



Tumuhimbise Moses

Follow

Feb 9, 2018 · 6 min read · Listen



Save



Sign in to Medium with Google



Giacomo Zacchia

giacomo.zacchia@gmail.com



Giacomo Zacchia

giacomo@nightmedia.co

1 more

YouTube Data Mining : Experiences of using R.



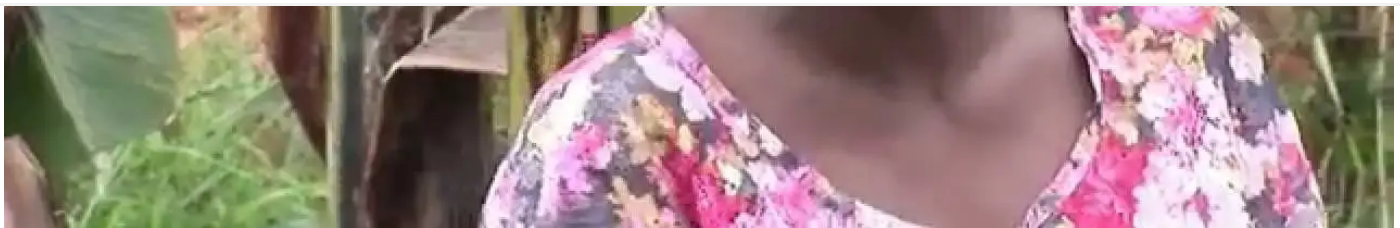
Open in app

Sign up

Sign In



Search Medium



Are you are risk taker ?

You Tube as any other social media platform provides a rich trove of data. One can put use of such information in many ways. Tickled by the curiosity of the popularity of well known domestic and international stand up comedian Anne Kansiime, I took upon the challenge of scraping data from her You Tube channel with the hope that I could provide some statistically driven insights into the notoriety of her online content. R provides a package known as tuber with the latest version available for download through the CRAN repository.

After installing the package, we load it into our development environment and then any tinkering with this impressive tool can begin.

```
library (tuber)
```

Before the tuber package can be used fully, prerequisites demand that one holds Google Developer Console credentials so as to allow one to access the YouTube API. I found the following [tutorial](#) (with step by step images) useful to setting up YouTube OAuth details.

I recommend testing whether you can connect to the YouTube API with ease as the first step before any data scraping. The following snippet below is helpful in determining whether you were able to correctly generate the required API credentials.

```
app_id <- "1088891801215-  
srtv9mvhb4dt1qr62254vfcgqfg9cp33.apps.googleusercontent.com"  
app_secret <- "hyJRTIMMvnBNj6WKE6jIzcXa"  
yt_oauth(app_id, app_secret)
```



100



4

After successfully connecting to the API, you should be redirected to the following web page with the following message.



Now on to the data scraping. I began by extracting the statistics of a single video. In order to do this, I got the video id. A YouTube video id entails the alphanumeric figures after the equals sign within the link.



```
get_stats(video_id='8XAfjB0fVas')
```

With the video id, we can now extract some statistics of the video using the line below. I chose to place the newly extracted data in a data frame for easy referencing.

```
## Auto-refreshing stale OAuth token.
```

From the information extracted, we can see that this particular video was viewed 479,534 times, was liked 2,319 times, was disliked 110 times and 185 comments were made. Additional details about the video can be extracted using the `get_video_details(video_id=)` snippet.

```
## $id
## [1] "8XAfjB0fVas"
##
## $viewCount
## [1] "479587"
##
## $likeCount
## [1] "2320"
##
## $dislikeCount
## [1] "110"
##
## $favoriteCount
## [1] "0"
##
## $commentCount
## [1] "185"
```

