Data project portfolio

Employee Retention

By Metee Srisawtvichai

Today's Agenda

1. Business Goal
2. Data understanding and EDA
3. Modeling
4. Recommendations

Business Goal

1.Find Current Employee Turnover Factors

2.Predicting employees about to leave

3. Recomendations

Data understanding & EDA

Features

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14999 entries, 0 to 14998
Data columns (total 10 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   satisfaction_level     14999 non-null  float64
 1   last_evaluation        14999 non-null  float64
 2   number_project         14999 non-null  int64
 3   average_montly_hours   14999 non-null  int64
 4   time_spend_company     14999 non-null  int64
 5   Work_accident          14999 non-null  int64
 6   left                   14999 non-null  int64
 7   promotion_last_5years  14999 non-null  int64
 8   Department             14999 non-null  object
 9   salary                 14999 non-null  object
dtypes: float64(2), int64(6), object(2)
```

Total 14,998 rows

10 Features

- 2 decimal
- 6 Integer
- 2 String

Check null and duplicated

```
#check null value
df_raw.isna().sum()
```

```
satisfaction_level        0
last_evaluation           0
number_project            0
average_montly_hours      0
time_spend_company        0
Work_accident             0
left                      0
promotion_last_5years     0
Department                0
salary                    0
dtype: int64
```

```
# Check for duplicates
df_raw.duplicated().sum()
```

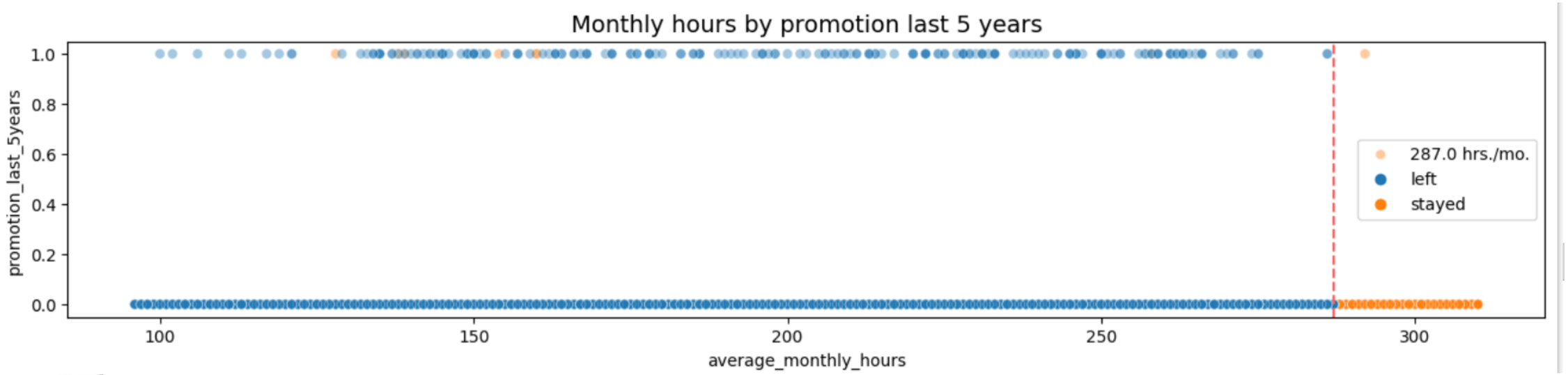
```
3008
```

