Twitter Data Extractor

User Manual

Author: Coşkun Deniz <codenineeight@gmail.com>

Created at: September 2022

Table of Contents

Introduction

This command-line tool extracts user and tweet data from Twitter and reports the results to CSV, Excel, Google Sheets documents or MongoDB, SQLite databases.

Supported Features

- Extract single/multiple user data.
- Extract user’s friends/followers data.
- Extract tweets data for a user.
- Extract tweets data for a search keyword.
- Report results to CSV, Excel or Google Sheets documents.
- Report results to MongoDB or SQLite databases.

Basic package only contains results in the requested output file/database.

Premium package contains the source code and installation of the script on the user side. User of the tool should have the Twitter API access.

How to Setup

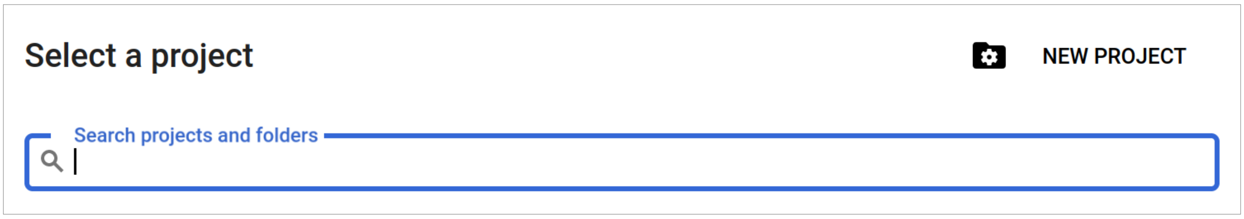
Installing Dependencies

Run the following commands to install required packages in the project directory.

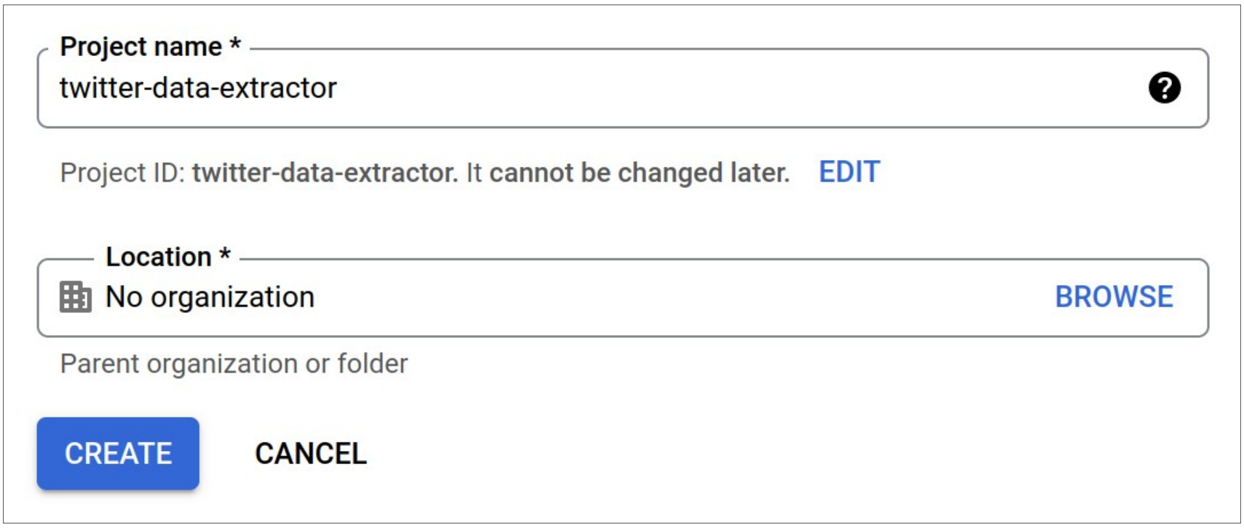
```
python -m venv env
source env/bin/activate
python -m pip install -r requirements.txt
```

Creating a Project on Google Cloud Platform for Google Sheets Usage

1. Go to Google Developers Console and create a new project (or select the one you already have).



Write a name for your project and click the *CREATE* button.



