LUISS

Department of Business and Management

Master's Degree in Corporate Finance

Chair of Advanced Corporate Finance

Algorithmic peer selection: Machine learning-driven comparable analysis for private companies valuation

Prof. Raffaele Oriani

SUPERVISOR

Prof. Marco Vulpiani

CO-SUPERVISOR

Matr. Michele Moreschini

CANDIDATE

Academic Year 2024/2025

Abstract

This thesis examines whether unsupervised machine learning can improve peer selection in comparable company analysis (CCA), a cornerstone method for private firm valuation. Traditional approaches, often based on industry codes or analyst discretion, suffer from subjectivity and systematic biases. To address these limitations, the study applies density-based clustering algorithms: DBSCAN, OPTICS and their NLP-augmented variants to a large sample of European and North American healthcare firms.

The empirical results show that clustering delivers valuations at least comparable to industry benchmarks, with mixed performance across methods. DBSCAN effectively identified idiosyncratic outliers but did not significantly outperform the baseline. NLP integration, in its current calibration, frequently collapsed observations into a dominant cluster, limiting its usefulness for comparability. By contrast, OPTICS consistently generated more homogeneous peer groups and significantly improved valuation accuracy. Compared to industry medians, OPTICS reduced mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE), while increasing the proportion of valuations falling within ±20–30% of observed M&A and IPO transaction prices. Statistical tests confirmed that these improvements were significant.

The study contributes to three strands of literature: (i) relative valuation by demonstrating that clustering-based valuations fall within accepted error bands; (ii) bias in peer selection by offering a replicable, data-driven alternative to analyst judgment; and (iii) machine learning in finance by providing evidence that OPTICS captures economic similarity more effectively than traditional classifications. For practitioners, the findings highlight both the potential and the limits of algorithmic peer selection.

Table of Contents

1. Introduction

1.1 Motivation

Valuation by comparables is widely used by practitioners in corporate finance, particularly when valuing private companies in M&A transactions. For all its popularity, however, the method continues to be dogged by ongoing criticisms: peer groups are frequently selected on an ad hoc basis, industry codes may fail to proxy for actual economic comparables and multiples are themselves subject to market cycle and analyst judgment. These weaknesses are especially problematic in the private company context where the absence of reliable market prices coupled with heterogeneous disclosure standards magnify risks of bias and improper valuation. At the same time, advancements in data science provide opportunities to systematize peer selection, boost statistical accuracy and mitigate analyst discretion using clustering algorithms, natural language processing (NLP) and predictive models. The thesis is motivated by tackling these shortcomings by testing algorithmic peer selection methods against traditional benchmarks and thus contributing to both valuation practice and the academic literature.

1.2 Research Questions and Objectives

The central research question is whether algorithmic methods can produce peer groups that improve the accuracy and robustness of comparable company analysis. This question leads to several sub-questions:

- Do clustering algorithms generate more homogeneous and reliable peer sets than industry classifications?
- In terms of accuracy and stability of valuation, how do the different clustering techniques (DBSCAN, OPTICS) compare?
- Do textual similarity measures based on NLP provide a richer notion of comparability than just a numeric financial ratios?
- Beyond valuation multiples, can algorithmic peers be used to estimate risk parameters such as beta and cost of capital for private firms?

The objectives are therefore threefold: (i) to design and test a replicable methodology for clustering-based peer selection; (ii) to benchmark its performance against industry medians using multiple accuracy metrics and statistical tests; and (iii) to extend the framework to risk estimation, assessing its relevance for corporate finance practice.

1.3 Original Contributions

This dissertation includes three contributions. First, it combines unsupervised learning with a traditional valuation, in that algorithmic clustering provides a transparent and replicable peer selection process for publicly traded companies, as opposed to freerein selection, wherein similar companies can be included at the editor's discretion without clear rationale. Second, it incorporates textual similarity through NLP technology in a new readable style to create a reasonable way to account for a qualitative aspect of comparability beyond industry codes in financial clustering. Third, it expands a comparable company analysis to the risk estimation exercise, suggesting that peer firms derived from a clustering analysis can better provide an estimate of betas and WACC for private firms. To my knowledge, there has never been a study that has unified the four dimensions illustrated in this presentation, namely: clustering, textual similarity, risk estimation and an empirical analysis rooted in the M&A transaction literature. The contributions, therefore, intend to provide a toolbox to practitioners while also contributing to the academic understanding of peer groups' formation.

Beyond these points, the thesis also contributes by systematically benchmarking algorithmic methods against conventional approaches using multiple error metrics and statistical tests. This comparative design not only quantifies accuracy gains but also highlights contexts where traditional benchmarks may still perform adequately, offering practical guidance for both researchers and practitioners.

1.4 Structure of the Thesis

The thesis is organized as follows. Chapter 2 reviews the literature on traditional peer selection, valuation multiples and recent advances in clustering and NLP. Chapter 3 presents the methodological framework including data collection, cleaning, clustering design, validation metrics and hypotheses. Chapter 4 reports the empirical results comparing algorithmic peer selection to industry-based benchmarks across valuation accuracy and theorizes on risk estimation tasks. Chapter 5 discusses the implications for valuation practice, limitations of the study and directions for further research.

2. Literature Review & Context:

2.1 Traditional Peer Selection and Valuation Multiples:

2.1.1 Introduction

Relative valuation compares a company's value with the market prices of similar businesses. It is widely used in practice because it is quick and appears to reflect what the market is willing to pay. Surveys by Morgan Stanley show that almost all equity analysts use a market-multiples approach when valuing companies and that price-to-earnings (P/E) and enterprise-value-to-EBITDA (EV/EBITDA) are the most popular multiples.[1]

As Damodaran (2002) notes, relative valuation requires fewer explicit assumptions than discounted cash flow (DCF) methods and its results tend to track the mood of the market. This popularity is part of the reason why multiples remain a cornerstone of sell-side research and investment banking but the method's simplicity hides significant assumptions about comparability, growth and risk. Academic work therefore scrutinises how peer groups are selected and how multiples behave.

2.1.2 Historical Development of Comparable Company Analysis

Valuing a firm by reference to comparable businesses is one of the oldest valuation techniques. Early practitioners used earnings multiples as simple guides to value and courts in the United States relied on comparable companies during the early twentieth century when adjudicating fair values. Economists at Harvard Business School formalised the approach in the 1930s and textbooks such as Graham and Dodd's Security Analysis (1934) popularised earnings-based multiples. The method was adopted by investment banks throughout the post-war era even as academic finance moved towards discounted cash flow methods. From the 1980s onward, scholars began to test and refine the technique.

Empirical studies found that selecting peers from the same industry improves accuracy and that further gains are achieved when firms are matched on fundamentals such as growth and risk (Alford, 1992). Over the last two decades, classification standards and large databases have made peer selection more systematic but the underlying idea remains: the value of a firm can be inferred from how similar companies trade.

[1] Morgan Stanley Global Insight, *Valuation Multiples: What They Miss, Why They Differ, and the Link to Fundamentals*, April 2024.

2.1.3 Valuation Multiples and Their Behaviour

Valuation multiples are standardized ratios relating a firm's market value to a financial metric and form the foundation of comparable company analysis; they are broadly classified into equity-based and enterprise value-based multiples.

Equity-based multiples, such as Price to Earnings and Price to Book ratios, compare the market value of equity to earnings or book value and are affected by leverage this makes them sensitive to differences in capital structures. Enterprise value-based multiples like EV/EBITDA and EV/Sales measure the value of the entire firm incorporating both debt and equity; EV/EBITDA links enterprise value with operating profits reducing distortions from capital structure while EV/Sales is useful for loss-making firms but less precise as it ignores profitability.

The distinction between equity-side and EV-side multiples is crucial when selecting benchmarks: equity multiples suit firms with similar leverage, while EV multiples allow comparisons independent of capital structure. Their distributions are typically positively skewed, meaning a small set of highly valued firms drives averages upward. Consequently, the median better represents central valuations than the mean.

Alford (1992) demonstrates that selecting peers based on industry significantly reduces valuation errors when using P/E multiples, though deviations from actual market prices remain substantial. Liu, Nissim, and Thomas (2002) show that forward-looking earnings multiples consistently achieve superior accuracy compared to trailing earnings, cash flow, book value, or sales-based measures. In their study, approximately half of firms valued using forward earnings multiples fall within ±15% of market prices. Lie and Lie (2002) reinforce these findings, observing that EBITDA-based multiples outperform EBIT-based ones, as excluding depreciation and amortization standardizes cash flows. They also find that asset-based multiples can provide less biased valuations than earnings or sales-based multiples, especially when combined with forward-looking forecasts. Schreiner and Spremann (2007) conclude that equity-based multiples tend to be more accurate than enterprise value multiples when estimating market prices, confirming the advantage of forward-looking measures and highlighting that integrating intangibles improves valuation precision.

Overall, multiples' effectiveness depends on aligning the choice of multiple with the valuation objective, accounting for leverage differences, favouring forward-looking earnings and working correctly with their statistical properties. These empirical results are synthesized in Table 1, which summarizes the accuracy of different valuation multiples in developed markets.

*Accuracy of Valuation Multiples in Developed Markets (**Table 1**)*

Study	Sample	Multiples tested	Key Findings	Key Insight
Alford (1992)	US firms, 1978–1987	P/E	Industry-based peers reduce error variance significantly	Industry restriction improves accuracy
Liu, Nissim & Thomas (2002)	US firms, 1979–1999	Forward vs trailing earnings, book, cash flow, sales	~50% of forward P/E valuations within ±15% of price	Forward earnings most accurate
Lie & Lie (2002)	US firms, 1982–1999	EBIT, EBITDA, sales, assets	EBITDA multiples outperform EBIT; asset mult. less biased	EBITDA superior for comparability
Schreiner & Spremann (2007)	European firms, 1996–2001	Equity vs EV multiples	Equity multiples closer to mk. prices	Equity-side measures more precise

Source: Author's elaboration based on Alford (1992), Liu et al. (2002), Lie & Lie (2002), Schreiner & Spremann (2007).

2.1.4 Industry Classification and the Role of SIC/GICS Codes

A fundamental question in peer selection is how to define "similar" companies. Traditionally analysts start by selecting firms in the same industry. Formal classification systems have standardized this process: The U.S. introduced the Standard Industrial Classification (SIC) in the 1930s and later replaced by the North American Industry Classification System (NAICS) in 1997. In 1999, Standard & Poor's and MSCI launched the Global Industry Classification Standard (GICS), designed for investors and now dominant worldwide.

Industry codes are a natural starting point as firms in the same sector usually share drivers of value such as growth, margins and risk factors. Academic research supports this approach: Alford (1992) showed that restricting comparables to the same industry improves valuation accuracy relative to selecting peers based on size or growth alone. Narrower industry definitions reduce valuation errors up to a point, though Alford also observed that larger firms exhibit lower errors because they are easier to match with stable peers.

However, classification codes are imperfect. Many corporations operate across segments and SIC or NAICS often group firms with different business models. Bhojraj, Lee, and Oler (2003) compared SIC, NAICS, Fama-French sectors and GICS, finding that GICS outperforms older schemes by producing peer groups with greater homogeneity in returns, valuation multiples and financial indicators.

Recent studies explore alternative approaches. Kaustia and Rantala (2021) define peers based on common analyst coverage, showing that companies followed by the same analysts are substantially more financially similar than those grouped only by industry codes. Similarly, Cooper and Cordeiro (2008) find that selecting just 5–10 peers matched on growth rates can achieve valuation accuracy comparable to using an entire industry group. Consistent with this, Liu et al. (2002) demonstrate that incorporating key value drivers such as growth and profitability when assembling peer sets improves accuracy significantly.

In summary, industry classification remains a crucial starting framework, with GICS generally preferred. However, best practice combines modern codes with firm-specific comparability criteria such as size, growth, leverage and business mix to construct peer groups that best reflect the target firm's economic reality.

2.1.5 Analyst Practices in Selecting Comparable Firms

Academic research often adopts broad peer groups, including all firms within an industry, whereas practitioners are far more selective. In investment banking and equity research, peer groups typically comprise 5 to 10 firms considered most similar to the target carefully chosen through a mix of qualitative and quantitative criteria: Qualitatively analysts focus on similar products, business models and geographic exposure. Quantitatively they screen for comparable size, growth, margins and leverage. The objective is to identify firms whose valuation multiples provide reliable benchmarks. Cooper and Cordeiro (2008) show that a small, well-matched peer set achieves an accuracy comparable to using a broad industry sample.

In practice the process often begins with industry peers and refine the list by excluding outliers or mismatched firms such as conglomerates when valuing a pure-play company. The following trade-off emerges: too few peers risk over-reliance on idiosyncratic results while too many dilute comparability. By contrast, academic studies select larger peer groups to avoid selection bias, whereas practitioners prioritize representativeness based on sector expertise.

Once the peer set is defined, analysts derive benchmark multiples, typically the median EV/EBITDA and P/E, using current prices and financial data. The median is preferred due to the positive skewness of multiples distributions. Applying these benchmarks to the target's metrics provides implied valuation ranges but practitioner judgment often introduces qualitative adjustments as, for example, a company with superior growth or profitability may justify a multiple above the median while weaker fundamentals can warrant a discount. Meitner (2006) notes that analysts frequently adjust valuations using quartile positioning or "rules of thumb" to reflect perceived relative strengths or weaknesses.

Context also matters. In overheated markets analysts may avoid peak multiples, whereas during downturns they may consider historical averages or resilient peers. Sell-side analysts often complement relative valuation with DCF models but Imam, Barker and Clubb (2008) find that multiples, especially P/E, remain highly influential in setting target prices because they reflect market consensus and are easily communicated to investors while DCF serves as a cross-check.

In summary, practitioner-driven peer selection is pragmatic and experience-based. It relies on small, representative groups, uses median multiples to establish benchmarks and integrates qualitative judgment to refine valuations. While relative valuation grounds analysis in market data, peer choice and multiple interpretation introduce subjectivity.

More recently, Cooper and Lambertides (2022) extend this line of research by formally analyzing the optimal number of comparables to include in multiples-based valuation. They find that there is a trade-off: using too few peers increases idiosyncratic noise, while including too many dilutes comparability. Their results suggest that intermediate peer group sizes balance accuracy and robustness, complementing earlier evidence by Cooper and Cordeiro (2008) that even a small, well-matched set of peers can rival industry-wide benchmarks.

2.1.6 Accuracy and Biases of Traditional Peer Selection

A crucial question for the comparable companies approach is: How accurate are peer-based valuations and what biases might affect them? Over decades, empirical research has shed light on these issues by testing how well multiples predict actual market values and by observing how practitioners choose peers in practice.

Traditional peer selection can be accurate, though valuation errors of 20–30% remain common yet researchers often remark that such error magnitudes are acceptable given the volatility in equity markets. Using multiple methods can narrow uncertainty.

Liu et al. (2002) provided an optimistic benchmark with forward earnings multiples, finding half of valuations within ±15% of market prices. Schreiner and Spremann (2007) similarly concluded that in European markets, well-chosen multiples "approximate market values reasonably well", especially when forward-looking metrics are used. These results suggest that when a company has clear comparables and stable expected performance, relative valuation can be quite powerful. Consistently, Alford (1992) found that valuation errors were smaller for larger firms, attributing this to their more predictable performance and easier comparability.

However, accuracy tends to deteriorate in certain scenarios. Peer valuation is less reliable for firms at extreme ends of performance e.g., very fast-growing firms, young unprofitable firms, or firms in distress. In these cases, standard multiples can mislead. Kim & Ritter's (1999) study of IPO valuations is illustrative: they found not only were multiples-based appraisals of IPOs biased upward (suggesting systematic overvaluation) but also the dispersion of errors was high. In their sample, only 12–27% of IPO valuations (depending on the multiple used) fell within 15% of the eventual market price, far worse than for mature companies. This reflects the fact that early-stage companies often lack meaningful earnings and their peer groups, if drawn from older, established firms they may not be truly comparable in growth outlook. Similarly, Gilson, Hotchkiss & Ruback (2000) examined companies emerging from bankruptcy and found that multiples had limited accuracy there (with only ~21% within 15% error bands). By contrast, Kaplan & Ruback (1995) looked at highly leveraged transaction (LBO) targets and found that a simple EBITDA multiple valuation was often as effective as a full DCF in explaining deal values. In fact, in their LBO sample, a healthy proportion (37–58%) of valuations by EBITDA multiples were within 15% of actual prices.

Beyond questions of accuracy, potential biases in peer selection have attracted attention. Table 2 synthesizes the main empirical findings on biases in analyst-selected peer groups (IPOs, equity research and M&A fairness opinions), providing effect sizes where available.

