A PRESENTATION ON "HR Analytics"

HENRY HARVIN EDUCATION

SECTOR-2, C-107, NOIDA, UP

TEACHER'S GUIDE BY

ANIL JADON
DHIRAJ UPADHAYAY
POOJA GUPTA (MENTOR)

SUBMITTED BY :

Pankaj Chaudhary

HUMAN RESOURCE ANALYTICS

SUBMISSION:

Logistic Regression Model to Predict if the employee is going to leave the company or not and determine the factors driving it.

- #importing necessary libraries
- import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns
- #reading the excel file and loading the data in hr_data
- hr_data = pd.read_excel('HR_data.xlsx')
- #viewing the data
- hr_data.head()
- #viewing the correlation between the columns of the dataframe
- hr_data.corr()

PROBLEM STATEMENT

The company have contacted a consulting company to understand the factors on which the decision of employees to leave the company depends. Specifically, they want to understand why people are leaving their company on a frequent basis.. The company wants to know:

- Which variables are significant in predicting the decision of leaving the company

- How well those variables describe the probability of the person leaving the company

STEPS INVOLVED

Data Reading and Understanding
Data Cleaning and Preparation
Visualizing the Data
Data exploration

ASSUMPTIONS AND DATA HANDLING

Technical: we have assumed that ‘technical’, ‘IT’ and ‘support’ are the same categories. Same with others as well.

Data Cleansing: We have renamed the ‘IT’ and ‘support’ columns to ‘technical’, as per our understanding. We have converted all the data to lower case to avoid any case errors. We renamed ‘average_montly_hours’ to it correct name. The **duplicated** function searched for any duplicate values in our data and found 3008 entries. Hence, we deleted them. After deletion, we have 9653 entries and 10 columns.

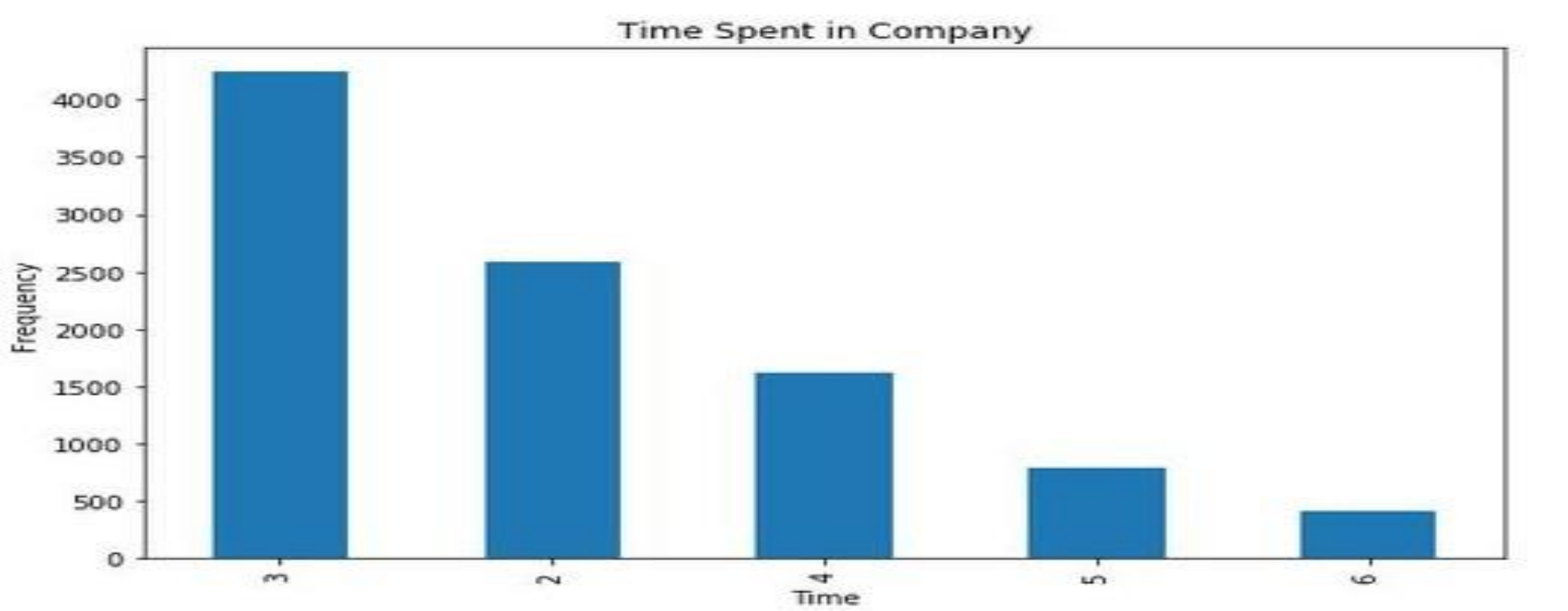
There were **no missing values** in the data and by performing the above steps, we prepared our data for analysis.

