



گزارش تکلیف دوم درس الگوریتم های علوم داده (بخش دوم)

نام و نام خانوادگی: فاطمه ترودی

شماره دانشجویی: ۴۰۳۴۲۲۰۴۸

نام استاد: دکتر سعیدرضا خردپیشه

نیمسال دوم ۱۴۰۳-۰۴

فهرست مطالب

۱. مقدمه.....	۴
۲. آماده سازی و پیش پردازش داده‌ها.....	۴
۲.۱. بارگذاری داده‌ها.....	۴
۲.۲. بررسی داده های تکراری.....	۴
۲.۳. بررسی داده های گم‌شده.....	۴
۲.۴. بررسی نوع داده‌ها.....	۵
۲.۵. نگاشت شناسه دسته (Category Id) ویدئو به نام دسته (Category Name).....	۵
۲.۶. مهندسی ویژگی ها (Feature Engineering).....	۵
۳. تحلیل سوالات اصلی.....	۶
۳.۱. معیار های تعامل (لایک، دیس لایک و کامنت) به طور کلی و در بین دسته های مختلف ویدئو چگونه توزیع می‌شوند؟.....	۶
۳.۲. کدام کانال‌ها و دسته‌های ویدئویی YouTube در هر کشور و در سطح جهانی ترند شده‌اند؟.....	۷
۳.۳. آیا الگوهای فصلی یا روز هفته در ویدئو های پرطرفدار وجود دارد؟ زمان آپلود روز/زمان بر تعامل ویدئو تاثیر می‌گذارد؟.....	۹
۳.۴. آیا ویدئو های با نرخ دیس لایک بالا بیشتر میزان تعامل بیشتری از ویدئو های عموماً لایک شده دارند؟.....	۱۱
۳.۵. برچسب‌های ویدئو چگونه بر تعامل تأثیر می‌گذارند و کدام برچسب‌ها بیشتر در ویدئوهای پرطرفدار استفاده می‌شوند؟.....	۱۲
۳.۶. طول عنوان ویدئو چگونه بر تعامل تأثیر می‌گذارد؟.....	۱۳
۳.۷. آیا رابطه ای بین احساس عنوان ویدئو و تعامل وجود دارد؟.....	۱۴
۳.۸. آیا عناوین کلیکی (به عنوان مثال، "تکان دهنده"، "باید تماشا کنید") منجر به تعامل بیشتر می‌شود؟.....	۱۴

۴. داستان سرایی و سوالات اضافی.....	۱۵
۴.۱. سوالات اضافی.....	۱۵
۵. سوالات آماری.....	۲۱
۵.۱. آزمون آماری اول.....	۲۱
۵.۲. آزمون آماری دوم.....	۲۱

۱. مقدمه

در دنیای امروز، پلتفرم یوتیوب به یکی از بزرگترین شبکه‌های اجتماعی و اشتراک‌گذاری ویدیو در سطح جهان تبدیل شده است. میلیون‌ها ویدیو روزانه در این پلتفرم آپلود می‌شوند و بسیاری از این ویدیوها به سرعت در سراسر دنیا به تریند تبدیل می‌شوند. در این پروژه، هدف ما تحلیل و بررسی عواملی است که باعث می‌شوند ویدیوها در یوتیوب به تریند برسند. ما با استفاده از تکنیک‌های تحلیل داده‌های اکتشافی (EDA) و تجسم داده‌ها به بررسی الگوها و روندهای مرتبط با ویژگی‌های ویدیوها، تعاملات آن‌ها، و همچنین عوامل تاثیرگذار بر تریند شدن ویدیوها پرداختیم.

۲. آماده سازی و پیش پردازش داده‌ها

۲/۱. بارگذاری داده‌ها

برای این پروژه از داده‌های ویدیوهای تریند یوتیوب استفاده کردیم که شامل اطلاعاتی مانند شناسه ویدیو، تاریخ انتشار، عنوان ویدیو، نام کانال، دسته‌بندی ویدیو، تعداد بازدیدها، لایک‌ها، دیس‌لایک‌ها، تعداد کامنت‌ها و سایر ویژگی‌های مربوط به ویدیوها است. این داده‌ها از پلتفرم یوتیوب جمع‌آوری شده‌اند. این مجموعه داده شامل ۱۰ مجموعه داده جدا به تفکیک کشور های مختلف بود که برای بررسی و تحلیل کامل تر و جامع تر این مجموعه داده ها را با هم ادغام کردیم زیرا در آینده نیاز به تحلیل های کلی و در سطح جهان داشتیم.

۲/۲. بررسی داده های تکراری

در مرحله بعد، داده‌ها بررسی شدند تا مقادیر تکراری شناسایی شوند. پس از شناسایی، تمام ردیف‌های تکراری از مجموعه داده حذف شدند تا از تاثیرگذاری نامناسب بر تحلیل‌ها جلوگیری کنیم.

۲/۳. بررسی داده های گم‌شده

ابتدا داده‌ها را بررسی کردیم تا مقادیر گم‌شده‌ای که ممکن است تحلیل‌های بعدی را تحت تاثیر قرار دهند، شناسایی کنیم. در این پروژه، برخی از ستون‌ها (مانند توضیحات ویدیو) مقادیر گم‌شده داشتند که این مقادیر با توجه به شرایط مختلف یا حذف شده یا با مقادیر مناسب پر شدند.

۲/۴. بررسی نوع داده‌ها

نوع داده‌های مختلف در مجموعه داده بررسی شد تا مطمئن شویم که تمام ستون‌ها در قالب‌های صحیح قرار دارند. به‌ویژه، ستون‌های مربوط به تاریخ‌ها و زمان‌ها به فرمت مناسب تبدیل شدند تا تحلیل‌های زمانی دقیق انجام شود.

۲/۵. نگاشت شناسه دسته (Category Id) ویدئو به نام دسته (Category Name)

در این مرحله با استفاده‌های از فایل‌های با فرمت json شناسه دسته (Category Id) هر ویدئو را به نام دسته آن مانند Gaming, Film & Animation و ... نگاشت کردیم و یک ستون به نام category_name به مجموعه داده خود اضافه کردیم زیرا در تعداد زیادی از تجزیه‌ها به این ستون نیاز داشتیم.

۲/۶. مهندسی ویژگی‌ها (Feature Engineering)

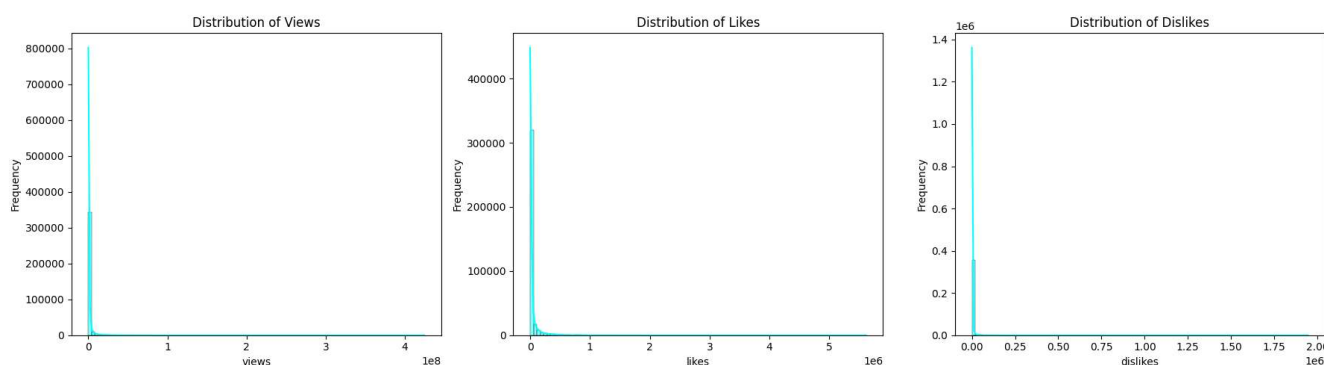
در این مرحله چند ستون جدید کاربردی برای تحلیل به مجموعه داده اضافه کردیم که به شرح زیر هستند:

- **pulish_dayofweek**: این ویژگی بیان کننده روز هفته تاریخ انتشار ویدئو است.
- **pulish_hour**: این ویژگی بیان کننده ساعت تاریخ انتشار ویدئو است.
- **publish_date**: این ویژگی بیان کننده تاریخ انتشار ویدئو است.
- **trending_dayweek**: این ویژگی بیان کننده روز هفته ترند شدن ویدئو می‌باشد.
- **like_ratio**: این ویژگی بیان کننده نرخ لایک یک ویدئو است و برابر تعداد لایک‌ها تقسیم بر تعداد بازدیدها می‌باشد.
- **dislike_ratio**: این ویژگی بیان کننده نرخ دیس‌لایک یک ویدئو است و برابر تعداد دیس‌لایک‌ها تقسیم بر تعداد بازدیدها می‌باشد.
- **engagement_ratio**: این ویژگی بیان کننده نرخ تعامل یک ویدئو است و برابر مجموع تعداد لایک‌ها، تعداد دیس‌لایک‌ها و تعداد کامنت‌ها می‌باشد.

