Employee Attrition Analysis Using Python

Raviteja Bodla

Metro College Of technology

SEP-2018

**Introduction:**

Employee turnover is not only a cumbersome process but also a financial burden. Employers can benefit from knowledge of an employee's likelihood of changing the company. Often, decisions for leaving a company do not emerge out of a sudden but are usually the outcome of careful pre-planning. Utilising a data-driven approach to analyse which employees are likely to leave the company soon, therefore, can be of great interest to managers. With the gained knowledge, human resource management teams can act upon these predictions to persuade employees to stay before they jump ship.

**Objective:**

The Objective of the study is to analyse various factors that are contributing to the employee attribution.

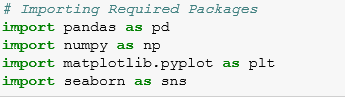
**Obtaining the data:**

The data was found from the “Human Resources Analytics” dataset provided by Kaggle’s website:

https://www.kaggle.com/ludobenistant/hr-analytics.

This dataset is a simulation of a hypothetical company, which means that the features and observations used are all made up to mimic a real-world scenario. The number of observations given from the dataset contains 14,999 employee information.

I will be using Python as the programming language for the analysis:





**Data preparation/cleaning:**

Typically, data preparation/cleaning requires a lot of work and can be a very tedious procedure. This dataset from Kaggle is clean and contains no missing values. But still, I will have to examine the dataset to make sure that everything else is readable and that the observation values match the feature names appropriately.

The following independent variables were used in the model:

• **Satisfaction**: An employee’s level of satisfaction in percentage

• **Evaluation**: An employee’s evaluation score in percentage

• **Project Count:** The number of projects the employee has done

• **Average Monthly Hours:** The total monthly hours an employee worked

• **Years At Company:** The number of years an employee was at the company

• **Work Accident:** Whether an employee had an accident or not. Where 0 (zero) means no and 1 (one) means yes

• **Promotion:** Whether an employee had a promotion within the last five years. Where 0 (zero) means no and 1 (one) means yes

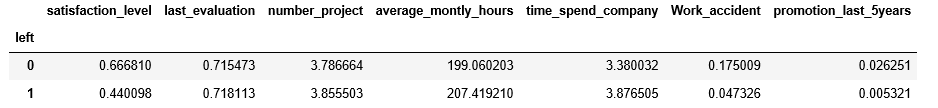
• **Department:** The type of department an employee worked under. Which includes sales, accounting, hr, technical, support, management, IT, product management, and marketing.

• **Salary:** The type of salary an employee got, which ranges from low, medium, or high.

**Exploratory Data Analysis:**

Here are some important numbers to keep in mind of the dataset:

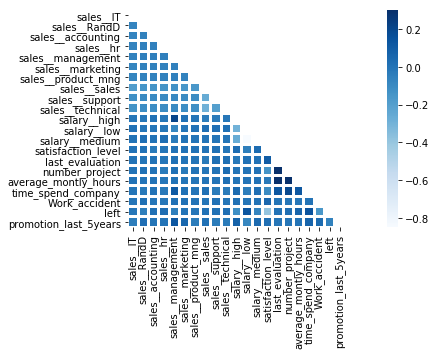
* There is 14,999 employees and 9 independent variables
* Attrition rate: 24%
* Mean satisfaction: 0.66



**Correlation Matrix & Heatmap:**

From the heatmap, there is a positive (+) correlation between the variables: project count, average monthly hours, and evaluation. Which means that the employees who spent worked more hours and did more projects had higher evaluations.

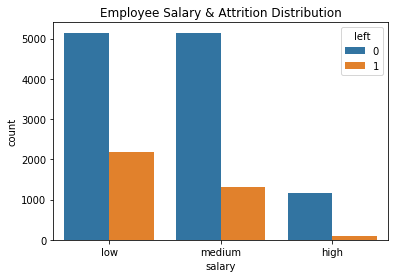
For the negative (-) relationships, the most important feature that correlated with our target variable (turnover) is satisfaction. This should support our initial intuition that employees who tend to quit would normally have lower satisfaction level.



