# 1 Introduction

This program is designed for use with the USDA website (<https://quickstats.nass.usda.gov/>). By default, it will retrieve county-level and state-level area, production, and yield data for a specific crop and merge them into a single file. In case of any inconsistency, such as differing rows between area and production data, the program also generates a file containing separate raw area, production, and yield data. Users have to detect the error themselves and then try manually merging them.

The program is constructed upon the APIs offered by the USDA. However, all API request fields are abstracted into an Excel file - simply enter your desired fields in the Excel spreadsheet to obtain the desired data. If you are seeking more customized queries, please consult the API details at <https://quickstats.nass.usda.gov/api> and modify the code accordingly.

# 2 Environment Setting

## 2.1. Python Interpreter

To run this program, you should first have Python installed on your laptop first.

### 2.1.1 Download Python on Mac

Python 3.9 or later is recommended to run the code. If you have an older version of Python installed, please delete it first.

1. If your macOS is older than Catalina, you might have Python already installed on your Mac. Here’s how to check with Terminal:

Open your Terminal and enter:

python3 --version

If Python is installed, you will see a message indicating Python’s version.

If it’s not installed, you will see “command not found: python”

If Python is not installed, go to the website <https://www.python.org/downloads/>. Click the highlighted button highlighted above to get the Python Installer.

A screenshot of a computer

Description automatically generated

2. After installing, please close and reopen your terminal.

**Verify**: We can use the terminal to check if your Python interpreter was installed correctly. Try the following command:

python3

If the installation worked, you should see some text printed out about the interpreter followed by >>> on the same line.

Note: If the ***python3*** command doesn't work, try using ***python*** or ***py.***

### 2.1.2 Download Python on Windows

Python 3.9 or later is recommended to run the code. If you have an older version of Python installed, please delete it first.

1. Go to the website <https://www.python.org/downloads/> to download the latest Python. Click the button highlighted to get the Python Installer.

A screenshot of a computer

Description automatically generated

2. Open the installer, just click “Install Now”. Please make sure you check the checkbox “**Add Python to PATH**”, and don’t change the download path, it will save you the hassle of setting the environment path manually.

A screenshot of a computer

Description automatically generated

3. After installing, please close and reopen your terminal. You can now check your Python version with the following command in the command prompt.

python3 --version

**Verify**: We can use the terminal to check if your Python interpreter was installed correctly. Try the following command:

python3

If the installation worked, you should see some text printed out about the interpreter followed by ***>>>*** on the same line. This is where you can type in Python code.

Note: If the ***python3*** command doesn't work, try using ***python*** or ***py.*** If none of these commands works, there might be something wrong with your environment path. Refer to this video to solve your problem. <https://www.youtube.com/watch?v=brX8x-qkACs>

## 2.2 Visual Studio Code (VS Code)

The Python interpreter you just installed allows you to run Python code. You will also need a text editor, where you can write Python code.

### 2.2.1 Download VSCode for MacOS

Open this website <https://code.visualstudio.com/docs/setup/mac> and follow the instructions to download VSCode for Mac

### 2.2.2 Download VSCode for Windows

Open this website <https://code.visualstudio.com/docs/setup/windows> and follow the instructions to download VSCode for Windows.

## 2.3 Python Extension

VSCode is nothing but a fancier notepad, we need a communication bridge between the VSCode and the Python Interpreter we just installed.

1. Click the extension icon on the left sidebar,

2. Enter “python” in the search box

3. install the first one (labeled Microsoft)

4. After installing, a run icon will appear in the upper right corner when you open a python file (.py file)

A screenshot of a computer program

Description automatically generated

## 2.4 Libraries

You need to download all of the following libraries to ensure the code can run successfully.

1. First off, make sure pip is installed on your MacOS. Open your terminal, enter:

pip -V

Pay attention, ***V is capitalized here***. If a message indicating the pip version shows up, it means the pip is already installed. Otherwise, enter the command below to install the pip

python3 -m ensurepip

2. After the pip is installed, you need to download the following libraries by entering the commands below.

Requests:

pip install requests

Numpy:

pip install numpy

Pandas:

pip install pandas

# 3. How to use the program

## 3.1 Input

On the website, you can see a list of filters.

A screenshot of a computer

Description automatically generated

Now, open the “input.xlsx” Excel file in the code folder, you can see the things below:

A screenshot of a computer

Description automatically generated

The fields in the Excel files correspond to the fields on the website.

**In the Excel file, there are no “Geographic Level” and “Category” fields** since they are of no use. First, the program will automatically download both state and county-level data. Second, the “Data Item” is more specific than the “Category” field.

### 3.1.1 Program

It corresponds to the program filter on the website, it can only be either “CENSUS” or “SURVEY.”

### 3.1.2 Sector

It corresponds to the sector filter on the website.

