

期末大作业报告

分析 NBA 高薪球员对球队的影响

数据 10225501454 何俊彦

一、选题

近期，金州勇士战绩不佳，球队的第二球星，手握 4300 万美元年薪的汤普森表现较上赛季下滑明显，并被曝拒绝了球队 2400 万美元年薪续约请求。这条新闻引发出了一个问题，高薪球员在本赛季的表现到底如何，又能否帮助球队取得很好的成绩？本次作业将分析 NBA 高薪球员对球队的影响以及其中一些异常数据。

二、数据选择与收集

本次分析使用的数据来自篮球数据网站 <https://www.basketball-reference.com/> 与 <https://hoopshype.com/salaries/players/>。后者提供球员年薪的数据，本次实验针对了本赛季年薪前 100 位的球员。前者提供了各类非常全面的球员数据，除了基础数据（得分、助攻、篮板等）之外，还包含了许多进阶数据。数据截止至北京时间 1 月 14 日 0 点，且均只计算球员本赛季在目前所在球队的表现，若球员在赛季中期被交易，则不计算在交易之前球队的表现。

本次实验选取的基础数据为球员及其所属球队、年薪、上场次数、缺席次数、缺席时球队胜场数以及球队的总战绩。由此可以算出球员上场时与缺席时球队的胜率，以提现球员对球队的重要性。

此外，为了衡量球员的进攻效率，本实验采取了球员真实命中率。由于篮球比赛每次出手的分值并不是完全相同的，真实命中率针对三分和罚球做了一定的修正。它的计算公式是： $TS\%(\text{真实命中率}) = 100\% * PTS(\text{得分}) / [2 * (FGA(\text{出手次数}) + 0.44 * FTA(\text{罚球次数}))]$ ，其中 0.44 为罚球的加权因子。它把每一次出手都调整成一次两分出手，体现球员每次出手的得分回报与进攻把握率。

本实验还使用了每 100 次进攻回合的正负值净值 (Plus/Minus Net Per 100 Possessions)。这是一种统计篮球球员在场上时球队得分和对手得分之间的差异，以每 100 次进攻回合为基准的指标。这个统计量被用来衡量球员在场上对球队整体表现的影响。其中，Plus/Minus (正负值)是指球队在球员上场时的得分减去对手得分。如果一个球员在场上时，他所在的球队比对手得分更多，那么他的 Plus/Minus 就是正值；如果球队在场上时被对手得分更多，那么 Plus/Minus 就是负值。Net Per 100 Possessions (每 100 次进攻回合的净值)将得分差异标准化，以每 100 次进攻回合为基准。这是为了消除不同球队的比赛速度差异。简而言之，如果一个球员的 Plus/Minus Net Per 100 Possessions 值是正的，说明球队在他上场时比对手更有优势；如果是负的，说明球队在他上场时对手更占优势。

本次实验使用了爬虫收集数据。表格形式如下：

Num	Name	Team	Salary	TS%	Games_played	Games_abs	Win2	PM_player
1	Stephen Curry	GSW	51915615	0.621	36	2	0	-7.9
Team	Win	Loss						
BOS	29	9						

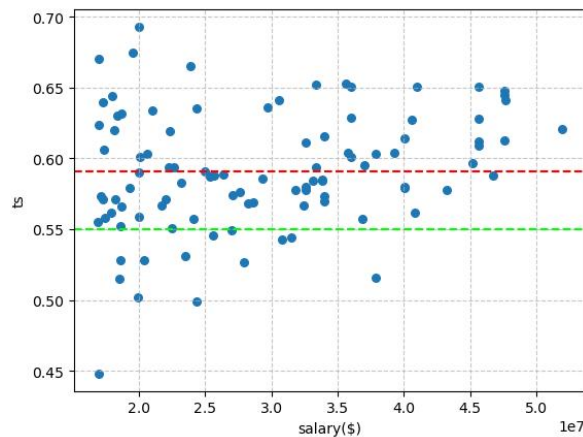
三、数据筛选

整理数据过程中，发现表中有几个异常数据：Lonzo Ball 本赛季未登场，Evan Fournier 本赛季仅登场 2 次，被其主帅弃用，显然这两位球员的数据不具有参考价值，从表格中删除。

