

SCRAPING AND ANALYSING YOUTUBE TRENDING VIDEOS FOR BI

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Technology Degree in

Information Technology

By

SOWMIYA K (Reg. No. 37120072)

SUPRIYA S(Reg. No. 37120075)



DEPARTMENT OF INFORMATION TECHNOLOGY

SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI,

CHENNAI – 600119. TAMIL NADU

MARCH 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" Grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119.

Phone: 044 - 2450 3150 / 51 / 52 / 54 / 55 Fax: 044 - 2450 2344

www.sathyabama.ac.in



SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Professional Training Report is the bonafide work of SOWMIYA K (Reg. No.37120072) & SUPRIYA S (Reg. no.37120075) who underwent the professional training in **"Scraping and Analysing YouTube Trending Videos for BI"** under our supervision from December 2020 to March 2021.

Internal Guide

Dr. R. SUBHASHINI, M.E., Ph.D.,

Head of the Department

Dr. R. SUBHASHINI, M.E., Ph.D.,

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

We, SOWMIYA (Reg. No. 37120072) & SUPRIYA (37120075) hereby declare that the Professional Training Report on “scraping and analysing “**Scraping and Analysing Youtube Trending Videos for BI**” done by us under the guidance of Dr.R.SUBHASHINI, ME., Ph.D., at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of technology degree in **Information Technology**.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATES

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to the Board of **Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

We convey our thanks to **Dr. T.Sasikala M.E., Ph.D, Dean**, School of Computing , **Dr.R.Subhashini M.E., Ph.D. , Head** of the Department of Information Technology for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr.R.Subhashini M.E., Ph.D. ,** for his valuable guidance, suggestions and constant encouragement paved the way for the successful completion of our project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the **Department of Information Technology** who were helpful in many ways for the completion of the project.

ABSTRACT

Analysis of structured data has seen tremendous success in the past. However, large scale of unstructured data has been analysed in the form of video format remains a challenging area. YouTube, a Google company, has over a billion users and it used to generate billions of views. Since YouTube data is getting created in a very huge amount with a lot of views and with an equally great speed, there is a huge demand to store the data, process the data and carefully study the data in this large amount of it usable. The project utilizes the YouTube Data API which allows the applications or websites to incorporate functions in which they are used by YouTube application to fetch and view the information. The Google Developers Console which is used to generate a unique access key which is further required to fetch the data from YouTube public channel. Process the data and finally data stored in AWS. This project extracts the meaningful output of which can be used by the management for analysis, these methodologies helpful for business intelligence.

CHAPTER	TITLE	PAGE NO
	ABSTRACT	IV
	LIST OF TABLES	V
	LIST OF FIGURES	VIII
	LIST OF ABBREVIATIONS	VIII
1	INTRODUCTION	1
2	LITERATURE SURVEY	3
3	SOCIAL MEDIA DATA AND CHALLENGES	6
	3.1 Types of data	6
	3.2 Text data formats	6
	3.3 Open-source databases	7
	3.4 Data access via tools	7
	3.5 Challenges of YouTube Trending Analytic	9
	3.5.1 Getting social media metrics to the right people	9
	3.5.2 Visualization	9
	3.5.3 Unstructured data is challenging	10
	3.5.4 Increasing signal to noise	10
	3.5.5 A wait and see attitude	10

4	METHODOLOGY	12
	4.1 System Requirements	12
	4.1.1 Software Requirements	12
	4.1.2 Hardware Requirements	12
	4.2 Packages	12
	4.3 Architecture Diagram	13
	4.3.1 System Architecture	13
	4.3.2 Collecting Data	14
	4.3.3 Storing	14
	4.3.4 Encryption	15
	4.4 Implementation	16
	4.4.1 Getting Goggle YouTube API	16
	4.4.2 Data Collection	16
	4.4.3 Data Processing	16
	4.4.4 Data Pre-Processing	18
	4.5 Indian Trending YouTube Data Analysis	19
	4.5.1 Exploratory Data Analysis	19
	4.5.2 Data Visualization	19
	4.6 Storing in Cloud	20

	4.6.1 Amazon S3 Security and Access Management	21
	4.6.2 Protecting Data Using Encryption	22
5	RESULT	24
6	CONCLUSION AND FUTURE WORK	25
	REFERENCE	27
	APPENDIX	29
	SCREENSHOTS	32
	PLAGIARISM REPORT	37
	PAPER ACCEPTANCE	38
	REGISTRATION	39

LIST OF FIGURES

FIGURES	TITLE	PAGE NO
4.3	ARCHITURE DIAGRAM	13
	4.3.1 SYSTEM ARCHITECTURE	14
	4.3.2 COLLECTING DATA	14
	4.3.3 STORING IN CLOUD	15
	4.3.4 ENCRYPTION	15

LIST OF ABBREVIATIONS

API	Application Programming Interface
NLP	Natural language processing
AES	Advanced Encryption Standard

CHAPTER 1

INTRODUCTION

Social media analytics is the process of gathering and analyzing data from social networks such as YouTube, Facebook, Instagram, LinkedIn and Twitter. It is commonly used by marketers to track online conversations about products and companies. One author defined it as "the art and science of extracting valuable hidden insights from vast amounts of semi-structured and unstructured social media data to enable informed and insightful decision making.

In this project we are going to scrape and analyze the YouTube data using Google API. YouTube as a user generated content is one of the largest and most popular video sharing websites. It hosts over four billion views a day. YouTube provides public statistics regarding its uploaded videos, most notably the number of views, which shows the aggregate number of times a video has been watched up to that point. Naturally, the number of views for a video indicates the level of popularity of that video; and it takes a varying amount of time for a video to become popular (if it becomes popular). Meanwhile, there are some videos that can attract users' attention in a relatively short time.

YouTube also supports a feature called trending, which represents content that has the potential of becoming popular in a relatively short time. Consequently, although trending videos are usually not popular (yet) when declared as trending by YouTube, they have the potential of becoming popular (eventually). For example, some videos are labeled trending while having only a few hundreds in viewership numbers. From another perspective, through trending videos, YouTube tries to highlight emerging trends developing within different viewership communities.

Meanwhile, the general attributes of the viewership of trending videos have not been studied thoroughly. To the best of our knowledge, basic statistics about YouTube trending videos have not been studied, analyzed, or even received any adequate attention. Considering the fact that more than one billion unique users visit YouTube each month and they upload 72 hours of video every minute, YouTube is the best place for e.g. brand engagement or advertising, but it is genuinely difficult and competitive to get the attention of users.

Therefore, when a video becomes popular, it is exposed to millions of users for free and has the opportunity of keeping their attention for a while. Finding these trends are significantly important as many different websites have emerged just to pick up on the latest trends on the web, such as “BuzzFeed” or “WhatTheTrend”. There are other kinds of websites such as “channel Factory” that tries to build an audience on YouTube for content owners or advertisers.

Understanding of YouTube trending videos and their statistics, and a deeper insight about their life cycles, can greatly affect the strategies for marketing, target advertising, recommendation systems and search engines, as was suggested by prior YouTube measurement studies. This represents a key motivation for our effort.

In this paper, we present our findings for measuring, analyzing, and comparing key aspects of YouTube trending videos. Our study is based on collecting and monitoring high-resolution time-series of the viewership and related statistics of more than 8,000 YouTube videos over an aggregate period. To put this number into perspective, YouTube declares only tens of videos as trending on a daily basis. This number is highly selective when compared to the thousands of videos uploaded on YouTube every minute.

Furthermore, and from the onset of our study, our goal has been to collect data viewership time-series of trending videos to achieve robust analysis of their entire life cycles. Since trending videos are declared as such just several hours after they are uploaded, we are able to analyze trending videos across their lifecycle; this provides an invaluable insight into their viewership time-series over a critical period of their lifetime.

