**B站用户行为习惯与特征分析**

前言

所有的图表都是用pyecharts做的，都是动态图，但是在word中没办法插入，所以如有需要请参照文件夹中的html文件或notebook里的图表，以下的图都是静态图。

1. 分析目的

该报告从UP主和观众两个角度出发，通过对视频信息的抓取，处理和分析，从多维度总结了B站用户行为习惯与特征。

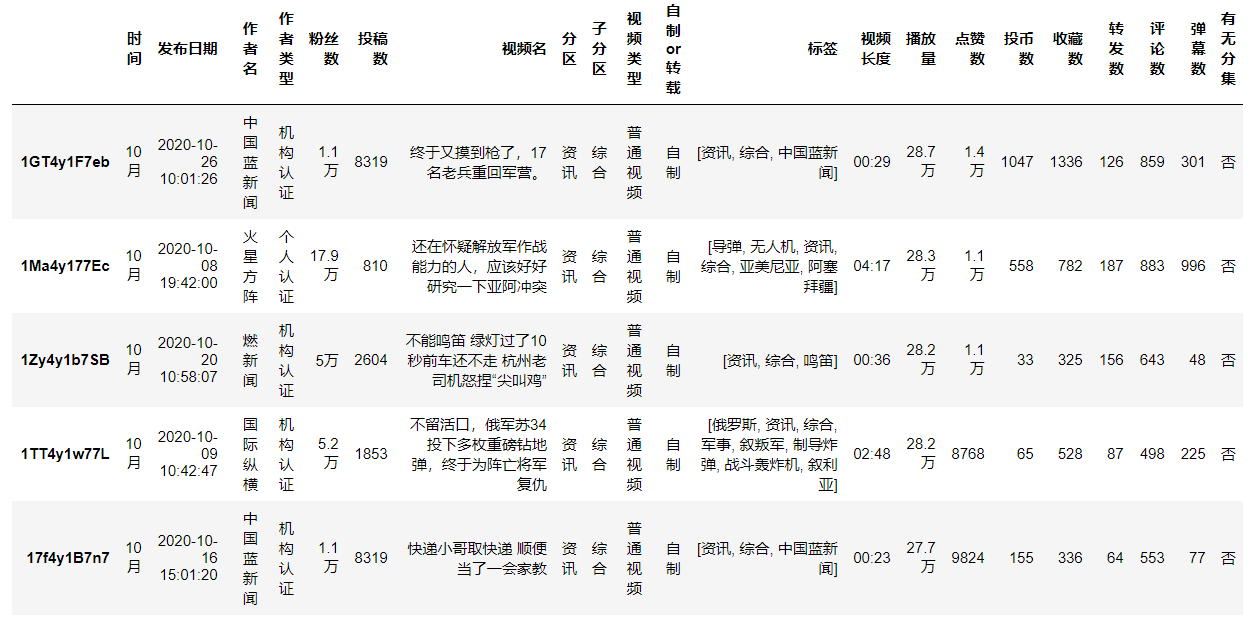
1. 数据描述

报告选取了B站生活、音乐、游戏、舞蹈、知识、动画、鬼畜、美食、时尚、资讯、数码、影视和娱乐分区共13个分区的视频作为数据来源（其他分区的视频多为B站官方发布，不是用户行为，因此不做考虑），共爬取了两份数据：

（一）各分区的各子分区1月到10月的投稿数（对应chart1-1到chart1-4的分析）



（二）各分区的各子分区1月到10月播放量前100的视频信息（对应chart2-1到chart6-2的分析）



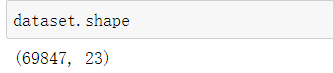
初步预处理包括：

1. 将“发布日期”分解为对应的一周中的某一天，如“Monday”，“Tuesday”等等
2. 将播放量、点赞数、投币数等字符串类型的数据转化成整数

结果如下：



总计69847条数据，23个特征：



1. 具体分析

该部分截图只有图表，代码已经在notebook里按顺序列好。

1. UP主行为习惯与特征
2. 投稿量
   1. 每个频道投稿数的时间变化

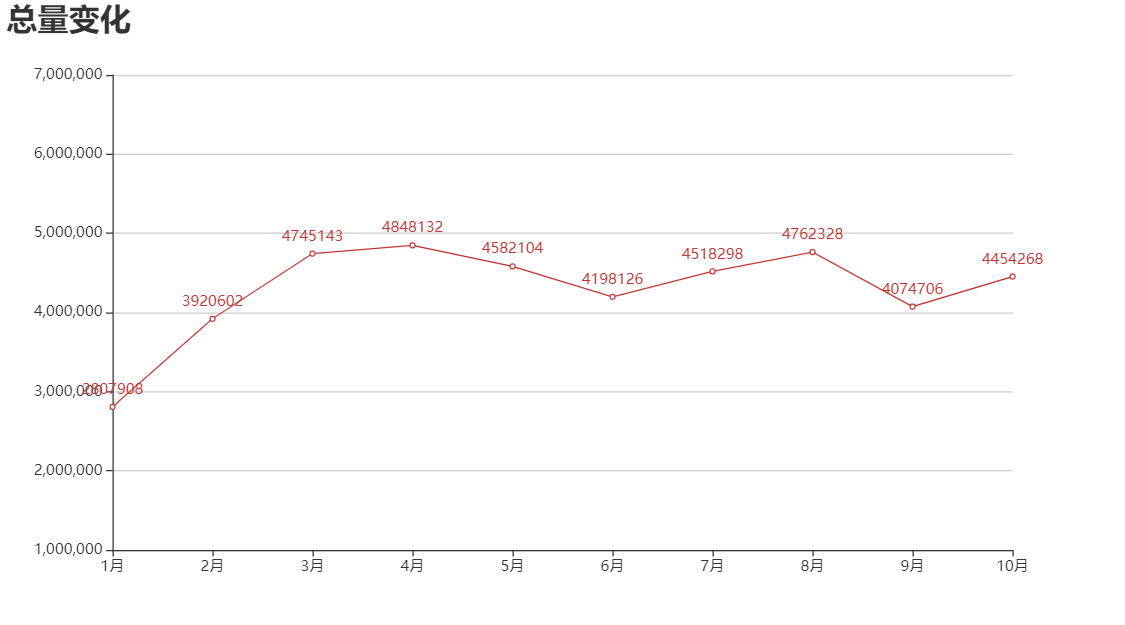


chart1-1

总投稿数在1-3月有一段快速且大幅的增长，之后有一定的波动，不过基本维持在450万这个水平左右。

* 1. 各频道投稿数的增量对比

这张图是为了能够对每个分区的投稿数的总体增长情况有一个更详细的了解。从图中可以看出，无论是相对增长还是绝对增长都呈现出了三个分区遥遥领先，所有分区共同进步的特点。

