

Human Resource Analytics Case Study

- Poojan Fegade

Overview Context

In a company ABC, every year, around 20% of the employees leave the company or are replaced with greater talent pool.

Problem Statement

The company faces ~ 20% attrition, which is not well received in the industry. We need a model to predict whether an employee will leave the company.

We also need to analyse the data to gain some useful information regarding the factors that lead to attrition to help the company avoid fatal errors in their operations.

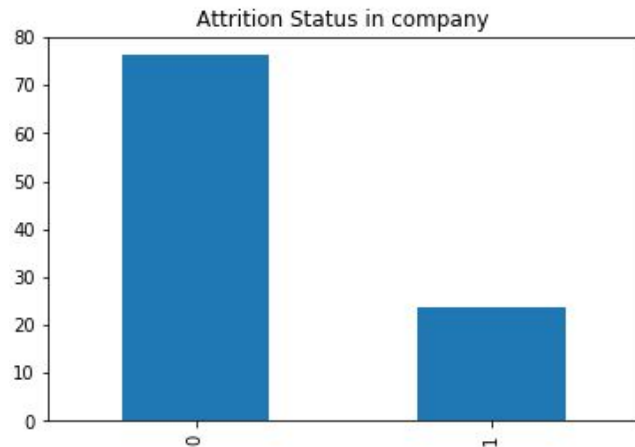
Methodology

- Understanding the data fields as well as their basic statistical properties
- Identifying the categorical data fields and understanding it's role
- Data cleaning by filling missing values or removing null record.
- Manipulate data by encoding categorical variables
- Split the data as training and testing data (7:3 ratio respectively)
- Fit model using features and validate using the test data.

Data Visualisation

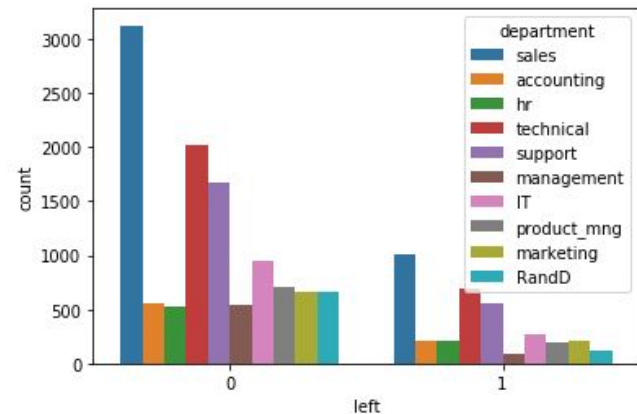
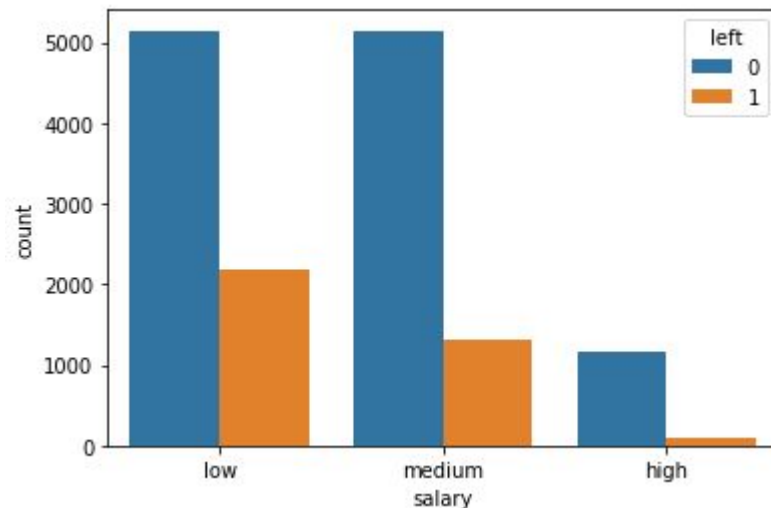
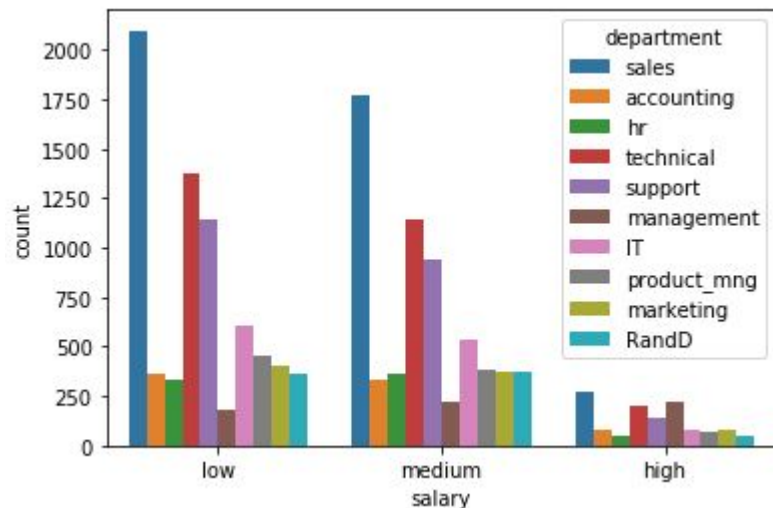
The attrition status is as shown.

