



Session 4: Getting Started with Commanalytic Toxicity Analysis of Twitter Data using Perspective API

June 24, 2021

SocialMediaLab.ca

CSS Bootcamp Schedule Summer 2021

Session #1 Getting Started with Communalystic: Data Collection from Reddit

May 13, 2021,
10:00-
11:30am (EDT)

Session #2 Toxicity Analysis with Reddit Data using Perspective API

May 27, 2021,
10:00-
11:30am (EDT)

Session #3 Getting Started with Communalystic: Data Collection from Twitter (Twitter Thread via API v2.0 and Twitter Academic Track)

June 10, 2021,
10:00-
11:30am (EDT)

Session #4 Toxicity Analysis of Twitter data using Perspective API

June 24, 2021,
10:00-
11:30am (EDT)

Session #5 Social Network Analysis of Signed Networks with Reddit and Twitter data

July 8, 2021,
10:00-
11:30am (EDT)

Session #6 Getting Started with Communalystic: Data Collection from Facebook & Instagram (via CrowdTangle API) + Social Network Analysis of Two-mode Semantic Networks with CrowdTangle data

July 22, 2021,
10:00-
11:30am (EDT)



Today's agenda



Guest talk by Dr. Paola Pascual-Ferra



Brief recap of how to collect Twitter data and run the Toxicity Analysis in Communalytic

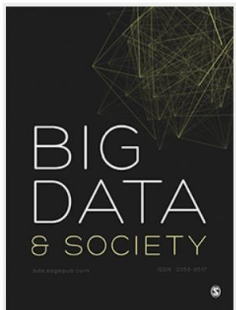


Analyzing Twitter data using Jupyter Notebook

Today's agenda



Guest talk by Dr. Paola Pascual-Ferra



Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic

Paola Pascual-Ferrá , Neil Alperstein , Daniel J Barnett , Rajiv N Rimal 

[Show less](#) ^

First Published June 10, 2021 | Research Article |



<https://doi.org/10.1177/20539517211023533>

[Article information](#) v



10



<https://doi.org/10.1177/20539517211023533>

Today's agenda



Guest talk by Dr. Paola Pascual-Ferra



Brief recap of how to collect Twitter data and run the Toxicity Analysis in Communalytic



Analyzing Twitter data using Jupyter Notebook

Brief Recap of Sessions 2 & 3



Data Collection from Twitter



How to request Twitter's Bearer authentication token



Toxicity Analysis & How to request Google's Perspective API key

Communalytic is a research tool for studying online communities and online discourse.

Communalytic can collect and analyze public data from social media platforms. It uses advanced text and social network analysis techniques to automatically pinpoint toxic and anti-social interactions, identify influencers, map shared interests and the spread of misinformation, and detect signs of possible coordination among seemingly disparate actors.



collect



extract





analyze



visualize

<https://communalytic.com/>

How to choose between Communalystic Edu and Pro.

	 communalystic EDU	 communalystic PRO
Account Type	Free	\$349/6-mo. to support site infrastructure (server-side data collection, storage, processing, analysis and visualization)
Designed For	Students and is ideal for teaching and learning about social media analytics	Academic researchers and is ideal for large scale academic research projects
Account Caps	≤ 30K records shared across 3 datasets	≤ 10M records shared across 50 datasets
Reddit	Live-collection* of public posts from any public subreddit for ≤ 7 consecutive days (Limit: ≤ 30K posts)	Live* & historical collection of public posts from any public subreddit for ≤ 31 consecutive days (Limit: Account Caps)
Twitter Threads (API ver.2) req. Twitter developer's account	Public replies to any public tweet posted within the previous 7 days (Limit: ≤ 30K tweets)	Public replies to any public tweet posted within the previous 7 days (Limit: ≤ 500K tweets/month)
Twitter Academic Track (API ver.2) req. Application to Twitter	Not supported	Full-archive historical-search of tweets back to 2006 (Limit: ≤ 10M tweets/month)
CrowdTangle (FB/IG) URL Search req. a CrowdTangle account	Public Facebook or Instagram posts that shared the same URL (Limit: ≤ 30K posts)	Public Facebook or Instagram posts that shared the same URL (Limit: Account Caps)

* Live-collection = the collection of posts/tweets posted on or after the date when you initiated the data collection.

My Datasets

Server Time: June 10, 2021 12:19 UTC

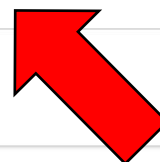
Collect data from

Reddit (LIVE)

Twitter Thread

CrowdTangle

CSV File



Search dataset list...



My Datasets

[Shared With Me](#)



Donald J. Trump @realDonaldTrump · Sep 10

Sending out 80 MILLION BALLOTS to people who aren't even asking for a Ballot is unfair and a total fraud in the making. Look at what's going on right now!

20.4K

30.5K

98.6K



David Former Trump Supporter Weissman @davidm... · Sep 10

Replying to @realDonaldTrump
Tweet reported for misinformation on the elections.

104

146

2.3K



Osha Davidson @OshaDavidson · Sep 10

Me too.

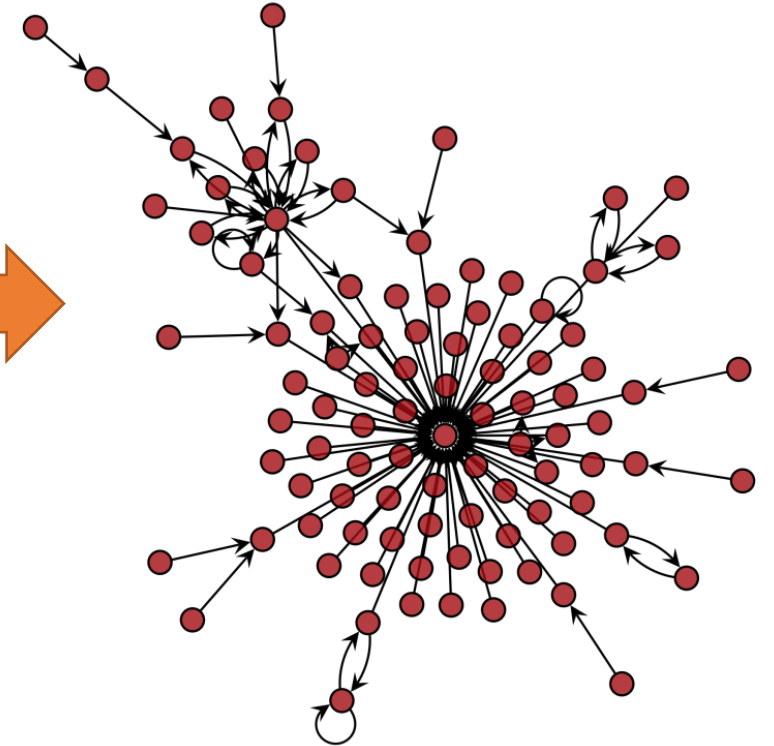
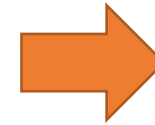
5

2

181



[Show replies](#)



A network representation of replies to a given tweet

Server Time: May 11, 2021 14:04 UTC

Collect data from

Reddit (LIVE)

Twitter Thread

CrowdTangle

CSV File



Search dataset list...