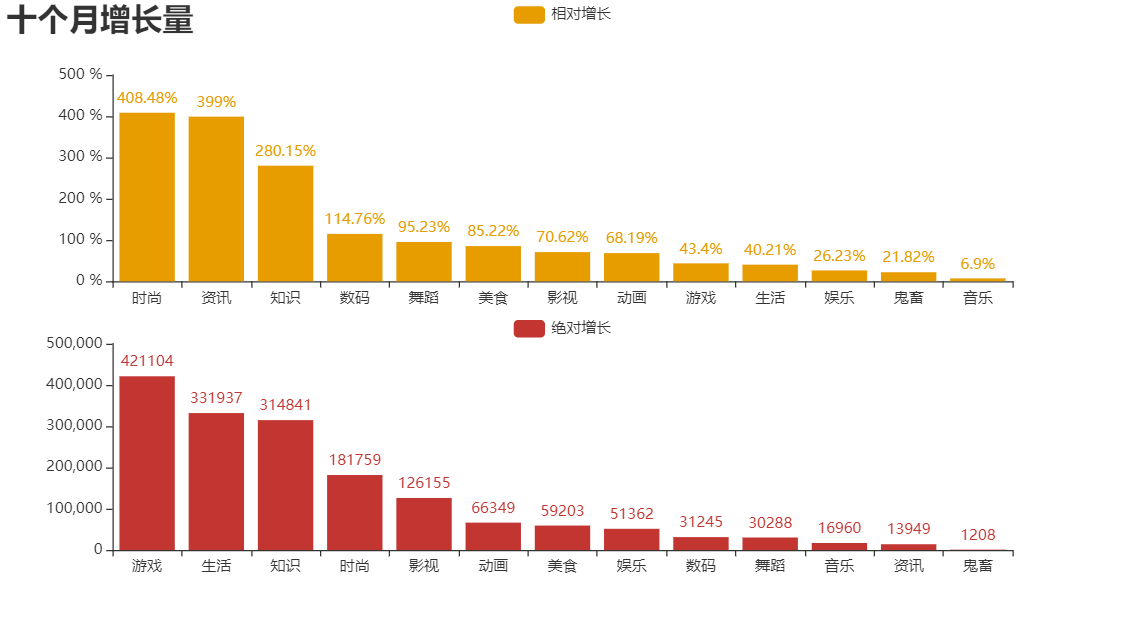


chart1-2

就相对增长来说，时尚、资讯和知识分区涨幅最大，最高的甚至已经突破400%，而B站主力频道的涨幅排名却比较靠后，这可能是因为时尚、资讯和知识分区的体量相对小很多。

就绝对增长而言，游戏、生活和知识分区的增长占比达到了65%，游戏区和生活区的霸主地位依然稳固。

值得注意的是，知识分区的两项指标排名都非常靠前，预示着该分区，至少从供给的角度来讲，未来可能会成为主要分区之一，因为接近300%的相对增长基本不可能单纯靠老UP主增加投稿量来实现，必然是新UP主的加入，而大幅的绝对增长就意味着新UP主的数量大大超过其他分区。

* 1. 各频道投稿数时间变化

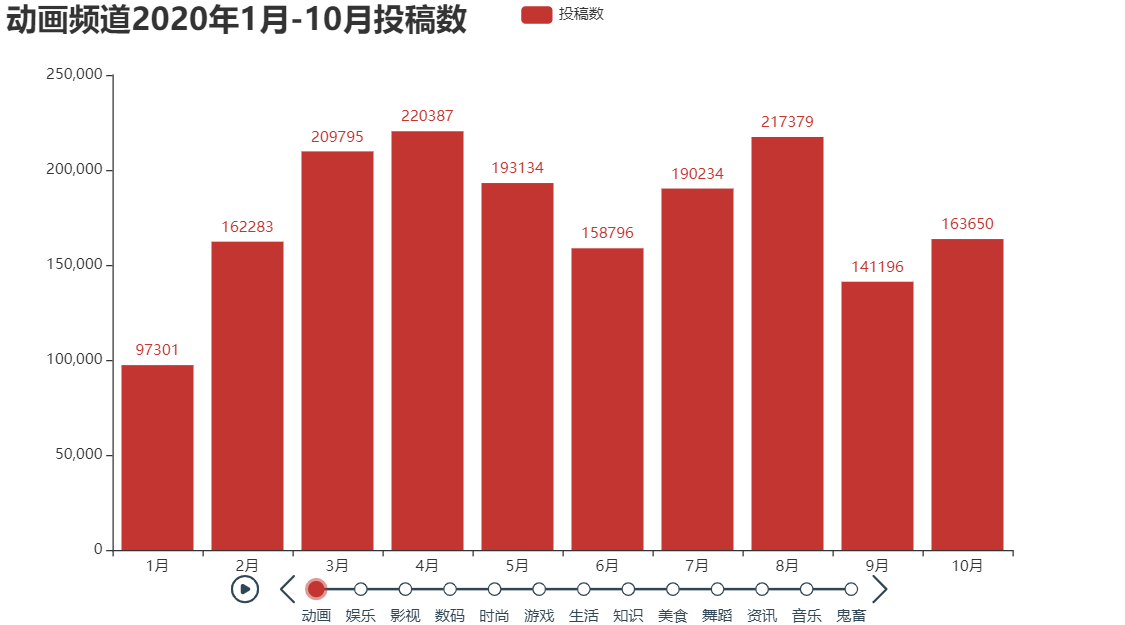


chart1-3

可以发现频道投稿数的变化有一个共同的趋势，即1月到4月快速增长，又在5、6月下滑。按照后续的变化趋势又可以细分成两类，一类是“M型”，即在8月左右达到新峰值，随后又呈现一定程度的下跌，这个趋势和总体变化趋势相近，主要体现在动画、娱乐、影视、数码、游戏、生活、美食、音乐和鬼畜区；另一类是整体上升型，即后续投稿量呈稳定增长趋势，包括时尚、知识、舞蹈和资讯区。

