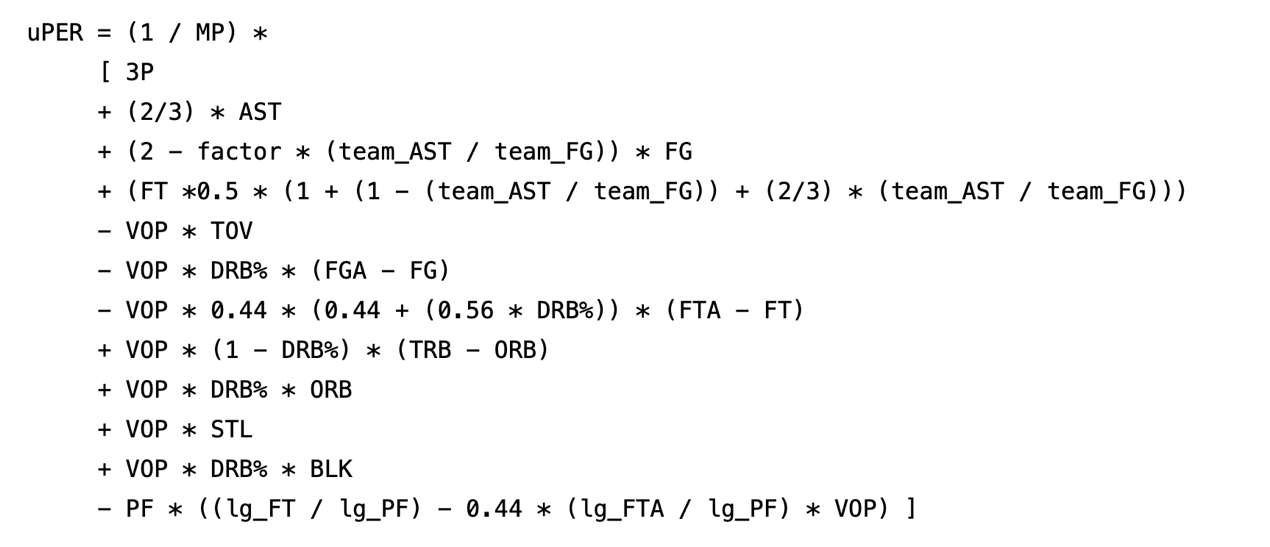
一些体育网站如（虎扑体育），NBA官网等利用简单指标之间的组合，计算出更有意义的指标，使用这些调整后的指标，对分析不同的队伍和不同风格的球员有更显著的意义。比如，一个在更偏防守的球队打球的球员的进攻数据，可能不会比一个在更偏进攻的球队打球的球员的进攻数据更好。然而这并不意味，第一个球员的进攻能力不如第二个球员强。为了削弱球队风格对球员数据的影响，分析者通过不同的方式将数据标准化。其中一种方式，就是根据控球时间标准化球员的各项数据。即根据每个球员每次控球时所产生的数据，可以更加精准的评判一个球员真正的实力。

PER：衡量球员效率的指标Per Efficiency Rating. 指标根据NBA官网公开的各球员、球队、联盟的基础数据计算得出。总结球员对球队的正向贡献和负向贡献，即在衡量球员得分的指标前赋予一个正的系数，在衡量球员失分的指标前赋予一个负的系数，再将所有加权后的指标相加，再除以根据球队和联盟数据构建的指标对球员的数据进行规范化。PER最终得到球员每分钟的贡献率，该指标不应该受到球员上场时间，球队进攻类型，整个联盟赛制变化的影响。该指标客观有效的衡量球员的比赛效率，可以成为衡量球员对球队贡献指标之一。但是由于在大多数开源的数据中更多记录了一个球员进攻表现（更容易记录）如进攻篮板球，三分球得分，罚球得分等；记录球员防守表现的数据更少（不容易被记录）。所以该指标对一个防守很强的球员更有倾向。在实际应用中，两个同样优秀的球员，其一打球风格更激进，其二防守表现突出，可能PER在得出第一位球员更优秀的结论。在本指标中用到的基础数据有

uPER = (1 / MP) \*

[ 3P

+ (2/3) \* AST

+ (2 - factor \* (team\_AST / team\_FG)) \* FG

+ (FT \*0.5 \* (1 + (1 - (team\_AST / team\_FG)) + (2/3) \* (team\_AST / team\_FG)))

- VOP \* TOV

- VOP \* DRB% \* (FGA - FG)

- VOP \* 0.44 \* (0.44 + (0.56 \* DRB%)) \* (FTA - FT)

+ VOP \* (1 - DRB%) \* (TRB - ORB)

+ VOP \* DRB% \* ORB

+ VOP \* STL

+ VOP \* DRB% \* BLK

- PF \* ((lg\_FT / lg\_PF) - 0.44 \* (lg\_FTA / lg\_PF) \* VOP) ]

因为我们的模型是找出球队表现和球员个统计量的 关系，我不仅会用到个体球员数据，也会用到球队统计数据。对于球队数据，我将用到球队的胜率为因变量来衡量球队在整一个赛季的表现，对于球员，我将用根据球员的基本数据构造全面衡量球员效率的霍林格PER，本模型中PER是十分重要的一个指标。

* Team\_PER:

1. in\_PER:一个十分常见并且广泛应用的指标是PER(PLAYER EFFICIENCY RATING)，球员效率，根据球员的基本数据生成的衡量球员能力的指标。PER是体育学家约翰·霍林格最初构建的。这个指标衡量的是球员每分钟的表现和根据球队节奏调整后的表现。这样在节奏慢的球队打球的球员不会因为其球队的节奏而表现出比其他同水平球员更差的数据。此外，PER是衡量每个球员每分钟表现的数据，这样就不会因为每个球员上场时间不同而有不同的数据表现。因为PER指标是各大数据网站公开的，在接下来的模型建立中，将直接使用这个数据。

