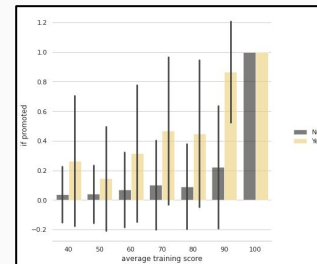
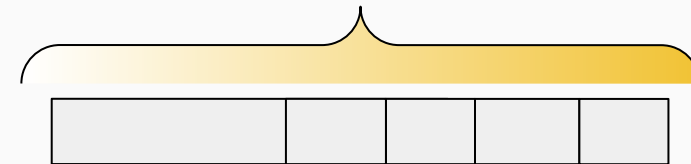


# HR PREDICTION ANALYSIS FOR PROMOTIONS

Group 9



	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	award
0	8	7	3	2	2	1	35	5.0	8	
1	5	22	1	1	1	1	30	5.0	4	
2	8	19	1	1	2	1	34	3.0	7	
3	8	23	1	1	1	2	39	1.0	10	
4	9	26	1	1	1	1	45	3.0	2	
...	...	...	...	...	...	...	...	...	...	...
54802	8	14	1	1	1	2	31	1.0	2	
54803	9	14	1	1	2	1	48	3.0	17	
54804	5	27	3	2	1	1	37	2.0	6	
54805	1	1	1	1	1	1	27	5.0	3	
54807	3	22	1	1	1	1	27	1.0	5	

48660 rows x 12 columns

**THE BOTTOM LINE**

**PROMOTION CYCLE**

Pro

**SOLUTION:  
PREDICTION ANALYSIS**



# AGENDA

**01**

**DATA**

**02**

**FEATURE  
SELECTION**

**03**

**MODELS**

**04**

**APPLICATION**

# 01. DATA



employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service
0	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	5.0	8
1	Operations	region_22	Bachelor's	m	other	1	30	5.0	4
2	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	3.0	7
3	Sales & Marketing	region_23	Bachelor's	m	other	2	39	1.0	10
4	Technology	region_26	Bachelor's	m	other	1	45	3.0	2
...	...	...	...	...	...	...	...	...	...
54803	Technology	region_14	Bachelor's	m	sourcing	1	48	3.0	17
54804	Operations	region_27	Master's & above	f	other	1	37	2.0	6
54805	Analytics	region_1	Bachelor's	m	other	1	27	5.0	3
54806	Sales & Marketing	region_9	NaN	m	sourcing	1	29	1.0	2
54807	HR	region_22	Bachelor's	m	other	1	27	1.0	5

54808 rows x 13 columns

department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	award
0	8	3	2	2	1	35	5.0	8	
1	5	1	1	1	1	30	5.0	4	
2	8	1	1	2	1	34	3.0	7	
3	8	1	1	1	2	39	1.0	10	
4	9	1	1	1	1	45	3.0	2	
...	...	...	...	...	...	...	...	...	...
54802	14	1	1	1	2	31	1.0	2	
54803	14	1	1	2	1	48	3.0	17	
54804	5	3	2	1	1	37	2.0	6	
54805	1	1	1	1	1	27	5.0	3	
54807	3	22	1	1	1	27	1.0	5	