۳. تحلیل سوالات اصلی

۳/۱. معیارهای تعامل (لایک، دیس لایک و کامنت) به طور کلی و در بین دسته های مختلف ویدئو چگونه توزیع می شوند؟

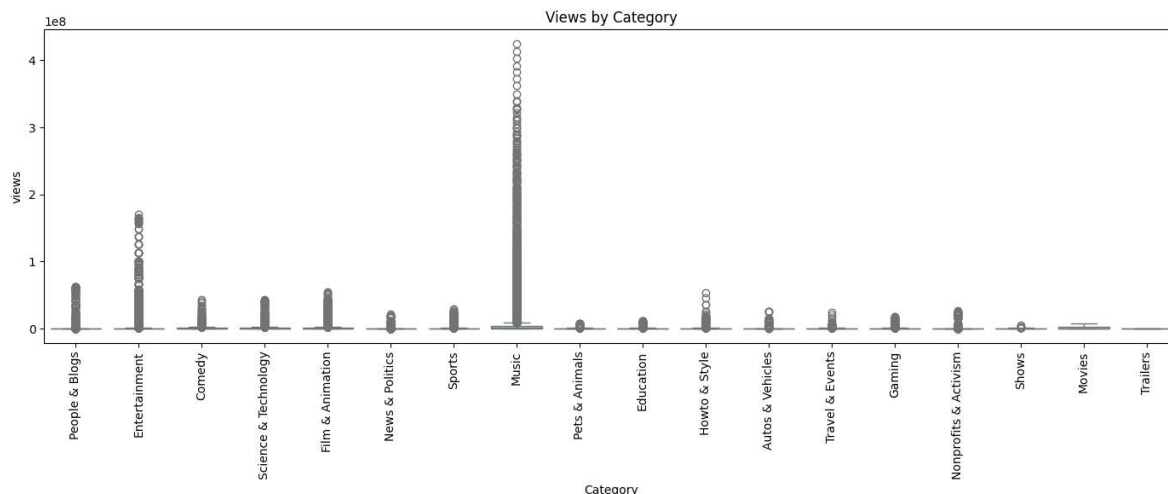
برای بررسی این سوال، ابتدا توزیع باز دیده ها، لایک ها و دیس لایک ها در کل داده ها به نمایش درآمد. سپس توزیع این معیارها در دسته بندی های مختلف ویدیوها بررسی شد. نتایج نشان داده شد که نمودار آن به شکل زیر می باشد.



شکل ۱- نمودار هیستوگرام توزیع بازدید، لایک و دیس لایک به طور کلی

توزیع باز دیده ها به شدت به سمت راست منحرف است، به طوری که اکثر ویدیوها کمتر از ۱ میلیون بازدید دارند. دم بلند تا حدود ۸۰ میلیون بازدید گسترش می یابد، که نشان دهنده چند ویدئوی با تعداد بازدید فوق العاده بالا است. این نشان می دهد که در حالی که اکثر ویدئوهای پرتعداد بینندگان متوسطی را به دست می آورند، تعداد کمی از ویدئوها محبوبیت زیادی پیدا می کنند.

لایک ها نیز دارای انحراف راست هستند و اکثر ویدیوها کمتر از ۰.۵ میلیون لایک دریافت می کنند. دم بلند در نمودار به حدود ۵ میلیون لایک می رسد، که نشان می دهد ویدیوهای بسیار پسندیده کمتر رایج هستند اما می توانند به تعامل قابل توجهی دست یابند. دیس لایک ها نیز به طور مشابه دارای انحراف راست هستند، به طوری که اکثر ویدیوها دیس لایک های بسیار کمی دارند. دم نمودار به حدود ۲ میلیون دیس لایک می رسد، که نشان می دهد تعداد کمی از ویدیوها بازخورد منفی قابل توجهی را به خود جلب می کند.



شکل ۲- نمودار جعبه ای توزیع بازدید، لایک و دیس لایک در بین دسته های مختلف

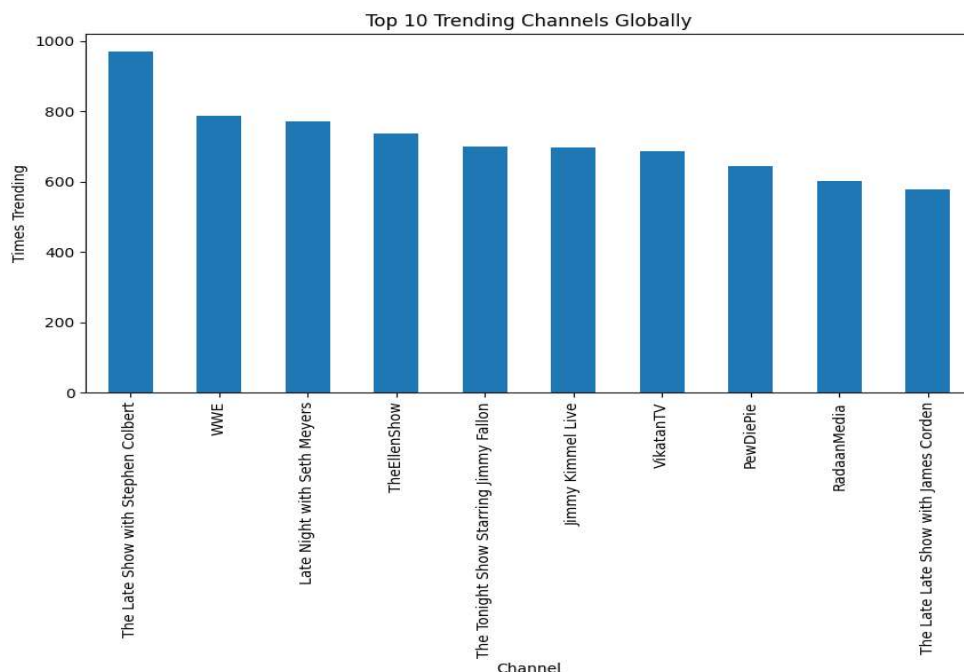
دسته "موسیقی" با بالاترین میانگین بازدیدها و دامنه وسیعی که به ۴ میلیارد بازدید می رسد برجسته است، که نشان می دهد موزیک ویدئوها اغلب در مقایسه با سایر دسته ها بیننده زیادی دارند. موارد پرت در این دسته پیشنهاد می کنند که برخی از موزیک ویدیوها با بازدیدهای فوق العاده بالا در فضای مجازی پخش می شوند.

بیشتر دسته ها (افراد و وبلاگ ها، کمدی و ...) دارای میانگین کمتر از ۰.۵ میلیارد بازدید هستند. دسته هایی مانند فیلم ها و سریال ها نقاط داده بسیار کمی دارند و بازدیدهای کمی دارند، احتمالاً به دلیل ویدئوهای پرطرفدار کمتر در این دسته ها.

موسیقی ها در بین ویدیوهای پرطرفدار، احتمالاً به دلیل جذابیت جهانی و قابلیت اشتراک گذاری بیشترین محبوبیت را دارند. دسته های دیگر اگرچه محبوب هستند، اما معمولاً بازدیدهای کمتری را جذب می کنند.

