

analyze_scope_correlations

February 4, 2023

```
[ ]: import sys

sys.path.append("..")

import pathlib

import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import statsmodels.api as sm
import statsmodels.formula.api as smf
import pathlib
import time
import pickle
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from sklearn.preprocessing import KBinsDiscretizer
from postprocessors.scope_imputers import ScopeImputerPostprocessor
from datasources.core import PreviousScopeFeaturesDataManager
from base.dataset_loader import OxariDataManager
from datasources import LocalDatasource
from IPython.display import display, Markdown, Latex
pd.set_option('display.float_format', lambda x: '%.5f' % x)
# pd.reset_option('~display.', silent=True)

[ ]: cwd = pathlib.Path(__file__).parent
meta_model = pickle.load((cwd.parent / 'local/objects/meta_model/
↳snapshot-03-02-2023.pkl').open('rb'))
meta_model

[ ]: OxariMetaModel(params={})

[ ]: dataset = PreviousScopeFeaturesDataManager(
    LocalDatasource(path=cwd.parent/'local/data/scopes_auto.csv'),
    LocalDatasource(path=cwd.parent/'local/data/financials_auto.csv'),
    LocalDatasource(path=cwd.parent/'local/data/categoricals_auto.csv'),
```

```

).run()
DATA = dataset.get_data_by_name(OxariDataManager.ORIGINAL)
DATA

```

2023-02-04 19:09:30,610 - ScopeLoader - INFO - From 21583 initial data points, 21583 are complete data points and 0 data points have missing or invalid scopes
 2023-02-04 19:09:30,610 - ScopeLoader - INFO - From 21583 initial data points, 21583 are complete data points and 0 data points have missing or invalid scopes
 2023-02-04 19:09:30,610 - ScopeLoader - INFO - From 21583 initial data points, 21583 are complete data points and 0 data points have missing or invalid scopes
 2023-02-04 19:09:30,610 - ScopeLoader - INFO - From 21583 initial data points, 21583 are complete data points and 0 data points have missing or invalid scopes
 Not prepending group keys to the result index of transform-like apply. In the future, the group keys will be included in the index, regardless of whether the applied function returns a like-indexed object.
 To preserve the previous behavior, use

```
>>> .groupby(..., group_keys=False)
```

To adopt the future behavior and silence this warning, use

```
>>> .groupby(..., group_keys=True)
```

```
[ ]:
```

| | ft_cat_company_name | ft_cat_country_code | \ |
|-------|---|---------------------|---|
| 0 | Abu Dhabi Islamic Bank PJSC | NaN | |
| 1 | Dubai Investments PJSC | ARE | |
| 2 | Dubai Investments PJSC | ARE | |
| 3 | Dubai Investments PJSC | ARE | |
| 4 | Emirates Integrated Telecommunications Company... | ARE | |
| ... | ... | ... | |
| 16483 | Tsogo Sun Hotels Ltd | ZAF | |
| 16484 | Tsogo Sun Gaming Ltd | ZAF | |
| 16485 | Tsogo Sun Gaming Ltd | ZAF | |
| 16486 | Tsogo Sun Gaming Ltd | ZAF | |
| 16487 | Tsogo Sun Gaming Ltd | ZAF | |

| | ft_cat_country_name | ft_cat_industry_name | ft_cat_sector_name | \ |
|-------|----------------------|----------------------|------------------------|---|
| 0 | NaN | Banks | Financials | |
| 1 | United Arab Emirates | Conglomerates | Industrials | |
| 2 | United Arab Emirates | Conglomerates | Industrials | |
| 3 | United Arab Emirates | Conglomerates | Industrials | |
| 4 | United Arab Emirates | Telecom Services | Communication | |
| ... | ... | ... | ... | |
| 16483 | South Africa | Lodging | Consumer Discretionary | |
| 16484 | South Africa | Resorts & Casinos | Consumer Discretionary | |
| 16485 | South Africa | Resorts & Casinos | Consumer Discretionary | |
| 16486 | South Africa | Resorts & Casinos | Consumer Discretionary | |

| | | | | |
|-------|--------------|-------------------|------------------------|--|
| 16487 | South Africa | Resorts & Casinos | Consumer Discretionary | |
|-------|--------------|-------------------|------------------------|--|

