SALFORT Overview

05 April 2024 19:42

1) Objective: Employee likely to quit ?

Factors that contribute to their leaving ?

Employer retention strategies.

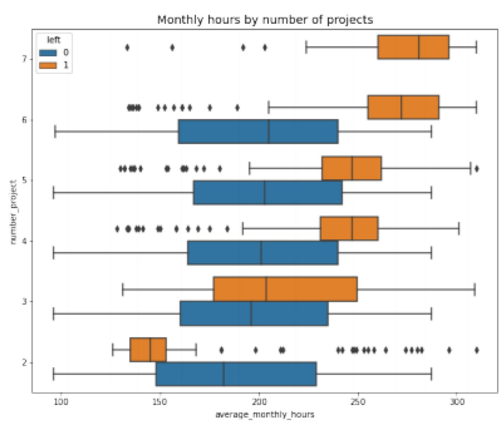
14999 observations/Rows , 10 columns.

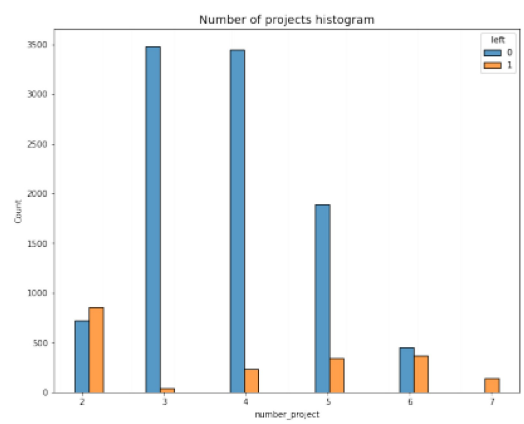
2) Dataset ✓

3) Exploratory Data Analysis (Some of the highlights of Analysis below)

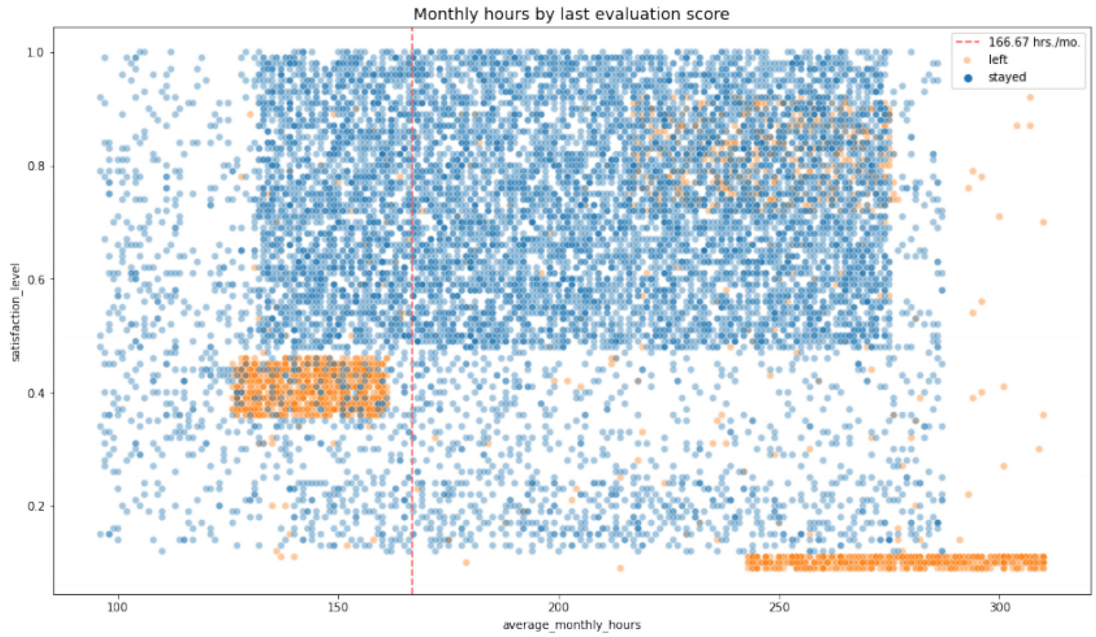
Relationships:

1) No. of projects, No. of working hours , Did they leave ?

