

YouTube 热门视频趋势与流量预测分析

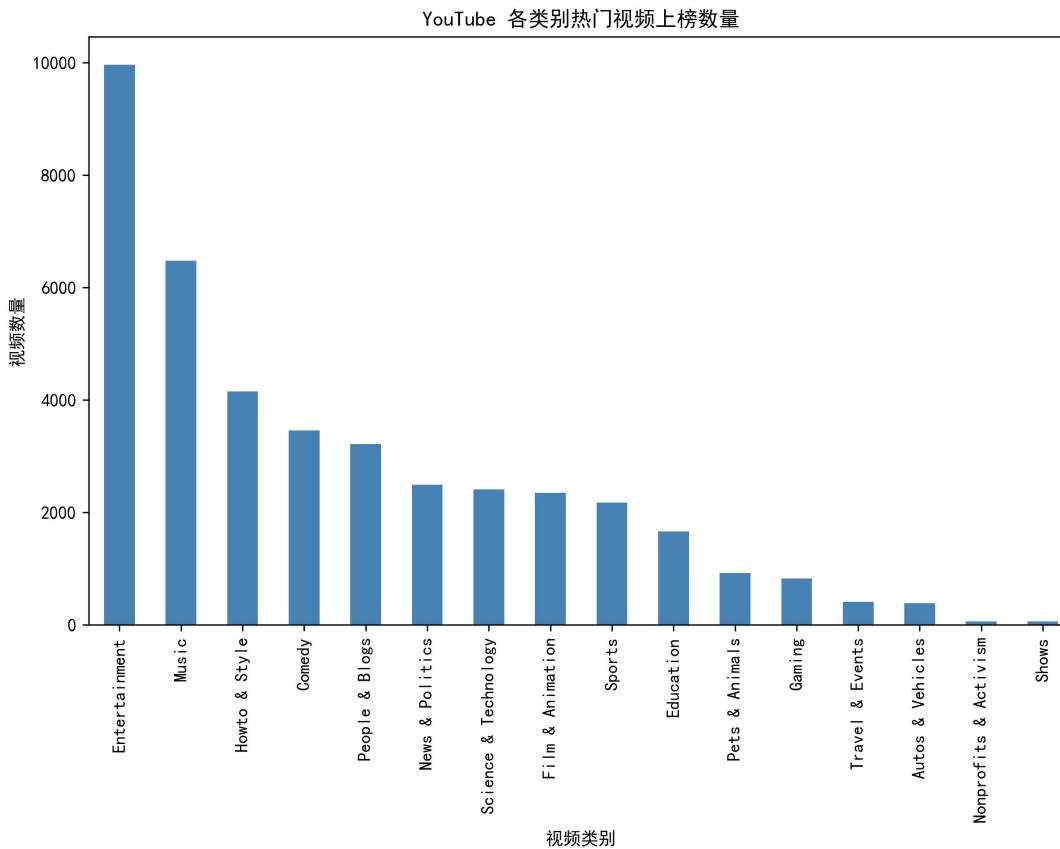
本项目是为了学习对 YouTube 热门视频数据的挖掘，识别不同内容分类的流量特征，并构建模型预测视频播放量，为内容创作者提供数据支持。

