

CS 6460 Project Proposal

Benjamin Wang

bwang421@gatech.edu

1 OVERVIEW

The main focus of my project will be centered around creating hands-on tutorial videos on YouTube to teach beginners how to accomplish tasks working with data from setting up their environment (Anaconda), explaining the purpose of tools and libraries (Excel Solver, Numpy, Pandas, Sklearn), and demonstrating through examples how to manipulate, filter, and clean data which will primarily be in CSV.

If there is time, I would also like to touch on image processing, and machine learning concepts such as neural networks, clustering, and PCA/SVD, but I believe the audience for these topic is much smaller so it is lower priority to be completed.

I will be employing various techniques I've learned about throughout the course in order to improve knowledge transfer, viewer retention, and learner outcomes.

The key methods I will be using are listed below:

1. Aim for videos to be no longer than 6 minutes in length through lesson planning and editing. (Guo, Kim, and Rubin, 2014)
2. Provide a mix of media throughout the video switching between a facecam to talk directly to the viewer, powerpoint presentations explaining , hand-drawn examples, and screen casts for coding. (Hone and Said, 2016)
3. Encourage communication between viewers through the comments section to try to build a sense of community and distributed teaching. (Picciano, 2002)
4. Encourage comments to also provide feedback in terms of what the viewers liked and disliked, and what they felt was not explained clearly. (Kim et al., 2014)
5. Help reinforce the lesson with a question at the end of the video pertaining to what was taught. (Ralston-Berg and Nath, 2008)

2 RELATED WORK

Educational channels on YouTube have the same premise as what I am aiming to accomplish. Some of these channels like [MATLAB](#) and [deeplearning.ai](#)

follow a more formal powerpoint-esque presentation format (Levasseur and Sawyer, 2006) while others use a less formal hand-drawn teaching style reminiscent of school teachers using overhead projectors like [Khan Academy](#). There are channels featuring recorded college lectures such as [MIT OpenCourseWare](#), and strictly spoken over animations like [Osmosis](#) (Tackett et al., 2018). It is not hard to see that there is already a huge saturation of sources for educational videos on YouTube with a large actively engaged user base.

What will my content contribute?

I've felt that despite the large amount of tutorials, many of them are not designed or optimized to be digested through a platform like YouTube, they are lacking in hands-on demonstrations, feed information through one channel, are poorly edited and unnecessarily long, and explanation of concepts and tools are glossed over leaving the viewer with a poor high-level understanding on how to splash the knowledge outside of what the video is showing.

I hope to provide content which tries to address all of these issues by utilizing the methods I've researched mentioned above.

3 CONTENT DESCRIPTION

My videos will have two tracks involving work being done on Python and Microsoft Excel, these videos will aim to be short and to the point (< 5 min), and completely self-contained so that learning a task from another video is not necessary, or if it is, a reference can easily be made tagging the necessary video.

By self-contained, I mean someone can simply follow along and copy the instructor in order to accomplish a task. This means I will not be referencing steps assuming the viewer is already familiar i.e. "Load the last video's dataset and run the function that we wrote for splitting apart the records." is the type of dialogue I will try my best to avoid even if it means having some redundancy of steps covered in some videos.

I will be covering basic tasks at first such as environment setup, string manipulation and regular expressions, and generating visuals for data. I will also have videos providing high level overviews of topics such as when Excel is a good tool to use for analysis versus python, how to determine what visuals to choose to best represent your results for others to understand easily, and selection of data

types.

Once I create these foundational videos, I will upload them to collect feedback from users and classmates. Based on the feedback, I will determine if I will continue building more content, improve the existing content, or supplement the content with technical videos to explain concepts such as data structures and algorithms.

4 DELIVERABLES

I will provide a YouTube channel hosting these videos, and organize videos into a playlist. My long-term goal is to have videos building on top of other videos such as **'How to scrap Yelp reviews and create your own recommendation system from scratch.'**, but due to the time frame of the class, I would be happy to at least complete all of the baseline videos covering basic skills to build familiarity.

