

YOUR LOGO

2025 公司员工流失预测分析 项目汇报

PowerPoint design



主讲人：马龙



时间：2025.4



CONTENTS

1. 项目背景与目标

2. 数据探索与清洗

3. 数据分析与可视化

4. 模型构建与评估

5. 结果解读与建议

6. 总结与展望

YOUR LOGO

项目背景与目标

PowerPoint design

PART 01



项目背景



数据现状与需求

目前，Salifort Motors公司的员工流失率很高。公司高级领导团队致力于创造一种支持员工成功和职业发展的企业文化，所以他们对高流失率非常担忧。此外，高流失率在财务上也是一项巨大成本，让公司在招聘、培训和提升员工技能方面投入了大量资金。如果能够预测员工是否会离职，并发现他们离职的原因，他们就能更好地理解问题并制定解决方案。



项目意义

- 降低员工流失率，减少招聘与培训成本。
- 提升员工满意度与忠诚度，增强企业竞争力。

项目目标

数据分析目标

- 深入分析员工数据，构建精准的离职预测模型。
- 识别影响员工离职的关键因素，为决策提供依据。

01

02

业务改进目标

- 提出针对性措施，提高员工留存率。
- 优化人力资源管理流程，提升管理效率。

YOUR LOGO

数据探索与清洗

PowerPoint design

PART 02



数据概览



数据规模与结构

- 数据集包含 15,000 行数据，10 列变量，涵盖员工满意度、绩效评估、工作时长等多维度信息。
- 数据类型多样，包括浮点型、整数型和对象型变量。

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 14999 entries, 0 to 14998
```

```
Data columns (total 10 columns):
```

#	Column	Non-Null Count	Dtype
0	satisfaction_level	14999 non-null	float64
1	last_evaluation	14999 non-null	float64
2	number_project	14999 non-null	int64
3	average_monthly_hours	14999 non-null	int64
4	time_spend_company	14999 non-null	int64
5	Work_accident	14999 non-null	int64
6	left	14999 non-null	int64
7	promotion_last_5years	14999 non-null	int64
8	Department	14999 non-null	object
9	salary	14999 non-null	object

