AI usage

Regex map 4A

You: I need to load data from a TSV file but I'm not sure how to handle potential issues. What should I consider?

Deepseek: When working with tabular data files, think about what could go wrong during loading. The file might be empty or formatted differently than expected. How would you verify the data loaded correctly? What pandas function would let you inspect the first few rows?

You: Okay, I've loaded the data. Now I need to get coordinates for placenames. I heard about geocoding but it seems complex. Where should I start?

Deepseek: Geocoding does have several components. First, you'll need a geocoding service - there are several Python libraries for this. Consider how you might handle rate limits since these services often have usage restrictions. How would you structure the code to process each placename while respecting those limits?

### ****2. Geocoding with Rate Limiting****

**You:** The geocoding API is slow. How do I avoid hitting rate limits while processing hundreds of placenames?

**Deepseek:** You can use RateLimiter to add delays between requests. Here’s how you might set it up:

python

Copy

Download

from geopy.extra.rate\_limiter import RateLimiter

geocode = RateLimiter(geolocator.geocode, min\_delay\_seconds=1) # 1 request/sec

You: I'm getting locations but some fail. How should I handle missing coordinates?

Deepseek: That's a common issue with geocoding. Think about the data flow - after attempting to geocode all locations, you'll have some with coordinates and some without. What pandas operations could help you filter out the unsuccessful ones? How would this affect your visualization later?

You: For the visualization, I want to show points on a map with sizes based on frequency. What approach would work best?

Deepseek: There are several mapping libraries available. Consider what features you need - interactive plots, different projections, hover information. How would you represent both the geographical position and the count data visually? What parameters in a scatter plot would control the point sizes?

**You:** What if the input file isn’t a TSV or has missing columns? How can I check that before processing?

**Deepseek:** You can validate the structure early. For example, check if required columns exist:

python

Copy

Download

required\_columns = ["placename", "count"]

if not all(col in df.columns for col in required\_columns):

raise ValueError("Missing required columns in the input file.")

You: The map looks good but I want to customize the appearance. How can I improve the base map?

Deepseek: Think about what makes a map readable. The background colors, landmass visibility, and projection type all affect interpretation. Look into the documentation for your chosen visualization library - what parameters control these aspects? How would you adjust them to make the data stand out?

You: Finally, how can I make sure the whole process is robust from start to finish?

Deepseek: Consider the entire pipeline as separate components. Where are the potential failure points? How would you add checks between steps? What feedback would help someone using this script understand if something went wrong? Think about both technical robustness and user communication.

**You:** How can I adjust the map’s style and point sizes dynamically?

**Deepseek:** The scatter\_geo function accepts parameters like size\_max for scaling. Example:

python

Copy

Download

fig = px.scatter\_geo(

size\_max=20, # Controls max point size

projection="natural earth" # Try "mercator" or "orthographic"

NER Map 4B

Here’s the conversation with all symbols removed:

You: I have two TSV files one with place mentions and another with coordinates How do I combine them correctly

Deepseek: Merging datasets requires matching column names First inspect the columns in both files to identify the common key likely the place name If the column names differ youll need to rename them How would you ensure the merge doesnt drop unmatched entries

You: The merge works but some rows disappear Why

Deepseek: By default merging keeps only rows with matches in both datasets If you want to keep all places even those missing coordinates youd use how equals left But here dropping unmatchable rows might be intentional why visualize data you cant map

You: How can I verify the merge worked as expected

Deepseek: After merging check the row count and sample the output For example print the number of merged entries and display the first few rows Are latitude and longitude populated Are counts numeric

You: Some coordinates are invalid How do I clean them

Deepseek: Force numeric conversion and drop missing values For example convert latitude and longitude columns to numeric types and remove rows where these values are missing What else might you check Extreme values like latitude 900

You: The map points are too small How do I adjust their size

Deepseek: The size parameter scales points by your data column but size max sets the upper limit Try increasing size max to 25 or higher How might you normalize sizes if counts vary wildly

You: Can I customize the color scheme

Deepseek: Plotly supports color scales like Viridis Plasma or custom lists For example use a sequential color scale like Plasma Should higher counts be darker or brighter Test sequential or diverging palettes

You: How do I save this map for sharing

Deepseek: Export to HTML for interactivity For example write the figure to an HTML file For static images use write image which requires additional libraries Which format suits your needs

You: The map looks cluttered Any way to declutter

Deepseek: Consider clustering nearby points or filtering low counts Alternatively adjust zoom to focus on dense regions Whats the priority showing all data or readability

You: What if the input files have typos in column names

Deepseek: Always validate column names before merging Print the columns from both datasets and rename mismatched ones programmatically How would you handle case sensitivity or whitespace differences

You: How do I handle places with the same name but different coordinates

Deepseek: Ambiguous placenames require disambiguation You might need to clean the data beforehand or merge on additional columns like country codes What context is missing from the raw data

You: The color scale doesnt reflect the data range well Any fixes

Deepseek: Adjust the color axis bounds using color continuous scale parameters You could set a fixed range or normalize the counts What visual cues best represent your data distribution

You: Can I add interactive elements like tooltips

Deepseek: Hover elements are controlled by hover name and hover data parameters Customize tooltips to show additional fields like country or count How much detail is useful without overwhelming users

You: How do I test this script on different datasets

Deepseek: Parameterize the input file paths and add validation checks For example read filenames from command-line arguments or a config file What edge cases should you anticipate missing files invalid formats