D Descriptive Analysis

1

```
In [5]: df_hr.shape #Number of Rows and Columns in the Dataset  
Out[5]: (12333, 17)
```

2

```
In [4]: percent_missing = df_hr.isnull().sum() * 100 / len(df_hr)  
percent_missing #Check for Missing Values  
  
Out[4]: Candidate Ref          0.000000  
DOJ Extended          0.000000  
Duration to accept offer    22.046542  
Notice period          0.000000  
Offered band          0.000000  
Percent hike expected in CTC  6.056920  
Percent hike offered in CTC  4.832563  
Percent difference CTC      6.900186  
Joining Bonus          0.000000  
Candidate relocate actual  0.000000  
Gender                  0.000000  
Candidate Source          0.000000  
Rex in Yrs              0.000000  
LOB                     0.000000  
Location                0.000000  
Age                     0.000000  
Status                  0.000000  
dtype: float64
```

D Descriptive Analysis

3

```
In [6]: df_hr.describe() #Mean, Median, Standard Deviation in Numeric Columns of the dataset
```

Out[6]:

	Candidate Ref	Duration to accept offer	Notice period	Percent hike expected in CTC	Percent hike offered in CTC	Percent difference CTC	Rex in Yrs	Age
count	1.233300e+04	9614.000000	12333.000000	11586.000000	11737.000000	11482.000000	12333.000000	12333.000000
mean	2.872888e+06	21.189619	37.943323	41.837663	37.772016	-2.078916	4.160788	30.232304
std	5.099726e+05	26.410351	24.526587	29.435844	35.593317	19.997159	2.810108	4.178853
min	2.109586e+06	-228.000000	0.000000	-68.830000	-60.530000	-68.950000	0.000000	19.000000
25%	2.378256e+06	2.000000	30.000000	25.000000	20.000000	-9.090000	3.000000	27.000000
50%	2.820675e+06	9.000000	30.000000	38.890000	34.480000	0.000000	4.000000	30.000000
75%	3.338197e+06	32.000000	60.000000	52.170000	47.060000	0.000000	6.000000	34.000000
max	3.836076e+06	224.000000	120.000000	359.770000	471.430000	300.000000	26.000000	64.000000

```
In [11]: df_hr.mode().head(1) #Mode of the Dataset
```

Out[11]:

	Candidate Ref	DOJ Extended	Duration to accept offer	Notice period	Offered band	Percent hike expected in CTC	Percent hike offered in CTC	Percent difference CTC	Joining Bonus	Candidate relocate actual	Gender	Candidate Source	Rex in Yrs	LOB	Location	Age	Status
0	2109586	No	1.0	30.0	E1	42.86	42.86	0.0	No	No	Male	Direct	3.0	INFRA	Chennai	34.0	Joined

D Descriptive Analysis

4

In [9]: `df_hr.cov() #Covariance in the Dataset`

Out[9]:

	Candidate Ref	Duration to accept offer	Notice period	Percent hike expected in CTC	Percent hike offered in CTC	Percent difference CTC	Rex in Yrs	Age
Candidate Ref	2.600721e+11	-344551.733268	-2.772006e+06	157253.220607	65993.932037	86864.832110	-179529.871010	758151.026446
Duration to accept offer	-3.445517e+05	697.506639	1.968370e+02	14.112901	8.478346	-3.982166	6.489460	1.126334
Notice period	-2.772006e+06	196.837029	6.015535e+02	33.450447	-4.279058	-29.356333	18.641615	-7.311679
Percent hike expected in CTC	1.572532e+05	14.112901	3.345045e+01	866.468923	684.261537	-105.130121	-13.502857	-20.422610
Percent hike offered in CTC	6.599393e+04	8.478346	-4.279058e+00	684.261537	1266.884223	429.960892	-10.176624	-14.684310
Percent difference CTC	8.686483e+04	-3.982166	-2.935633e+01	-105.130121	429.960892	399.886379	1.483615	3.675312
Rex in Yrs	-1.795299e+05	6.489460	1.864161e+01	-13.502857	-10.176624	1.483615	7.896704	5.655801
Age	7.581510e+05	1.126334	-7.311679e+00	-20.422610	-14.684310	3.675312	5.655801	17.462816