We see more than 20% attrition, which is worrying.



Next we observe that the attrition levels in higher in case of employees with lower salaries.

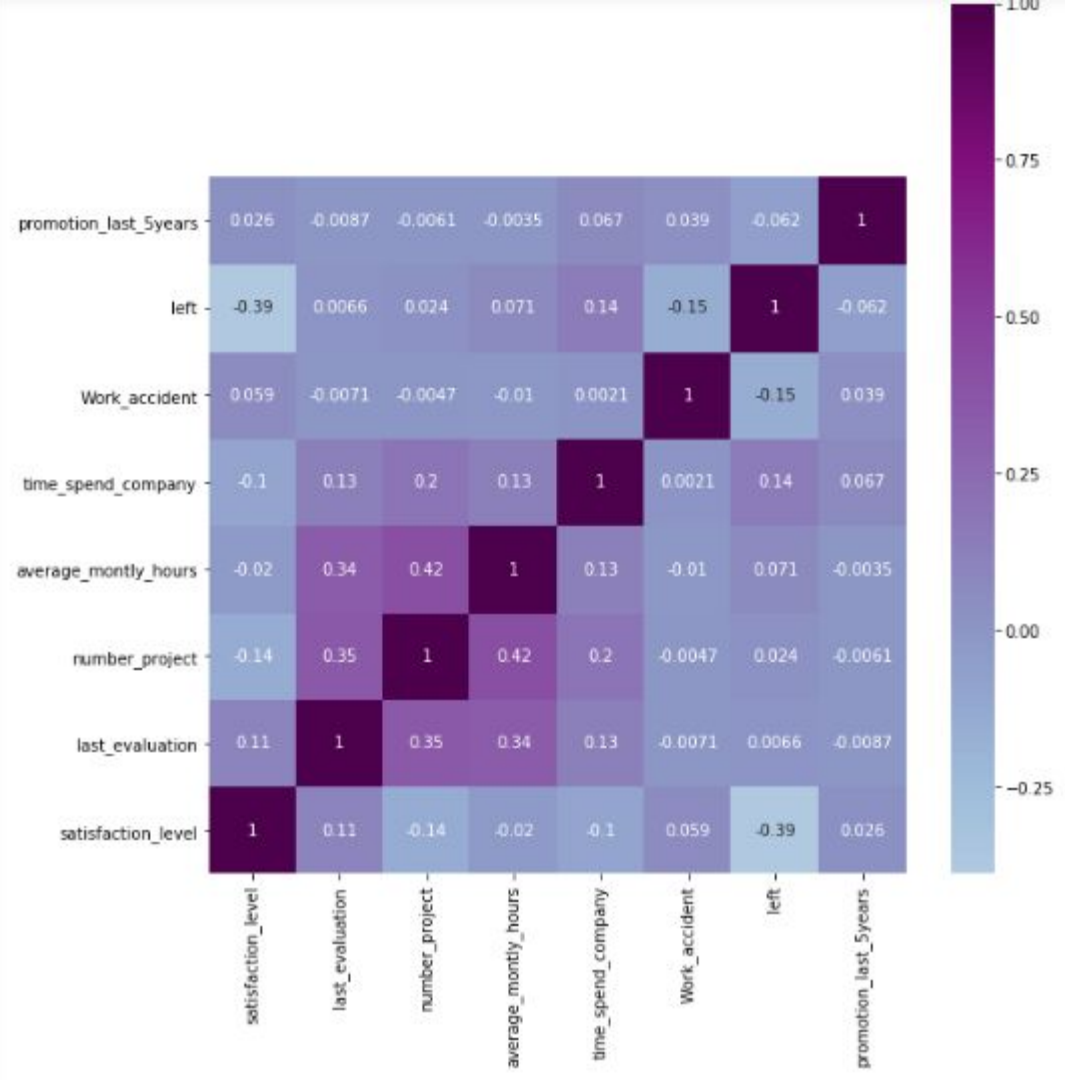
We then come to an understanding that department wise sales leads in attrition percentage due to various reasons like salary, work hours, number of projects



The cross- correlation matrix for the dataset is as shown.