48660 rows x 12 columns

## CLEANING DATA

### STEP 01

Categorical to Numerical

### STEP 02

Removed Employee ID's

### STEP 03

Removed NaN's and blanks

NOT PROMOTED

PROMOTED

## HANDLING IMBALANCED DATA - SMOTE

44428

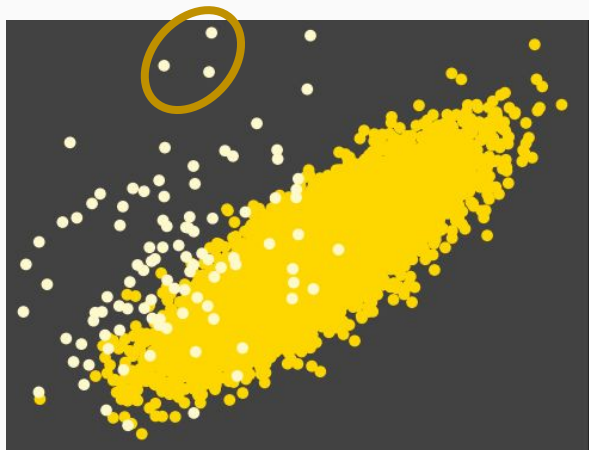
4232

BEFORE SMOTE

44428

44428

AFTER SMOTE



## 02. FEATURE SELECTION



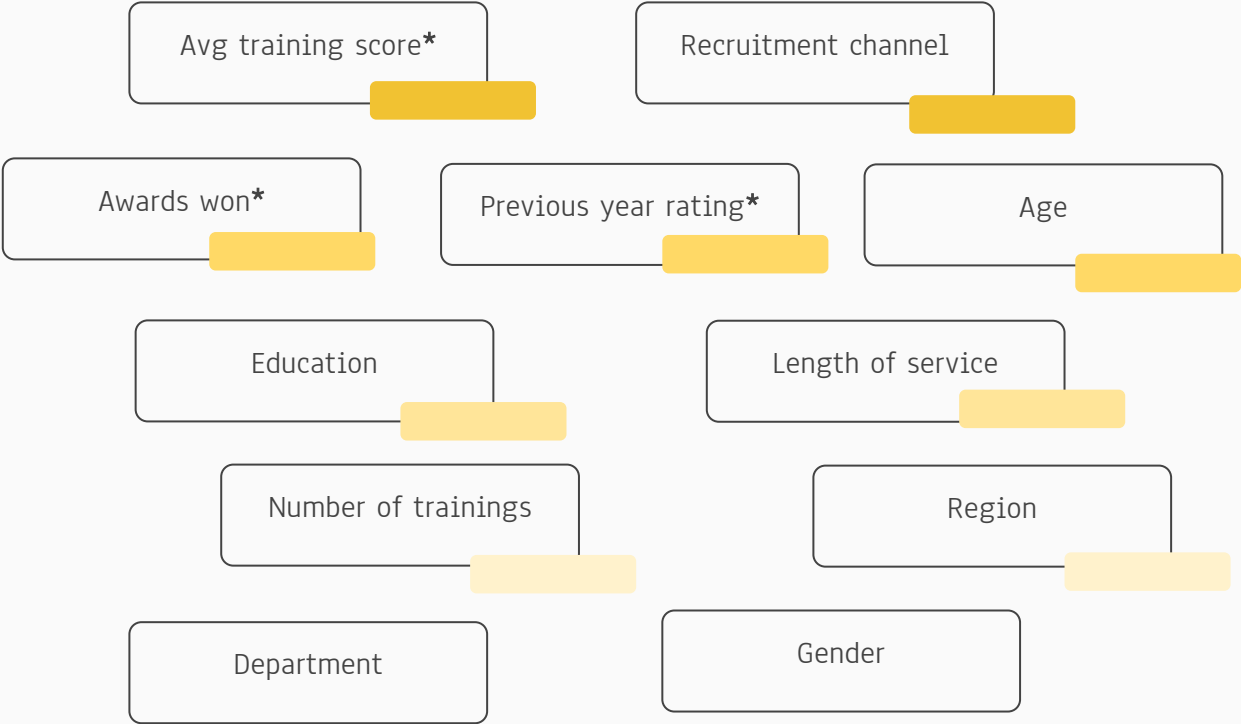
(\*) indicates top 3 chi-squared statistics