```
# Drop duplicates and save resulting dataframe in a new variable as needed
df_clean = df_raw.drop_duplicates(keep='first')
```

Not found null value

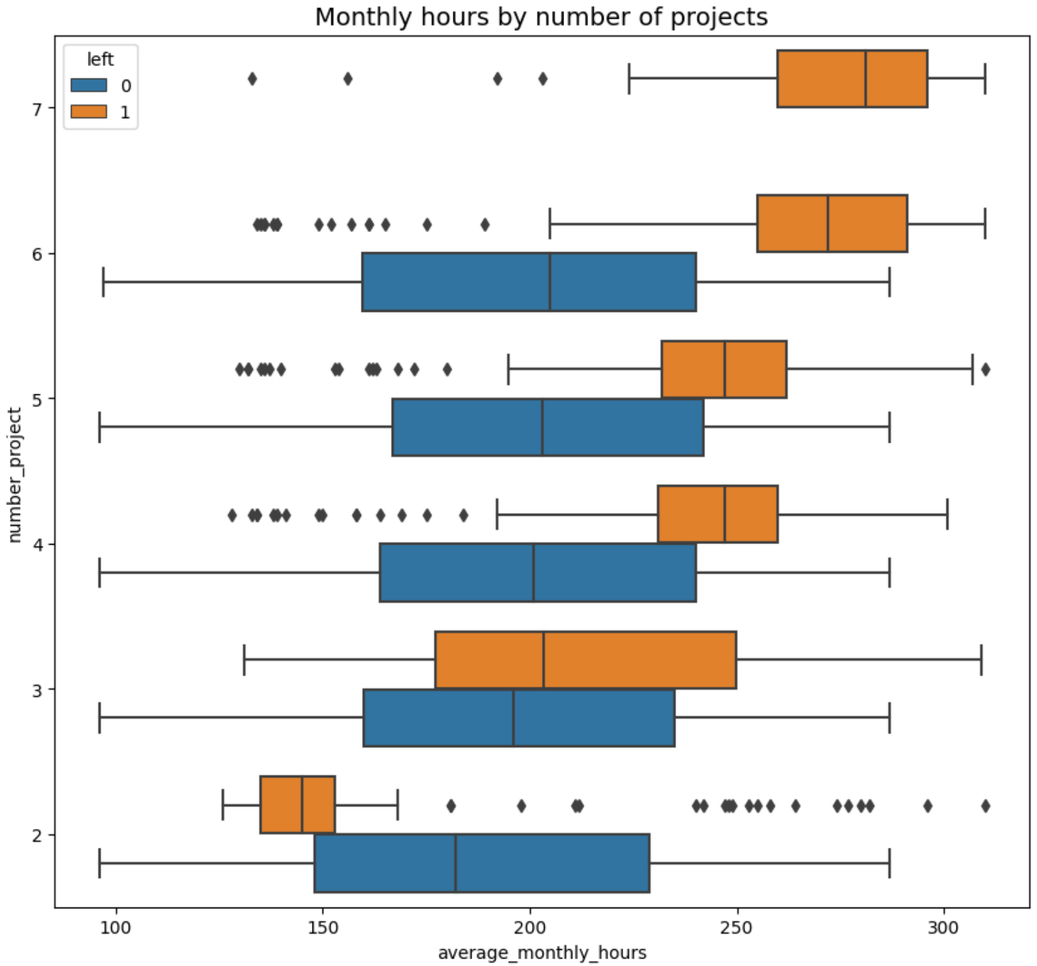
Found 3,000 duplicated rows
Solution remove duplicated rows

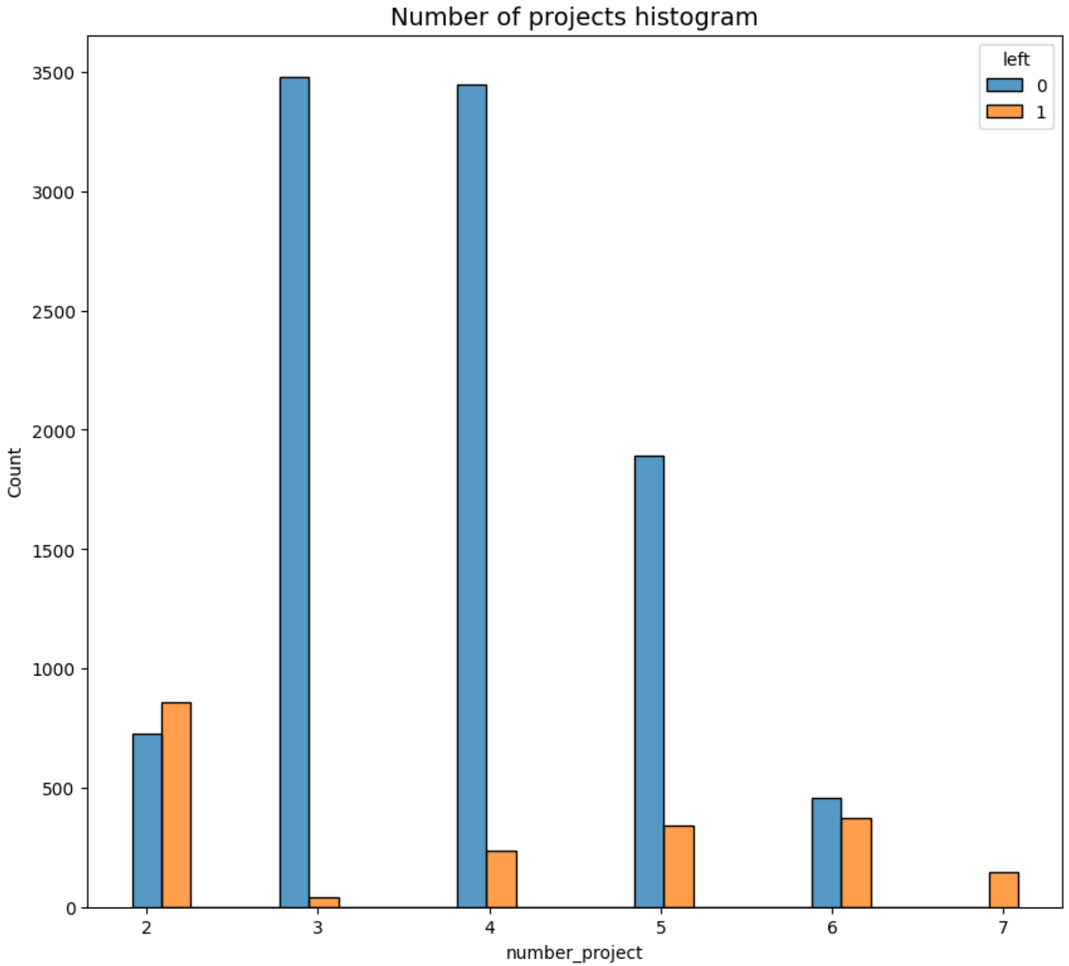
EDA Monthly hours & Promoted



