Using SAS to extract data from the Census Data API

Ahmed Al-Attar

AnA Data Warehousing Consulting, LLC

ahmed.al-attar@anadwc.com

SAS EXPLORE

Using Census Data API with SAS

AGENDA

- Motivation
- Glance at available alternatives
- How is this solution/approach different?
- Demo

MOTIVATION

Over the past few years, the Census Data Application Programming Interface (API) team have been hard at work, aligning themselves with the Digital Government Strategy, and growing their list of publicly available data sets, providing access to key U.S. statistics.

The Census API makes key demographic, socio-economic and housing statistics more accessible than ever before. More data sets are continuously added over time.

Please note: The API does not include any information that could identify an individual; such information is kept strictly confidential by law. The API only uses statistics that the Census Bureau has already released publicly and in aggregate form.

GLANCE AT AVAILABLE ALTERNATIVES

- Videos:
  - Demystifying the Census API
  - Using the API to Get All Results for an ACS Table
  - Accessing and Analyzing U.S. Census Data in R
  - Analyzing 2020 Census data with R and tidycensus
  - 2022 R Webinar Resources: Analyzing Census Data Using R
  - Using the New 2016-20 American Community Survey Estimates in R
  - How to use US Census API with Python for Economics Data | Easy Tutorial
  - Census Data Analysis and Mapping with Python
  - Big Public Data **(using SAS)**
  - Analyzing Census Bureau Data in SAS Studio | A 6 Video Course **(Census Academy Course)**
  - Using American Community Survey Data with Open-Source Software **(WebEx back on 04/13/2022)**

GLANCE AT AVAILABLE ALTERNATIVES

- JMP
  - Census Data Import Example via HTTP Request
    - Importing SAS data file to JMP with the SAS formats - JMP User Community

- Papers:
  - Resources for getting the 2010 US Census Summary Files into SAS®
  - Get your Census Data for Free
  - Data Visualizations using Census Shapefiles, PROC GMAP, SAS/GRAPH Animation, and BISG

HOW IS THIS SOLUTION/APPROACH DIFFERENT?

- **First – My personal analysis of the previous two (2) SAS examples**
  - **Big Public Data**
    - Manual and specific data manipulation/conversion
    - Does not tackle extracting large amount of data columns, bypassing the API limits in a single query!
      - Census Data API User Guide:
        - You can include up to 50 variables in a single API query and can make up to 500 queries per IP address per day. More than 500 queries per IP address per day requires that you register for a Census key. That key will be part of your data request URL string
  - **Analyzing Census Bureau Data in SAS Studio | A 6 Video Course**
    - Using CSV downloads from data.census.gov rather than processing the JSON response from HTTPS query
    - Dives into manual data manipulation/conversion using Excel (using CSV source file)
    - Generates Charts and performs Statistical Analysis against the imported Census data

HOW IS THIS SOLUTION/APPROACH DIFFERENT?

- Second – What makes this approach/solution different and better? 😜
  1. Everything is consolidated into single directory, with its sub-directories.
  2. Out of the box Pre-developed custom SAS programs and SAS Macro programs to streamline data acquisition, and manipulation.
  3. Opinionated approach, yet extendable. It follows the [Census Data API - Four Easy Steps to Building a Request]

```
H:\SASCENSUSAPI
+---code
|  +---macros
|  |  +---censusapi : Census Api specific custom macros
|  |  +---etl : Etl specific custom macros
|  |  \---util : Utility custom macros
|  \---programs
|        step1-Build_DataApiCollection.sas
|        step2-SubmitDataApiQuery.sas
+---config
|     setup.sas : Initialization program
+---data  : Location to store api metadata locally
\---output : Location to store generated SAS output file(s)
   \---graphics
   \---images
   \---data
```

HOW IS THIS SOLUTION/APPROACH DIFFERENT?

- Third – How does it work?

1. Update your settings, **and Submit** the **setup.sas** program ---- One time only every time you start new SAS session

```
%LET g_projRootPath = <Put your value here>;          *<---- Specify installation path. Do not include trailing slash!;
%LET g_apiKey = <Put your value here>;                *<---- Specify your Developer API Key.;
```

- You can use this link (https://api.census.gov/data/key_signup.html) to request a key;

2. Update **step1-Build_DataApiCollection.sas**, if default values are not satisfactory, **and submit it**

```
/* Get the latest list of available tables under the Census Data API */
%censusapi_getalldatasets(p_outLibName=apilib
, p_outDsName=_api_all_data
, p_dataJsonURL=%str(https://api.census.gov/data.json)
, p_reportOutputPath=&g_outputRoot)
```

- Run it periodically to pick updated list of registered tables under the Census Data API.
  - Expect the unique_ds_id to change after each run, as the registered data sets keep changing*
- This program will collect tables listing metadata and generates a SAS Data set and an Excel Workbook output

Using Census Data API with SAS

HOW IS THIS SOLUTION/APPROACH DIFFERENT?

