

CLEANING DATA SET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 54808 entries, 0 to 54807
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   department                            54808 non-null  object
1   region                                54808 non-null  object
2   education                             52399 non-null  object
3   gender                                54808 non-null  object
4   recruitment_channel                   54808 non-null  object
5   no_of_trainings                       54808 non-null  int64
6   age                                   54808 non-null  int64
7   previous_year_rating                  50684 non-null  float64
8   length_of_service                     54808 non-null  int64
9   KPIs_met >80%                        54808 non-null  int64
10  awards_won?                           54808 non-null  int64
11  avg_training_score                     54808 non-null  int64
12  is_promoted                           54808 non-null  int64
dtypes: float64(1), int64(7), object(5)
memory usage: 5.4+ MB
```

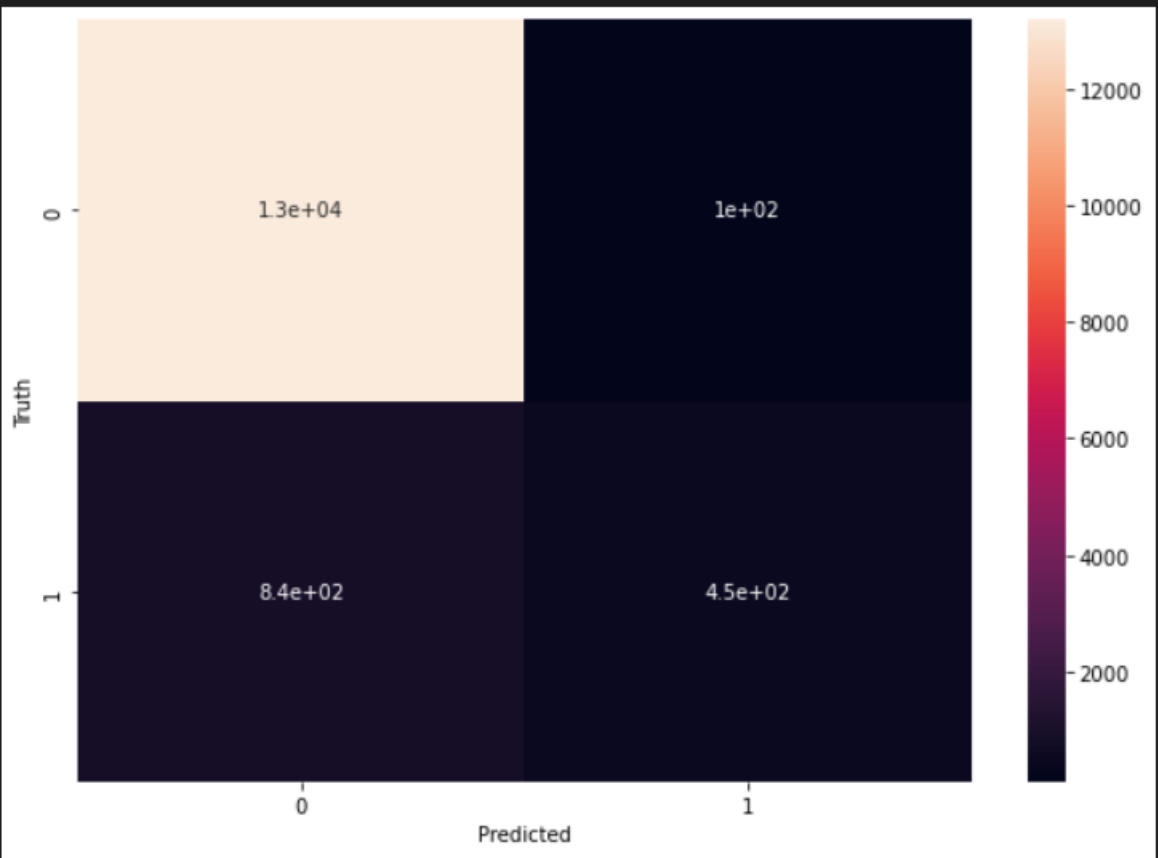
ACCURACY OF TEST_DATA USING DECISION TREE CLASSIFIER

```
Accuracy: 0.9349910946705028
```

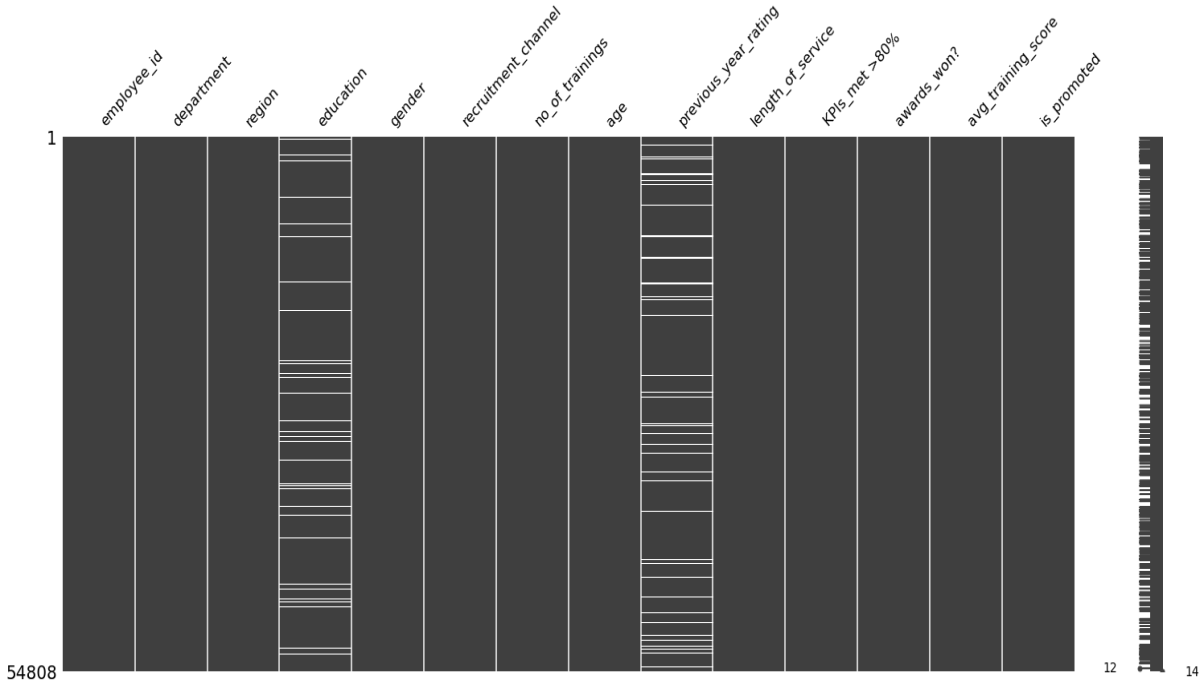
ACCURACY OF TEST_DATA USING RANDOM FOREST CLASSIFIER

```
Accuracy: 0.9253322372927798
```

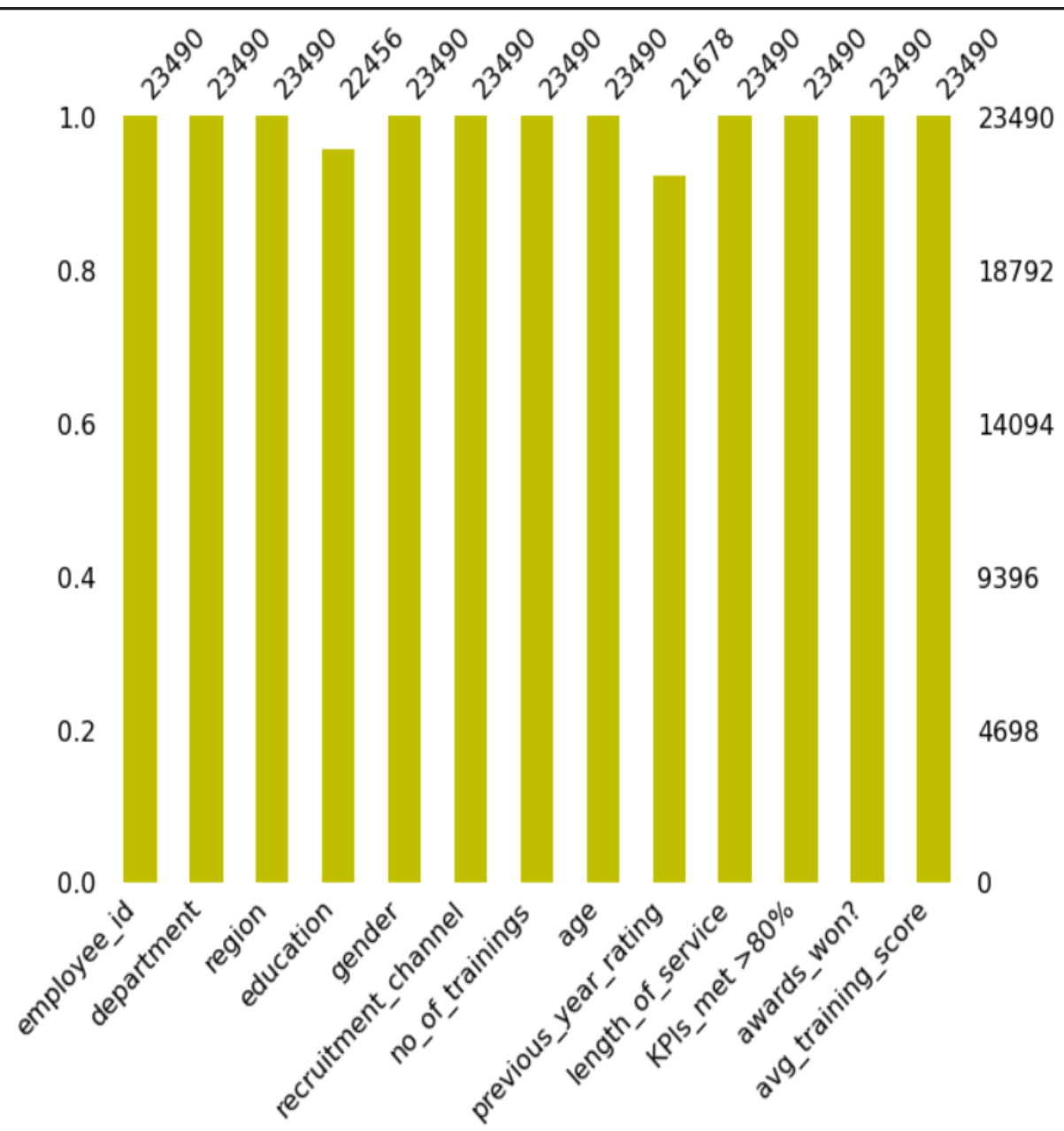
PLOT OF CONFUSION MATRIX



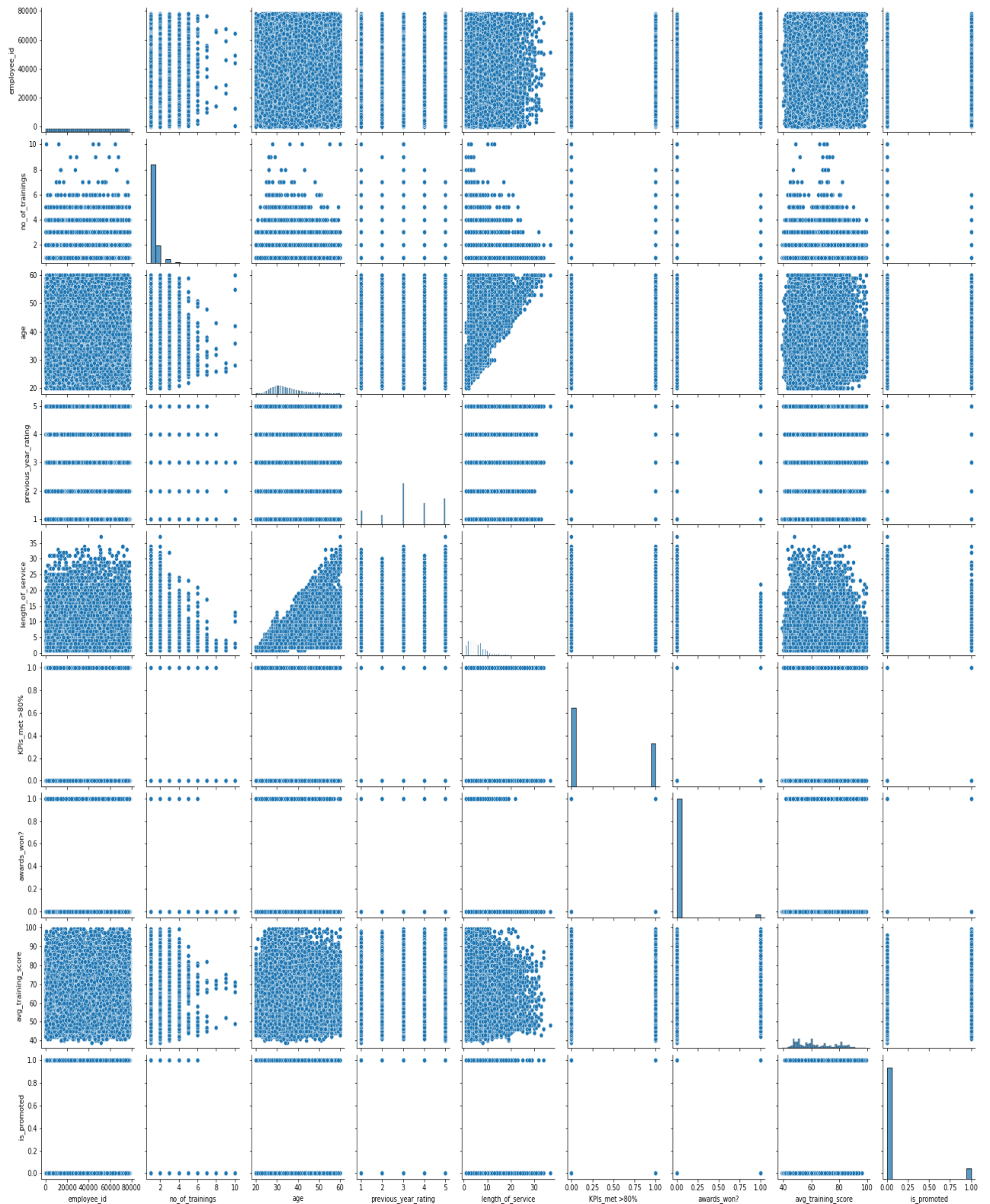
PLOT OF NULL VALUES IN THE DATASET



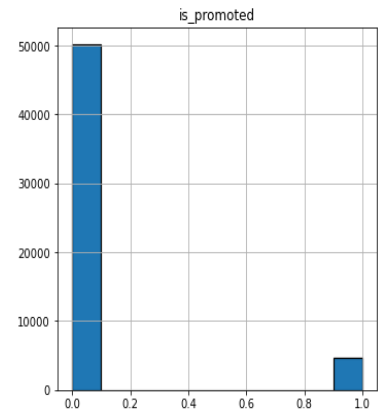
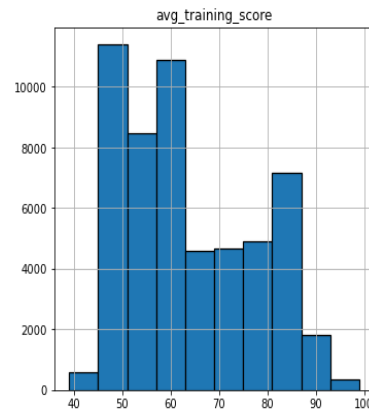
