1. Make a git clone of repo https://github.com/kharrigian/smgeo into your environment
2. Contact kharrigian@jhu.edu and ask for the pretrained files
3. After receiving them, I found I was unable to download them from dropbox in Ubuntu so it's better to have them on google drive and then download from there
4. Each file has a specific location that you need to put within the smgeo folder which is given in configurations/settings.json (make sure this part is done correctly)
5. Edit the config.json file and put in your Reddit API credentials
6. Then, run the two command lines to test if everything is in the right place:
   python -m pytest tests/ -Wignore -v
   python -m pytest tests/ --cov=smgeo/ --cov-report=html -v -Wignore

7. Do **touch nameoffile.csv** to make a csv file for the output
8. Then, run this line to get the geolocation for the global model:

python -m scripts/model/reddit/infer.py <model_path> <user_list> <output_csv> --reverse_geocode --end_date 2019-12-30

Where model_path should be ./models/Global_Text_SubredditTime/model.joblib

User_list should be the name of the file with the usernames and should end in .txt

python -m scripts/model/reddit/infer.py ./models/Global_Text_SubredditTime/model.joblib <user_list> <output_csv> --reverse_geocode --end_date 2019-12-30

The parameters at the end can be adjusted to your liking. Use the following to see all the settings

python scripts/model/reddit/infer.py --help

7. The pushshift.io request will take a while but after, the vectorization will occur. If you get an error saying "JSONDecodeError("Expecting value", s, err.value) from None"

Then it is likely some of the usernames are banned and you need to remove those. You can do this by either seeing at which number the vectorization stops and then deleting that user or a try catch method.

8. If you run into a memory error, just allot more memory.

9. After the process is complete, the .csv file will contain your results