D Descriptive Analysis

5

```
In [8]: df_hr.corr() #Correlations in the Dataset
```

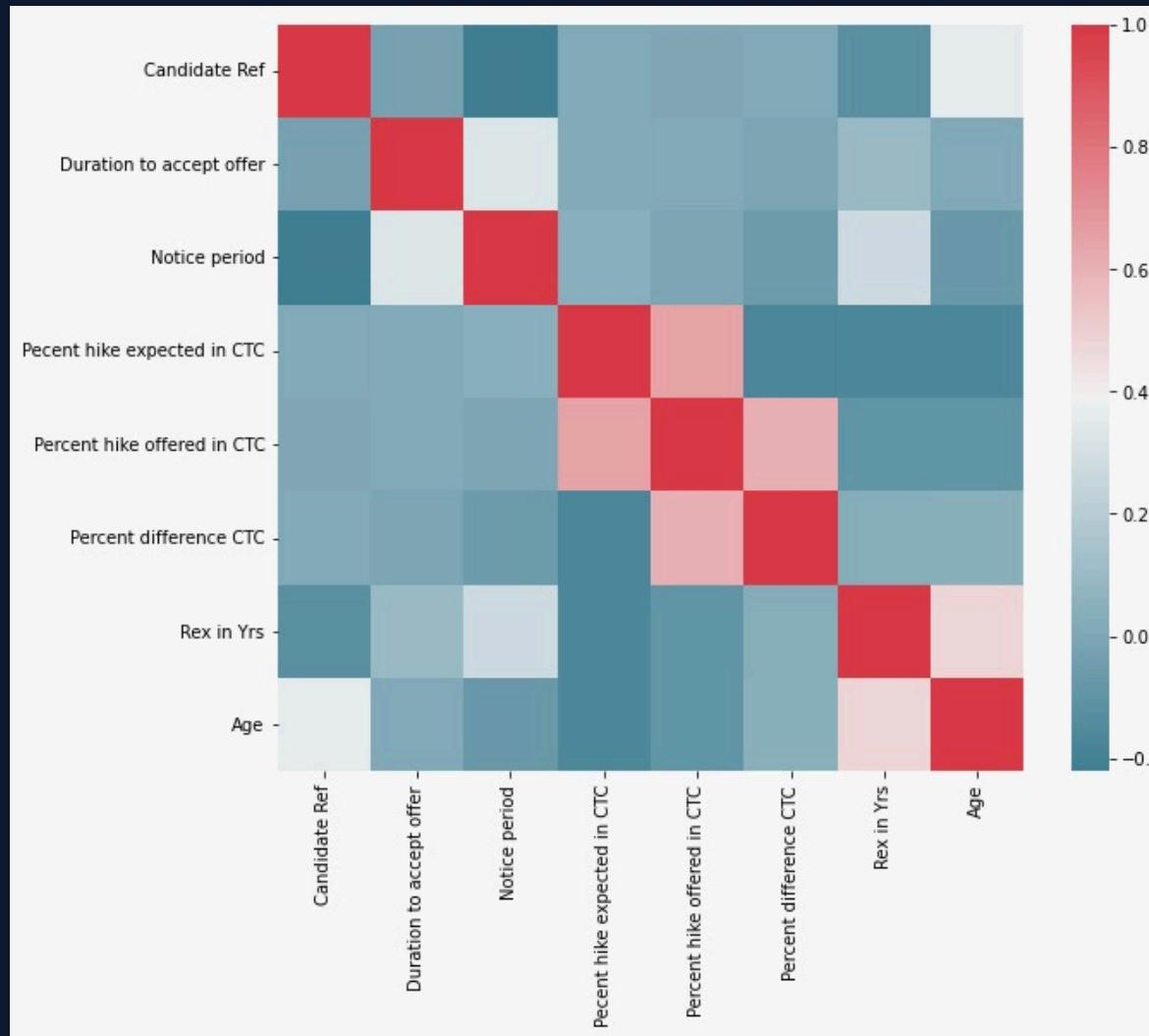
Out[8]:

	Candidate Ref	Duration to accept offer	Notice period	Percent hike expected in CTC	Percent hike offered in CTC	Percent difference CTC	Rex in Yrs	Age
Candidate Ref	1.000000	-0.026692	-0.221621	0.010554	0.003662	0.008575	-0.125276	0.355756
Duration to accept offer	-0.026692	1.000000	0.326540	0.018304	0.009044	-0.007847	0.095212	0.010302
Notice period	-0.221621	0.326540	1.000000	0.047160	-0.004985	-0.061238	0.270473	-0.071338
Percent hike expected in CTC	0.010554	0.018304	0.047160	1.000000	0.651497	-0.177674	-0.165577	-0.168325
Percent hike offered in CTC	0.003662	0.009044	-0.004985	0.651497	1.000000	0.603027	-0.102241	-0.099320
Percent difference CTC	0.008575	-0.007847	-0.061238	-0.177674	0.603027	1.000000	0.026684	0.044472
Rex in Yrs	-0.125276	0.095212	0.270473	-0.165577	-0.102241	0.026684	1.000000	0.481631
Age	0.355756	0.010302	-0.071338	-0.168325	-0.099320	0.044472	0.481631	1.000000

D Descriptive Analysis

6

Correlation Matrix

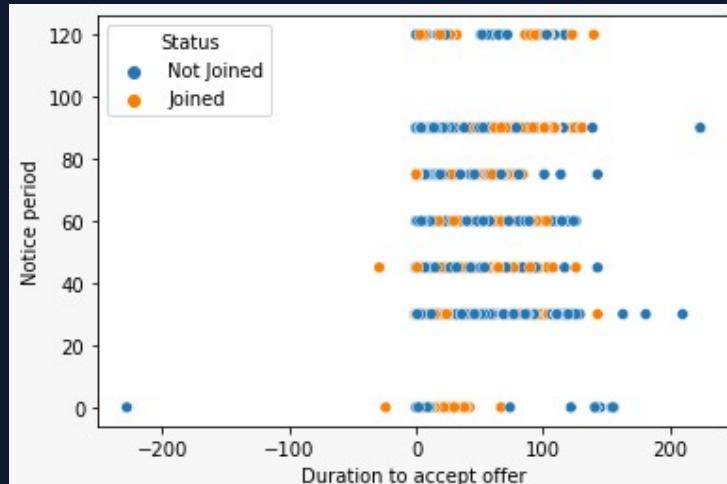


D

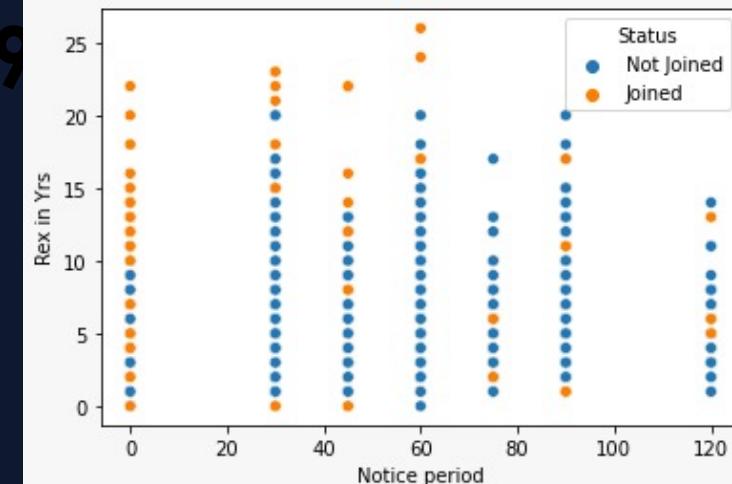
Descriptive Analysis

7

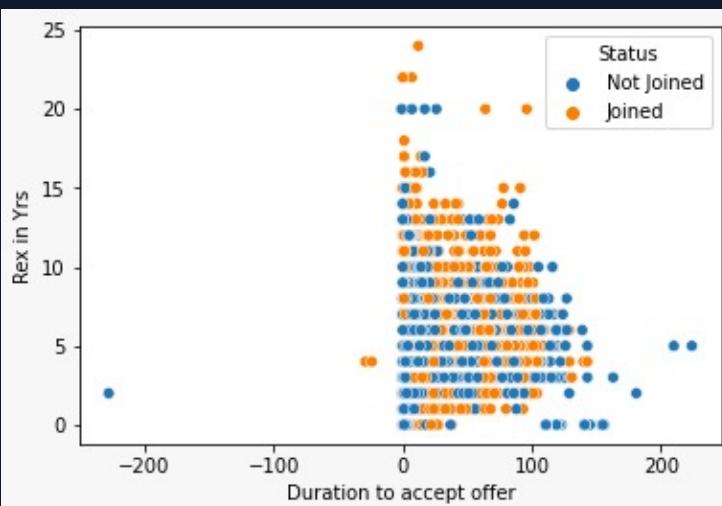
Scatter Plot - Notice Period + Duration to accept v/s Joining Status



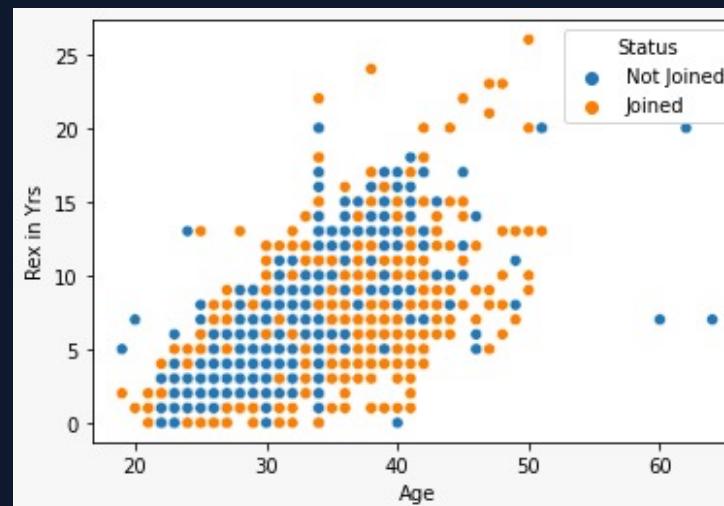
Scatter Plot - Relevant Exp + Notice Period v/s Joining Status



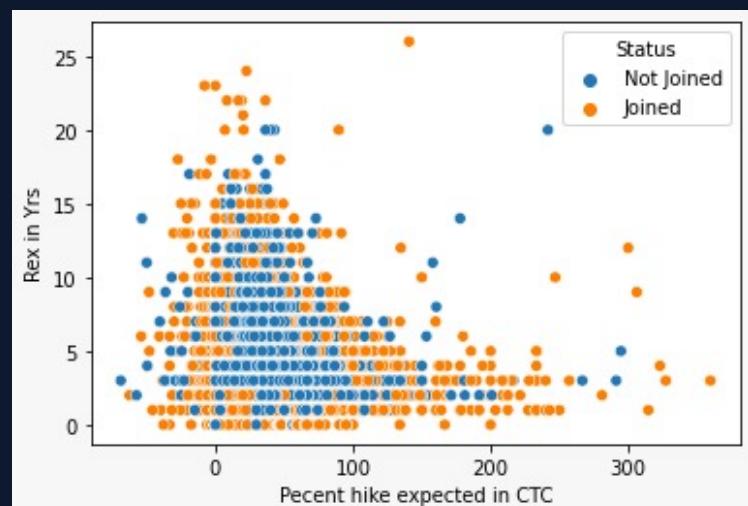
Scatter Plot - Relevant Exp + Duration to accept v/s Joining Status



