**120年奥运数据集**

**背景描述**

该数据集(kaggle平台用户分享)整理了从1896年雅典奥运会至2016年里约热内卢奥运会120年的奥林匹克运动会的历史数据。需要注意的是，在1896年-1992年期间，冬季奥运会与夏季奥运会都是在同一年举行的。在这之后，冬季与夏季的奥运会才被错开举办，冬季奥运会从1994年开始4年举办一次，夏季奥运会从1996开始4年举办一次。应避免认为夏季与冬季奥运会一直错开举办。

**数据说明**

* 该数据集包含两个文件：
  + athlete\_events.csv：参赛运动员基本生物数据和奖牌结果
  + noc\_regions.csv：国家奥委会3个字母的代码与对应国家信息
* 数据集的整体特征

| **数据集名称** | **数据类型** | **特征数** | **实例数** | **值缺失** | **相关任务** |
| --- | --- | --- | --- | --- | --- |
| 120年奥运历史数据集：运动员和成绩 | 字符、数值数据 | 15 | 271,116 | 有 | 可视化 |

* **文件athlete\_events.csv中包含15个字段，具体信息如下：**
  + 每一行代表的是一个参加个人比赛运动员

| **No** | **属性** | **数据类型** | **字段描述** |
| --- | --- | --- | --- |
| 1 | ID | Integer | 给每个运动员的唯一ID |
| 2 | Name | String | 运动员名字 |
| 3 | Sex | Integer | 性别 |
| 4 | Age | Float | 年龄 |
| 5 | Height | Float | 身高 |
| 6 | Weight | Float | 体重 |
| 7 | Team | String | 所代表的国家队 |
| 8 | NOC | String | 国家奥委会3个字母的代码 |
| 9 | Games | String | 年份与季节 |
| 10 | Year | Integer | 比赛年份 |
| 11 | Season | String | 比赛季节 |
| 12 | City | String | 举办城市 |
| 13 | Sport | String | 运动类别 |
| 14 | Event | String | 比赛项目 |
| 15 | Medal | Sring | 奖牌 |

**文件noc\_regions.csv中包含3个字段，具体信息如下：**

| **No** | **属性** | **数据类型** | **字段描述** |
| --- | --- | --- | --- |
| 1 | NOC | String | 国家奥委会3个字母的代码 |
| 2 | Region | String | 国家 |
| 3 | Notes | String | 地区 |

**可探索研究的方向：**

1、分析 120 年的 Kaggle 数据，中国何时站上奥运之巅？<https://blog.csdn.net/csdnnews/article/details/82948265>

2、奥林匹克历史数据分析

<https://zhuanlan.zhihu.com/p/42163938>

3、从1896年说起，用数据看 120 年奥运变迁：

https://zhuanlan.zhihu.com/p/55521946?utm\_source=QQ\_article\_bottom

4、可视化参考

http://fms.news.cn/swf/2016\_sjxw/olympic\_0802/index.html