* 1. 同一个时间段不同频道投稿数对比

从下面这张图可以得到的最直观的信息就是投稿数的分布在不同的月份里有惊人的相似程度，游戏区和生活区的投稿数都大大超过其他分区，而资讯和鬼畜区非常地惨淡，而鬼畜区涨幅也非常地低（见chart1-1），想当年鬼畜区也算一个大区，现在却没落了。

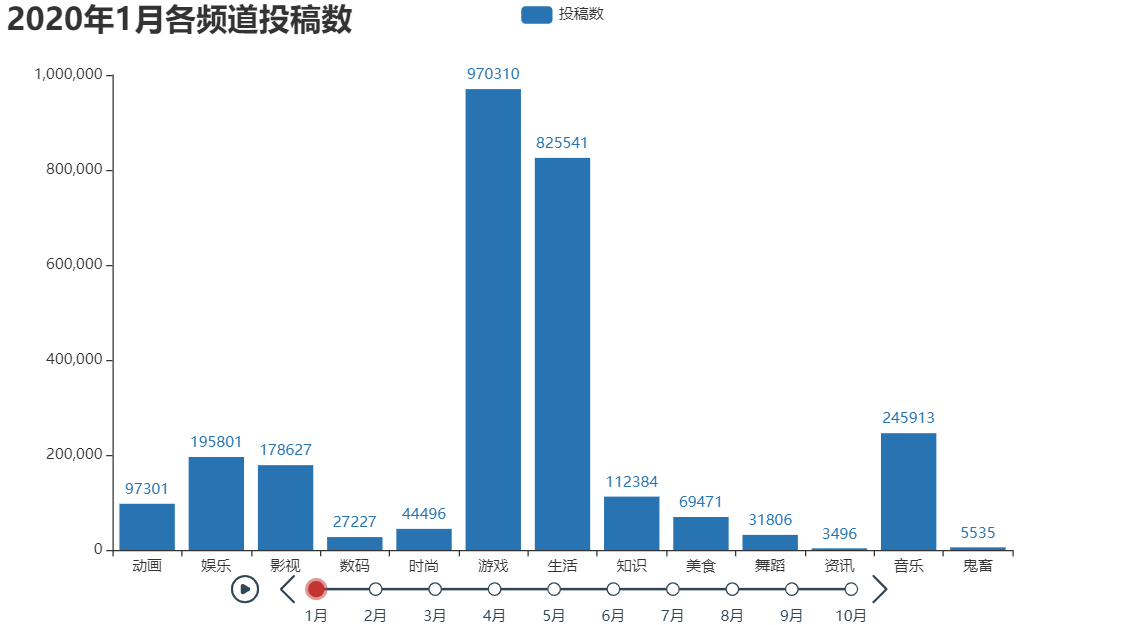


chart1-4

1. 投稿习惯
   1. UP主行为习惯与特征

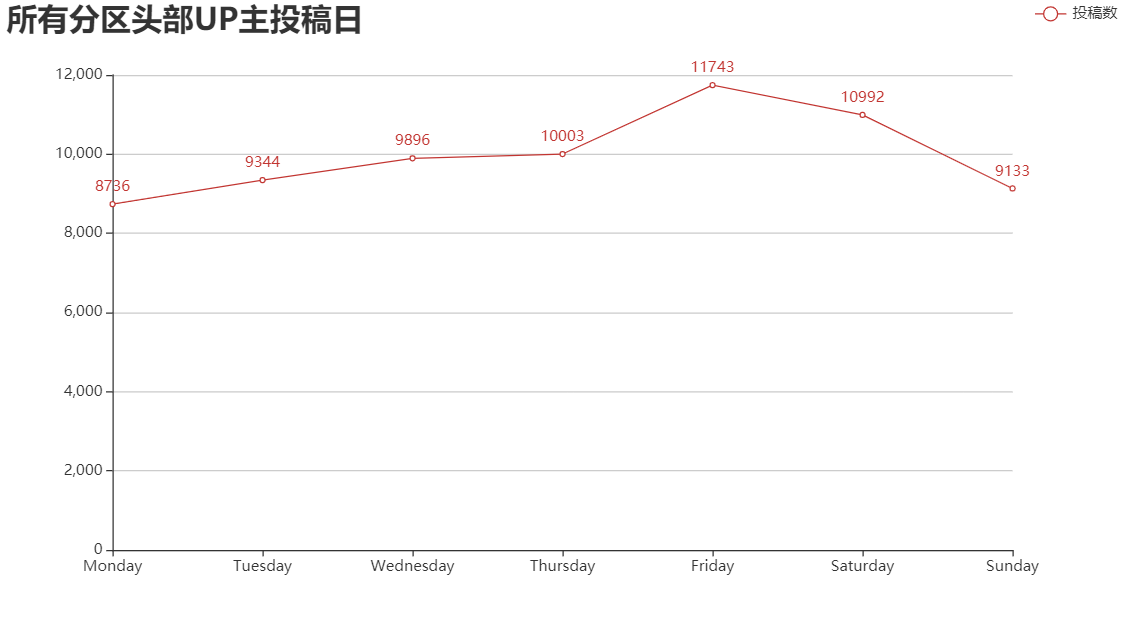


chart2-1

图中可以看出有更多的UP主会在周五和周六发布视频，但每天的差距其实不是特别大，所以可以猜测头部UP主中非全职的比较少，但还是有一部分，从而导致了周一最少，周五、周六最多的情况，毕竟周一一般来说是最忙的，周五、周六时间就相对多一些。

