

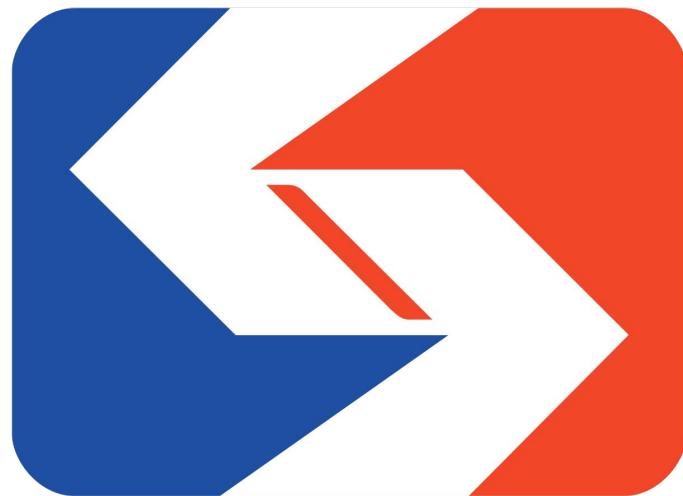
SEPTA

File Miner Script Deck

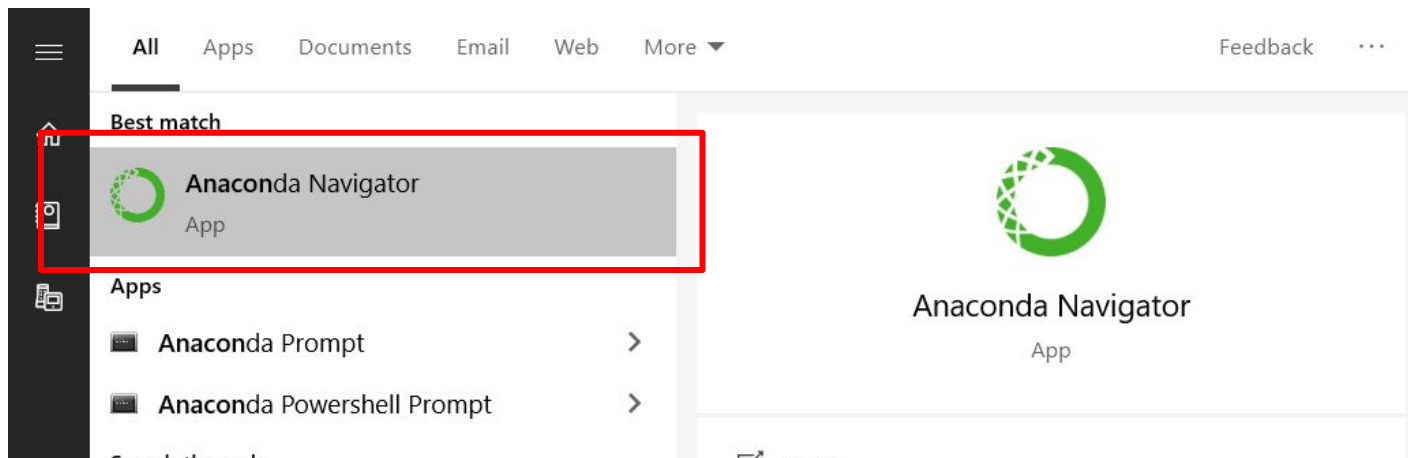
Author:

Sarvesh Shah

sarveshshah00@gmail.com



Open Anaconda Navigator



Use JupyterLab

The screenshot displays the JupyterLab application interface. On the left is a sidebar with navigation links: Home (with a green house icon), Environments (with a cube icon), Learning (with a book icon), and Community (with a group of people icon). The main area is titled 'Applications on' followed by a dropdown menu showing 'base (root)' and a 'Channels' button. Below this, there are two application cards. The first card is for 'JupyterLab' (version 1.1.4), featuring a circular logo with the word 'lab'. Its description is 'An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.' At the bottom of this card is a 'Launch' button, which is highlighted with a red rectangular border. The second card is for 'Jupyter Notebook' (version 6.0.2), featuring a logo with the word 'jupyter' and an orange circular graphic. Its description is 'Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.' It also has a 'Launch' button at the bottom. To the right of the Jupyter Notebook card, there is a partially visible card for 'Sci EnvRor' with the text 'advan debug'. Each application card has a gear icon in its top right corner for settings.

