

Where are the Cute Kittens and what happened to your attention spa...

P2 Deliverable, project proposal, and initial analyses for [ADA - Fall 2022](#)

Table of Contents

- [Where are the Cute Kittens and what happened to your attention spa...](#)
 - [Abstract - Where are the cute kittens?](#)
 - [Research Question](#)
 - [Additional Datasets](#)
 - [Methods](#)
 - [Proposed Timeline](#)
 - [Organisation of work](#)
 - [Questions to TA](#)
 - [Notebook](#)
 - [Abstract](#)
 - [Dependencies](#)
 - [Authors](#)
 - [References](#)

Abstract - Where are the cute kittens?

You sit comfortably on your couch and pick up your phone. As you scroll through the latest TikTok trend, past that funny Reel about Gen Z humor, you snugg into a blanket, ready to dive into the new 20min episode by 3Blue1Brown about why 'PI Is Everywhere', but once on YouTube a Short with Elon Musk taking a puff with Joe Rogan captures your attention, and you end up captivated by a Vox documentary on the Rise to Power of Vladimir Putin. ¹

You stop and think: *Where are the cute kittens?*

What happened to YouTube? How did we go from funny cat videos and epic fail montages to long-form podcasts, product reviews, and political analyses with production quality on par with high-budget mainstream media?

In this project, we will travel through time, and explore YouTube's trends over the years, using the YouNiverse Dataset [2](#).

Research Question

Our aim is *not* to establish a causal relationship between particular features and engagement, nor to create a recipe for future channel success. Our focus is to present interesting trends that can be digested and appreciated regardless of how intertwined Social media as a whole have become.

- What is the behavior of the length of videos over the years?
- What are the most popular topics and categories over the years and how are they distributed?
- How did the description and title of videos change over the years?
- How are spikes of interest in certain topics correlated to real-life events at various points in time?
- Are there clear shifts in the behavior of the creators related to shifts in the policy of the platform? Can we see significant changes in topics, lengths of the videos, language used, and tags near changes in the monetization policies of YouTube?
- How did the frequency of uploads change over the years?

Additional Datasets

Methods

Proposed Timeline

- **[14/11 - 20/11] - Week 9:** P2 - showing the feasibility, clarity on research questions, initial exploration, and analysis.
- **[21/11 - 04/12] - Week 10-11:** Breadth: look for interesting trends, research YouTube monetization policies, and do a deeper analysis on the rise of long-format videos on YouTube alongside short-format videos on other platforms. Clean results on lengths of videos on YouTube, analysis on tags, descriptions, and titles to present results and refine with TA. Sentiment analysis, graph and connections, word cloud, and all NLP should be tried out during these two weeks.
- **[05/12 - 11/12] - Week 12:** outline of results, a draft of data story, feedback from TA, and refine based on feedback. Finding signal over noise, and clearing up results narrowing down on what really seems interesting and surprising. NLP over titles and descriptions will need more revision, ask TA. Length of videos and background on YouTube Policies and significant events should be there to assess it clearly.
- **[12/12 - 18/12] - Week 13:** Website mockup and design, final touches, and captivating angles to create an interesting story. Show TA + get feedback from colleagues about the overall feel and look. Is it digestible and entertaining while also being clear?
- **[19/12 - 23/12] - Week 14:** P3 - Cleaned up code, keep one main notebook, add useful comments, final touches on data story, redact final README with notes gathered during previous weeks and clean it up. Make the data story so clean that grandparents understand it.

Organization of work

- **[21/11 - 04/12] - Week 10-11:**
 - **Mariia and Daria:** research on big real-life events and hypotheses on the effect of trends on YouTube. Setup of the notebook to run on the full dataset. Research on NLP and methods to do sentiment analysis and exploring the idea of connected graphs to show correlations.

- **Rouby:** Run clean notebook on the full dataset, further research on additional datasets that can be used, exploration of the contrast between emerging long-format videos on YouTube alongside short-term format around. Exploration of time series.
- **Jacopo:** research on YouTube monetization policies, exploration of the length of videos. Trends at the top and the worst engagement. Exploration of Frequency of uploads. Research on animation and word clouds. Setup of corresponding notebooks.
- **All:** meeting, share notes on Google Docs, GitHub and WhatsApp. Individual setup of corresponding notebooks for each research and synchronization with the main, cleaner notebook. TA Q&A and feedback.
- **[05/12 - 11/12] - Week 12:**
 - **Mariia and Daria:** correct NLP implementation, analysis of tags, and descriptions to make top and worst topics emerge. A clear relationship between real-life events and YouTube trends.
 - **Rouby:** clear results on the frequency of uploads, curation of the main notebook, clear results from time series.
 - **Jacopo:** clear results from the length of videos, animations, initial implementation of the website, and potential interactive elements like word cloud and graph.
 - **All:** draft of data story, eliminate uninteresting and insignificant results, external resources, and ideas on how to make interesting results pop out. Make sure feedback from TA is clear, ask how to refine and find signal over noise. Don't fall in love with uninteresting yet complex results. Double-check results on sentiment analysis, use of regular expression and all NLP related.
- **[12/12 - 18/12] - Week 13:**
 - **All:** website mockup, design, a clear data story, and notebook run. Draft of README evolved from notes + P2.
 - **Jacopo:** website coding and initial deployment.
- **[19/12 - 23/12] - Week 14:**

- **All:** last cleanup and external feedback, cleanup of the main notebook and add comments. The full run-up on meetings and synchronization of work. Official deployment of the website. Correct grammar, and make sure the story is clear and flows. Clean README.

Questions to TA

Notebook

Abstract

Dependencies

Authors

- [@mariia](#)
- [@daria](#)
- [@rouby](#)
- [@jacopo](#)

References

1. [3Blue1Brown](#) - [Joe Rogan](#) - [Elon Smoking Weed](#) - [Vox](#) - [Putin doc](#) ↩
2. [YouNiverse dataset](#) - [GitHub](#) - [Paper](#) ↩