

## Conference call analysis - Step 2

November 22, 2025

```
[1]: import pandas as pd
import swifter
import re
import string
import pickle
```

```
[2]: from swifter import set_defaults
set_defaults(
    npartitions=None,
    dask_threshold=1,
    scheduler="processes",
    progress_bar=True,
    progress_bar_desc=None,
    allow_dask_on_strings=True,
    force_parallel=False,
)
```

```
[12]: import os
import pandas as pd
import swifter
import re
import string
import pickle

def clean_text(text):
    text = str(text).lower()
    text = re.sub(f"[{re.escape(string.punctuation)}]", " ", text)
    text = re.sub(r'\s+', ' ', text)
    return text.strip()

dei_keywords_lowtype1 = [
    'diversity inclusion', 'equal opportunity', 'equal employment', 'DEI',
    'EEO', 'EEOC',
    'discrimination', 'pay equity', 'diverse workforce', 'gender diversity',
    'ethnicity diversity', 'inclusive culture', 'equal pay', 'fair treatment',
    'equitable practices', 'inclusive workplace', 'equal rights',
    'discrimination prevention', 'diverse representation', 'inclusive policies',
    'affirmative action', 'anti-discrimination laws', 'diversity training',
```

```

    'equal employment opportunity', 'gender pay gap',
    'minority representation', 'racial equality', 'sexual harassment',
    'prevention',
    'unconscious bias training', 'workforce diversity metrics',
    'diversity strategy', 'inclusion strategy', 'accessibility in the',
    'workplace',
    'cultural competence training', 'LGBT', 'LGBTQ', 'equal opportunity',
    'employer',
    'equity audits', 'equity audit', 'bias incident reporting', 'diverse',
    'leadership',
    'inclusive recruitment', 'equal employment',
    'diversity management', 'gender equality', 'age diversity',
    'racial discrimination', 'ethnic inclusion', 'inclusive policies',
    'workplace harassment prevention', 'diversity assessment',
    'gender bias', 'disability inclusion', 'inclusive benefits',
    'gender identity inclusion', 'minority support', 'cultural diversity',
    'initiatives',
    'inclusive training', 'gender mainstreaming', 'disability accommodation',
    'diversity benchmarking', 'sexual orientation inclusion', 'inclusive',
    'communication',
    'racial bias training', 'gender representation', 'inclusive performance',
    'reviews',
    'equity and inclusion strategy', 'inclusive organizational culture',
    'inclusive culture', 'diversity goals',
    'gender sensitivity training', 'inclusive leadership',
    'diversity and inclusion metrics', 'racial equity', 'inclusive HR',
    'diversity hiring initiatives', 'gender neutral', 'diversity programs',
    'workplace diversity', 'disability access', 'inclusive policy development',
    'diversity initiatives', 'diversity in leadership', 'inclusive language',
    'equal opportunities',
    'gender-based discrimination', 'women empowerment', 'female leadership',
    'women in leadership', 'women in the workplace', 'gender equity',
    'women in STEM', 'gender parity', 'gender-responsive', 'gender inclusivity',
    'racial justice', 'ethnic minority support', 'ethnic representation',
    'racial inclusivity', 'race-based discrimination', 'racial diversity',
    'ethnic diversity', 'minority empowerment', 'people of color',
    'black lives matter', 'racial disparities', 'anti-racism training',
    'ethnic equality'
]

dei_keywords_lowtype1_with_boundaries = [f'\\b{keyword}\\b' for keyword in
    dei_keywords_lowtype1]

dei_pattern_lowtype1 = re.compile('|'.
    join(dei_keywords_lowtype1_with_boundaries), re.IGNORECASE)

```

```

def get_dei_info(text, pattern):
    matches = set(re.findall(pattern, text))
    return bool(matches), ', '.join(matches)

input_dir = 'Batches'
output_dir = 'Batches_processed'

os.makedirs(output_dir, exist_ok=True)

for filename in os.listdir(input_dir):
    if filename.endswith('.pkl'):
        output_file_path = os.path.join(output_dir, filename)

        if os.path.exists(output_file_path):
            print(f"File {output_file_path} already exists. Skipping...")
            continue

        print(f"Processing {filename}...")
        batch_file_path = os.path.join(input_dir, filename)

        with open(batch_file_path, 'rb') as f:
            data = pickle.load(f)

            data['text_clean'] = data['componenttextpreview'].swifter.
↳ apply(clean_text)

            data[['DEI_indicator_lowtype1', 'DEI_lowtype1_content']] =
↳ data['text_clean'].swifter.progress_bar(enable=True).
↳ allow_dask_on_strings(enable=True).apply(lambda x: pd.Series(get_dei_info(x,
↳ dei_pattern_lowtype1)))

            with open(output_file_path, 'wb') as f:
                pickle.dump(data, f)

            print(f"Processed and saved {output_file_path}.")

```

```

File Batches_processed\batch_0.pkl already exists. Skipping...
File Batches_processed\batch_1.pkl already exists. Skipping...
File Batches_processed\batch_10.pkl already exists. Skipping...
File Batches_processed\batch_100.pkl already exists. Skipping...
File Batches_processed\batch_101.pkl already exists. Skipping...
File Batches_processed\batch_102.pkl already exists. Skipping...
File Batches_processed\batch_103.pkl already exists. Skipping...
File Batches_processed\batch_104.pkl already exists. Skipping...
File Batches_processed\batch_105.pkl already exists. Skipping...
File Batches_processed\batch_106.pkl already exists. Skipping...

```

[illegible]

[illegible]



Processed and saved Batches\_processed\batch\_233.pkl.  
Processing batch\_234.pkl...

Pandas Apply: 0%| | 0/47888 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_234.pkl.  
Processing batch\_235.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_235.pkl.  
Processing batch\_236.pkl...

Pandas Apply: 0%| | 0/71960 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_236.pkl.  
Processing batch\_237.pkl...

Pandas Apply: 0%| | 0/70936 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_237.pkl.  
Processing batch\_238.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_238.pkl.  
Processing batch\_239.pkl...

Pandas Apply: 0%| | 0/72209 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_239.pkl.  
File Batches\_processed\batch\_24.pkl already exists. Skipping...  
Processing batch\_240.pkl...

Pandas Apply: 0%| | 0/58084 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_240.pkl.  
Processing batch\_241.pkl...

Pandas Apply: 0%| | 0/58943 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_241.pkl.  
Processing batch\_242.pkl...

Pandas Apply: 0%| | 0/55450 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_242.pkl.  
Processing batch\_243.pkl...

Pandas Apply: 0%| | 0/54145 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_243.pkl.  
Processing batch\_244.pkl...

Pandas Apply: 0%| | 0/58767 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_244.pkl.  
Processing batch\_245.pkl...

Pandas Apply: 0%| | 0/62133 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_245.pkl.  
Processing batch\_246.pkl...

Pandas Apply: 0%| | 0/49735 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_246.pkl.  
Processing batch\_247.pkl...

Pandas Apply: 0%| | 0/59586 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_247.pkl.  
Processing batch\_248.pkl...

Pandas Apply: 0%| | 0/62984 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_248.pkl.  
Processing batch\_249.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_249.pkl.  
File Batches\_processed\batch\_25.pkl already exists. Skipping...  
Processing batch\_250.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_250.pkl.  
Processing batch\_251.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_251.pkl.  
Processing batch\_252.pkl...

Pandas Apply: 0%| | 0/60002 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_252.pkl.  
Processing batch\_253.pkl...

Pandas Apply: 0%| | 0/61530 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_253.pkl.  
Processing batch\_254.pkl...

Pandas Apply: 0%| | 0/54872 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_254.pkl.  
Processing batch\_255.pkl...

Pandas Apply: 0%| | 0/57340 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_255.pkl.  
Processing batch\_256.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_256.pkl.  
Processing batch\_257.pkl...

Pandas Apply: 0%| | 0/64072 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_257.pkl.  
Processing batch\_258.pkl...

Pandas Apply: 0%| | 0/45826 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_258.pkl.  
Processing batch\_259.pkl...

Pandas Apply: 0%| | 0/55907 [00:00<?, ?it/s]

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_259.pkl.  
 File Batches\_processed\batch\_26.pkl already exists. Skipping...  
 Processing batch\_260.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_260.pkl.  
 Processing batch\_261.pkl...

Pandas Apply: 0%| | 0/48074 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_261.pkl.  
 Processing batch\_262.pkl...

Pandas Apply: 0%| | 0/59531 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_262.pkl.  
 Processing batch\_263.pkl...