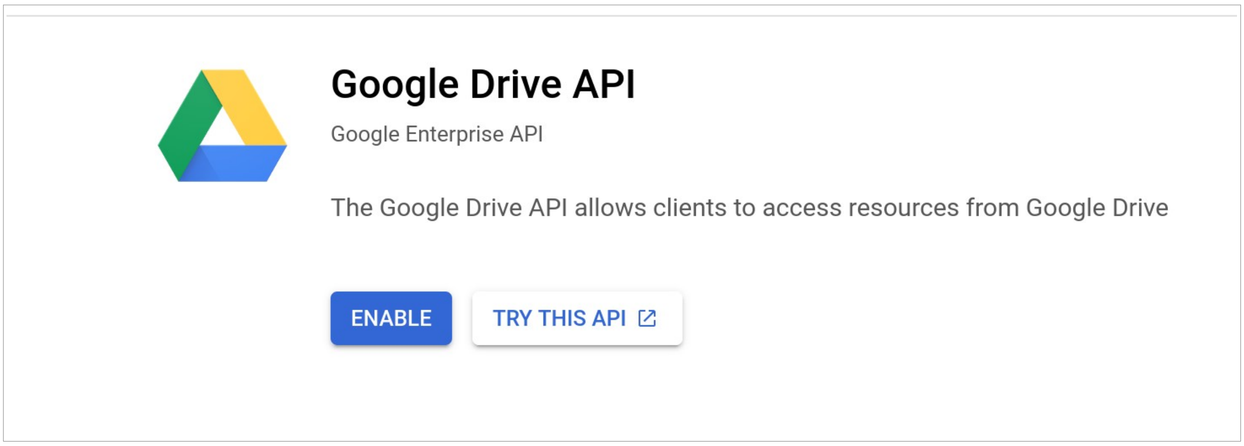
Project name *
twitter-data-extractor

Project ID: twitter-data-extractor. It cannot be changed later. EDIT

Location *
No organization

Parent organization or folder

2. Select APIs & Services → Library from the side menu. First search for “Google Drive API” and enable it.

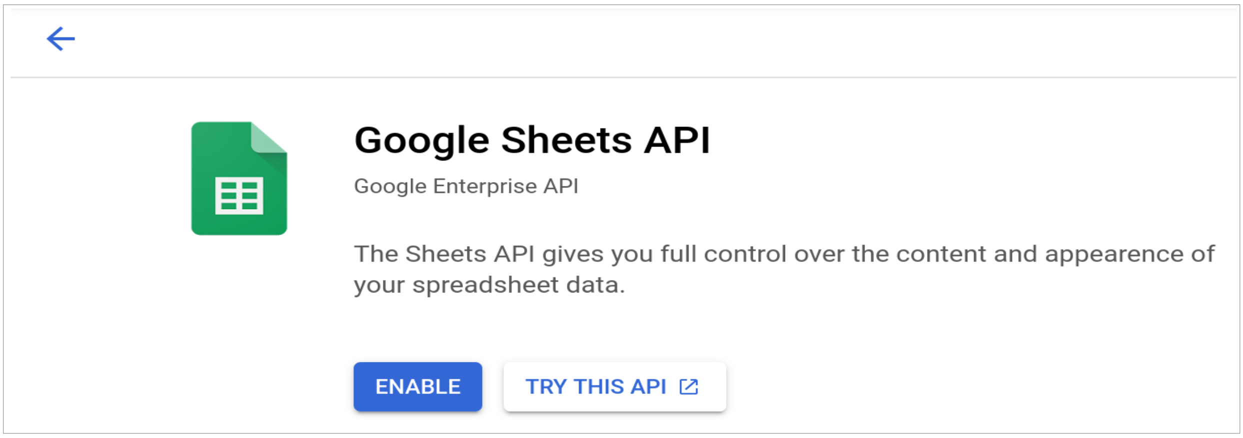


Google Drive API

Google Enterprise API

The Google Drive API allows clients to access resources from Google Drive

Then search for “Google Sheets API” and enable it.