```
dtypes: float64(2), int64(6), object(2)
```

```
memory usage: 1.1+ MB
```

数据概览



基本统计信息

展示关键变量的计数、均值、标准差等统计指标，为后续分析提供基础数据支持。

通过统计分析初步了解数据分布特征。

	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	left	promotion_last_5years
count	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000	14999.000000
mean	0.612834	0.716102	3.803054	201.050337	3.498233	0.144610	0.238083	0.021268
std	0.248631	0.171169	1.232592	49.943099	1.460136	0.351719	0.425924	0.144281
min	0.090000	0.360000	2.000000	96.000000	2.000000	0.000000	0.000000	0.000000
25%	0.440000	0.560000	3.000000	156.000000	3.000000	0.000000	0.000000	0.000000
50%	0.640000	0.720000	4.000000	200.000000	3.000000	0.000000	0.000000	0.000000
75%	0.820000	0.870000	5.000000	245.000000	4.000000	0.000000	0.000000	0.000000
max	1.000000	1.000000	7.000000	310.000000	10.000000	1.000000	1.000000	1.000000

数据清洗过程



列名标准化

将列名统一为蛇形命名法，修正拼写错误，如将“average_montly_hours”改为“avg_hours”，提升数据可读性。
标准化列名便于后续数据处理与分析。



缺失值与重复值处理

检查发现数据集中无缺失值，但存在重复行数据。
删除重复行后数据集从 14999 行减少到 11991 行，确保数据的唯一性与准确性。



异常值检测

通过箱线图等方法检测到部分变量存在异常值，如“avg_hours”和“time_spend”。
根据模型需求决定是否去除异常值，以避免对分析结果产生干扰。

YOUR LOGO



数据分析与可视化

PowerPoint design

PART 03

离职员工占比



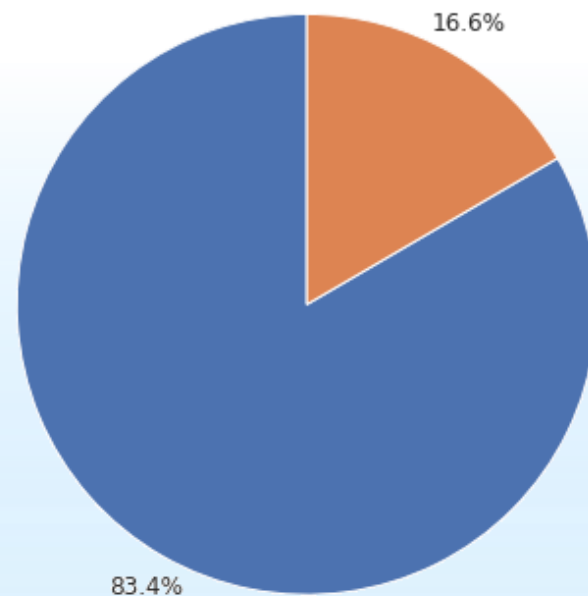
离职情况统计

离职员工数量为 1991 人，离职员工占比为 16.60%。存在明显的类别不平衡现象，不过需要根据后续的建模情况来判断是否需要采取类别平衡。

Ratio of active to separated employees



left



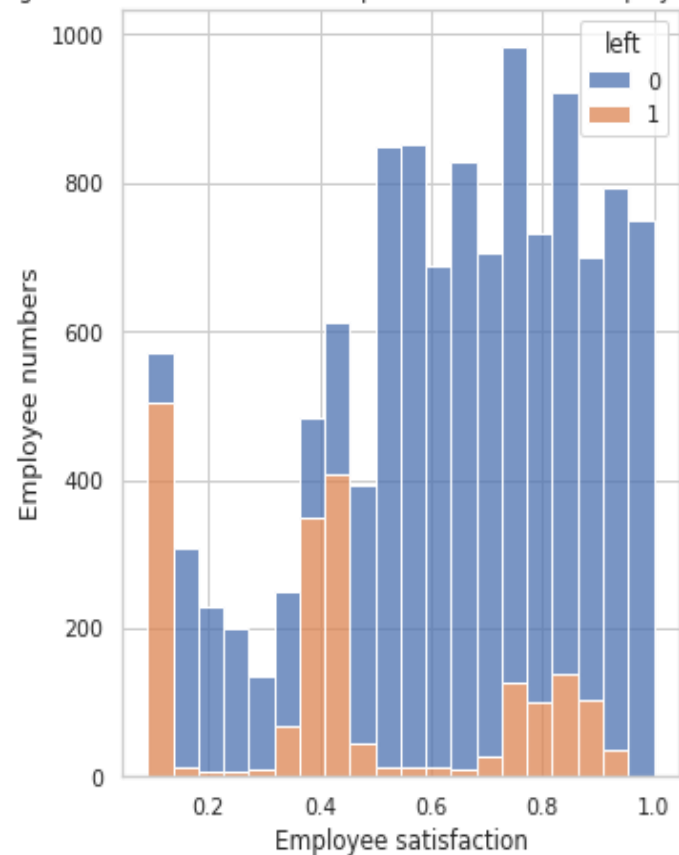
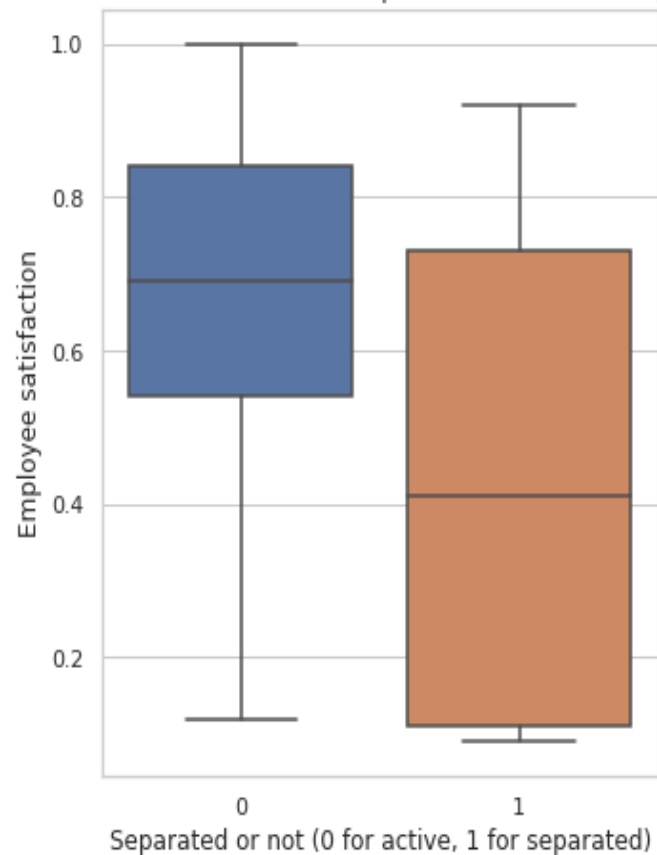
变量关系探索（一）



满意度与离职关系

