# 数据库综合作业

## 作业进度

1. 爬取榜单数据
2. 数据清洗
3. 将爬取数据直接写入数据库



## 可视化

预计将数据库的数据读取到python中并在python进行数据可视化。

1. 综合得分计算指标可视化

哔哩哔哩综合得分，是视频是否能排上排行榜的依据，若能知道其规则，对于视频内容的倾向是有很大的价值的，所以在此首先进行综合得分计算指标的分析及其可视化。预计使用一种多因素统计分析方法进行分析，如灰色关联度分析。分析出权重后用关系图可视化。

1. 白嫖占比分析图

白嫖数目=播放量-赞-收藏-投币

考虑到有观众可能会直接一键三连（评论、点赞、收藏），包括可以重复播放但是赞、收藏、投币无法重复，结果不会特别的准确。因此数据仅供娱乐。预计使用饼图进行比例可视化。

1. 视频时长与视频的受欢迎程度分析

目前视频行业可以分为爱优腾为代表的长视频赛道，快手抖音为代表的短视频赛道。B站作为多种元素结合的网站，从中可以分析出B站用户喜欢的视频类型。预计使用散点图可视化。

1. B站最受欢迎前十名综合得分可视化

通过综合得分的差异与名次的排序，这张图试图体现b站视频的竞争激烈程度。预计使用柱状图可视化。