| | ft_numc_cash | ft_numc_equity | ft_numc_inventories | ft_numc_market_cap \ |
|-------|--------------|----------------|---------------------|----------------------|
| 0 | NaN | NaN | NaN | NaN |
| 1 | 217.30946 | 3244.61239 | 667.88046 | NaN |
| 2 | 206.48441 | 3261.02982 | 727.20000 | NaN |
| 3 | 241.48999 | 3217.00694 | 679.22097 | NaN |
| 4 | NaN | 2138.36728 | 10.77595 | 7716.75173 |
| ... | ... | ... | ... | ... |
| 16483 | 67.35266 | 333.50346 | 3.04954 | NaN |
| 16484 | NaN | 581.50179 | 8.91365 | 1638.55256 |
| 16485 | NaN | 802.62599 | 8.57649 | 2257.05147 |
| 16486 | 223.15674 | 919.27986 | 10.04324 | 1709.71879 |
| 16487 | 177.37534 | 724.73675 | 5.20066 | 931.65516 |

| | ft_numc_net_income | ... | ft_numc_roe | ft_numc_stock_return \ |
|-------|--------------------|-----|-------------|------------------------|
| 0 | NaN | ... | NaN | NaN |
| 1 | 177.38080 | ... | 5.46693 | -0.02080 |
| 2 | 179.01130 | ... | 5.48941 | 0.00299 |
| 3 | 94.63581 | ... | 2.94173 | 0.00929 |
| 4 | 477.14658 | ... | 22.36602 | 0.27992 |
| ... | ... | ... | ... | ... |
| 16483 | -47.11006 | ... | -14.12581 | -1.26713 |
| 16484 | 130.78663 | ... | 23.44980 | -0.13227 |
| 16485 | 83.84508 | ... | 10.07068 | -0.07639 |
| 16486 | 105.07088 | ... | 13.52237 | -0.03575 |
| 16487 | 137.52401 | ... | 0.14882 | -0.37745 |

| | ft_numc_total_assets | ft_numc_total_liab | ft_numd_employees \ |
|-------|----------------------|--------------------|---------------------|
| 0 | NaN | NaN | NaN |
| 1 | 5324.49204 | 1979.27624 | NaN |
| 2 | 5706.52280 | 2382.87270 | NaN |
| 3 | 5925.54037 | 2644.78176 | NaN |
| 4 | 5012.75909 | NaN | NaN |
| ... | ... | ... | ... |
| 16483 | 808.75773 | 351.59036 | NaN |
| 16484 | 1745.10462 | NaN | 9285.50000 |
| 16485 | 2850.92950 | NaN | 26014.00000 |
| 16486 | 2432.31333 | 1573.34341 | 25844.00000 |
| 16487 | 2422.02658 | 1503.04837 | 10935.50000 |

| | key_isin | key_year | tg_numc_scope_1 | tg_numc_scope_2 \ |
|---|--------------|----------|-----------------|-------------------|
| 0 | AEA000801018 | 2020 | 202.00000 | 6291.00000 |
| 1 | AED000601016 | 2018 | 24259.00000 | 123944.00000 |
| 2 | AED000601016 | 2019 | 26483.00000 | 130473.00000 |
| 3 | AED000601016 | 2020 | 8386.00000 | 134191.00000 |
| 4 | AEE000701012 | 2016 | 32439.20000 | 277777.31000 |

| | | | | |
|-------|--------------|------|------------|--------------|
| ... | ... | ... | ... | ... |
| 16483 | ZAE000272522 | 2020 | 3955.00000 | 83187.00000 |
| 16484 | ZAE000273116 | 2015 | 5715.00000 | 212337.00000 |
| 16485 | ZAE000273116 | 2017 | 5198.00000 | 216516.00000 |
| 16486 | ZAE000273116 | 2018 | 5578.00000 | 217394.00000 |
| 16487 | ZAE000273116 | 2019 | 4302.00000 | 145846.00000 |

| | |
|---|-----------------|
| | tg_numc_scope_3 |
| 0 | 131.00000 |
| 1 | 917.00000 |
| 2 | 1032.00000 |
| 3 | 977.00000 |
| 4 | 2127.87000 |

| | |
|-------|--------------|
| ... | ... |
| 16483 | 69019.00000 |
| 16484 | 81804.00000 |
| 16485 | 103504.00000 |
| 16486 | 103516.00000 |
| 16487 | 22991.00000 |

