



YouTube

Trending Video Analysis

Group Name: Happy Data Programming

Members: Wei Zhang, Wanyi Su, Zeyu Hu,
Jinyao Lu, Yin Yu Kevani Chow

Apr 9, 2022

Motivation

Why is it an important project?



1st Video Sharing Platform
2nd Largest Search Engine
2nd Most Popular Platform
for Influencer Marketing



2.3 Billions Users



197 Millions



50 Millions



18 Millions

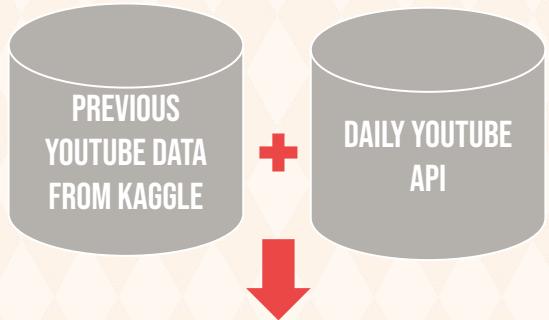


Digital Marketing
US\$237.15 millions



YouTuber Annual Revenue
US\$29.5 Millions

Dataset



Eastern Time Zone

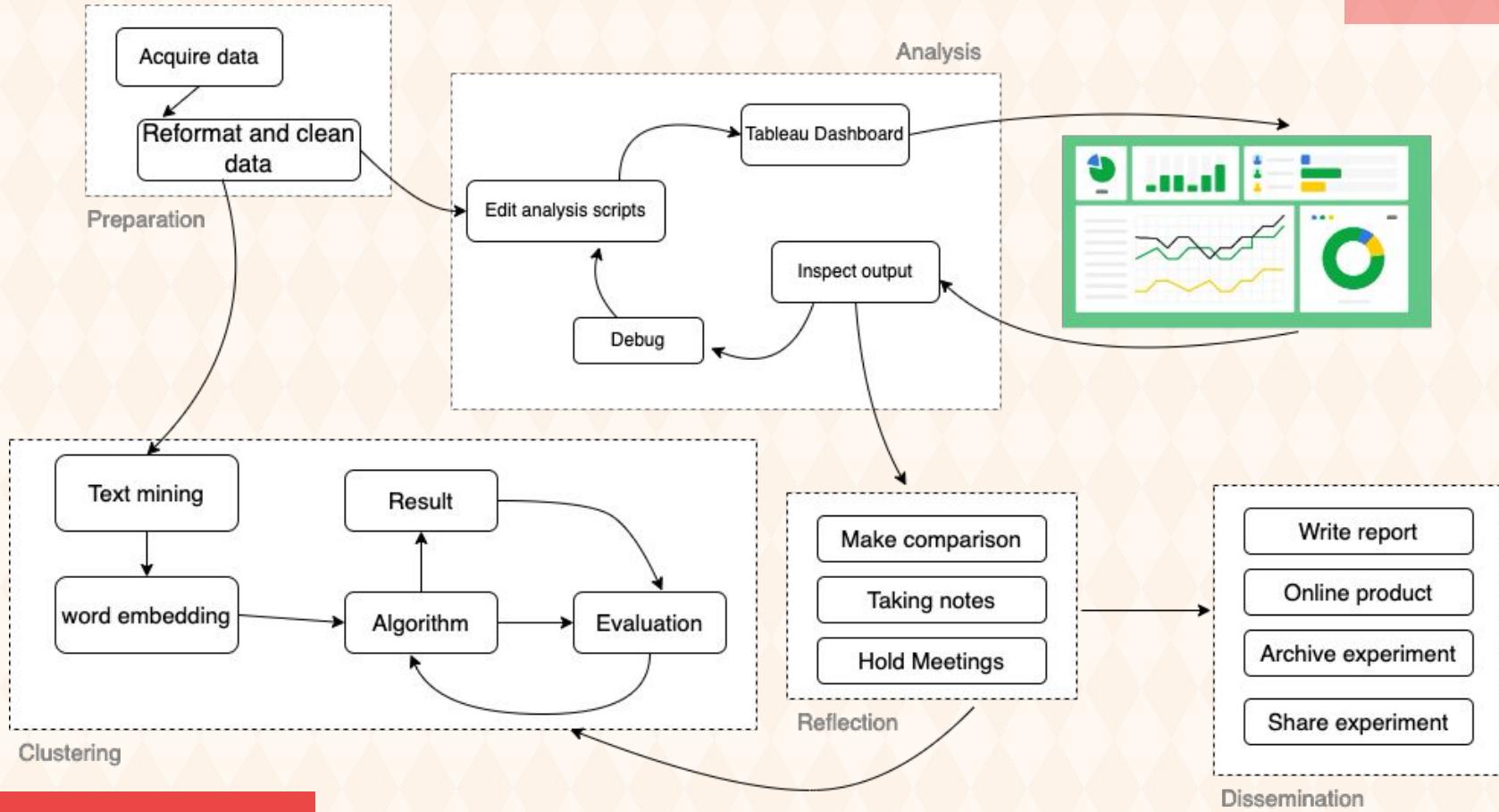


Eastern Time Zone



Greenwich Mean Time

Data Science Workflow



Exploratory Data Analysis



Problem Statement

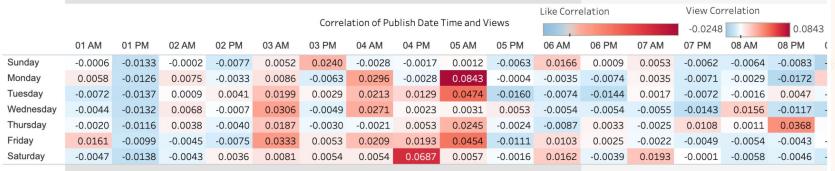
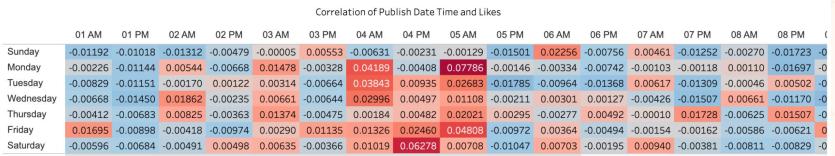
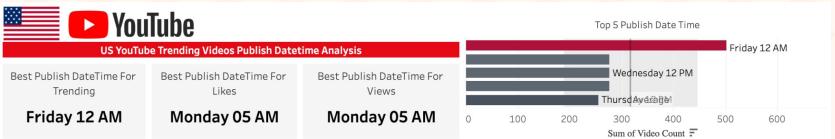
- 1 Publish date and time of trending videos
- 2 Cultural differences reflected in trending videos
- 3 Cultural similarities reflected in trending videos

Interactive Tableau Dashboards

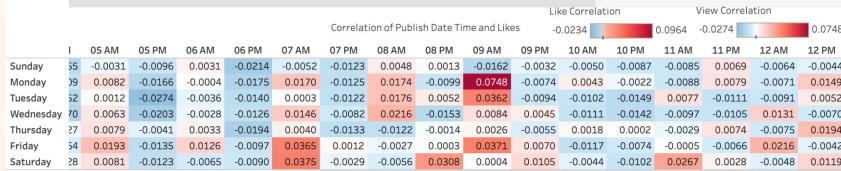
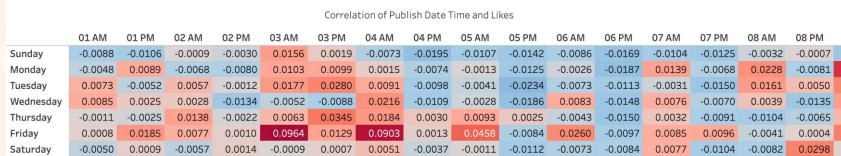
Problem Solving

