This dataset was created for research purposes associated with the University Maastricht Data Science Research Project. No commercial use is intended. The research is intended to explore the potential gender gap present on youtube, further sentiment analysis research on youtube comment sections among youtube communities, and develop the authors' research skills further. The scope of the analysis is limited to beauty channels on english-speaking youtube. This dataset was created by Aline Mellersh, Katarina Bartekova, Rinske Jongma, and Stephen McCarthy as members of the Data Science Research Project in the Youtube:Gender Gap research group. The creation of this dataset was and is not funded, and in part, is self-funded indirectly through our tuition payments and time spent on this research project. No ethical review process was conducted.

The motivation behind limiting the scope of our research to beauty youtubers, is two-fold. Firstly, we are a small team with a short deadline and need a reasonably narrow deliverable. Secondly, we have identified a potential gap in research on youtube and gender, where prior research has focused on the treatment of non-male content creators in male-dominated communities, we intend to explore the treatment of content creators (of any pronoun) on a traditionally female-dominated community such as the beauty youtuber community. This leads to basic assumptions, being, that the beauty youtuber community is largely female and/or non-cis. We took thirty (approximately the largest as of June 2021) channels from within this community with about ten channels for each gender category studied. Rather than lexically assign gender like some previous research has done, we aimed for confirmation of self-identification by screening each channel for a video, description, interview, post, or other source where they refer to their own gender identity and only include them in the pool of channels when confirmed. We provide a link to each source used for each channel in the table in this datasheet. These sources and information are up to date as of 01/07/2021 and will not be updated or maintained in the future, which may lead to inaccuracies in the future.

Each instance represents a sampled comment from a video on a channel sampled from the beauty youtube community. Each instance within the dataset is a top-level comment comprised of raw comment text, the comment author's username, the author id string, the comment's like count, the time and date of when the comment was posted and when/if it was edited, the reply count, the video id string, and the channel id string. This is all pulled from the YouTube API. This process is conducted by university students and paid student supervisors. Collection of the data took place in June of 2021 with data ranging from early 2020 to June 2021. After fitting these instances in the dataframe, we cleaned, tokenized, and filtered text, each attributing to a new column. All tools used are present in the provided code sample. We add resulting sentiment analysis scores, word counts, compounded scores derived from the sentiment analysis scores, toxicity measures, and other measures to the dataframe. The data is public communications and is publically available and non-confidential. We own no rights privileged beyond what is publicly available. The dataset includes top-level comment usernames, which are generally considered to be people (when not bots) and sub-populations could theoretically be identified from comment author usernames. Our research focuses on comments pertaining to the content creator and does not use comment usernames. The dataset contains data that might be considered sensitive concerning use of strong language, hate speech, and whatever else a public comment may contain. Commenters were not notified or solicited to include their comment data, as it was/is publicly available.

We sampled roughly 769,000 comments from 30 youtube channels and then sampled 30,000 comments from that sample with about 10,000 comments per gender category. The dataset is thus a sample of instances from a larger dataset. The sampled dataset does not appear to be missing instances, but there are unrecognizable emojis and characters used that the sentiment scoring tools such as Emosent cannot recognize. The sample of 30,000 comments is provided, though recreation of such a dataset relies upon external sources such as Youtube's API and this project remaining posted on GitHub. Use of the Youtube API can be learned at https://developers.google.com/youtube/v3/getting-started. All analysis is conducted in a Python environment with a large assortment of installed and imported packages, found in the code sample.

This dataset should not be used for non-sentiment analysis purposes directly pertaining to the thirty sampled youtube channels and should not be used for any illegal or unethical means. This dataset does not have a DOI, but will be distributed alongside our code and datasheet on Github in the beginning of July, 2021. The dataset will not be distributed under a copyright or other intellectual property. No support, hosting, or maintenance of the datasheet, code, or provided links are guaranteed for the foreseeable future. There are no planned updates. The channels sampled can be found on the following page.

| | | | | |
|---|---|---|---|---|
| Jeffreestar | UCkvK_5omS-42Ovgah8KRKtg | male | https://www.youtube.com/user/jeffreestar/videos | https://www.youtube.com/watch?v=essJYTvt5rQ |
| Manny Mua | UCbO9bltbkYwa56nZFQx6XJg | male | https://www.youtube.com/user/MannyMua733/videos | https://twitter.com/mannymua733/status/1295483233568686089 |
| Wayne Goss | UCCvoAe__WFYMNAEN-C-CtYA | male | https://www.youtube.com/user/gossmakeupartist/videos | https://www.youtube.com/watch?v=a9ctb9RLCm4 |
| James Charles | UCucot-Zp428OwkyRm2I7v2Q | male | https://www.youtube.com/c/JamesCharles/featured | https://abcnews.go.com/Entertainment/make-artist-influencer-james-charles-opens-beauty-career/story?id=61924967 |
| PatrickStarrr | UCDHQbU57NZilrhbuZNbQcRA | male | https://www.youtube.com/c/patrickstarrr/featured | https://fashionista.com/2016/02/patrick-starr-instagram-youtube#:~:text=I'm%20a%20boy%2C%20I,and%20over%20800%2C000%20YouTube%20subscribers. |
| Bretman Rock | UC3EFKdXAU99j3ppGgvTz7XQ | male | https://www.youtube.com/c/BretmanRock/featured | https://www.youtube.com/watch?v=I2WfMjssJLM |
| Thomas Halbert | UCFn4TEi42U-WHYjiqaxpp3w | male | https://www.youtube.com/channel/UCFn4TEi42U-WHYjiqaxpp3w/videos | https://www.youtube.com/watch?v=v8DsSwvhRwM |
| Michael Finch | UCXPbZbUPaNCcr5iyDetqsxg | male | https://www.youtube.com/c/MichaelFinch/featured | https://www.youtube.com/watch?v=wyEhda5WaWc |
| Skelotim | UC7FVvYGiBpgjGx7rqb5SIPg | male | https://www.youtube.com/c/Skelotims/about | https://www.nytimes.com/2016/10/19/arts/design/those-lips-those-eyes-that-stubble-the-transformative-power-of-men-in-makeup.html |
| Jake Warden | UCMYXusJ8Tcocq3AtT23T0eA | male | https://www.youtube.com/c/JakeWarden/featured | https://www.youtube.com/watch?v=GpfjAhGCkE0&t=2s |
| Edward Avila | UCNM5-NRrDxL0PWC5H52smaQ | male | https://www.youtube.com/mrpanda101/featured | https://www.youtube.com/watch?v=x2VtyMbqhWw |
| NikkieTutorials | UCzTKskwIc_-a0cGvCXA848 | female | https://www.youtube.com/c/nikkietutorials/videos | https://www.youtube.com/watch?v=DQdXV |

