In [ ]:
```
###necessary libraries
import pandas as pd
import glob
import os
from datetime import datetime, timezone
import re
import numpy as np
import itertools
from functools import reduce

# file where csv files of flair analysis lies
path_flair = r'C:\Users\victo\Master_Thesis\semanticanalysis\analysis_with_flair\bmw\outcome_using_flair'
all_files_flair = glob.glob(os.path.join(path_flair, "*.csv"))

# read files to pandas frame
list_of_files_flair = []

for filename in all_files_flair:
    list_of_files_flair.append(pd.read_csv(filename,
                                           sep=',',
                                           )
                               )

# Concatenate all content of files into one DataFrames
concatenate_list_of_files_flair = pd.concat(list_of_files_flair,
                                            ignore_index=True,
                                            axis=0,
                                            )

# removing duplicates
cleaned_dataframe_flair = concatenate_list_of_files_flair.sort_values(by='url', ascending=False)
cleaned_dataframe_flair = cleaned_dataframe_flair.drop_duplicates(subset=["url"], keep='first', ignore_index=True)

print(cleaned_dataframe_flair)

# file where csv files of vader analysis lies
path_vader = r'C:\Users\victo\Master_Thesis\semanticanalysis\analysis_with_vader\bmw\outcome_using_vader'
all_files_vader = glob.glob(os.path.join(path_vader, "*.csv"))

# read files to pandas frame
list_of_files_vader = []

for filename in all_files_vader:
    list_of_files_vader.append(pd.read_csv(filename,
                                           sep=',',
                                           )
                               )

# Concatenate all content of files into one DataFrames
concatenate_list_of_files_vader = pd.concat(list_of_files_vader,
                                            ignore_index=True,
                                            axis=0,
                                            )

# removing duplicates
cleaned_dataframe_vader = concatenate_list_of_files_vader.sort_values(by='url', ascending=False)
cleaned_dataframe_vader = cleaned_dataframe_vader.drop_duplicates(subset=["url"], k
```

```
eep='first', ignore_index=True)

print(cleaned_dataframe_vader)

# file where csv files of textblob analysis lies
path_textblob = r'C:\Users\victo\Master_Thesis\semanticanalysis\analysis_with_textb
lob\bmw\outcome_using_texblob'
all_files_textblob = glob.glob(os.path.join(path_textblob, "*.csv"))

# read files to pandas frame
list_of_files_textblob = []

for filename in all_files_textblob:
    list_of_files_textblob.append(pd.read_csv(filename,
                                              sep=',',
                                              )
                                  )

# Concatenate all content of files into one DataFrames
concatenate_list_of_files_textblob = pd.concat(list_of_files_textblob,
                                               ignore_index=True,
                                               axis=0,
                                               )


# removing duplicates
cleaned_dataframe_textblob = concatenate_list_of_files_textblob.sort_values(by='url
', ascending=False)
cleaned_dataframe_textblob = cleaned_dataframe_textblob.drop_duplicates(subset=["ur
l"], keep='first', ignore_index=True)

print(cleaned_dataframe_textblob)

##merging files together
merged_df = pd.merge(cleaned_dataframe_flair, cleaned_dataframe_vader, on=['url'])
merged_df = pd.merge(merged_df, cleaned_dataframe_textblob, on=['url'])
merged_df['formatted date'] = pd.to_datetime(merged_df['formatted date'])
merged_df.rename(columns={'formatted date': 'formatteddate'}, inplace=True)

path_stockprices = r'C:\Users\victo\Master_Thesis\stockprice_data\bmw\minutely_stoc
kpricefiles_with_return'

for file in glob.iglob(path_stockprices + '\*.csv'):
    date = re.search('\d{4}-\d{2}-\d{2}', file)
    date = date.group()
    df_daily_stock_prices = pd.read_csv(file,
                                        sep=',',
                                        )
    df_daily_stock_prices['Date'] = pd.DatetimeIndex(pd.to_datetime(df_daily_stock_
prices['Date'])).tz_localize('GMT').tz_convert('Europe/Berlin')
    df_daily_stock_prices['Date'] = pd.to_datetime(df_daily_stock_prices['Date'].d
t.strftime('%Y-%m-%d %H:%M:%S'))

    df_stock_prices_semantics = df_daily_stock_prices.merge(merged_df,
                                                            left_on='Date',
                                                            right_on='formatteddate
',
                                                            how='left')
    df_stock_prices_semantics.to_csv(r'C:\Users\victo\Master_Thesis\merging_data\bm
w\minutely\merged_files\bmwprices_with_semantics_' + date + '.csv', index=False)
    print('File of ' + date + ' has been saved!')
```