[← Back to My Datasets](#)

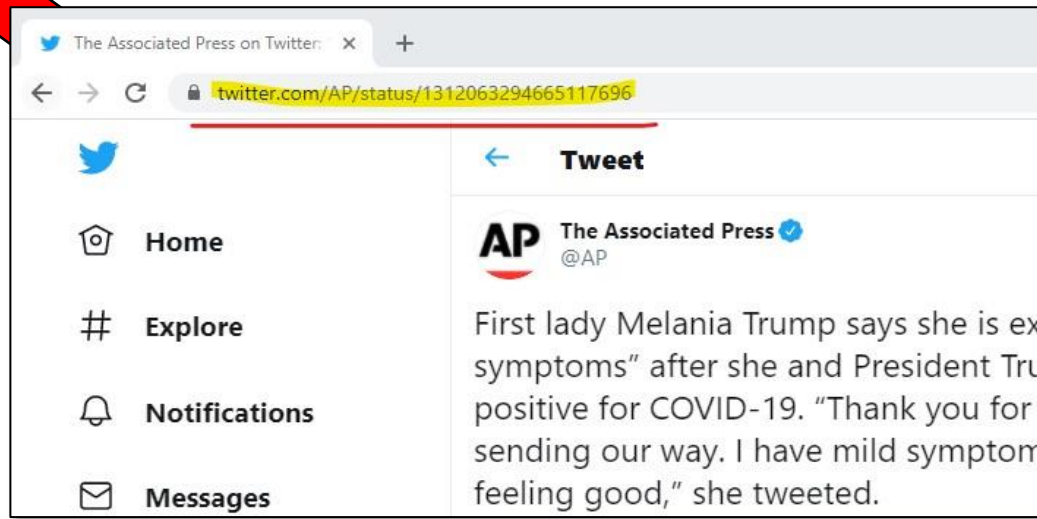
Twitter Thread Data Collection

Please enter the following details to begin collection

 ?

Email me once job completes

 Start Data Collection



Reddit (LIVE)

Twitter Thread

CrowdTangle

CSV File

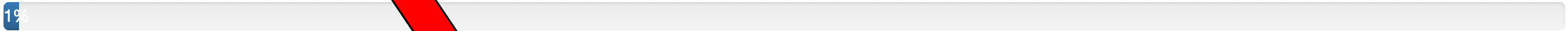
Search dataset list...



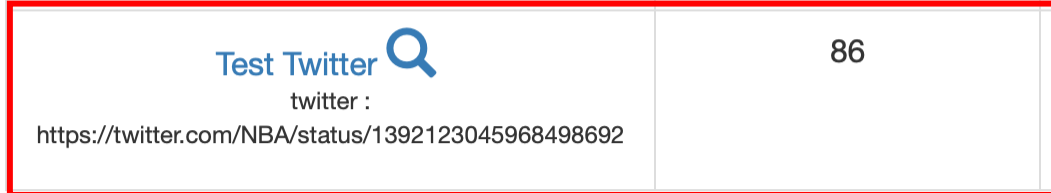
My Datasets

Shared With Me

Total records in the account: 186 / 30000



Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
Test Reddit reddit : CanadaPolitics Click to check progress	Extracting until 2021-05-12 100					
Test Twitter twitter : https://twitter.com/NBA/status/1392123045968498692	86					





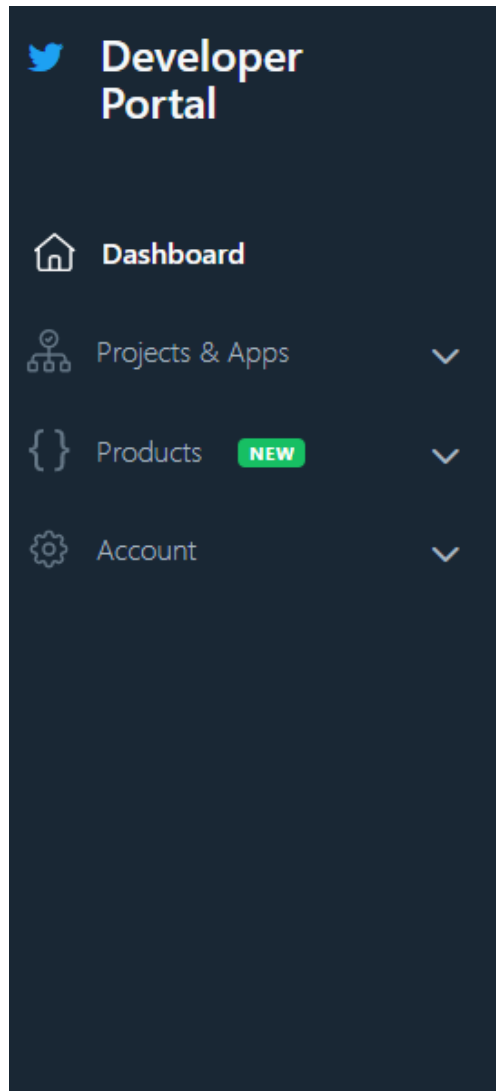
Working with Twitter's Academic Product Track in **Communalytic Pro**

