EMPLOYEE PROMOTION ANALYSIS

EXPLORATARY DATA ANALYSIS

GOPE SAI KUMAR
Sai.gope2002@gmailcom

Employee Promotion Analysis

Project Overview

This project analyzes employee promotion trends using a dataset containing various employee attributes. The goal is to identify patterns and visualize the promotion distribution using a pie chart.

Dataset Information

- **Filename:** HR Analytics Dataset
- **Total Records:** 54,808
- **Columns:** 14
- **Target Variable** (1 = Promoted, 0 = Not Promoted)

Project Overview:-

Problem Statement: Predict employee promotion eligibility based on certain attributes/features.

Dataset Overview:

The dataset consists of multiple employee-related attributes, including demographic, performance, and training data. The target variable is is_promoted, which indicates whether an employee was promoted (1) or not (0).

Key Columns:

1. employee_id - Unique identifier for each employee
2. department - Department the employee belongs to
3. region - Geographic region of the employee
4. education - Employee's education level
5. gender - Gender of the employee
6. recruitment_channel - Source of recruitment
7. no_of_trainings - Number of training programs attended

8.age - Employee age

9.previous_year_rating - Performance rating from the previous year

10. length_of_service - Years of service in the company
11. KPIs_met >80% - Whether performance KPIs exceeded 80%
12. awards_won? - Whether the employee has won any awards
13. avg_training_score - Average score in training programs
14. is_promoted - **Target Variable** (1 = Promoted, 0 = Not Promoted)

Objectives

- Load and explore the dataset
- Analyze the distribution of promotions
- develop a predictive model for employee promotion. The model utilizes employee-related features to predict whether an employee will be promoted.

Implementation Steps

1. **Load the Dataset**

- Read the CSV file using pandas
- Inspect the dataset structure and missing values

2. **Data preprocessing**

Features:

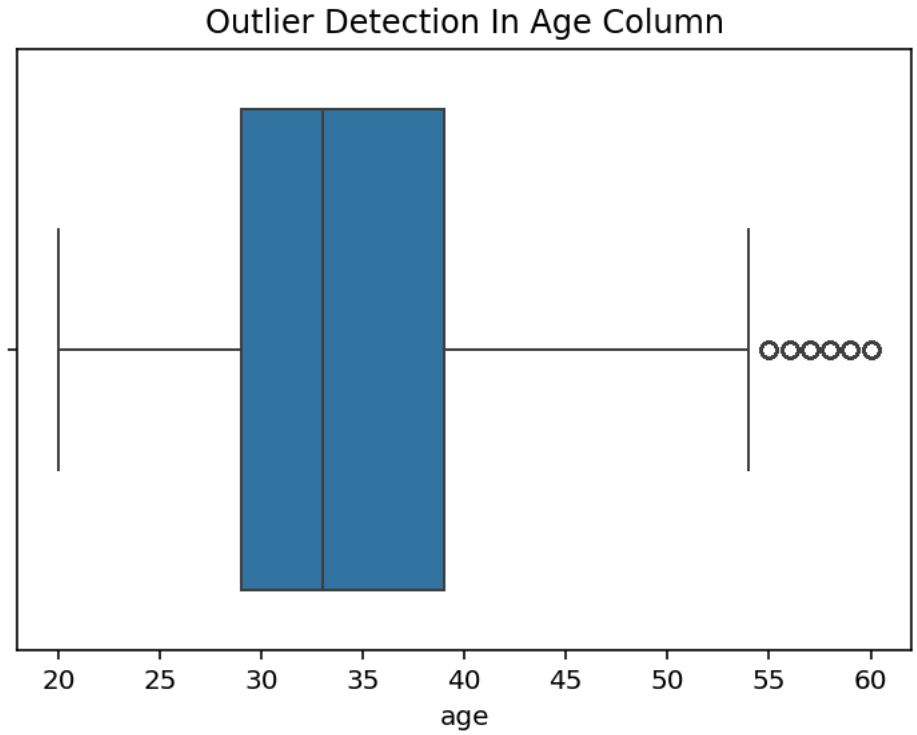
- Categorical: department, region, education, gender, recruitment_channel
- Numerical: no_of_trainings, age, previous_year_rating, length_of_service, KPIs_met >80%, awards_won?, avg_training_score

Handle missing values:

- education: Impute missing values using the mode (most frequent value in the columns
- previous_year_rating: Fill missing values with 0, as employees with length_of_service equal to 1 do not have a previous rating.

Outlier detection :

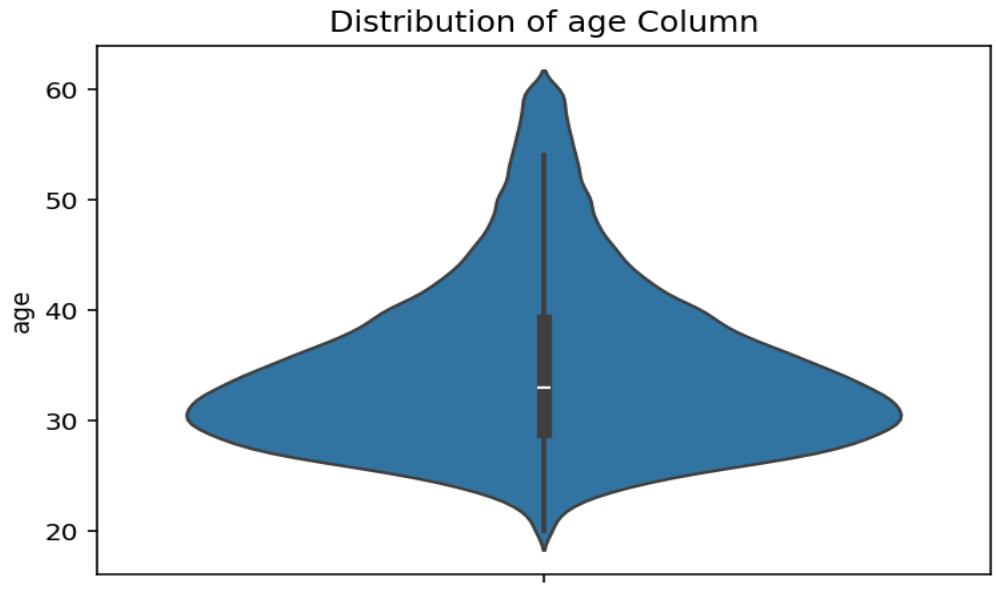
- In age column we are checking outliers



