针对指定网站抓取特定数据，并根据格式要求输出，初步实现可导出到excel格式即可。后续进阶需求为对已导出数据进行制定分析，并可以调节相关参数，来测试模型和算法的有效性。

本文档主要包括几方面内容：

1. 数据源、特定的数据、抓取范围和特定网址
2. 输出格式要求
3. 计算公式和筛选条件

### 数据来源及抓取范围说明

所需采集数据说明



数据来源网址及位置示意图

目标网址1

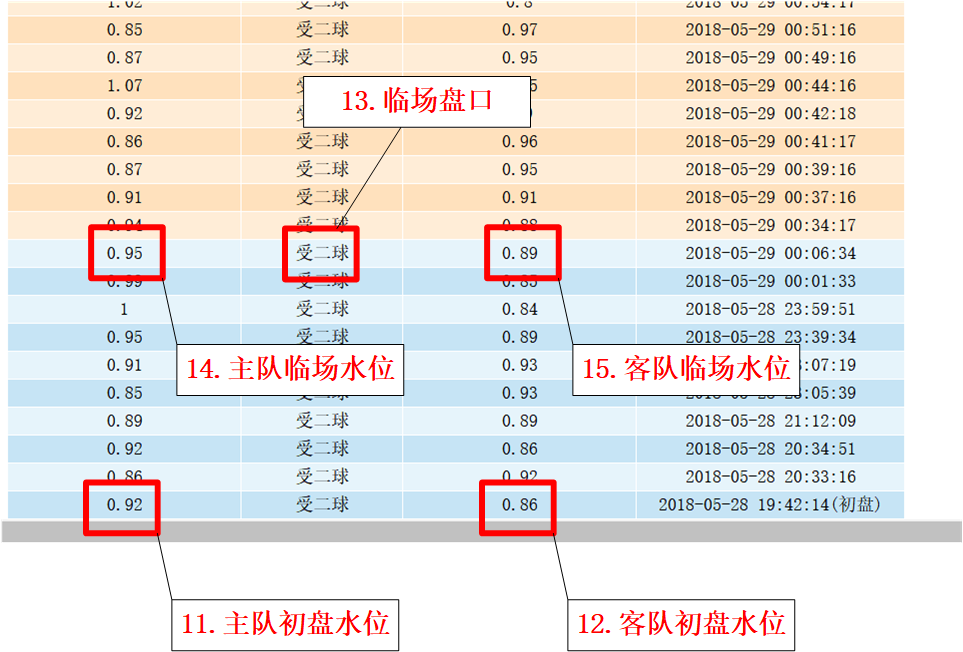
<http://data.7m.com.cn/result_data/default_gb.shtml?date=2018-05-28>

图片中红框代表所需数据显示位置



目标网址2

<http://s2.7m.hk/chart.html?ecd=1&id=3241816>



目标网址2为目标网址1的三级详情页，具体进入路径如下

第一步，点击红框链接，进入下面的网址

http://data.7m.com.cn/goaldata/ft/3720743.shtml



第二步点击红框位置进入如下网址

http://s2.7m.hk/chart.htm?ecd=0&id=3241816



### 输出格式需求

见excel附件 第二sheet

### 计算公式

初期按照异常盘口和关注球队两个维度对目标比赛进行初步筛选，后续需求待细化

1. 问题盘口计算

问题盘口定义为庄家所开出的盘口与我们所谓的正常盘口的偏离度来衡量，偏离度越大我们则认为盘口值得关注的程度就越大。

初期计算公式如下： (（主队赛事排名-客队赛事排名）/5)+（-0.5）-(初盘盘口)

结果计算出来，结果的述职越大，则说明盘口和预期的偏离度越大，结果为正值，则偏向主队，结果为负值责偏向为客队。

1. 关注球队

根据历史经验，一些球队可以特别关注，关注名单后续补充