Pandas Apply: 0%| | 0/56110 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_263.pkl.  
 Processing batch\_264.pkl...

Pandas Apply: 0%| | 0/59844 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_264.pkl.  
 Processing batch\_265.pkl...

Pandas Apply: 0%| | 0/58179 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_265.pkl.  
 Processing batch\_266.pkl...

Pandas Apply: 0%| | 0/54673 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_266.pkl.  
 Processing batch\_267.pkl...

Pandas Apply: 0%| | 0/58410 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_267.pkl.  
Processing batch\_268.pkl...

Pandas Apply: 0%| | 0/63298 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_268.pkl.  
Processing batch\_269.pkl...

Pandas Apply: 0%| | 0/47759 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_269.pkl.  
File Batches\_processed\batch\_27.pkl already exists. Skipping...  
Processing batch\_270.pkl...

Pandas Apply: 0%| | 0/54362 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_270.pkl.  
Processing batch\_271.pkl...

Pandas Apply: 0%| | 0/57475 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_271.pkl.  
Processing batch\_272.pkl...

Pandas Apply: 0%| | 0/51928 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_272.pkl.  
Processing batch\_273.pkl...

Pandas Apply: 0%| | 0/54757 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_273.pkl.  
Processing batch\_274.pkl...

Pandas Apply: 0%| | 0/60368 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_274.pkl.  
Processing batch\_275.pkl...

Pandas Apply: 0%| | 0/52131 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_275.pkl.  
Processing batch\_276.pkl...

Pandas Apply: 0%| | 0/47842 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_276.pkl.  
Processing batch\_277.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_277.pkl.  
Processing batch\_278.pkl...

Pandas Apply: 0%| | 0/56484 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_278.pkl.  
Processing batch\_279.pkl...

Pandas Apply: 0%| | 0/47394 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_279.pkl.  
File Batches\_processed\batch\_28.pkl already exists. Skipping...  
Processing batch\_280.pkl...

Pandas Apply: 0%| | 0/49678 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_280.pkl.  
Processing batch\_281.pkl...

Pandas Apply: 0%| | 0/57663 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_281.pkl.  
Processing batch\_282.pkl...

Pandas Apply: 0%| | 0/50517 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_282.pkl.  
Processing batch\_283.pkl...

Pandas Apply: 0%| | 0/49799 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_283.pkl.  
Processing batch\_284.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_284.pkl.  
Processing batch\_285.pkl...

Pandas Apply: 0%| | 0/51895 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_285.pkl.  
Processing batch\_286.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_286.pkl.  
Processing batch\_287.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_287.pkl.  
Processing batch\_288.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_288.pkl.  
Processing batch\_289.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_289.pkl.  
File Batches\_processed\batch\_29.pkl already exists. Skipping...  
Processing batch\_290.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_290.pkl.  
Processing batch\_291.pkl...

Pandas Apply: 0%| | 0/48969 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_291.pkl.  
Processing batch\_292.pkl...

Pandas Apply: 0%| | 0/50088 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_292.pkl.  
Processing batch\_293.pkl...

Pandas Apply: 0%| | 0/49993 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_293.pkl.  
 Processing batch\_294.pkl...  
 Pandas Apply: 0%| | 0/44951 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_294.pkl.  
 Processing batch\_295.pkl...  
 Pandas Apply: 0%| | 0/52927 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_295.pkl.  
 Processing batch\_296.pkl...  
 Pandas Apply: 0%| | 0/50640 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_296.pkl.  
 Processing batch\_297.pkl...  
 Pandas Apply: 0%| | 0/45490 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_297.pkl.  
 Processing batch\_298.pkl...  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_298.pkl.  
 Processing batch\_299.pkl...  
 Pandas Apply: 0%| | 0/45854 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_299.pkl.  
 File Batches\_processed\batch\_3.pkl already exists. Skipping...  
 File Batches\_processed\batch\_30.pkl already exists. Skipping...  
 Processing batch\_300.pkl...  
 Pandas Apply: 0%| | 0/50403 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_300.pkl.  
 Processing batch\_301.pkl...  
 Pandas Apply: 0%| | 0/44385 [00:00<?, ?it/s]

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_301.pkl.  
 Processing batch\_302.pkl...

Pandas Apply: 0%| | 0/52821 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_302.pkl.  
 Processing batch\_303.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_303.pkl.  
 Processing batch\_304.pkl...

Pandas Apply: 0%| | 0/48958 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_304.pkl.  
 Processing batch\_305.pkl...

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_305.pkl.  
 Processing batch\_306.pkl...

Pandas Apply: 0%| | 0/47701 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_306.pkl.  
 Processing batch\_307.pkl...

Pandas Apply: 0%| | 0/56274 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_307.pkl.  
 Processing batch\_308.pkl...

Pandas Apply: 0%| | 0/44590 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_308.pkl.  
 Processing batch\_309.pkl...

Pandas Apply: 0%| | 0/45957 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_309.pkl.  
File Batches\_processed\batch\_31.pkl already exists. Skipping...  
Processing batch\_310.pkl...

Pandas Apply: 0%| | 0/49902 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_310.pkl.  
Processing batch\_311.pkl...

Pandas Apply: 0%| | 0/49310 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_311.pkl.  
Processing batch\_312.pkl...

Pandas Apply: 0%| | 0/43716 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_312.pkl.  
Processing batch\_313.pkl...

Pandas Apply: 0%| | 0/45672 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_313.pkl.  
Processing batch\_314.pkl...

Pandas Apply: 0%| | 0/45815 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_314.pkl.  
Processing batch\_315.pkl...

Pandas Apply: 0%| | 0/43401 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_315.pkl.  
Processing batch\_316.pkl...

Pandas Apply: 0%| | 0/48509 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_316.pkl.  
Processing batch\_317.pkl...

Pandas Apply: 0%| | 0/51296 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_317.pkl.  
Processing batch\_318.pkl...

Pandas Apply: 0%| | 0/45158 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_318.pkl.  
Processing batch\_319.pkl...

Pandas Apply: 0%| | 0/47587 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_319.pkl.  
File Batches\_processed\batch\_32.pkl already exists. Skipping...  
Processing batch\_320.pkl...

Pandas Apply: 0%| | 0/40630 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_320.pkl.  
Processing batch\_321.pkl...

Pandas Apply: 0%| | 0/46919 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_321.pkl.  
Processing batch\_322.pkl...

Pandas Apply: 0%| | 0/44481 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_322.pkl.  
Processing batch\_323.pkl...

Pandas Apply: 0%| | 0/50415 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_323.pkl.  
Processing batch\_324.pkl...

Pandas Apply: 0%| | 0/46322 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_324.pkl.  
Processing batch\_325.pkl...

Pandas Apply: 0%| | 0/43921 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
Processed and saved Batches\_processed\batch\_325.pkl.  
Processing batch\_326.pkl...

Pandas Apply: 0%| | 0/42082 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_326.pkl.  
Processing batch\_327.pkl...

Pandas Apply: 0%| | 0/46756 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_327.pkl.  
Processing batch\_328.pkl...

Pandas Apply: 0%| | 0/45183 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_328.pkl.  
Processing batch\_329.pkl...

Pandas Apply: 0%| | 0/44379 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_329.pkl.  
File Batches\_processed\batch\_33.pkl already exists. Skipping...  
Processing batch\_330.pkl...

Pandas Apply: 0%| | 0/40262 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_330.pkl.  
Processing batch\_331.pkl...

Pandas Apply: 0%| | 0/48477 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_331.pkl.  
Processing batch\_332.pkl...

Pandas Apply: 0%| | 0/44570 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_332.pkl.  
Processing batch\_333.pkl...

Pandas Apply: 0%| | 0/45310 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_333.pkl.  
Processing batch\_334.pkl...

Pandas Apply: 0%| | 0/48548 [00:00<?, ?it/s]  
Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_334.pkl.  
Processing batch\_335.pkl...

Pandas Apply: 0%| | 0/40344 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_335.pkl.  
 Processing batch\_336.pkl...  
 Pandas Apply: 0%| | 0/39528 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_336.pkl.  
 Processing batch\_337.pkl...  
 Pandas Apply: 0%| | 0/44706 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_337.pkl.  
 Processing batch\_338.pkl...  
 Pandas Apply: 0%| | 0/38996 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_338.pkl.  
 Processing batch\_339.pkl...  
 Pandas Apply: 0%| | 0/43885 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_339.pkl.  
 File Batches\_processed\batch\_34.pkl already exists. Skipping...  
 Processing batch\_340.pkl...  
 Pandas Apply: 0%| | 0/39764 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_340.pkl.  
 Processing batch\_341.pkl...  
 Pandas Apply: 0%| | 0/40191 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_341.pkl.  
 Processing batch\_342.pkl...  
 Pandas Apply: 0%| | 0/34972 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]  
 Processed and saved Batches\_processed\batch\_342.pkl.  
 Processing batch\_343.pkl...  
 Pandas Apply: 0%| | 0/35882 [00:00<?, ?it/s]  
 Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_343.pkl.  
Processing batch\_344.pkl...