1. 数据处理

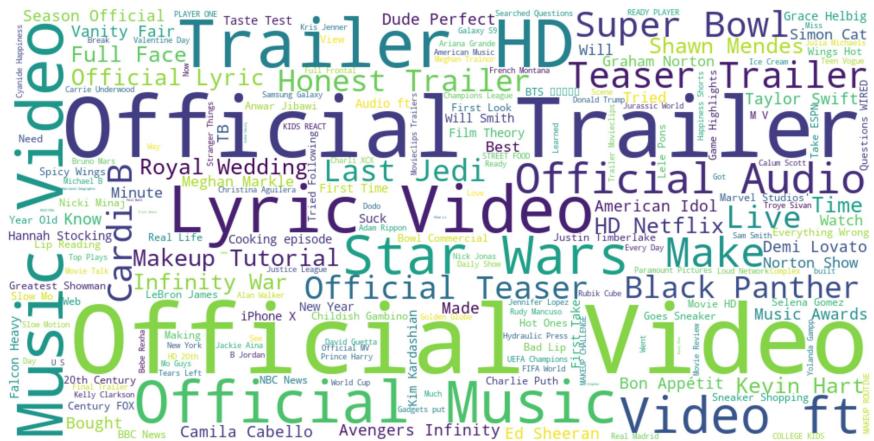
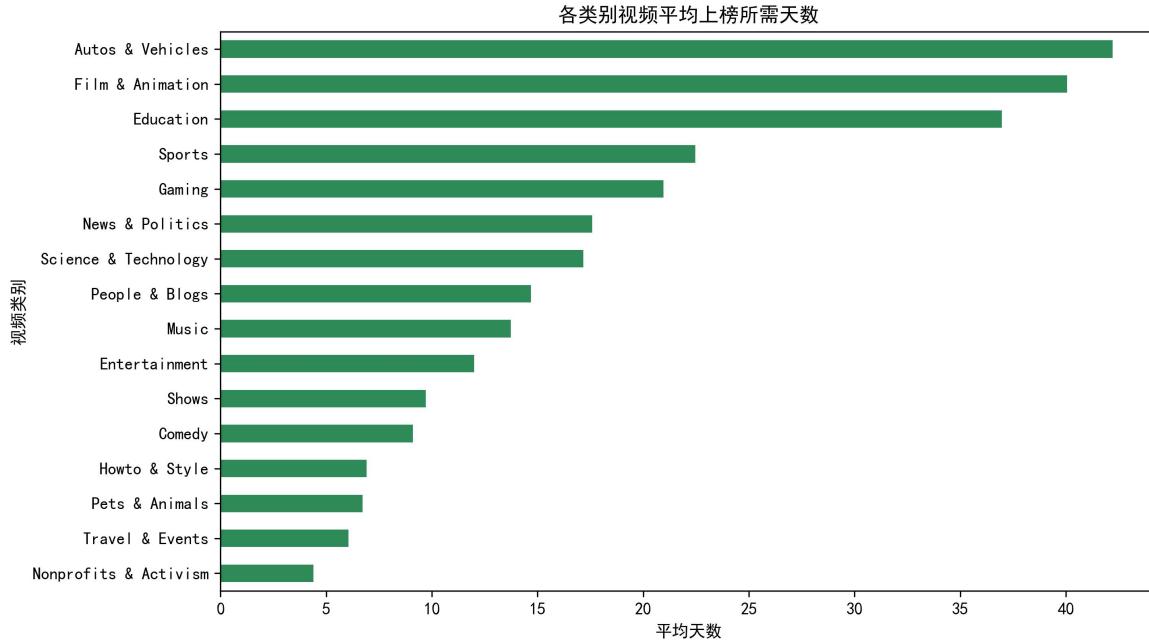
利用 Python 解析 JSON 格式的分类字典，并将其与 4 万余条 CSV 视频记录进行关联，实现了分类 ID 到业务语义（如 Music, Gaming）的映射。统一了日期格式，并计算出关键指标 `days_to_trend`（视频发布至登上热门的天数）。

2. 核心发现

Entertainment 和 Music 是上榜数量最多的类别，是平台的流量核心。新闻类视频上榜最快（平均 1-2 天），而汽车、电影类内容发酵周期较长（平均 40 天）。