Scatter Plot - Relevant Experience + Age v/s Joining Status



Scatter Plot – Relevant Experience + Percent hike v/s Joining Status

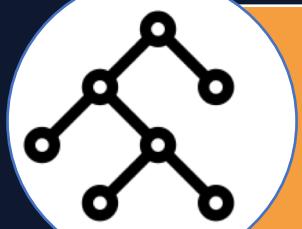


Fitting Models



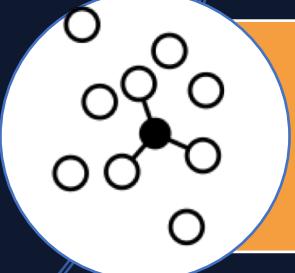
Logistic Regression:

```
random_state=0, solver='liblinear', multi_class='auto'
```



Random Forest:

```
n_estimators=10000, max_depth=5,  
random_state=0
```



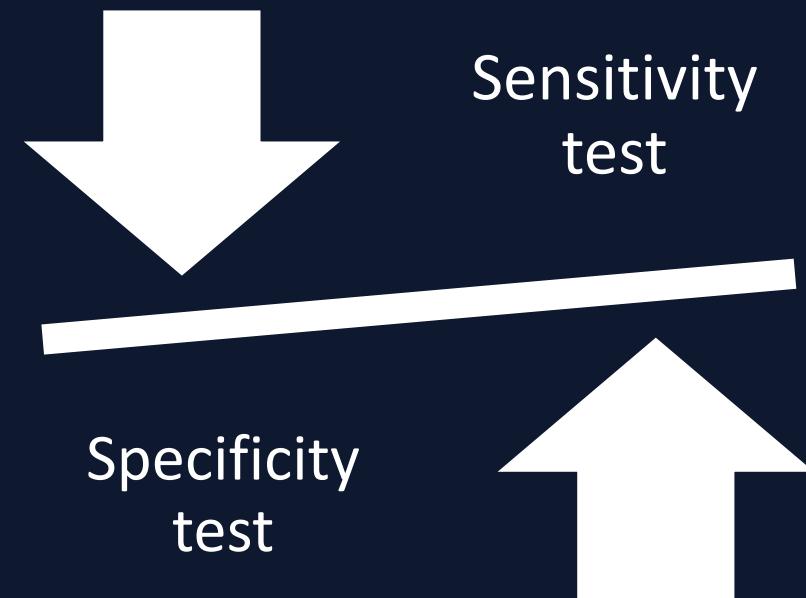
K-NN:

```
n_neighbors = 3
```

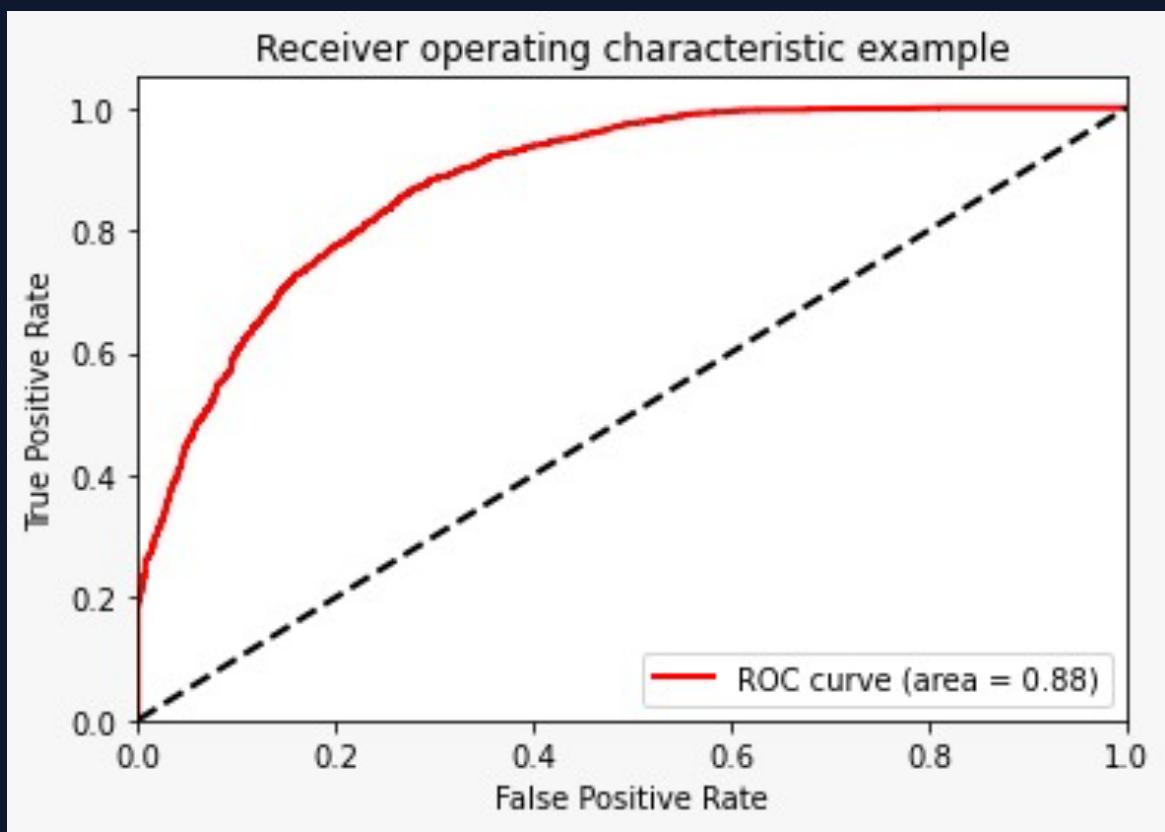
- Logistic Regression : With solver as 'liblinear' and multi_class as 'auto', we get a perfect accuracy, sensitivity and specificity.