A separate data set **corr** was created that dealt only with the correlation of our target variable, left. This was done in order to select the best response variables for our study.

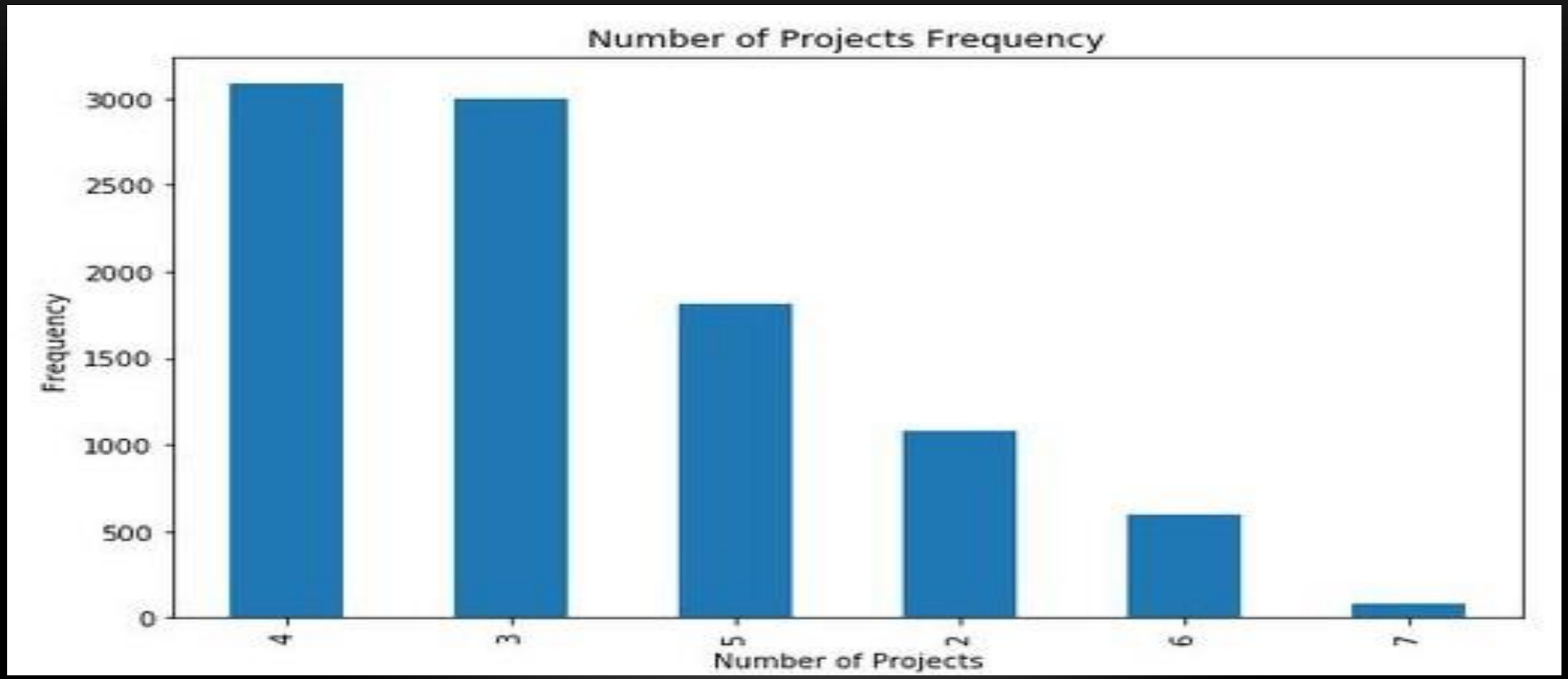
Another dataset **hr** was created that included only the columns that we selected based on our data exploration.

