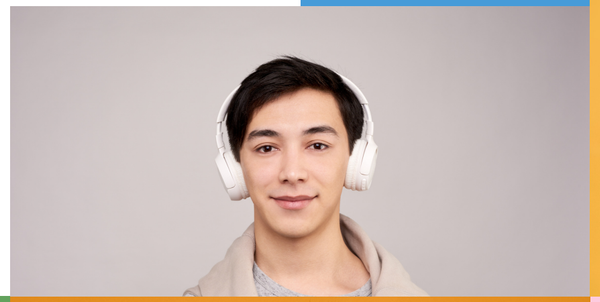
Leave Or Stay



ML Project
- By Ryan -

Turnover Problem!

Every company has a problem with employees leaving

If we have data for this problem. Can we predict it ?

Employee Turnover Data

https://www.kaggle.com/datasets/tejashvi14/employee-future-prediction

Scope

- Employee in 3 City in India (Banglalore, Pune, New Delhi)
- Data Only in Year 2012 - 2018
- Total Data 4653

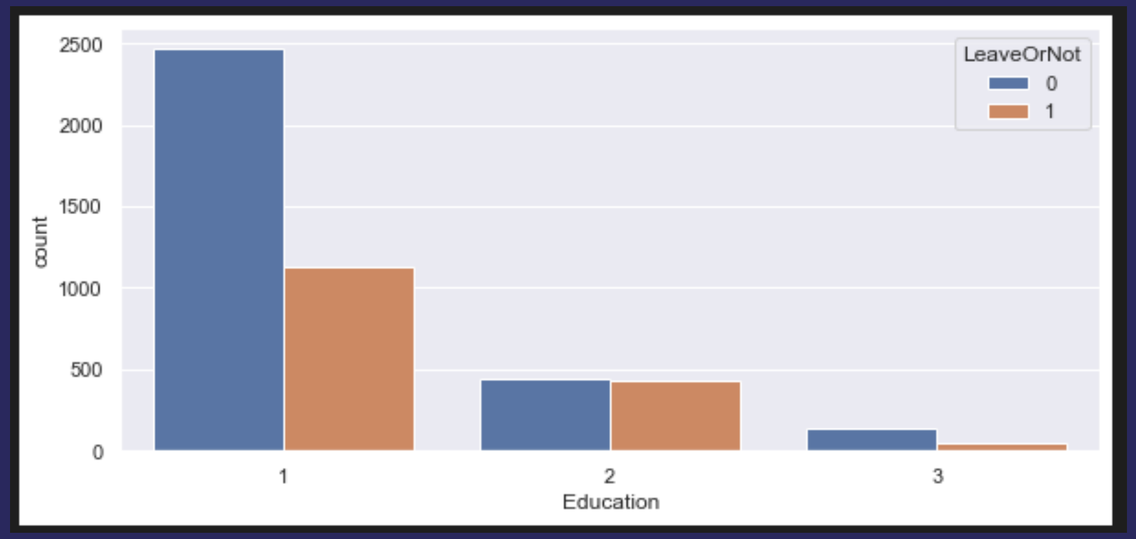
Feature

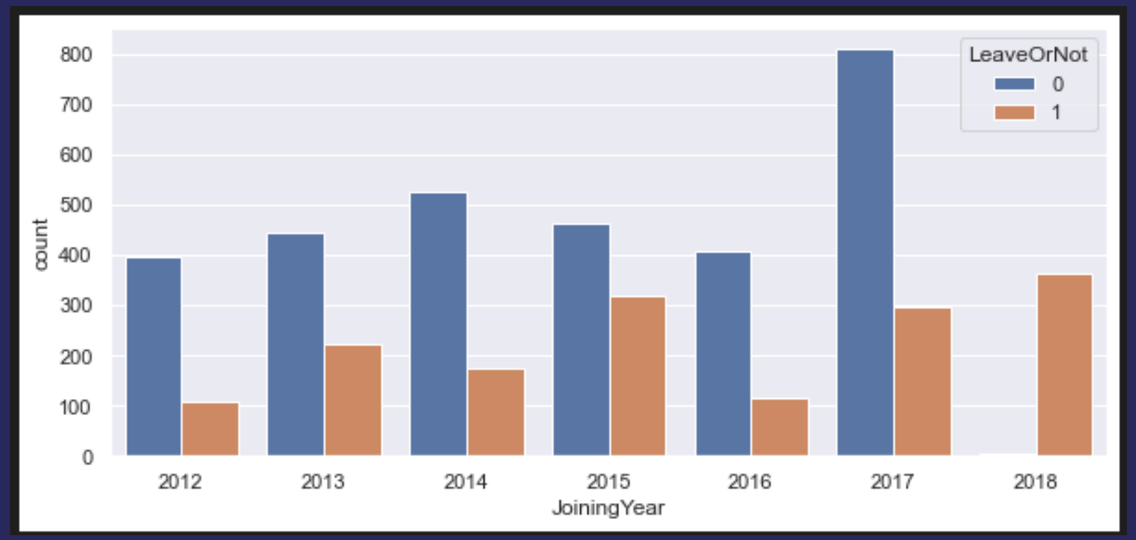
- Education
- Join Year
- City
- Payment Tier
- Age
- Gender
- Ever Benched
- Experience Work