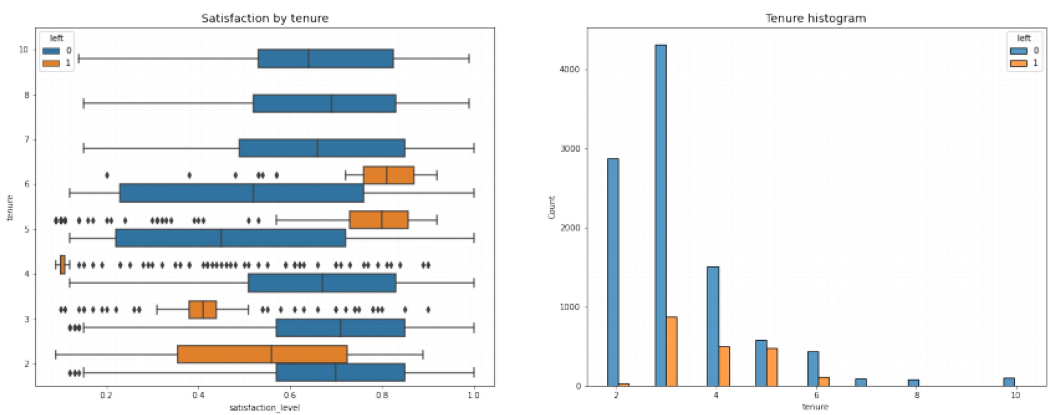




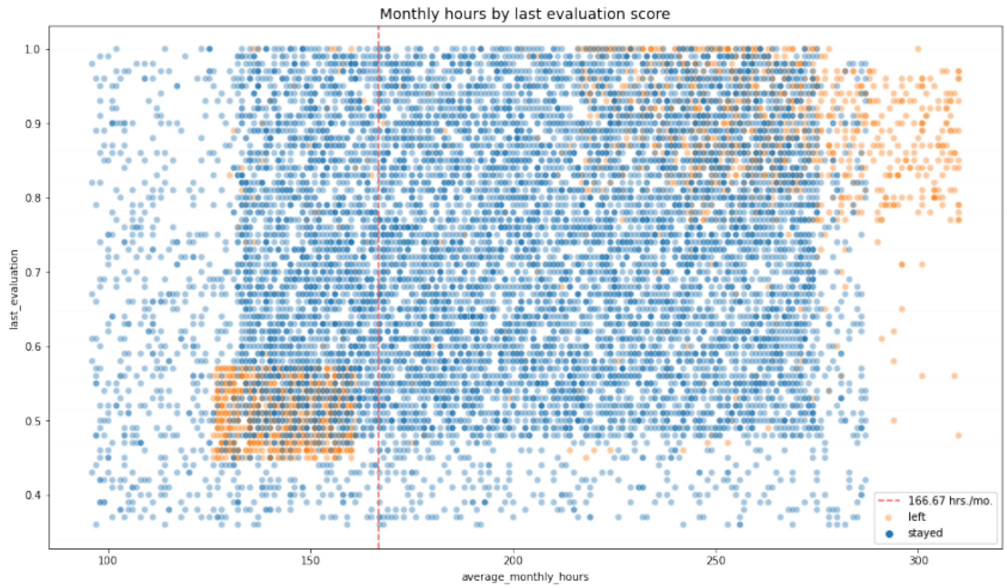
2) No. of avg. working hours (monthly) + satisfaction levels ?