* 1. 各分区头部UP主投稿日

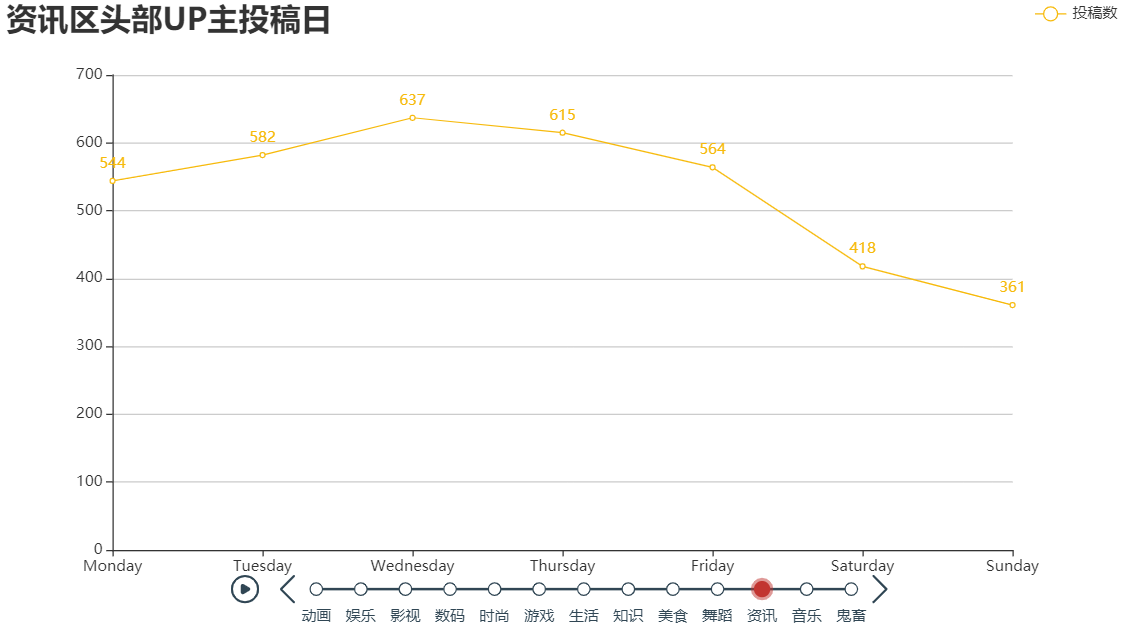


chart2-2

总体来说，周五、周六依然是投稿数最多的时候，周一是最少的时候。但也有一些新特征，比如，娱乐区有很多UP主选择在周二发布，影视和数码区的高峰在周四。但其中最奇怪的是资讯区，周末反而是最少的，而且差距有点大，这一点会在后文中给出解释。

* 1. 所有分区头部UP主投稿时间点

（横坐标说明：“0”指的就是0点到1点发的，“1” 指的就是1点到2点发的，以此类推，以下都简称为某某时点发的）

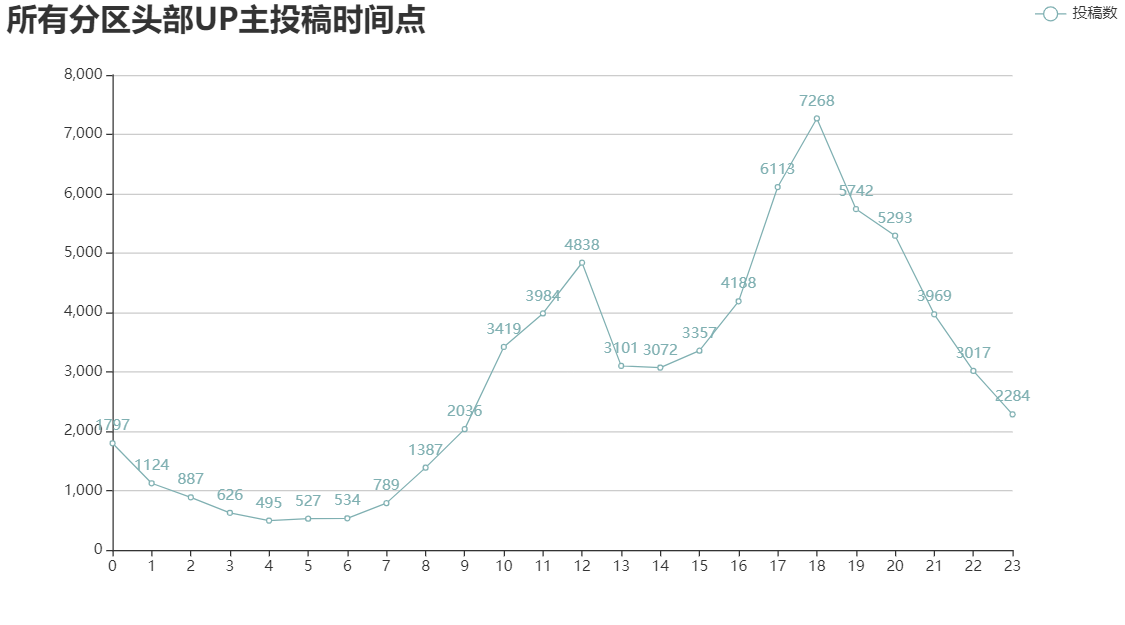


chart2-3

可以发现，投稿数在午饭饭点达到一波小高峰，到1点到3点又快速下降，在晚饭饭点又达到新高，随后也是一直下降直到凌晨4点降至最低点。所以我猜测头部UP主的行为模式是这样的：一些UP主会早上起来做视频，做到中午大部分的UP主就做完了，大部分人中午会选择休息，而更多的UP主会在下午才开始工作，再加上早上就开始工作的UP主可能有的早上没完成，下午接着做，所以出现了1点到6点比7点到12点高一截的情况。到了晚上，大家陆陆续续也都做完了，所以数量直线下滑，但是最低也有500左右，说明还是有一小部分头部UP主是连夜肝视频的。