*Documented Biases in Analyst-Selected Peer Groups (**Table 2**)*

Study	**Context**	**Bias type**	**Quantitative effect**	**Implication**
Paleari, Signori & Vismara (2014)	IPOs (Europe)	Optimistic bias by underwriters in selecting peers	Underwriter-selected peers have valuation multiples 13–38% higher than algorithmic	Inflates IPO valuation ranges and justifies higher offer prices
Vismara, Signori & Paleari (2014)	IPOs (Europe)	Post-IPO peer switching (selection not stable over time)	Pre-IPO disclosed peers differ from post-IPO; pre-IPO sets tilted toward higher-multiple firms	Supports optimistic pricing narratives before listing
Franco, Hope & Larocque (2015)	Equity research (US)	Optimism bias in sell-side peer selection	Analysts tilt toward high-valuation peers; association with optimistic target prices	Upward-biased target prices; reduced objectivity
Eaton, Guo, Liu & Officer (2022)	M&A fairness opinions (US)	Incentive-consistent strategic peer selection	Advisors select comparables beyond SIC; shift of valuation ranges depending on incentives/litigation	Makes negotiated deal prices appear “fair” within constructed ranges

Source: Author’s elaboration based on Paleari, Signori & Vismara (2014), Vismara, Signori & Paleari (2014), Franco, Hope & Larocque (2015), and Eaton, Guo, Liu & Officer (2022).

One concern is confirmation bias or motivated reasoning: an analyst might subconsciously (or deliberately) select peers that make their target look favourably valued by selecting high-multiple peers to inflate, or low-multiple peers to depress valuations. Recent empirical evidence confirms that such strategic behaviour occurs. Eaton et al. (2023, JFE) examine M&A fairness opinions and find systematic bias in peer selection by investment banks. When target companies faced heightened risk of shareholder litigation for undervaluation, banks responded by cherry-picking lower-valued peer firms to use in their comparable analysis. By selecting peers with lower EV/Sales multiples, bankers made deal prices appear at the high end of 'fair'. This downward bias in peer selection was shown to reduce the likelihood of lawsuits by making the negotiated price look generous. While this behaviour may protect the deal from legal challenges, it raises obvious concerns about objectivity: the comp analysis in such cases is not an unbiased estimate of value but rather a defensive tool shaped by advisor incentives.

Beyond classification codes, econometric approaches have also been proposed to improve peer group formation. Asche and Misund (2007), focusing on oil and gas companies, apply Chow tests of structural similarity to ensure that selected peers share consistent value drivers. Their evidence shows that algorithmic, econometric selection yields more homogeneous and comparable groups than traditional industry-based methods, thereby reducing systematic valuation errors.

Even without intent, subjective peer choice leads to variability, as analysts may emphasize either high-growth or conservative peers. Herding can also occur, as in the dot-com bubble with unconventional multiples. The history of valuation is replete with such market-driven biases in multiples: what investors deem relevant (and thus which comps they consider) can swing with market sentiment. Academic observers caution that analysts should remain vigilant about these distortions. For example, Bradshaw (2002) documented that U.S. sell-side analysts often justify target prices by invoking certain multiples that align with their earnings forecasts, potentially leading to optimistic price targets in bull markets. Consistent with this, Imam and colleagues (2008) found UK analysts frequently use DCF as a presentational veneer while their actual target prices are grounded in multiples and subjective judgment a practice that can conceal biases under the appearance of rigorous analysis.

Regional and cross-country differences also shape the behaviour of multiples. Although their general properties are consistent, average levels and valuation accuracy vary with institutional settings. European firms, for example, have historically traded at lower P/E ratios than U.S.

firms so analysts often apply a discount when using U.S. peers looking at emerging markets, the challenges are greater: fewer comparable listed companies, uneven reporting standards and divergent risk profiles reduce reliability. Empirical evidence shows that locally derived multiples provide better accuracy than foreign benchmarks (Pinto et al., 2019). As a result, practitioners increasingly adjust multiples for country risk or apply global peer groups cautiously, sometimes incorporating macroeconomic corrections.

2.1.7 Conclusion

Traditional Peer group valuation remains a widely used technique in corporate finance because it links a firm's value to the market pricing of comparable businesses. Its strength lies in practicality and ease of communication: investors and managers readily understand valuations expressed through common multiples, unlike discounted cash flow models which rely on explicit forecasts and discount rates, the multiples build directly on observable market data.

Research shows that, when applied carefully, this method can achieve reasonable accuracy. Liu, Nissim, and Thomas (2002) report that forward earnings multiples reduce valuation errors, with about half of estimates within 15 percent of market prices. Alford (1992) demonstrates that choosing peers from the same industry lowers errors compared to broader peer groups. Studies also highlight the benefits of using medians instead of means to correct for skewed distributions of multiples. Together, these findings outline clear best practices: focus on forward-looking metrics, restrict peers to similar industries and rely on robust statistical measures.

However, limitations are equally well documented. Accuracy is weaker for IPOs, distressed firms, or young high growth companies where comparability is limited. Kim and Ritter (1999) show that only a minority of IPO valuations fall within acceptable error bands. Gilson, Hotchkiss, and Ruback (2000) find that multiples perform poorly for firms emerging from bankruptcy. By contrast, Kaplan and Ruback (1995) find that EBITDA multiples are effective in leveraged buyouts, where companies are mature and cash generating.

Bias is another concern. Analysts may select peers strategically to support specific outcomes, as documented by Eaton et al. (2023) in M&A fairness opinions. Even without intent, subjectivity leads to variation and market sentiment can drive which multiples are applied. Examples such as the dot com bubble illustrate how prevailing norms can distort valuations.

Cross country research adds further nuance. Pinto et al. (2019) find that local benchmarks improve accuracy in emerging markets, where institutional settings, reporting standards and risk profiles differ from developed economies.

2.2 Machine Learning Approaches to CCA:

2.2.1 Introduction

Machine learning (ML) has begun to significantly influence corporate valuation and comparable company analysis offering data-driven solutions to longstanding challenges such as peer selection bias, inconsistent multiple estimates and unreliable cost of capital inputs. Recent research demonstrates that ML techniques from clustering algorithms to advanced predictive models, can enhance the accuracy and objectivity of relative valuation. In this section, we review how ML is being applied to CCA, organized into specific areas: general applications of ML in valuation, clustering methods for peer group formation, unsupervised learning for multiples estimation, clustering for beta and WACC estimation, NLP-based textual similarity for peer identification and validation of these approaches against market outcomes.

2.2.2 ML in Corporate Valuation

Machine learning is increasingly used to improve valuation models by detecting complex patterns in financial data. In out-of-sample tests ML models have been shown to outperform traditional valuation in accuracy and consistency. For example, Geertsema and Lu (2023) use gradient boosting for relative valuation and report substantially higher valuation accuracy than standard multiples regressions. Their ML-predicted valuations behave like fundamental values: stocks that the model flags as over-valued tend to experience negative returns while under-valued stocks see price increases. Moreover, the key drivers identified by the ML model are profitability, growth and efficiency metrics, align with theoretical expectations.

Another innovation by Geertsema & Lu is an interpretable framework to express the ML valuation output in terms of comparable firms. They derive each firm's predicted multiple as a weighted average of peer firm multiples, where the weights indicate each peer's importance or comparability. In other words, rather than analysts manually picking comparables, the ML model implicitly learns which companies are most comparable and how much influence each should have on the valuation. This approach helps address the "black box"[2] concern by explaining valuations in terms of peer contributions. Overall, the use of ML techniques, including tree-based ensembles and neural networks, in corporate valuation is proving to improve accuracy and reduce human bias, while still grounding results in fundamental drivers.

[2] A key limitation of machine learning in valuation is the *black box problem*, where models deliver predictions without transparent reasoning, reducing interpretability for analysts and investors.

Table 3 summarizes the empirical evidence comparing machine learning with traditional comparable company analysis across different settings. The studies consistently show that ML models, especially tree-based and ensemble methods, reduce valuation errors and enhance predictive accuracy compared to analyst-driven or regression-based approaches.

*Machine Learning vs. Traditional Comparable Company Analysis (**Table 3**)*

Study	Context / Dataset	Method compared	Key Findings	Key insight
Geertsema & Lu (2023)	Listed firms, relative valuation	Gradient boosting vs. multiples regression	Substantially higher valuation accuracy: mispriced firms show subsequent return reversals	ML valuations behave like fundamentals; interpretable via peer-weights.
Jagrič et al. (2024)	20,000+ private-firm transactions	Self-Organizing Maps (SOM) vs. linear baselines	SOM wins the vast majority of head-to-head comparisons; lower bias and error across multiples	Unsupervised ML learns better peer groups for multiples.
Alanis et al. (2024)	Small, young and private firms	ML models vs. comparable-company beta estimation	MAE reduced by >42% vs. CCA (out-of-sample, 1990–2021)	Better beta inputs tighten valuation ranges.
Staňková (2024)	U.S. public firms	Gradient boosting decision trees vs. industry multiples	~24 p.p. decrease in median percentage error	Boosted trees improve multiples-based valuation.
Hanauer et al. (2022)	European stock markets	Tree-based ML vs. linear regression	48–66 bps/month risk-adjusted returns vs 11–36 bps for linear	ML enhances fundamental analysis and valuation relevance.
Vayas-Ortega et al. (2020)	Cross-sector, emphasis on utilities	Ensembles (bagging, SVR, GPR) vs. DCF	RMSE as low as 1.21 (utilities), outperforming DCF in that sector	Ensembles robust for EV prediction.

Source: Author's elaboration based on Geertsema & Lu (2023), Jagrič et al. (2024), Alanis et al. (2024), Staňková (2024), Hanauer et al. (2022), and Vayas-Ortega et al. (2020).

As highlighted in Table 3, the improvements are both statistically and economically significant, ranging from error reductions of 24–42% to superior risk-adjusted returns. These results confirm the potential of ML methods to provide more robust and objective valuation benchmarks than traditional CCA.

2.2.3 Clustering for Peer Groups

A critical step in CCA is selecting a set of truly comparable peer companies. Traditionally analysts rely on industry classifications (e.g. SIC or NAICS codes), growth, profitability and subjective judgment to choose peers which often introduces bias and inconsistency. Studies have documented that human-selected peer groups can be systematically biased, for instance, underwriters and equity analysts tend to cherry-pick peers with inflated valuation multiples to justify higher target prices or IPO valuations. Paleari, Signori & Vismara (2014) found that the companies chosen by underwriters for IPO comps had valuation multiples 13–38% higher than those of ML selected peer groups and this indicates a strong optimistic bias in manual selection.

Clustering algorithms offer a way to form peer groups based on measurable similarity, reducing reliance on crude industry codes or gut feeling. In a pioneering example, Ingram and Margetis (2010) applied cluster analysis on firm fundamentals to group similar companies and derive "proxy" betas for private firms. By clustering companies with analogous financial characteristics, they could assign a non-public firm to a peer cluster and use that cluster's average beta as an estimate of its systematic risk. This demonstrated the feasibility of clustering to replicate market-based metrics for firms lacking market data. Recent research has introduced systematic approaches to automate peer selection using unsupervised clustering. For instance, Jagrič et al. (2024) apply self-organizing maps on a large universe of private-firm transactions, showing that algorithmic clustering can consistently outperform traditional regression-based peer grouping in terms of valuation accuracy. Similarly, Gerling (2023) employs text embeddings of corporate website content combined with k-means clustering to form economically intuitive peer groups, demonstrating that such methods produce more homogeneous and replicable sets of comparables than traditional industry codes. These data-driven approaches yield objective, repeatable peer groups without analyst intervention, reducing the time required to construct comps while improving consistency and reliability.

Crucially, clustering for peer groups also mitigates the problem of overly broad industry classes that "often fail to capture true economic similarity", as noted in valuation research. Two firms with the same 4-digit SIC code might have very different business models, while two firms in different industries could in fact compete closely or have similar financial profiles. Clustering algorithms can detect such nuanced similarities. Eaton et al. (2022) find that investment banks in M&A contexts pay close attention to product market space and other similarities when choosing comps and that standard industrial classifications do a poor job of capturing related firms in these situations. This evidence supports using data-driven similarity rather than relying solely on industry labels.

In summary, clustering techniques enable data-defined peer groups that are more homogeneous in size, growth, profitability and other value drivers, thereby improving the comparability of multiples and reducing bias in CCA outcomes.

2.2.4 Unsupervised Clustering for Multiples

Unsupervised learning methods have shown promise in improving how valuation multiples are derived from peer groups. Rather than pre-specifying comparison criteria, unsupervised algorithms can learn the optimal dimensions of similarity and group firms accordingly. A notable example is the application of self-organizing maps (SOMs), a type of neural network that performs clustering. Jagrič et al. (2024) conduct an extensive study on private firm valuation using multiples, asking whether AI algorithms can "learn" better peer groups. Using a dataset of over 20,000 private company transactions, they compare a SOM-based peer selection model against a traditional linear regression approach that uses a fixed set of peer selection criteria. The results are striking: the SOM consistently outperforms the traditional model, achieving lower valuation errors and bias across five common valuation multiples. In direct head-to-head tests, the machine learning approach produced more accurate estimates of transaction multiples and statistical tests confirmed its prediction superiority in roughly 95% of cases (i.e. the ML valuation was closer to the actual deal price in most transactions). These findings strongly indicate that unsupervised clustering can optimize peer groups and valuation estimates beyond what is possible with predefined industry or financial filters. The SOM was essentially able to determine which past transactions (peers) provide the best valuation benchmark, without human bias or an arbitrary rule limiting the peer set.

*Unsupervised Clustering in Peer Group Formation (**Table 4**)*

Study	Context / Dataset	Method	Key Findings	Key insight
Jagrič et al. (2024)	20,000+ private-firm transactions (OECD, Zephyr/Orbis)	Self-Organizing Maps (SOM) vs. OLS	SOM outperforms OLS across 5 multiples; consistently lower error and bias	SOM identifies more homogeneous peer groups, improving valuation accuracy
Ding et al. (2019)	US firms, accounting misstatement detection	k-medians clustering on financial ratios	Adding cluster-based peer variables improves detection accuracy	Clustering captures firm similarity beyond standard industry codes
Nanda et al. (2010)	Indian stock market	k-means, fuzzy c-means, SOM	k-means produced most compact and stable clusters for portfolio grouping	Cluster methods can define coherent economic cohorts
Gerling (2023)	German firms, Company2Vec embeddings	k-means on text embeddings	Peer groups more intuitive and homogeneous than SIC/GICS	Text+clustering improves peer comparability vs. static industry labels
Ingram & Margetis (2010)	Private firm beta estimation (US)	Cluster analysis on fundamentals	Clusters used to proxy betas for non-public firms	Clustering extends to risk estimation where market data is missing

Source: Author's elaboration based on Jagrič et al. (2024), Ding et al. (2019), Nanda et al. (2010), Gerling (2023), and Ingram & Margetis (2010).

Other clustering approaches have similarly been used to systematize multiples analysis. Jagrič et al. (2024) combine unsupervised clustering with nearest-neighbour selection to derive valuation multiples: after grouping firms by similarity, they identify the closest peers within each cluster and use their median multiples for valuation. Applied across sectors (e.g. Technology, Health Care, Industrials), this approach captures groupings that traditional industry-based methods might overlook. Evidence indicates that such unsupervised grouping yields more coherent peer cohorts and thus more stable multiple estimates. The benefit is twofold: it reduces subjectivity (the algorithm defines the peer group instead of an analyst picking companies to fit a narrative) and can dynamically adjust peer groups as data changes (new companies or shifting fundamentals automatically alter cluster membership).

In summary, research to date suggests that unsupervised clustering methods (such as hierarchical algorithms, k-means variants and neural clustering like SOMs) can improve the accuracy of multiples-based valuations by finding optimal peer groups and weighting schemes.

2.2.5 Clustering for Beta and WACC

Estimating a private or thinly traded company's cost of capital (cost of equity, beta and by extension WACC) is another area where machine learning and clustering are making inroads. Traditional beta estimation for private firms often relies on selecting public comparables (by industry and size) and unlevering/relevering betas. ML approaches have demonstrated the ability to improve these estimates. Ingram and Margetis (2010) provided an early example by using cluster analysis on accounting fundamentals to find public companies similar to a given private firm and then using the cluster's average beta as a proxy. While they did not report numerical performance metrics, the conceptual result was that clustering-based peer betas closely replicated market-derived cost of equity for the target firm. This suggests clustering can stratify companies into risk buckets more finely than broad industry groupings, yielding more appropriate beta benchmarks for otherwise hard-to-value firms.