BASED ON OUR DATA EXPLORATION, WE CHOSE THE BEST FEATURES THAT WOULD HELP US PREDICT THE EMPLOYEE LEAVING. THESE ARE :

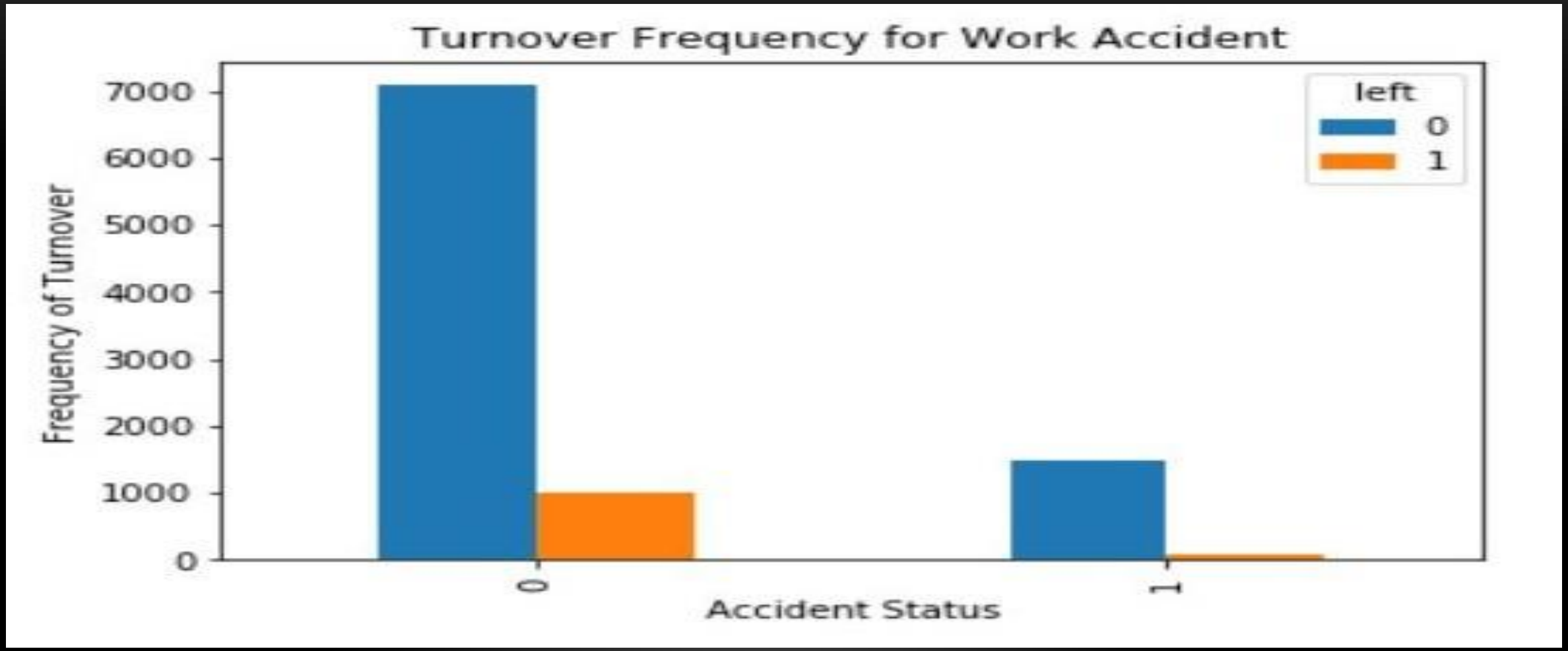
SATISFACTION LEVEL, TIME SPEND COMPANY, LAST EVALUATION, NUMBER OF PROJECTS, WORK ACCIDENT, PROMOTION LAST 5 YEARS, SALARY, DEPARTMENT.

