

CS 6460 Project Paper

Benjamin Wang

bwang421@gatech.edu

Abstract—

The internet has ushered in an era where information on any topic can be found through a simple web search. Despite the ease access for most people to find information, most of this information is in reading format thus there is a need for lessons and concepts to be made available on more modern channels such as video and audio platforms. Adding more layers and depth to how information is presented is a double edged sword which can either provide more control to the instructor and better understanding to the listener when properly used, or only add to the confusion due to poor design or execution. There has been much research done and data collected in what works best for online learning, however these papers focused around video content understandably source their data from formal learning settings such as online course websites such as [edX](#), [Udacity](#), and [Coursera](#). There is little research when it comes to what works best at achieving success in an informal learning setting like YouTube. In this paper, I will discuss my project which cover this gap by collecting data through video creation on topics related to data science.

1 INTRODUCTION

For this project, I wanted to create a video series to fill gaps in the area of data science on the video platform YouTube. The focus of my efforts were around creating and editing videos that ranged from hands-on coding tutorials to informal hand-drawn illustrations meant to tie theoretical concepts with real-world examples.

I created my video series utilizing features and ideas highlighted in research papers I found on the topic of online learning, and experimented with a mix of different methods to try to see which ones would net the best results.

Despite the lower than desired traffic to the channel, the like to dislike ratio on the channel indicated that viewers found the information and presentation

helpful.

The remainder of the paper is organized as follows: Section 2 discusses related works that inspired some of the work done on this project, Section 3 talks about the methodology and tools that were chosen for content creation, Section 4 covers the content that was created while the discussion on success metrics, challenges, and analytics are covered in Section 5. Concluding comments and thoughts then follow in section 6.

2 RELATED WORK

2.1 Khan Academy

Khan Academy is an American non-profit educational organization created in 2008 by Sal Khan, with the goal of creating a set of online tools that help educate students (Noer, 2012). The organization produces short lessons in the form of videos. Its website also includes supplementary practice exercises and materials for educators. All resources are available for free to users of the website and app (Martí, 2015)

Khan Academy popularized the electronic blackboard-style of teaching along with using YouTube as a learning platform. As of writing this paper, the [YouTube channel](#) currently has 5.94 million subscribers and over 7,800 videos uploaded on the channel.

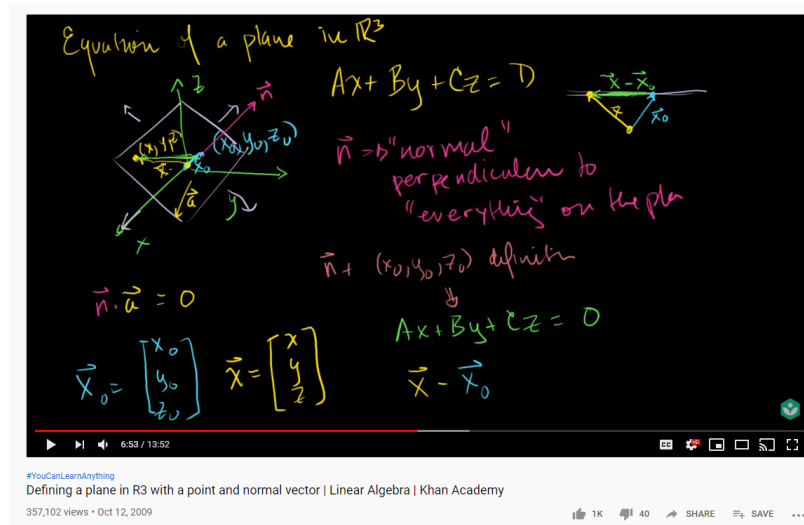


Figure 1—An example of the tablet-style video tutorials that were popularized by Khan Academy

Besides the teaching style, Khan employs other popular methods to keep users engaged such as gamification in the form of earning badges (Antonaci, Klemke, and Specht, 2019)

The majority of Khan's content lives on YouTube, but its course structure and supplemental material is maintained through the Khan Academy website.

2.2 Coursera, edX, and Udacity

Coursera, edX, and Udacity were one of the first providers of massive open on-line courses founded in 2012.