[16488 rows x 27 columns]

```
[ ]: print("Impute scopes with Model")
scope_imputer = ScopeImputerPostprocessor(estimator=meta_model).run(X=DATA).
    ↪ evaluate()
scope_imputer
```

Impute scopes with Model

2023-02-04 19:09:40,193 - ScopeImputerPostprocessor - DEBUG - Number of unique companies in the data: 2778

2023-02-04 19:09:40,193 - ScopeImputerPostprocessor - DEBUG - Number of unique companies in the data: 2778

['year_transition'] did not aggregate successfully. If any error is raised this will raise in a future version of pandas. Drop these columns/ops to avoid this warning.

```
[ ]: @ScopeImputerPostprocessor[ Count of imputed {'tg_numc_scope_1': 45,
'tg_numc_scope_2': 273, 'tg_numc_scope_3': 3700} ]
```

```
[ ]: data_merged_with_input = scope_imputer.data.merge(DATA, on=["key_year",
    ↪ "key_isin"])
data_merged_with_input
```

| | | | | | |
|------|--------------|----------|-------------------|-------------------|---|
| []: | key_isin | key_year | tg_numc_scope_1_x | tg_numc_scope_2_x | \ |
| 0 | AEA000801018 | 2020 | 202.00000 | 6291.00000 | |
| 1 | AED000601016 | 2018 | 24259.00000 | 123944.00000 | |
| 2 | AED000601016 | 2019 | 26483.00000 | 130473.00000 | |

| | | | | |
|------|--------------|------|-------------|--------------|
| 3 | AED000601016 | 2020 | 8386.00000 | 134191.00000 |
| 4 | AEE000701012 | 2016 | 32439.20000 | 277777.31000 |
| ... | ... | ... | ... | ... |
| 8294 | ZAE000272522 | 2019 | 3675.00000 | 86233.00000 |
| 8295 | ZAE000272522 | 2020 | 3955.00000 | 83187.00000 |
| 8296 | ZAE000273116 | 2017 | 5198.00000 | 216516.00000 |
| 8297 | ZAE000273116 | 2018 | 5578.00000 | 217394.00000 |
| 8298 | ZAE000273116 | 2019 | 4302.00000 | 145846.00000 |

| | tg_numc_scope_3_x | predicted_s1 | predicted_s2 | predicted_s3 | \ |
|------|-------------------|--------------|--------------|--------------|---|
| 0 | 131.00000 | False | False | False | |
| 1 | 917.00000 | False | False | False | |
| 2 | 1032.00000 | False | False | False | |
| 3 | 977.00000 | False | False | False | |
| 4 | 2127.87000 | False | False | False | |
| ... | ... | ... | ... | ... | |
| 8294 | 76101.00000 | False | False | False | |
| 8295 | 69019.00000 | False | False | False | |
| 8296 | 103504.00000 | False | False | False | |
| 8297 | 103516.00000 | False | False | False | |
| 8298 | 22991.00000 | False | False | False | |

| | ft_cat_company_name | ft_cat_country_code | \ |
|------|---|---------------------|---|
| 0 | Abu Dhabi Islamic Bank PJSC | NaN | |
| 1 | Dubai Investments PJSC | ARE | |
| 2 | Dubai Investments PJSC | ARE | |
| 3 | Dubai Investments PJSC | ARE | |
| 4 | Emirates Integrated Telecommunications Company... | ARE | |
| ... | ... | ... | |
| 8294 | Tsogo Sun Hotels Ltd | ZAF | |
| 8295 | Tsogo Sun Hotels Ltd | ZAF | |
| 8296 | Tsogo Sun Gaming Ltd | ZAF | |
| 8297 | Tsogo Sun Gaming Ltd | ZAF | |
| 8298 | Tsogo Sun Gaming Ltd | ZAF | |

