ETF Project – Group 2

Group 2 is analyzing trending YouTube videos in the U.S., created by the YouTube API. The term “trending” is based on a daily algorithm that scans short term activity regarding views, comments, likes, dislikes, and shares.

**Extract**

In order to do this we found out 2 data sources. A CSV that contained over 40,000 trending videos and a JSON file that has the associated category ID tag for video genre.

**Transform**

We want to understand the impact of engagements with a video (likes, dislikes, and comments). To do this we needed to account for video views so adjust information on a per view basis. We divided views by each out of engagement categories to understand the number of likes, dislikes, and comments per view for each of the 31 genres.

Example 1

* Hypothesis:
  + Viewers are more likely to comment on a negative video, therefore, videos with higher dislikes per views will also have a higher number of comments per view
* Null Hypothesis:
  + Viewers are not more likely to comment on a negative video, therefore, videos with higher dislikes per views will not have a higher number of comments per view

Example 2

* Videos that

**Load**

Once out CSV and JSON files we’re