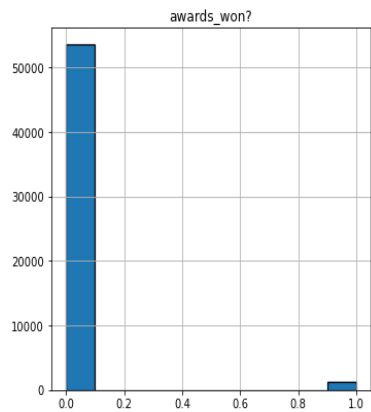
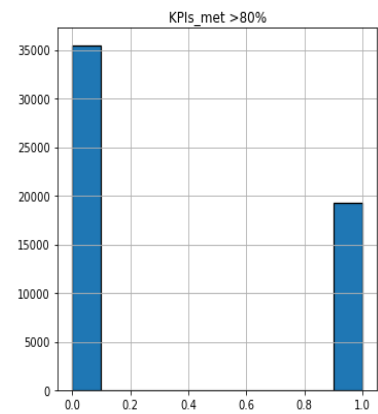
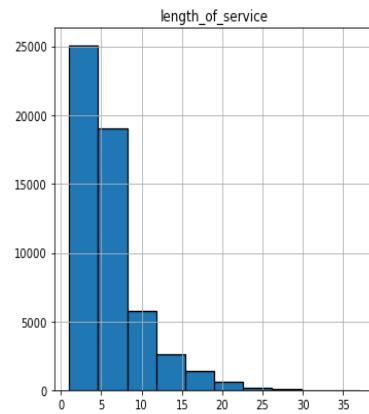
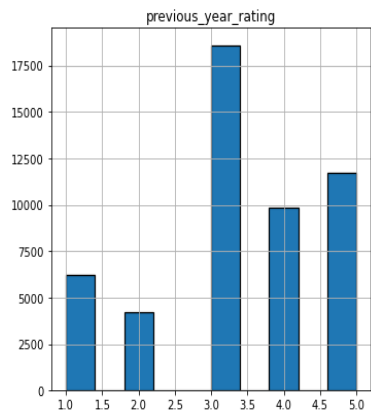
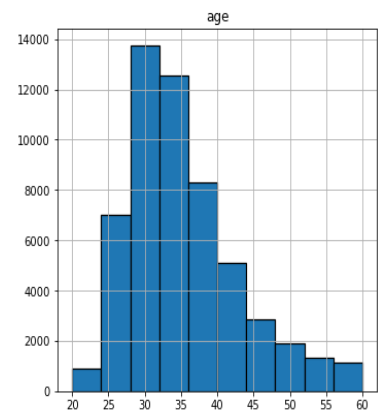
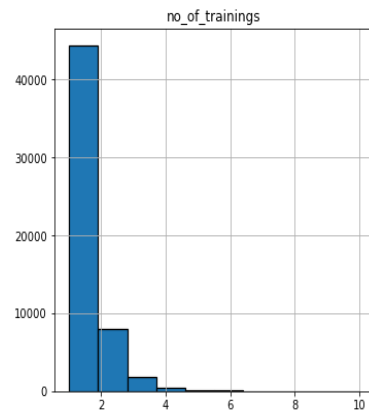
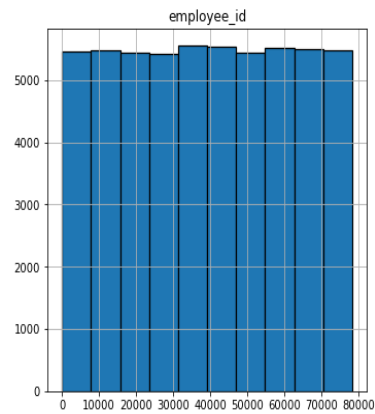
PLOT OF MISSING VALUES IN TEST DATA



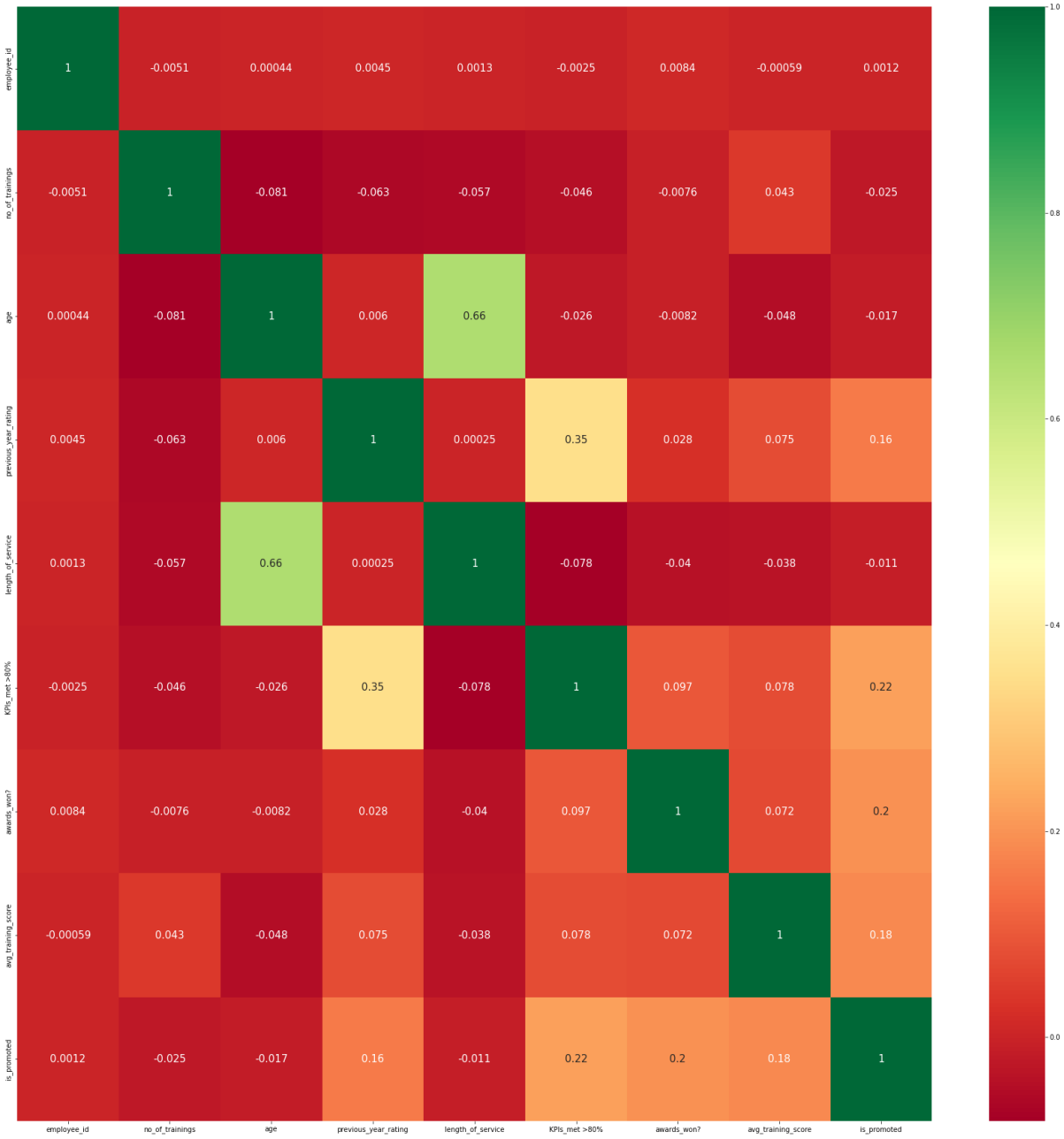
PAIRPLOT ON TRAIN DATA



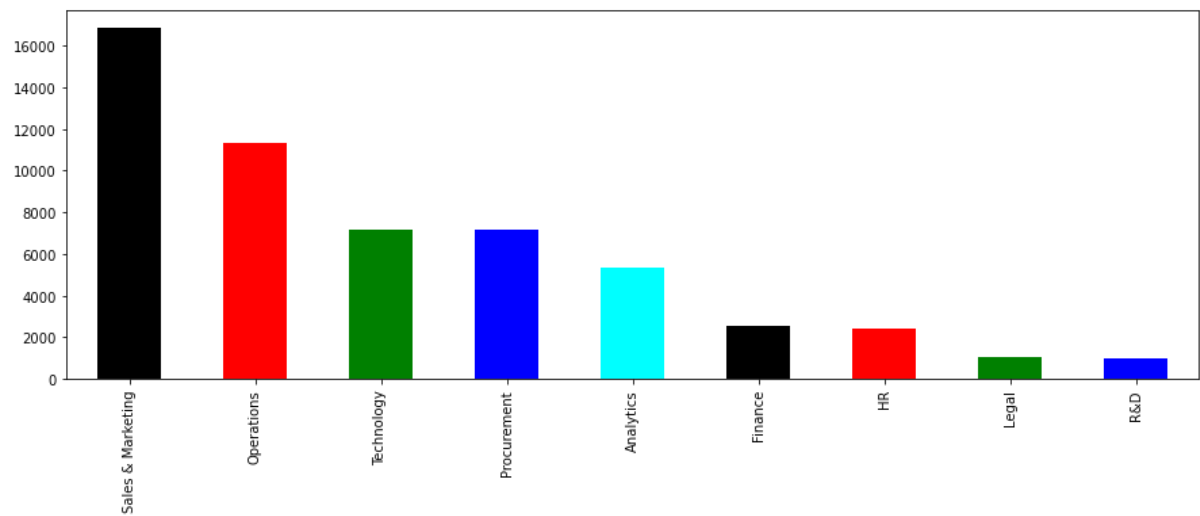
PLOT OF DISTRIBUTION OF DATA FOR EACH FEATURE OF EMPLOYEE



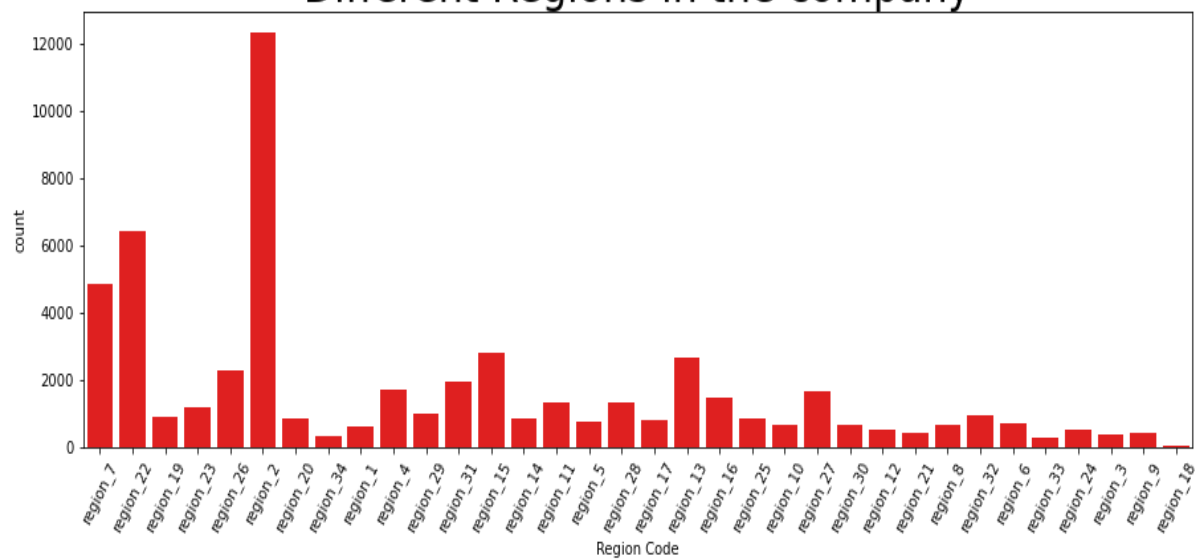
