HR Case Study

A deeper analysis on Churn

The problem

Churn???

Employee churn is the overall turnover in an organization's staff as existing employees leave and new ones are hired. The churn rate is usually calculated as the percentage of employees leaving the company over some specified time period.

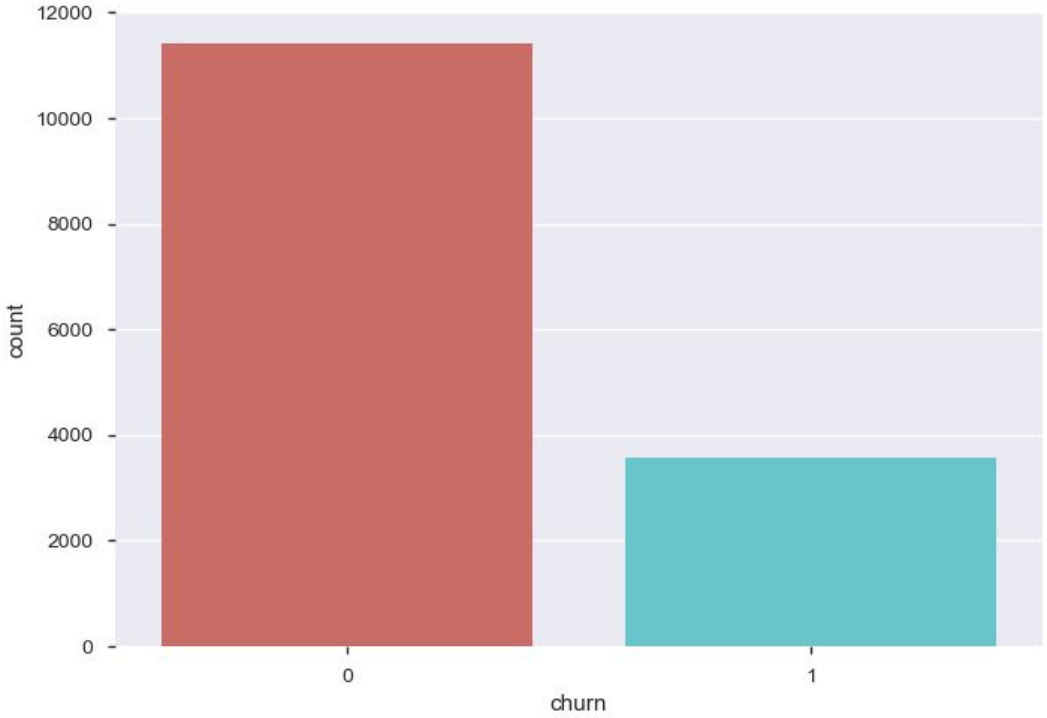
Problem statement

- HR- Data is slowly losing employees to competitors in the market. Unfortunately management has not been able to figure out the reason for employee churn.

Deliverable

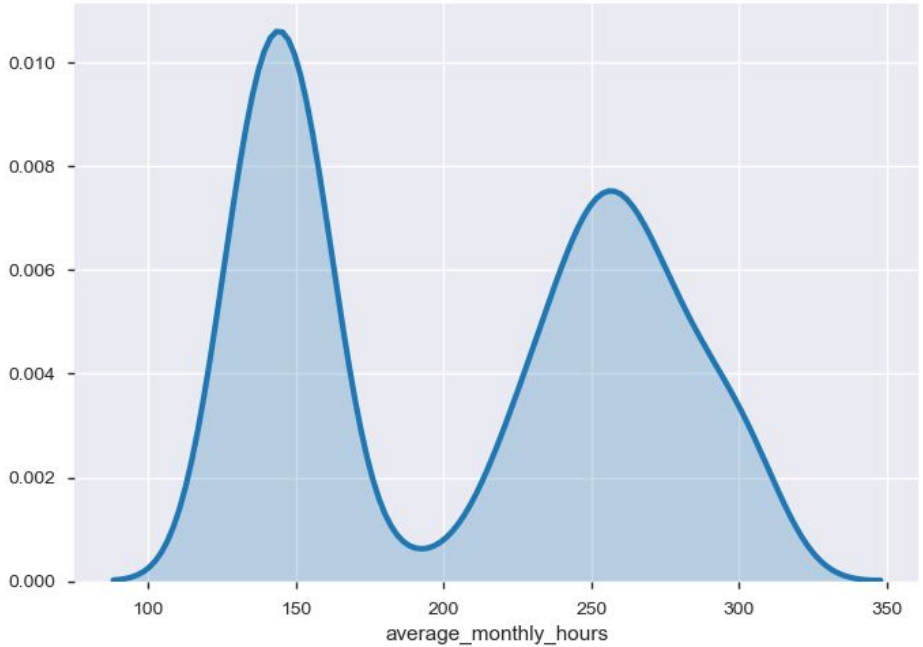
HR-Data would like to know which factors are highly affecting churn. This would help the organization take precautionary measures to minimize the rate at which employees are churning.

