

绝地求生游戏数据分析

项目简介：

此次的项目分析数据基于 Kaggle 平台的开源数据集，数据集地址记录在代码处，分析工具为 python。主要进行的是探索性数据分析（EDA），工作主要有前期的数据准备和预处理，如对数据的获取、清洗、整合等；然后对数据进行一定程度上的挖掘并进行数据可视化；完成数据分析后将选取的数据上传至 mysql 数据库中。最后利用机器学习对数据集做了一个二分类的模型。

数据集介绍：

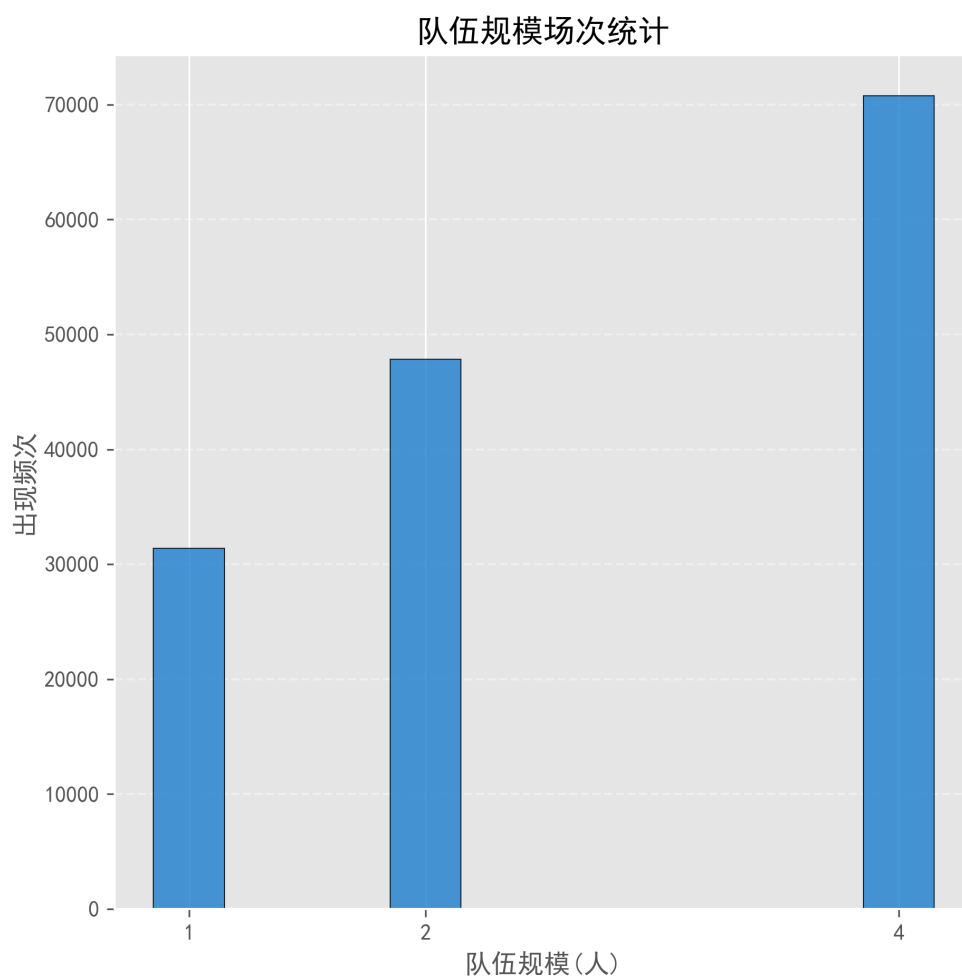
数据集信息来自是一款叫做《绝地求生》（PUBG）的即时多人的战术竞技类游戏，在该游戏中，玩家需要在游戏地图上收集各种资源，并在不断缩小的安全区域对抗其他玩家，让自己生存到最后。根据数据集作者介绍，数据源来自开放的第三方网站 pubg.op.gg，记录了大概 720000 场比赛的数据，包含了赛事概况数据和玩家死亡的具体数据两部分，本次项目取自于大概 150000 场的比赛概况数据，主要有赛事 id、比赛日期、游戏规模、比赛模式、队伍规模，最终排名、击杀数、生存时间等比赛信息。