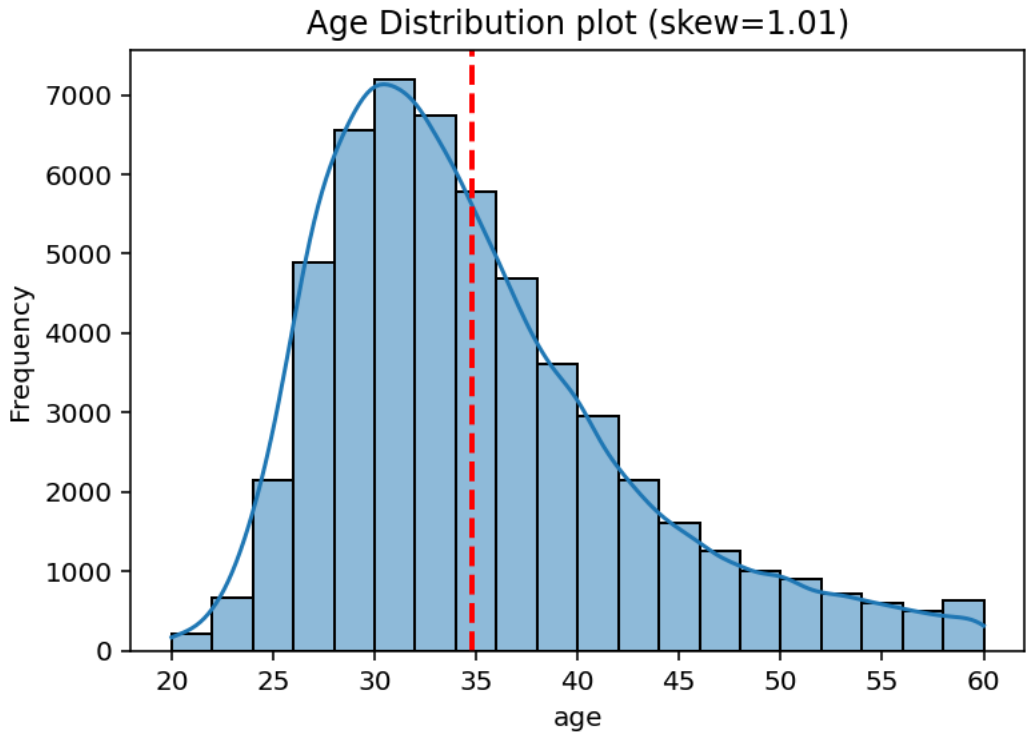
- We have 1435 outliers in age column
- Most of the outliers are starting from age 55 to 60

Checking distribution :

- Checking distribution of age column by violin plot

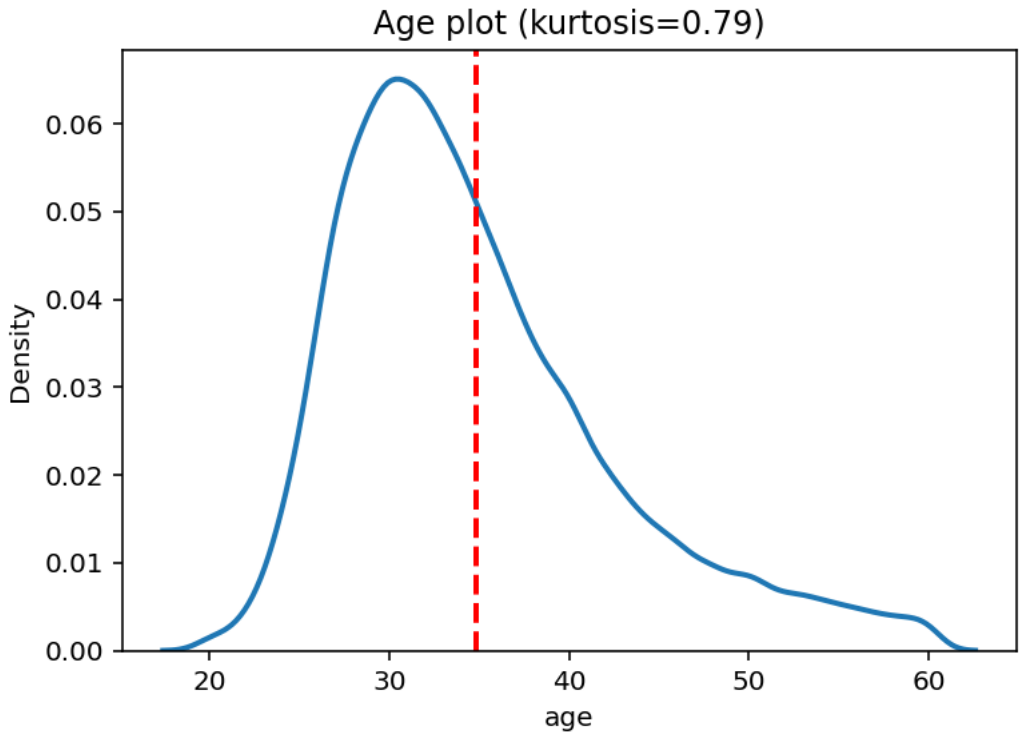


Histplot For Age Column :



- By This Plot We Can Say There Are More People Whose Age Is Below The Mean Age(33)
- Skew is is 1 it has positive skewness
- Mean age is 34.8

