Internship Report

Data Pipeline for Geographic Information of Four Crop Types and Data Pipeline for Geographic Information of Perennial Plants (Palm and Rubber)

Presented to

Sirikul Hutasavi

By

Pongphat Mitrathanun

Jatenipat Ausanarassamee

This report is part of the internship at the

Geo-Informatics Application and Management Department of Agriculture Economics and Geo-Social

Geo-Informatics and Space Technology Development Agency (Public Organization)

June - July 2024

Technologies Used

1. Visual Studio Code

Visual Studio Code is a popular text editor developed by Microsoft, widely used by programmers and coders. VS Code is highly versatile and flexible, making it suitable for developing a wide range of applications and programs.

2. Python Programming

Python is a programming language known for its easy-to-read and easy-to-write syntax. It supports multiple platforms, has a comprehensive standard library, and a large user community. Python supports object-oriented programming and can be integrated with other languages effectively. It is commonly used for web development, data analysis, artificial intelligence development, and many other applications.

3. QGIS

QGIS (Quantum GIS) is a free and open-source geographic information system software. It is continuously developed and includes various plugins to support specific functionalities.

Libraries Used

1. GeoPandas

GeoPandas is a Python library designed specifically for working with geospatial data. It extends the capabilities of pandas, which is used for managing tabular data, to handle geospatial data as well. GeoPandas allows reading, writing, and analyzing geospatial data, including managing geometric shapes such as points, lines, and polygons.

Installation of GeoPandas

Install GeoPandas using the command “pip install geopandas” then using the command
“import geopandas as gpd”

2. Pandas

Pandas is a Python library developed for managing and analyzing tabular data. It has two main data structures, Series and DataFrame, which are efficient for handling large and complex data.

Installation of Pandas:

Install Pandas using the command “pip install pandas” then using the command
“import pandas as pd”

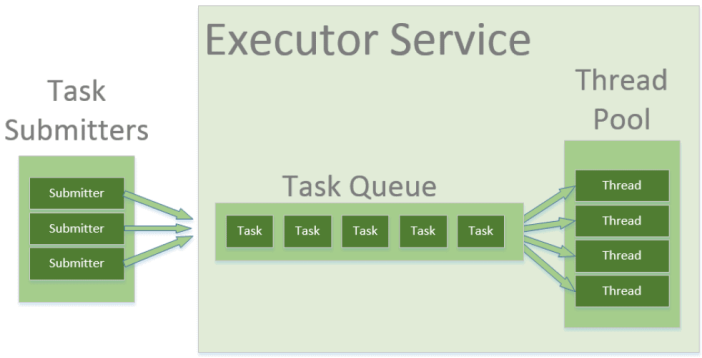
3. TQDM

TQDM, derived from the Arabic word "taqadum" meaning progress, is used to display progress bars in the terminal of the editor being used.

Installation of TQDM:

Install tqdm using the command “pip install tqdm” then using the command “from tqdm import tqdm”

4. ThreadPoolExecutor and as_completed from concurrent.futures



ThreadPoolExecutor and as_completed from concurrent.futures

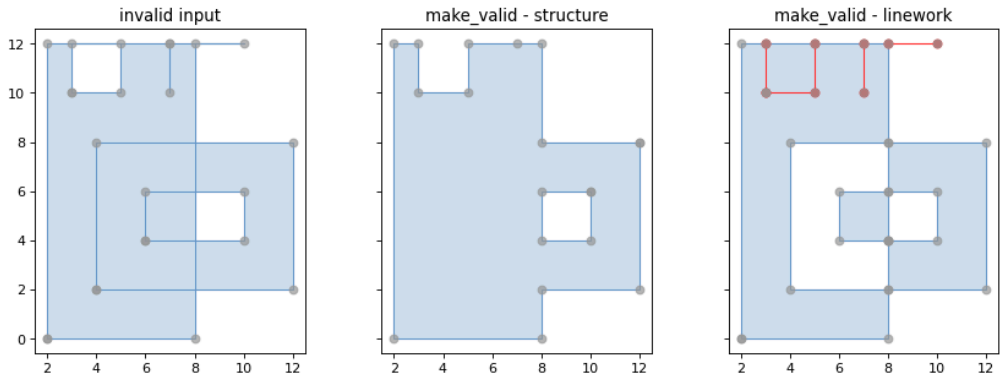
ThreadPoolExecutor is a class in the concurrent.futures module in Python used for managing parallel execution of tasks. It controls the creation and usage of multiple threads to process multiple tasks simultaneously, enhancing data processing efficiency and reducing processing time. It is suitable for data processing, I/O operations, or computations requiring significant time.

as_completed is a built-in function in the concurrent.futures module for managing parallel execution tasks. It returns an iterator that yields completed tasks as they finish, not in the order they were submitted. This function allows handling results as soon as they are ready without waiting for all tasks to complete.

Installation of ThreadPoolExecutor and as_completed from concurrent.futures

Install ThreadPoolExecutor and as_completed from concurrent.futures using the command “from concurrent.futures import ThreadPoolExecutor, as_completed
from shapely.validation import make_valid”

5. make_valid from shapely.validation

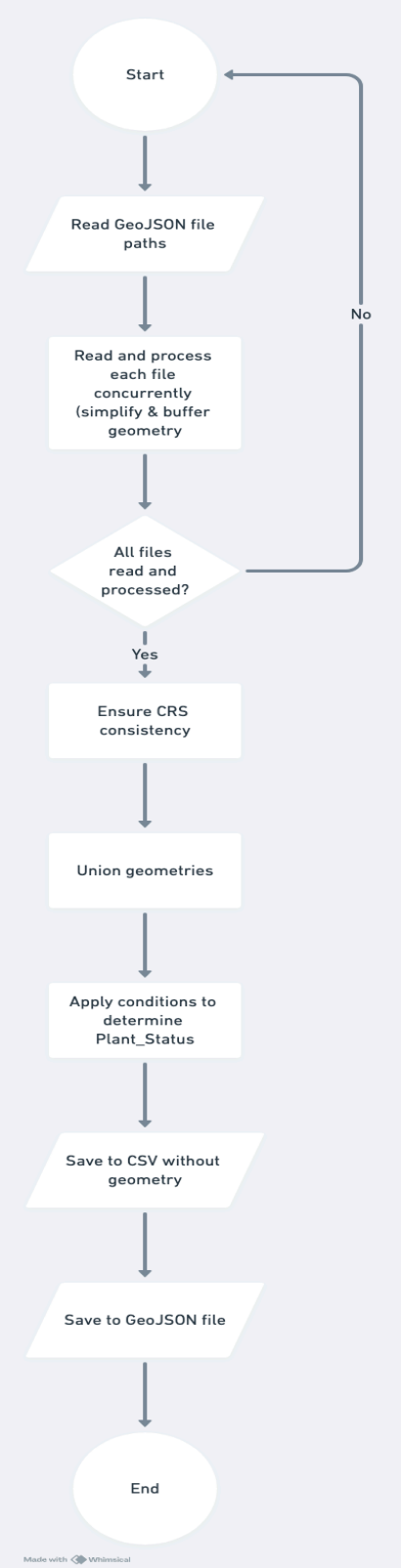


make_valid from the shapely.validation library is used to correct potentially invalid geometric objects, making them valid according to geometric processing rules. This can resolve issues arising from geometries that do not conform to geometric rules, such as overlaying geographic data.

