

Exp 3 英雄联盟游戏数据分析

和泳毅 PB19010450

Exp 3 英雄联盟游戏数据分析

实验描述

数据分析

1. Team1/2胜利分布
 - 1.1 全部数据中Team1/2胜利分布
 - 1.2 不同地图中Team1/2胜利分布
 - 1.3 不同队列中Team1/2胜利分布
2. 相关系数热图
3. 一血
 - 3.1 一血的分布
 - 3.2 一血对获胜的影响
 - 3.3 拿下一血对拿下一龙的影响
 - 3.4 拿下一血对拿下一先锋的影响
 - 3.5 拿下一血对拿下一塔的影响
 - 3.6 拿下一血对拿下一男爵的影响
 - 3.7 拿下一血对拿下一水晶的影响
4. 一龙
 - 4.1 一龙的分布
 - 4.2 击杀第一条大龙对获胜的影响
 - 4.3 拿下一龙对拿下一先锋的影响
 - 4.4 拿下一龙对拿下一塔的影响
 - 4.5 拿下一龙对拿下一男爵的影响
 - 4.6 拿下一龙对拿下一水晶的影响
 - 4.7 同时拿下一血一龙对获胜的影响
5. 一先锋
 - 5.1 一先锋的分布
 - 5.2 拿下第一个先锋对获胜的影响
 - 5.3 拿下一先锋对拿下一塔的影响
 - 5.4 拿下一先锋对拿下一男爵的影响
 - 5.5 拿下一先锋对拿下一水晶的影响
 - 5.6 拿下一先锋对拿下一龙的影响
 - 5.7 同时拿下一龙一先锋对获胜的影响
 - 5.8 拿下一先锋而对手拿下一龙对获胜的影响
6. 一塔
 - 6.1 一塔的分布
 - 6.2 拿下一塔对获胜的影响
 - 6.3 拿下一塔对拿下一男爵的影响
 - 6.4 拿下一塔对拿下一水晶的影响
7. 一男爵
 - 7.1 一男爵的分布
 - 7.2 拿下一个男爵对获胜的影响

- 7.3 拿下一个男爵对拿下一水晶的影响
 - 7.4 同时拿下一男爵一先锋对获胜的影响
 - 7.5 拿下一男爵而对手拿下一先锋
- 8. 一水晶
 - 8.1 一水晶的分布
 - 8.2 拿下一个水晶对获胜的影响
 - 8.3 同时拿下一塔一水晶对获胜的影响
 - 8.4 拿下一塔而对手拿下一水晶对获胜的影响
- 9. 最大击杀数
 - 9.1 获胜队伍最大玩家击杀数分布
 - 9.2 最大玩家击杀数卡方拟合和泊松拟合
 - 9.3 最大击杀数对获胜影响
- 10. 场均击杀与击杀差
 - 10.1 获胜队伍场均击杀数分布
 - 10.2 获胜队伍场均击杀差分布
 - 10.3 场均击杀数正态拟合
 - 10.3.1 拟合图
 - 10.3.2 QQ图
 - 10.3.3 Kolmogorov-Smirnov 检验
 - 10.4 场均杀敌数对获胜影响
- 11. 最大助攻数
 - 11.1 获胜队伍玩家最大助攻数分布
 - 11.2 玩家最大助攻数卡方拟合和泊松拟合
 - 11.3 最大助攻数对获胜影响
- 12. 场均助攻与助攻差
 - 12.1 获胜队伍场均助攻数分布
 - 12.2 获胜队伍场均助攻差分布
 - 12.3 场均助攻数正态拟合和卡方拟合
 - 12.4 场均助攻数对获胜影响
- 13. 最大死亡数
 - 13.1 获胜队伍玩家最大死亡数分布
 - 13.2 玩家最大死亡数泊松拟合
 - 13.3 最大死亡数对获胜影响
- 14. 场均死亡与死亡差
 - 14.1 获胜队伍场均死亡数分布
 - 14.2 获胜队伍场均死亡差分布
 - 14.3 场均死亡数卡方拟合
 - 14.4 场均死亡数对获胜影响
- 15. KDA
 - 15.1 获胜队伍总KDA分布
 - 15.2 KDA卡方分布和伽玛分布拟合
- 16. 总经济和经济差
 - 16.1 获胜队伍总经济分布
 - 16.2 获胜队伍经济差分布
 - 16.3 经济差正态拟合
 - 16.3.1 拟合图
 - 16.3.2 QQ图

- 16.3.3 Kolmogorov–Smirnov 检验
- 16.4 经济差T分布拟合
 - 16.4.1 拟合图
 - 16.4.2 Kolmogorov–Smirnov 检验
- 16.5 经济差对获胜影响
- 17. 总英雄出场率
 - 17.1 Team1英雄出场数排行
 - 17.2 Team2英雄出场数排行
 - 17.3 英雄总出场率排行前十
- 18. 胜利队伍英雄出场率
 - 18.1 Team1获胜英雄出场数排行
 - 18.2 Team2获胜英雄出场数排行
 - 18.3 胜利队伍英雄总出场率排行前十
- 19. 英雄胜率排行
 - 19.1 前十
 - 19.2 后十
- 20. 英雄KDA排行
- 21. 英雄K/D/A排行
 - 21.1 K值
 - 21.2 A值
 - 21.3 D值
- 22. 获胜队伍中英雄常见位置和分路
- 23. 获胜队伍中英雄常见位置+分路搭配
- 总结
- 感想

实验描述

实验背景：

英雄联盟（LoL）是美国游戏开发商Riot Games（2011年由腾讯收购）开发和发行的一款多人在线战斗竞技游戏。

在游戏中，玩家扮演一个"召唤师"角色，每个召唤师控制一个拥有独特技能的"英雄"，并与一组其他玩家或电脑控制的英雄战斗。游戏的目标是摧毁对方的防御塔和基地。

召唤者峡谷是英雄联盟中最受欢迎的地图，在这种地图类型中，两队五名玩家竞争摧毁一个被称为基地的敌人建筑，这个建筑由敌人的队伍和一些防御塔护卫。每个队伍都希望保卫自己的建筑，同时摧毁对方的建筑。这些主要包括：

- Towers（防御塔）：每支队伍总共有11座防御塔
- Inhibitor（水晶）：每条道有一个水晶
- Elder Dragon（大龙）
- Rift Herald（峡谷先锋）
- Baron Nasho（纳什男爵）

- Nexus (基地)

英雄联盟最具争议的元素之一，就是其严重的滚雪球效应。许多职业选手接受赛后采访时都提到其输赢都因为“滚雪球”，我们研究各方面各指标的数据，来看这些因素的发展是否真的影响了比赛的成败。在这个实验中，我们分析了8万场英雄联盟的排名比赛，尝试得出有效的结论。

实验要求：

- 给定一个数据集和预测任务，需要分析数据以及抽取特征
- 数据分析
 - 单个特征的分布
 - 异常样本
 - 特征间的相关性
 - 特征与预测目标的相关性 ...
- 特征抽取
 - 特征的变换，如：str -> int,取log
 - 尝试组合特征
 - 特征 aggregation

数据集说明：

- team1_win代表team1是否取得胜利
- team1_firstBlood代表team1是否取得一血，其他类似特征同理
- player1_championId代表玩家1选择的英雄ID
- player1_kills代表玩家1的击杀数，其他类似特征同理
- player1 ~ player5属于team1，player6 ~ player10属于team2

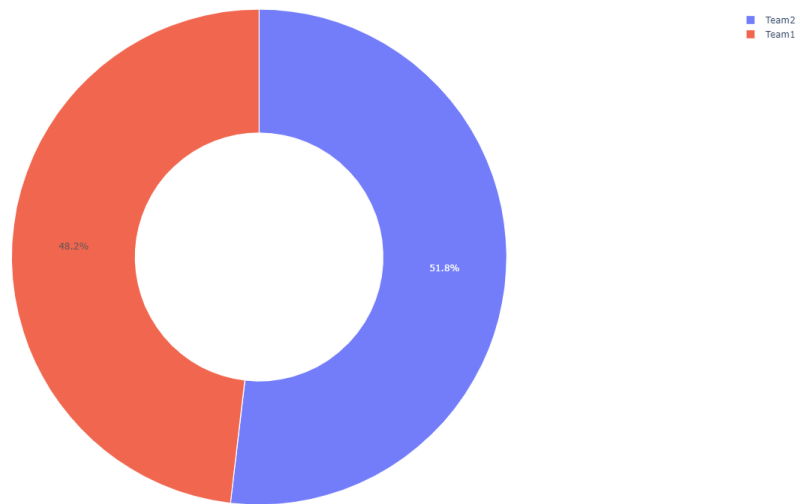
数据分析

1. Team1/2胜利分布

首先我们要对样本的分布平衡性进行分析，我们的目标是获胜，如果样本的两队获胜情况不平衡，说明该样本是有问题的。

1.1 全部数据中Team1/2胜利分布

全部比赛里Team1/2获胜分布



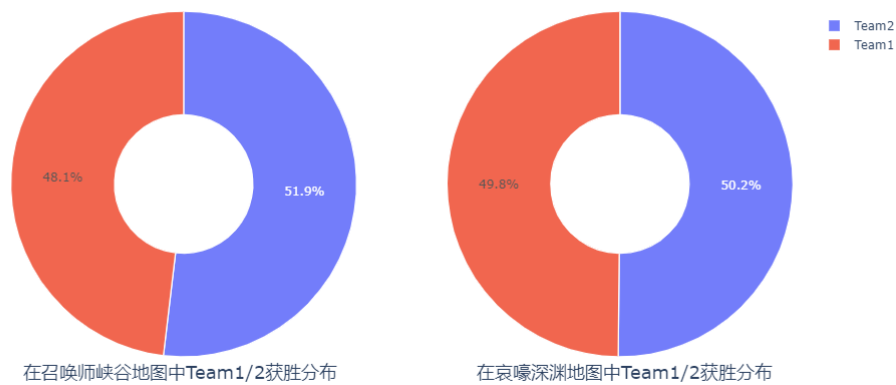
样本一共有80000条记录，其中team1获胜38560次，占比48.2%，team2获胜41440次，占比51.8%。因此可以认为样本胜负分布是平衡的。

1.2 不同地图中Team1/2胜利分布

地图名	数量
Summoner's Rift	79001
Howling Abyss	999

80000条记录中，地图只有11号和12号两个情况，分别对应召唤师峡谷（*Summoner's Rift*）和嚎哭深渊（*Howling Abyss*）。其中召唤师峡谷地图出现79001次，占比98.75%，嚎哭深渊出现99次，占比1.25%。可以认为样本的绝大部分地图选择都是召唤师峡谷。

不同地图中Team1/2胜利分布



并且在召唤师峡谷中，team1获胜38024次，占比48.1%，team2获胜40977次，占比51.9%，可以认为在该地图中胜负分布是平衡的，并且和全样本高度相似。

在嚎哭深渊中，team1获胜498次，占比49.8%，team2获胜501次，占比50.2%，可以认为在该地图中胜负分布是平衡的。

1.3 不同队列中Team1/2胜利分布

模式	数量
5v5 Ranked Solo games	50146
5v5 Blind Pick games	20966
5v5 Ranked Flex games	4016
Clash games	2594
URF games	1023
5v5 ARAM games	999
One for All games	256

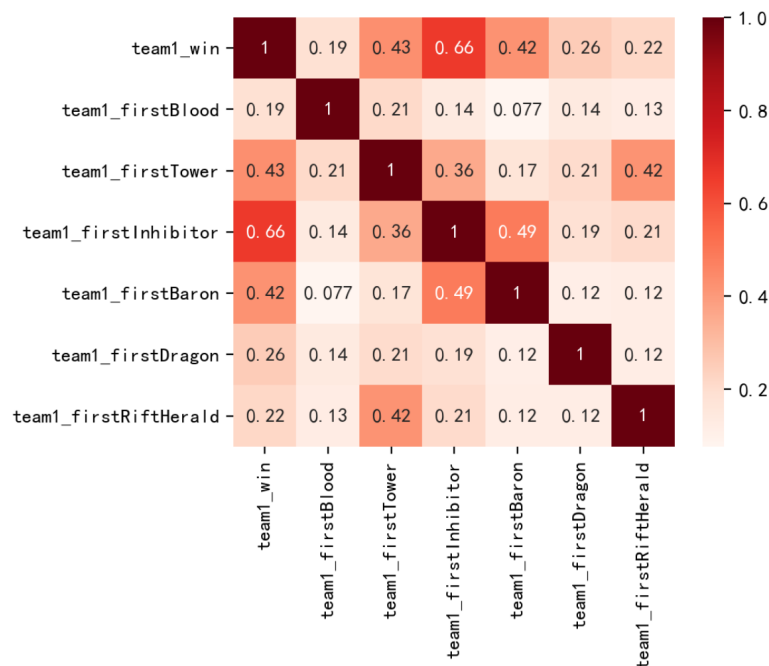
80000条记录中，队列420，430，440，700，900，450，1020 分别对应单排（*Ranked Solo games*），匹配（*Blind Pick games*），灵活组排（*Ranked Flex games*），冠军杯（*Clash games*），无限火力（*URF games*），大乱斗（*ARAM games*），克隆（*One for All games*）。分布占比62.68%，26.21%，5.02%，3.24%，1.28%，1.25%，0.32%。

不同队列中Team1/2胜利分布



对不同地图的获胜分布做分析，注意到在灵活组排模式中的获胜分布为38%和62%，这是非常不平衡的，虽然灵活组排只占总记录的5.02%，但为了接下来针对获胜情况分析的准确性，我们在所有记录中删去模式为灵活组排的4016条。处理后的记录有75984条。

2. 相关系数热图



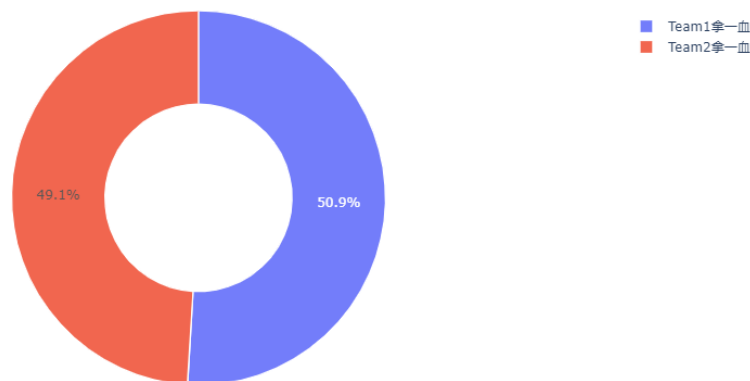
英雄联盟比赛中，最外显的目标是一血、一塔、一水晶、一男爵、一龙、一先锋。因此我们用这几项数据对获胜做出相关热力图，可以大概地看出相关性的高低。其中拿下一水晶对获胜的相关性最大，之后我们会重点关注一水晶。

3. 一血

在一般比赛中，这些目标出现的大致顺序为：**一血** → **(一龙、一先锋)** → **一塔** → **一男爵** → **一水晶**，其中一龙和一先锋的顺序一般是不确定的。故我们按照这个顺序去分析，并且也要注意前向对后向对影响。

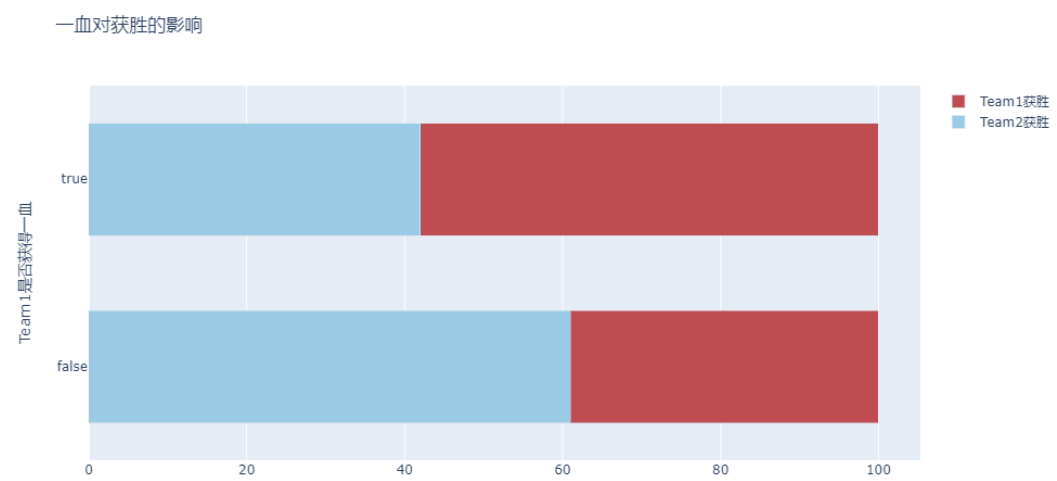
3.1 一血的分布

Team1/2拿一血的分布



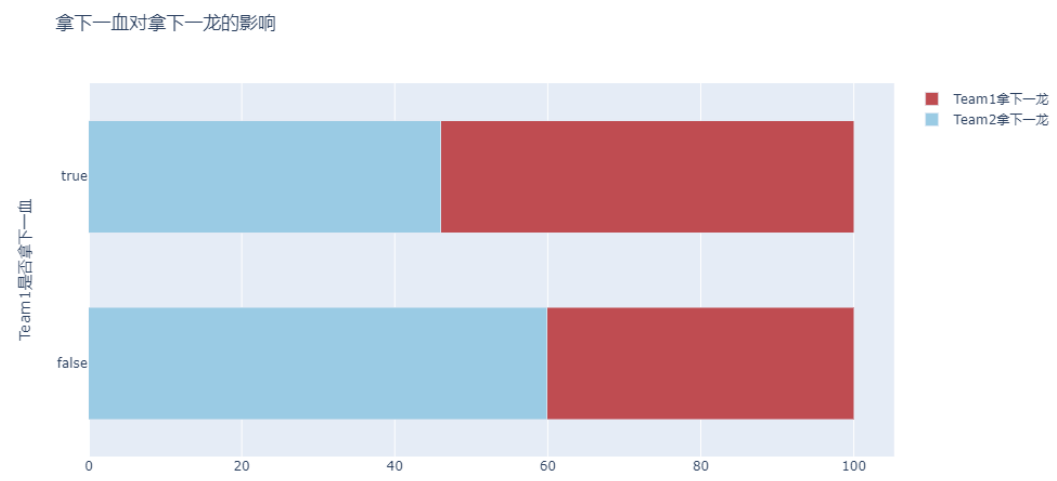
75984条记录中，其中team1拿一血38704次，占比50.9%，team2拿一血37280次，占比49.1%，是平衡的。

3.2 一血对获胜的影响



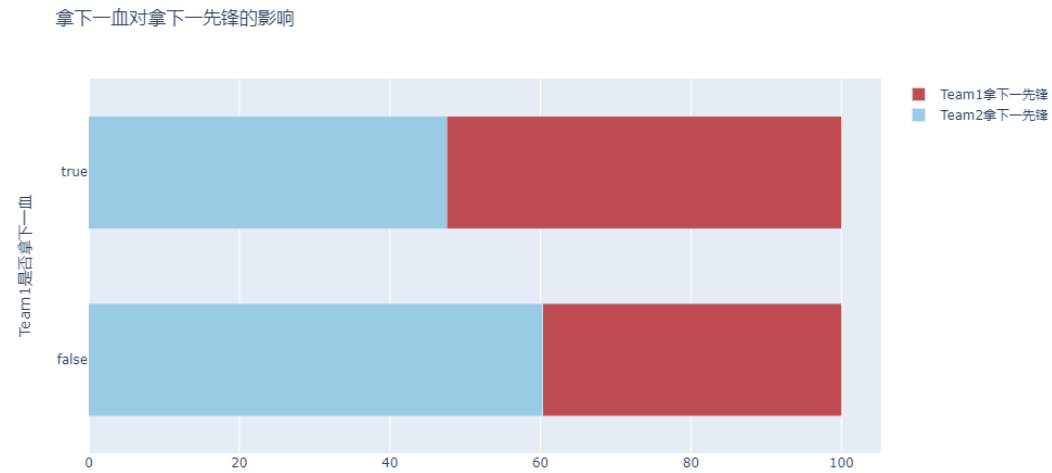
由对比图得知拿下一血的队伍胜率相对较高，在team1的比赛中，首先拿下一血时的胜率为58.04%，相较未拿下一血的比赛高19.05%。在team2的比赛中，拿下一血时的胜率为61.01%，相较未获得一血的比赛高19.05%。**平均看来，拿下一血时的胜率为59.525%，拿下一血与获胜有一定的相关性但相关性并不是很高。**