This provides information such as the video initial upload was on 3rd October 2015. Other information includes some of the tags embedded to increase search engine optimization.

```
## $kind
## [1] "youtube#videoListResponse"
##
```

```
## $etag
## [1] "\"_gJQceDMxJ8gP-8T2HLXUoURK8c/ayTSBFkx2upWmbCv6oiJb0G2whI\""
##
## $pageInfo
## $pageInfo$totalResults
## [1] 1
##
## $pageInfo$resultsPerPage
## [1] 1
##
##
## $items
## $items[[1]]
## $items[[1]]$kind
## [1] "youtube#video"
##
## $items[[1]]$etag
## [1] "\"_gJQceDMxJ8gP-8T2HLXUoURK8c/nkwqeVQVP71TP-Q6Q19srEiC2gM\""
##
## $items[[1]]$id
## [1] "8XAfjB0fVas"
##
## $items[[1]]$snippet
## $items[[1]]$snippet$publishedAt
## [1] "2015-10-03T06:32:44.000Z"
##
## $items[[1]]$snippet$channelId
## [1] "UCd3zZSCsHS-7smn2HwgM0FA"
##
## $items[[1]]$snippet$title
## [1] "Kansiime will not be donating any blood. Kansiime Anne.
African comedy."
##
## $items[[1]]$snippet$description
## [1] "What!?! U have not Subscribed yet??? Eh nga you are a RISK -
TAKER!!\nNOW NEW CLIPS EVERY WEEK !!!\npls also like me on
facebook\nhttps://www.facebook.com/pages/Kansiime-Anne-
Entertainer/263758240421584?ref=ts&fref=ts"
##
## $items[[1]]$snippet$thumbnails
## $items[[1]]$snippet$thumbnails$default
## $items[[1]]$snippet$thumbnails$default$url
## [1] "https://i.ytimg.com/vi/8XAfjB0fVas/default.jpg"
##
## $items[[1]]$snippet$thumbnails$default$width
## [1] 120
##
## $items[[1]]$snippet$thumbnails$default$height
## [1] 90
##
##
```

```
## $items[[1]]$snippet$thumbnails$medium
## $items[[1]]$snippet$thumbnails$medium$url
## [1] "https://i.ytimg.com/vi/8XAfjB0fVas/mqdefault.jpg"
##
## $items[[1]]$snippet$thumbnails$medium$width
## [1] 320
##
## $items[[1]]$snippet$thumbnails$medium$height
## [1] 180
##
##
## $items[[1]]$snippet$thumbnails$high
## $items[[1]]$snippet$thumbnails$high$url
## [1] "https://i.ytimg.com/vi/8XAfjB0fVas/hqdefault.jpg"
##
## $items[[1]]$snippet$thumbnails$high$width
## [1] 480
##
## $items[[1]]$snippet$thumbnails$high$height
## [1] 360
##
##
## $items[[1]]$snippet$thumbnails$standard
## $items[[1]]$snippet$thumbnails$standard$url
## [1] "https://i.ytimg.com/vi/8XAfjB0fVas/sddefault.jpg"
##
## $items[[1]]$snippet$thumbnails$standard$width
## [1] 640
##
## $items[[1]]$snippet$thumbnails$standard$height
## [1] 480
##
##
## $items[[1]]$snippet$thumbnails$maxres
## $items[[1]]$snippet$thumbnails$maxres$url
## [1] "https://i.ytimg.com/vi/8XAfjB0fVas/maxresdefault.jpg"
##
## $items[[1]]$snippet$thumbnails$maxres$width
## [1] 1280
##
## $items[[1]]$snippet$thumbnails$maxres$height
## [1] 720
##
##
##
## $items[[1]]$snippet$channelTitle
## [1] "Kansiime Anne"
##
## $items[[1]]$snippet$tags
## $items[[1]]$snippet$tags[[1]]
## [1] "kansime"
```

```
##
## $items[[1]]$snippet$tags[[2]]
## [1] "Anne"
##
## $items[[1]]$snippet$tags[[3]]
## [1] "Kansiime"
##
## $items[[1]]$snippet$tags[[4]]
## [1] "ann"
##
## $items[[1]]$snippet$tags[[5]]
## [1] "Ugandan"
##
## $items[[1]]$snippet$tags[[6]]
## [1] "african"
##
## $items[[1]]$snippet$tags[[7]]
## [1] "comedy"
##
## $items[[1]]$snippet$tags[[8]]
## [1] "just"
##
## $items[[1]]$snippet$tags[[9]]
## [1] "for"
##
## $items[[1]]$snippet$tags[[10]]
## [1] "laughs"
##
## $items[[1]]$snippet$tags[[11]]
## [1] "funny"
##
## $items[[1]]$snippet$tags[[12]]
## [1] "gags"
##
##
## $items[[1]]$snippet$categoryId
## [1] "23"
##
## $items[[1]]$snippet$liveBroadcastContent
## [1] "none"
##
## $items[[1]]$snippet$localized
## $items[[1]]$snippet$localized$title
## [1] "Kansiime will not be donating any blood. Kansiime Anne.
African comedy."
##
## $items[[1]]$snippet$localized$description
## [1] "What!?! U have not Subscribed yet??? Eh nga you are a RISK -
TAKER!!\nNOW NEW CLIPS EVERY WEEK !!!\npls also like me on
facebook\nhttps://www.facebook.com/pages/Kansiime-Anne-
Entertainer/263758240421584?ref=ts&fref=ts"
```

```
##  
##  
## $items[[1]]$snippet$defaultAudioLanguage  
## [1] "en"
```