**Salary Vs Turnover:**

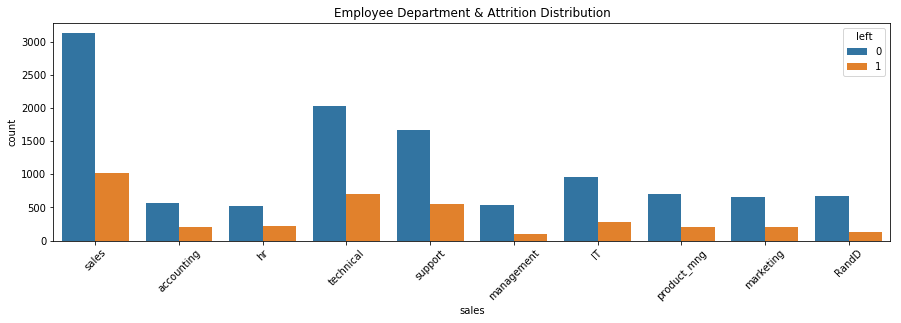
**-**Majority of employees who left either had low or medium salary

-Only a few employees left with high salary



**Department Vs Attrition:**

The top three departments with the most attrition are sales, technical, and support.



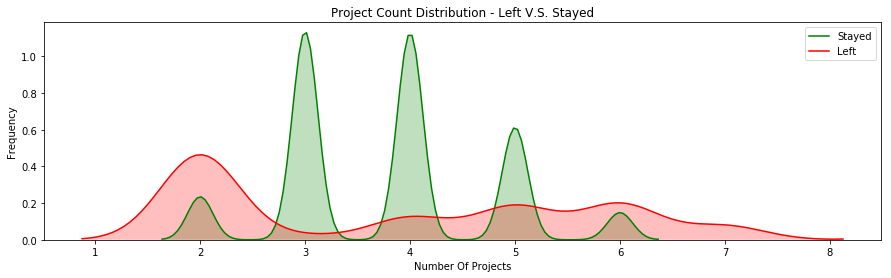
**Project Count Vs Attrition:**

-More than half of employees with 2,6, and 7 projects left the company

-Majority of the employees who did not leave had 3,4, and 5 projects

-All employees with 7 projects left the company

-There is an increase in turnover as project count increases

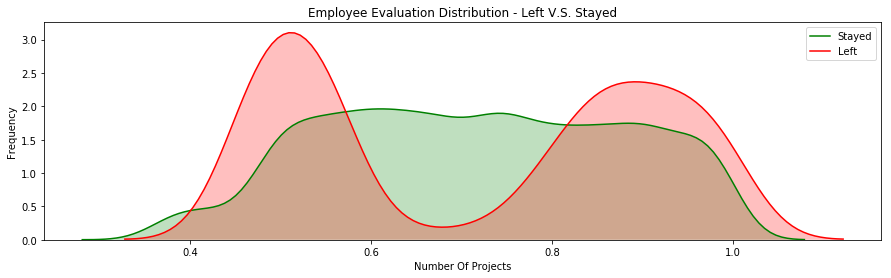


**Attrition Vs Evaluation:**

-There is a bimodal distribution for employees that left the company

-Employees with low evaluation levels (0.2-0.6) and high evaluation levels (0.8-1) were the bulk of employee turnover

-Employees with evaluation levels (0.6-0.8) had the smallest turnover rate.



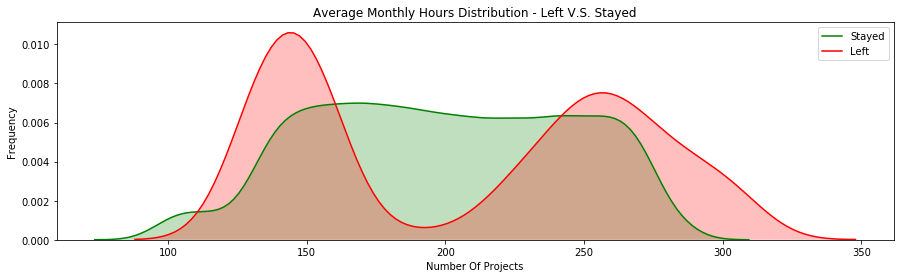
**Average Monthly Hours Vs Attrition:**

-There is another bimodal distribution for employees that left the company.