Pandas Apply: 0%| | 0/36537 [00:00<?, ?it/s]

Dask Apply: 0%| | 0/10 [00:00<?, ?it/s]

Processed and saved Batches\_processed\batch\_344.pkl.

File Batches\_processed\batch\_35.pkl already exists. Skipping...  
File Batches\_processed\batch\_36.pkl already exists. Skipping...  
File Batches\_processed\batch\_37.pkl already exists. Skipping...  
File Batches\_processed\batch\_38.pkl already exists. Skipping...  
File Batches\_processed\batch\_39.pkl already exists. Skipping...  
File Batches\_processed\batch\_4.pkl already exists. Skipping...  
File Batches\_processed\batch\_40.pkl already exists. Skipping...  
File Batches\_processed\batch\_41.pkl already exists. Skipping...  
File Batches\_processed\batch\_42.pkl already exists. Skipping...  
File Batches\_processed\batch\_43.pkl already exists. Skipping...  
File Batches\_processed\batch\_44.pkl already exists. Skipping...  
File Batches\_processed\batch\_45.pkl already exists. Skipping...  
File Batches\_processed\batch\_46.pkl already exists. Skipping...  
File Batches\_processed\batch\_47.pkl already exists. Skipping...  
File Batches\_processed\batch\_48.pkl already exists. Skipping...  
File Batches\_processed\batch\_49.pkl already exists. Skipping...  
File Batches\_processed\batch\_5.pkl already exists. Skipping...  
File Batches\_processed\batch\_50.pkl already exists. Skipping...  
File Batches\_processed\batch\_51.pkl already exists. Skipping...  
File Batches\_processed\batch\_52.pkl already exists. Skipping...  
File Batches\_processed\batch\_53.pkl already exists. Skipping...  
File Batches\_processed\batch\_54.pkl already exists. Skipping...  
File Batches\_processed\batch\_55.pkl already exists. Skipping...  
File Batches\_processed\batch\_56.pkl already exists. Skipping...  
File Batches\_processed\batch\_57.pkl already exists. Skipping...  
File Batches\_processed\batch\_58.pkl already exists. Skipping...  
File Batches\_processed\batch\_59.pkl already exists. Skipping...  
File Batches\_processed\batch\_6.pkl already exists. Skipping...  
File Batches\_processed\batch\_60.pkl already exists. Skipping...  
File Batches\_processed\batch\_61.pkl already exists. Skipping...  
File Batches\_processed\batch\_62.pkl already exists. Skipping...  
File Batches\_processed\batch\_63.pkl already exists. Skipping...  
File Batches\_processed\batch\_64.pkl already exists. Skipping...  
File Batches\_processed\batch\_65.pkl already exists. Skipping...  
File Batches\_processed\batch\_66.pkl already exists. Skipping...  
File Batches\_processed\batch\_67.pkl already exists. Skipping...  
File Batches\_processed\batch\_68.pkl already exists. Skipping...  
File Batches\_processed\batch\_69.pkl already exists. Skipping...  
File Batches\_processed\batch\_7.pkl already exists. Skipping...  
File Batches\_processed\batch\_70.pkl already exists. Skipping...  
File Batches\_processed\batch\_71.pkl already exists. Skipping...

File Batches\_processed\batch\_72.pkl already exists. Skipping...  
File Batches\_processed\batch\_73.pkl already exists. Skipping...  
File Batches\_processed\batch\_74.pkl already exists. Skipping...  
File Batches\_processed\batch\_75.pkl already exists. Skipping...  
File Batches\_processed\batch\_76.pkl already exists. Skipping...  
File Batches\_processed\batch\_77.pkl already exists. Skipping...  
File Batches\_processed\batch\_78.pkl already exists. Skipping...  
File Batches\_processed\batch\_79.pkl already exists. Skipping...  
File Batches\_processed\batch\_8.pkl already exists. Skipping...  
File Batches\_processed\batch\_80.pkl already exists. Skipping...  
File Batches\_processed\batch\_81.pkl already exists. Skipping...  
File Batches\_processed\batch\_82.pkl already exists. Skipping...  
File Batches\_processed\batch\_83.pkl already exists. Skipping...  
File Batches\_processed\batch\_84.pkl already exists. Skipping...  
File Batches\_processed\batch\_85.pkl already exists. Skipping...  
File Batches\_processed\batch\_86.pkl already exists. Skipping...  
File Batches\_processed\batch\_87.pkl already exists. Skipping...  
File Batches\_processed\batch\_88.pkl already exists. Skipping...  
File Batches\_processed\batch\_89.pkl already exists. Skipping...  
File Batches\_processed\batch\_9.pkl already exists. Skipping...  
File Batches\_processed\batch\_90.pkl already exists. Skipping...  
File Batches\_processed\batch\_91.pkl already exists. Skipping...  
File Batches\_processed\batch\_92.pkl already exists. Skipping...  
File Batches\_processed\batch\_93.pkl already exists. Skipping...  
File Batches\_processed\batch\_94.pkl already exists. Skipping...  
File Batches\_processed\batch\_95.pkl already exists. Skipping...  
File Batches\_processed\batch\_96.pkl already exists. Skipping...  
File Batches\_processed\batch\_97.pkl already exists. Skipping...  
File Batches\_processed\batch\_98.pkl already exists. Skipping...  
File Batches\_processed\batch\_99.pkl already exists. Skipping...

```
[2]: import os
import pandas as pd
import pickle
from tqdm import tqdm

output_dir = 'Batches_processed'
merged_file_path = 'merged_data.pkl'

data_frames = []

files_to_process = [f for f in os.listdir(output_dir) if f.endswith('.pkl')]

for filename in tqdm(files_to_process, desc="Merging files"):
    if filename.endswith('.pkl'):
        file_path = os.path.join(output_dir, filename)
        print(f"Reading {file_path}...")
```

```

        with open(file_path, 'rb') as f:
            data = pickle.load(f)
            data = data[["companyid", "keydevid", "componenttext",
↪ "mostimportantdateutc", "speakertypename",
                        "DEI_indicator_lowtype1", "DEI_lowtype1_content"]]
            data_frames.append(data)

if data_frames:
    merged_data = pd.concat(data_frames, ignore_index=True)
    print(f"Merged {len(data_frames)} files.")

    with open(merged_file_path, 'wb') as f:
        pickle.dump(merged_data, f)
    print(f"Saved merged data to {merged_file_path}.")
else:
    print("No files to merge.")

```

```

Merging files:  0%|
| 0/345 [00:00<?, ?it/s]

Reading Batches_processed_new\batch_0.pkl...

Merging files:  1%|
| 2/345 [00:00<01:07,  5.10it/s]

Reading Batches_processed_new\batch_1.pkl...
Reading Batches_processed_new\batch_10.pkl...

Merging files:  1%|
| 4/345 [00:00<01:04,  5.28it/s]

Reading Batches_processed_new\batch_100.pkl...

Merging files:  1%|
| 5/345 [00:00<01:05,  5.21it/s]

Reading Batches_processed_new\batch_101.pkl...

Merging files:  2%|
| 6/345 [00:01<01:01,  5.48it/s]

Reading Batches_processed_new\batch_102.pkl...
Reading Batches_processed_new\batch_103.pkl...

Merging files:  2%|
| 8/345 [00:01<01:00,  5.58it/s]

Reading Batches_processed_new\batch_104.pkl...
Reading Batches_processed_new\batch_105.pkl...

Merging files:  3%|
| 10/345 [00:01<01:01,  5.46it/s]

```

Reading Batches\_processed\_new\batch\_106.pkl...  
Reading Batches\_processed\_new\batch\_107.pkl...  
Merging files: 3%|  
| 12/345 [00:02<01:05, 5.12it/s]  
  
Reading Batches\_processed\_new\batch\_108.pkl...  
Reading Batches\_processed\_new\batch\_109.pkl...  
Merging files: 4%|  
| 14/345 [00:02<01:02, 5.27it/s]  
  
