YouTube Videos Likeability Predictor

Data Science Programming (ISM 6930)
University of South Florida

Adrian Texeira

Ankit Singh

Tarjani Shah

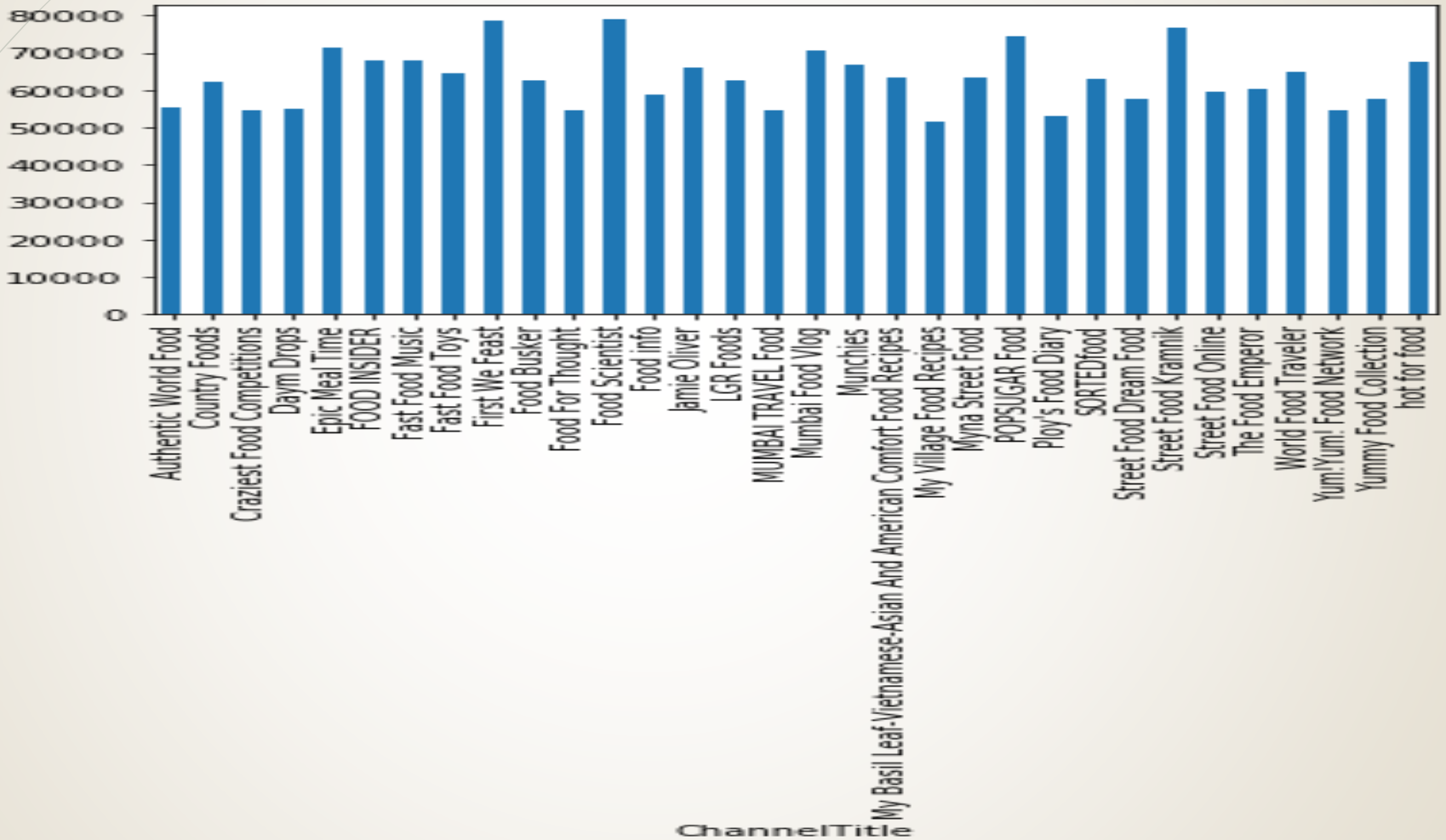
Motivation

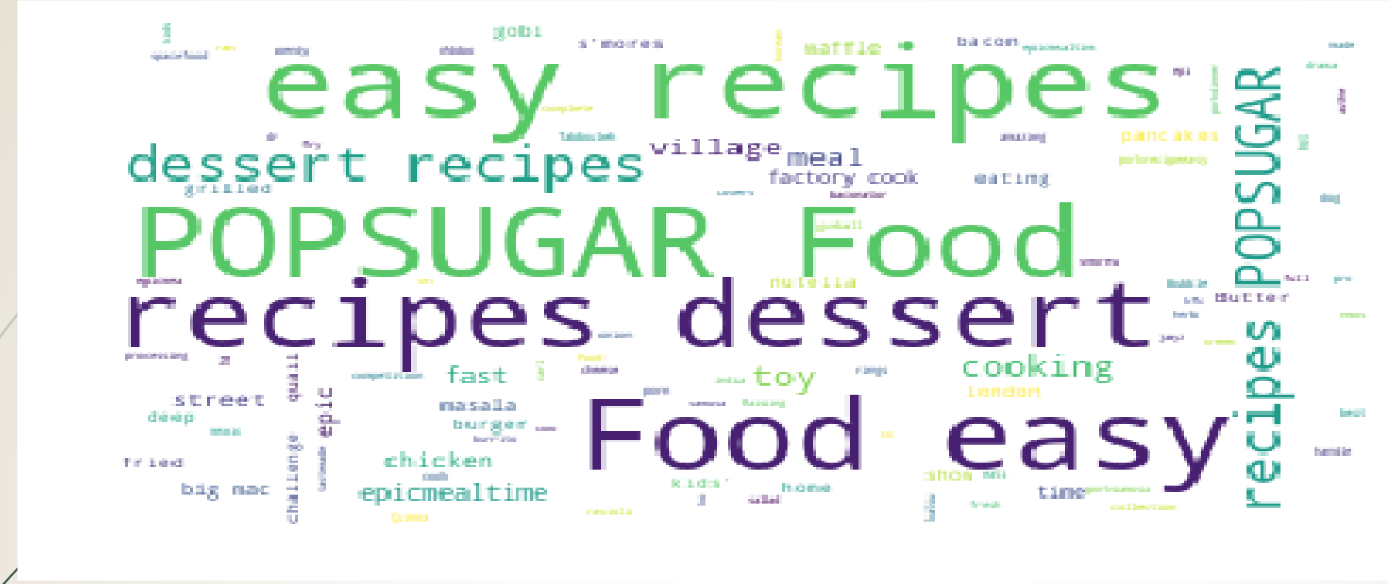
- Over the past 5 years YouTube has paid out more than $5 billion to YouTube content creators.
- With more and more companies turning to YouTube influencers to capture the millennial audience, getting people to watch your videos on YouTube is becoming increasingly lucrative.
- Our goal is to create a model that can help influencers predict the likeability for their next video.