These providers host a wide range of courses and subjects from different instructors, but their content style still follows the same framework of instructor-led lectures, with a mix of presentation styles such as the khan-blackboard drawings, screen-sharing hands-on examples, power point presentations, animations, and such depending on the instructor leading the class. The sources cited in this paper studying the efficacy of online learning and video content primarily sourced their data from one of these three MOOC providers. (Guo, Kim, and Rubin, Kim et al., Hone and Said, 2014, 2014, 2016)

3 METHODOLOGY

The production of my video content was centered around two areas which I will cover: First, was the design and delivery of each video, and second was the tools used for video recording and editing.

3.1 Design and Delivery

My research and readings highlighted many pedagogies in educational video creation that I could use (ralson4, Guo, Kim, and Rubin, Kim et al., Levasseur and Sawyer, ralsom4, 2014, 2014, 2006). In order to narrow down this list, I manually selected features to focus on including in my videos which I thought would be better suited for the task of engaging viewers on YouTube:

1. Video Editing
2. Video Length
3. Variety in presentation mediums
4. Personalizing the content

The features above were selected because they were commonly found in class-

room videos which had the highest viewer retention, lowest dropouts, and highest completion rates.

Thus the initial effort was made to create content that fit the framework of having a short length, was well edited and cut together to maintain a fast and engaging pace, utilized multiple presentation styles such as screen sharing, power point, and blackboard drawings, and also added a personal style to videos to build rapport with viewers.

3.2 Tools and Environment

I initially began recording videos in the room of my apartment, as research supported the idea that investing heavily in a studio setup did not have a meaningful impact on viewer retention and engagement (Kim et al., 2014)

I used the following hardware to record content:

1. Windows 10 Desktop
2. Logitech C920s Pro HD Webcam
3. Monoprice 10 x 6.25-inch Graphic Drawing Tablet
4. Samson Meteor Mic USB Studio Microphone

And the following software was used for capturing video and editing:

1. OBS Studio (Video Recording)
2. DaVinci Resolve 16 (Video Editing)

Using OBS Studio involved an initial hurdle of testing and finding the best settings for frame rate, filters, video, and sound, and then creating and saving scenes for window capture.

DaVinci Resolve 16 was chosen because it was a robust and free software to use for video editing. With that said, video editing is a very involved process and despite Resolve having the same capabilities as a paid software, I found it came with a trade-off in terms of performance and ease-of-use.

To elaborate further, Resolve is a huge memory hog. Even for my desktop with an overclocked intel i7, 32 GB RAM, and an Nvidia Geforce 1080 GTX GPU, switching between panels, cropping and cutting video scenes, and other common tasks caused a 'hanging' to occur which only became worse over time and required a full restart on the application to restore performance.

4 CONTENT

4.1 Learn, Install, and Use Anaconda for Data Science in 5 minutes!

[Video Link](#)

This video walks the viewer through the steps in getting Anaconda installed on their computer and running a jupyter notebook with some simple coding samples.

Despite the simplicity of being an intro video, this video incorporated almost all of the features highlighted to increase viewer retention and engagement: (Guo, Kim, and Rubin, 2014) (Kim et al., 2014)

1. Pre-written script to improve the flow of the lesson and video.
2. Short video length of under 10 minutes to reduce viewer dropout
3. Facial camera and introduction to create a personalized feel with the viewer
4. A mix of presentation mediums to keep viewers engaged (PowerPoint, screen share, and hands-on examples)
5. Large video editing to cut to highlighted points for instruction.

4.2 How to use Regular Expressions in Excel VBA!

[Video Link](#)

I took the framework from my first introductory video and tried to employ the same features to a real tutorial covering areas that I could not find on YouTube. The idea was that because this video was going over an uncovered topic, the traffic would be much higher than the first and be a good data point for the project's study.

The next two videos below followed this same format:

4.3 Understanding the SQL Join and Union Operation

[Video Link](#)

This video incorporates the drawing board and colored excel spreadsheet to visually illustrate how the SQL join and union operator work on data.

4.4 Beginner's Guide to SQL Queries with Examples

[Video Link](#)

This video provides a detailed introduction to people with no SQL experience on how queries are written, and uses an excel spreadsheet with color coding to illustrate examples of various query operations.

