Understanding Middle Market Tech Companies using `THEMA`

An introduction to our *Topological Hyperparameter Evaluation Mapping Algorithm*

Abstract

This report summarizes our work on a selection of middle market technology companies. In particular, we generate graph models using `THEMA`, a Topological Hyper-parameter Evaluation Mapping Algorithm developed at Krv Analytics, and perform unsupervised *classification* and *analysis* of companies.

1 Motivation

The current arsenal of artificial intelligence methods fall short when it comes to comparing and analyzing complex objects across different data disciplines. By leveraging *topology* and *graph learning*, `THEMA` can extract meaningful and tractable insights from sparse, high dimensional data sets, which are notoriously difficult to navigate with standard machine learning and statistical approaches. These datasets represent only a small subset of intricate objects with numerous features, which still pose significant challenges to even the most advanced AI techniques available today.

This is especially prevalent in *venture capitalism* (VC), where the ever increasing complexity of evaluation criteria is pushing the industry towards higher and higher dimensional representations of companies that incorporate internal, financial, environmental, and geo-political features (just to name a few). Thus, it has become imperative for the industry to understand and contextualize company similarity across multiple facets.

We demonstrate our utility in the VC space by using `THEMA` to generate graph models of a selection of middle market tech companies. These models provide a ”map” of company similarity in high dimensional space, and automatically generate clusters of companies based on these multidimensional relationships. We believe that our models have the potential to significantly improve data driven investment strategies.

2 Data Set

Our data set contains financial, geo-political, public opinion, and environmental data on **30 middle market** companies in the **Technology Sector**. This information was queried from public, open source, and academic resources– for a full description we refer the reader to the Appendix 6. Also, see Table 4 and Table 3 for a list of companies and description of our non-financial data fields, respectively.

We acknowledge the limitations of our data set and ask the reader to keep this in mind as they examine the results of our demo. `THEMA`'s ability to produce novel findings that enhance investment strategy will continue to improve with higher quality and more descriptive data.

3 Model

Our goal is to *map* out the landscape of middle market tech companies. To that end, we use `THEMA` to generate a graph model of our data set that captures the relationships between *all* 30 companies in question with respect to the available data fields.

How to interpret the model below:

- *Nodes* (circles) represent clusters of similar companies, sized by the number of companies per cluster.
- *Edges* (lines) connect similar clusters based on multidimensional relationships in the data.
- *Groups* (color blocks) are connected components, or isolated sections of the graph. These represent groups of companies in the data set with distinct characteristics across the many data fields available.

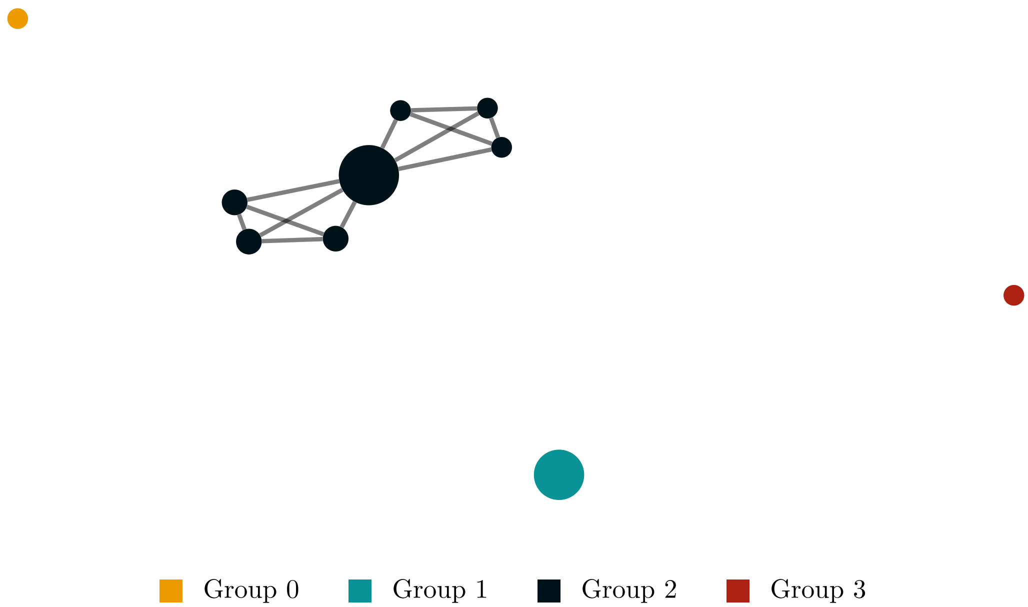


Figure 1: Graph Model of Middle Market Technology Companies

4 Results

The groups generated by `THEMA` give an *unsupervised classification* of the 30 middle market tech companies. These groups arise from the structure of our high dimensional set, whose complex relationships are distilled into a graph model. Here are the companies contained in each group, as segmented by `THEMA`:

Group 0	**Group 1**
Amtech Systems, Inc.	Airgain, Inc.
Camtek Ltd.	AXT, Inc.
Marine Products Corporation	Kopin Corporation
Haynes International, Inc.	NVE Corporation
	Vicor Corporation
	inTEST Corporation
	Qualstar Corporation
	Richardson Electronics, Ltd.
	Clearfield, Inc.
	Aehr Test Systems

Group 2	**Group 3**
Cohu, Inc.	Ambarella, Inc.
CTS Corporation	Iridium Communications Inc.
FormFactor, Inc.	MACOM Technology Solutions Holdings, Inc.
Lattice Semiconductor Corporation	Rambus Inc.
Mercury Systems, Inc.	
Nova Ltd.	
Power Integrations, Inc.	
Semtech Corporation	
Veeco Instruments Inc.	
Photronics, Inc.	
Silicon Laboratories Inc.	
Universal Display Corporation	

Table 1: Company Groupings

Unlike traditional clustering methods, `THEMA` uses the graph model's structure to build digestible descriptions of these groups. This allows us to illustrate which features connect the companies in each group, and compare group profiles across all available data fields.

We address *why these companies are grouped together* and *what makes these groups interesting and/or significant* in the following subsections.

4.1 Group Connectivity

We start by labeling nodes based on the most homogeneous data field across it's members. These are the features where the companies in a given node are *most similar*. Combining this information across nodes in a group gives rise to a concise description of a *group's connectivity*, which we display in Figure 2. See Table 3 for full feature descriptions.

Figure 2: Group Connectivity

Our pie charts can be used in combination with the box plots to quickly determine characteristic features of each group. The pie charts identify important data fields within a group and the box plots allow for easy comparison between other groups, as well as the global average. This helps situate a groups profile within a global context.

As an example, consider Group 1. With a single defining characteristic in it's pie chart, we can quickly check the box plot to investigate Group 1's *Total Liabilities Net Minority Interest*. These 10 companies are very tightly grouped in this feature *and* are significantly lower than all other groups, and the global average. When a group has more complex graph structure, the connectivity may be attributed to multiple data columns, as is the case for Group 2.

Summarizing Group Connectivity:

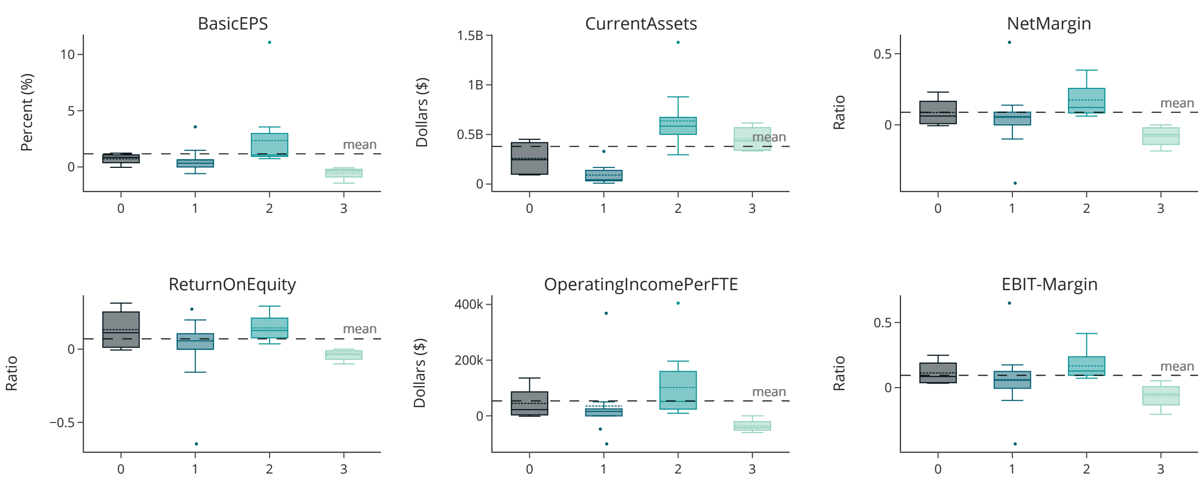
- **Group 0:** Companies with *high* Capex To Assets ratios.
- **Group 1:** Companies with *low* Total Liabilities Net Minority Interest.
- **Group 2**: Companies with *high* Operating Margin and Return On Equity, as well as *above average* Capex To CFI ratios.
- **Group 3**: Companies located in counties where a *high* percentage of people believe global warming (GW) is real.