Reading Batches\_processed\_new\batch\_11.pkl...  
Reading Batches\_processed\_new\batch\_110.pkl...  
Merging files: 5%|  
| 16/345 [00:03<01:03, 5.19it/s]  
  
Reading Batches\_processed\_new\batch\_111.pkl...  
Merging files: 5%|  
| 17/345 [00:03<01:02, 5.23it/s]  
  
Reading Batches\_processed\_new\batch\_112.pkl...  
Merging files: 5%|  
| 18/345 [00:03<00:59, 5.50it/s]  
  
Reading Batches\_processed\_new\batch\_113.pkl...  
Reading Batches\_processed\_new\batch\_114.pkl...  
Merging files: 6%|  
| 20/345 [00:03<00:54, 5.97it/s]  
  
Reading Batches\_processed\_new\batch\_115.pkl...  
Reading Batches\_processed\_new\batch\_116.pkl...  
Merging files: 6%|  
| 22/345 [00:04<00:51, 6.33it/s]  
  
Reading Batches\_processed\_new\batch\_117.pkl...  
Reading Batches\_processed\_new\batch\_118.pkl...  
Merging files: 7%|  
| 24/345 [00:04<00:49, 6.46it/s]  
  
Reading Batches\_processed\_new\batch\_119.pkl...  
Reading Batches\_processed\_new\batch\_12.pkl...  
Merging files: 8%|  
| 26/345 [00:04<00:50, 6.31it/s]  
  
Reading Batches\_processed\_new\batch\_120.pkl...  
Reading Batches\_processed\_new\batch\_121.pkl...  
Merging files: 8%|  
| 28/345 [00:04<00:51, 6.20it/s]

Reading Batches\_processed\_new\batch\_122.pkl...  
Reading Batches\_processed\_new\batch\_123.pkl...  
Merging files: 9%|  
| 30/345 [00:05<00:50, 6.25it/s]  
  
Reading Batches\_processed\_new\batch\_124.pkl...  
Reading Batches\_processed\_new\batch\_125.pkl...  
Merging files: 9%|  
| 32/345 [00:05<00:48, 6.47it/s]  
  
Reading Batches\_processed\_new\batch\_126.pkl...  
Reading Batches\_processed\_new\batch\_127.pkl...  
Merging files: 10%|  
| 34/345 [00:05<00:49, 6.33it/s]  
  
Reading Batches\_processed\_new\batch\_128.pkl...  
Reading Batches\_processed\_new\batch\_129.pkl...  
Merging files: 10%|  
| 36/345 [00:06<00:49, 6.19it/s]  
  
Reading Batches\_processed\_new\batch\_13.pkl...  
Reading Batches\_processed\_new\batch\_130.pkl...  
Merging files: 11%|  
| 38/345 [00:06<00:50, 6.04it/s]  
  
Reading Batches\_processed\_new\batch\_131.pkl...  
Reading Batches\_processed\_new\batch\_132.pkl...  
Merging files: 12%|  
| 40/345 [00:06<00:51, 5.88it/s]  
  
Reading Batches\_processed\_new\batch\_133.pkl...  
Reading Batches\_processed\_new\batch\_134.pkl...  
Merging files: 12%|  
| 42/345 [00:07<00:50, 6.04it/s]  
  
Reading Batches\_processed\_new\batch\_135.pkl...  
Reading Batches\_processed\_new\batch\_136.pkl...  
Merging files: 13%|  
| 44/345 [00:07<00:49, 6.06it/s]  
  
Reading Batches\_processed\_new\batch\_137.pkl...  
Reading Batches\_processed\_new\batch\_138.pkl...  
Merging files: 13%|  
| 46/345 [00:07<00:46, 6.43it/s]  
  
Reading Batches\_processed\_new\batch\_139.pkl...  
Reading Batches\_processed\_new\batch\_14.pkl...

Merging files: 14%|  
| 48/345 [00:08<00:46, 6.37it/s]  
  
Reading Batches\_processed\_new\batch\_140.pkl...  
Reading Batches\_processed\_new\batch\_141.pkl...  
  
Merging files: 14%|  
| 50/345 [00:08<00:46, 6.37it/s]  
  
Reading Batches\_processed\_new\batch\_142.pkl...  
Reading Batches\_processed\_new\batch\_143.pkl...  
  
Merging files: 15%|  
| 52/345 [00:08<00:43, 6.67it/s]  
  
Reading Batches\_processed\_new\batch\_144.pkl...  
Reading Batches\_processed\_new\batch\_145.pkl...  
  
Merging files: 16%|  
| 54/345 [00:09<00:45, 6.46it/s]  
  
Reading Batches\_processed\_new\batch\_146.pkl...  
Reading Batches\_processed\_new\batch\_147.pkl...  
  
Merging files: 16%|  
| 56/345 [00:09<00:45, 6.40it/s]  
  
Reading Batches\_processed\_new\batch\_148.pkl...  
Reading Batches\_processed\_new\batch\_149.pkl...  
  
Merging files: 17%|  
| 58/345 [00:09<00:45, 6.26it/s]  
  
Reading Batches\_processed\_new\batch\_150.pkl...  
Reading Batches\_processed\_new\batch\_150.pkl...  
  
Merging files: 17%|  
| 60/345 [00:10<00:47, 5.97it/s]  
  
Reading Batches\_processed\_new\batch\_151.pkl...  
Reading Batches\_processed\_new\batch\_152.pkl...  
  
Merging files: 18%|  
| 62/345 [00:10<00:47, 5.93it/s]  
  
Reading Batches\_processed\_new\batch\_153.pkl...  
Reading Batches\_processed\_new\batch\_154.pkl...  
  
Merging files: 19%|  
| 64/345 [00:10<00:45, 6.12it/s]  
  
Reading Batches\_processed\_new\batch\_155.pkl...  
Reading Batches\_processed\_new\batch\_156.pkl...  
  
Merging files: 19%|  
| 66/345 [00:11<00:44, 6.27it/s]

Reading Batches\_processed\_new\batch\_157.pkl...  
Reading Batches\_processed\_new\batch\_158.pkl...  
Merging files: 20%|  
| 68/345 [00:11<00:41, 6.67it/s]  
  
Reading Batches\_processed\_new\batch\_159.pkl...  
Reading Batches\_processed\_new\batch\_160.pkl...  
Merging files: 20%|  
| 70/345 [00:11<00:42, 6.46it/s]  
  
Reading Batches\_processed\_new\batch\_161.pkl...  
Reading Batches\_processed\_new\batch\_162.pkl...  
Merging files: 21%|  
| 72/345 [00:12<00:43, 6.21it/s]  
  
Reading Batches\_processed\_new\batch\_163.pkl...  
Reading Batches\_processed\_new\batch\_164.pkl...  
Merging files: 21%|  
| 74/345 [00:12<00:41, 6.60it/s]  
  
Reading Batches\_processed\_new\batch\_165.pkl...  
Reading Batches\_processed\_new\batch\_166.pkl...  
Merging files: 22%|  
| 76/345 [00:12<00:39, 6.84it/s]  
  
Reading Batches\_processed\_new\batch\_167.pkl...  
Reading Batches\_processed\_new\batch\_168.pkl...  
Merging files: 23%|  
| 78/345 [00:12<00:39, 6.84it/s]  
  
Reading Batches\_processed\_new\batch\_169.pkl...  
Reading Batches\_processed\_new\batch\_170.pkl...  
Merging files: 23%|  
| 80/345 [00:13<00:41, 6.33it/s]  
  
Reading Batches\_processed\_new\batch\_171.pkl...  
Reading Batches\_processed\_new\batch\_172.pkl...  
Merging files: 24%|  
| 82/345 [00:13<00:39, 6.63it/s]  
  
Reading Batches\_processed\_new\batch\_173.pkl...  
Reading Batches\_processed\_new\batch\_174.pkl...  
Merging files: 24%|  
| 84/345 [00:13<00:37, 6.92it/s]

Merging files: 25%|  
| 86/345 [00:14<00:36, 7.12it/s]  
  
Reading Batches\_processed\_new\batch\_175.pkl...  
Reading Batches\_processed\_new\batch\_176.pkl...  
  
Merging files: 26%|  
| 88/345 [00:14<00:34, 7.37it/s]  
  
Reading Batches\_processed\_new\batch\_177.pkl...  
Reading Batches\_processed\_new\batch\_178.pkl...  
  
Merging files: 26%|  
| 89/345 [00:14<00:38, 6.71it/s]  
  
Reading Batches\_processed\_new\batch\_179.pkl...  
  