Current Churn



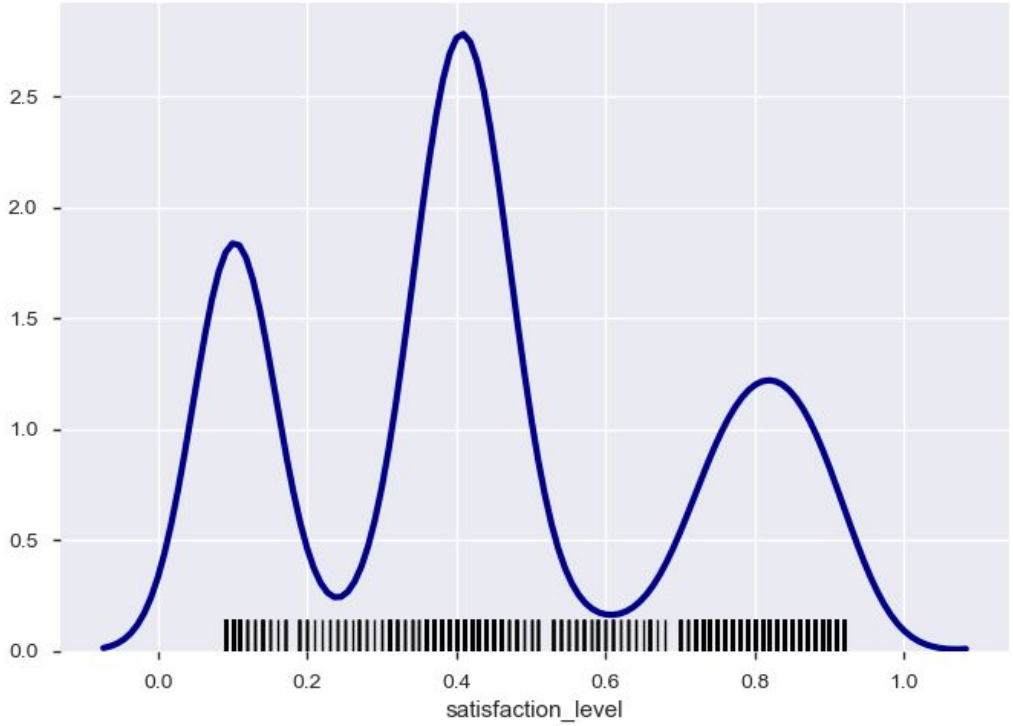
- Churn value is comparatively high, 23% approximately.
- Roughly 800 employees out of the total.
- 23% may seem like a low percentage. However, 800 employees leaving a company is substantial reason to find out why.
- Although this churn level is low it is still necessary for HR-Data to find out the factors driving this churn.

Average monthly hours



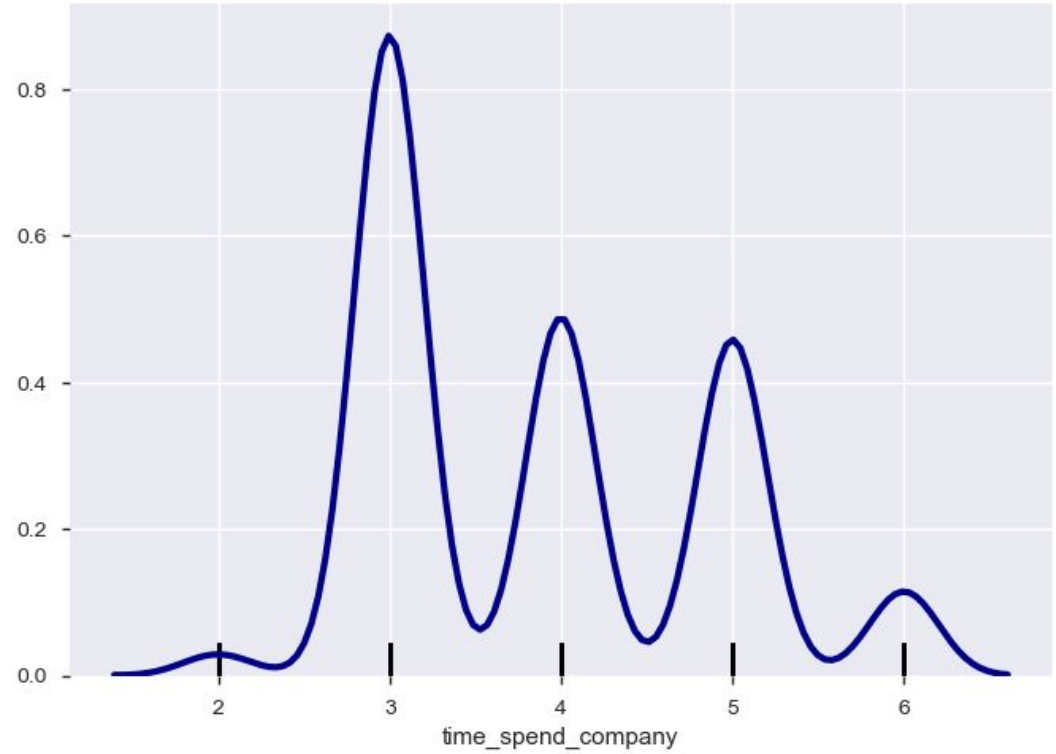
- The average number of hours worked by employees who churn is centered at 150 hours.
- There is also a significant amount of employees churning at 250 hours.