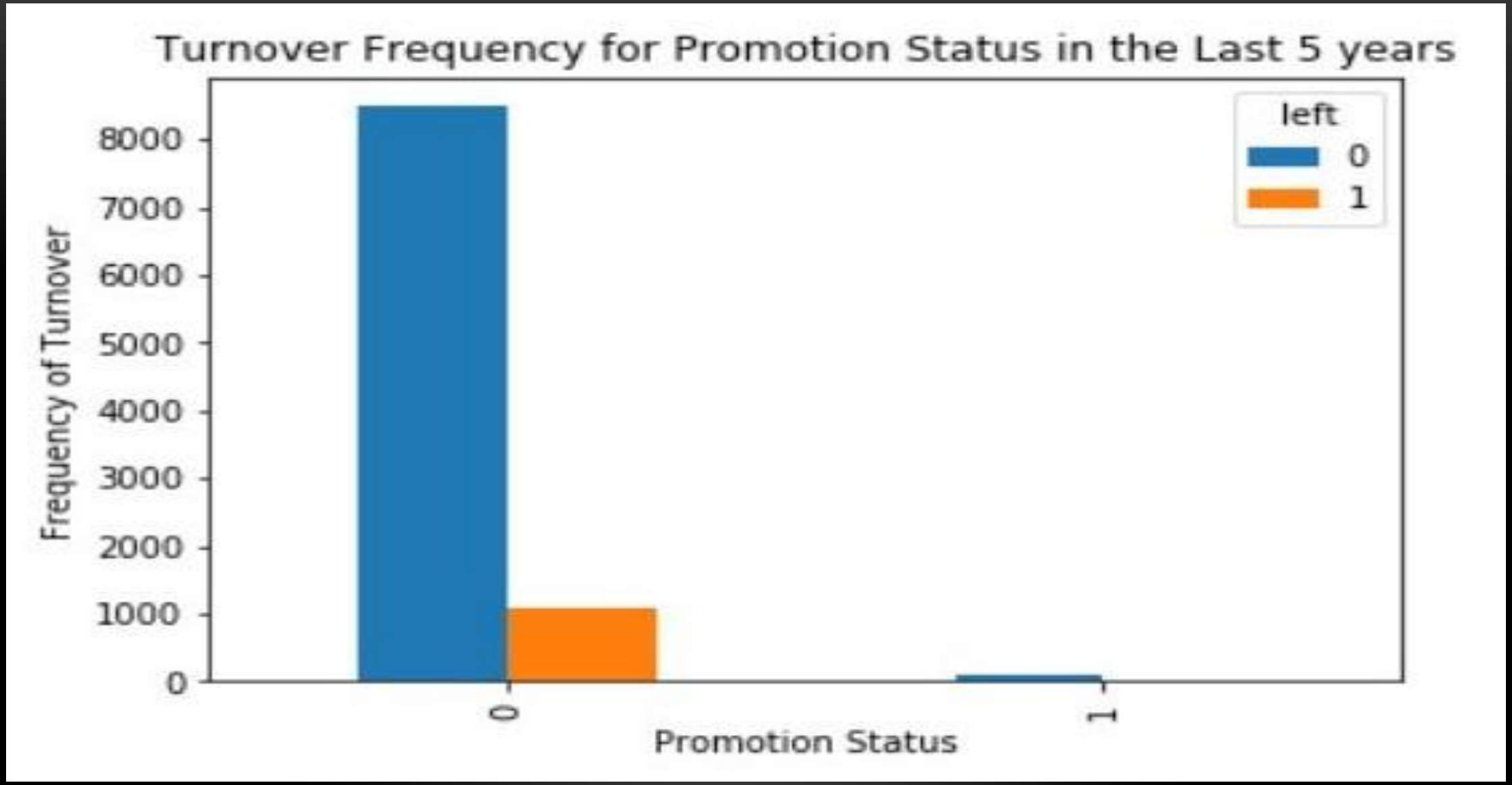


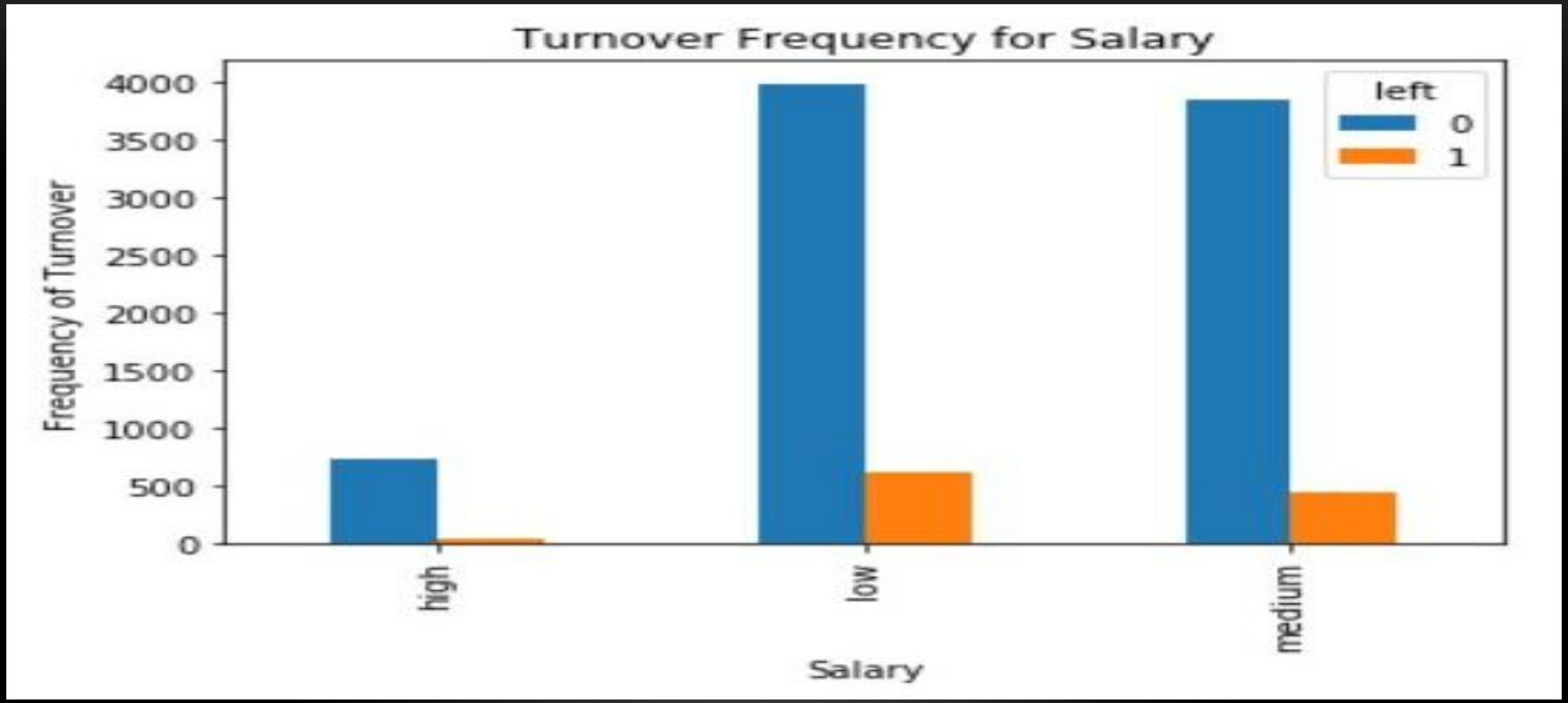
DATA EXPLORATION