While our groupings are derived from every feature of the data set, looking at the most homogeneous data fields within each group gives a high-level and digestible overview of why certain companies are grouped together.

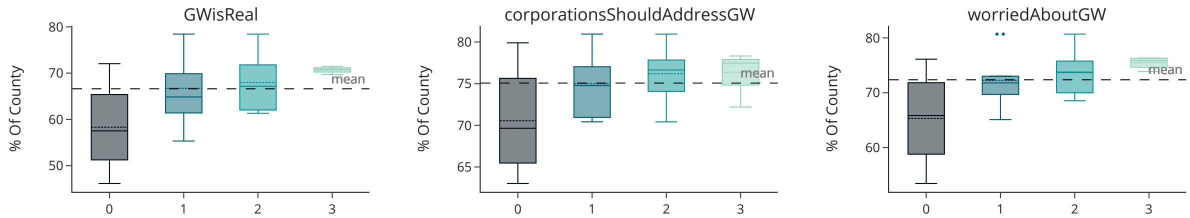
4.2 Group Descriptions

In addition to the connectivity described above, we can look at ensembles of features to get a better sense of the unique properties of each group. To understand the significance of each group of companies, one must understand the interplay between their financial, geo-political, and environmental identities as well as relationships between groups.

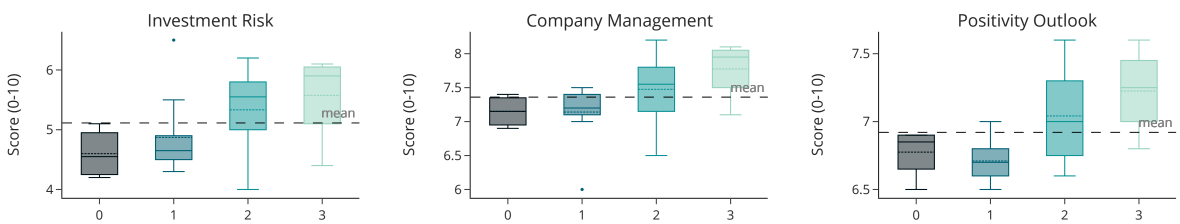
4.2.1 Selected Financial Attributes



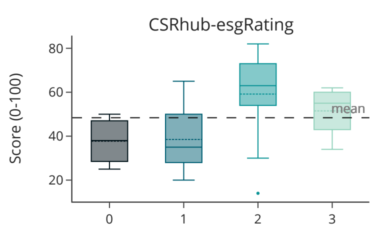
4.2.2 County-level Climate Opinions



4.2.3 NLP-Generated Press Opinion



4.2.4 Environmental Attributes



Touching on Group Comparison:

- **Group 0 vs Group 2:** While these groups are very similar across their Net Margin and EBIT-Margin, indicating a higher degree of profitability and operating efficiency, they differ significantly across the perceived investment risks and ESG scores.
- **Group 1 vs Group 2:** While drastically different across environmental and financial metrics, these groups hover near our dataset-average in terms of their county-level climate opinions.

5 Use Cases

We believe that among the most powerful contributions of `THEMA` is a shift in the ways that we approach company data: we should use the structure of our data to design more accurate evaluation metrics within the context of specific groups. Many modeling techniques in the VC space could benefit from this approach– we include two example applications below to motivate the reader and foreshadow the utility of our method.

> Although clustering companies is not a new idea, our model is far superior in capturing and reporting complex relationships at multiple scales. We believe this is crucial for getting the best results out of the data available, especially for domains that consistently depend on sparse, high dimensional data sets.