四、数据分析

1) 真实命中率

在读取表格之后，选取球员真实命中率为 y 轴指标，年薪为 x 轴指标，画出散点图，再以所有高薪球员（即本实验包括的 98 位球员）平均真实命中率与全联盟球员平均真实命中率作为参照。得到散点图：



其中红色虚线为高薪球员平均真实命中率（59.1204081632653%），绿色虚线为全联盟球员平均真实命中率（55%）。根据图表，可以按年薪把这些球员分成两类并计算，得到结果：高薪球员真实命中率不足全联盟平均水平的有 13 个，占全部高薪球员的 13.26530612244898%，其中年薪超过 3500 万的只有 1 名球员；年薪超过 3500 万的球员真实命中率超过高薪球员平均水平的有 21 个，占全部年薪超过 3500 万的球员的 75.0%；年薪低于 3500 万的高薪球员真实命中率超过高薪球员平均水平的有 24 个，占全部年薪低于 3500 万的高薪球员的 34.285714285714285%。

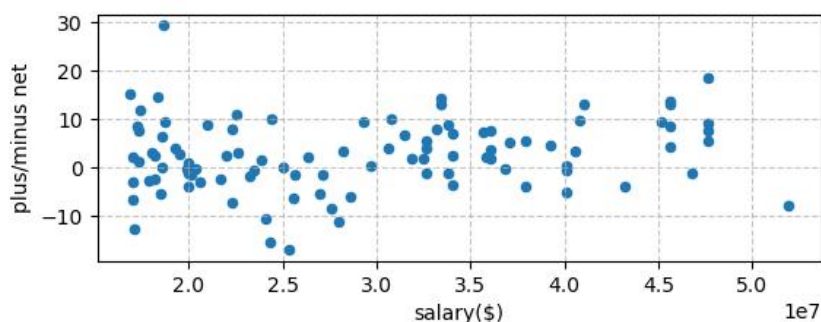
由此可见，年薪高于 3500 万的球员大多能以顶级的真实命中率帮助球队，而年薪低于 3500 万的高薪球员表现参差不齐，而在所有高薪球员中真实命中率最高的两位年薪均小于 3500 万。对这两位球员分析可得：

Jarrett Allen	CLE	20000000	0.693
Jakob Poeltl	TOR	19500000	0.675

这两位球员都是知名的蓝领型内线，有着稳固的护框能力，而在进攻中基本以吃饼，篮下的攻框或二次进攻得分，不占用太多出手次数，保证近框区域的高命中率，因此真实命中率极高，体现出两位球员把握机会能力强，是非常优秀的蓝领内线。

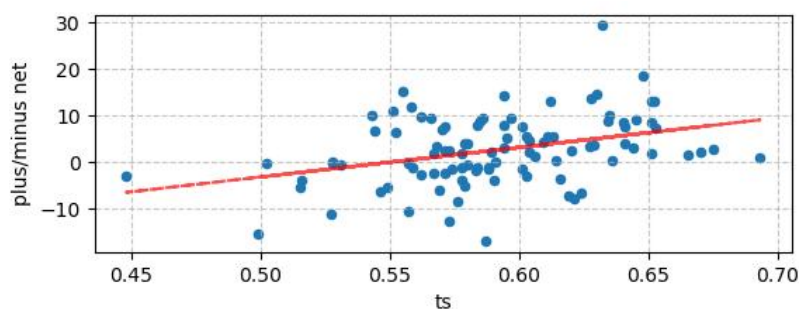
2) 每 100 次进攻回合的正负值净值（PM 值）

选取球员每 100 次进攻回合的正负值净值为 y 轴指标，年薪为 x 轴指标，画出散点图：



可以发现分布十分的分散。以 3000 万年薪、PM 值等于 0 为界线进行计算，可得年薪高于 3000 万的球员 PM 值 ≥ 0 的有 77.27%，而年薪低于 3000 万的高薪球员 PM 值 ≥ 0 的只有 48.15%。可见年薪低于 3000 万的高薪球员大多能为球队做出稳定的贡献，制造优势。