Target

Modeling 'Leave Or Stay' Prediction

Exploring Data



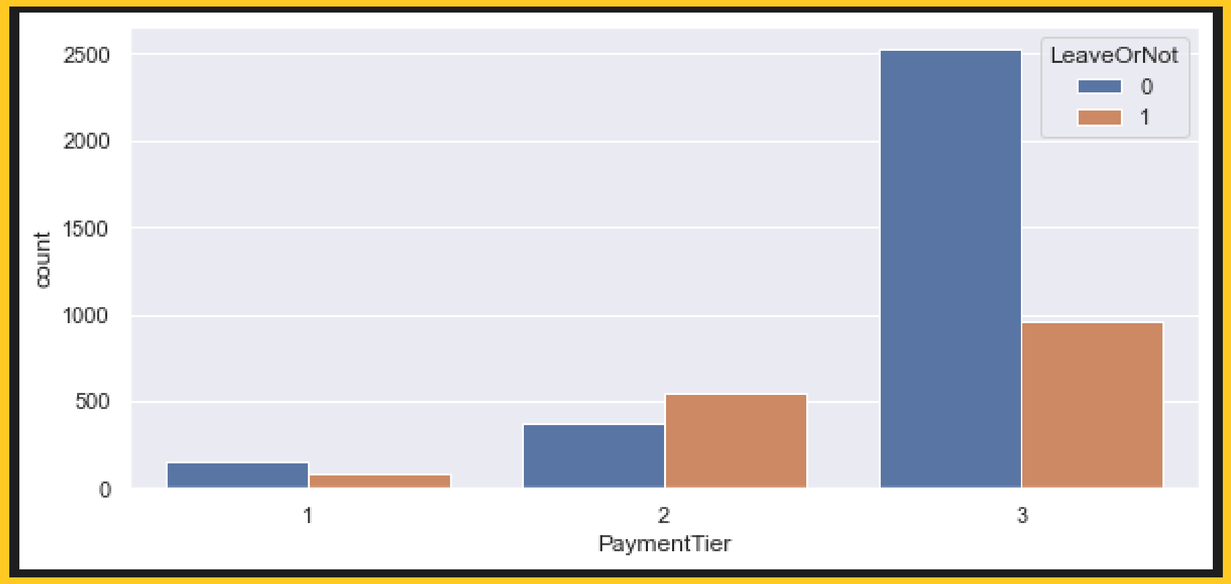


49% of master degree left

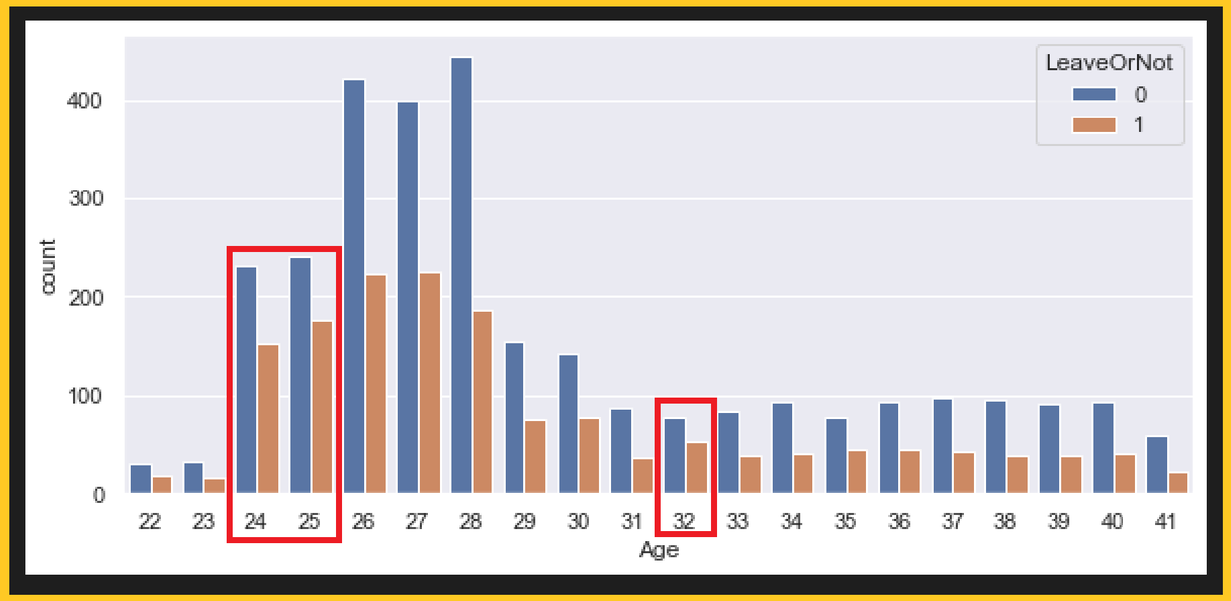