More recently, a body of work has emerged applying modern ML algorithms to beta and cost of capital estimation. A review by Drobetz et al. (2021, 2024) finds that techniques such as random forests, neural networks and gradient-boosted trees significantly outperform ordinary least squares in predicting stock betas. Across various markets, these ML models achieved 20–59% lower mean squared errors in beta estimates compared to the conventional regression approach, the improved beta accuracy also translated into better portfolio outcomes: for example, portfolios constructed with ML-estimated betas had higher Sharpe ratios and closer to market-neutral risk profiles than those using traditional betas. Alanis et al. (2024) likewise report that machine learning can sharply reduce beta estimation error. In their study focusing on small, young, or private firms, an ML model (incorporating nonlinear relationships and broader data) reduced mean absolute error by about 42% versus the standard comparable-company technique. In practical terms, this could mean a much more reliable unlevered beta input for computing WACC in private firm valuations.

Clustering itself directly features in some approaches to cost of capital. As noted, cluster analysis can group firms by risk characteristics; one application is to identify a cluster of analogues for a private company and use the cluster's average cost of equity or WACC as a proxy. This kind of method ensures that the size, growth, leverage and industry risk factors of the peers closely mirror the target firm. The evidence from Sakouvogui and Nganje (2019) is also relevant: while not a clustering study per se, they show that applying cross-validation (a machine learning validation technique) to CAPM regression yields smaller error metrics in estimating beta than the usual full-sample regression. This implies that more robust, data-driven techniques help avoid overfitting and noise in risk estimation. Another innovative example is Wang and Chen (2023), who integrate a four-layer neural network into a dynamic asset pricing model; their model produced excess return predictions that beat a static CAPM by 118%.

In summary the research consensus is that ML and clustering can improve both equity risk estimates and overall cost of capital assessments. By leveraging richer data (including accounting ratios, macro variables and even textual risk disclosures) and flexible algorithms these methods outperform blunt heuristics like "use the industry beta." The upshot for CCA is that not only can ML yield better multiples but it can also refine the discount rates used in valuation. For instance, Alanis et al. (2024) demonstrate that an ML-based beta input would substantially tighten the valuation range for a private company compared to a traditional peer beta approach.

2.2.6 NLP and Textual Similarity

A rapidly developing frontier in comparable analysis is the use of natural language processing (NLP) to gauge similarity between companies based on textual disclosures and descriptions. The intuition is that financial statements, company descriptions and filings contain rich information about a firm's markets, products and strategy, aspects often missed by numeric factors alone. Traditional peer selection rarely considers such qualitative information, beyond the coarse industry category. For example, Hoberg and Phillips (2016) introduced "text-based network industry" classifications using business descriptions in 10-K filings, finding that companies connected via textual similarity are indeed economically closer than those just sharing a SIC code. In the context of M&A, Eaton et al. (2022) find that product market similarity (measured via text analysis of business descriptions) is one of the most important factors in investment banks' peer selection, whereas standard SIC industry codes "do a poor job" of capturing related firms. In many merger deals, the chosen comparable companies differ in SIC code from the target, yet have high overlap in product/service language, indicating that bankers implicitly prioritize textual/strategic similarity over naive industry classification. This underscores the value of NLP-driven approaches: by parsing the language companies use about themselves, algorithms can uncover comparability along dimensions like technology, customer base, or supply chain position that numeric metrics or industry labels might not reveal.

Accordingly, recent studies and tools have applied NLP to construct better peer groups. One approach is to compute textual similarity scores between firms (using techniques like TF-IDF, word embeddings, or BERT on corporate descriptions) and then cluster or match firms with high similarity. A practical example is provided by Gerling (2023), who applies company embeddings ("Company2Vec") derived from corporate websites and clusters them with k-means, producing peer groups that align more closely with economic reality than SIC/GICS classifications. Similarly, Covas (2023) demonstrates that Wikipedia business descriptions can be processed with GPT-based named entity recognition to identify comparable firms, highlighting the potential of textual data for systematic peer selection. The resulting clusters tend to align with intuitive peer sets for instance, if two companies both heavily mention "cloud software" and "enterprise clients" in their descriptions, an NLP clustering will group them even if they belong to different official industries. These methods can capture comparables that a purely financial screen might miss. Hybrid approaches further combine textual clustering with financial similarity measures (e.g., KNN on ratios within each cluster) to ensure peers are aligned on both qualitative business model and quantitative characteristics.

Evidence is mounting that textual similarity correlates with valuation multiples and outcomes. Firms identified as close peers via text have more comparable valuation ratios and transaction multiples than those just picked by industry. In M&A deal pricing, targets with highly text-similar peers tend to have those peers used as comps and those comps lead to more justified (less biased) valuations. Additionally, textual analysis can highlight aspects like business model risk or innovation, which affect valuation multiples (e.g., two biotech companies might both be pre-revenue but a text analysis could reveal they focus on different therapeutic areas, affecting how comparable they truly are). By incorporating NLP, ML-driven CCA models become "context-aware," capturing strategic comparability. Eaton et al. (2021/2022) even suggest that using NLP-based peer identification could improve the fairness and accuracy of merger valuations by reducing the reliance on broad industry benchmarks. In sum, NLP offers a powerful complement to clustering: it adds a semantic layer of similarity, helping algorithms respect the qualitative nuances of what makes two companies comparable. The convergence of clustering and NLP in peer group formation is a very recent development.

2.2.7 Validation Against Market Outcomes

No matter how sophisticated the methodology, valuation techniques ultimately must be tested against real market prices to gauge their effectiveness. In the context of comparable company analysis (CCA) and machine learning–driven approaches, a critical question is whether these methods produce valuations that are closer to actual transaction values or market prices than traditional methods.

In IPOs, peer-based multiples tend to track actual offer prices closely. For example, Najar and Paré (2019) report that in French IPOs, earnings multiples and comparable transactions yielded average valuation ratios of 1.04 and 1.06 relative to offer prices, while discounted cash flow methods systematically overvalued firms by 30–47%. Similarly, Ong et al. (2021) find that in Malaysian IPOs, the price-to-sales multiple most accurately explained book-built offer prices, outperforming P/E and P/B multiples. In an Indonesian sample, Handaya, Warganegara and Warganegara (2009) show that selecting peers based on the closest return on assets significantly reduced valuation errors compared to simple industry averages. Furthermore, Tizniti and Aasri (2021) demonstrate that combining multiples, dividend discount models and discounted cash flow in a weighted approach reduced mean valuation error from 21.24% to 14.25% and improved explanatory power (R^2 from 86.9% to 91.7%). Collectively, these findings affirm that carefully chosen comparables can yield valuations very near actual market prices in IPO.

In mergers and acquisitions, validation results are more mixed and highly context dependent. In private company takeovers, Grbenic (2021) finds that enterprise value to total assets outperforms other multiples, while Dessaint, Olivier, Otto, and Thesmar (2017) show that CAPM-based valuations overestimated deal values by 12–33%, particularly for low-beta targets. By contrast, Seppä and Laamanen (2001) report that a binomial option-based valuation model achieved an R^2 of 39.3% and an average error of –9%, outperforming a DCF benchmark.

Importantly, studies such as Handaya et al. (2009) also highlight that algorithmic selection of comparables, based on metrics like return on assets, outperforms traditional approaches. This suggests that clustering or ML-based peer selection may further improve valuation accuracy when tested against realized transactions.

To date, however, most ML-driven valuation frameworks have not been fully validated against actual IPO or M&A prices, representing a significant gap in the literature. While some evidence indicates promise, such as ML valuations predicting subsequent stock returns (Geertsema & Lu, 2023) or outperforming linear benchmarks in transaction samples (Jagrič et al., 2024), systematic ex-post validation remains scarce. Robust evaluation using realized transaction multiples, IPO offer prices and post-deal outcomes is essential to demonstrate external validity. Metrics such as mean absolute percentage error, valuation bias and rank-order accuracy should be employed consistently to assess whether ML or clustering-based peer selection methods outperform traditional CCA in practice.

Table 5 summarizes the main validation studies comparing estimated values with realized IPO and M&A prices across different contexts.

*Validation of Valuation Methods in M&A Settings (**Table 5**)*

Study	Context / Dataset	Method	Key Findings	Key insight
Tizniti & Aasri (2021)	IPOs (multi-country)	Weighted mix: DCF, DDM, multiples	Weighted approach cuts mean error (21.2% to 14.3%); R^2 from 86.9% to 91.7%	Blending methods improves robustness
Grbenic (2021)	Private-firm M&A	EV/Total Assets vs. EV/EBITDA, EV/Sales	EV/Total Assets most accurate; EV/Sales weakest	Context-specific multiples matter in M&A
Dessaint, Olivier, Otto & Thesmar (2017)	US M&A	CAPM vs. market values	CAPM overestimates by 12–33% (esp. low-beta firms)	CAPM misvalues targets; multiples often closer
Seppä & Laamanen (2001)	Venture/M&A deals	Binomial option model vs. DCF	Binomial model R^2 = 39.3%, avg error –9%; DCF biased and less accurate	Option-based approaches outperform DCF in uncertain settings

Source: Author's elaboration based on Tizniti & Aasri (2021), Grbenic (2021), Dessaint et al. (2017) and Seppä & Laamanen (2001).

As shown in Table 5, multiples generally perform best in IPO settings, while M&A outcomes are more heterogeneous and context dependent. These findings reinforce the need for systematic validation of new ML-based methods against realized market outcomes.

2.3 Research Gaps and Chapter Conclusions

2.3.1 Identified Research Gaps

While the intersection of machine learning and comparable company analysis is a growing field, several important gaps remain in the literature. First, no study so far has combined unsupervised clustering and NLP-based firm profiling into a unified peer selection framework validated against market outcomes. Clustering techniques have been explored (Section 2.2.3) and textual similarity metrics have been proposed (Section 2.2.5) but researchers have yet to jointly leverage these tools to select comparables for private firm valuation. The potential synergies, using financial data clustering to ensure quantitative similarity and NLP to ensure strategic similarity, are currently untapped. Moreover, existing studies typically examine one aspect at a time (e.g. clustering for multiples, or ML for beta, or text-based similarity in M&A) rather than an end-to-end valuation process. This siloed approach leaves open the question of how an integrated ML-driven CCA system would perform.

Secondly, external validation of ML approaches in valuation is scant. As noted in Section 2.2.6, we have evidence that traditional comparables can be accurate against market prices but we lack published studies showing that a clustering/NLP/ML-based peer selection yields statistically significant improvements in prediction error when tested on actual deals or IPOs. Many ML valuation papers focus on error metrics within their sample or predictive power for stock returns (a form of validation) but not on realized transaction comparisons. This is a crucial gap to fill in order to convince practitioners of the reliability of these new methods.

Another gap relates to algorithmic transparency and practicality. Finance professionals may be wary of purely black-box models. Geertsema & Lu (2023) make progress on explainability by deriving peer weights but not all ML valuation studies offer such insights. Research has yet to fully address how to make clustering-based valuations explainable and auditable. For instance, providing intuitive reasons why certain peers were selected or why a model's estimate differs from market consensus. Bridging this interpretability gap is important for real-world adoption.

In summary, the literature would benefit from studies that integrate multiple ML techniques (clustering + NLP + advanced modelling) into a cohesive valuation approach and test it against real market outcomes. The promise of ML in CCA is clear from individual pieces of evidence but demonstrating a full-stack solution: an explainable, algorithmic peer selection method that consistently beats traditional practice in accuracy and bias, remains an open challenge. Addressing these gaps is the motivation for this thesis.

2.3.2 Chapter Conclusion

Machine learning approaches are revolutionizing comparable company analysis by addressing many of its historical pain points. Clustering is replacing subjective peer group selection, yielding more homogeneous and unbiased peer sets than those chosen by human analysts. Unsupervised learning algorithms like SOMs and spectral clustering have demonstrated the ability to improve multiple valuation accuracy for private firms, often outperforming traditional regression-based methods. ML models are not only enhancing the valuation multiples side of CCA but also improving key inputs: studies show significantly better beta and cost of capital estimates when using ML techniques as compared to conventional comparables or CAPM regressions. Furthermore, natural language processing is adding a new dimension by capturing qualitative similarity, which strengthens peer identification beyond simplistic industry codes.

Together, these innovations contribute to a more objective and rigorous CCA process. Early evidence of validation is encouraging though more work is needed to firmly establish outperformance in realized transactions. As it stands, the convergence of clustering, NLP and predictive modelling is pushing the frontiers of how we estimate value by comparables. The biases inherent in traditional CCA can be mitigated by algorithmic selection that transparently applies consistent criteria. The limitations of rigid industry classifications are overcome by richer similarity metrics from text and data. And the perennial question of "which multiples and comparables are best?" is increasingly answerable through ML techniques.

In conclusion, machine learning is set to augment and partly automate the CCA toolkit, making valuations more data-driven, reproducible and accurate. The research reviewed in this section lays a solid foundation but also highlights the need for integrated approaches and thorough validation. The subsequent chapters will build on these insights, aiming to design and test a framework that combines unsupervised clustering and NLP for peer selection and evaluates its performance against actual market outcomes. Future research should also consider the behaviour of these models for different market environments, as during periods of volatility their robustness might be questioned. Cross-market applications might explore whether the results generalize to contexts other than European and North American ones, thus offering a global perspective on valuation practices in M&A. Finally, comparative studies of ML methods against expert-driven benchmarks will be crucial to demonstrate not only statistical gains but also practical usability in professional finance.

3. Methodology

3.1 Research Design and Objectives

3.1.1 Restating the research gap

The choice of comparable firms is a critical step in relative valuation, yet academic and practitioner evidence shows that it is also one of the most problematic. Traditional approaches depend largely on industry codes (such as SIC, NAICS, or GICS) and on analyst discretion, often resulting in peer groups that are heterogeneous in size, growth, or profitability (Alford, 1992; Bhojraj et al., 2003). Empirical studies document both accuracy limitations, typical errors of 20–30 percent in multiples-based valuation and systematic biases. For instance, underwriters in IPOs select high-multiple peers to justify optimistic offer prices (Paleari et al., 2014), while advisors in M&A fairness opinions adjust peer selection strategically to align with litigation incentives (Eaton et al., 2022).

Machine learning (ML) has emerged as a promising alternative by forming peer groups algorithmically, thereby reducing subjectivity and improving reproducibility. Research demonstrates that unsupervised clustering can generate peer groups that are more homogeneous than industry codes and often lead to lower valuation errors (Jagrič et al., 2024). Similarly, textual similarity measures based on NLP capture economic comparability that industry classifications miss (Hoberg & Phillips, 2010; Gerling, 2023). Yet, despite these advances, existing work typically examines one dimension in isolation, either clustering, or textual similarity, or predictive modelling rather than developing an integrated pipeline for peer selection and valuation. Moreover, while many studies report improvements in error metrics within their sample, systematic validation against realized market outcomes (such as actual IPO prices or M&A deal values) remains scarce.

Another underexplored dimension is the role of clustering algorithms beyond the standard k-means and hierarchical approaches. Density-based methods such as DBSCAN and HDBSCAN/OPTICS have proven effective in other domains for identifying non-spherical clusters and excluding noise points but they have not yet been applied in comparable company analysis. This omission is notable, as outlier exclusion is particularly relevant in valuation: forcing a unique conglomerate or distressed firm into a peer group can distort multiples and propagate error. Addressing these methodological and empirical gaps provides the motivation for this thesis.

3.1.2 Overall methodological framework

The empirical design of this thesis follows a sequential pipeline that mirrors the steps of comparable company analysis while replacing subjective judgment with algorithmic procedures. The process begins with the construction of a dataset of more than 1260 European and U.S. listed firms from Refinitiv. Each firm is represented by standardized features capturing revenues, enterprise value, profitability ratios, growth rates, leverage indicators and categorical variables for industry and geography. Valuation multiples are excluded at this stage to avoid circularity, as they are later introduced only as outputs extracted from peer groups.

On this dataset, three unsupervised clustering algorithms are applied. K-means partitions the universe into compact groups by minimizing intra-cluster variance. Agglomerative hierarchical clustering builds nested groups, offering flexibility in granularity. DBSCAN and HDBSCAN/OPTICS identify dense regions in the feature space and classify atypical firms as noise, preventing the inclusion of outliers that distort valuation multiples. Using different clustering philosophies enables a comparative assessment of their suitability for peer selection.

Once clusters are formed, about 200 private M&A targets are assigned out-of-sample. For centroid-based methods, assignment relies on Euclidean distance to cluster centroids, while in the density-based setting targets are linked if they fall within the reachability radius of a cluster. This ensures that each private firm inherits the peer set most aligned with its financial and structural profile.

From these clusters, valuation multiples are extracted using the median of peer observations. The median mitigates skewness and the influence of extreme values, aligning with best practice (Liu et al., 2002; Schreiner & Spremann, 2007). Applying these multiples to the financials of the target firms generates algorithmic estimates of enterprise value that replicate practitioner logic while avoiding subjective peer selection.

The predicted valuations are then benchmarked against realized M&A deal prices. Accuracy is assessed through mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE).

