

SOCIAL MEDIA ANALYSIS USING NATURAL LANGUAGE PROCESSING TECHNIQUES

JYOTIKA SINGH

Jyotika Singh

Director of Data Science
at
ICX Media, Inc.



<https://twitter.com/jyotikasingh>



<https://github.com/jsingh811>



<https://www.linkedin.com/in/jyotikasingh/>

SOCIAL MEDIA LANGUAGE



Jen Mil 9 months ago

THANK YOU FOR HELPING HIM 😢♥



REPLY



Jave Flores 9 months ago



REPLY



Lazarew. Live 8 months ago

Ребят вы просто класс по больше бы таких людей!!!



REPLY



Clarice Roma 4 months ago

Vida longa para estes anjos que salvam os animais 🙏💖🙏🙏🙏



REPLY



Joana Beltran 9 months ago

God bless you beautiful dog 😳😍💕



1



REPLY



shreya khandelwal 2 months ago

Uh r a very gud teacher.... must say uh make topics very esy... I learned constitution from ur video only... Thank uh for these videos



REPLY

SOLUTIONS

- URL REMOVAL REGEX EXPRESSIONS
- EMOJI REMOVAL REGEX EXPRESSIONS
- PUNCTUATION REMOVAL
- ONLY RETAIN ALPHABETS AND SPACES
- SPELLING CORRECTORS
- TERM NORMALIZATION MAPPING (DELICIOUS -> SCRUMPTIOUS, DELISH, SCRUMMY, YUMMY, YUM-YUM, YUMYUM, YUMM, MOREISH, PENG, NUMMY, NOMNOM, FNIGER LICKING, NOM, NOM NOM; WASHINGTON DC -> DC, WASH DC, DISTRICT OF COLUMBIA, D.C.)
 - CO-OCCURRENCE MEASURE FOR A TOPIC
 - COSINE DISTANCE
 - ALTERNATE NAMES DATASETS
- LANGUAGE DETECTION
- LANGUAGE TRANSLATION
- SEGMENT DATA BY TOPIC

YOUTUBE API : GETTING STARTED

YOUTUBE DATA API DOCUMENTATION

[HTTPS://DEVELOPERS.GOOGLE.COM/YOUTUBE/V3/GETTING-STARTED](https://developers.google.com/youtube/v3/getting-started)

[HTTPS://DEVELOPERS.GOOGLE.COM/YOUTUBE/V3/DOCS](https://developers.google.com/youtube/v3/docs)

REGISTER A PROJECT

ENABLE A PROJECT

USE YOUR API KEY!

*Use the data in compliance
with their terms and services

The screenshot shows the YouTube Data API's "Try this API" interface. On the left, there are dropdown menus for "videoSyndicated" and "videoType". Below them are sections for "Show standard parameters" and "Credentials". Under "Credentials", two checkboxes are checked: "Google OAuth 2.0" and "API key". At the bottom of this section is a blue "EXECUTE" button. To the right, there are tabs for JAVASCRIPT, JAVA, PHP, and PYTHON, with PYTHON selected. The Python code example is:

```
def main():
    # Disable OAuthlib's HTTPS verification when running locally.
    # *DO NOT* leave this option enabled in production.
    os.environ["OAUTHLIB_INSECURE_TRANSPORT"] = "1"

    api_service_name = "youtube"
    api_version = "v3"
    client_secrets_file = "YOUR_CLIENT_SECRET_FILE.json"
```

Below the code are two tabs: "application/json" (selected) and "Raw HTTP Response". The "application/json" tab shows a green status bar with "200" and the JSON response:

```
{"snippet": { "publishedAt": "2018-09-16T15:00:00Z", "channelId": "UCHdhriklar625Rky7Jdn7A", "title": "How To Paint A Room • Nifty 101", "description": "00:54 - Picking a color 02:33 - Picking a sheen 04:51 - Materials & ...", "thumbnails": { "default": {}}
```

GOTCHAS!

