

This guide from the perspective of a Windows user, so some steps could be different if you use a different OS.

Setting Up Python

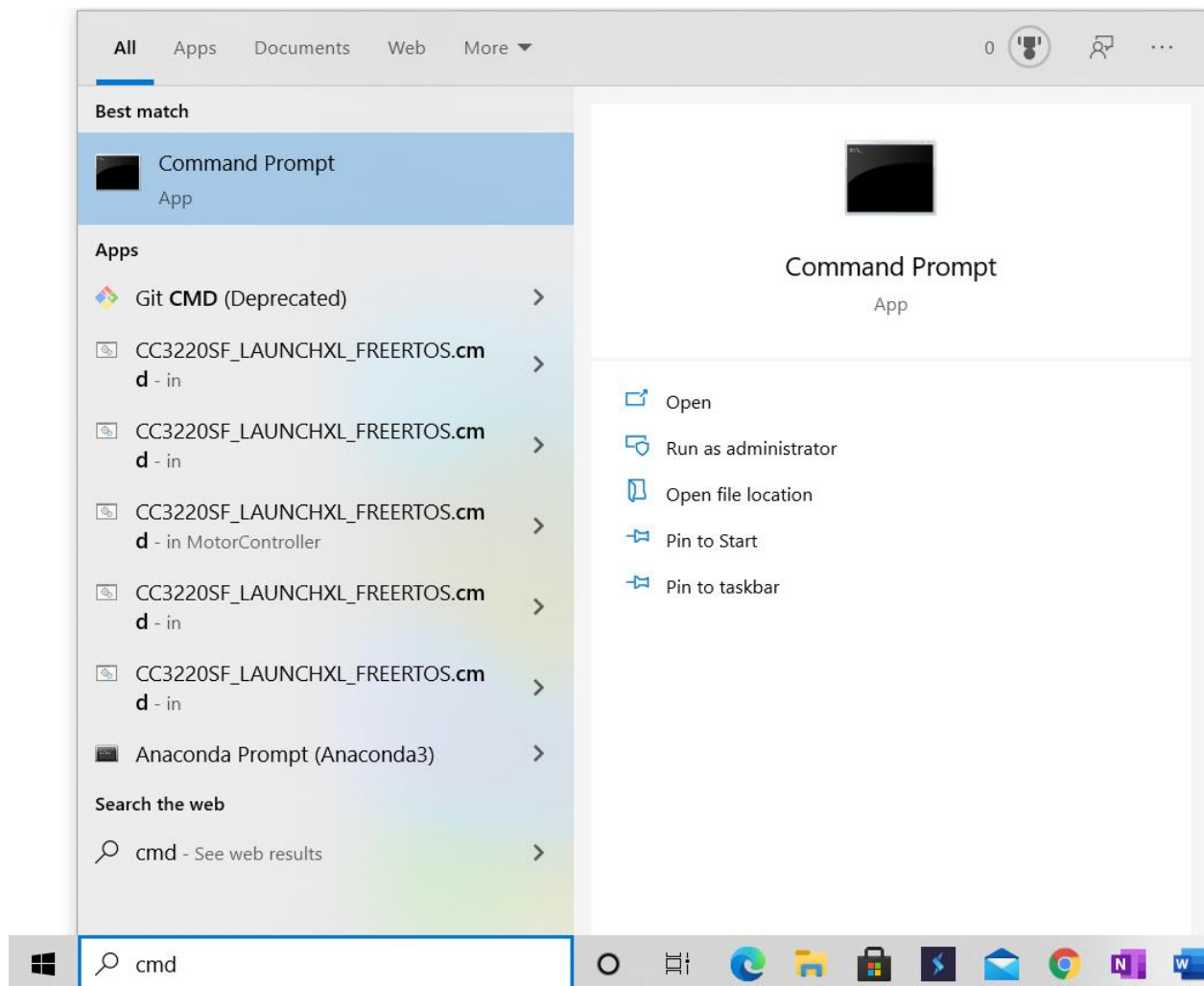
To set python up on your computer first install python from python.org

<https://www.python.org/downloads/>

When installing, it is important to check the “Add Python 3.9 to PATH” box.



