



HR Promotion Prediction

Muneerah Albadr





Agenda

- Project Goal
- Dataset
- Process Data
- Modeling

Project Goal

HR analytics revolutionises the way human resources departments operate, leading to higher efficiency and better overall results. However, one of the problems most HR departments face is identifying the right people for promotion and preparing them in time.

- The goal is to predict the identifying the employees most likely to get promoted.

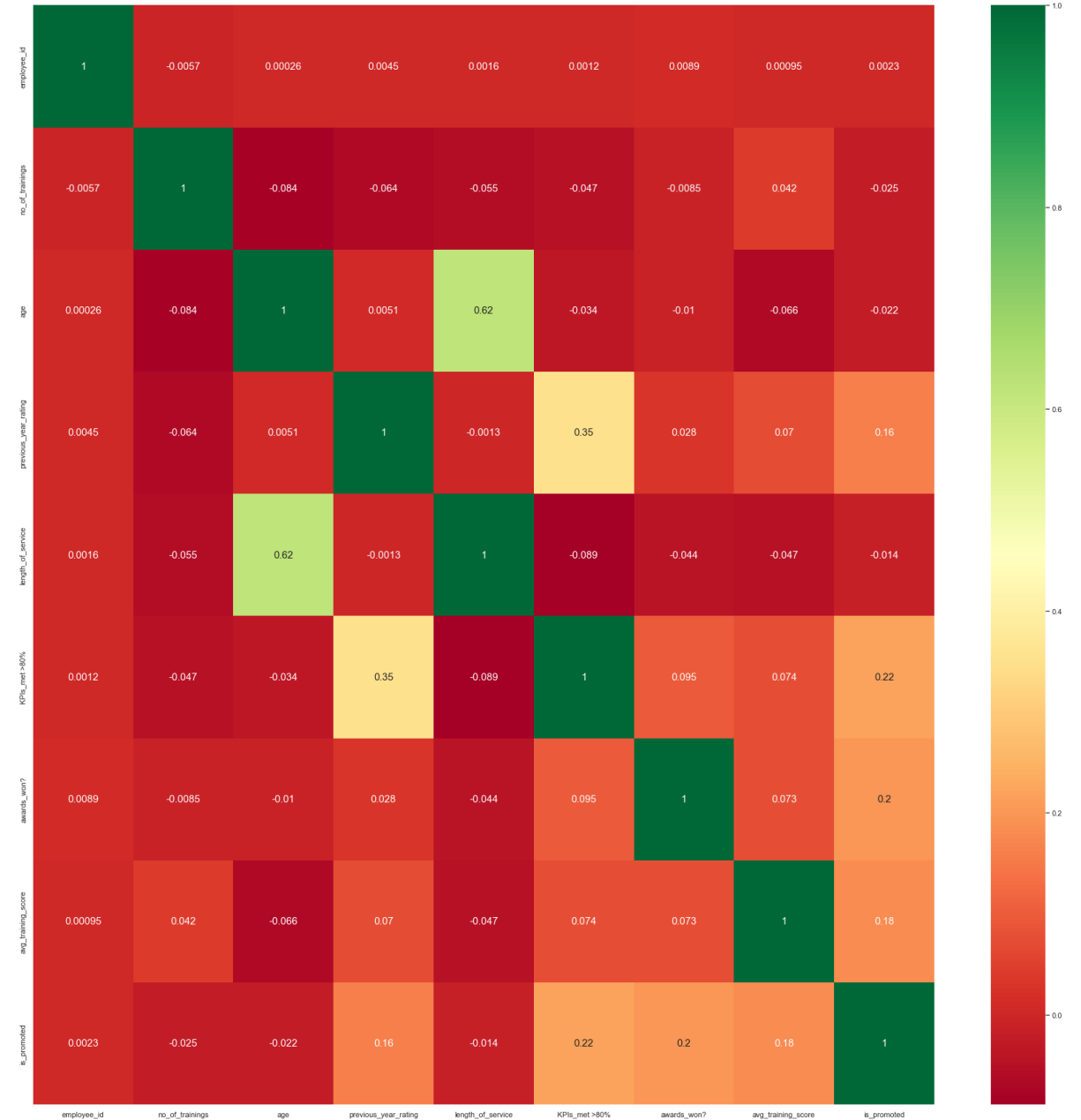
Dataset

This dataset contains about +23,400 employees' records from a large multinational corporation (MNC), and they have 9 broad verticals across the organisation. and 13 features.

