Calculating ratios with python

ToDo

1. Load data
2. Preprocess and merge

1. Load data

```
ratio_file = os.path.join("data", "investments_lowcarbon+GS+JB.xlsx")
ratio = pd.read_excel(ratio_file)
ratio = ratio[['company', 'year', 'capex_low-carbon_as_pct_calc']]
ratio['low_carbon_ratio'] = ratio['capex_low-carbon_as_pct_calc'] / 100
ratio = ratio.drop(columns=['capex_low-carbon_as_pct_calc'])
ratio = ratio.dropna(subset=['low_carbon_ratio'])
ratio.head()
```

```
##    company  year  low_carbon_ratio
## 10   exxon  2010          0.014789
## 11   exxon  2011          0.012940
## 12   exxon  2012          0.011965
## 13   exxon  2013          0.011204
## 14   exxon  2014          0.012369
```

```
posts_file = os.path.join("data", "1_labels_and_data.csv")
posts = pd.read_csv(posts_file)
posts['published_at'] = pd.to_datetime(posts['published_at'])
posts['year'] = posts['published_at'].dt.year
posts = posts[['channel_name', 'green', 'brown', 'misc', 'year']]
posts.head()
```

```
##   channel_name  green  brown   misc  year
## 0   ExxonMobil  False   True  False  2021
## 1   ExxonMobil  False  False   True  2021
## 2   ExxonMobil   True   True  False  2021
## 3   ExxonMobil   True  False  False  2021
## 4   ExxonMobil  False  False   True  2021
```

```
matching_file = os.path.join("data", "1_matching.csv")
matching = pd.read_csv(matching_file)
matching
```

```
##                channel company
## 0           ExxonMobil   exxon
## 1       ExxonMobil LNG   exxon
## 2   ExxonMobil Nat Gas   exxon
## 3  ExxonMobil Chemical   exxon
## 4    ExxonMobil Marine   exxon
## 5                 Esso   exxon
## 6       ExxonMobil PNG   exxon
```

2. Preprocess and merge

```
posts_merged = posts.merge(matching, left_on='channel_name', right_on='channel', how='left')
posts_merged.head()
```

```
##   channel_name  green  brown   misc  year     channel company
## 0   ExxonMobil  False   True  False  2021  ExxonMobil   exxon
## 1   ExxonMobil  False  False   True  2021  ExxonMobil   exxon
## 2   ExxonMobil   True   True  False  2021  ExxonMobil   exxon
## 3   ExxonMobil   True  False  False  2021  ExxonMobil   exxon
## 4   ExxonMobil  False  False   True  2021  ExxonMobil   exxon
```

```
def compute_green_ratio(group):
    green_sum = group['green'].sum()
    posts_count = ((group['green'] != 0) | (group['brown'] != 0)).sum()
    return pd.Series({'green_ratio': green_sum / posts_count})
aggregated_posts = posts_merged.groupby(
    ['company', 'year']
).apply(compute_green_ratio, include_groups = False)
aggregated_posts
```

```
##               green_ratio
## company year
## exxon   2017     0.000000
##         2018     0.692308
##         2019     0.466667
##         2020     0.600000
##         2021     0.435897
##         2022     0.589744
##         2023     0.827381
##         2024     0.884615
```

```
ratios = aggregated_posts.merge(ratio, on=['company', 'year'], how='left')
ratios['posts_to_capex'] = ratios['green_ratio'] / ratios['low_carbon_ratio']
ratios['posts_to_capex_normalized'] = (ratios['green_ratio'] - ratios['low_carbon_ratio']) / (ratios['g
ratios
```

```
##   company  year  ...  posts_to_capex  posts_to_capex_normalized
## 0   exxon  2017  ...        0.000000                  -1.000000
## 1   exxon  2018  ...       37.654615                   0.948260
## 2   exxon  2019  ...       30.478000                   0.936464
## 3   exxon  2020  ...       26.964000                   0.928479
## 4   exxon  2021  ...       19.589231                   0.902862
## 5   exxon  2022  ...        4.020174                   0.601607
## 6   exxon  2023  ...        6.211569                   0.722668
## 7   exxon  2024  ...        6.375606                   0.728836
##
## [8 rows x 6 columns]
```

```
output_file = os.path.join("data", "2_ratios.csv")
ratios.to_csv(output_file, index=False)
```