数据分析：

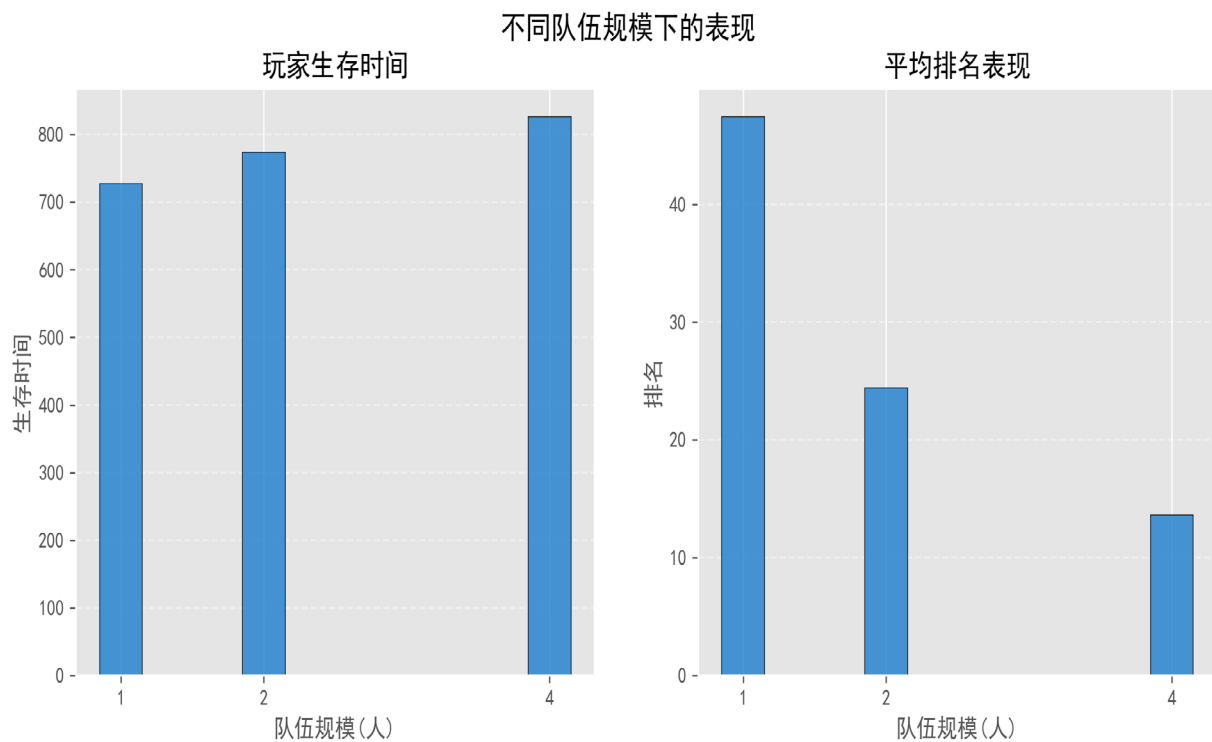
数据集信息

源数据包含 13849287 场不唯一的比赛数据，每场比赛有 15 类数据信息，分别是比赛日期、游戏规模、赛事 id、比赛模式、队伍规矩、玩家助攻、击倒敌人次数、车辆驾驶距离、玩家行走距离、玩家造成伤害量、击杀玩家数、玩家名称、生存时间、队伍 id、最终排名。

在数据选择上，使用了唯一的赛事 id 数据，即去除了重复的比赛场次，最终共有 149994 场比赛数据。同时还去除了相对不太重要的数据维度，有队伍 id，比赛日期，玩家名称；并且在游戏模式分布中，所有玩家都是第三人称视角（TPP），因此也筛除出分析维度里。源数据并没有缺失值，所以最终以筛选后的数据来进行后面的数据分析与挖掘。