It was found that many employees resigned from the company. If not promoted and the average monthly working time is more than 287 hours.

EDA Number of projects

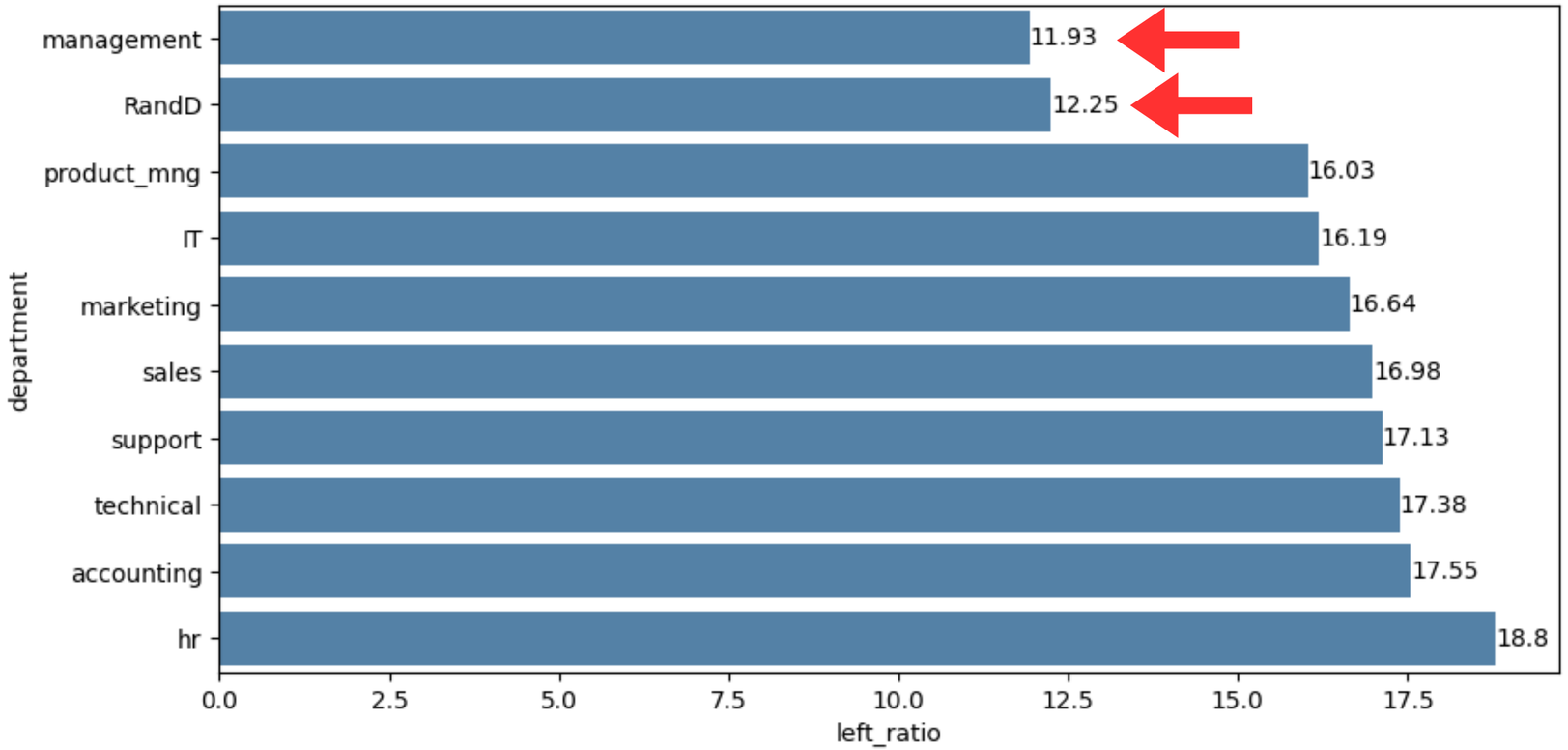




EDA Number of projects (cont)

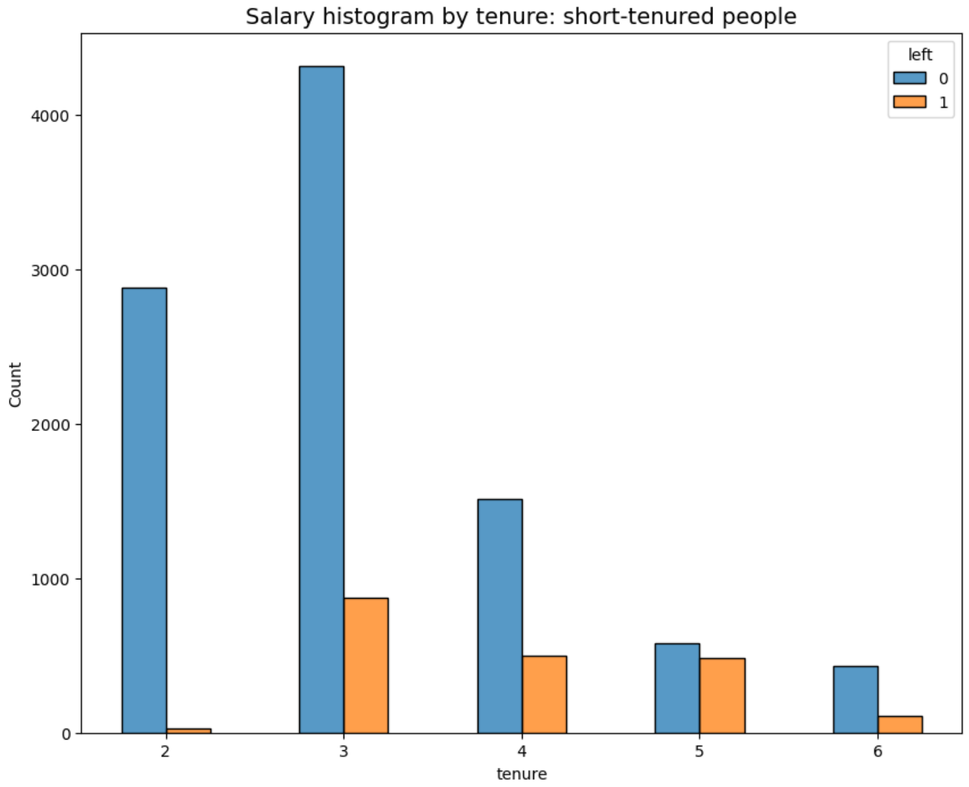
It was found that employees who resigned from the company had the number of projects they