Home

Environments

Learning

Community

Applications on base (root) Channels

lab

JupyterLab

1.1.4

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch

jupyter

Notebook

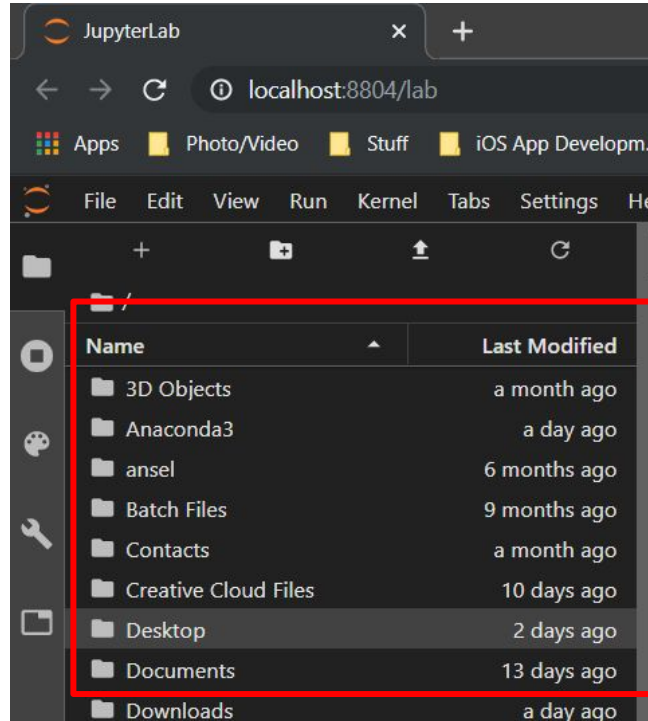
6.0.2

Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.