- RATE LIMITS!
- ERROR HANDLING

```
File "/opt/anaconda3/lib/python3.7/site-packages/googleapiclient/http.py", line 856, in execute
    raise httperror(resp, content, uri=uri)
googleapiclient.errors.HttpError: <httperror 403 when requesting
https://www.googleapis.com/youtube/v3/commentThreads?part=snippet&maxResults=50&videoId=#####&key#####
#####&alt=json returned "the video identified by the <code><a href="/youtube/v3/docs/commentThreads/list#videoId">videoId</a></code> parameter has disabled comments.">
```

HOW TO INTERACT WITH YOUTUBE API

- USE THE API WEB EXPLORER
- BUILD YOUR OWN CODE USING API DOCUMENTATION EXAMPLES

<https://developers.google.com/youtube/v3/quickstart/python>

- OPEN-SOURCE

- PYYOUTUBEANALYSIS <https://github.com/jsingh811/pyyoutubeanalysis>

- RUN SEARCHES, COLLECT COMMENTS
- DEFINE SEARCH PARAMS - KEYWORDS, TIMEFRAME, TYPE
- ERROR HANDLING
- RUNS LOCATION EXTRACTION ON TEXT
- RUNS KEYPHRASE EXTRACTION ON TEXT
- SOCIAL MEDIA DATA CLEANING METHODS
- AUTOMATIC REPORT GENERATION

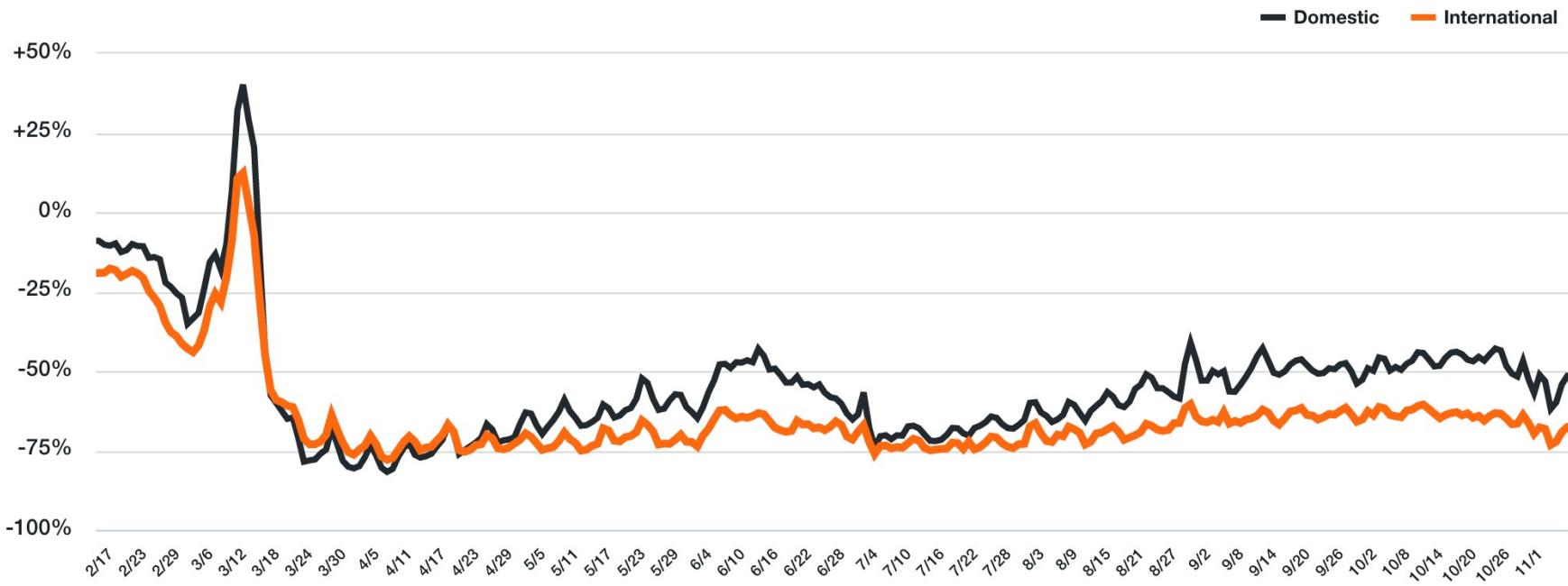
} YouTube API related
}
} NLP on social media language

TREND ANALYSIS USING STATISTICS AND NLP

TRAVEL IN 2020 - USA

Domestic vs. international flight searches

A day by day look at domestic and international flight search interest in the country selected, compared to the same day one year prior.