Installation of shapely

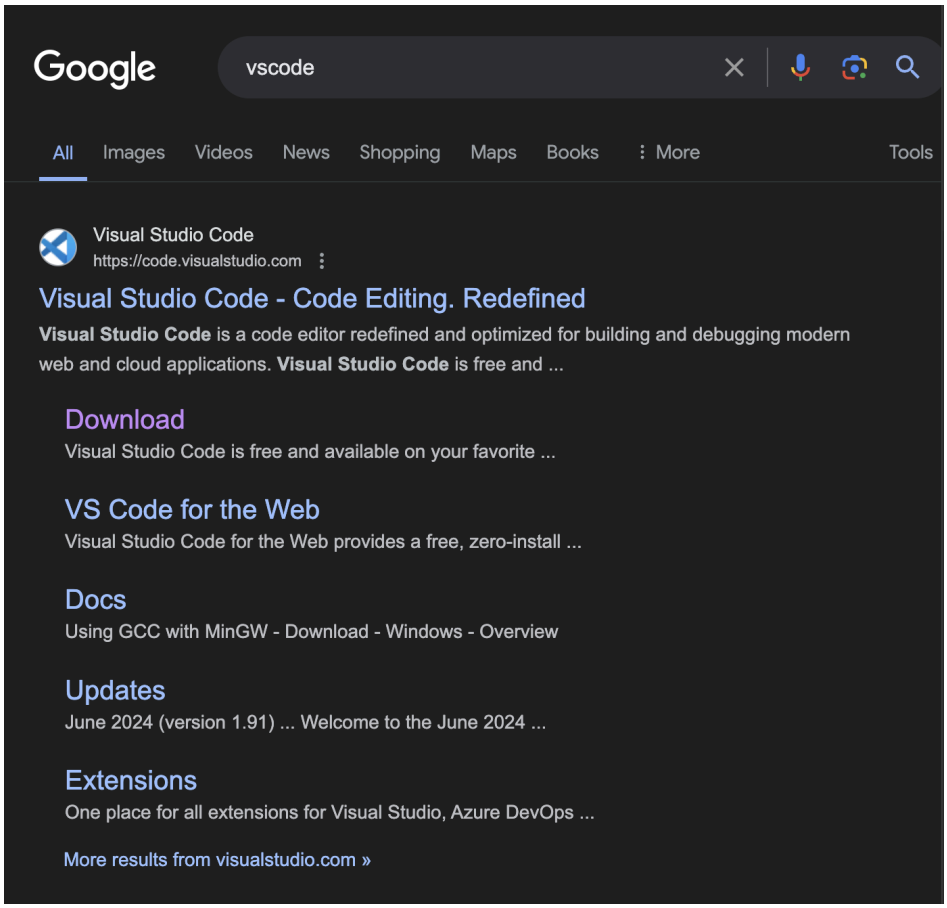
Install make_valid from the shapely.validation using the command “pip install shapely” then using the command “from shapely.validation import make_valid”

Project Workflow



How to use the program

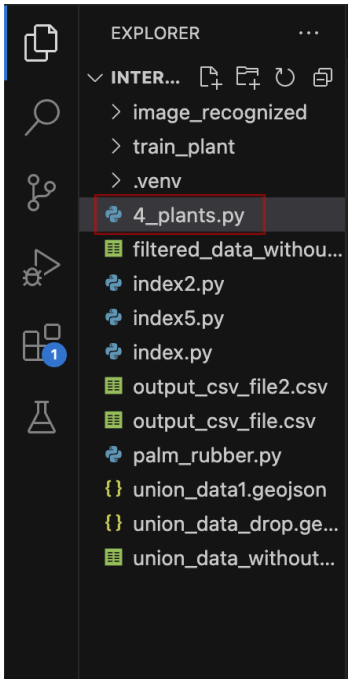
Step 1: Install Vscode



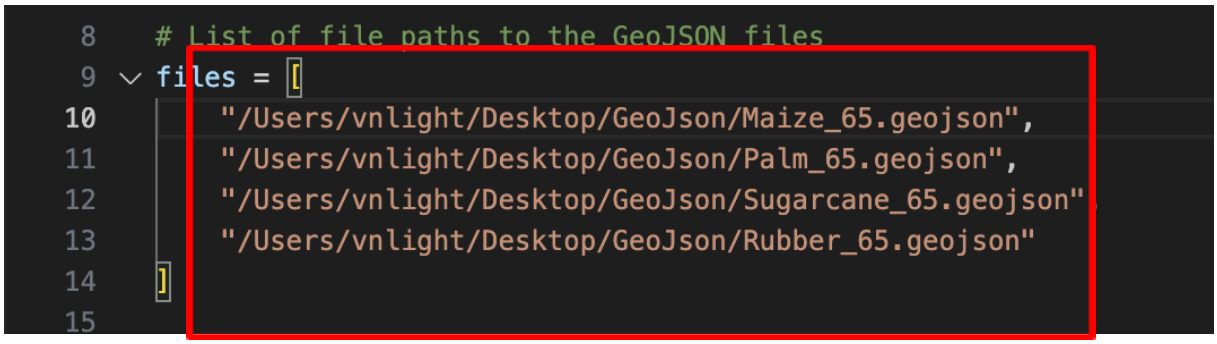
https://code.visualstudio.com/download

Step 2: Open Vscode and Install extension

Step 3: For the Four crop type , Find the file name “4_plants.py”



Step 4: Replace file path with the file that you want.

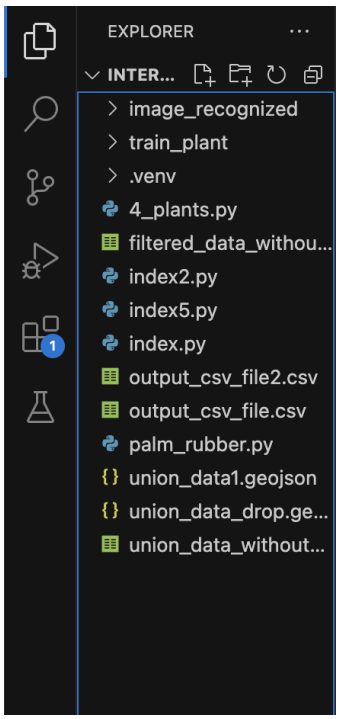


Step 5: You can change the name of the csv file and geojson file by changing the output_csv_file and output_geojson_file to the file name that you want.