| | ... | ft_numc_revenue | ft_numc_roa | ft_numc_roe | ft_numc_stock_return | \ |
|------|-----|-----------------|-------------|-------------|----------------------|---|
| 0 | ... | NaN | NaN | NaN | NaN | |
| 1 | ... | 773.36882 | 0.03331 | 5.46693 | -0.02080 | |
| 2 | ... | 733.36447 | 0.03137 | 5.48941 | 0.00299 | |
| 3 | ... | 753.62914 | 0.01597 | 2.94173 | 0.00929 | |
| 4 | ... | 3464.92583 | 0.09642 | 22.36602 | 0.27992 | |
| ... | ... | ... | ... | ... | ... | |
| 8294 | ... | 306.37559 | -0.00672 | -1.40200 | NaN | |
| 8295 | ... | 235.28740 | -0.05825 | -14.12581 | -1.26713 | |
| 8296 | ... | 940.89520 | 0.10478 | 10.07068 | -0.07639 | |
| 8297 | ... | 757.06720 | 0.04518 | 13.52237 | -0.03575 | |
| 8298 | ... | 844.85714 | 0.04883 | 0.14882 | -0.37745 | |

| | ft_numc_total_assets | ft_numc_total_liab | ft_numd_employees | \ |
|------|----------------------|--------------------|-------------------|---|
| 0 | NaN | NaN | NaN | |
| 1 | 5324.49204 | 1979.27624 | NaN | |
| 2 | 5706.52280 | 2382.87270 | NaN | |
| 3 | 5925.54037 | 2644.78176 | NaN | |
| 4 | 5012.75909 | NaN | NaN | |
| ... | ... | ... | ... | |
| 8294 | 1018.18054 | 325.08342 | NaN | |
| 8295 | 808.75773 | 351.59036 | NaN | |
| 8296 | 2850.92950 | NaN | 26014.00000 | |
| 8297 | 2432.31333 | 1573.34341 | 25844.00000 | |
| 8298 | 2422.02658 | 1503.04837 | 10935.50000 | |

| | tg_numc_scope_1_y | tg_numc_scope_2_y | tg_numc_scope_3_y |
|------|-------------------|-------------------|-------------------|
| 0 | 202.00000 | 6291.00000 | 131.00000 |
| 1 | 24259.00000 | 123944.00000 | 917.00000 |
| 2 | 26483.00000 | 130473.00000 | 1032.00000 |
| 3 | 8386.00000 | 134191.00000 | 977.00000 |
| 4 | 32439.20000 | 277777.31000 | 2127.87000 |
| ... | ... | ... | ... |
| 8294 | 3675.00000 | 86233.00000 | 76101.00000 |
| 8295 | 3955.00000 | 83187.00000 | 69019.00000 |
| 8296 | 5198.00000 | 216516.00000 | 103504.00000 |
| 8297 | 5578.00000 | 217394.00000 | 103516.00000 |
| 8298 | 4302.00000 | 145846.00000 | 22991.00000 |

[8299 rows x 33 columns]

Dataset ### Description of the training Dataset This table describes key characteristics of the dataset used for training.

```
[ ]: display(Markdown(f"The data contained {DATA.shape[0]} data points with {DATA.\
    ↪shape[1]} columns."))
display(Markdown(f"Of the seventeen columns {DATA.columns.str.startswith('tg_').\
    ↪sum()} acted as dependent variables, {DATA.columns.str.startswith('ft_').\
    ↪sum()} as independent variables and {DATA.columns.str.startswith('key_').\
    ↪sum()} as columns with meta information."))
display(DATA.describe().T)
```

The data contained 16488 data points with 27 columns.