- Third – How does it work?

3. Open the Excel workbook and navigate through the tables listing. Navigate through all four (4) HTML Link columns and examine/go-to the URL listed within it.

4. To illustrate usage example, open the **step2-SubmitDataApiQuery.sas** program **and submit it**.

```
/* Compose a complete Profile of the specified data set */
%GLOBAL g_dsRowId g_dsUniqueId;

SYSECHO "Calling the util_getDsIdValues macro to extract the data set's unique ID variables from the API metadata table ";
%util_getDsIdValues(p_apiBaseUrl=%str(http://api.census.gov/data/2000/dec/sf1) ,p_rtrnRowIDMacVarName=g_dsRowId
,p_rtrnUniqueIDMacVarName=g_dsUniqueId) 1

SYSECHO "Calling the censusapi_getDsFullInfo macro to compose a complete Profile of the specified data set";
%censusapi_getDsFullInfo(p_apiListingLibName=APILIB, p_apiListingDsName=_API_ALL_DATA, p_dsRowId=&g_dsRowId) 2

SYSECHO "Calling the censusapi_submitDataApiQuery macro to submit a Census Data Api query and generate SAS data set out of it";
%censusapi_submitDataApiQuery(p_apiBaseURL=%STR(https://api.census.gov/data/2000/dec/sf1?)
, p_apiGetClause=
%STR(get=P010014,P010015,P010010,P010011,P010012,P010013,P010003,P010004,P010005,P010006,P010001,P010002,P010007,P010008,P010009,NAME)
, p_apiForClause=%bquote(for=zip%20code%20tabulation%20area%20(3%20digit)%20(or%20part):*)
, p_apiInClause=%bquote(in=state:09,23,25,33,44,50,34,36,42,17,18,26,39,55,19,20,27,29,31,38,46)
, p_dsUniqueId=&g_dsUniqueId
, p_outDsName=work.sf1_response
, p_dataApiKey =&g_apiKey
, p_maxVarCount=48) 3
```

SAS EXPLORE

sas

HOW IS THIS SOLUTION/APPROACH DIFFERENT?

- Third – How does it work?

1 Pick desired Dataset Base URL. This could be any of the four (4) HTML Link columns, but without the trailing <type>.html portion. Example:

Group list (HTML): http://api.census.gov/data/timeseries/eits/qfr/groups.html

Base URL: http://api.census.gov/data/timeseries/eits/qfr/

2 This step Must be executed for the desired data set prior to be queried via the Census Data API. This macro call will

- Collect the data set’s information (Variables, Groups, Examples), and creates SAS data sets of that information
- Generates Excel workbook for easy navigation and table filtering functionality
- Generates data set specific SAS Formats, to be used for manipulating the Data API JSON response

3 After opening the data set’s specific Excel workbook and navigating through the Variables, Groups, and Examples worksheets. A sample query from the Examples worksheet was picked and used its separate components/parts as inputs to the enclosed SAS macro call.

Original sample query:

https://api.census.gov/data/2000/dec/sf1?**get=**P010014,P010015,P010010,P010011,P010012,P010013,P010003,P010004,P010005,P010006,P010001,P010002,P010007,P010008,P010009,NAME**&for=**zip%20code%20tabulation%20area%20(3%20digit)%20(or%20part**):*&in=**state:09,23,25,33,44,50,34,36,42,17,18,26,39,55,19,20,27,29,31,38,46

Using Census Data API with SAS

DEMONSTRATION

- **DEMO**
  - How to make your data look worse (than it really is) - Graphically Speaking (sas.com)
  - robslink.com/SAS/book/example19_info.htm
  - Visualize Census Data in SAS Studio | Video 4 of 6

- **Open-Source SAS**
  - SAS® OnDemand for Academics

- **Open-Source SAS code**
  - Github repo

```
DATA _NULL_;
  RC = GITFN_CLONE("https://github.com/ahmedanadwc/sasCensusApi.git",  /* This is the location of the Repository -- Don't Change this */
  "/home/<SASProfile>/sasCensusApiLocalGitRepo"); /* This is the location of your local copy of the Repository -- Change this as you see fit */
RUN;
```

PS. You can run the above code from your **SAS 9.4 M6** and above, on your desired Platform (Windows/Linux) and even SAS Studio, as a replacement to downloading the zip file from www.github.com and then FTP it to your Linux Server.

Using Census Data API with SAS

- Questions?