When we deal with data we need more places to store the data so for this project we are storing the mass media data in Amazon S3(simple storage system), and encrypted by using amazon server-side encryption S3. Amazon s3 is high durability and secure.

CHAPTER 2

LITERATURE SURVEY

TITLE : IEEE 4TH International Conference On Big Data Analytics (ICBDA)

AUTHOR : SUN Yi, Marius Ngaboyamahina

YEAR : 2019

DESCRIPTION:

Customer satisfaction is an essential area of the industry in this 21st century sometimes as known as the information age. However, the perception of customer expectation remains a problem in today's businesses. The internet has enabled people to spread out their thoughts through Social Media (SM) platforms, forums, news comments, and blogs. Consequently, those platforms are generating exponentially the immense amounts of data. The extraction of opinions from those big data can actively allow to rate organizations, learn the consumer needs, and adjust the business's strategies. This paper presents a concept of building a rating system, using Big Data Analytics (BDA) techniques, that apply the existing Sentiment Analysis (SA) algorithms to gain insight into reviews gathered from SM applications. The system will allow to list the various categories of services and evaluate them based on the obtained the customers' reactions. Also, this study aims to manage a large volume of information to rank the institutions and provide a practical solution for competitive, marketing analysis, and track the improvement of customer satisfaction within both the public and private sectors to boost the excellent service delivery in Rwanda.

TITLE : Retrieving YouTube Video By Sentiment Analysis On User Comment

AUTHOR : Hanif Bhuiyan, Jinat Ara, Rajon Bardhan, Md. Rashedul Islam

YEAR : 2017

DESCRIPTION:

YouTube is one of the comprehensive video information source on the web where video is uploading continuously in real time. It is one of the most popular site in social media, where users interact with sharing, commenting and rating (like/views) videos. Generally the quality, relevancy and popularity of the video is maintained based on this rating. Sometimes irrelevant and low quality videos ranked higher in the search result due to the number of views or likes, which seems untenable. To minimize this issue, we present a Natural Language processing (NLP) based sentiment analysis approach on user comments. This analysis helps to find out the most relevant and popular video of YouTube according to the search. The effectiveness of the proposed scheme has been proved by a data driven experiment in terms of accuracy of finding relevant, popular and high-quality video.

TITLE : Machine Learning (ML) Can Develop Future Developments And Support Decision Making.

AUTHOR : Andrej Miklosik, Martin Kuchta.

YEAR : 2019

DESCRIPTION:

Exponential technological expansion creates opportunities for competitive advantage by applying new data-oriented approaches to digital marketing practices. Machine learning (ML) can predict future developments and support decision-making by extracting insights from large amounts of generated data. This functionality greatly impacts and streamlines the strategic decision-making process of organizations. The research gap analysis revealed that a little is known about marketers' attitude toward, and knowledge about, ML tools and their adoption and utilization to support strategic and operational management. The research presented here focuses on the selection and adoption of the ML-driven analytical tools by three distinct groups: marketing agencies, media companies, and advertisers. Qualitative and quantitative research was conducted on a sample of these organizations operating in Slovakia. The findings highlight: 1) the important role of intelligent analytical tools in the creation and deployment of marketing strategies; 2) the lack of knowledge about emerging technologies, such as ML and artificial intelligence (AI); 3) the potential application of the ML tools in marketing, and; 4) the low level of adoption and utilization of the ML-driven analytical tools in marketing management. A framework consisting of enablers and a process map was developed to help organizations identify the opportunities and successfully execute projects that are oriented toward the deployment and adoption of the analytical ML tools in digital marketing.

CHAPTER 3

SOCIAL MEDIA DATA AND CHALLENGES:

Clearly, there is a large and increasing number of (commercial) services providing access to social networking media (e.g., Twitter, Facebook and Wikipedia) and news services (e.g., Thomson Reuters Machine Readable News). Equivalent major academic services are scarce. We start by discussing types of data and formats produced by these services.

3.1 Types of data

- Historic data sets—previously accumulated and stored social/news, financial and economic data.
- Real-time feeds—live data feeds from streamed social media, news services, financial exchanges, telecoms services, GPS devices and speech. And into:
 - Raw data—unprocessed computer data straight from source that may contain errors or may be unanalyzed.
 - Cleaned data—correction or removal of erroneous (dirty) data caused by disparities, keying mistakes, missing bits, outliers, etc.
 - Value-added data—data that has been cleaned, analyzed, tagged and augmented with knowledge.

3.2 Text data formats

The four most common formats used to markup text are: HTML, XML, JSON and CSV.

- HTML—Hypertext Markup Language (HTML) as well-known is the markup language for web pages and other information that can be viewed in a web browser. HTML consists of HTML elements, which include tags enclosed in angle brackets (e.g., `<div>`), within the content of the web page.
- XML—Extensible Markup Language (XML)—the markup language for structuring textual data using `<tag[...</tag[` to define elements.
- JSON—JavaScript Object Notation (JSON) is a text based open standard designed for human-readable data interchange and is derived from JavaScript.
- CSV—a comma-separated values (CSV) file contains the values in a table as a series of ASCII text lines organized such that each column value is separated by a comma from the next column's value and each row starts a new line. For

completeness, HTML and XML are so-called markup languages (markup and content) that define a set of simple syntactic rules for encoding documents in a format both human readable and machine readable. A markup comprises start-tags (e.g., <tag>), content text and endtags (e.g., </tag>).

3.3 Open-source databases

A major open source of social media is Wikipedia, which offers free copies of all available content to interested users (Wikimedia Foundation 2014). These databases can be used for mirroring, database queries and social media analytics. They include dumps from any Wikimedia Foundation project: <http://dumps.wikimedia.org/>, English Wikipedia dumps in SQL and XML: <http://dumps.wikimedia.org/enwiki/>, etc.

3.4 Data access via tools

As discussed, most commercial services provide access to social media data via online tools, both to control access to the raw data and increasingly to monetize the data.

Freely accessible sources

Google with tools such as Trends and Insights is a good example of this category. Google is the largest ‘scraper’ in the world, but they do their best to ‘discourage’ scraping of their own pages. (For an introduction of how-to surreptitious scrape Google—and avoid being ‘banned’—see <http://google-scraper.squabbel.com>.) Google’s strategy is to provide a wide range of packages, such as Google Analytics, rather than from a researchers’ viewpoint the more useful programmable HTTP-based APIs.

Trends displays a particular search term, in this case ‘libor.’ Using Google Trends, you can compare up to five topics at a time and also see how often those topics have been mentioned and in which geographic regions the topics have been searched for the most.

Commercial sources

There is an increasing number of commercial services that scrape social networking media and then provide paid-for access via simple analytics tools. (The more comprehensive platforms with extensive analytics are reviewed in Sect. 8.) In

addition, companies such as Twitter are both restricting free access to their data and licensing their data to commercial data resellers, such as Gnip and DataSift. Gnip is the world's largest provider of social data.

Different types of feeds serve different types of use cases and correspond to different types of queries and API endpoints on the publisher's source API. After selecting the feed, the user is assisted by Gnip to configure it with any required parameters before it begins collecting data. This includes adding at least one rule. Under 'Get Data' –['Advanced Settings' you can also configure how often your feed queries the source API for data (the 'query rate'). Choose between the publisher's native data format and Gnip's Activity Streams format (XML for Enterprise Data Collector feeds).

Data feed access via APIs

For researchers, arguably the most useful sources of social media data are those that provide programmable access. Google Trends AI & Soc (2015) 30:89–116 95 123 APIs, typically using HTTP-based protocols. Given their importance to academics, here, we review individual wikis, social networking media, RSS feeds, news, etc.