Google Sheets API

Google Enterprise API

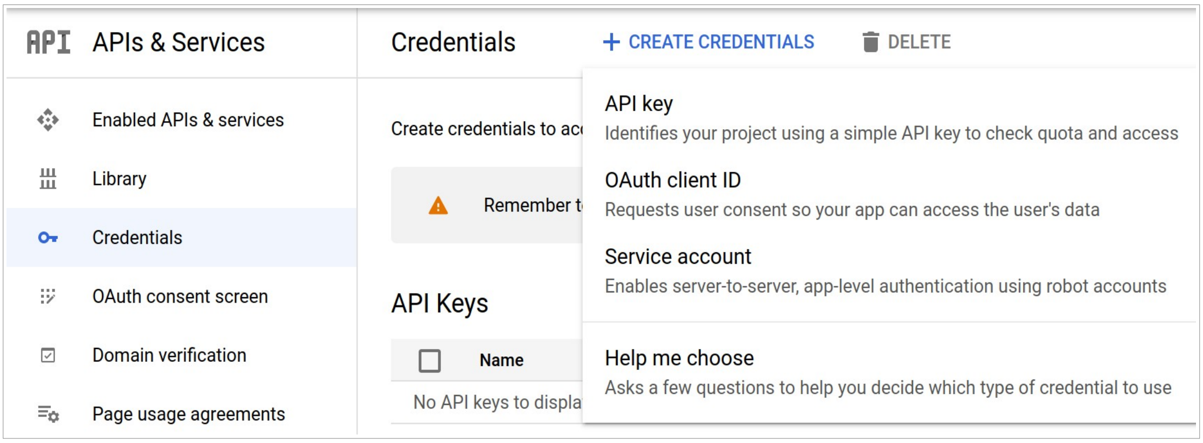
The Sheets API gives you full control over the content and appearence of your spreadsheet data.

3. Create credentials for service account.

   Since it’s a separate account, by default it does not have access to any spreadsheet until you share it with this account. Just like any other Google account.