* 1. 各分区头部UP主投稿时间点

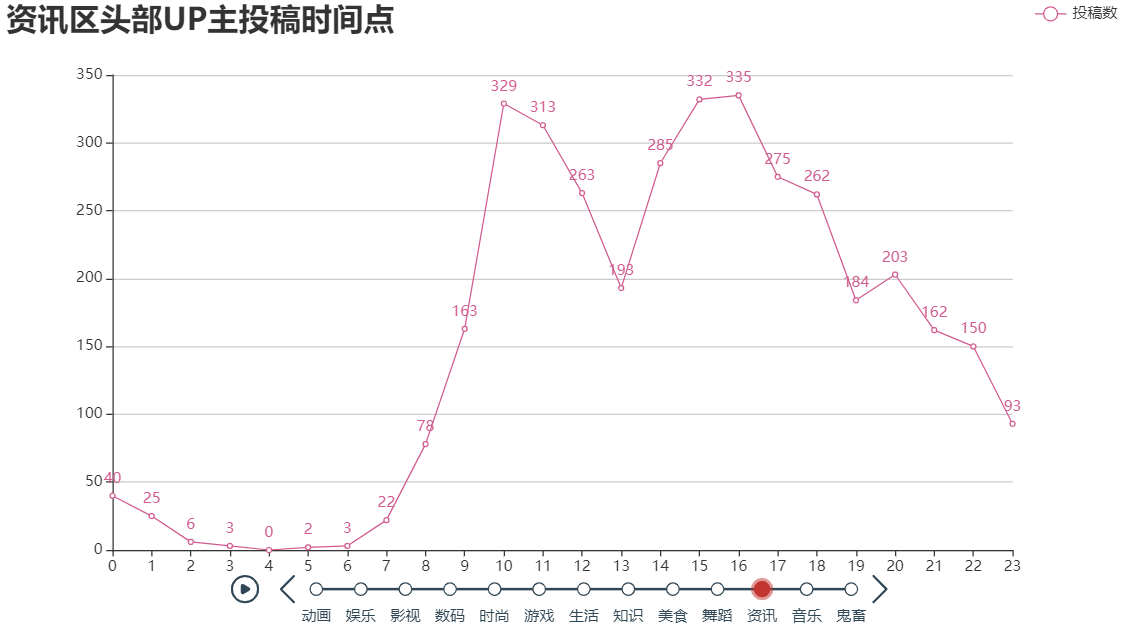


chart2-4

大部分分区的特征与总体特征类似，即12点和18点为极大值，4、5点和13点为极小值。但也有少数分区的趋势有点偏离，如娱乐区，它在20点的投稿数远大于其他时间段，前面也提到有很多娱乐区UP主会选择在周二投稿，可能他们有一些独特的做事风格吧。此外，资讯区偏差也非常明显，早上主要集中在10点和11点，下午主要集中在15点到16点，结合chart2-2，也可以发现这个分区的做事风格也很独特，在下一张图我会阐述我的原因猜想。

1. 头部UP主和视频类型
   1. 头部UP主类型

我认为下面这张图可以在一定程度上解释上述现象。可以明显看到资讯区几乎全是机构认证，而大部分机构都是有一个比较固定的工作时间，可能是早上12点休息，下午6点下班，所以才会出现大部分稿件集中在中午休息和下午下班之前发，而且是在工作日发的现象。但是也看得出来休息时段发稿的数量也不少，所以可能他们加班也比较频繁。

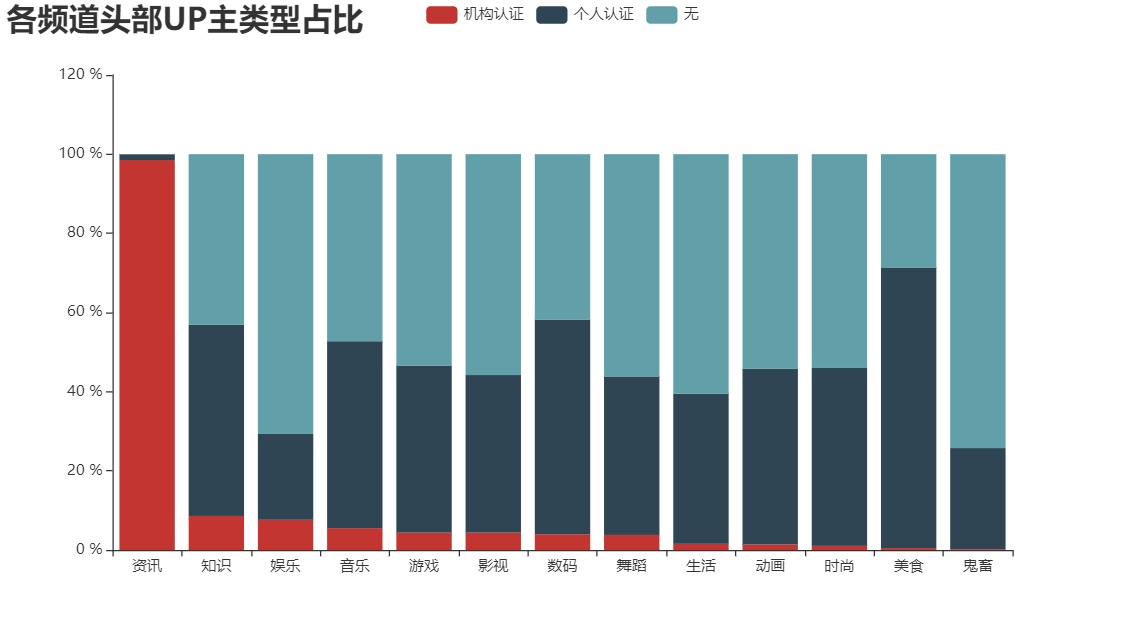


chart3-1

此外，可以看到在其他分区，机构认证很少，无认证和有个人认证的UP主都有一定的比例。在此解释一下个人认证的条件。个人认证一般要么B站粉丝数累计≥10万，要么站外粉丝数>=50w。其中一个好处是搜索优先权，意味着有更多机会曝光作品。但是我们还是能看到有相对一部分头部视频的UP主没有认证，说明是金子总会发光的，只要视频做的好，就算是无名之辈，也能有百万播放量。