# FEATURE INFLUENCE

HIGH

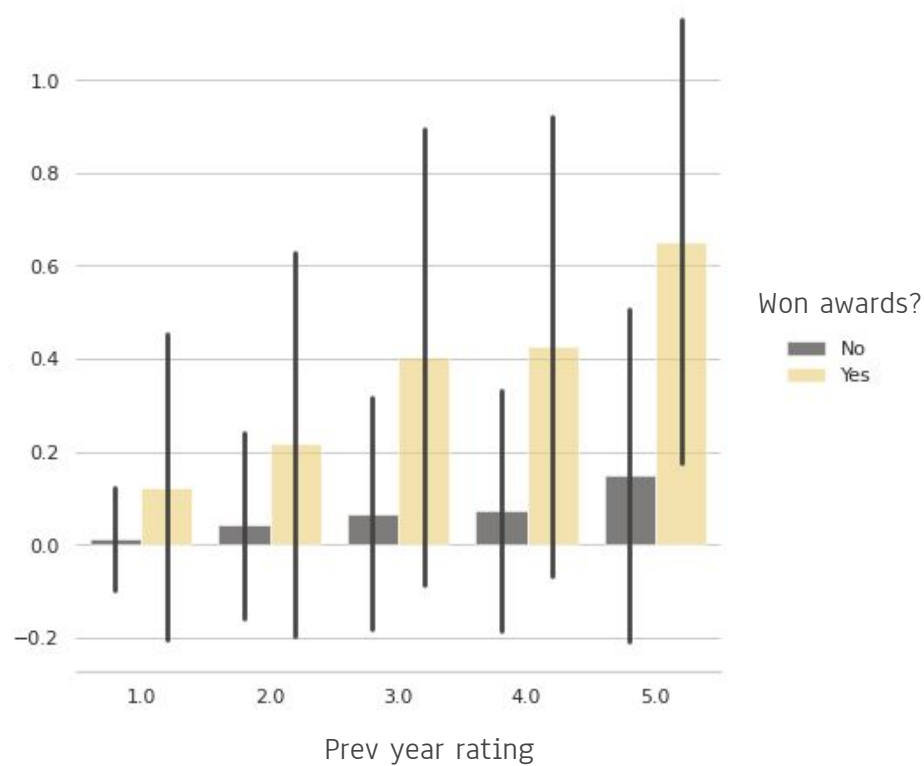
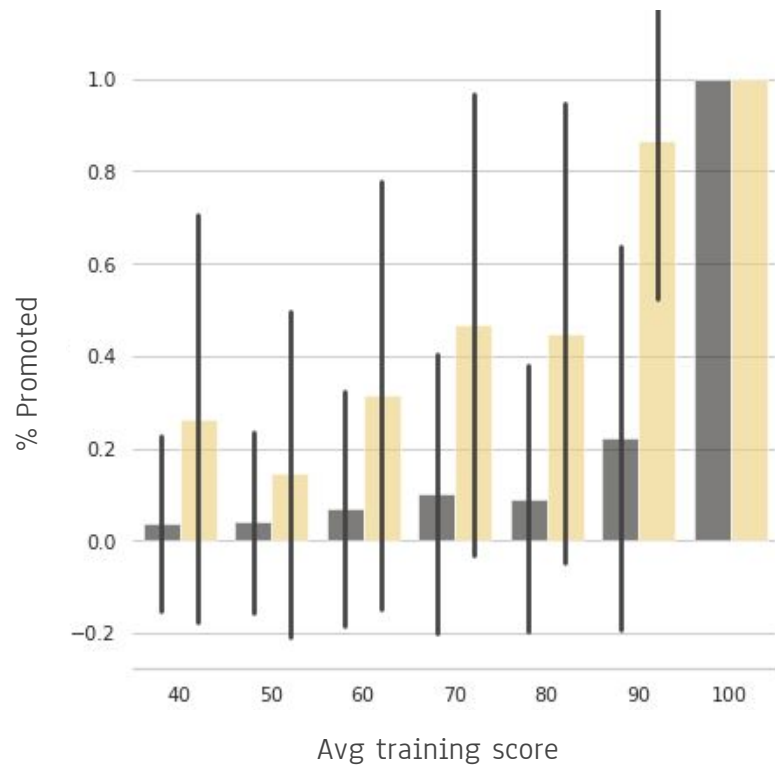