### 3.1.3 Group

It corresponds to the group filter on the website.

### 3.1.4 Commodity

It corresponds to the commodity filter on the website, in other words, the crop name.

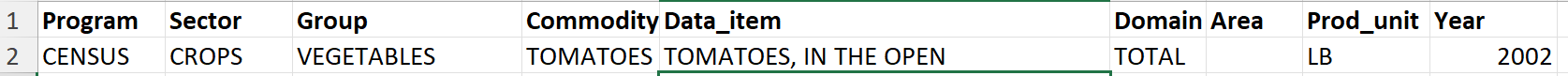
### 3.1.5 Data\_item

The Data\_item field is the most important in the Excel file. This will determine what data you will download in the end.

A screenshot of a computer

Description automatically generated

For example, you want to download “TOMATOES, IN THE OPEN - ACRES HARVESTED” as area data and “TOMATOES, IN THE OPEN – PRODUCTION, MEASURED IN LB” as production data. You need to enter “TOMATOES, IN THE OPEN” in the Excel “Data\_item” column, as shown below.



In other words, you need to copy the text before the hyphen in the Data item entry on the website and paste it into the Excel data\_item field. More examples:

|  |  |
| --- | --- |
| **Data item entry on the website** | **The thing you need to enter in the Data\_item in Excel** |
| TOMATOES, IN THE OPEN –  ACRES HARVESTED | TOMATOES, IN THE OPEN |
| SUGARCANE, SUGAR –  ACRES HARVESTED | SUGARCANE, SUGAR |
| BARLEY – ACRES HARVESTED | BARLEY |
| COTTON, PIMA, IRRIGATED –  ACRES HARVESTED | COTTON, PIMA, IRRIGATED |
| CORN, GRAIN, NON-IRRIGATED –  ACRES HARVESTED | CORN, GRAIN, NON-IRRIGATED |

Note:

(1) In this Data\_item field, you can specify whether you prefer irrigated, non-irrigated, or total data. The website offers both irrigated and non-irrigated data for most crops, accessible by selecting the appropriate data item field, as demonstrated by the last two rows in the table above.

(2) The program can only merge data with matching data item fields for both area and production data, such as "TOMATOES, IN THE OPEN - ACRES HARVESTED" and "TOMATOES, IN THE OPEN – PRODUCTION, MEASURED IN LB". If there is no production data corresponding to the area data's data item field, only the area data will be presented in the final result.

### 3.1.6 Domain

Normally we choose “TOTAL”.

### 3.1.7 Prod\_unit

A crop sometimes has more than one production unit. For example, corn has BU and LB as its production units. You need to specify the one you want. Sample inputs are ***“LB”, “BU”, “CWT”,*** etc.

A black text on a white background

Description automatically generated

You have to go to the website to examine the Data Item filter to see what production units the crop you want to select has. If you enter an incorrect production unit, no production data will be returned and the program will possibly crash.

## 3.2 Execution

After finishing the input Excel file, it is time to run the program

1. Open the VScode, click the “File” in the upper left corner -> click “Open the folder” and select the root folder containing the code. (Please don’t use “Open the file” since the program might have problems detecting the “input.xlsx”)

2. Select the “API.py” file, and click the run icon in the upper right corner.

Now, you are supposed to see a terminal shows up in the VSCode, and “Program starts” is printed on the terminal. During the code execution, you will also see some prompt if some data you are requesting doesn’t exist. When the program finishes, “Program ends” will be printed.

A screenshot of a computer program

Description automatically generated

## 3.3 Output

### 3.3.1 Where is the processed file?

When the program finishes, two folders (CENSUS and Survey) will be generated automatically within the directory where the code is located, below is the directory structure:

- Root

|- CENSUS

| |- State\_Level

| |- County\_Level

|- SURVEY

| |- State\_Level

| |- County\_Level

|- api.py

|- service.py

|- infrastructure.py

|- input.xlsx

|- README.docx

Final csv files will be in the “CENSUS/State\_Level”, “CENSUS/State\_County”, “SURVEY/State\_County”, and “SURVEY/State\_County” depending on your input.

### 3.3.2 Naming

The naming pattern is

<geographic level>\_<year>\_<crop\_name>\_<irr/nonirr>.csv

such as “State\_2022\_TOMOTOES”. If you select the irrigated or non-irrigated data item field, the name of the final result will be “State\_2022\_TOMOTOES-irr.csv” or “State\_2022\_TOMOTOES-nonirr.csv”.

## 3.4 Error Handling

If no error happens, there shouldn’t be any Excel file in the root directory. Otherwise, It means some inconsistency happened, such as differing rows between area and production data,

Open the file, it contains separate raw area, production, and yield data. you have to manually figure out what the problem is and merge them.