Social networking media

As with Wikipedia, popular social networks, such as Facebook, Twitter and Foursquare, make a proportion of their data accessible via APIs. Although many social networking media sites provide APIs, not all sites (e.g., Bing, LinkedIn and Skype) provide API access for scraping data. While more and more social Fig. 3 Gnip Dashboard, Publishers and Feeds 96 AI & Soc (2015) 30:89–116 123 networks are shifting to publicly available content, many leading networks are restricting free access, even to academics. For example, Foursquare announced in December 2013 that it will no longer allow private check-ins on iOS 7, and has now partnered with Gnip to provide a continuous stream of anonymized check-in data. The data is available in two packages: the full Firehose access level and a filtered version via Gnip's PowerTrack service. Here, we briefly discuss the APIs provided by Twitter and Facebook.

3.5 Challenges of YouTube Trending Analytic

3.5.1 Getting social media metrics to the right people

Often, social media is treated like the ugly stepchild within the marketing department and real-time social media analytics are either absent or ignored. Real-time social media analytics creates serious challenges for many organizations. Often, organizations are married to an old paradigm — a vestige of by-gone days when data was hard to get, taking months of data gathering and analysis. These organizations didn't integrate data gathering into tactical and strategic decision-making because they couldn't. Incorporating real-time analytics just isn't possible with their existing bureaucracy.

For one thing, real-time analytics requires moving analysts closer to decision-makers and enabling decision-makers with analytic skills for ad hoc data analysis. But, that's not what most businesses look like. Many decision-makers lack analytics skills necessary for ad hoc analysis.

New arrival, Uber, which runs a ride-sharing program that competes with taxi companies and car services, uses real-time analytics to show how people move around a city at any give time, allowing Uber to optimize their customer service. Placing cars nearby reduces competition with local cab companies and real-time analytics provide insights necessary to do that. To do this, Uber uses real-time data to incentivize more drivers to provide services by raising the price of a ride.

Others such as Samsung and NASCAR do a great job of providing real-time social media analytics to guide decision-makers. Take a look at this command centre NASCAR uses to monitor chatter surrounding their events.

3.5.2 Visualization

Visualizing real-time social media analytics is another key element involved in developing insights that matter. Face it, human beings don't do a great job of processing long tables of numbers. Notice on NASCAR's command centre, much of the date is displayed visually.

Simply displaying values graphically helps in making the kinds of fast interpretations necessary for making decisions with real-time data, but adding more complex algorithms and using models provides deeper insights, especially when visualized.

3.5.3. Unstructured data is challenging

Unlike the survey data firms are used to dealing with, most (IBM estimates 80%) is unstructured — meaning it consists of words rather than numbers. And, text analytics lags seriously behind numeric analysis.

While unstructured data tends to muck-up any kind of analysis, it's especially challenging in the context of real-time analytics, because you want interpretations IN REAL TIME. Handling text in real time often means using computer-generated translations of the written word. However, no computer can effectively categorize much of what's written in social media where “bad” might mean bad or it might mean good, depending on context, relationship, and other variables.

3.5.4 Increasing Signal to Noise

Social media data is inherently noisy. Reducing noise to even detect signal is challenging — especially in real time. Sure, with enough time, new analytics tools can ferret out the few meaningful comments across various social networks, but few can handle this in Real time

3.5.5 A Wait and See Attitude

Again, businesses are used to a certain operational model that makes real-time social media analytics challenging. For instance, we listened to a presentation by an analyst from NPR. He showed complex A/B testing used to determine the effectiveness of headlines, even whole articles online. As a statistician, he's concerned about achieving statistical significance in his testing before making decisions.

That's great if you're talking about putting \$100 million into building and marketing a product, but doesn't make much sense in the fast-paced world of social media. Real-time analytics require real-time decisions. Period.

If it'll take you several days to gather enough data for statistical significance, forget it. Especially if you're only trying to determine which headline does better, by the time you have a statistically significant answer, no one cares anymore. The news trend has moved to another topic.

CHAPTER 4

METHODOLOGY

4.1 SYSTEM REQUIREMENTS:

4.1.1 Software Requirements:

Software requirements deals with defining resources requirements and prerequisites that needs to be installed on a computer to provide functioning of an application the minimal software requirements are as follows,

Front end : Python
Dataset : CSV
IDE : Anaconda Jupyter
OS : Windows 10,7.

4.1.2 Hardware Requirements:

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware. The minimal hardware requirements are as follows.

Processor : Pentium IV
RAM : 8GB
Processor : 2.4GHz
MainMemory : 8GB RAM
Hard Disk : 1tb

4.2 PACKAGES :

MATPLOT LIB:

Matplotlib is plotting library for the python programming language and its numerical mathematics extension numpy it provides an object oriented API for embedding plots into applications using general purpose GUI toolkits like Tkinter,wxpython,qtetc.pyplot function makes some changes for a created figures,

creates a plotting area in figure, decorate the plots with labels. And its multi-platform data visualization library built on numpy arrays.

REQUESTS:

- The requests module allows you to send HTTP requests using Python.
- The HTTP request returns a Response Object with all the response data (content, encoding, status, etc).

OS:

The **OS module in Python** provides functions for interacting with the operating system. **OS** comes under **Python's** standard utility **modules**. This **module** provides a portable way of using operating system-dependent functionality. The `*os*` and `*os.path*` **modules** include many functions to interact with the file system.

TIME:

The **Python time module** provides many ways of representing **time** in code, such as objects, numbers, and strings. It also provides functionality other than representing **time**, like waiting during code execution and measuring the efficiency of your code.

PANDAS:

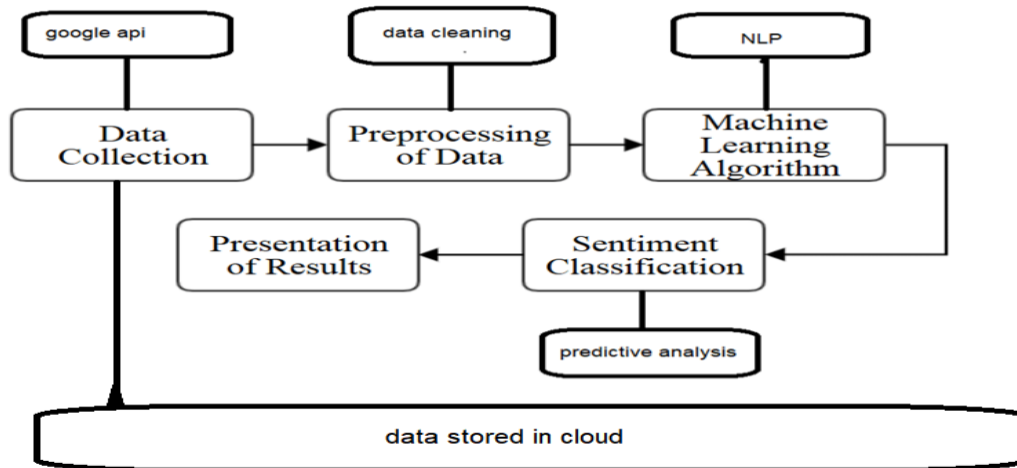
In computer programming, pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

4.3 ARCITECTURE DIAGRAM

4.3.1 System Architecture:

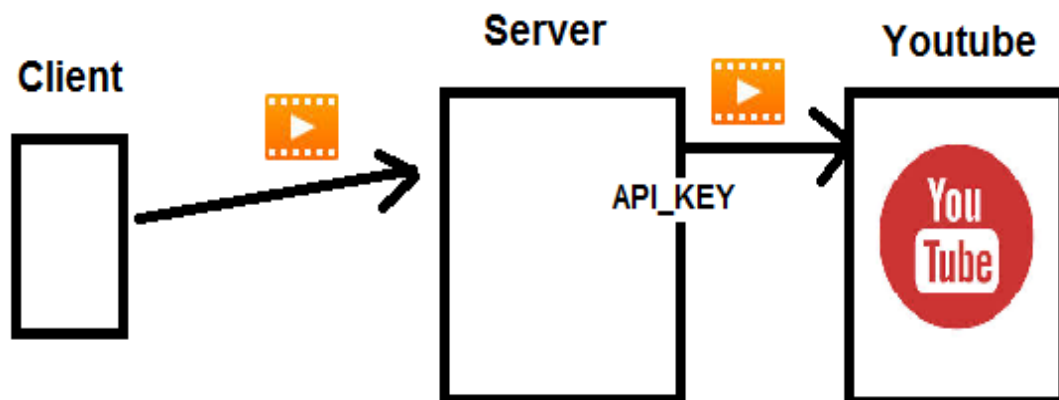
Collecting data from YouTube with the help of YouTube API v3 is the version of collected API. For analyzing first step to preprocess the data to make the analyzing part more accurate, to cleaning using the

NLP and doing the sentimental to find the people opinion. Finally stored and encrypt the data in Amazon S3.