4.5 Hypothesis Testing, Error Types, and Likelihood Explained!

[Video Link](#)

In this video, I diverged from my previous video formats to try to experiment and see if I could get better results from previous videos. The reason for this was because previously created videos were receiving very low views and no feedback from viewers.

In this video, I did not incorporate a shorter video span, and I used a drawing board for the entire video.

4.6 Gentle Introduction to VBA in Excel

[Video Link](#)

Based on feedback I received from viewers on previous videos, I incorporated more video editing methods such as zoom and panning. One of the issues I did not predict was that my video might be viewed on smaller screen devices such as phones, so the text and drawings I created would be very difficult to view and follow.

The video itself provides viewers with a walkthrough and some examples on how to get started with writing their own macros in VBA and some tips and advice on how to find info on things like environment variables built into VBA.

4.7 Understanding what a matrix determinant is

[Video Link](#)

This video provides users with an understanding of what a matrix determinant represents in the physical world when a matrix multiplication occurs, such as increase in size, reduction in dimensions, and such.

4.8 Beginner's Guide to Getting Started with Image Files in Python!

[Video Link](#)

This video format follows a hands-on screen sharing tutorial introducing the

viewer to the OpenCV library in python by covering syntax for basic expressions and operations involving images such as reading and writing, splitting color channels, and changing size and orientation.

5 RESULTS

The aggregated data for each video can be found in the following spreadsheet:

[Video Spreadsheet](#)

5.1 Success Metrics

5.2 Channel Data

A big issue for this project was the lack of data in the form of view traffic.

In the span of 6 months since the channel was opened and the videos published, my project channel received 135 unique views from 373 impressions with an 11.5% click-through rate.

With the little views data there was, the information indicated positive reception such as averaging 50% viewer for videos a quarter of the way in, and 25% for videos 3/4 of the way completed. The channel received 5 likes, 0 dislikes, and 8 subscribers. Presuming these metrics such as viewer retention and views to subscriber ratio could scale out and remain close to the current sample, this project could be considered a success as far as viewer engagement is concerned (Kim et al., Guo, Kim, and Rubin, 2014, 2014)

5.3 'Best' videos

3 of the 10 videos I created accounted for roughly 70% of the views on the channel. Interestingly, all 3 of the videos each employed a separate framework.

Video 1

The first video "Learn, Install, and Use Anaconda for Data Science in 5 minutes!" was the shortest at 5 minutes and 52 seconds, but also one of the most detailed. I went beyond showing how to simply install the program, and gave a high-level explanation of what Anaconda is, why it is relevant to data science, and explained how to run and use jupyter notebook.

This video could have been dramatically shortened, and was one of the reasons

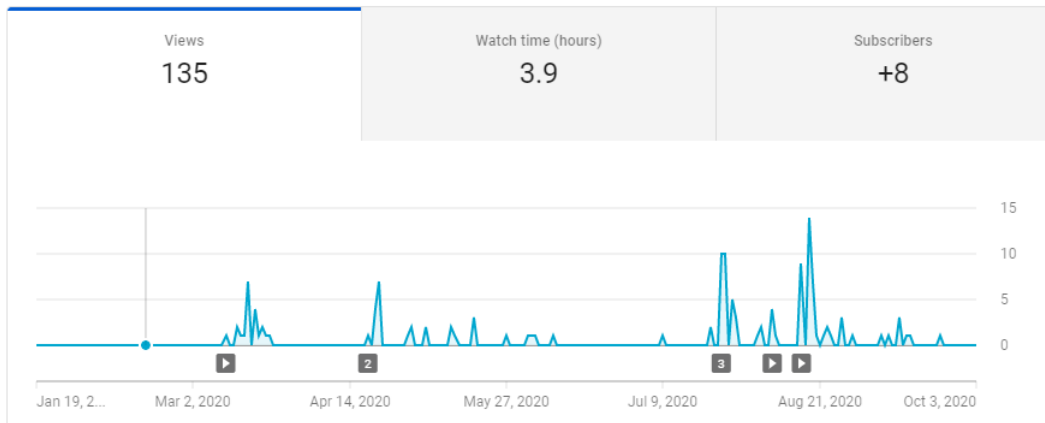


Figure 2—Channel Views from March 2020 to September 2020

Impressions and how they led to watch time

Data available Jan 19 – Oct 3, 2020 (259 days)