1

Publish day and time of trending videos



Correlation of Publish Time

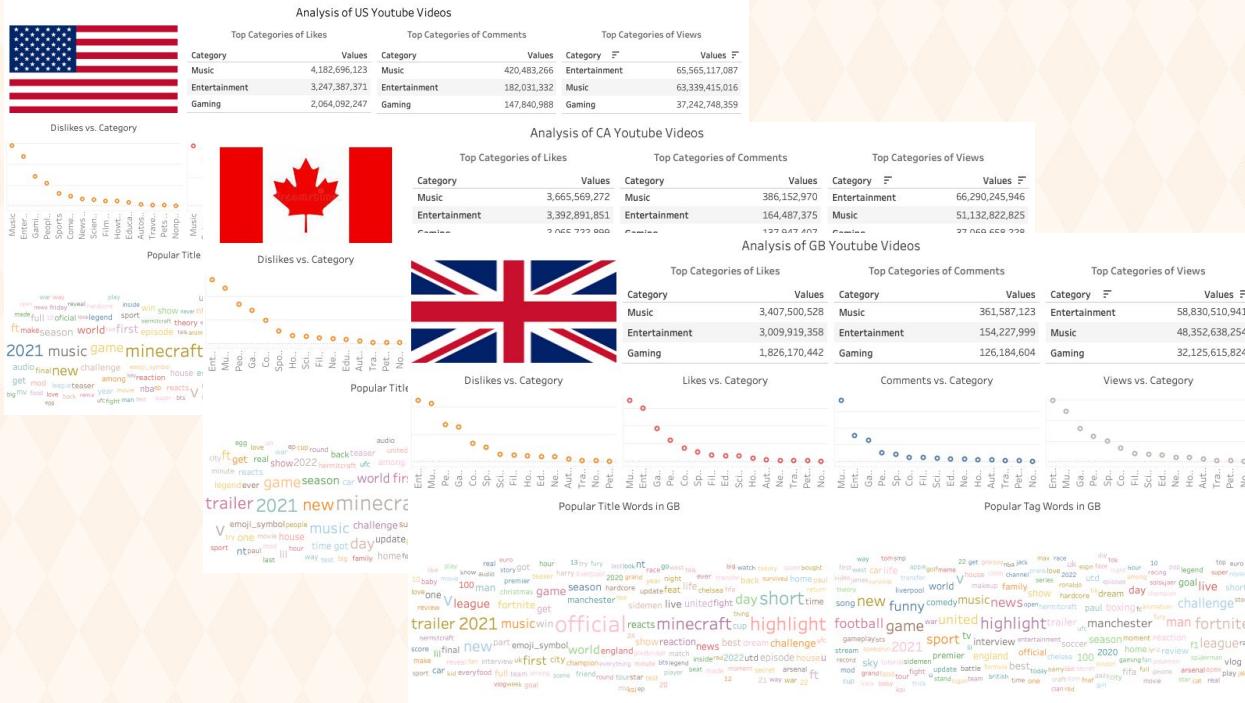


vs. Trending
vs. Likes
vs. Views

Problem Solving

2

Cultural differences reflected in trending videos



Media

Disney, Fox / Twitch / BBC

Person

Jake / Harry, Solskjær, Ronaldo

Sports

NBA, UFC, WWE / Chelsea, Manchester, Liverpool

Geographical
UK, euro, England

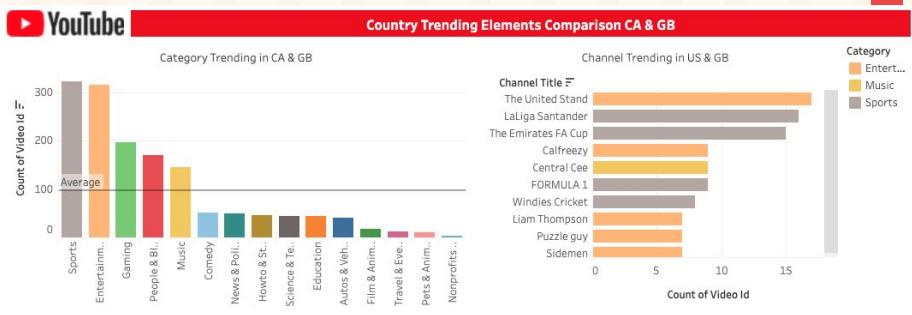
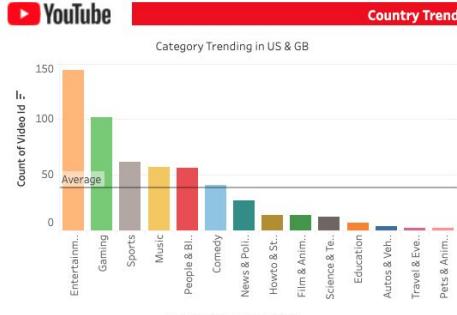
Holiday
Christmas