5.1 Target Matching: Axcelis Inc

We use `THEMA` to develop a target matching scheme. This places a new company, not yet seen by the model, into a particular group, thus mapping the target into its location within the data's landscape. This example was inspired by the need to compare private and public companies. In many cases, private company data is limited, and thus it is very informative to map them into the landscape of public companies where you can design more nuanced evaluation criteria or even attempt to predict performance.

In this experiment we look at Axcelis Inc - missing ESG data and NLP-generated press opinion data - and demonstrate `THEMA`'s ability to predict which group this company would fall into based on the available inputs.

Group	Matching Score (↓)
Group 0	64.3
Group 1	13,700
Group 2	**10.4**
Group 3	106
Target Placement	**Group 2**

Table 2: Axcelis Inc Matching Scores

Table 2 displays the target's matching scores for each group, where a group's score is defined as the sum of mean variance across the target's available data fields. The group with the lowest value corresponds to the most tightly matched group, which for this particular target/model pair results in Axcelis Inc being placed in Group 2. Figure 3 shows Axelis's data fields in relation to each group.

5.1.1 Leveraging Target Matching

As a thought experiment, lets assume the companies in Group 2 are found to be more resilient to supply-chain shortages than the rest of our groups. By target matching and placing Axcelis Inc solidly in Group 2, we can assume it will react in a similar way when subjected to similar factors, i.e. will be less affected by supply chain shortages than similar companies in other groups.

While we acknowledge our dataset lacks sufficient information for ESG scoring, target matching can also be used to predict the ESG impacts of a company without a score or in cases where it is believed the ESG score is inaccurate. Here, we match Axcelis Inc into Group 2 – additional research shows Axcelis Inc does indeed have an above average ESG score comparable to those of companies within Group 2.

> Target matching allows us to extract and apply insights gained from our global dataset to inform us about companies we lack data on (private companies for instance). For example, one could use `THEMA` to predict ESG impacts or resiliency to supply-chain shortages.