Satisfaction level



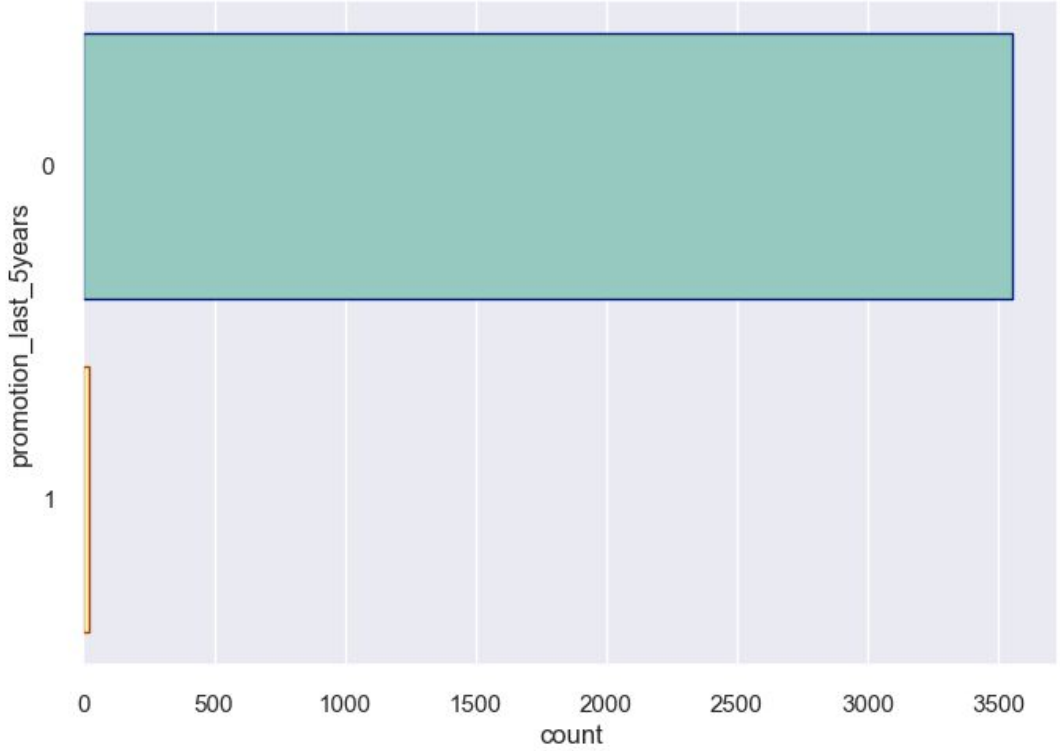
- Most employees churn at a satisfaction level of 0.4.
- This tells us employees were most likely unhappy with the company. This could be due to a number of factors such as micromanagement and happiness at the workplace.