Episode
Spiderman / Squid Game

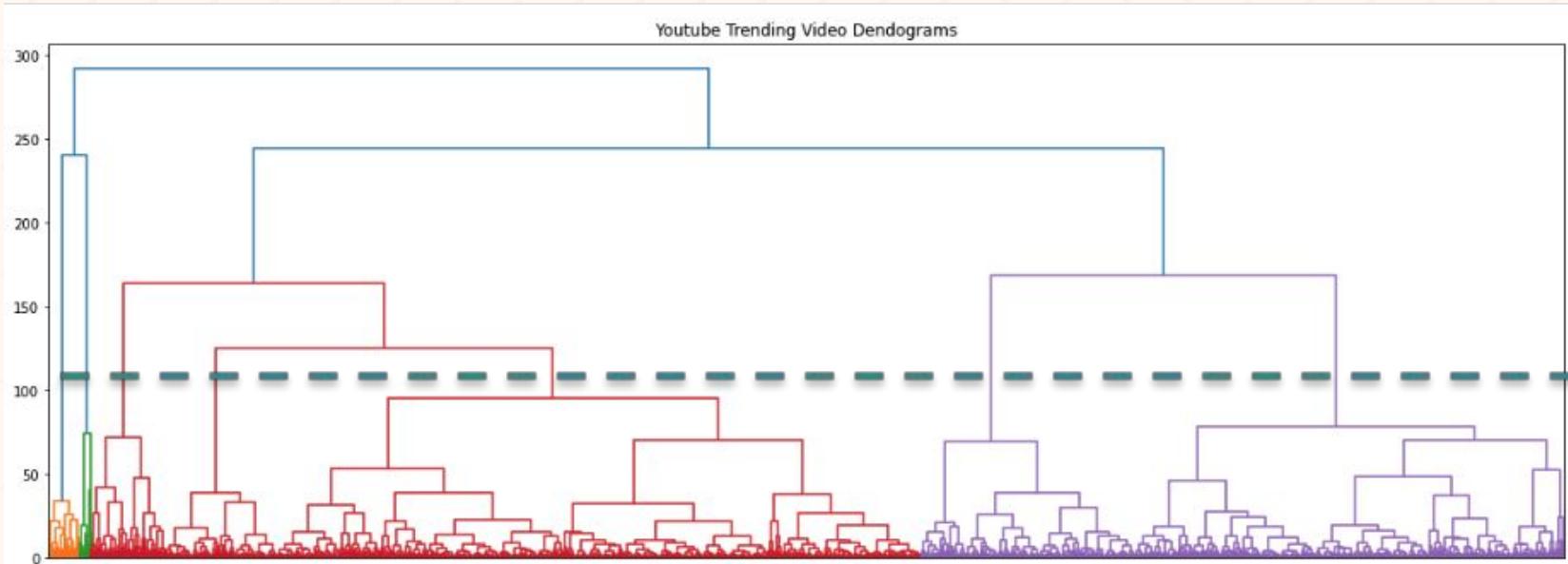
Problem Solving

3

Cultural similarity reflected in trending videos



Machine Learning - Hierarchical Clustering

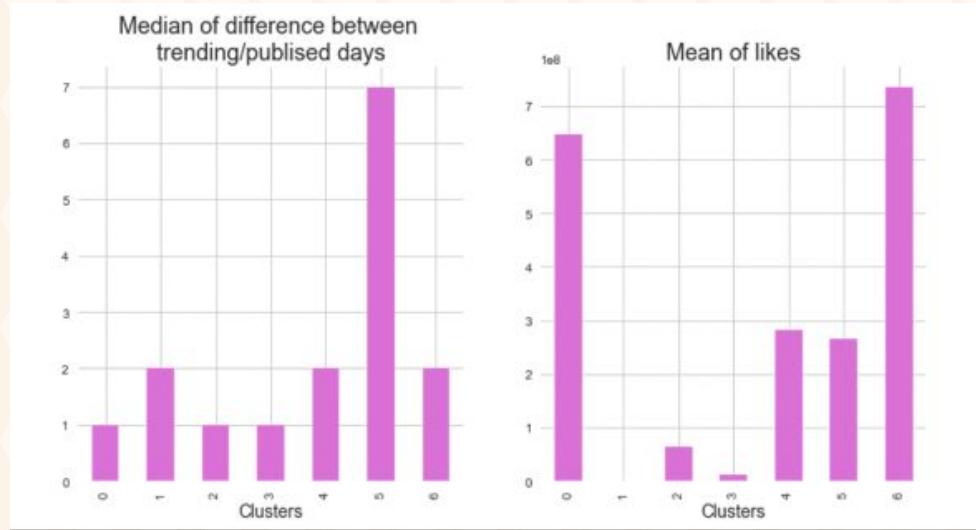


Dendrogram with Numerical/Categorical Features

View_count, likes, dislikes
Comment_count, pubToTreDays

Comments_disabled
ratings_disabled

Machine Learning - Hierarchical Clustering



Dendrogram with Numerical/Categorical Features

View_count, likes, dislikes
Comment_count, pubToTreDays

Comments_disabled
ratings_disabled

Results ->
focus on 4 Categories:
Entertainment, Gaming,
Music, Sports

Machine Learning - KMeans

The Gaming Category



Tokenization

Lowercase, remove urls, emojis, special characters



Lemmatization

Words base form caring -> care



Remove stopwords

Repeated but not useful words (tailored)

Word embedding

Word -> vector

Word2Vec

Website

Home Introduction EDA Findings Product Contact

YouTube Trending Video Analysis

Wei Zhang | Wanyi Su | Zeyu Hu | Jinyao Lu | Yin Yu Kevani Chow

Introduction EDA Findings Product Contact

Canada Video Count 93,745

The United States Video Count 93,795

Great Britain Video Count 92,795

EDA

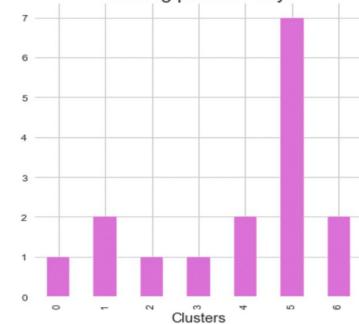
Link: <https://yt-trending-video-analysis.herokuapp.com/>

Home Introduction EDA Findings Product Contact

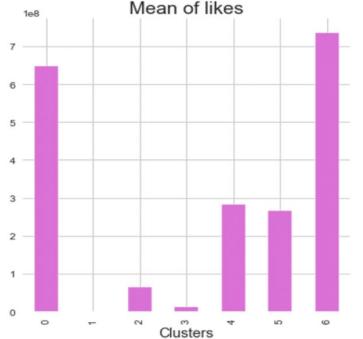
Modeling Result

- "Difference between trending/published days" represents the number of days it takes the video to become trending.
- "Likes" represent the number of likes in the day when it becomes trending.
- Cluster 0 trends within 1 day (median) rapidly and outperform others by winning high number of likes (>600 million on average). Cluster 6 shares the similar behavior to trend in 2 days (median) and attract more than 700 million likes on average.