Here we observe that the features don't have significant correlation with left.

This may be due to outliers and other noise in the data.

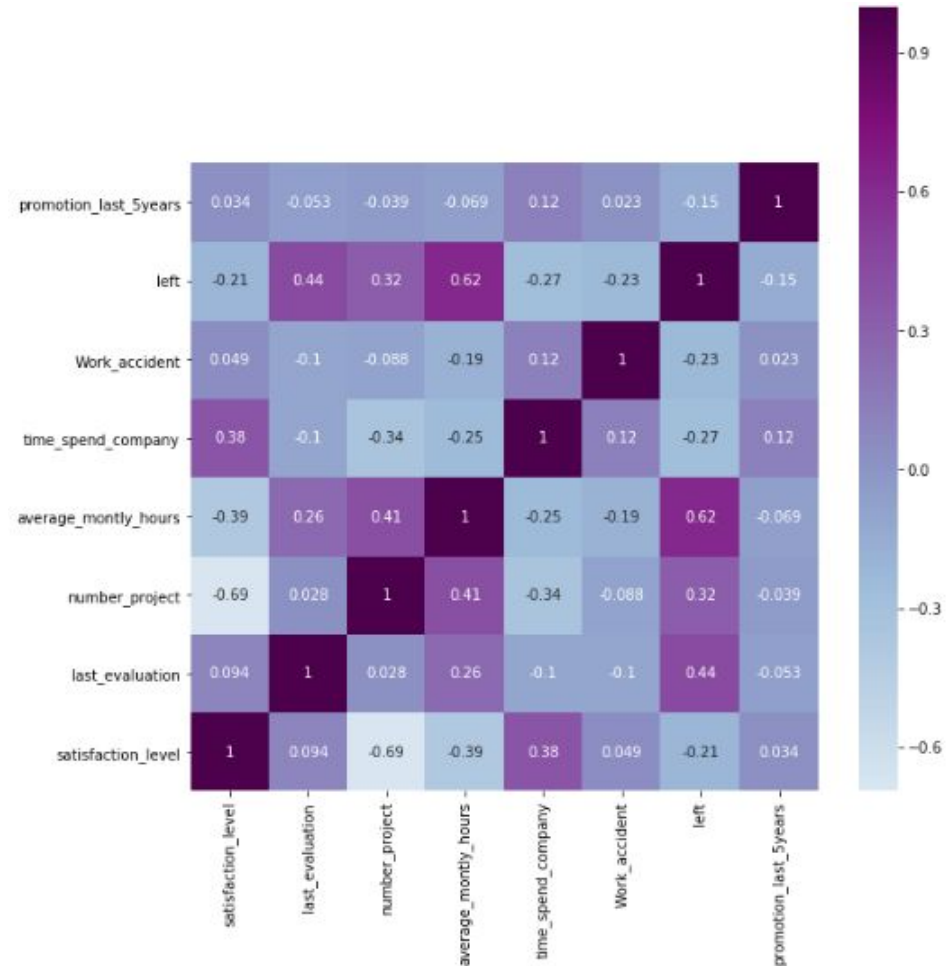


Segregating the data

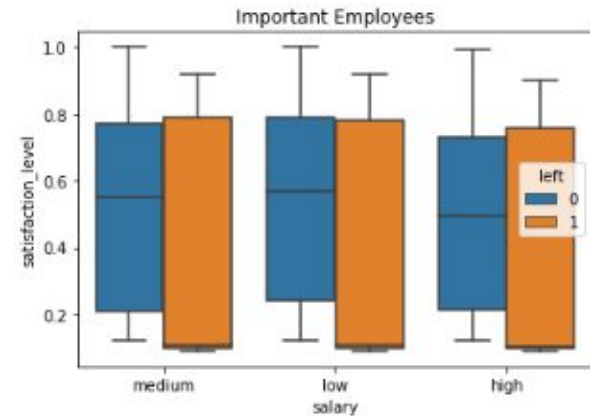
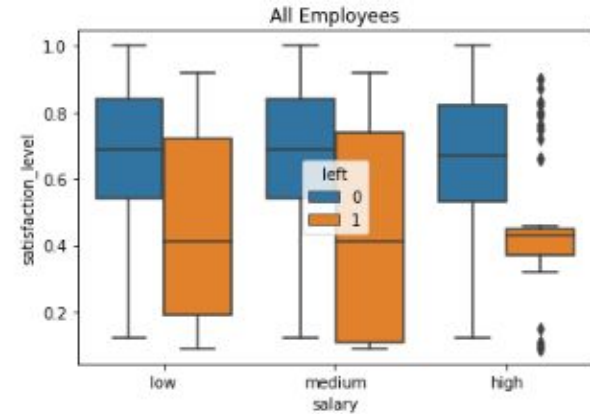
We then segregate the data by recognizing valuable employees to the company by setting bars on last_evaluation, number_projects and time_spent_company