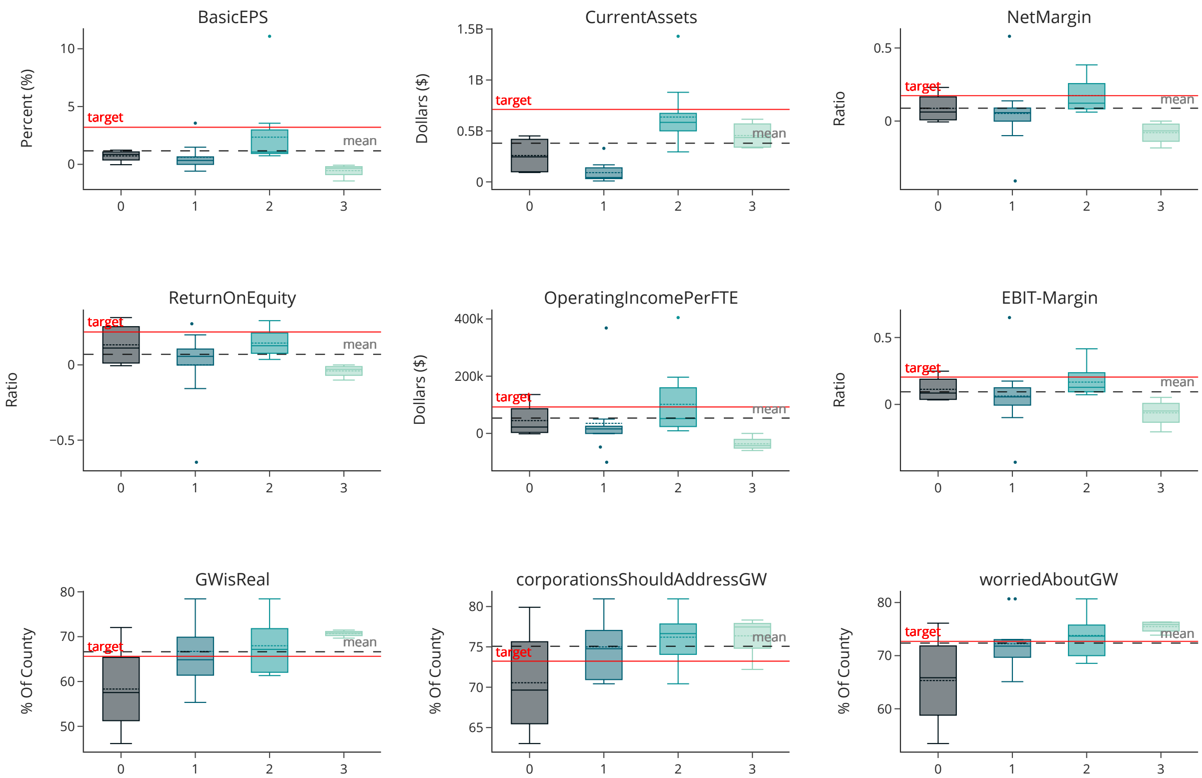


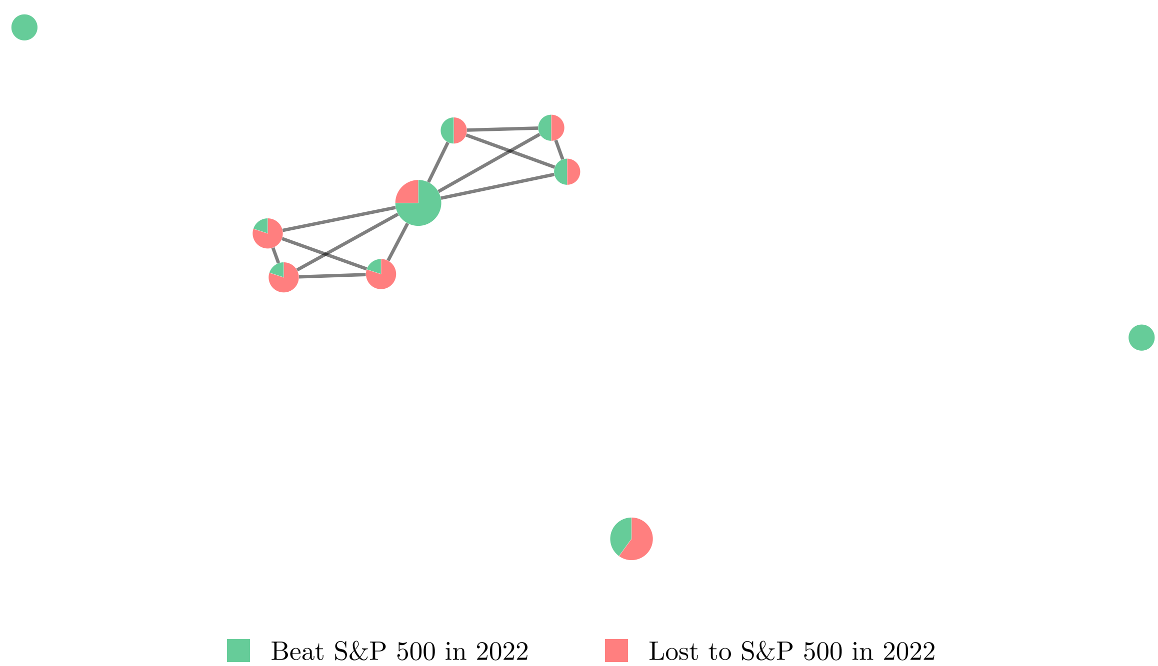
Figure 3: Axcelis Target Matching

5.2 Daily Double

The following figure shows an alternative method for group evaluation. Rather than being interpreted to extract any particular meaningful analysis, this bonus material has been included to expand the notion of group characterization beyond analyzing average behavior.

Visualizing Market Performance:

- *Nodes* are from our original graph model.
- *Edges* are from our original graph model.
- *Pie Charts* display the fraction of companies within each node that either beat or lost to the market in 2022.



While we understand past results are not indicative of future performance, our method is plug-and-play in terms of evaluating group performance using additional metrics. A more advanced performance analysis or stock prediction index could easily be subbed in for stock performance here.

6 Conclusion

In this report, we have analyzed the landscape of 30 middle market technology companies. Without any prior expertise, labels, or training, our algorithm can model multidimensional relationships between these companies and generate informative groups. As mentioned previously, our model goes above and beyond traditional clustering techniques in its ability to capture and thus leverage structural information of our data set. We even use the outputs of `THEMA` to design elementary methods for *target matching* and *evaluating market performance*, which when paired with more sophisticated financial metrics and higher quality data could make `THEMA` a powerful addition to any financial analysis software stack.