99% of people joined in 2018 left

Exploring Data

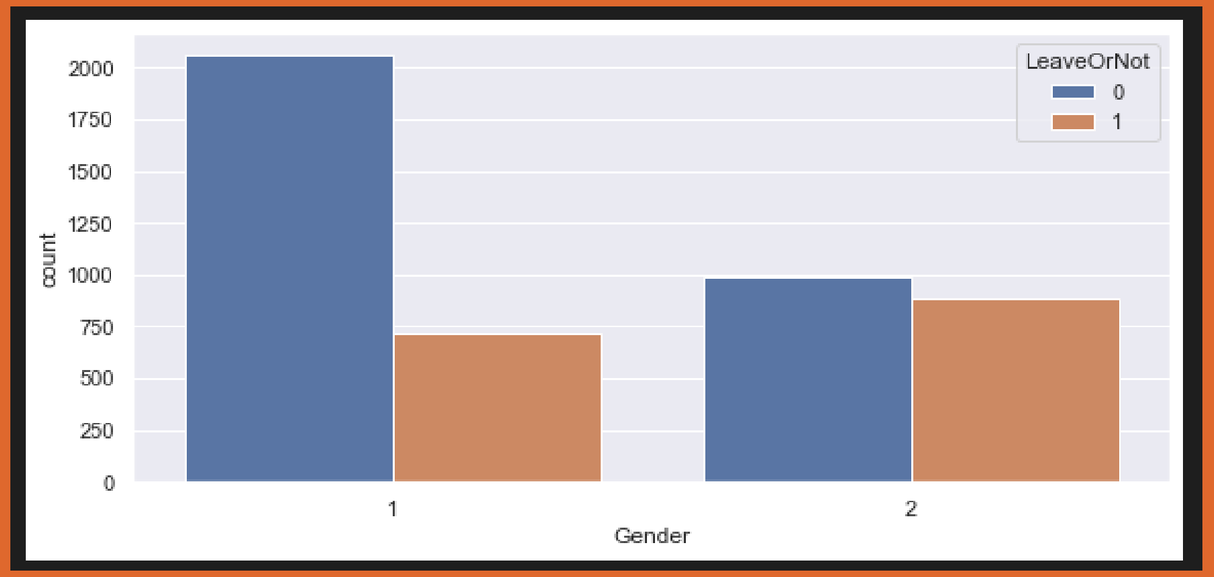
60% of people with PaymentTier Grade 2 left

