# ANALYZING YOUTUBE TRENDS
## (from August 2020 to April 2024) AND TRAINING A CONTENT CATEGORY PREDICTION MODEL

**G2**

Timofei Šinšakov (Gr 7)
Natalja Frantikova (Gr 4)
Andrei Potrebin (Gr 2)

# INTRODUCTION

### MOTIVATION & DATA OVERVIEW

YouTube's "trending" section showcases videos that have garnered significant attention. Given the platform's vast popularity, one can draw conclusions about societal interests based on trends. By analyzing data from two major markets, insights into media consumption and cultural trends can be obtained.

The datasets used for this project (from Kaggle) include two datasets covering trending YouTube videos from USA and UK from 2020-09-04 to 2024-04-14. Each dataset has 16 columns representing video metadata and engagement metrics. Each row represents a trending video on just one specific day (hence duplicates).

### GOALS & METHODOLOGY

Goal 1: Train an ML model to predict the category of a video

Goal 2: Compare trending videos from US and UK

Goal 3: Analyze the evolution of trending topics over time

Our approach consisted of:

- Data cleaning
- Analytics and Data visualization using frequency plots and tables, pie graphs and word clouds
- Temporal analysis using time-series
- Hypothesis testing using the binomial test, p-value computation

# ANALYSIS RESULTS

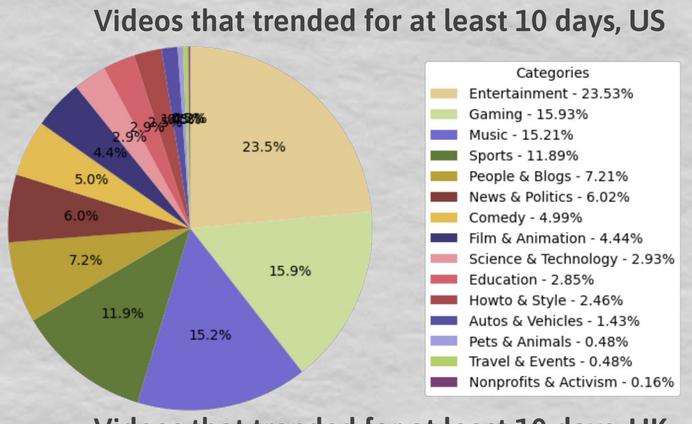
## US vs UK CONTENT

There are differences in the distributions of content categories (all videos). Categories whose popularity differences are significant (with $p < 0.05$) are:

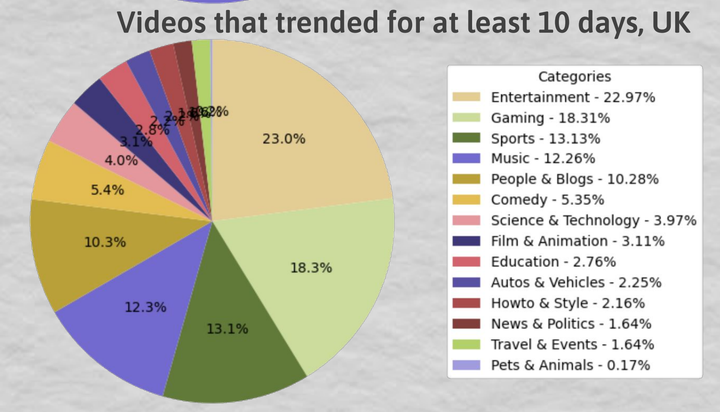
**(More prevalent in US)**
- Music
- Gaming
- News & Politics
- Film & Animation
- HowTo & Style
- Comedy

**(More prevalent in UK)**
- Sports
- Entertainment
- Travel & Events
- Autos & Vehicles
- People & Blogs

Videos that trended for at least 10 days, US




Videos that trended for at least 10 days, UK




In US, some videos stayed in trends for upto 37 days (MrBeast), while only upto 17 days in UK (2 music videos).

## US vs UK CHANNELS

Out of 8294 channels in the US and 7566 in the UK, a total of 5032 channels trended in both countries. So they share a lot of common trends, but many channels are still distinctive to only one county.