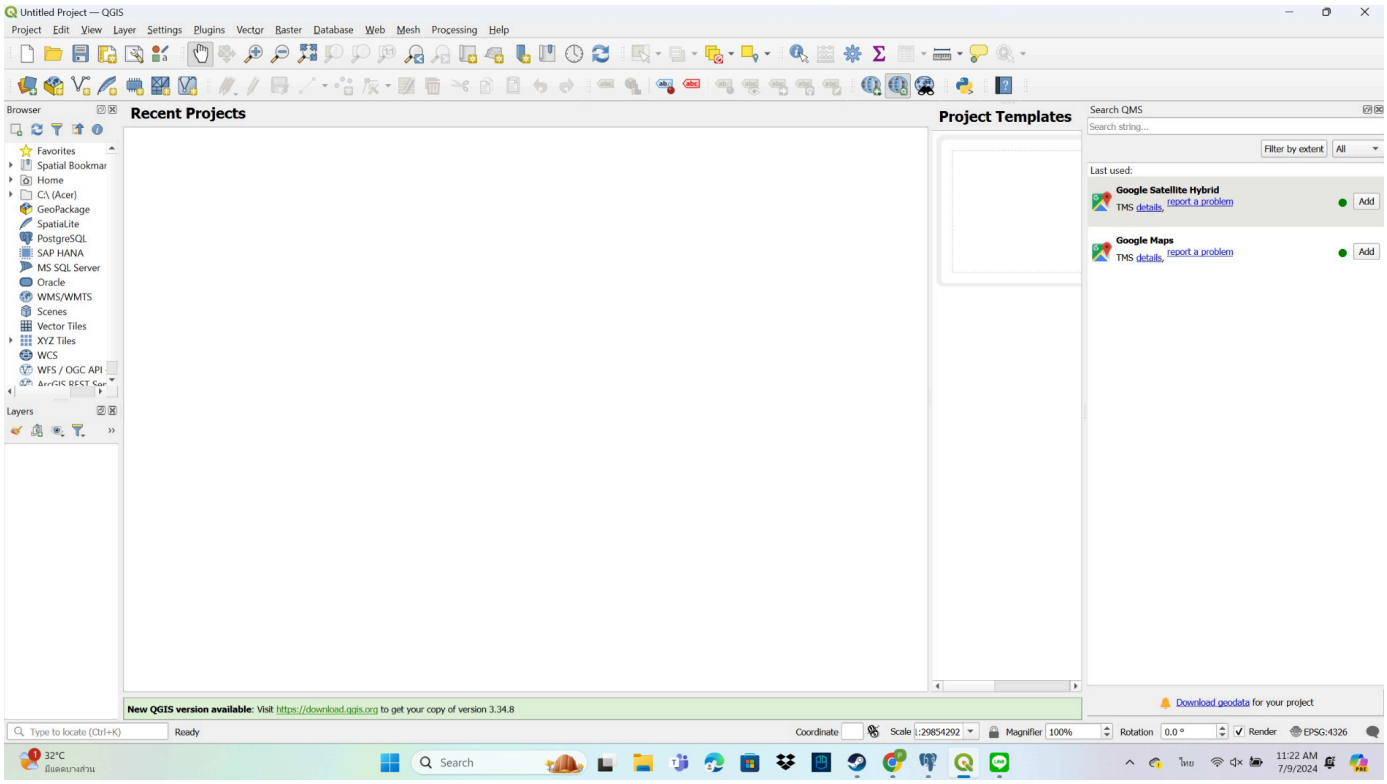
```python
# Save the GeoDataFrame to a CSV file without the geometry column
output_csv_file = 'union_data_without_geometry_4_plant.csv'
gdf_union.drop(columns=['geometry']).to_csv(output_csv_file, index=False)
print(f"Union data without geometry saved to {output_csv_file}")


# Save to geojson
geojson_data = gdf_union.to_json()
output_geojson_file = 'union_data_4_plant.geojson'
with open(output_geojson_file, 'w') as f:
    f.write(geojson_data)
print(f"Union data saved to {output_geojson_file}")

```

Step 6: After you change all things that are important to run. You can press on run button to run this code (Please ensure that you already have installed all necessary libraries)

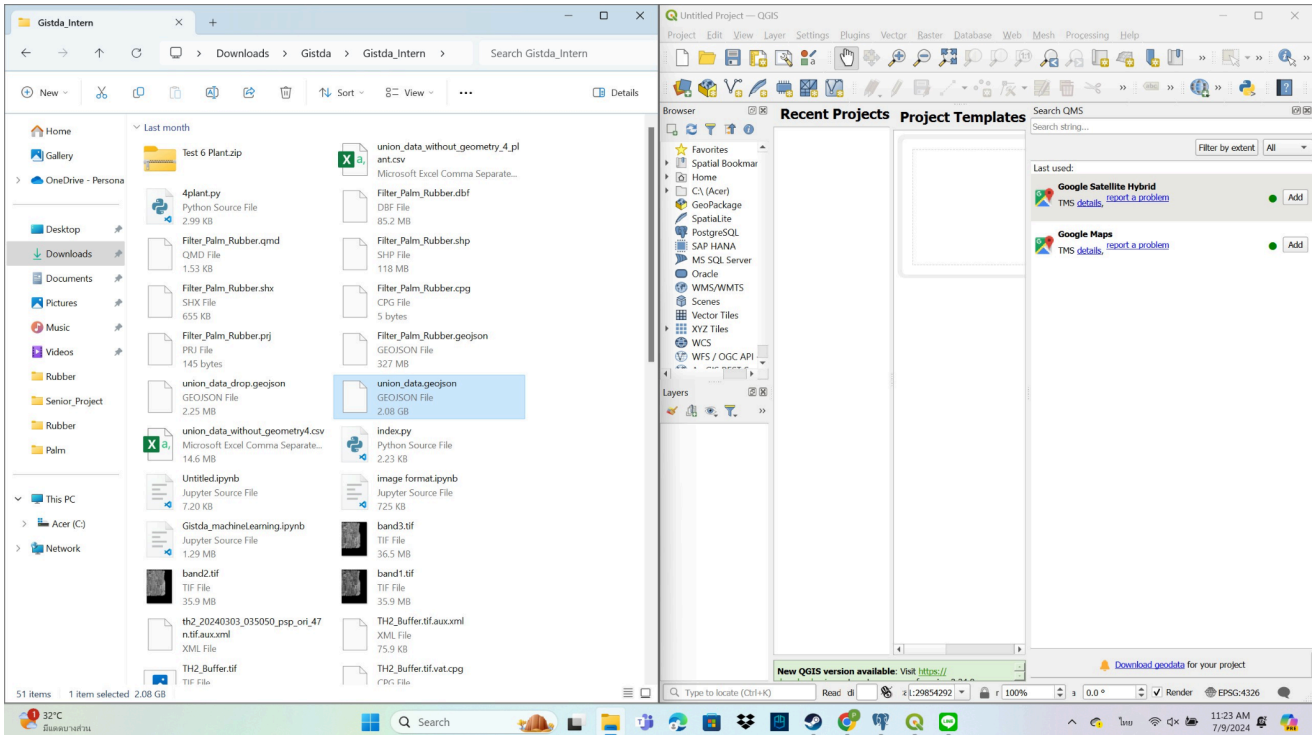
Step 7: You can see the output on the left side of the screen