TRAVEL VLOGS ON YOUTUBE



IS THIS THE BEST PLACE IN EUROPE?! | Amalfi Travel Vlog
Haylsa & Kyle • 279K views • 11 months ago
We recently spent 10 days exploring the Amalfi Coast and boy was it a DREAM! Rooftop sunrises in Positano, exploring the coast ...



I spent 24 Hours on a plane! 30 hour travel vlog!
Jazzy Vlogs • 380K views • 5 months ago
I spent 24 hours on a plane travelling to the other side of the world! 30 hours total travel time and two long haul flights!! At the time ...



JAMAICA TRAVEL VLOG: BEST TRIP EVER | Montego Bay
Maryam Hampton • 113K views • 1 year ago
Greece Travel Vlog DREAM TRIP- <https://youtu.be/mp18Mn35u-4> FOLLOW my ...

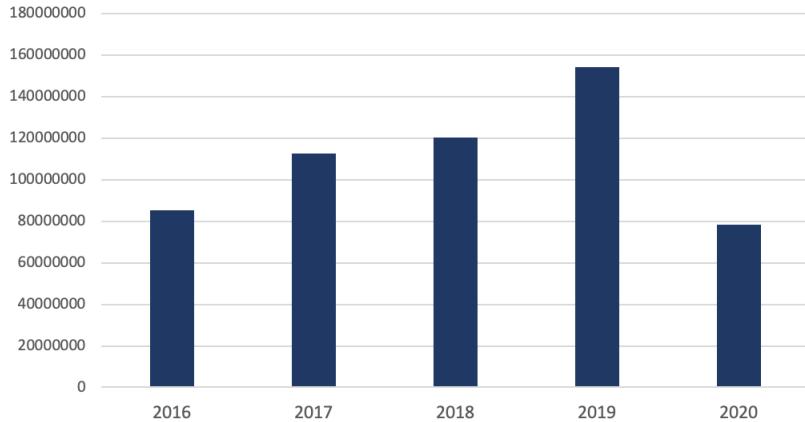


ZANZIBAR, TANZANIA PART 1 | TRAVEL VLOG
Being Neiicey • 195K views • 1 year ago
Don't Forget to Subscribe! New Vlogs Every Monday & Friday! #zanzibar #tanzania #travelwithneiicey
Subscribe to Main ...

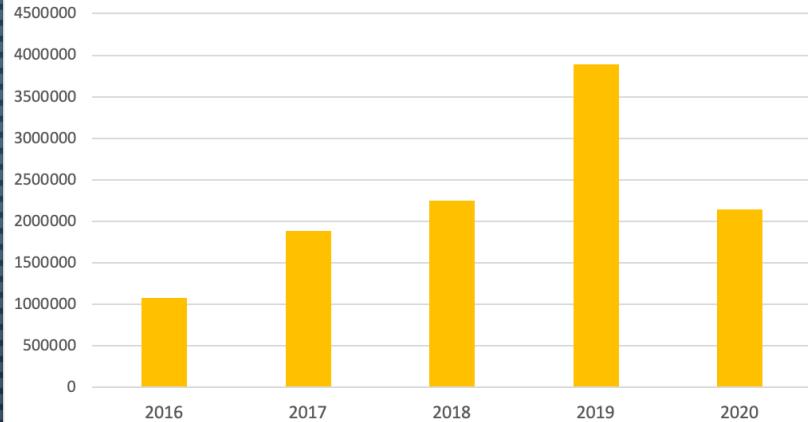


Brunei Is Not What I Expected | Solo Travel Vlog
In Between Lattes • 457K views • 9 months ago
Wow, Brunei is not what I expected at all, especially after reading all the recent controversy! Join me on a two-part ultimate solo ...