箱线图显示离职员工满意度普遍较低，直方图进一步对比不同离职状态下员工满意度的分布。分析结果表明，员工满意度是影响离职的重要因素之一。

Distribution of satisfaction of separated and current employeesHistogram of the Distribution of Separated and Active Employee Satisfaction

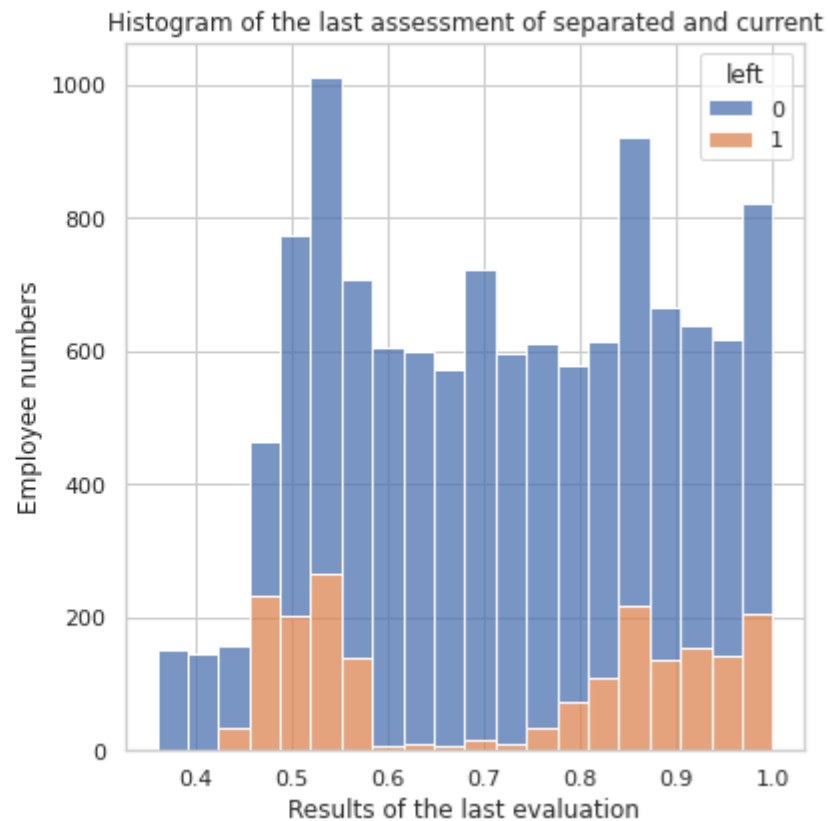
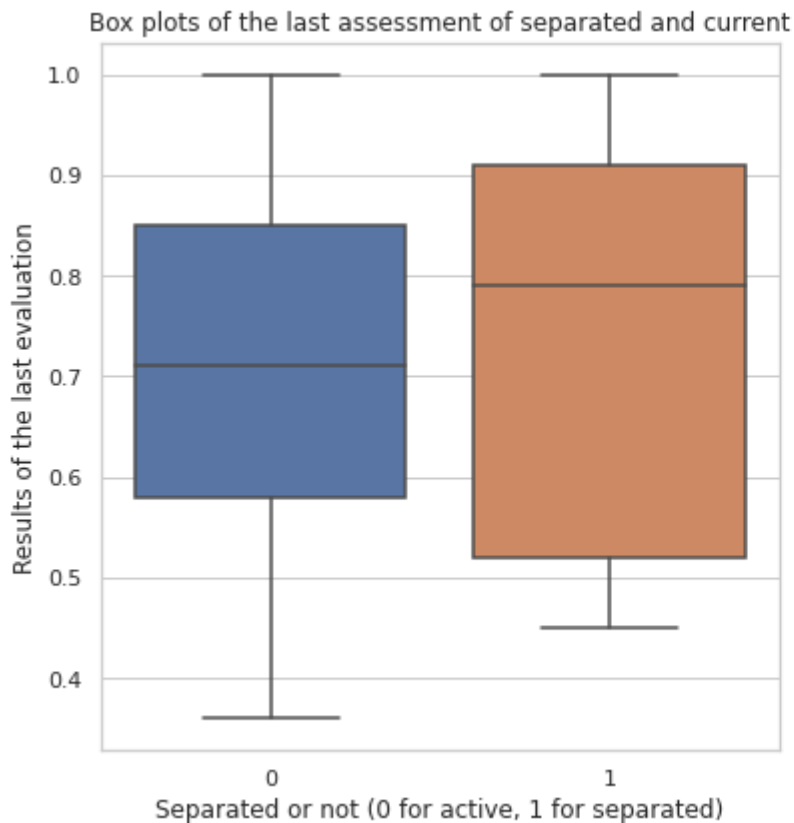


变量关系探索（一）



评估结果与离职关系

箱线图呈现离职与在职员工最近一次评估结果的分布差异，直方图分析两者关系，可以看到评估结果过低和过高时，离职率都会有上升。绩效评估得分与离职情况存在一定的关联性



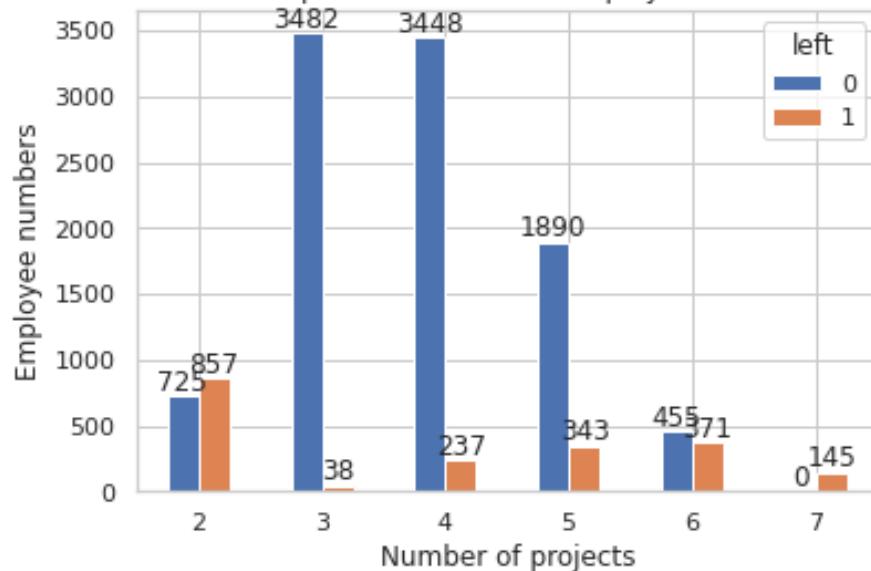
变量关系探索（二）



项目数量与离职关系

柱状图展示不同项目数量下离职员工和在职员工的数量对比，发现项目数量为2时，离职员工比例超过50%，项目数量为7时，所有员工全部离职了，说明让员工参与过低和过高数量的项目都是不合适的。合理分配项目任务对降低员工离职率具有重要意义。

Comparison of the number of separated and active employees under different project numbers



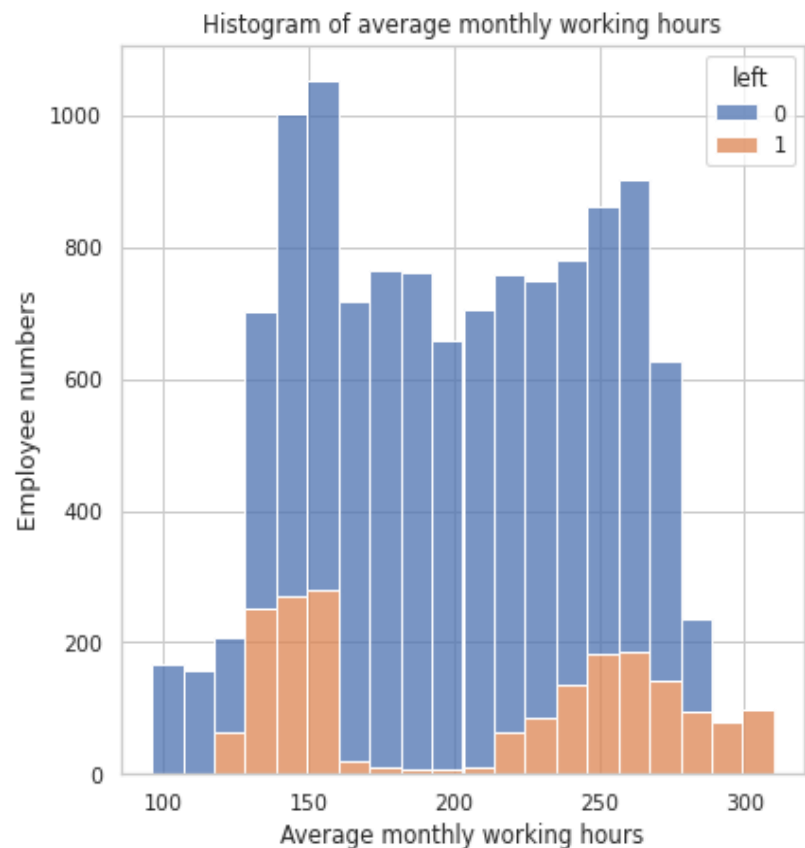
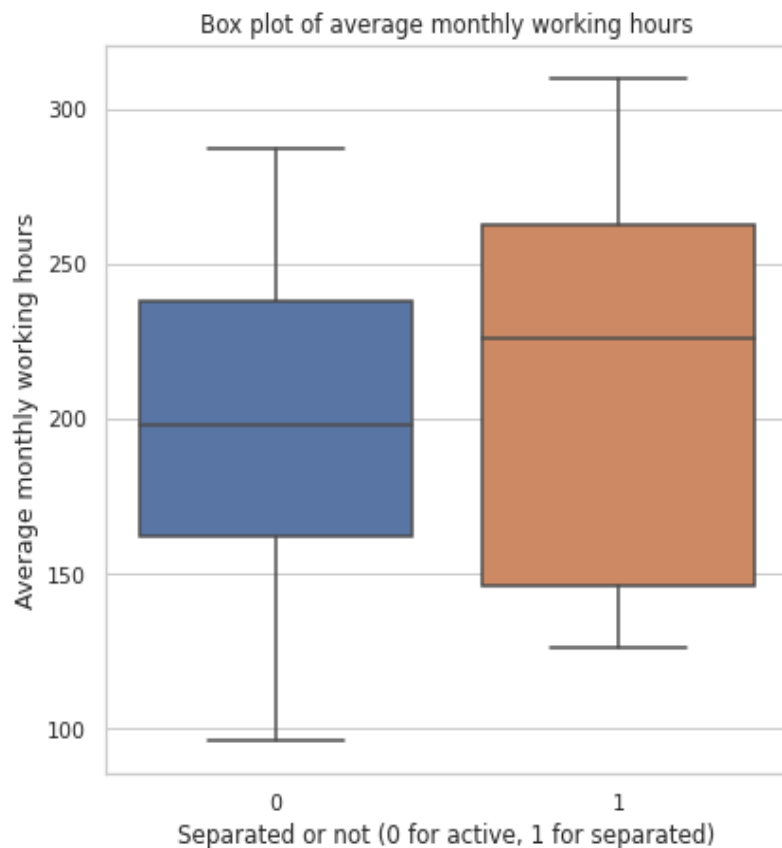
变量关系探索（二）



工作时长与离职关系