Of the seventeen columns 3 acted as dependent variables, 22 as independent variables and 2 as columns with meta information.

| | count | mean | std | \ |
|---------------------|-------------|-------------|---------------|---|
| ft_numc_cash | 4394.00000 | 4248.04557 | 26526.89408 | |
| ft_numc_equity | 10211.00000 | 71929.66043 | 2395941.82430 | |
| ft_numc_inventories | 14289.00000 | 1932.58130 | 17726.63462 | |

| | | | |
|---------------------------------|-------------|----------------|----------------|
| ft_numc_market_cap | 15822.00000 | 19894.21753 | 42089.91725 |
| ft_numc_net_income | 16159.00000 | 3754.35817 | 137508.01207 |
| ft_numc_ppe | 16126.00000 | 13252.86434 | 235111.37598 |
| ft_numc_preyear_tg_numc_scope_1 | 13623.00000 | 3994624.81874 | 15891403.60688 |
| ft_numc_preyear_tg_numc_scope_2 | 13036.00000 | 755897.63198 | 2540475.16461 |
| ft_numc_preyear_tg_numc_scope_3 | 7873.00000 | 13236917.79286 | 71977826.68407 |
| ft_numc_rd | 8224.00000 | 2711.78295 | 56254.92903 |
| ft_numc_revenue | 16179.00000 | 22785.67336 | 309159.25115 |
| ft_numc_roa | 10183.00000 | 0.06949 | 0.82135 |
| ft_numc_roe | 16047.00000 | 232.52816 | 24975.06607 |
| ft_numc_stock_return | 10201.00000 | -0.01111 | 13.17676 |
| ft_numc_total_assets | 16174.00000 | 112590.60315 | 3139612.07409 |
| ft_numc_total_liab | 4409.00000 | 49553.65118 | 211078.88563 |
| ft_numd_employees | 13725.00000 | 39166.29493 | 67076.39651 |
| key_year | 16488.00000 | 2014.91794 | 3.37614 |
| tg_numc_scope_1 | 16391.00000 | 3854435.72230 | 16322286.51366 |
| tg_numc_scope_2 | 15736.00000 | 730656.52479 | 2618576.10176 |
| tg_numc_scope_3 | 9414.00000 | 13184398.40554 | 74707532.24870 |

| | min | 25% | 50% \ |
|---------------------------------|----------------|-------------|--------------|
| ft_numc_cash | -0.00000 | 154.16548 | 567.77764 |
| ft_numc_equity | -12459.00000 | 1159.91428 | 3056.75762 |
| ft_numc_inventories | -183.52404 | 93.05000 | 435.41603 |
| ft_numc_market_cap | 0.29692 | 2812.80926 | 7200.51421 |
| ft_numc_net_income | -8946924.88098 | 107.91485 | 358.62992 |
| ft_numc_ppe | 0.00000 | 519.42047 | 1883.98764 |
| ft_numc_preyear_tg_numc_scope_1 | 5.00000 | 11474.50000 | 91852.00000 |
| ft_numc_preyear_tg_numc_scope_2 | 7.00000 | 32479.00000 | 134772.91000 |
| ft_numc_preyear_tg_numc_scope_3 | 5.00000 | 9600.00000 | 55900.00000 |
| ft_numc_rd | -99.35493 | 17.79868 | 89.59962 |
| ft_numc_revenue | -29410.14274 | 2081.30127 | 5872.34520 |
| ft_numc_roa | -28.06794 | 0.01340 | 0.04127 |
| ft_numc_roe | -25581.41698 | 5.10603 | 11.00474 |
| ft_numc_stock_return | -744.31113 | -0.13776 | 0.06624 |
| ft_numc_total_assets | 0.00000 | 3855.56722 | 10642.64176 |
| ft_numc_total_liab | 2.83230 | 1558.59797 | 5122.33201 |
| ft_numd_employees | 0.00000 | 5507.00000 | 15116.00000 |
| key_year | 2007.00000 | 2012.00000 | 2016.00000 |
| tg_numc_scope_1 | 5.00000 | 10101.18500 | 79759.00000 |
| tg_numc_scope_2 | 6.50000 | 28914.25000 | 121333.70400 |
| tg_numc_scope_3 | 5.00000 | 9199.82500 | 56313.50000 |

| | 75% | max |
|---------------------|-------------|-----------------|
| ft_numc_cash | 1717.69469 | 669580.21647 |
| ft_numc_equity | 7919.30809 | 220102283.09395 |
| ft_numc_inventories | 1418.70082 | 1639174.80331 |
| ft_numc_market_cap | 18688.64000 | 920224.30298 |
| ft_numc_net_income | 1132.48693 | 11083000.00000 |

