

# Tutorial Documentation for Automating Data Retrieval from GA Universal – Violet Zewei You

## Introduction

The document is designed to guide users on how to use a Python script for automating the retrieval of landing page session data from Google Analytics Universal. Typically, data is retrieved using council codes from "Audience > Custom > Custom Variables: DatasetID". However, this often necessitates a cross-check with data retrieved using the organization ID from "Behavior > Site Content > Landing Page". Manually retrieving data using organization IDs for each entry is time-consuming and tedious. This Python script automates the process, saving time and reducing errors.

## HOW TO (After you finish everything in Step 2 video)

### 1. Handling N/A Values in Power BI Excel

In the 'Primary Category' sheet of the Power BI Excel file, some cells under a certain column may display 'N/A'.

This 'N/A' indicates either:

- 1) The organization ID is valid and can find data in Google Analytics (GA), but it didn't appear in the retrieved data.
- 2) During the given time period, there were no visits to that organization's page, hence the session count is zero, and it does not appear in GA.

The goal is to ensure that all relevant organization IDs, including those marked as 'N/A', are accounted for in our analysis. This involves adding these organization IDs to Dataset 1's 'Organization ID' column.

Step-by-Step Process:

- 1) Open the 'Primary Category' sheet in the Power BI Excel file.
  - 2) Locate the column with 'N/A' values.
  - 3) For each 'N/A' entry, identify the corresponding organization ID.
  - 4) Open Dataset 1, specifically the 'Organization ID' column.
  - 5) Append the extracted organization IDs from the 'Primary Category' sheet to this column.
  - 6) Ensure that this is done accurately to maintain data integrity.
- ### 2. Running the Python Script for Data Retrieval and Cleaning
- 1) Using Chrome, log in with the account [sacommunity.volunteers@gmail.com](mailto:sacommunity.volunteers@gmail.com).
  - 2) Navigate to My Drive > Colab Notebooks.

The image shows a Google Drive interface. The top section is 'My Drive' with a search bar and navigation options. Below it, a message says 'Change your start page!'. A table lists files in 'My Drive':

Name	Owner	Last modified	File size
Colab Notebooks	me	Feb 7, 2023	-
Karaoke	me	Jan 6, 2023	-
Untitled Document	me	Jan 11, 2023	101 KB
Untitled Document	me	Jul 4, 2022	1 KB

The bottom section is 'Colab Notebooks' with a similar table:

Name	Owner	Last modified	File size
2.py	me	Feb 20, 2024	616 bytes
credentials.json	me	Feb 6, 2024	417 bytes
Data Cleaning Metadata Norwood.ipynb	me	Feb 12, 2024	72 KB
Data CleaningSteps.ipynb	me	Feb 20, 2024	8 KB
Find Council.ipynb	me	Feb 14, 2024	35 KB
token.json	me	10:48 AM	433 bytes
Untitled0.ipynb	me	Feb 14, 2024	10 KB

- 3) Find the Data CleaningSteps.ipynb, credentials.json, token.json, and 2.py files.
- 4) Open Data CleaningSteps.ipynb.
- 5) Download credentials.json and token.json files to your local machine.
- 6) On the left side of the Colab interface, locate the 'Files' section.



- 2) Open your local console or command prompt.
- 3) Navigate to the directory where 2.py is downloaded.
- 4) Run the script by typing `python 2.py` or `python3 2.py` (depending on your Python version).
- 5) This process will initiate the creation of a new token.
  - The script will likely open a browser window or provide a link for authentication.
  - Log in with the appropriate Google account ([sacommunity.volunteers@gmail.com](mailto:sacommunity.volunteers@gmail.com)).
  - Grant any necessary permissions to generate a new token.
  - Upon successful authentication, a new `token.json` file will be created or updated on your local machine.
- 6) Return to the Google Colab environment.
- 7) In the 'Files' section, delete the old `token.json` file.
- 8) Upload the newly generated `token.json` file.
- 9) After uploading the new token, re-run the `Data Cleaning Steps.ipynb` notebook.

### 3. Processing and Updating Data in analytics\_data Excel File

- 1) After the Colab script execution, `analytics_data.xlsx` will be available in the Colab environment.
- 2) Download this file to your local machine.
- 3) Access your Teams platform.
- 4) Navigate to the appropriate file storage area.
- 5) Upload the `analytics_data.xlsx` file for shared access and future reference.
- 6) Open the `analytics_data.xlsx` file.
- 7) Manually review the session links.
- 8) Identify and delete rows that do not correspond to the correct SACommunity landing pages.
- 9) Identify all rows with the same 'Organization ID'.
- 10) Sum up the sessions for these rows and consolidate them into a single row per 'Organization ID'.
- 11) Delete the additional rows once their session counts have been added to the consolidated row. This step ensures each organization ID is represented only once with the total number of sessions.
- 12) Open your Power BI Excel file, specifically the 'Dataset1' sheet.
- 13) Update the session data in 'Dataset1' accordingly.