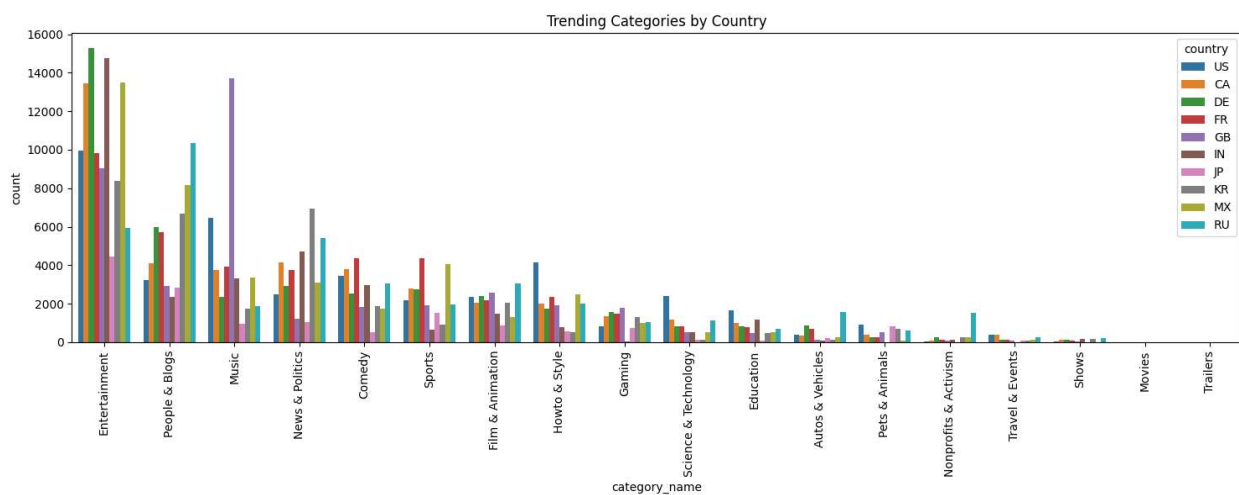
۳/۲. کدام کانال ها و دسته های ویدئویی YouTube در هر کشور و در سطح جهانی ترند شده اند؟

در این بخش به بررسی ترندترین کانال ها و دسته های ویدئویی در سطح جهانی و کشوری پرداختیم که نتایج آن به صورت زیر است.



شکل ۳- نمودار میله ای ترندترین کانال های جهان

نمودار میله ای ۱۰ کانال برتر جهانی پرطرفدار YouTube را بر اساس تعداد ویدیوهای پرطرفدار نشان می دهد. «The Tonight Show Starring Jimmy Fallon» با حدود ۹۰۰ ویدیوی پرطرفدار پیشواز است و پس از آن «WWE» و «TheEllenShow» هر کدام با تقریباً ۸۰۰ ویدیو قرار دارند. این نمودار نشان می دهد که کانال های متمرکز بر سرگرمی، به ویژه برنامه های گفتگوی آخر شب و محتوای کشتی، بر ترند شدن جهانی تسلط دارند که نشان دهنده جذابیت آنهاست.

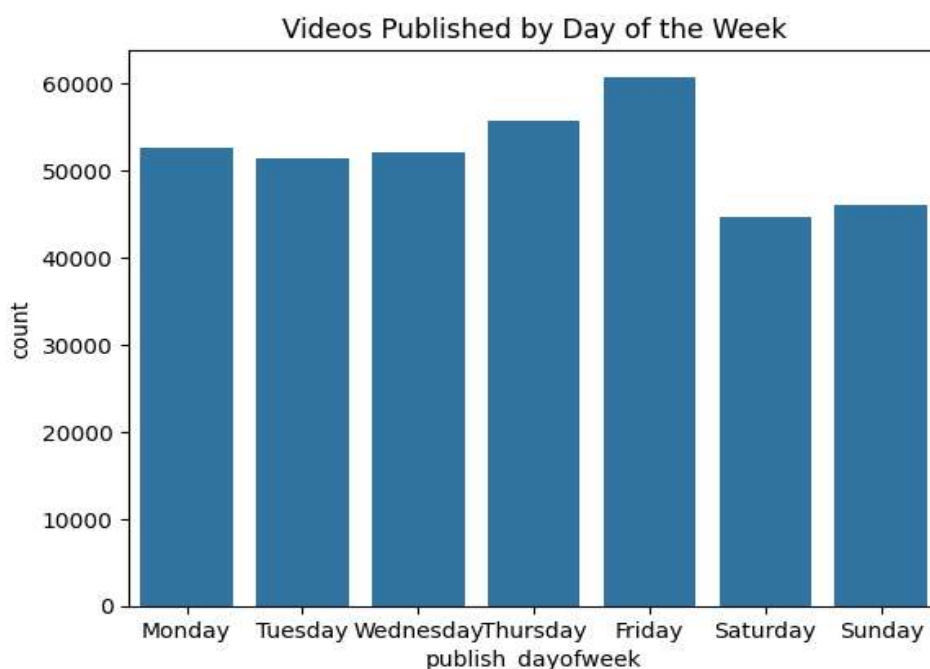


شکل ۴- نمودار میله ای ترندترین دسته های ویدئویی در کشور های مختلف

نمودار میله ای تعداد ویدیوهای پرتعداد YouTube را بر اساس دسته بندی در ۱۰ کشور نشان می دهد. دسته سرگرمی در اکثر کشورها غالب است، در آلمان با حدود ۱۴۰۰۰ ویدیو به اوج خود می رسد. دسته موسیقی نیز دارای تعداد بسیار زیادی ویدئو به ویژه در کشور های بریتانیا و ایالات متحده آمریکا می باشد. مردم و وبلاگ ها و اخبار و سیاست در سراسر مناطق، به ویژه در روسیه و کره، محبوب هستند. دسته هایی مانند کمدی، ورزش، و بازی روندهای متوسطی را نشان می دهند، در حالی که دسته های خاص (مانند خودروها و وسایل نقلیه، تریلرها، فیلم ها و ...) کمترین نمایش را دارند.

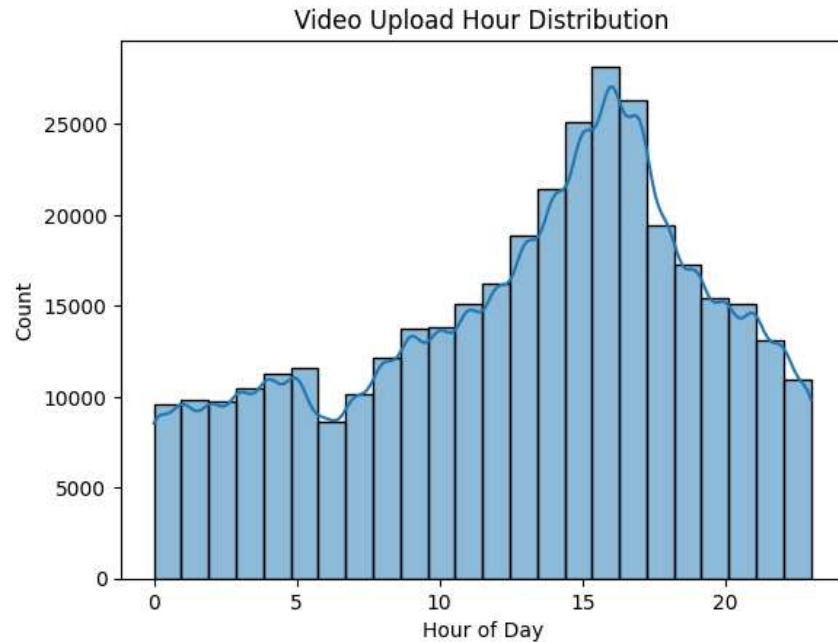
۳/۳. آیا الگوهای فصلی یا روز هفته در ویدئوهای پرتعداد وجود دارد؟ زمان آپلود روز/زمان بر تعامل ویدئو تأثیر می گذارد؟

در این بخش، تحلیل کردیم که آیا ویدیوها در روزهای خاصی از هفته یا فصول خاص سال ترند می شوند یا خیر.



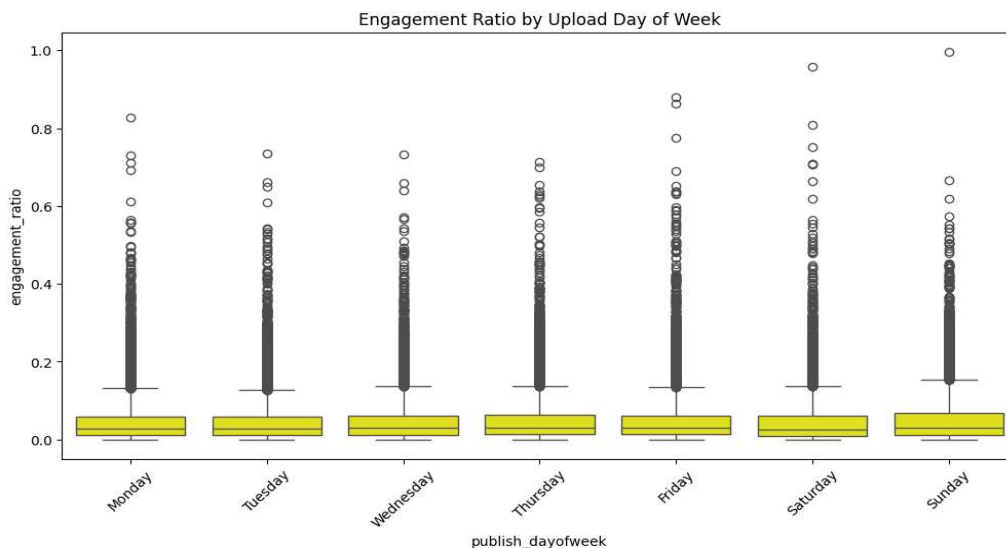
شکل ۵- نمودار میله ای تعداد ویدئو های ترند بر اساس روز هفته

نمودار میله ای تعداد ویدیوهای ترند یوتیوب را که بر اساس روزهای هفته را نشان می دهد. پنج شنبه و جمعه بیشترین تعداد ویدیو را دارند، هر کدام حدود ۶۰۰۰۰ ویدیو و پس از آن دوشنبه، سه شنبه و چهارشنبه هر کدام نزدیک به ۵۰۰۰۰ ویدیو قرار دارند. شنبه و یکشنبه ویدیوهای کمتری با تعداد حدود ۴۵۰۰۰ مشاهده می شود. این نشان می دهد که سازندگان تمایل دارند بیشتر در اواسط تا آخر هفته منتشر کنند.