3.3 拿下一血对拿下一龙的影响



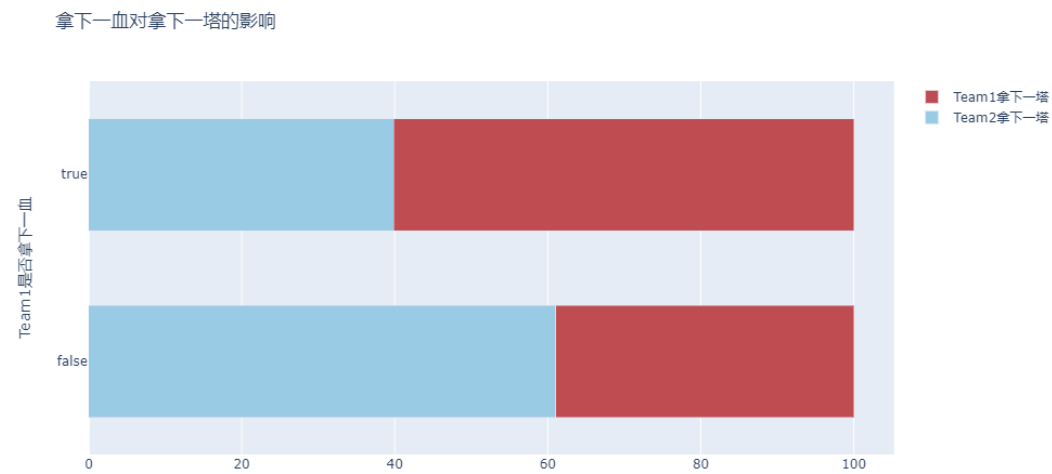
由对比图得知拿下一血的队伍后续拿下一龙的概率相对较高，在team1的比赛中，首先拿下一血后拿下一龙的概率为53.97%，相较未拿下一血的比赛高13.89%。在team2的比赛中，拿下一血后拿下一龙的概率为59.92%。**平均看来，拿下一血时后续拿下一龙的概率为56.915%，拿下一血与拿下一龙的相关性并不是很高。**

3.4 拿下一血对拿下一先锋的影响



由对比图得知拿下一血的队伍后续拿下一先锋的概率相对较高，在team1的比赛中，首先拿下一血后拿下一先锋的概率为52.41%，相较未拿下一血的比赛高12.74%。在team2的比赛中，拿下一血后拿下一先锋的概率为60.33%。**平均看来，拿下一血时后续拿下一先锋的概率为56.37%，拿下一血与拿下一先锋的相关性并不是很高。**

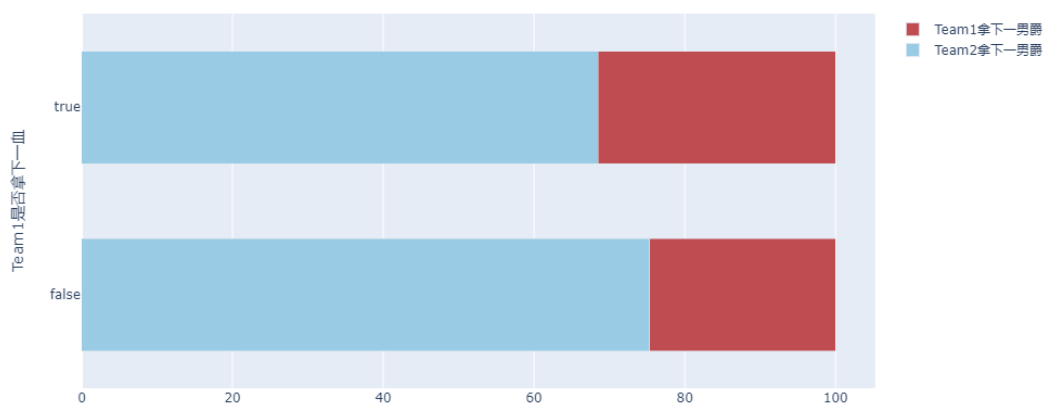
3.5 拿下一血对拿下一塔的影响



由对比图得知拿下一血的队伍后续拿下一塔的概率相对较高，在team1的比赛中，首先拿下一血后拿下一塔的概率为60.14%，相较未拿下一血的比赛高21.27%。在team2的比赛中，拿下一血后拿下一塔的概率为61.03%。**平均看来，拿下一血时后续拿下一塔的概率为60.585%，拿下一血与拿下一塔有一定的相关性。**

3.6 拿下一血对拿下一男爵的影响

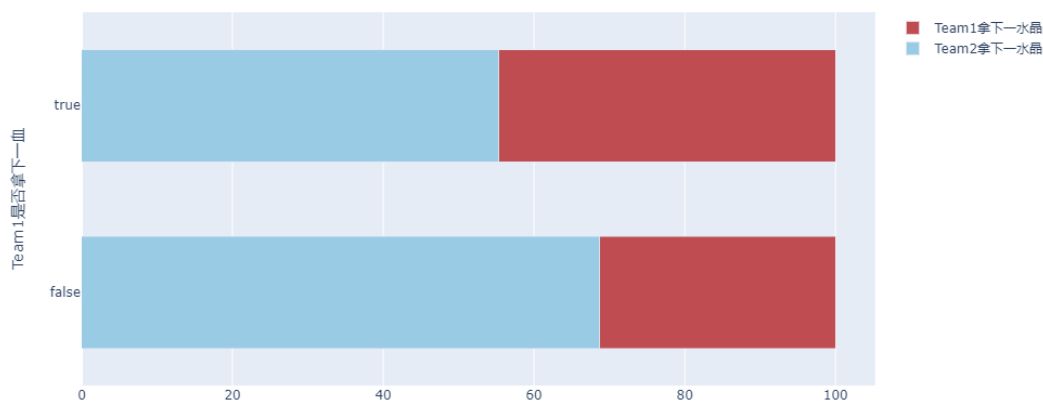
拿下一血对拿下一男爵的影响



在team1的比赛中，首先拿下一血后拿下一男爵的概率为31.49%，相较未拿下一血的比赛高6.85%。在team2的比赛中，拿下一血后拿下一男爵的概率为75.36%。**平均看来，拿下一血时后续拿下一男爵的概率为53.425%，拿下一血与拿下一男爵的相关性不高。**同时我们注意到无论team1是否拿一血，team2拿一男爵的概率都要远大于team1，这是不正常的，后续分析一男爵时要注意这个问题。

3.7 拿下一血对拿下一水晶的影响

拿下一血对拿下一水晶的影响

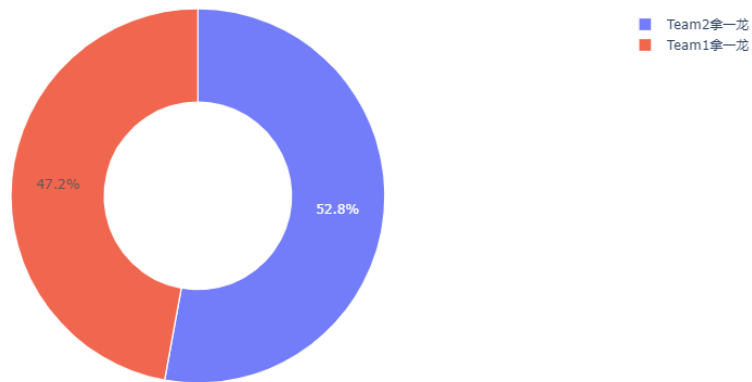


在team1的比赛中，首先拿下一血后拿下一水晶的概率为44.65%，相较未拿下一血的比赛高13.35%。在team2的比赛中，拿下一血后拿下一水晶的概率为68.7%。**平均看来，拿下一血时后续拿下一水晶的概率为56.675%，拿下一血与拿下一水晶的相关性并不是很高。**

4. 一龙

4.1 一龙的分布

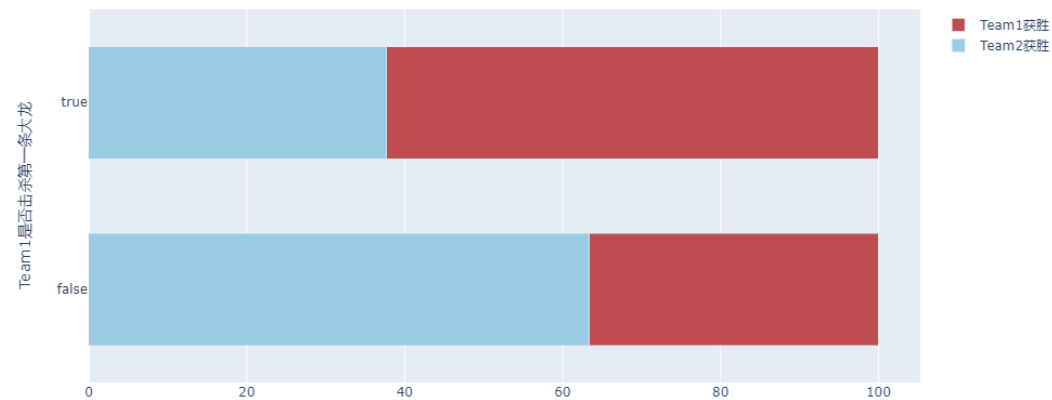
Team1/2拿一龙的分布



75984条记录中，其中team1拿一龙35830次，占比47.2%，team2拿一龙40154次，占比52.8%，是平衡的。

4.2 击杀第一条大龙对获胜的影响

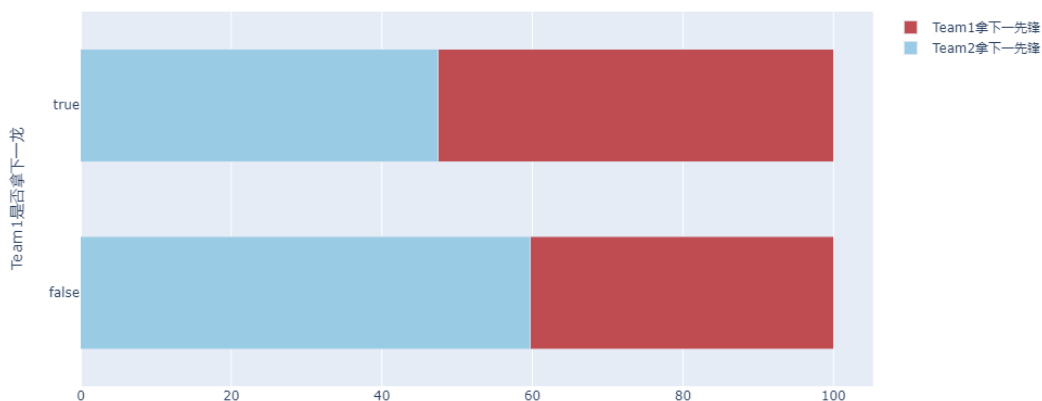
击杀第一条大龙对获胜的影响



由对比图得知拿下一龙的队伍胜率相对较高，在team1的比赛中，拿下一龙时的胜率为62.26%，相较未拿下一龙的比赛高25.68%。在team2的比赛中，拿下一龙时的胜率为63.42%。**平均看来，拿下一龙时的胜率为62.84%，拿下一龙与获胜有一定的相关性。**

4.3 拿下一龙对拿下一先锋的影响

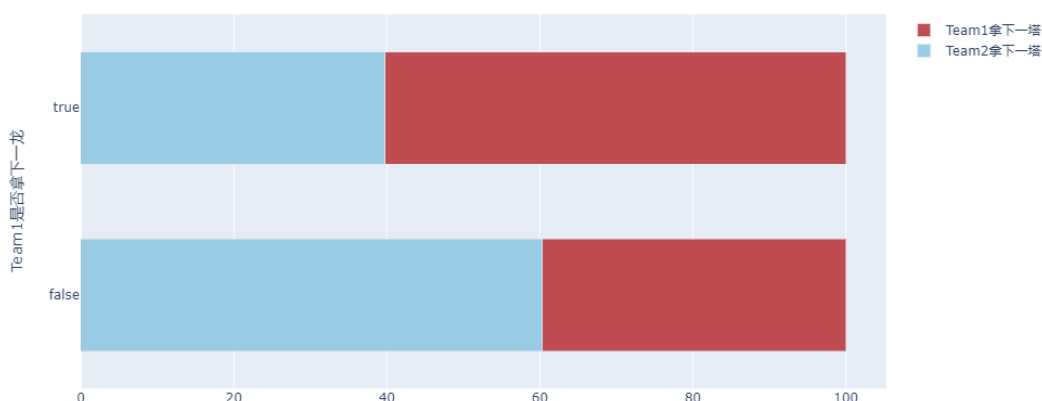
拿下一龙对拿下一先锋的影响



由对比图得知拿下一龙的队伍后续拿下一先锋的概率相对较高，在team1的比赛中，拿下一龙后拿下一先锋的概率为52.66%，相较未拿下一龙的比赛高12.3%。在team2的比赛中，拿下一龙后拿下一先锋的概率为59.64%。**平均看来，拿下一龙时后续拿下一先锋的概率为56.15%，拿下一龙与拿下一先锋的相关性并不是很高。**

4.4 拿下一龙对拿下一塔的影响

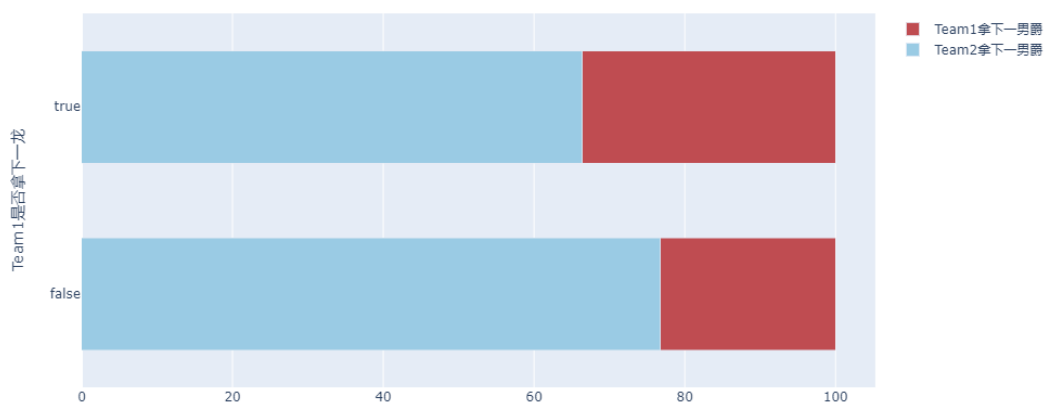
拿下一龙对拿下一塔的影响



由对比图得知拿下一龙的队伍后续拿下一塔的概率相对较高，在team1的比赛中，拿下一龙后拿下一塔的概率为60.59%，相较未拿下一龙的比赛高20.51%。在team2的比赛中，拿下一龙后拿下一塔的概率为59.92%。**平均看来，拿下一龙时后续拿下一塔的概率为60.255%，拿下一龙与拿下一塔有一定的相关性。**

4.5 拿下一龙对拿下一男爵的影响

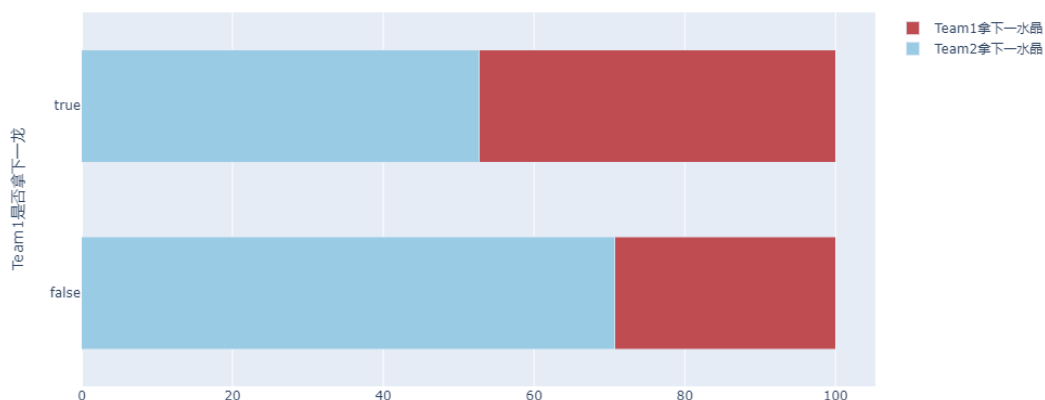
拿下一龙对拿下一男爵的影响



在team1的比赛中，拿下一龙后拿下一男爵的概率为33.6%，相较未拿下一龙的比赛高10.35%。在team2的比赛中，拿下一龙后拿下一男爵的概率为76.75%。
平均看来，拿下一龙时后续拿下一男爵的概率为55.175%，拿下一龙与拿下一男爵的相关性不高。

4.6 拿下一龙对拿下一水晶的影响

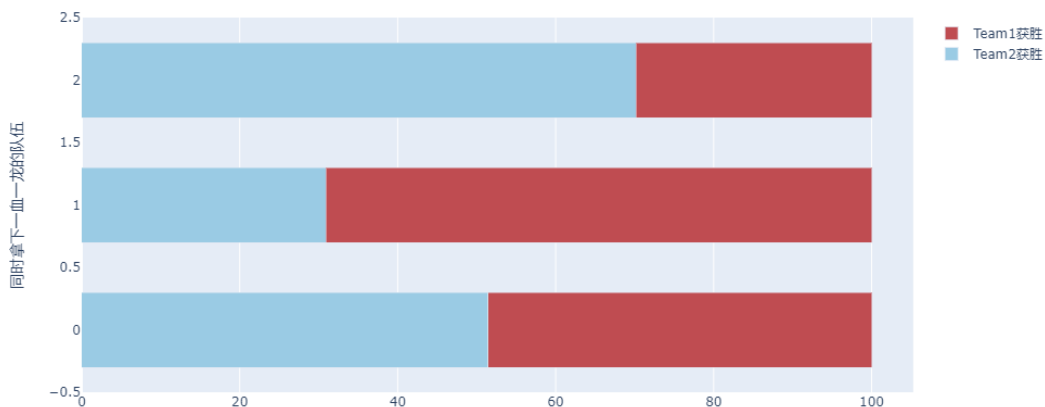
拿下一龙对拿下一水晶的影响



在team1的比赛中，拿下一龙后拿下一水晶的概率为47.61%，相较未拿下一龙的比赛高17.99%。在team2的比赛中，拿下一龙后拿下一水晶的概率为70.38%。
平均看来，拿下一龙时后续拿下一水晶的概率为58.995%，拿下一龙与拿下一水晶有一定的相关性。

4.7 同时拿下一血一龙对获胜的影响

同时拿下一血一龙对获胜的影响

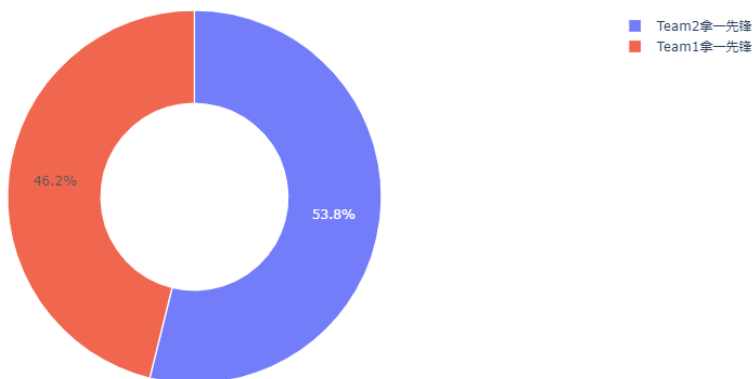


由对比图得知拿下一血和一龙的队伍胜率相对较高，在team1的比赛中，拿下一血和一龙时的胜率为69.1%。在team2的比赛中，拿下一血和一龙时的胜率为70.19%。平均看来，拿下一血和一龙时的胜率为69.645%，拿下一血和一龙与获胜有较高的相关性。