4.3.2 Collecting Data:

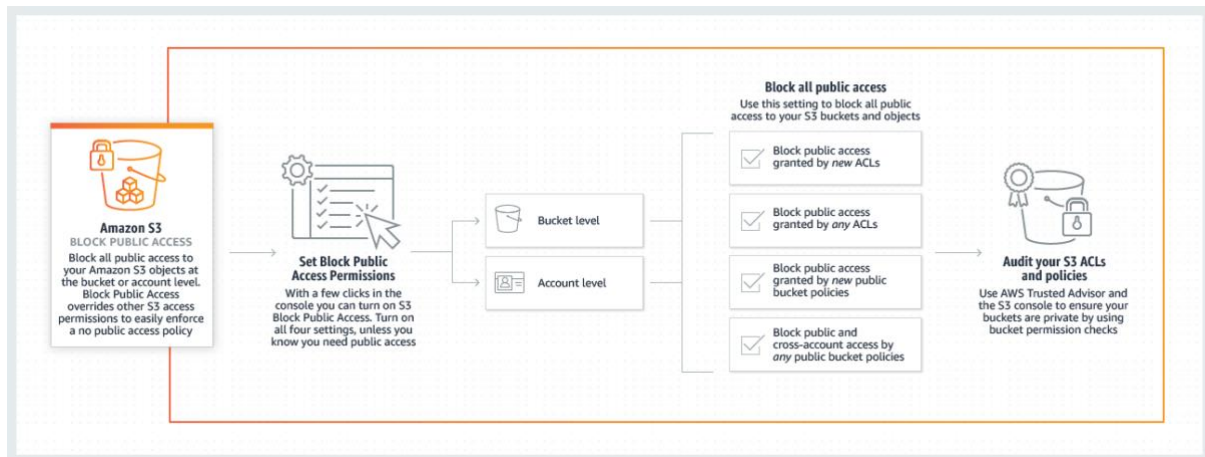
Collecting the data by request the YouTube API V3. YouTube API is an alphanumeric code.



4.3.3 Storing:

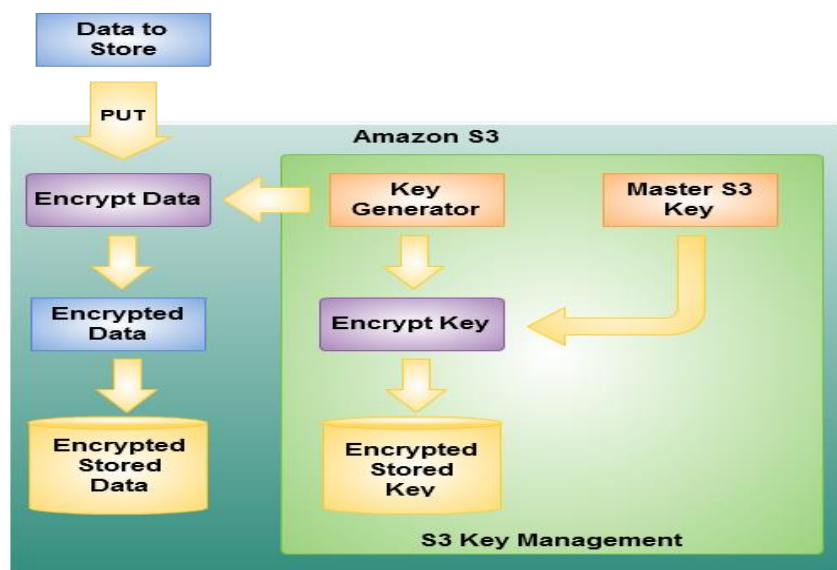
S3 Block Public Access provides controls across an entire AWS Account, or at the individual S3 bucket level to ensure that objects never have public access, now and in the future. Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, or both. In order to ensure that public access to all your S3 buckets and objects is blocked, turn on block all public access

at the account level. These settings apply account-wide for all current and future buckets. S3 Block Public Access settings override S3 permissions that allow public access, making it easy for the account administrator to set up a centralized control to prevent variation in security configuration regardless of how an object is added or a bucket is created.



4.3.4 Encryption:

Server-side encryption protects data at rest. Amazon S3 encrypts each object with a unique key. As an additional safeguard, it encrypts the key itself with a master key that it rotates regularly. Amazon S3 server-side encryption uses one of the strongest block ciphers available to encrypt your data, 256-bit Advanced Encryption Standard (AES-256).



4.4 IMPLEMENTATION:

4.4.1 Getting Google YouTube API

Google YouTube APIs are application programming interfaces (APIs) developed by Google which allow communication with Google Services and their integration to other services. The **YouTube Data API** (v3) lets you incorporate **YouTube** functionality into your own application. You can use the **API** to fetch search results and to retrieve, insert, update, and delete resources like videos or playlists.

1. Go to <https://developers.google.com/> and log in or create an account, if necessary.
2. After logging in go to this link <https://console.developers.google.com/project> and click on the blue CREATE PROJECT button as depicted in the photo below. Wait a moment as google prepares your project.
3. Fill in whatever Project Name you want.
4. Then click Google APIs link in the top left corner and then click the link option called "YouTube Data API." It's under YouTube API's. You can see it highlighted in the photo below, bottom right.
5. Now click on the "ENABLE" button.
6. Next click on the blue 'Go to Credentials' button to the right.
7. Choose the select option YouTube Data API v3 for the first select option and Web server(e.g. node js. Tomcat) for the second selection. Then choose Public data. Now click the blue button, "What credentials do I need?."
8. Almost done, wait for google to create your new project and you should see the screen below where you can copy your API Key.
9. Paste the API Key in our YouTube Options menu page as depicted below.

4.4.2 Data Collection

Data collection plays the most important role in Data Analytics. The Internet provides almost unlimited sources of data for a variety of topics. The importance of this area depends on the type of business, but traditional industries can acquire a diverse source of external data and combine those with their transactional data. data collection will be done by using the Google API and countries code (example: 'IN' for India). These two help to get the data in the format of semi-structured data.

For **data** collected from different sources and methods need specific **data** analysis methods and tools to process and get insights from them. Intended for developers who want to write applications that interact with YouTube. It explains about the basic concepts of YouTube and of the API itself and it also provides an overview of the different functions that the API supports and using some of the libraries like request, time, etc., to get the dataset of different countries trending youtube semi-structured data in spyder.

4.4.3 Data Processing

The phases of the intelligence cycle used to convert raw information into the actionable intelligence or the knowledge are conceptually similar to the phases in data analysis. Data that is initially obtained, must be processed or organized for the analysis. For instance, these may involve placing data into rows and into columns in a table format for further analysis, often through the use of spreadsheet or statistical software.

