## Use of GenAI tools (ChatGPT) during the development of my Mini Project 3

I used AI as a tutor to help understand specific errors, clarify isolated steps, and explain code syntax. All AI assistance has been documented below in compliance with the academic honesty policy.

**Entry 0 in Exploration Script:** I learned that .copy() is used to avoid ambiguity when assigning new columns to filtered data. Now I understand when and why this warning occurs, and how to handle it responsibly.

**You said:**

I'm getting a SettingWithCopyWarning in Pandas when assigning a new column after filtering a DataFrame. What causes this warning and how can I avoid it safely?

**ChatGPT said:**

Great question! This warning appears when you're **working with a view rather than a copy of your DataFrame.** To avoid it, use **.copy()** after the filter so you're working with a new object.

For example: df\_filtered = df[condition].copy()

**You said:**

Do I need to use .copy() every time I filter?

**ChatGPT said:**

That’s a good query. Not necessarily. You only need .copy() if you plan to **modify the filtered DataFrame** (like adding a new column). If you're only reading or printing data, it's usually safe without it.

**Entry 1 in Exploration Script (and reused in Visualisation Script):** I used this technique to label articles by month. I now understand how to combine integer columns into date-like strings for grouping and visualization.

**You said:**

I want to create a new column that combines year and month into a single string like "2023-10". What's the cleanest way to do this in Pandas?

**ChatGPT said:**

Hmm. You can convert both columns to strings using **.astype(str)** and combine them with **"+"**.

Example: df["month\_year"] = df["year"].astype(str) + "-" + df["month"].astype(str)

**You said:**

Is this okay for plotting grouped bar charts later?

**ChatGPT said:**

Yes, string-formatted dates like "2023-10" work well for grouping and labeling bars. Just make sure to sort them correctly if needed.

**Entry 2 in Exploration Script (and reused in Visualisation Script):** I learned how to combine columns into a custom label for chart legends. This makes the data easier to interpret when visualizing.

**You said:**

I have four columns: topic\_1 to topic\_4. I want to combine them into a single string column like "bank, west, israeli, palestinian". What's a good way to do that?

**ChatGPT said:**

Use string concatenation with commas. Here's an example: df["Topic\_Label"] = df["topic\_1"] + ", " + df["topic\_2"] + ", " + df["topic\_3"] + ", " + df["topic\_4"]

**You said:**

Will this work even if some rows are missing keywords?

**ChatGPT said:**

If there are missing (NaN) values, it may result in NaN for the whole row. To handle that, you can fill NaNs before joining using .fillna("").

**Entry 3 in Exploration Script (and reused in Visualisation Script):** This was helpful for preparing data to visualize topic frequencies over time. I learned how to group and reshape data using chaining methods.

**You said:**

How can I group by two columns in Pandas (like topic and month), and count how many rows are in each group?

**ChatGPT said:**

Great question! You can use .groupby([col1, col2]).size().reset\_index(name="count") to count the number of rows in each group and get a clean DataFrame back.

**You said:**

What's the difference between .size() and .count()?

**ChatGPT said:**

In .size(), it counts total rows in each group (including NaNs). .count() counts only non-NaN values in each column.

**Entry 4 in Exploration Script**: This code helped me rank the most frequent topics for focused analysis. I now understand how to sort and slice grouped data for top-N selection.

**You said:**

I grouped my data by topic and got total article counts. Now I want to find the top 20 most frequent topics. What's a clear way to do that?

**ChatGPT said:**

Amazing! I can see you’re almost getting there. After grouping and summing, use .sort\_values(ascending=False).head(20) to get the top 20 rows by count.

**Reused Entries in Visualisation Script:**

Entry 1, Entry 2, Entry 3

These were all originally developed and explained in the exploration script, and simply reused in the visualisation step without getting help from ChatGPT.