5. 一先锋

5.1 一先锋的分布

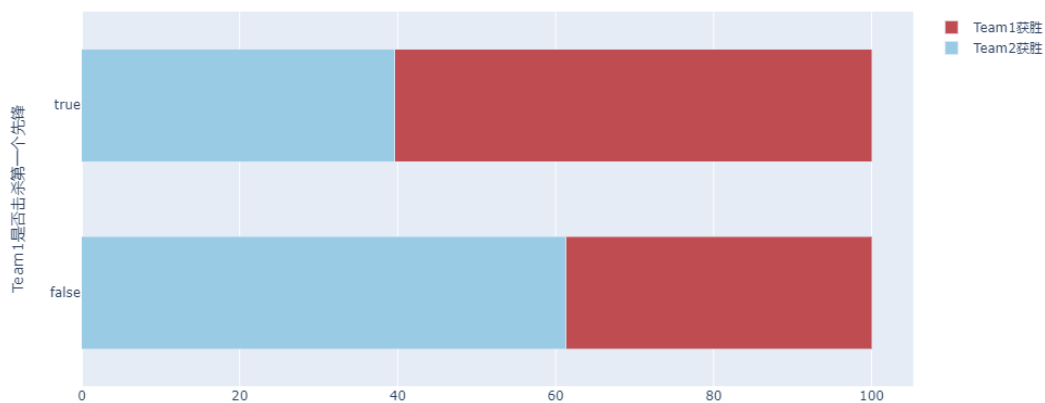
Team1/2拿一先锋的分布



75984条记录中，其中team1拿一先锋35074次，占比46.2%，team2拿一先锋40910次，占比53.8%，是平衡的。

5.2 拿下第一个先锋对获胜的影响

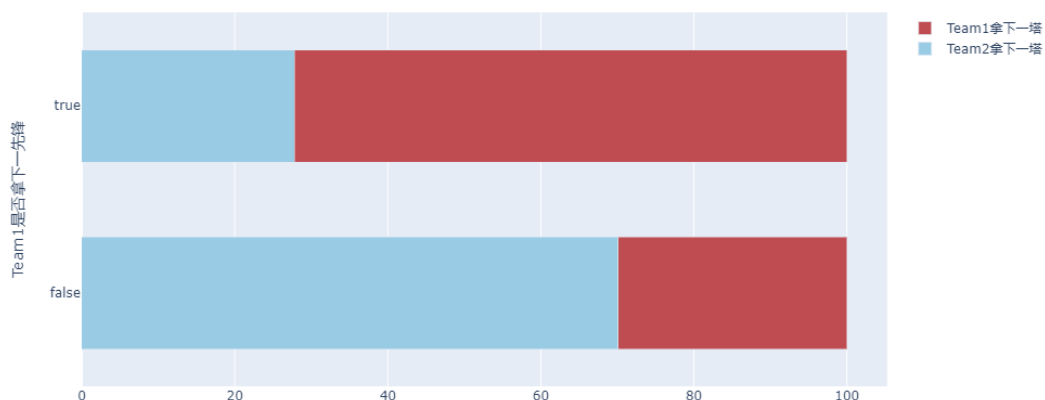
击杀第一个先锋对获胜的影响



由对比图得知拿下一先锋的队伍胜率相对较高，在team1的比赛中，拿下一先锋时的胜率为60.38%，相较未拿下一先锋的比赛高21.71%。在team2的比赛中，拿下一先锋时的胜率为61.33%。**平均看来，拿下一先锋时的胜率为60.85%，拿下一先锋与获胜有一定的相关性。**

5.3 拿下一先锋对拿下一塔的影响

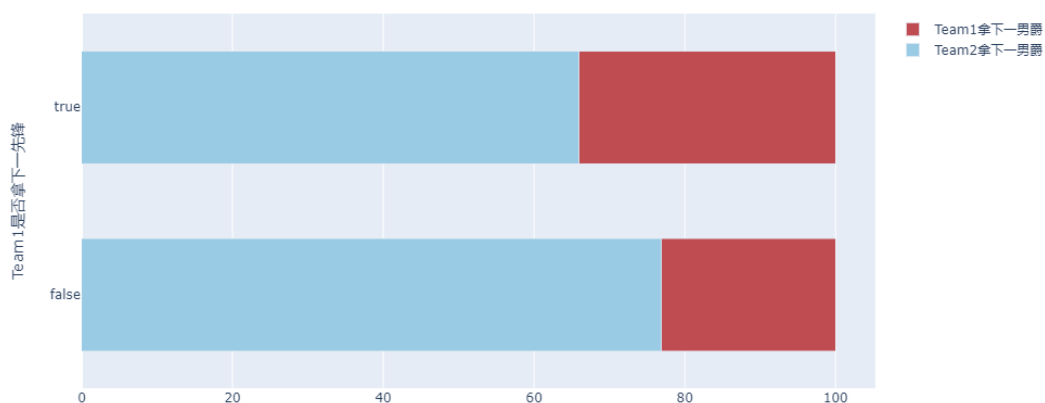
拿下一先锋对拿下一塔的影响



由对比图得知拿下一先锋的队伍后续拿下一塔的概率相对较高，在team1的比赛中，拿下一先锋后拿下一塔的概率为72.44%，相较未拿下一先锋的比赛高42.14%。在team2的比赛中，拿下一先锋后拿下一塔的概率为69.6%。**平均看来，拿下一先锋时后续拿下一塔的概率为71.07%，拿下一先锋与拿下一塔有较高的相关性。**

5.4 拿下一先锋对拿下一男爵的影响

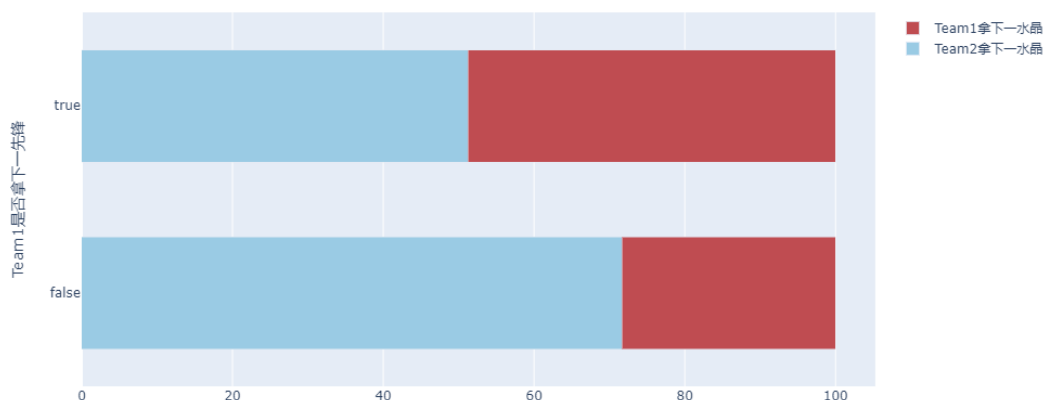
拿下一先锋对拿下一男爵的影响



在team1的比赛中，拿下一先锋后拿下一男爵的概率为34.04%，相较未拿下一先锋的比赛高10.97%。在team2的比赛中，拿下一先锋后拿下一男爵的概率为76.93%。平均看来，拿下一先锋时后续拿下一男爵的概率为55.485%，拿下一先锋与拿下一男爵的相关性不是很高。

5.5 拿下一先锋对拿下一水晶的影响

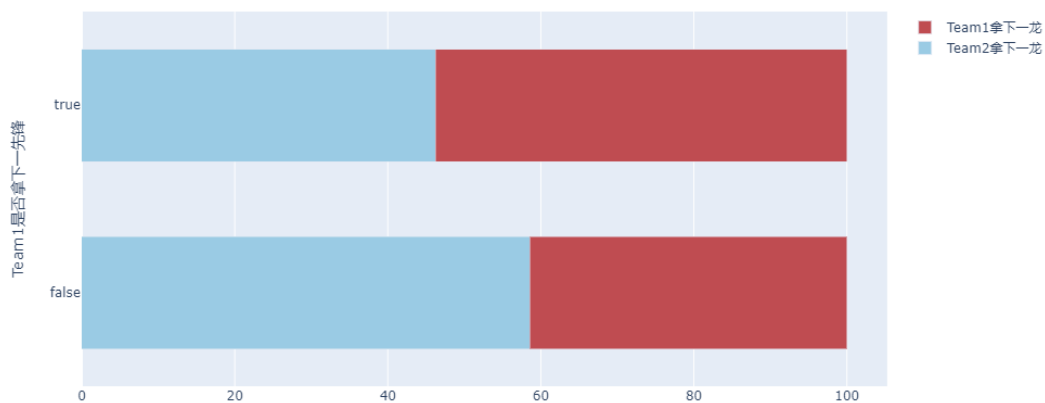
拿下一先锋对拿下一水晶的影响



在team1的比赛中，拿下一先锋后拿下一水晶的概率为49.03%，相较未拿下一先锋的比赛高20.3%。在team2的比赛中，拿下一先锋后拿下一水晶的概率为71.27%。平均看来，拿下一先锋时后续拿下一水晶的概率为60.15%，拿下一先锋与拿下一水晶的相关性较高。

5.6 拿下一先锋对拿下一龙的影响

拿下一先锋对拿下一龙的影响

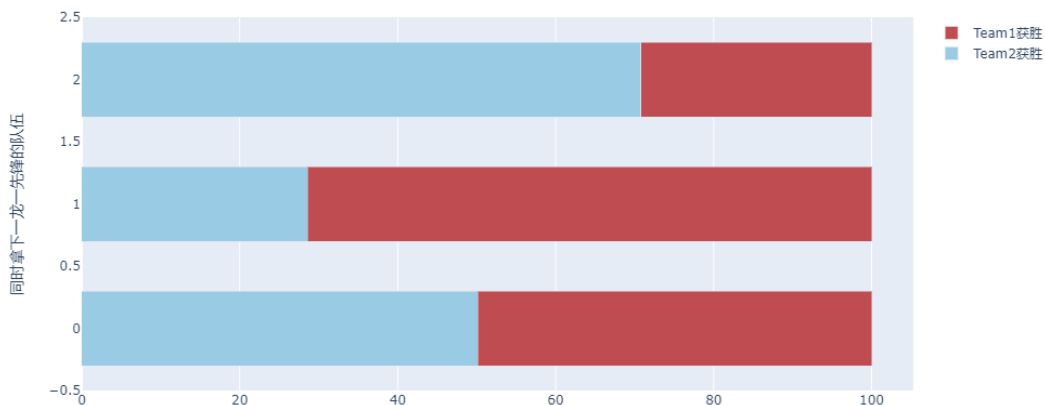


因为一先锋和一龙的先后关系不是很清晰，常常属于伴随关系，这里对一先锋和一龙多做一些结合分析。

由对比图得知拿下一先锋的队伍后续拿下一龙的概率相对较高，在team1的比赛中，拿下一先锋后拿下一龙的概率为53.79%，相较未拿下一先锋的比赛高12.33%。在team2的比赛中，拿下一先锋后拿下一龙的概率为58.54%。**平均看来，拿下一先锋时后续拿下一龙的概率为56.165%，拿下一先锋与拿下一龙的相关性不是很高。**

5.7 同时拿下一龙一先锋对获胜的影响

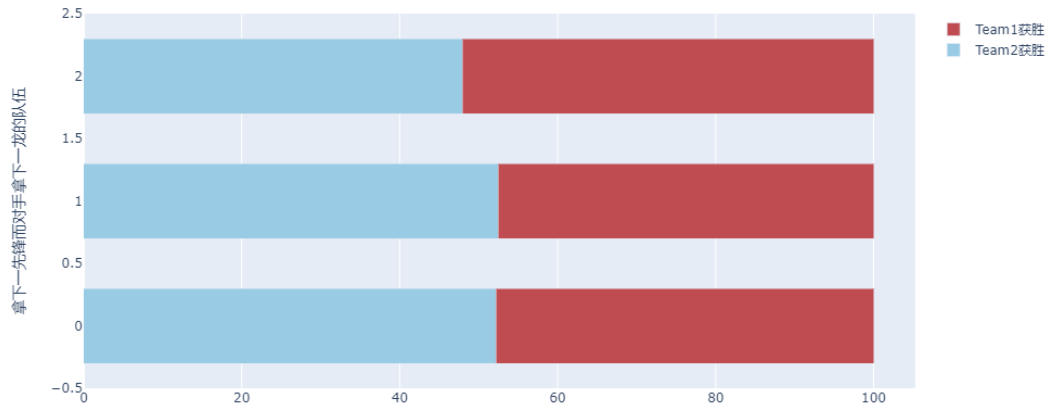
同时拿下一龙一先锋对获胜的影响



由对比图得知拿下一龙和一先锋的队伍胜率相对较高，在team1的比赛中，拿下一龙和一先锋时的胜率为71.42%。在team2的比赛中，拿下一龙和一先锋时的胜率为70.83%。**平均看来，拿下一龙和一先锋时的胜率为71.125%，拿下一龙和一先锋与获胜有较高的相关性。**

5.8 拿下一先锋而对手拿下一龙对获胜的影响

拿下一先锋而对手拿下一龙对获胜的影响

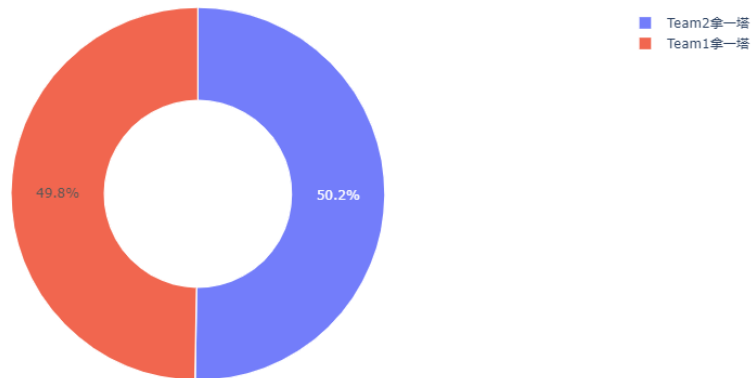


在team1的比赛中，拿下一先锋而对手拿下一龙时的胜率为47.54%。在team2的比赛中，拿下一先锋而对手拿下一龙时的胜率为47.93%。**平均看来，拿下一先锋而对手拿下一龙时的胜率为47.735%，拿下一先锋而对手拿下一龙与获胜没有什么相关性。反而胜率不高，说明一龙的重要性略高于一先锋。**

6. 一塔

6.1 一塔的分布

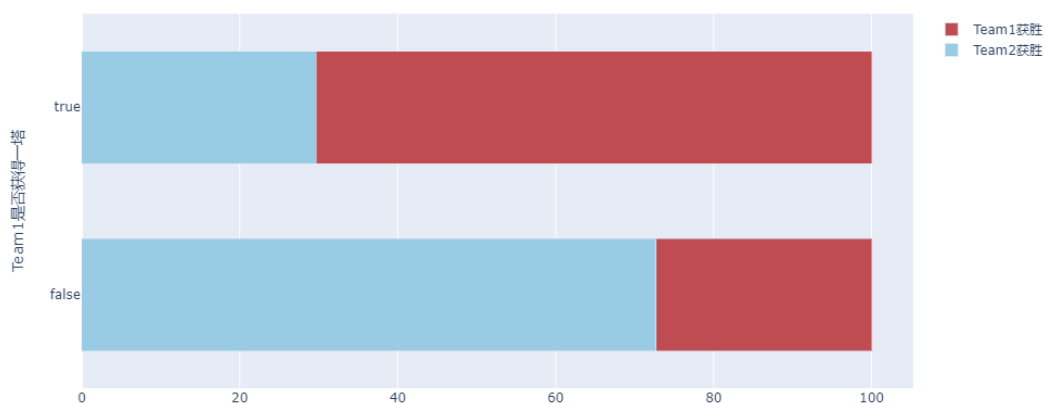
Team1/2拿一塔的分布



75984条记录中，其中team1拿一塔37804次，占比49.8%，team2拿一塔38180次，占比50.2%，是平衡的。

6.2 拿下一塔对获胜的影响

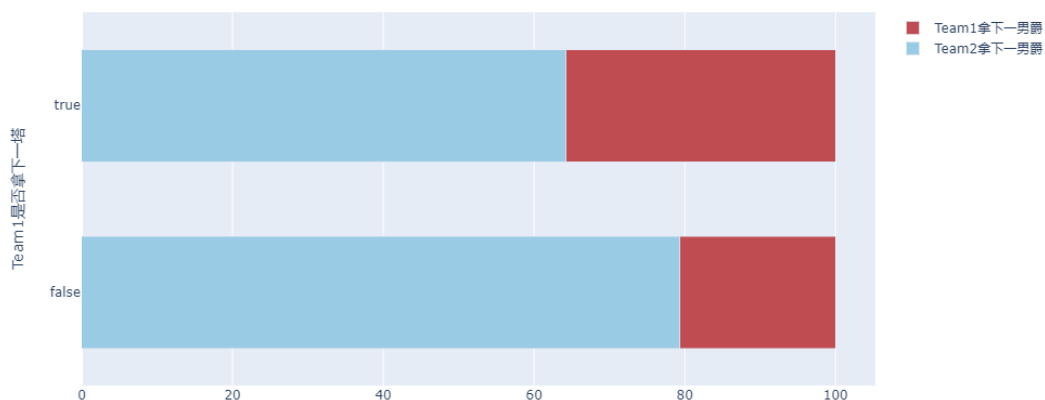
一塔对获胜的影响



由对比图得知拿下一塔的队伍胜率相对较高，在team1的比赛中，拿下一塔时的胜率为70.33%，相较未拿下一塔的比赛高43.06%。在team2的比赛中，拿下一塔时的胜率为72.73%。**平均看来，拿下一塔时的胜率为71.53%，拿下一塔与获胜有很高的相关性**

6.3 拿下一塔对拿下一男爵的影响

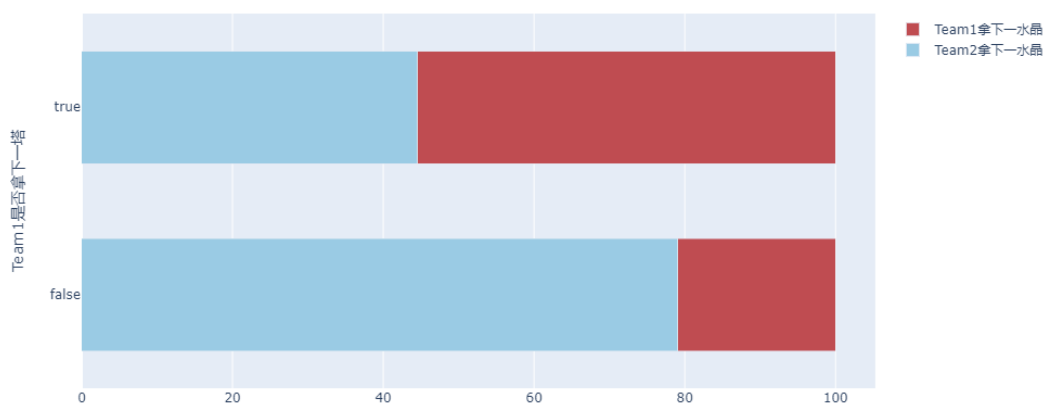
拿下一塔对拿下一男爵的影响



在team1的比赛中，拿下一塔后拿下一男爵的概率为35.71%，相较未拿下一塔的比赛高15.08%。在team2的比赛中，拿下一塔后拿下一男爵的概率为79.37%。**平均看来，拿下一塔时后续拿下一男爵的概率为57.54%，拿下一塔与拿下一男爵的相关性不是很高。**

6.4 拿下一塔对拿下一水晶的影响

拿下一塔对拿下一水晶的影响

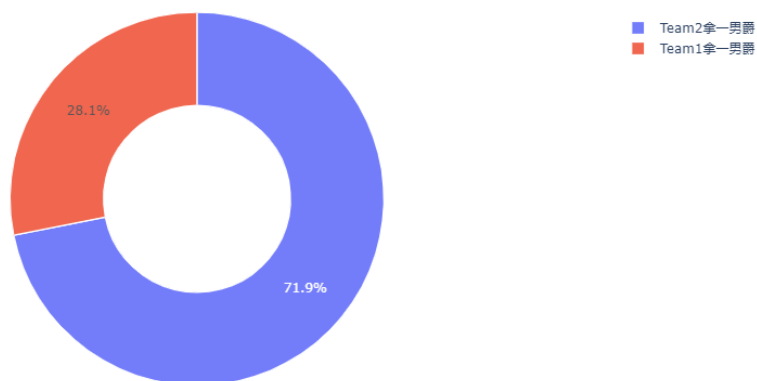


在team1的比赛中，拿下一塔后拿下一水晶的概率为55.46%，相较未拿下一塔的比赛高34.55%。在team2的比赛中，拿下一塔后拿下一水晶的概率为79.09%。
平均看来，拿下一塔时后续拿下一水晶的概率为67.275%，拿下一塔与拿下一水晶的相关性较高。

7. 一男爵

7.1 一男爵的分布

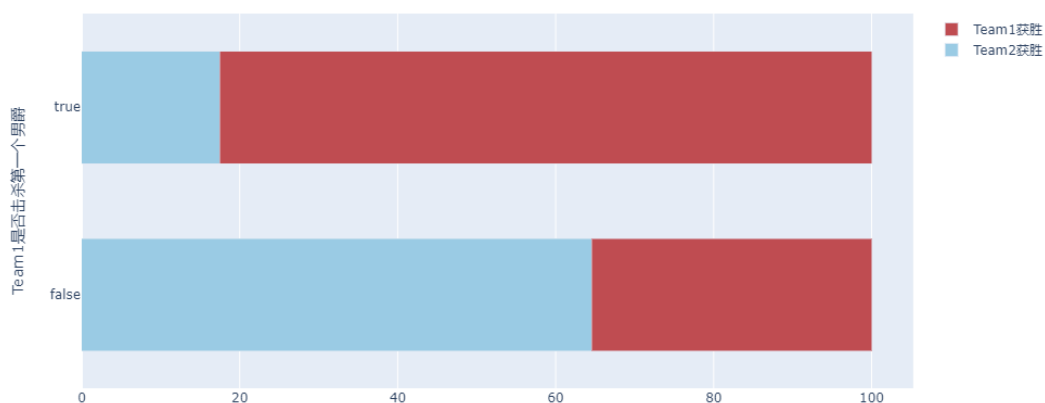
Team1/2拿一男爵的分布



75984条记录中，其中team1拿一男爵21374次，占比28.1%，team2拿一男爵54610次，占比71.9%，是不平衡的。所以后续我们基于一男爵的所有分析得出的结论，使用时要谨慎。

