

AI 对就业机会的结构性重塑：教育、技能与公平的新挑战

1. 作业目标

本次作业的目标是帮助同学们在真实数据情境中锻炼 数据思维 和 分析能力。通过对“AI对就业机会的结构性重塑”这一主题的数据集进行探索，同学们将学习如何从多维度数据中发现有意义的模式与趋势，并形成有逻辑的分析报告。

熟悉数据分析的基本流程

- 掌握数据加载与初步检查方法，理解各字段的含义与数据类型。
- 能够处理缺失值、异常值，对数据进行初步清洗。
- 学会使用多个字段进行交叉分析，而不是依赖单一指标。
- 结合可视化（如柱状图、折线图、散点图等）揭示趋势与变量之间的关系。

为后续大作业打下基础

- 通过这次练习掌握数据预处理、可视化和多维度分析的基本方法。
- 积累如何从宏观主题切入，再聚焦到若干子问题进行分析的经验。
- 为后续更大规模、更复杂的数据研究和写作做好准备。

2. 数据集介绍

AI对就业市场的影响：增加与减少的工作岗位（2024-2030年）

 数据集描述：本数据集探讨了人工智能（AI）如何改变全球就业市场。重点识别哪些工作岗位因AI采用而增加或减少，该数据集提供了跨行业和国家的就业趋势、自动化风险、教育要求、性别多样性和其他劳动力相关因素的洞察。

该数据集包含30,000行数据和13个有价值的列，基于持续研究和公共数据洞察生成，以反映真实的劳动力市场模式。可用于数据分析、预测建模、AI政策规划、职位推荐系统和经济预测。

 数据来源：这是一个合成数据集，使用现实建模、公共就业数据模式（美国劳工统计局、经合组织、麦肯锡、世界经济论坛报告）和AI模拟生成，以反映2024-2030年的合理情境（未来的部分均属于预测）。适用于教育、研究和AI项目目的。

 列描述：

列名	描述

Job Title	工作/职位名称（如数据分析师、收银员等）
Industry	工作所属的行业部门（如IT、医疗保健、制造业）
Job Status	指示该工作因AI采用而增加或减少
AI Impact Level	AI对该工作的预估影响程度：低、中等或高
Median Salary (USD)	该工作的年度中位薪资（美元）
Required Education	该工作通常要求的最低教育水平
Experience Required (Years)	所需的平均工作经验年数
Job Openings (2024)	2024年当前职位空缺数
Projected Openings (2030)	预计2030年的职位空缺数
Remote Work Ratio (%)	可远程工作的职位估计百分比
Automation Risk (%)	该工作被自动化或AI替代的概率
Location	工作数据所基于的国家（如美国、印度、英国等）
Gender Diversity (%)	该工作中非男性性别的大致代表性百分比

3. 作业任务

为了帮助大家循序渐进地开展分析，我们已经设计了一个统一的宏观主题，并准备了若干子主题作为参考。你们的任务是围绕这些主题展开数据分析与讨论。

主题：AI 对就业机会的结构性重塑：教育、技能与公平的新挑战

本次作业分为**必做任务**（共 85 分）和**选做任务**（额外 20 分），总分最高为 105 分。

注意：本次作业以小组形式完成，每组人数不超过3人。

3.1 必做任务：基础与探索性分析（共85分）

A. 数据预处理与概览（20分）

该部分主要考察数据加载、数据清洗、描述性统计。需要在给定数据集上完成：加载数据，检查数据类型、缺失值和明显异常值。

- 清晰说明你对缺失值或异常值（如果发现）的处理方法（如删除、填充等）。
- 对关键数值型字段（如 Median Salary, Automation Risk）和分类型字段（如 Industry, Required Education）进行描述性统计和频次统计。

B. 指定子主题分析（40分，每项 20 分，3选2）

1. 技能与教育门槛

- 列组合推荐： Required Education + Experience Required (Years) +

Automation Risk (%) + Median Salary (USD)

- 分析方向：教育与经验在岗位稳定性和薪资水平中的作用，高学历/高技能是否更能抵御AI冲击。

2. 行业与地区差异

- 列组合推荐：Industry + Location + Job Status + Openings Abs Change

- 分析方向：比较不同行业、不同地区的岗位净增/净减，识别就业“热点”和“冷点”。

3. 就业创造 vs 替代

- 列组合推荐：Job Status + AI Impact Level + Openings Abs Change

- 分析方向：AI创造的新岗位是否能抵消被替代的岗位？新增岗位集中在哪些领域？

C. 自由子主题分析（自选1个，25分）

该部分作为后续大作业的预热环节。同学们可以尝试在现有数据集的基础上，自主提出并探索一个感兴趣的分析方向。这样做的目的是帮助大家逐步积累提出研究问题、设计分析思路的经验。在之后的大作业中，你们将需要结合自己选择的研究领域，独立提出主题并完成更系统的分析。所以现在请你们自己思考一个子主题并且给出分析和结论。

- 你们可以结合数据集中的任意字段，提出新的研究问题，例如：

- AI对不同薪资水平岗位的影响差异？高薪工作是否更安全？低薪岗位是否更容易被AI替代？

- 远程工作与岗位安全性之间的关系。远程比例高的岗位是否更具增长潜力？自己选择感兴趣的一个子主题，进行类似B中的分析。

3.2 选做任务：进阶挑战（额外20分）

任选以下一项挑战任务，展示更深入的数据处理或建模能力。需在报告中单独开辟章节说明你的方法和结果。

1. 预测建模。请你构建一个模型（如线性回归、决策树或随机森林）来预测 Median Salary (USD) 或 Automation Risk (%)。需说明你选择了哪些特征（Features），为什么选择它们，以及模型的简要评估结果（如 R^2 或 RMSE）。

2. 聚类分析。尝试使用 K-Means 或其他聚类算法，基于 Automation Risk (%), Median Salary (USD), Experience Required (Years) 和 Remote Work Ratio (%) 等特征，对 Job Title 进行聚类。解释你分成了几类（K值如何选择），并尝试解读每一类工作（“工作集群”）的共同特征（例如：“高薪-高风险-高经验”集群，“低薪-高风险-低经验”集群等）。

4. 作业要求

提交形式

- 每组提交一份分析报告（Word 或 PDF 均可），不限字数，简洁即可，这是个小作业~
 - 报告需包含：标题、研究的子主题、分析过程（涉及到的统计图表）、结果与结论。
- 分析数据的代码（.py or .ipynb均可）

内容要求

- 报告主题需围绕“宏观主题：AI 对就业机会的结构性重塑”。
- 必做部分（3.1）需要使用至少3种图表（柱状图、折线图、散点图等）进行辅助说明，图文结合。
- 体现基本的数据思维与可视化呈现，希望能在统计学结果呈现的基础上深入分析，尽量不要浮于表面。

提交时间

- 2025年11月5日23：59

提交方式

- OBE提交

评分方式

- 人工评分。各组自愿报名参与“优秀报告”评选。由人工评分选出3–5组（总人数不超过班级人数的30%）优秀报告，获评优秀报告的组将依次汇报展示，由班级同学投票，确定优秀报告的最终次序。