4.1 Milestone 1

For milestone #1, I will provide a link to the finalized and uploaded versions of the YouTube videos for my first 3 task items: Installing and Running Anaconda, Pros and Cons of Excel vs Python, and Why VBA is worth learning for analysis.

4.2 Milestone 2

Milestone #2 will consist of my next outline items, presuming things are on schedule, I will deliver links to items 4-9 outlined in the task list below, or whatever videos are completed if I record and edit out of order.

4.3 Final Project Deliverable

I will submit a link to a playlist of all recorded videos that have been uploaded to YouTube, and also provide a report detailing the analytics collected from the videos, feedback from viewers, highlighted comments, and my own analysis of the results collected from implementing the ideas that were researched. The video presentation will provide an outline of all the items mentioned as well and highlight parts of my videos showing how different aspects from papers like (Guo, Kim, and Rubin, 2014) (Kim et al., 2014) were incorporated to maximize viewer retention and reduce viewer dropout.

5 CONTENT OUTLINE AND TASK LIST

5.1 Outline

What I lay out here is not set in stone, but I hope to see the content I create follow the general framework laid out here:

1. Introductory video - Explaining channel content and format
2. Installing and Running Anaconda - Walkthrough explaining Anaconda and how to install it.
3. Pros and Cons: Excel versus Python for data analysis
4. VLOOKUP vs VBA: Why you should learn to use VBA - Overview of what the VLOOKUP function does, and why it is important to learn how to use VBA.
5. Dictionaries and Lists explained
6. Explaining numpy, pandas, and sklearn - Overview of what each library specializes in.
7. String Manipulation, Search, and Regular Expressions with python
8. String Manipulation, Search, and Regular Expressions with Excel VBA
9. Working with CSVs in python
10. Quick introduction to VBA - Overview of VBA, common commands, and internal variables.
11. Visualizing and graphing data, and why it is important - High level explanation of the business need to create visual summaries of analysis.
12. Graphing data in python - Introducing matplotlib
13. Graphing data in Excel VBA - Introducing VBA commands to automatically generate visuals
14. (Optional) Explaining Solver in Excel - High level overview of the solver add-in and what it can be used for.
15. (Optional) How to use Solver with Excel VBA
16. (Optional) Binary Search with Excel VBA
17. (Optional) Any time remaining: Cover SQL, logistic regression, k-means clustering, PCA/SVD, and CNN.

Between each of these big topics, I will examine the feedback I receive and see where additional technical discussions can be made to explain what is going on 'under-the-hood' and behind the scenes to give learners a better understanding and intuition of the materials so they can go out on their own to accomplish tasks.

I also hope to provide small projects and quizzes at the end of each video to test their learning.

Below is the schedule I hope to follow, I've modeled each week as a 'Check-list' of the tasks I'm aiming to complete:

Week #	Task #	Description	Est Hours	Member
7	1	Research video editing	5	Benjamin
7	2	Record #1: Introduction Video	1	Benjamin
7	3	Record #2: Install and setup Anaconda	1	Benjamin
7	4	Record #3: Pros and Cons of Excel vs Python	1	Benjamin
7	5	Edit #1	1	Benjamin
7	6	Edit #2	1	Benjamin
7	7	Edit #3	1	Benjamin
7	8	Upload items #1,#2,#3	0.1	Benjamin
7	9	Weekly Status Check	0.25	Benjamin
8	10	Record #4: VLOOKUP vs VBA	1	Benjamin
8	11	Record #5: Dictionaries and Lists explained	1	Benjamin
8	12	Record #6: Numpy, Pandas, and Sklearn explained	3	Benjamin
8	13	Edit #4	1	Benjamin
8	14	Edit #5	1	Benjamin
8	15	Edit #6	2	Benjamin
8	16	Upload #4,#5,#6	0.25	Benjamin
8	17	Deliver YouTube links to items #1,#2,#3 for intermediate milestone #1	0.25	Benjamin
8	18	Weekly Status Check	0.25	Benjamin
9	19	Record #7: String Manipulation, Search, and Regular Expressions with Python	2	Benjamin
9	20	Record #8: String Manipulation, Search, and Regular Expressions with Excel	2	Benjamin
9	21	Edit #7	2	Benjamin
9	22	Edit #8	2	Benjamin
9	23	Upload #7,#8 to YouTube	.25	Benjamin
9	24	Request class feedback for items #1-#6	0.25	Benjamin
9	25	Examine YouTube analytics for trends	3	Benjamin
9	26	Weekly Status Check	0.25	Benjamin