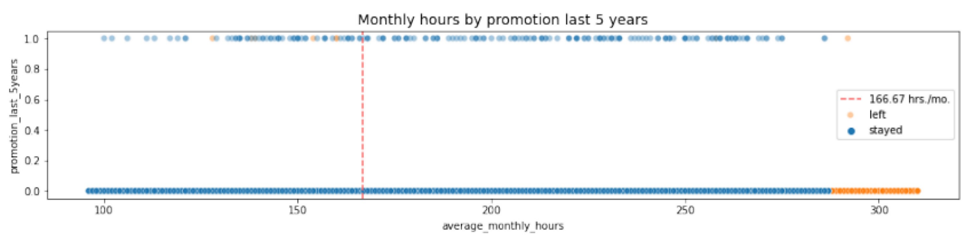
3) Satisfaction levels by tenure



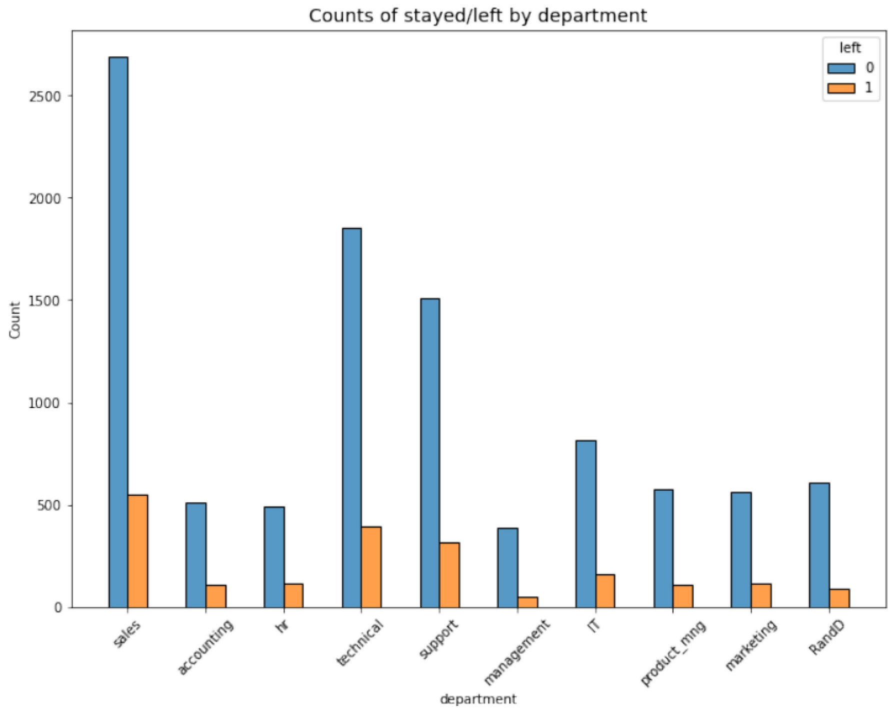
4) No. of monthly working hours & last evaluation score


Monthly hours by last evaluation score
last_evaluation
average_monthly_hours
166.67 hrs./mo.
left
stayed

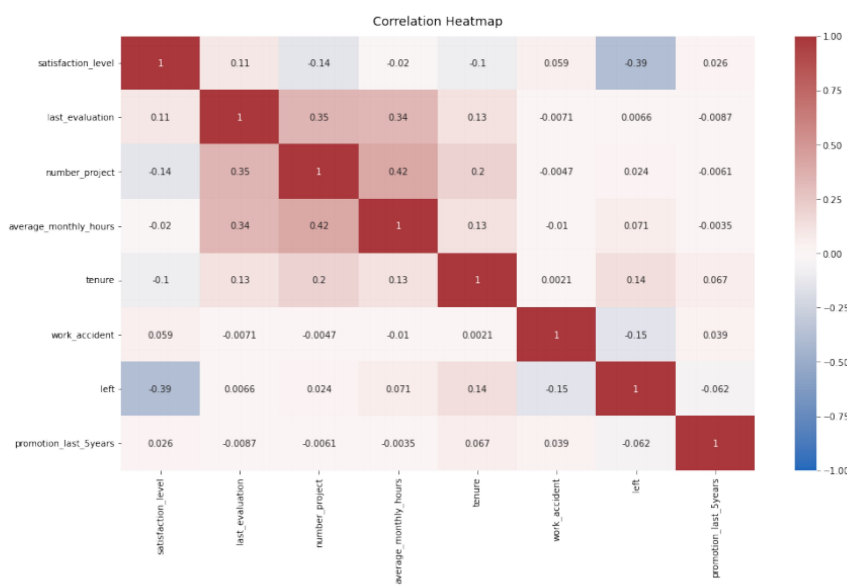
5)


Monthly hours by promotion last 5 years
promotion_last_5years
average_monthly_hours
166.67 hrs./mo.
left
stayed

6)



7)



MODELS :

1) logistic Model (Data split : 83% : 17%)

1) logistic Model (Data split : 85% : 15%)

