Employee Attrition

TakenMind Project - Proof of Concept

By: Jack Praveen Raj Ilango

Problem Statement:

Employee Attrition Problem

- The data is for Company “X” which is trying to control attrition.
- There are two sets of data: “Existing employees” and “Employees who have left”.
- What type of employees are leaving?
- Determine which employees are prone to leave next.

Data Analysis:

1. Salary:

Salary	Number Of Employees
low	2172
medium	1317
high	82
Grand Total	**3571**

Lower salaries

- Lower salaries were found to be one of the main reasons
- Mainly in low or medium range

Data Analysis:

1. Salary:

- Using the bar graph, it is easy to see that salary plays an important role
- Concentrated in low or medium range

Data Analysis:

2. Number of Promotions:

Promotion In 5 years	Number Of Employees
0	3552
1	19
Grand Total	**3571**

- Number of promotions was also found to play an important role in attrition
- Almost all the employees had 0 promotions over 5 years

Data Analysis:

3. Last Evaluation: (Interesting finding)

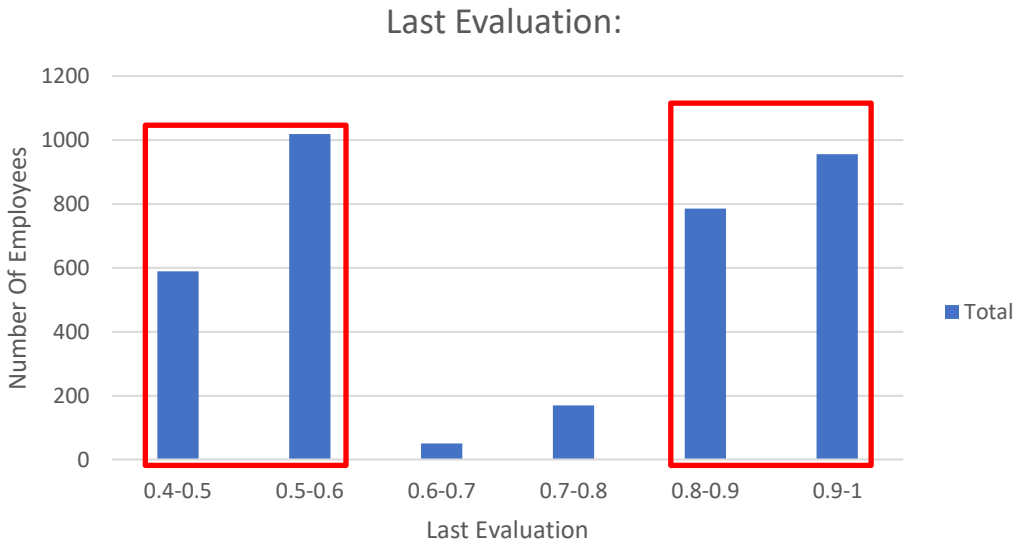
Last Evaluation	Number Of Employees
0.4-0.5	589
0.5-0.6	1019
0.6-0.7	51
0.7-0.8	170
0.8-0.9	786
0.9-1	956
Grand Total	**3571**

- Employees at lower and higher ranges of evaluation scores tend to leave the company

Data Analysis:

3. Last Evaluation: (Interesting finding)

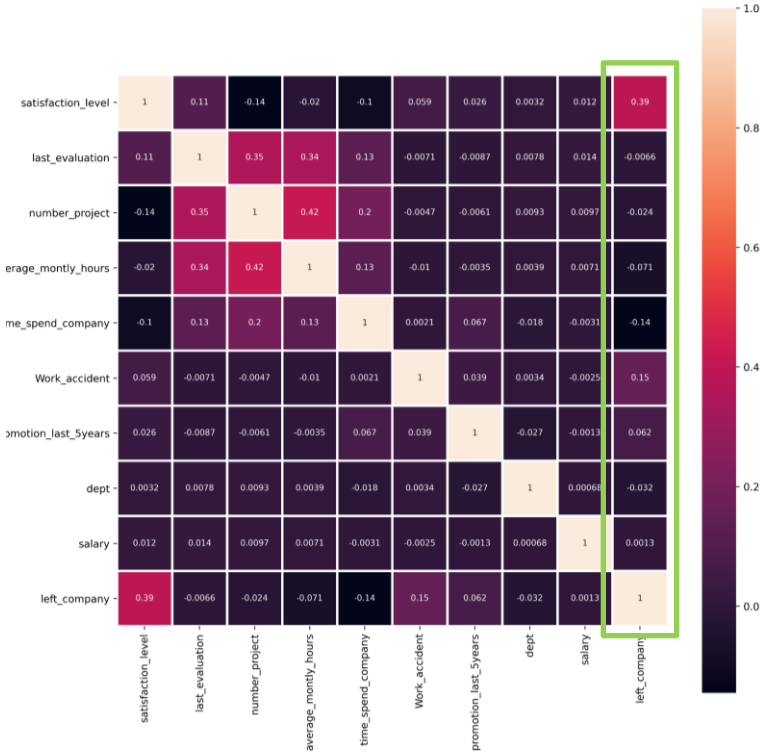
- Using the bar graph, we can see that there is a pattern in evaluation scores
- Concentrated in 0.4 - 0.6 and 0.8 - 1.0 range