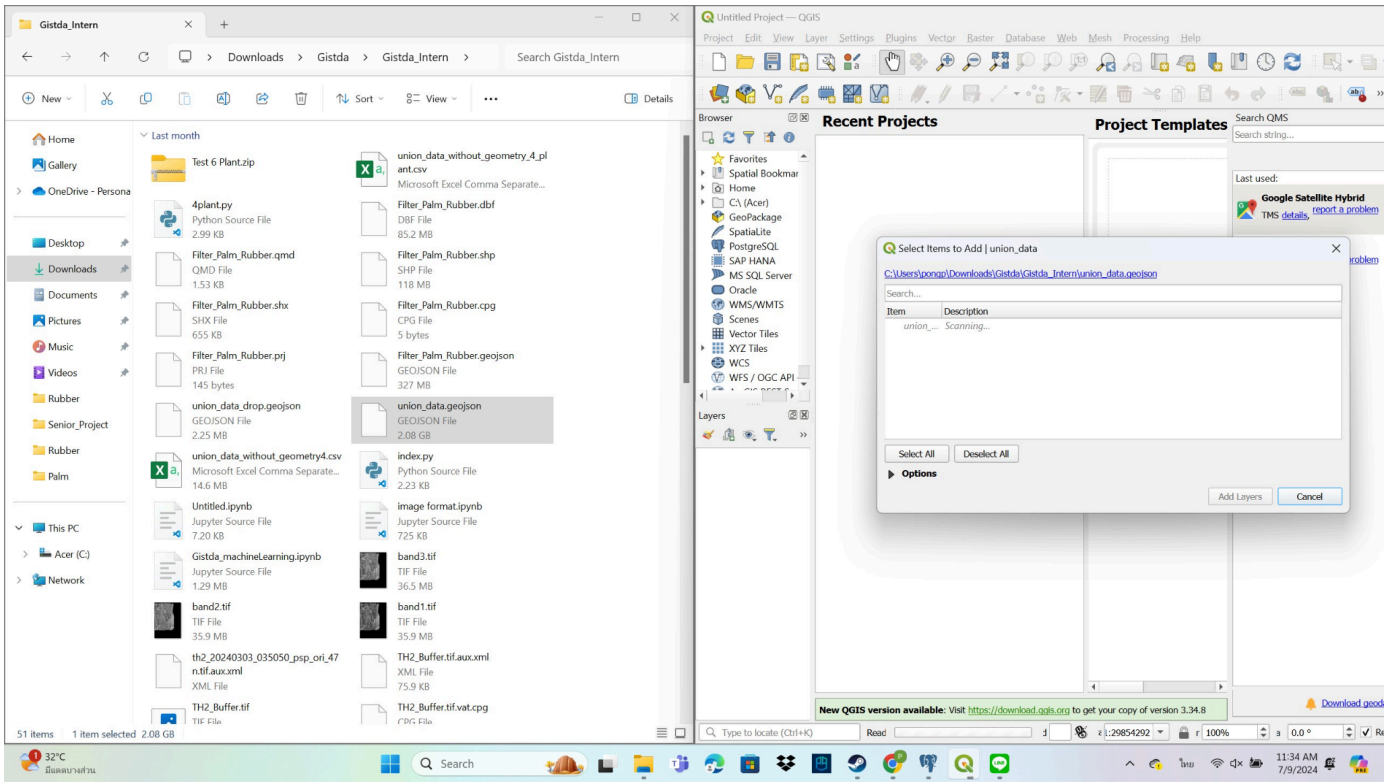
Step 8: Go to QGIS



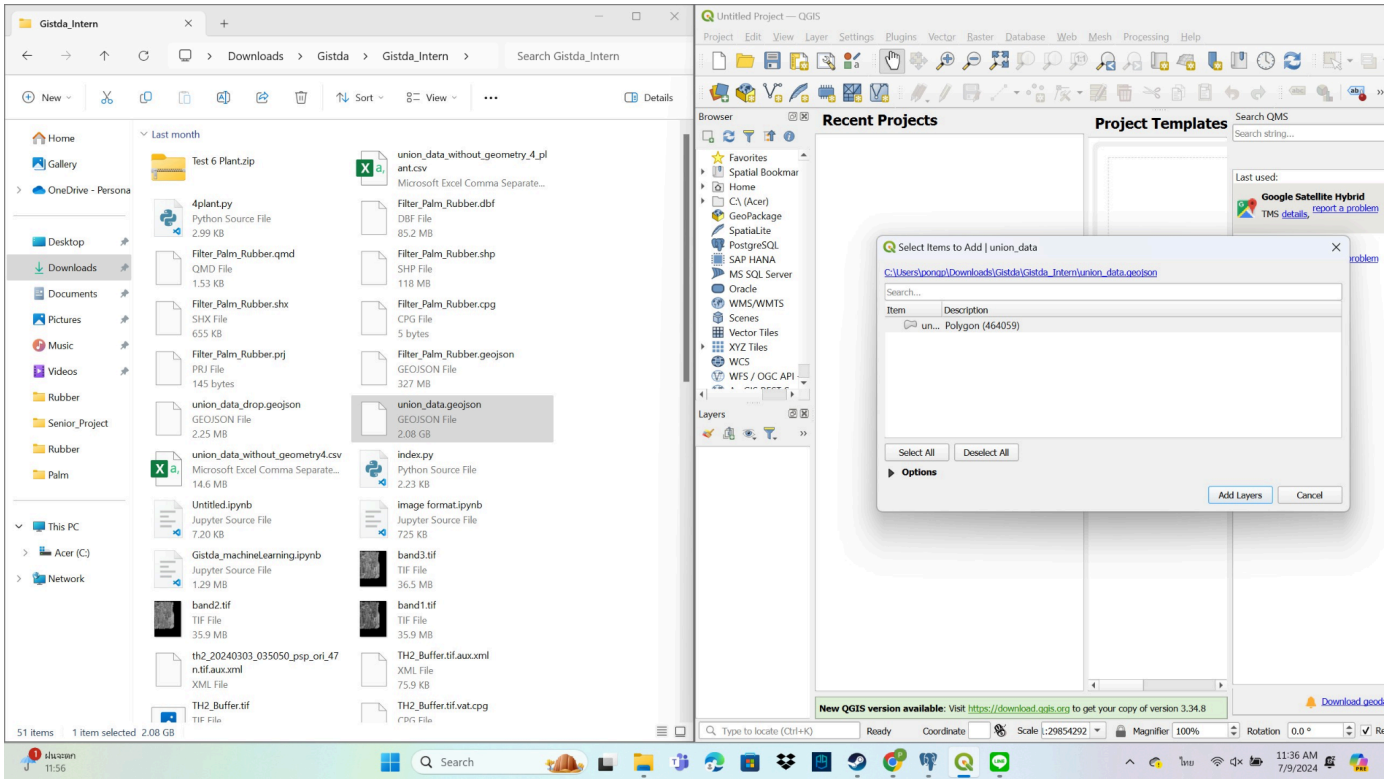
Step 9: Drag and drop the geojson file that you want to use to QGIS. (In this case, I assume to use union_data.geojson file



