

# 数据说明

## 原始数据说明

### 人格数据集 (data-final.csv)

这些数据是通过互动式在线性格测试收集的：<https://openpsychometrics.org/tests/IPIP-BFFM/>

该人格测试是用IPIP的 "大五因子标记" 构建的：<https://ipip.ori.org/newBigFive5broadKey.htm>

在测试开始时，参与者被告知他们的回答将被记录并用于研究，并在测试结束时被要求确认他们的同意。

以下项目被放在一页上，每个项目都用单选按钮进行五分制评分。

页面上的顺序是：EXT1、AGR1、CSN1、EST1、OPN1、EXT2等等。

评分标准是：1=不同意，3=中立，5=同意。

- EXT - Extraversion - 外向性，指个体对外部世界的积极投入程度。
- EST - Neuroticism - 神经质，指个体体验消极情绪的倾向。
- AGR - Agreeableness - 宜人性，指个体在合作与社会和谐性方面的差异。
- CSN - Conscientiousness - 尽责性，指个体在目标导向行为上的组织、坚持和动机。
- OPN - Openness to experience - 开放性，指个体对经验持开放、探求的态度。

由于数据量过大，我们只选取了数量最多的前五个国家数据（美国、加拿大、澳大利亚、印度、英国），并对每个特质的十个问题进行聚合得到output\_\*\*.csv数据集，每项大五人格指标记录为十个问题倾向(0-5)的总值。

### 社会数据集

数据集来源：<https://www.oecdbetterlifeindex.org/#/11111111111>

良好生活指数，由经济合作与发展组织(OECD)提供，旨在综合评估不同国家在生活质量方面的表现。

social.xls

这个数据集从不同方面衡量幸福度水平，具体数据包括如下：

- Country Name: 国家名称
- Ladder score: 梯子分数，通常是指主观幸福感的评分，衡量一个国家的整体幸福水平。
- upperwhisker 和 lowerwhisker: 通常指的是置信区间的上限和下限，表明梯子分数的统计误差范围。
- Explained by: Log GDP per capita: 通过人均GDP的对数解释，表示经济因素对幸福感评分的贡献。
- Explained by: Social support: 通过社会支持解释，表示社会关系和支持网络对幸福感评分的贡献。

- Explained by: Healthy life expectancy: 通过健康预期寿命解释，表示健康和预期寿命对幸福感评分的贡献。
- Explained by: Freedom to make life choices: 通过自由选择生活的自由度解释，表示自由度对幸福感评分的贡献。
- Explained by: Generosity: 通过慷慨解释，表示人们的慷慨和互助对幸福感评分的贡献。
- Explained by: Perceptions of corruption: 通过对腐败感知的解释，表示对腐败的感知对幸福感评分的贡献。
- Dystopia + residual: 反乌托邦+残差，通常是指最糟糕情况下的假设分数加上模型中无法解释的部分，用于校准模型的基础评分。

## global.xlsx

这个数据集包含不同国家了11个维度的指标：

- 住房
  - 无基本设施的住宅 (Percentage 百分比)
  - 住房支出 (Percentage 百分比)
  - 每人房间数 (Ratio 比例)
- 收入
  - 家庭净调整后可支配收入 (US Dollar 美元)
  - 家庭净财富 (US Dollar 美元)
- 就业
  - 劳动力市场不安全性 (Percentage 百分比)
  - 就业率 (Percentage 百分比)
  - 长期失业率 (Percentage 百分比)
  - 个人收入 (US Dollar 美元)
- 社区
  - 社会支持质量 (Percentage 百分比)
- 教育
  - 教育程度 (Percentage 百分比)
  - 学生技能 (Average score 平均分)
  - 受教育年限 (Years 年)
- 环境
  - 空气污染 (Micrograms per cubic metre 每立方米微克)
  - 水质 (Percentage 百分比)
- 公民参与
  - 制定法规的利益相关者参与度 (Average score 平均分)
  - 选民投票率 (Percentage 百分比)
- 健康
  - 预期寿命 (Years 年)
  - 自我报告的健康状况 (Percentage 百分比)

- 生活满意度
  - 生活满意度 (Average score 平均分)
- 安全
  - 独自在夜间行走的安全感 (Percentage 百分比)
  - 凶杀率 (Ratio 比例)
- 工作与生活平衡
  - 超长时间工作的雇员比例 (Percentage 百分比)
  - 用于休闲和个人护理的时间 (Hours 小时)

## 产出数据

---

我们提取出来自五个国家的数据集

output\_\*文件为中间数据文件

进一步与社会数据集做数据连接，得到产出数据在person\_global.csv中，同时包含人格数据和社会数据