The Data frames The data folder contains a series of csv files that can be loaded in Python using pandas. Each data frame has been produced using different text analysis methods: ● counting the lengths of the articles, with separate tables grouping lengths by year and month this is my task tell me the process without giving me the code

**ChatGPT said:**

Here’s a clear explanation of the **process** you should follow to complete this task of analyzing article lengths using pandas, **without code**:

**🔍 Step-by-Step Process**

**1. Access the Data Folder**

* Locate the folder that contains all the .csv files.
* Identify the files related to article length, especially those that summarize **lengths by year** and **lengths by month**.

**2. Load the CSV Files**

* Use pandas to read each .csv file into a separate **DataFrame**.
* Give each DataFrame a meaningful name based on its content (e.g., yearly, monthly).

**3. Understand the Structure**

* Look at the columns in each DataFrame to understand what data is available.
  + Typical columns might include: article ID, year, month, word count or character count.

**4. Clean the Data (if needed)**

* Check for any missing or inconsistent values in the article length columns.
* Ensure that columns like "year" or "month" are in the correct format (e.g., integers or datetime).

**5. Group and Summarize**

* For the yearly DataFrame:
  + Group the data by **year**, then calculate summary statistics (like average length, total length, etc.).
* For the monthly DataFrame:
  + Group the data by **both year and month** for a more granular view.

**6. Compare or Visualize (optional but helpful)**

* Once the grouped data is ready, you can compare trends across years or months.
* Consider plotting a graph or chart to visualize how article lengths change over time.

**7. Save or Use the Results**

* Save your grouped summaries to new files (if needed).
* Or use the grouped data for further text analysis or interpretation.

..................................................................................................

tell me the use of imprting os

**ChatGPT said:**

Sure! Here's a simple explanation of why and **when we import os in Python**:

**📦 What is os?**

* os stands for **Operating System**.
* It is a **built-in Python module** that lets you interact with your computer’s file system and operating system features.

**🧠 Why Import os?**

You import os to **perform tasks related to files and folders**, such as:

**1. Accessing Files in a Folder**

* Example: You want to get a list of all .csv files in a folder.
* Use: os.listdir("folder\_path") gives you all files in that folder.

**2. Building File Paths**

* Combines folder names and file names correctly across operating systems (Windows, macOS, Linux).
* Use: os.path.join("folder", "file.csv")

**3. Checking File or Folder Status**

* Example: Check if a file exists before opening it.
* Use: os.path.exists("file.csv")

**4. Looping Through Files**

* You can loop through all files in a directory and do something with each (like loading .csv files into pandas).

................................................................................................

Traceback (most recent call last): File "C:/Users/ATIYA/Downloads/Visualisation.py", line 13, in <module> df = pd.read\_csv("C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv") File "C:\Users\ATIYA\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\ATIYA\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\ATIYA\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\ATIYA\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\ATIYA\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: 'C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv' .. this is the error i am getting after using this code:import pandas as pd import os # path to the folder which contains the CSV files folder\_path = ('C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length') # loop through all the CSV files for file in os.listdir(folder\_path): if file.endswith('.csv'): file\_path = os.path.join(folder\_path, file) # load the csv file df = pd.read\_csv("C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv") print(f"\nProcessing file: {file}") # Check columns (to identify 'text' and 'date' columns) print("Columns in this file:", df.columns.tolist()) # Calculate article length if 'text' in df.columns: df['article\_length'] = df['text'].astype(str).str.len() # Character count else: print("No 'text' column found.") # Convert date column to datetime and extract year/month if 'date' in df.columns: df['date'] = pd.to\_datetime(df['date'], errors='coerce') df['year'] = df['date'].dt.year df['month'] = df['date'].dt.month else: print("No 'date' column found.") # Group by year and month grouped = df.groupby(['year', 'month'])['article\_length'].mean().reset\_index() # Save the grouped data output\_name = f"grouped\_lengths\_{file}" grouped.to\_csv(output\_name, index=False) print(f"Saved grouped data to: {output\_name}"). how to fix it

You're very close, Atiya! 🌟 The error is due to one key issue:

**❌ Problem:**

You're looping through files **correctly**, but still trying to load a **hardcoded file** named:

swift

CopyEdit

'C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv'

That file **doesn’t exist**, which is why you're getting:

FileNotFoundError: [Errno 2] No such file or directory

You **already have a loop** to go through all CSV files in the folder — so you should **read each file inside the loop** using file\_path.

Also, all the processing (reading, printing, calculating) should happen **inside the loop** — not outside.

**🔧 Corrected Step-by-Step Plan:**

1. Use the correct folder path.
2. Loop through files.
3. Read each file **inside** the loop using file\_path.
4. Then do the processing.

...........................................................................................

Processing file: length-year-month.csv Processing file: length-year.csv Processing file: length.csv Traceback (most recent call last): File "C:/Users/ATIYA/Downloads/Visualisation.py", line 15, in <module> df = pd.read\_csv("C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv") File "C:\Users\ATIYA\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\ATIYA\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\ATIYA\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\ATIYA\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\ATIYA\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: 'C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv' now showing this error

**ChatGPT said:**

Thanks for sharing the error. Let's fix this step by step.

**❗ Problem:**

You're still using this **hardcoded line** at the bottom of your script:

python

CopyEdit

df = pd.read\_csv("C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv")

This **line is NOT needed anymore** because:

* You're already loading each file dynamically inside the loop using file\_path.
* That specific file length.csv may **not even exist**, or may have a different name (e.g. length-year.csv or length-year-month.csv).

You **must delete** this line:

python

CopyEdit

df = pd.read\_csv("C:/Users/ATIYA/Downloads/FASDH25/FASDH25-portfolio3/data/dataframes/length.csv")

Why? Because your loop already does this:

python

CopyEdit

df = pd.read\_csv(file\_path)

That line is correct — it uses the current file from the loop.

Processing file: length-year-month.csv Processing file: length-year.csv Processing file: length.csv Columns in this file: ['year', 'month', 'day', 'length'] No 'text' column found. No 'date' column found. Traceback (most recent call last): File "C:/Users/ATIYA/Downloads/Visualisation.py", line 38, in <module> grouped = df.groupby(['year', 'month'])['article\_length'].mean().reset\_index() File "C:\Users\ATIYA\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\groupby\generic.py", line 1951, in \_\_getitem\_\_ return super().\_\_getitem\_\_(key) File "C:\Users\ATIYA\AppData\Local\Programs\Python\Python313\Lib\site-packages\pandas\core\base.py", line 244, in \_\_getitem\_\_ raise KeyError(f"Column not found: {key}") KeyError: 'Column not found: article\_length' now showing tis error

**ChatGPT said:**

Thanks for sharing the full message — this is actually **expected** based on the columns your file has.

Let me explain exactly what's going on and how to fix it.

**❗ Problem Summary:**

You're seeing this error:

vbnet

CopyEdit

KeyError: 'Column not found: article\_length'

Why?

Because **your CSV already has a column named length**, and you’re trying to use a column article\_length which was never created. That’s why this line fails:

python

CopyEdit

grouped = df.groupby(['year', 'month'])['article\_length'].mean().reset\_index()

**What's in your file?**

From the printout:

kotlin

CopyEdit

Columns in this file: ['year', 'month', 'day', 'length']

So:

* You don’t need to calculate article\_length — it's **already there** as length.
* You also don’t need to extract date columns (year, month) — they’re already separate.

..........................................................................................