<https://pro.communalytic.com>

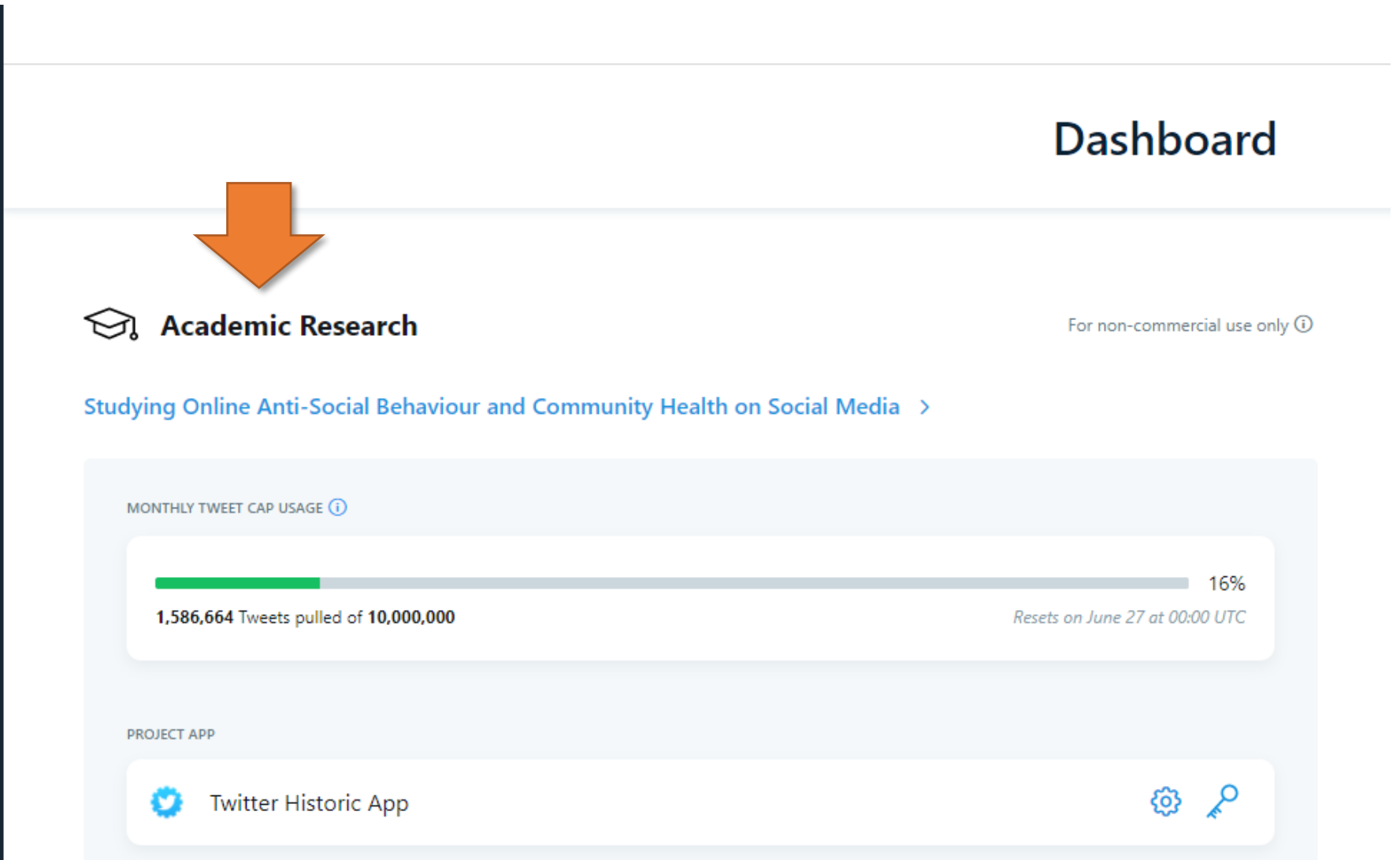
Eligibility requirements:

- Master's student, doctoral candidate, postdoc, faculty, or research-focused employee at an academic institution
- A clearly defined research project (i.e., you have a specific plan for Twitter data use, analysis, and sharing)
- Non-commercial use
- **Apply at <https://developer.twitter.com/en/portal/petition/academic/is-it-right-for-you>**

Twitter's Developer Dashboard: Confirm Access to the Academic API



- Developer Portal
- Dashboard
- Projects & Apps
- Products **NEW**
- Account



Dashboard

Academic Research For non-commercial use only ⓘ

[Studying Online Anti-Social Behaviour and Community Health on Social Media >](#)

MONTHLY TWEET CAP USAGE ⓘ

1,586,664 Tweets pulled of 10,000,000 16% Resets on June 27 at 00:00 UTC

PROJECT APP

Twitter Historic App ⚙️ 🔑

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Twitter's Developer Dashboard: Retrieve your "Bearer" access token

The screenshot displays the Twitter Developer Dashboard interface. On the left is a dark sidebar with the Twitter logo and the text 'Developer Portal'. Below this are navigation links: 'Dashboard', 'Projects & Apps' (with a dropdown arrow), 'Overview', 'ThreadExplorer', 'App Name' (highlighted with a red arrow), and 'STANDALONE APPS'. At the bottom of the sidebar are 'Products' (with a 'NEW' badge and a dropdown arrow) and 'Account' (with a dropdown arrow). The main content area is white and shows the 'App Name' header with 'THREDEXPLORER' above it. Below the header are two tabs: 'Settings' and 'Keys and tokens' (selected, with an orange arrow pointing to it). The 'Keys and tokens' section is divided into two main areas: 'Consumer Keys' and 'Authentication Tokens'. Under 'Consumer Keys', there is a card for 'API Key & Secret' for '@polidashboard', with 'View Keys' and 'Regenerate' buttons. Under 'Authentication Tokens', there is a card for 'Bearer Token' generated on August 13, 2020, with 'Regenerate' and 'Revoke' buttons. Below that is a card for 'Access Token & Secret' for '@polidashboard' with a 'Generate' button.

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Adding your “Bearer” access token to your Communalytic PRO account

The screenshot shows the Communalytic PRO user interface. At the top left is the logo and the text "(PRO)". Navigation links for "FAQ", "Tutorials", and "Publications" are in the top right. A user menu is open, showing options like "My Datasets", "My Profile", "Share Requests", "Share History", and "Logout".

The main content area is titled "Personal Details" and "API Keys". A modal dialog titled "Enter Twitter Bearer API Key" is open, featuring a text input field labeled "Twitter Bearer Key" and "Submit" and "Close" buttons. The "Enter Key" button in the "API Keys" section is highlighted with an orange box and a yellow "2" callout. The input field in the modal is also highlighted with an orange box and a yellow "3" callout.