-Employees who had less hours of work (~150 hours or less) left the company more.

-Employees who had more hours of work (~250 hours or more) left the company more.

-Employees who left generally were underworked or overworked.



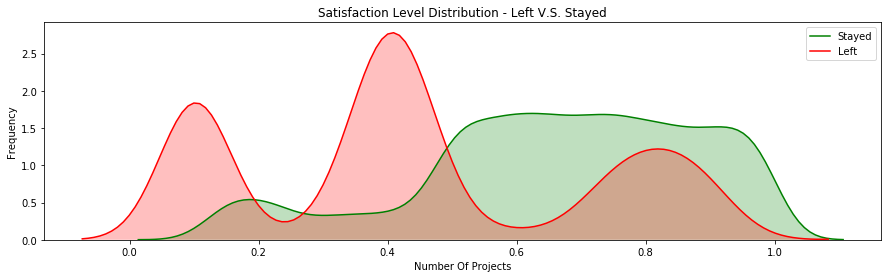
**Satisfaction Vs Attrition:**

-There is a tri-modal distribution for employees that left the company

-Employees left with really low satisfaction levels of (0-0.2)

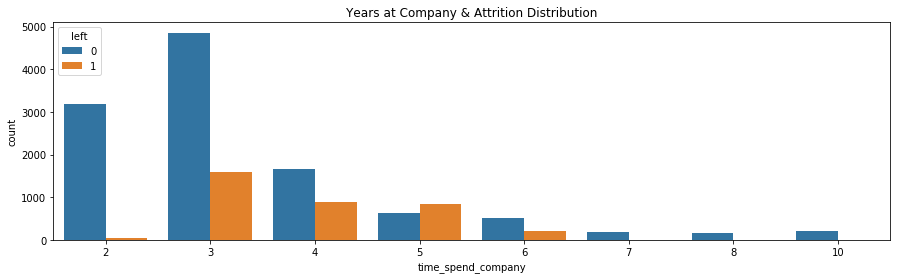
-Employees left with low satisfaction levels of (0.3-0.5)

- Employees left with high satisfaction levels of (0.7-1)



**Time Spent in Company Vs Attribution:**

-More than half of the employees with 4 and 5 years left the company.



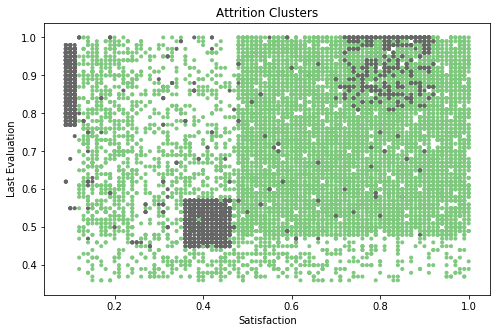
**Attrition Clusters:**

-There are three distinct clusters for employees who left the company

Cluster 1 (“Overworked” Employees): These employees had satisfaction scores below 0.2 and evaluation scores above 0.75. Employees here were evaluated highly and felt bad at work.

Cluster 2 (“Under Performing Employees”): These employees had satisfaction scores between (0.35-0.5) and evaluation scores below 0.6. Employees here were evaluated poorly and felt bad at work. This is a typical reason why employees leave.

Cluster 3 (“Ideal Worker”): These employees had satisfaction scores between (0.7-1) and evaluation scores of (0.8-1). Employees here were evaluated highly and felt satisfied at work.



**Conclusion:**

* Employees generally left when they are underworked (less than 150hr/month or 6hr/day).
* Employees generally left when they are overworked (more than 250hr/month or 10hr/day).
* Employees with either really high or low evaluations should be taken into consideration for high attrition rate.
* Employees with low to medium salaries are the bulk of employee attrition.
* Employees that had 2,6, or 7 project count was at risk of leaving the company
* Employee satisfaction is the highest indicator for employee attrition.
* Employee that had 4 and 5 years at Company should be taken into consideration for high attrition rate
* Employee satisfaction, years at company, and evaluation were the three biggest factors in determining attrition.