   1. Go to APIs & Services → Credentials from the side menu.

   2. Choose “CREATE CREDENTIALS > Service account”.



API key

Identifies your project using a simple API key to check quota and access

OAuth client ID

Requests user consent so your app can access the user's data

Service account

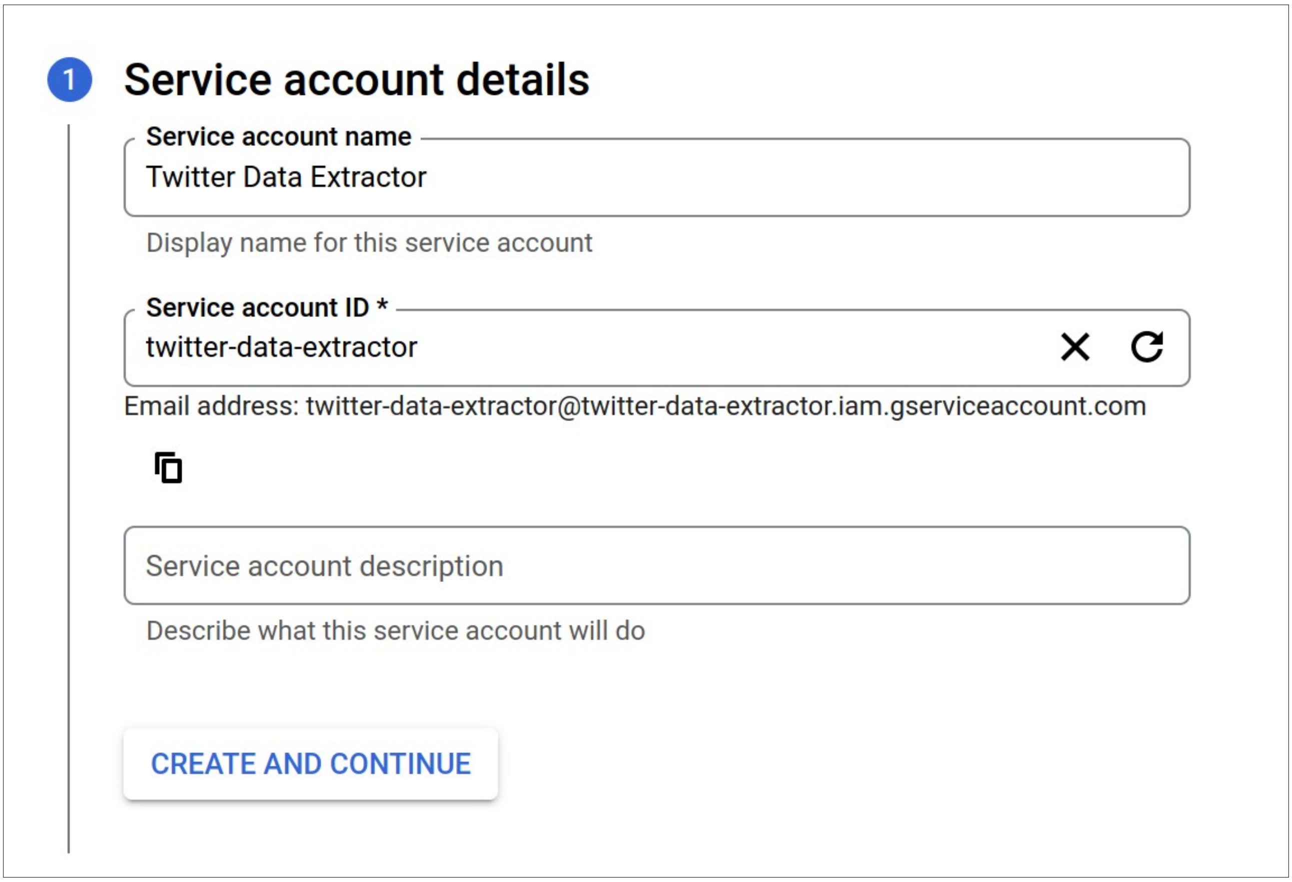
Enables server-to-server, app-level authentication using robot accounts

Help me choose

Asks a few questions to help you decide which type of credential to use

   3. Fill out the form.

   4. Click “CREATE AND CONTINUE” and “DONE”.



Service account details

Service account name: Twitter Data Extractor

Display name for this service account

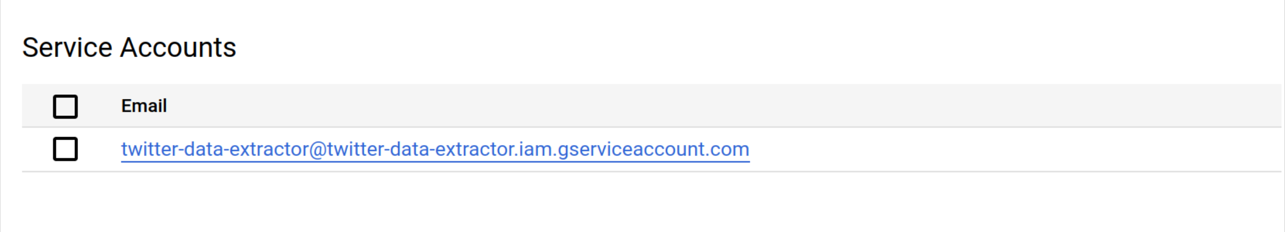
Service account ID *: twitter-data-extractor

Email address: twitter-data-extractor@twitter-data-extractor.iam.gserviceaccount.com

