Salifort Motors

Employee Retention Project

<Name> 林 龍蔵
<Data> 2023/10/16



目次

- 概要 & 課題
- 探索的データ分析
- モデル選択
- モデル構築と評価
- 結論

概要&課題

- Salifort Motors は、従業員のデータを使用して 従業員の離職理由を判断する方法を模索してい る
- Salifort Motors の人事部は、従業員の満足度を 向上させる取り組みをしたい

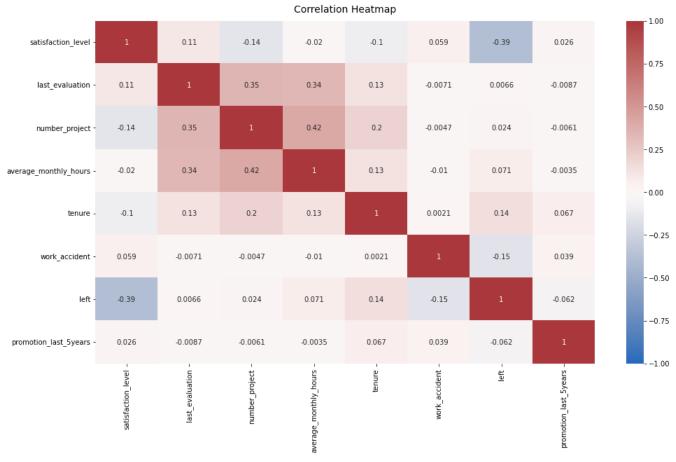
• 従業員を離職される原因とは?



SECTION1

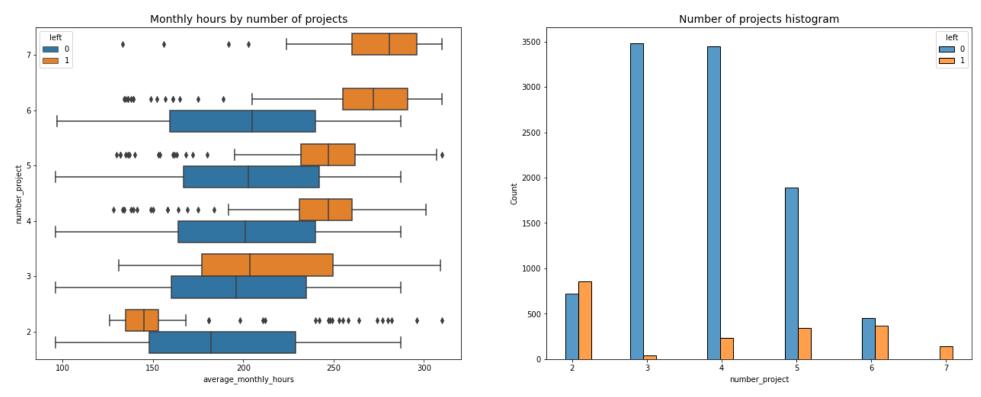
探索的データ分析

相関ヒートマップ



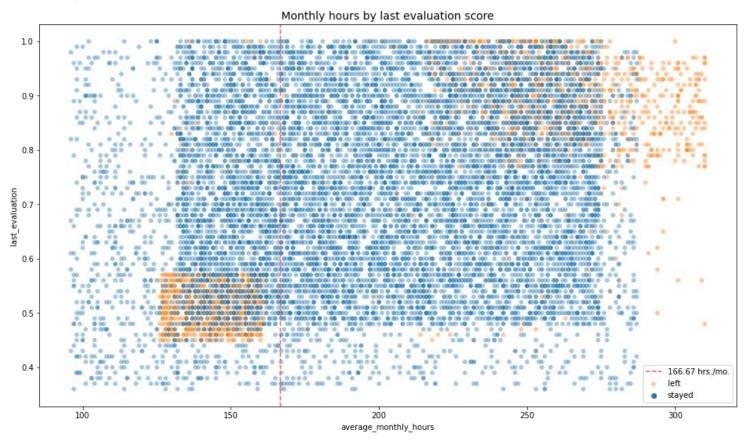
- ・最近の評価、プロジェクトの数、月平均労働時間に正の相関がみられる
- ・満足度と退職には、負の相関がみられる