* 1. 头部UP主类型

可以看出，一般来说，视频类型占比大小是普通视频>活动作品>合作视频>互动视频。出现该现象应该是因为：

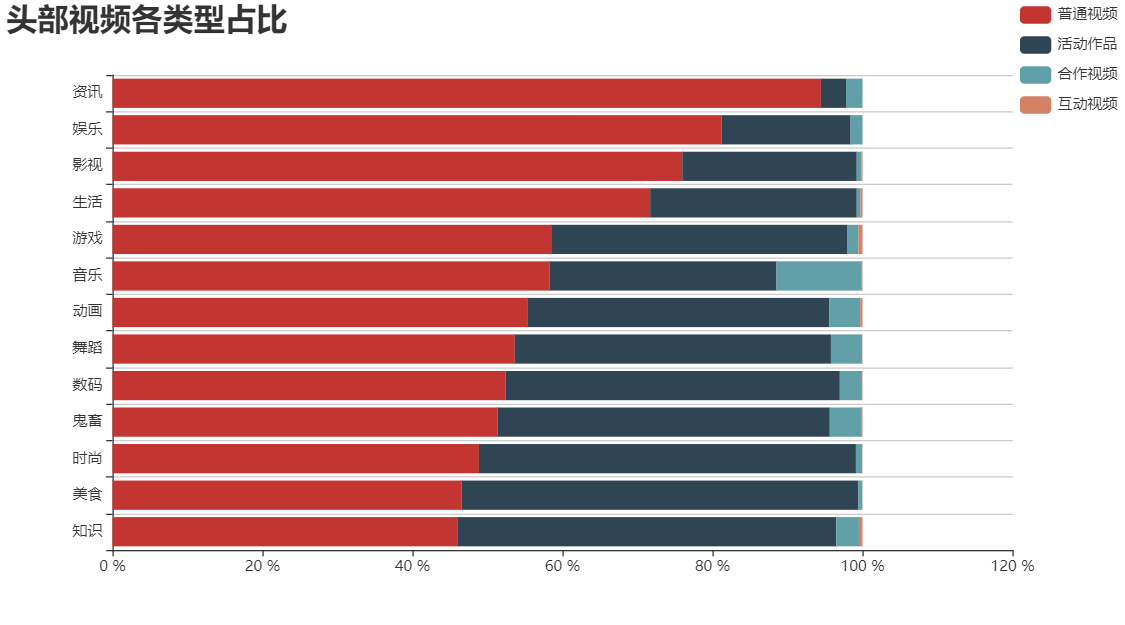


chart3-2

B站活动依然有限，而且各分区活动数量不一，所以虽然活动作品可以获得更多收益和曝光，活动视频一般都少于，有时甚至远少于普通视频。

要想制作合作视频有一定的门槛，即UP主的创作力或影响力要达到一定程度，因此占比较少。而且这种类型可能只在某些分区比较实用，比如音乐区，因为可能需要有人谱曲，有人作词，有人演唱，有人伴奏，可分工性强一些，所以该类型占比比其他分区的多一些。

互动视频制作成本高，并且我猜测可能大多数人其实并不喜欢这种形式，观众可能更多的只想好好看视频，根本就懒得去点击，因此互动视频一开始的热门可能只是因为新鲜，到现在头部视频里很少能看见互动视频的身影了。

1. 头部视频时长
   1. 所有分区头部视频长度分布

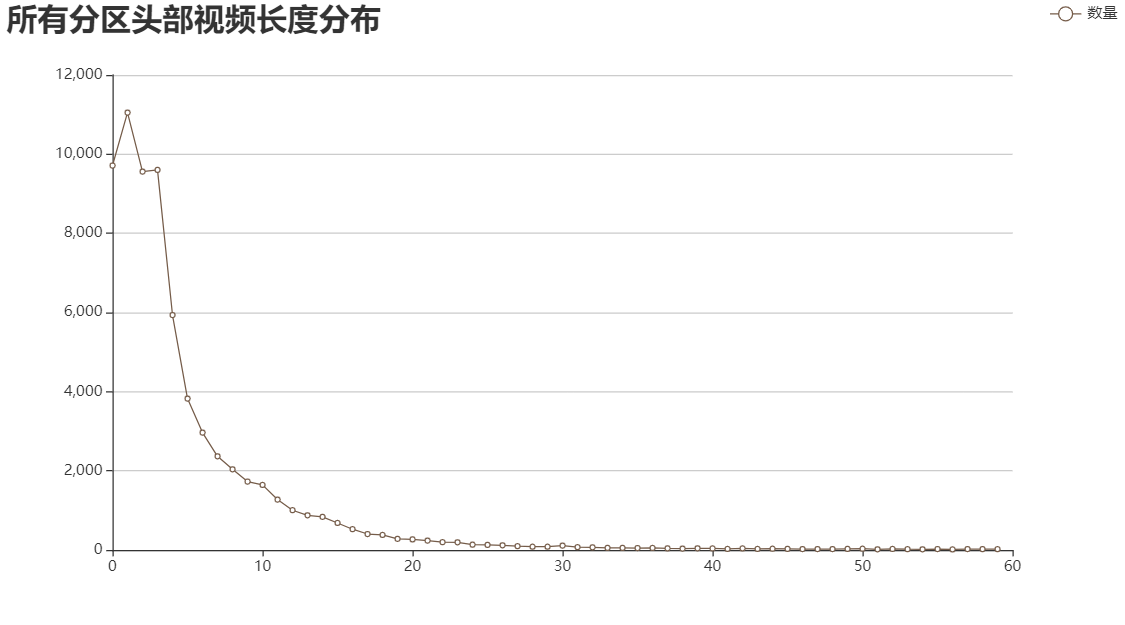


chart4-1

从图中可以得出一个总体特征，B站头部视频长度主要集中在0到4分钟，5到20分钟的视频数量下降速度非常之快，多于20分钟的基本没有。

* 1. 各分区头部视频长度分布

看chart4-2可以发现，各分区的长度分布呈现两种趋势，一种是和总体趋势类似的，包括动画、娱乐、影视、生活、舞蹈、资讯、鬼畜，而另一种类型的分区在5到20分钟的视频数量下降速度明显要慢很多，也就是说有相当多的视频时长在这段区间。这主要是因为这些分区的视频往往需要UP主加上一定时间的解说。比如数码区UP主介绍设备，教你装机；时尚区UP主教你化妆穿搭；知识区UP主给你解释一个科学现象都是需要一段时间才能说清楚的。