Initial Column Observations:

- **video_id**: contains alphanumeric code for video identification. However, this is not helpful for our data exploration/analysis
- **trending_date**: contains the date the video started trending in YYDDMM format
- **title**: contains the title of the video. Some contain non-ascii characters and emojis which we will need to remove
- **channel_title**: contains the title of the channel

- **category_id**: contains the id number for each category. We will need to match the id number to its corresponding category name
- **publish_time**: contains the date and time the video was published. It is formatted differently from trending date, which we plan to clean
- **tags**: contains the tags in one long string, which we will need to separate into a list
- **views, likes, dislikes, comment_count**: contains numerical values
- **thumbnail_link**: contains url for picture of the thumbnail.
- **comments_disabled, ratings_disabled, video_error_or_removed**: contains boolean values
- **description**: contains the video description as a string. Some contain non-ascii characters, emojis, and urls which we will need to remove if we use

4.4.4 Data Pre-Processing

Once processed and organized, the data might be incomplete and it contains duplicates, or contains errors. The need for the data cleaning process, will arise from problems in the way that the data is entered and stored. Data cleaning is the process of preventing the data and correcting the errors of those data. Common tasks are to include the record matching, identifying inaccuracy of data, overall quality of existing data, deduplication, and column segmentation. Such data problems that can also be identified through a variety of analytical techniques.

- Each YouTube video has a **video_id**, which is a unique alphanumeric code.
- Each YouTube video also includes a thumbnail that is shown with their video.
- The link to the video thumbnails is contained in the **thumbnail_link** column and although it would be interesting to explore thumbnails, as it may be that users are more likely to view a video for certain types of thumbnails.
- Natural Language Processing for title.
- We developed a function to clean the **video titles** by removing stop words, non-alphanumeric tokens, and money amounts to better understand the occurrence of certain keywords being used in the titles.

4.5 INDIAN TRENDING YOUTUBE DATA ANALYSIS

4.5.1 Exploratory Data Analysis

Once the datasets are cleaned, it can be analyzed. Analysts can apply a variety of techniques, referred to as exploratory data analysis, to begin understanding the messages that contained within the obtained data. The process of data in exploration may result in additional data cleaning or additional requests for data; thus, the initialization of the iterative phases that are mentioned in the lead paragraph of this section.

Descriptive statistics, such as, the average or medium, can be generated in the understanding of the data. Data visualization is also a technique that is used, in which the analyst is able to examine the data in a graphical format in order to obtain additional insights, regarding the messages within the data.

4.5.2 Data Visualization

Now that we have successfully cleaned our data, we will start exploring our data. For analyzing we use the Indian trending videos. We will start by looking at what words are commonly used in video titles, as we predicted that word choice will be an important factor for Trending videos. We will start by looking at what words are commonly used in video titles, as we predicted that word choice will be an important factor for Trending videos.

Figure 5, helps us understand why the top 10 channels with trending videos may not coincide with the exploration of categories we did earlier as the distribution is extremely rightly skewed. Around 1350 channels have under 10 trending videos, while we have few channels with over 100 trending videos. This tells us that looking too deeply into the top 10 channels won't really help us as the majority have well below that number of trending videos. We will calculate the percentages of the number of trending videos that channels have to look at the distribution more closely.

This demo, that more than 3/4 of channels have created under 20 youtube trending videos, with 34.5 percent of channels having between 1 - 5 youtube trending videos. This shows the top channels with well over 100 trending videos.

4.6 STORING IN CLOUD

Amazon Simple Storage Service (Amazon S3) is especially created for storing. Amazon S3 is an object storage service that offers industry-leading scalability, data availability, security, and performance. This means customers of all sizes and industries can use it to store and protect any amount of data for a range of use cases, such as data lakes, websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics.

Amazon S3 provides easy-to-use management features so you can organize your data and configure finely-tuned access controls to meet your specific business, organizational, and compliance requirements. Amazon S3 is designed for 99.999999999% (11 9's) of durability, and stores data for millions of applications for companies all around the world.

- Amazon S3 stores data as objects within buckets.
- An object is composed of a file and any metadata that describes that file. To store an object in Amazon S3, you upload the file you want to store into a bucket. When you upload a file, you can set permissions on the object and add any metadata.
- 1. Get started with AWS management console and move to S3.
- 2. Creating bucket name is **socialmd** in US East (Ohio)us-east-2region and stored the objects.(every bucket name must be unique)
- Every object they provide the URL link to access the authorized user by using ACL

4.6.1 Amazon S3 Security and Access Management

To protect your data in Amazon S3, by default, users only have access to the S3 resources they create. You can grant access to other users by using one or a

combination of the following access management features: AWS Identity and Access Management (IAM) to create users and manage their respective access.

Access Control Lists (ACLs) to make individual objects accessible to authorized users; bucket policies to configure permissions for all objects within a single S3 bucket; and Query String Authentication to grant time-limited access to others with temporary URLs. Amazon S3 also supports Audit Logs that list the requests made against your S3 resources for complete visibility into who is accessing what data.

4.6.2 Protecting Data Using Encryption

Data protection refers to protecting data while in-transit (as it travels to and from Amazon S3) and at rest (while it is stored on disks in Amazon S3 data centres). You can protect data in transit using Secure Socket Layer/Transport Layer Security (SSL/TLS) or client-side encryption. You have the following options for protecting data at rest in Amazon S3:

- **Server-Side Encryption** – Request Amazon S3 to encrypt your object before saving it on disks in its data centres and then decrypt it when you download the objects.(For here we used the server-side encryption)
- **Client-Side Encryption** – Encrypt data client-side and upload the encrypted data to Amazon S3. In this case, you manage the encryption process, the encryption keys, and related tools.

Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3)

- When you use Server-Side Encryption with Amazon S3-Managed Keys (SSE-S3), each object is encrypted with a unique key. As an additional safeguard, it encrypts the key itself with a master key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt your data. For more information, see Protecting data using server-side encryption with Amazon S3-managed encryption keys (SSE-S3).

Server-Side Encryption with Customer Master Keys (CMKs) Stored in AWS Key Management Service (SSE-KMS)

- Server-Side Encryption with Customer Master Keys (CMKs) Stored in AWS Key Management Service (SSE-KMS) is similar to SSE-S3, but with some additional benefits and charges for using this service. There are separate permissions for the use of a CMK that provides added protection against unauthorized access of your objects in Amazon S3.
- SSE-KMS also provides you with an audit trail that shows when your CMK was used and by whom. Additionally, you can create and manage customer managed CMKs or use AWS managed CMKs that are unique to you, your service, and your Region. For more information, see [Protecting Data Using Server-Side Encryption with CMKs Stored in AWS Key Management Service \(SSE-KMS\)](#).

Server-Side Encryption with Customer-Provided Keys (SSE-C)

- With Server-Side Encryption with Customer-Provided Keys (SSE-C), you manage the encryption keys and Amazon S3 manages the encryption, as it writes to disks, and decryption, when you access your objects. For more information, see [Protecting data using server-side encryption with customer-provided encryption keys \(SSE-C\)](#).

You can use the Amazon S3 Encryption Client in the AWS SDK in your own application to encrypt objects and upload them to Amazon S3. This method allows you to encrypt your data locally to ensure its security as it passes to the Amazon S3 service. The Amazon S3 service receives your encrypted data; it does not play a role in encrypting or decrypting it.

The Amazon S3 Encryption Client encrypts the object by using envelope encryption. The client calls AWS KMS as a part of the encryption call you make when you pass your data to the client. AWS KMS verifies that you are authorized to

use the customer master key (CMK) that you specify and, if so, returns a new plaintext data key and the data key encrypted under the CMK.