More data can be extracted from the You Tube channel, however in order to achieve this, one needs to know the channel id. I found it difficult to find this information. However, after a few minutes of Stack Overflow browsing, I stumbled upon a You Tube channel name converter by [Johnny the Tank](#).

You simply type in the channel name and it generates the channel id.

Not secure | johnnythetank.github.io/youtube-channel-name-converter/

Youtube Username to Channel ID Converter

Star 45

Convert your Youtube Username to Channel ID

Channel ID:

Channel URL:

```
get_channel_stats(channel_id='UCd3zZSCsHS-7smn2HwgM0FA')
```

With the channel id, one can tinker around with functions such as extracting channel statistics. From the output yielded, we observe the channel was first created on 17th April 2012.

```
## Channel Title: Kansiime Anne
## No. of Views: 119184892
## No. of Subscribers: 439579
## No. of Videos: 561
```

More video details provided by the function such as date of publish, country of origin, number of overall channel views and number of subscriptions

```
## $kind
## [1] "youtube#channel"
##
## $etag
## [1] "\"_gJQceDMxJ8gP-8T2HLXUoURK8c/M_9jAjNkQ6sZhxHyvE2wf59z0qk\""
##
## $id
## [1] "UCd3zZSCsHS-7smn2HwgM0FA"
##
## $snippet
## $snippet$title
## [1] "Kansiime Anne"
##
## $snippet$description
## [1] "NEW CLIPS EVERY WEEK!!!\nAnne Kansiime is a multiple award
winning talented comedian, TV presenter, Actress and writer in fun
factory, and currently hosts a weekly show on citizen TV Kenya 'Dont
Mess With Kansiime'.\n\n\"One's dream job is the kind of job you do
without constant reminder that you are working.\""
##
## $snippet$publishedAt
## [1] "2012-04-17T12:43:23.000Z"
##
## $snippet$thumbnails
## $snippet$thumbnails$default
## $snippet$thumbnails$default$url
## [1] "https://yt3.ggpht.com/-
kJUq6Pvr0bo/AAAAAAAAAI/AAAAAAAAAA/8hrTYF7WZnA/s88-c-k-no-mo-rj-
```



```
c0xffffffff/photo.jpg"
##
##
## $snippet$thumbnails$medium
## $snippet$thumbnails$medium$url
## [1] "https://yt3.ggpht.com/-
kJUq6Pvr0bo/AAAAAAAAAI/AAAAAAAAAA/8hrTYF7WZnA/s240-c-k-no-mo-rj-
c0xffffffff/photo.jpg"
##
##
## $snippet$thumbnails$high
## $snippet$thumbnails$high$url
## [1] "https://yt3.ggpht.com/-
kJUq6Pvr0bo/AAAAAAAAAI/AAAAAAAAAA/8hrTYF7WZnA/s240-c-k-no-mo-rj-
c0xffffffff/photo.jpg"
##
##
##
## $snippet$localized
## $snippet$localized$title
## [1] "Kansiime Anne"
##
## $snippet$localized$description
## [1] "NEW CLIPS EVERY WEEK!!!\nAnne Kansiime is a multiple award
winning talented comedian, TV presenter, Actress and writer in fun
factory, and currently hosts a weekly show on citizen TV Kenya 'Dont
Mess With Kansiime'. \n\n\"One's dream job is the kind of job you do
without constant reminder that you are working.\""
##
##
## $snippet$country
## [1] "UG"
##
##
## $statistics
## $statistics$viewCount
## [1] "119184892"
##
## $statistics$commentCount
## [1] "28"
##
## $statistics$subscriberCount
## [1] "439579"
##
## $statistics$hiddenSubscriberCount
## [1] FALSE
##
## $statistics$videoCount
## [1] "561"
```