HEAT MAP FOR CORRELATION MATRIX



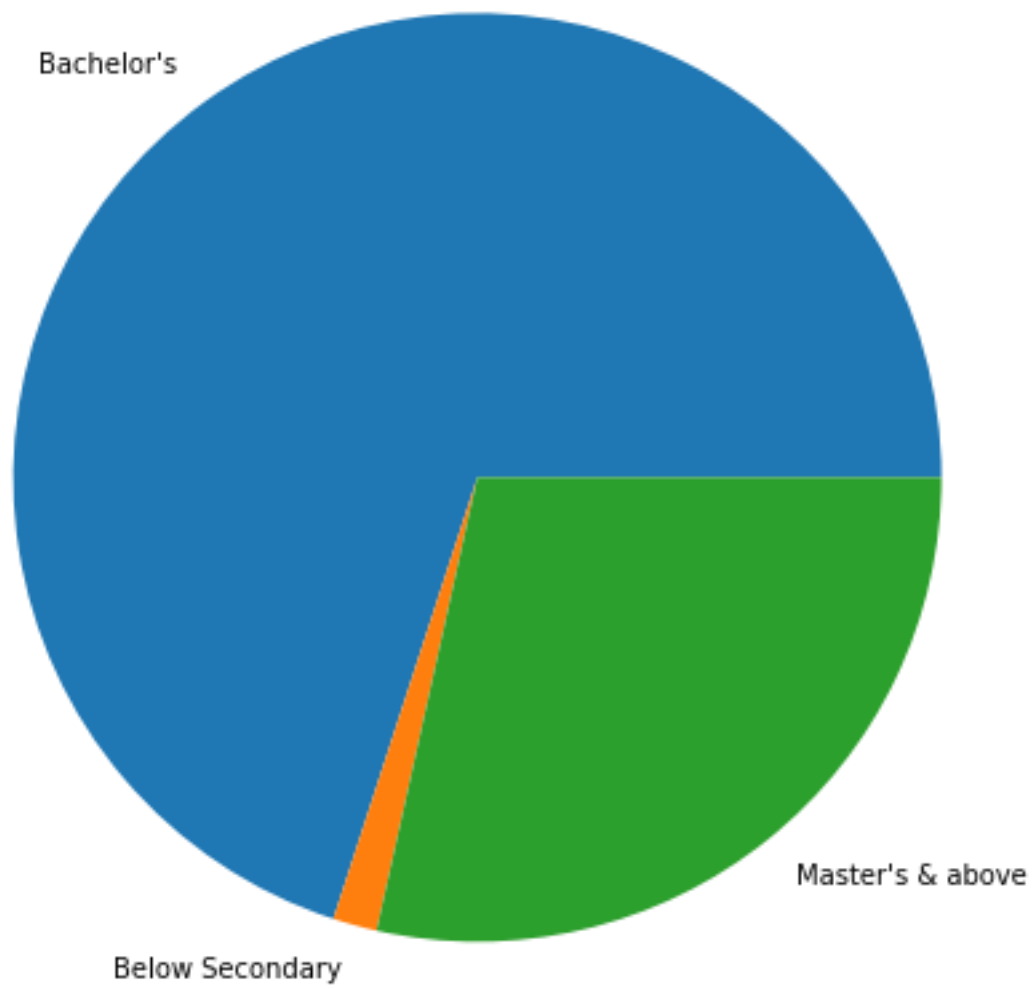
PLOT OF DIFFERENT GROUPS IN THE DATASET



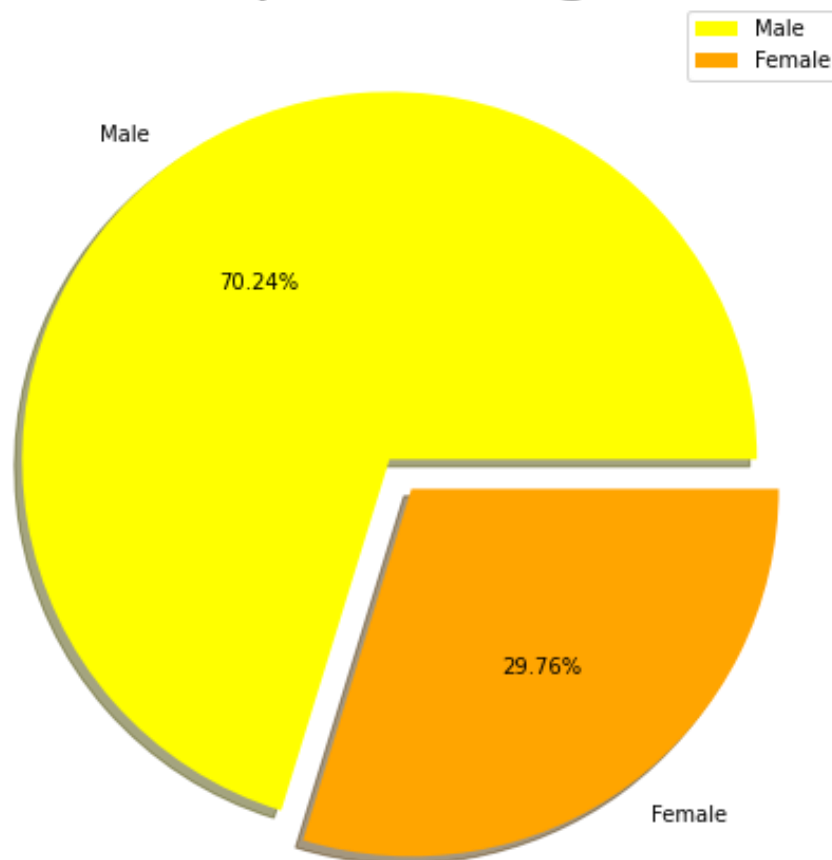
Different Regions in the company



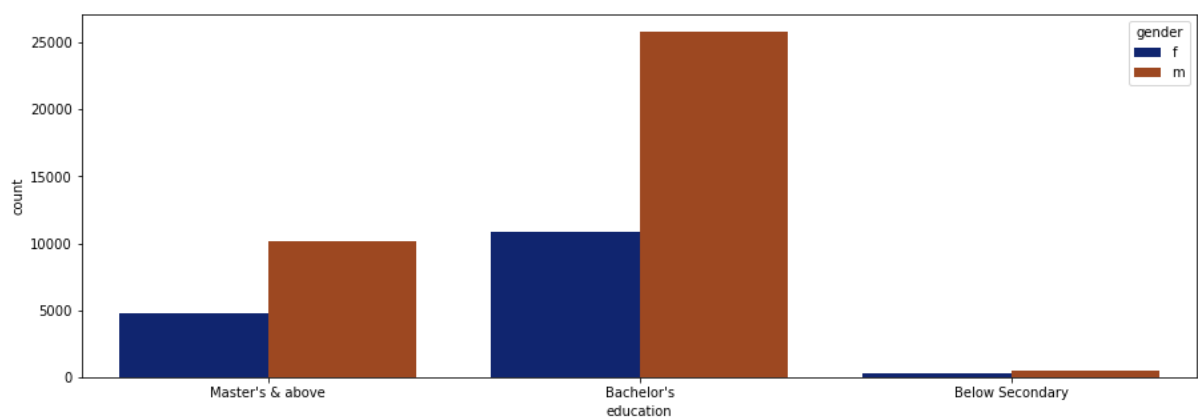
Pie Chart of different types of education

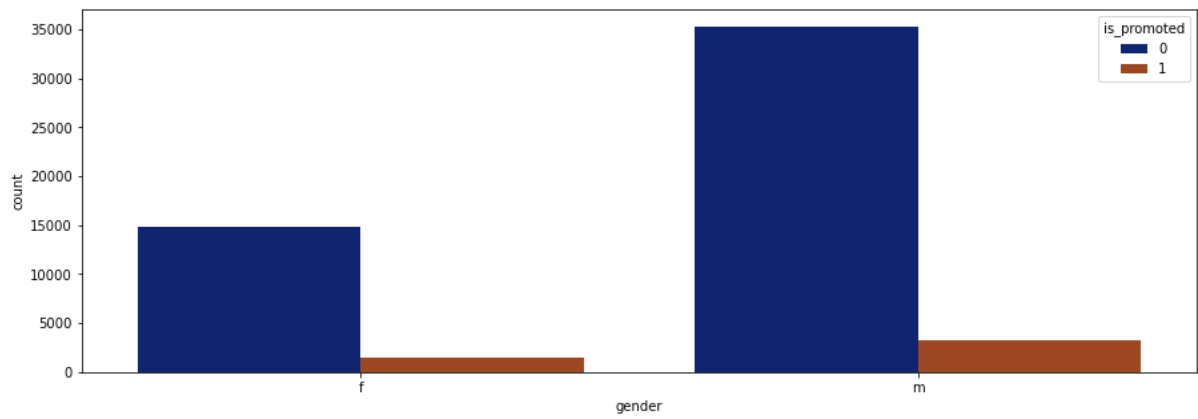


A Pie Chart Representing GenderGap

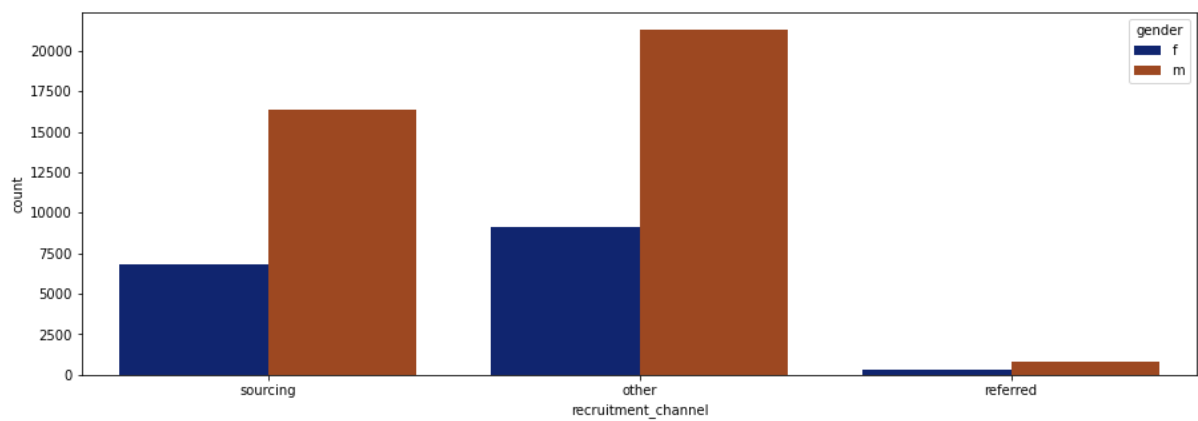


COMPARISON OF PROMOTED GENDER (MALE AND FEMALE)

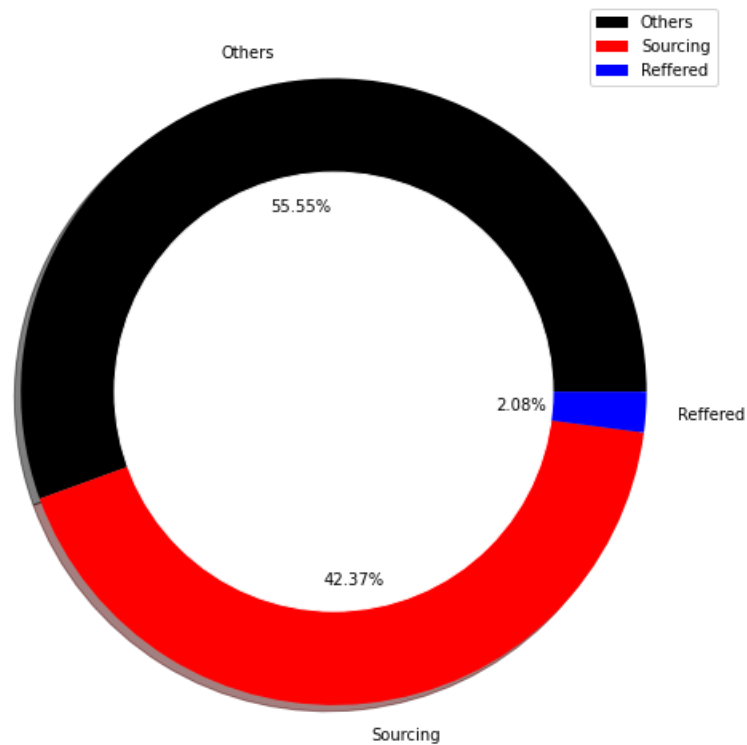




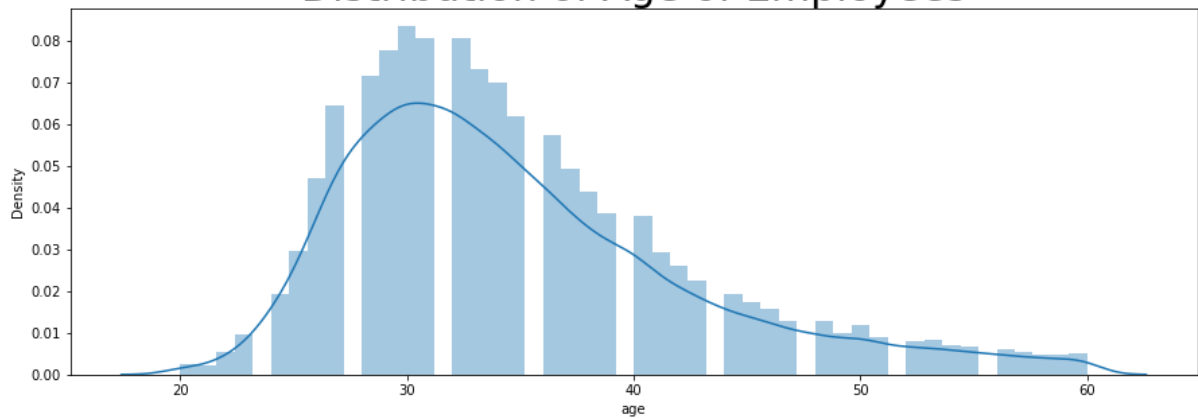