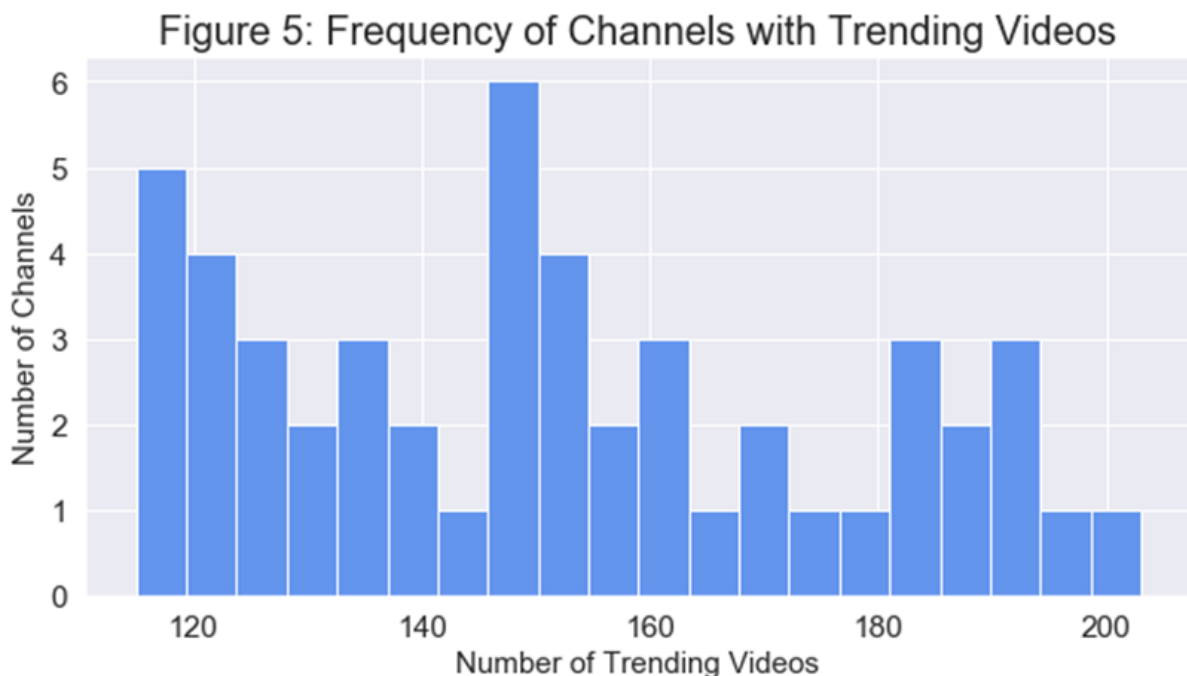
The Amazon S3 Encryption Client encrypts the data by using the plaintext key and then deletes the key from memory. The encrypted data key is sent to Amazon S3 to store alongside your encrypted data.

CHAPTER 5

RESULT

Data analytics can help to reduce bad decisions to a minimum by providing a fact-based information platform before the decision-making process. In other words, data analytics can help answer questions like: What do customers and business partners need? In what quantities will a specific product be demanded in the future or can I forecast the time before material abrasion leads to damage? If you have the right data and the appropriate analysis method, you can answer all these questions. The below graph shows the analytics, which involves in business intelligence using YouTube trending videos

Percentage of number of trending videos by channel



percentage = (part / total) × 100

```
above_20 = len([i for i in channel_freq if i > 20]) / len(channel_freq) * 100
```

```
percent_15_20 = len([i for i in channel_freq if (i <= 20) & (i > 15)]) / len(channel_freq) * 100
```

```
percent_11_15 = len([i for i in channel_freq if (i <= 15) & (i > 10)]) / len(channel_freq) * 100
```

```
percent_5_10 = len([i for i in channel_freq if (i <= 10) & (i > 5)]) / len(channel_freq) * 100  
percent_1_5 = len([i for i in channel_freq if (i <= 5)]) / len(channel_freq) * 100
```

Number of Trending Videos by Channel

- 1 to 5 -> Videos - 34.5%
- 5 to 11 -> Videos - 24.8%
- 11 to 15 -> Videos - 10.8%
- 15 to 20 -> Videos - 7.1%
- 21+ -> Videos - 22.9%

This demo, that more than 3/4 of channels have created under 20 youtube trending videos, with 34.5 percent of channels having between 1 - 5 youtube trending videos. This shows the top channels with well over 100 trending videos.

Trending videos help the marketer to deploy their products in these channels because trending videos will change everything 15 minutes according to the like, share, comments so this graph is showing which channel is coming frequently in trending list. This will help to deploy and create a product using these channels. When we deal with data we need more place to store the data so, data are stored in cloud. Finally data sets are stored in Amazon S3.

Business intelligence is the process of collecting, storing and analyzing data from business operations. BI provides comprehensive business metrics, in near-real time, to support better decision making.

Illustrates to get the right decision for digital marketing for company growth, Especially marketer can market they product using smart analytical process. YouTube trending video helps businesses understand how customers feel about their brand, giving them first-hand information to improve their products, make data-driven decisions, and deliver better customer experiences.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Do find a positive correlation between controversial of a video and viewer engagement in categories concerning politics and social issues. However, there are limitations of this project due to the incompleteness of dataset, confounding variables in real world scenario, and assumptions to simplify a complicated problem. We will handle this potential issue by recognizing the impact of the bias in the conclusion and discussion of our results.

As for keywords in video titles. Although we found that there are some correlations between keywords and the number of views, the impact is not completely clear. We observed that in some category some words dominate the context of video title, but videos whose title contains some of those top words does not mercenarily become more popular than those who do not.

Since the Trending YouTube Video dataset includes only top trending videos, there is a potential for survivor bias in the results. Videos with similar characteristics, that did not trend, are not included in the dataset and thus will not be studied.

We assume that video statistics such as views are independent of category, title, and many other variables presented in this dataset. Each view count is assumed to have equal weights. However, we do not consider the replay ability of a video because it cannot be distinguished from the dataset. In addition, we assume that there is no underlying association between individual ratings of a video. However, ratings in this dataset can inaccurately reflect the hidden distribution as comment and rating bots might be involved.

For convenience, we assume that frequent words in each category has higher impact on the number of views. However, some less frequent words that appear in only a few videos might also have great impact (by the assumption of tf-idf). In our study, we only focus on more general cases that covers the majority videos of the category.

Regarding future work, our group will try more advanced techniques on video statistics analysis, such as regression models and learning algorithms. We will also employ more language models in the realm of Natural Language Processing, such as word2vec, topic models and tied. Besides working on numerical and string-based data, we will extend this analysis to image data by encoding thumbnails. As for content-based analysis, we could obtain newspaper headlines from other dataset or using topic model to study the effect of trending news on trending videos.

PRIVACY/ETHICS CONSIDERATION

This Trending YouTube Video Statistics dataset is a collected from google built using YouTube's API Services. Developers have permission to use data collected from YouTube's API Services as long as they follow the policies listed in the YouTube API Services Policies document. Overall, online video research does raise valid ethical and data privacy concerns regarding informed consent. However, the unique potential of such a dataset outweighs the cost since this report is not for a commercial purpose and in our analysis, we have minimized the use of unique identifiers of the YouTube Video by using a bag-of-words model to analyze common terms in the title, tags, and description.