شکل ۶- نمودار هیستوگرام توزیع ویدئوهای ترند بر حسب ساعت روز

هیستوگرام توزیع ویدئوهای ترند یوتیوب را بر حسب ساعت آپلود از روز نشان می دهد. بیشتر آپلودها بین ساعت ۱۵:۰۰ تا ۲۰:۰۰ انجام می شود، که در حدود ساعت ۱۷:۰۰ تا ۱۸:۰۰ با بیش از ۲۵۰۰۰ ویدیو به اوج خود می رسد، که نشان دهنده ترجیح قوی برای آپلود در اواخر بعد از ظهر است. این الگو ممکن است با سازندگانی که بیشترین در دسترس بودن مخاطب را هدف قرار می دهند، مطابقت داشته باشد و احتمال ترند شدن را افزایش دهد.

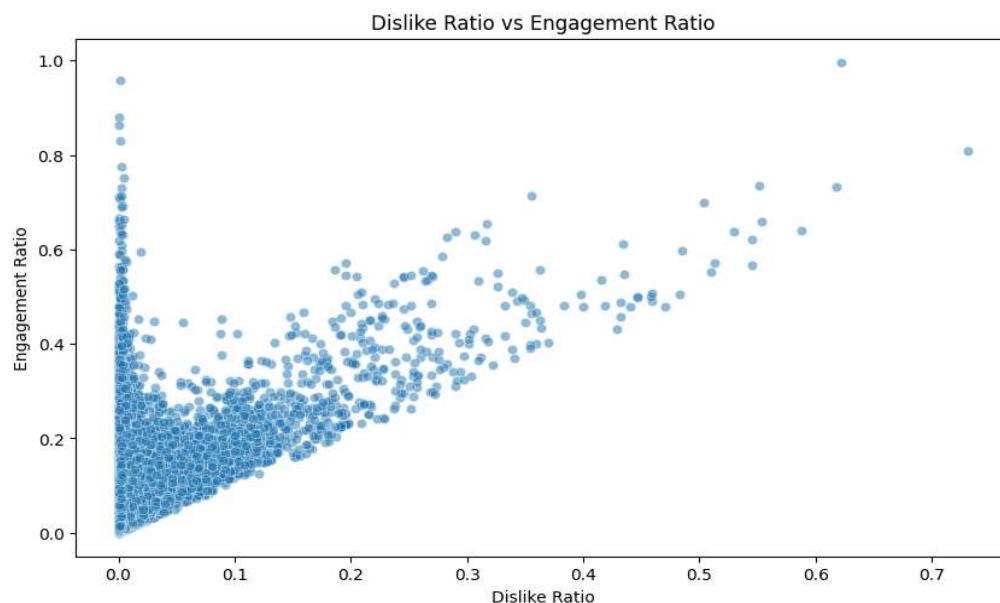


شکل ۷- نمودار جعبه‌ای نرخ تعامل ویدئوهای ترند بر اساس روز آپلود

نمودار جعبه‌ای نرخ تعامل ویدیوهای ترند YouTube را بر اساس روز آپلود در هفته را نشان می‌دهد. میانه نرخ تعامل به طور مداوم در تمام روزها حدود ۰.۰۵ است. موارد دورافتاده در هر روز وجود دارند، با برخی از ویدیوها به نسبت تعامل تا ۱.۰ می‌رسند که نشان‌دهنده محتوای بسیار جذاب است. هیچ روند واضحی در تعامل روزانه وجود ندارد، که نشان می‌دهد روز آپلود تأثیر کمی بر نرخ تعامل دارد.

۳/۴. آیا ویدئوهای با نرخ دیس‌لایک بالا بیشتر میزان تعامل بیشتری از ویدئوهای عموماً لایک شده دارند؟

در این بخش، رابطه بین نسبت دیس‌لایک‌ها و تعاملات ویدیوها بررسی شد.

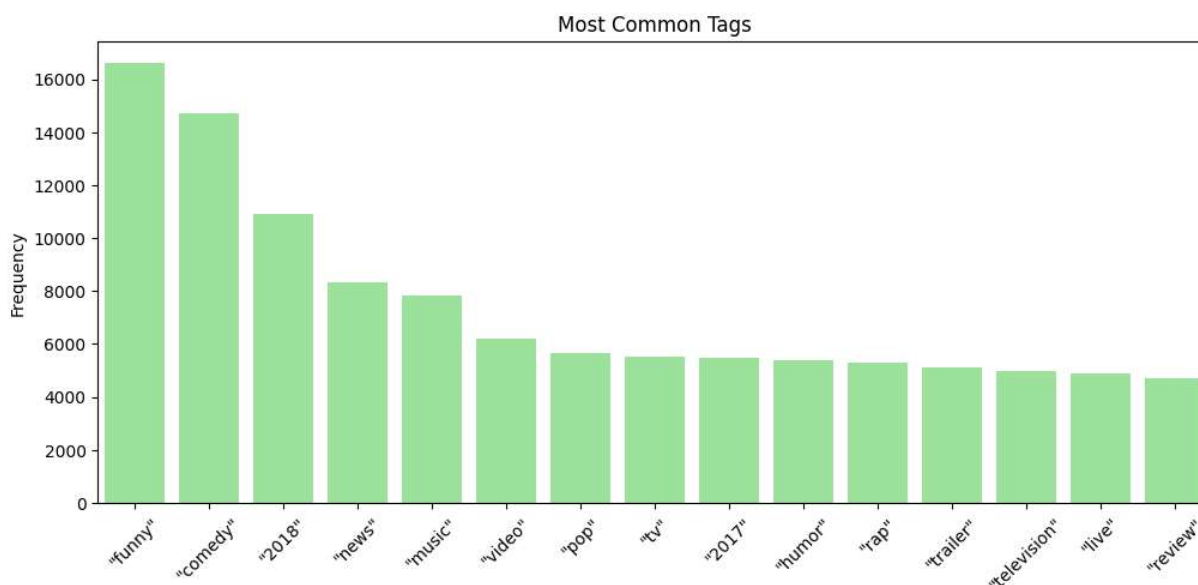


شکل ۸- نمودار پراکندگی رابطه بین دیس‌لایک‌ها و تعاملات

نمودار پراکندگی رابطه بین نسبت نرخ دیس‌لایک و نرخ تعامل را برای ویدیوهای ترند یوتیوب بررسی می‌کند. بیشتر ویدیوها با نسبت دیس‌لایک پایین با نسبت تعامل بین ۰.۰ تا ۰.۲ خوشه‌بندی می‌شوند، که نشان می‌دهد ویدیوهایی با دیس‌لایک‌های کمتر عموماً تعامل متوسطی دارند. از آنجایی که نرخ دیس‌لایک بیش از ۰.۲ افزایش می‌یابد، نسبت‌های تعامل بسیار متفاوت است، به طوری که برخی از ویدیوها به ۰.۸ می‌رسند، که نشان می‌دهد ویدیوهای جنجالی (نرخ دیس‌لایک بالا) همچنان می‌توانند تعامل بالایی داشته باشند. در کل ارتباط زیادی بین این دو ویژگی برقرار نیست و همبستگی محاسبه شده آنها نیز ۰.۳۳ می‌باشد.