Top 30 channels that only trended in US and not UK

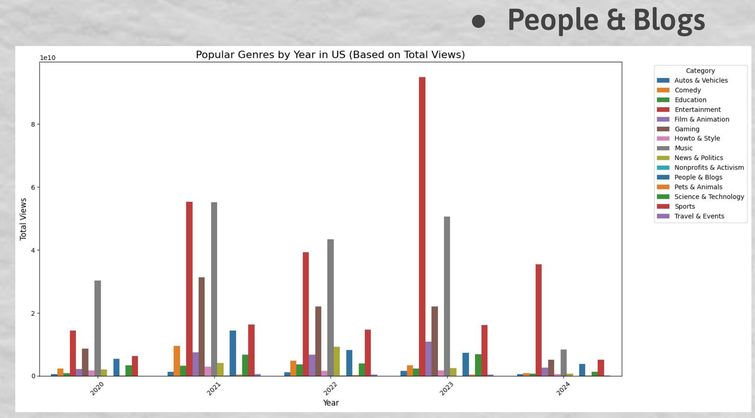

Top 30 channels that only trended in UK and not US

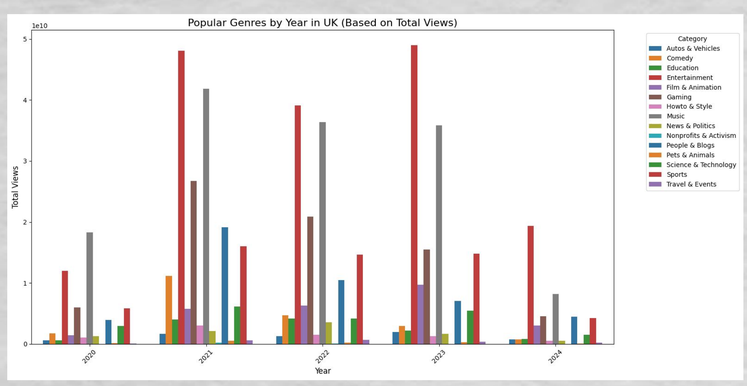

## EVOLUTION OF TRENDING TOPICS

We found that Youtube's main topics like entertainment, gaming, music and sports remain relevant, while the popularity of other topics is fickle.

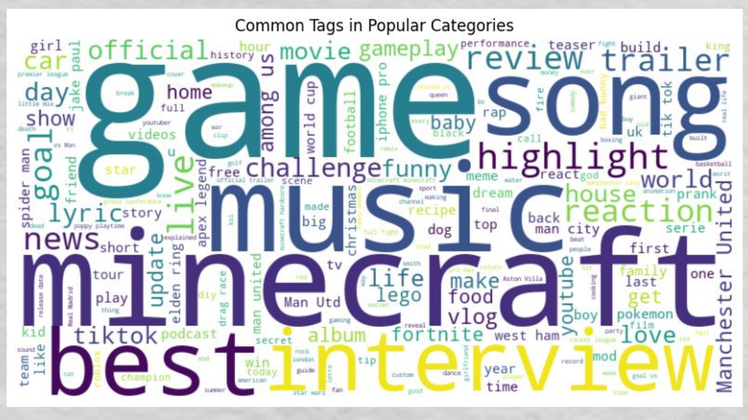
**Growth in views**
- Entertainment
- Film & Animation (UK)
- Politics (2022)

**Decline in views**
- Comedy
- Music (UK)
- Gaming
- Education
- Howto & Style
- People & Blogs




Most popular trending video tags within the database timeframe




# CATEGORY PREDICTION

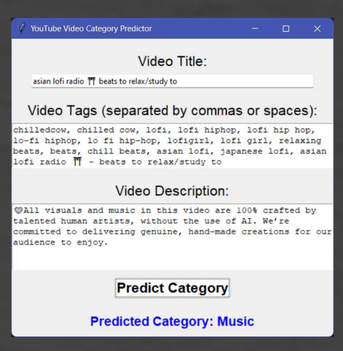
## APPROACH & RESULTS

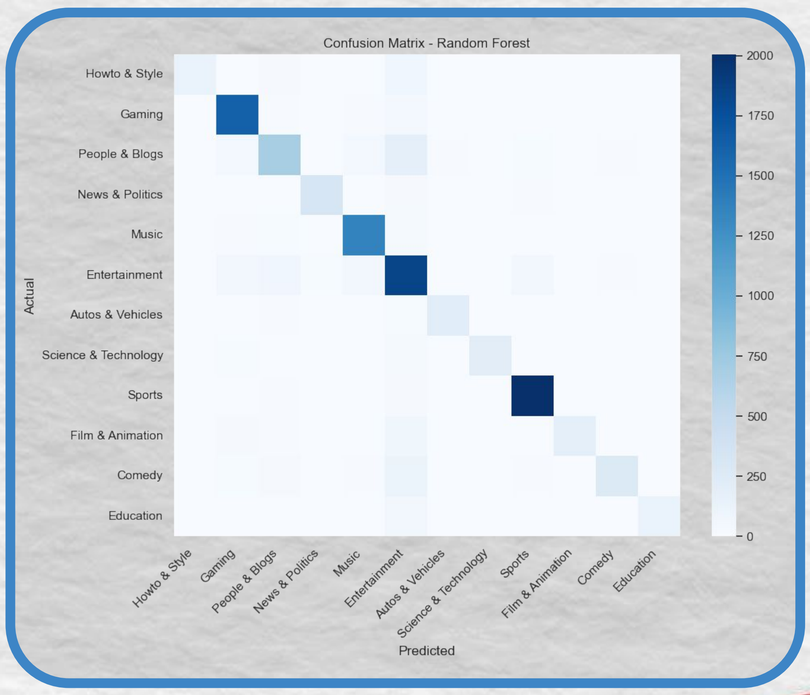
The data underwent a thorough cleaning process, removing links, HTML tags, and non-alphabetic characters, followed by the removal of English stopwords.

By combining titles, tags, and descriptions into a single text feature and vectorizing them, meaningful numerical features were extracted for classification.

A Random Forest model was then trained, achieving an accuracy of approximately 85,7% in predicting video categories.

*We've also created a simple python app for you, so you can try to play with the model yourself!*




**Our results are more than this. Find out more on the project's GitHub!**

Link: https://github.com/Eldern45/DSYoutubeAnalytics

Data source: https://www.kaggle.com/datasets/rsrishav/youtube-trending-video-dataset
Files used: US_category_id.json, US_youtube_trending_data.csv, GB_category_id.json, GB_youtube_trending_data.csv