

1. 结合你目前的公司业务场景或者从你感兴趣的一家公司业务出发，尝试站在用户角度思考问题，初步建立标签体系（可参考第3节/第4节标签体系搭建方法）？
2. Kaggle 是众所周知的数据竞赛平台，为社区用户举办了一次问卷调查，调查内容涉及人工智能领域的不同维度，收集到 16000 份关于从业者的人员属性、工作背景，学习特征和技能特征等方面的信息，帮助打算进入人工智能领域的有识之士更好的了解行业现状，做到有的放矢，不走弯路。基于 kaggle 比赛中给到数据:multipleChoiceResponses.csv 对用户进行**描述性统计分析**，并**对你感兴趣用户群体**（如男性或者数据分析师）进行进一步分析，查看这类群体的特征与其他特征的差异（提示：TGI），这个过程可以参考抖音客群洞察分析报告。

大家可以重点关注以下维度:

Country	国籍
Age	受访者年龄
EmploymentStatus	就业情况
CurrentJobTitleSelect	当前职称
MLToolNextYearSelect	明年将学习机器学习工具
MLMethodNextYearSelect	明年将学习机器学习方法
LanguageRecommendationSelect	首选推荐编程语言
JobSkillImportanceBigData	大数据在工作中的重要度
JobSkillImportanceDegree	学位在工作中的重要程度
JobSkillImportanceEnterpriseTools	企业工具在工作中的重要度
JobSkillImportancePython	python 在工作中的重要度
JobSkillImportanceR	R 在工作中的重要度
JobSkillImportanceSQL	SQL 在工作中的重要度
JobSkillImportanceKaggleRanking	Kaggle 排名在工作中的重要度
JobSkillImportanceVisualizations	可视化排名在工作中的重要度
DataScienceIdentitySelect	数据科学
FormalEducation	教育背景
Tenure	写代码从事数据分析年限
JobSatisfaction	工作满意度

【补充】数据处理方式：1) 缺失率大于80 %可以考虑直接去掉，或者对于重要字段，说明不同人群在这个维度上填写情况，比如缺失的部分为该用户“不愿填写”/非缺失的部分为该用户“愿意填写”，最后计算愿意填写人员占比；2) 缺失率小于80%，说明该标签非缺失部分数量占比，可以成为“有效用户占比”。3) TGI在人群占比小的时候波动较大，此时TGI值不具有代表性，可以考虑将占比小的组群合并，名为“其他”，或者在后续说明时解释一下。

3. 模拟面试场景，准备 1 分钟介绍用户画像的内容（200个字左右）。