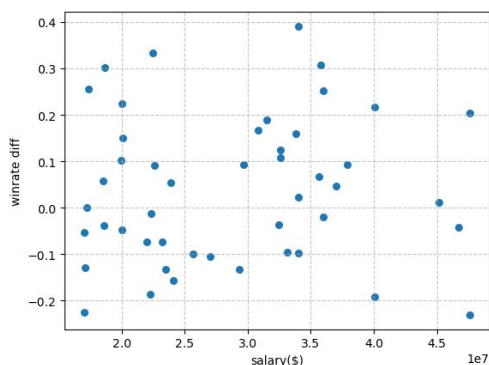
本实验还尝试分析真实命中率对 PM 值的影响。选取球员每 100 次进攻回合的正负值净值为 y 轴指标，真实命中率为 x 轴指标，画出散点图，同时引入线性拟合的模块帮助分析两者关系。得到图表：



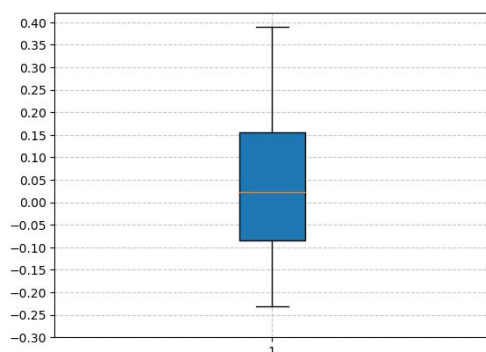
可见散点分布并不成线性，线性拟合得出的直线只能提供一定的趋势参考，真实命中率高的球员一般 PM 值相对较高，其中真实命中率超过 60% 的球员大多数 PM 值都为正。

3) 球队胜率提升

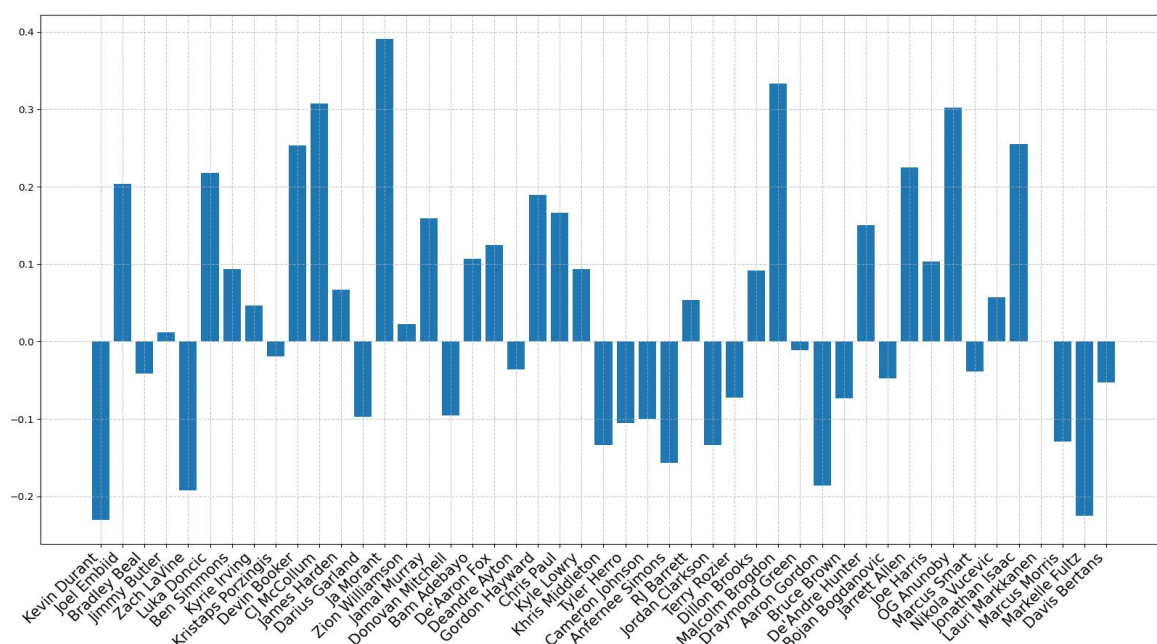
读取球队战绩的表格，转换成字典，可以结合球员上场次数、缺席次数、缺席时球队胜场数简单地将球员上场时胜率，缺席时胜率，以及两者差值计算出来。将新得到的数据写入一个新 csv 文件中，并选取球员上场与缺席时球队胜率差值为 y 轴指标，年薪为 x 轴指标，画出散点图，来分析球员对球队胜率的提升。此处仅选择了缺席次数 ≥ 5 次的球员来提升数据的参考价值。



