YOUTUBE ASMR DATA MINING

CSPB Data Mining

Emma Crawford

August 14, 2023

AGENDA

Questions

Data Preparation

Tools Used

Models Used

Knowledge Gained
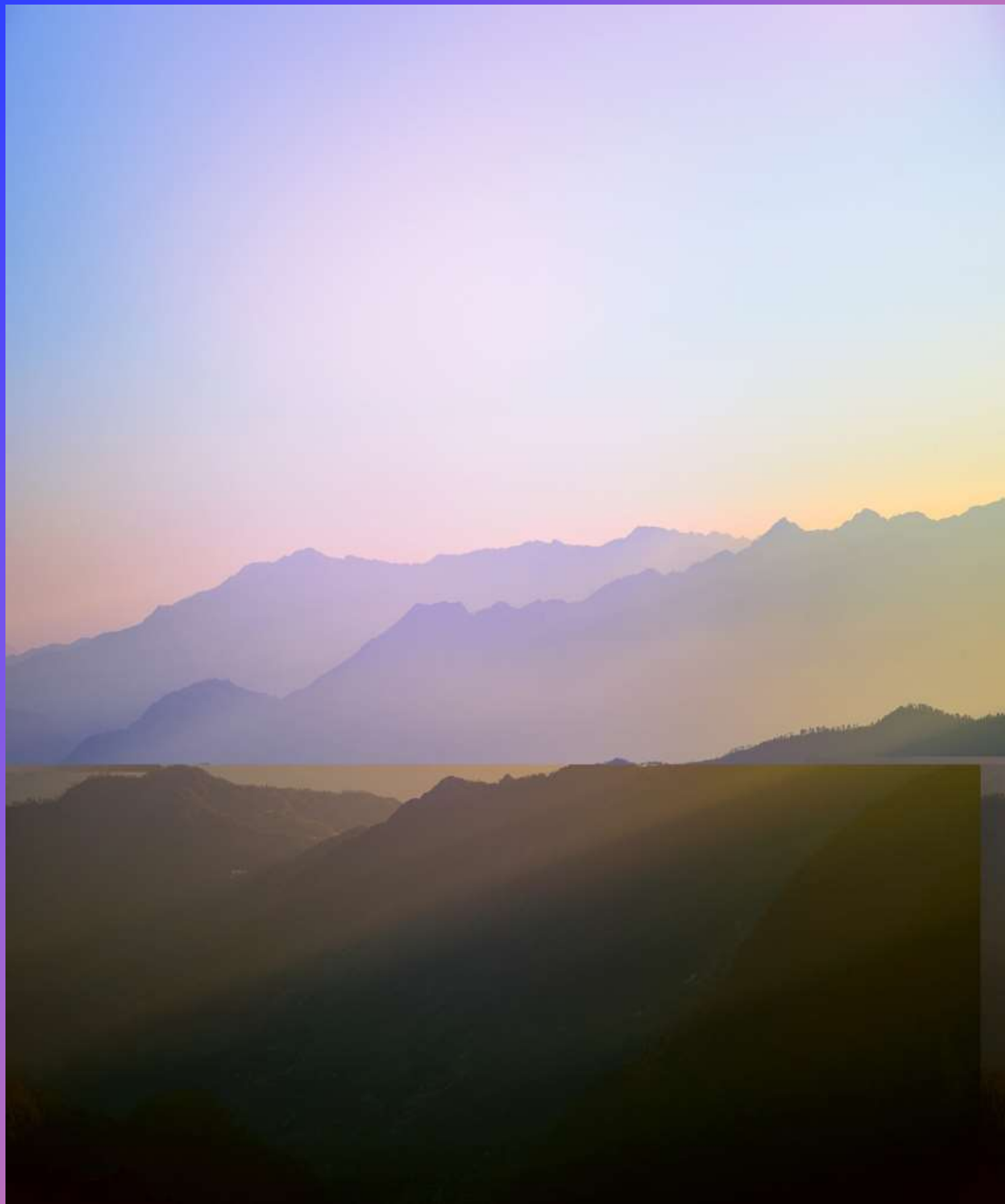
Applications

What is ASMR?