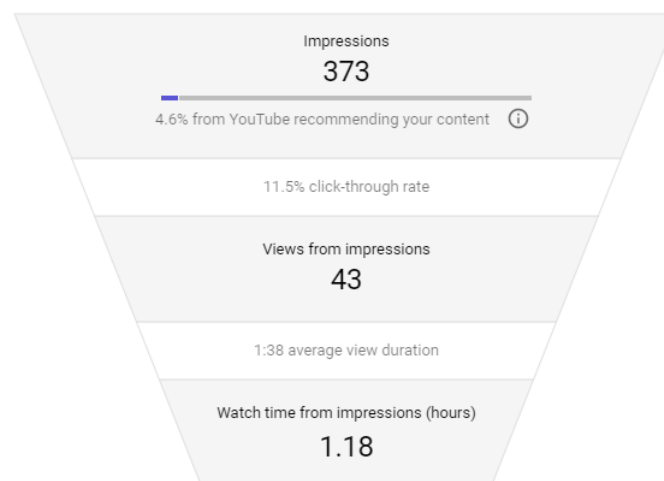


Figure 3—Impressions from March 2020 to September 2020

Traffic source types ▲

Views · Lifetime

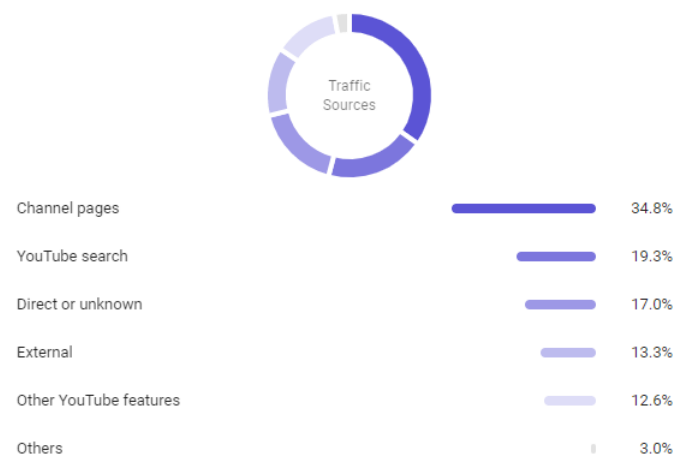


Figure 4—Sources of traffic for the channel

I tried to cut the fluff of explaining and jump right into example in my other videos.

5.3.1 Video 2

The second video "How to use Regular Expressions in Excel VBA!" was longer at 9 minutes and 12 seconds, and followed a similar format of the first video. There was an introduction and a review of regular expressions which took up the first 1:53 of the video.

The remaining time was spent explaining how to import the regular expressions library for VBA, and showing multiple examples with increasing complexity.

5.3.2 Video 3

The longest video "Hypothesis Testing, Error Types, and Likelihood Explained!" runs at 19 minutes and 47 seconds and is presented entirely on an electronic blackboard.

Despite being one of the most viewed videos, the viewer retention was poor at 8% retention by the 75% completion point. This can be attributed primarily to video length. The original idea behind this video was to explain and tie in all of the topics of the title together to help the viewer understand how everything was related to one another, however it would have probably been better to split apart each topic into a separate video.

5.4 Challenges

Initially starting this project, I presumed that lesson design and planning was going to be the most difficult and time consuming portion of the project. While it did take up a large part of my time, I think the largest amount of time was spent on recording and editing video.

5.5 Sound Noise and Video Editing

I underestimated the difficulty involved in the process of video editing, and recording. During recording sessions, the quarantine caused my entire apartment complex and surrounding neighborhood to stay at home. There was a large amount of background noise in many videos I had, causing me to delete and restart the recording process while also trying to find opportune windows for recording.

I eventually chose to move to a quieter complex which allowed me to dramatically improve the production quality and ease of producing videos.

5.6 Low Traffic

I presumed that the videos I created would be in niche areas which I felt did not provide adequate information on the topics I was covering.

Despite this, or perhaps because of it, the traffic to these videos has been low and as a result I have very little viewer data to draw meaningful conclusions.

The small amount of data I do have indicates a positive reception, with all videos receiving only likes and no dislikes.