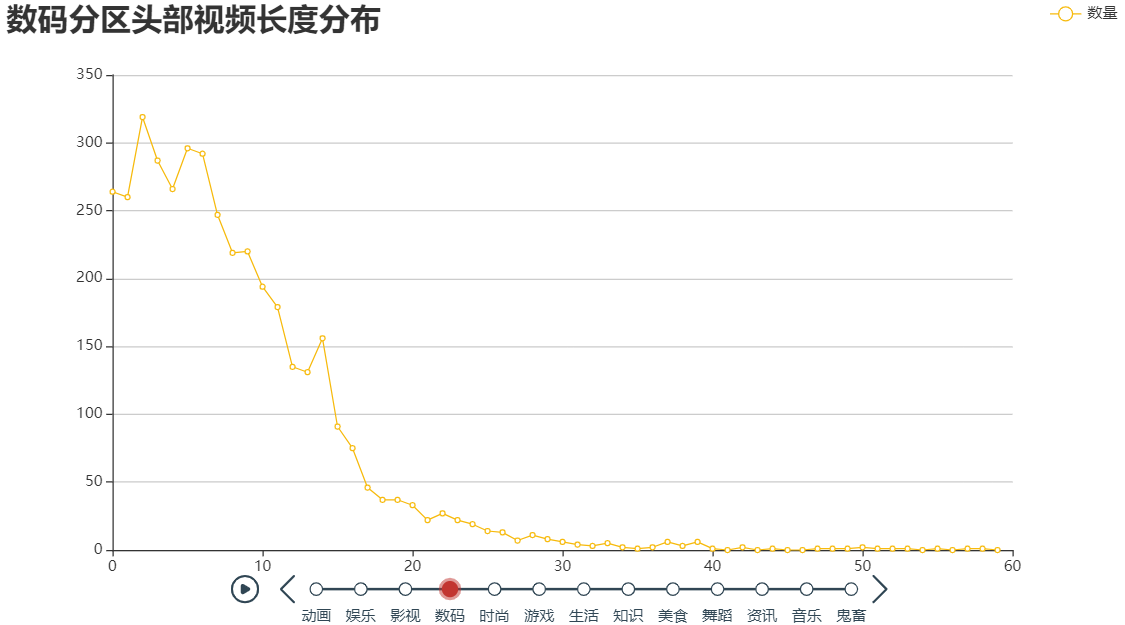


chart4-2

1. 观众行为习惯与特征

1、 头部视频平均播放量

1. 各分区头部视频平均播放量

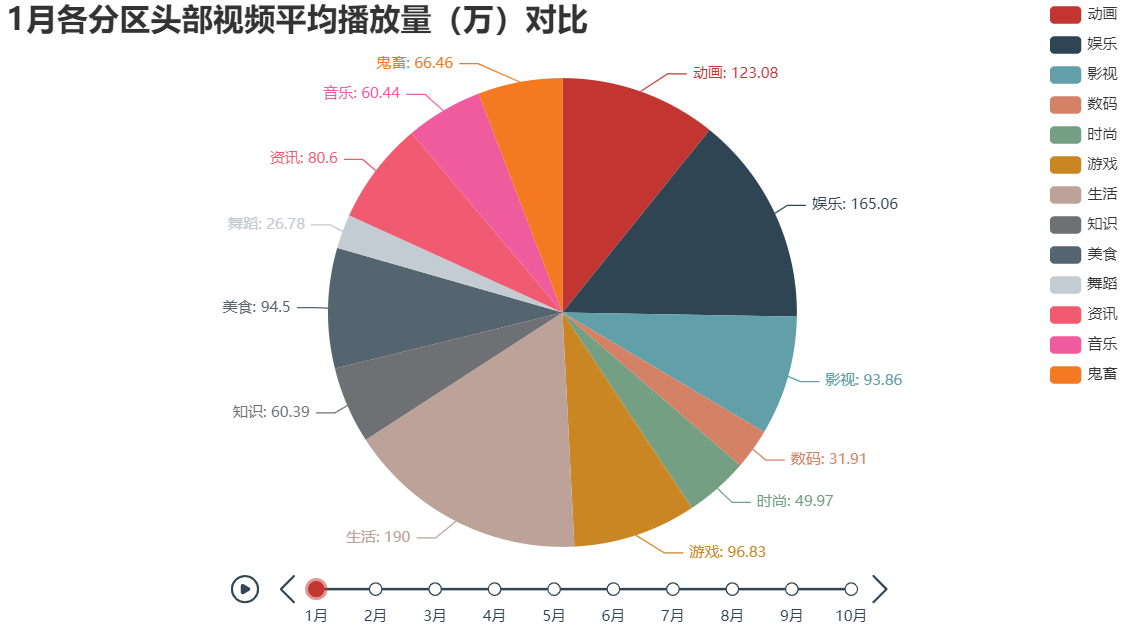


chart5-1

通过不同时间段的对比可以发现，从播放量的角度来看，分区地位非常稳定，观众最喜欢的是生活和娱乐区，关注度最少的舞蹈和数码区，这应该是因为舞蹈和数码本来受众就比较小。