Service account description

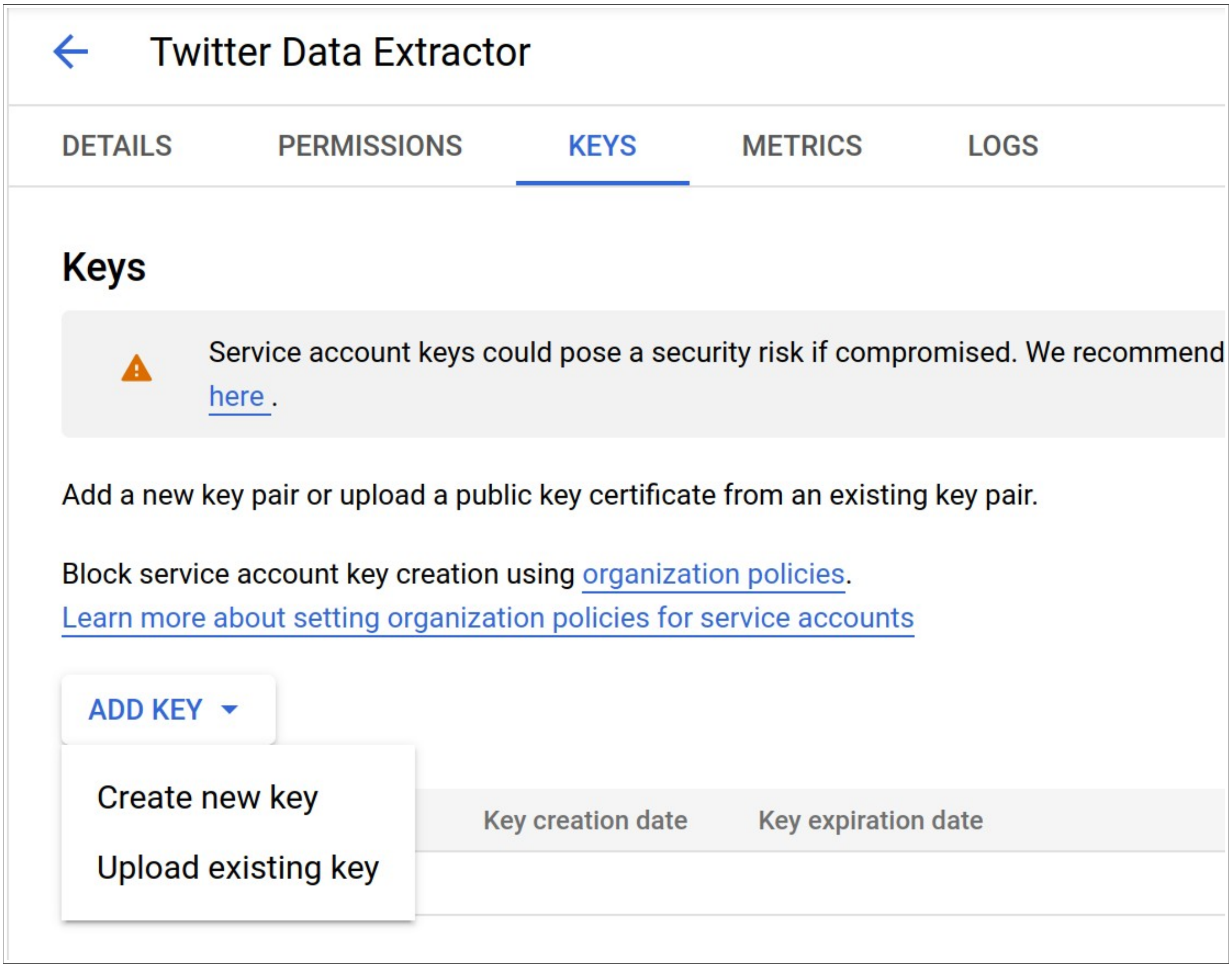
Describe what this service account will do