Number of years spent at the company



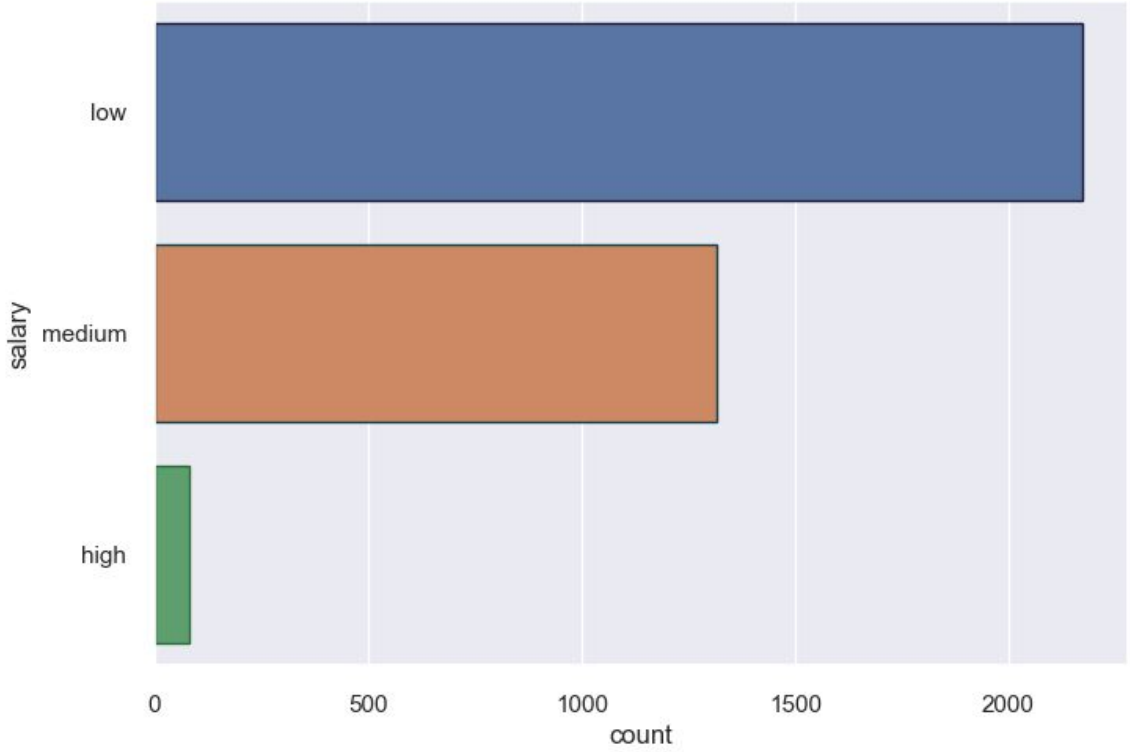
- Majority of employees churn after spending 3 years at HR-Data

Promotion



- Employees who churn have not been promoted in the last 5 years

Salary



- Employees who churn have a low salary.

A deep-dive

Recursive Feature Elimination

- A feature selection method that fits a model and removes the weakest feature (or features) until the specified number of features is reached.