These along other snippets of information are returned by the code `get_channel_stats` function. The function below lists all the channel videos. However, since the channel has over 560 videos, we include the `max_results` parameter to output only 3 results to make a more lean post.

```
list_channel_videos(channel_id = 'UCd3zZSCsHS-
7smn2HwgM0FA',max_results = 3)
```

Unfortunately or fortunately, the information provide by this function does not return the videos by title but rather id.

```
##      .id                                kind
## 1 items1 youtube#playlistItem
## 2 items2 youtube#playlistItem
## 3 items3 youtube#playlistItem
##
##                                     etag
## 1 "_gJQceDMxJ8gP-8T2HLXUoURK8c/fTdSB30hp0g2uxqyWYBJWzMIZzY"
## 2 "_gJQceDMxJ8gP-8T2HLXUoURK8c/0t8YcVKimicXqSPDbjgUtV6Pmdg"
## 3 "_gJQceDMxJ8gP-8T2HLXUoURK8c/KCd-nfQNjwkf_gYq9k2Q9-p3AXA"
##                                     id
contentDetails.videoId
## 1 VVVkM3paU0NzSFMtN3NtbjJId2dNMEZBLjVELWRoVmt2a1pZ          5D-
dhVkvkZY
## 2 VVVkM3paU0NzSFMtN3NtbjJId2dNMEZBLk9UN0V3QmRYeEFZ
OT7EwBdXxAY
## 3 VVVkM3paU0NzSFMtN3NtbjJId2dNMEZBLjFMUV9XUjRSdW9F
1LQ_WR4RuoE
##      contentDetails.videoPublishedAt
## 1      2018-02-08T15:07:27.000Z
## 2      2018-02-09T05:14:26.000Z
## 3      2018-02-05T07:04:42.000Z
```

It would only take one familiar with the package to interpret this output. For sake of easy identification, I have marked the video id.

```
$items[[49]]
$items[[49]]$kind
[1] "youtube#video"

$items[[49]]$etag
[1] "\"_gJQceDMxJ8gP-8T2HLXUoURK8c/EvMc1qv26hsv52-RYjw9KQQRsN8\""

$items[[49]]$id
[1] "tJFk9dPdx00"

$items[[49]]$contentDetails
$items[[49]]$contentDetails$duration
[1] "PT9M49S"

$items[[49]]$contentDetails$dimension
[1] "2d"

$items[[49]]$contentDetails$definition
[1] "hd"

$items[[49]]$contentDetails$caption
[1] "true"

$items[[49]]$contentDetails$licensedContent
[1] TRUE

$items[[49]]$contentDetails$projection
[1] "rectangular"
```

In my next post, I will show how extracted data can be visualized through plots and how to make these statistical findings more meaningful.

[API](#)[You Tube Data Mining](#)[R](#)[R Tutorial](#)[Tuber](#)[About](#) [Help](#) [Terms](#) [Privacy](#)[Get the Medium app](#)