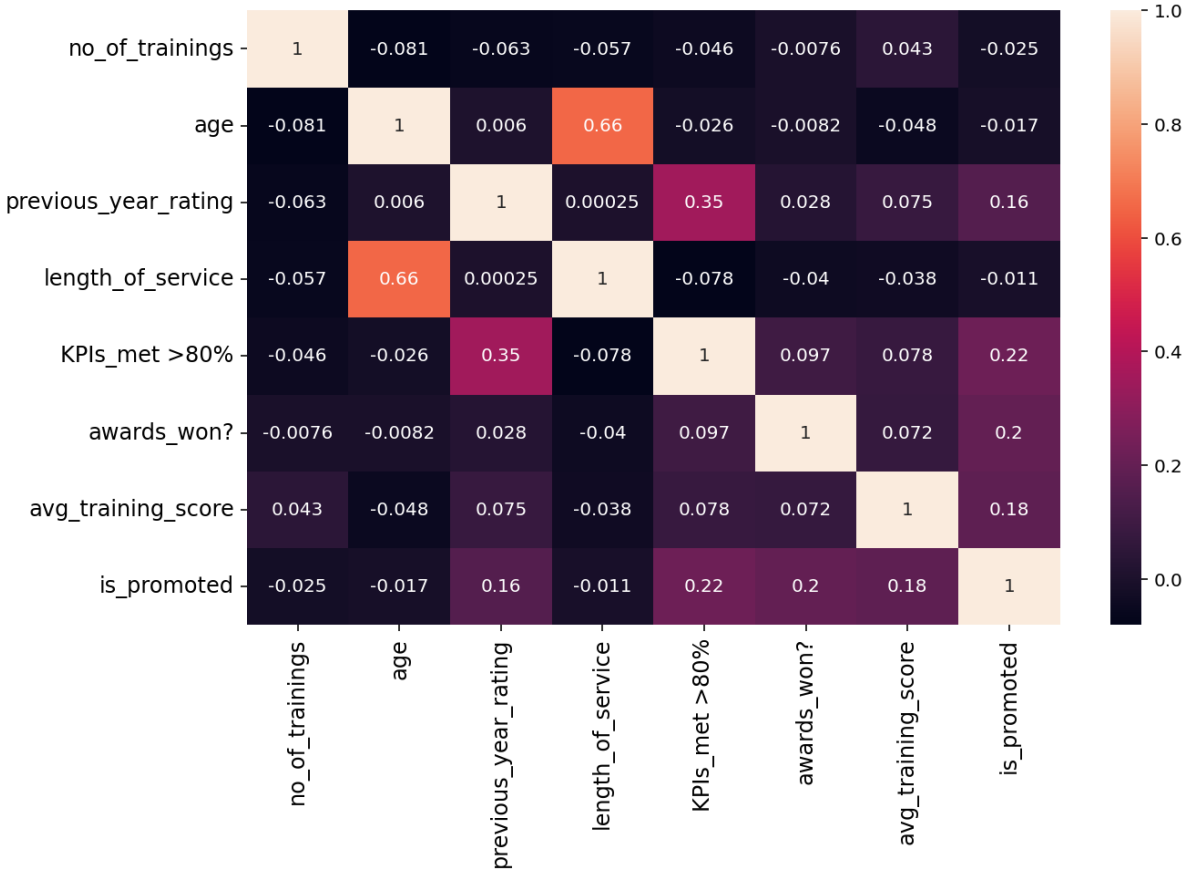
Plot For Kurtosis :



- The kurtosis is platkutic less extreme values
- It means the data mostly normal distributed
- It has standard deviation of 7.6 it was deviated from mean

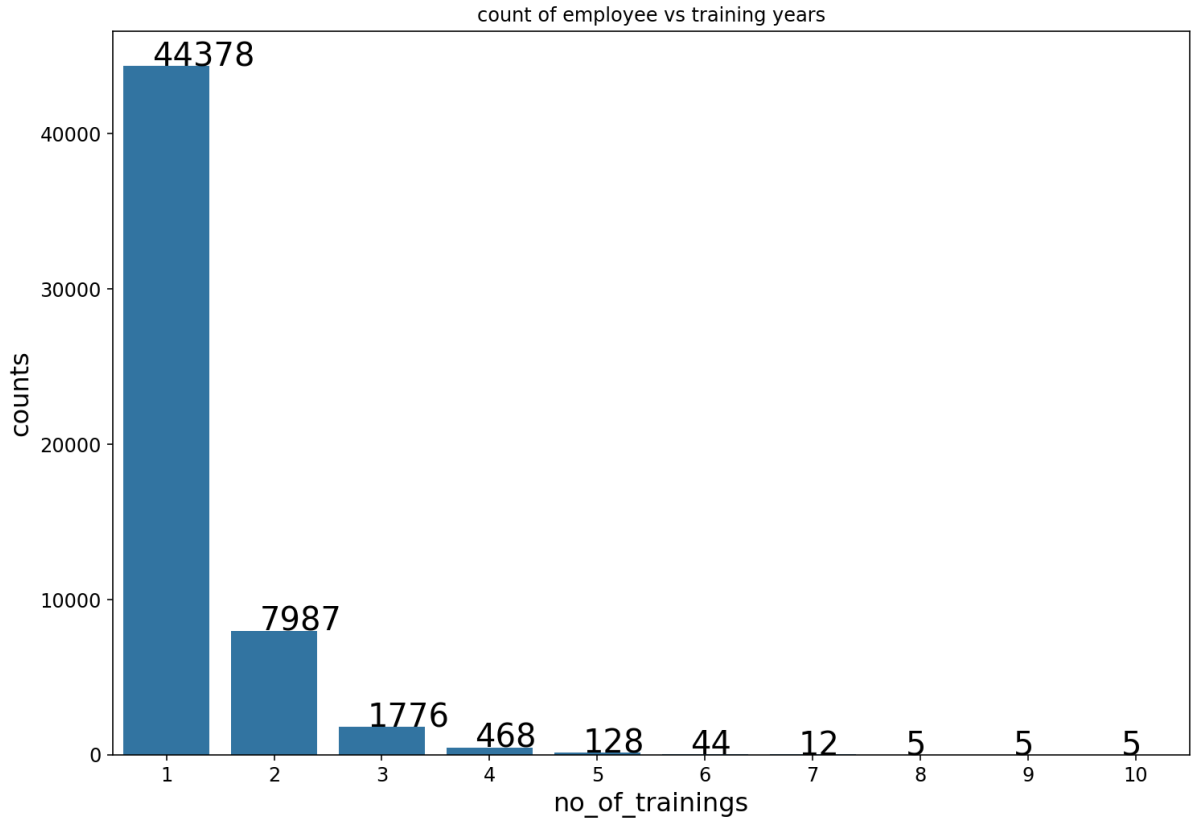
EXPLORATARY DATA ANALYSIS & VISUALIZATION