Data Analysis:

Heatmap analysis

- A Heatmap of Correlation values of all the columns against employment status can be useful to find important factors.



Prediction:

Overview:

1. Adding Column “left_company” so that model could learn and predict who wil leave.
2. Merging both Datasets.
3. Splitting into training and testing datasets (80% and 20% respectively)
4. Testing accuracies of different models.

Prediction:

Prediction Models:

Linear Regression Accuracy Score: 0.9777777777777777

Decision Tree Accuracy Score: 0.9782222222222222

Random Forest Accuracy Score: 0.9884444444444444

Accuracy = 98.44% (Highest)

Predicted Data:

Results:

Emp Id	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	dept	salary
4520	0.44	0.45	2	124	3	0	0	sales	low
5698	0.81	0.98	6	196	2	0	0	support	low
5832	0.28	0.46	4	260	2	0	0	accounting	medium
6264	0.9	0.87	4	231	5	0	0	management	low
6359	0.81	0.98	5	243	6	0	0	sales	medium
7425	0.23	0.99	5	176	4	1	0	sales	low
7990	0.36	0.45	2	135	3	1	0	support	high
8473	0.23	0.47	4	277	5	0	0	RandD	medium
8678	0.2	1	3	123	4	0	0	support	low
8972	0.58	0.54	3	287	6	0	0	technical	medium
9176	0.64	0.99	5	262	5	0	0	sales	low
9537	0.28	0.94	6	167	3	1	0	RandD	low
9583	0.43	0.51	2	123	3	0	0	technical	medium
10099	0.73	0.83	5	266	5	0	0	sales	low

14 employees might leave the company

Predicted Data:

1. Salary:

Salary	No. of Employees
low	8
medium	5
high	1
Grand Total	**14**

Predicted Data:

2. Number Of Promotions:

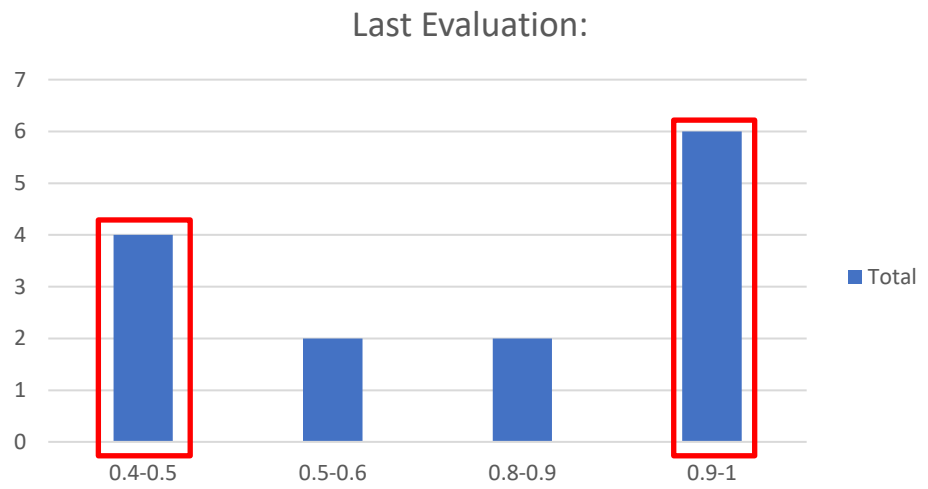
Number of Promotions = 0

Salary	No. of Employees
low	**8**
0	8
medium	**5**
0	5
high	**1**
0	1
Grand Total	**14**

Predicted Data:

3. Last Evaluation:

Last Evaluation	No. of Employees
0.4-0.5	4
0.5-0.6	2
0.8-0.9	2
0.9-1	6
Grand Total	**14**



Suggestions:

Overview:

1. From the given and the predicted data, we can see that employees with low salaries tend to leave the company. Thus, to prevent that the company may increase their salaries.
2. Similarly, Promoting employees who have not received any promotions in the last 5 years will prevent them from leaving.
3. Employees who have high evaluation scores tend to leave since they have not received a promotion or higher salary.
4. Similarly, Employees who have low evaluation scores tend to either leave or get fired.