# Introduction

This program is for the India area data <https://agcensus.da.gov.in/> (previously <https://agcensus.dacnet.nic.in/StateCharacteristic.aspx>).

It is just a straightforward crawler to help you download the files.

# 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 laptop. 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

Selenium:

pip install selenium

Retry:

pip install retry

## 2.5 Chrome Driver

Now, we need to let your program control your chrome. To achieve this, we need to download a chrome driver.

1. Check your Chrome browser version:

Open Chrome -> click 3 dot menu -> settings -> About Chrome. Now you can get your Chrome version number.

2. Download Chrome Driver

If your Chrome version is greater than 114, go to <https://googlechromelabs.github.io/chrome-for-testing/>. You will see a table as shown below. If the leading digit of your Chrome version matches the leading digit in the "Stable" row, click "Stable" to download the Chrome driver.

For instance, if your Chrome version is 122.0.5615.121 and the driver version displayed is 122.0.6261.94, they are compatible. The subsequent digits don’t matter.

If they do not match, ensure your Chrome is updated before proceeding.A green and blue table with black text

Description automatically generated

Now, we can move the driver to the default environment path. You can check them with the below command and normally the path returned is **/usr/local/bin**

sudo nano /etc/paths

Now, go to where the chrome driver is downloaded, find the chromedriver\_mac64.zip file, and unpack it. You will see a Chromedriver executable file. Now, we need to move the Chrome driver to the environment path, you can do it by using the following command

mv chromedriver /usr/local/bin

# 3. How to use the program

## 3.1 Input

Open the “input.xlsx” file, you will see things below

A screenshot of a computer

Description automatically generated

You need to enter the “Year”, “State” and “Crop”. District and Tehsil can be leave blank. If they are blank, the program will automatically download all districts and tehsils.

## 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 “main.py” file and click the run icon in the upper right corner.

3. **After the downloading finishes, open the “check\_name.py” and click the run icon in the upper right corner. This will rename the files with incorrect names due to the Chrome driver glitches during the downloading**.

## 3.3 Output

### 3.3.1 Where is the downloaded file?

When the program starts, a folder called CSV will be generated automatically within the directory where the code is located, all the files downloaded will go there.

### 3.3.2 Naming

The naming pattern is

<year>\_<state\_name>\_<district\_name>\_<tehsil\_name>\_<crop\_name>.csv

If tehsil\_name is “all”, it means this file is a district-level file.

If both district\_name and name are “all”, it means this file is a state-level file

## 3.4 Error Handling

When an error happens, an err.txt will be generated and the failed downloading will be recorded. You can re-download the corresponding data later.