# Human Resource Analytics Case Study

## SUBMISSION

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

# CASE STUDY OVERVIEW

# PROBLEM STATEMENT

A company XYZ has around 9000 employees. However, in every 3-4 months, their employees are leaving and they have to hire new (raw) talent from the market. Training them and guiding them becomes extremely difficult when they have to hire new person in every 3-4 months.

They 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

# BUSINESS GOAL

I am required to model the decision of leaving with the available independent variables. It will be used by the management to understand how exactly the decision varies with the independent variables. They can accordingly manipulate the company policies, the business strategy etc. to meet certain employee satisfaction level. Further, the model will be a good way for management to understand the shortcomings in their current system.

# DATA UNDERSTANDING

# Data Reading and Understanding

The data set contains 14999 entries and 10 columns. The names of those columns are:

Number of projects, work_accident, left, promotion last 5 years, department, salary, time_spend_company, average_monthly_hours, satisfaction_level, last_evaluation

Here, our **target variable** is **LEFT.**

# ASSUMPTIONS AND DATA HANDLING

# 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.

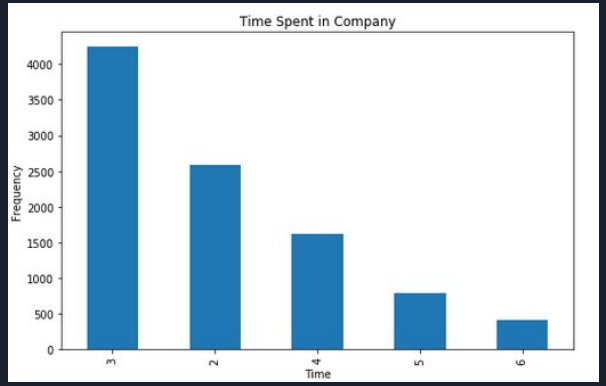
# DATA EXPLORATION

# Exploratory Data Analysis

Based on our data exploration, we chose the best features that would help us predict the prices of the cars. These are :

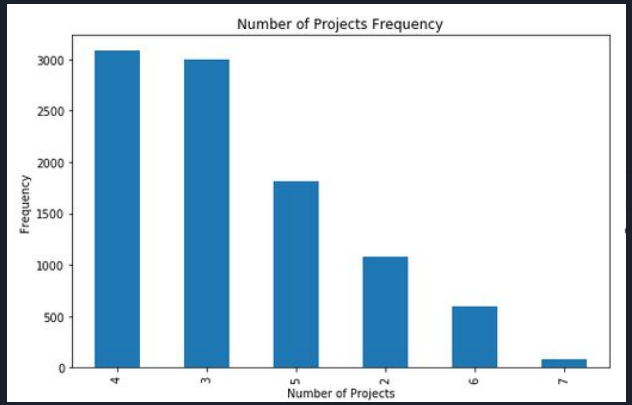
Satisfaction Level, Time Spend Company, Last Evaluation, Number of Projects, Work Accident, Promotion last 5 years, Salary, Department

1. Time Spend Company




It seems that most employees spent just around 3 hours in the company.

2. Number of Projects



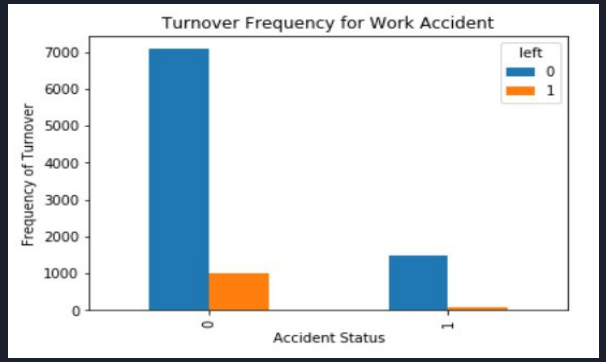

It seems that most employees are able to complete only 4 projects on time.

# Exploratory Data Analysis

Based on our data exploration, we chose the best features that would help us predict the prices of the cars. These are :

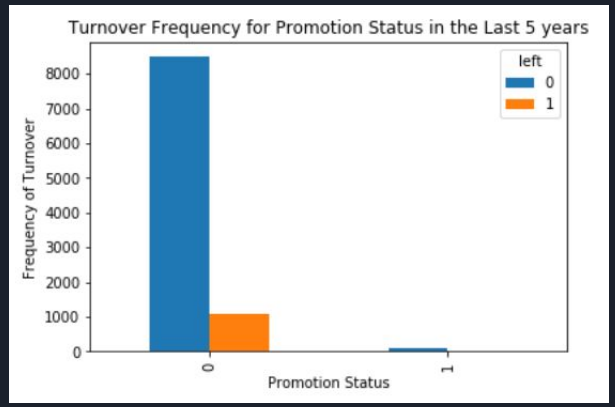
Satisfaction Level, Time Spend Company, Last Evaluation, Number of Projects, Work Accident, Promotion last 5 years, Salary, Department

## 3. Work Accident




It seems that people who don't have any work accidents are leaving the company more.