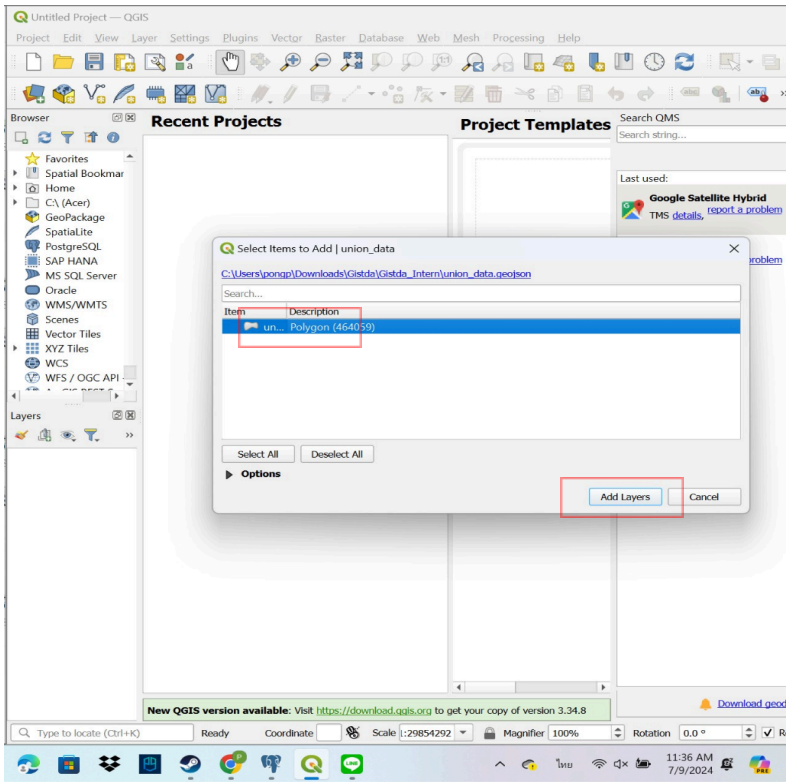
Step 10: Wait for the file to download



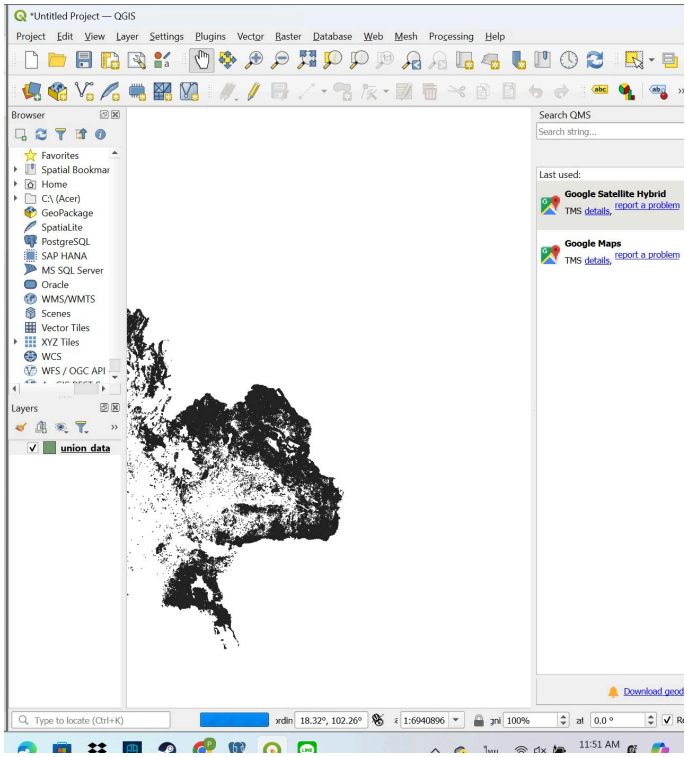
Step 11: After download, File will appear.



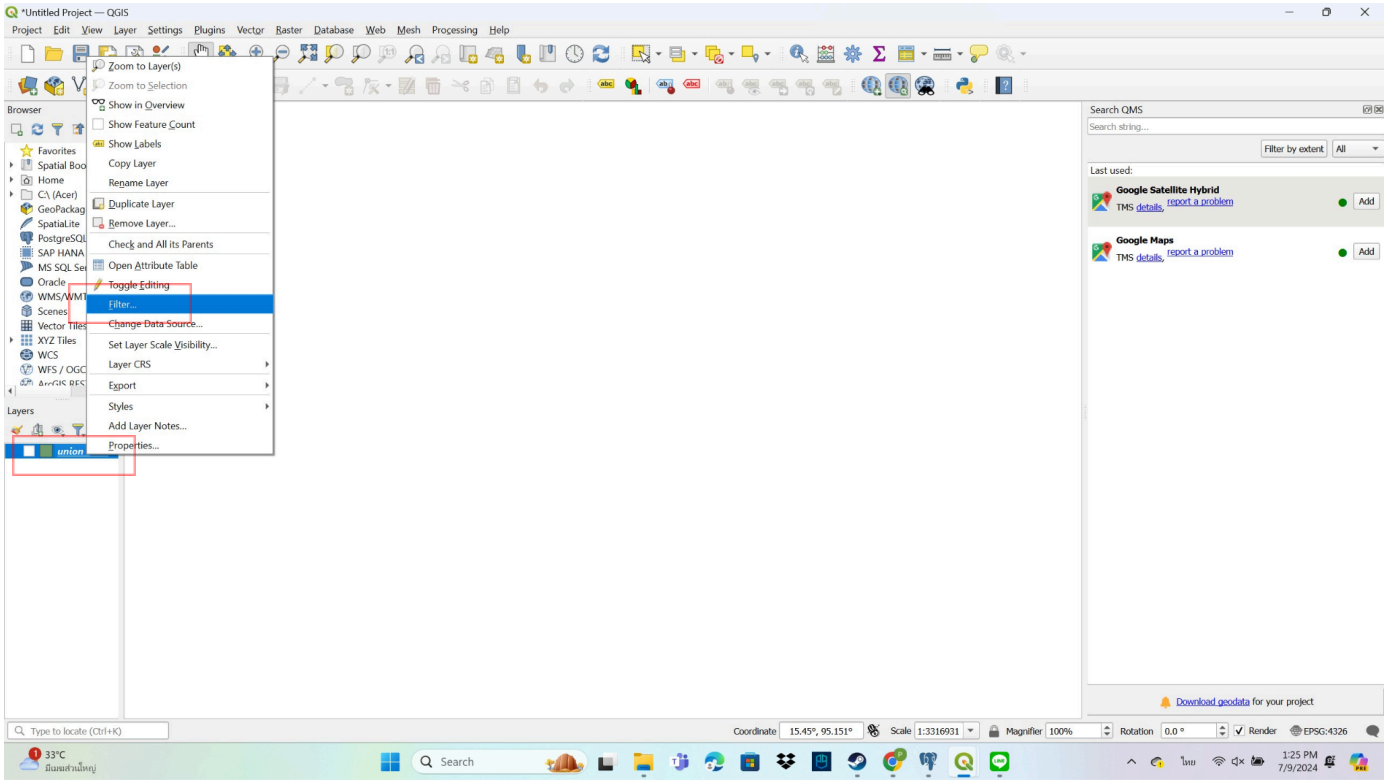
Step 12: Click to the file and Click on the Add layers button.