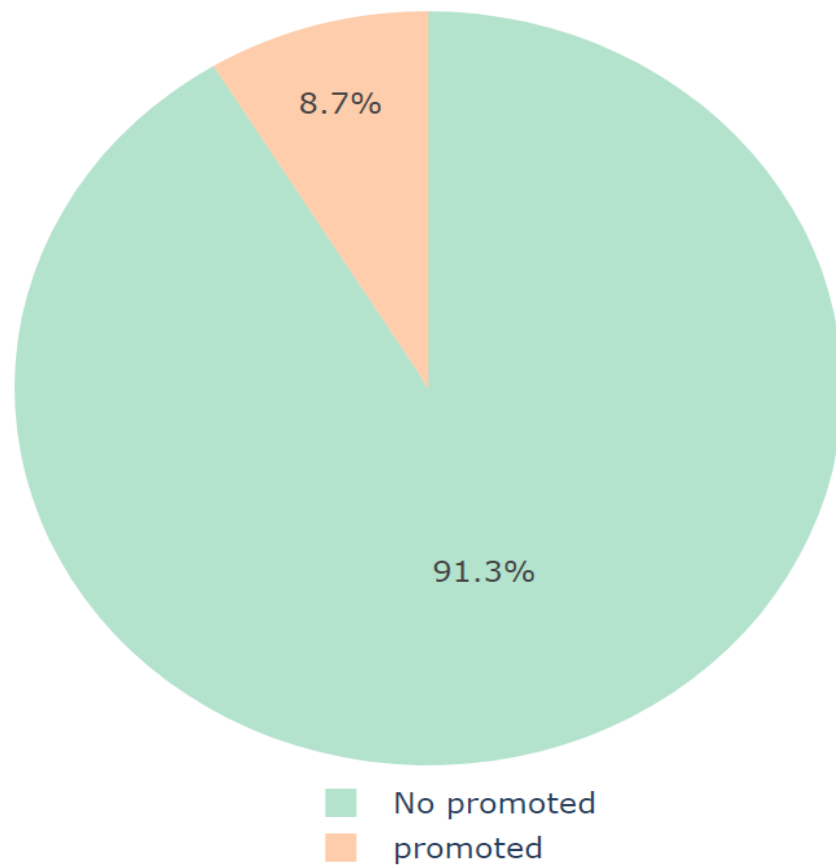
	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	KPIs_met>80%	awards_won?	avg_training_score
0	65438	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	5.0	8	1	0	49
1	65141	Operations	region_22	Bachelor's	m	other	1	30	5.0	4	0	0	60
2	7513	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	3.0	7	0	0	50
3	2542	Sales & Marketing	region_23	Bachelor's	m	other	2	39	1.0	10	0	0	50
4	48945	Technology	region_26	Bachelor's	m	other	1	45	3.0	2	0	0	73