可见数据非常分散，再画出箱型图：



由图可得球员对球队胜率提升的中位数为 2.50%，数值大多分布在-10%到+15%之间。由于本赛季已进行的比赛场数还未过半，且高薪球员大多为球队核心，缺席场数普遍较少，因此此项数据波动程度很大，缺席场数更多的球员此项数据更具有参考性。因此，本实验选择列出柱状图以便分析一些异常数据，球员年薪从左到右依次降低。



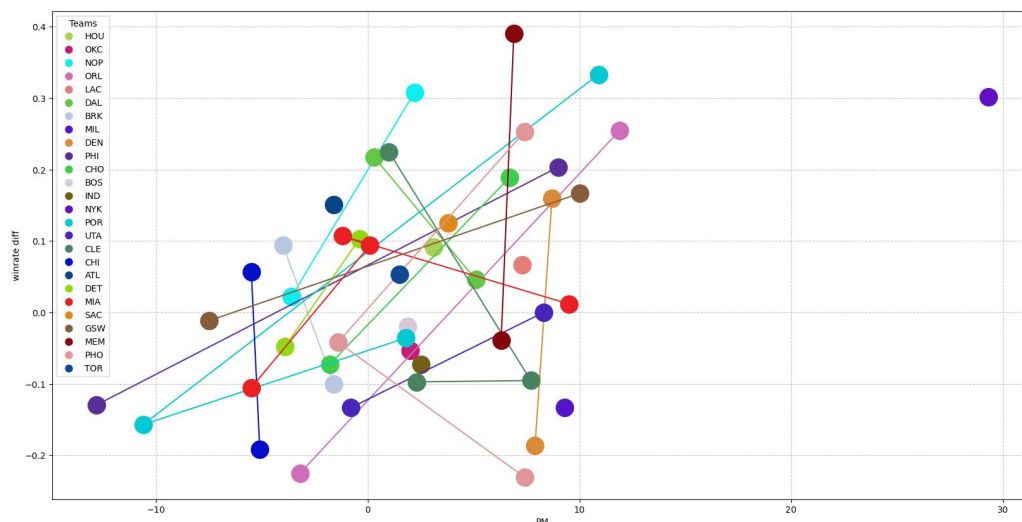
尤其值得注意的是 Kevin Durant 以及 Ja Morant 两位球员。pre_winrate 表示出场时球队胜率，abs_winrate 表示缺席时球队胜率，diff 表示两者差值。

Name	Team	Salary	pre_winrate	abs_winrate	diff
Kevin Durant	PHO	47649433	0.4839	0.7143	-0.2304
Ja Morant	MEM	34005250	0.6667	0.2759	0.3908

Durant 本赛季出战 31 场，球队获胜 15 场，胜率 48.39%；未出战 7 场，球队获胜 5 场，胜率高达 71.43%。虽然本赛季 Durant 真实命中率高达 64.1%，但是胜率的差距还是一定程度上体现他的单打对球队进攻的负面影响。

Morant 由于 25 场的禁赛，本赛季只在近期解禁后出战 9 场，赢下其中 6 场。而球队在没有 Morant 时在 29 场中仅赢下 8 场，胜率只有 27.59%，比他出战时胜率少了约 40%，可见 Morant 在球队的核心地位。