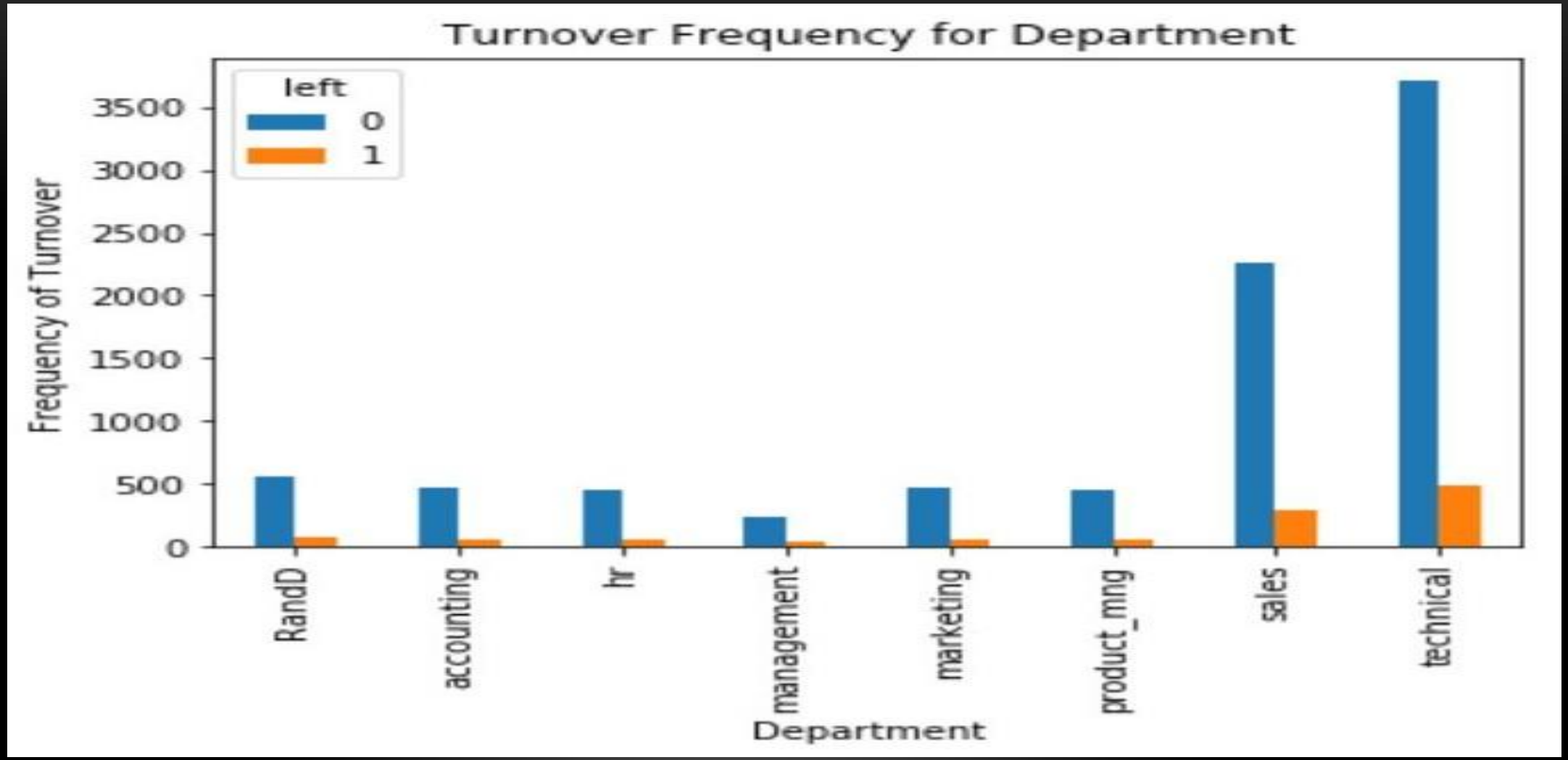
EXPLORATORY DATA ANALYSIS




Turnover Frequency for Promotion Status in the Last 5 years
left
0
1
Frequency of Turnover
8000
7000
6000
5000
4000
3000
2000
1000
0
0
1
Promotion Status