Process Data

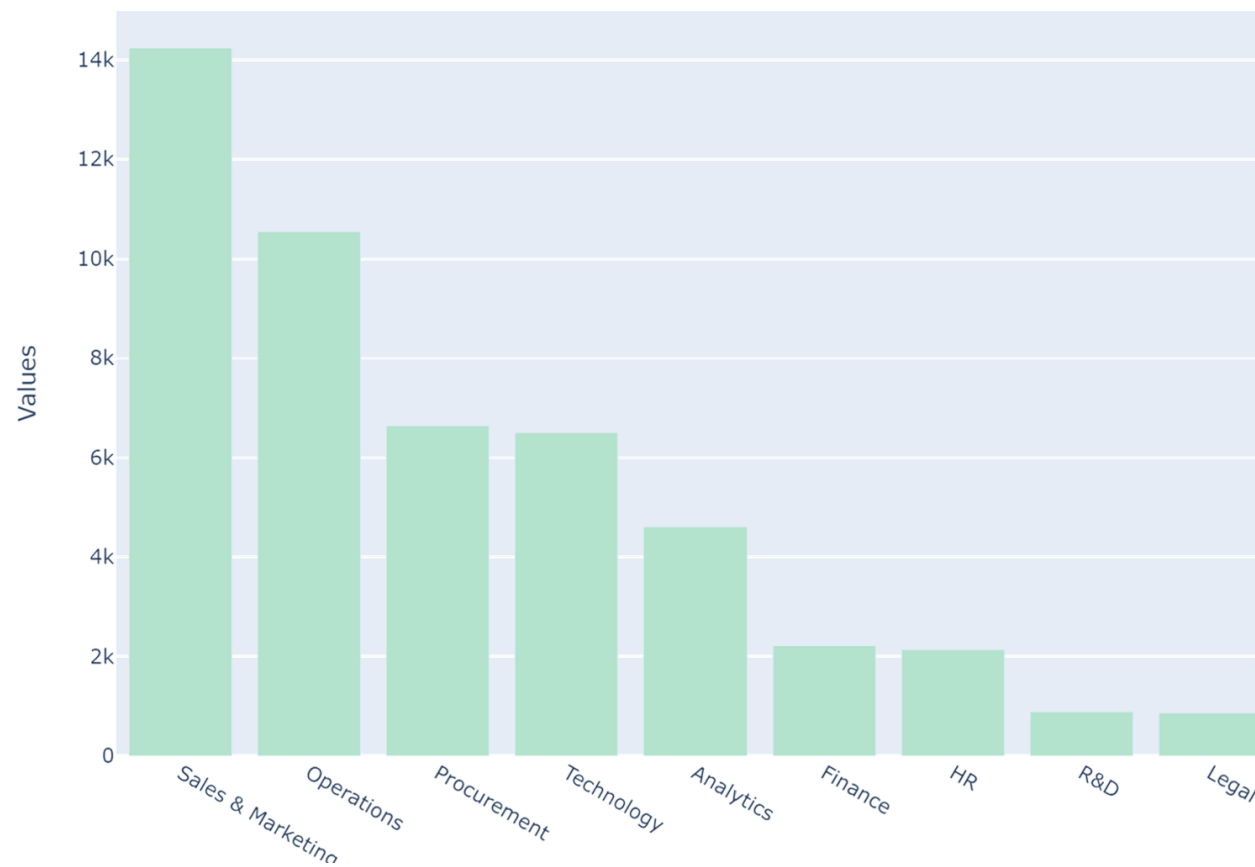
- ❖ Explore the features
- ❖ Clean the data
- ❖ Exploratory Data Analysis and Visualization