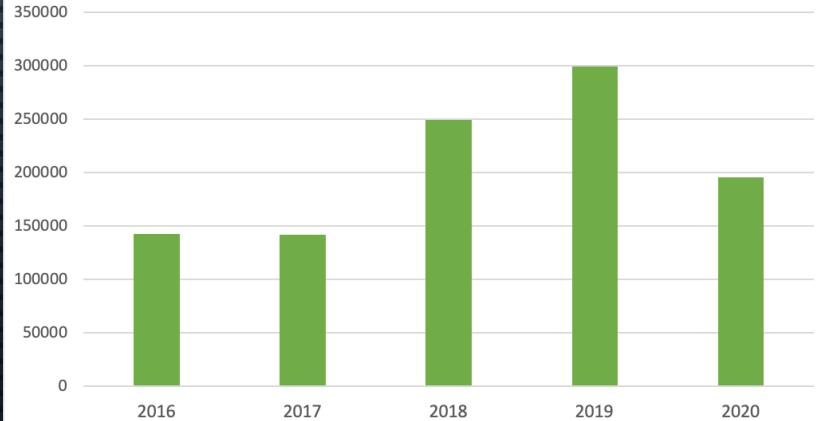
Views



Likes

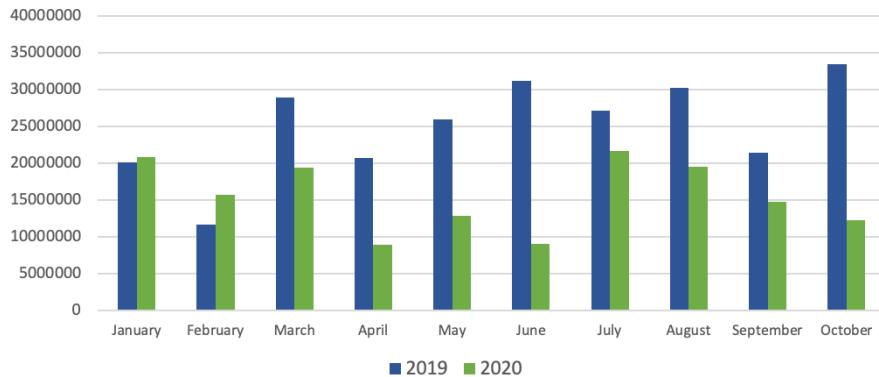


Comments



YEARLY STATS

Views

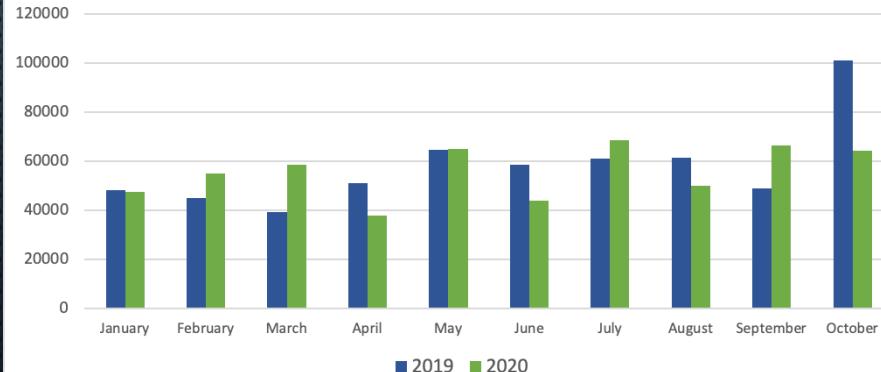


Likes

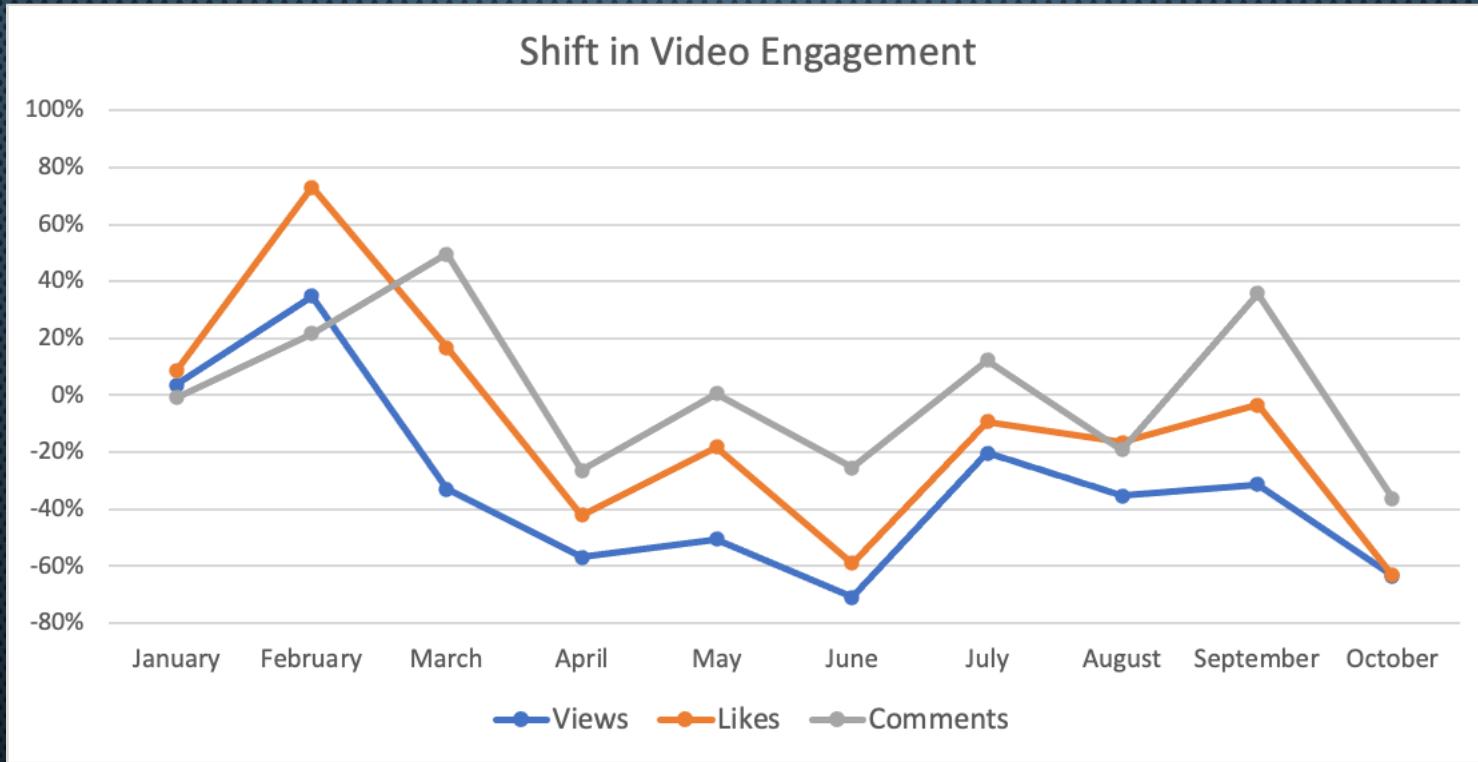


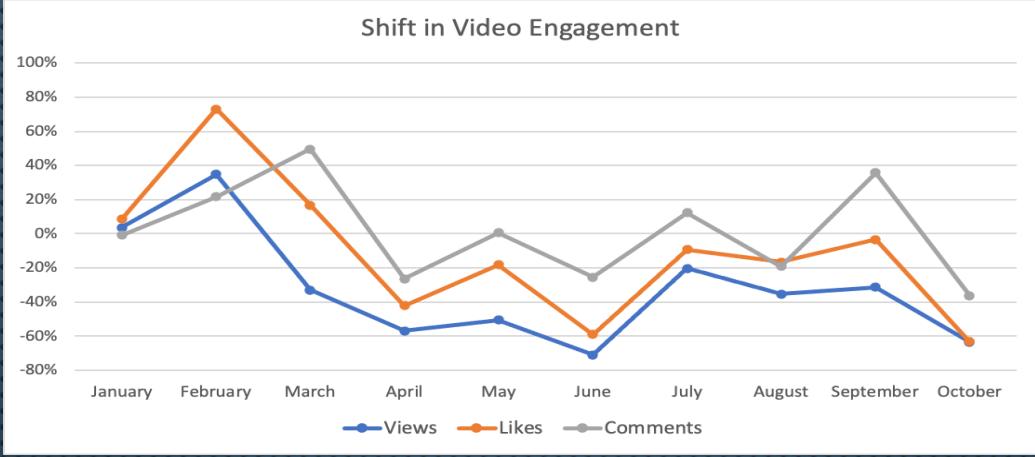
2019 VS 2020

Comments



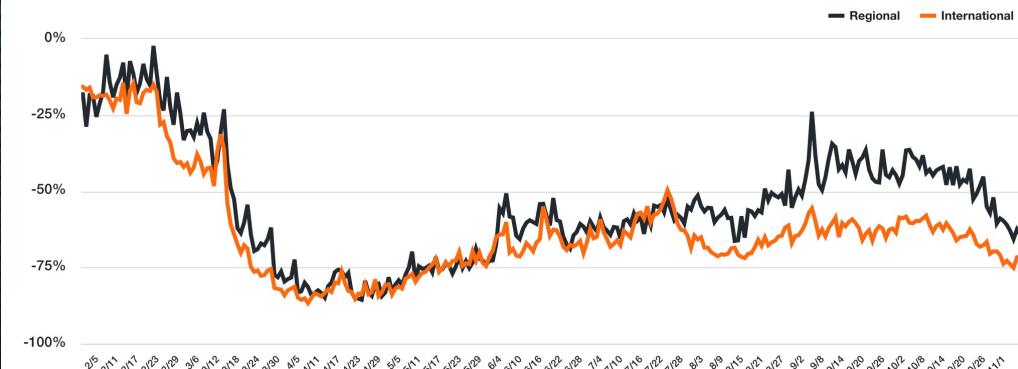
2019 VS 2020





Regional vs. international flight searches

A day by day look at regional (Denmark, Sweden, Norway, Finland) and international flight search interest in the country selected, compared to the same day one year prior.



WHAT TRAVEL WERE PEOPLE TALKING ABOUT?

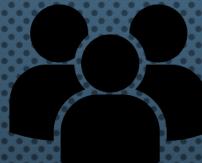
WHAT DID PEOPLE COMMENT ABOUT?

A dense, colorful word cloud centered around travel and tourism, featuring large, bold words like "london", "italy", "india", "japan", "australia", and "canada" in various colors, surrounded by smaller city and country names.

WHAT WERE PEOPLE COMMENTING?

i'm going to sorrento in 10 days and i'm so excited. i've been watching tonnes of sorrento and italy vlogs and yours are so lush 😍😍

