Youtube Dataset

Description of the dataset:
The dataset contains several months of data on daily trending YouTube videos in the United States. This data was collected by YouTube, which was used to analyze the top trending videos for the year. Variables that were used to analyze this are listed below:
Trending date: Date on which the video is trending.
Title: Title of the video.
Channel title: Channel that uploaded the video.
Category id: Category (genre) video fits in.
Publish time: Date and time video was uploaded.
Tags: Specialized words that the Youtube algorithm uses to suggest content to individual users.
Views: Total number of views the video got.
Likes: Total number of likes the video got.
Dislikes: Total number of dislikes the video got.
Comment count: Total number of comments the video got.
Thumbnail link: Sharable link to the thumbnail of the video.
Comments disabled: To check if the comment section is disabled or not.
Ratings disabled: To check if the ratings of the video are disabled.
Video removed: To check if the video is still available on the channel.

Description:
We aim to analyze the given dataset to determine the best possible combination of the above mentioned variables to produce a trending video on YouTube. YouTube's ad revenue in the year 2023 was $31.5 billion. We would also like to analyze which combination of variables form the most watched or 'interacted with' video, which can be used by companies to advertise their product. We have listed down a few business questions that we would like to answer by the end of this project.

Business Questions:

1) What factors affect how popular a YouTube video can be in the USA?
2) What regions (category ID) have the most views that could be used for advertising?
3) Does engagement being disabled significantly affect the view count on a video?
4) How would the title length affect the view count of a video?
5) Based on this data, what strategy can an advertiser use to pick a type of youtube video to advertise their product on?