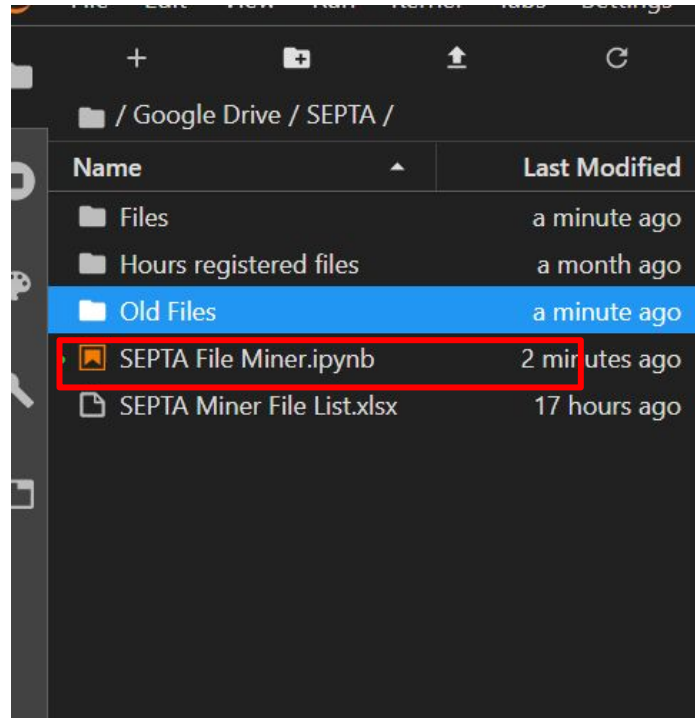
Launch

Sci EnvRor
advan debug

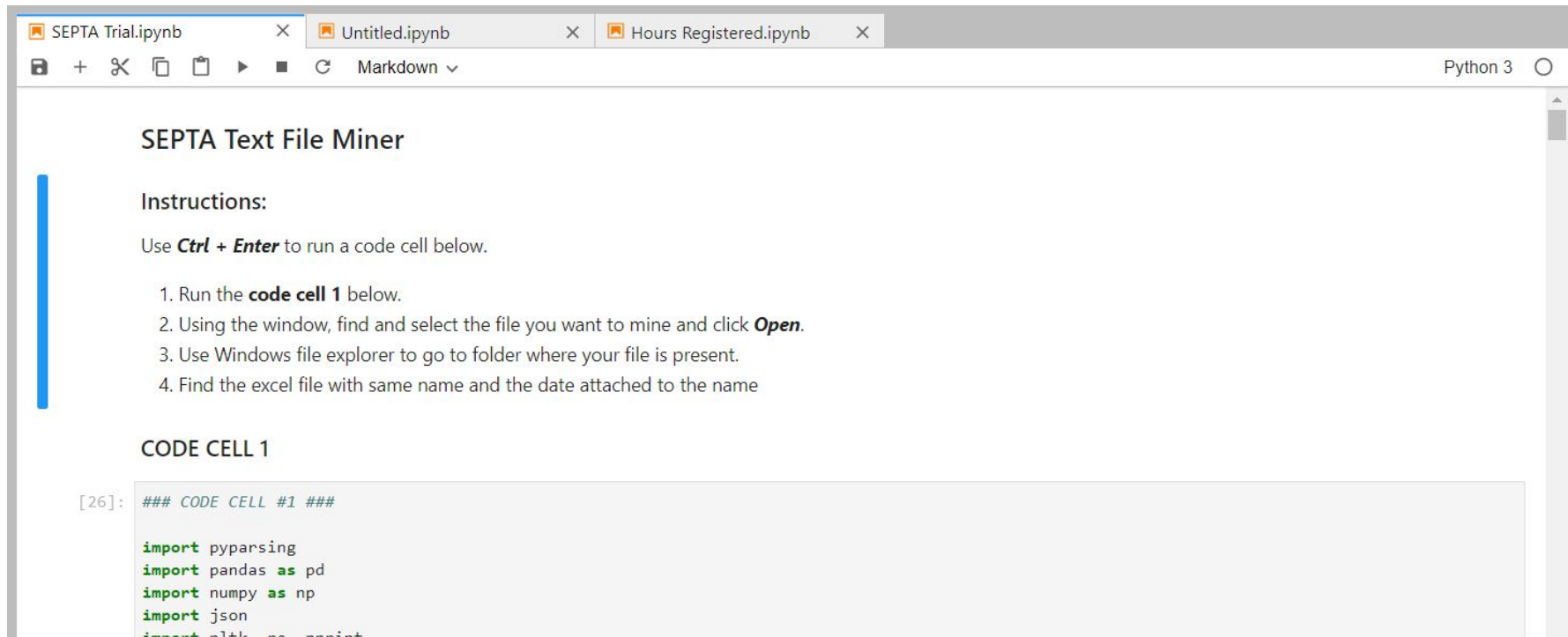
Use the file browser to open the python file



Open the python file with double-click



Python notebook



The screenshot shows a Jupyter Notebook interface with three tabs: 'SEPTA Trial.ipynb', 'Untitled.ipynb', and 'Hours Registered.ipynb'. The 'Untitled.ipynb' tab is active. The notebook content includes a title 'SEPTA Text File Miner', a section 'Instructions:', and a list of four steps. Below the instructions is a code cell labeled 'CODE CELL 1' containing Python code for imports.