To check that python was installed correctly, open the command prompt on your computer



Then, type and enter “python”. If the installation was successful, your command prompt should look something like this

```
Command Prompt - python
Microsoft Windows [Version 10.0.18363.1198]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\richa>python
Python 3.9.0 (tags/v3.9.0:9cf6752, Oct 5 2020, 15:34:40) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

The three arrows (>>>) indicate that you are in python mode. For now, you will need to exit python mode to install modules for it. To exit python mode, type and enter “exit()”

After exiting python mode, you will need to install modules that are used in the program. These modules are, numpy, pandas, xlrd(version 1.2), and openpyxl

To install these modules, enter the following

“pip install numpy”

“pip install pandas”

“pip install xlrd==1.2.0”

“pip install openpyxl”

The result should look something like this:

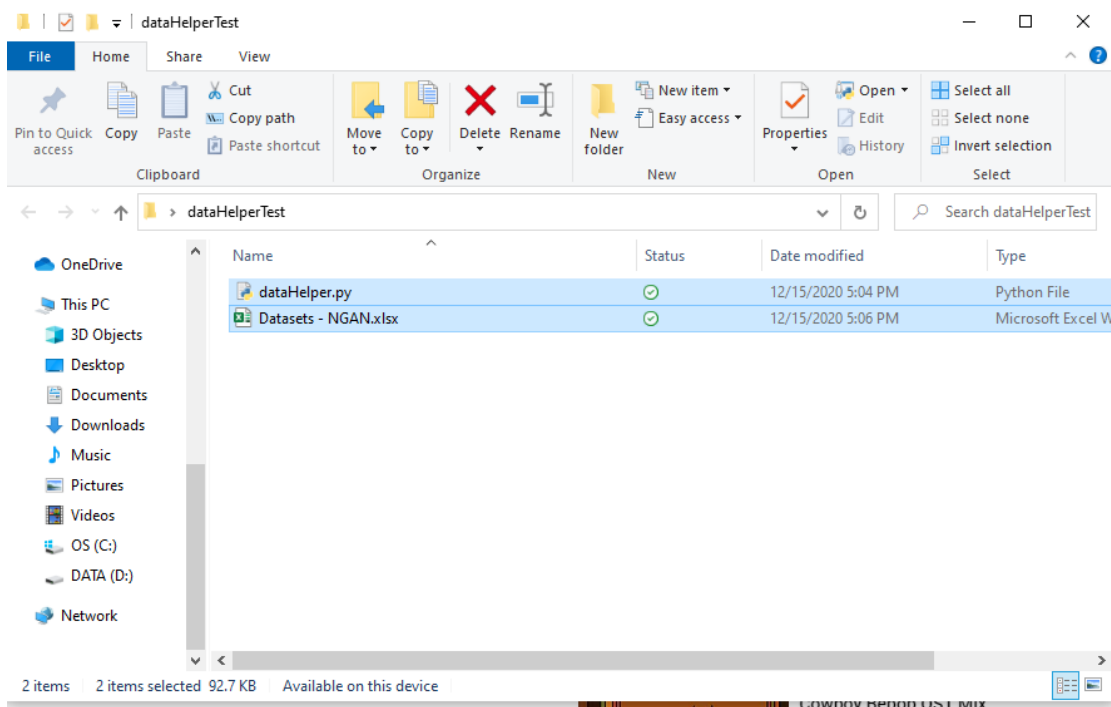
```
C:\Users\richa>python
Python 3.9.0 (tags/v3.9.0:9cf6752, Oct 5 2020, 15:34:40) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> exit()

C:\Users\richa>pip install numpy
Collecting numpy
  Downloading numpy-1.19.4-cp39-cp39-win_amd64.whl (13.0 MB)
    |#####| 13.0 MB 6.4 MB/s
Installing collected packages: numpy
Successfully installed numpy-1.19.4
WARNING: You are using pip version 20.2.3; however, version 20.3.3 is available.
You should consider upgrading via the 'c:\users\richa\appdata\local\programs\python\python39\python.exe -m pip install --upgrade pip' command.