Merging files: 26%|  
| 90/345 [00:14<00:49, 5.16it/s]  
  
Reading Batches\_processed\_new\batch\_18.pkl...  
  
Merging files: 27%|  
| 92/345 [00:15<00:50, 5.00it/s]  
  
Reading Batches\_processed\_new\batch\_180.pkl...  
  
Merging files: 27%|  
| 93/345 [00:15<00:45, 5.53it/s]  
  
Reading Batches\_processed\_new\batch\_181.pkl...  
Reading Batches\_processed\_new\batch\_182.pkl...  
  
Merging files: 28%|  
| 95/345 [00:15<00:42, 5.86it/s]  
  
Reading Batches\_processed\_new\batch\_183.pkl...  
Reading Batches\_processed\_new\batch\_184.pkl...  
  
Merging files: 28%|  
| 97/345 [00:16<00:46, 5.30it/s]  
  
Reading Batches\_processed\_new\batch\_185.pkl...  
  
Merging files: 28%|  
| 98/345 [00:16<00:43, 5.70it/s]  
  
Reading Batches\_processed\_new\batch\_186.pkl...  
Reading Batches\_processed\_new\batch\_187.pkl...  
  
Merging files: 29%|  
| 100/345 [00:16<00:43, 5.65it/s]  
  
Reading Batches\_processed\_new\batch\_188.pkl...  
Reading Batches\_processed\_new\batch\_189.pkl...  
  
Merging files: 29%|  
| 101/345 [00:16<00:40, 6.01it/s]

Reading Batches\_processed\_new\batch\_19.pkl...  
Merging files: 30%|  
| 103/345 [00:17<00:41, 5.86it/s]  
Reading Batches\_processed\_new\batch\_190.pkl...  
Reading Batches\_processed\_new\batch\_191.pkl...  
Merging files: 30%|  
| 105/345 [00:17<00:41, 5.76it/s]  
Reading Batches\_processed\_new\batch\_192.pkl...  
Reading Batches\_processed\_new\batch\_193.pkl...  
Merging files: 31%|  
| 107/345 [00:17<00:40, 5.88it/s]  
Reading Batches\_processed\_new\batch\_194.pkl...  
Reading Batches\_processed\_new\batch\_195.pkl...  
Merging files: 32%|  
| 109/345 [00:18<00:36, 6.49it/s]  
Reading Batches\_processed\_new\batch\_196.pkl...  
Reading Batches\_processed\_new\batch\_197.pkl...  
Merging files: 32%|  
| 111/345 [00:18<00:36, 6.42it/s]  
Reading Batches\_processed\_new\batch\_198.pkl...  
Reading Batches\_processed\_new\batch\_199.pkl...  
Merging files: 33%|  
| 113/345 [00:18<00:38, 5.99it/s]  
Reading Batches\_processed\_new\batch\_2.pkl...  
Reading Batches\_processed\_new\batch\_20.pkl...  
Merging files: 33%|  
| 115/345 [00:19<00:38, 5.99it/s]  
Reading Batches\_processed\_new\batch\_200.pkl...  
Reading Batches\_processed\_new\batch\_201.pkl...  
Merging files: 34%|  
| 117/345 [00:19<00:35, 6.46it/s]  
Reading Batches\_processed\_new\batch\_202.pkl...  
Reading Batches\_processed\_new\batch\_203.pkl...  
Merging files: 34%|  
| 119/345 [00:19<00:33, 6.71it/s]  
Reading Batches\_processed\_new\batch\_204.pkl...  
Reading Batches\_processed\_new\batch\_205.pkl...

Merging files: 35%|  
| 121/345 [00:19<00:33, 6.63it/s]  
  
Reading Batches\_processed\_new\batch\_206.pkl...  
Reading Batches\_processed\_new\batch\_207.pkl...  
  
Merging files: 36%|  
| 123/345 [00:20<00:32, 6.76it/s]  
  
Reading Batches\_processed\_new\batch\_208.pkl...  
Reading Batches\_processed\_new\batch\_209.pkl...  
  
Merging files: 36%|  
| 125/345 [00:20<00:34, 6.29it/s]  
  
Reading Batches\_processed\_new\batch\_21.pkl...  
Reading Batches\_processed\_new\batch\_210.pkl...  
  
Merging files: 37%|  
| 127/345 [00:20<00:33, 6.45it/s]  
  
Reading Batches\_processed\_new\batch\_211.pkl...  
Reading Batches\_processed\_new\batch\_212.pkl...  
  
Merging files: 37%|  
| 129/345 [00:21<00:36, 5.99it/s]  
  
Reading Batches\_processed\_new\batch\_213.pkl...  
  
Merging files: 38%|  
| 130/345 [00:21<00:35, 6.02it/s]  
  
Reading Batches\_processed\_new\batch\_214.pkl...  
Reading Batches\_processed\_new\batch\_215.pkl...  
  
Merging files: 38%|  
| 132/345 [00:21<00:33, 6.42it/s]  
  
Reading Batches\_processed\_new\batch\_216.pkl...  
Reading Batches\_processed\_new\batch\_217.pkl...  
  
Merging files: 39%|  
| 134/345 [00:22<00:33, 6.35it/s]  
  
Reading Batches\_processed\_new\batch\_218.pkl...  
Reading Batches\_processed\_new\batch\_219.pkl...  
  
Merging files: 39%|  
| 136/345 [00:22<00:34, 6.01it/s]  
  
Reading Batches\_processed\_new\batch\_22.pkl...  
  
Merging files: 40%|  
| 137/345 [00:22<00:33, 6.13it/s]  
  
Reading Batches\_processed\_new\batch\_220.pkl...  
Reading Batches\_processed\_new\batch\_221.pkl...

Merging files: 40%|  
| 139/345 [00:22<00:30, 6.71it/s]  
  
Reading Batches\_processed\_new\batch\_222.pkl...  
Reading Batches\_processed\_new\batch\_223.pkl...  
  
Merging files: 41%|  
| 141/345 [00:23<00:29, 7.03it/s]  
  
Reading Batches\_processed\_new\batch\_224.pkl...  
Reading Batches\_processed\_new\batch\_225.pkl...  
  
Merging files: 41%|  
| 143/345 [00:23<00:30, 6.61it/s]  
  
Reading Batches\_processed\_new\batch\_226.pkl...  
Reading Batches\_processed\_new\batch\_227.pkl...  
  
Merging files: 42%|  
| 145/345 [00:23<00:31, 6.35it/s]  
  
Reading Batches\_processed\_new\batch\_228.pkl...  
Reading Batches\_processed\_new\batch\_229.pkl...  
  
Merging files: 43%|  
| 147/345 [00:24<00:31, 6.31it/s]  
  
Reading Batches\_processed\_new\batch\_230.pkl...  
Reading Batches\_processed\_new\batch\_230.pkl...  
  
Merging files: 43%|  
| 149/345 [00:24<00:29, 6.72it/s]  
  
Reading Batches\_processed\_new\batch\_231.pkl...  
Reading Batches\_processed\_new\batch\_232.pkl...  
  
Merging files: 44%|  
| 151/345 [00:24<00:29, 6.65it/s]  
  
Reading Batches\_processed\_new\batch\_233.pkl...  
Reading Batches\_processed\_new\batch\_234.pkl...  
  
Merging files: 44%|  
| 152/345 [00:24<00:28, 6.72it/s]  
  
Reading Batches\_processed\_new\batch\_235.pkl...  
  
Merging files: 45%|  
| 154/345 [00:25<00:34, 5.54it/s]  
  
Reading Batches\_processed\_new\batch\_236.pkl...  
  
Merging files: 45%|  
| 155/345 [00:25<00:35, 5.35it/s]  
  
Reading Batches\_processed\_new\batch\_237.pkl...  
Reading Batches\_processed\_new\batch\_238.pkl...