プロジェクトの数に対する月平均労働時間



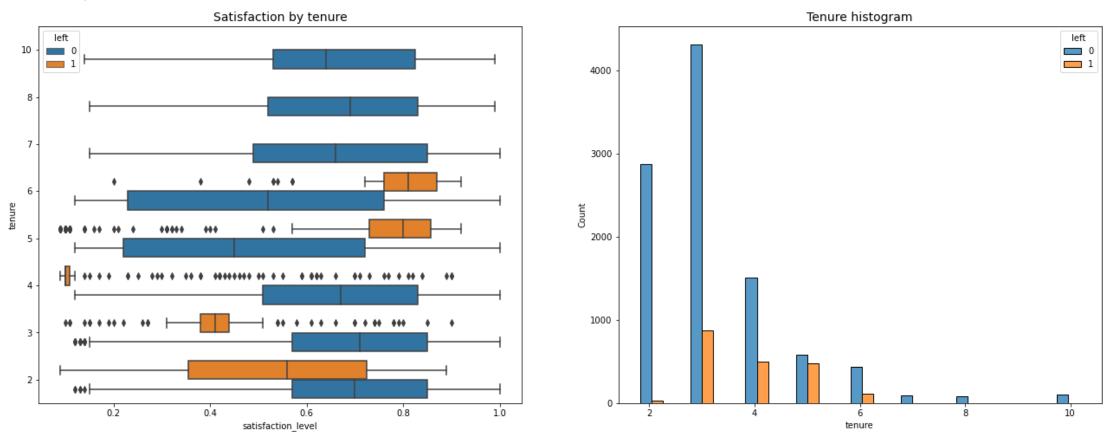
- ・プロジェクトが7つあった人は全員退社しており、6つあった人も月平均労働時間は、 ほかのグループと比べはるかに長い
- ・最適なプロジェクト数は、3~4であることがわかる

最近の評価に対する月平均労働時間



- ・退職した人たちには、月平均労働時間が長く評価が高いグループと月平均労働時間が短く評価が低いグループに分けられる
- ・全体的に週40時間労働を超えて働いている

勤続年数に対する満足度



- ・勤続年数が短く満足度が低いグループと勤続年数が長く満足度が高いグループに分けられる
- ・4年目の退職者の満足度が非常に低い



モデル選択

- 予測したい変数が従業員が退職するか、退職しないかのカテゴ リー変数
- よって、ツリーベースの機械学習モデルを構築する方が適切
- 今回は、ディシジョンツリーモデルとランダムフォレストモデルを実装



SECTION3

モデル構築と評価

モデル比較

評価スコア

	model	precision	recall	F1	accuracy	auc
0	decision tree cv	0. 914552	0. 916949	0. 915707	0. 971978	0. 969819
	model	precision	recall	F1	accuracy	auc
0	random forest cv	0. 950023	0. 915614	0. 932467	0. 977983	0. 980425

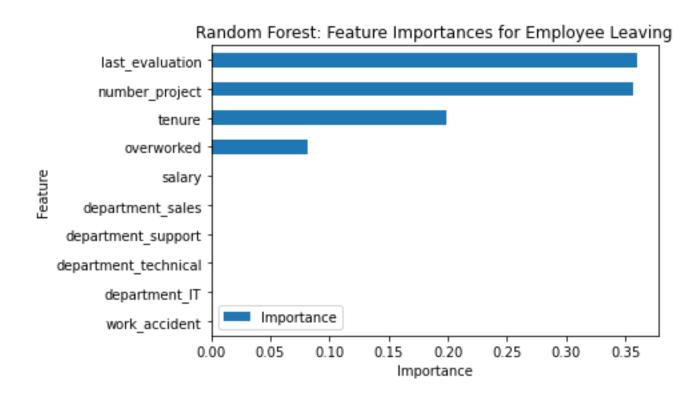
テストスコア

	model	precision	recall	F1	accuracy	auc
0	decision tree cv	0.914552	0.916949	0.915707	0.971978	0.969819
	mod	del precisi	on rec	all	f1 accura	cy Al

0 random forest1 test 0.964211 0.919679 0.941418 0.980987 0.956439

・わずかにランダムフォレストモデルの方が今回のデータに適している

変数の重要度



・評価とプロジェクトの数が退職に直結していることがわかる **※overwork=monthly_average_hours**



結論

プロジェクトの数を平均3~4つになるようマネジメントをするべき

•全体的に労働時間を少なくし、特に長時間労働はなくすべき

・勤務4年目の従業員に対して、不満足の理由をさらに調 査する必要あり