SEPTA Text File Miner

Instructions:

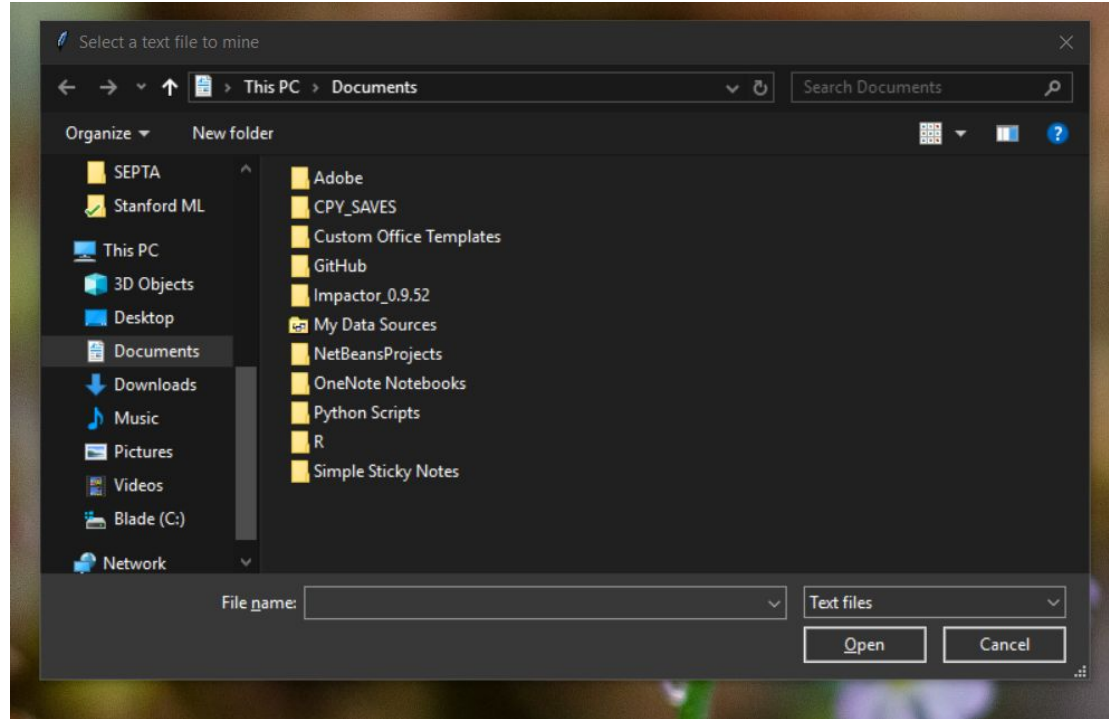
Use **Ctrl + Enter** to run a code cell below.

1. Run the **code cell 1** below.
2. Using the window, find and select the file you want to mine and click **Open**.
3. Use Windows file explorer to go to folder where your file is present.
4. Find the excel file with same name and the date attached to the name

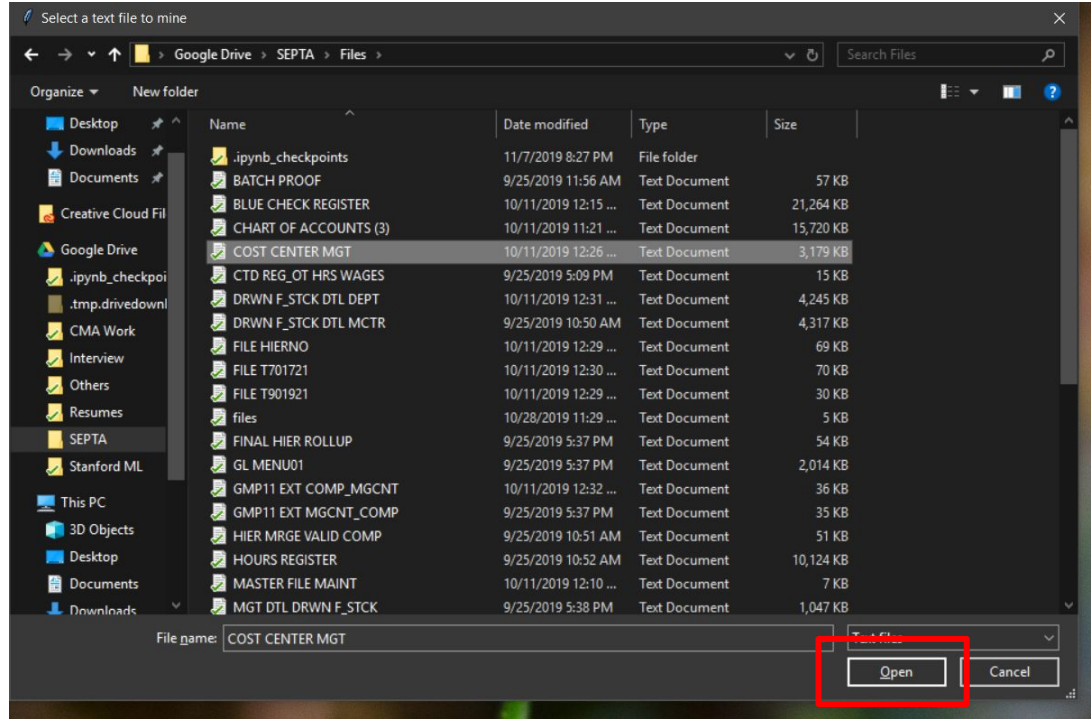
CODE CELL 1

```
[26]: ### CODE CELL #1 ###  
  
import pyarsing  
import pandas as pd  
import numpy as np  
import json  
import nltk as nlp
```