## 4. Promotion Status in the last 5 years



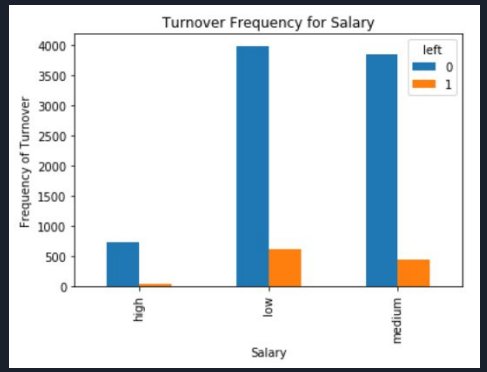

It is clearly visible that people are often not getting promotions and hence they are leaving the company.

# Exploratory Data Analysis

Based on our data exploration, we chose the best features that would help us predict the prices of the cars. These are :

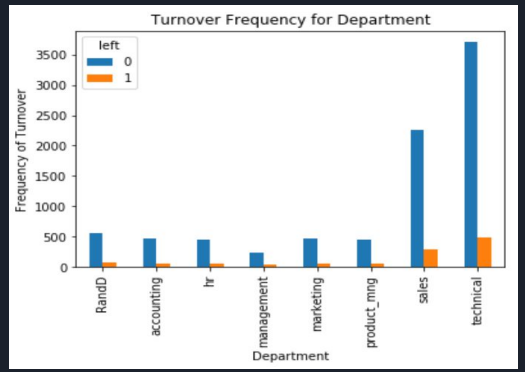
Satisfaction Level, Time Spend Company, Last Evaluation, Number of Projects, Work Accident, Promotion last 5 years, Salary, Department

5. Salary




It seems that people with low salary range are leaving the company more frequently.

6. Department




It seems that most people leaving the company are from technical department.

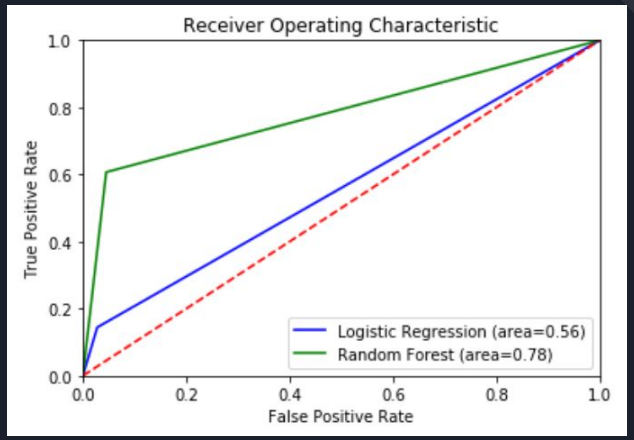
# TECHNIQUE COMPARISON

Simple Logistic Regression vs Random Forest Classifier

# Technique Comparison

| Simple Logistic Regression | Random Forest Classifier |
| --- | --- |
| In Simple Logistic Regression, Our Accuracy Score was **88.607%.**<br><br>The Receiver Operative Characteristic (ROC) Curve was covering area=0.56. | Using Random Forest, Our Accuracy Score was **91.25%.**<br><br>The Receiver Operating Characteristic (ROC) Curve was covering area=0.78. |

ROC Curve:



# SUGGESTIONS AND INSIGHTS

# Suggestions and Insights

1. One of the most important thing for an employee is his salary. Most employees leaving the company have low or medium salary.
2. Promotion is something that motivates an employee to give his 100% for the company. However, the people leaving the company did not get promotion for half a decade.
3. The company is inflicting workload too much on the employees. Most employees are able to complete only 4 projects on time.
4. Training could be done more efficiently. Most people leaving are from technical background.
5. Satisfaction level of employees is very important for a company. Hence, they should work on keeping the satisfaction level of the employees high.

NOTE:

An employee is an asset to the company. They define the future and present of the company.

Thanks! You can access the project in the following GITHUB Repository:

**https://github.com/yashj1301/Python-Projects/tree/master/HR%20Analytics**

Contact details:

**Yash Jain**
9582617525
Greater Noida, UP-201306

**yash.jain106@gmail.com**
**https://www.linkedin.com/in/yash-j-5537a0146/**