To determine whether differences are meaningful, the framework incorporates statistical testing. Paired t-tests and Wilcoxon signed-rank tests compare clustering-based valuations with industry-median benchmarks, while Friedman tests evaluate relative performance across algorithms. Robustness checks assess sensitivity to the number of clusters, cluster stability and performance consistency across industries and regions. An extension evaluates whether clusters also provide coherent estimates of unlevered betas and cost of capital, broadening their application beyond multiples.

This methodological framework preserves the intuitive structure of comparable company analysis but systematizes each stage. By integrating unsupervised clustering, validation against market outcomes and robustness analysis, it addresses the weaknesses of traditional peer selection and provides a transparent, replicable basis for private firm valuation.

3.1.3 Hypotheses

The methodological framework translates into a set of testable hypotheses that reflect both the central research question and its extensions. They capture the expectation that algorithmic peer selection enhances valuation accuracy, that the choice of clustering algorithm influences results and that cluster-based peers may also serve as reliable inputs for risk estimation. In particular, the study is designed to assess both predictive accuracy and statistical robustness across alternative methods, while also exploring the broader applicability of clustering in estimating risk parameters:

- **H_1:** Clustering-based peer groups improve valuation accuracy compared to traditional industry-based benchmarks.
- **H_2:** Different clustering algorithms yield heterogeneous performance; density-based clustering (DBSCAN/OPTICS) is expected to outperform by filtering outliers.
- **H_3:** Cluster-derived multiples lead to lower MAE, RMSE and MAPE.
- **H_4:** Performance differences are statistically significant across paired and non-parametric tests.
- **H_5:** Peer groups also provide valid inputs for estimating beta and WACC, extending their application beyond multiples.

Together, these hypotheses provide the foundation for the empirical analysis that follows and guide the structure of Chapter 3. They ensure that the study contributes simultaneously to valuation practice and to the academic literature on algorithmic peer selection. Moreover, they establish a clear link between methodological innovation and measurable improvements in empirical accuracy. By articulating both valuation and risk-focused hypotheses, the framework connects practical relevance with theoretical advancement.

3.2 Data and Variables

3.2.1 Sources & Samples

The empirical analysis is built on two datasets sourced from Refinitiv's Market Screener, one covering public firms and the other private M&A transactions.

The first dataset consists of 1,260 publicly listed healthcare firms headquartered in Europe and North America, observed over five fiscal years (2021–2025). The focus on healthcare was a deliberate choice. On the one hand, the industry is fragmented in homogeneous sub-industries; on the other, healthcare represents a data-rich industry with consistent disclosure and a high level of M&A activity. Restricting the scope to a single sector also enhances comparability by reducing cross-industry heterogeneity in valuation multiples and ensures a balanced distribution of deals over time. Within this sector, companies were included only if revenues were strictly positive, excluding distressed or pre-revenue firms. While this narrows coverage, it avoids substantial missing data problems and minimizes the need for imputation, thereby preserving data quality. Importantly, even micro-cap companies are retained, which increases heterogeneity within the sample without compromising reliability.

The second dataset contains 199 completed M&A transactions in the healthcare sector between 1 January 2021 and the present. Transactions were filtered by geography (Europe and North America), completion status (completed only) and disclosure requirements: enterprise value and revenues at the effective date had to be strictly positive and available. This ensures that all deals can be consistently benchmarked against multiples. Private firms are included alongside public targets, provided they meet disclosure thresholds. The reliance on recent transactions reflects both practical and methodological considerations: Refinitiv limits exports to 200 deals per query and focusing on the 2021–2025 period ensures alignment with the public-company dataset while capturing the post-COVID market environment.

*Dataset Construction (**Table 6**)*

Dataset	Source	Sample size	Period	Filters	Final obs.
Public firms	Refinitiv Market Screener	1,260 firms	2021–2025 (5Y)	HQ in EU/NA; Healthcare (GICS); Revenues > 0	6,300 firm-years
M&A transactions	Refinitiv Market Screener – Deals	199 deals	2021–2025	EU/NA;Healthcare (TRBC);Completed; EV > 0; Revenues > 0; Public & private	199 deals

Source: Author's elaboration based on data from Refinitiv Market Screener (public firms) and Refinitiv Market Screener – Deals (M&A transactions).

The table highlights how the construction of the datasets balances scope and quality. The public firm panel ensures sufficient scale for clustering, while the M&A sample provides a validation benchmark grounded in realized market outcomes. Both datasets are aligned in sector, geography and reporting standards, ensuring consistency across the empirical framework.

Both datasets were standardized into a common reporting currency (EUR millions). Although much of the academic literature defaults to USD, EUR was adopted for uniformity and ease of reporting, given the European focus of the thesis. Industry comparability across the two samples was ensured through consistent use of the GICS classification. Temporal alignment was enforced by using multi-year data for public companies but only the year of the transaction for private targets, consistent with how valuations are conducted at the deal date.

The use of Refinitiv's Market Screener platform, rather than combining multiple underlying databases such as Worldscope and SDC, was primarily motivated by accessibility and efficiency. It provides a unified environment to extract both public company fundamentals and M&A deal data, which simplifies the process and ensures consistency across datasets. While more specialized databases might offer additional depth, this choice reflects the practical constraints of data access and computational resources while still providing reliable coverage for the research objectives.

Together, the two datasets form a coherent empirical setting: the public firm universe provides the fundamentals required for algorithmic clustering, while the M&A transactions provide an externally validated benchmark against which to test the accuracy of cluster-based valuations.

3.2.2 Data Preprocessing

i) Raw Database features:

The preprocessing stage begins with the extraction of raw variables from Refinitiv's Market Screener, separately for the public companies' dataset and the M&A transaction dataset. These native fields constitute the foundation upon which derived ratios and valuation multiples are later constructed.

i-a) Public companies:

The dataset of listed firms includes both identifiers and financial fundamentals. Identifiers comprise the Refinitiv RIC code, company name, country of headquarters, GICS industry classification and business description. Financial data are provided on a five-year panel basis and cover revenues, EBITDA, EBIT, net income, total assets, cash, total debt, net debt, market capitalization and enterprise value. In addition, Refinitiv supplies pre-computed indicators such as EBITDA margin, net profit margin, ROE, ROA, beta, debt-to-equity percentage and historical net debt/EBITDA. Historical valuation multiples (EV/Revenue, EV/EBITDA, EV/EBIT, P/E) are also reported, although these are excluded from the clustering stage to prevent circularity and are retained only for validation purposes.

*Public Dataset – native variables (**Table 7a**)*

Category	Variables
Identifiers	Identifier (RIC), Company name, Country, GICS industry, Business description
Fundamentals	Revenues, EBITDA, EBIT, Net income, Total assets, Cash, Total debt, Net debt, Market capitalization, Enterprise value
Pre-computed indicators	EBITDA margin, Net profit margin, ROE, ROA, Beta, Debt/Equity %, Net debt/EBITDA
Historical multiples	EV/Revenue, EV/EBITDA, EV/EBIT, P/E

Source: Author's elaboration based on Refinitiv Market Screener – Public Companies Database.

i-b) M&A transactions:

The deals dataset combines transaction identifiers, qualitative descriptors and firm-level financials. Transaction-level information includes a unique deal number, completion year, deal type and percentage of shares acquired. Firm descriptors capture the target's name, nation, industry and public status, as well as the acquiror's name, nation, macro- and mid-level industry classification and a textual business description of the target. Financial disclosure at the deal date includes revenues, EBITDA, EBIT, EBITDA margin, net income, EPS, book value per share, total assets, total debt, net debt, equity value at announcement, enterprise value at announcement, deal value and rank value including net debt. All figures are expressed in millions of euros.

*M&A Dataset – native variables (**Table 7b**)*

Category	Variables
Transaction info	Deal number, Year completed, Deal type, % shares acquired
Target descriptors	Target name, Nation, Industry, Public status, Business description
Acquiror descriptors	Acquiror name, Nation, Macro industry, Mid industry
Financials (deal date)	Revenues, EBITDA, EBIT, EBITDA margin, Net income, EPS, Book value per share, Total assets, Total debt, Net debt
Value measures	Equity value at announcement, Enterprise value at announcement, Deal value, Rank value incl. net debt

Source: Author's elaboration based on Refinitiv Market Screener – Deals Database.

ii) Feature engineering:

ii-a) Public DB:

For listed firms, derived variables are constructed by applying standard financial definitions consistently across all fiscal years.

Profitability margins are defined as:

$$\boldsymbol{EBITDA\ margin = \frac{EBITDA}{Revenues}}, \qquad \boldsymbol{EBIT\ margin = \frac{EBIT}{Revenues}},$$
$$\boldsymbol{Net\ income\ margin = \frac{Net\ income}{Revenues}}$$

Performance indicators are derived as :

$$\boldsymbol{ROA = \frac{Net\ income}{Total\ assets}}, \qquad \boldsymbol{ROE = \frac{Net\ income}{Equity}}$$

Capital structure is further characterized through:

$$\boldsymbol{Net\ debt/EBITDA = \frac{Net\ debt}{EBITDA}}, \qquad \boldsymbol{Debt/Equity}\ \% = \boldsymbol{\frac{Total\ debt}{Equity}},$$
$$\boldsymbol{Assets/Equity = \frac{Total\ assets}{Equity}}$$

Valuation multiples are also constructed from the fundamental variables, namely:

$$\boldsymbol{EV/Rvenues = \frac{Enterprise\ value}{Rvenues}}, \qquad \boldsymbol{EV/EBITDA = \frac{Enterpire\ value}{EBITDA}},$$
$$\boldsymbol{P/E = \frac{Market\ Cap}{Net\ t\ Income}}$$

As a result, the engineered variables are summarized in Public Dataset – Derived Variables (Table 8a).

*Public Dataset – Derived variables (**Table 8a**)*

Profitability	Leverage	Efficiency / Liquidity	Valuation
EBITDA margin, EBIT margin, Net profit margin, Net income	Net debt/EBITDA, Debt/Equity %, Equity	Asset turnover, Cash/Assets	EV/Revenue, EV/EBITDA, EV/EBIT, P/E

Source: Author's elaboration based on Refinitiv Market Screener – Public Database.

ii-b) Deals (transaction-date accounting):

To reconcile valuation fields and ensure internal consistency, equity and enterprise values are constructed through an ordered hierarchy. Equity value is taken from the announced figure when available; otherwise, it is backed out from enterprise value and net debt (adding absolute net cash when net debt is negative), or, when necessary, inferred from deal value and the acquired stake; as a last resort, rank value including net debt is used as a proxy. Enterprise value follows the reverse hierarchy: the announced figure when available; else equity value plus net debt (subtracting net cash if net debt is negative); else the rank value including net debt. The percentage of shares acquired is retained if disclosed; when missing but deal value and equity value are available and consistent, it is inferred as:

$$\%Shares\ Acquired\ = \frac{100\ x\ Deal\ value}{equity\ value}$$

From native financials (revenues, EBITDA, EBIT, net income, assets, debt, net debt), standard ratios are then computed to normalize across firm size and capital structure: profitability (EBITDA margin, EBIT margin, net income margin), leverage (net debt/EBITDA, net debt/assets, net debt/equity; debt/EBITDA, debt/equity; assets/equity) and efficiency/liquidity (asset turnover, cash/assets). Valuation multiples (EV/EBITDA, EV/EBIT, EV/Revenues, P/E) are derived from the reconciled enterprise/equity values and used solely at the validation stage, not as clustering inputs, to avoid circularity (Liu et al., 2002; Schreiner & Spremann, 2007; Kim & Ritter, 1999; Gilson et al., 2000).

*M&A Dataset – Derived variables (**Table 8b**)*

Profitability	Leverage	Efficiency / Liquidity	Valuation
EBITDA margin, EBIT margin, Net income margin	Net debt/EBITDA, Net debt/Assets, Net debt/Equity, Debt/EBITDA, Debt/Equity, Assets/Equity	Asset turnover, Cash/Assets	EV/EBITDA, EV/EBIT, EV/Revenues, P/E; Equity value (reconciled), Enterprise value, % shares acquired

Source: Author's elaboration based on Refinitiv Market Screener – Deals Database.

iii) Data Cleaning:

To begin with, the first step in preparing the dataset was a thorough data cleaning procedure. Given the presence of reporting errors, observations showing impossible values (such as negative revenues, non-sensical ratios, or out-of-range accounting figures) were identified and removed. This ensured that subsequent analyses relied only on consistent and reliable information.

iii-a) Public DB:

The public company dataset initially presented a high proportion of missing values, amounting to 22.3% of all numerical entries (30,905 missing cells out of 138,600). Figure 11 illustrates the distribution of NaN across variables and years: disclosure gaps were particularly severe for cash (≈70%), net debt/EBITDA (≈73%), valuation multiples (EV/EBITDA, EV/EBIT above 60%) and performance indicators such as ROE (≈52%) and net income (≈24% in 2023).

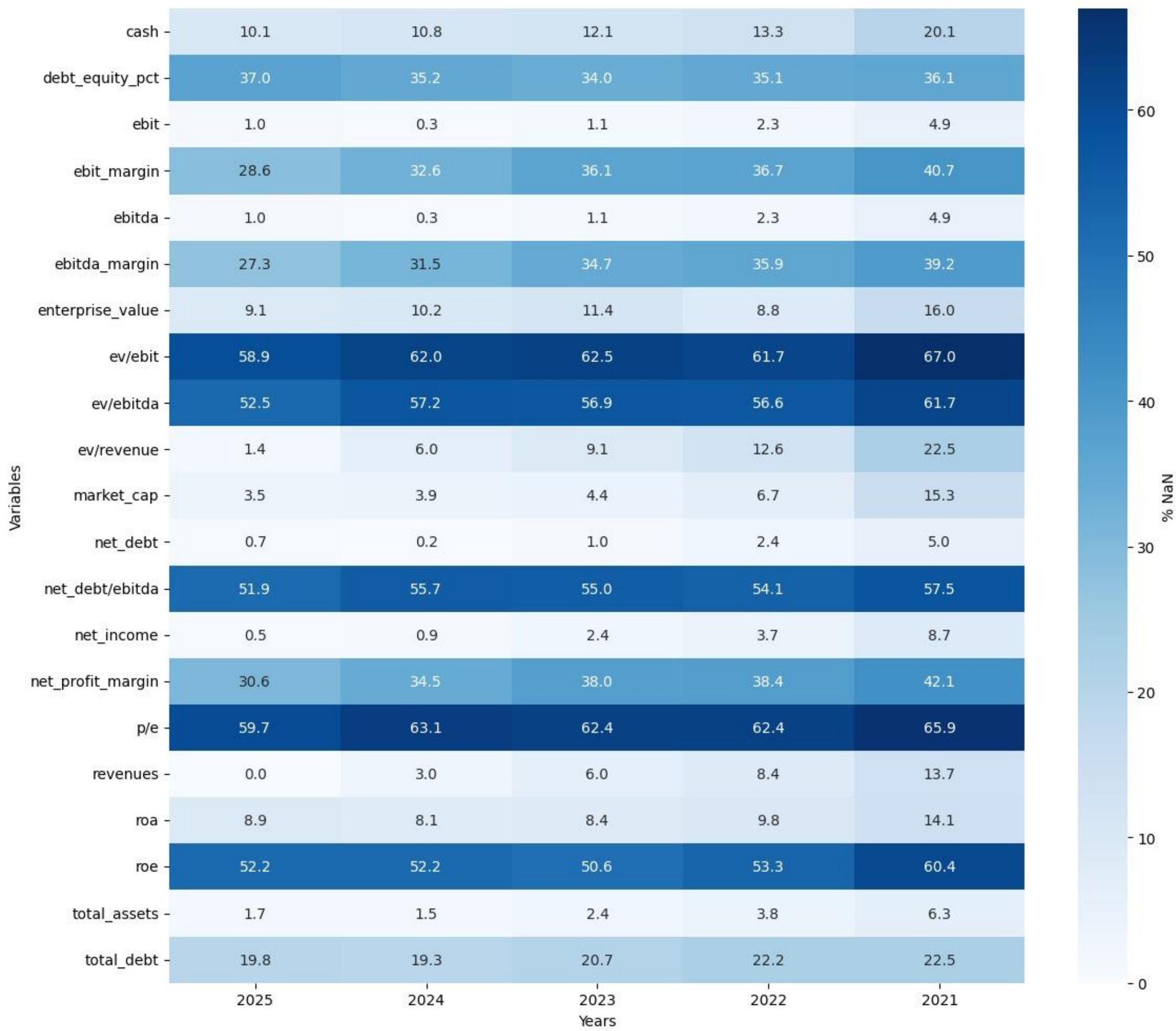
Heatmap of missing values – Public DB before cleaning ***(Figure1)***

Instead of dropping observations, missing values were addressed through a systematic relational framework. Standard accounting and valuation identities were applied to reconstruct variables wherever inputs were available.