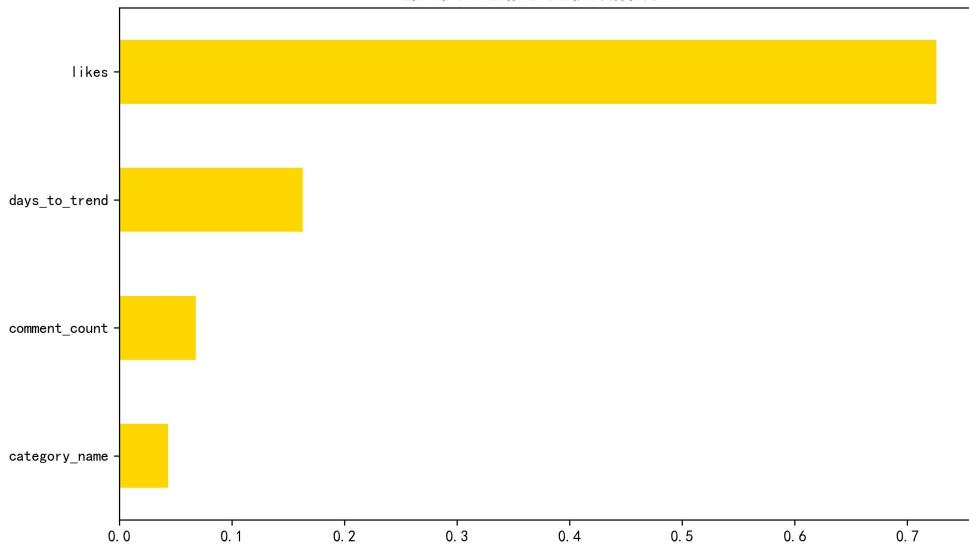
词云图显示，“Official”、“Video”、“Trailer”等词汇在爆款标题中高频出现，说明权威性标注对点击率有显著贡献。



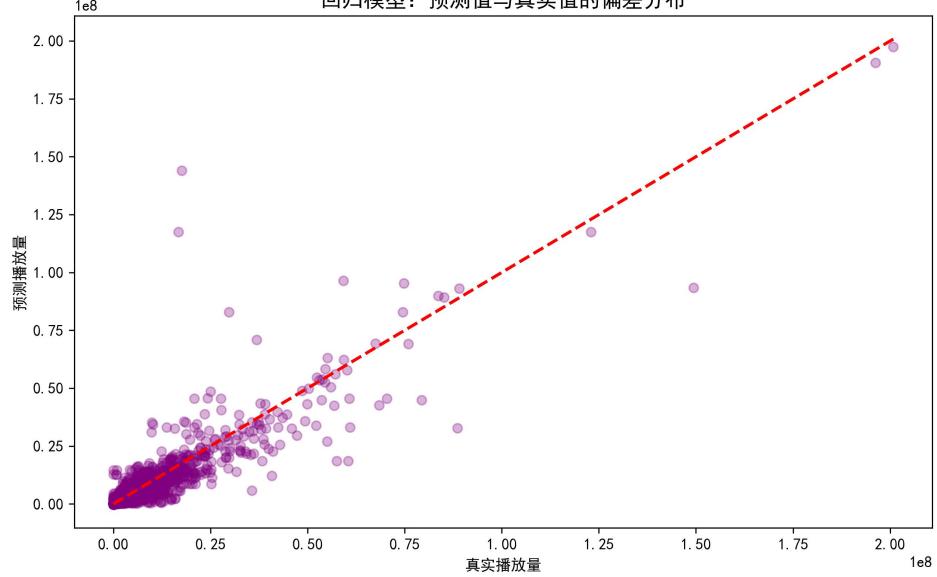
3. 流量预测模型

使用 XGBoost 回归算法，模型在验证集上的 R^2 达到了 0.8044。特征重要性显示，“点赞数”对播放量的解释力最强（权重> 0.7），远超评论数。

哪些因素最能决定视频播放量？



回归模型：预测值与真实值的偏差分布



4. 总结建议

建议博主在运营初期重点引导用户“点赞”，其对平台推流的贡献大于评论。针对不同品类（如教育 vs 新闻）设定不同的时效考核指标。