Week #	Task #	Description	Est Hours	Member
10	27a	If good feedback: Record/Edit #9: Working with CSVs in Python	4	Benjamin
10	28a	If good feedback: Record/Edit #10: Introduction to VBA in Excel	4	Benjamin
10	27b	If bad feedback: Re-record/Edit previous item #s	15	Benjamin
10	28b	Collect feedback again on redone materials	0.25	Benjamin
10	29	Upload content to YouTube	0.25	Benjamin
10	30	Collect and document viewer feedback/comments	2	Benjamin
10	31	Weekly Status Check	0.25	Benjamin
11	32	Record #11: Why visualizations are important	1	Benjamin
11	33	Record #12: Graphing/Visualizing Data in Python	2	Benjamin
11	34	Record #13: Graphing/Visualizing Data in Excel	2	Benjamin
11	35	Edit #11	1	Benjamin
11	36	Edit #12	2	Benjamin
11	37	Edit #13	2	Benjamin
11	38	Upload #11,#12,#13 to YouTube	0.25	Benjamin
11	39	Provide links for items #4-#13 for intermediate milestone #2	0.25	Benjamin
11	40	Weekly Status Check	0.25	Benjamin
12	41	Examine viewer feedback and identify areas that need clarification	5	Benjamin
12	42	Embed supplementary materials in description and playback for past content	5	Benjamin
12	43	Look through YouTube analytics	2	Benjamin
12	44	Weekly Status Check	0.25	Benjamin

Week #	Task #	Description	Est Hours	Member
13	45	Record any optional outline items based on viewer feedback	4	Benjamin
13	46	Edit optional outline items	4	Benjamin
13	47	Upload optional items to YouTube	0.25	Benjamin
13	48	Begin writing final report	2	Benjamin
13	49	Record project presentation video	3	Benjamin
13	50	Edit project presentation video	2	Benjamin
13	51	Weekly Status Check	0.25	Benjamin
14	52	Complete final project report	8	Benjamin
14	53	Submit final project, presentation, and paper	0.25	Benjamin

Total Est Hours
105.35

6 CONCLUSION

With my main outline and objective decided, the next step is for me to obtain the necessary video editing knowledge to become familiar with the work so that I can efficiently churn out the content in a short period of time and gather feedback.

I will be spending the week learning how to use my recording software, recording multiple channels to different files i.e. screen cast, web cam, and tablet drawing, then learning how to edit them all into one video.

7 REFERENCES

- [1] 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).
- [2] 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>.
- [3] 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. URL: <https://doi.org/10.1145/2556325.2566237>.*
- [4] 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).
 - [5] Picciano, Anthony G (2002). “Beyond student perceptions: Issues of interaction, presence, and performance in an online course”. In: *Journal of Asynchronous learning networks* 6.1, pp. 21–40.
 - [6] Ralston-Berg, Penny and Nath, Leda (2008). “What makes a quality online course? The student perspective”. In: *Proceedings of the 24th annual conference on distance teaching and learning*.
 - [7] Tackett, Sean, Slinn, Kyle, Marshall, Tanner, Gaglani, Shiv, Waldman, Vincent, and Desai, Rishi (2018). “Medical Education Videos for the World: An Analysis of Viewing Patterns for a YouTube Channel”. In: *Academic Medicine* 93, p. 1. DOI: [10.1097/ACM.0000000000002118](https://doi.org/10.1097/ACM.0000000000002118).