| | | |
|---------------------------------|--------------|------------------|
| ft_numc_ppe | 6421.71122 | 18338964.85267 |
| ft_numc_preyear_tg_numc_scope_1 | 885196.00000 | 377600000.00000 |
| ft_numc_preyear_tg_numc_scope_2 | 504856.75000 | 68647841.00000 |
| ft_numc_preyear_tg_numc_scope_3 | 666500.00000 | 2298400000.00000 |
| ft_numc_rd | 361.15807 | 3005639.71794 |
| ft_numc_revenue | 16171.34169 | 24681241.32815 |
| ft_numc_roa | 0.07715 | 34.16375 |
| ft_numc_roe | 18.21770 | 3156000.00000 |
| ft_numc_stock_return | 0.31471 | 213.39993 |
| ft_numc_total_assets | 34698.73068 | 377421932.57182 |
| ft_numc_total_liab | 17496.46732 | 3447903.62915 |
| ft_numd_employees | 42699.00000 | 798000.00000 |
| key_year | 2018.00000 | 2020.00000 |
| tg_numc_scope_1 | 787030.00000 | 619271085.00000 |
| tg_numc_scope_2 | 462757.75000 | 68647841.00000 |
| tg_numc_scope_3 | 749728.25000 | 2687800000.00000 |

Description of the Dataset imputed

```
[ ]: display(Markdown("ATTENTION: This correlation analysis is based on a subset of_
↳data."))
display(Markdown("The data was selected based on two criteria."))
display(Markdown("1. Data without information about any of the 3 scope targets_
↳where removed."))
display(Markdown("2. Data from years prior 2016 where removed."))
display(data_merged_with_input.describe().T)
```

ATTENTION: This correlation analysis is based on a subset of data.

The data was selected based on two criteria.

1. Data without information about any of the 3 scope targets where removed.
2. Data from years prior 2016 where removed.

| | count | mean | std \ |
|---------------------------------|------------|----------------|----------------|
| key_year | 8299.00000 | 2017.74503 | 1.17408 |
| tg_numc_scope_1_x | 8299.00000 | 3355589.04040 | 16657296.09221 |
| tg_numc_scope_2_x | 8299.00000 | 645352.06698 | 2795484.37174 |
| tg_numc_scope_3_x | 8299.00000 | 8210864.58858 | 64395906.28620 |
| ft_numc_cash | 4394.00000 | 4248.04557 | 26526.89408 |
| ft_numc_equity | 6394.00000 | 110018.59214 | 3027202.14341 |
| ft_numc_inventories | 7319.00000 | 2136.05071 | 24525.48371 |
| ft_numc_market_cap | 7815.00000 | 18318.94300 | 44386.64078 |
| ft_numc_net_income | 8161.00000 | 6157.36052 | 193437.02901 |
| ft_numc_ppe | 8146.00000 | 18390.45795 | 330282.36375 |
| ft_numc_preyear_tg_numc_scope_1 | 7127.00000 | 3479766.80486 | 15345459.06632 |
| ft_numc_preyear_tg_numc_scope_2 | 6913.00000 | 668707.62622 | 2613653.27720 |
| ft_numc_preyear_tg_numc_scope_3 | 3982.00000 | 13691401.39689 | 80559011.28284 |
| ft_numc_rd | 3788.00000 | 5114.80026 | 82812.55838 |

| | | | |
|----------------------|------------|----------------|----------------|
| ft_numc_revenue | 8169.00000 | 25985.55353 | 433671.89767 |
| ft_numc_roa | 6386.00000 | 0.07837 | 1.03488 |
| ft_numc_roe | 8112.00000 | 444.13457 | 35126.41907 |
| ft_numc_stock_return | 6392.00000 | -0.08186 | 16.64349 |
| ft_numc_total_assets | 8165.00000 | 139526.50802 | 4410226.72669 |
| ft_numc_total_liab | 4409.00000 | 49553.65118 | 211078.88563 |
| ft_numd_employees | 6654.00000 | 35972.56927 | 66379.11316 |
| tg_numc_scope_1_y | 8254.00000 | 3372994.01805 | 16700894.15046 |
| tg_numc_scope_2_y | 8026.00000 | 652531.69174 | 2839099.23956 |
| tg_numc_scope_3_y | 4599.00000 | 13600830.18752 | 85380009.42948 |