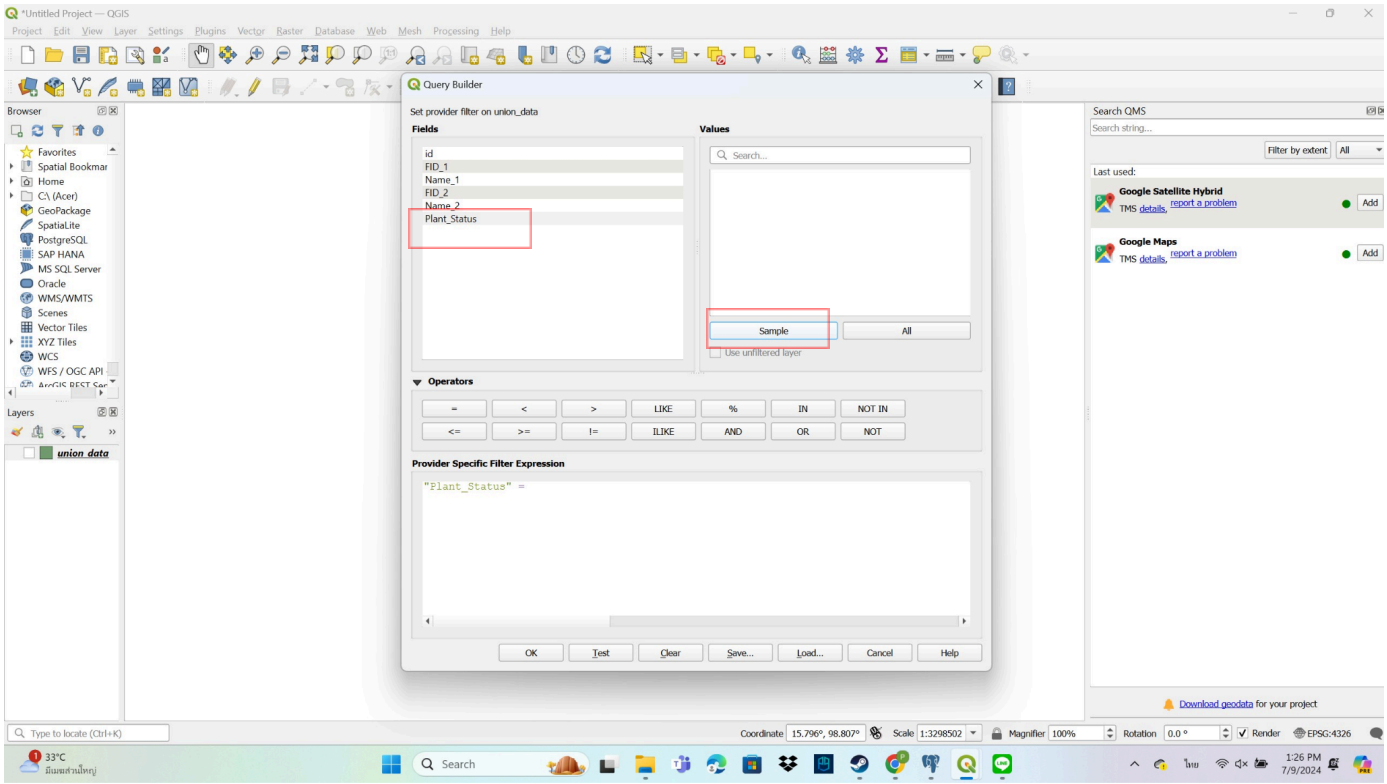
Step 13: It will display an image as shown below



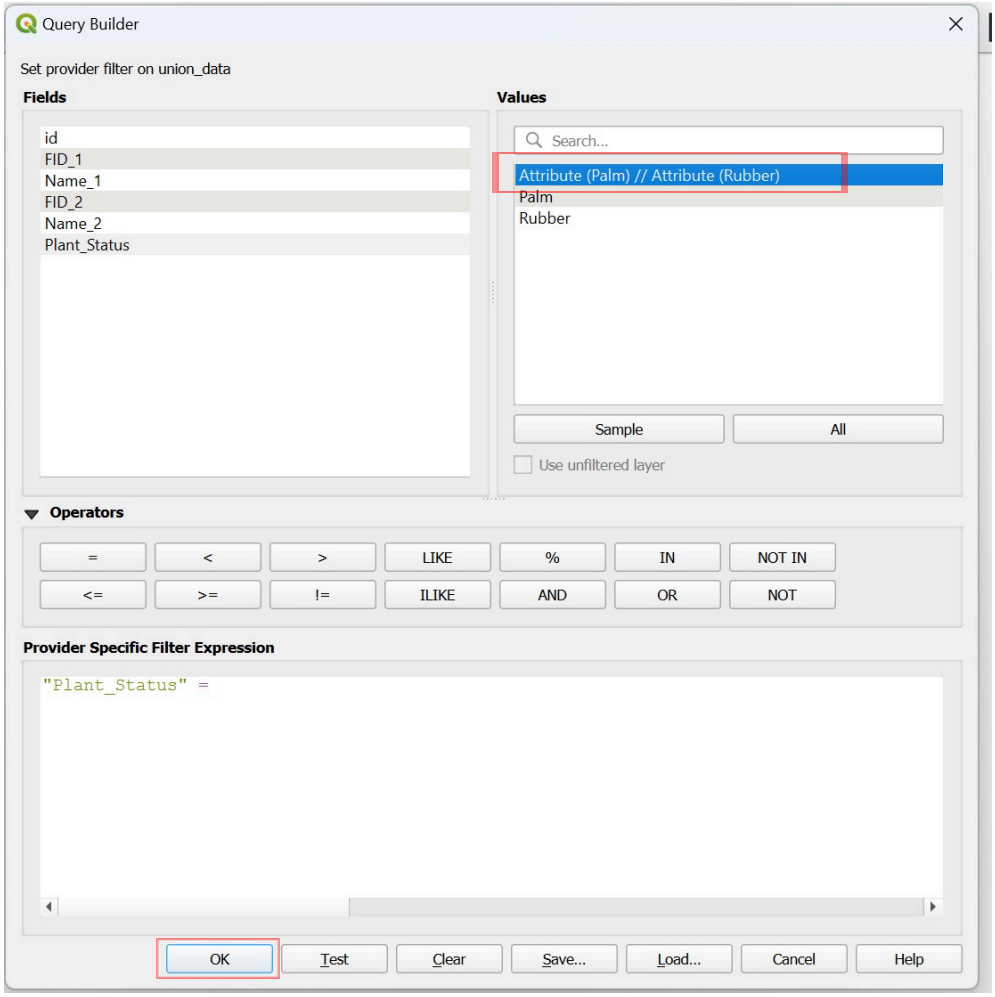
Step 14: Click to the union_data and click to the filter



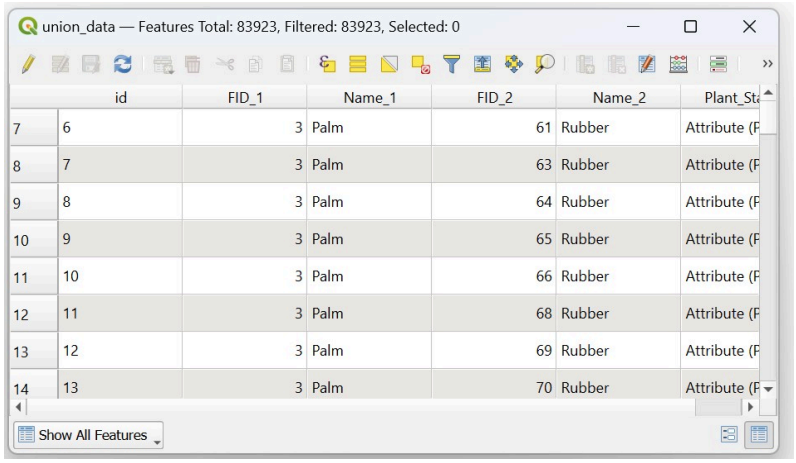
Step 15: Click to the Plant_Status then Click to the Sample button



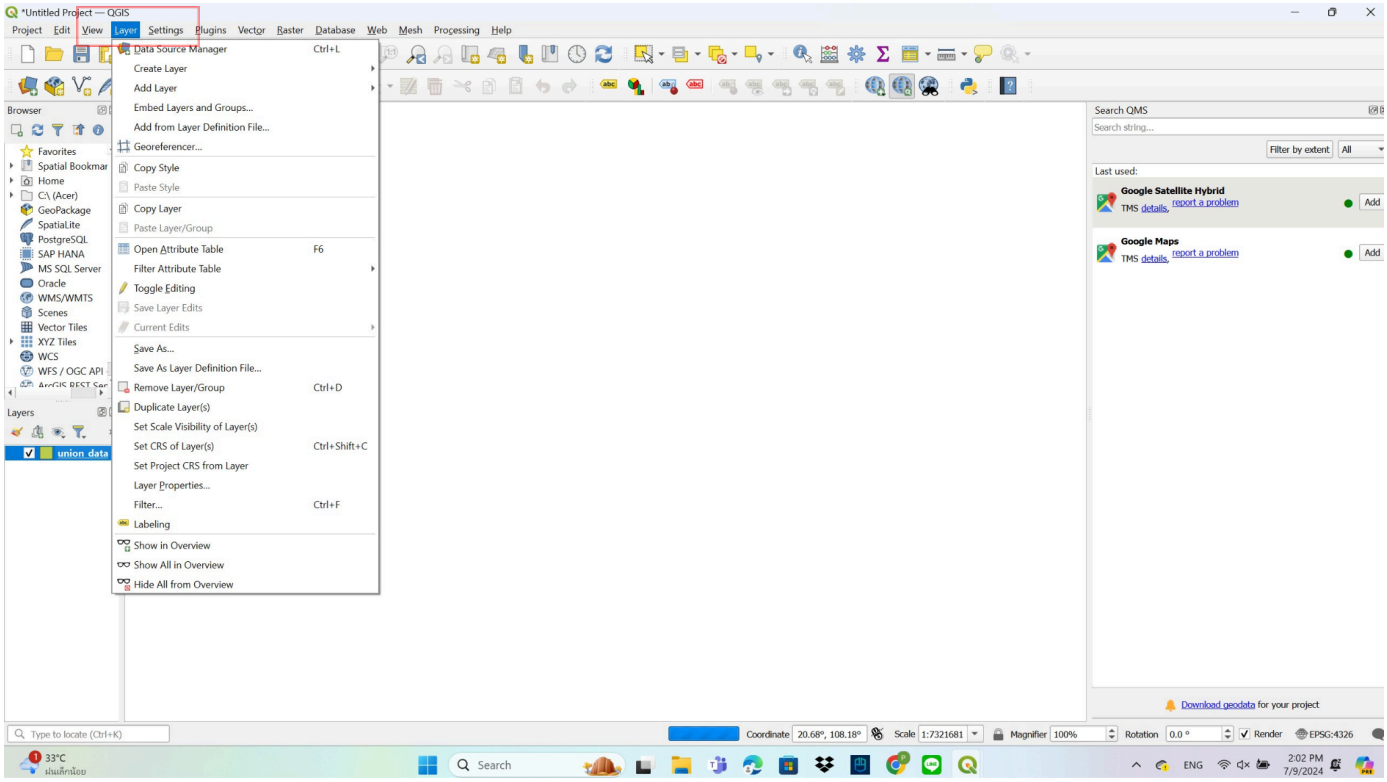
Step 16: Then, it will display the attribute as shown below. Click to the Attribute and Click OK.



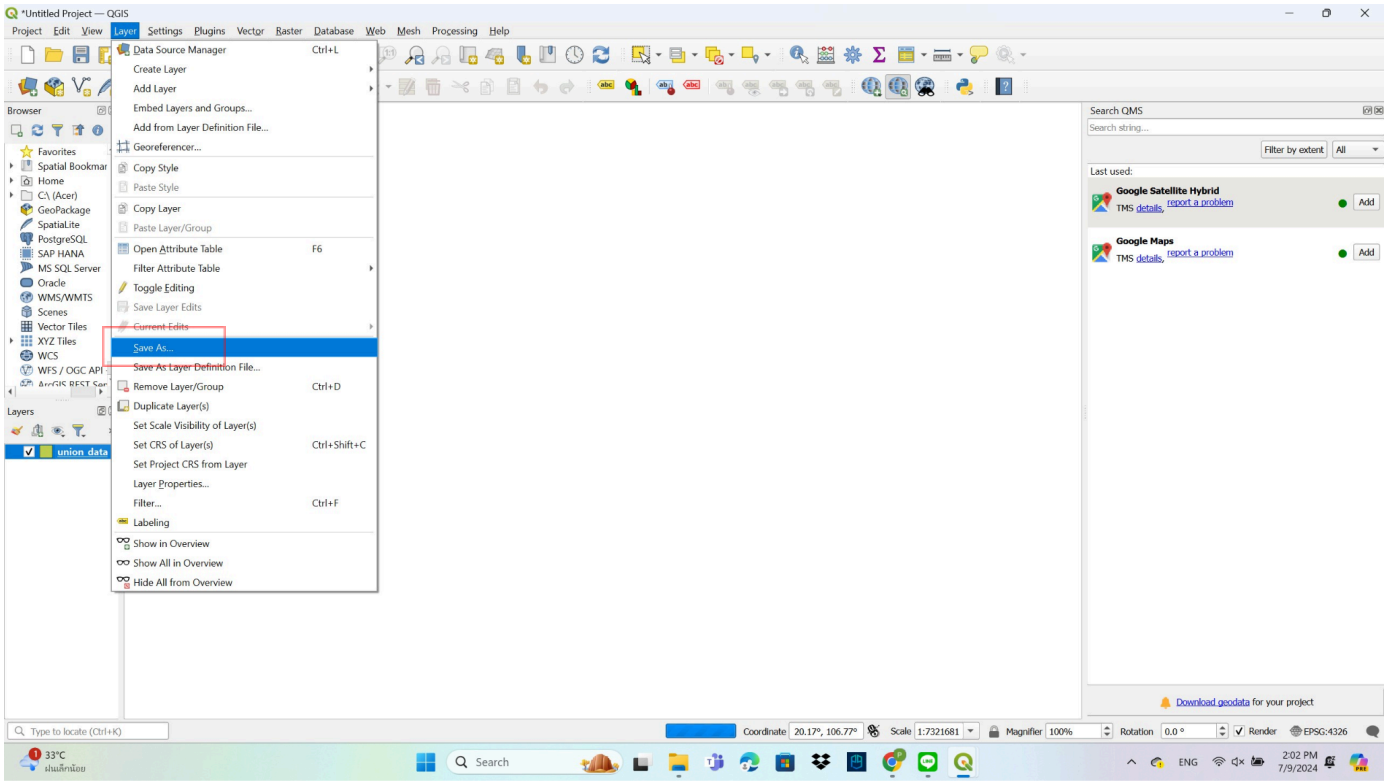
Step 17: After clicking the OK button, It will display the table that shows the result after the filter.