1. 各子分区头部视频平均播放量

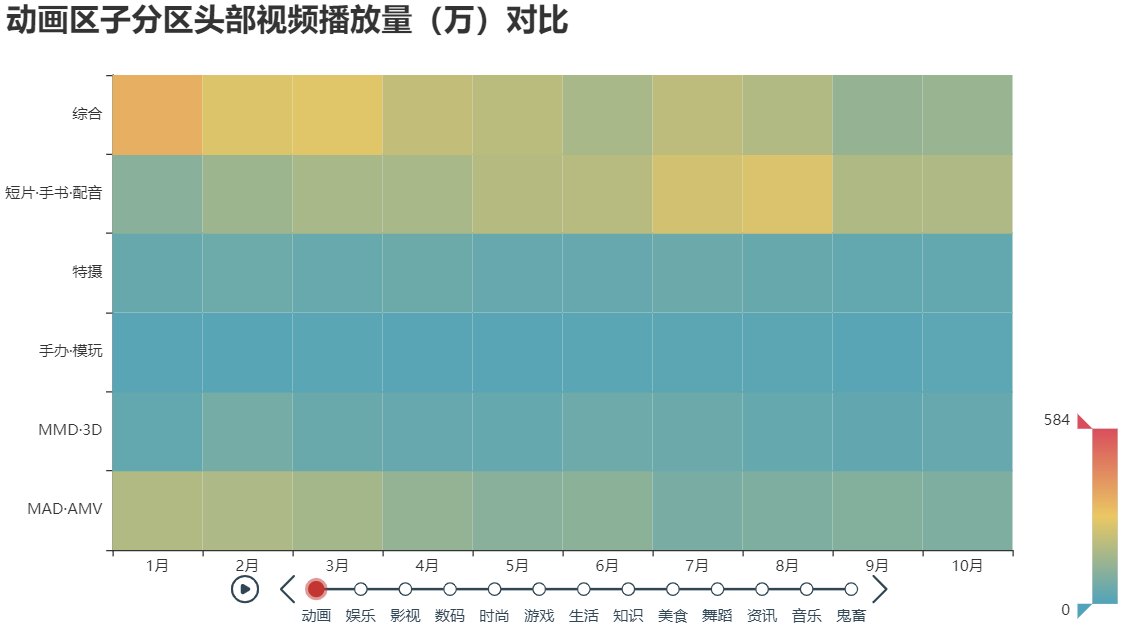


chart5-2

这张图主要是为了了解各个分区内部是否也存在垄断现象，即某些子分区的播放量远大于其他子分区。判断的方法就是看颜色的对比程度，差距越大，垄断现象越严重。

可以发现，每个分区都存在一定程度的垄断，这里因为色差的跨度有点大，所以即使颜色看起来相近，也可能存在一定的差距。

综合来看垄断最严重的区是生活区，日常和搞笑子分区之和比其他子分区之和还要大；其次是动画、影视、游戏、知识、美食和鬼畜区，它们的黄蓝相间的程度都比较明显。

2、 白嫖与话痨程度

1. 白嫖程度对比

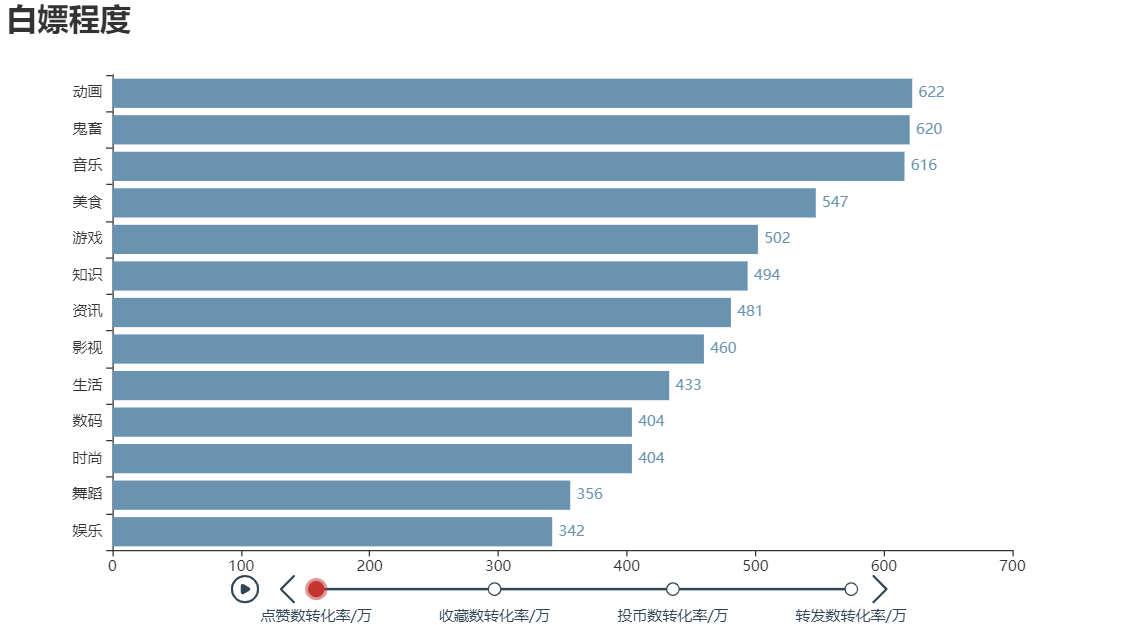


chart6-1

首先可以发现转化率上，点赞>收藏>投币>转发，道理很简单，点赞最不麻烦，代价最小；收藏没什么代价，但是麻烦一点而且很看视频性质；投币的代价比较大，毕竟对大多数人来说币是有限的；而转发既麻烦而且可能要付出让别人觉得自己很奇怪的代价。

其次，白嫖现象非常严重，情况最好的分区点赞转化率不到十分之一，收藏和投币转化率不到二十分之一。对比之下动画、鬼畜和音乐区在一键三连的三个方面排名都很靠前，说明这三个区的观众是最善良的。

1. 话痨程度对比

从图chart6-1可以得出两点结论：

首先，数码和知识区的观众最话痨，可能是因为这些观众展现自己知识储备的欲望非常强烈，而舞蹈区的观众最安静，可能他们的双手早已离开手机或者键盘。

其次，几乎所有分区的弹幕数都大于评论数，说明观众比较喜欢弹幕发言，毕竟头部视频的评论一般比较多，评论区留言很大概率没人会看到，但是弹幕被看到的概率还是相当高的。

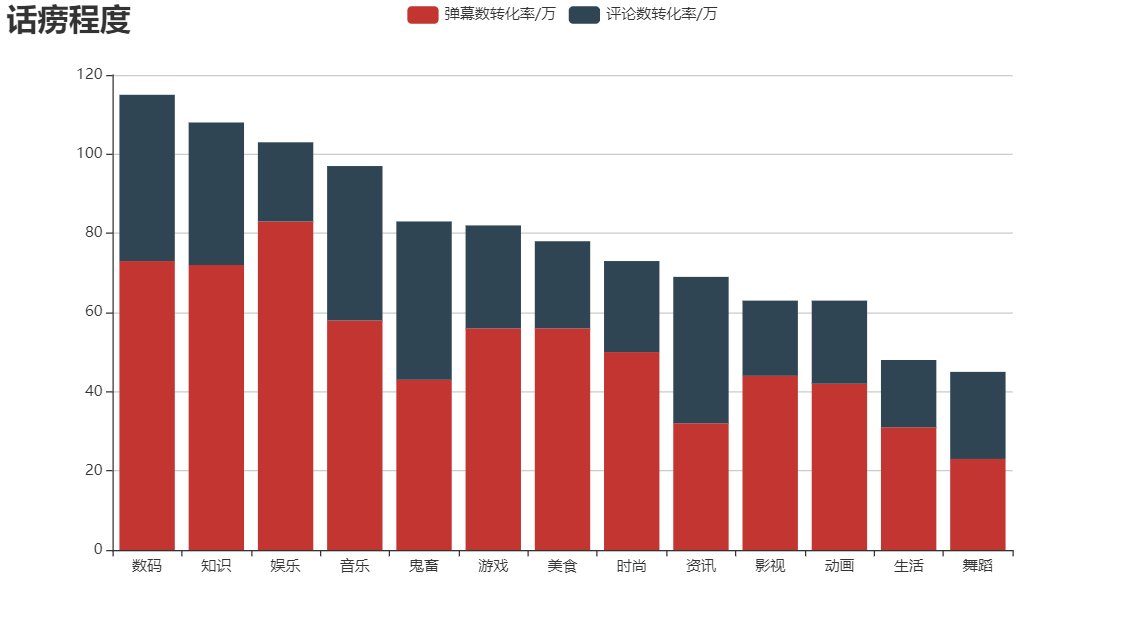


chart6-2