LOW

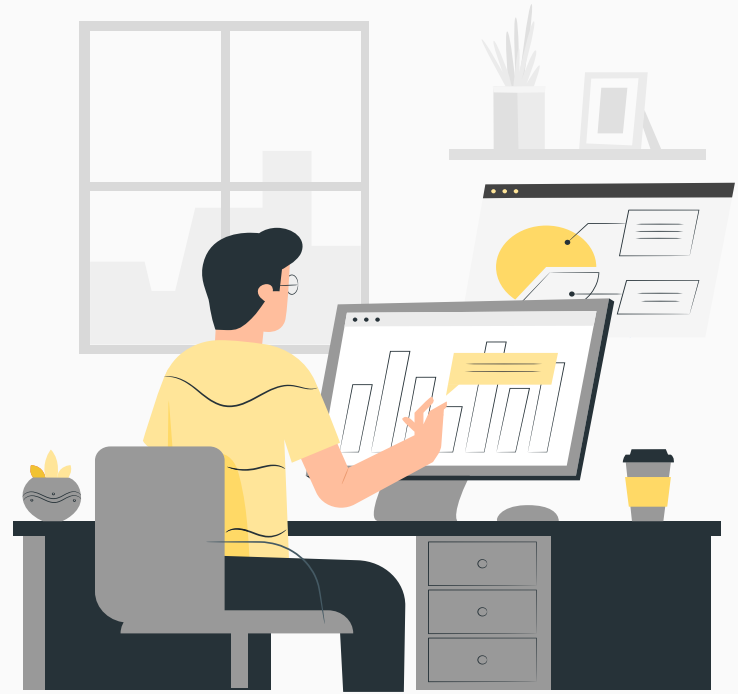




## FEATURE VISUALIZATION



## 03. MODELS



				MODEL COMPARISON	
INTERPRET- ABILITY	COMPUTE SPEED	RECALL	PRECISION		
High	Fast	68.5%	70.4%	—	LOGISTIC REGRESSION
Avg	Fast	68.3%	72.4%	—	LINEAR SVM
Low	Slow	72.8%	75.7%	—	POLYNOMIAL SVM
Low	Slow	75.8%	80.3%	—	RBF SVM
High	Fast	65.8%	77.4%	—	NAIVE BAYES
Low	Slow	78.9%	60.7%	—	MULTI-LAYER PERCEPTRON

- Precision mitigates false positives (desired)
- Features removed: Department and Gender
- Used Cross Validation

## 04. APPLICATION



(\*) indicates top 3 chi-squared statistics

# PERSON A



Gender	Female
Age	36
Department	R&D
Education	MS & Above
Recruitment	Sourcing
Length of service	9
Awards won *	0
Number of trainings	1
Avg training score *	80
Prev year rating *	5.0/5.0
Prediction (Yes)	76.8%

# PERSON B

Gender	Male
Age	60
Department	Technology
Education	MS & Above
Recruitment	Sourcing
Length of service	9
Awards won *	0
Number of trainings	1
Avg training score *	61
Prev year rating *	5.0/5.0
Prediction (Yes)	19.4%



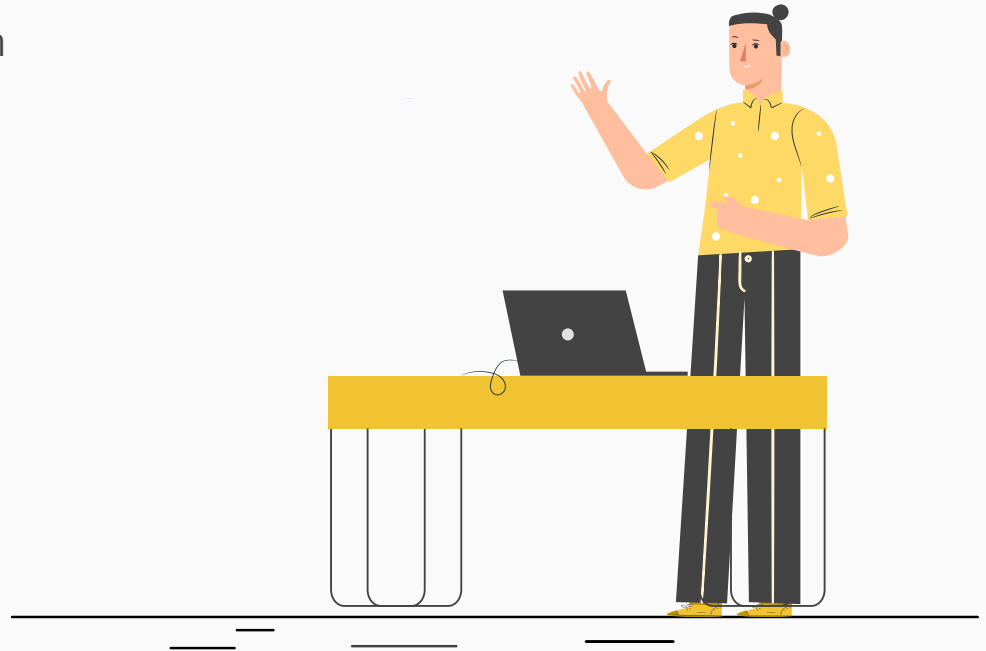
# PERSON C

Gender	Male
Age	34
Department	Analytics
Education	BS
Recruitment	Sourcing
Length of service	7
Awards won *	1
Number of trainings	1
Avg training score *	90
Prev year rating *	5.0/5.0
Prediction (Yes)	100%



