

# HOW TO ANALYZE YOUTUBE

**Professor Ernesto Lee**

# YOUTUBE

- Owned by Google
- Hires tons of Analysts
- Has one of the best Algorithms in the world



**YOUTUBE**  
VIDEO

# GET THE VIDEO DATA

<https://github.com/fenago/ISM3545>

18.01.11\_US\_videos.csv



# REVIEW THE DATA

- Counts
- Nulls / NaNs
- Data Types
- Any and all cleaning / pre-processing

# FEATURE ENGINEERING

**In this case, I'm going to drop these fields:**

- **video\_id, channelId,**
- **categoryId. trending\_date,**
- **tags,**
- **thumbnail\_link,**
- **comments\_disabled,**
- **ratings\_disabled,**
- **description**

# UNIVARIATE ANALYSIS

Use the DA Template and answer questions like:

- Average of
  - Views
  - Likes
  - Dislikes
  - comments

# SIDEBAR - HOW DOES THE ALGORITHM WORK?

1. View count
2. How quickly the video is generating views (i.e. “temperature”)
3. Where views are coming from, including outside of YouTube
4. The age of the video
5. How the video performs compared to other recent uploads from the same channel

<https://support.google.com/youtube/answer/7239739?hl=en#:~:text=Trending%20helps%20viewers%20see%20what's,of%20viewers%20would%20find%20interesting.&text=The%20list%20of%20trending%20videos,same%20position%20in%20the%20list.>

# BIVARIATE ANALYSIS / MULTIVARIATE ANALYSIS AND CORRELATION

Can you find a pattern between 2 columns in this dataset?

Are there any obvious correlations?

If you add back in the categorical columns, can you discover any new insights?