此外，本实验对 PM 值与胜率提升的关系进行了探究。选取球员上场与缺席时球队胜率差值为 y 轴指标，PM 值为 x 轴指标，并把不同球队的球员用不同颜色显示，再把之前的表格按照球队名排序，以便分析同一队伍球员的数据。



可见 PM 值小于 0 的球员大多出战胜率比缺席胜率低，而球队有多位球员处在图表左下角时战绩相对惨淡，例如活塞（DET）3 胜 36 负，开拓者（POR）10 胜 28 负，篮网（BRK）16 胜 22 负，公牛（CHI）18 胜 22 负。而处于相对靠外的球员所在的球队战绩相对较好，例如骑士（CLE）22 胜 15 负，独行侠（DAL）23 胜 16 负，掘金（DEN）27 胜 13 负。但灰熊（MEM）的两位高新球员都处于靠外的位置，却只取得 14 胜 24 负的战绩，这主要时由于主将 Morant 被长期禁赛导致的。

4) 对勇士队（GSW）的分析

对于备受瞩目的金州勇士（GSW），本赛季目前只取得 18 胜 20 负位列西区倒数第 4，以下数据能体现其中原因：

Name	Team	salary	ts	PM
Stephen Curry	GSW	51915615	0.621	-7.9
Klay Thompson	GSW	43219440	0.578	-4.1
Chris Paul	GSW	30800000	0.543	10
Andrew Wiggins	GSW	24330357	0.499	-15.7
Draymond Green	GSW	22321429	0.619	-7.5

其中 Andrew Wiggins 状态尤其低迷，近期已经被剥夺首发位置。他的真实命中率甚至低于五成，PM 值低至-15.7，可谓严重伤害了球队。此外，Klay Thompson 本赛季真实命中率低于高薪球员平均值，这并不符合顶级射手的水准；Chris Paul 此项数据也相对惨淡，未达 55%；Draymond Green 因暴力行为被长期禁赛，这使球队的防守收到严重打击。再深入分析，结合核心球员前几赛季的发挥，可以发现 Stephen Curry 的真实命中率处于近 10 年的低位。

2013-14 ★	25	GSW	NBA	PG	.610
2014-15 ★	26	GSW	NBA	PG	.638
2015-16 ★	27	GSW	NBA	PG	.669
2016-17 ★	28	GSW	NBA	PG	.624
2017-18 ★	29	GSW	NBA	PG	.675
2018-19 ★	30	GSW	NBA	PG	.641
2019-20	31	GSW	NBA	PG	.557
2020-21 ★	32	GSW	NBA	PG	.655
2021-22 ★	33	GSW	NBA	PG	.601
2022-23 ★	34	GSW	NBA	PG	.656

这些因素导致了本赛季 Stephen Curry, Klay Thompson, Andrew Wiggins, Draymond Green 的组合每百回合平均净负 9.6 分（下表 PTS 一栏）。显然，这是勇士本赛季如此低迷的原因。

Rk	Lineup	Tm	MP	FG	FGA	FG%	3P	3PA	3P%	eFG%	FT	FTA	FT%	PTS
1	S. Curry K. Looney K. Thompson A. Wiggins	GSW	335:23	-3.4	+1.0	-.042	+1.6	+4.2	-.001	-.034	-2.4	-2.7	-.011	-7.5
2	S. Curry D. Green K. Thompson A. Wiggins	GSW	211:57	-4.6	+1.3	-.060	+1.2	+3.5	-.004	-.055	-1.5	-1.7	-.004	-9.6

五、总结

本次实验所采用的数据指标都一定程度反映了高薪球员对球队的影响，核心球员的表现很大程度决定了球队的实力。然而一个球员还有很多其他方面的数据会影响球队表现，本次实验并未涉及。就本次实验所涉及数据而言，年薪超过 3500 万的球员有出色的进攻效率，帮助球队在场上取得优势，他们对球队胜率的提升也体现出他们对于球队不可缺少，占据核心地位。