7.2 拿下一个男爵对获胜的影响

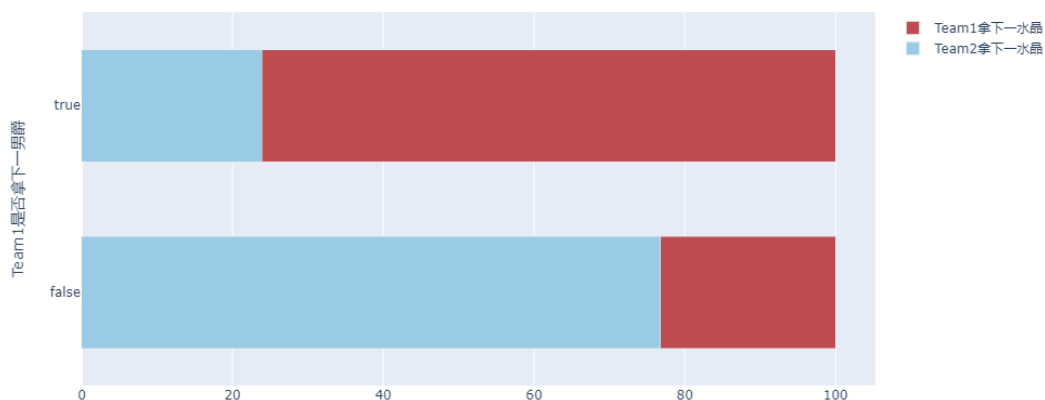
击杀第一个男爵对获胜的影响



由对比图得知拿下一男爵的队伍胜率相对较高，在team1的比赛中，拿下一男爵时的胜率为82.5%，相较未拿下一男爵的比赛高47.04%。在team2的比赛中，拿下一男爵时的胜率为64.54%。**平均看来，拿下一男爵时的胜率为72.52%，拿下一男爵与获胜有很高的相关性。**

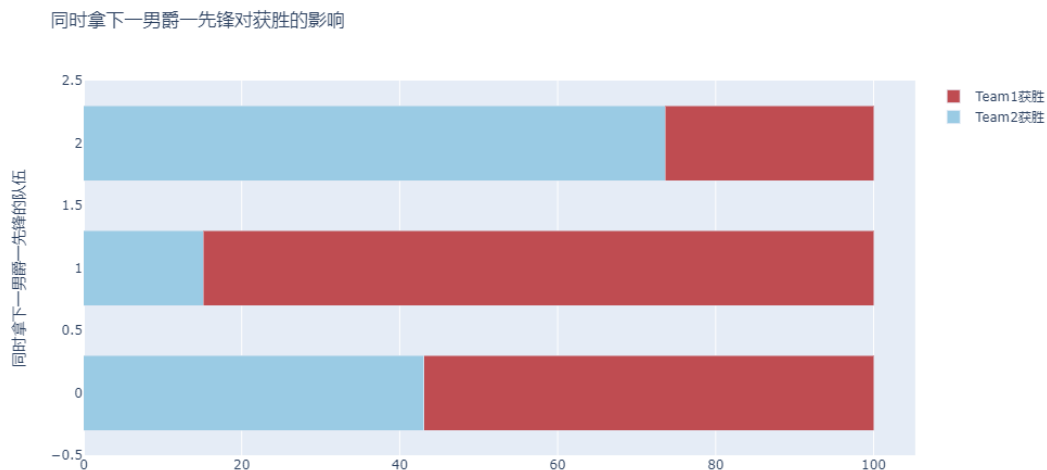
7.3 拿下一个男爵对拿下一水晶的影响

拿下一男爵对拿下一水晶的影响



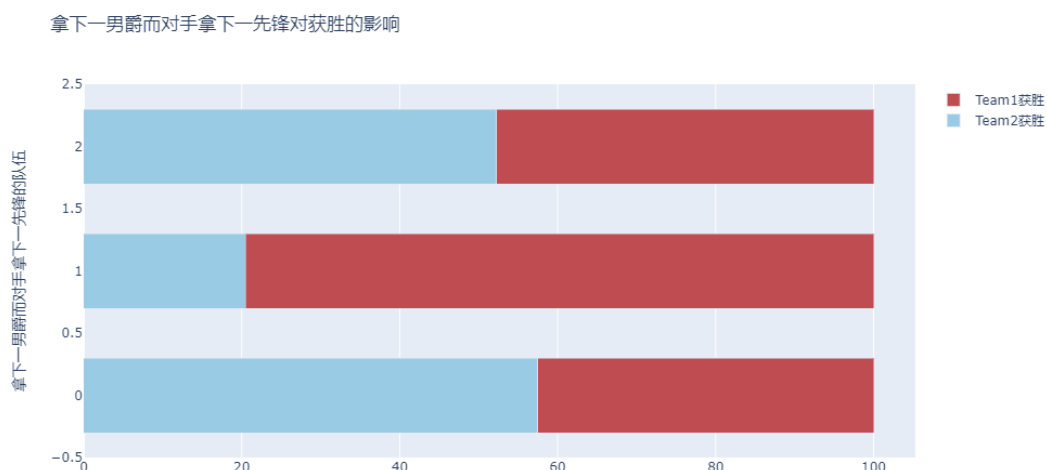
在team1的比赛中，拿下一男爵后拿下一水晶的概率为76.1%，相较未拿下一男爵的比赛高52.87%。在team2的比赛中，拿下一男爵后拿下一水晶的概率为76.77%。**平均看来，拿下一男爵时后续拿下一水晶的概率为76.435%，拿下一男爵与拿下一水晶有很高的相关性。**

7.4 同时拿下一男爵一先锋对获胜的影响



由对比图得知拿下一男爵和一先锋的队伍胜率相对较高，在team1的比赛中，拿下一男爵和一先锋时的胜率为84.87%。在team2的比赛中，拿下一男爵和一先锋时的胜率为73.58%。**平均看来，拿下一男爵和一先锋时的胜率为79.225%，拿下一男爵和一先锋与获胜有非常高的相关性。**

7.5 拿下一男爵而对手拿下一先锋

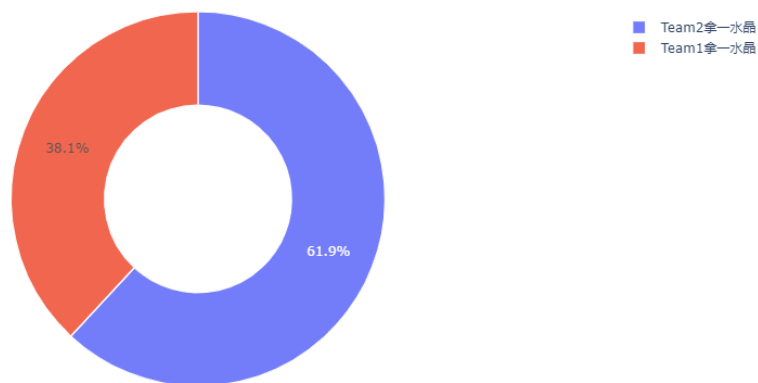


在team1的比赛中，拿下一男爵而对手拿下一先锋时的胜率为79.5%。在team2的比赛中，拿下一男爵而对手拿下一先锋时的胜率为52.25%。**平均看来，拿下一男爵而对手拿下一先锋时的胜率为65.875%，拿下一男爵而对手拿下一先锋与获胜有较高的相关性。**而且胜率较高，说明一男爵的重要性要高于一先锋。

8. 一水晶

8.1 一水晶的分布

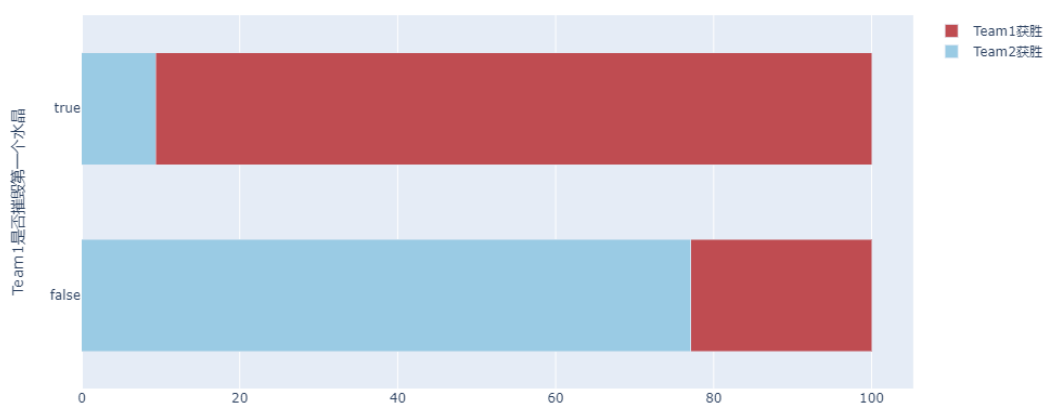
Team1/2拿一水晶的分布



75984条记录中，其中team1拿一水晶28949次，占比38.1%，team2拿一水晶47035次，占比61.9%，是不平衡的。所以后续我们基于一水晶的所有分析得出的结论，使用时要谨慎。

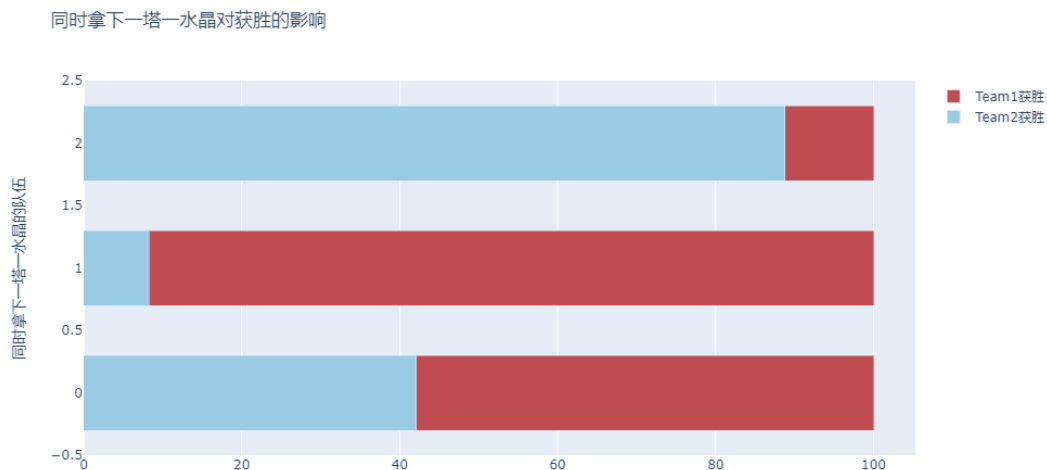
8.2 拿下一个水晶对获胜的影响

拿下一个水晶对获胜的影响



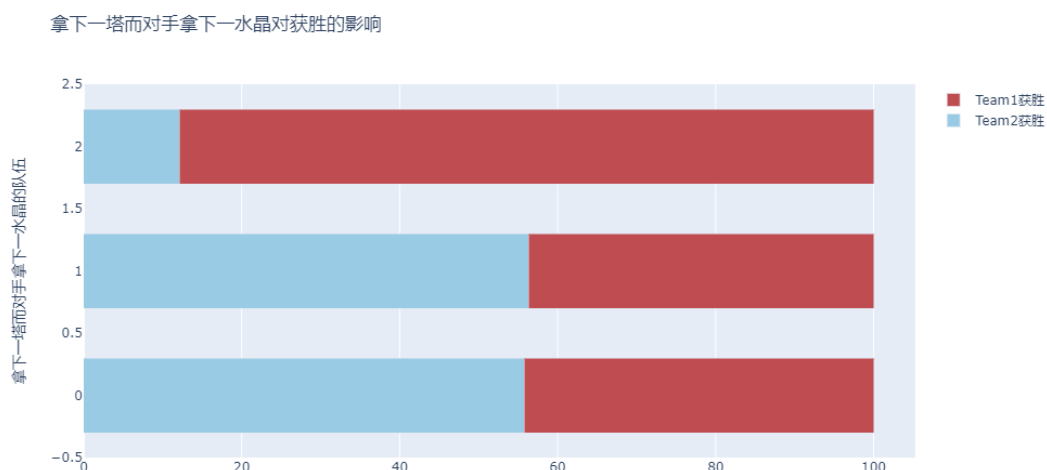
由对比图得知拿下一个水晶的队伍胜率相对较高，在team1的比赛中，拿下一个水晶时的胜率为90.66%，相较未拿下一个水晶的比赛高67.8%。在team2的比赛中，拿下一个水晶时的胜率为77.14%。**平均看来，拿下一个水晶时的胜率为83.9%，拿下一个水晶与获胜有非常高的相关性。**这一点在某种程度上是可以预见的，因为首先拿下一个水晶代表队伍已经积累的足够的优势。

8.3 同时拿下一塔一水晶对获胜的影响



由对比图得知拿下一塔和一水晶的队伍胜率相对较高，在team1的比赛中，拿下一塔和一水晶时的胜率为91.72%。在team2的比赛中，拿下一塔和一水晶时的胜率为88.75%。**平均看来，拿下一塔和一水晶时的胜率为90.235%，拿下一塔和一水晶与获胜有非常高的相关性。**

8.4 拿下一塔而对手拿下一水晶对获胜的影响

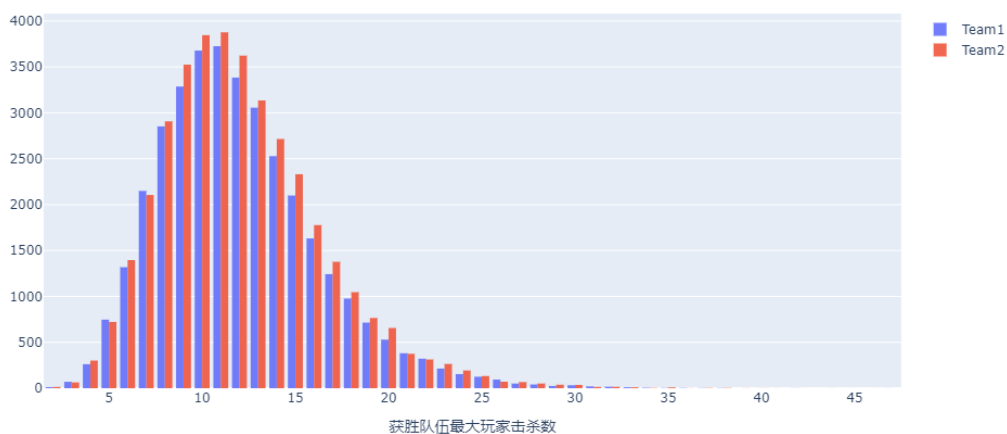


在team1的比赛中，拿下一塔而对手拿下一水晶时的胜率为43.69%。在team2的比赛中，拿下一塔而对手拿下一水晶时的胜率为12.14%。**平均看来，拿下一塔而对手拿下一水晶时的胜率为27.915%，拿下一塔而对手拿下一水晶与获胜没有什么相关性。**而且胜率很低，说明一水晶的重要性远高于一塔。

9. 最大击杀数

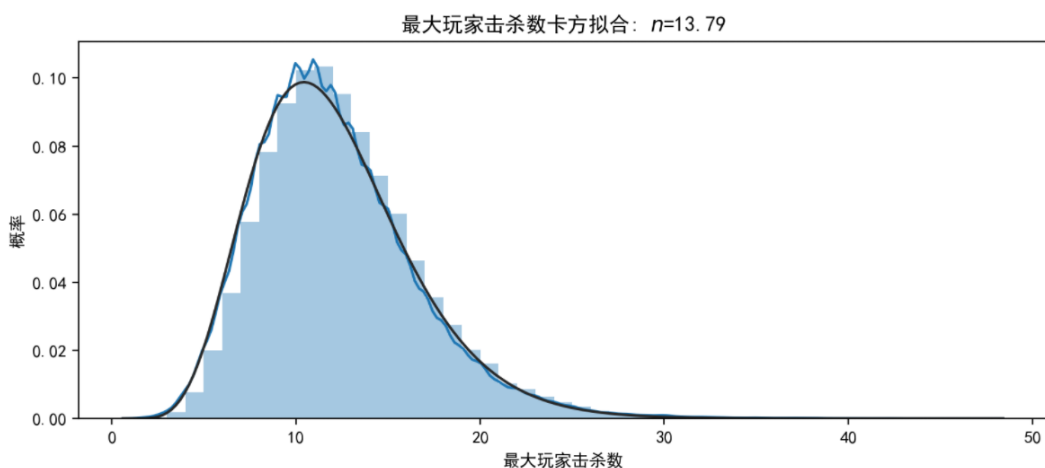
9.1 获胜队伍最大玩家击杀数分布

获胜队伍最大玩家击杀数分布

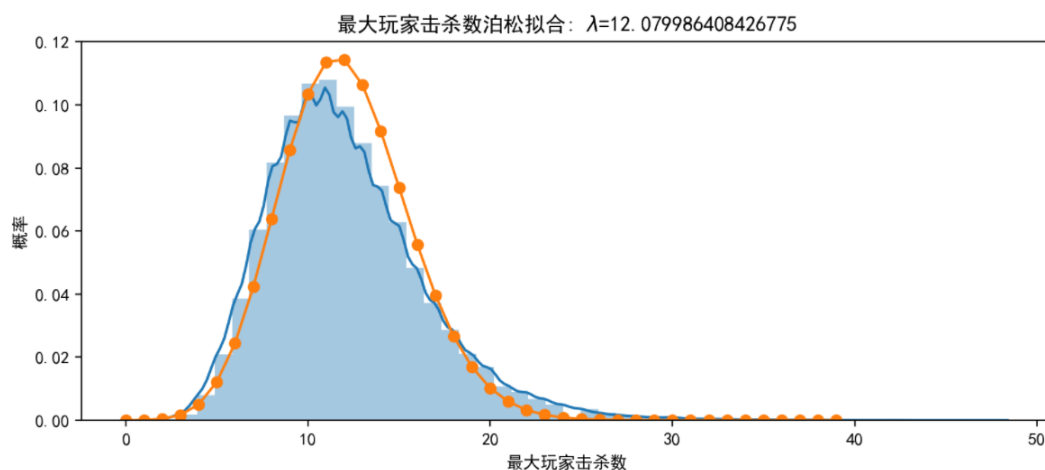


样本中存在获胜队伍最大击杀数为0，即全队击杀均为0的比赛，很可能属于开局掉线的比赛。所以在统计最大击杀规律之前要处理这些数据，这里选择去除这些异常比赛。**可以看出获胜队伍的最大击杀数大部分集中在区间[8,13]，且均值为12.08。**初步判断分布相似于卡方分布或者泊松分布。

9.2 最大玩家击杀数卡方拟合和泊松拟合



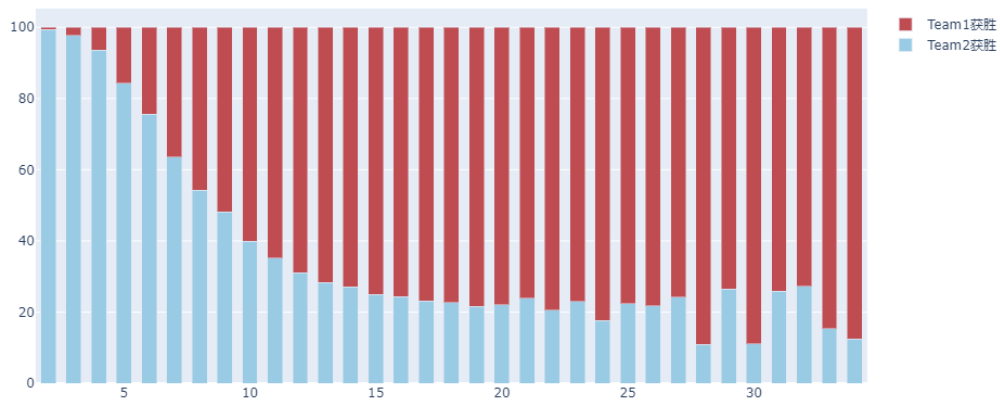
得到自由度约为14的卡方分布的拟合，但是注意图像的左半边，拟合的效果并不好。



用均值12.08作为参数绘制泊松分布图像，明显效果不好。

9.3 最大击杀数对获胜影响

最大杀敌数对获胜影响

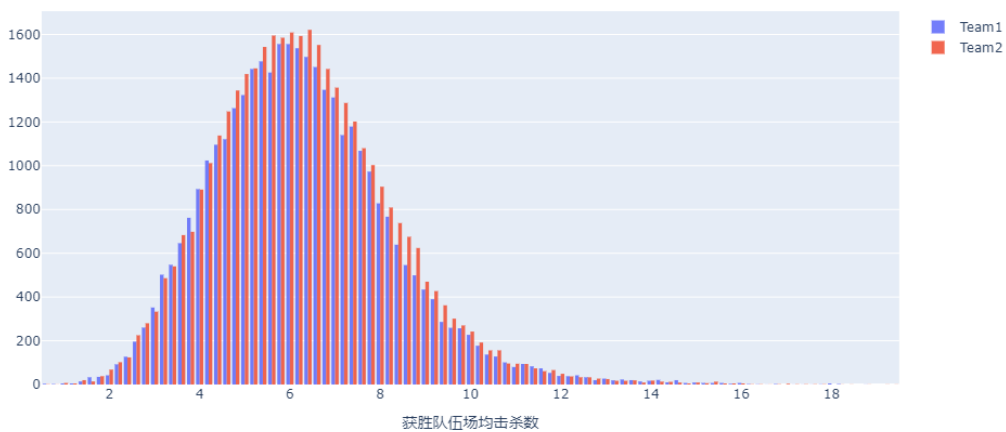


我们关注常见比赛的规律，于是做出限制，关注最大击杀数小于35的部分。可以发现开始时，最大击杀数越大，胜率越大。到最大击杀数为17时，胜率已为76.82%。而之后的胜率不再随着最大击杀数增加，反而有一定下降的趋势。我们从9.1的分布图中看出，在常规比赛中，最大击杀数大于20的场次很少。而在这种比赛中，很可能存在双方都有高端玩家的情况。