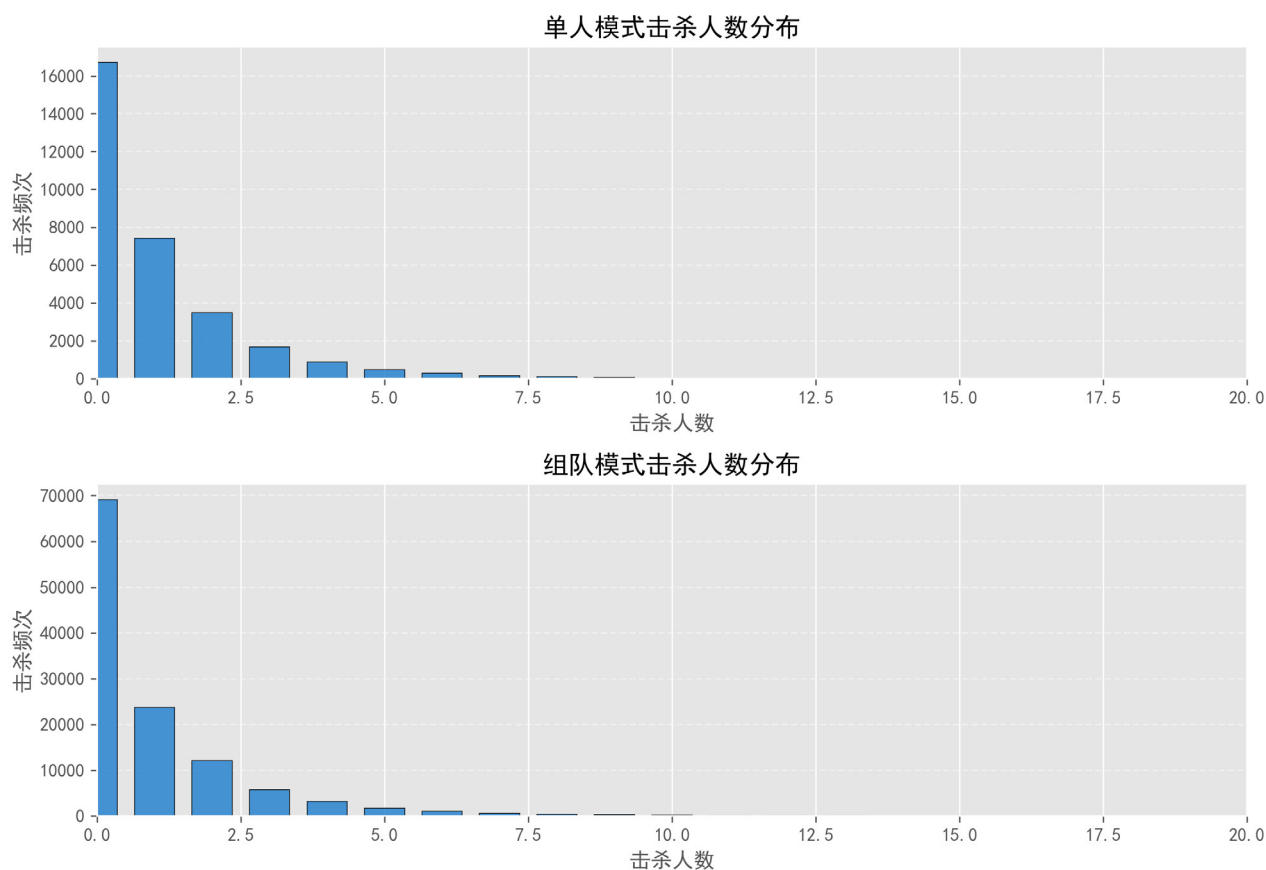


从队伍规模的分布中可以看出，单人排位的赛事最少，大概 30000 场左右；随着队伍规模的扩大，比赛场次也在增加，四人组队的比例最高，大概在 70000 场左右。玩家或许更加青睐多人组队下的比赛，又或是更喜欢与朋友们进行游戏。