- Heatmap For Numerical Columns



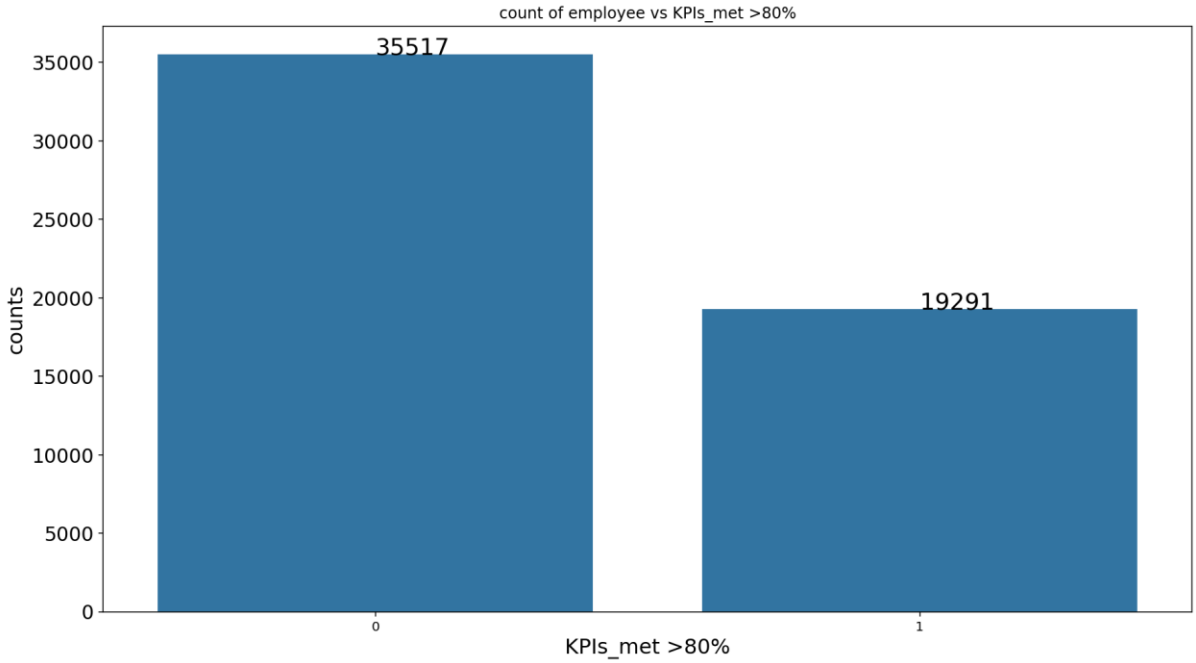
- Heatmap is to find corelations between numerical columns
- Age and length_of_service are mostly correlated
- Previous_year_rating and is_promoted corelation
- Kpi and previous year rating is corelated
- Most of values are not corelated much the data is non linear data

Countplot For no_of_trainings :



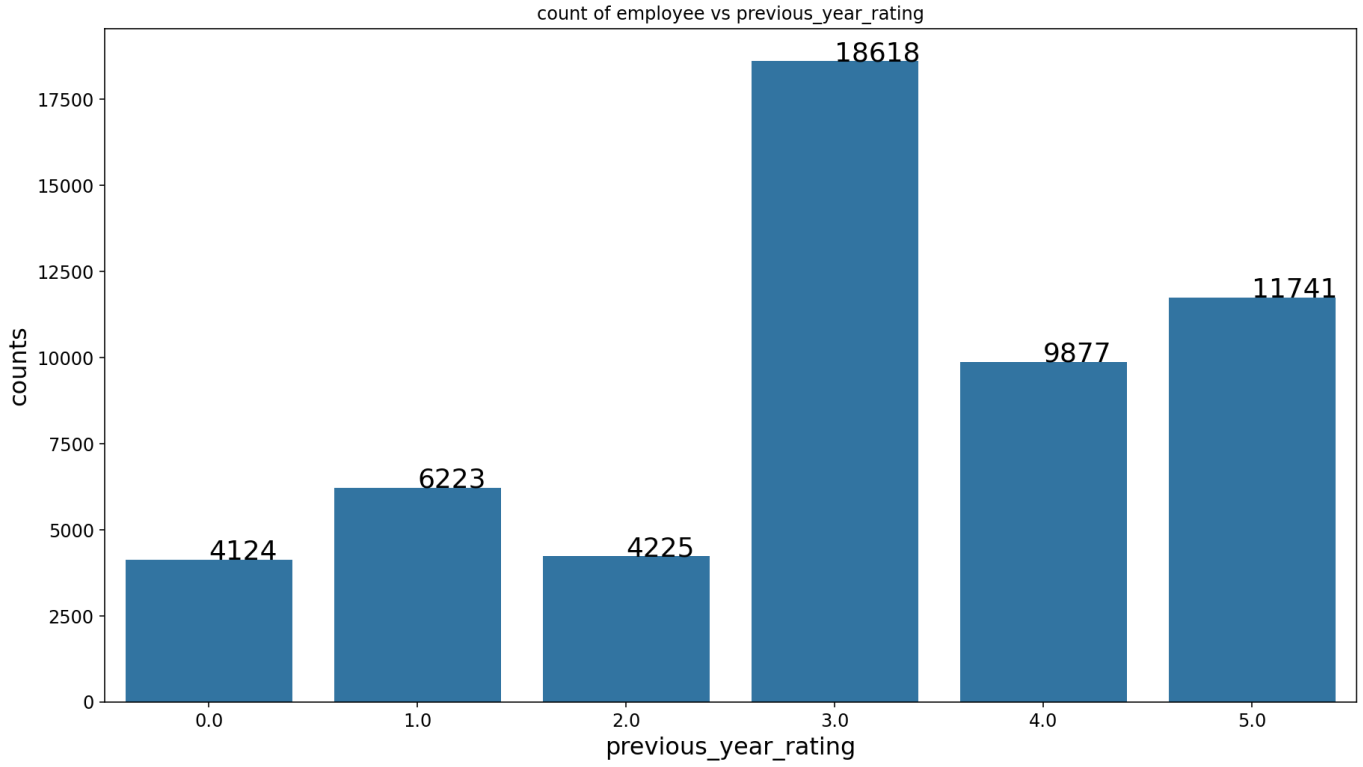
- Most Of Them Have 1 Year Of Training(44378)
- As 1 year suggest that may be they new joiners in the company