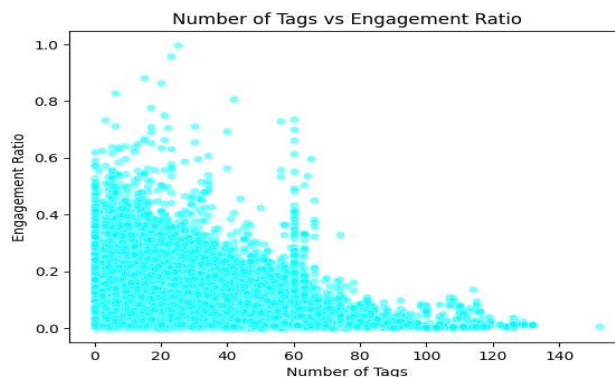
۳/۵. برچسب‌های ویدیو چگونه بر تعامل تأثیر می‌گذارند و کدام برچسب‌ها بیشتر در ویدیوهای طرفدار استفاده می‌شوند؟

تگ‌ها نقش مهمی در شناسایی محتوای ویدیوها دارند. در این بخش، بررسی کردیم که تگ‌های ویدیویی چگونه بر تعاملات تأثیر می‌گذارند



شکل ۹- نمودار میله‌ای رایج‌ترین برچسب‌ها

نمودار میله‌ای تعداد رایج‌ترین برچسب‌ها را در ویدیوهای ترند یوتیوب نشان می‌دهد. "Funny" پربازدیدترین برچسب است که در حدود ۱۶۰۰۰ ویدئو ظاهر می‌شود و پس از آن "کمدی" (۱۴۰۰۰) و "۲۰۱۸" (۱۲۰۰۰) قرار دارند. سایر برچسب‌های محبوب عبارتند از «اخبار»، «موسیقی» و «ویدئو» که هر کدام حدود ۸۰۰۰ تا ۱۰۰۰۰ است.

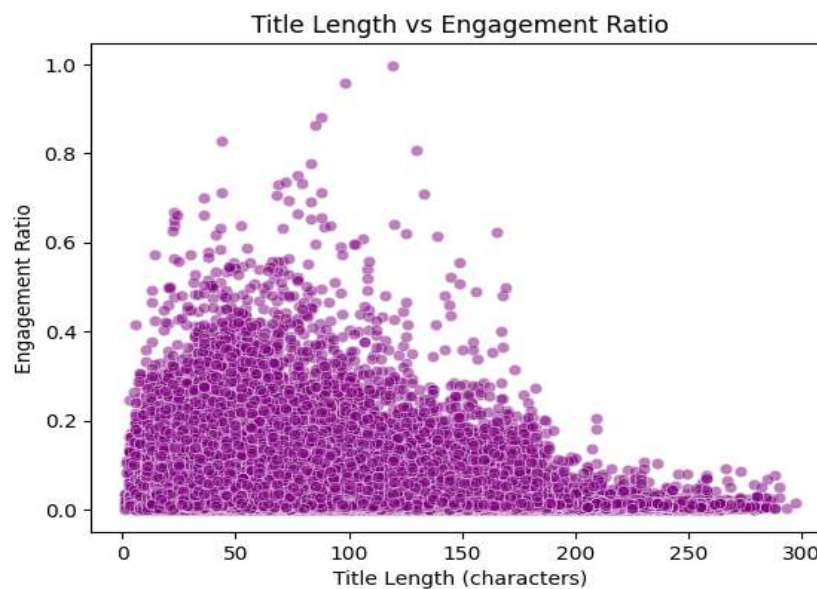


شکل ۱۰- نمودار پراکندگی رابطه بین تعداد برچسب‌ها و نرخ تعامل

نمودار پراکندگی رابطه بین تعداد تگ ها و نرخ تعامل را برای ویدیوهای ترند یوتیوب بررسی می کند. بیشتر ویدیوها با کمتر از ۴۰ برچسب و نسبت تعامل زیر ۰.۲ خوشه می شوند، که نشان می دهد تعداد بیشتر برچسب ها با تعامل بهتر ارتباط قوی ندارد. ویدیوهایی با برچسب های بیشتر (تا ۱۴۰) کم هستند و نرخ تعامل متفاوتی را نشان می دهند، با برخی از نقاط پرت به نسبت های بالای ۰.۸ می رسند. این نشان می دهد که اگرچه تعداد متوسطی از برچسب ها در میان ویدیوهای ترند رایج است، برچسب گذاری بیش از حد، تعامل بالاتر را تضمین نمی کند.

۳/۶. طول عنوان ویدیو چگونه بر تعامل تأثیر می گذارد؟

در این بخش، رابطه بین طول عنوان ویدیوها و میزان تعامل آن ها بررسی شد.

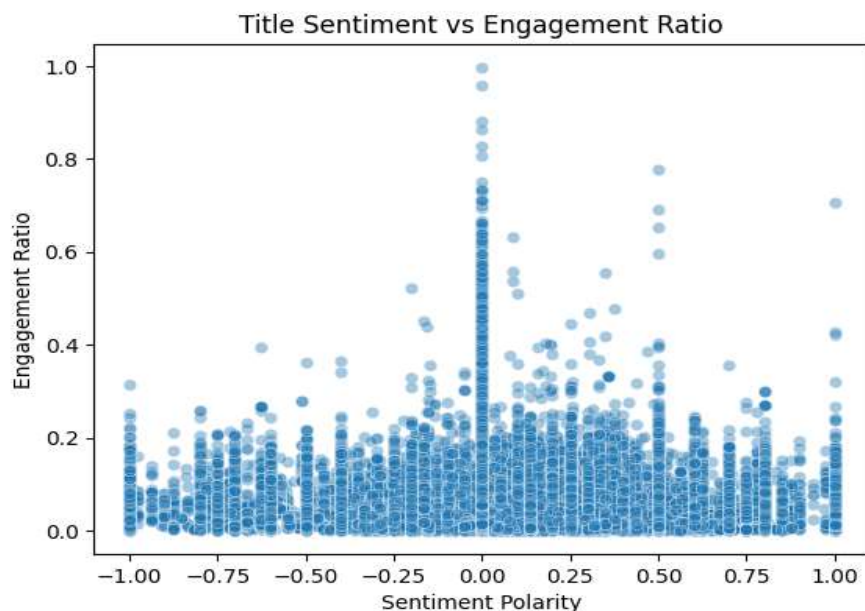


شکل ۱۱- نمودار پراکندگی رابطه بین طول عنوان ویدئو و نرخ تعامل

نمودار پراکندگی رابطه بین طول عنوان (تعداد کاراکتر) و نرخ تعامل را برای ویدئوهای ترند یوتیوب بررسی می کند. اکثر ویدیوها عناوین کمتر از ۱۰۰ کاراکتر دارند، با نرخ تعامل زیر ۰.۲ خوشه بندی می شوند، که نشان می دهد عناوین کوتاه تر در میان ویدیوهای ترند رایج تر هستند. با افزایش طول عنوان بیش از ۱۰۰ کاراکتر، نقاط داده پراکنده تر می شوند و نسبت های تعامل به طور گسترده ای متفاوت هستند، برخی به ۱.۰ می رسند. این نشان می دهد که هیچ ارتباط قوی بین طول عنوان و تعامل وجود ندارد. در حالی که عناوین کوتاه تر غالب هستند.

۳/۷. آیا رابطه ای بین احساس عنوان ویدیو و تعامل وجود دارد؟

در این قسمت، احساسات عنوان‌های ویدیویی بررسی شد و ارتباط آن‌ها با میزان تعامل تحلیل شد.

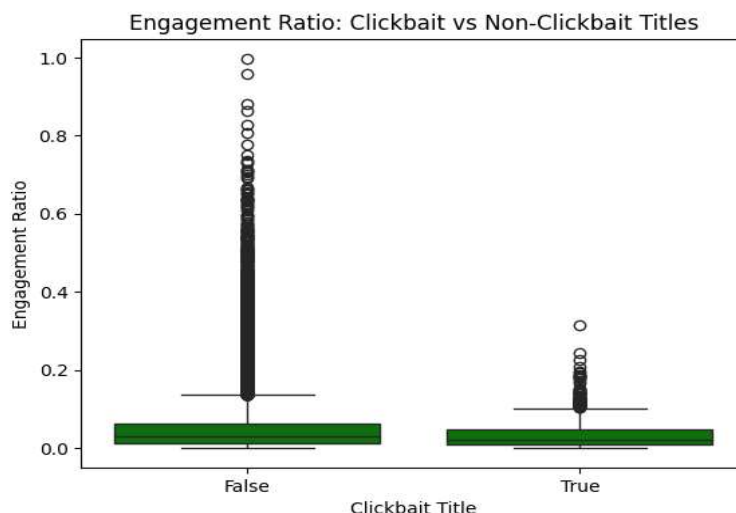


شکل ۱۲- نمودار پراکندگی رابطه بین احساسات عنوان ویدئو و نرخ تعامل

نمودار پراکندگی رابطه بین قطب احساسات عنوان (از -۱ تا ۱) و نرخ تعامل را برای ویدیوهای ترند یوتیوب بررسی می‌کند. بیشتر ویدئوها حول احساسات خنثی (۰.۰) با نسبت تعامل زیر ۰.۲ جمع می‌شوند، که نشان می‌دهد عناوین خنثی در بین ویدیوهای پرتیرافدار رایج هستند. ویدیوهای دارای احساسات مثبت (۰.۲۵ تا ۱.۰) تراکم کمی بالاتر را نشان می‌دهند، با برخی از موارد دور از نسبت تعامل تا ۱.۰. عناوین احساسات منفی (-۱.۰ تا -۰.۲۵) کمتر متداول هستند. نمودار نشان دهنده یک رابطه ضعیف است.