使用箱线图和直方图分析平均每月工作时长与离职的关系，可以看到工作时长过长或过短都让离职率变得更高。

所以，工作时长的合理性直接影响员工的工作满意度和留存意愿

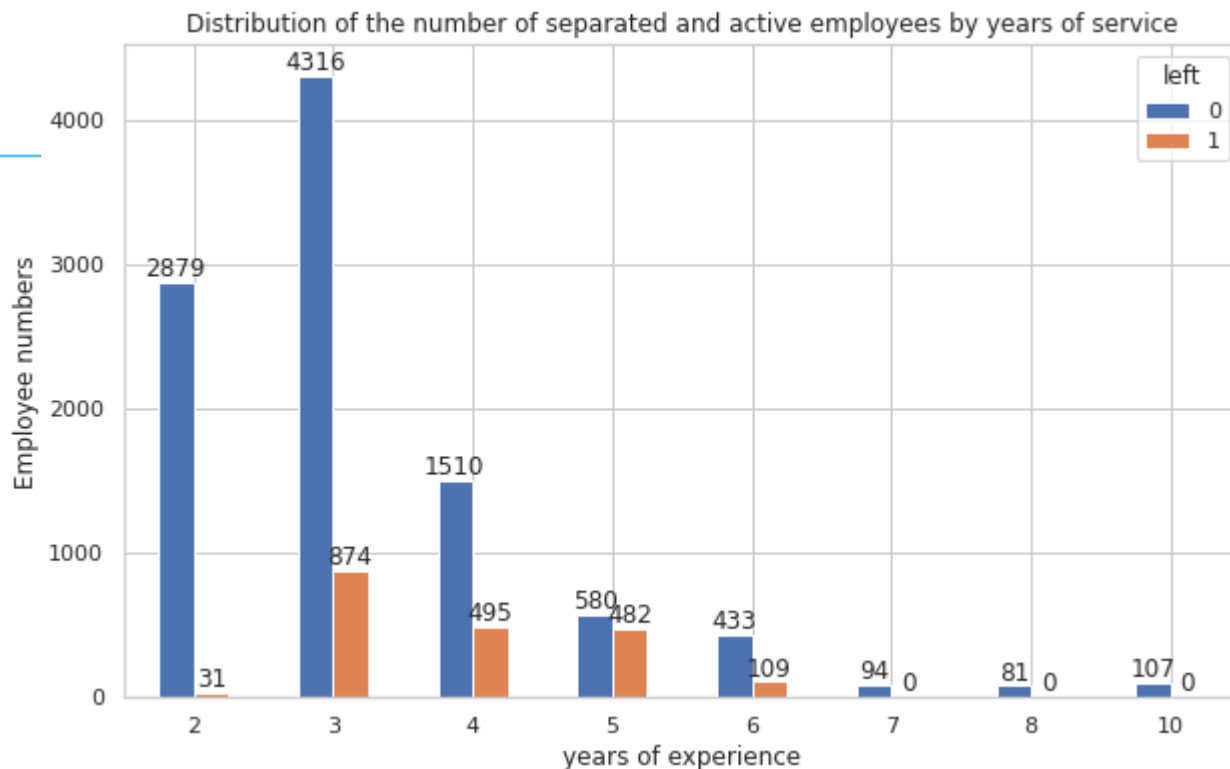


变量关系探索（三）

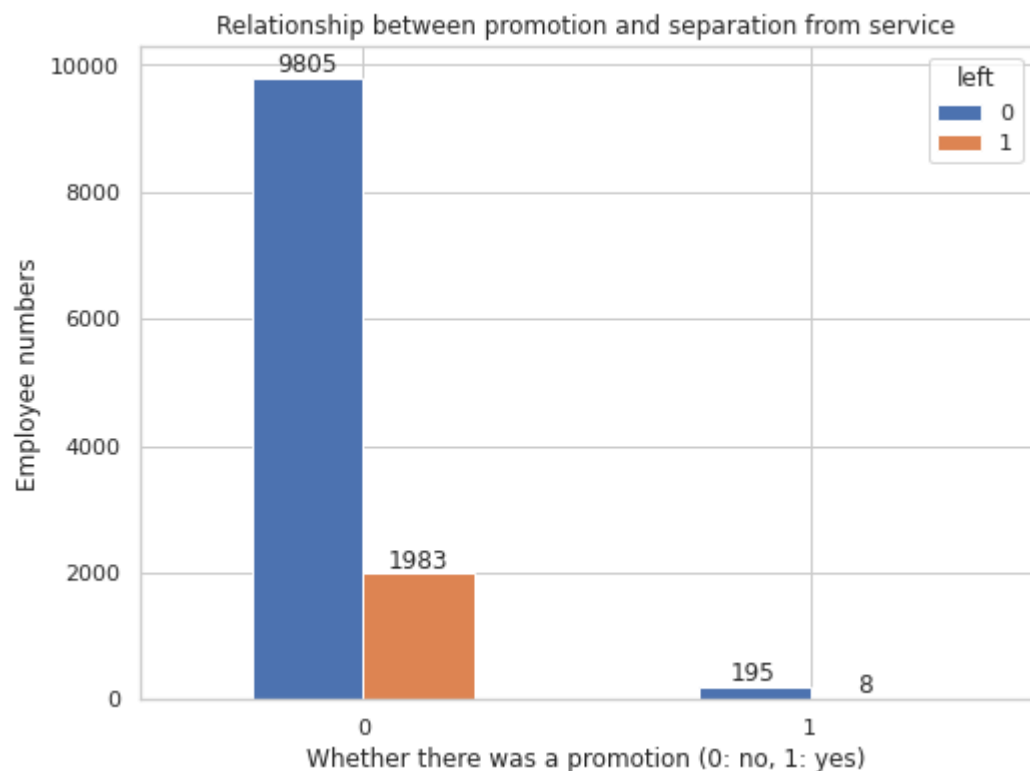
01

工作年限与离职关系

柱状图展示随着员工工作年限的变化，员工离职变化的情况。可以看到离职较多发生在员工入职后的第三、第四和第五年。员工入职时间较短以已经长期稳定下来的这部分的员工离职率相当低。所以，工作年限是员工留存的重要相关因素之一。



变量关系探索（三）



02

晋升与离职关系

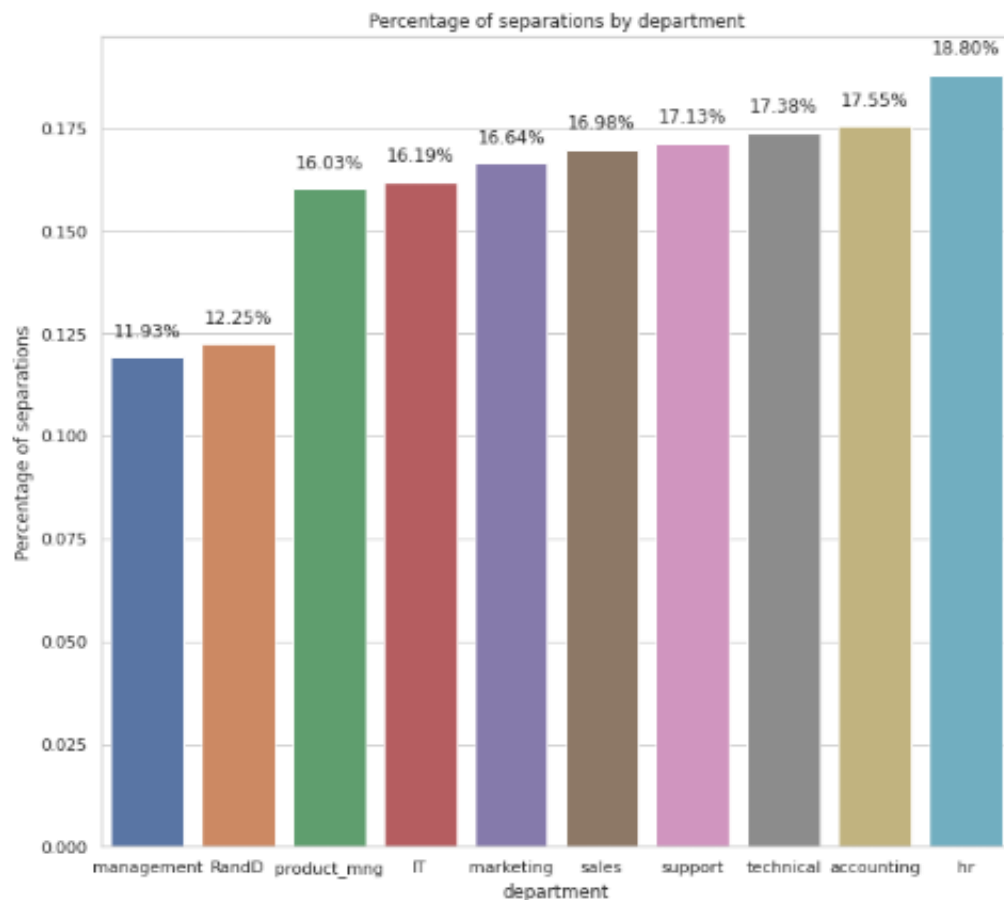
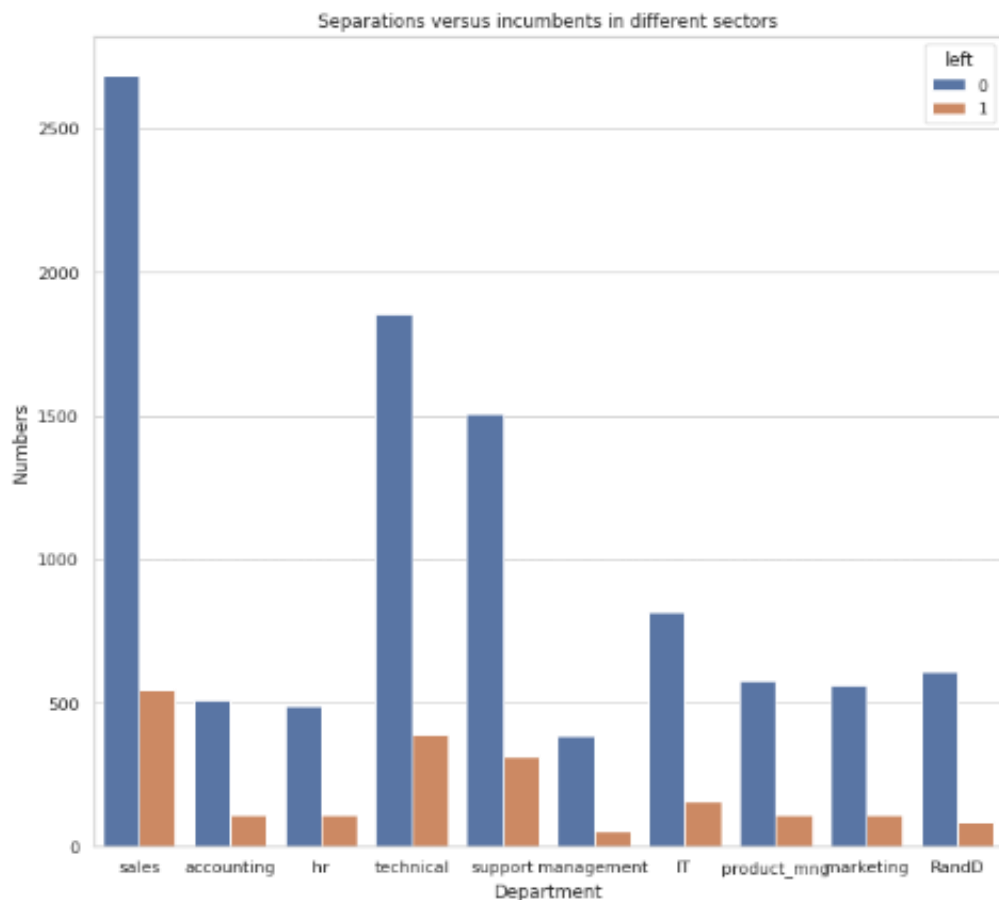
柱状图呈现近 5 年有晋升和无晋升员工的离职数量对比，可以看到公司在最近五年晋升的人数非常少，而且未晋升的员工离职率显著高于晋升员工。所以，晋升机会对员工的职业发展和留存意愿具有显著影响。

变量关系探索（四）

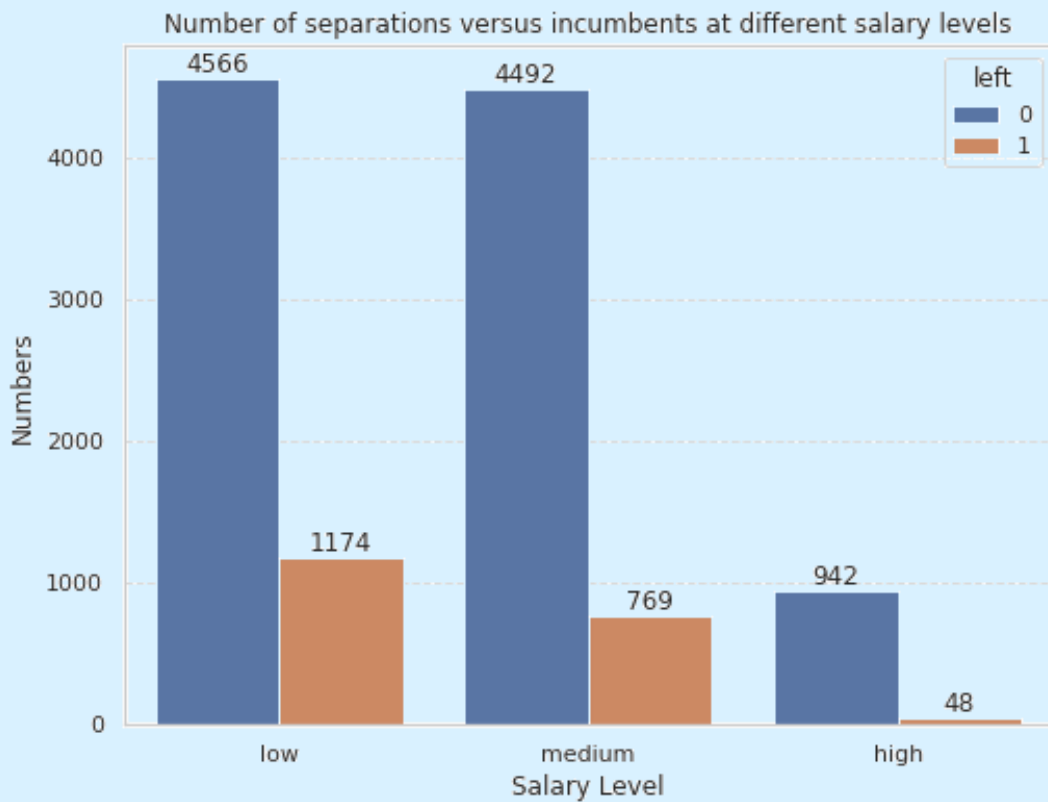


部门与离职关系

计数图展示不同部门离职员工和在职员工的数量分布，柱状图呈现各部门的离职率。可以看到管理层和科研部门的离职率最低，HR部门的离职率最高。几个人数比较多的部门，比如销售部、技术部和支持部的离职率都偏高，可以重点关注。



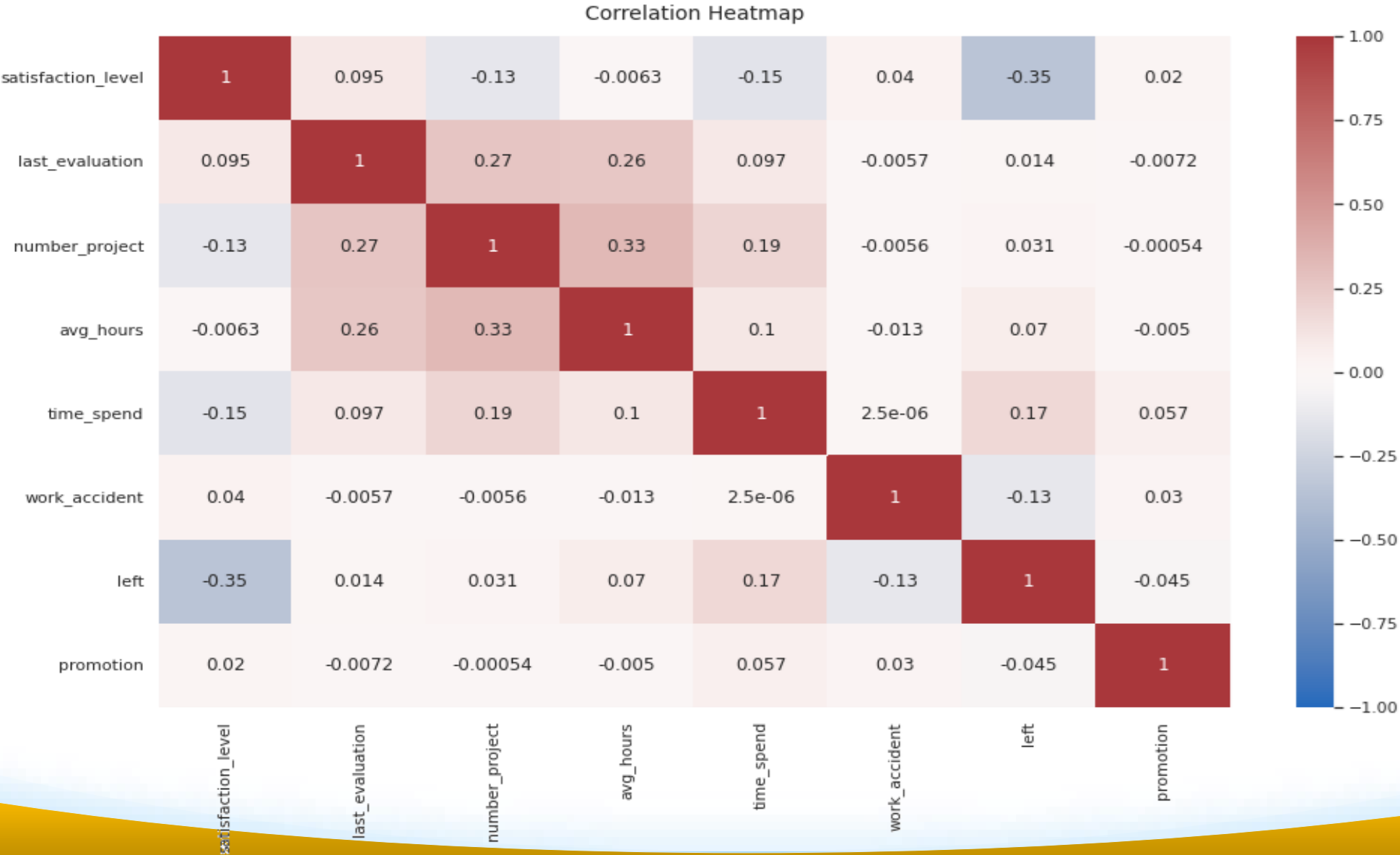
变量关系探索（四）



薪资与离职关系