- The categories in these two clusters are presented in the word cloud. In

Median of difference between trending/publisied days



Mean of likes



Modeling Result

Data Product

Home Introduction EDA Findings ▾ Product ▾ Contact

Clustering Result
Find out your category

Product

Select the category you are interested in

Entertainment Gaming Music Sports

The dashboard features a navigation bar at the top with links for Home, Introduction, EDA, Findings, Product, and Contact. A 'Clustering Result' section with a 'Find out your category' button is located above the main content area. The main area is titled 'Product' and contains a sub-instruction 'Select the category you are interested in'. Below this are four category icons: Entertainment (TV), Gaming (diamond), Music (earphones), and Sports (soccer ball). Each icon has a corresponding word cloud below it. The word clouds are color-coded and contain various terms related to their respective categories.

Gaming is a more specific topic that ranges from games info, playing to other related things. This category has been

<https://trending-video-analysis.herokuapp.com/product.html>

Select a Category

- WordCloud
- Text Analysis
- Relative Video link

Reflection And Future Work



Organize the scattered data into a more streamlined platform



More analysis on the thumbnail image



Ethical issues in videos analysis



In text mining, test and compare with stemming methods for preprocess



Break down big question into smaller ones



Add additional categories such as: Film & Animation, Autos & Vehicles, Pets & Animals

Thank you!

Questions?