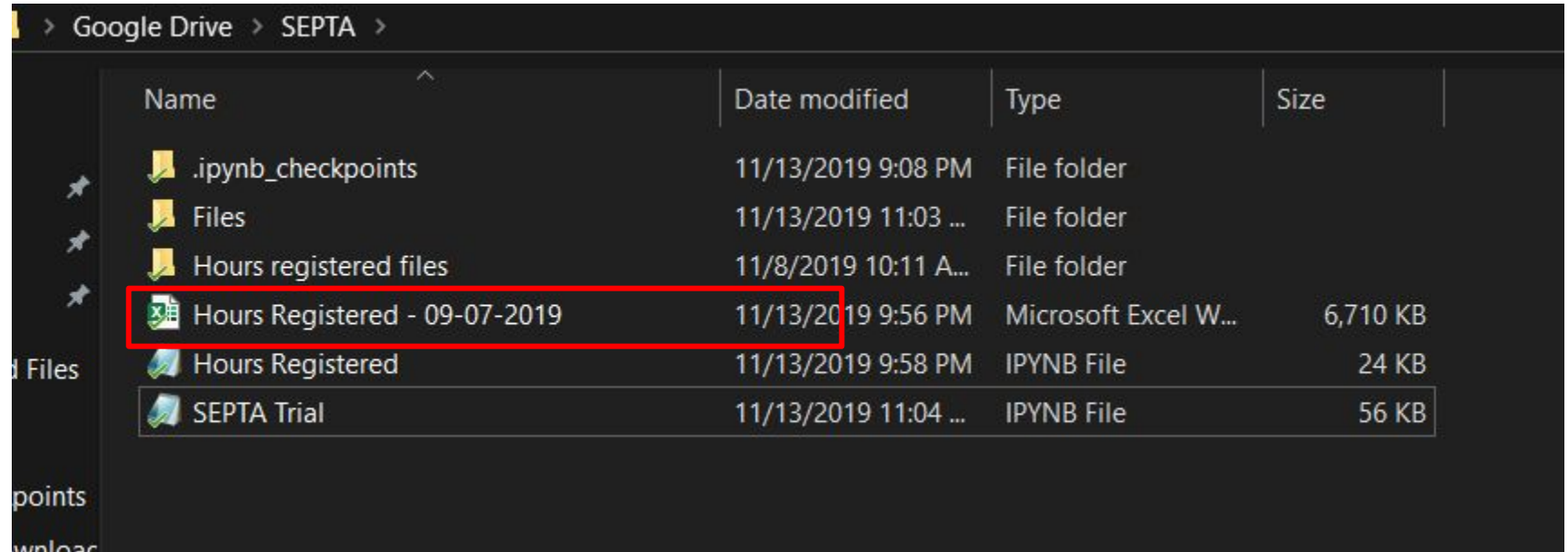
Use this window to find your file



Select the file and click 'Open'