10. 场均击杀与击杀差

10.1 获胜队伍场均击杀数分布

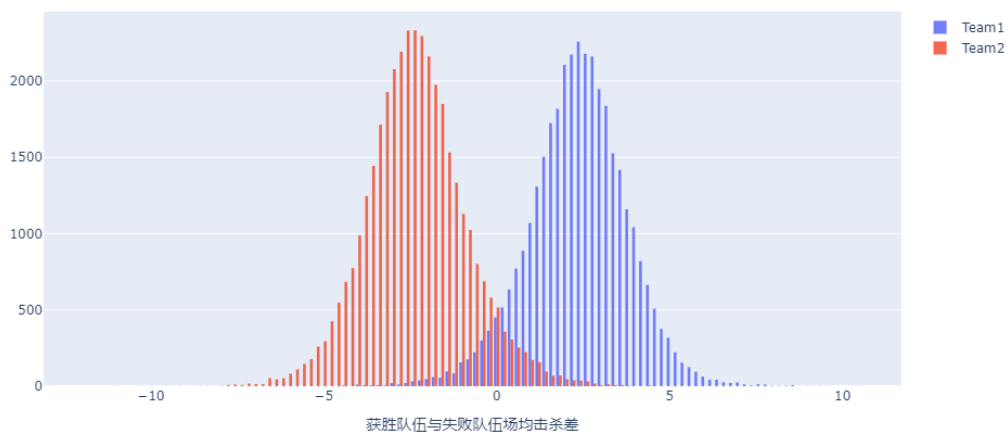
获胜队伍场均击杀数分布



统计场均击杀时，在去除无效比赛的基础上，我们关注更常规的比赛，令获胜队伍的场均击杀大于0且小于20。发现其大部分集中在[4.4,7.4]的区间，且均值为6.27。初步判断分布相似于正态分布。

10.2 获胜队伍场均击杀差分布

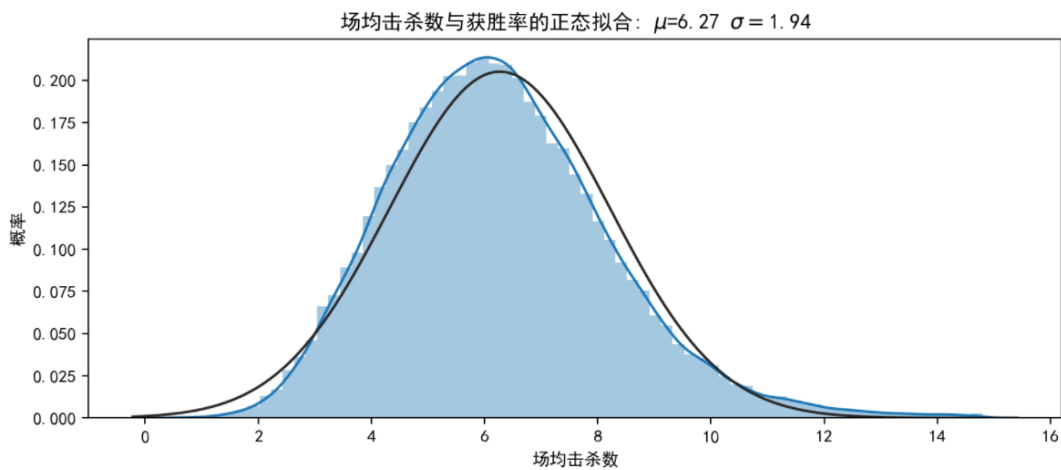
获胜队伍场均击杀差分布



有时候我们也关注两队的击杀差，两队的场均击杀差分布十分对称，**大多数相差的绝对值集中在[2.1,3.1]**。

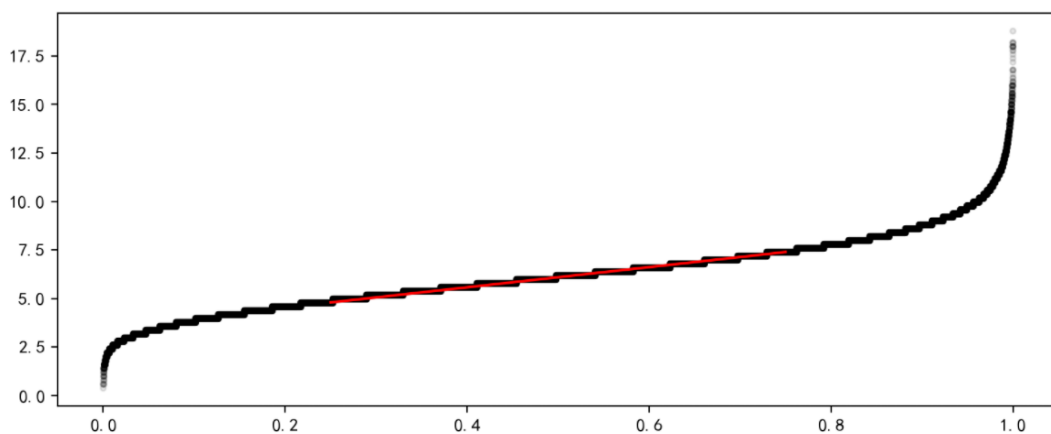
10.3 场均击杀数正态拟合

10.3.1 拟合图



得到 $\mu = 6.27$, $\sigma = 1.94$ 的正态分布(黑色曲线)，效果一般。

10.3.2 QQ图



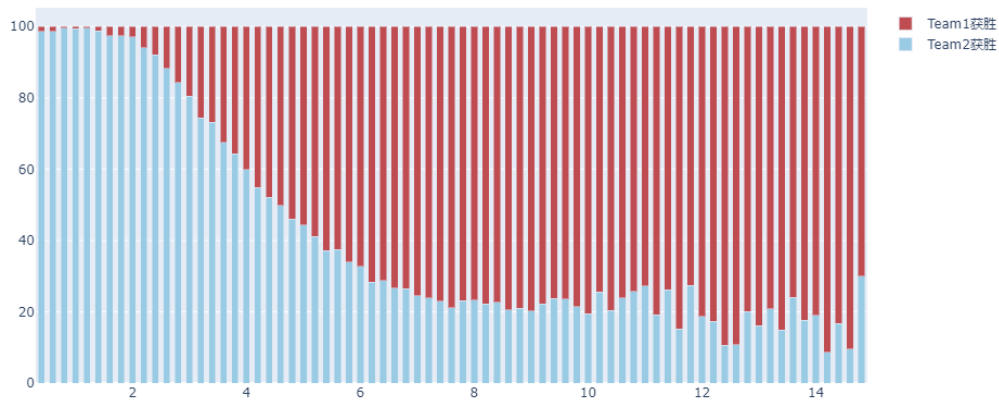
QQ图反映效果一般。

10.3.3 Kolmogorov-Smirnov 检验

进行K-S检验, $p - value < 0.05$, 我们拒绝假设, 因此**场均击杀数不符合正态分布**。

10.4 场均杀敌数对获胜影响

场均杀敌数对获胜影响

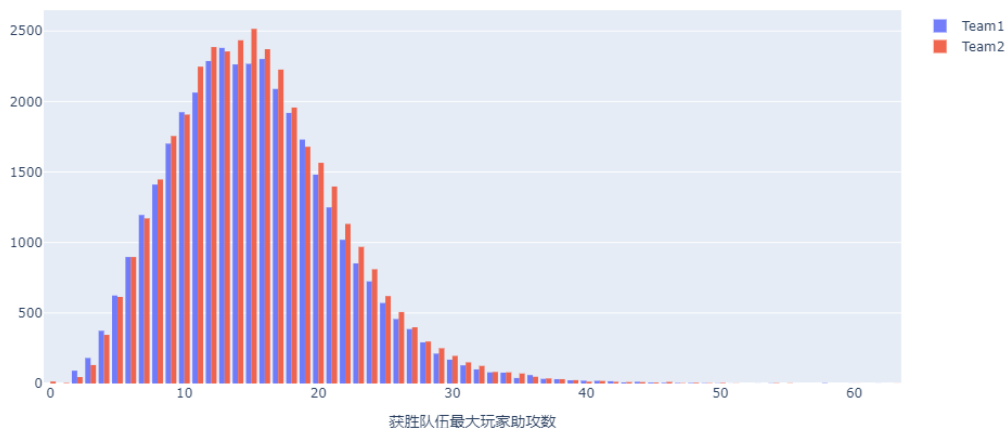


关注场均击杀数小于15的部分。可以发现开始时, 场均击杀数越大, 胜率越大。到场均击杀数为7.6时, 胜率已为78.94%。而之后的胜率不再随着场均击杀数增加, 反而有一定下降的趋势。

11. 最大助攻数

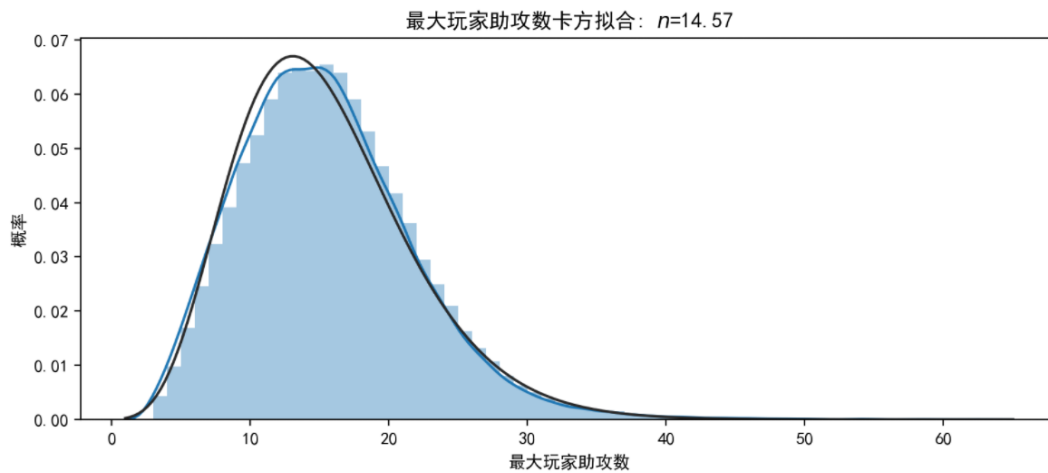
11.1 获胜队伍玩家最大助攻数分布

获胜队伍最大玩家助攻数分布

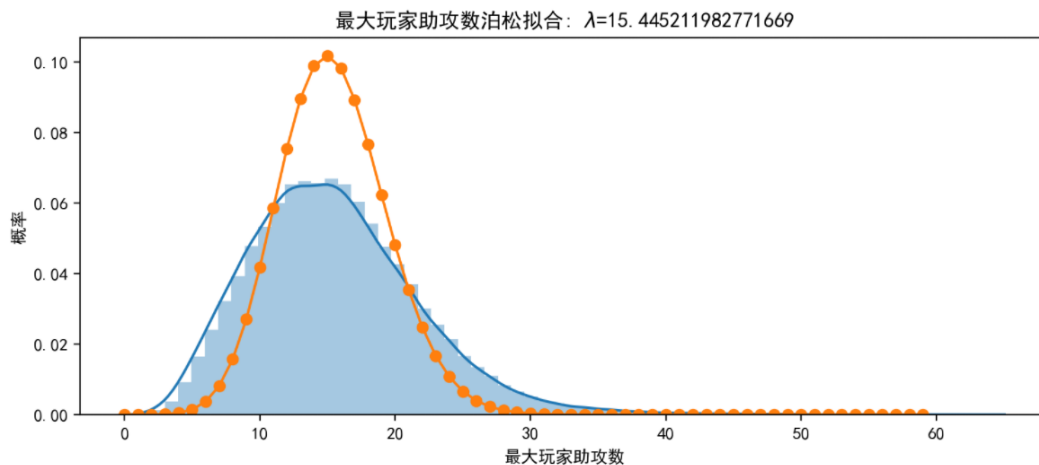


样本中存在获胜队伍最大助攻数为0, 即全队助攻均为0的比赛, 很可能属于开局掉线的比赛。所以在统计最大助攻规律之前要处理这些数据, 这里选择去除这些异常比赛。可以看出获胜队伍的最大助攻数大部分集中在区间[10,18], 且均值为15.45。初步判断分布相似于卡方分布或者泊松分布。

11.2 玩家最大助攻数卡方拟合和泊松拟合



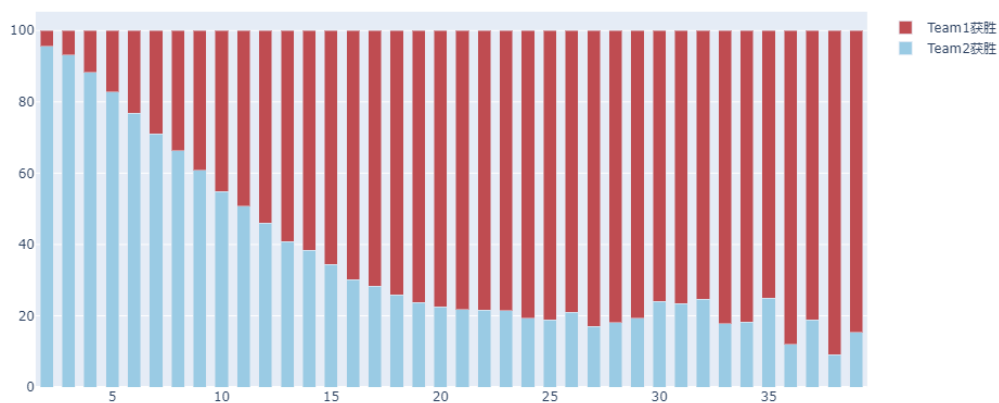
得到自由度约为15的卡方分布的拟合，但是注意图像的左半边，拟合的效果并不好。



用均值15.45作为参数绘制泊松分布图像，明显效果不好。

11.3 最大助攻数对获胜影响

最大助攻数对获胜影响

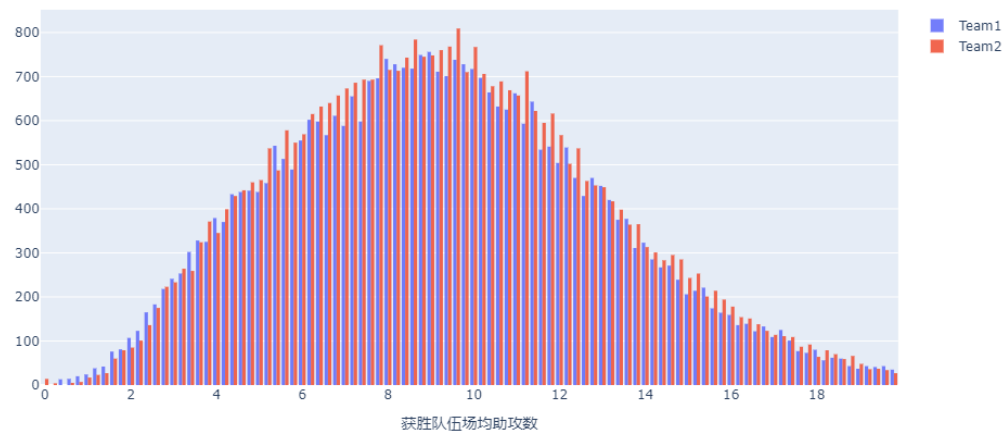


关注最大助攻数小于40的部分。可以发现开始时，最大助攻数越大，胜率越大。到最大助攻数为27时，胜率已为83.01%。而之后的胜率不再随着最大助攻数增加，反而有一定下降的趋势。我们从11.1的分布图中看出，在常规比赛中，最大助攻数大于30的场次很少。而在这种比赛中，很可能存在双方都有高端玩家的情况。

12. 场均助攻与助攻差

12.1 获胜队伍场均助攻数分布

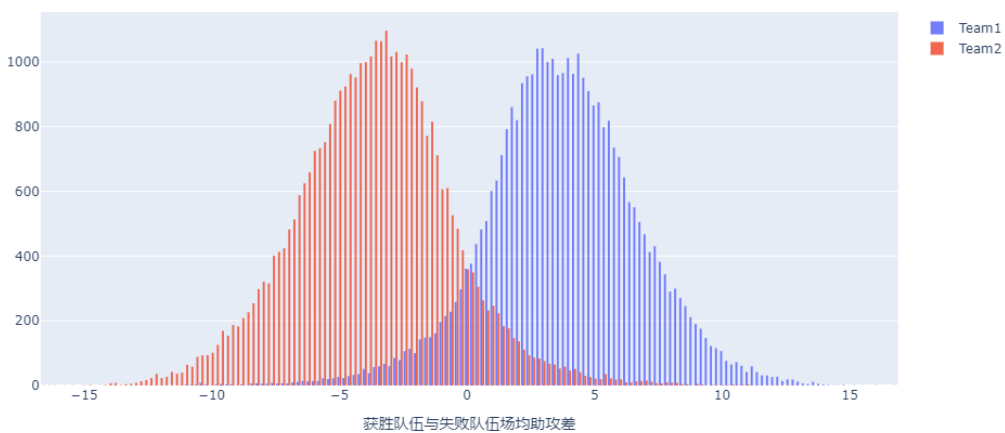
获胜队伍场均助攻数分布



统计场均助攻时，在去除无效比赛的基础上，我们关注更常规的比赛，令获胜队伍的场均助攻大于0且小于20。发现其大部分集中在[6.8,11.4]的区间，且均值为6.27。初步判断分布相似于正态分布或者卡方分布。

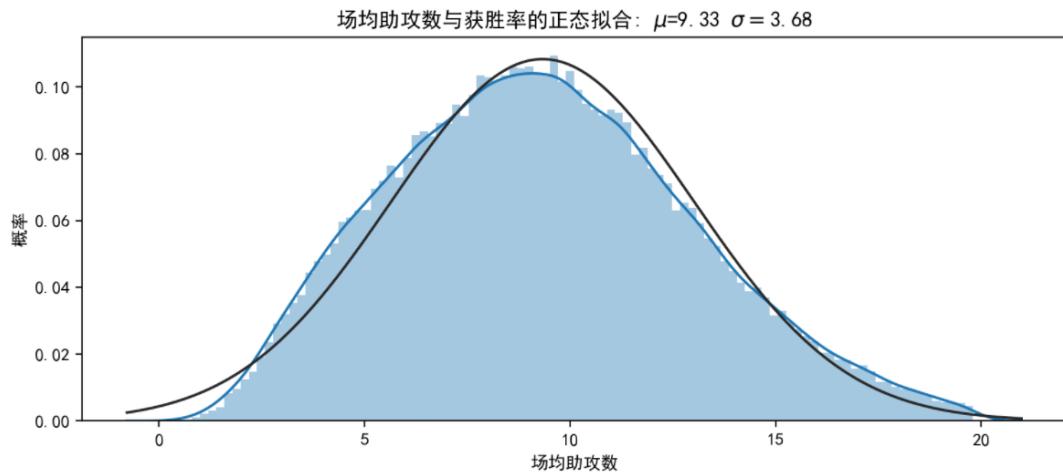
12.2 获胜队伍场均助攻差分布

获胜队伍场均助攻差分布

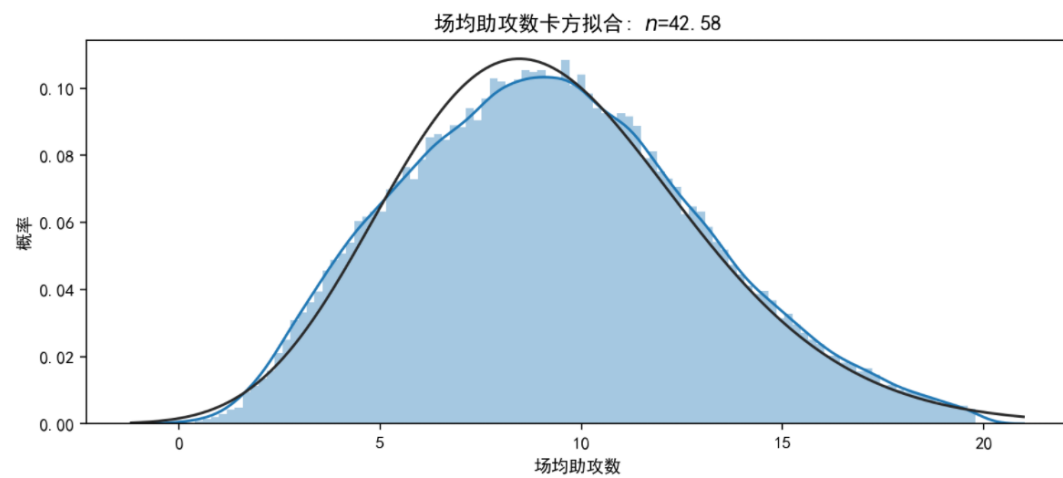


有时候我们也关注两队的助攻差，两队的场均助攻差分布十分对称，大多数相差的绝对值集中在[1.5,5.3]。

12.3 场均助攻数正态拟合和卡方拟合



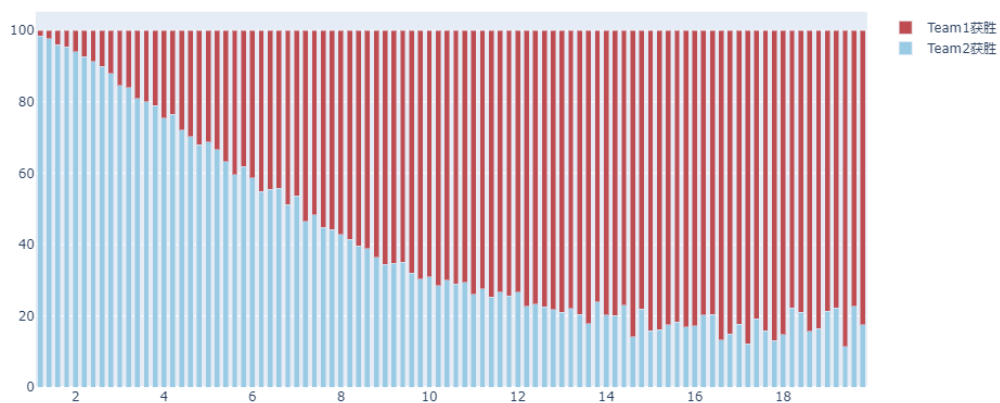
得到 $\mu = 9.33$, $\sigma = 3.68$ 的正态分布(黑色曲线), 效果一般。