| | | | | |
|---|---|---|---|---|
| | Q | | | LSYu-0&t=302s |
| Tati | UC4qk9TtGhBKCkoWz5qGJcGg | female | https://www.youtube.com/c/Tati/videos | https://www.youtube.com/watch?v=uFvtCUzfyL4 |
| Sarabeautycorner | UC0YvTCy1I4_a-3pn47_5DBA | female | https://www.youtube.com/c/SaraBeautyCorner/about | https://www.youtube.com/c/SaraBeautyCorner/about |
| Michelle Phan | UCuYx81nzzz4OFQrhbKDzTng | female | https://www.youtube.com/c/MichellePhan/about | https://www.youtube.com/watch?v=2OGa3NBzDZM |
| Bethany Mota | UCc6W7efUSkd9YYoxOnctlFg | female | https://www.youtube.com/c/bethanymota/videos | https://www.youtube.com/watch?v=0WKewlb1HOM |
| Carli Bybel | UC21yq4sq8uxTcfgIxxyE9VQ | female | https://www.youtube.com/c/CarliBel55/videos | https://www.youtube.com/watch?v=TE-9xELzK6s |
| Safyia Nygaard | UCbAwSkqJ1W_Eg7wr3cp5BUA | female | https://www.youtube.com/c/SafiyaNygaard/videos | https://www.youtube.com/watch?v=t7VP5_REfFQ |
| grav3yardgirl | UCGwPbAQdGA3_88WBuGtg9tw | female | https://www.youtube.com/user/grav3yardgirl/videos | She calls herself a girl in her youtube name and description. |
| Nikyta Dragun | UCNBvzAJI3N92Sgl0guRxSxQ | female | https://www.youtube.com/c/NikitaDragun/videos | https://www.youtube.com/c/NikitaDragun/about |
| Glam&Gore | UCoziFm3M4sHDq1kkx0UwtRw | female | https://www.youtube.com/c/GlamAndGoreMakeup/about | https://www.youtube.com/watch?v=jvcfj0x2pJA |
| Princess Jules (Julie Vu) | UCT9lRRTBWIqMIfVgSyfsg7Q | female - REWRITE HER ID into FEMALE | https://www.youtube.com/channel/UCT9lRRTBWIqMIfVgSyfsg7Q | https://www.youtube.com/watch?v=dRCMAdCMfrU |
| Simply Nailogical | UCGCVyTWogzQ4D170BLy2Arw | female | https://www.youtube.com/c/simplynailogical/videos | https://www.youtube.com/watch?v=OylDdm-6ZH0 |
| J aka thaibrows | UCzN3iACIG5HSQpY-hudis1Q | non-binary | https://www.youtube.com/channel/UCzN3iACIG5HSQpY-hudis1Q | https://www.youtube.com/channel/UCzN3iACIG5HSQpY-hudis1Q/about |
| Chella Man | UCa1vUXV2WMRobPo-ZfEeRhg | genderqueer, trans-masculine | https://www.youtube.com/c/ChellaManArt/featured | https://www.youtube.com/c/ChellaManArt/about |
| YuhuaHamasaki | UCbLunbaq3ia4-FLfOHrHngQ | non-binary | https://www.youtube.com/c/YuhuaHamasaki/featured | https://www.seventeen.com/celebrity/movies-tv/a21734217/rupauls-drag-race-yuhua-hamasaki-it-gets-better |

| | | | | |
|---|---|---|---|---|
| | | | | -gender-neutral/ |
| Brendan Jordan | UCDQGFnUC1iw_oQ5Vs1KoVlg | genderqueer, trans-masculine | https://www.youtube.com/c/BrendanJordan1/featured | https://www.instagram.com/brendanwjordan/ |
| Miles Jay | UCeNgRHpH7OHZetYjC5JZXGw | non-binary | https://www.youtube.com/user/MilesJaiProductions | https://tomboyx.com/blogs/news/a-conversation-with-miles-jai / https://twitter.com/milesjai |
| clawdeena9 | UCJyw_54aiqanw8s39eq6HbA | genderqueer, non-binary | https://www.youtube.com/user/clawdeena9/videos | https://theunsealed.com/from-closeted-to-courageous-how-makeup-transformed-my-life/ |