Youtube Statistics Analysis Project Proposal

Daphne Magsby

I. INTRODUCTION

As it has practically become inherent in our culture and society today, YouTube is a social platform that we all have heard of. YouTube is a video sharing website that allows users to upload, view, rate, share, favorite, report, comment on videos, and subscribe to other users. The website was founded in 2005 and was then acquired by Google in 2006 and is now run as a Google subsidiary. The slogan of the YouTube website is "Broadcast Yourself." This implies that the YouTube service is designed primarily for ordinary people who want to publish videos they have created. While several companies and organizations also use YouTube to promote their business, the vast majority of YouTube videos are created and uploaded by amateurs. Available content includes video clips, TV show clips, music videos, short and documentary films, audio recordings, movie trailers, live streams, and other content such as video blogging, short original videos, instructional videos, such as step-by-step, do-it-yourself guides, and other how-to and educational videos. Videos that have been uploaded to YouTube may appear on the YouTube website and can also be posted on other websites, though the files are hosted on the YouTube server. People also use YouTube as a profitable enterprise, since Google offers revenue sharing for advertisement clicks generated on video pages. This has led to a large community of YouTube content creators, and is the motivation behind this project.

II. MOTIVATION

As technology continues to take over causing more and more people to have the equipment needed to record and edit videos, more people turn toward YouTube and are becoming content creators. Having a break down of some the most sucessful creators on the platform could be beneficial. For example, Knowing the geographical region where creators are most likely to have success, the geographical region most of creator viewers will come from, having an idea of the type of content that will propel a channels success, and knowing a rough estimate of long it will take to reach specific milestones will be useful information.

III. DATA

I plan to gather information about the top YouTube channels such as the geographical location of each channel, the number of videos posted by that channel, the number of subscribers that channel has, the date that channel was formed, and the ratio of female to male viewers, as well as the percentage of each age group of the viewers. I will be excluding Vevo music channels because these channels include high budget music videos that are distributed by record labels and other channels that are associated with recording artists.

IV. TOOLS

To gather the information necessary for the analysis I will be using the SocialBlade api to scrape information from the site. SocialBlade is a website that has the primary purpose of tracking YouTube statistics. The site provides the number of uploads, subscribers, country, channel type, subscriber growth by day and month, and the date created for the channel. It also provides each video that channel has uploaded along with the date uploaded, number of views, number of comments, and number of likes. The site gathers its information from YouTube's public API and then examines the data across multiple days and aggregates it into a user frienldly display format. According to the YouTube API Services Terms of Service, it would be a contract violation to scrape YouTube for this data. Since SocialBlade has a compiled list of YouTube statistics it will be easier to scrape than YouTube itself. The scraping of the data will be done using Python. I will be storing the data using Amazon's DynomoDB database.

V. ANALYSIS

I will be using the gathered information to try and draw correlations between subscriber count and the number of lifetime views as well as individual video views. Correlation between the longevity of the channel and the subscribers. Can relatively newer channels subscriber count outgrow older channels based on the type of channel it is/ the type of videos that are posted.

VI. EXPECTED OUTCOME

I expect to be able to find a connection between the age of a channel, the number of videos that channel has uploaded, and the amount of subscribers that channel has. I also expect to be able to find a connection between the type pf channel/videos that and the amount of views.

VII. TIMELINE

- Week 1:
  Finalize criteria for analysis and tools used to gather data

- Weeks 2, 3, & 4:
  Begin gathering and storing data

TABLE I
8 Weeks

Week 1	10/9 - 10/15
Week 2	10/16 - 10/22
Week 3	10/23 - 10/29
Week 4	10/30 - 11/5
Week 5	11/6 - 11/12
Week 6	11/13 - 11/19
Week 7	11/20 - 11/26
Week 8	11/27 - 12/3

- Weeks 5, 6 & 7:
  Begin creating meaningful relations

- Week 8
  Finalize results

REFERENCES

[1] Christensson, Per. "YouTube Definition." TechTerms. Sharpened Productions, 07 October 2009. Web. 02 October 2017. ¡https://techterms.com/definition/youtube¿.
[2] Urgo. Frequently Asked Questions (FAQ). Social BLADE / Socialblade, 2017. Web. 02 October 2017. ¡socialblade.com/youtube/help¿.