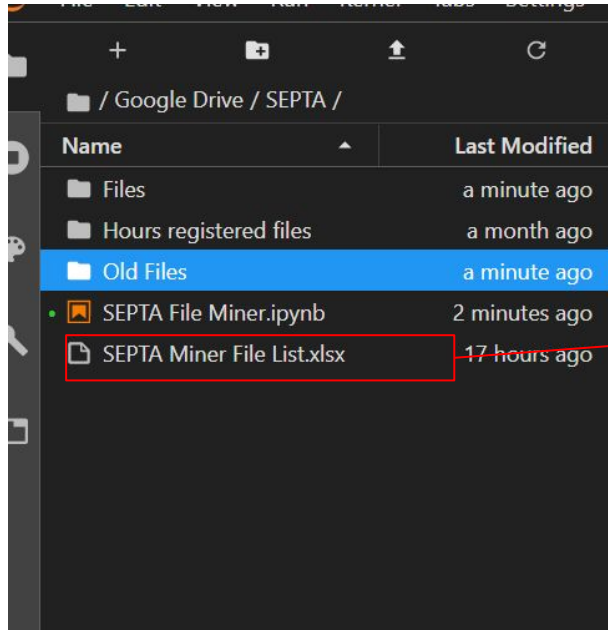
‘Go back to the folder and find your output’



The screenshot shows a Google Drive interface with a dark theme. The breadcrumb navigation at the top reads 'Google Drive > SEPTA >'. Below this is a table of files and folders. The file 'Hours Registered - 09-07-2019' is highlighted with a red rectangular box. To the left of the table, there is a sidebar with a search bar and several pinned items, including 'd Files', 'points', and 'wnload'.

Name	Date modified	Type	Size
.ipynb_checkpoints	11/13/2019 9:08 PM	File folder	
Files	11/13/2019 11:03 ...	File folder	
Hours registered files	11/8/2019 10:11 A...	File folder	
Hours Registered - 09-07-2019	11/13/2019 9:56 PM	Microsoft Excel W...	6,710 KB
Hours Registered	11/13/2019 9:58 PM	IPYNB File	24 KB
SEPTA Trial	11/13/2019 11:04 ...	IPYNB File	56 KB

The files that will work with this script are stored in an excel file named 'SEPTA Miner File List.xlsx'



A screenshot of an Excel spreadsheet with a single column labeled 'A'. The first row contains the header 'File Name'. The subsequent rows contain a list of file names, starting with 'BATCHPROOF.TXT' and ending with 'MPCTREAWORKREPORT.TXT'. A red arrow from the Google Drive screenshot points to the row containing 'HOURSREGISTER.TXT'.

A
File Name
BATCHPROOF.TXT
CHARTOFACCOUNTS.TXT
CTDREG_OTHRSWAGES.TXT
FINALHIERROLLUP.TXT
GLMENU01.TXT
GMP11EXTCOMP_MGCNT.TXT
GMP11EXTMGCNT_COMP.TXT
HOURSREGISTER.TXT
MGTDTLDRWNF_STCK.TXT
MPCACCTPAYABLE.TXT
MPCACCTRECVRREPORT.TXT
MPCAPCLAIMSREPORT.TXT
MPCGLJOURNALS.TXT
MPCMATANDSUPP.TXT
MPCPAYROLL.TXT
MPCPROVLIAB.TXT
MPCREVENUE.TXT
MPCTREAWORKREPORT.TXT

Things to know

Ensure the file name is named as shown in the excel file list.

If not, please ensure that the file is renamed before you feed it to the script.

The script will display appropriate messages if something goes wrong, kindly follow the message instructions and re-run the script.

If you would like to explore the script, please create a backup copy of the script and then tweak the script.