I loved the tour looked like you guys had fun. im going there next week, how long ago were you there and were there lots of restrictions and closing due to covid



Did they require you to have a prior covid test?

Great video man, this place looks amazing. I have never been to Iceland, would love to visit some day. Honestly can't wait for the lockdown to be lifted so I can start travelling again. Thanks for sharing your experience. :)

HOW DID I GET ALL THIS ANALYSIS?

- PYOUTUBEANALYSIS TO PULL VIDEO DATA FOR DIFFERENT YEARS (2016, 2017, 2018, 2019, 2020) FOR YEARLY STATISTICAL ANALYSIS
- PYOUTUBEANALYSIS TO PULL COMMENTS AND VIDEO DATA FOR AUGUST, SEPTEMBER AND OCTOBER 2020
- PYOUTUBEANALYSIS FOR TEXT CLEANING
- PYOUTUBEANALYSIS FOR USING SPACY AND NLTK BASED LOCATION EXTRACTION AND KEYPHRASE EXTRACTION FROM COMMENTS AND VIDEO TEXT
- WORD CLOUD GENERATORS USING WEB TOOLS

RESOURCES

LANGDETECT <https://pypi.org/project/langdetect/>

WORDCLOUD <https://pypi.org/project/wordcloud/>

SKLEARN <https://scikit-learn.org/>

NLTK <https://www.nltk.org/>