| | min | 25% | 50% \ |
|---------------------------------|----------------|-------------|--------------|
| key_year | 2016.00000 | 2017.00000 | 2018.00000 |
| tg_numc_scope_1_x | 5.00000 | 7057.00000 | 52996.20305 |
| tg_numc_scope_2_x | 6.50000 | 19961.39550 | 93364.00000 |
| tg_numc_scope_3_x | 7.10000 | 5236.40922 | 31870.00000 |
| ft_numc_cash | -0.00000 | 154.16548 | 567.77764 |
| ft_numc_equity | -12459.00000 | 1107.82950 | 2971.73901 |
| ft_numc_inventories | 0.00000 | 64.10462 | 332.82187 |
| ft_numc_market_cap | 0.29692 | 2312.07167 | 6167.21278 |
| ft_numc_net_income | -8946924.88098 | 89.99918 | 314.03310 |
| ft_numc_ppe | 0.00000 | 416.11212 | 1491.81167 |
| ft_numc_preyear_tg_numc_scope_1 | 5.00000 | 8962.50000 | 64706.00000 |
| ft_numc_preyear_tg_numc_scope_2 | 7.00000 | 25673.00000 | 107365.00000 |
| ft_numc_preyear_tg_numc_scope_3 | 8.00000 | 9636.25000 | 66748.00000 |
| ft_numc_rd | -0.35079 | 14.80747 | 81.49547 |
| ft_numc_revenue | -1968.77163 | 1663.00002 | 4504.53587 |
| ft_numc_roa | -28.06794 | 0.01334 | 0.03914 |
| ft_numc_roe | -25581.41698 | 4.52623 | 10.39193 |
| ft_numc_stock_return | -744.31113 | -0.15127 | 0.05892 |
| ft_numc_total_assets | 0.00000 | 3252.57538 | 9132.39699 |
| ft_numc_total_liab | 2.83230 | 1558.59797 | 5122.33201 |
| ft_numd_employees | 0.00000 | 4832.00000 | 13003.50000 |
| tg_numc_scope_1_y | 5.00000 | 7259.92500 | 53862.50000 |
| tg_numc_scope_2_y | 6.50000 | 20869.25000 | 94068.00000 |
| tg_numc_scope_3_y | 7.10000 | 8850.50000 | 69341.00000 |

| | 75% | max |
|---------------------|--------------|------------------|
| key_year | 2019.00000 | 2020.00000 |
| tg_numc_scope_1_x | 548500.00000 | 619271085.00000 |
| tg_numc_scope_2_x | 372046.51000 | 68647841.00000 |
| tg_numc_scope_3_x | 524817.50000 | 2687800000.00000 |
| ft_numc_cash | 1717.69469 | 669580.21647 |
| ft_numc_equity | 7634.97500 | 220102283.09395 |
| ft_numc_inventories | 1219.40644 | 1639174.80331 |
| ft_numc_market_cap | 16117.38654 | 920224.30298 |
| ft_numc_net_income | 963.56002 | 11083000.00000 |
| ft_numc_ppe | 5229.42968 | 18338964.85267 |

| | | |
|---------------------------------|---------------|------------------|
| ft_numc_preyear_tg_numc_scope_1 | 668619.00000 | 377600000.00000 |
| ft_numc_preyear_tg_numc_scope_2 | 422851.00000 | 68647841.00000 |
| ft_numc_preyear_tg_numc_scope_3 | 1001500.00000 | 2298400000.00000 |
| ft_numc_rd | 323.94462 | 3005639.71794 |
| ft_numc_revenue | 12930.89440 | 24681241.32815 |
| ft_numc_roa | 0.07324 | 34.16375 |
| ft_numc_roe | 17.05182 | 3156000.00000 |
| ft_numc_stock_return | 0.33351 | 213.39993 |
| ft_numc_total_assets | 29310.00000 | 377421932.57182 |
| ft_numc_total_liab | 17496.46732 | 3447903.62915 |
| ft_numd_employees | 36255.87500 | 798000.00000 |
| tg_numc_scope_1_y | 552750.00000 | 619271085.00000 |
| tg_numc_scope_2_y | 372000.00000 | 68647841.00000 |
| tg_numc_scope_3_y | 1235696.50000 | 2687800000.00000 |