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Full-archive Search API: Create a new search (1)

commuanalytic (PRO) FAQ Tutorials Publications My Datasets Profile

My Datasets

Server Time: June 07, 2021 13:18 UTC

Collect Data from

Reddit (HISTORICAL) | Reddit (LIVE) | CSV File | Twitter Thread | Twitter Archive (Academic Track) | CrowdTangle

* You can simultaneously run two Reddit, one Twitter and one CrowdTangle collectors.

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Full-archive Search API: Create a new search (2)

Search Query
(up to 1024
chars)



Collect Historical Tweets

via [Twitter's Academic Research Product Track](#)

Please enter the following details to begin collection

Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss)

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss)

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Toxicity Analysis with Perspective API

To access the API, review the [Get Started](#) documentation.

Overview

Key Concepts

Attributes & Languages

Best Practices & Risks

Methods

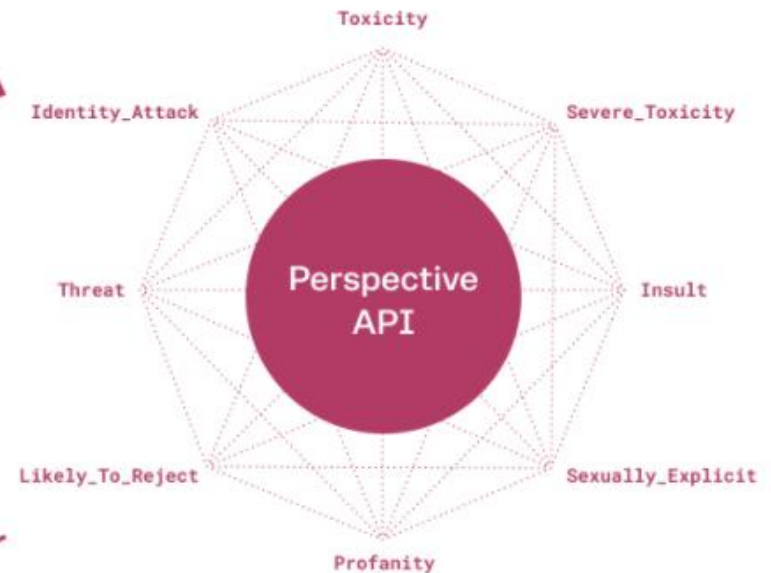
Limits & Errors

FAQs

INPUT: TEXT
"Shut up. You're an idiot!"

OUTPUT: SCORE

Toxicity	0.99
Severe_Toxicity	0.75
Insult	1.0
Sexually_Explicit	0.04
Profanity	0.93
Likely_To_Reject	0.99
Threat	0.15
Identity_Attack	0.03



<https://support.perspectiveapi.com/s/docs-get-started>

1. Twitter API



2. Collection



3. Analysis



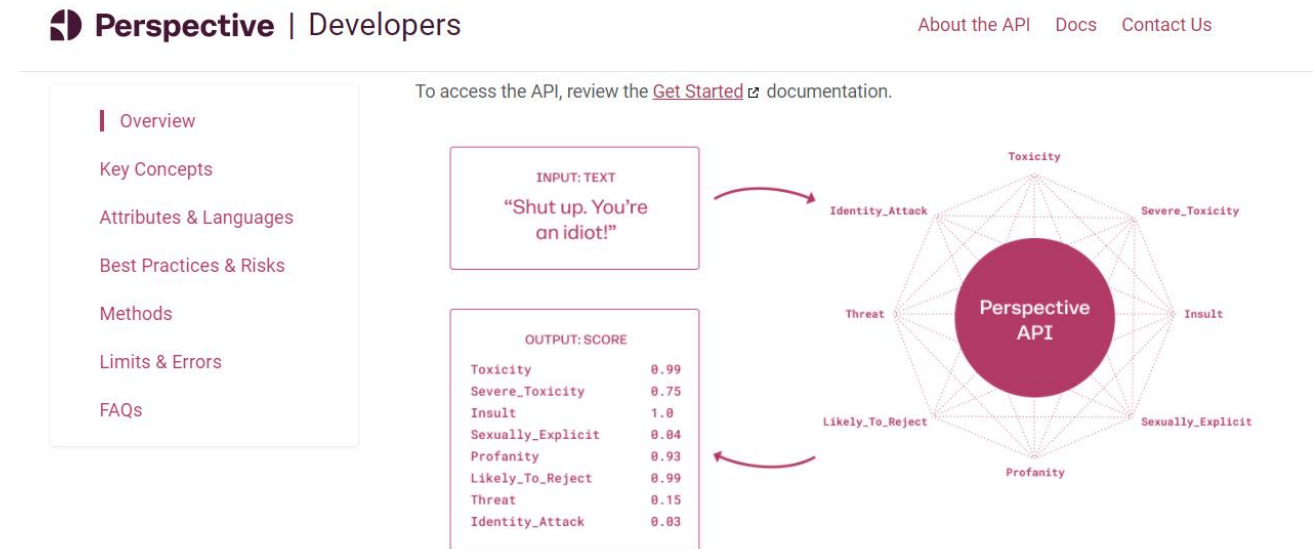
4. Export



5. Sharing

Toxicity Analysis with Perspective API

- **Training data:** online forums including Wikipedia & New York Times comments, with crowdsourced labels (e.g., a comment is “toxic” or not)
- **Machine Learning:** Convolutional Neural Network (CNN) trained with GloVe word embeddings



<https://support.perspectiveapi.com/s/docs-get-started>

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Toxicity Scores from Perspective API

	Definition	Sample post
Toxicity	Rude, disrespectful, or unreasonable post	“This is one of stupidest things I've read but fuck me I laughed at the second line”
Severe toxicity	Very hateful, aggressive, disrespectful post. This score is less sensitive to posts that include positive uses of curse words	“Fuck off pathetic loser, no one cares about your worthless opinion”
Insult	Insulting, inflammatory, or negative post toward an individual or a group	“How fucking stupid is [Name]? That is pretty fucking stupid. What's next - a deep fake having him say racist things as a "social experiment"?”
Identity attack	Negative post attacking someone because of their identity (including race, gender, sexual orientation, ideology, religion, nationality, etc.)	“You people are a bunch of fags. And I voted for [Political Party Name]”
Profanity	Post with swear words or other obscene language	“Why vote for the [Political Party Name] when you know they won't win shit.”
Threat	Post with an intention to inflict pain, injury, or violence against an individual or group	“Shoot all yellow vests! We have to kill all Nazis!”