计数图展示不同薪资水平下离职员工和在职员工的数量对比，低薪资员工离职率相对较高，高薪资员工离职率非常低。薪资水平是影响员工离职的重要因素之一。

相关性分析



YOUR LOGO

模型构建与评估

PowerPoint design

PART 04



模型选择依据



01

预测任务类型

本次任务为二分类预测任务，目标是判断员工是否离职。存在类别不平衡现象，不过并不极端，初期可以不做处理，后续如果预测效果较差，可以尝试采用升采样处理
明确任务类型为模型选择提供依据。



02

随机森林模型优势

随机森林能处理复杂的非线性关系，对数据分布要求不高，不易过拟合，可评估特征重要性。这些优势使其成为本次预测任务的理想选择。同时，在采用了xgboost模型和随机森林预测做出对比。

数据预处理

数值变量标准化

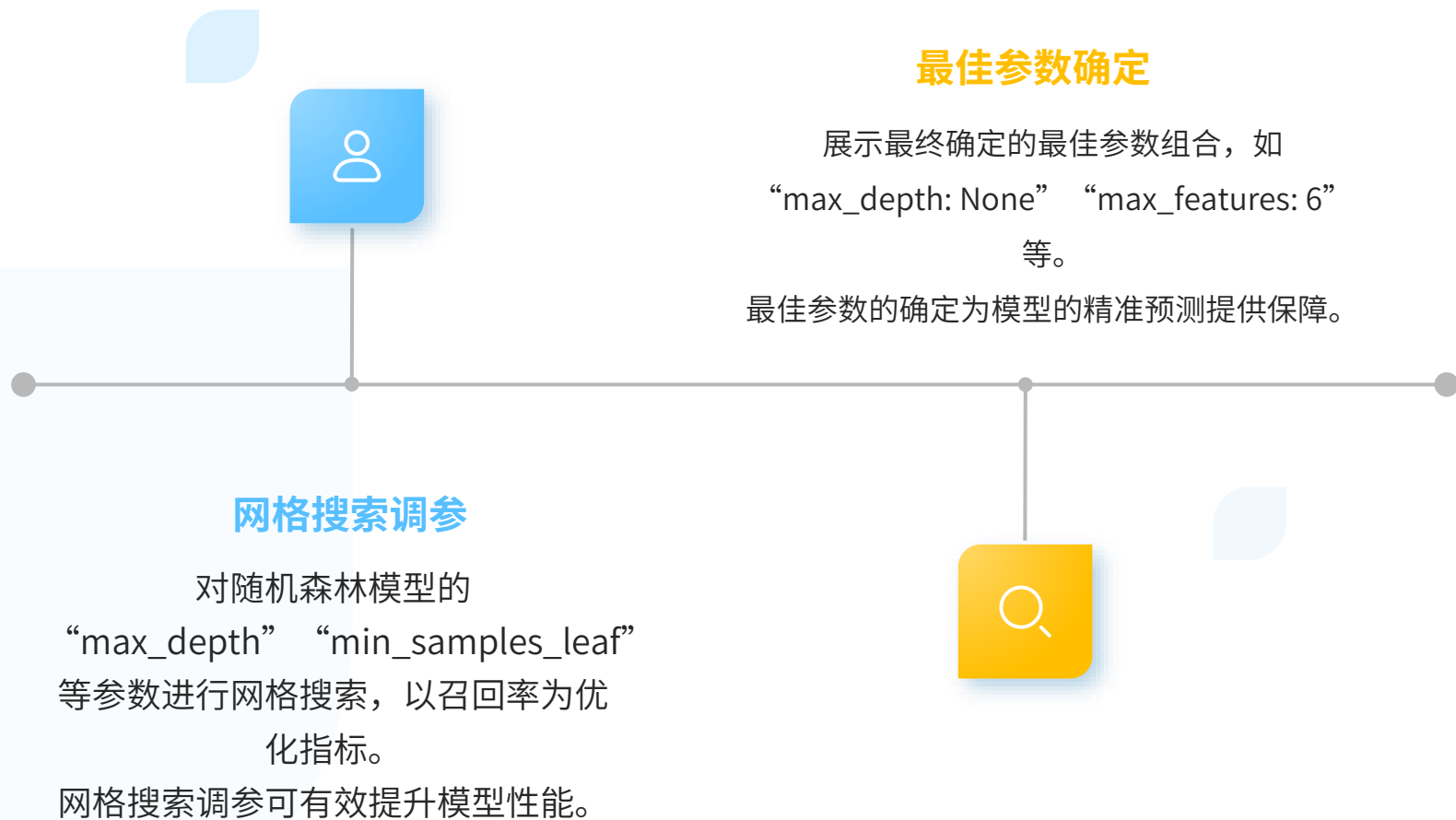
对“satisfaction_level” “last_evaluation” 等数值变量使用标准化方法处理，使其具有相同尺度。虽然基于树的模型对缩放并不敏感，同时数据存在少量异常值，基于业务实际判断，并没有删除。标准化处理有助于减少其对树分裂过程的干扰，从而提升模型训练效果。



分类变量编码

对“salary” 变量使用序数编码，将薪资等级转换为数值。
对“Department” 变量使用独热编码，将部门类别转换为数值向量。

模型训练与调优



模型评估结果

✓ 评估指标展示

展示随机森林模型的 F1 值、召回率、精确率和准确率，评估模型性能。

通过多指标评估全面了解模型的预测能力。

✓ 混淆矩阵呈现

呈现模型预测结果的混淆矩阵，直观展示模型对离职和在职员工的预测准确性。