5. Click on the email link under Service Accounts section.


Service Accounts

□	Email
□	twitter-data-extractor@twitter-data-extractor.iam.gserviceaccount.com

6. Go to KEYS tab and select “Create new key” from “ADD KEY” drop down menu.



Twitter Data Extractor

DETAILS PERMISSIONS KEYS METRICS LOGS

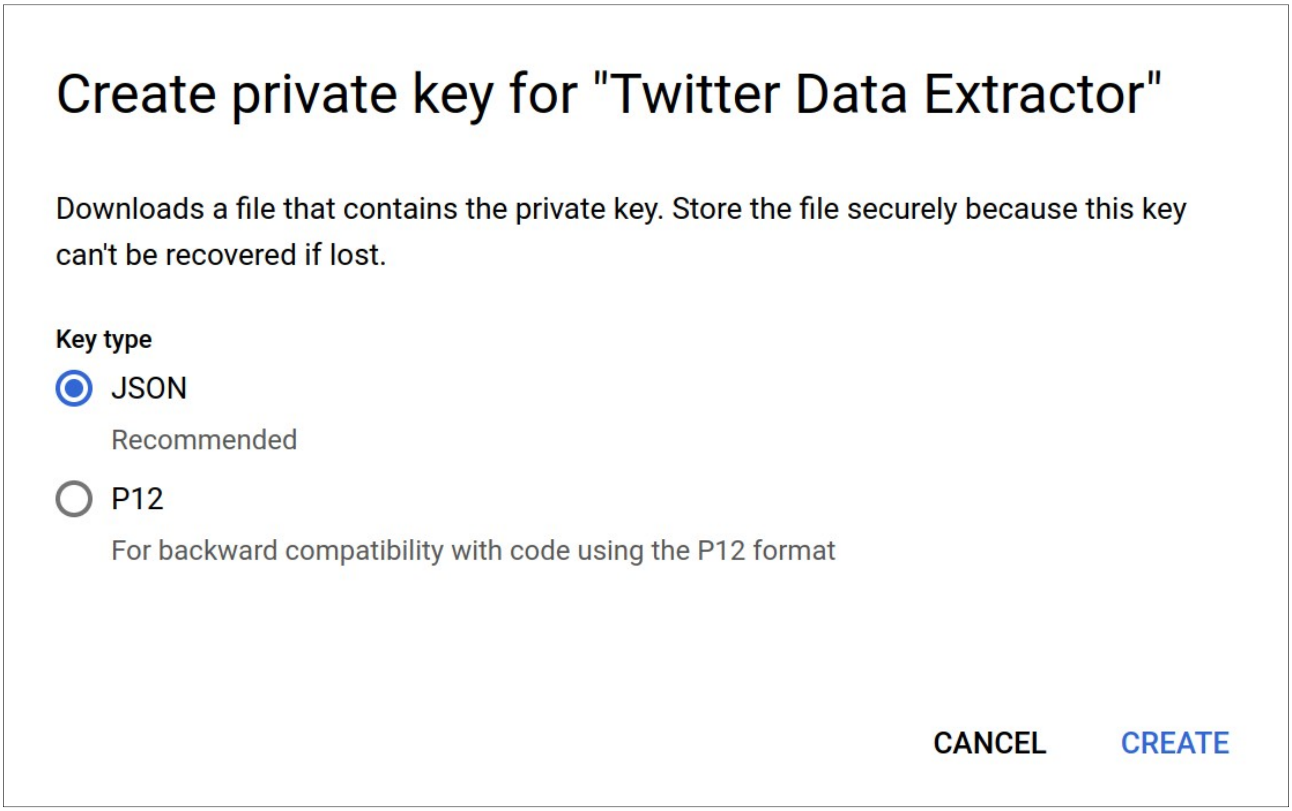
Keys

Service account keys could pose a security risk if compromised. We recommend here .

Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using organization policies.
Learn more about setting organization policies for service accounts

7. Select JSON from the popup menu and click CREATE.



Create private key for "Twitter Data Extractor"

Downloads a file that contains the private key. Store the file securely because this key can't be recovered if lost.