得到自由度约为43的卡方分布的拟合, 但是注意图像的左上部分, 拟合的效果并不好。

12.4 场均助攻数对获胜影响

场均助攻数对获胜影响

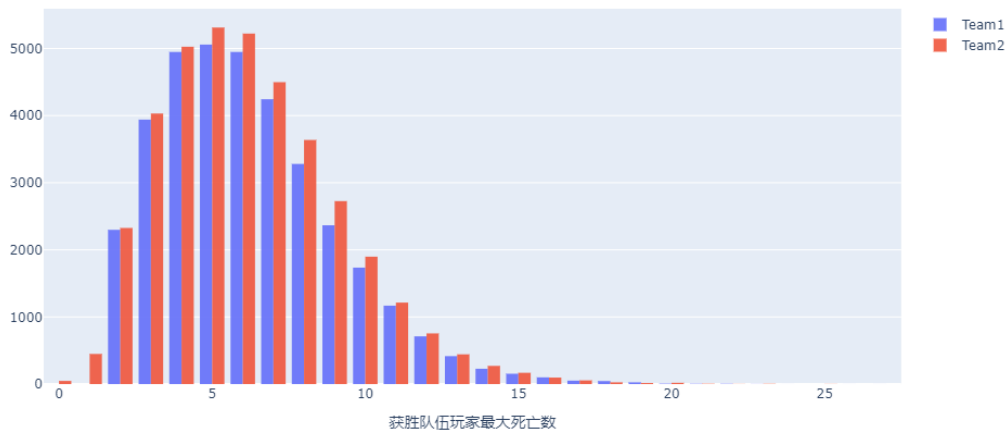


关注场均助攻数小于20的部分。可以发现开始时, 场均助攻数越大, 胜率越大。到场均助攻数为13.6时, 胜率已为82.28%。而之后的胜率不再随着场均助攻数增加, 反而有一定下降的趋势。

13. 最大死亡数

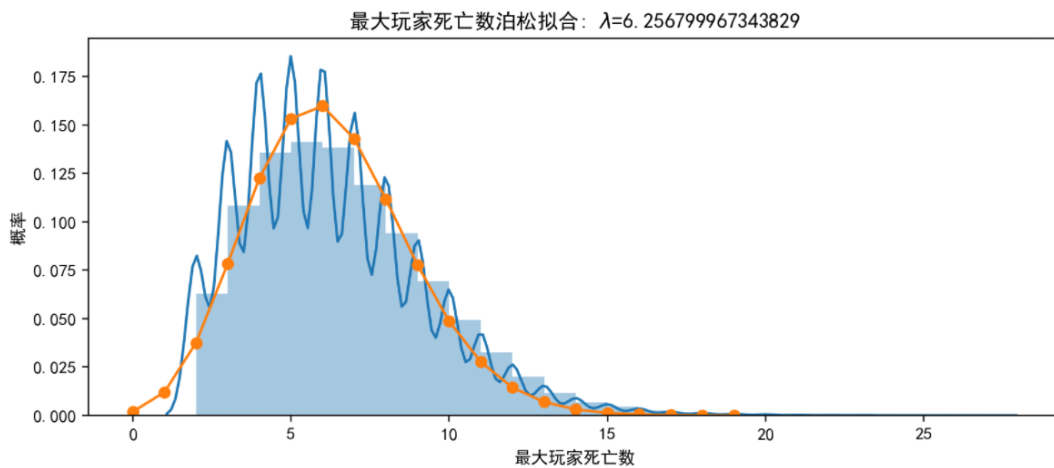
13.1 获胜队伍玩家最大死亡数分布

获胜队伍玩家最大死亡数分布



样本中存在获胜队伍最大死亡数为0，即全队死亡均为0的比赛，很可能属于开局掉线的比赛。所以在统计最大死亡规律之前要处理这些数据，这里选择去除这些异常比赛。可以看出获胜队伍的最大死亡数大部分集中在区间[3,7]，且均值为6.26。初步判断分布相似于泊松分布。

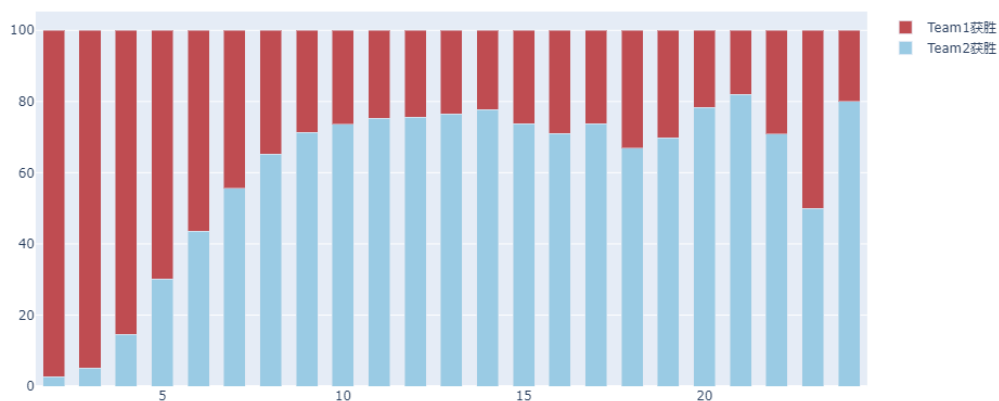
13.2 玩家最大死亡数泊松拟合



用均值6.26作为参数绘制泊松分布图像，明显效果不好。

13.3 最大死亡数对获胜影响

最大死亡数对获胜影响

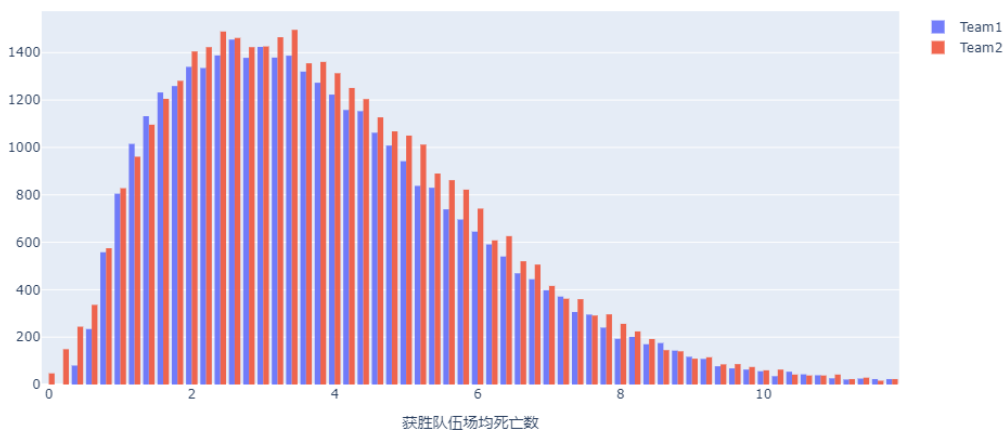


关注最大死亡数小于25的部分。可以发现开始时，最大死亡数越大，胜率越小。到最大死亡数为14时，胜率已为22.22%。而之后的胜率不再随着最大死亡数减小，反而有一定上升的趋势。我们从13.1的分布图中看出，在常规比赛中，最大死亡数大于15的场次很少。而在这种比赛中，很可能存在双方都有高端玩家的情况。

14. 场均死亡与死亡差

14.1 获胜队伍场均死亡数分布

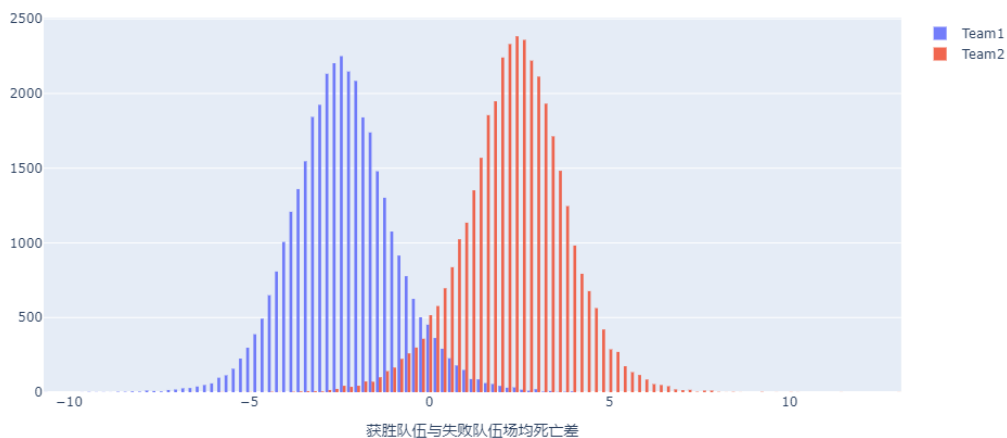
获胜队伍场均死亡数分布



统计场均死亡时，在去除无效比赛的基础上，我们关注更常规的比赛，令获胜队伍的场均死亡大于0且小于12。发现其大部分集中在[1.2,5.2]的区间，且均值为6.27。初步判断分布相似于卡方分布。

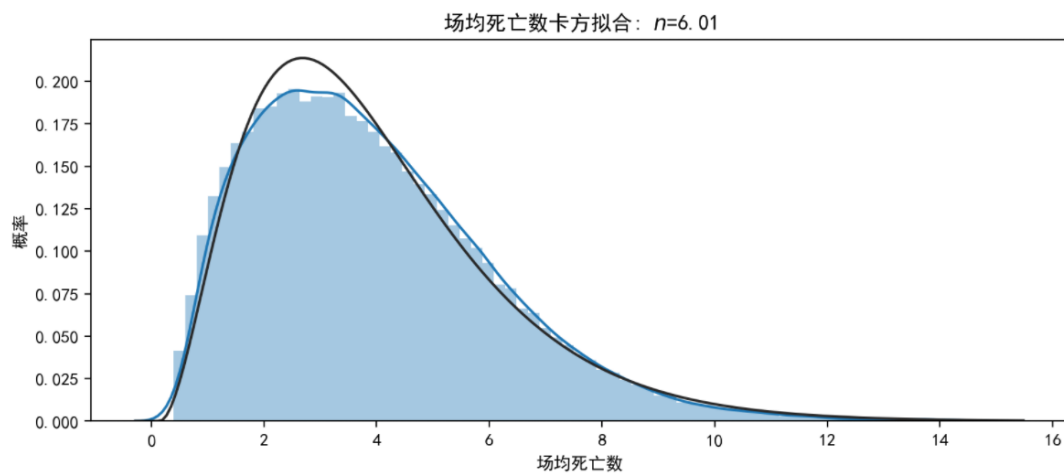
14.2 获胜队伍场均死亡差分布

获胜队伍场均死亡差分布



有时候我们也关注两队的死亡差，两队的场均死亡差分布十分对称，**大多数相差的绝对值集中在[1.7,3.3]**。

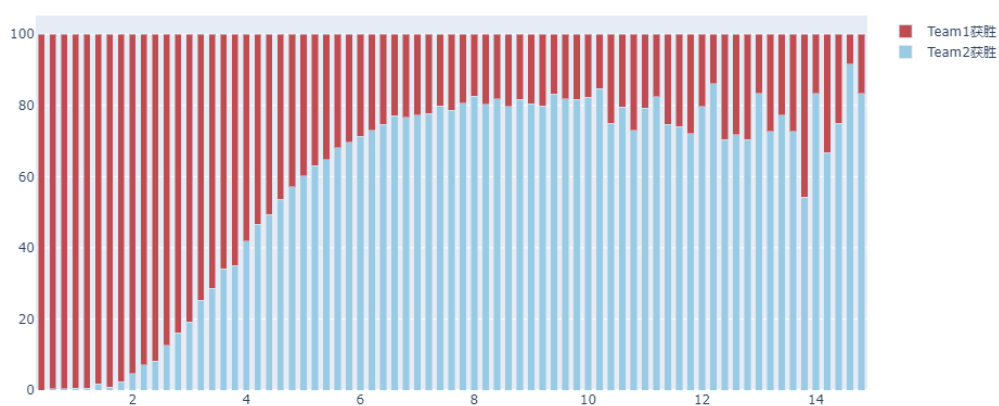
14.3 场均死亡数卡方拟合



得到自由度约为6的卡方分布的拟合，但是注意图像的最上部分，拟合的效果并不好

14.4 场均死亡数对获胜影响

场均死亡数对获胜影响

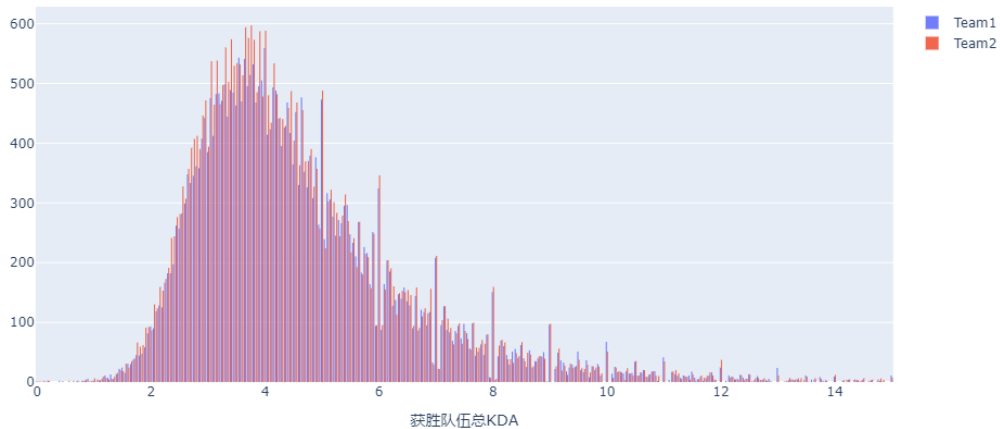


关注场均死亡数小于15的部分。可以发现开始时，场均死亡数越大，胜率越小。到场均死亡数为9时，胜率已为19.59%。而之后的胜率不再随着场均死亡数减小，反而有一定上升的趋势。

15. KDA

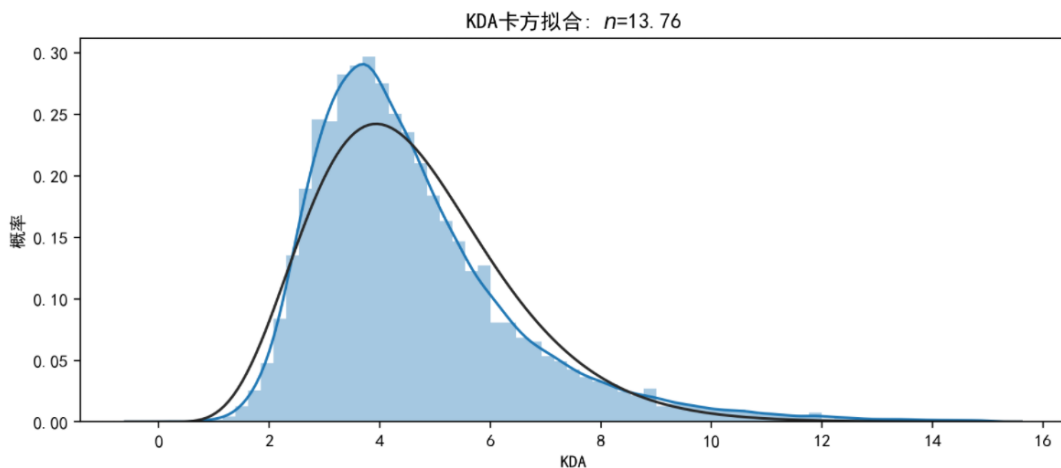
15.1 获胜队伍总KDA分布

获胜队伍总KDA分布

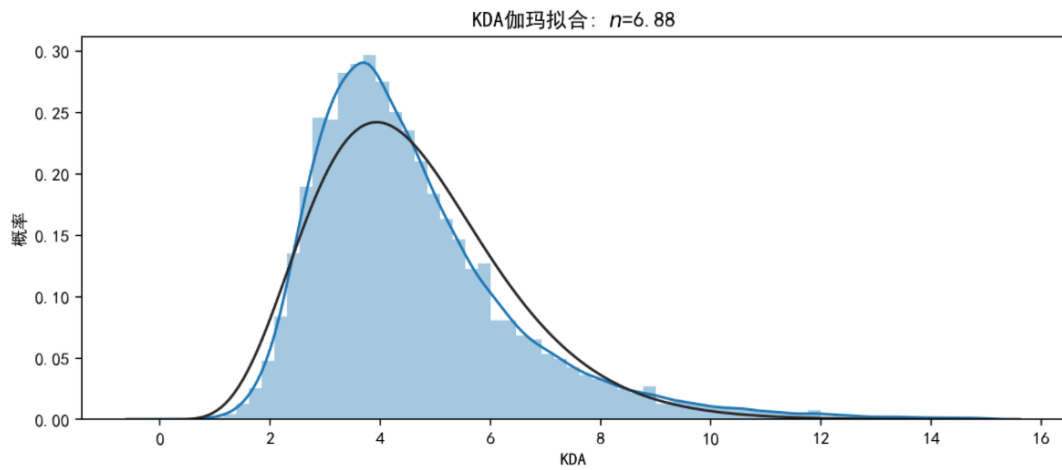


统计KDA时，我们要先处理掉无效局防止KDA计算出现错误值，要去除全队击杀为0/助攻为0/死亡为0的比赛。然后**发现其大部分集中在[2.475,5.525]的区间**。初步判断分布相似于卡方分布和伽玛分布。