Margins, multiples and leverage ratios were recomputed following the same logic, ensuring internal consistency across the five fiscal years. The implementation filled only NaN values, leaving all disclosed data untouched.

After this relational filling, disclosure improved substantially for several fundamentals: missingness in cash fell from about 67–69% to roughly 10–13%, revenues from ~10% to <3% and margins from 30–40% to ~27–39%. Net profit margin, ROA and EBIT margin also showed clear reductions. However, some valuation multiples (EV/EBIT, EV/EBITDA, P/E) and leverage measures (Net debt/EBITDA, ROE) still display more than 50% missing values across years, limiting their reliability for peer group calibration. Overall, the dataset retained all 1,260 firms and achieved a marked reduction in missingness for core operating and accounting variables, even though several market-based ratios remain sparsely disclosed.

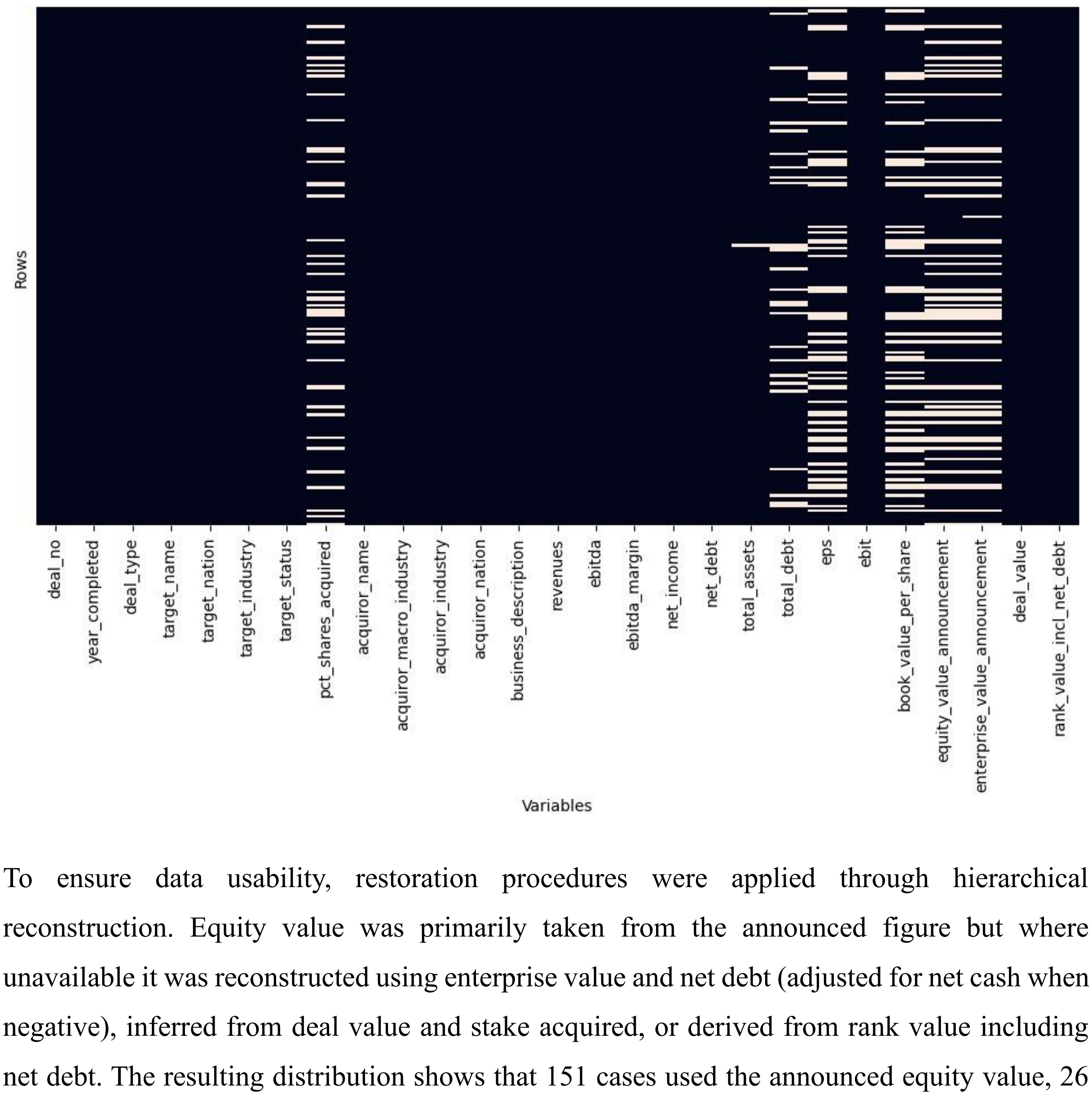
Heatmap of missing values – Public DB after relational filling ***(Figure2)***



iii-b) Deal DB:

The deal dataset required an extensive cleaning procedure to address disclosure gaps and inconsistent reporting. Several key fields exhibited substantial non-disclosure: book value per share (27.6% missing), EPS (27.1%), enterprise value at announcement (24.6%), equity value at announcement (24.1%), percentage of shares acquired (20.1%), total assets (0.5%) and total debt (13.1%). In contrast, core accounting items such as revenues, EBITDA and total assets were consistently available.

Heatmap of missing values – raw deal database ***(Figure3)***

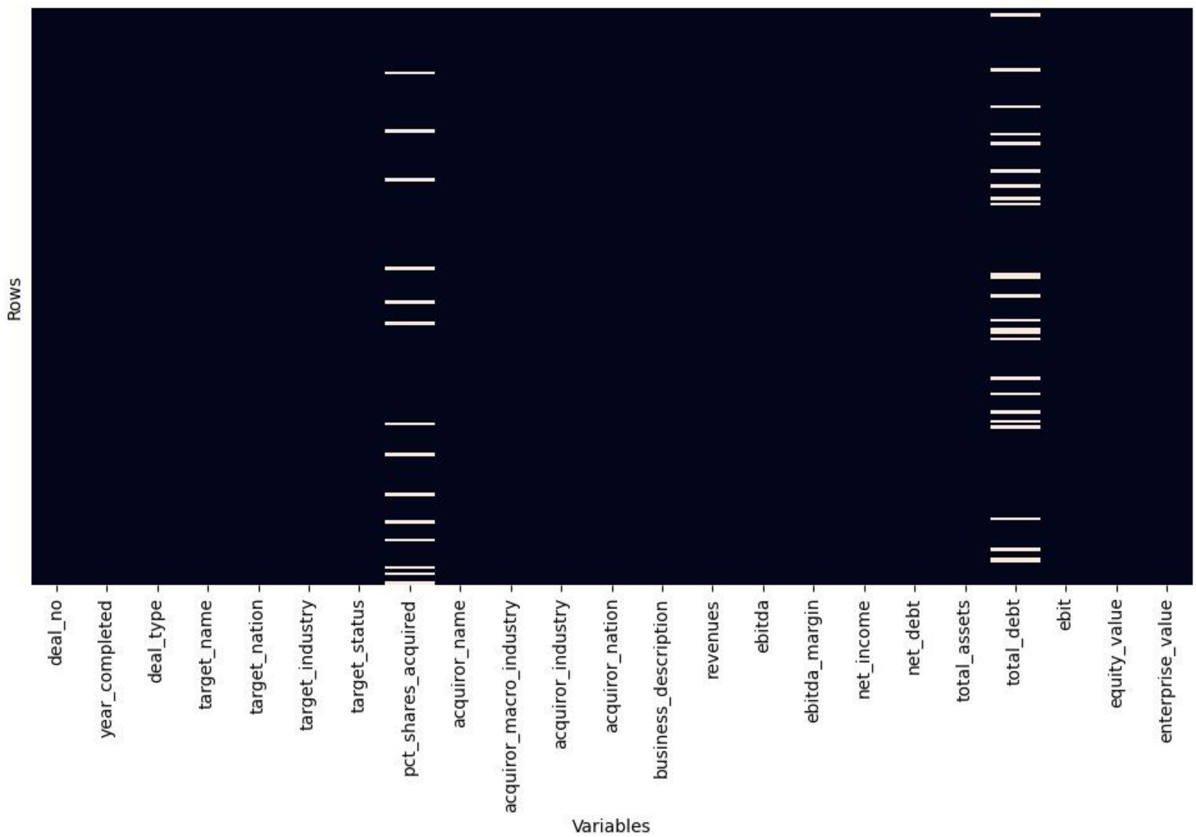


To ensure data usability, restoration procedures were applied through hierarchical reconstruction. Equity value was primarily taken from the announced figure but where unavailable it was reconstructed using enterprise value and net debt (adjusted for net cash when negative), inferred from deal value and stake acquired, or derived from rank value including net debt. The resulting distribution shows that 151 cases used the announced equity value, 26 were inferred from deal value and shares acquired and 22 were reconstructed from rank value less net debt.

Similarly, enterprise value was recovered following a reverse hierarchy: 150 cases directly from the announced value, 27 from equity plus net debt and 22 from the rank value. The percentage of shares acquired was available in 84 deals, reconstructed in 92 cases from the ratio of deal value to equity value and unavailable in 23. Cash was reconstructed indirectly in 173 cases as total debt minus net debt, with only 26 observations missing after this step.

After variable restoration, further filters were applied. Observations were dropped if both revenues and EBITDA margin were missing, or if total assets were unavailable. Deals with negative reconstructed equity or enterprise values were also excluded. This process reduced the dataset from 199 to 189 observations.

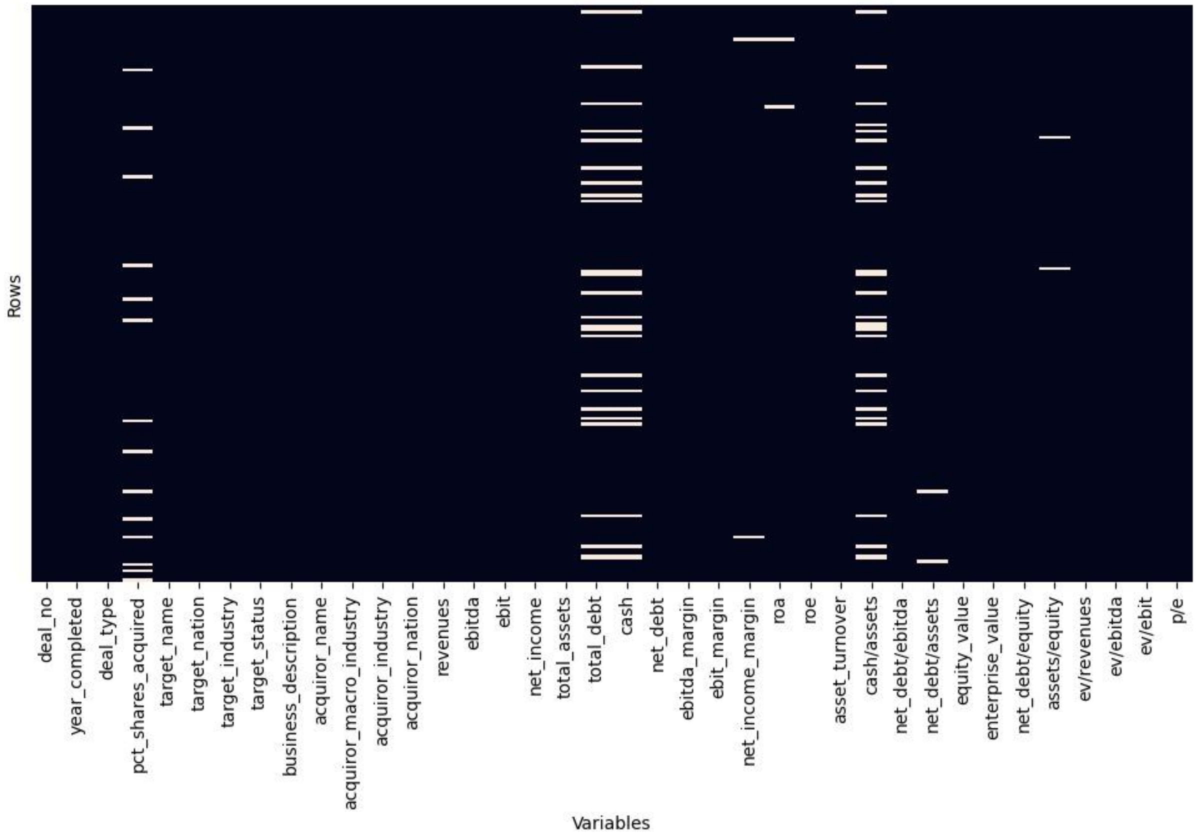
Heatmap of missing values – after restoration and pruning ***(Figure4)***



Subsequently, redundant fields (EPS, book value per share) and high-NaN variables that lost relevance after reconciliation were removed. At this stage, the dataset contained 23 variables, with only 39 missing values across 4,347 cells (0.9%). The refined dataset preserved key financials, ratios and reconstructed valuation measures, ensuring internal consistency while maintaining representativeness.

Following feature engineering (Section 3.2.2 (ii)), derived ratios were added, further reducing the relative weight of missing data. The final dataset reached 189 observations and 27 variables, with the share of missing values concentrated in cash and related ratios (13.2%) and in the reconstructed stake acquired (7.4%).

Heatmap of missing values – final dataset after feature engineering ***(Figure5)***



Finally, year-level diagnostics confirmed a balanced time distribution (21 deals in 2021, 52 in 2022, 50 in 2023, 41 in 2024 and 25 in 2025) and the percentage of missing numerical values declined over time (from 14 in 2021 to only 7 in 2025).

This sequence of cleaning steps ensured that the deal dataset achieved a high level of integrity, with 1.9% overall missingness, robust reconstructed valuation fields and a structure fully comparable with the public dataset for subsequent clustering and validation. Additional sensitivity checks confirmed that imputations did not bias central tendencies, while outlier removal further improved the stability of statistical outputs. As a result, the dataset provides a reliable empirical foundation for the clustering exercises that follow, minimizing the risk of distortions from data quality issues.

iv) Outliers Handling

An important preprocessing step concerns the treatment of outliers. Financial datasets are typically characterized by extreme values, particularly in leverage and valuation ratios, which may reflect genuine heterogeneity (e.g., distressed firms, micro-cap companies) rather than data errors. Arbitrary trimming or winsorization was deliberately avoided in order to preserve the representativeness of the sample and to maintain small and micro-cap observations, which are relevant in M&A markets.

To mitigate distortions in distance-based clustering algorithms (such as k-means and hierarchical clustering), all numerical variables were standardized prior to analysis. Robust scaling was applied, transforming each variable according to its median and interquartile range:

$$x' = \frac{x - median(x)}{IQR}$$

This approach reduces the disproportionate influence of extreme values without discarding observations, ensuring that ratios and financial magnitudes contribute comparably to distance computations.

For density-based clustering methods (DBSCAN and HDBSCAN), outliers are handled natively by the algorithm: observations that do not belong to any high-density region are classified as noise points and excluded from peer groups. This feature makes density-based methods particularly well suited for financial applications where occasional extreme firms can otherwise distort group formation.

3.2.3 Variables

Following the extraction, feature engineering and cleaning procedures, the dataset was consolidated into a set of variables organized into two groups: clustering inputs and validation outputs. This distinction reflects the dual objective of the empirical design: (i) forming economically homogeneous peer groups based on fundamentals and (ii) benchmarking valuations against realized transaction prices using market multiples.

Clustering inputs. Peer groups are constructed on the basis of variables that capture firms' size, profitability, capital structure and efficiency. Size is proxied by revenues, total assets and, for public firms, market capitalization. Profitability is measured through operating and net margins (EBITDA margin, EBIT margin, net income margin), complemented by return ratios (ROA, ROE). Capital structure is reflected in both flow-based and stock-based indicators: net debt/EBITDA, net debt/assets, net debt/equity, debt/equity % and assets/equity. Finally, efficiency and liquidity are captured through asset turnover and cash/assets. These variables are standard in valuation practice and ensure comparability across firms of different scale (Alford, 1992; Liu et al., 2002). Importantly, all inputs are backward-looking and based on realized accounting figures, ensuring consistency with ex ante valuation settings.

Validation outputs. Valuation performance is assessed by applying market multiples derived from reconciled equity and enterprise values. The key outputs are EV/EBITDA, EV/EBIT, EV/Revenues, and P/E, alongside enterprise value, equity value and, in the deal's dataset, the percentage of shares acquired. These variables are used exclusively in the validation stage to benchmark predicted valuations against actual deal prices. By construction, they are excluded from the clustering process to avoid circularity in peer group formation (Schreiner & Spremann, 2007).