Key type

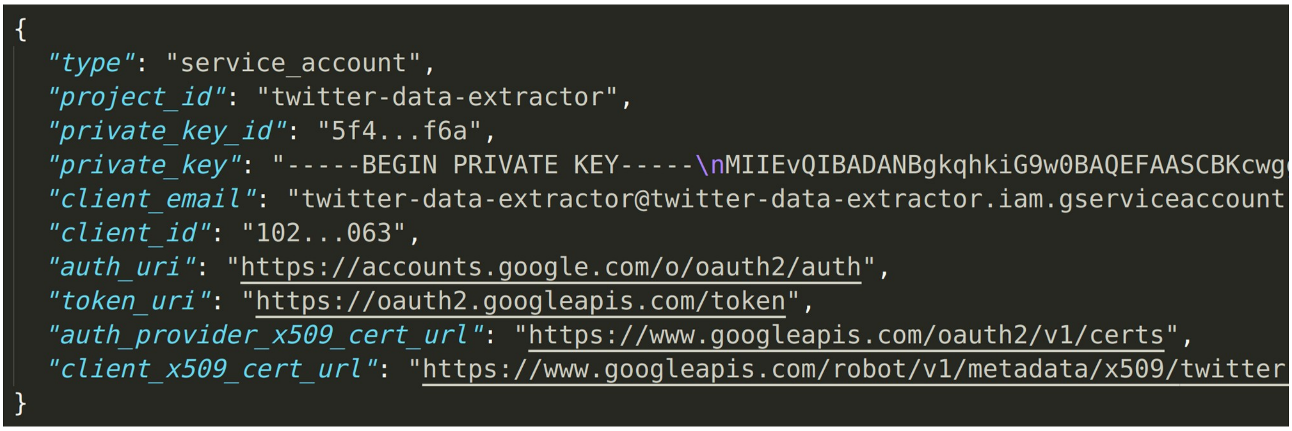
● JSON

Recommended

○ P12

For backward compatibility with code using the P12 format

8. JSON file with credentials will be downloaded automatically. It may look like this



```
{
  "type": "service_account",
  "project_id": "twitter-data-extractor",
  "private_key_id": "5f4...f6a",
  "private_key": "-----BEGIN PRIVATE KEY-----\nMIIEvQIBADANBgkqhkiG9w0BAQEFAASCBKcwg
  "client_email": "twitter-data-extractor@twitter-data-extractor.iam.gserviceaccount
  "client_id": "102...063",
  "auth_uri": "https://accounts.google.com/o/oauth2/auth",
  "token_uri": "https://oauth2.googleapis.com/token",
  "auth_provider_x509_cert_url": "https://www.googleapis.com/oauth2/v1/certs",
  "client_x509_cert_url": "https://www.googleapis.com/robot/v1/metadata/x509/twitter
}
```

4. Rename the downloaded file as “credentials.json” and copy it to the project directory.

MongoDB Installation

If you will use MongoDB to save users/tweets data, install it from here.

You can check the running status after installation and start the database server with the following commands on Linux.

```
sudo service mongod status
```

```
sudo service mongod start
```

How to Use

Basic Usage

The following commands are a few examples of getting user data, user’s friends, tweets or tweets of a given keyword.

```
python twitter_data_extractor.py -u gvanrossum
python twitter_data_extractor.py --forme -ul "gvanrossum,nedbat"
python twitter_data_extractor.py -u gvanrossum -fr
python twitter_data_extractor.py --forme -u gvanrossum -ut
python twitter_data_extractor.py -s python
```

Results are written to *results.xlsx* file by default. Logs can be seen in the *tw_extractor.log* file in the project directory.

Supported Arguments

Short Argument	Long Argument	Description	Default Value
-h	--help	Show help message and exit.	
-c	--useconfig	Read configuration from json file.	False
-cf	--configfile	Read configuration from given json file.	config.json
-	--forme	Determine API user (account owner or on behalf of a user)	False
-u	--user	Extract user data for the given username	
-ul	--users	Extract user data for the given comma separated usernames	
-fr	--friends	Extract friends data for the given username	False
-fl	--followers	Extract followers data for the given username	False
-ut	--user_tweets	Extract tweets of user with the given username	False
-s	--search	Extract latest tweets for the given search keyword	

-tc	--tweet_count	Limit the number of tweets gathered	20
-e	--excludes	Fields to exclude from tweets queried as comma separated values (replies,retweets)	retweets
-ot	--output_type	Output file type (csv, xlsx, gsheets, mongodb or sqlite)	xlsx
-of	--output_file	Output file name	results.xlsx
-sm	--share_mail	Mail address to share Google Sheets document	

TODO: add config file usage explanations

Example Commands

Support

If you need support, you can contact me by emailing to codenineeight@gmail.com with the “twitter_data_extractor” prefix in the subject. You can also see my Upwork profile here.