Taking a step back, we hope the reader will take note of the generality of this demonstration. The type of analysis that we support and make accessible through `THEMA` is completely general. We can capture high dimensional relationships of *any* type of object, not just companies: patients, users, customers, institutions, are just some of the other complex, high dimensional objects who can benefit from `THEMA`. We *specifically* target problems and data sets where standard techniques in machine learning and statistics fall short, but graph learning and topology excel. However, our methodology can *also* work hand in hand with traditional pipelines, deep learning architectures, and large language models to help improve the performance of your current analysis. We hope this example motivates the power of understanding the global structure of your data and challenge you to improve your ability to define powerful and contextualized evaluation metrics by using `THEMA`.

Appendix

Data Set Description

The companies in this technology middle market cohort were categorized based on their **average total revenue from 2019 to 2023** between $10,000,000 and $1,000,000,000, and were selected based on data availability. Below is a list of our sources, broken down by category:

Financial data

- Yahoo Query
- yfinance

Geo-political data

- MIT Election Data + Science Lab
- The Yale Program on Climate Change Communication
- The Rocky Mountain Institute

ESG data

- CSRHUB ESG Scores

NLP Generated Press Opinion Data

- Yahoo Finance[1]

Data Field	Description
Investment Risk	LLM generated impression from most recent yahoo finance news articles scored 0 to 10 (best=10)
Company Management	LLM generated impression of management from most recent yahoo finance news articles scored 0 to 10 (best=10)
Positivity Outlook	LLM generated impression of how from most recent yahoo finance news articles scored 0 to 10 (best=10)
GW is Real	Estimated percentage per county who believe that most scientists think global warming is happening
Corporations Should Address GW	Estimated percentage per county who think corporations and industy should be doing more/much more to address global warming
Worried About GW	Estimated percentage per county who are somewhat/very worried about global warming
Winning Party	2020 presidential election county majority
Winning Party Votes	2020 presidential total votes in county for the winning party
Winning Party Percent	2020 presidential winning part votes divided by total county votes
Governor Party	State Governor Party
Legislation Majority Party	State Legislature Majority Party

Table 3: Non-Financial Data Field Break Down

[1] We use large language models (LLMs) to generate press opinions scores for companies, based on the most recent news articles available from Yahoo finance APIs in late June 2023. We acknowledge the limitations of this approach, but point out that our model is fully compatible with more sophisticated metrics on press opinion.

Company Name	Ticker
Airgain, Inc.	AIRG
Ambarella, Inc.	AMBA
Amtech Systems, Inc.	ASYS
AXT, Inc.	AXTI
Camtek Ltd.	CAMT
Cohu, Inc.	COHU
CTS Corporation	CTS
FormFactor, Inc.	FORM
Iridium Communications Inc.	IRDM
Kopin Corporation	KOPN
Lattice Semiconductor Corporation	LSCC
MACOM Technology Solutions Holdings, Inc.	MTSI
Mercury Systems, Inc.	MRCY
Nova Ltd.	NVMI
NVE Corporation	NVEC
Power Integrations, Inc.	POWI
Rambus Inc.	RMBS
Semtech Corporation	SMTC
Veeco Instruments Inc.	VECO
Vicor Corporation	VICR
inTEST Corporation	INTT
Marine Products Corporation	MPX
Photronics, Inc.	PLAB
Qualstar Corporation	QBAK
Silicon Laboratories Inc.	SLAB
Haynes International, Inc.	HAYN
Richardson Electronics, Ltd.	RELL
Clearfield, Inc.	CLFD
Universal Display Corporation	OLED
Aehr Test Systems	AEHR

Table 4: Company List

Our dataset contained the following *Financial Variables*:

OperatingCashFlow, FreeCashFlow, CashFlowFromContinuingInvestingActivities, CashFlowFromContinuingFinancingActivities, CapitalExpenditure, NetIncome, BasicEPS, TotalRevenue, EBITDA, OperatingIncome, TotalExpenses, EBIT, OperatingRevenue, TotalAssets, TotalLiabilitiesNetMinorityInterest, StockholdersEquity, CurrentAssets, CurrentLiabilities, OperatingMargin, NetMargin, ReturnOnEquity, ReturnOnAssets, CurrentAssetsToLiabilities, CapexToAssets, CapexToCFI, CapexToRevenue, OperatingIncomePerFTE, EBIT-Margin,