were responsible for equal to 7 projects, followed by 1 project.

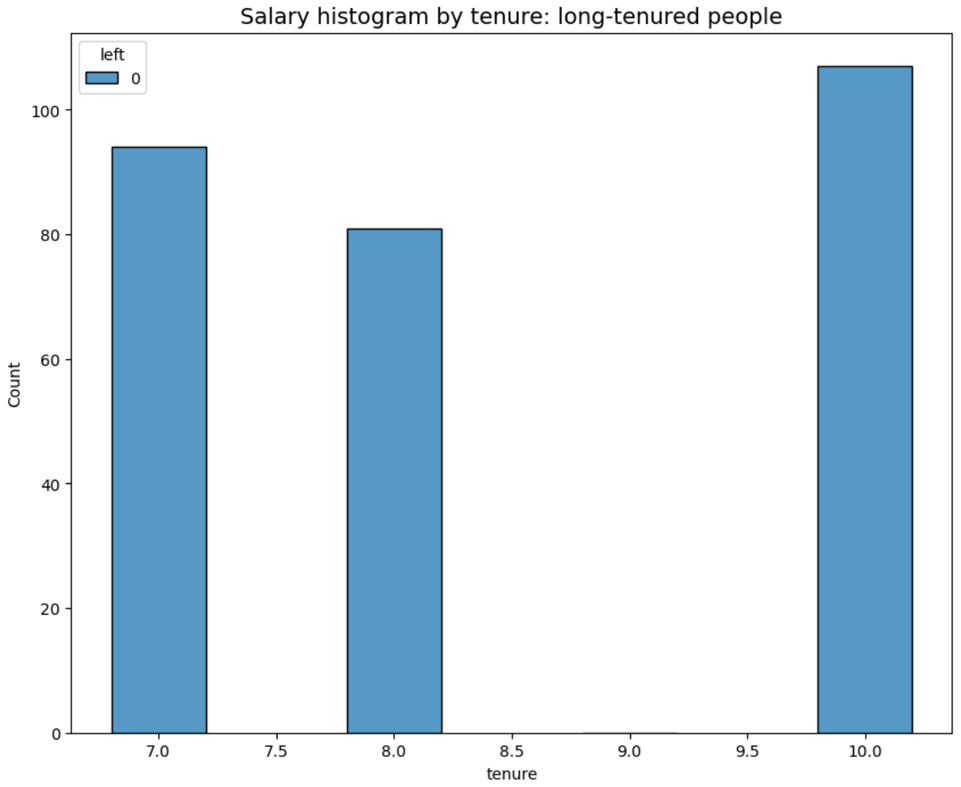
EDA Department



Found management and RandD have less left ratio

EDA Tenure

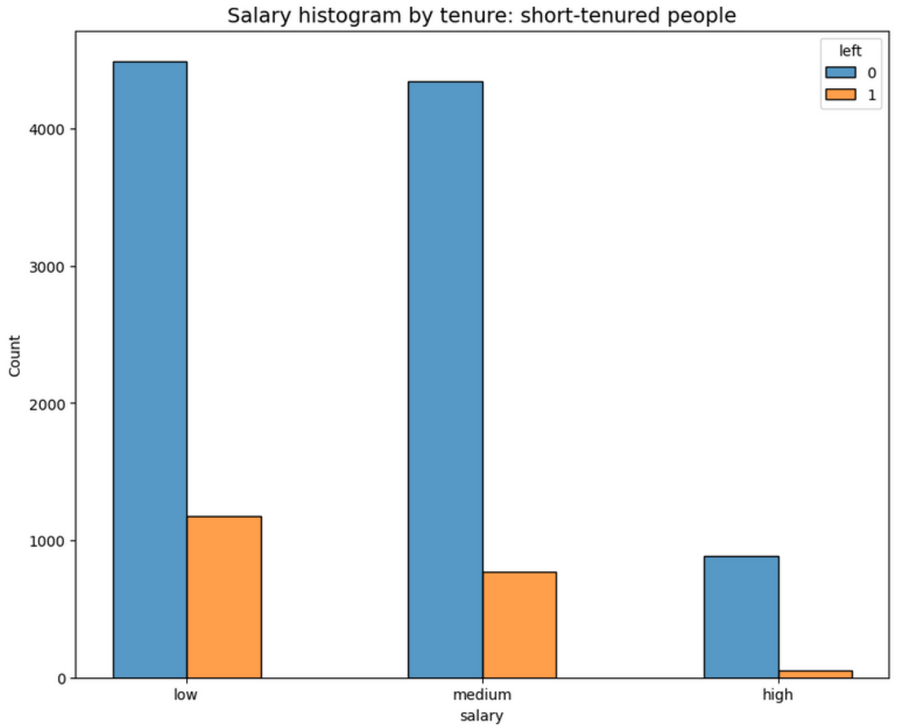


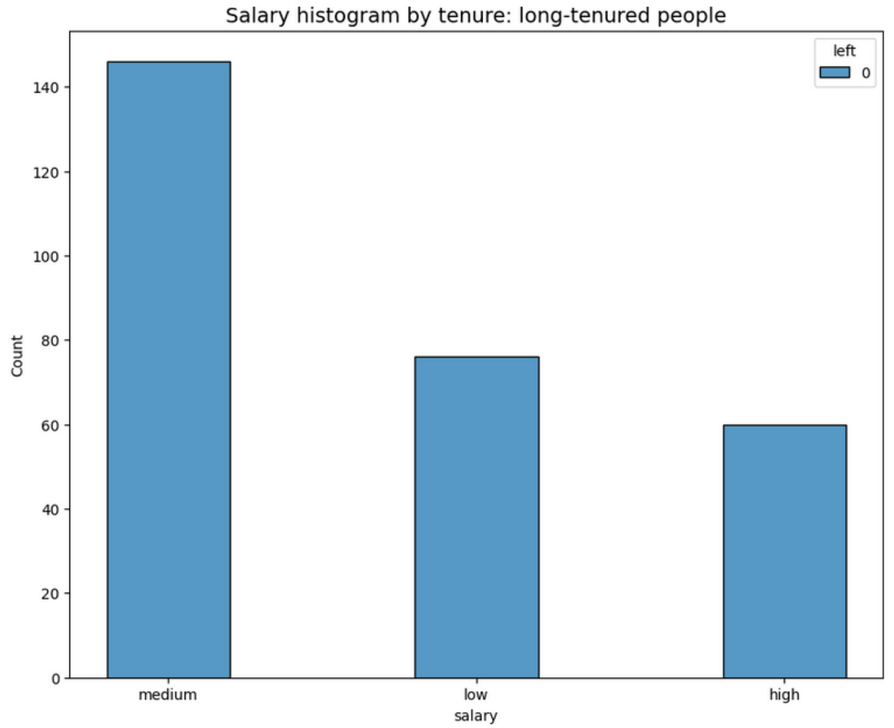


Found number of Employee leff ratio have Tenure 5>4>6>3>2

Found Tenure >=7 not have employee left