COMPARISON OF REQUIRED GENDER (MALE AND FEMALE)



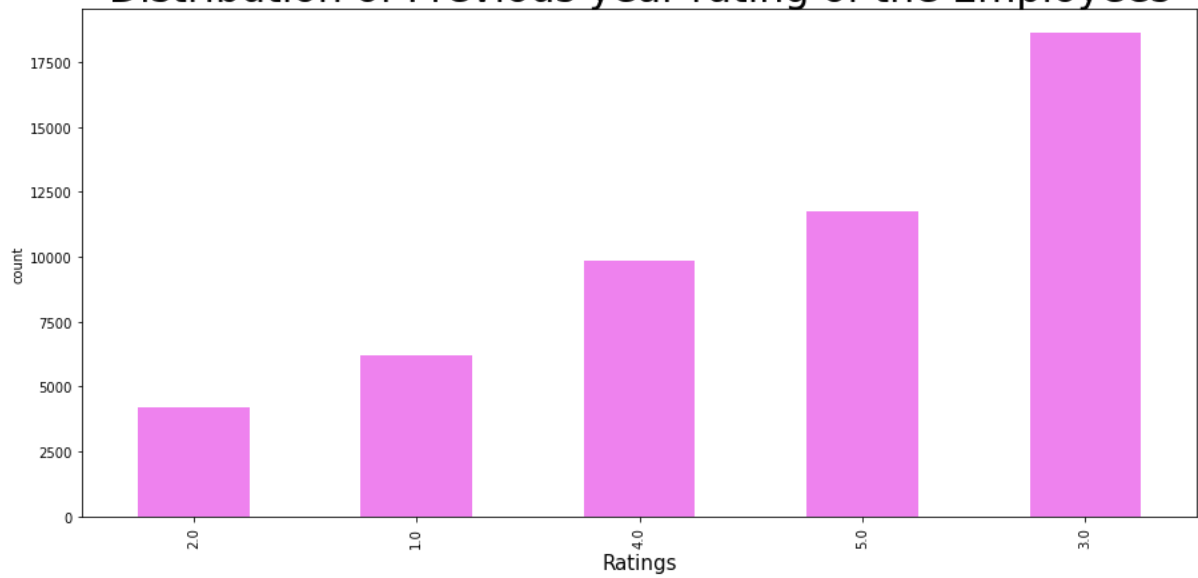
Showing share of different Recruitment Channels



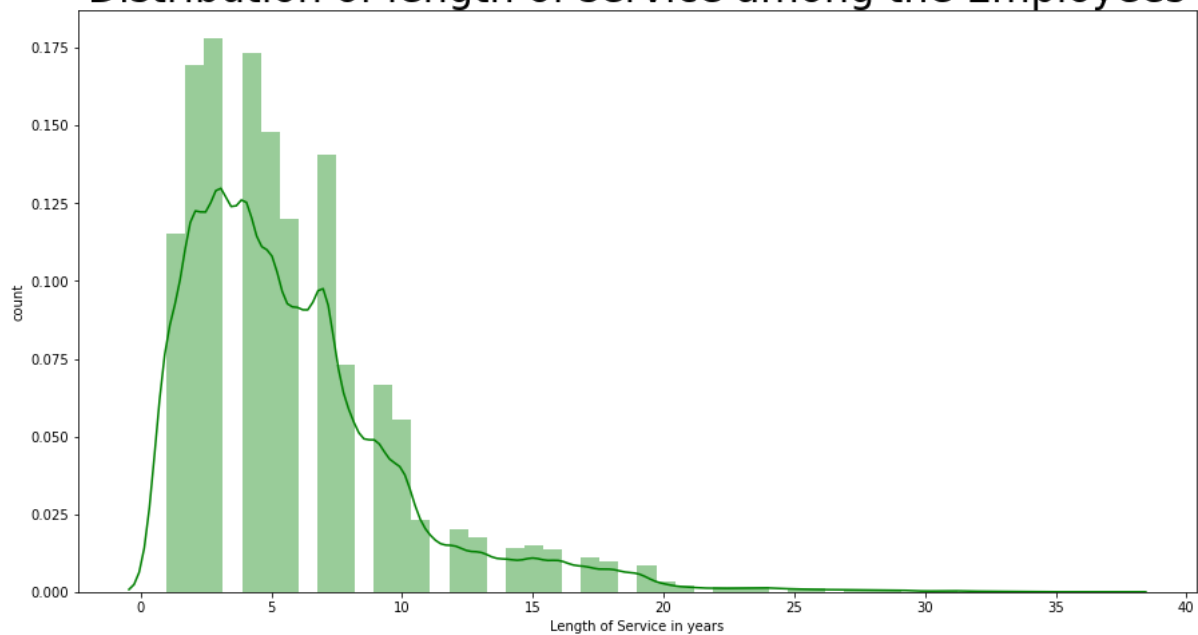
Distribution of Age of Employees



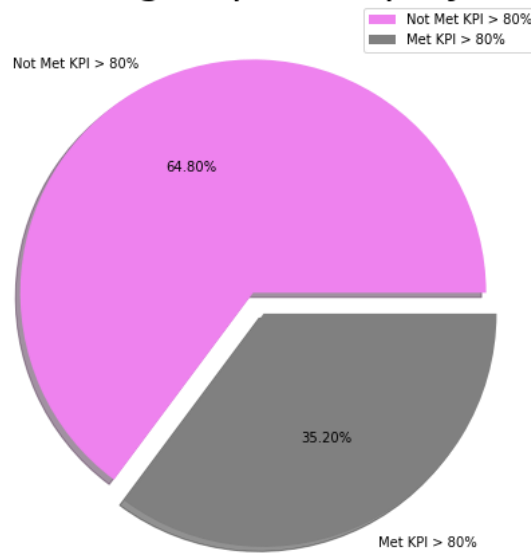
Distribution of Previous year rating of the Employees



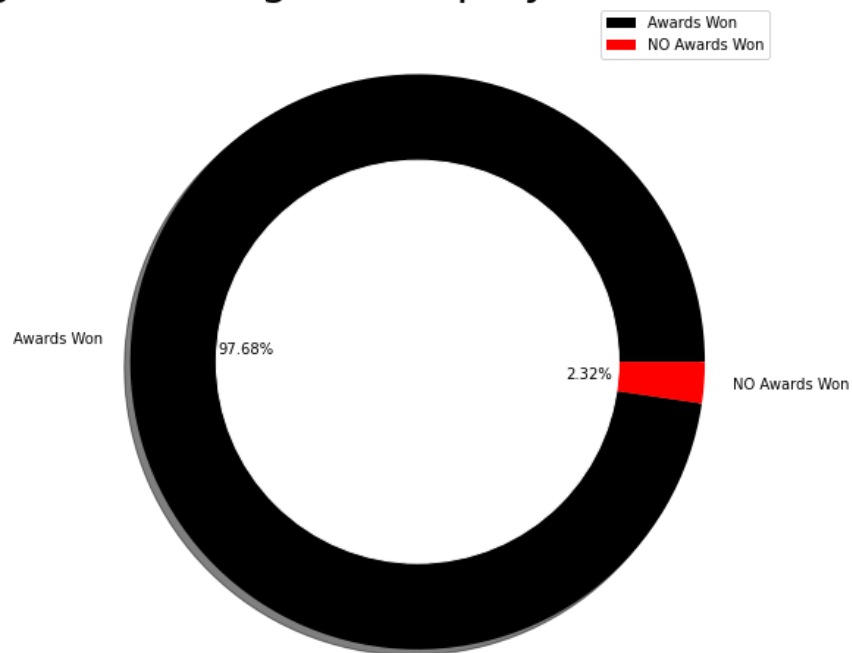
Distribution of length of service among the Employees



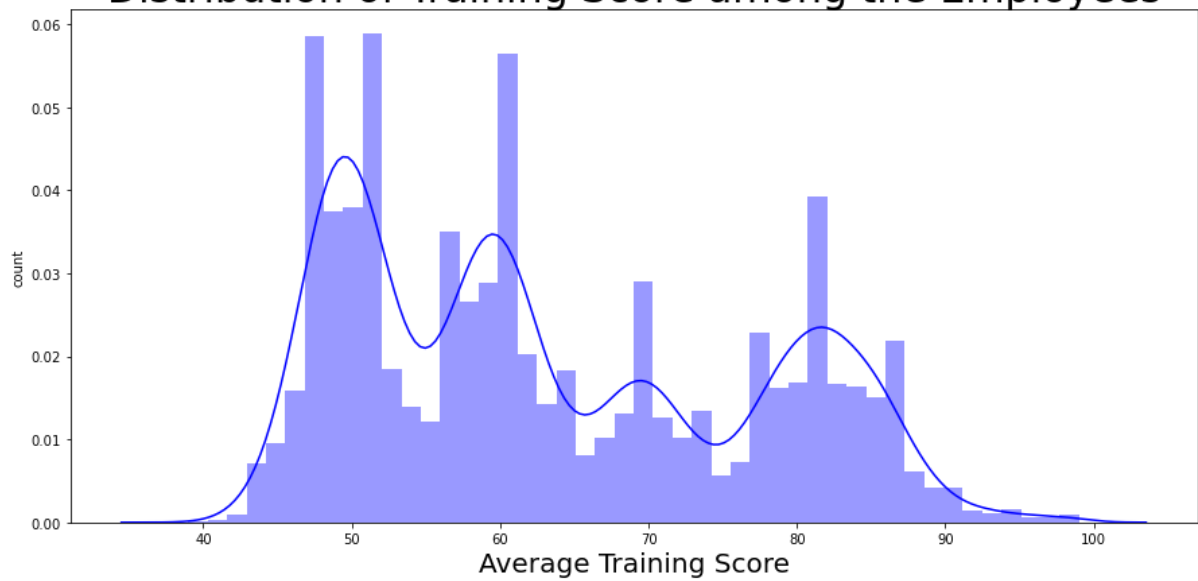
A Pie Chart Representing Gap in Employees in terms of KPI



