

Conference call analysis - Step 1

November 22, 2025

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

```
[2]: # pip install wrds
```

```
[3]: import wrds
db = wrds.Connection()
```

Enter your WRDS username [Dan]: dxl902

Enter your password:

WRDS recommends setting up a .pgpass file.

Create .pgpass file now [y/n]?: y

pgpass file created at C:\Users\Dan\AppData\Roaming\postgresql\pgpass.conf

Created .pgpass file successfully.

You can create this file yourself at any time with the create_pgpass_file() function.

Loading library list...

Done

1 Create Sample

```
[4]: import pandas as pd

file_path = 'Raw_Data/ciqtranscript.csv'
ciqtranscriptmeta = pd.read_csv(file_path, encoding='latin-1')
```

```
[5]: ciqtranscriptmeta["transcriptid"] = ciqtranscriptmeta["transcriptid"].astype(int)
```

```
[6]: ciqtranscriptmeta = ciqtranscriptmeta[ciqtranscriptmeta["keydeveventtypename"]_
↳ == "Earnings Calls"]
```

```
[7]: max_transcriptid_by_keydevid = ciqtranscriptmeta.
↳ groupby('keydevid')['transcriptid'].max().reset_index()
```

```
[8]: max_transcriptid_by_keydevid = ciqtranscriptmeta.
      ↳groupby('keydevid')['transcriptid'].max().reset_index()
max_transcriptid_by_keydevid.columns = ['keydevid', 'max_transcriptid']
ciqtranscriptmeta2 = pd.merge(ciqtranscriptmeta, max_transcriptid_by_keydevid,
      ↳on='keydevid', how='left')
```

```
[9]: ciqtranscriptmeta2 = ciqtranscriptmeta2[ciqtranscriptmeta2["transcriptid"] ==
      ↳ciqtranscriptmeta2["max_transcriptid"]]
```

```
[10]: ciqtranscriptmeta2.head()
```

```
[10]:
```

	companyid	keydevid	transcriptid	\
4	18511	115648395	94097	
7	18511	132825577	137571	
12	18511	142695855	190899	
15	18511	206859802	309627	
18	18511	224174603	387778	

	headline	mostimportantdateutc	\
4	3i Group plc, H1 2011 Earnings Call, Nov 11, 2010	2010-11-11	
7	3i Group plc, 2011 Earnings Call, May 12, 2011	2011-05-12	
12	3i Group plc, H1 2012 Earnings Call, Nov 10, 2011	2011-11-10	
15	3i Group plc, 2012 Earnings Call, May 17, 2012	2012-05-17	
18	3i Group plc, H1 2013 Earnings Call, Nov 15, 2012	2012-11-15	

	mostimportanttimeutc	keydeveventypeid	keydeveventtypename	companyname	\
4	9:45:00	48	Earnings Calls	3i Group plc	
7	9:00:00	48	Earnings Calls	3i Group plc	
12	10:00:00	48	Earnings Calls	3i Group plc	
15	9:00:00	48	Earnings Calls	3i Group plc	
18	9:45:00	48	Earnings Calls	3i Group plc	

	transcriptcollectiontypeid	transcriptcollectiontypename	\
4	1	Proofed Copy	
7	1	Proofed Copy	
12	8	Audited Copy	
15	1	Proofed Copy	
18	1	Proofed Copy	

	transcriptpresentationtypeid	transcriptpresentationtypename	\
4	5	Final	
7	5	Final	
12	5	Final	
15	5	Final	
18	5	Final	

	transcriptcreationdate_utc	transcriptcreationtime_utc	audiolengthsec	\
--	----------------------------	----------------------------	----------------	---

4	2011-01-22	0:53:36	2945.0
7	2011-06-23	2:30:41	3646.0
12	2011-11-12	6:28:58	3667.0
15	2012-05-17	23:37:16	3059.0
18	2012-11-15	16:33:51	3400.0

	max_transcriptid
4	94097
7	137571
12	190899
15	309627
18	387778

```
[11]: import pandas as pd
from tqdm import tqdm
import pickle

batch_size = 1000

import os
if not os.path.exists('Batches'):
    os.makedirs('Batches')

for i in tqdm(range(0, len(ciqtranscriptmeta2), batch_size), desc="Processing_
→batches"):
    batch_ids = ciqtranscriptmeta2['transcriptid'].iloc[i:i+batch_size].tolist()
    batch_ids_str = ', '.join(map(str, batch_ids))

    sql_query = f'''SELECT a.*, b.*, c.componenttext
                    FROM (SELECT * FROM ciq.wrds_transcript_detail
                          WHERE transcriptid in ({batch_ids_str})) AS a,
                          ciq.wrds_transcript_person as b,
                          ciq.ciqtranscriptcomponent AS c
                    WHERE a.transcriptid=b.transcriptid AND b.
→transcriptcomponentid=c.transcriptcomponentid
                    ORDER by a.transcriptid, b.componentorder;'''

    data = db.raw_sql(sql_query)

    batch_filename = f'Batches/batch_{i//batch_size}.pkl'
    with open(batch_filename, 'wb') as f:
        pickle.dump(data, f)
```

Processing batches:
100%|| 345/345
[3:33:51<00:00, 37.19s/it]

[]:

[]: