**AI Documentation**

**Prompts**

I’m a student of DH(Digital Humanities) at FAS and I am working on a project where I need to look at how certain word pairs appear over time using data in a table. I already know the simple stuff like opening files and grouping words into categories.

What I really want help with are the harder parts, like:

* How to pick data only from certain months and years when the date info is spread across separate columns.
* How to combine those separate month and year pieces into one date so I can use it easily.
* How to add up the counts of word pairs for each month and category.
* How to organize and label this information so it makes sense when shown.
* How to create a side-by-side bar chart that compares two groups over time.
* How to save this chart as something interactive I can open in a web browser.

Can you explain the steps I should follow to figure out these parts on my own? Please don’t give me the exact instructions or code — just guide me through what to think about and how to approach each challenge so I can try it myself.

done. next step

**🔹 Step 1: Load and Inspect Your CSV with pandas**

**🧠 Think:**

* What pandas function do you use to read a CSV file?
* Once it's loaded, how do you view the first few rows?

**📝 Try to write:**

1. The code to **import pandas**.
2. The code to **load your CSV file** into a DataFrame (let's call it df).
3. The code to **see the first 5 rows** of data.

👉 You try writing that first and share it here. Then we’ll clean the data together.

Great! Since you’ve loaded and inspected your data, let’s move on.

**🔹 Step 2: Clean the Data (Lowercase and Strip Whitespaces)**

**Why?**

* Lowercasing makes matching easier since “Peace Talks” and “peace talks” are treated the same.
* Stripping whitespace removes unwanted spaces around text that could break matching.

**Questions for you:**

* How can you convert all text in a pandas column (like 'ngram') to lowercase?
* How can you remove spaces before and after text in a pandas column?

**Hint:**

Pandas string methods use .str accessor, like .str.lower() or .str.strip().

Okay Got it

**🔹 Step 3: Combine Month and Year into a Date**

**Why?**

* We want to analyze frequency over time.
* It’s easier to group and plot data by a proper datetime format than by separate month/year strings.

**Questions:**

* How can you combine two columns (like 'month' and 'year') into one column?
* How do you convert a text date (e.g., “October 2023”) into a pandas datetime object?

**Hint:**

* You can combine two string columns using + operator (with a space in between).
* Use pd.to\_datetime() to parse the combined string into a datetime.

Umm, okay ill try to do this it. Please state nect step

**🔹 Step 4: Categorize 2-Grams into Themes**

**Why?**

* You want to separate the phrases into meaningful groups like **“Destruction”** and **“Harmony”**.
* This helps analyze and compare trends by theme.

**Questions:**

* How can you create a list of phrases for each theme?
* How can you check if each row’s 2-gram belongs to one of these theme lists?
* How can you assign a new column 'category' that labels each row based on these lists?

**Hint:**

* Use Python lists for each theme.
* Use apply() with a function to check membership and assign categories.

**Your task:**

1. Define two lists of 2-grams, e.g., category1\_phrases and category2\_phrases.
2. Write a function that takes a 2-gram, checks which list it’s in, and returns the category name.
3. Use .apply() to create a new 'category' column with the result.

if i have to apply 2 categories such as i have made one category as Destruction and given it value and another is Harmony and also putting values to it. Then?

Got it! You want to **assign categories (like “Destruction” and “Harmony”) and also associate some kind of value** with each category in your data. Let’s clarify this step by step:

**1. You already have a 'category' column with values "Destruction" or "Harmony" for each row (or maybe None for others).**

**2. Now, if by “putting values” you mean:**

* **Using the existing count column** as the numeric value related to each category (frequency of the 2-gram),
* Or creating a new column with a numeric code/weight for categories (like Destruction = 1, Harmony = 2),

the approach differs.

**If you want to keep using the counts per category for grouping and summing (typical case):**

* You don’t need to create extra values for categories.
* Just group by 'date' and 'category' and sum the 'count'.

**Quick questions for you to decide:**

* Do you want to **sum the counts per category over time** (likely yes for frequency analysis)?
* Or do you want to assign **numeric codes/weights to categories** for something else?

Let me know, and I can guide you on the exact next step!

I practiced these earlier too in my class so, i did it right on the first place and you revived my memory. And the further steps I can do on my own. I will also look into my earlier codes practiced in class. Got it!

That’s fantastic! 🎉 I’m glad this helped you reconnect with what you already know — that’s the best feeling when learning.

If you want, I’m here anytime to help you troubleshoot or explore the visualization part or anything else related to your project. Just drop a message!

Good luck with your n-gram analysis and your Digital Humanities journey. You’ve got this! 🚀✨