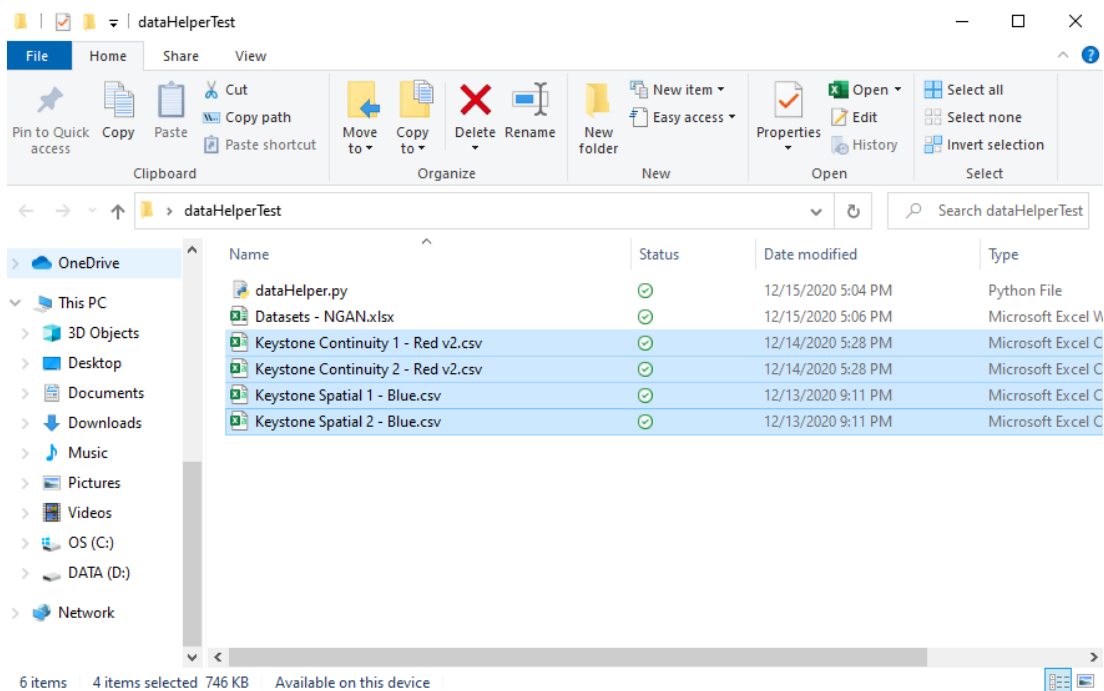
C:\Users\richa>pip install pandas
Collecting pandas
  Downloading pandas-1.1.5-cp39-cp39-win_amd64.whl (8.9 MB)
    |#####| 8.9 MB 945 kB/s
Collecting python-dateutil>=2.7.3
  Downloading python_dateutil-2.8.1-py2.py3-none-any.whl (227 kB)
    |#####| 227 kB ...
Requirement already satisfied: numpy>=1.15.4 in c:\users\richa\appdata\local\programs\python\python39\lib\site-packages
(from pandas) (1.19.4)
Collecting pytz>=2017.2
  Downloading pytz-2020.4-py2.py3-none-any.whl (509 kB)
    |#####| 509 kB 6.8 MB/s
```

Setting Up the Script's Environment

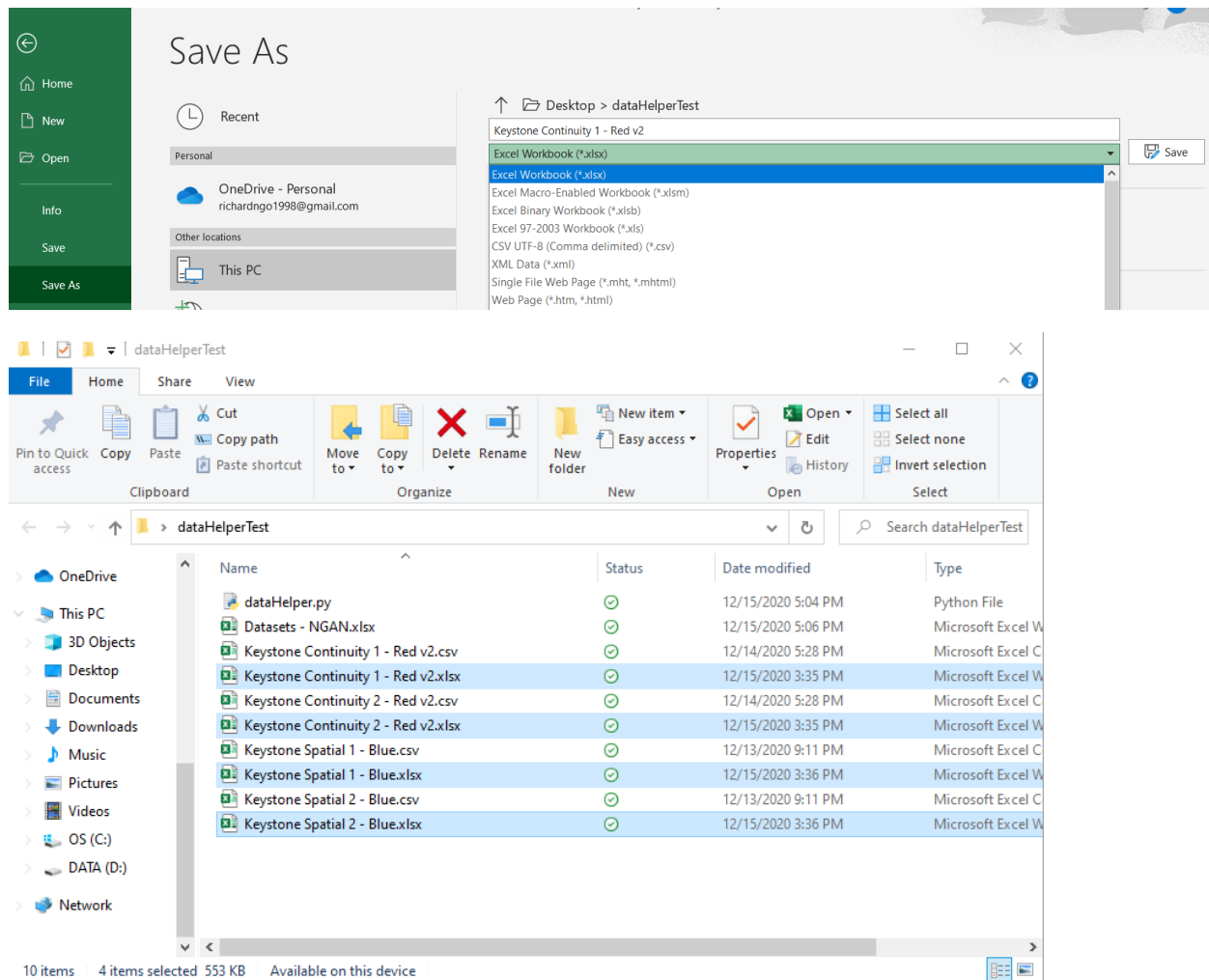
First, create a folder that you want to run the program in, and drag all the files inside the dataHelper.zip file into this folder



Next, drag in all the spatial and continuity files relevant to your current dataset.