۳/۸. آیا عناوین کلیک (به عنوان مثال، "تکان دهنده"، "باید تماشا کنید") منجر به تعامل بیشتر می‌شود؟

در این قسمت، ارتباط بین ویدئوهای با عنوان clickbait-style و میزان تعامل تحلیل شد. با استفاده از کلمات کلیدی از جمله shocking, must watch, unbelievable, insane و ... عنوان‌هایی که شامل این کلمات کلیدی بودند شناسایی شدند و در نهایت ارتباط آن‌ها با نرخ تعامل بررسی شد که به صورت زیر می‌باشد.



شکل ۱۳- نمودار جعبه‌ای رابطه بین ویدئوهای clickbait-style و نرخ تعامل

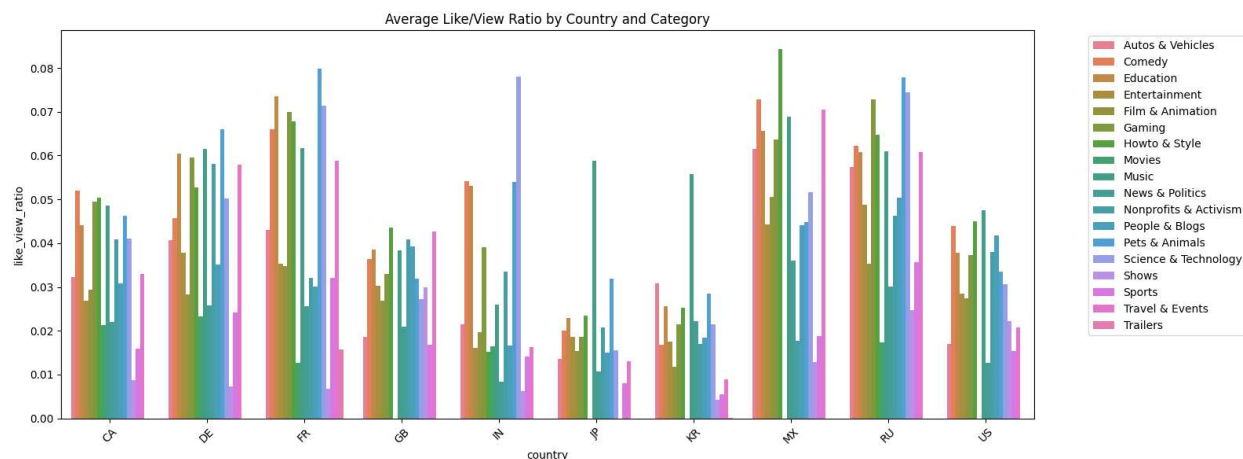
نمودار جعبه‌ای نرخ‌های تعامل ویدئوهای پرتعداد یوتیوب را با عناوین clickbait (به عنوان مثال، «تکان‌دهنده»، «باید تماشا کنید») در مقابل عناوین غیرکلیک مقایسه می‌کند. هر دو گروه دارای نرخ تعامل میانه مشابه در حدود ۰.۰۵ هستند، که نشان دهنده تفاوت معنی داری در تعامل نیست. با این حال، عناوین غیرکلیکی دارای نقاط پرت بیشتری هستند، با برخی از آنها نسبت تعامل تا ۱.۰ می‌رسد، در حالی که عناوین کلیکی عناوین پرت کمتری دارند و در حدود ۰.۴ به اوج می‌رسند. این نشان می‌دهد که عناوین clickbait به طور مداوم منجر به تعامل بالاتر نمی‌شوند.

۴. داستان سرایی و سوالات اضافی

در این مرحله، یک داستان اصلی در مورد "عوامل تاثیرگذار بر ترند شدن ویدیوها" ایجاد شد. سپس سوالات اضافی و مرتبط با این داستان مطرح شد و پاسخ‌های آن‌ها از طریق تجزیه و تحلیل داده‌ها به دست آمد.

۴/۱. سوالات اضافی

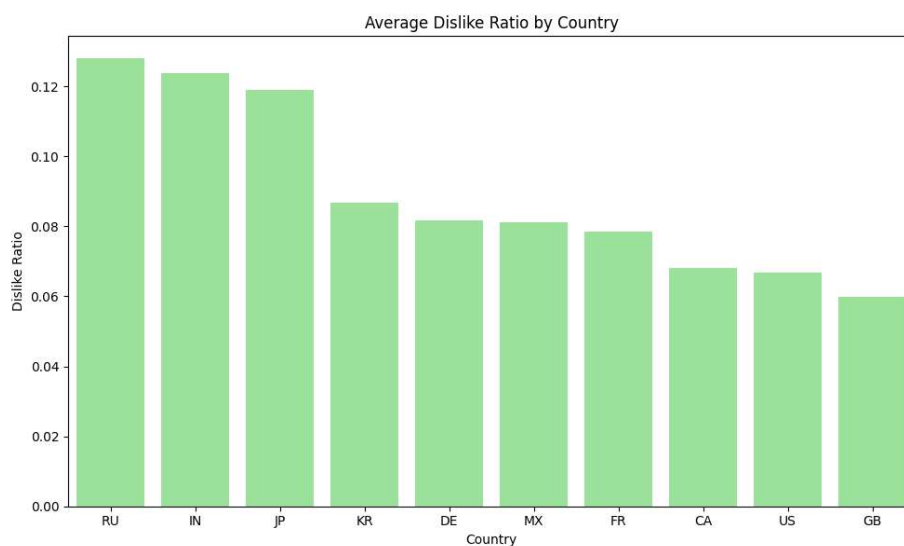
۱. میانگین تعامل کشورهای مختلف برای دسته های ویدیویی مشابه چگونه است؟



شکل ۱۴- نمودار میله‌ای میانگین تعامل (بازدید، لایک) کشورها برای دسته‌های ویدیویی

نمودار میله‌ای میانگین نسبت تعامل ویدیوهای ترند یوتیوب را در کشورها و دسته‌ها نشان می‌دهد. در همه کشورها، دسته‌هایی مانند فیلم و انیمیشن (زرد) و موسیقی (سبز) اغلب تعامل بیشتری را نشان می‌دهند، با نسبت‌هایی که در کشورهایی مانند آلمان و بریتانیا به ۰.۰۸ می‌رسد. به طور کلی، نسبت‌های تعامل بر اساس دسته‌بندی بیشتر از کشورها متفاوت است، با موسیقی و فیلم و انیمیشن به طور مداوم مخاطبان را به طور مؤثرتری درگیر می‌کنند، در حالی که اخبار و سیاست تمایل به مشارکت کمتری در سطح جهانی دارد.

