**First off all I asked chat GPt to make the arrangement for my Trello so that I can make the work as sequential as possible:**

This is the question for my assignment to help me to sequence and step by process of the work don't generate code just make the steps clear:

Create a gazetteer for the NER places For all the places in the ner\_counts.tsv file, use geocoding to find their coordinates; use “NA” for places for which you did not find coordinates. Write your results to a tsv file in the gazetteer folder called “NER\_gazetteer.tsv”, which will contain 3 columns: placename, latitude and longitude. Save the script that generated the gazetteer as build\_gazetteer.py (or build\_gazetteer.ipynb if you did this on Google Colab). If the ner\_counts.tsv file contains “NA” values, look up the coordinates manually (using Google) and add them to the tsv file. Make a note of all the place names for which you had to look up the coordinates manually in your README file.

**ChatGPT said:**

Here’s a clear **step-by-step process** to follow for your assignment — without generating code, just explaining what you need to do in logical sequence:

**Step-by-Step Workflow for Creating a NER Gazetteer**

**Step 1: Understand Your Input File**

* Open and examine the ner\_counts.tsv file.
* This file should contain the list of place names identified using NER (Named Entity Recognition).
* These are the names you'll attempt to geocode.

**Step 2: Set Up Your Script**

* Create a Python script (or Colab notebook) and name it:  
  build\_gazetteer.py or build\_gazetteer.ipynb.

**Step 3: Read Place Names**

* Write code to **read the place names** from ner\_counts.tsv.
* You’ll extract just the names — ignore counts or other columns.

**Step 4: Use Geocoding to Get Coordinates**

* For each place name:
  + Attempt to **automatically geocode** the name (e.g., using geopy or another API).
  + If coordinates are found, record them.
  + If not found, record "NA" for both latitude and longitude.

**Step 5: Save to TSV File**

* Write the results into a new TSV file:  
  NER\_gazetteer.tsv inside the gazetteer folder.
* This file should have three columns:
  1. **placename**
  2. **latitude**
  3. **longitude**

**Step 6: Handle Missing Coordinates Manually**

* Open the NER\_gazetteer.tsv file and:
  + Search for any "NA" values in latitude/longitude.
  + For these, **manually look up coordinates on Google** (Maps or Search).
  + Add the correct coordinates into the file.

**Step 7: Update README File**

* In your project folder, open or create a README.md file.
* Add a list of **all place names** for which you had to look up coordinates manually.

Would you like help planning the README entry or checking how to geocode in Python?

4o

Follow up

Outline the process for geocoding places in Python using geopy library.

I need help structuring the 'README.md' with the manual lookups.

Look up best practices for writing a README with manual lookup info for project.

**2nd point**

* I asked for **clarification on how to approach geocoding** a list of place names programmatically.
* I asked for examples of how to **structure a script** that reads a TSV file, geocodes place names, and writes a new TSV output.
* I asked how to **handle missing geocoding results (NA)** and how to document manual lookup of coordinates.

At no point did I ask GenAI to write the complete script or solution for the task. All code in build\_gazetteer.ipynb was written and tested by me based on my understanding, with appropriate comments to explain the logic.

**Manual Geocoding:**

The following place names were not resolved automatically by the geocoding script. I looked up their coordinates manually using Google Search, as instructed, and included them in the NER\_gazetteer.tsv file:

These entries were marked as "NA" by the geocoding script and required manual intervention.

Morocco’s, Israel’s, Indonesia’s, Colombo’s Kollupitiya, Nairoukh, Shawawra, Mazzeh, Houthis’, Thameen Darby, Mercator, Darby’s Nakba, Yemen’s, Hamas’s, Lebanon’s Nabatieh, Yir’on, Israel’s Goren, Dahiyeh, Washington’s, Yemen’s coast, Lebanon’s, Khreis’s, north📍#Gaza, Iran’s Sistan-Baluchestan province, Iran’s, Salameh’s, Beirut’s, Beirut’s Shatila, France’s.

**3rd point.**

**1. Geocoding a List of Place Names Programmatically**

🧠 **What you asked:**  
*How do I approach geocoding a list of place names using code?*

✅ **How your script does it:**  
You pass a list of place names from a .tsv file to two geocoding services: Geonames (first) and Nominatim (fallback), then extract latitude and longitude.

💻 **Code Example:**

python

CopyEdit

coords = get\_coordinates\_geonames(place)

if not coords:

coords = get\_coordinates\_nominatim(place)

📤 **Sample Output in Console:**

makefile

CopyEdit

Processing: Gaza City

Processing: Tel Aviv

Processing: Rafah

Processing: Jabalia

**✅ 2. Reading from TSV → Geocoding → Writing Output to New TSV**

🧠 **What you asked:**  
*How should I structure a script that reads a TSV, geocodes names, and writes results to another TSV?*

✅ **How your script does it:**

* Reads from ner\_counts.tsv
* Stores each result in a DataFrame
* Writes the DataFrame to NER\_gazetteer.tsv

💻 **Code Example:**

python

CopyEdit

df = pd.read\_csv(input\_file, sep='\t') # Read TSV

gazetteer = pd.DataFrame(columns=['placename', 'latitude', 'longitude'])

gazetteer.loc[idx] = [place, coords['latitude'], coords['longitude']] # Add result

gazetteer.to\_csv(output\_file, sep='\t', index=False) # Write output

📄 **Input (ner\_counts.tsv) sample:**

nginx

CopyEdit

name count

Gaza City 15

Tel Aviv 9

Rafah 6

📄 **Output (NER\_gazetteer.tsv) sample:**

nginx

CopyEdit

placename latitude longitude

Gaza City 31.5 34.46667

Tel Aviv 32.06667 34.76667

Rafah 31.287 34.257

**✅ 3. Handling Missing Results & Manual Lookup**

🧠 **What you asked:**  
*How do I deal with missing geocoding results and document manual lookup?*

✅ **How your script does it:**

* If both geocoders fail, it writes 'NA' in latitude and longitude.
* Also saves those place names in manual\_lookup list and prints them out.

💻 **Code Example:**

python

CopyEdit

if coords:

gazetteer.loc[idx] = [place, coords['latitude'], coords['longitude']]

else:

gazetteer.loc[idx] = [place, 'NA', 'NA']

manual\_lookup.append(place)

📤 **Console Output:**

vbnet

CopyEdit

Processing: Al-Mawasi

Error with Geonames for Al-Mawasi: ...

Error with Nominatim for Al-Mawasi: ...

Manual lookup needed for:

Al-Mawasi

📄 **Output (NER\_gazetteer.tsv) for missing place:**

r

CopyEdit

placename latitude longitude

Al-Mawasi NA NA

📝 **README note example:**

We attempted to geocode all place names using the Geonames API and Nominatim as a fallback. For entries where no coordinates were found (marked as “NA”), we performed a manual lookup using Google Maps and updated the TSV accordingly.

Would you like a sample README.md file with this explanation included, formatted professionally?