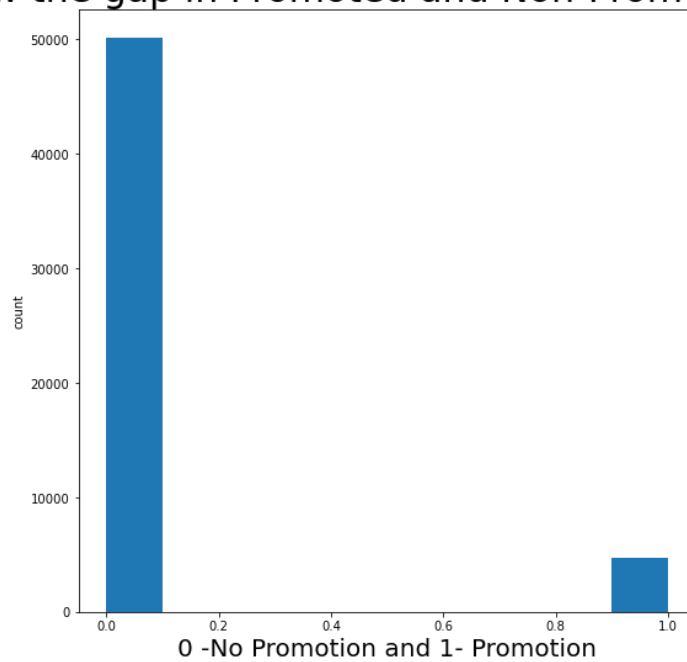
Showing a Percentage of employees who won awards



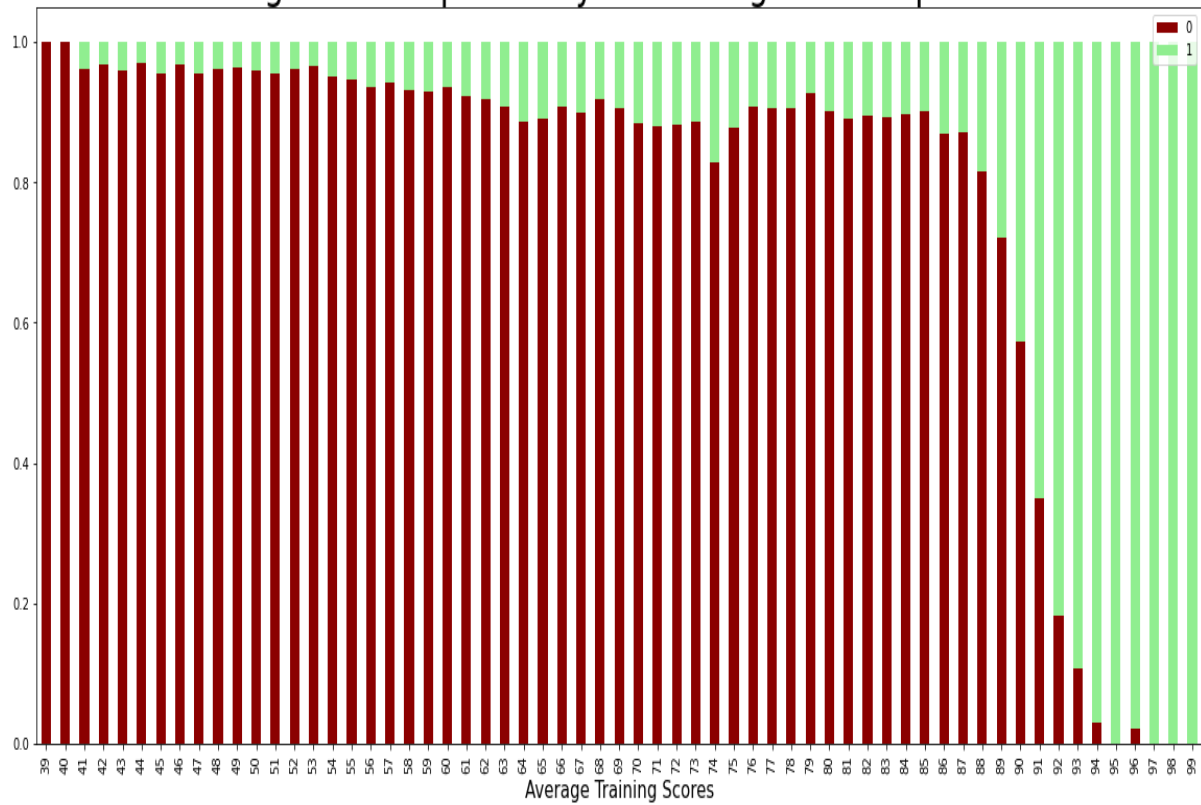
Distribution of Training Score among the Employees



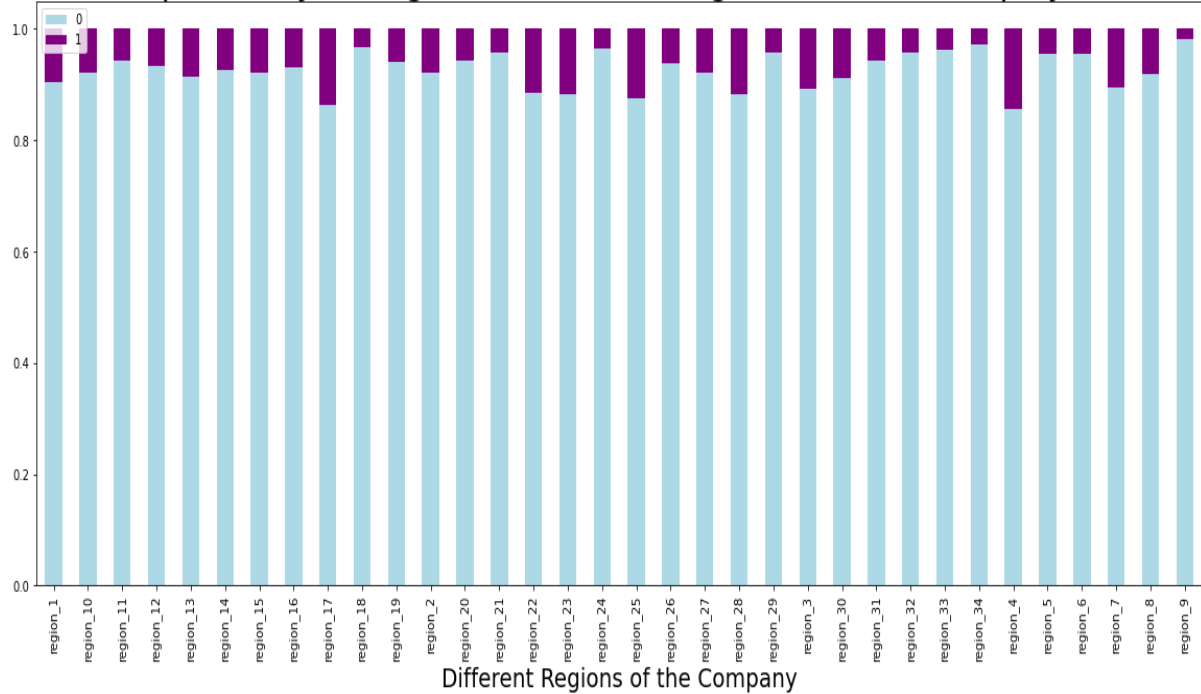
Plot to show the gap in Promoted and Non-Promoted Employees



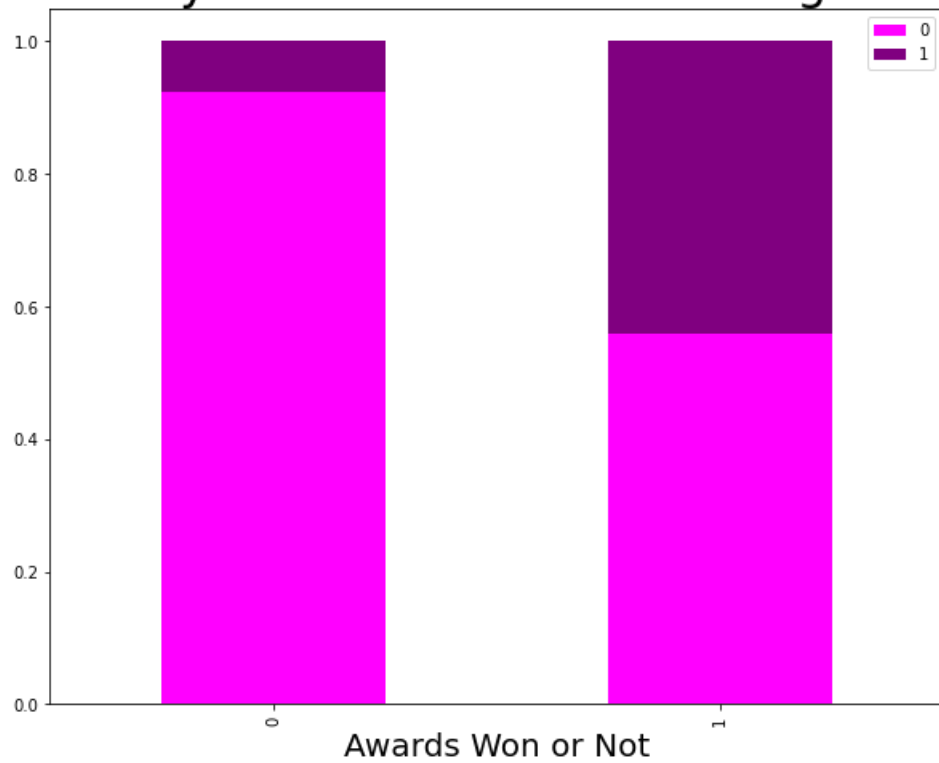
Looking at the Dependency of Training Score in promotion



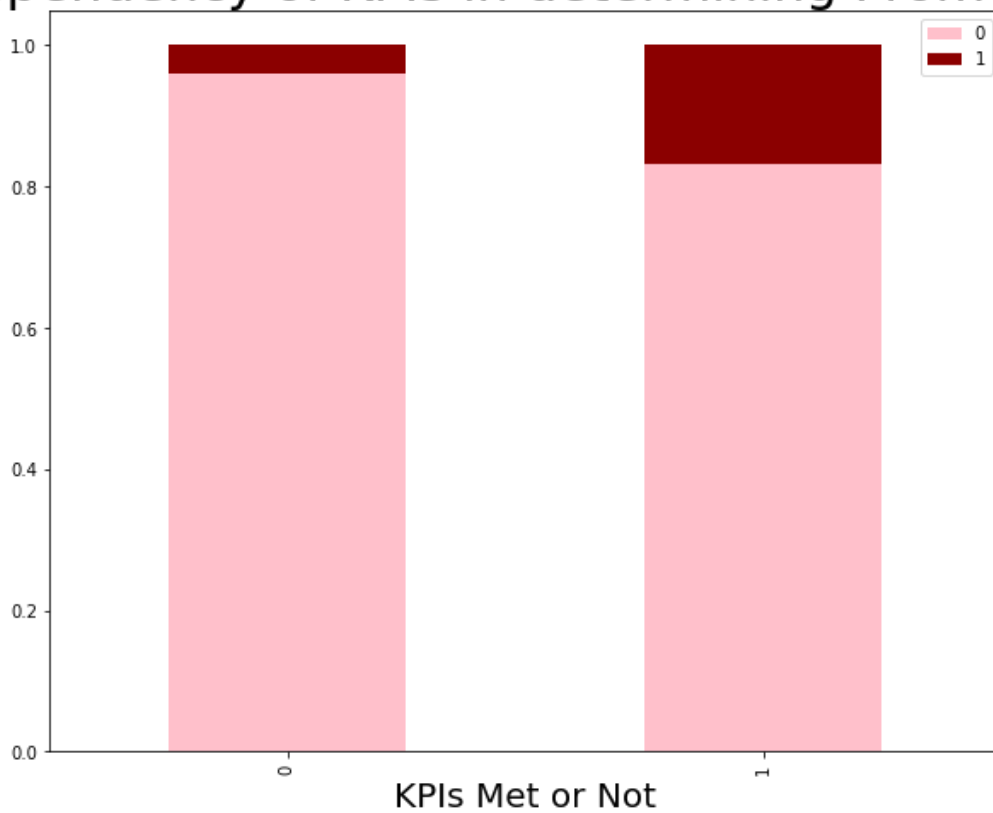
Dependency of Regions in determining Promotion of Employees