We find that in case of valuable employees :

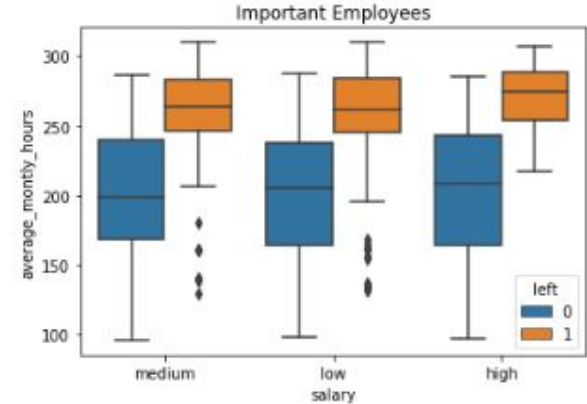
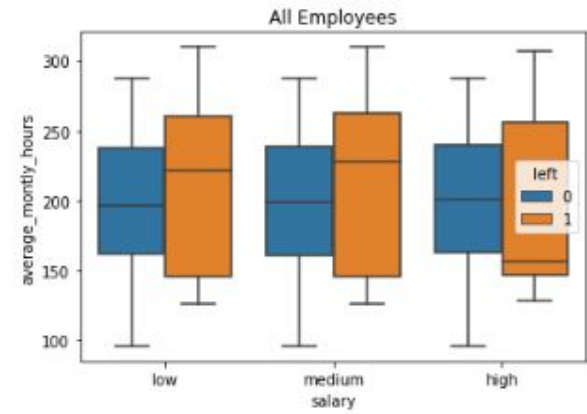
Average_monthly_hours, last_evaluation, number_projects and promotion_last_5_years are important fields to consider.



Comparative analysis of data (valuable vs general)

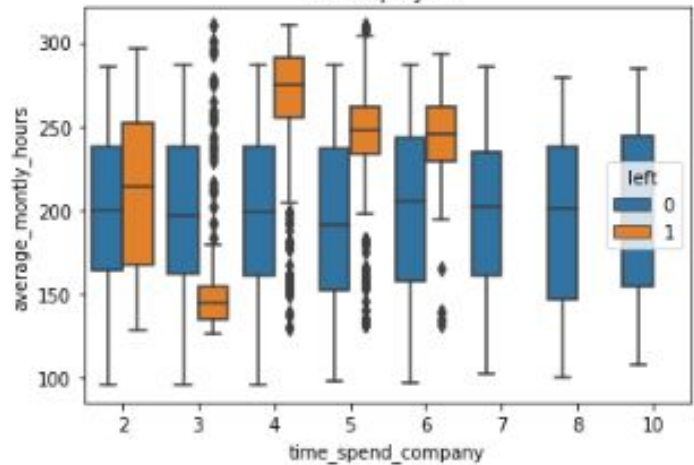


Satisfaction levels of employees is consistently lower in case of employees that have left the company.