Countplot For KPIs_met >80% :



- As we see only 19291 who kpi are met are promoted

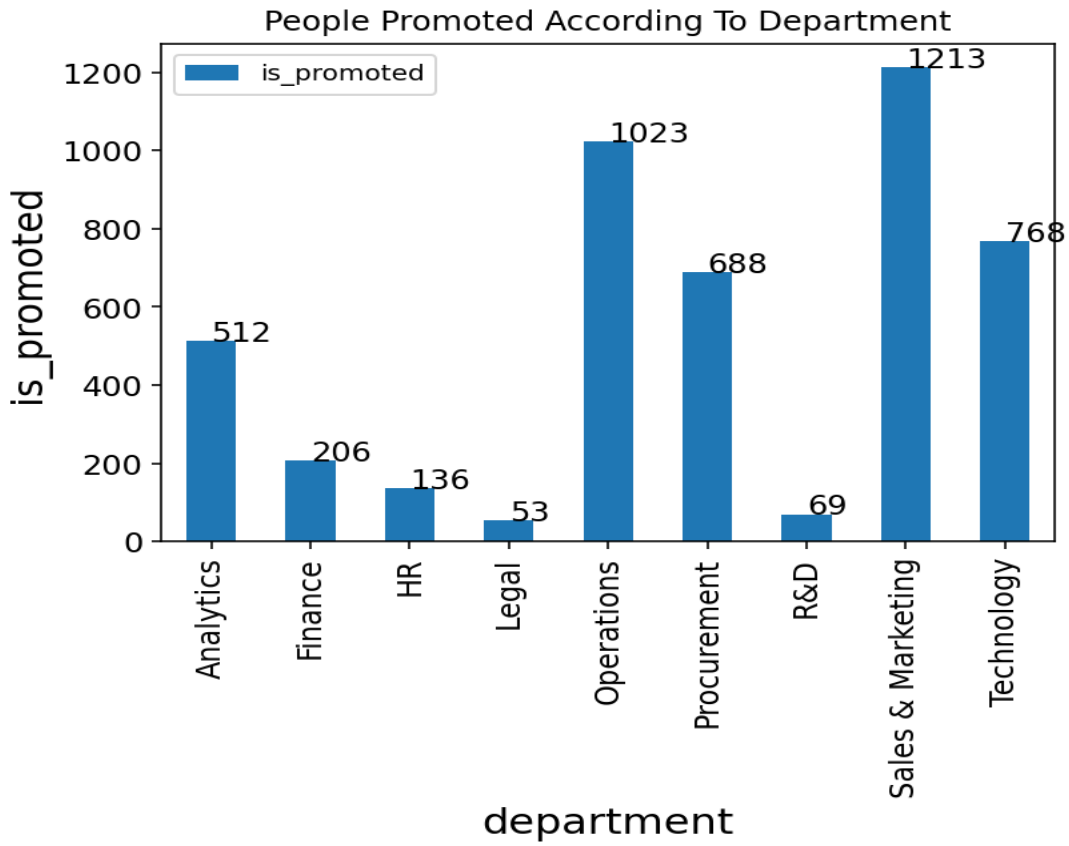
Countplot For previous_year_rating :



- Most of people previous year rating is 3

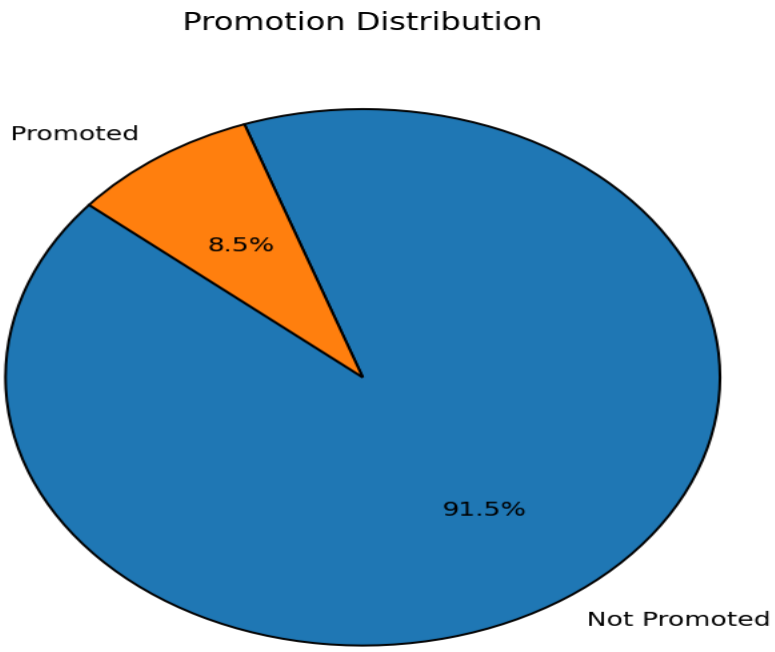
- We have 4124 employee with zero rating

Bar Plot For "department" vs "is_promoted" :



- Most of employees from sales marketing and operations is likely to get promoted

Pie Chart To Check Percentage Of People Promoted Vs Not Promoted:



It looks like most of people are not likely to get promoted

Pivot Table For Education vs Is_promoted:

group_edu_prom - DataFrame

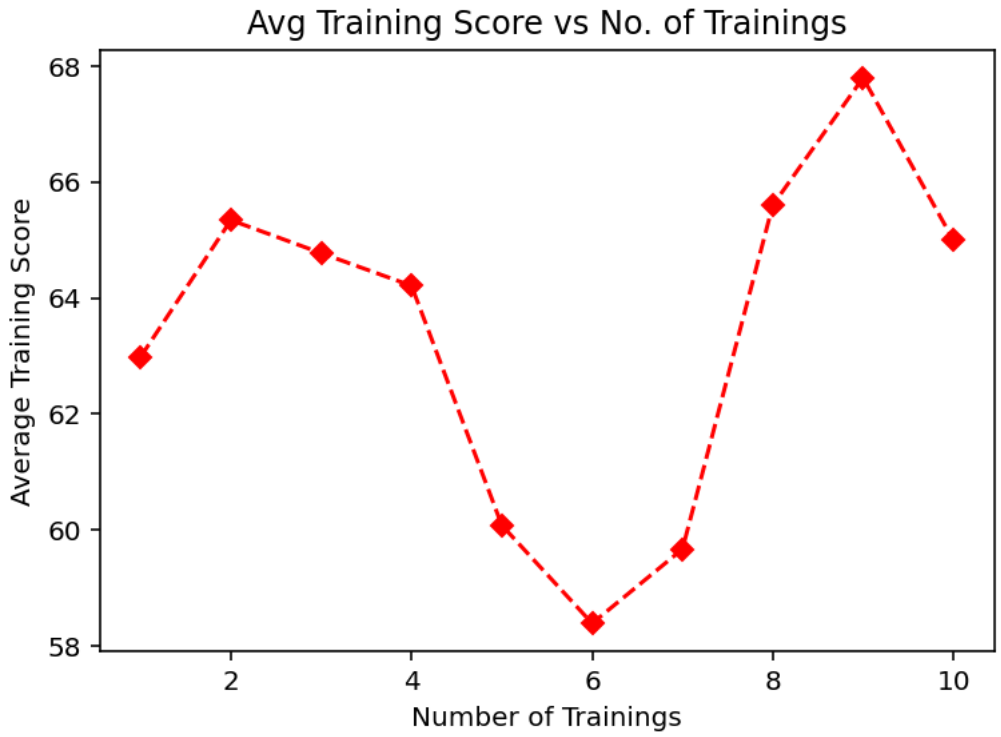
Index	education	not_promoted	is_promoted	promotion_per_edu	promotion_per_total
0	Bachelors	35948	3130	8.00962	5.71085
1	BelowSecondary	738	67	8.32298	0.122245
2	Mastersabove	13454	1471	9.85595	2.68391

- We have more promotions in employee where education background bachelors
- But when we go to percentages of promotions according education and the mastersabove had more promotion rate with 9 percentage
- And according total employees the bachelors have more promotions

Line plot For avg_training_score vs length_of_service :

- The average training score is between 60-65 upto the experience of 24 years
- And we have highest score around experience of 32 to 34 years and drop in scores after that

Line plot For avg_training_score vs no_of_trainings:

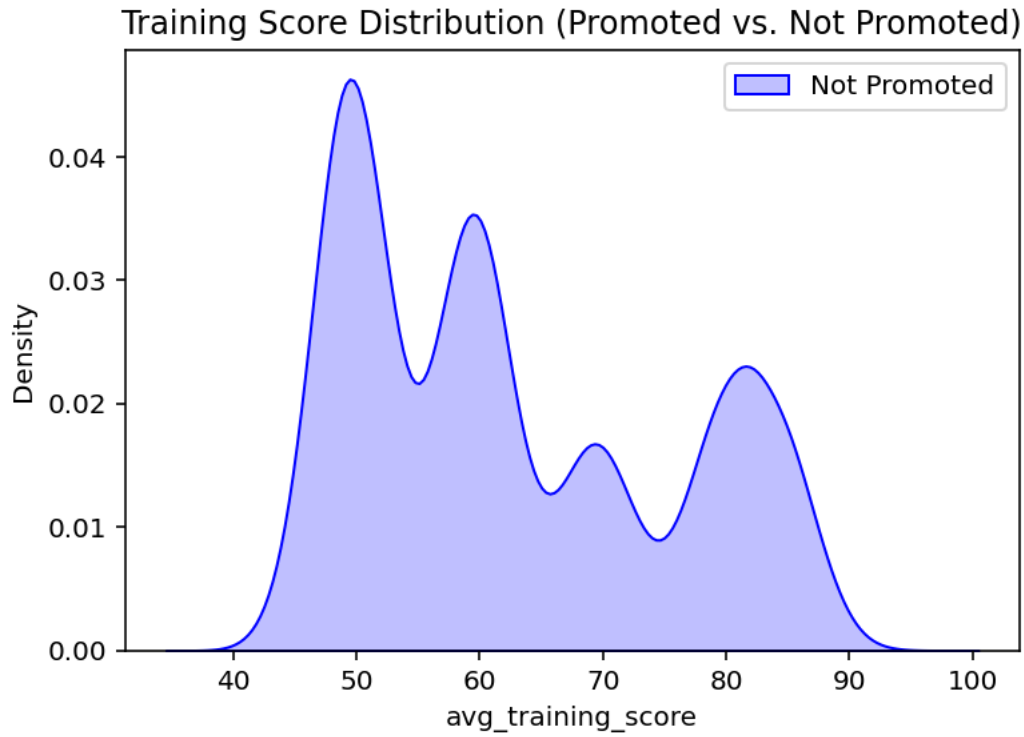


- The average score is more for employee who trainings of between 8 to 10
- There significant drop average training score at 5 to 7 number of trainings