3.3 Clustering Methods and Model Selection

3.3.1 Algorithms

The clustering stage applies three classes of unsupervised learning algorithms: centroid-based (k-means), hierarchical and density-based (DBSCAN/HDBSCAN). These methods differ in their assumptions about cluster geometry, sensitivity to noise and parameterization, enabling a comprehensive assessment of their suitability for peer group formation.

i) K-means:

K-means partitions the dataset into k groups by minimizing the within-cluster sum of squared distances (WCSS). Formally, given a set of firms X = {x_1, …, x_n} ⊂ R^p, the objective is:

$$arg\ min \sum_{j=1}^{k} \sum_{x_i \in C_j} \left|\left|x_i - \mu_j\right|\right|^2$$

where C_j is cluster j and μ_j its centroid. The algorithm proceeds iteratively: (i) initialize k centroids, (ii) assign each point to the nearest centroid, (iii) recompute centroids as cluster means and (iv) repeat until convergence. While computationally efficient (Jagrič et al., 2024), k-means assumes spherical clusters of similar size and requires pre-specifying k, which is often non-trivial in financial data.

ii) Hierarchical clustering:

Hierarchical methods build a nested structure of clusters based on pairwise distances. In the agglomerative approach, each observation starts as a singleton cluster and at each step the two closest clusters are merged until all firms are in a single group. The distance between clusters A and B under average linkage is defined as:

$$d(A,B) = \frac{1}{|A|\,|B|}\sum_{x_i \in \mathrm{A}}\sum_{x_j \in \mathrm{B}} d(x_i, x_j)$$

where d(·,·) denotes Euclidean distance. The result is a dendrogram, from which the optimal number of clusters can be chosen ex post. Hierarchical clustering offers interpretability and does not require an initial guess for k but it is computationally more expensive ($O(n^2 \log n)$).

iii) DBSCAN:

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) identifies clusters as dense regions of the feature space. Given parameters $\varepsilon > 0$ (radius) and $MinPts \in N$ (minimum points), the ε-neighborhood of x_i is:

$$N_\varepsilon(x_i) = \{x_j \in X \colon d(x_i, x_j) \leq \varepsilon\}$$

- x_i is a core point if $|N_\varepsilon(x_i)| \geq MinPts$.
- x_i is a border point if it is within $N_\varepsilon(x_j)$ of a core point but not itself a core point.
- Otherwise, x_i is noise.

Two points are density-connected if there exists a chain of density-reachable points linking them. A cluster is a maximal set of density-connected points. DBSCAN thus detects clusters of arbitrary shape without requiring k and naturally excludes sparse observations as noise.

iv) OPTICS/HDBSCAN:

HDBSCAN extends DBSCAN by removing the need to fix ε. It defines the mutual reachability distance:

$$d_{mreach}(x_i, x_j) = max\{core_k(x_i), core_k(x_j), d(x_i, x_j)\}$$

where $core_k(x)$ is the distance from x to its k-th nearest neighbor. A minimum spanning tree is then built over all observations with this distance metric, producing a hierarchy of clusters across multiple density levels. The algorithm extracts the most stable clusters according to persistence scores, enabling identification of groups with heterogeneous size and density (Staňková, 2024).

vi) Comparison:

K-means provides a baseline centroid partition that is efficient but assumes homogeneity in cluster shape and size. Hierarchical clustering offers interpretability and flexibility in ex post selection of k, at higher computational cost and noise sensitivity. DBSCAN and OPTICS provide a density-based perspective, automatically detecting non-spherical clusters and excluding noise, which is particularly advantageous in financial datasets with distressed or conglomerate firms.

vii) NLP:

Natural Language Processing (NLP) techniques allow the integration of qualitative information, such as business descriptions, into the clustering framework. In this study, textual data are vectorized through Term Frequency–Inverse Document Frequency (TF-IDF), which assigns higher weight to terms that are frequent within a document but rare across the corpus. To mitigate the high dimensionality and sparsity of TF-IDF matrices, a Truncated Singular Value Decomposition (SVD) is applied, extracting 50 latent semantic dimensions that capture the dominant topics across firms.

The resulting text embeddings are concatenated with standardized numerical features (revenues, ROA, net debt/EBITDA, EBITDA margin), producing a joint representation of firms that incorporates both financial and qualitative characteristics. Clustering is then performed with DBSCAN and OPTICS on the combined feature space.

This procedure allows the algorithm to consider similarity in both quantitative fundamentals and business model descriptions. However, preliminary results show that the textual embeddings tend to dominate the feature space, often collapsing firms into a small number of large clusters. This highlights the trade-off between exploiting rich unstructured data and preserving granularity in peer identification.

3.3.2 Validity indices

A critical step in clustering is the evaluation of partition quality and the selection of parameters such as the number of clusters (k) in k-means or the neighborhood radius (ε) and minimum points (MinPts) in DBSCAN. Since clustering is unsupervised, validity must be assessed through internal indices that measure compactness and separation, complemented by stability criteria.

i) Silhouette score:

For each observation i, let a(i) denote the mean intra-cluster distance and b(i) the lowest mean distance to points in another cluster. The silhouette coefficient is defined as:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

with s(i) ∈ [-1,1]. Values close to 1 indicate well-separated clusters, while negative values suggest misclassification. The overall silhouette score is the average across all observations.

ii) Calinski–Harabasz index:

Also known as the variance ratio criterion, this index measures the ratio of between-cluster to within-cluster dispersion:

$$CH = \frac{Tr(B_k)}{Tr(W_k)} \cdot \frac{n - k}{k - 1}$$

where B_k is the between-cluster scatter matrix and W_k the within-cluster scatter matrix. Higher values indicate better-defined clusters.

iii) Davies–Bouldin index:

This index evaluates average similarity between clusters, defined as the ratio of within-cluster scatter to between-cluster separation:

$$DB = \frac{1}{k}\sum_{i=1}^{k} max\frac{S_i + S_j}{M_{ij}}$$

where S_i is the average distance of points in cluster i to its centroid and M_{ij} the distance between centroids of clusters i and j. Lower values indicate better clustering.

iv) DBSCAN-specific criteria:

For density-based methods, cluster validity cannot be captured fully by centroid-based indices. The selection of ε is guided by the k-distance plot, where a sharp change in slope ("knee") indicates the appropriate radius. HDBSCAN further evaluates clusters through stability scores, measuring the persistence of a cluster across density levels. A cluster with stability S can be expressed as:

$$S(C) = \int_{\lambda\mathrm{birth}}^{\lambda\mathrm{death}} |C(\lambda)|\, d\lambda$$

where λ denotes inverse density and λ_{birth}, λ_{death} represent the range of densities for which cluster C exists. More stable clusters are considered more meaningful.

v) Multi-index comparison:

Each index captures a different aspect of clustering quality: silhouette emphasizes separability, Calinski–Harabasz rewards high inter-cluster variance, Davies–Bouldin penalizes overlap and HDBSCAN stability highlights persistence. Since no single criterion is universally superior, model selection in this thesis is based on the joint evaluation of multiple indices, ensuring robustness and reducing the arbitrariness of parameter choice (Hanauer et al., 2022; Jagrič et al., 2024).

3.3.3 Model Selection

Based on the comparative properties of clustering algorithms and the evaluation criteria discussed above, the empirical analysis focuses primarily on density-based methods. These approaches are particularly suited to financial datasets characterized by heterogeneity, irregular cluster shapes and the presence of outliers. Unlike k-means and hierarchical clustering, which impose restrictive assumptions on cluster geometry and require the ex-ante specification of k, DBSCAN/OPTICS identify clusters of varying density and automatically classify idiosyncratic firms as noise, preventing distortions in peer multiples. Model selection is guided by a combination of validity indices to ensure robustness and replicability.

3.4 Target Assignment and Peer Extraction

Once the clustering of public firms has been performed, the next step is to associate private M&A targets with these groups to construct peer sets. This procedure ensures that valuation is based on the most economically comparable firms identified by the unsupervised learning framework. Two tasks are required: (i) the assignment of each target to an existing cluster and (ii) the extraction of the corresponding peer set of public firms.

3.4.1 Assignment of targets

Targets in the deal database are observed at the transaction date and therefore provide a single accounting snapshot rather than a five-year panel. To assign them consistently, the feature vector of each target is constructed using the same set of ratios and fundamentals employed in clustering the public firms (revenues, profitability, leverage and efficiency indicators).

Assignment proceeds through nearest-cluster matching. For centroid-based algorithms such as k-means, each target is allocated to the cluster whose centroid minimizes the Euclidean distance:

$$\hat{C}(x_{target}) = arg\ min\ d(x_{target}, \mu j)$$

where μ_j denotes the centroid of cluster C_j

For density-based methods (DBSCAN and HDBSCAN), targets are mapped using the distance to the nearest core point in the cluster. If a target lies within the ε-neighborhood of a core observation, it is assigned to that cluster. If not, it is treated as noise. To preserve sample size, a secondary rule is applied whereby noise targets are reassigned to the nearest cluster centroid, ensuring that each transaction contributes to validation.

Crucially, the assignment is conducted in an out-of-sample fashion: targets do not affect the formation of clusters but are mapped onto the pre-computed structure, thereby avoiding look-ahead bias.

3.4.2 Peer set construction

Once a target has been assigned, the peer set is defined as the collection of all public firms belonging to the same cluster. This ensures that multiples are extracted from a group of companies that share homogeneous financial characteristics with the target, rather than from arbitrary or industry-code-based peers.

For each target, the peer set provides the distribution of valuation multiples EV/EBITDA, EV/EBIT, EV/Revenues and P/E derived from the financials of the public firms in that cluster. These multiples form the basis for the subsequent valuation exercise in Section 3.5.

3.5 Valuation Procedure

After assigning private M&A targets to cluster-derived peer groups, the next step is to compute their estimated values. This requires extracting valuation multiples from the peer set and applying them to the target's financials. The procedure is designed to replicate standard comparable company analysis (CCA) while eliminating the subjectivity of manual peer selection.

3.5.1 Valuation computation

For each target t, a peer set P(t) is identified as described in Section 3.4. From this set, valuation multiples are computed as:

The multiples are aggregated across peers (using the median to mitigate the impact of outliers) to obtain cluster-level benchmarks:

$$\widetilde{M}_{cluster} = median_{\boldsymbol{i \in P(t)}}(M_i)$$

These median multiples are then applied to the target's financial fundamentals (revenues, EBITDA, EBIT, net income) to compute implied enterprise and equity values:

$$\widehat{EV}_t^{(M)} = \widetilde{M}_{cluster} \cdot \text{Foundamental}_t$$

where Fundamental_t corresponds to the denominator of the chosen multiple. For instance, EBITDA is used for EV/EBITDA.

When multiple valuation metrics are available, the central estimate is constructed as the median across the set of implied enterprise values. Equity values are then obtained by subtracting net debt from enterprise value:

$$\widehat{EquityValue}_t = \widehat{EV}_t - NetDebt_t$$

This ensures consistency across methods and comparability with transaction prices.

3.5.2 Baseline

To benchmark performance, predicted values from cluster-derived peer sets are compared to a baseline industry-median approach, which reflects standard practice in both academia and investment banking (Alford, 1992; Liu et al., 2002). Specifically, for each target, the corresponding GICS industry median multiple is extracted from the public firm dataset and applied to the same financial fundamentals:

$$\widehat{EV}_t^{(baseline)} = \widetilde{M}_{industry} \cdot \text{Foundamental}_t$$

This baseline is subject to the same computation rules (median multiples, enterprise-to-equity reconciliation) but peers are defined by industry codes rather than clustering. By contrasting cluster-based valuations with the baseline, the analysis isolates the incremental value of algorithmic peer selection.

3.6 Validation and Statistical Testing

The validation stage compares predicted values with actual transaction prices (enterprise or equity value) to assess whether clustering-based peer selection improves accuracy relative to the baseline industry-median approach.

3.6.1 Error metrics

Valuation errors are computed by comparing estimated values $\hat{V}_t$ with observed deal values V_t:

- Absolute Error (AE):

$$AE_t = \left|\hat{V}_t - V_t\right|$$

- Relative Error (RE):

$$RE_t = \frac{\hat{V}_t - V_t}{V_t}$$

- Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|\hat{V}_t - V_t\right|$$

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{V}_t - V_t)^2}$$

- Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{\hat{V}_t - V_t}{V_t}\right|$$

- Percentage within bounds: share of deals where the estimated value lies within ±10%, ±20%, ±30% of the observed transaction price (Handaya et al., 2009; Ong et al., 2021).

These indicators capture complementary aspects of accuracy: absolute deviation (MAE, RMSE), proportional deviation (MAPE) and practical robustness (% within bounds).

3.6.2 Statistical tests

To assess whether improvements are statistically significant, paired comparisons are conducted between clustering-based predictions and baseline estimates:

- Paired t-test: evaluates mean differences in errors under the assumption of normally distributed differences.

$$t = \frac{\bar{d}}{s_d/\sqrt{N}}$$

where $\bar{d}$ is the mean difference and s_d the standard deviation.

- Wilcoxon signed-rank test: a non-parametric alternative that ranks absolute differences and tests whether the median difference is zero, robust to non-normal error distributions.
- Friedman test: a non-parametric test for repeated measures, applied when comparing multiple clustering algorithms simultaneously (k-means, DBSCAN, HDBSCAN).

Combining parametric and non-parametric tests mitigates the risk of spurious inference in the presence of heavy-tailed or skewed financial data (Dessaint et al., 2017).

3.7 Extension: Beta & WACC

While the primary focus of this thesis is on valuation multiples, the clustering framework can be extended to risk estimation. In particular, peer groups derived from unsupervised learning provide a natural basis for estimating the systematic risk of private firms, which in turn is necessary to compute the cost of capital.

i) Unlevered beta:

For each cluster C_j, the unlevered beta is computed as the median across public firms in the group:

$$\beta_{U,j} = median_{i \in Cj} \left(\frac{\boldsymbol{\beta L, i}}{1 + (1 - \tau) \cdot \frac{D_i}{E_t}} \right)$$

where $\beta_{L,i}$ is the levered beta of firm i, τ the corporate tax rate and D_i/E_i the debt-to-equity ratio. This cluster-level unlevered beta represents the systematic risk of an average firm with similar fundamentals.

ii) Relevering for the target:

The private target's relevered beta is then obtained by applying its observed (or pro forma) leverage ratio:

$$\beta_{L,t} = \beta_{U,j} \cdot \left[1 + (1 - \tau) \cdot \frac{D_t}{E_t}\right]$$

where D_t/E_t is the target's capital structure.

iii) Cost of equity and WACC:

The cost of equity follows the Capital Asset Pricing Model (CAPM):

$$K_{e,t} = R_f + \beta_{L,t} \cdot ERP$$

where R_f is the risk-free rate and ERP the equity risk premium. The weighted average cost of capital (WACC) is then derived as:

$$WACC_t = \frac{E_t}{D_t + E_t} k_{e,t} + \frac{D_t}{D_t + E_t} K_d (1 - \tau)$$

where K_d denotes the cost of debt, typically proxied by the firm's yield to maturity or an industry spread over government bonds.

iv) Contribution:

It is possible to extend cluster-based peer groups to the estimation of betas and WACC, the methodology provides a unified pipeline that supports both valuation by multiples and risk-adjusted discounting. This addresses one of the key challenges in private firm valuation, namely the absence of market data for estimating systematic risk (Alanis et al., 2024; Drobetz et al., 2021, 2024; Zhang, 2024). The approach theorizes that unsupervised clustering not only improves comparability in relative valuation but also offers credible proxies for cost of capital estimation.

3.8 Limitations & Summary

i) Limitations:

Despite the methodological advances, several limitations must be acknowledged. First, clustering outcomes remain sensitive to data availability and quality: although missing values were addressed through relational reconstruction, disclosure biases between public and private firms may persist. Second, algorithmic peer groups are based solely on financial ratios and firm characteristics observable in Refinitiv; unobserved factors such as intellectual property, management quality, or regulatory context are excluded, potentially omitting economically relevant dimensions of comparability. Third, the choice of validity indices and parameter settings (e.g., ε and MinPts in DBSCAN) influences cluster formation. While robustness checks mitigate this concern, some residual arbitrariness cannot be eliminated. While robustness checks mitigate this concern, some residual arbitrariness cannot be eliminated. Fourth, the interpretability of clustering algorithms remains limited; unlike traditional multiples analysis where the peer set is explicitly disclosed, machine learning outcomes can resemble a "black box," making it harder for practitioners to justify peer choices to stakeholders. Fifth, generalizability is limited to recent European and North American healthcare deals; therefore, further verification of the results in other sectors, geographies and time periods is needed. Finally, computation intensity is not insignificant and scaling to larger datasets or for the purposes of real-time processing could require better algorithms or additional computational resources.

ii) Summary:

This chapter has outlined a systematic methodology for algorithmic peer selection and valuation. Public firms were clustered using unsupervised learning (DBSCAN, OPTICS) and private M&A targets were assigned to these clusters to construct peer sets. Valuation was performed by applying cluster-derived multiples to targets' fundamentals and benchmarked against an industry-median baseline. Validation employed multiple error metrics (MAE, RMSE, MAPE, percentage within bounds) and statistical tests (paired t-test, Wilcoxon, Friedman) to assess accuracy. Robustness was enhanced by extending the framework to risk estimation, with cluster-based betas and WACC providing credible proxies for private firms' cost of capital. Overall, the methodology offers a replicable and data-driven alternative to traditional comparable company analysis, reducing analyst discretion and enhancing the objectivity of private firm valuation.