For some reason, the program sometimes has a hard time reading some files as csv files. To be on the safe side, you should resave xlsx versions of these files

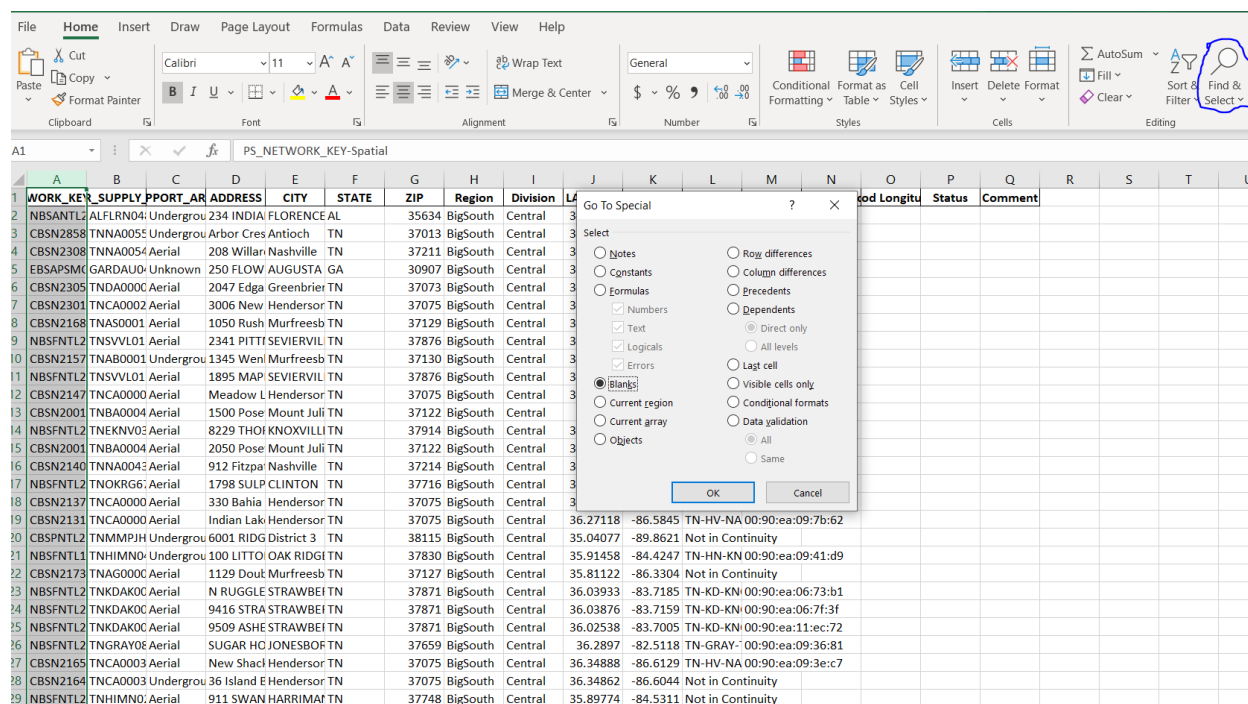


To add your own dataset in, open up the Datasets - NGAN.xlsx file, and replace the information in the dataset with your own. (The program uses the column names to understand what is what, so make sure to keep those intact)

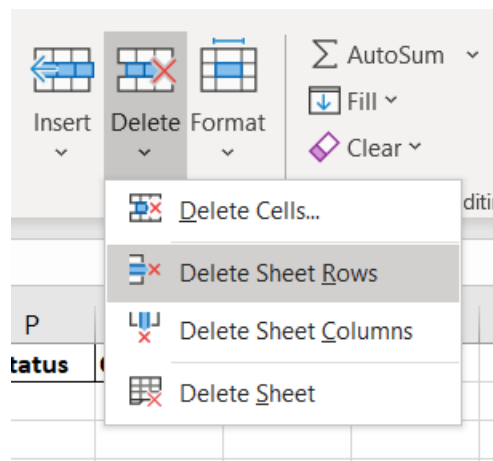
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	WORK KEY	SUPPLY POINT	ADDRESS	CITY	STATE	ZIP	Region	Division	LATITUDE	LONGITUDE	Inuity PS	Mac Address	Address	
2	FRKBPSM(WV0278A)	Aerial	County Hv	FAIRMON	WV	26554	Keystone	Northeast	39.4492	-80.2074				
3	FRKDPZ11 PARD00C2	Undergro	7th Ave	TEMPLE	PA	19605	Keystone	Northeast	40.39308	-75.9215				
4	FRKDPZ11 PARD00D	Aerial	Cumberla	Mount Pe	PA	19606	Keystone	Northeast	40.32884	-75.8908				