- **A**utonomous **S**ensory **M**eridian **R**esponse

- Pleasant visuals and audio

- Promotes relaxation and sleep

- Common subcategories
  - Appealing object manipulation
  - Spa treatments
  - Cooking or eating

QUESTIONS

Key Questions

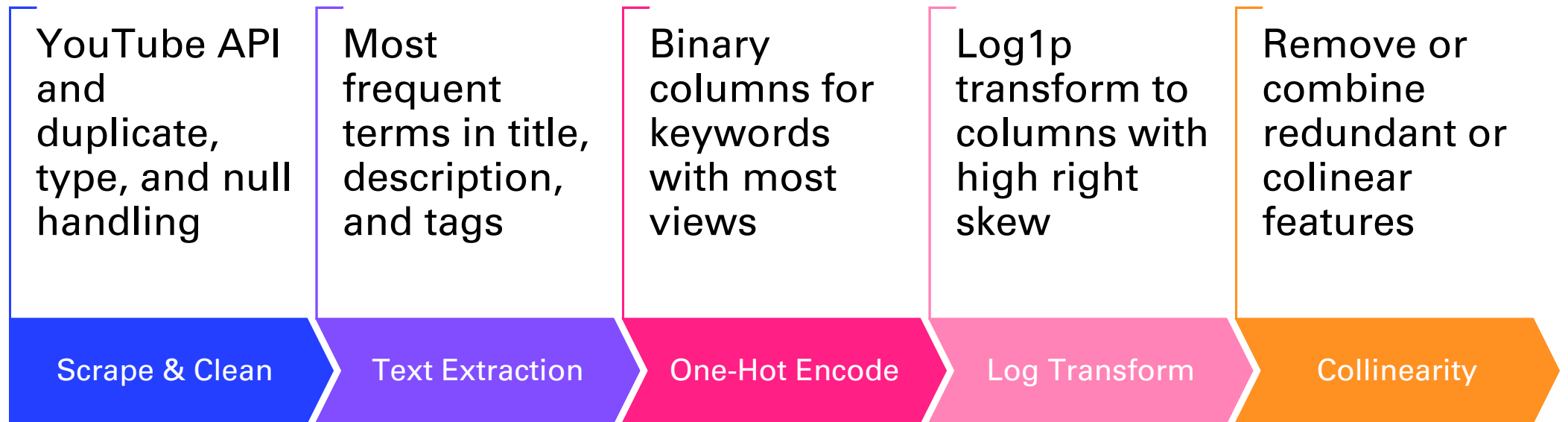What are the top predictors of high view count for a video?

What attribute combinations or niches contribute most to high view counts?

DATA PREPARATION

Process

Scrape & Clean	Text Extraction	One-Hot Encode	Log Transform	Collinearity
YouTube API and duplicate, type, and null handling	Most frequent terms in title, description, and tags	Binary columns for keywords with most views	Log1p transform to columns with high right skew	Remove or combine redundant or colinear features

TOOLS

- Matplotlib/Seaborn
- Sklearn
- SciPy
- Statsmodels

- Pandas
- NumPy
- NLTK
- YouTube API V3

MODELS

Prediction Model Results

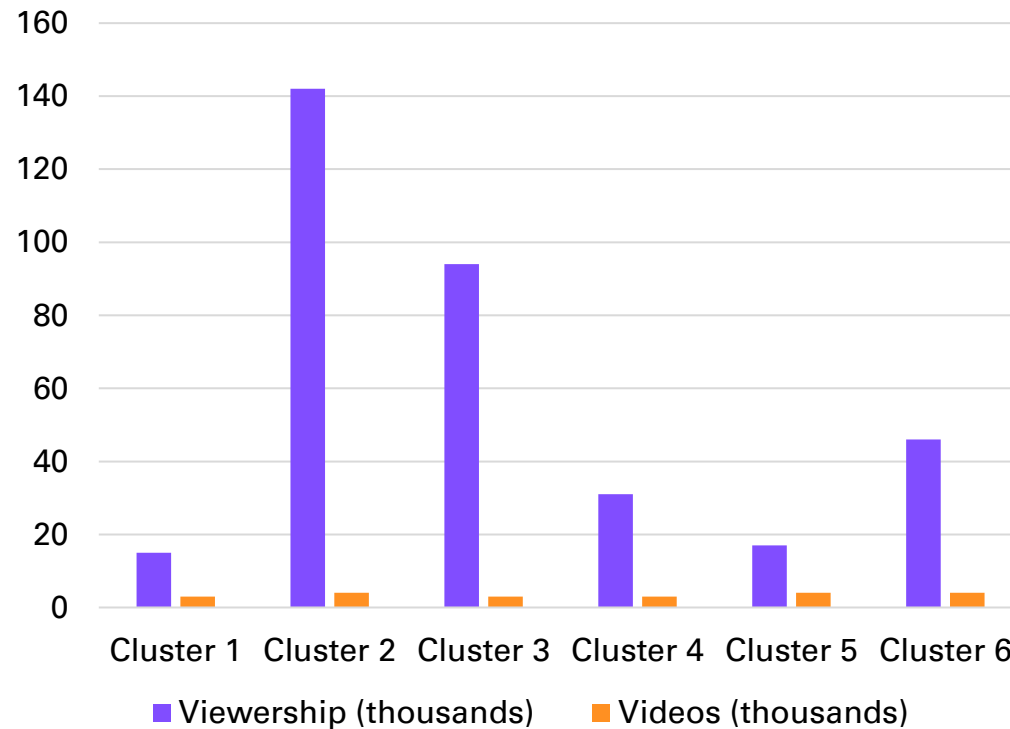| Model	Include Likes & Comments	Clustered*	MSE/MAS	R-squared
Linear Regression	yes	no	0.117	0.968
Linear Regression	no	no	1.344	0.636
Random Forest	yes	no	0.172	0.984
Random Forest (tuned)	no	no	0.627	0.807
Random Forest (tuned)	no	yes	0.625	0.809