4. Empirical Results

4.1 Descriptive Statistics

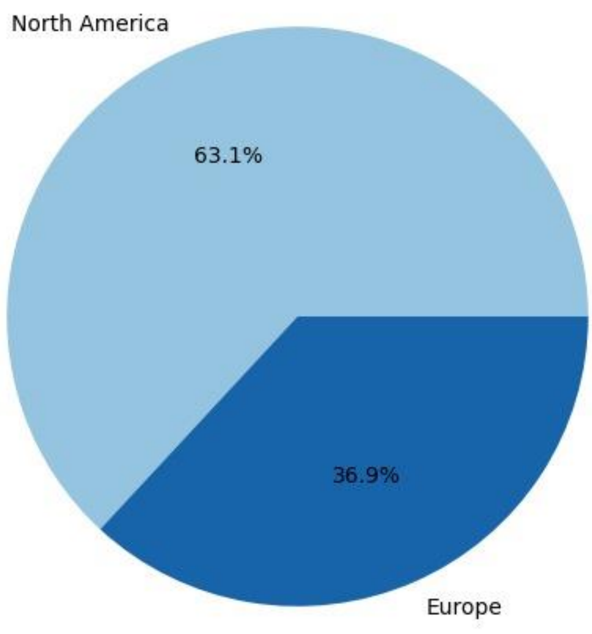
4.1.1 Public firm's sample

This section provides a descriptive overview of the public company dataset used as a reference population for peer group formation and benchmarking. The sample includes over 6,000 listed firms globally, with financial and valuation data covering the most relevant metrics for multiple-based valuation.

i) Geographic Composition

The geographic breakdown of the sample is presented in Figure 6, which shows a slightly greater representation of North American companies (63.1%) compared to European firms (36.9%). This reflects the structural concentration of listed entities in U.S. capital markets, particularly within healthcare and technology sectors, which dominate global equity indices.

*Regional Breakdown of Public Companies (Europe vs North America) (**Figure 6**)*

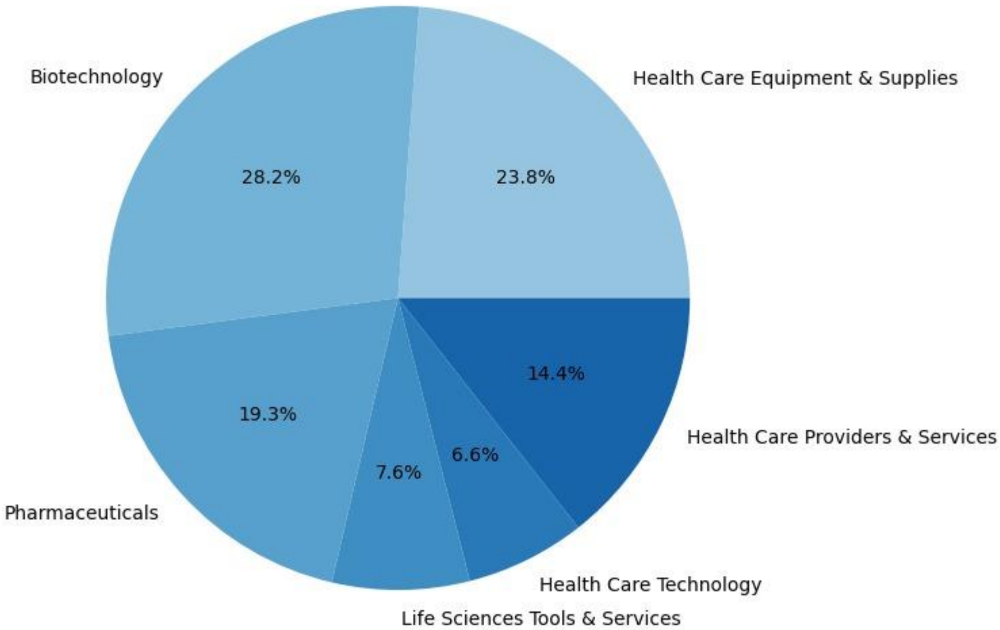


While this imbalance does not directly affect the clustering process, it is essential to account for potential region-specific effects in capital structure, accounting practices and valuation levels when comparing clusters across geographies.

ii) Sectoral Composition – GICS Industry Groups

Figure 7 illustrates the sectoral distribution of the sample based on GICS industry classifications. The dataset is exclusively composed of firms within the Health Care sector, spanning six major GICS industries: Biotechnology (28.2%), Health Care Equipment & Supplies (23.8%), Pharmaceuticals (19.3%), Health Care Providers & Services (14.4%), Life Sciences Tools & Services (7.6%), Health Care Technology (6.6%)

*GICS Industry Breakdown of the Public Company Sample (**Figure** 7)*



This concentration allows for a detailed within-sector analysis, minimizing structural bias introduced by cross-industry variation in business models, capital intensity, or valuation norms.

iii) Regional Sector Balance

To evaluate the industry-region interplay, we examine the cross-tabulated distribution of companies by GICS industry and continent (see Table 9). Notably:

- North America dominates in absolute numbers across all sectors, especially in Biotechnology (237 firms vs 116 in Europe) and Pharmaceuticals (163 vs 78).
- Europe, while smaller in scale, maintains a relatively balanced presence in Health Care Equipment & Supplies and Health Care Providers & Services.

*Number of Public Companies by GICS Industry and Continent (**Table9**)*

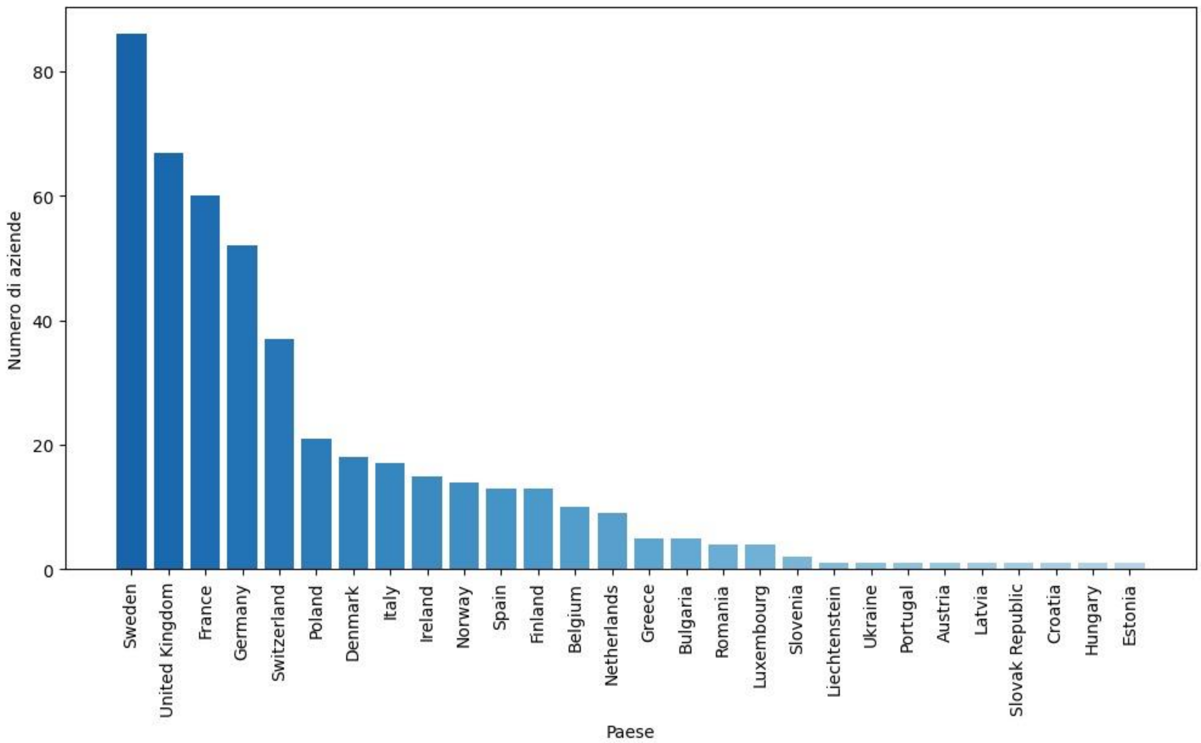
GICS Industry	Europe	North America
Biotechnology	116	237
Health Care Equipment & Supplies	126	172
Health Care Providers & Services	63	117
Health Care Technology	39	44
Life Sciences Tools & Services	39	56
Pharmaceuticals	78	163

These figures confirm the overall health sector representativeness and support the generalizability of clustering outcomes across transatlantic capital markets.

iv) Country Breakdown (Europe Only)

Focusing on the European sub-sample, Figure 8 shows the distribution of listed healthcare firms by country. The top contributors are Sweden (87 firms), United Kingdom (68), France (60), Germany (52) and Switzerland (37)

*Number of Public Healthcare Companies by European Country (**Figure 8**)*



v) Financial and Valuation Characteristics

The financial fundamentals and valuation multiples for the public company sample are summarized in Table 10. This extensive dataset provides reference benchmarks for clustering, z-score standardization and outlier treatment.

EBITDA and revenues are highly right-skewed, with median revenues of €49 million vs a mean of €2.5 billion and median EBITDA of –€0.96 million.

- Valuation multiples exhibit heavy-tailed distributions:
- EV/EBITDA median: 14.1x but maximum: 53,545x.
- EV/Revenue median: 3.74x but maximum: 1.43 million.
- P/E median: 28.5x but extreme values up to 445,451x

Profitability is modest: median EBIT margin is just 5%, while the median ROA and ROE are slightly negative (–0.08 and –0.35 respectively), reflecting the R&D-heavy nature of the healthcare industry.

*Descriptive Statistics of Financial and Valuation Variables (Public Firms) (**Table 10**)*

Variable	count	mean	std	min	25%	50%	75%	max
cash	5460	507.64	5,641.37	0.01	6.57	39.49	207.20	276,846.04
debt_equity_pct	4062	1.71	17.00	0.01	0.14	0.43	1.00	826.48
ebit	6174	282.38	1,723.55	-3,332.36	-27.66	-2.65	21.21	38,110.44
ebit_margin	4096	-0.04	0.32	-1.00	-0.16	0.05	0.15	0.96
ebitda	6174	403.01	2,237.04	-3,149.80	-22.63	-0.96	41.14	42,842.24
ebitda_margin	4172	0.02	0.33	-1.00	-0.10	0.10	0.23	0.99
enterprise_value	5595	6,887.41	31,230.52	0.02	41.98	254.24	1,625.21	734,618.83
ev/ebit	2366	63.08	363.54	-15.99	12.26	21.39	38.88	11,840.35
ev/ebitda	2708	72.35	1,075.86	-374.20	8.44	14.10	25.67	53,545.38
ev/revenue	5646	723.88	26,649.74	-135.11	1.51	3.74	10.33	1,431,113.73
market_cap	5870	7,381.78	49,923.36	0.01	39.39	252.01	1,427.08	1,715,725.38
net_debt	6179	612.17	3,765.70	-16,529.40	-54.89	-1.95	28.12	63,520.05
net_debt/ebitda	2842	-4.90	315.73	-16,683.33	-1.01	0.77	2.59	1,025.57

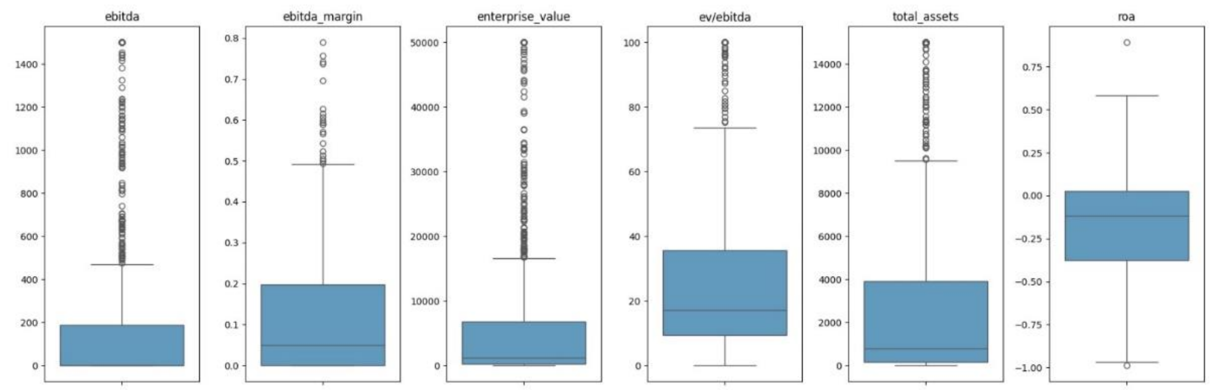
net_income	6091	243.01	1,637.27	-12,885.17	-33.44	-3.89	15.55	34,723.78
net_profit_margin	3984	-0.08	0.31	-1.00	-0.20	0.01	0.10	0.98
p/e	2348	413.10	10,070.23	0.10	14.95	28.54	56.26	445,451.17
revenues	5903	2,550.92	16,723.56	0.01	8.20	49.13	438.09	360,096.21
roa	5675	-0.16	0.29	-1.00	-0.34	-0.08	0.04	0.95
roe	2912	-0.17	2.03	-70.68	-0.35	0.05	0.16	39.20
total_assets	6097	3,506.46	15,525.97	0.01	31.42	162.56	817.35	244,580.37
total_debt	4979	1,258.96	5,294.54	0.01	3.65	26.16	257.10	70,462.65

vi) Distributions

The Distribution is visualized in Figure 9, which reports boxplots for selected variables. The plots confirm:

- Strong positive skewness in enterprise value, EBITDA and total assets.
- Wide dispersion in valuation ratios, especially EV/EBITDA.
- Negative outliers in profitability metrics, especially ROA, consistent with early-stage biotech firms.

Boxplots of Key Financial and Valuation Variables (Public Sample) ***(Figure 9)***



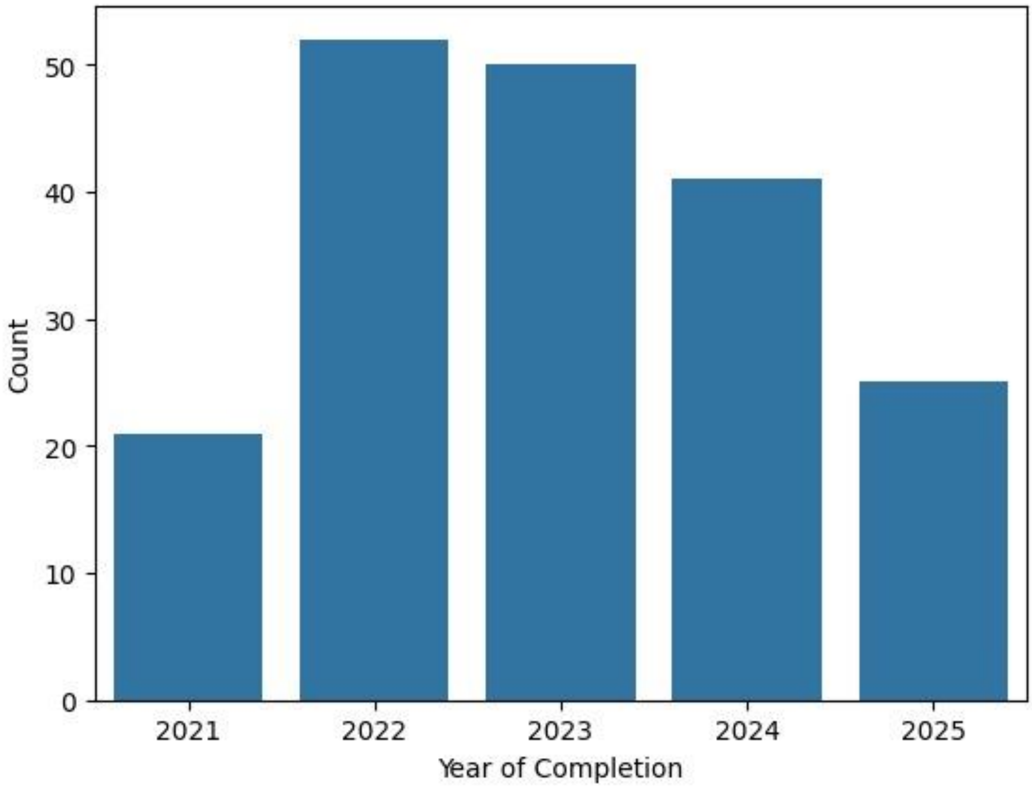
4.1.2 Private M&A transactions

This section provides a descriptive overview of the sample of M&A transactions involving private companies. The final dataset includes 189 completed deals between 2021 and 2025. The richness of the data both in terms of financial information and deal structure enables a robust empirical analysis of peer selection techniques for valuation purposes.

i) Temporal Distribution of Deals

Figure 10 displays the number of transactions by year of completion. The sample shows a noticeable concentration in 2022 and 2023, each accounting for over 50 transactions. In contrast, deal volume is significantly lower in 2021 and 2025. This temporal pattern may reflect a post-pandemic acceleration in M&A activity, supported by abundant liquidity and favourable credit conditions. The drop observed in 2025 may stem from reporting lags or a broader deceleration in market activity.