15.2 KDA卡方分布和伽玛分布拟合



得到自由度约为14的卡方分布的拟合，但是注意图像的左上部分，拟合的效果并不好。

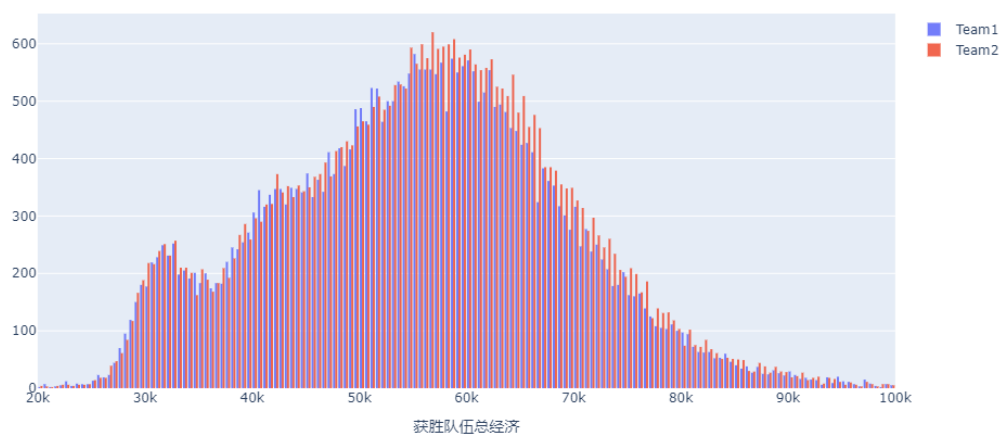


得到参数约为7的卡方分布的拟合，但是注意图像的左上部分，拟合的效果并不好。

16. 总经济和经济差

16.1 获胜队伍总经济分布

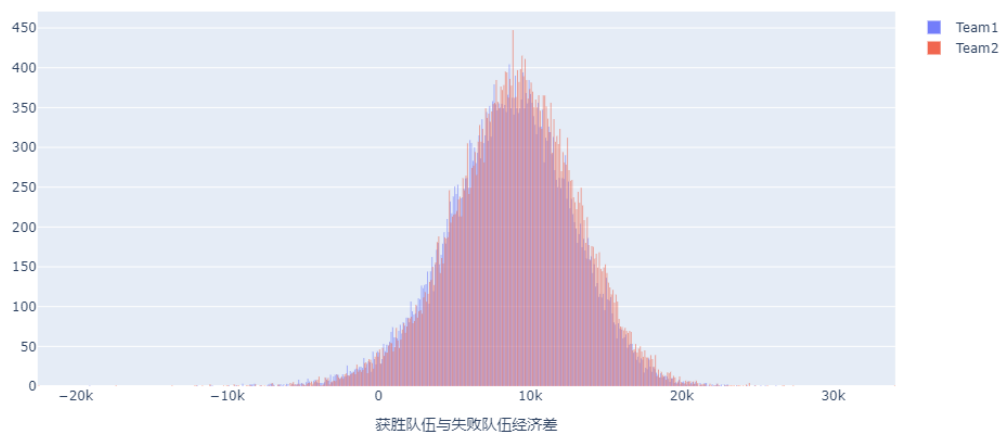
获胜队伍总经济分布



我们关注更常规的比赛，令获胜队伍的总经济小于100k且大于20k。发现其大部分集中在[40k,70k]的区间。

16.2 获胜队伍经济差分布

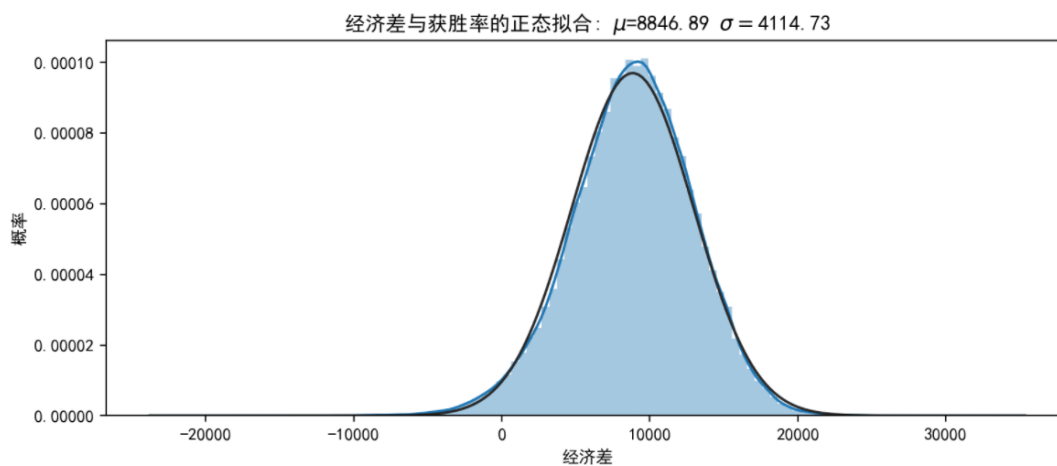
获胜队伍经济差分布



实际上我们常常关注的是两队的经济差，经济差更能反映两队的差距。初步判断分布相似于正态分布或者t分布。

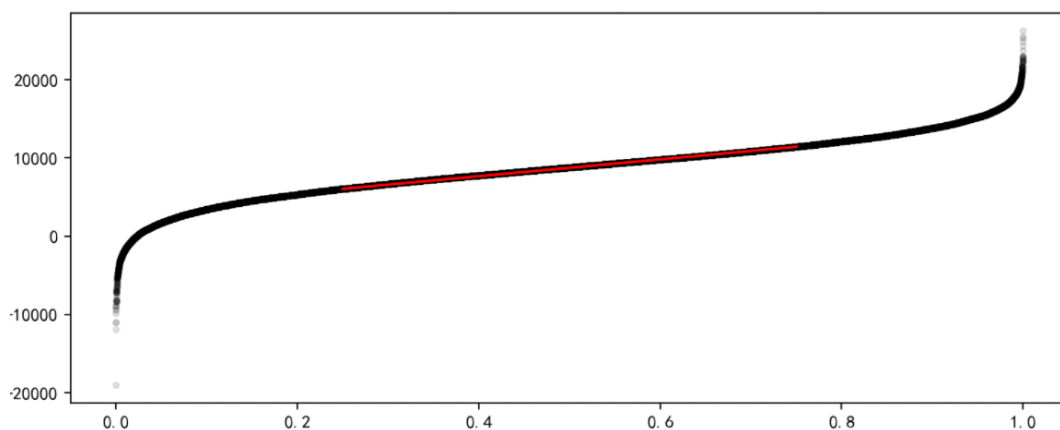
16.3 经济差正态拟合

14.3.1 拟合图



得到 $\mu = 8846.89$, $\sigma = 4114.73$ 的正态分布(黑色曲线), 效果不错。

16.3.2 QQ图



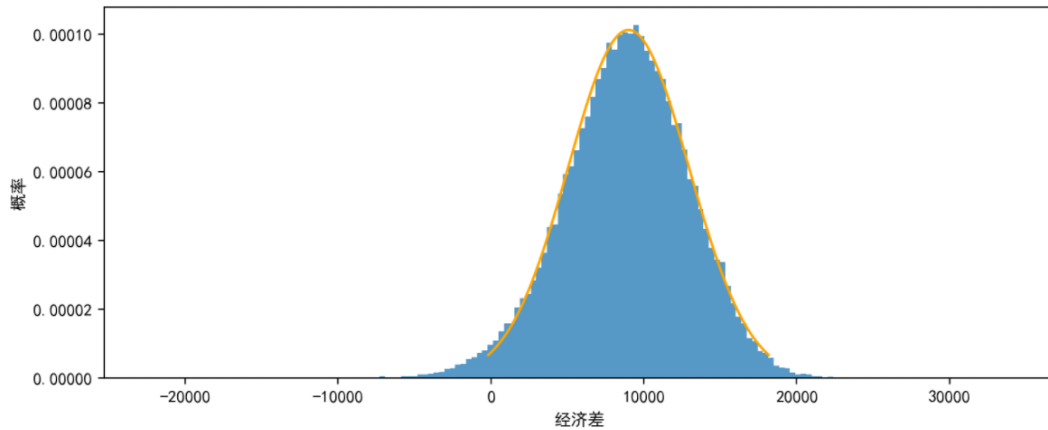
QQ图反映效果不错。

16.3.3 Kolmogorov–Smirnov 检验

进行K-S检验, $p - value = 0.075 > 0.05$, 我们拒结束假设, 因此**经济差符合正态分布**。

16.4 经济差T分布拟合

16.4.1 拟合图



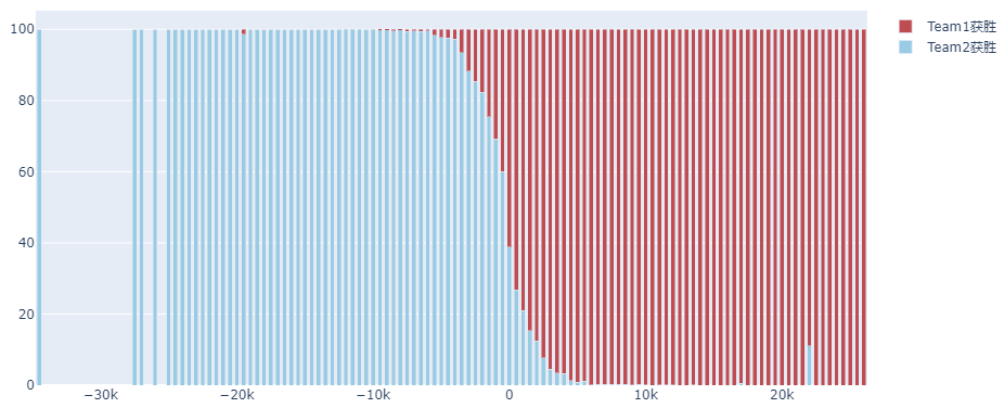
得到t分布, 效果不错。

16.4.2 Kolmogorov–Smirnov 检验

进行K-S检验, $p - value = 0.074 > 0.05$, 我们拒结束假设, 因此**经济差符合t分布**。

16.5 经济差对获胜影响

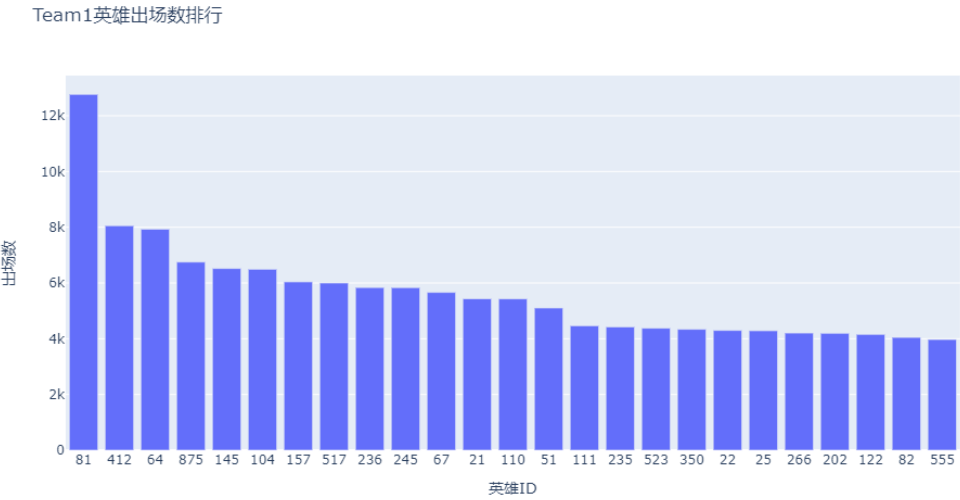
经济差对获胜影响



一般情况下两队的经济差不会太大, **集中在相差绝对值[0,5k]的范围内, 经济越高, 胜率越大**, 而**超过这个范围并不意味着有决定性的胜率**, 因为样本数据记录的是结束时的经济差。

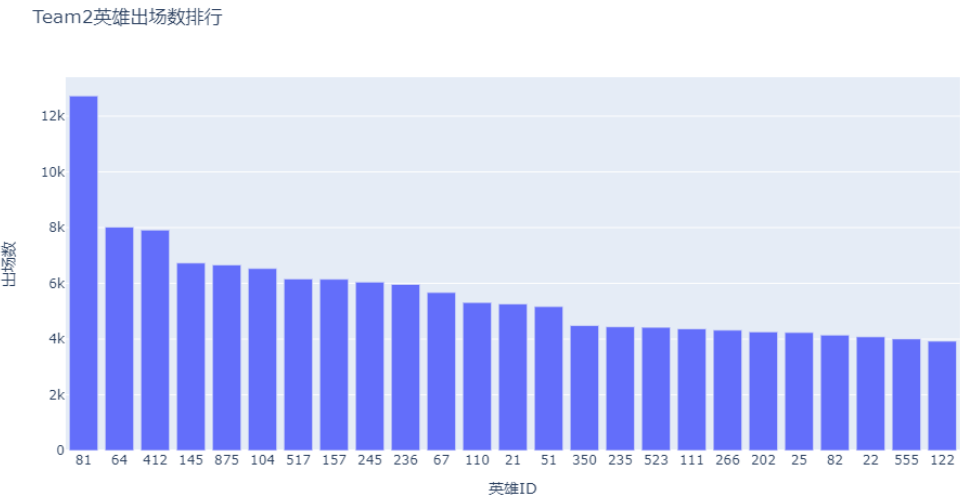
17. 总英雄出场率

17.1 Team1英雄出场数排行



team1比赛的英雄出场数排行前25。

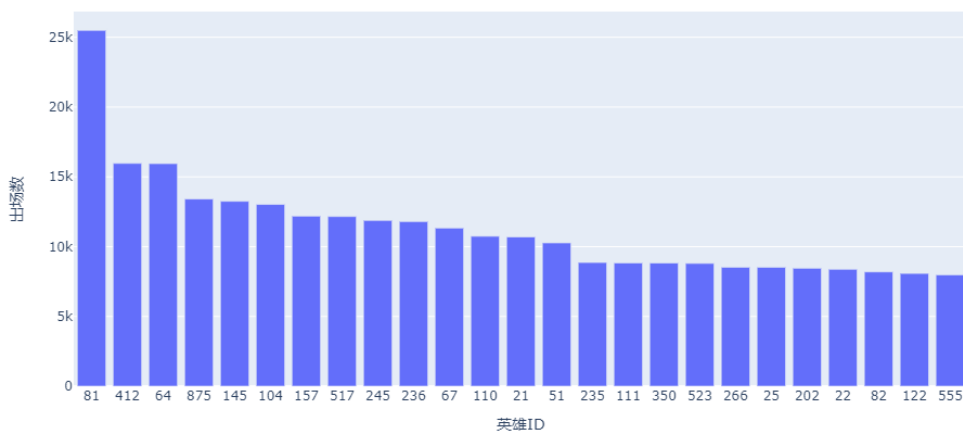
17.2 Team2英雄出场数排行



team2比赛的英雄出场数排行前25，与team1类似。

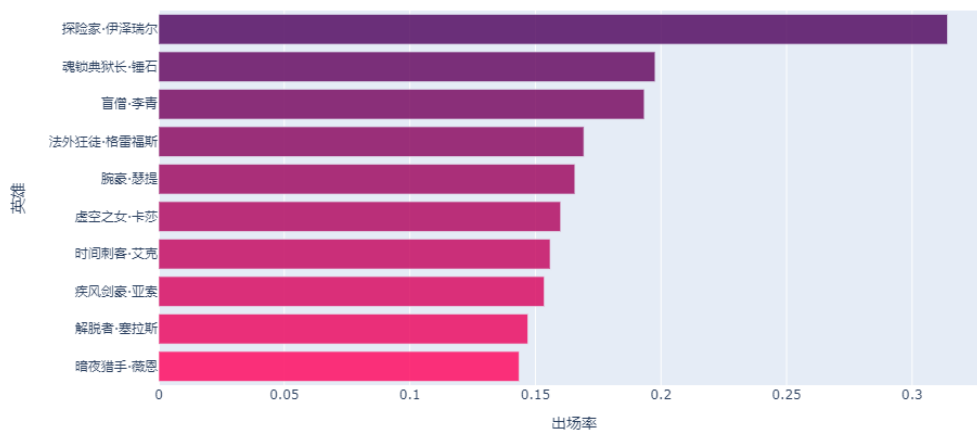
17.3 英雄总出场率排行前十

英雄总出场数排行



全部比赛中英雄的出场数排行前25。

胜利队伍英雄出场率排行前十

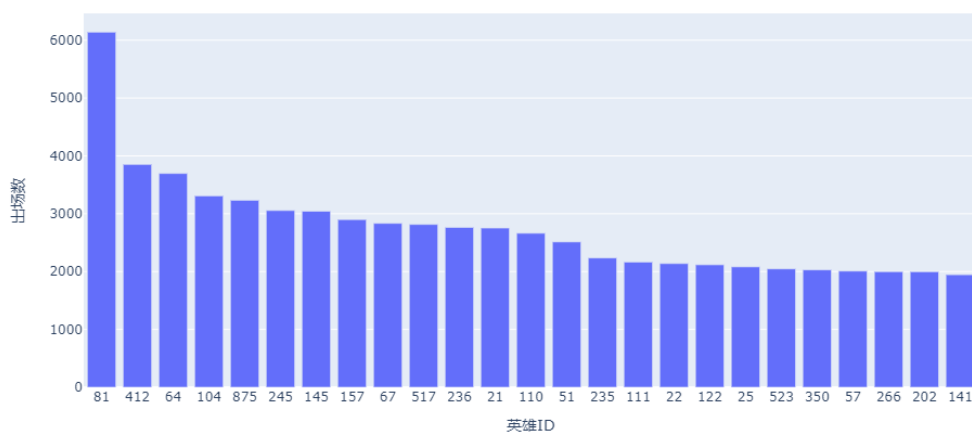


全部比赛中英雄的出场率排行前10。注意这里定义的**出场率**有所不同，在所有样本中，queued为700（冠军杯）、450（大乱斗）、1020（克隆）的模式不允许双方选择重复英雄，而在其他模式中，发现双方选择重复英雄的局数非常少，近似的认为在双方选择重复英雄的比赛中，该英雄出场算一次。

18. 胜利队伍英雄出场率

18.1 Team1获胜英雄出场数排行

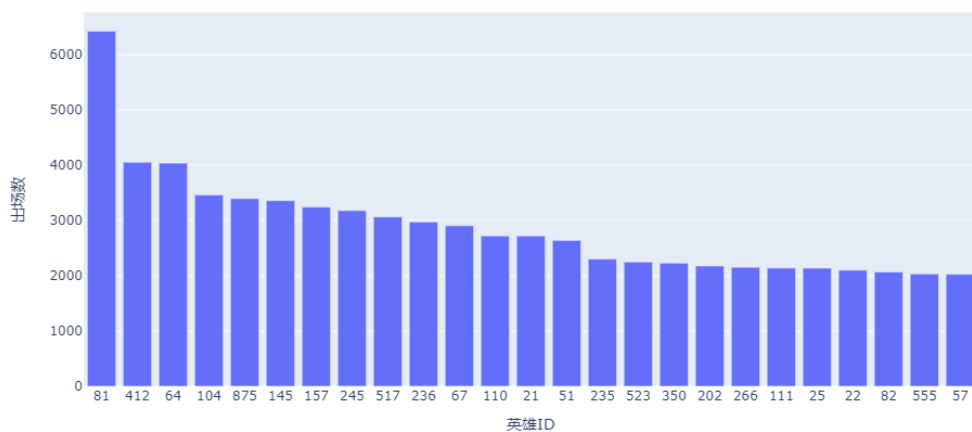
Team1获胜英雄出场数排行



team1获胜比赛的英雄出场数排行前25，这里已经与team1的全部比赛有所不同，说明还需要分析英雄的胜率。

18.2 Team2获胜英雄出场数排行

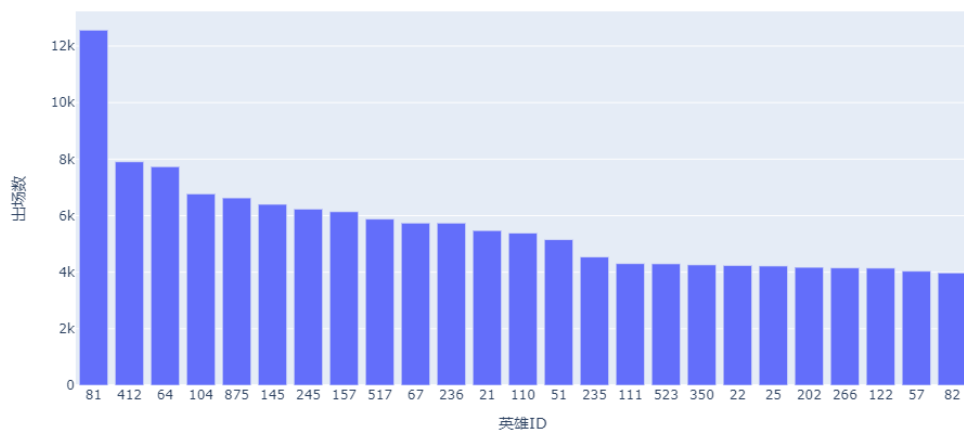
Team2获胜英雄出场数排行



team2获胜比赛的英雄出场数排行前25。

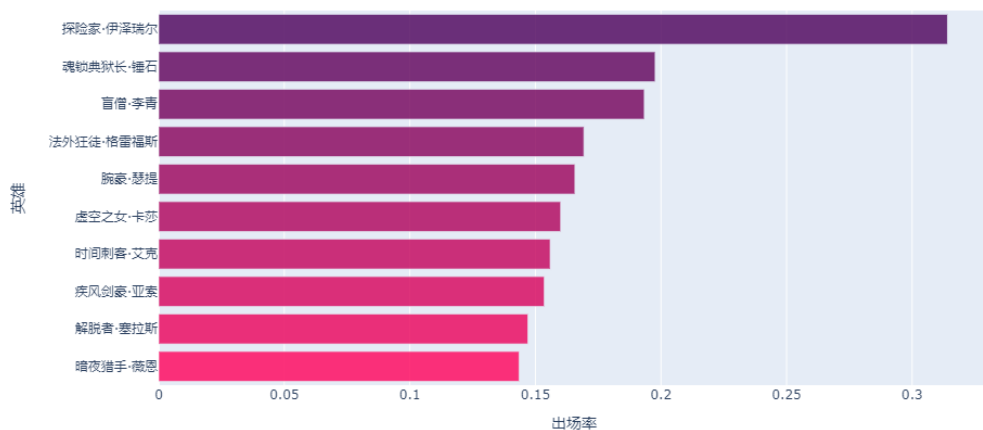
18.3 胜利队伍英雄总出场率排行前十

胜利队伍英雄出场数排行



全部获胜队伍的英雄出场数排行前25。

胜利队伍英雄出场率排行前十



全部获胜队伍比赛中英雄的出场率排行前10。

19. 英雄胜率排行

19.1 前十

英雄	胜率
琴瑟仙女·娑娜	0.5410462008400153
掘墓者·约里克	0.5322997416020672
星界游神·巴德	0.5294224482382855
铸星龙王·奥瑞利安·索尔	0.5286697247706422
无双剑姬·菲奥娜	0.5272644653135413
暴怒骑士·克烈	0.5266934557979334
炼金术士·辛吉德	0.5263852242744064
唤潮鲛姬·娜美	0.5260631001371742
时间刺客·艾克	0.5247874758017002
风暴之怒·迦娜	0.5247621371066113