<https://support.perspectiveapi.com/s/about-the-api-attributes-and-languages>

1. Twitter API



2. Collection



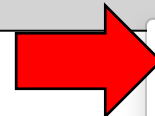
3. Analysis



4. Export



5. Sharing



My Datasets

Server Time: May 11, 2021 15:04 UTC

Collect data from

Reddit (LIVE)

Twitter Thread

CrowdTangle

CSV File

Search dataset list...



1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

API Keys

Twitter Bearer
Token:

Remove Key

CrowdTangle
API:

Apply for academic/research access to Facebook's
CrowdTangle [here](#)

Enter Key

Perspective
API:

-

Enter Key

Perspective is a machine learning API by Jigsaw and Google designed to conduct a 'toxicity' analysis of online comments. To use this API within Communalytic, please follow [these steps](#) to generate an API key. Once generated, enter your API key in the text field above.



***Note:** We recommend using a personal Gmail account to request an API key, some institutional emails may block Google Cloud console / API key requests

Account tier: STUDENT

Datasets: 2/3

Co

Enter Perspective API Key ✕HRTK

Submit

Close

Pe

Per

ana

to generate an API key. Once generated, enter your API key in the text field above.

***Note:** We recommend using a personal Gmail account to request an API key, some institutional emails may block Google Cloud console / API key requests

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Reddit (LIVE)

Twitter Thread

CrowdTangle

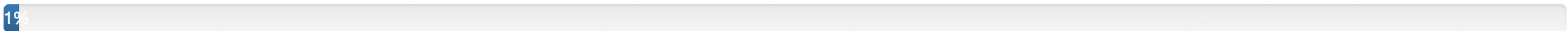
CSV File























Search dataset list... 

My Datasets

Shared With Me

Total records in the account: 186 / 30000



Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
metacanada till Feb 1  reddit : metacanada	6,675 					
metacanada 4 book chapter  reddit : -	22,560  					
metacanada  reddit : metacanada	6,703 					



Overview	
Dataset Name:	metacanada 4 book chapter
Platform:	reddit
Subreddit:	-
Collection started:	2020-02-11 02:11
Collection before:	2020-02-11 02:11
Records:	22,560

🕒 Posts Per Day

☁ Word Cloud

📄 Top Ten Posters

📊 Toxicity Analysis

Toxicity Analysis

The Toxicity analysis can only run with one language mode at a time. Please select the primary language of your dataset using the drop-down menu below.

English

📊 Start Analysis

- Toxicity Analysis is based on a machine learning API called [Perspective API](#) by Google. If you already have an API key for Perspective, it can be added to your account under [My Profile](#); otherwise, see [this guide](#) on how to generate a new Perspective API.
- Due to the post length restriction imposed by the API, Communalystic can only analyze the first 3000 characters of each post. Posts longer than 3000 characters will be automatically truncated.
- Link's and URLs will be removed before text is sent to API for analysis.
- The resulting toxicity scores will be added to the export files and available for download via the "Export Posts" and "Export Network" options.

1. Twitter API

2. Collection

3. Analysis

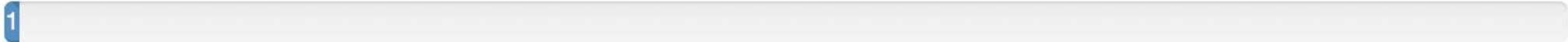
4. Export

5. Sharing

Toxicity Analysis

Analysis in Progress ...

You may close this window and visit it later.



Progress: 59 / 4451

Estimated Time Left: 1h 20m

Check progress in **43** sec

Cancel Analysis

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Toxicity Analysis

There were 22560 comments analyzed in English

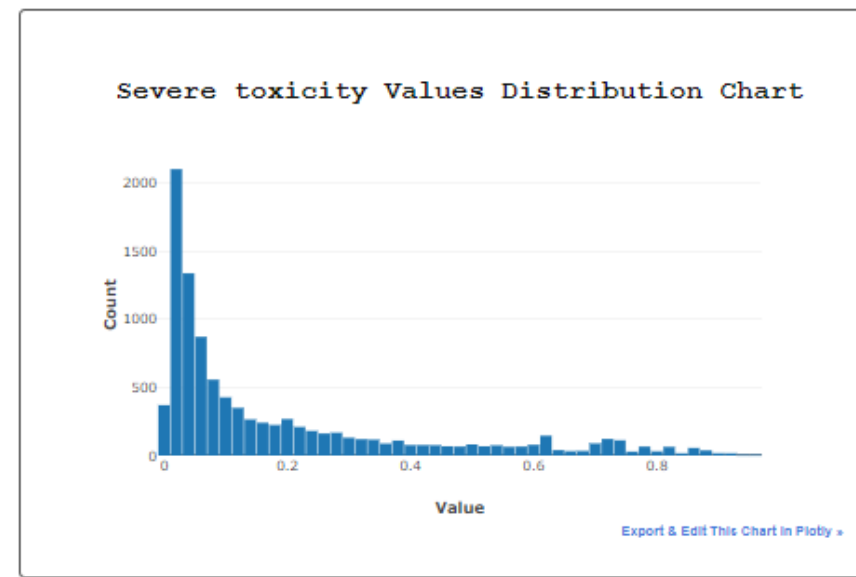
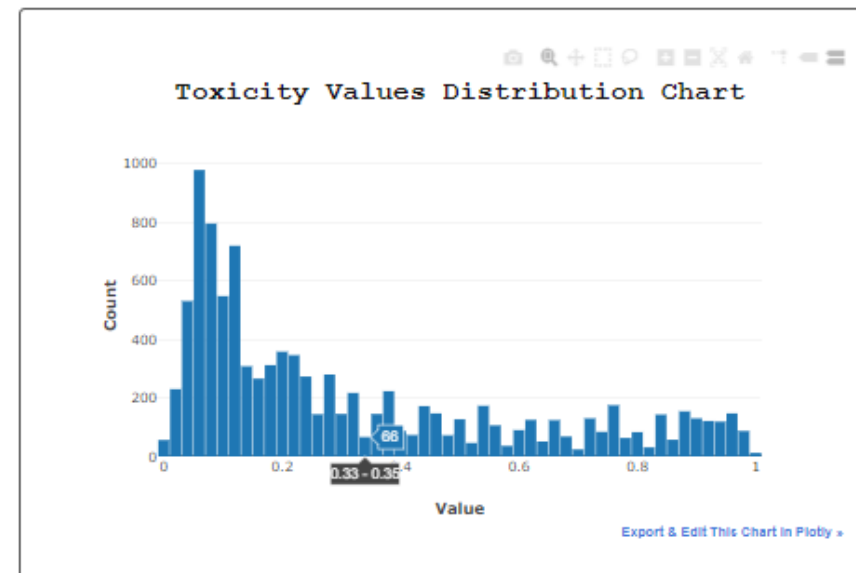
Reset Analysis

⚠ The Toxicity analysis can only run with one language mode at a time. To change the primary language selection for this dataset, click on the Reset Analysis button and rerun the analysis as needed.