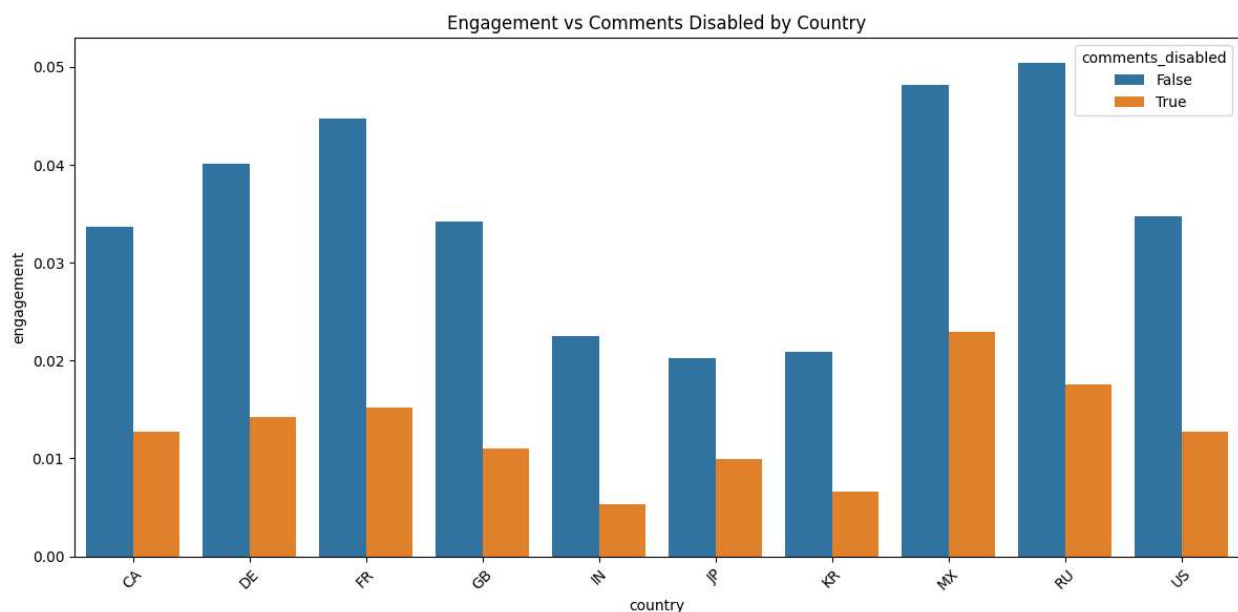
۲. کدام کشورها بالاترین تحمل یا واکنش را به ویدیوهای جنجالی نشان می‌دهند (بر اساس نسبت دیس لایک)؟



شکل ۱۵- نمودار میله‌ای نسبت دیس لایک بر اساس کشور

نمودار میله ای میانگین نسبت دیس لایک ویدئوهای ترند یوتیوب را بر اساس کشور نشان می دهد. روسیه دارای بالاترین نسبت دیس لایک با حدود ۰.۱۲ است، پس از آن هند و ژاپن با تقریباً ۰.۱۱ قرار دارند، که نشان می دهد ویدئوهای ترند در این کشورها واکنش های منفی بیشتری را به خود جلب می کنند که نشان دهنده تحمل یا تعداد بیشتر محتوای جنجالی است. کره جنوبی و آلمان با نسبت های حدود ۰.۰۸-۰.۰۹ دنبال می شوند. این نشان می دهد که تفاوت های فرهنگی بر نحوه واکنش مخاطبان به ویدئوهای ترند تأثیر می گذارد.

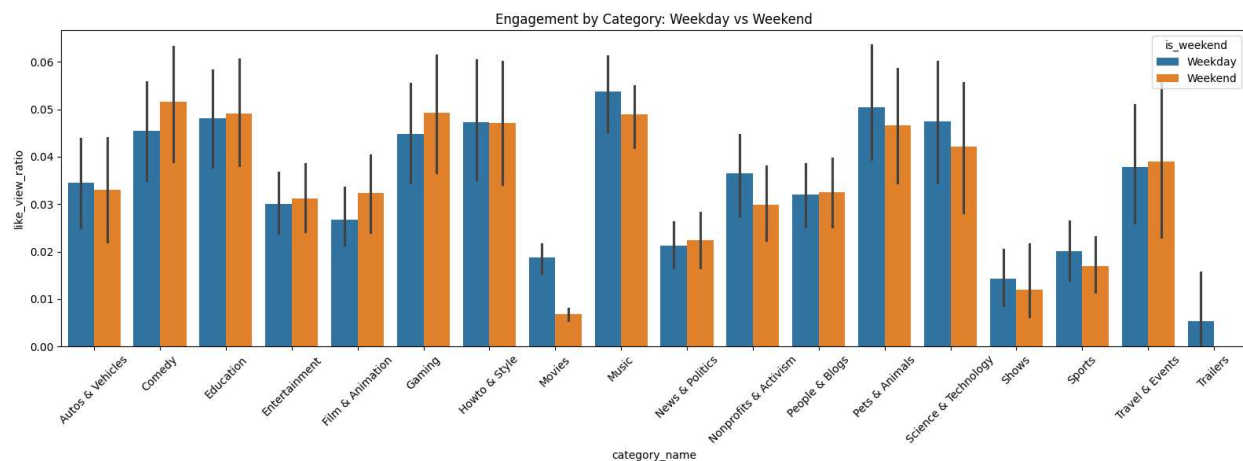
۳. آیا غیرفعال کردن نظرات بر تعامل در برخی کشورها بیشتر از کشورهای دیگر تأثیر می گذارد؟



شکل ۱۶- نمودار میله ای تعامل بر اساس فعال بودن یا نبودن نظرات

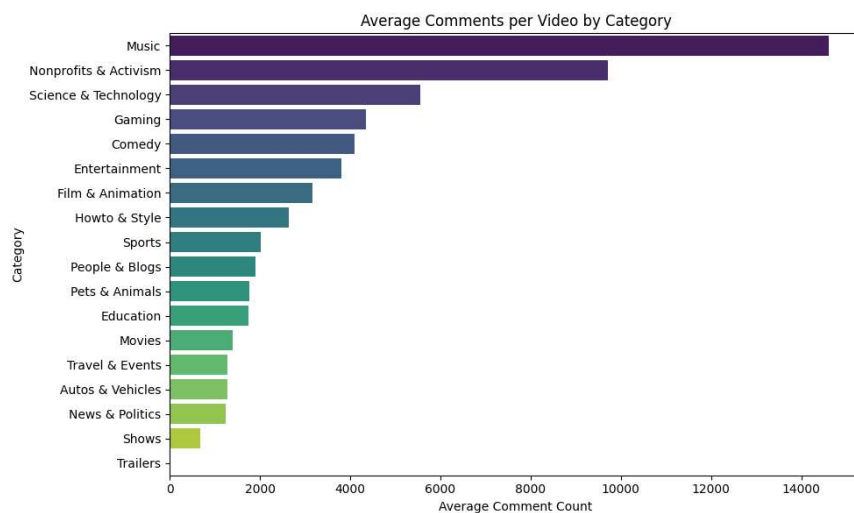
نمودار میله ای تعامل را برای ویدئوهای ترند یوتیوب با نظرات غیرفعال (نارنجی) در مقابل فعال (آبی) در سراسر کشورها مقایسه می کند. در برخی کشورها تعامل بدون در نظر گرفتن وضعیت نظر مشابه است، در اطراف ۰.۰۳-۰.۰۵ قرار دارد، که حداقل تأثیر را نشان می دهد. به طور کلی به نظر می رسد که فعال بودن یا نبودن نظرات بر تعامل ویدئو بسیار تأثیر گذار است.

۴. آیا الگوهای تعامل برای دسته بندی های ویدئویی مشابه زمانی که در روزهای هفته در مقایسه با آخر هفته در کشورهای مختلف پست می شوند، متفاوت است؟



شکل ۱۷- نمودار میله‌ای تعامل بر اساس دسته و زمان پست شدن ویدئو (وسط هفته و آخر هفته)

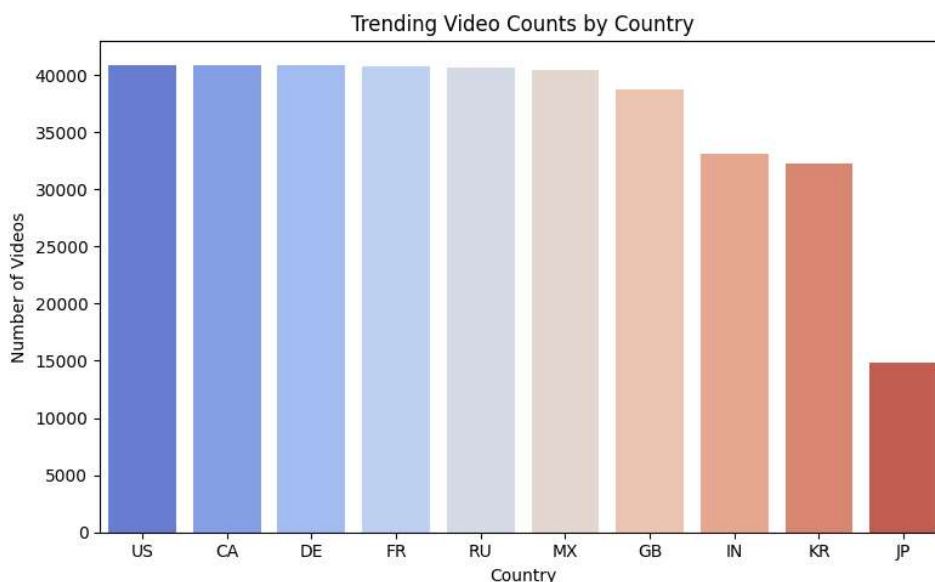
۵. کدام دسته های ویدئویی به طور متوسط بیشترین نظر را دریافت می کنند؟



شکل ۱۸- نمودار میله‌ای میانگین تعداد نظرات بر اساس دسته بندی ویدئوها

نمودار میله ای میانگین تعداد نظرات در هر ویدئو را بر اساس دسته بندی ویدئوهای ترند نشان می دهد. موسیقی با بالاترین میانگین با حدود ۱۴۰۰۰ نظر در هر ویدیو پیشتاز است و پس از آن سازمان‌های غیرانتفاعی و فعالیت‌ها و علم و فناوری هر کدام با میانگین ۱۰۰۰۰ تا ۱۲۰۰۰ نظر قرار دارند. بازی و کمدی نیز با حدود ۸۰۰۰ تا ۹۰۰۰ نظر در رتبه های بالایی قرار دارند.