- Random Forest: For n_estimators from 100 to 20000, the sensitivity is higher, and the specificity is lower.
- K-NN : For n_neighbor = 3, the sensitivity is higher, and the specificity is lower.

After multiple trials, we conclude that Logistic regression model describes the best relationship the status and other variables.

What we are looking for :



Logistic Regression - ROC Curve

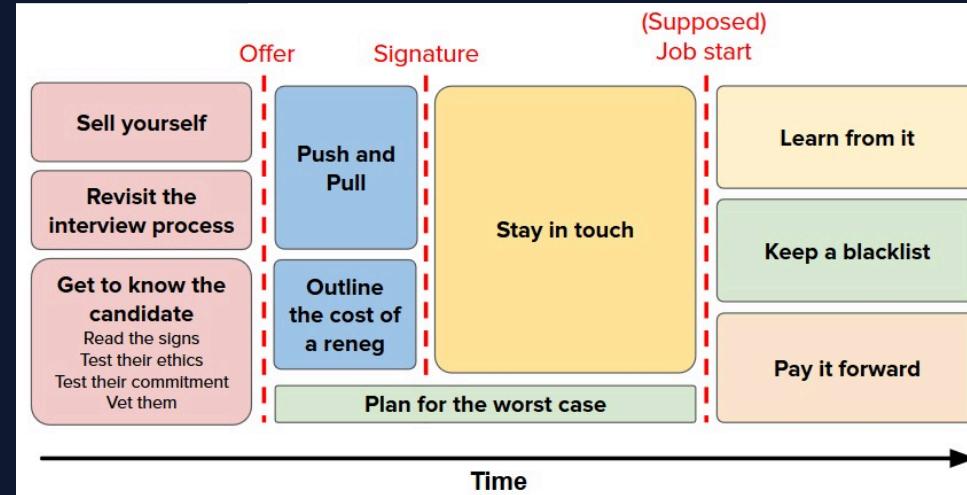


	accuracy	sensitivity	specificity
Logistic Regression	0.788342	0.778256	0.798239
Random Forest	0.777501	0.806243	0.749296
K-NN	0.765061	0.844995	0.686620