The production issues I had with my videos were resolved by changing my recording setting. The traffic on my initial videos incorporating the features highlighted in (Guo, Kim, and Rubin, 2014) and (Kim et al., 2014) performed at about the same level as my other videos which all had relatively low traffic.

6 CONCLUSION

What can be drawn from this is that the features mentioned (Guo, Kim, and Rubin, Kim et al., 2014, 2014) might be key in keeping audience retention for a classroom setting, but for a more casual environment like YouTube, the first challenge in building a successful project is driving viewers to the channel to view the content.

For future attempts at producing educational videos on YouTube or other entertainment platforms, more research needs to be done on understanding viewer behavior, how to optimize keywords, title, pop-up screens, and banners for maximizing views, impressions, and click-through rate.

The YouTube search algorithm also needs to be taken into consideration for collecting data (PETERSEN, 2019). Content creation techniques cannot be thoroughly vetted and tested without an adequate sample size of traffic and viewer behaviors. Artificially driving traffic to a channel or video through self-promotion and advertising would add a level of bias in the data that would be difficult to account for, thus the ideal scenario would be to find an optimal formula which rewards the channel and videos with a higher rank in the search.

Another area to consider is the topics being covered. The materials I reviewed

were for a small target audience where other videos in the same area seldom garner more than 10k views over a span of years, due to this fact, it should have been expected that viewership would have been low. Thus, when starting a channel it seems wise to create content on a consistent schedule, covering topics relevant to as wide reaching of an audience as possible to build the initial momentum necessary to earn rank in search.

The scripts and materials I produced are located at the following repository: [here](#)

7 REFERENCES

- [1] Antonaci, A., Klemke, R., and Specht, M. (2019). "The Effects of Gamification in Online Learning Environments: A Systematic Literature Review." In:
- [2] Guo, Philip, Kim, Juho, and Rubin, Rob (2014). "How video production affects student engagement: An empirical study of MOOC videos". In: pp. 41–50. DOI: [10.1145/2556325.2566239](https://doi.org/10.1145/2556325.2566239).
- [3] Hone, Kate S and Said, Ghada R El (2016). "Exploring the factors affecting MOOC retention: A survey study". In: *Computers & Education* 98, pp. 157–168. ISSN: 0360-1315. DOI: <https://doi.org/10.1016/j.compedu.2016.03.016>. URL: <http://www.sciencedirect.com/science/article/pii/S0360131516300793>.
- [4] Kim, Juho, Guo, Philip J, Seaton, Daniel T, Mitros, Piotr, Gajos, Krzysztof Z, and Miller, Robert C (2014). "Understanding In-Video Dropouts and Interaction Peaks In online Lecture Videos". In: *Proceedings of the First ACM Conference on Learning @ Scale Conference*. L@S '14. New York, NY, USA: Association for Computing Machinery, pp. 31–40. ISBN: 9781450326698. DOI: [10.1145/2556325.2566237](https://doi.org/10.1145/2556325.2566237). URL: <https://doi.org/10.1145/2556325.2566237>.
- [5] Levasseur, David and Sawyer, Kanan (2006). "Pedagogy Meets PowerPoint: A Research Review of the Effects of Computer-Generated Slides in the Classroom". In: *Review of Communication* 6, pp. 101–123. DOI: [10.1080/15358590600763383](https://doi.org/10.1080/15358590600763383).
- [6] Martí, Laura (Oct. 2015). "Digital systems for open access to formal and informal learning". In: *International Journal of Educational Psychology* 4, p. 329. DOI: [10.17583/ijep.2015.1720](https://doi.org/10.17583/ijep.2015.1720).
- [7] Noer, Michael (2012). *One Man, One Computer, 10 Million Students: How Khan Academy Is Reinventing Education*. <https://www.forbes.com/sites/michaelnoer/2012/11/02/one-man-one-computer-10-million-students-how-khan-academy-is-reinventing-education/#5a94949444e0>.

- [8] PETERSEN, Emily (2019). *My videos are at the mercy of YouTube algorithm”: how content creators craft algorithmic personas and perceive the algorithm that dictates their work*. Tech. rep. Technical Report No. UCB/EECS-2019-48, Electrical Engineering and Computer ...