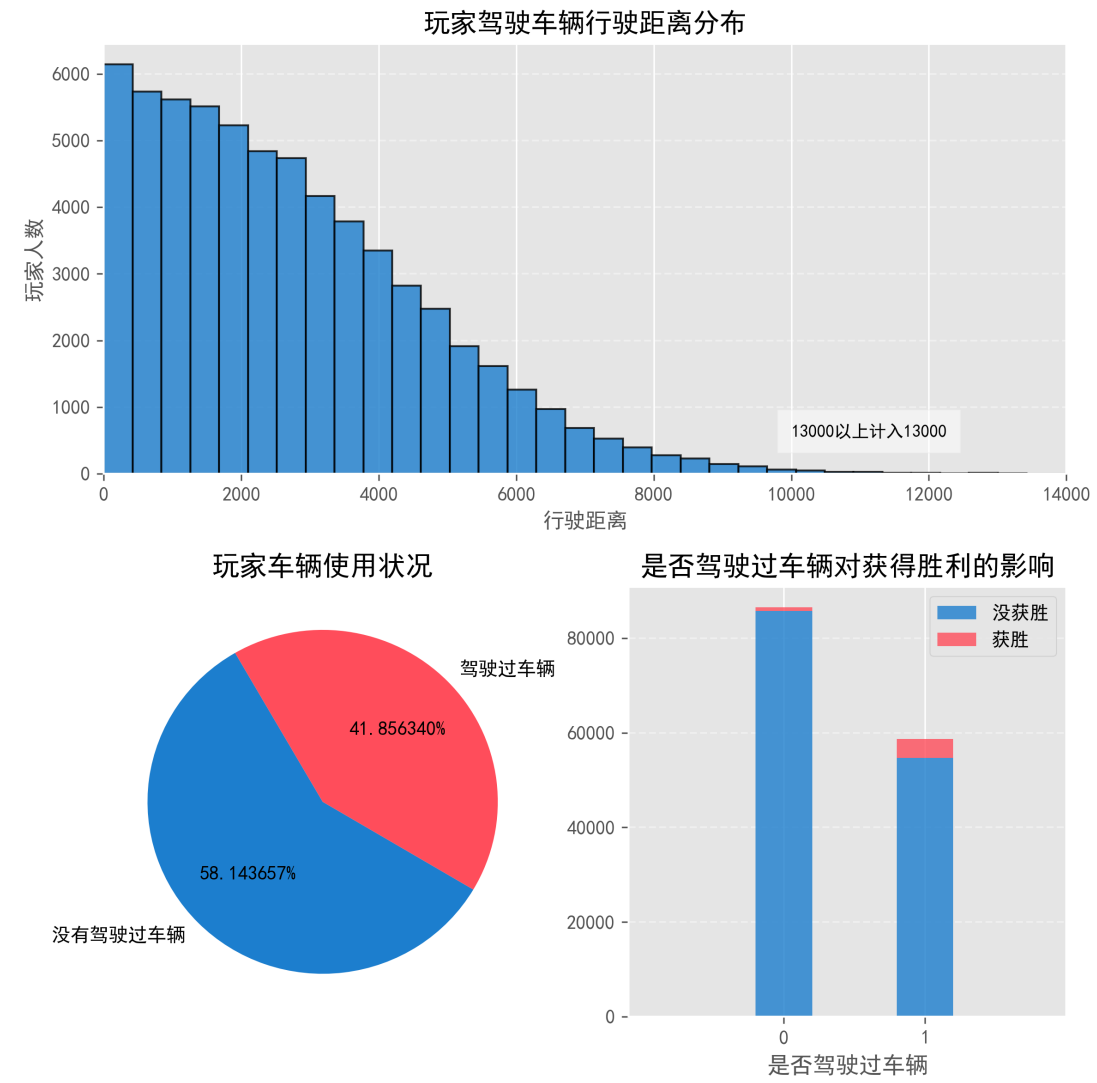
上图是不同队伍规模下的玩家平均生存时间和最终的平均排名表现。在生存时间上，各队伍规模下的表现相差不大，4人组队模式下稍微表现得更好。而在最终的排名表现里，单人平均在48名左右，而二人组组队是23，四人组队则为13。但是考虑到多人模式下玩家的总数也是一样的，所以双人排位和四人排位的最终排名转化过来大概在46和52名，所以是否进行组队并不是决定你能否获胜的关键因素。