5	FRKDPZ11 PARD00J4	Aerial	503 Eisent	Hyde Park	PA	19605	Keystone	Northeast	40.37761	-75.9296				
6	FRKEPSM(PA0121T6)	Aerial	Jackman R	SAINT BEN	PA	15773	Keystone	Northeast	40.62911	-78.7292				
7	FRKEPSM(PA0121T2)	Aerial	Beaver St	Hastings	PA	16646	Keystone	Northeast	40.66452	-78.7126				
8	FRKEPSM(PA0121T2)	Aerial	Municipal	HASTINGS	PA	16646	Keystone	Northeast	40.65998	-78.7203				
9	FRKEPSM(PA0121T9)	Aerial	Brick Rd	CARROLLT	PA	15722	Keystone	Northeast	40.58278	-78.7111				
10	FRKEPSM(PA0121T8)	Aerial	Deveaux S	CARROLLT	PA	15722	Keystone	Northeast	40.59746	-78.7226				
11	FRKEPSM(PA0121T8)	Aerial	W Campb	Carrolltow	PA	15722	Keystone	Northeast	40.60583	-78.7103				
12	FRKEPSM(PA0121T8)	Aerial	Myers St	EBENSBUF	PA	15931	Keystone	Northeast	40.48062	-78.7411				
13	FRKEPSM(PA0121T7)	Aerial	Township	CARROLLT	PA	15722	Keystone	Northeast	40.60558	-78.7527				
14	FRDS4840 LX0041AC	Unknown			PA	17112	Keystone	Northeast	40.32477	-76.8338				