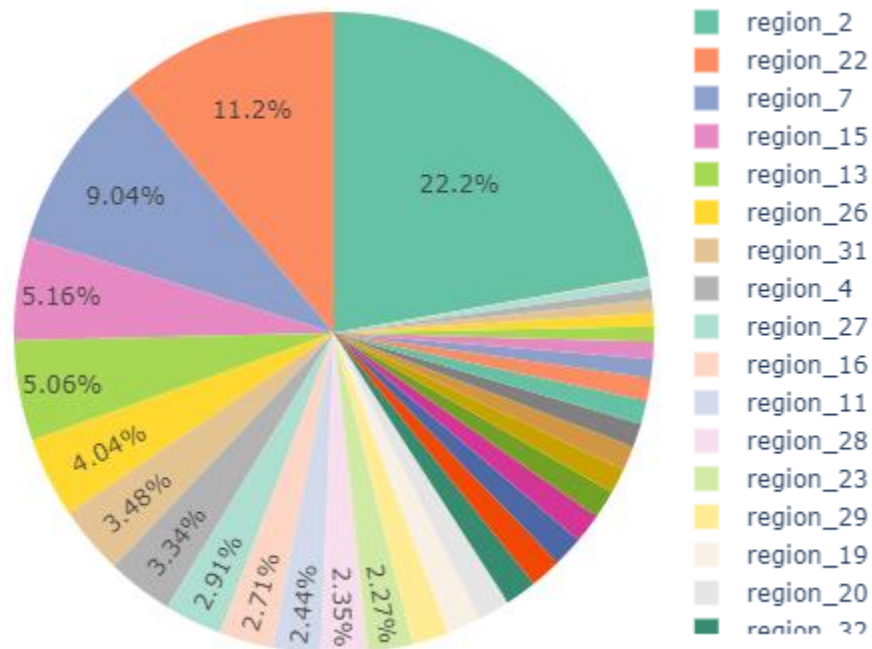
Employees Promoted



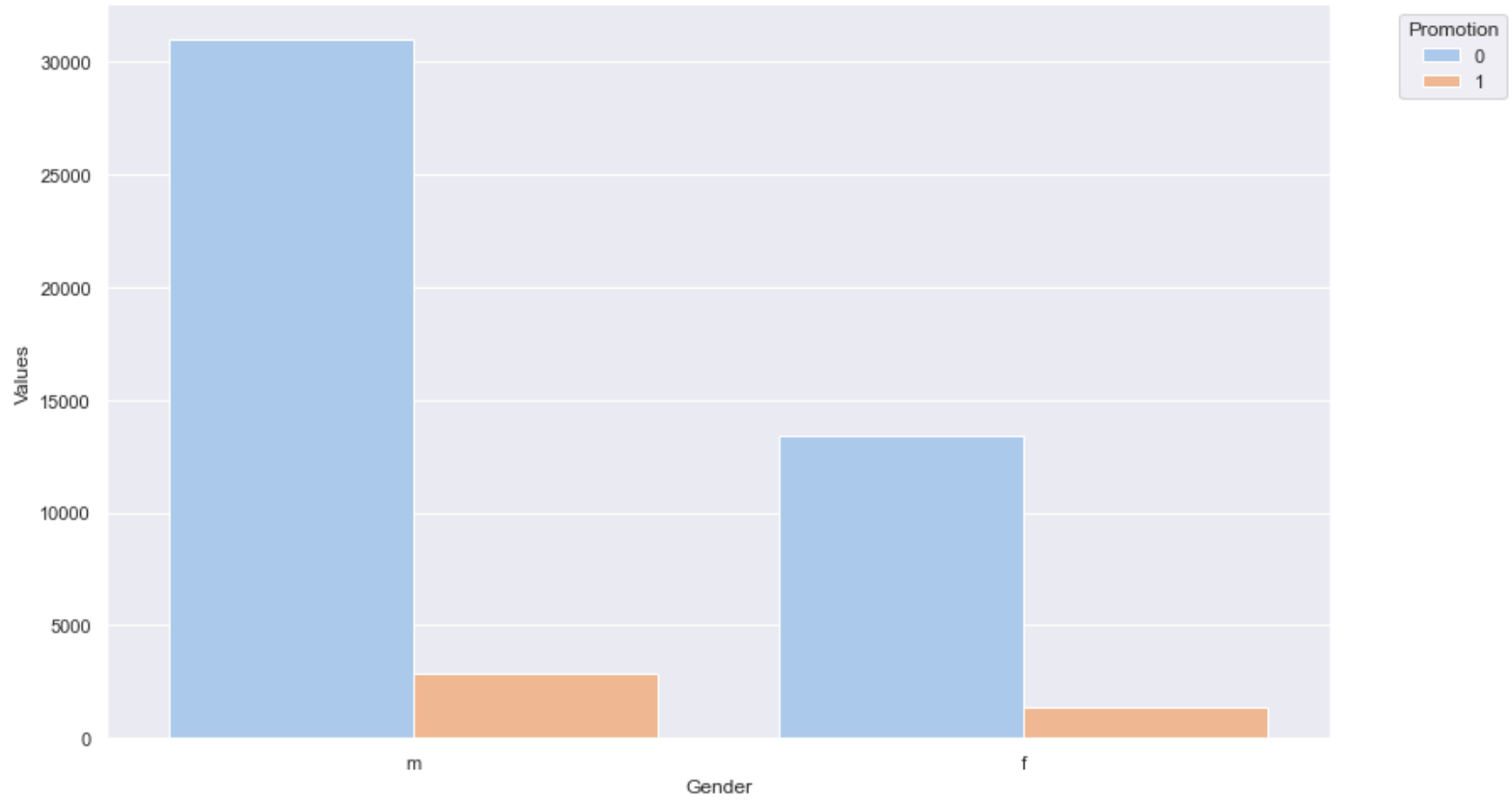
Number of Employees by Department



Number of Employees by Region

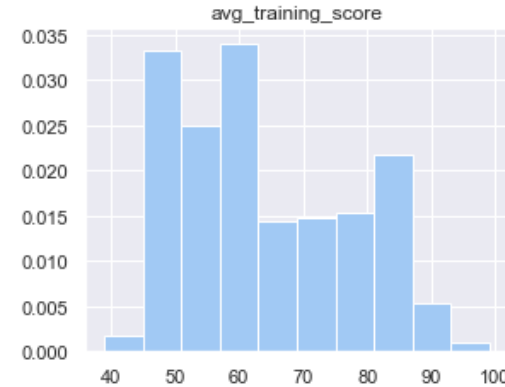
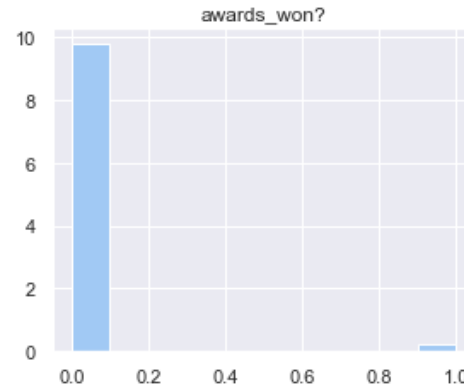
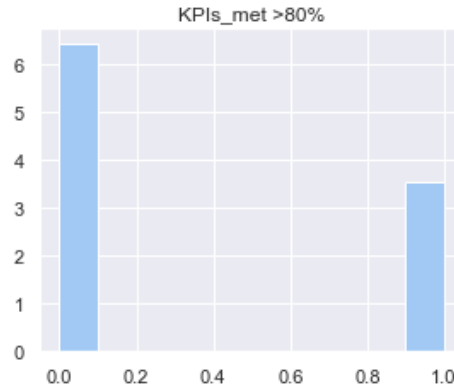
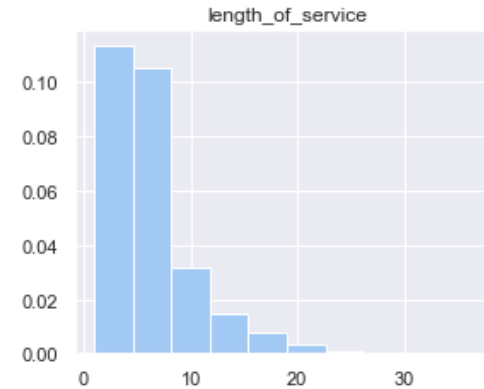
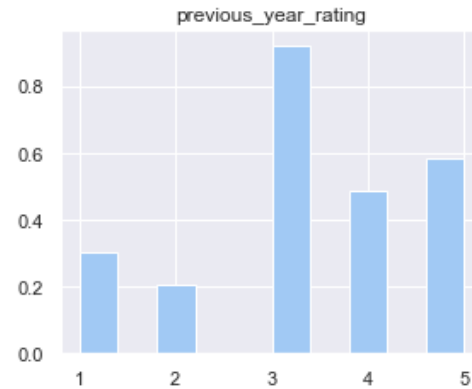
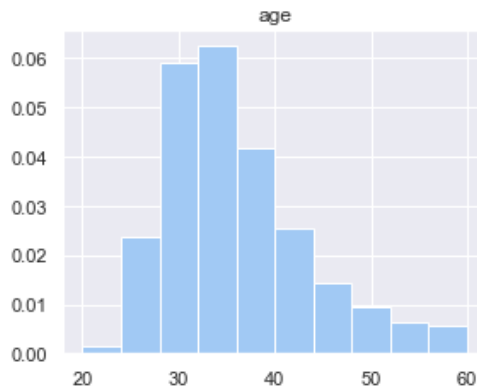


Promoted and Non-Promoted Employees by Gender



Modeling

- ❖ Data cleansing to train the model
- ❖ Convert the categories variables into Numerical variables



Modeling

- ❖ Data cleansing to train the model
- ❖ Convert the categories variables into Numerical variables
- ❖ Train and compare the scores of multiple ranking algorithms (Logistic Regression, Linear SVC, Decision Tree Classifier, Random Forest Classifier, Naive Bayes Classifier, K-nearest Neighbors Classifier)

Modeling

❖ A Random Forest Classifier was used to classify the data.

* Accuracy: 93 %

* Recall: 32 %

* Precision: 75 %

* F1: 45 %

	Decision Tree	Gaussian Naive Bayes	Logistic Regression	Support Vector Classifier	K-nearest Neighbors	Random Forest	Best Score
Accuracy	0.893896	0.911344	0.916482	0.915228	0.923448	0.932203	Random Forest
Precision	0.403075	0.464391	0.668991	0.662146	0.750171	0.752686	Random Forest
Recall	0.457697	0.128066	0.078688	0.052930	0.180997	0.328211	Decision Tree
F1 Score	0.428641	0.200695	0.140691	0.097949	0.291202	0.457073	Random Forest



Conclusion

So in our findings, for this particular data set and classifiers used, random forest is by far the most performing and accurate classifier.



Thanks..

