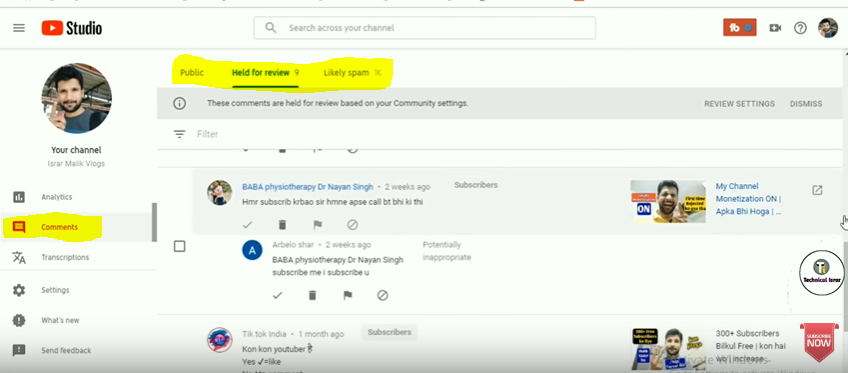
Capstone 2: Project proposal

# Problem Statement:

Comments analytics for YouTube channels.

# What problem I am trying to solve:

The YouTube channels owners can access analytics on their videos on YouTube studio. There is a section for comments as well but it doesn’t have any analytics on the comments. It only shows a list of comments, comments held for review and likely spam comments. (Snapshot: [Source](https://youtu.be/QGW8mvAv20Q?t=1507))



I am planning to come up with an enhanced version of analysis on comments which can be useful to the channel owner to grow their views and revenue.

# Datasets

1. ~~Use selenium to scrape comments from the channel.~~ Use YouTube API instead.
2. Utilize UCI’s data set to filter out spam comments and then perform rest of the analysis on non-spam comments.

[http://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection#](http://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection)

# Who is the client?

1. YouTube channel owners
2. This tool can be made available as a stand-alone tool as well and can be used by anyone to perform analysis on channel/video of their interest.

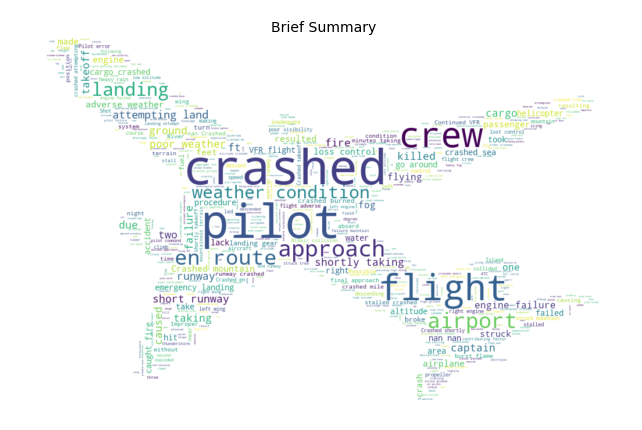
# Justification how clients can utilize it

1. Clients can use this tool to analysis what is overall sentiment of people on a particular video.
2. They can know what people are talking most about. They can utilize this information in their upcoming videos.
3. They can analyze most liked comments and most disliked comments on their highly viewed videos.

# Process flow

# Visualization:

1. Word cloud / bar chart on most frequent words. If time permits I’d like to make a word cloud like below in some other shape.([Photo source](https://www.kaggle.com/ruslankl/airplane-crashes-data-visualization))



1. Cluster analysis to gauge what people are mostly talking about (Need to brainstorm this on how to implement, how to assign them categories)
2. Perform cluster analysis on most engaging comments (use list of top 100 comments with most up votes/ down votes/comments).
3. Perform cluster analysis on positive and negative comments separately to know what liked or disliked the most.
4. List of top 10 most active users/commenters.