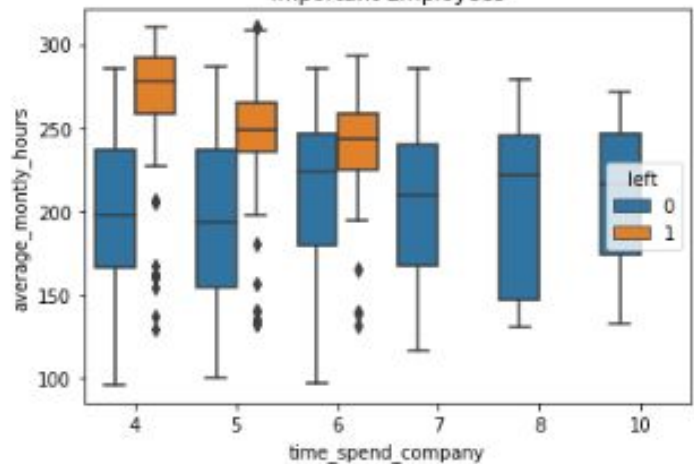
This tells us that higher working hours with low salary are more prone to leave.

All Employees



Employees with higher working hours leave the company.

Important Employees



Model

We combine two models in this case for better results. We treat valuable employees more seriously than temporary employees. Hence we first train two different models and then combine into one.

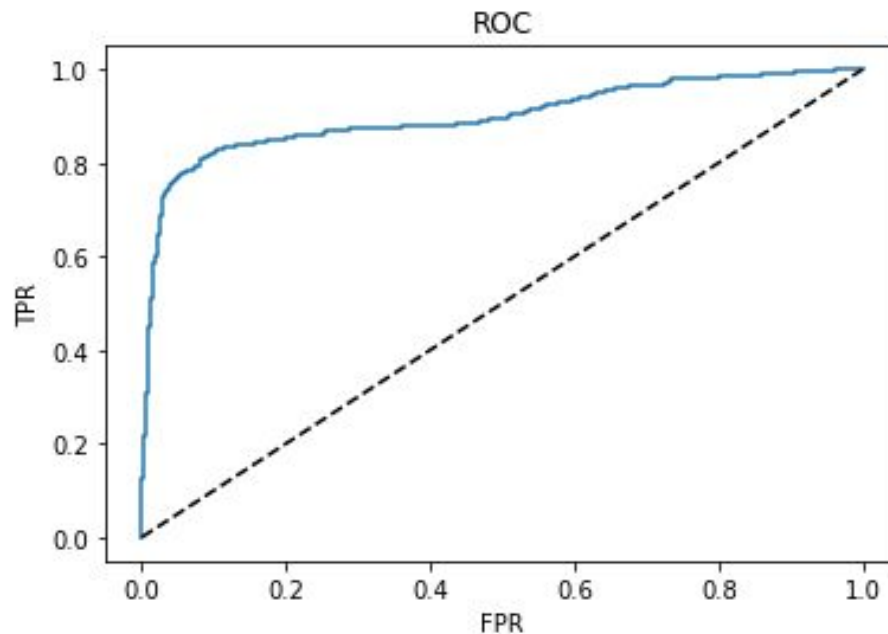
We use Logistic regression in both the cases as it is a classification problem.

	Model 1	Model2
Constant	2.7597185	-6.23156193
satisfaction_level	-3.11E+00	-0.08001637
last_evaluation	-8.37E-01	4.03662824
number_project	-1.17E+00	0.03457364
average_monthly_hours	-2.84E-03	0.0307354
time_spend_company	1.84E-01	-0.45682995
Work_accident	-1.36E+00	-0.95251866
promotion_last_5_years	-7.88E-01	-1.28536759
department_1	1.41E-01	-0.49786641
department_2	-2.25E-01	-0.92733148
department_3	4.66E-01	-0.733681
department_4	7.33E-01	-0.26084151
department_5	8.41E-02	-0.9188091
department_6	4.77E-01	-0.58752876
department_7	9.97E-04	-0.57317549
department_8	3.72E-01	-0.83515914
department_9	2.91E-01	-0.27942236
department_10	4.20E-01	-0.61774667
salary_1	-8.88E-03	-3.10146272
salary_2	1.62E+00	-1.36162395
salary_3	1.15E+00	-1.7684752

Confusion matrix of the whole model is:

	Predicted No	Predicted Yes
Actual No	3285	146
Actual Yes	260	810

ROC curve is as shown :



Insights

- Work life balance is important for any employee, hence the management needs to see to it that working hours don't exceed a lot.
- Employees with high job satisfaction are likely to stay as they are happy with the current job
- Employees in the sales department need to be paid more or should be cut on their daily hours spent in the company. These two fields are exceptionally high in the sales departments and they tend towards attrition
- Promotion in 5 years has positive effect in attrition. This indicator leads to demotivation in the employees for not receiving promotion in the longer duration of time. The company revise its promotion policies and metrics.