Merging files: 46%|  
| 157/345 [00:25<00:39, 4.76it/s]  
Reading Batches\_processed\_new\batch\_239.pkl...  
Merging files: 46%|  
| 158/345 [00:26<00:36, 5.08it/s]  
Reading Batches\_processed\_new\batch\_24.pkl...  
Reading Batches\_processed\_new\batch\_240.pkl...  
Merging files: 46%|  
| 160/345 [00:26<00:33, 5.44it/s]  
Reading Batches\_processed\_new\batch\_241.pkl...  
Merging files: 47%|  
| 161/345 [00:26<00:32, 5.59it/s]  
Reading Batches\_processed\_new\batch\_242.pkl...  
Reading Batches\_processed\_new\batch\_243.pkl...  
Merging files: 47%|  
| 163/345 [00:26<00:31, 5.73it/s]  
Reading Batches\_processed\_new\batch\_244.pkl...  
Reading Batches\_processed\_new\batch\_245.pkl...  
Merging files: 48%|  
| 165/345 [00:27<00:30, 5.90it/s]  
Reading Batches\_processed\_new\batch\_246.pkl...  
Reading Batches\_processed\_new\batch\_247.pkl...  
Merging files: 48%|  
| 166/345 [00:27<00:33, 5.38it/s]  
Reading Batches\_processed\_new\batch\_248.pkl...  
Merging files: 49%|  
| 168/345 [00:27<00:32, 5.40it/s]  
Reading Batches\_processed\_new\batch\_249.pkl...  
Reading Batches\_processed\_new\batch\_25.pkl...  
Merging files: 49%|  
| 169/345 [00:28<00:34, 5.10it/s]  
Reading Batches\_processed\_new\batch\_250.pkl...  
Merging files: 49%|  
| 170/345 [00:28<00:37, 4.64it/s]  
Reading Batches\_processed\_new\batch\_251.pkl...  
Merging files: 50%|  
| 171/345 [00:28<00:40, 4.27it/s]

Reading Batches\_processed\_new\batch\_252.pkl...  
Merging files: 50%|  
| 172/345 [00:28<00:42, 4.08it/s]  
Reading Batches\_processed\_new\batch\_253.pkl...  
Merging files: 50%|  
| 174/345 [00:29<00:39, 4.36it/s]  
Reading Batches\_processed\_new\batch\_254.pkl...  
Merging files: 51%|  
| 175/345 [00:29<00:36, 4.69it/s]  
Reading Batches\_processed\_new\batch\_255.pkl...  
Reading Batches\_processed\_new\batch\_256.pkl...  
Merging files: 51%|  
| 177/345 [00:29<00:34, 4.93it/s]  
Reading Batches\_processed\_new\batch\_257.pkl...  
Merging files: 52%|  
| 178/345 [00:29<00:30, 5.45it/s]  
Reading Batches\_processed\_new\batch\_258.pkl...  
Reading Batches\_processed\_new\batch\_259.pkl...  
Merging files: 52%|  
| 179/345 [00:30<00:29, 5.57it/s]  
Reading Batches\_processed\_new\batch\_26.pkl...  
Merging files: 52%|  
| 181/345 [00:30<00:31, 5.25it/s]  
Reading Batches\_processed\_new\batch\_260.pkl...  
Merging files: 53%|  
| 182/345 [00:30<00:29, 5.57it/s]  
Reading Batches\_processed\_new\batch\_261.pkl...  
Reading Batches\_processed\_new\batch\_262.pkl...  
Merging files: 53%|  
| 184/345 [00:31<00:28, 5.69it/s]  
Reading Batches\_processed\_new\batch\_263.pkl...  
Reading Batches\_processed\_new\batch\_264.pkl...  
Merging files: 54%|  
| 186/345 [00:31<00:27, 5.71it/s]  
Reading Batches\_processed\_new\batch\_265.pkl...  
Reading Batches\_processed\_new\batch\_266.pkl...

Merging files: 54%|  
| 188/345 [00:31<00:27, 5.80it/s]  
  
Reading Batches\_processed\_new\batch\_267.pkl...  
Reading Batches\_processed\_new\batch\_268.pkl...  
  
Merging files: 55%|  
| 190/345 [00:32<00:25, 5.99it/s]  
  
Reading Batches\_processed\_new\batch\_269.pkl...  
Reading Batches\_processed\_new\batch\_270.pkl...  
  
Merging files: 56%|  
| 192/345 [00:32<00:26, 5.84it/s]  
  
Reading Batches\_processed\_new\batch\_271.pkl...  
Reading Batches\_processed\_new\batch\_272.pkl...  
  
Merging files: 56%|  
| 194/345 [00:32<00:25, 6.03it/s]  
  
Reading Batches\_processed\_new\batch\_273.pkl...  
Reading Batches\_processed\_new\batch\_274.pkl...  
  
Merging files: 57%|  
| 196/345 [00:33<00:25, 5.93it/s]  
  
Reading Batches\_processed\_new\batch\_275.pkl...  
Reading Batches\_processed\_new\batch\_276.pkl...  
  
Merging files: 57%|  
| 198/345 [00:33<00:23, 6.29it/s]  
  
Reading Batches\_processed\_new\batch\_277.pkl...  
Reading Batches\_processed\_new\batch\_278.pkl...  
  
Merging files: 58%|  
| 200/345 [00:33<00:23, 6.10it/s]  
  
Reading Batches\_processed\_new\batch\_279.pkl...  
Reading Batches\_processed\_new\batch\_280.pkl...  
  
Merging files: 59%|  
| 202/345 [00:34<00:23, 6.10it/s]  
  
Reading Batches\_processed\_new\batch\_281.pkl...  
Reading Batches\_processed\_new\batch\_282.pkl...  
  
Merging files: 59%|  
| 203/345 [00:34<00:22, 6.29it/s]  
  
Reading Batches\_processed\_new\batch\_283.pkl...  
Reading Batches\_processed\_new\batch\_284.pkl...  
  
Merging files: 59%|  
| 205/345 [00:34<00:22, 6.26it/s]

Reading Batches\_processed\_new\batch\_282.pkl...  
Reading Batches\_processed\_new\batch\_283.pkl...  
Merging files: 60%|  
| 207/345 [00:34<00:23, 5.98it/s]  
Reading Batches\_processed\_new\batch\_284.pkl...  
Merging files: 60%|  
| 208/345 [00:35<00:22, 6.01it/s]  
Reading Batches\_processed\_new\batch\_285.pkl...  
Reading Batches\_processed\_new\batch\_286.pkl...  
Merging files: 61%|  
| 210/345 [00:35<00:23, 5.83it/s]  
Reading Batches\_processed\_new\batch\_287.pkl...  
Reading Batches\_processed\_new\batch\_288.pkl...  
Merging files: 61%|  
| 212/345 [00:35<00:22, 5.86it/s]  
Reading Batches\_processed\_new\batch\_289.pkl...  
Reading Batches\_processed\_new\batch\_29.pkl...  
Merging files: 62%|  
| 214/345 [00:36<00:21, 6.03it/s]  
Reading Batches\_processed\_new\batch\_290.pkl...  
Reading Batches\_processed\_new\batch\_291.pkl...  
Merging files: 63%|  
| 216/345 [00:36<00:20, 6.32it/s]  
Reading Batches\_processed\_new\batch\_292.pkl...  
Reading Batches\_processed\_new\batch\_293.pkl...  
Merging files: 63%|  
| 218/345 [00:36<00:19, 6.64it/s]  
Reading Batches\_processed\_new\batch\_294.pkl...  
Reading Batches\_processed\_new\batch\_295.pkl...  
Merging files: 64%|  
| 220/345 [00:36<00:19, 6.53it/s]  
Reading Batches\_processed\_new\batch\_296.pkl...  
Reading Batches\_processed\_new\batch\_297.pkl...  
Merging files: 64%|  
| 222/345 [00:37<00:19, 6.47it/s]  
Reading Batches\_processed\_new\batch\_298.pkl...  
Reading Batches\_processed\_new\batch\_299.pkl...

Merging files: 65%|  
| 224/345 [00:37<00:19, 6.11it/s]  
Reading Batches\_processed\_new\batch\_3.pkl...  
Merging files: 65%|  
| 225/345 [00:37<00:19, 6.01it/s]  
Reading Batches\_processed\_new\batch\_30.pkl...  
Reading Batches\_processed\_new\batch\_300.pkl...  
Merging files: 66%|  
| 227/345 [00:38<00:17, 6.59it/s]  
Reading Batches\_processed\_new\batch\_301.pkl...  
Reading Batches\_processed\_new\batch\_302.pkl...  
Merging files: 66%|  
| 229/345 [00:38<00:17, 6.54it/s]  
Reading Batches\_processed\_new\batch\_303.pkl...  
Reading Batches\_processed\_new\batch\_304.pkl...  
Merging files: 67%|  
| 231/345 [00:38<00:17, 6.62it/s]  
Reading Batches\_processed\_new\batch\_305.pkl...  
Reading Batches\_processed\_new\batch\_306.pkl...  
Merging files: 68%|  
| 233/345 [00:38<00:17, 6.32it/s]  
Reading Batches\_processed\_new\batch\_307.pkl...  
Reading Batches\_processed\_new\batch\_308.pkl...  
Merging files: 68%|  
| 235/345 [00:39<00:16, 6.77it/s]  
Reading Batches\_processed\_new\batch\_309.pkl...  
Reading Batches\_processed\_new\batch\_31.pkl...  
Merging files: 69%|  
| 237/345 [00:39<00:19, 5.67it/s]  
Reading Batches\_processed\_new\batch\_310.pkl...  
Merging files: 69%|  
| 238/345 [00:39<00:18, 5.93it/s]  
Reading Batches\_processed\_new\batch\_311.pkl...  
Reading Batches\_processed\_new\batch\_312.pkl...  
Merging files: 70%|  
| 240/345 [00:40<00:16, 6.54it/s]  
Reading Batches\_processed\_new\batch\_313.pkl...  
Reading Batches\_processed\_new\batch\_314.pkl...