*K-means clustering

KNOWLEDGE

Views by K-means Cluster

Top Clusters (k=6)

[2]
6am, high subscribers, non-english characters
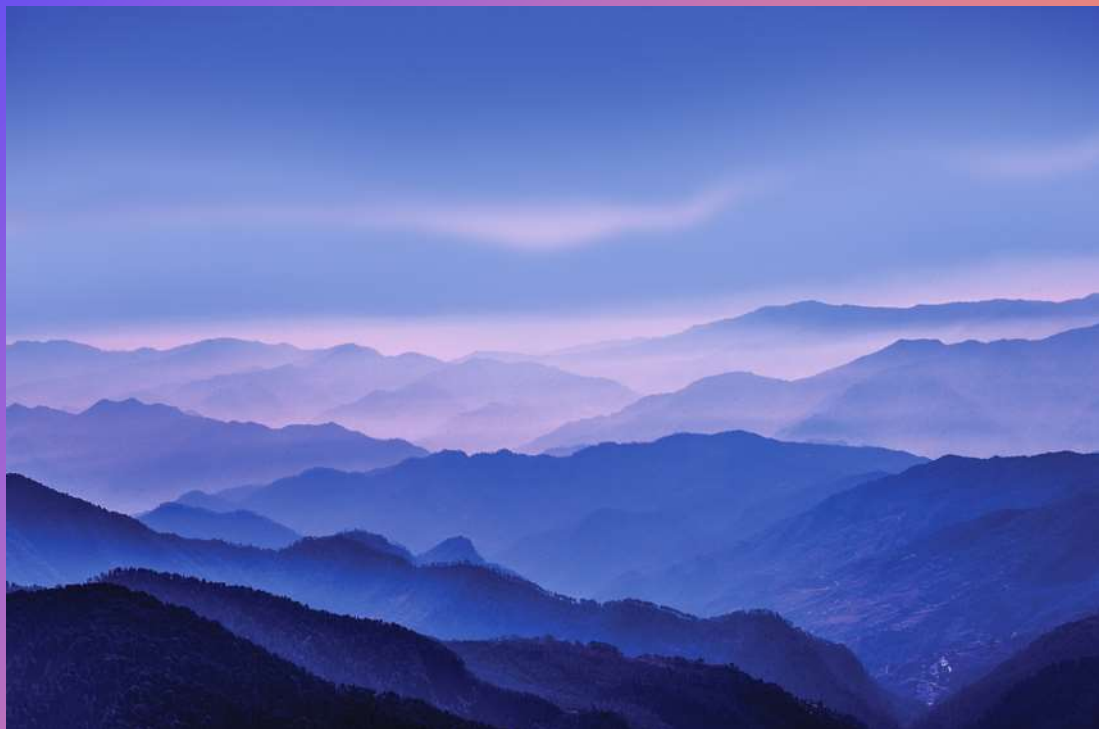
[3]
2pm, sleep & massage

[6]
2pm, sleep only

APPLICATIONS

Takeaways

- **Top Predictors (in order)**
  likes, comments, subscribers

- **Best Time**
  Sunday 6am-2pm

- **Best Topics**
  Sleep & Massage

- **Best Text**
  Title: 10 words
  Tags: >20

- **Bonus Tips**
  - The longer the better
  - Delete old videos with low views

THANK YOU

Emma Crawford