Salifort Motors Data Project

Solving high employee turnover with data-driven insights

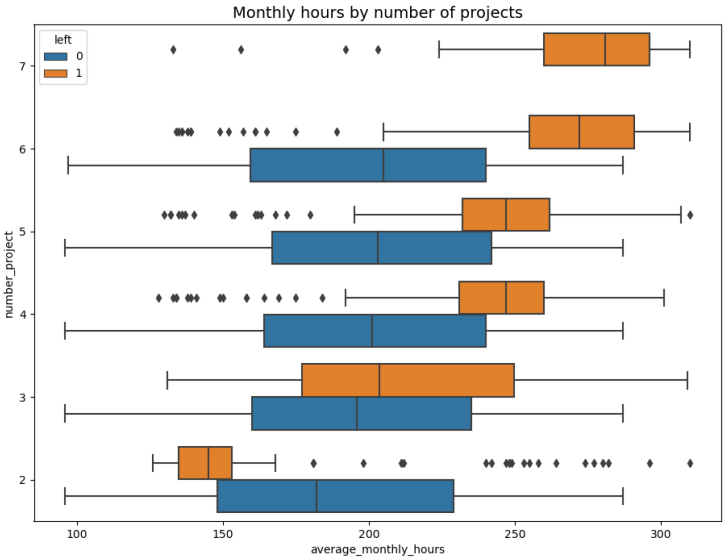
OVERVIEW

Salifort Motors has been experiencing high turnover, leading to increased costs in recruitment, training, and upskilling. Through exploratory data analysis and predictive models on employee survey data, this project aims to find insights into the reasons for the high turnover, and to build models that accurately predict which employees might leave.

METHODOLOGY

The data was first cleaned by checking for column datatypes, nulls, duplicates, and outliers. Pairplots and heatmaps were then used to explore and identify potential relationships between variables. Boxplots and scatterplots were then used to analyse these relationships further.

As the task was supervised binary classification (whether an employee left or not), decision tree models were chosen (Random Forest and XGBoost), to be trained, tuned, and evaluated.



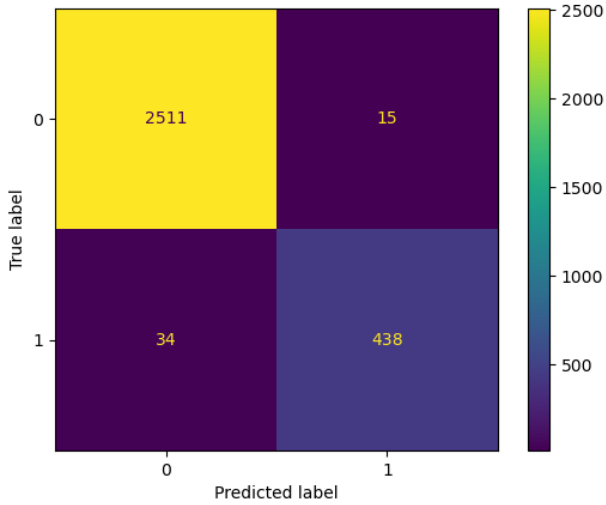
KEY INSIGHTS

EXPLORATORY DATA ANALYSIS

It was found that employee satisfaction levels, last evaluation scores, hours worked, and number of projects had strong influence on employees who left.

Three clusters of employees who left were also identified, based on their features:

1. Poor performers who tended to have weak last evaluation scores, few projects, and low work hours
2. Overworked, disatisfied employees who had too many projects assigned or worked too many hours each month
3. Strong performers who put in long hours and reflected that they were satisfied, but still left



PREDICTIVE MODEL

Random Forest and XGBoost models were trained and tested with GridsearchCV, and were both found to have high precision, recall, and accuracy (RF – 0.975, 0.924, 0.984 | XGB – 0.982, 0.928, 0.986). The XGBoost model was marginally better. Feature importance reinforced that hours worked, satisfaction level, last evaluation, and number of projects affected were good predictors.

SUGGESTIONS & NEXT STEPS

- Investigation into the company working culture, with overworking at the forefront.
- Deeper exploration into internal and external reasons for high-value, satisfied employees leaving the company.
- Utilisation of the XGBoost model on new employee surveys to identify potential leavers.
- Train new models to predict satisfaction levels and evaluation scores, using additional data.