**AI DOCUMENTATION**

**You said:**

Traceback (most recent call last): File "C:\Users\HAJI LAPTOP g55\Downloads\FASDH25-portfolio3\kamil-ahmad-n-gram-assignment\kamil-ahmad-1-gram-year-exploration.py", line 5, in <module> df = pd.read\_csv("data/dataframes/n-grams/1-gram/1-gram-year.csv") File "C:\Users\HAJI LAPTOP g55\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\io\parsers\readers.py", line 1026, in read\_csv return \_read(filepath\_or\_buffer, kwds) File "C:\Users\HAJI LAPTOP g55\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\io\parsers\readers.py", line 620, in \_read parser = TextFileReader(filepath\_or\_buffer, \*\*kwds) File "C:\Users\HAJI LAPTOP g55\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\io\parsers\readers.py", line 1620, in \_\_init\_\_ self.\_engine = self.\_make\_engine(f, self.engine) File "C:\Users\HAJI LAPTOP g55\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\io\parsers\readers.py", line 1880, in \_make\_engine self.handles = get\_handle( File "C:\Users\HAJI LAPTOP g55\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\io\common.py", line 873, in get\_handle handle = open( FileNotFoundError: [Errno 2] No such file or directory: 'data/dataframes/n-grams/1-gram/1-gram-year.csv'

**ChatGPT said:**

This error means Python **can’t find your CSV file** at the path you provided:

swift

CopyEdit

data/dataframes/n-grams/1-gram/1-gram-year.csv

**You said:**

What is the best way to analyze the 1-gram-year dataset for meaningful trends?

**ChatGPT said:**  
To analyze the 1-gram-year dataset effectively, start by cleaning the data. One of the first things you’ll want to do is remove common stop words like "the," "and," and "is," because they appear frequently but don’t carry much meaning. Once cleaned, you can group the data by year and sum the frequencies of meaningful content words. Plotting a line chart or bar chart of the most frequent words over time can help you visualize trends and identify spikes during key events.

**You said:**

should i remove the stop words

**ChatGPT said:**

Yes, you **should remove stop words** — especially for **1-gram** analysis — because:

**Why remove stop words?**

* **Stop words** like "the," "is," "and," "of" appear very frequently but **don’t carry meaningful content**.
* Including them can **distract from more informative words** like *Israel*, *Gaza*, *Palestinian*, etc.
* Your charts and frequency counts will be **more insightful and focused** if they highlight content-specific vocabulary.

**But also remember:**

* For **2-gram** and **3-gram**, be careful — some meaningful phrases (e.g., "State of Palestine") **might include stop words**. So you can **filter more selectively** or check if the phrase still makes sense after removal.
* For **critical analysis**, always mention that you applied stop word filtering and briefly explain why.

So in summary:  
✅ Yes for 1-gram  
🤔 Selectively for 2-gram and 3-gram

**You said:**

**I removed stop words but now my chart is empty. Why?**

python

CopyEdit

stop\_words = set(["the", "and", "of", "to"])

df = df[~df["1-gram"].isin(stop\_words)]

top\_words = df.groupby("1-gram")["count"].sum().nlargest(10)

df\_top = df[df["1-gram"].isin(top\_words)]

**ChatGPT said:**  
Good attempt! The issue is in the last line. You're trying to filter the DataFrame using a Series of frequencies instead of a list of words. You should use:

python

CopyEdit

df\_top = df[df["1-gram"].isin(top\_words.index)]

top\_words.index contains the actual word strings, not the frequency values. That should fix your empty chart problem.

**You said:**

= RESTART: C:\Users\HAJI LAPTOP g55\Downloads\FASDH25-portfolio3\data\dataframes\n-grams\kamil-ahmad-1-gram-year-exploration.py year 1-gram count-sum count-mean 0 2023 the 94046 42.135305 1 2024 the 47119 45.263208 2 2023 of 41879 18.915537 3 2023 to 38874 18.511429 4 2023 and 38431 18.068171 Index(['year', '1-gram', 'count-sum', 'count-mean'], dtype='object') 43225 [2023 2024 2021 2022 2017] Top 10 most frequent 1-grams: 1-gram the 188145 of 83497 to 78486 and 76984 in 73915 a 55489 on 33660 that 32070 s 30517 israel 29688 Name: count-sum, dtype: int64 Matplotlib is building the font cache; this may take a moment. Traceback (most recent call last): File "C:\Users\HAJI LAPTOP g55\Downloads\FASDH25-portfolio3\data\dataframes\n-grams\kamil-ahmad-1-gram-year-exploration.py", line 29, in <module> plt.title('Frequency of Top 1-Grams Over Time') NameError: name 'plt' is not defined

**ChatGPT said:**

You're very close! That last error:

NameError: name 'plt' is not defined

...means you used plt in your code but **forgot to import it**.