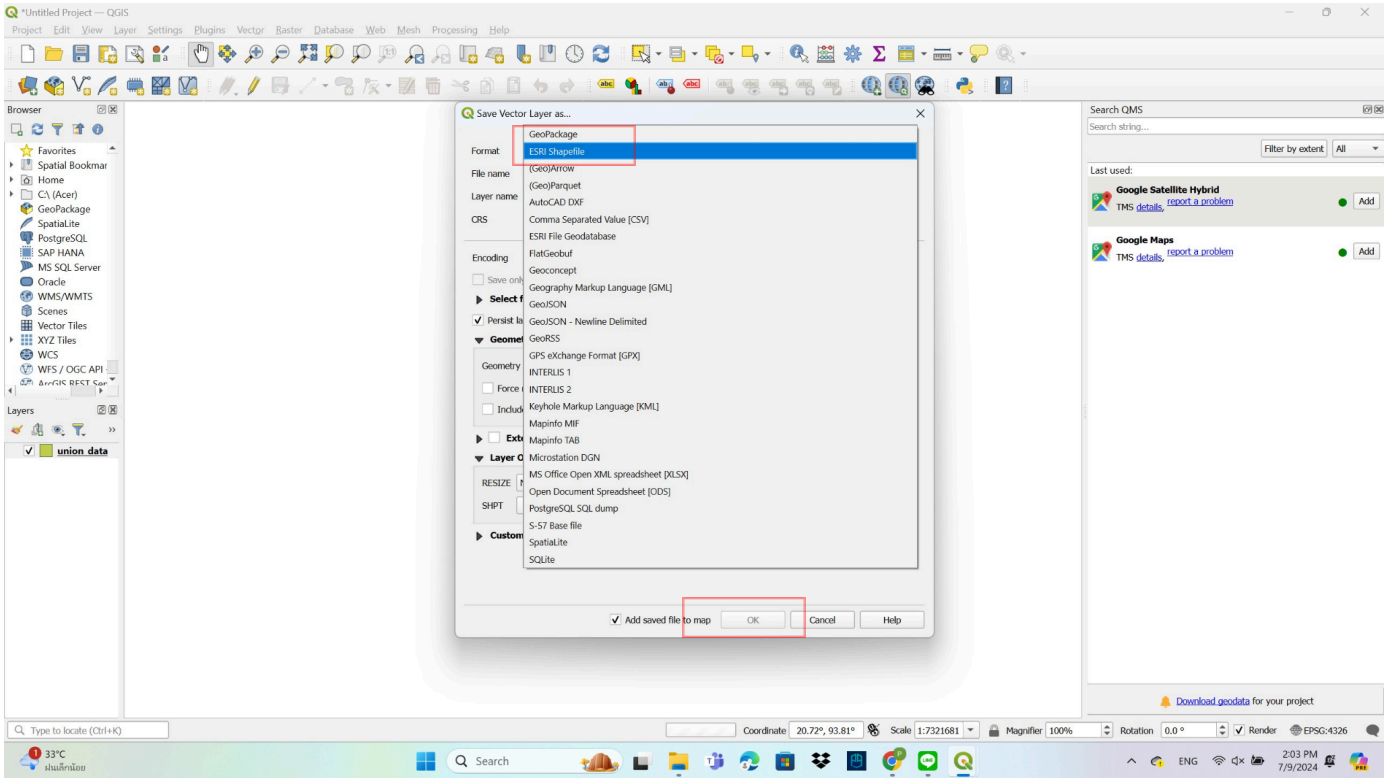
Step 18: Click on the Layer button



Step 19: Save as

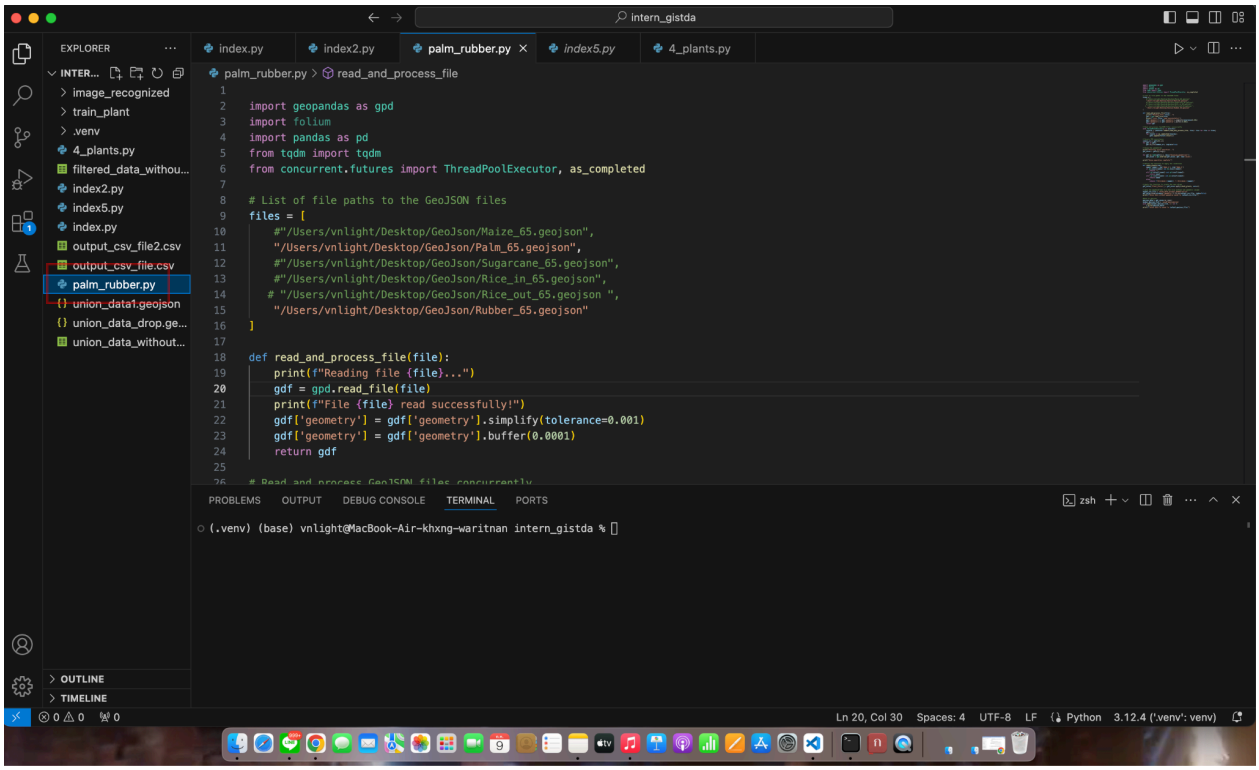


Step 20: Choose the file type that you want to export and then click the OK button.

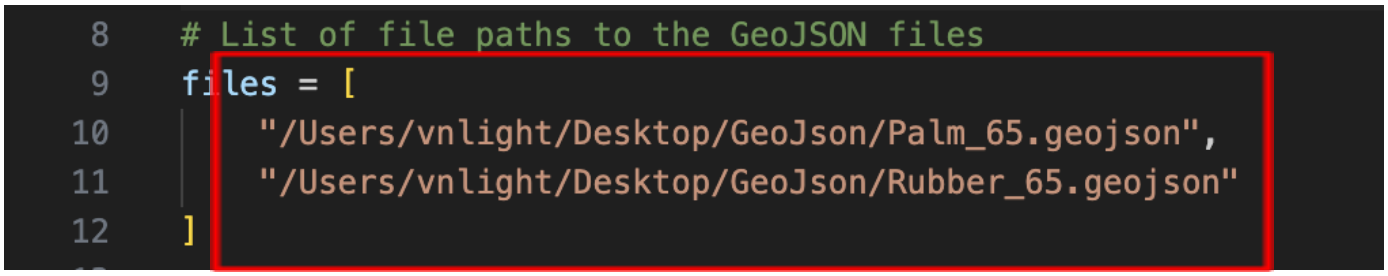


Note: All of these steps are for four crop types. If you want to make Perennial Plants (Palm and Rubber) , you can do the following step as shown below.

Step 1: For the Four crop type , Find the file name “palm_rubber.py”



Step 2: Replace file path with the file that you want.



In addition to this, you can follow all the steps above because for four types of crop and perennial plants (palm , rubber) , everything is the same. Things that need to be done is to change the file paths.