Merging files: 70%|  
| 242/345 [00:40<00:14, 6.95it/s]  
  
Reading Batches\_processed\_new\batch\_315.pkl...  
Reading Batches\_processed\_new\batch\_316.pkl...  
  
Merging files: 71%|  
| 244/345 [00:40<00:14, 6.78it/s]  
  
Reading Batches\_processed\_new\batch\_317.pkl...  
Reading Batches\_processed\_new\batch\_318.pkl...  
  
Merging files: 71%|  
| 246/345 [00:40<00:14, 6.90it/s]  
  
Reading Batches\_processed\_new\batch\_319.pkl...  
Reading Batches\_processed\_new\batch\_320.pkl...  
  
Merging files: 72%|  
| 248/345 [00:41<00:14, 6.69it/s]  
  
Reading Batches\_processed\_new\batch\_320.pkl...  
Reading Batches\_processed\_new\batch\_321.pkl...  
  
Merging files: 72%|  
| 250/345 [00:41<00:13, 6.99it/s]  
  
Reading Batches\_processed\_new\batch\_322.pkl...  
Reading Batches\_processed\_new\batch\_323.pkl...  
  
Merging files: 73%|  
| 252/345 [00:41<00:13, 6.83it/s]  
  
Reading Batches\_processed\_new\batch\_324.pkl...  
Reading Batches\_processed\_new\batch\_325.pkl...  
  
Merging files: 74%|  
| 254/345 [00:42<00:12, 7.27it/s]  
  
Reading Batches\_processed\_new\batch\_326.pkl...  
Reading Batches\_processed\_new\batch\_327.pkl...  
  
Merging files: 74%|  
| 256/345 [00:42<00:12, 7.18it/s]  
  
Reading Batches\_processed\_new\batch\_328.pkl...  
Reading Batches\_processed\_new\batch\_329.pkl...  
  
Merging files: 75%|  
| 258/345 [00:42<00:13, 6.50it/s]  
  
Reading Batches\_processed\_new\batch\_330.pkl...  
  
Merging files: 75%|  
| 259/345 [00:42<00:12, 6.88it/s]

Reading Batches\_processed\_new\batch\_330.pkl...  
Reading Batches\_processed\_new\batch\_331.pkl...  
Merging files: 76%|  
| 261/345 [00:43<00:12, 6.89it/s]  
  
Reading Batches\_processed\_new\batch\_332.pkl...  
Reading Batches\_processed\_new\batch\_333.pkl...  
Merging files: 76%|  
| 263/345 [00:43<00:12, 6.72it/s]  
  
Reading Batches\_processed\_new\batch\_334.pkl...  
Reading Batches\_processed\_new\batch\_335.pkl...  
Merging files: 77%|  
| 265/345 [00:43<00:10, 7.41it/s]  
  
Reading Batches\_processed\_new\batch\_336.pkl...  
Reading Batches\_processed\_new\batch\_337.pkl...  
Merging files: 77%|  
| 267/345 [00:43<00:10, 7.58it/s]  
  
Reading Batches\_processed\_new\batch\_338.pkl...  
Reading Batches\_processed\_new\batch\_339.pkl...  
Merging files: 78%|  
| 269/345 [00:44<00:11, 6.61it/s]  
  
Reading Batches\_processed\_new\batch\_340.pkl...  
Merging files: 78%|  
| 270/345 [00:44<00:11, 6.80it/s]  
  
Reading Batches\_processed\_new\batch\_340.pkl...  
Reading Batches\_processed\_new\batch\_341.pkl...  
Merging files: 79%|  
| 272/345 [00:44<00:10, 7.27it/s]  
  
Reading Batches\_processed\_new\batch\_342.pkl...  
Reading Batches\_processed\_new\batch\_343.pkl...  
Merging files: 79%|  
| 274/345 [00:44<00:09, 7.50it/s]  
  
Reading Batches\_processed\_new\batch\_344.pkl...  
Reading Batches\_processed\_new\batch\_345.pkl...  
Merging files: 80%|  
| 275/345 [00:45<00:10, 6.63it/s]  
  
Reading Batches\_processed\_new\batch\_346.pkl...  
Merging files: 80%|  
| 277/345 [00:45<00:12, 5.53it/s]

Reading Batches\_processed\_new\batch\_37.pkl...  
Merging files: 81%|  
| 278/345 [00:45<00:12, 5.40it/s]  
Reading Batches\_processed\_new\batch\_38.pkl...  
Reading Batches\_processed\_new\batch\_39.pkl...  
Merging files: 81%|  
| 280/345 [00:46<00:12, 5.21it/s]  
Reading Batches\_processed\_new\batch\_4.pkl...  
Reading Batches\_processed\_new\batch\_40.pkl...  
Merging files: 82%|  
| 282/345 [00:46<00:12, 5.14it/s]  
Reading Batches\_processed\_new\batch\_41.pkl...  
Merging files: 82%|  
| 283/345 [00:46<00:12, 5.10it/s]  
Reading Batches\_processed\_new\batch\_42.pkl...  
Merging files: 82%|  
| 284/345 [00:46<00:11, 5.32it/s]  
Reading Batches\_processed\_new\batch\_43.pkl...  
Reading Batches\_processed\_new\batch\_44.pkl...  
Merging files: 83%|  
| 286/345 [00:47<00:10, 5.62it/s]  
Reading Batches\_processed\_new\batch\_45.pkl...  
Reading Batches\_processed\_new\batch\_46.pkl...  
Merging files: 83%|  
| 288/345 [00:47<00:10, 5.69it/s]  
Reading Batches\_processed\_new\batch\_47.pkl...  
Reading Batches\_processed\_new\batch\_48.pkl...  
Merging files: 84%|  
| 290/345 [00:47<00:10, 5.34it/s]  
Reading Batches\_processed\_new\batch\_49.pkl...  
Merging files: 84%|  
| 291/345 [00:48<00:09, 5.43it/s]  
Reading Batches\_processed\_new\batch\_5.pkl...  
Reading Batches\_processed\_new\batch\_50.pkl...  
Merging files: 85%|  
| 293/345 [00:48<00:10, 5.08it/s]  
Reading Batches\_processed\_new\batch\_51.pkl...  
Reading Batches\_processed\_new\batch\_52.pkl...

Merging files: 86%|  
| 295/345 [00:49<00:10, 4.92it/s]  
  
Reading Batches\_processed\_new\batch\_53.pkl...  
Reading Batches\_processed\_new\batch\_54.pkl...  
  
Merging files: 86%|  
| 297/345 [00:49<00:09, 4.95it/s]  
  
Reading Batches\_processed\_new\batch\_55.pkl...  
  
Merging files: 86%|  
| 298/345 [00:49<00:09, 5.05it/s]  
  
Reading Batches\_processed\_new\batch\_56.pkl...  
  
Merging files: 87%|  
| 299/345 [00:49<00:09, 5.08it/s]  
  
Reading Batches\_processed\_new\batch\_57.pkl...  
  
Merging files: 87%|  
| 300/345 [00:49<00:08, 5.22it/s]  
  
Reading Batches\_processed\_new\batch\_58.pkl...  
Reading Batches\_processed\_new\batch\_59.pkl...  
  
Merging files: 88%|  
| 302/345 [00:50<00:08, 5.15it/s]  
  
Reading Batches\_processed\_new\batch\_60.pkl...  
  
Merging files: 88%|  
| 303/345 [00:50<00:07, 5.34it/s]  
  
Reading Batches\_processed\_new\batch\_61.pkl...  
  
Merging files: 88%|  
| 304/345 [00:50<00:07, 5.36it/s]  
  
Reading Batches\_processed\_new\batch\_62.pkl...  
  