REFERENCE

- [1] Yessenov K, Misailovic S, Sentiment analysis of movie review comments, pp 1–17, (2009), . <http://people.csail.mit.edu/kuat/courses/6.863/report.pdf>.
- [2] Sun Yi, Marius Ngaboyamahina, IEEE 4th International Conference on Big Data Analytics (ICBDA), (2019).
- [3] Hanif Bhuiyan, Jinat Ara, Rajon Bardhan, Md. Rashedul Islam, Retrieving YouTube Video by Sentiment Analysis on User Comment, (2017).
- [4] Andrej Miklosik, Martin Kuchta, Machine Learning (ML) can Develop Future Developments and Support Decision Making, (2019).
- [5] A.Severyn,A.A.Moschitt.Uryupina, B.Plank and K. Filippova,“Multi-lingual opinion mining on YouTube,” Information Processing & Management, 52(1), pp. 46-60, (2016).
- [6] S. Chelaru, C. Orellana-Rodriguez and I. S. Altingovde, “How useful is social feedback for learning to rank YouTube videos?” In World Wide Web, 17(5), pp. 1-29, (2013).
- [7] P. Schultes, V. Dorner and F. Lehner, “Leave a Comment! An InDepth Analysis of User Comments on YouTube,” Wirtschaftsinformatik, pp. 659-673, (2013).
- [8] S. Siersdorfer, S. Chelaru, J. S. Pedro, I. S. Altingovde and W. Nejdl, “Analyzing and mining comments and comment ratings on the social web,” ACM Transactions on the Web (TWEB), 8(3), pp. 1-39, (2014).
- [9] University of Edinburgh, pp 1–74,

<http://homepages.inf.ed.ac.uk/miles/mscprojects/wolfram.pdf>.

[10] YouTube Statistics,

<http://www.youtube.com/yt/press/statistics.html>.

[11] YouTube data API documentation,

<https://developers.google.com/youtube/2.0/reference>.

[12] Wikimedia Foundation (2014) Wikipedia:Database download.

http://en.wikipedia.org/wiki/Wikipedia:Database_download.

[13] "What is structured data?-Definition from WhatIs.com".

WhatIs.com. Retrieved 2016-12-06.

[14] https://github.com/meyhaa/trendingyoutubevideos/blob/master/full_trending_video_analysis.ipynb.

[15] <https://www.kaggle.com/ammr111/youtube-trending-videos-analysis>.

[16]Trending-YouTube-Scraper

<https://github.com/mitchelljy/Trending-YouTube-Scraper>.

[17] <https://github.com/rahulanand1103/Web-Scraping-Youtube>.

[18] YouTube data API documentation,

<https://developers.google.com/youtube/2.0/reference>

APPENDIX

```
if request.status_code == 429:
    print("Temp-Banned due to excess requests, please wait and continue later")
    sys.exit()
return request.json()

def get_tags(tags_list):
    # Takes a list of tags, prepares each tag and joins them into a string by the pipe character
    return prepare_feature("|".join(tags_list))

def get_videos(items):
    lines = []
    for video in items:
        comments_disabled = False
        ratings_disabled = False

        # We can assume something is wrong with the video if it has no statistics, often this means it has been c
        # so we can just skip it
        if "statistics" not in video:
            continue

        # A full explanation of all of these features can be found on the GitHub page for this project
        video_id = prepare_feature(video['id'])

        # Snippet and statistics are sub-dicts of video, containing the most useful info
        snippet = video['snippet']
        statistics = video['statistics']

        # This list contains all of the features in snippet that are 1 deep and require no special processing
        features = [prepare_feature(snippet.get(feature, "")) for feature in snippet_features]

        # The following are special case features which require unique processing, or are not within the snippet
        description = snippet.get("description", "")
        thumbnail_link = snippet.get("thumbnails", dict()).get("default", dict()).get("url", "")
        trending_date = time.strftime("%y.%d.%m")
        tags = get_tags(snippet.get("tags", ["none"]))
        view_count = statistics.get("viewCount", 0)

        # This may be unclear, essentially the way the API works is that if a video has comments or ratings disab
        # then it has no feature for it, thus if they don't exist in the statistics dict we know they are disable
```

```

# then it has no feature for it, thus if they don't exist in the statistics dict we know they are disabled
if 'likeCount' in statistics and 'dislikeCount' in statistics:
    likes = statistics['likeCount']
    dislikes = statistics['dislikeCount']
else:
    ratings_disabled = True
    likes = 0
    dislikes = 0

if 'commentCount' in statistics:
    comment_count = statistics['commentCount']
else:
    comments_disabled = True
    comment_count = 0

# Compiles all of the various bits of info into one consistently formatted line
line = [video_id] + features + [prepare_feature(x) for x in [trending_date, tags, view_count, likes, dislikes,
comment_count, thumbnail_link, comments_disabled, ratings_disabled, description]]

lines.append(",".join(line))
return lines

def get_pages(country_code, next_page_token="&"):
    country_data = []

    # Because the API uses page tokens (which are literally just the same function of numbers everywhere) it is much
    # more inconvenient to iterate over pages, but that is what is done here.
    while next_page_token is not None:
        # A page of data i.e. a list of videos and all needed data
        video_data_page = api_request(next_page_token, country_code)

        # Get the next page token and build a string which can be injected into the request with it, unless it's None
        # then let the whole thing be None so that the loop ends after this cycle
        next_page_token = video_data_page.get("nextPageToken", None)
        next_page_token = f"&pageToken={next_page_token}&" if next_page_token is not None else next_page_token

        # Get all of the items as a list and let get_videos return the needed features
        items = video_data_page.get('items', [])
        country_data += get_videos(items)

```

```

next_page_token = video_data_page.get("nextPageToken", None)
next_page_token = f"&pageToken={next_page_token}&" if next_page_token is not None else next_page_token

# Get all of the items as a list and let get_videos return the needed features
items = video_data_page.get('items', [])
country_data += get_videos(items)

return country_data

def write_to_file(country_code, country_data):

    print(f"Writing {country_code} data to file...")

    if not os.path.exists(output_dir):
        os.makedirs(output_dir)

    with open(f"{output_dir}/{time.strftime('%y.%d.%m')}_{country_code}_videos.csv", "w+", encoding='utf-8') as file:
        for row in country_data:
            file.write(f"{row}\n")

def get_data():
    for country_code in country_codes:
        country_data = [", ".join(header)] + get_pages(country_code)
        write_to_file(country_code, country_data)

if __name__ == "__main__":

    parser = argparse.ArgumentParser()
    parser.add_argument('--key_path', help='Path to the file containing the api key, by default will use api_key.txt')
    parser.add_argument('--country_code_path', help='Path to the file containing the list of country codes to scrape')
    parser.add_argument('--output_dir', help='Path to save the outputted files in', default='output/')

    args = parser.parse_args()

    output_dir = args.output_dir
    api_key, country_codes = setup(args.key_path, args.country_code_path)

    get_data()

```

SCREENSHOTS

Feed Options

Add a follow button and position it using the options below. This option will not work for combined feeds.

YouTube API Key

This is required to make the feed work. Simply click the button below and it will connect to your YouTube account to get an access token and access token secret, and it will return it in the input below. Then just click the save button and you will now be able to generate your YouTube feed.

 **YouTube** | Log in and get my Access Token (API key) [Button not working?](#)

☒ Add your own API Key?