Regression Summary

	coef	std err	z	P> z	[0.025	0.975]
const	-5.3801	0.223	-24.153	0.000	-5.817	-4.944
LOB_INFRA	2.8060	0.086	32.548	0.000	2.637	2.975
LOB_BSERV	4.2296	0.158	26.729	0.000	3.919	4.540
Candidate relocate actual_Yes	26.4871	1.33e+04	0.002	0.998	-2.61e+04	2.61e+04
Percent hike offered in CTC	0.0123	0.003	4.787	0.000	0.007	0.017
Offered band_E1	2.2226	0.102	21.773	0.000	2.023	2.423
LOB_SALES	2.9076	0.368	7.910	0.000	2.187	3.628
Offered band_E2	2.0313	0.116	17.586	0.000	1.805	2.258
Candidate Source_Employee Referral	1.2508	0.073	17.231	0.000	1.109	1.393
Duration to accept offer	0.0032	0.001	3.552	0.000	0.001	0.005
Offered band_E6	-13.7821	1639.939	-0.008	0.993	-3228.003	3200.438
Age	0.0410	0.007	5.978	0.000	0.028	0.054
Candidate Source_Direct	0.6870	0.047	14.534	0.000	0.594	0.780
LOB_EAS	1.2816	0.134	9.582	0.000	1.019	1.544
LOB_ETS	1.8859	0.102	18.568	0.000	1.687	2.085
Offered band_E3	2.0724	0.169	12.254	0.000	1.741	2.404
LOB_Healthcare	2.0450	0.184	11.133	0.000	1.685	2.405
Offered band_E5	1.8692	0.892	2.095	0.036	0.120	3.618
Percent difference CTC	-0.0097	0.004	-2.719	0.007	-0.017	-0.003
Offered band_E4	2.7172	0.418	6.497	0.000	1.897	3.537
Pecent hike expected in CTC	-0.0078	0.002	-3.172	0.002	-0.013	-0.003
Rex in Yrs	-0.0378	0.015	-2.452	0.014	-0.068	-0.008
LOB_ERS	1.9060	0.077	24.596	0.000	1.754	2.058
Notice period	-0.0218	0.001	-20.515	0.000	-0.024	-0.020
DOJ_Extended_Yes	1.1379	0.049	23.439	0.000	1.043	1.233
LOB_CORP	3.0693	0.212	14.447	0.000	2.653	3.486
LOB_MMS	3.5508	0.663	5.354	0.000	2.251	4.851
Gender_Male	0.0955	0.051	1.855	0.064	-0.005	0.196
LOB_CSMP	2.3351	0.112	20.823	0.000	2.115	2.555
LOB_BFSI	2.2298	0.084	26.675	0.000	2.066	2.394
Joining Bonus_Yes	0.2176	0.124	1.760	0.078	-0.025	0.460

F Final Recommendations



Most significant factors that influence candidate's decision to **NOT** join

Percent difference in Salary



- Set realistic expectation with the candidate
- Highlight the other parts of the compensation – benefits, variable comp, bonus, allowance etc.

Relevant Experience



- Understand the candidate and their professional background – industry, reasons for leaving etc.
- Candidates with less experience are more likely to renege. To combat this challenge recruiter must sell the future – scope for career progression and personal development.

Notice Period



- Offer to buy "buy back" from current employer to reduce notice period
- Back up candidates for candidates who have 30 – 120 days notice period