Merging files: 89%|  
| 306/345 [00:51<00:07, 5.33it/s]  
  
Reading Batches\_processed\_new\batch\_63.pkl...  
  
Merging files: 89%|  
| 307/345 [00:51<00:07, 5.26it/s]  
  
Reading Batches\_processed\_new\batch\_64.pkl...  
  
Merging files: 89%|  
| 308/345 [00:51<00:07, 5.26it/s]  
  
Reading Batches\_processed\_new\batch\_65.pkl...

```

Merging files: 90%|
| 309/345 [00:51<00:06, 5.34it/s]

Reading Batches_processed_new\batch_66.pkl...

Merging files: 90%|
| 310/345 [00:51<00:06, 5.34it/s]

Reading Batches_processed_new\batch_67.pkl...

Merging files: 90%|
| 311/345 [00:52<00:06, 5.35it/s]

Reading Batches_processed_new\batch_68.pkl...

Merging files: 90%|
| 312/345 [00:52<00:06, 5.47it/s]

Reading Batches_processed_new\batch_69.pkl...
Reading Batches_processed_new\batch_7.pkl...

Merging files: 91%|
| 314/345 [00:52<00:05, 5.26it/s]

Reading Batches_processed_new\batch_70.pkl...

Merging files:
91%|                                     | 315/345
[00:52<00:05, 5.28it/s]

Reading Batches_processed_new\batch_71.pkl...

Merging files:
92%|                                     | 316/345
[00:52<00:05, 5.42it/s]

Reading Batches_processed_new\batch_72.pkl...
Reading Batches_processed_new\batch_73.pkl...

Merging files: 92%|
| 318/345 [00:53<00:04, 5.40it/s]

Reading Batches_processed_new\batch_74.pkl...
Reading Batches_processed_new\batch_75.pkl...

Merging files:
93%|                                     | 320/345
[00:53<00:04, 5.27it/s]

Reading Batches_processed_new\batch_76.pkl...

Merging files:
93%|                                     | 321/345
[00:53<00:04, 5.14it/s]

Reading Batches_processed_new\batch_77.pkl...
Reading Batches_processed_new\batch_78.pkl...

```

Merging files:  
94%| | 323/345  
[00:54<00:04, 4.99it/s]  
Reading Batches\_processed\_new\batch\_79.pkl...  
Merging files:  
94%| | 324/345  
[00:54<00:04, 5.03it/s]  
Reading Batches\_processed\_new\batch\_8.pkl...  
Merging files:  
94%| | 325/345  
[00:54<00:03, 5.15it/s]  
Reading Batches\_processed\_new\batch\_80.pkl...  
Merging files:  
94%| | 326/345  
[00:54<00:03, 5.33it/s]  
Reading Batches\_processed\_new\batch\_81.pkl...  
Reading Batches\_processed\_new\batch\_82.pkl...  
Merging files:  
95%| | 328/345  
[00:55<00:03, 5.57it/s]  
Reading Batches\_processed\_new\batch\_83.pkl...  
Reading Batches\_processed\_new\batch\_84.pkl...  
Merging files:  
95%| | 329/345  
[00:55<00:02, 5.43it/s]  
Reading Batches\_processed\_new\batch\_85.pkl...  
Merging files:  
96%| | 330/345  
[00:55<00:03, 4.82it/s]  
Reading Batches\_processed\_new\batch\_86.pkl...  
Merging files:  
96%| | 331/345  
[00:56<00:03, 4.22it/s]  
Reading Batches\_processed\_new\batch\_87.pkl...  
Merging files:  
96%| | 332/345  
[00:56<00:03, 3.81it/s]  
Reading Batches\_processed\_new\batch\_88.pkl...

Merging files:  
97%| | 333/345  
[00:56<00:03, 3.50it/s]  
Reading Batches\_processed\_new\batch\_89.pkl...

Merging files:  
97%| | 334/345  
[00:56<00:02, 3.67it/s]  
Reading Batches\_processed\_new\batch\_9.pkl...

Merging files:  
97%| | 336/345  
[00:57<00:02, 4.02it/s]  
Reading Batches\_processed\_new\batch\_90.pkl...

Merging files:  
98%| | 337/345  
[00:57<00:01, 4.28it/s]  
Reading Batches\_processed\_new\batch\_91.pkl...

Merging files:  
98%| | 338/345  
[00:57<00:01, 4.63it/s]  
Reading Batches\_processed\_new\batch\_92.pkl...  
Reading Batches\_processed\_new\batch\_93.pkl...

Merging files:  
98%| | 339/345  
[00:57<00:01, 4.71it/s]  
Reading Batches\_processed\_new\batch\_94.pkl...

Merging files:  
99%| | 341/345  
[00:58<00:00, 4.51it/s]  
Reading Batches\_processed\_new\batch\_95.pkl...

Merging files:  
99%| | 342/345  
[00:58<00:00, 4.70it/s]  
Reading Batches\_processed\_new\batch\_96.pkl...

Merging files:  
99%| | 343/345  
[00:58<00:00, 4.83it/s]  
Reading Batches\_processed\_new\batch\_97.pkl...  
Reading Batches\_processed\_new\batch\_98.pkl...

```

Merging files:
100%|                               | 345/345
[00:59<00:00,  5.83it/s]

Reading Batches_processed_new\batch_99.pkl...
Merged 345 files.
Saved merged data to merged_data.pkl.

```

## 1 Merge gvkey

```

[3]: file_path = 'Raw_Data/us_gvkey.csv'
    us_gvkey = pd.read_csv(file_path, encoding='latin-1')
    us_gvkey = us_gvkey[["gvkey", "conm"]].drop_duplicates("gvkey")

[4]: ciq_common = pd.read_csv("Raw_Data/ciq_common.csv")

[5]: ciq_common_gvkey = ciq_common[ciq_common["symboltypecat"] == "gvkey"]

[6]: ciq_common_gvkey = ciq_common_gvkey[ciq_common_gvkey["symboltypename"] == "S&P_
    ↪GVKey"]

[8]: ciq_common_gvkey = ciq_common_gvkey[["companyid", "gvkey", "startdate",
    ↪"enddate"]]

[9]: merged_data_filtered = merged_data[merged_data["DEI_indicator_lowtype1"]]
    filtered_df_nondup = merged_data_filtered.drop_duplicates("keydevid")

[10]: merged_data_filtered2 = merged_data_filtered.merge(ciq_common_gvkey, on =
    ↪["companyid"], how = "left")

    # Replace 'B' and 'E' with actual dates
    merged_data_filtered2['startdate'] = merged_data_filtered2['startdate'].
    ↪replace('B', '1900-01-01')
    merged_data_filtered2['enddate'] = merged_data_filtered2['enddate'].
    ↪replace('E', '2099-12-31')

    merged_data_filtered2['mostimportantdateutc'] = pd.
    ↪to_datetime(merged_data_filtered2['mostimportantdateutc'])
    merged_data_filtered2['startdate'] = pd.
    ↪to_datetime(merged_data_filtered2['startdate'])
    merged_data_filtered2['enddate'] = pd.
    ↪to_datetime(merged_data_filtered2['enddate'])

    condition1 = merged_data_filtered2['mostimportantdateutc'] >=
    ↪merged_data_filtered2['startdate']
    condition2 = merged_data_filtered2['mostimportantdateutc'] <=
    ↪merged_data_filtered2['enddate']

```

```
merged_filtered_df = merged_data_filtered2[condition1 & condition2]
merged_filtered_df = merged_filtered_df.merge(us_gvkey, on = ["gvkey"], how = "left")
merged_filtered_df = merged_filtered_df[merged_filtered_df["conm"].notna()]
```

```
[11]: date_conditions = [
        ('2020-08-25', '2015-08-25', 'exp_5y'),
    ]

    aggregated_dfs = []

    for end_date, start_date, column_name in date_conditions:
        condition3 = merged_filtered_df['mostimportantdateutc'] < pd.
            to_datetime(end_date)
        condition4 = merged_filtered_df['mostimportantdateutc'] > pd.
            to_datetime(start_date)
        merged_filtered_df_temp = merged_filtered_df[condition3 & condition4]

        aggregated_df = merged_filtered_df_temp.groupby('gvkey').size().
            reset_index(name=column_name)

        aggregated_dfs.append(aggregated_df)

    final_merged_df = aggregated_dfs[0]
    for df in aggregated_dfs[1:]:
        final_merged_df = final_merged_df.merge(df, on='gvkey', how='outer')
```

```
[12]: final_merged_df.to_csv("Raw_Data/exposure.csv")
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[222]:
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```