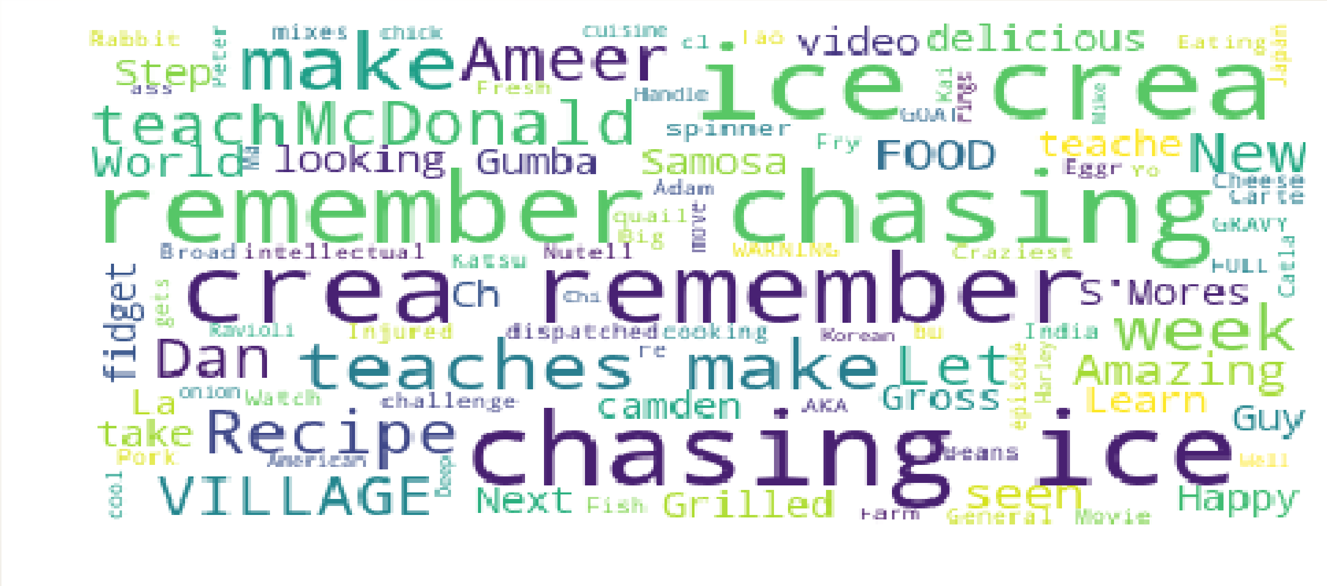
Data Preprocessing

- We leveraged Google Developer Console's YouTube Data API functions.
- Using various functions for fetching Video and video related metadata through channels and playlists.
- Performed Sentiment Analysis for the comments of the videos and added count of comments as positive, negative and neutral to the existing data as new FEATURES.
- Calculated the dependent variable i.e LikePerView and View-Difference using features like LikeCount, ViewCount and PreviousViewCount

Exploratory Data Analysis




easy recipes
dessert recipes
POPSUGAR Food
recipes dessert
Food easy
POPSUGAR
recipes
village
meal
factory cook
eating
cooking
fast
toy
street
masala
burger
chicken
epicmealtime
epic
big mac
time


make
Ameer
ice crea
video
delicious
Step
teach
McDonald
World
looking
Gumba
Samosa
FOOD
teache
New
remember
chasing
crea
remember
S'Mores
week
fidget
Dan
teaches
make
Let
Amazing
camden
Gross
Learn
take
Recipe
chasing
ice
Guy
VILLAGE
Next
Grilled
seen
Happy

Machine Learning Models

- Using models like GradientBoostingRegresor and XGBoostRegressor we were able to predict Likeability of a video based on features like:
  - Channel View count
  - Channel subscriber count
  - Channel age
  - Previous video views
  - Video age
  - Duration

OBSERVATION & RECOMMENDATIONS

- Predicting Views was what we started with initially soon to realize it depends on many other factors like content etc.
- One of the important observations were that just uploading a video and waiting for days to pass by so that views can increase is a bad idea, one factor which affects views or likeability of a video in general is the CHANNEL.
- Another point would be the DURATION of the video, if the video is of a shorter duration and conveys sufficient information it will get views and likes. That's kind of obvious!

Future Scope

- Thumbnails and Title are features that we initially planned to used but couldn't. If running deep learning models can extract hidden features then that might be useful to increase likes.
- Another factor contributing to likes of videos would be content, by using models like topic modelling and finding its relevance with respect to the topic can be used as a feature.
- This methodology can be extended to different genres of videos.

References

- https://towardsdatascience.com/youtube-views-predictor-9ec573090acb
- All API calls are referred from this source: https://developers.google.com/youtube/v3/docs/videos
- Progress Bar code referenced from: https://github.com/allenwang28/YouTube-Virality-Predictor/blob/master/scripts/get_last_video_count.ipynb
- Emoji removal: http://stackoverflow.com/a/13752628/6762004
- Code:

  https://github.com/allenwang28/YouTube-Virality-Predictor

THANK YOU!!!