混淆矩阵可清晰呈现模型的预测误差分布。

✓ 特征重要性分析

列出影响员工离职的重要特征及重要性排序，如满意度、项目数量等。

特征重要性分析为业务决策提供关键依据。

✓ 特征工程

由于对真实部署后，是否可以正确收集到满意度指标（主观性较高），所以构建工作负荷指数特征（（项目数*月均工作时长）/部门均值），进行随机森林和 xgboost 建模对比，预测效果稍有下滑，但在可接受范围内。

YOUR LOGO

结果解读与建议

PowerPoint design

PART 05



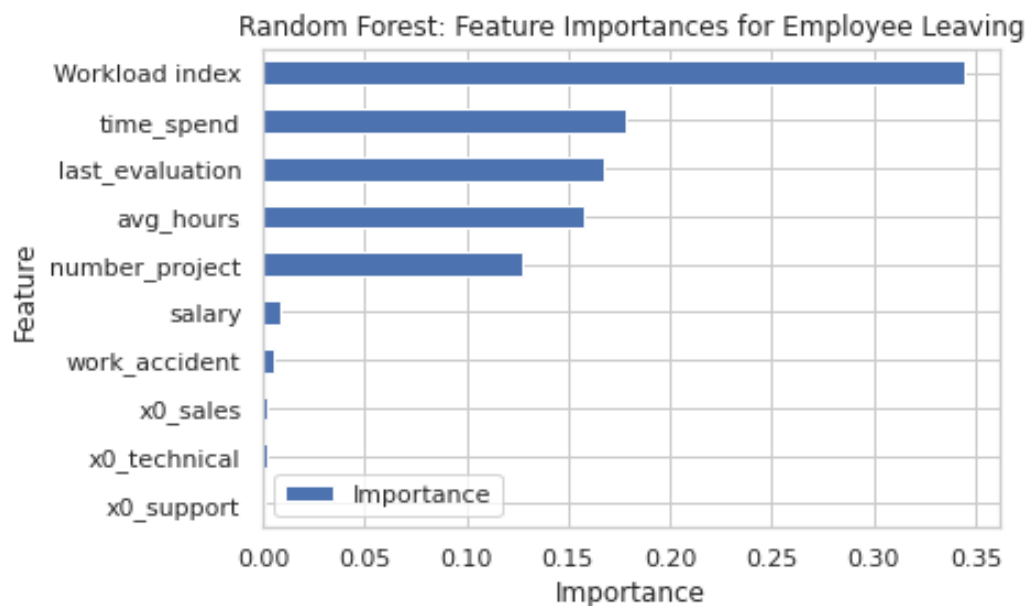
模型结果解读



关键因素总结

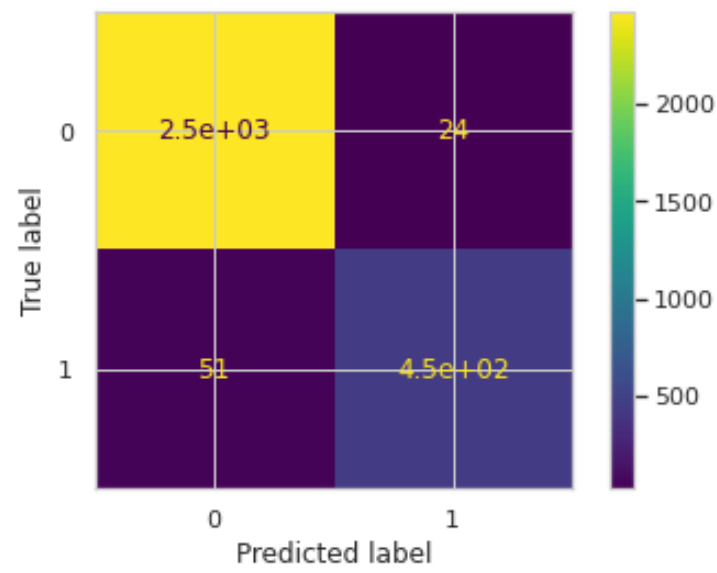
从特征重要性来看，对员工离职与否影响最大的几个因素是工作压力、工作年限、绩效评估结果、月均工作时长和项目数量。薪资水平、部门以及其他因素的影响非常小。

关键因素的识别为后续改进措施提供方向。



模型结果解读

Model	F1	Recall	Precision	Accuracy
Random Forest Validated2	0.925844	0.906289	0.946449	0.975871



模型性能分析

我们建模的目的是在预测中尽可能多的捕捉那些确实将要离职的员工，以便及时采取措施对他们进行补救和挽留。所以我们更关心召回率而不是精确率。在经过特征工程和网格搜索调优后，我们得到的最终模型在测试集上的召回率为 89.82%，可以很好地执行预测员工流失的任务。

业务建议

01

减轻员工压力

如果员工承担了太多的工作任务就需要引起注意。这对是否离职是很直接的影响。