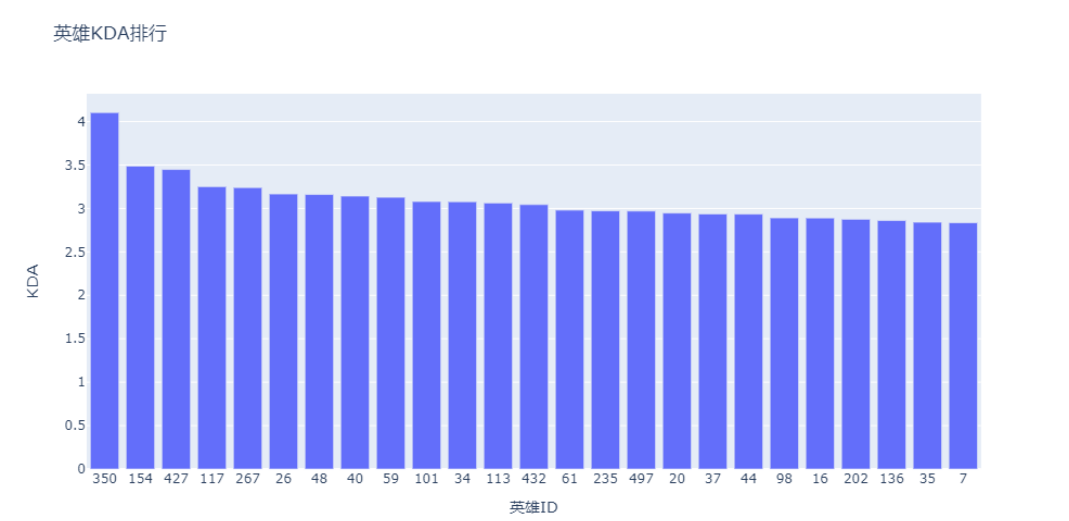
可以看到胜率第一的英雄为琴瑟仙女·娑娜，但她并不在出场率前十。除了第一名，其他英雄直接的胜率差距并不大。

19.2 后十

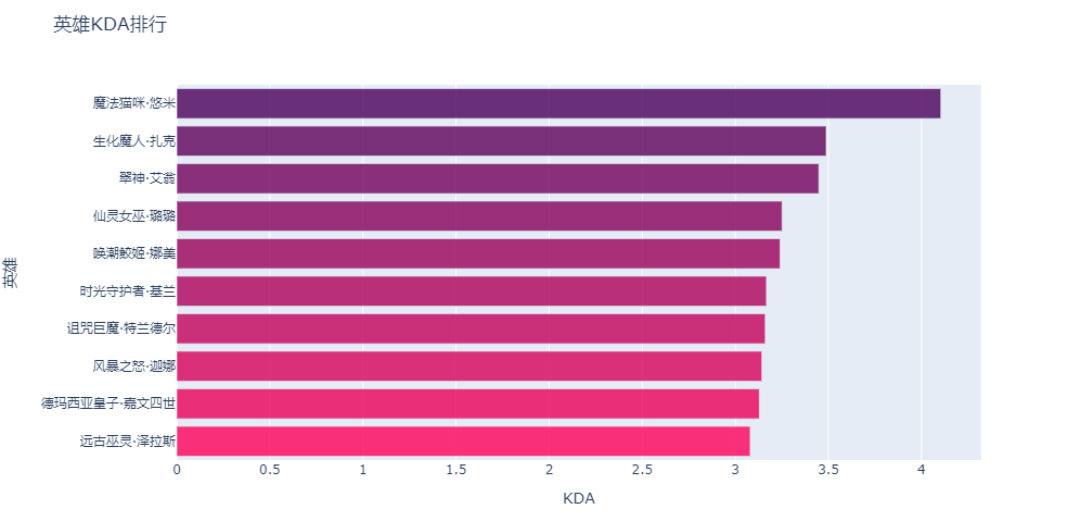
英雄	胜率
刀锋舞者·艾瑞莉娅	0.47215295095594345
弗雷尔卓德之心·布隆	0.47143427886536154
麦林炮手·崔丝塔娜	0.4711038071761161
元素女皇·琪亚娜	0.47016599371915657
诡术妖姬·乐芙兰	0.46993143976493634
北地之怒·瑟庄妮	0.4671968190854871
沙漠皇帝·阿兹尔	0.4647887323943662
英勇投弹手·库奇	0.458197104156936
狂暴之心·凯南	0.45287958115183247
离群之刺·阿卡丽	0.45163073521282476

胜率后十位的英雄胜率均低于50%，但也都不在出场率前十。

20. 英雄KDA排行



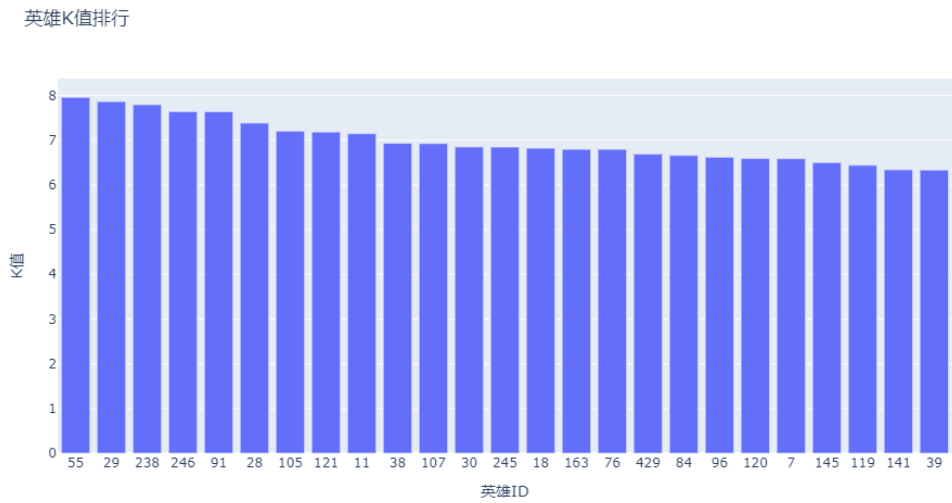
全部获胜队伍的KDA排行前25。



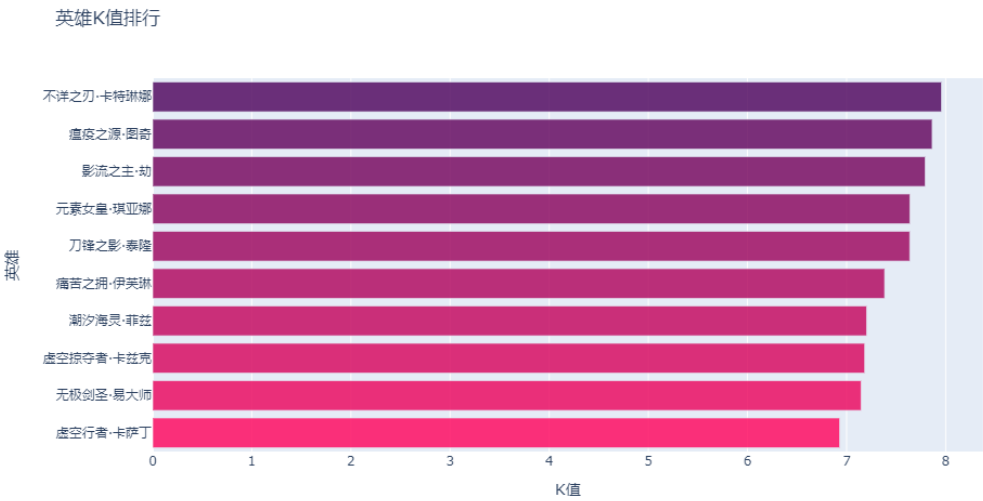
全部获胜队伍的KDA排行前10。这与胜利队伍英雄总出场率排行前十没有重合。而唤潮鲛姬·娜美和风暴之怒·迦娜位于英雄胜率排行前十中。

21. 英雄K/D/A排行

21.1 K值



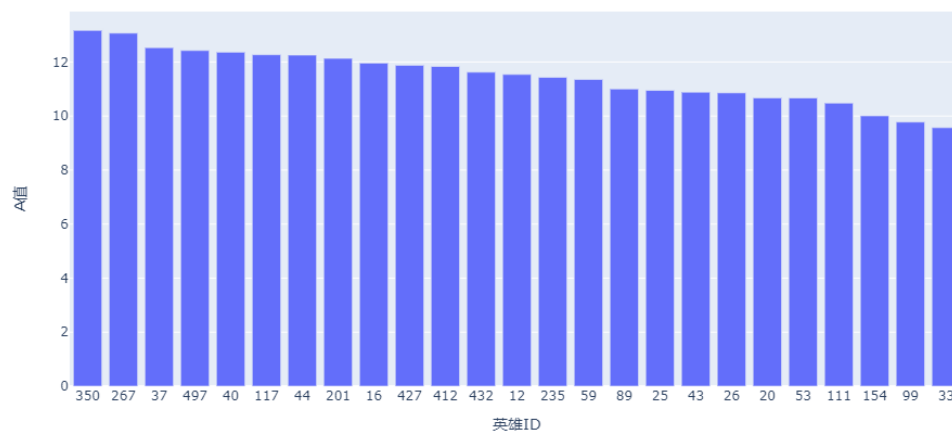
全部获胜队伍的K值排行前25。



全部获胜队伍的K值排行前10。

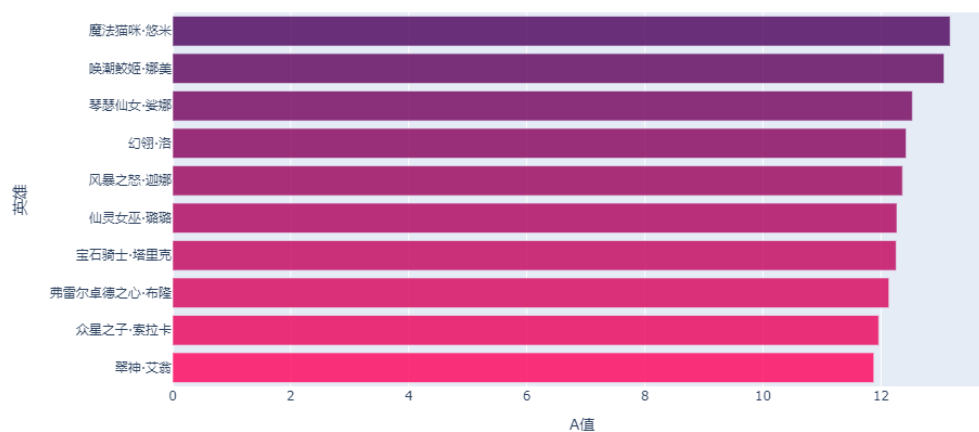
21.2 A值

英雄A值排行



全部获胜队伍的A值排行前25。

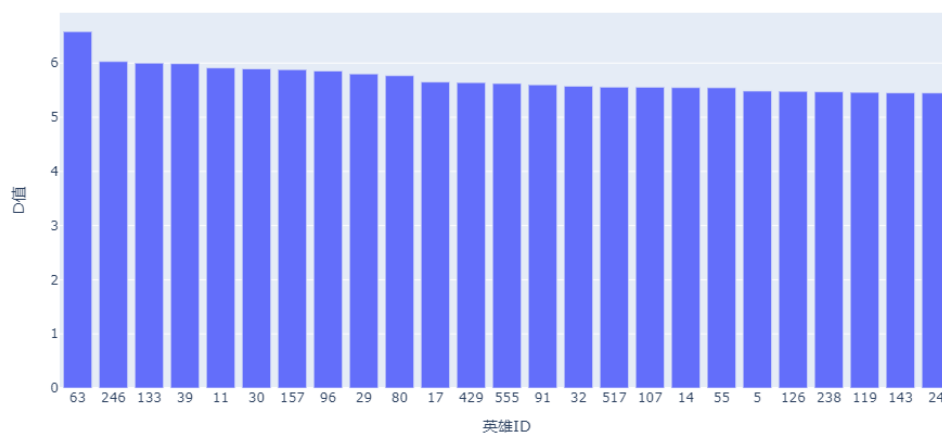
英雄A值排行



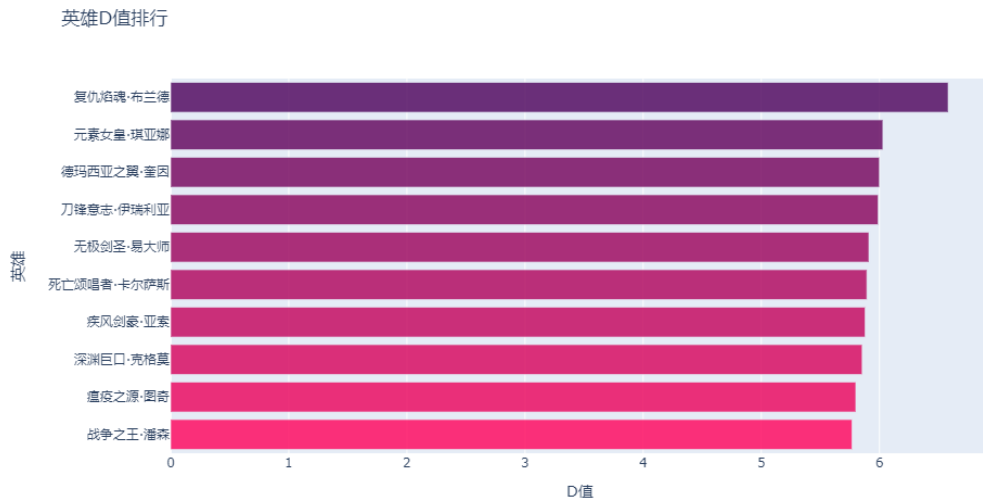
全部获胜队伍的A值排行前10。

21.3 D值

英雄D值排行



全部获胜队伍的D值排行前25。



全部获胜队伍的D值排行前10。

22. 获胜队伍中英雄常见位置和分路

英雄ID	胜利最常见位置
1	SOLO
2	DUO_SUPPORT
3	SOLO
4	SOLO
5	DUO_SUPPORT
6	SOLO
7	SOLO
8	SOLO
9	DUO_SUPPORT
10	SOLO

在所有获胜中，所有英雄出现频率最高的位置，全部内容请见jupyter notebook。

英雄ID	胜利最常见分路
1	MIDDLE
2	JUNGLE
3	MIDDLE
4	MIDDLE
5	JUNGLE
6	TOP
7	MIDDLE
8	MIDDLE
9	JUNGLE
10	TOP

在所有获胜中，所有英雄出现频率最高的分路，全部内容请见jupyter notebook。

23. 获胜队伍中英雄常见位置+分路搭配

英雄ID	胜利最常见搭配
1	SOLO + MIDDLE
2	SOLO + TOP
3	SOLO + MIDDLE
4	SOLO + MIDDLE
5	SOLO + TOP
6	SOLO + TOP
7	SOLO + MIDDLE
8	SOLO + MIDDLE
9	DUO_SUPPORT + BOTTOM
10	SOLO + TOP

在所有获胜中，所有英雄出现频率最高的位置+分路的搭配，全部内容请见 jupyter notebook。

总结

- 先关注最外显的目标：一血、一塔、一水晶、一男爵、一龙、一先锋。按照独立对获胜的影响有排序：一水晶>一男爵>一塔>一龙>一先锋>一血。所以在比赛中我们的首要目标应该是拿下一水晶，或者是实现一水晶的途径。当然能拿到一水晶也意味着队伍已经有了一定量的优势，累积优势并且发展本身就是比赛获胜的途径。
- 并且因为在一般比赛中，这些目标出现的大致顺序为：一血 →（一龙、一先锋）→ 一塔 → 一男爵 → 一水晶，我们要分析前向对后向的影响。考虑影响大于60%的内容有：拿下一血/一龙/一先锋都对拿下一塔也有着很强的助推力，拿下一先锋/一塔/一男爵对拿下一水晶也有很强的助推力。在未取得一水晶前，我们可以考虑通过一水晶前的目标来发展，可以根据具体情况来选择合适的路径，比如将一水晶的前向目标设为一塔，一塔的前向目标设为一龙等等。
- 同时也要考虑部分目标同时出现在同一个队伍中的情况，这里只考虑了两个目标共同影响的情况。发现无论是哪两个目标聚集出现在同一个队伍，这个队伍获胜的概率都极大提升。甚至如果一塔与一水晶同时被某队伍拿下，他们将有90%的概率获胜。这也是可以预见到，如果拿下一塔与一水晶，也意味着队伍本身发展已经极具优势，并且从拿下一塔到拿下一水晶可能相对顺利。所以无论是什么目标，我们都应该合理地争取，并且注意不同目标搭配带来的提升效果。
- 也有这些目标先后被两队拿下的情况，我们从这些情况里也可以发现不同目标对获胜影响的大小关系。比如拿下一龙和拿下一先锋带来的效用相差不多但一龙略大于一先锋，拿下一男爵的效用要大于拿下一先锋，拿下一水晶对获胜的影响占据最高地位。
- 一个队伍中最突出的人的表现一定程度上也能影响比赛的获胜，通过观察获胜队伍玩家最大击杀数，我们也能将最大击杀数作为影响获胜的指标，同时提高队伍的最大击杀数至10以上将获得大于60%的胜率，这也可以是玩家的目标。而场均击杀更能反映队伍的水平，它一定程度上代表了队伍的发展，

提升场均击杀至6以上将获得大于60%的胜率。这意味着队伍中需要没有短板或者分工要极其明确。

- 同理，助攻的多少除了反映玩家的水平与发展，也能反映其积极程度。提高队伍的最大助攻数至14以上将获得大于60%的胜率，提高队伍的场均助攻数至8以上将获得大于60%的胜率。这要求队伍分工明确，并且玩家都要积极比赛。
- 死亡数能反映自身的发展于对手的发展，在最大死亡数不超过7的情况下还能保证50%的胜率，场均死亡数不超过5的情况下保证50%的胜率。死亡意味着提升对手的发展以及对比赛的暂时失控，这是要值得注意的。
- KDA是衡量一个玩家/队伍的主要指标，KDA越高，此玩家/队伍的当下表现越优异。获胜队伍的场均KDA大部分集中在[2.475,5.525]的区间，但在比赛中时刻计算KDA并不现实，不过我们只需要关注KDA的三个参数，提高KDA意味着提升击杀+助攻并且注意控制死亡。
- 经济差是最直观反映双方发展差距的指标，当经济差过大，将影响双方基础实力的平衡。通过拟合，发现经济差的分布服从正态分布或者t分布，这将很好的为预测做指导。一般情况下两队的经济差不会太大，集中在相差绝对值[0,5k]的范围内，所以注意经济差也是比赛时的一项重要任务。
- 每一个赛季结束后都会有游戏出场率、胜率、KDA的统计信息，这将反映的玩家的偏好、地图的特点、当前赛季英雄的平衡性等等。同时如果有获胜英雄的最常见位置、分路、搭配，也能更好的为玩家之后的比赛提供指导性的思考。

感想

在本次实验以前，为并没有接触过英雄联盟这款游戏，从了解游戏名词到结合数据来分析的确花了不少时间。在实验的过程中遇到了很多困难，实验中每一个使用的库都是未曾接触过的，甚至为了提升可视化效果还在不断寻找更好用的库，在这个过程中的收获的喜悦是远大于解决困难的痛苦的。我也对python对大规模数据的处理有了更深的认识，往往是一个函数或者是几句逻辑语言，就能解决需要很多步骤或者看上去非常复杂的问题。对jupyter notebook的使用也让我感受到了极大的便利，它相较于pycharm更加的轻量，对于代码某一部分的检验更加便捷。通过大量数据的处理而得到直观结果的过程也让我对数据分析更加地感兴趣，挖掘数据后的规律和信息也十分有趣。最后感谢老师对于数据分析和统计方法的教学，助教对实验三的精心设计！