Predicting Youtube Video View Count Performance Based on Upload Timing

Jelica Bornath

Problem Definition

- What are we trying to solve?
  - Can a creator maximize their video's view count through strategic upload timing?
- What does success look like?
  - Establish if a video's performance can be predicted through historical video performance data on daily trending Youtube videos
- What is the scope?
  - Daily Trending Videos data from 2014, including channel level details like subscriber count, and video details like view count, video publishing time etc.

Key Outcomes

- **Goal** - determine if the video view count can be maximized through strategic upload timing based by Weekday
- **Outcome** - Incorporating video attributes like number of likes and dislikes in addition to upload timing show promising results, but Weekday and Subscriber count alone were not sufficient
- **Recommendation** - Select specific creator's channels from better performing video categories to identify trends by the individual creator

Data Collection - Attributes for Analysis

Channel Level Attributes	Video Level Attributes
Channel View Count	Video View Count
Video Count	Video Like Count
Subscriber Count	Video Dislike Count
Channel Elapsed Time	Video Elapsed Time
Channel Comment Count	Video Comment Count
	Video Published Date/Time

Target Variable

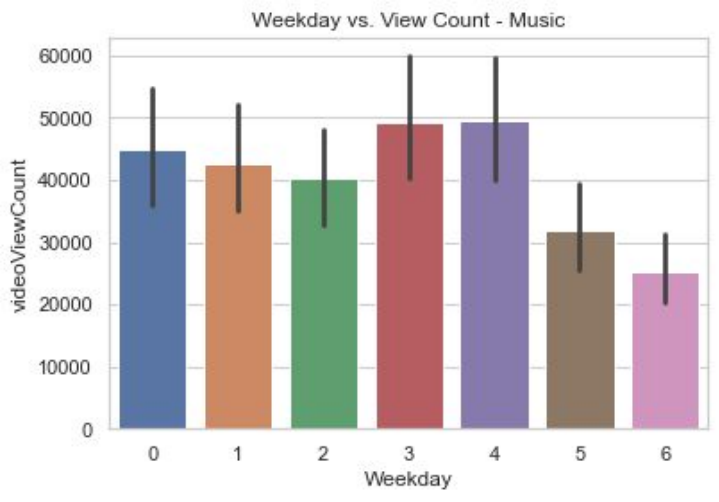
Feature used in modelling

Data Collection - Data

- 78,588 trending videos from 2014 used for analysis
- Target Feature : Video View Count
- 5 Video Categories Prioritized

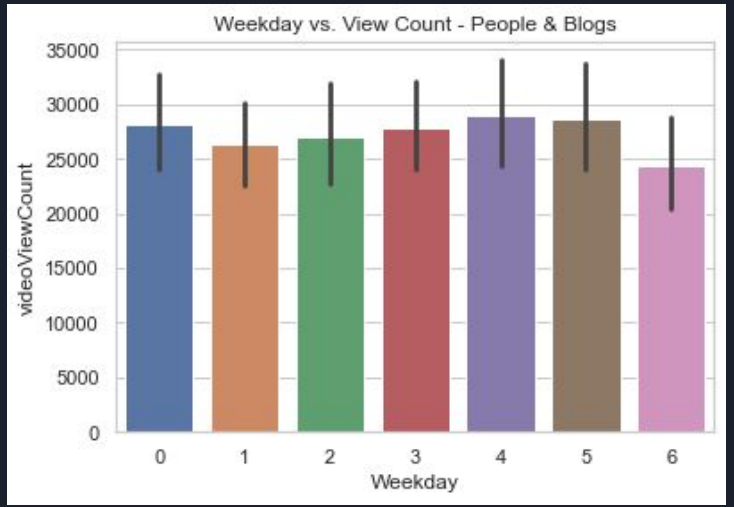
Video Category ID	Video Category Name	# of Videos
22	People & Blogs	16,795
20	Gaming	12,307
10	Music	11,052
24	Entertainment	8,667
17	Sports	6,507