各模式下击杀人数分布



在击杀人数上，单场比赛能够完成 20 杀以上的玩家占比过于的少，并且在 25 杀以上的人数几乎没有，甚至对于一些极端的击杀数怀疑是来自于外挂的表现，因此在分布上只选择了 0~20 的人数分布。从图中可观察到单人和组队模式的表现非常的相似，大部分人处于无击杀或者是击杀数少于 2 的区间中，随着需要击杀更多的敌人，难度也是不断的上升，能够做到单场比赛完成 10 杀的玩家凤毛麟角。

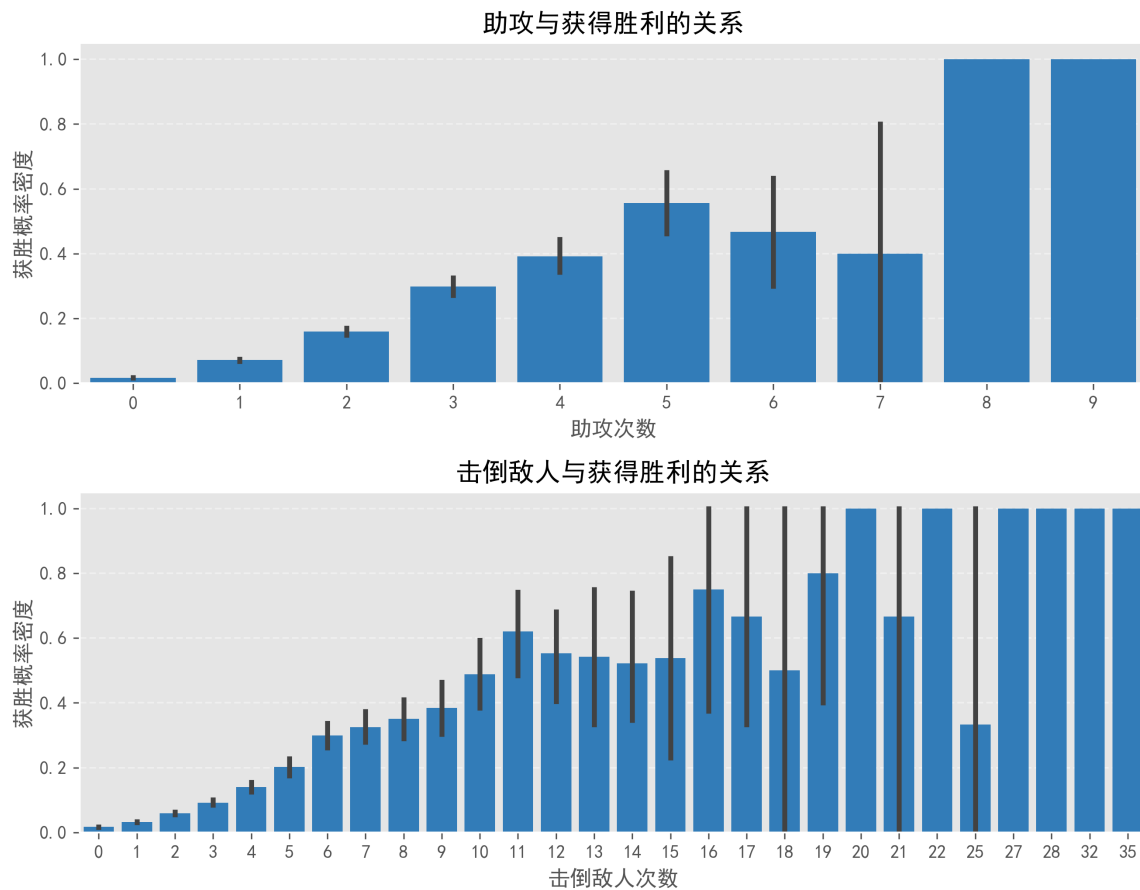
玩家车辆驾驶信息



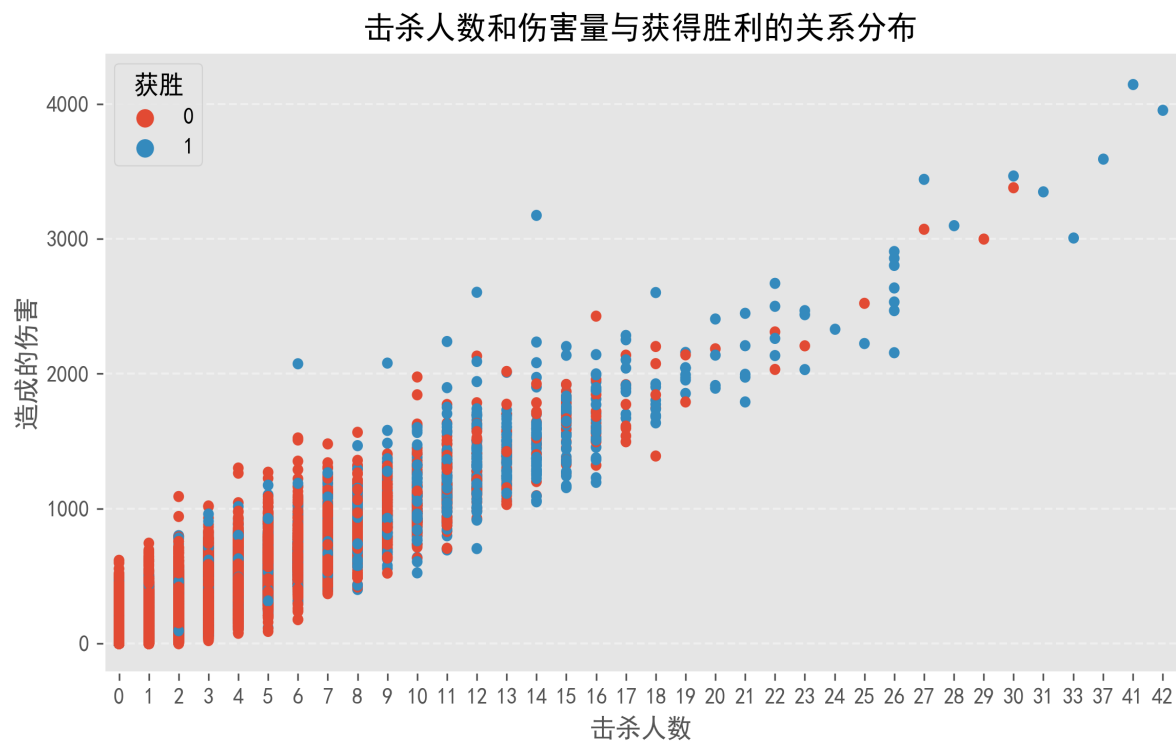
在游戏中不断缩小的安全区里，车辆的使用对于玩家尽可能存活得更久有着重要的作用。车辆的稀少导致了驾驶过的玩家占比只有 41%左右；而在驾驶过车辆的行驶距离上，0~2000 的占比最多，并且随着距离的增加，能够持久地使用车辆的玩家也在不断减少。

是否驾驶过车辆	无	有
是否获胜		
无	86488	58729
有	724	4053

作为游戏里的稀缺品，车辆也在一定程度上影响着玩家的胜率，没有驾驶过车具的玩家获胜概率为 0.8%，驾驶过车辆的玩家胜率为 7%左右。



助攻和击倒敌人的次数同时也是影响胜率的一部分，当助攻在 5 次左右的时候，获胜的几率相对比较高，而在 7 次以上则几乎确定胜局。击倒敌人方面，随着次数的增多，胜率也相应地提高，到 15 次左右时胜率波动较为稳定，当击倒数超越 25 时，获胜的概率可以近乎 100%。



在击杀人数与伤害量和获胜的分布图中，可以发现随着击杀人数与伤害量不断地增加，获胜的几率也在明显上升，所以想要增加胜率，仅仅为了存活而回避战斗的“苟活”方式并不是一个好的选择，真正要取得胜利还是需要加强战斗的技巧。

以上为探索性分析的部分，完成后将数据分析中筛选并清洗过的数据集上传至 mysql 数据库中以便保存和方便日后使用。然后第二步，对清洗过的数据进行机器学习和建模，按照特征工程、模型选择与训练、效果评估等步骤，对此次绝地求生游戏的数据根据获胜与否建立了一个二分类模型。最后将所有项目上传至 github 便于管理和分享。

项目地址: <https://github.com/KKwave/pubg-analyzing>

作者：龚伟星

日期：2018-5-2