15	FRKBUI01 WV0432N	Aerial	Filmore St	New Cum	WV	26047	Keystone	Northeast	40.49752	-80.6027				
16	FRKBPSM(WV0278A)	Aerial	County Hv	FAIRMON	WV	26554	Keystone	Northeast	39.44116	-80.0579				
17	FRKBPSM(WV1167Y)	Aerial	County Hv	FLEMINGT	WV	26354	Keystone	Northeast	39.28813	-80.0451				
18	FRKBPSM(WV1167Y4)	Aerial	County Hv	GRAFTON	WV	26354	Keystone	Northeast	39.30027	-79.9588				

Average: 6235.462487 Count: 372 Sum: 636017.1736 Display Settings

Currently, empty rows will also throw the program off. To remove empty rows within the Datasets-NGAN.xlsx file, highlight just the first column (click on A), go to Home->Find & Select->Go To Special, and choose Blanks



Then go to Delete Sheet Rows to remove them.



Your environment should be set up

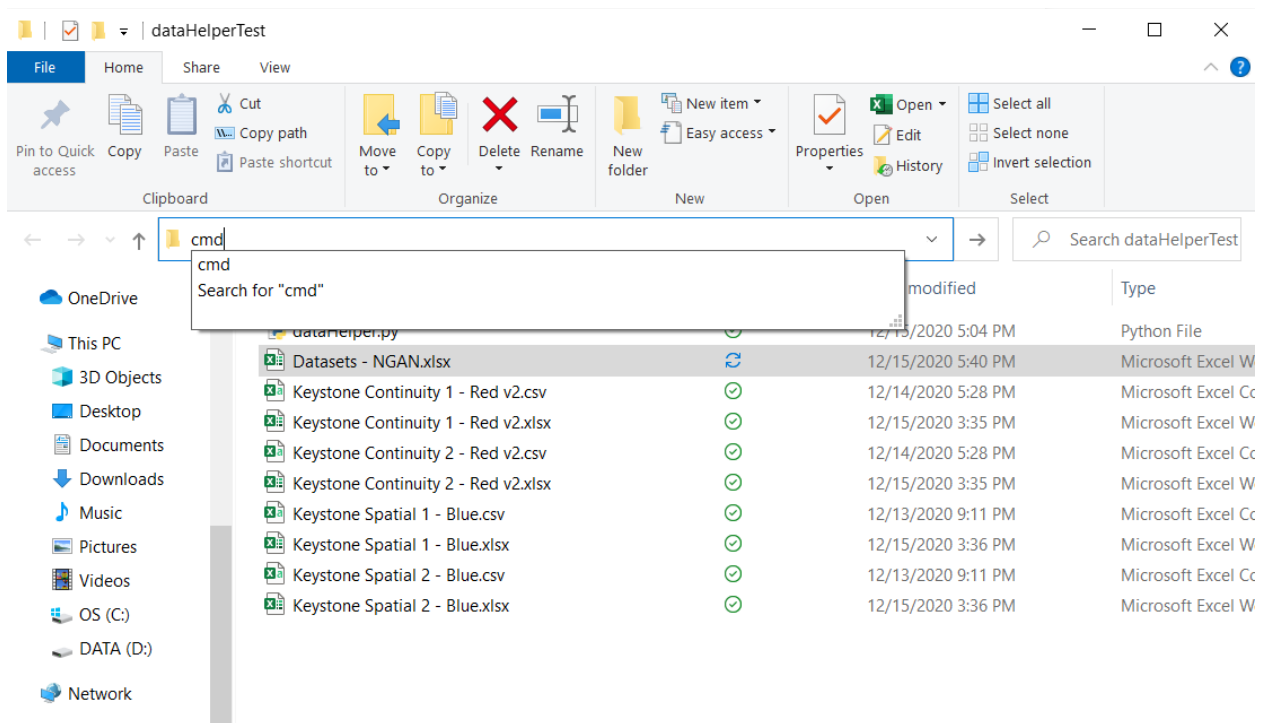
Using the Script with your Dataset

Within dataHelper.py, the only portion that needs to be changed is the Settings section.

```
10
11 import pandas as pd
12 import numpy as np
13
14 global myFile
15 global spaFiles
16 global conFiles
17 global myDB
18 global spaDB
19 global conDB
20 global KeepInput
21 global maxRange
22
23 #Settings
24 #your dataset file name goes below
25 myFile = 'Datasets - NGAN.xlsx'
26 spaFiles = ['Keystone Spatial 1 - Blue.xlsx', 'Keystone Spatial 2 - Blue.xlsx']
27 conFiles = ['Keystone Continuity 1 - Red v2.xlsx', 'Keystone Continuity 2 - Red v2.xlsx']
28
```

Open this file with notepad or any other text editor, and replace the filenames inside of spaFiles with the spatial files of your dataset, and conFiles with the continuity files of your dataset.

Now, to finally run the script, open up the folder that you are using for this program, and type and enter cmd into the file location bar to open the command prompt within your folder.



In this command prompt, enter python mode by entering “python”.

Then, enter “import dataHelper” to load the script, and then run the functions inside by entering the following:

“dataHelper.findClosestReds()” -Finds and saves the closest continuity to each spatial point

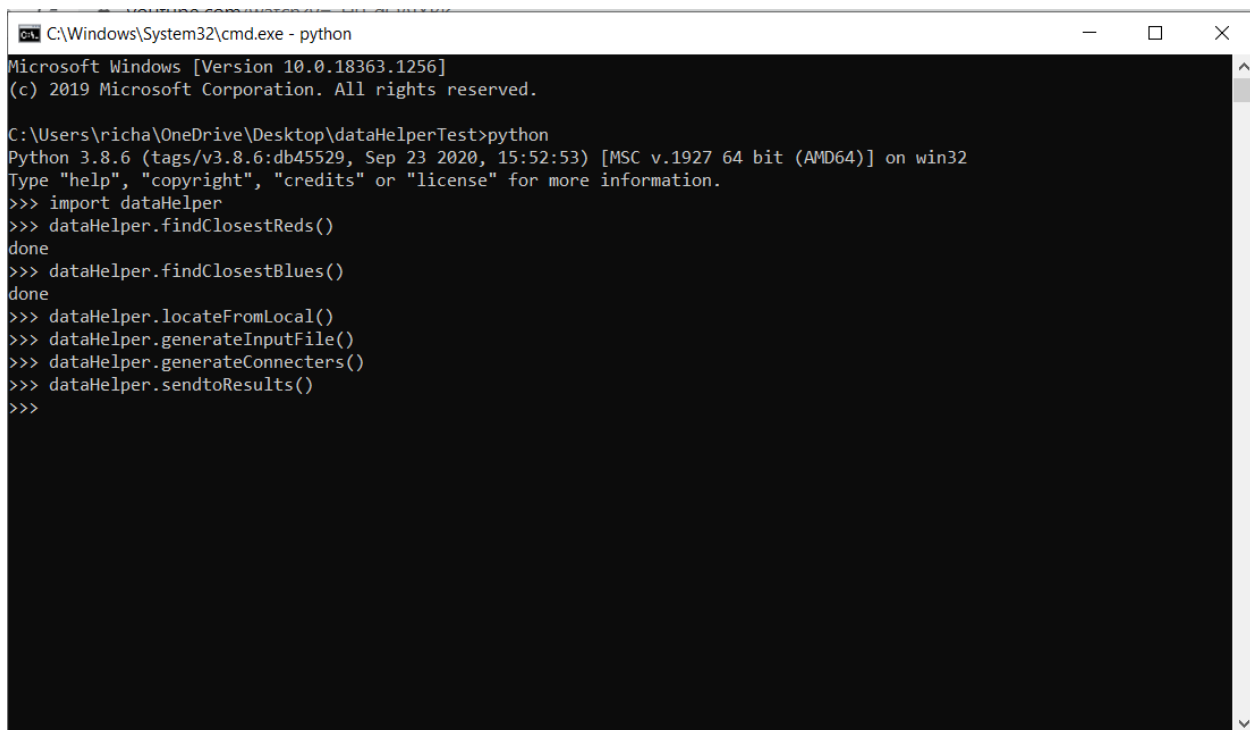
“dataHelper.findClosestBlues()” -Finds and saves the spatial closest to each continuity point

“dataHelper.locateFromLocal()” -Gets pairs for your dataset in Datasets-NGAN.xlsx

“dataHelper.generateInputFile()” -Creates a file needed for the next two functions below

“dataHelper.generateConnecters()” – Creates a kml file that can be imported to Google Earth, visually shows what pairs were predicted. (when importing, make sure you change the file filter to include all files and not just .txt and .csv, otherwise you won’t be able to see kml files)

“dataHelper.sendtoResults()” – Sends results back to the Datasets-NGAN.xlsx file

A screenshot of a Windows command prompt window titled "C:\Windows\System32\cmd.exe - python". The window shows the execution of a Python script. The prompt is "C:\Users\richa\OneDrive\Desktop\dataHelperTest>python". The output shows the Python version (3.8.6) and the script's location. The user enters "import dataHelper", and the prompt changes to ">>>". The user then enters a series of function calls: "dataHelper.findClosestReds()", "dataHelper.findClosestBlues()", "dataHelper.locateFromLocal()", "dataHelper.generateInputFile()", "dataHelper.generateConnecters()", and "dataHelper.sendtoResults()". Each function call is followed by a "done" message. The prompt returns to ">>>" after each function call. The window has standard Windows window controls (minimize, maximize, close) in the top right corner.

```
C:\Windows\System32\cmd.exe - python