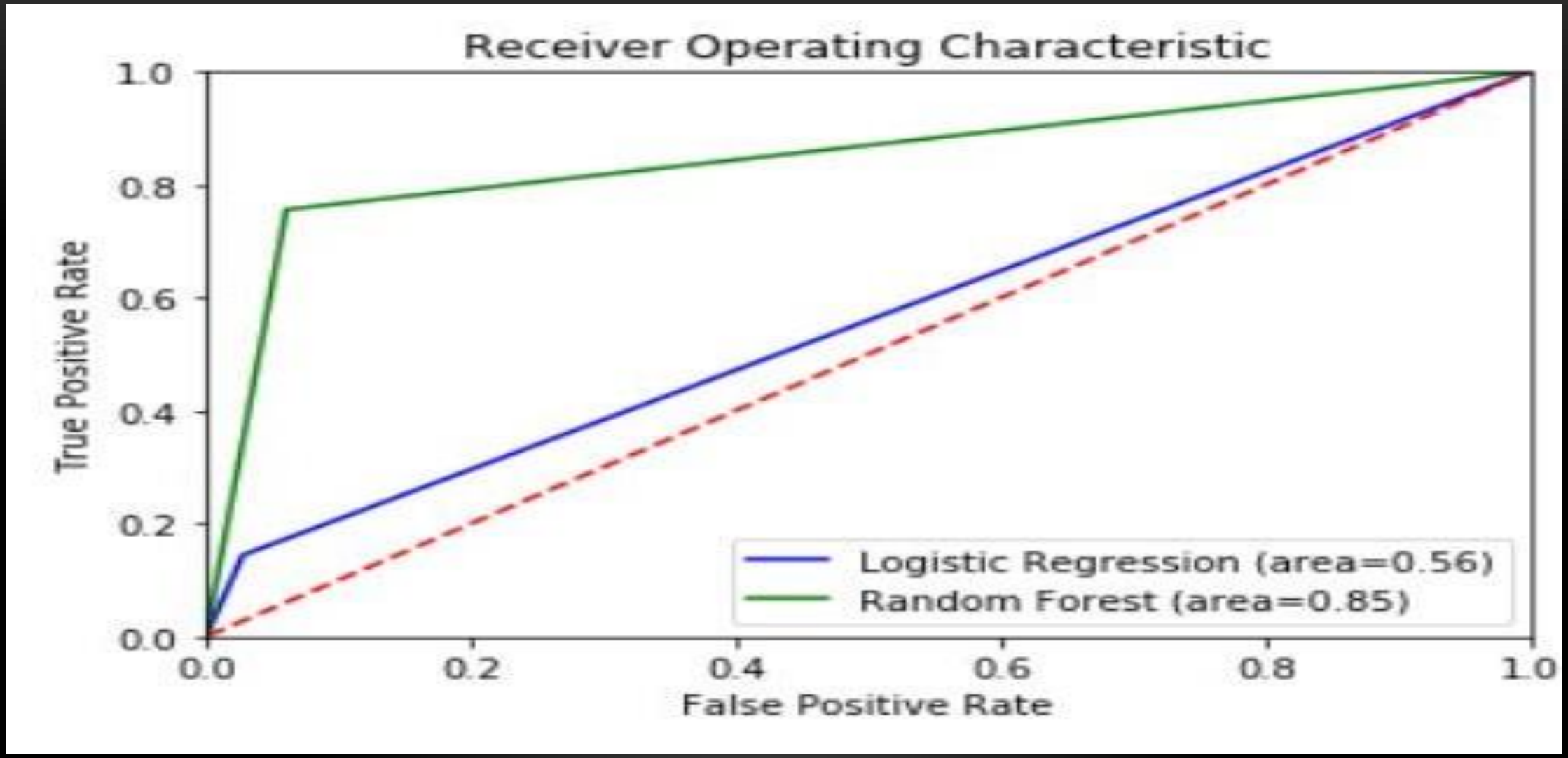

Turnover Frequency for Salary
Frequency of Turnover
4000
3500
3000
2500
2000
1500
1000
500
0
left
0
1
high
low
medium
Salary

EXPLORATORY DATA ANALYSIS



TECHNIQUE COMPARISON

Simple Logistic Regression	Random Forest Regression
In simple linear regression our accuracy score was 88.607%	Using random forest our accuracy was 92.025%
The receiver Operative characteristic (ROC) Curve was covering area=0.56	The receiver Operative characteristic (ROC) Curve was covering area=0.85.


Receiver Operating Characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (area=0.56)
Random Forest (area=0.85)

SUGGESTIONS AND INSIGHTS

1. Salary, Most employees leaving the company have low or medium salary.
2. Promotion, the people leaving the company did not get promotion for half a decade.
3. The workload too much on the employees. Most employees are able to complete only 4 projects on time.
4. They should provide gift vouchers and other small small things like foregine tours etc.
5. Training could be done more efficiently. Most people leaving are from technical background.
6. Satisfaction level of employees is very important for a company. Hence, they should work on keeping the satisfaction level of the employees high.

NOTE: AFTER DOING THIS ANALYSIS HENCE, WE CAN PREDICT THAT AN EMPLOYEE IS AN ASSET OF THE COMPANY.
THEY DEFINE THE FUTURE AND PRESENT OF THE COMPANY.

THANK
YOU!

You can access the project in the following GITHUB Repository:
https://github.com/pankajchaudhary1/sip-Project/tree/master/HR%20prediction

Contact details:

Pankaj Chaudhary
9911931247
Govind Puram, Ghaziabad
(201013), UP.

pankajchudhary1247@gmail.com
https://github.com/pankajchaudhary1

THANK YOU