Click on the highest/lowest values to see the top posts for each category

Download the Toxicity analysis results as a CSV file.

	Average for dataset	Highest value	Lowest value
Toxicity ⓘ	0.32	1.00	0.00
Severe toxicity ⓘ	0.19	0.95	0.00
Identity attack ⓘ	0.25	0.98	0.00
Insult ⓘ	0.28	0.99	0.00
Profanity ⓘ	0.26	0.99	0.00
Threat ⓘ	0.22	0.99	0.01
Sexually Explicit ⓘ	0.17	1.00	0.00
Flirtation ⓘ	0.30	0.98	0.02



1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

Export Twitter Dataset from Commanalytic

BREAKING: US Senate votes to acquit former President Trum 
twitter :
<https://twitter.com/NBCPolitics/status/1360693216601849857>

1,922



FAQ

Tutorials

Publications

My Datasets

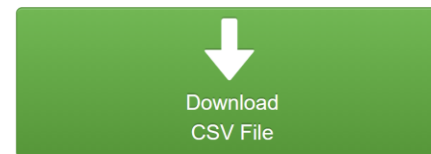



← Back to My Datasets

Overview

Dataset Name: **BREAKING: US Senate votes to acquit former President Trum**
Platform: **twitter**
Search Criteria: **<https://twitter.com/NBCPolitics/status/1360693216601849857>**
Collection started: **2021-02-13 20:59**
Collection ended: **2021-02-13 20:59**
Records: **1,922**

Download Dataset



 If this is the first time that you are working with Twitter data, please review the [Twitter attributes guide](#) prior to downloading and working with the CSV file of your dataset.

1. Twitter API



2. Collection



3. Analysis



4. Export



5. Sharing

How to open the exported CSV file in Excel

The screenshot shows the Microsoft Excel interface with the 'Data' tab selected. The ribbon includes 'Get & Transform Data', 'Queries & Connections', 'Data Types', and 'Sort & Filter'. The 'From Text/CSV' option is highlighted. Below the ribbon, a table of tweet data is displayed with columns for tweet_id, date, text, and name.

	tweet_id	date	text	name
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to get citizens	Vivekanand Singh
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the subject. Tha	Vinny D
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just reminder leave	Vinny D
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy has buried our children in debt. He hates Cana	Lioness
6				
7				
8				
9				
10				
11				
12				
13				

1. Twitter API



2. Collection



3. Metadata



4. Export










5. Sharing

Data Sharing via Tweet Ids

Collect / Export tweets as CSV

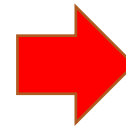
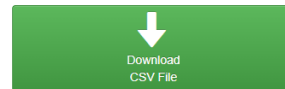
My Datasets Shared With Me

Total records in the account: 186 / 30000

Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
Test Reddit reddit : CanadaPolitics Click to check progress	Extracting until 2021-05-12 100					
Test Twitter twitter : https://twitter.com/NBA/status/1392123045968498692	86					



Download Dataset



	A	B	
1	tweet_id	date	text
2	1401598862201180162	2021-06-06T17:56:23.000Z	@JustinTrudeau
3	1401528927466639360	2021-06-06T13:18:30.000Z	This is real #genocide.https://t.co/ZA
4	1401599908059267073	2021-06-06T18:00:33.000Z	@JustinTrudeau My hobby is a barbe
5	1401598610551181312	2021-06-06T17:55:23.000Z	@JustinTrudeau @JustinTrudeau you
6	1401597964896722945	2021-06-06T17:52:49.000Z	@capano_rosa @JustinTrudeau @Jon
7	1401597739339718659	2021-06-06T17:51:56.000Z	@JustinTrudeau @marcomendicino
8	1401597661833142275	2021-06-06T17:51:37.000Z	@capano_rosa @JustinTrudeau @Jon
9	1401596728000892933	2021-06-06T17:47:55.000Z	@JustinTrudeau Open the fucking co
10	1401591680587812865	2021-06-06T17:27:51.000Z	@JustinTrudeau Everyone is pretty ti
11	1401589438581415936	2021-06-06T17:18:57.000Z	@JustinTrudeau Yeah maybe... but Ca
12	1401589398773374983	2021-06-06T17:18:47.000Z	@JustinTrudeau Science is a process
13	1401586720320090117	2021-06-06T17:08:09.000Z	@JustinTrudeau how many of the 2 b
14	1401586531563827200	2021-06-06T17:07:24.000Z	@JustinTrudeau Shame on you and y
15	1401579144132108290	2021-06-06T16:38:02.000Z	@JustinTrudeau NDP & Conserv
16	1401578885427482624	2021-06-06T16:37:01.000Z	@JustinTrudeau Taxing Cdns to death
17	1401577177846538245	2021-06-06T16:30:13.000Z	@JustinTrudeau Thank you PM Trude
18	1401574285643157504	2021-06-06T16:18:44.000Z	@JustinTrudeau https://t.co/na4uAR
19	1401573815537176576	2021-06-06T16:16:52.000Z	@JustinTrudeau https://t.co/6ez0K7T
20	1401573306952728577	2021-06-06T16:14:51.000Z	@JustinTrudeau This is a great effort
21	1401572148536221705	2021-06-06T16:10:14.000Z	@JustinTrudeau It's a humble reques
22	1401571865877913606	2021-06-06T16:09:07.000Z	@JustinTrudeau It's a humble reques
23	1401571804200673281	2021-06-06T16:08:52.000Z	@JustinTrudeau @marcomendicino
24	1401571047640555528	2021-06-06T16:05:52.000Z	@JustinTrudeau Those trees were to
25	1401568121698373641	2021-06-06T15:54:14.000Z	@JustinTrudeau So Happyyyy dearrrr

We only shared **tweet_ids** + **Toxicity Scores**
+ **tweet statuses** (as checked using the **twarc** library)

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

Today's agenda



Guest talk by Dr. Paola Pascual-Ferra



Brief recap of how to collect Twitter data and run the Toxicity Analysis in Commanalytic



Analyzing Twitter data using Jupyter Notebook
with Shahnawaz Attarwala

Sample Dataset of Tweet IDs



communalytic FAQ TUTORIALS PUBLICATIONS

Dataset

On October 2nd, 2020, U.S. President Donald Trump tweeted that he and the First Lady of the United States (Melania Trump), both tested positive for COVID-19 (see Figure 1). Within seconds, his tweet received thousands of replies.

Donald J. Trump @realDonaldTrump
Tonight, @FLOTUS and I tested positive for COVID-19. We will begin our quarantine and recovery process immediately. We will get through this TOGETHER!
12:54 AM · Oct 2, 2020 · Twitter for iPhone

404.3K Retweets 502.2K Quote Tweets 1.8M Likes

Figure 1: Donald Trump's tweet announcing his COVID-19 diagnosis

<https://communalytic.com/2021/01/13/toxicity-analysis-of-a-twitter-thread/>



Scholars Portal **Dataverse** Search User Guide Support English Log In

Social Media Lab (Ryerson University)

Scholars Portal Dataverse > Ryerson University Dataverse > Social Media Lab >

Twitter Dataset: Replies to President Trump's Tweet about contracting the Novel Coronavirus

Version 1.1

Gruzd, Anatoliy, 2021, "Twitter Dataset: Replies to President Trump's Tweet about contracting the Novel Coronavirus", <https://doi.org/10.5683/SP2/BMCGI3>, Scholars Portal Dataverse, V1, UNF:6:ZoGBrMO4f8/aPycUi+bk+A== [fileUNF]

Cite Dataset Learn about Data Citation Standards.

Access Dataset

Download Options

- Original Format ZIP (34.3 MB)
- Archival Format (.tab) ZIP (35.1 MB)

Description

On October 2nd, 2020, U.S. President Donald Trump tweeted that he and the First Lady of the United States (FLOTUS), Melania Trump, both tested positive for COVID-19 (<https://twitter.com/realDonaldTrump/status/1311892190680014849>). Within seconds, his tweet received thousands of replies. The dataset consists of 298,172 replies to Donald Trump's tweet announcing his COVID-19 diagnosis, posted on October 2nd between 6am and 12:30pm (ET). It contains tweet ids, the toxicity scores (as calculated by Google's Perspective API via <https://Communalytic.com>) and tweet availability status values (via twarc library). Following Twitter's API policy, we stripped metadata associated with each tweet. So, if you'd like to examine potential relationships between other metadata elements, you would need to recollect original tweets using tools like DocNow's Hydrator first. The only downside of this approach is that tweets that have been blocked or deleted will not be recollect. To help you get started, we also shared our Exploratory Data Analysis (EDA) Python script at <https://github.com/RUSocialMediaLab/toxicityanalysis>

Subject Computer and Information Science

Keyword COVID, Donald Trump, Toxicity Analysis, Twitter

<https://doi.org/10.5683/SP2/BMCGI3>

You can recollect tweets using programs like Hydrator

tweet_id

1401598862201180162

1401528927466639360

1401599908059267073

1401598610551181312

1401597964896722945

1401597739339718659

1401597661833142275

1401596728000892933

1401591680587812865

1401589438581415936

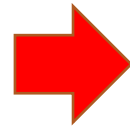
1401589398773374983

1401586720320090117

1401586531563827200

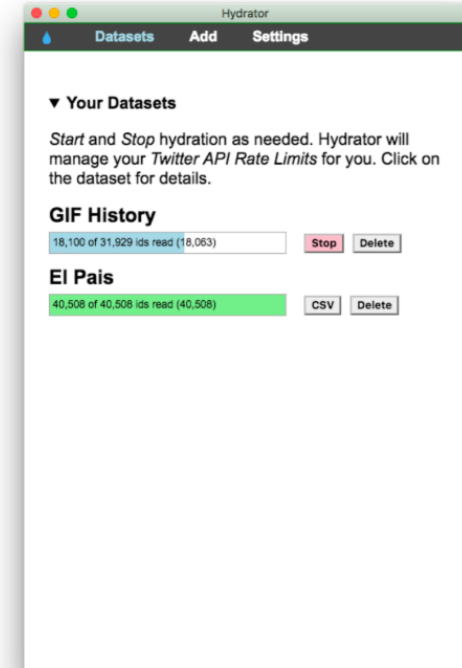
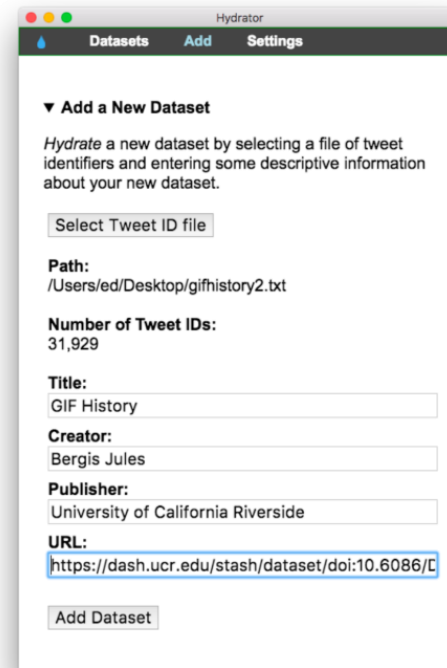
1401579144132108290

1401578885427482624



Hydrator

<https://github.com/DocNow/hydrator>



Why use Python for Data Science projects..