SPACY <https://spacy.io/usage>

VADER <https://pypi.org/project/vaderSentiment/>

TEXTBLOB <https://textblob.readthedocs.io/en/dev/quickstart.html>

POLYGLOT <https://pypi.org/project/polyglot/>

CHARDET <https://pypi.org/project/chardet/>

STANFORD CORENLP <https://stanfordnlp.github.io/stanfordnlp/>

WORD CLOUD WEB <https://www.wordclouds.com/>

GENSIM <https://pypi.org/project/gensim/>

KEYPHRASE EXTRACTION: RAKE-NLTK <https://pypi.org/project/rake-nltk/>

KEYPHRASE EXTRACTION: PKE <https://github.com/bovdinfl/pke>

<https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>

YOUTUBE DATA API <https://developers.google.com/youtube/v3/getting-started>

<https://developers.google.com/youtube/v3/docs>

YOUTUBE API WRAPPERS

<https://github.com/rohitkhatri/youtube-python>

<https://github.com/sns-sdks/python-youtube>

PYOUTUBEANALYSIS <https://github.com/is Singh811/pyYouTubeAnalysis>

TRAVEL VLOG REFERENCES

<https://www.sojern.com/blog/covid-19-insights-on-travel-impact-global-hotel-16/>

<https://www.kayak.com/flight-trends>

Thank You!