Amount of imputed data counts The amount of data points that where imputed.

```
[ ]: display(scope_imputer.data[["predicted_s1", "predicted_s2", "predicted_s3"]].
      ↪sum())
```

```
predicted_s1      45
predicted_s2     273
predicted_s3    3700
dtype: int64
```

Correlation Analysis ### No predictions

```
[ ]: display("Scope 1 Correlation (no predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s1"]==False][["tg_numc_scope_1_y",
      ↪"ft_numc_revenue"]].corr())
display("Scope 2 Correlation (no predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s2"]==False][["tg_numc_scope_2_y",
      ↪"ft_numc_revenue"]].corr())
display("Scope 3 Correlation (no predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s3"]==False][["tg_numc_scope_3_y",
      ↪"ft_numc_revenue"]].corr())
```

'Scope 1 Correlation (no predictions)'

| | tg_numc_scope_1_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_1_x | 1.00000 | 0.00690 |
| ft_numc_revenue | 0.00690 | 1.00000 |

'Scope 2 Correlation (no predictions)'

| | tg_numc_scope_2_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_2_x | 1.00000 | 0.01390 |
| ft_numc_revenue | 0.01390 | 1.00000 |

'Scope 3 Correlation (no predictions)'

| | tg_numc_scope_3_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_3_x | 1.00000 | 0.01070 |
| ft_numc_revenue | 0.01070 | 1.00000 |

Only predictions

```
[ ]: display("Scope 1 Correlation (only predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s1"]==True][["tg_numc_scope_1_x",
↪ "ft_numc_revenue"]].corr())
display("Scope 2 Correlation (only predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s2"]==True][["tg_numc_scope_2_x",
↪ "ft_numc_revenue"]].corr())
display("Scope 3 Correlation (only predictions)")
display(data_merged_with_input[data_merged_with_input["predicted_s3"]==True][["tg_numc_scope_3_x",
↪ "ft_numc_revenue"]].corr())
```

'Scope 1 Correlation (only predictions)'

| | tg_numc_scope_1_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_1_x | 1.00000 | -0.06304 |
| ft_numc_revenue | -0.06304 | 1.00000 |

'Scope 2 Correlation (only predictions)'

| | tg_numc_scope_2_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_2_x | 1.00000 | 0.37948 |
| ft_numc_revenue | 0.37948 | 1.00000 |

'Scope 3 Correlation (only predictions)'

| | tg_numc_scope_3_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_3_x | 1.00000 | 0.03353 |
| ft_numc_revenue | 0.03353 | 1.00000 |

Overall

```
[ ]: display("Scope 1 Correlation (overall)")
display(data_merged_with_input[["tg_numc_scope_1_x", "ft_numc_revenue"]].corr())
display("Scope 2 Correlation (overall)")
display(data_merged_with_input[["tg_numc_scope_2_x", "ft_numc_revenue"]].corr())
display("Scope 3 Correlation (overall)")
display(data_merged_with_input[["tg_numc_scope_3_x", "ft_numc_revenue"]].corr())
```

'Scope 1 Correlation (overall)'

| | tg_numc_scope_1_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_1_x | 1.00000 | 0.00696 |
| ft_numc_revenue | 0.00696 | 1.00000 |

'Scope 2 Correlation (overall)'

| | tg_numc_scope_2_x | ft_numc_revenue |
|-------------------|-------------------|-----------------|
| tg_numc_scope_2_x | 1.00000 | 0.01431 |

| | | |
|---------------------------------|-------------------|-----------------|
| ft_numc_revenue | 0.01431 | 1.00000 |
| 'Scope 3 Correlation (overall)' | | |
| | tg_numc_scope_3_x | ft_numc_revenue |
| tg_numc_scope_3_x | 1.00000 | 0.01371 |
| ft_numc_revenue | 0.01371 | 1.00000 |