Microsoft Windows [Version 10.0.18363.1256]
(c) 2019 Microsoft Corporation. All rights reserved.

C:\Users\richa\OneDrive\Desktop\dataHelperTest>python
Python 3.8.6 (tags/v3.8.6:db45529, Sep 23 2020, 15:52:53) [MSC v.1927 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import dataHelper
>>> dataHelper.findClosestReds()
done
>>> dataHelper.findClosestBlues()
done
>>> dataHelper.locateFromLocal()
>>> dataHelper.generateInputFile()
>>> dataHelper.generateConnecters()
>>> dataHelper.sendtoResults()
>>>
```

The Datasets-NGAN.xlsx file should now be updated with all the predicted pairs for your set.

Depending on how many spatial or continuity files you are using, the functions findClosestReds() and findClosestBlues() may take a while. However, these two functions only need to be run once for every set of spatial and continuity files you use.

Whenever your own dataset in Datasets-NGAN.xlsx is changed or updated, functions from locateFromLocal() and below will need to be run again.

Note: if you get an error similar to whats shown below, try using numpy (version 1.19.3) instead

pip install numpy==1.19.3

```
>>> (import dataHelper)
File "<stdin>", line 1
  (import dataHelper)
  ^
SyntaxError: invalid syntax
>>> import dataHelper
** On entry to DGEBAL parameter number 3 had an illegal value
** On entry to DGEHRD parameter number 2 had an illegal value
** On entry to DORGHR DORGQR parameter number 2 had an illegal value
** On entry to DHSEQR parameter number 4 had an illegal value
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "C:\Users\chris\Desktop\PI\dataHelper.py", line 11, in <module>
    import pandas as pd
  File "C:\Users\chris\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.8_qbz5n2kfra8p0\LocalCache\local-package
s\Python38\site-packages\pandas\__init__.py", line 11, in <module>
    import __dependency__
  File "C:\Users\chris\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.8_qbz5n2kfra8p0\LocalCache\local-package
s\Python38\site-packages\numpy\__init__.py", line 305, in <module>
    _win_os_check()
  File "C:\Users\chris\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.8_qbz5n2kfra8p0\LocalCache\local-package
s\Python38\site-packages\numpy\__init__.py", line 302, in _win_os_check
    raise RuntimeError(msg.format(__file__)) from None
RuntimeError: The current Numpy installation ('C:\\Users\\chris\\AppData\\Local\\Packages\\PythonSoftwareFoundation.Pyth
on.3.8_qbz5n2kfra8p0\\LocalCache\\local-packages\\Python38\\site-packages\\numpy\\__init__.py') fails to pass a sanity c
heck due to a bug in the windows runtime. See this issue for more information: https://tinyurl.com/y3dm3h86
>>>
```