Source: <https://www.nobledesktop.com/learn/python/python-vs-excel>

Benefits of using Jupyter Notebook



Source: Jupyter logo

- Jupyter notebook is an Integrated Development Environment (IDE) that acts as an interface between user and python.
- It is an open-source project.
- Supports multiple programming languages including Python and R.
- It allows users to create and share code and documents easily.
- It supports code, text, and interactive data visualization.
- Support single code cell execution, independent of other code cells

```
▶ 1 #Display brief info about the dataset  
2 pdf.info()
```

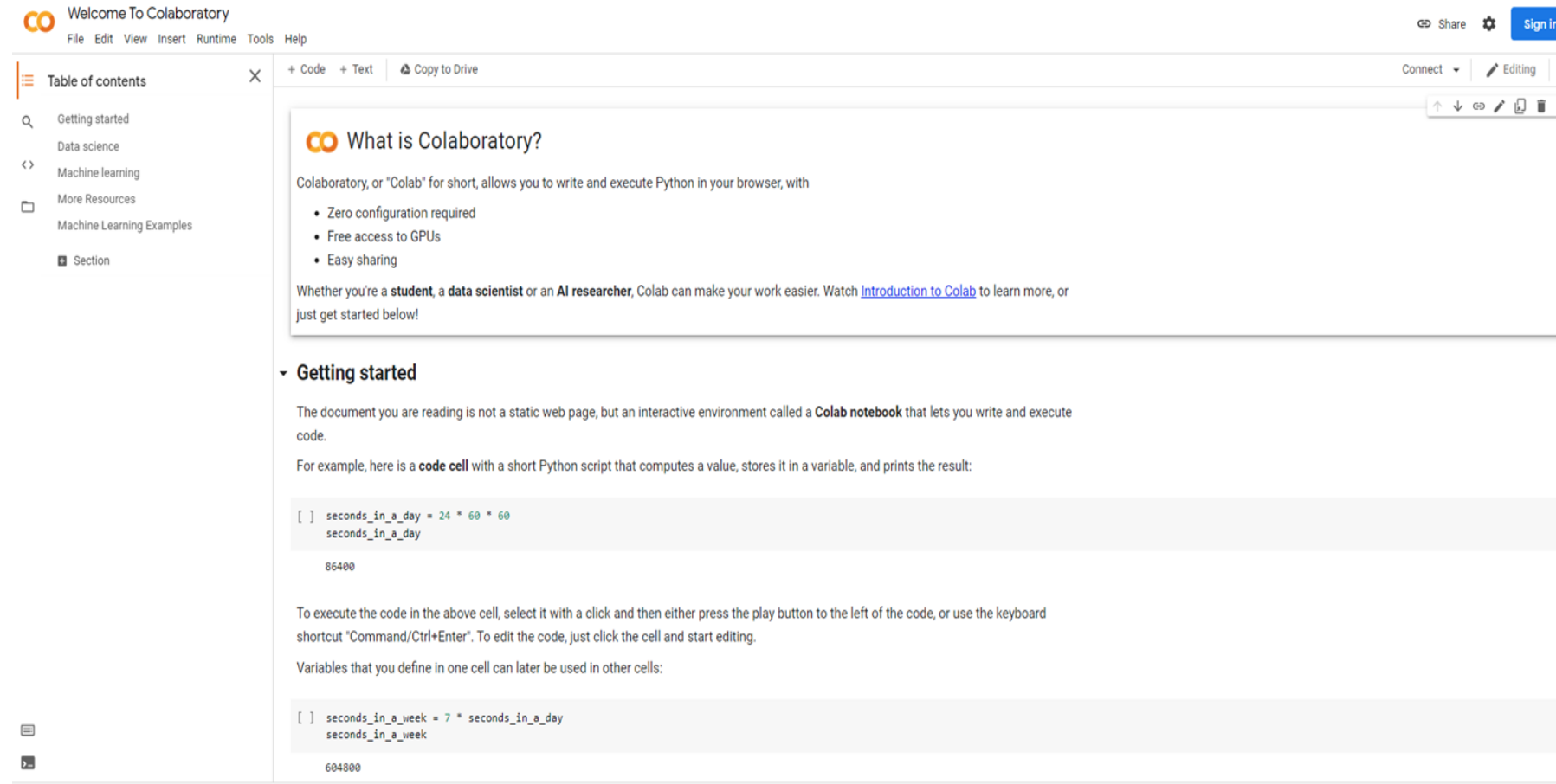
```
↳ <class 'pandas.core.frame.DataFrame'>  
RangeIndex: 298173 entries, 0 to 298172  
Data columns (total 11 columns):  
#   Column                Non-Null Count  Dtype  
---  ---  
0   tweet_id              298173 non-null  int64  
1   tweetstatus          298173 non-null  object  
2   language              298173 non-null  object  
3   Toxicity              298173 non-null  float64  
4   Severe Toxicity      298173 non-null  float64  
5   Profane               298173 non-null  float64  
6   Identity Attack      298173 non-null  float64  
7   Insult                298173 non-null  float64  
8   Threat                298173 non-null  float64  
9   Sexually Explicit    298173 non-null  float64  
10  Flirtation            298173 non-null  float64  
dtypes: float64(8), int64(1), object(2)  
memory usage: 25.0+ MB
```

```
[5] 1 #How many different language used in the tweets?  
2 len(pdf["language"].unique())
```

Google's free version of Jupyter Notebook

Benefits of using [Google Colab's](#) Jupyter Notebook:

- Pre-Installed Libraries
- Saved on the Cloud
- Collaboration
- Free RAM and Storage



The screenshot displays the Google Colaboratory web interface. At the top, it says 'Welcome To Colaboratory' with a menu bar including 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. On the right, there are 'Share', 'Settings', and 'Sign In' buttons. A left sidebar contains a 'Table of contents' with links to 'Getting started', 'Data science', 'Machine learning', 'More Resources', 'Machine Learning Examples', and 'Section'. The main content area is titled 'What is Colaboratory?' and explains that Colab allows writing and executing Python in a browser. It lists three benefits: zero configuration required, free access to GPUs, and easy sharing. Below this, a 'Getting started' section explains that the environment is interactive and provides an example of a code cell. The code cell contains a Python script that calculates the number of seconds in a day (24 * 60 * 60) and prints the result, 86400. A second code cell shows how to calculate the number of seconds in a week (7 * seconds_in_a_day) and prints the result, 604800.

Table of contents

- Getting started
- Data science
- Machine learning
- More Resources
- Machine Learning Examples
- Section

What is Colaboratory?

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Whether you're a **student**, a **data scientist** or an **AI researcher**, Colab can make your work easier. Watch [Introduction to Colab](#) to learn more, or just get started below!

Getting started

The document you are reading is not a static web page, but an interactive environment called a **Colab notebook** that lets you write and execute code.

For example, here is a **code cell** with a short Python script that computes a value, stores it in a variable, and prints the result:

```
[ ] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day

86400
```

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut "Command/Ctrl+Enter". To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

```
[ ] seconds_in_a_week = 7 * seconds_in_a_day
seconds_in_a_week

604800
```

Sample Google colab's Jupyter Notebook

The complete Jupyter Notebook of Toxicity Analysis is at this [link](#).