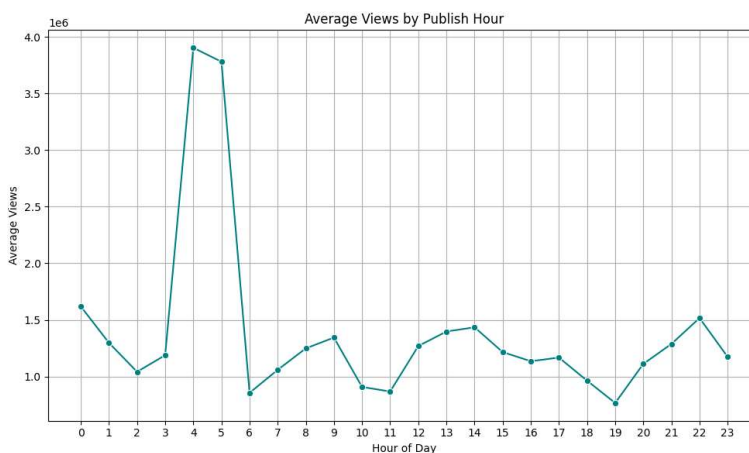
۶. کدام کشورها ترندترین ویدئوها را دارند؟



شکل ۱۹- نمودار میله‌ای کشورها با بیشترین تعداد ویدئو ترند

نمودار میله‌ای تعداد ویدئوهای ترند را بر اساس کشور نشان می‌دهد. ایالات متحده، کانادا، آلمان، فرانسه و روسیه با حدود ۴۰,۰۰۰ ویدئوی ترند هر کدام پیش‌تاز هستند که نشان‌دهنده حجم بالای محتوای ترند در این مناطق است. مکزیک و بریتانیا تقریباً ۳۵,۰۰۰ ویدئو دارند، در حالی که هند و کره جنوبی حدود ۳۰,۰۰۰ ویدئو دارند.

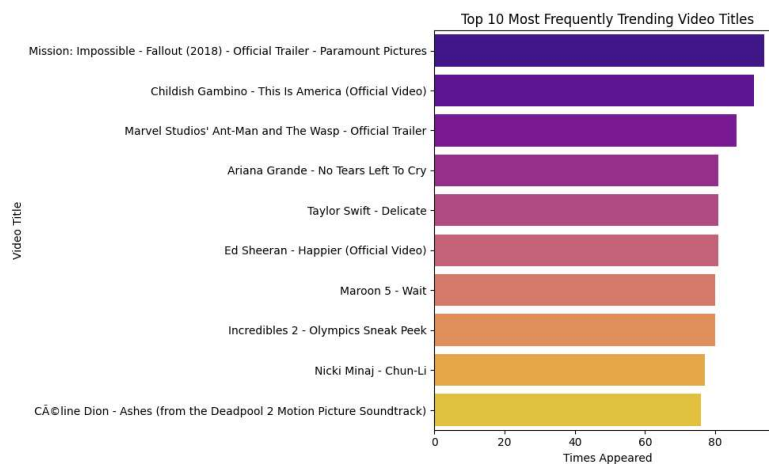
۷. کدام ساعات از روز منجر به بالاترین میانگین بازدید می‌شود؟



شکل ۲۰- نمودار خطی میانگین بازدیدها بر اساس ساعت انتشار

نمودار خط میانگین بازدیدها برای ویدئوهای ترند یوتیوب را بر اساس ساعت انتشار نشان می دهد. بالاترین میانگین بازدیدها در ساعت ۴:۰۰ رخ می دهد که در حدود ۴ میلیون بازدید به اوج خود می رسد و پس از آن افت شدیدی رخ می دهد. سایر اوج های قابل توجه در ساعت ۱۰:۰۰ و ۱۹:۰۰ هستند که هر کدام حدود ۱.۵ میلیون بازدید دارند.

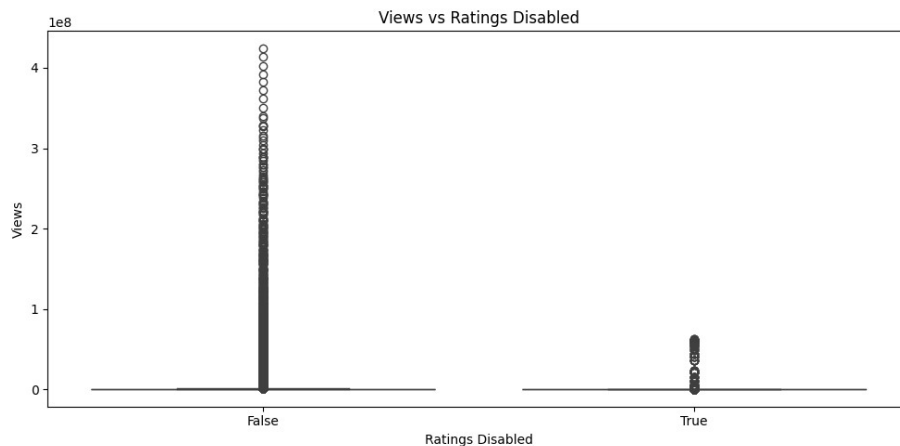
۸. ترندترین عناوین ویدئویی کدامند؟



شکل ۲۱- نمودار میله ای ترندترین عناوین ویدئویی

نمودار خط میانگین بازدیدها برای ویدئوهای ترند یوتیوب را بر اساس ساعت انتشار نشان می دهد. بالاترین میانگین بازدیدها در ساعت ۴:۰۰ رخ می دهد که در حدود ۴ میلیون بازدید به اوج خود می رسد و پس از آن افت شدیدی رخ می دهد. سایر اوج های قابل توجه در ساعت ۱۰:۰۰ و ۱۹:۰۰ هستند که هر کدام حدود ۱.۵ میلیون بازدید دارند.

۹. آیا غیرفعال کردن امتیاز (rating) روی بازدیدها تأثیر می گذارد؟



شکل ۲۲- نمودار جعبه‌ای تعداد بازدیدها بر اساس فعال بودن یا نبودن امتیاز

نمودار جعبه‌ای تعداد بازدیدها بر اساس فعال بودن یا نبودن امتیاز نشان می‌دهد که تفاوت به سزایی در تعداد بازدید در این دو گروه وجود دارد و این عامل می‌تواند تاثیرگذار باشد.

۵. سوالات آماری

۵/۱. آزمون آماری اول

آیا ارتباط معنی داری بین روز هفته انتشار یک ویدیو و احتمال ترند بودن آن وجود دارد؟

برای بررسی ارتباط بین روز هفته انتشار ویدیو و احتمال ترند شدن آن، آزمون خی-دو انجام شد. نتایج نشان داد که بین روز هفته انتشار و احتمال ترند شدن ویدیو ارتباط معنی داری وجود دارد.

- آماره خی-دو: ۶۹۵۰.۷۹
- مقدار p-value: صفر

۵/۲. آزمون آماری دوم

آیا تفاوت قابل توجهی در تعامل بیننده (نسبت لایک به بازدید) در دسته های مختلف ویدئو وجود دارد؟

برای بررسی تفاوت تعاملات بین دسته‌بندی‌های مختلف ویدئوها، آزمون Kruskal-Wallis H انجام شد. نتایج نشان داد که تفاوت معنی‌داری بین تعاملات در دسته‌بندی‌های مختلف ویدئوها وجود دارد.

- آماره H: ۴۵۰۴۲.۸۴

- مقدار p-value: صفر