其中：

PER指标的优缺点：

优点：PER指标充分运用了球员基本数据的特征，该指标比网上所有公开的原始数据更全面和精准的衡量每场比赛球员的得分情况，从而反映该球员的效率。

缺点：一个最显著的缺点是，该指标在球员防守指标上缺乏体现。上述公式主要反应球员进攻效率，尽管公式中包含球员的盖帽和抢断数据，但这些数据还是相对片面。

1. team\_possesions:每队的的进攻次数，从一个球队拿到控球权算做一次进攻开始，当球队因失误将球丢失或者投篮命中或者被抢断和盖帽，算做一次进攻的停止。在一场比赛中，两队的进攻次数大致相同，所以利用进攻次数可以更好的衡量球队的优秀程度。为了赢得比赛，每支球队应该在进攻中取得最多的得分。
2. team\_pace:每48分钟的球队的进攻效率。
3. pace\_adjustment：根据比赛双方进攻强度进行调整；
4. aPER: 调整后的PER：
5. team\_PER:衡量球队打球效率的指标，根据球员aPER构成球队team\_PER.
6. 按照每支球队整个赛季中各个球员在场时间将球员排序
7. 将每个球员的PER和上场时间相乘
8. 将每队上场时间最长的12个球员的PER与上场时间相乘后的指标相加

该计算方法的意义：

1. 由于PER是一个以分钟为单位的指标，通过与球员的上场时长相乘可以估计出球员在整个赛季中对球队贡献。
2. 一些球员通过选秀等方式会被交易到其他球队，所以要计算球员上场时间最长的12个球员，更充分可以反映球员对球队的贡献。因为这12个球员是球队中比较稳定的中流砥柱.
3. 该计算方法也可以自动给PER更高的球员赋予更大的权重，因为PER更高的球员上场时间更长，而PER较低的球员的上场时间会相对较短，所以时长作为权重，可以很好的平衡球员之间的效率。就算是某个PER很高的球员在赛季中受伤，他的时长也会限制他对球队的贡献。如史蒂芬库里在19-20赛季中因伤暂停比赛半年，尽管他是一个价值很高的球员，他的上场时长受到影响，他对球队的贡献

* Team\_Rank：球队排名，依据球队赢得比赛的次数，与球队整个赛季中各个方面的表现得到的整体排名。排名综合考虑球队的进攻，防守，总得分和整体数据得出。
* Adjust\_MoV：Margin of Victory 比分差距是用来衡量比赛中获胜队伍和失败队伍比赛分数差距的统计量，参考MOV可以快速看出比赛是否激烈，获胜队伍胜利的是否显著，大的比分差距代表比赛中获胜方在比赛中强势压制对手，而小的比分差距说明两个球队水平接近，比赛相当激烈。在本数据集中，MOV指标是取一个队在整个赛季比赛中球队MOV指标的平均值。如果一个球队的Margin of Victory指标是正数，说明该球队在整个赛季中表现更好。然而在比赛中由于每个球队的进攻风格不同，需要按照进攻强度来调整对应的MOV值。尤其是不同的队伍会根据遇到的对手球队调整进攻步伐和阵容等，所以我们利用更好反映球队强弱的Adjust\_Mov数据来建模。历史上最大的比分差是68分，是1992年克利夫兰骑士队和迈阿密热火队的比赛，当时骑士队是最强的球队之一。
* Adjust\_Ortg：Offensive Rating是指球队的进攻效率，由于每场比赛下来，球队的进攻节奏会受到很多因素的影响，包括教练风格，球队首发阵容，球员打球习惯，团队合作，对手球队的节奏等。所以很难用一场比赛的总得分作为其进攻效率的判断。为了方便起见，规定将每一百次进攻得分作为球队的进攻效率，并且根据球队的进攻节奏进行调整。一支进攻效率很高的球队，说明球队的进攻水平越高。根据得分制度，一支球队的进攻效率在每100次进攻得100分左右。
* Adjust\_DRtg: Defensive Rating是衡量一个球队防守效率的指标。由于在比赛进程中，不仅有进攻，还要考虑防守。根据对方球队的进攻效率不同，每个球队的防守程度也不可同日而语，在一支球队面临100次进攻时，对方球队的得分就是本球队的失分。我们采用根据进攻节奏步伐调整后的防守效率，更可以客观的反映球队的防守效率。每支球队在比赛时的平均失分在100分左右。失分越高，说明球队的防守水平越低。
* Adjust\_NRtg：如果只单纯根据Offensive Rating和Defensive Rating来衡量一个球队的优秀程度多少有一些偏差，这时需要引入球队净得分来更全面的衡量。Net Rating是球队每100次进攻中得分与失分的差。用来衡量球队风格，和更全面的评判球队的表现。一个球队如果在整个赛季中净得分大于0，说明进攻性更强。当然球队的净得分和球员的表现有很大的关系。球队的净得分均值是0.
* Win\_ratio:在整个赛季的比赛中，赢得的比赛与参加的比赛之比作为胜率。胜率对衡量球队在下一个赛季的表现有着很重要的作用。可以帮助教练调整训练节奏和步伐，帮助球员在选秀中找到一个更合适自己的队伍，可以帮助粉丝们在赌球中有一个更好的判断。胜率的平均值在0.5.