Hupu Spider

数据准备

data/

games/

赛事的基本信息,如时间、参与双方详情、比赛结果、相关比赛数据以及相关精彩瞬间图片等。



虎扑的比赛赛程(<u>https://nba.hupu.com/schedule/2020-03-22</u>) 中也会展示比赛信息。对于尚未进行的比赛,可以通过"比赛前瞻"查看两支球队的历史数据,以及**预测胜率**。对于已经结束的比赛,可以通过"数据统计"



• 2020-3-1

o 勇士vs太阳

■ game_base_info.csv, 比赛的基本信息,包括:

■ 比赛时间: 2020年03月01日 10:00

■ 耗时: 02:13

■ 球馆:托金斯迪克度假酒店球馆

■ 上座: 16395人 ■ game_score_info.csv

■ 比分情况:包括各小节与总比分

away team score table.csv

- 各个球员在该场比赛中的技术统计,包括:姓名,位置,时间,投篮,3分,发球,前场,后场,篮板,助攻,犯规,抢断,事务,封盖,得分,+/-(净胜分)等信息的记录。
- 对整个队的在该场比赛中的技术统计。
- home_team_score_table.csv
 - 同上
- game_recap.csv
- capture.csv
- 2020-3-2

players/

统计以下30支球队信息。



参考下图,包括球队基本信息,各个赛季的技术统计,以及球队阵容与各球员的详细数据(基本数据与比赛数据)。

休斯顿火箭(Houston Rockets) 赛程 | 球员 | 论坛 | 新闻



进入NBA: 1967年

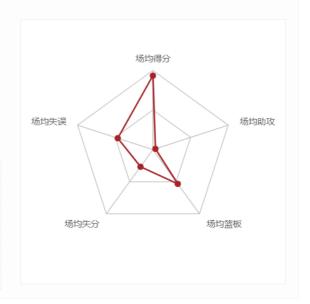
主场: 丰田中心 分区: 西南赛区

官网: http://www.nba.com/rockets/

主教练:迈克-德安东尼

体斯顿火箭队是一支有着高大中锋传统的球队,从早期的摩西-马龙、拉尔夫-桑普森到哈基姆-奥拉朱旺,以及2002年来自中国的状元秀姚明。

技术统计	数值	联盟排名
场均得分	118.1	2
场均助攻	21.5	29
场均篮板	44. 92	14
场均失分	114.4	22
场均失误	14.67	16



火箭本赛季表现

得分	出手数	命中率	3分得分	3分出手数	3分命中率
罚球出手数	罚球命中次数	罚球命中率	防守篮板	进攻篮板	篮板球数
助攻	抢断	盖帽	失误	犯规	

火箭阵容与数据											
		基本数据	比赛数	效据							
号码	球员	场数	首发数	时间	得分	助攻					
0	拉塞尔-威斯布鲁克	53	53	35.9	27.5	7.0					
13	詹姆斯-哈登	61	61	36.7	34.4	7.4					
10	埃里克-戈登	34	13	28.6	14.5	1.5					
33	罗伯特-科温顿	14	13	32.6	12.8	1.3					
17	PJ-塔克	64	64	34.4	7.1	1.6					
4	丹纽尔-豪斯	58	47	30.0	10.2	1.3					
19	泰森-钱德勒	25	5	8.8	1.4	0.2					
18	萨博-塞福洛沙	41	0	10.5	2.2	0.6					
25	奥斯汀-里弗斯	60	4	23.4	8.4	1.6					
16	本-麦克勒摩	63	19	22.9	9.8	0.9					
5	布鲁诺-卡博克洛	5	0	5.6	2.0	0.4					
55	以賽亚-哈尔滕施泰因	23	2	11.5	4.7	0.8					
3	克里斯-克莱蒙斯	29	0	8.7	4.9	0.8					
9	德马雷-卡罗尔	6	0	15.7	6.7	1.2					
32	杰夫-格林	10	0	20.1	10.4	1.2					
52	威廉-霍华德	2	0	6.5	0	0.5					
21	迈克尔-弗雷泽	7	0	9.3	2.0	0.1					

teams/

包括球员基本信息,以及更加详细的技术统计数据(加入对每场比赛的数据统计)。

詹姆斯-哈登 (James Harden)



位置: G(13号)