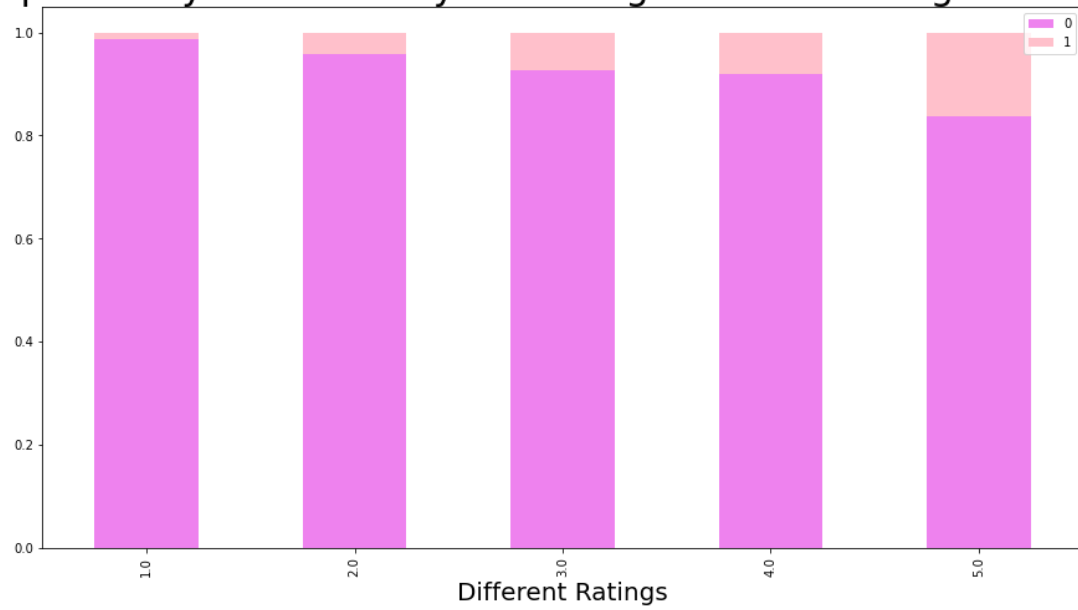
Dependency of Awards in determining Promotion



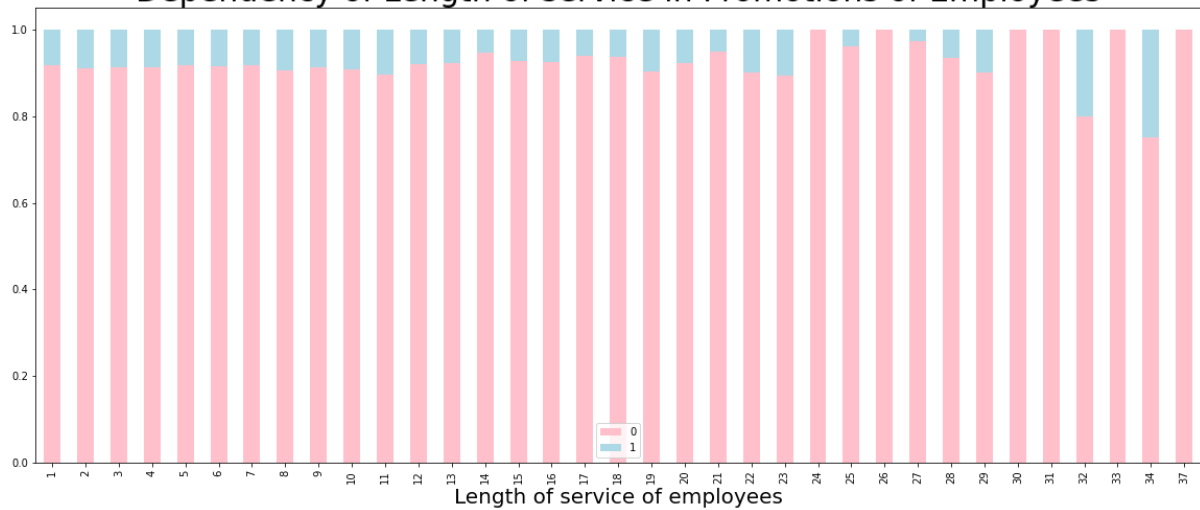
Dependency of KPIs in determining Promotion



Dependency of Previous year Ratings in determining Promotion



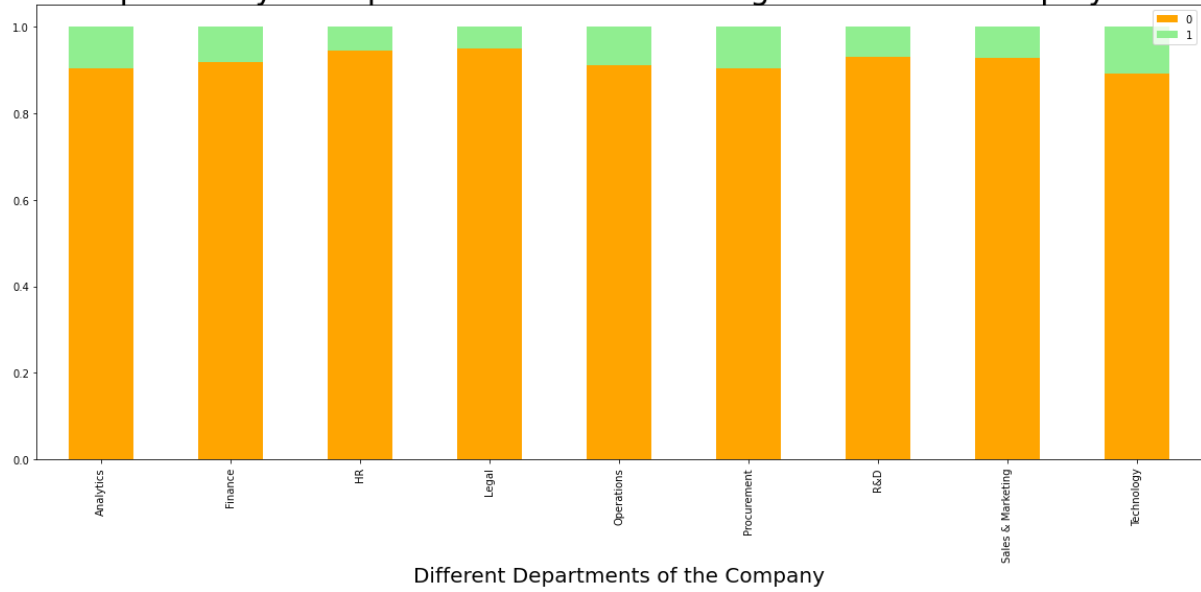
Dependency of Length of service in Promotions of Employees



Dependency of Age in determining Promotion of Employees



Dependency of Departments in determining Promotion of Employees



Dependency of Genders in determining Promotion of Employees