## LIMITATIONS AND NEXT STEPS

- Lack of identifiable information
- Need more data/datasets
- Productize codebase

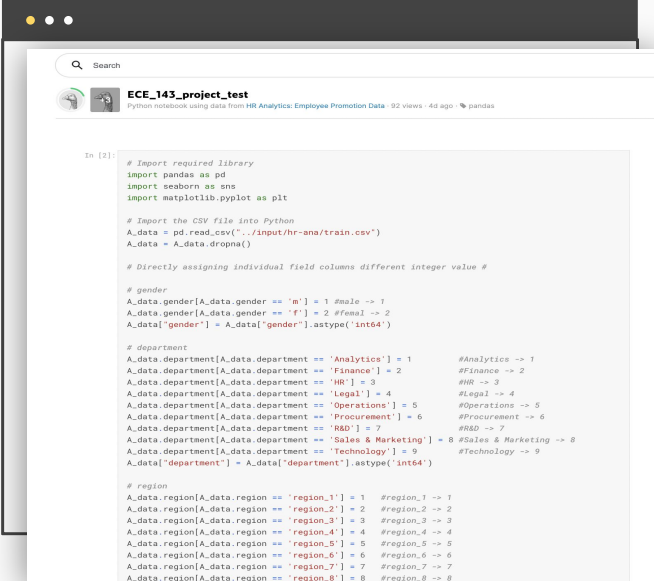


# THANKS

Does anyone have any questions?

Group 9

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik** and illustrations by **Storyset**



```
In [2]:  
  
# Import required library  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
# Import the CSV file into Python  
A_data = pd.read_csv("../input/hr-ana/train.csv")  
A_data = A_data.dropna()  
  
# Directly assigning individual field columns different integer value #  
  
# gender  
A_data.gender[A_data.gender == 'm'] = 1 #male -> 1  
A_data.gender[A_data.gender == 'f'] = 2 #female -> 2  
A_data['gender'] = A_data['gender'].astype('int64')  
  
# department  
A_data.department[A_data.department == 'Analytics'] = 1 #Analytics -> 1  
A_data.department[A_data.department == 'Finance'] = 2 #Finance -> 2  
A_data.department[A_data.department == 'HR'] = 3 #HR -> 3  
A_data.department[A_data.department == 'Legal'] = 4 #Legal -> 4  
A_data.department[A_data.department == 'Operations'] = 5 #Operations -> 5  
A_data.department[A_data.department == 'Procurement'] = 6 #Procurement -> 6  
A_data.department[A_data.department == 'R&D'] = 7 #R&D -> 7  
A_data.department[A_data.department == 'Sales & Marketing'] = 8 #Sales & Marketing -> 8  
A_data.department[A_data.department == 'Technology'] = 9 #Technology -> 9  
A_data['department'] = A_data['department'].astype('int64')  
  
# region  
A_data.region[A_data.region == 'region_1'] = 1 #region_1 -> 1  
A_data.region[A_data.region == 'region_2'] = 2 #region_2 -> 2  
A_data.region[A_data.region == 'region_3'] = 3 #region_3 -> 3  
A_data.region[A_data.region == 'region_4'] = 4 #region_4 -> 4  
A_data.region[A_data.region == 'region_5'] = 5 #region_5 -> 5  
A_data.region[A_data.region == 'region_6'] = 6 #region_6 -> 6  
A_data.region[A_data.region == 'region_7'] = 7 #region_7 -> 7  
A_data.region[A_data.region == 'region_8'] = 8 #region_8 -> 8
```

Kaggle Notebook:

<https://www.kaggle.com/eugeneli9/ece-143-project-test>