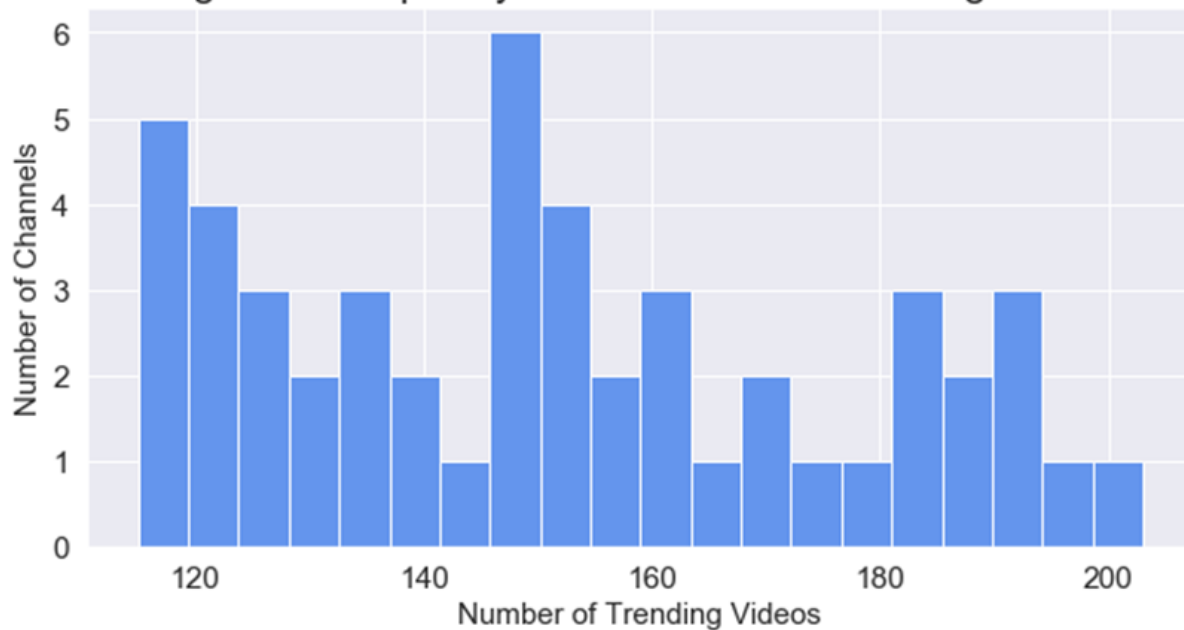
API Key Required

Your API key is working! Generate your shortcode on the [settings page](#).

	20.09.10_IN_videos	09-10-2020 21:04	Microsoft Excel Co...	348 KB
	20.09.10_JP_videos	09-10-2020 21:04	Microsoft Excel Co...	342 KB
	20.09.10_KR_videos	09-10-2020 21:04	Microsoft Excel Co...	247 KB
	20.09.10_MX_videos	09-10-2020 21:04	Microsoft Excel Co...	263 KB
	20.09.10_RU_videos	09-10-2020 21:04	Microsoft Excel Co...	409 KB
	20.09.10_US_videos	09-10-2020 21:03	Microsoft Excel Co...	297 KB
	20.19.08_IN_videos	25-12-2020 13:05	JSON File	395 KB
	21.18.01_BR_videos	18-01-2021 15:06	Microsoft Excel Co...	266 KB
	21.18.01_CA_videos	18-01-2021 15:05	Microsoft Excel Co...	294 KB
	21.18.01_DE_videos	18-01-2021 15:05	Microsoft Excel Co...	328 KB
	21.18.01_FR_videos	18-01-2021 15:06	Microsoft Excel Co...	272 KB
	21.18.01_GB_videos	18-01-2021 15:05	Microsoft Excel Co...	282 KB
	21.18.01_IN_videos	18-01-2021 15:05	Microsoft Excel Co...	352 KB
	21.18.01_JP_videos	18-01-2021 15:06	Microsoft Excel Co...	409 KB
	21.18.01_KR_videos	18-01-2021 15:06	Microsoft Excel Co...	247 KB
	21.18.01_MX_videos	18-01-2021 15:06	Microsoft Excel Co...	301 KB
	21.18.01_RU_videos	18-01-2021 15:06	Microsoft Excel Co...	408 KB
	21.18.01_US_videos	18-01-2021 15:05	Microsoft Excel Co...	296 KB
	INvideos		Type: Microsoft Excel Comma Separated Values File Size: 295 KB Date modified: 18-01-2021 15:05	Excel Co... 58,204 KB

DATA VISUALIZATION

Figure 5: Frequency of Channels with Trending Videos



STORAGE

The screenshot shows the AWS S3 Management Console interface. The top navigation bar includes the AWS logo, 'Services', a search bar, and user information. The main content area shows the 'Files and folders' tab for a bucket named 'socialmd'. The bucket contains 21 files and folders, totaling 86.5 MB (70.32% of the bucket's capacity). The files are listed in a table with columns for Name, Folder, Type, Size, Status, and Error.

Name	Folder	Type	Size	Status	Error
20.09.10_IN_videos.csv	Trending-YouTube-Scraper-master/output/	-	347.6 KB	Succeeded	-
20.09.10_IP_videos.csv	Trending-YouTube-Scraper-master/output/	-	341.0 KB	Succeeded	-
20.09.10_KR_videos.csv	Trending-YouTube-Scraper-master/output/	-	246.6 KB	Succeeded	-
20.09.10_MX_videos.csv	Trending-YouTube-Scraper-master/output/	-	262.1 KB	Succeeded	-
20.09.10_RU_videos.csv	Trending-YouTube-Scraper-master/output/	-	408.3 KB	Succeeded	-
20.09.10_US_videos.csv	Trending-YouTube-Scraper-master/output/	-	296.4 KB	Succeeded	-
20.19.08_IN_videos.json	Trending-YouTube-Scraper-master/output/	application/json	395.0 KB	Succeeded	-
IN_category_id.json	Trending-YouTube-Scraper-master/output/	application/json	8.0 KB	Succeeded	-
api_key.txt	Trending-YouTube-Scraper-master/	text/plain	39.0 B	Succeeded	-
country_codes.txt	Trending-YouTube-Scraper-master/	text/plain	32.0 B	Succeeded	-



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8" ?>
<Error>
  <Code>AccessDenied</Code>
  <Message>Access Denied</Message>
  <RequestId>SY3THEZHP7X0Q1Y7</RequestId>
  <HostId>dEqu4/jzoFL+cymgSb0kaORVt4H8eb2BemoAI1vPGAUdZniUM/LzTHDq/bFn/Svg/1GXq7saID0=</HostId>
</Error>
```

PLAGIARISM REPORT

ORIGINALITY REPORT

14%

SIMILARITY INDEX

12%

INTERNET SOURCES

12%

PUBLICATIONS

7%

STUDENT PAPERS

PRIMARY SOURCES

1

www.researchgate.net

Internet Source

10%

2

Hanif Bhuiyan, Jinat Ara, Rajon Bardhan, Md. Rashedul Islam. "Retrieving YouTube video by sentiment analysis on user comment", 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2017

Publication

1%

3

www.check-plagiarism.com

Internet Source

1%

4

www.tsjdev.de

Internet Source

1%

5

arxiv.org

Internet Source

<1%

6

Submitted to Amity University

Student Paper

<1%

PAPER ACCEPTANCE



IConIC 2K21 notification for paper 222

Inbox



IConIC 2K21 2:29 pm
to me ▾



Dear Sowmiya K,

Congratulations! Your paper 222 is selected for oral presentation. As per the mail you received earlier, go through the reviewer comments and modify accordingly/rectify the deficiencies in your Camera-ready paper.

For the format of Camera ready paper Refer:
(<http://pecteam.in/comet2/registration.html>)

1. Camera Ready Paper
2. Scanned Registration Form
3. Payment Proof (Scanned Copy/ Online transaction)

Subject of email should be "Final Submission PaperID". For example: Final Submission 101

Attach scanned copies of duly signed registration form, camera ready paper and send e-mail with attachments to "pecteamiconic2k21@gmail.com"

Note: Make the registration on or before 18/03/2021.

Best Regards

Prof.Dr.V.D.Ambeth Kumar

SUBMISSION: 222

TITLE: Scraping and Analysing Youtube Trending Vedios for BI

Transactions



11/03/2021 - 18/03/2021

Date**Issued Date:** 18-03-2021**Debit Account:** 097005004689**Name**

PEC CONFERENCE AND RESEARCH ⌵

Credit Account: 01570110103951**Credit Ifsc:** UCBA0000157**Refno****Utr No:** 107721388853**Ref ID:** 3566255**Amount**

₹7,500.00

Mode: IMPS ⚡**Remarks**

4th international conference

Success

Status

COMPLETED ▾