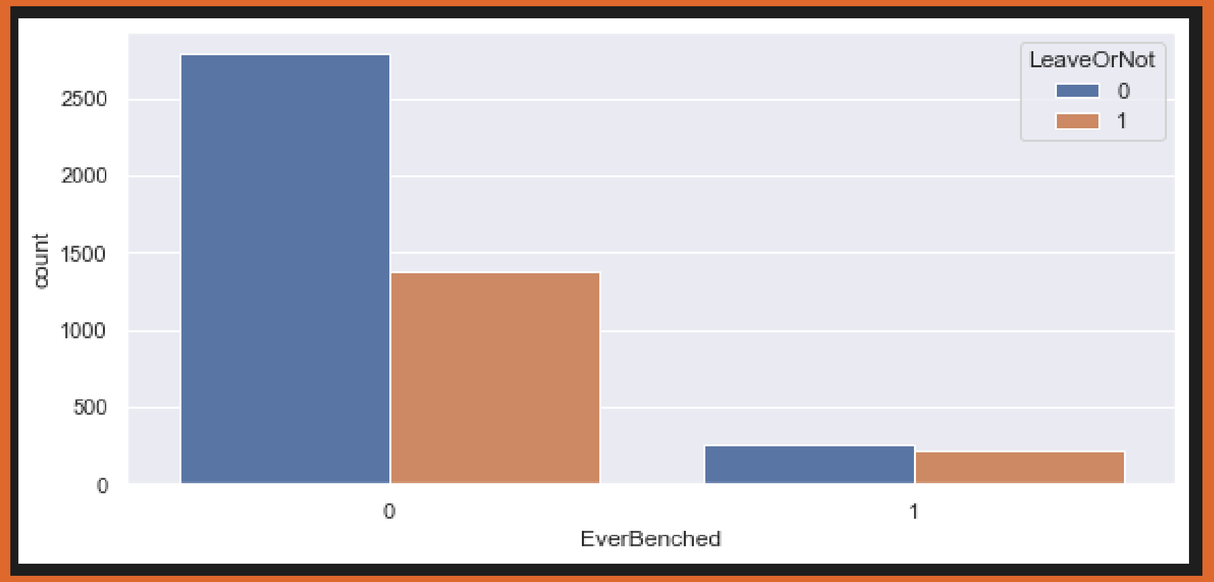


more than 40% of people in age 24,25, and 32 left



Exploring Data





47% of female employees left

45% of people who were ever Benched left

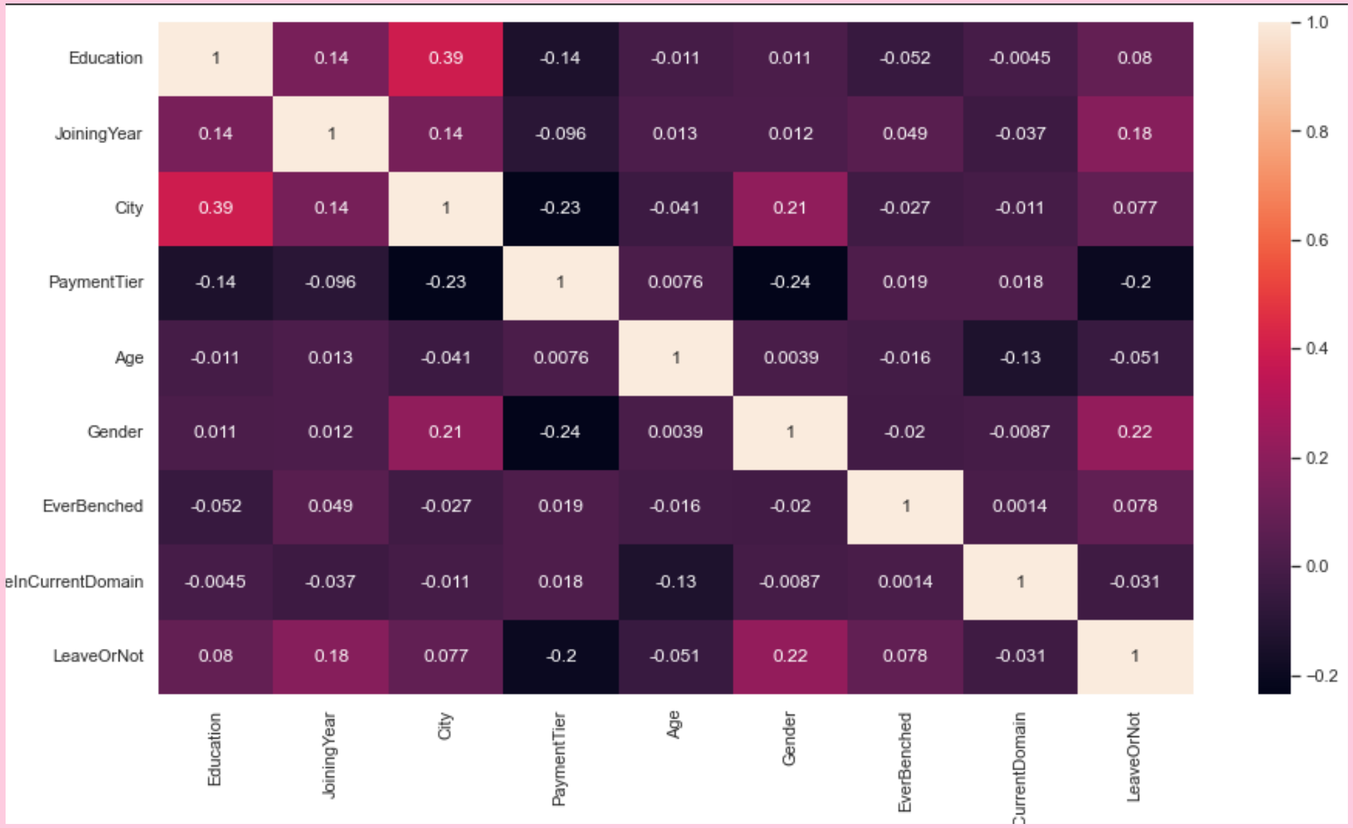
Feature Engineering

Feature Importances

	X	importance
0	Education	0.115956
1	JoiningYear	0.374869
2	City	0.085712
3	PaymentTier	0.194077
4	Age	0.043583
5	Gender	0.150517
6	EverBenched	0.009892
7	ExperienceInCurrentDomain	0.025394

Distribution Data

Correlation Each Feature



Modeling

SVM

```
from sklearn.svm import SVC

svm = SVC(kernel='rbf', C=1, probability=True)

svm.fit(X_train_scaled,y_train_samp)

y_pred_train_svm = svm.predict(X_train_scaled)
y_pred_test_svm = svm.predict(X_test_scaled)

y_proba_train_svm = svm.predict_proba(X_train_scaled)
y_proba_test_svm = svm.predict_proba(X_test_scaled)
```

✓ 1.9s

Random Forest

```
from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=100,criterion='gini',
        max_depth=5, min_samples_split=2, min_samples_leaf=1)

RF.fit(X_train_samp,y_train_samp)

y_pred_train_rf = RF.predict(X_train_samp)
y_pred_test_rf = RF.predict(X_test)

y_proba_train_rf = RF.predict_proba(X_train_samp)
y_proba_test_rf = RF.predict_proba(X_test)
```

✓ 0.4s

Evaluation

SVM

```
Training
              precision    recall  f1-score   support

           0       0.75      0.84      0.79      1280
           1       0.82      0.72      0.77      1280

    accuracy                           0.78      2560
   macro avg       0.78      0.78      0.78      2560
weighted avg       0.78      0.78      0.78      2560

ROC AUC Score: 0.8485964965820314

============================

Test
              precision    recall  f1-score   support

           0       0.85      0.81      0.83       611
           1       0.67      0.73      0.70       320

    accuracy                           0.78       931
   macro avg       0.76      0.77      0.77       931
weighted avg       0.79      0.78      0.79       931

ROC AUC Score: 0.81649191898527
```

Random Forest

```
Training
              precision    recall  f1-score   support

           0       0.77      0.88      0.82      1280
           1       0.86      0.74      0.79      1280

    accuracy                           0.81      2560
   macro avg       0.81      0.81      0.81      2560
weighted avg       0.81      0.81      0.81      2560

ROC AUC Score: 0.8691412353515625

============================

Test
              precision    recall  f1-score   support

           0       0.85      0.86      0.86       611
           1       0.73      0.70      0.72       320

    accuracy                           0.81       931
   macro avg       0.79      0.78      0.79       931
weighted avg       0.81      0.81      0.81       931

ROC AUC Score: 0.8439494680851065
```

Conclusion

- Accuracy Value of the Model obtained <= 85%
- The best model when viewed from the ROC AUC Score is Random Forest