可以尝试设置工作时长预警机制，主动帮助工作时长过长的员工减轻压力，并且可以完善加班奖金制度。

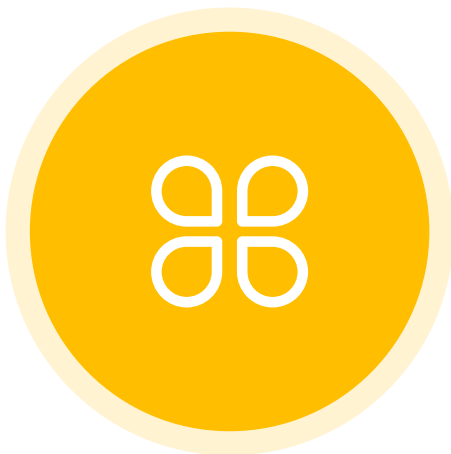


02

关注员工工作年限

重点关注入职3~5年的员工，尤其是绩效评估比较一般的这部分员工，给予他们更多的培训或转岗机会。

对于入职三年以上的员工，尝试设置老员工奖励，以表彰他们对公司发展持续长久的贡献。



03

关注员工绩效评估

对于绩效评估较低的员工给予更多得工作指导帮助他们对工作的信心，对于绩效评估较高的员工要注意他们是否已经不满足于当前的工作内容，尝试给他们提供更多的晋升机会，给予他们更大的舞台去施展能力。

YOUR LOGO

总结与展望

PowerPoint design

PART 06

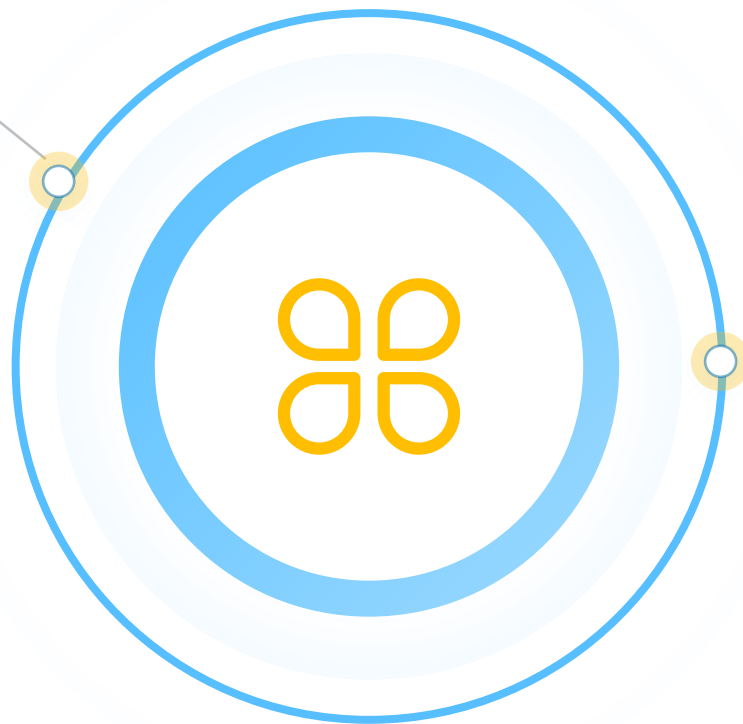


项目总结

项目成果

完成数据探索、清洗、分析、可视化及模型构建与评估。

项目成果为公司人力资源管理提供了有力的数据支持。



项目价值

为公司人力资源管理提供数据支持，帮助公司制定针对性措施降低员工流失率。
项目价值体现在提升人力资源管理效率和降低运营成本。

未来展望

模型改进方向

尝试其他模型，如支持向量机、神经网络；增加数据特征，如员工技能水平、团队氛围等。
模型改进方向为后续研究提供思路。

进一步研究问题

分析不同业务部门员工离职因素的差异；研究市场环境变化对员工离职的影响。
进一步研究问题有助于深入理解员工离职的复杂因素。

YOUR LOGO

2025

谢谢大家

PowerPoint design



主讲人：马龙



时间：2025.4