EDA Salary





Found Salary impact Employee short tenure (< 7 years)left
Salary low>medium>high

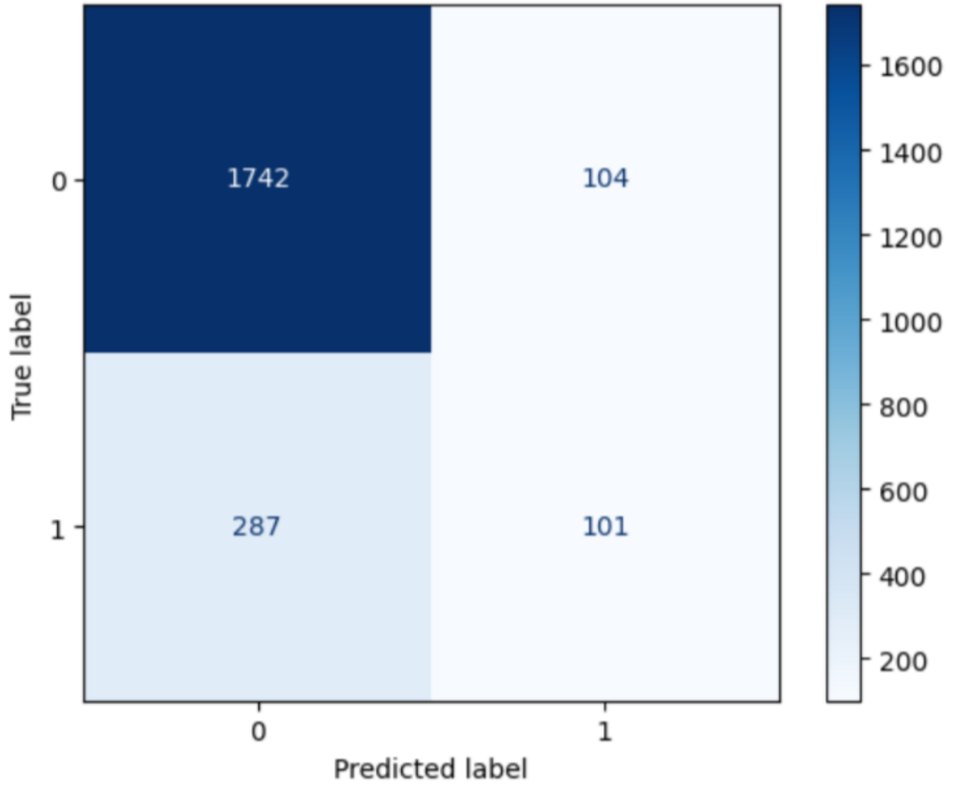
Found Salary not impact Employee long tenure (>=7 years)employee left

EDA Salary & Tenure

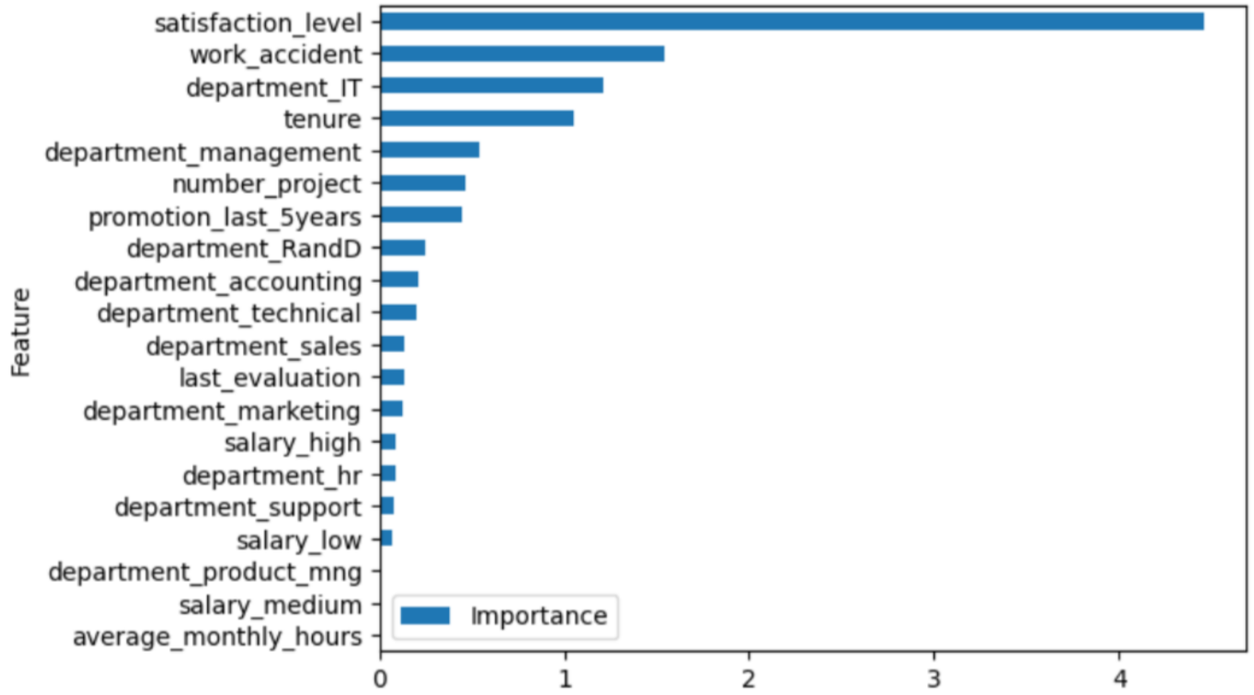
Found Salary = low and tenure = 5
100% left

Modeling for Prediction