→ precision = 80%

→ Recall = 83%

→ f1 score = 80%

→ Accuracy = 83%

2) Decision Tree

i) Hyper parameter Tuning (Best parameters)

Max depth = 4

Min Samples leaf = 1

Min samples split = 2

3) Random forest

Hyper parameter Tuning

```
{'max_depth': 5,
 'max_features': 1.0,
 'max_samples': 0.7,
 'min_samples_leaf': 1,
 'min_samples_split': 3,
 'n_estimators': 500}
```

[57]:

	model	AUC	precision	recall	f1	accuracy
0	random forest1 val	0.954	0.955	0.917	0.936	0.979
0	decision tree1 val	0.952	0.924	0.920	0.922	0.974

ROUND -2 :

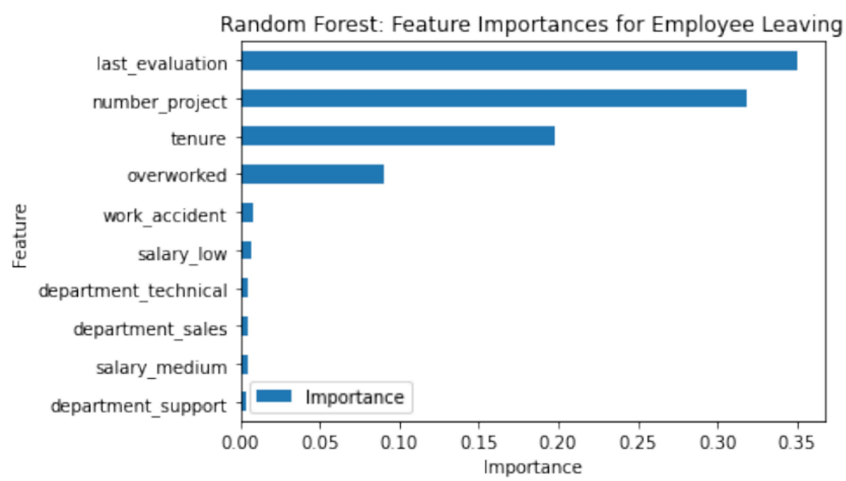
After feature Engineering considering data leaks

(dropped) Satisfaction level

New feature (overworked & not)

	model	AUC	precision	recall	f1	accuracy
0	decision tree2 val	0.942	0.883	0.907	0.895	0.965
0	random forest2 val	0.933	0.905	0.884	0.895	0.965

Random Forest Model seems to be the best 4 final



RESULTS, INTERPRETATION & INSIGHTS PRESENTATION

Logistic Regression

The logistic regression model achieved precision of 80%, recall of 83%, f1-score of 80% (all weighted averages), and accuracy of 83%, on the test set.

Tree-based Machine Learning

After conducting feature engineering, the decision tree model achieved AUC of 94.3%, precision of 86.5%, recall of 91.5%, f1-score of 88.9%, and accuracy of 96.2%, on the test set. The random forest model slightly outperformed the decision tree model.

2.5.2 Conclusion, Recommendations, Next Steps
The models and the feature importances extracted from the models confirm that employees at the company are overworked.
To retain employees, the following recommendations could be presented to the stakeholders:

- Cap the number of projects that employees can work on.
- Consider promoting employees who have been with the company for atleast four years, or conduct further investigation about why four-year tenured employees are so dissatisfied.
- Either reward employees for working longer hours, or don’t require them to do so.
- If employees aren’t familiar with the company’s overtime pay policies, inform them about this. If the expectations around workload and time off aren’t explicit, make them clear.
- Hold company-wide and within-team discussions to understand and address the company work culture, across the board and in specific contexts.
- High evaluation scores should not be reserved for employees who work 200+ hours per month.

Consider a proportionate scale for rewarding employees who contribute more/put in more effort.

Next Steps
It may be justified to still have some concern about data leakage. It could be prudent to consider how predictions change when last_evaluation is removed from the data. It’s possible that evaluations aren’t performed very frequently, in which case it would be useful to be able to predict employee retention without this feature. It’s also possible that the evaluation score determines whether an employee leaves or stays, in which case it could be useful to pivot and try to predict performance score. The same could be said for satisfaction score.