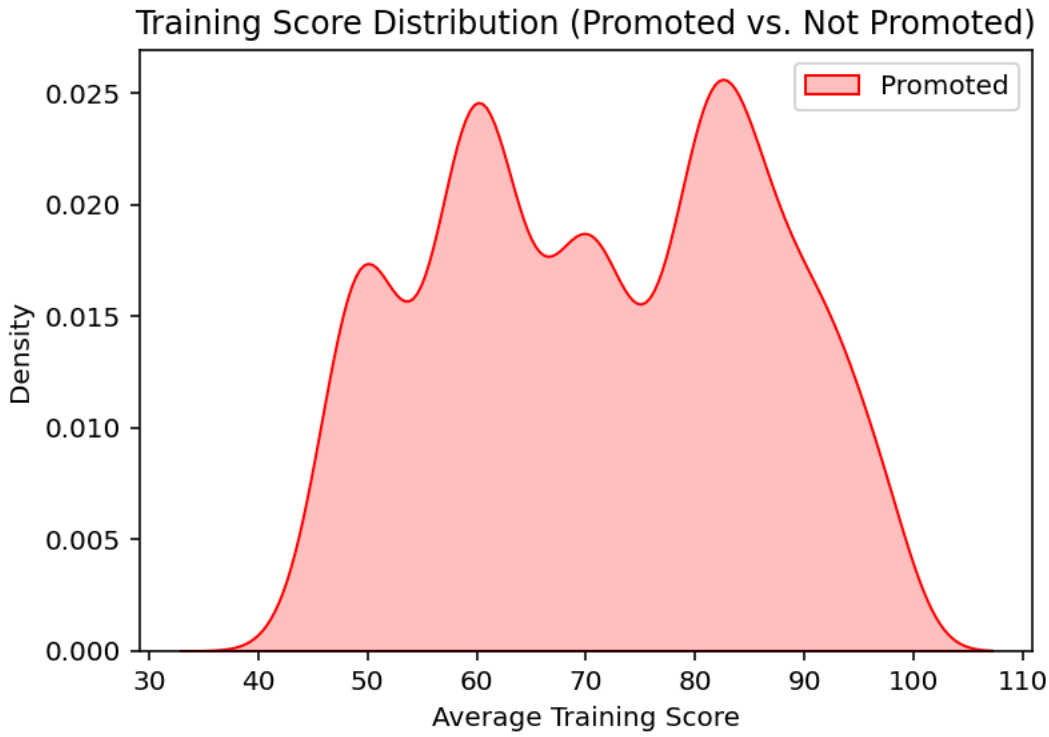
Line plot For avg_training_score vs no_of_trainings:

- From age 26 to 35 there is avg training score is more

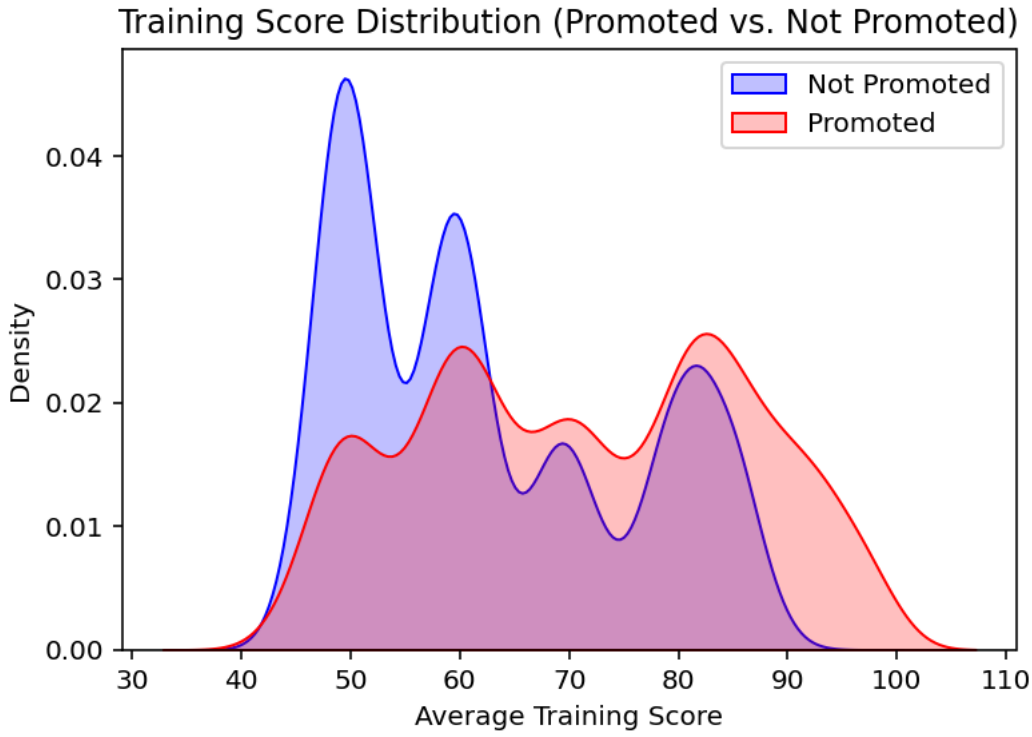
KDE Plot For Training Score Distribution for Non-Promoted Employee:



KDE Plot For Training Score Distribution for Promoted Employee:



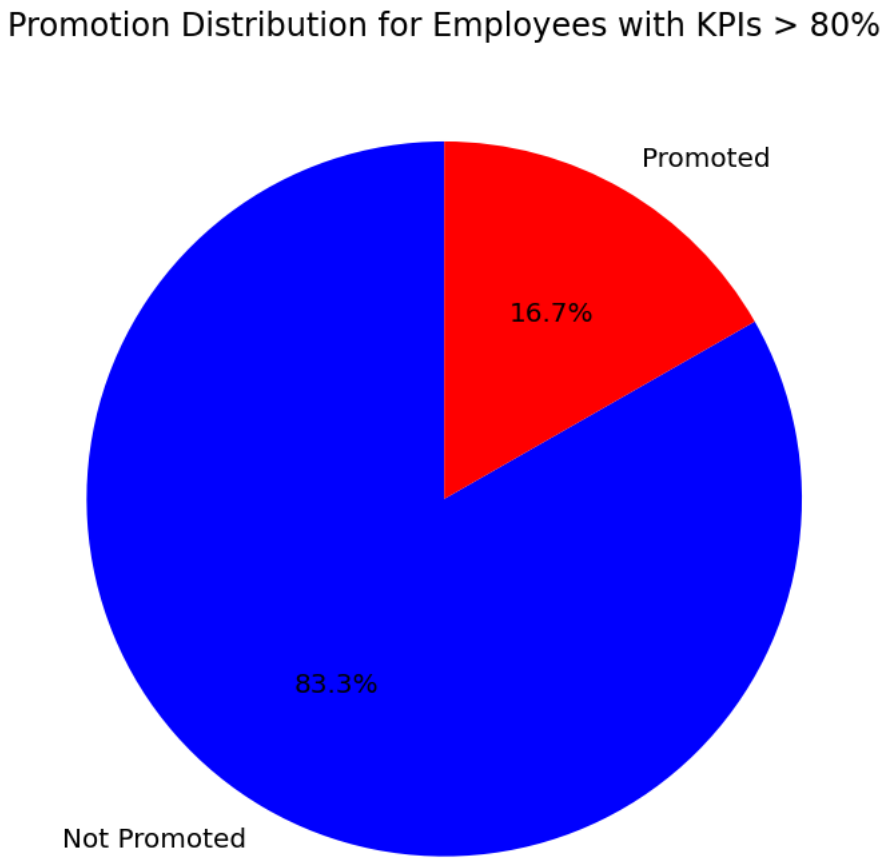
KDE Plot For Training Score Distribution for Promoted vs. Non-Promoted Employee:



- The range promotions are more where avg training score is between 80 to 100
- By this we know which range of people is likely to get promoted

KPIs_met >80% vs promotion :

- To analyize if even kpi is met the percentage employee getting promoted or not



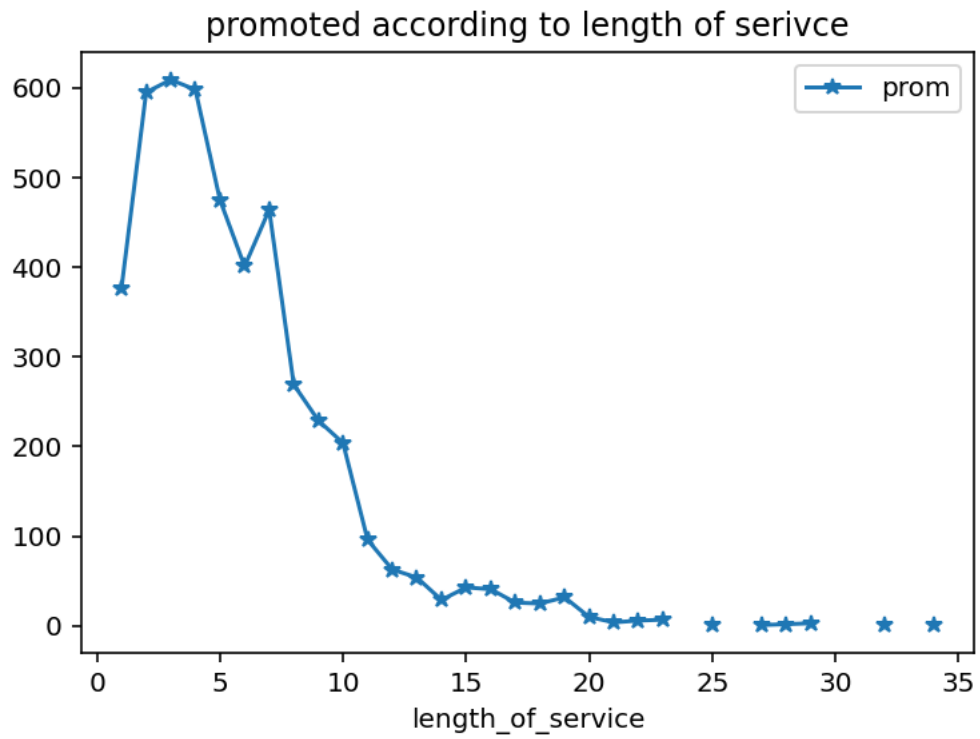
- Most of employees are not getting promoted even if kpi is greater than 80
- But there 16 percent of getting promotion if kpi is met

previous year rating vs promotions:

Index	previous_year_rating	not_promoted	promoted	promoted_acc_total	promoted_acc_rating
0	0	3785	339	0.618523	8.22017
1	1	6135	88	0.160561	1.41411
2	2	4044	181	0.330244	4.28402
3	3	17263	1355	2.47227	7.2779
4	4	9093	784	1.43045	7.93763
5	5	9820	1921	3.50496	16.3615

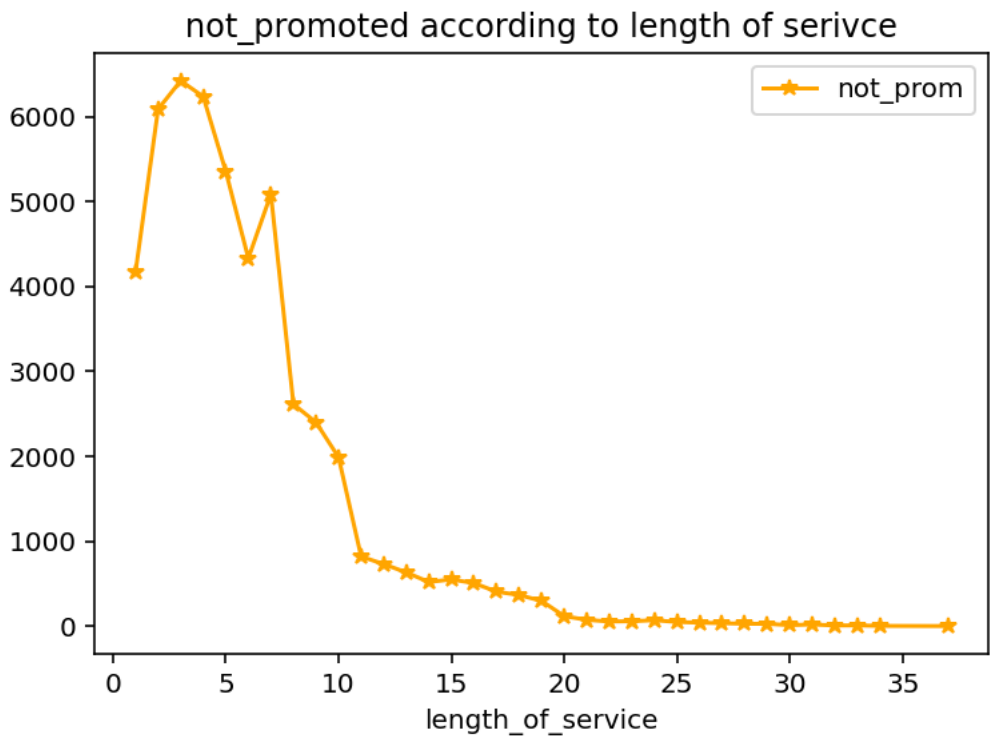
- People with rating 5 is getting promoted more

Length Of serivce vs is promoted:



- Promotions count is more in service 3 to 7

Length tof serivce vs is not_promoted:


not_promoted according to length of serivce
not_prom
6000
5000
4000
3000
2000
1000
0
0
5
10
15
20
25
30
35
length_of_service