身高: 1.96米/6尺5 体重: 100公斤/220磅

生日: 1989-08-26

球队: 休斯顿火箭

学校:亚利桑那州立大学 选秀: 2009年第1轮第3顺位

国籍:美国

本赛季薪金:3819万美元

合同: 2016年夏天在旧合同期内以4年1.18 亿美元提前续约至2020年,2017年夏天再次

以4年1.7亿美元提前续约至2023年,2022-

23赛季球员选项

技术统计	数值	联盟排名
场均得分	34.4	1
场均助攻	7.4	9
场均篮板	6.4	43
投篮命中率	43.5%	88
三分命中率	35.2%	103
罚球命中率	86.1%	21
场均盖帽	0.9	26
场均抢断	1.7	5

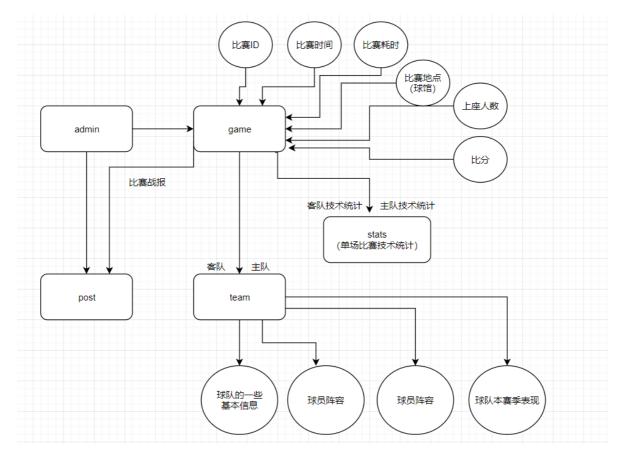
根据球员61场比赛计算得到排名

詹姆斯-哈登常规赛表现 詹姆斯-哈登季后赛表现						l							~	7)换至	图表模	式			
本赛季常规赛平均数据																			
	场次	时间	投篮	命中率	\$	三分	命中图	<u>\$</u>	罚球	命中図	壑 篮板	助攻	て 抢!	断 盖	輔 结	失误	犯规	得分	
	61	36.7	9.9-22.7	43.5%	, 4	1.4-12.6	35.29	% 1	0.1-11.8	86.19	% 6.3	7.4	1.7	72 0.	87 4	1.43	3.38	34.4	
	03月数据																		
	日期	对手	比分	E	时间	投篮	命中率	三分	命中率	罚球	命中率	篮板	助攻	抢断	盖帽	失误	犯规	得分	
	03/11	森林狼	111-117	(胜)	36	11-19	57.9%	5-10	50.0%	10-14	71.4%	4	7	2	1	3	1	37	
	03/09	魔术	126-106	6(败)	32	6-19	31.6%	3-10	30.0%	8-10	80.0%	7	6	3	0	4	3	23	
	03/08	黄蜂	99-108(败)	45	8-22	36.4%	2-11	18.2%	12-14	85.7%	10	14	2	0	10	5	30	
	03/06	快船	120-105	5(败)	28	4-17	23.5%	0-8	0%	8-11	72.7%	7	4	1	0	4	3	16	

03/03 尼克斯 123-125(败) 39 8-22 36.4% 3-13 23.1% 16-16 100% 2 8 4 0 3 4 35

数据库设计

• 简版如下



• game

- 。 比赛ID (主键)
- 。 比赛的基本信息: 时间, 耗时, 地点, 上座人数, 比分 (115:120) 。
- 。 比赛双方: 客队, 主队, 这里仅存球队的简称(比如LAL)作为外键, 以链接team表。
- 。 比赛技术统计: 客队技术统计, 主队技术统计, 这里也仅存ID, 作为外键, 以链接stats表。

team

- 球队名简称 (主键)
- 。 球队的基本信息:
- 。 球员阵容: 这个我还没考虑好。
- 。 球队本赛季表现:这里有两种实现。
 - 1. 不存数据库,在最后需要展示该数据时,利用stats表,筛选出来进行累加即可,同时也能保证数据一致性。
 - 2. 无条件信任虎扑,直接从虎扑上爬取本赛季的表现,并参考stats做成实体表。
- stats (单场比赛技术统计)
- 说明
 - o game中比较特殊的是stats,由于包含的信息较多,因此我想的是考虑作为一个实体,game 表中只存ID,但是技术统计的信息包含的太杂了,具体怎么设计我没想好,可能需要大家集 思广益一下。
 - 所有比赛相关的信息,可以通过 ./data/比赛时间/客队vs主队/ 获得。
 - 例, ./data/2019-01-01/GSWvsPHX/
 - 包括比赛基本信息,比分信息,客队技术统计,主队技术统计,比赛战报。

BUG记录

2020-02-13

这里能采集到精彩瞬间的图片,但是链接失效,导致HTTPs Error



2019-02-21

数据缺失



2019-10-02

当天没有比赛,导致将10-04的比赛数据错误的给了10-02。