Exploratory Data Analysis



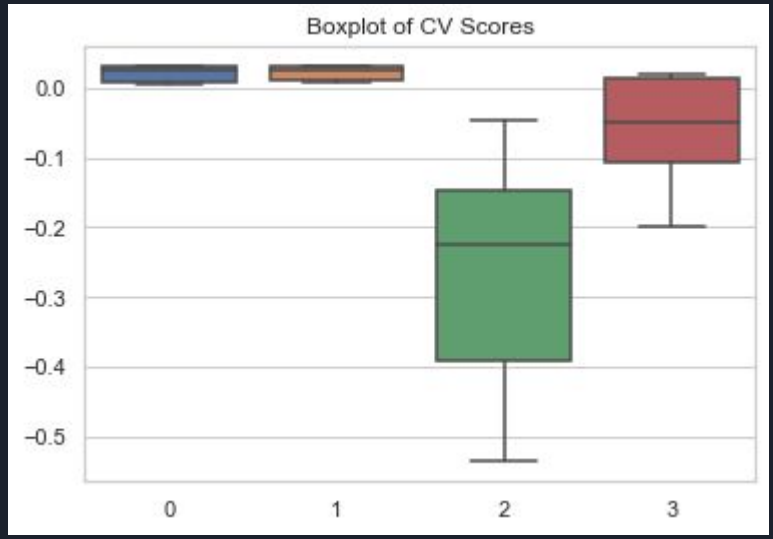
Music, Entertainment and Sports categories varied in performance across weekday

People & Blogs and Gaming showed little variation in performance across weekday.



Building the Regression Model

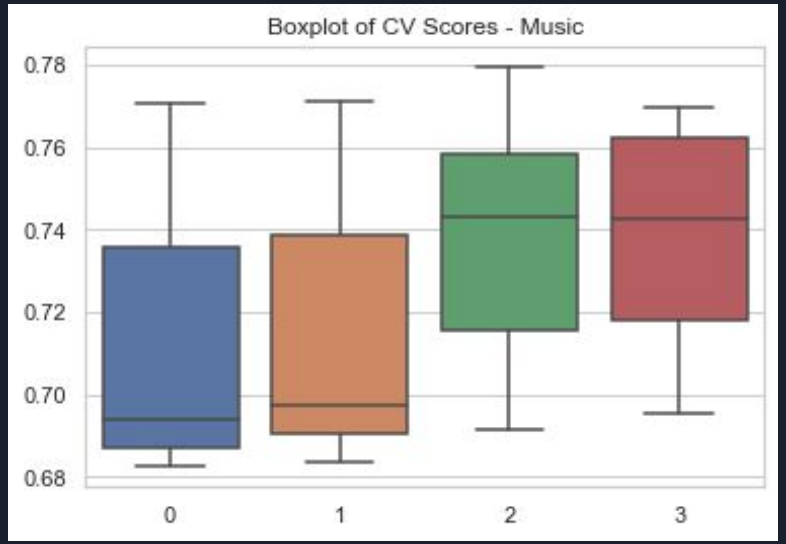
- The first attempt used Subscriber Count and Weekday of upload to predict view count
- These features achieved very poor CV scores across Linear Regression and Random Forest Models



Linear Regression (0), K Best Regression (1), Random Forest (2), Hypertuned Random Forest (3)

Improving & Selecting the Model

- Adding basic video performance metrics like Likes, Dislikes, Comment Count in addition to Subscriber Count and Weekday of upload improved results for specific categories like Music



Linear Regression (0), K Best Regression (1), Random Forest (2), Hypertuned Random Forest (3)

Recommendation

- Reco #1 - Further data analysis and modelling on specific channels from the categories indicating video performance variation by Weekday (Music, Entertainment, Sports etc.)
  - Establish if a model can predict a content creator’s best upload day based on their historical videos
- Reco #2 - Further analysis on categories of interest with Weekday as the target variable
  - Establish if a model can predict the best Weekday for upload by category, focus as a classification problem

Conclusion & Next Steps

- **Recap** - What does success look like?
  - Establish if a video's performance can be predicted through historical video performance data on daily trending Youtube videos
- **Conclusion** - Historical data could be used to determine a video's potential success, but the features required for it can only be collected after video is published (i.e. likes, dislikes, comment count) - upload timing alone is not enough
- **Next Steps** - analyze Music, Entertainment and Sports categories at a channel level, establish optimal timing by specific content creator