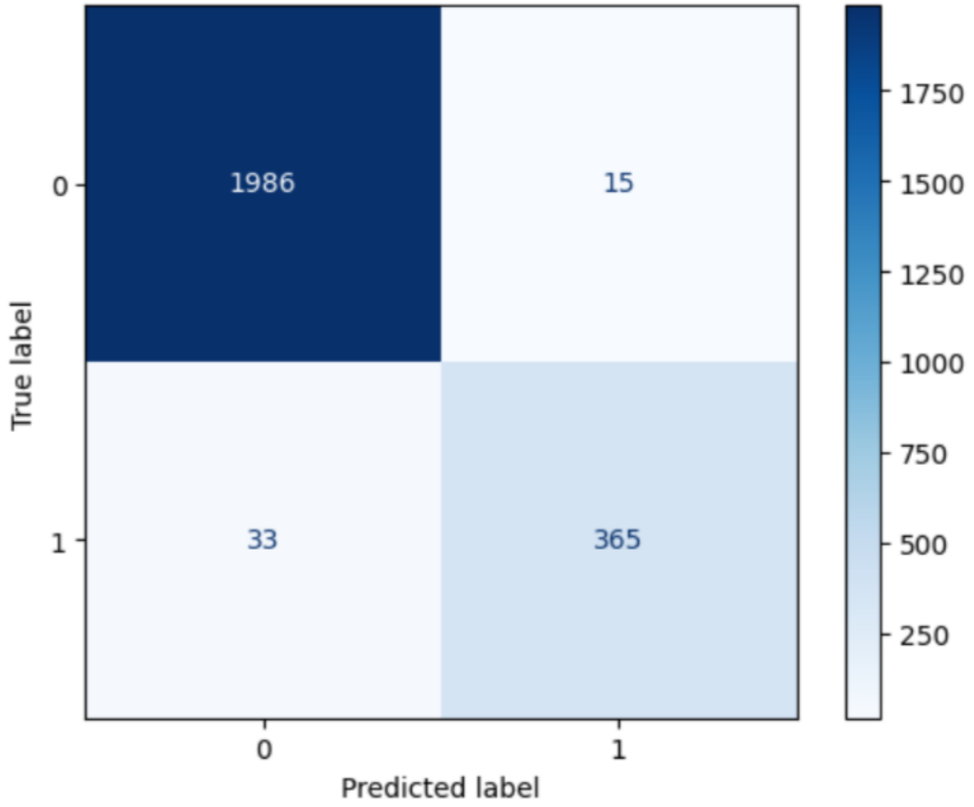
Logistic Regression



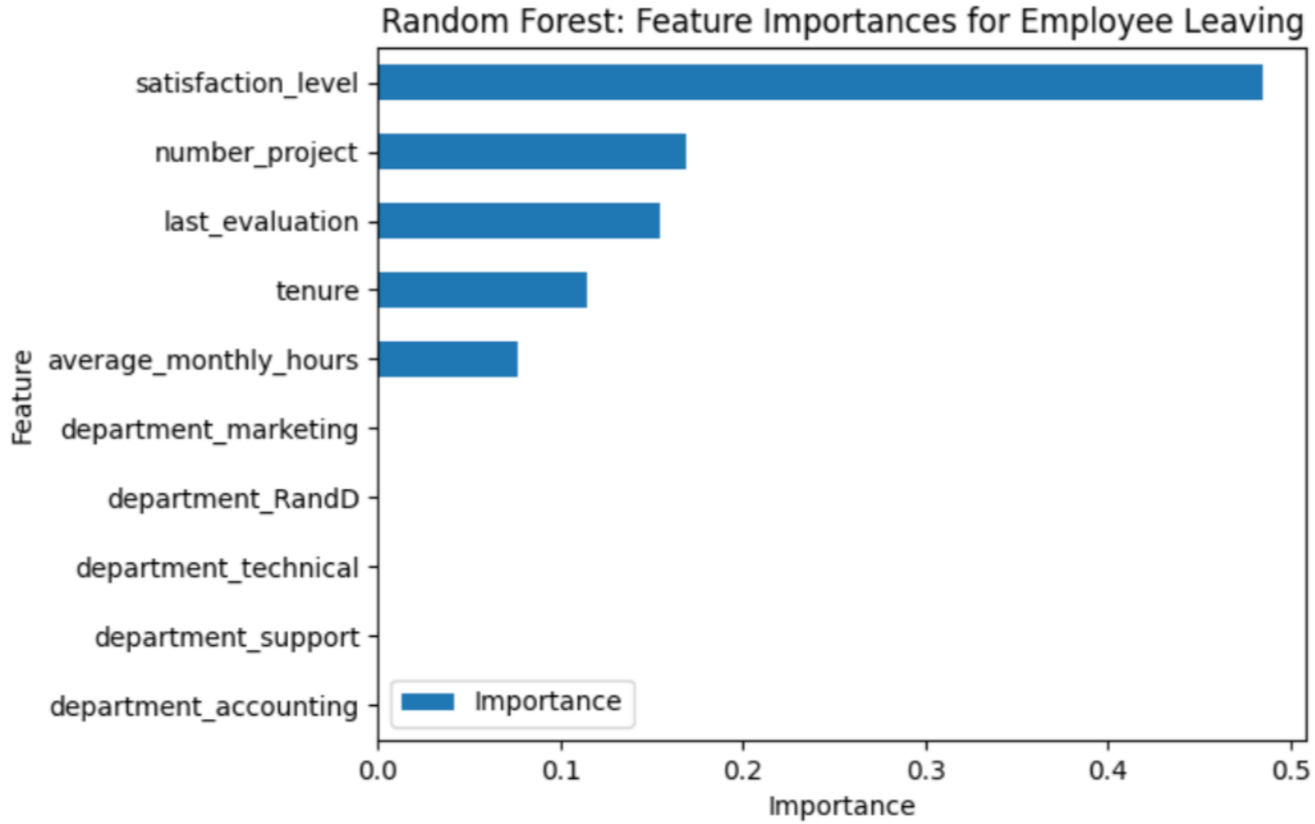
	precision	recall	f1-score	support
Predicted would not leave	0.86	0.94	0.90	1846
Predicted would leave	0.49	0.26	0.34	388
accuracy			0.82	2234
macro avg	0.68	0.60	0.62	2234
weighted avg	0.80	0.82	0.80	2234



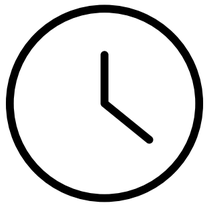
Random Forest



	precision	recall	f1-score	support
Predicted would not leave	0.98	0.99	0.99	2001
Predicted would leave	0.96	0.92	0.94	398
accuracy			0.98	2399
macro avg	0.97	0.95	0.96	2399
weighted avg	0.98	0.98	0.98	2399

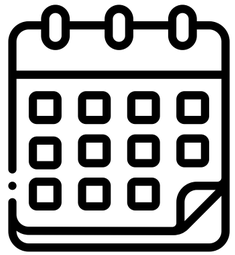


Recommendation to BU



<287

- working < 287 hours /month

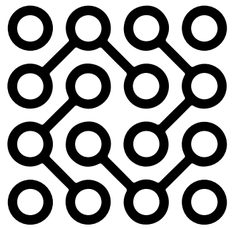


5

years tenure employee have to consider adjusting salary and promoted

2 – 6

- not Assign project >= 7 projects
- not Assign project = 1 projects



F1 94%

- Use ML model predicted proactive valuable employee left
  Random forest (Active)
  Logistic Regression (Challenge)

Thank you!

Have a great day ahead.