**Python3 Geocoder Instructions**

**Requirements:** Python 3, Google API Key

**Location:** I:\OBI\Staff Working Folders\MSchwartz\Python3Geocoder

**Introduction**

This program allows you to convert a set of addresses into coordinate points consisting of latitude and longitude values. This process is called ‘geocoding.’ Most GIS software and web mapping applications can handle coordinate points much more accurately and efficiently than addresses, allowing for a wider range of spatial analysis. You will choose a file containing addresses that you would like to geocode, and the program will create a new csv containing the coordinates and the name of the county for that address.

**Notes**

This program makes use of a Google API (Application Programming Interface), which only allows for 2500 requests per day (and 50 per second). This means that the dataset you use with this program must have no more than 2500 records. If it is larger, consider splitting the dataset into chunks and geocoding one chunk a day, or take a look at some of Google’s API usage [plans](https://www.google.com/url?q=https%3A%2F%2Fdevelopers.google.com%2Fmaps%2Fdocumentation%2Fgeocoding%2Fusage-limits) that allow for more requests.

**Safety**

The Google Maps API utilized in this program uses HTTPS (the Hypertext Transfer Protocol Secure). This means that a protocol called SSL (Secure Sockets Layer) is used to encrypt the data being transmitted between users and their website, preventing hackers from intercepting the data and being able to see any information. Most online banking sites use this same encryption method to protect their clients and their data. You can learn more about SSL [here](https://www.digicert.com/ssl/). Furthermore, the only data being sent to Google are the actual addresses; none of the other information in your dataset is sent.

**Setup**

First of all, make sure you have Python **3** installed on the computer you are using. It will not work with Python 2. Downloads are available [here](https://www.python.org/downloads/).

Next, make sure you have downloaded the files accompanying this document. There should be three others within the ‘Python3Geocoder’ folder this document came in: ‘setup.py’, ‘geocode.py’ and ‘apikey.csv’. Move ‘Python3Geocoder’ file to a convenient location on your computer. Make sure you do not have ‘apikey.csv’ open on your desktop while running the program.

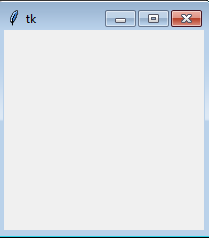
Once that is done, go to your command prompt. In Windows, you can find this by clicking on the start button and typing ‘cmd’ into the search box. In the command prompt window, type ‘python’ (without quotes) to make sure python is working. If you get an error message similar to “'python' is not recognized as an internal or external command, operable program or batch file.”, this [link](https://stackoverflow.com/questions/7054424/python-not-recognised-as-a-command) walks you through the steps to fix it.

Next, navigate to your ‘Python3Geocoder’ folder, and get the path to it. (Right click → Properties → General → Location). In the command prompt, type ‘cd’ and the path. [Example: cd C:\Users\Python3Geocoder]. Hit enter, then type ‘python setup.py’. This will run the setup.py file which installs all the Python modules needed for geocoding. If you get a ‘SyntaxError: invalid syntax’ message, enter CTRL + Z, and then try again.

**Running the Program**

In the command prompt, after you have run ‘python setup.py’, enter ‘python geocode.py’ and the program will start. A short message describing the program will display. If this is your first time running the program, it will ask you to type in your Google API key. If you do not already have one, you can request one [here](https://developers.google.com/maps/documentation/javascript/get-api-key). The program will automatically save the API key you enter for future use. If you have used the program before but would like to use a different API key, enter “1” (without quotes) when the program prompts you to do so. Also, to keep other people from using your API key and taking up some of your daily quota (explained in the ‘Notes’ section of this document), it is best if you keep it out of emails and other web services.

Once you input your API key, the program will ask you to select the file containing the data you want to geocode. A file window should pop up. Navigate to the file you want to use, then double click on it. Do NOT close the dialog box(es) that may pop up looking like the image below, as this will end the Python program early.



Once you’ve selected the correct file, look at the dataset and make a note of the names of the columns that contain a primary key for each record (such as an ID number) and the address component(s). The program will ask you to type in each of these names. If you misspell or capitalize the name differently than it appears in your dataset, the program will prompt you to enter it again. The program will also ask you about how your data is structured. If your address is contained in a single column, enter “1” (without quotes), and if it is split up into multiple columns, enter “2”. You will then enter the name(s) of the column(s) that contains your address data.

One column Multiple columns

Once you have done that, the geocoding process will begin. This may take several minutes, so do not close the program. If you see a message that says ‘A geocoding error occurred,’ this means that one or more of your addresses was not able to be geocoded. This could be because the address was misspelled or improperly recorded in your dataset, or the cell may not contain a valid address. These addresses will still be recorded in the csv created at the end of the program.

If you see a message that says ‘There is already a file(s) in your folder that contains geocode results. Please change its name or move it.’, this means you have used this geocoding program before in the same folder as your original dataset and it is trying to overwrite the files it created previously. The program will pause and ask you to either rename the old files or move them to a different folder.

When the geocoder has completed, you will see a message that says ‘A new file called "geocoderesults.csv" has been added to the same folder as your original dataset.’ The program will then ask you if you would like this new dataset to be joined to the original dataset you selected at the beginning of the program. If you enter “yes”, it will perform a left join between the original and geocoded data and move this to a **new** csv called “geocodemerge.csv”. It will **not** change the original dataset you began with.

When you open the new csv(s) created, you will see that one column holds the primary key you indicated at the beginning of the program. The other columns contain the address, latitude, longitude, and county information obtained from the geocoding, as well as an ‘Error Type’ column that describes the type of error that occured, if applicable. The most common error type is ‘AttributeError,’ which means that the address is invalid. Usually, this means it is misspelled, has a typo, or is not actually an address. You may want to filter these rows out of your csv to avoid issues with GIS software (in Excel: Data → Filter → Click arrow next to name of column containing errors, then uncheck boxes next to cells with ‘ERROR’s). If the Error Type is ‘GeocoderQueryError’, this most likely means the API key you entered at the beginning of the program is invalid. If the Error Type is ‘GeocoderQuotaExceeded’, this means you have gone over your 24-hour request limit. You must either store these records in a new file and wait a day to finish geocoding, or upgrade your Google API usage plan (link above).

**Use with ArcGIS Software**

The csv created by this program can now be used to create maps or other visualizations of geographic data. One option is to use ArcGIS Pro. Tutorials for beginners can be found [here](http://pro.arcgis.com/en/pro-app/get-started/introducing-arcgis-pro.htm). When you get to the point where you want to plot the latitude and longitude points in your csv file, use the [Make XY Event Layer tool](http://pro.arcgis.com/en/pro-app/tool-reference/data-management/make-xy-event-layer.htm). (You may need to go into the csv file and change the Cell Format for latitude and longitude to ‘Number’ with 5 decimal points before ArcGIS Pro will read it as a set of coordinates.)

**Links**

<https://developers.google.com/maps/documentation/geocoding/usage-limits>

<https://www.digicert.com/ssl/>

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

<https://stackoverflow.com/questions/7054424/python-not-recognised-as-a-command>

<https://developers.google.com/maps/documentation/javascript/get-api-key>

<http://pro.arcgis.com/en/pro-app/get-started/introducing-arcgis-pro.htm>

<http://pro.arcgis.com/en/pro-app/tool-reference/data-management/make-xy-event-layer.htm>