# 1. 数据集信息

该数据集根据墨西哥、秘鲁和哥伦比亚等国的个人的饮食习惯和身体状况估计其肥胖水平的数据。数据包含 17 个属性和 2111 条记录，记录用类变量 NObesity（肥胖级别）标记，允许使用体重不足、正常体重、超重级别I、超重级别II、肥胖类型 I、肥胖II型和肥胖III型的值对数据进行分类。77% 的数据是使用 Weka 工具和 SMOTE 过滤器综合生成的，23% 的数据是通过网络平台直接从用户那里收集的。

# 2. 属性信息

|  |  |  |  |
| --- | --- | --- | --- |
| **属性** | **说明** | **属性** | **说明** |
| Gender | 性别 | Age | 年龄 |
| Height | 身高(单位：米) | Weight | 体重(单位：公斤) |
| family\_history\_  with\_overweight | 家族肥胖史 | FAVC | 经常食用高热量食物 |
| FCVC | 食用蔬菜的频率 | NCP | 主餐数量 |
| CAEC | 在两餐之间进食食物 | SMOKE | 是否吸烟 |
| CH2O | 每日用水量 | SCC | 卡路里消耗监控 |
| FAF | 身体活动频率 | TUE | 使用科技设备的时间 |
| CALC | 酒精消耗 | MTRANS | 交通工具使用 |
| NObeyesdad | 肥胖级别 |  |  |

# 3. 数据分析任务要求

* 参照课堂所讲数据分析实例，利用Jupyter Notebook完成该数据集的数据分析功能，最终提交学号-姓名.ipynb文件到FTP服务器；
* 熟悉数据集的专业领域背景与属性特征，利用Pandas读取该数据集，并进行相应的数据预处理工作，包括数据清洗、数据规范化等内容；
* 综合运用所学Python知识，完成初步数据分析功能，要求至少完成3个具有实际意义的分析结论（用Matplotlib进行图像化的展示，要求图表类型选择合理），并能够进行科学的解释。