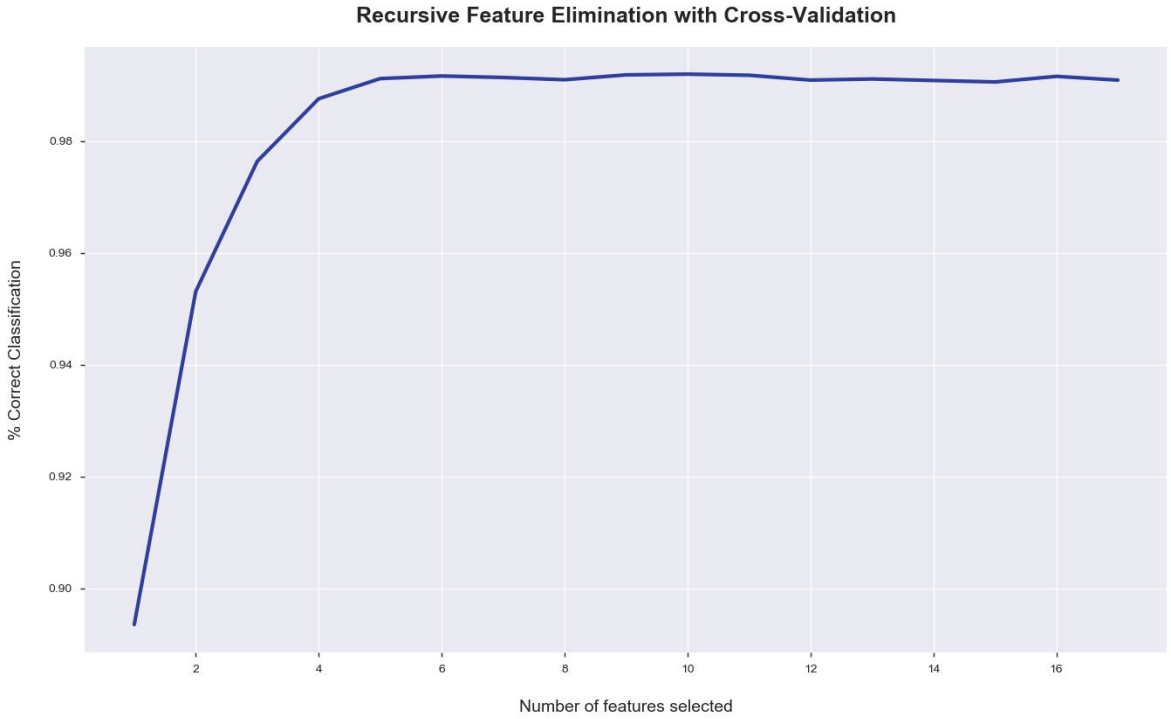
Recursive Feature Elimination

- Features are ranked by the model's coef or feature importance attributes, and by recursively eliminating a small number of features per loop.

Recursive Feature Elimination

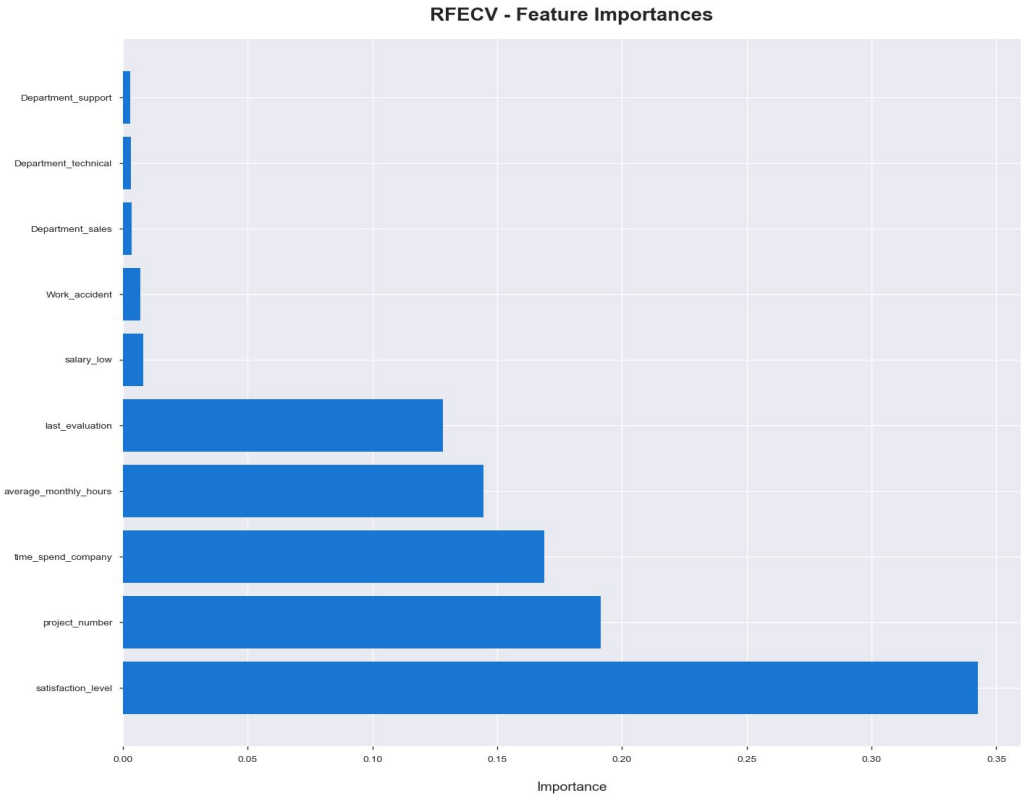
- To find the optimal number of features cross-validation is used with RFE to score different feature subsets and select the best scoring collection of features.

RFE



- It is visible that with 5-16 features the accuracy of the model is approximately 99%

RFE features in order of importance



Accuracy of the model

Precision

- Precision
- Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
- High precision relates to a low false positive rate.
- We have a precision of 0.49 which is low
- Precision = TP/TP+FP

Accuracy of the model

Recall

- Recall
- **Recall** (Sensitivity) - Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes
- We have a recall of 0.80 which is above 0.5. Our recall value is good for this model
- Recall = TP/TP+FN

Accuracy of the model

F1-score

- F1-Score
- F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.
- More useful than accuracy with an uneven class distribution
- We have an f1-score of 0.60
- F1 Score = 2*(Recall * Precision) / (Recall + Precision)

Recommendations

- Teleworking
- Consider promoting employees after 2 years
- Increase employee salary
- Reduce employee work load

THANK YOU