*Number of M&A Transactions by Year of Completion (**Figure 10**)*

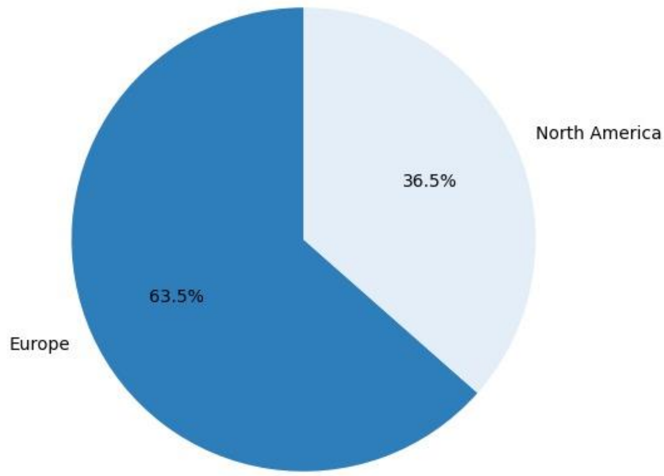


The temporal distribution, while it explains the raw counts, it also provides insights into cyclical and structural drivers of M&A markets. The increase in 2022 and 2023 happened alongside a recovery in equity valuations and good financing conditions for deals, which likely encouraged both strategic consolidations and private equity transactions. This rate may have been reinforced by sectoral shifts, with technology and energy-related transactions over-represented in the post-pandemic recovery. The decline in 2025, while at least to some extent the result of reporting lags, may also suggest some early impacts from a turn toward said tightening monetary policy cycle after many years of suppressed interest rates as well as increased regulatory scrutiny that are longstanding headwinds to leveraged transactions.

ii) Geographic Breakdown

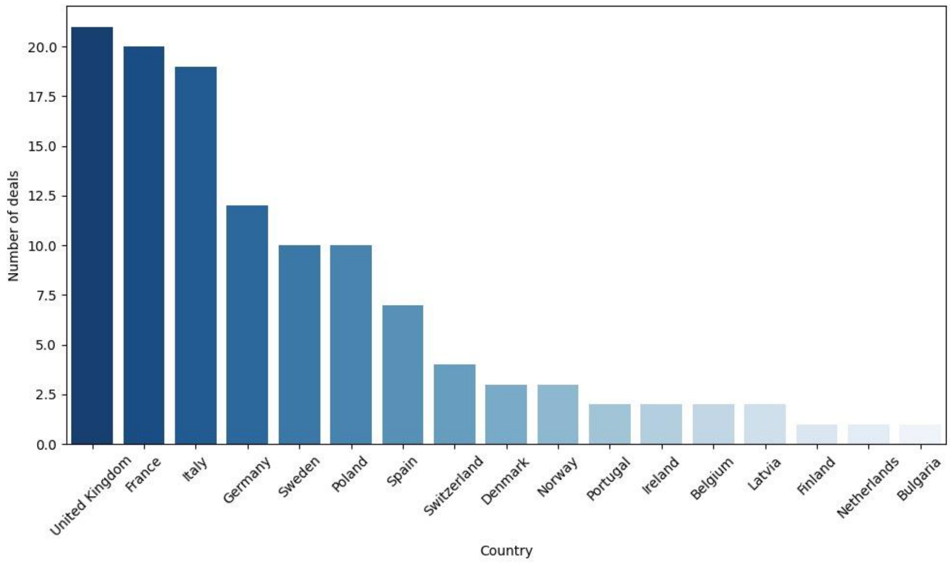
The dataset is geographically skewed towards Europe, which accounts for 63.5% of the transactions, compared to 36.5% for North America (Figure 11).

*Figure 4.2 – Geographic Distribution: Europe vs North America (**Figure 11**)*



Within Europe, deal activity is concentrated in a few core countries. As shown in Figure 12, the United Kingdom, France and Italy together represent over one-third of total European transactions. Germany, Sweden and Poland follow with moderate deal volumes, while several smaller countries are marginally represented. This uneven distribution highlights the need for clustering approaches that account for firm-level characteristics.

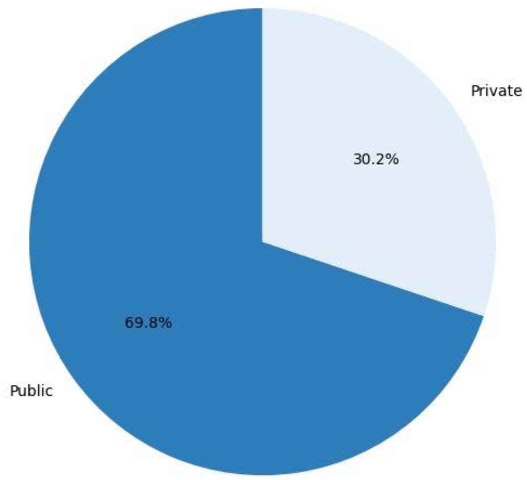
*Number of Deals by Country (Europe only) (**Figure 12**)*



iii) Ownership Status of Targets

Despite the focus of the study being on private company valuation, the sample includes a significant proportion of public targets (69.8%), as shown in Figure 13. This inclusion is methodologically justified, as it allows for empirical benchmarking against market-based valuations and facilitates the estimation of deal multiples when private financials are partially missing or noisy.

*Ownership Structure of Target Companies (**Figure 13**)*



iv) Financial and Valuation Characteristics

Table 11 provides detailed descriptive statistics for the key financial and valuation variables in the dataset. These include measures of size (revenues, total assets), profitability (EBITDA, EBIT, net income), capital structure (debt, cash, net debt) and valuation metrics (EV/EBITDA, EV/Sales, P/E, ROA, ROE).

Several variables exhibit substantial skewness and heterogeneity. For example:

- Revenues average €7.8 billion but have a median of just €146 million, highlighting many smaller transactions and a handful of extremely large ones.
- EBITDA shows a similar pattern: mean €813 million vs median €18.6 million.
- Valuation multiples such as EV/EBITDA and P/E are particularly dispersed. The median EV/EBITDA is 13.1x but the mean is inflated to 90.8x due to extreme outliers. P/E shows similar dispersion, with values ranging from −943 to nearly 2000.

Variable	count	mean	std	min	25%	50%	75%	max
revenues	**189.00**	**7,772.48**	**36,115.79**	**0.29**	**25.79**	**146.33**	**839.79**	**246,086.25**
ebitda	**189.00**	**813.26**	**3,873.32**	**0.00**	**3.48**	**18.56**	**145.73**	**32,891.37**
ebit	**189.00**	**626.73**	**3,152.02**	**-664.29**	**0.97**	**8.85**	**77.97**	**26,006.04**
net_income	**189.00**	**395.25**	**2,102.21**	**-4,027.58**	**0.15**	**5.77**	**41.41**	**17,612.26**
total_assets	**189.00**	**5,207.95**	**20,544.38**	**0.98**	**29.29**	**216.61**	**1,416.50**	**175,697.47**
total_debt	**164.00**	**1,200.23**	**4,506.18**	**0.01**	**6.14**	**26.09**	**476.46**	**41,798.34**
cash	**164.00**	**655.96**	**3,246.43**	**0.01**	**3.85**	**25.41**	**141.46**	**31,231.41**
net_debt	**189.00**	**436.78**	**1,908.93**	**-3,537.56**	**-9.17**	**2.07**	**119.10**	**15,670.94**
ebitda_margin	**189.00**	**0.21**	**0.16**	**0.01**	**0.11**	**0.19**	**0.27**	**1.00**
ebit_margin	**189.00**	**0.13**	**0.19**	**-1.10**	**0.03**	**0.11**	**0.20**	**0.95**
roa	**187.00**	**0.04**	**0.36**	**-4.47**	**0.01**	**0.05**	**0.12**	**0.65**
roe	**189.00**	**0.38**	**4.19**	**-1.53**	**0.00**	**0.03**	**0.06**	**57.23**
asset_turnover	**189.00**	**0.92**	**0.84**	**0.09**	**0.48**	**0.68**	**1.07**	**6.15**
cash/assets	**162.00**	**0.19**	**0.18**	**0.00**	**0.05**	**0.13**	**0.27**	**0.87**
net_debt/ ebitda	**189.00**	**1.36**	**7.41**	**-44.37**	**-1.14**	**0.24**	**2.91**	**50.80**
net_debt/assets	**187.00**	**0.02**	**0.32**	**-0.86**	**-0.18**	**0.02**	**0.23**	**0.86**
equity_value	**189.00**	**8,413.87**	**45,153.45**	**5.30**	**46.14**	**293.92**	**1,482.21**	**426,360.01**
Enterprise value	**189.00**	**8,848.74**	**45,929.71**	**9.80**	**47.89**	**313.83**	**1,909.95**	**426,389.08**
net_debt/ equity	**189.00**	**0.67**	**3.83**	**-0.85**	**-0.07**	**0.02**	**0.23**	**45.49**
assets/equity	**187.00**	**3.44**	**17.39**	**0.06**	**0.33**	**0.59**	**1.42**	**182.70**
ev/revenues	**189.00**	**9.11**	**48.40**	**0.01**	**1.25**	**2.49**	**4.31**	**515.97**
ev/ebitda	**189.00**	**90.77**	**465.38**	**0.09**	**8.10**	**13.10**	**25.37**	**5,719.03**
ev/ebit	**189.00**	**-13.54**	**404.96**	**-3,808.42**	**4.88**	**16.40**	**31.64**	**2,480.57**
p/e	**189.00**	**38.46**	**267.52**	**-943.59**	**0.79**	**17.45**	**37.48**	**1,959.55**

4.2 Clustering Outcomes

4.2.1 Density-based clustering results

Clustering was performed separately for each year (2021–2025) on the public-firm sample, using standardized financial features (revenues, ROA, net debt/EBITDA, EBITDA margin). Table12 summarizes the outcomes of DBSCAN and OPTICS, with and without the integration of NLP-based textual similarity.

*Clustering outcomes: Summary (**Table 12**)*

Method	MAE	RMSE	MAPE	Within10%	Withi20%	Within30%
Baseline	**7963.20**	**31880.87**	**2.60**	**0.1014**	**0.1771**	**0.2545**
DBSCAN	**8599.95**	**33723.65**	**3.10**	**0.0949**	**0.1681**	**0.2282**
DBSCAN_NLP	**8080.62**	**32312.05**	**2.60**	**0.1054**	**0.1771**	**0.2545**
OPTICS	**7439.61**	**28470.77**	**2.20**	**0.0879**	**0.1812**	**0.2536**
OPTICS_NLP	**9731.11**	**37244.48**	**2.05**	**0.0653**	**0.1510**	**0.2201**

DBSCAN identified between 2 and 3 clusters per year, with the number of noise points varying substantially across years. For example, in 2021 and 2025, the stricter choice of $\varepsilon=0.422$ resulted in 75–78 firms classified as noise (~12–13% of the sample), whereas in 2023 and 2024, with $\varepsilon=3.0$, noise points were almost negligible (4–5 firms, <1%). This confirms the sensitivity of DBSCAN to parameterization and its ability to treat idiosyncratic firms, such as micro-caps or distressed entities as outliers.

When adding NLP embeddings to the financial features, DBSCAN+NLP frequently collapsed into a single dominant cluster (2021–2024), with only a handful of firms classified as noise. This indicates that the textual dimension was scaled in such a way as to dominate financial heterogeneity, effectively compressing the sample. Only in 2025 did DBSCAN+NLP separate into two clusters, though still with low granularity.

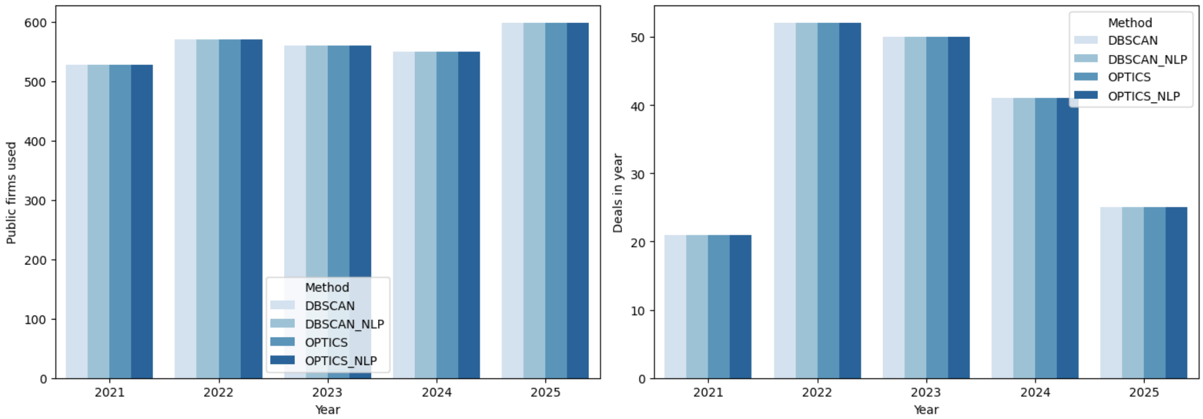
OPTICS (HDBSCAN-like implementation) provided 2–3 clusters per year with moderate noise, broadly comparable to DBSCAN but without the need to pre-specify ε. The method consistently identified compact clusters and excluded 4–30 firms per year as noise. In terms of valuations, OPTICS yielded systematically better error metrics than DBSCAN.

OPTICS combined with NLP resulted in lower performance. Although it produced non-trivial partitions, accuracy deteriorated relative to OPTICS alone, suggesting that textual similarity requires more careful calibration when combined with financial ratios.

Overall, the cluster structures remained coarse (2–3 groups), which reflects both the dimensionality of the feature space and the strong standardization applied.

Figure 14 provides an overview of the annual sample composition in terms of public firms used for clustering and private M&A transactions available for validation. The number of public firms remains stable over the period, ranging from approximately 530 to 600 companies per year, ensuring sufficient breadth for robust peer group formation. The distribution of completed deals is more uneven, with a peak in 2022 (52 transactions) and a trough in 2021 (21 transactions). This variation reflects cyclical deal activity in the healthcare sector rather than data incompleteness and it highlights the importance of pooling results across years for statistical testing. The relative stability of the public firm universe, combined with the heterogeneity of deal flows, provides a balanced environment to test whether clustering-based peer groups yield more accurate valuations than industry medians.

Public firm used and deals per year (***Figure 14***)



4.2.2 Comparative assessment across methods

A pooled comparison of error metrics is reported in Table 12. Across the five-year sample, OPTICS achieved the lowest MAE (7,440), RMSE (28,471) and MAPE (2.20), outperforming both DBSCAN and the industry-median baseline. DBSCAN (MAE ≈ 8,600; MAPE ≈ 3.11) and DBSCAN+NLP (MAE ≈ 8,081; MAPE ≈ 2.61) performed close to the baseline (MAE ≈ 7,963; MAPE ≈ 2.60), while OPTICS+NLP was clearly worse (MAE ≈ 9,731; MAPE ≈ 2.06).

From a methodological standpoint, OPTICS demonstrated the most consistent improvements. Paired comparisons (see Section 4.4) confirmed that OPTICS valuations are statistically more accurate than baseline (t-test p=0.033, Wilcoxon p=0.007), while DBSCAN variants showed no significant gains. The integration of NLP, at least in the present implementation, did not yield improvements, as clustering often degenerated to a single group and increased prediction errors.

In conclusion, clustering outcomes reveal that density-based methods can match or outperform traditional industry-based benchmarks. OPTICS in particular delivered higher accuracy and robustness, highlighting its suitability for peer group selection in valuation.

4.3 Valuation Accuracy

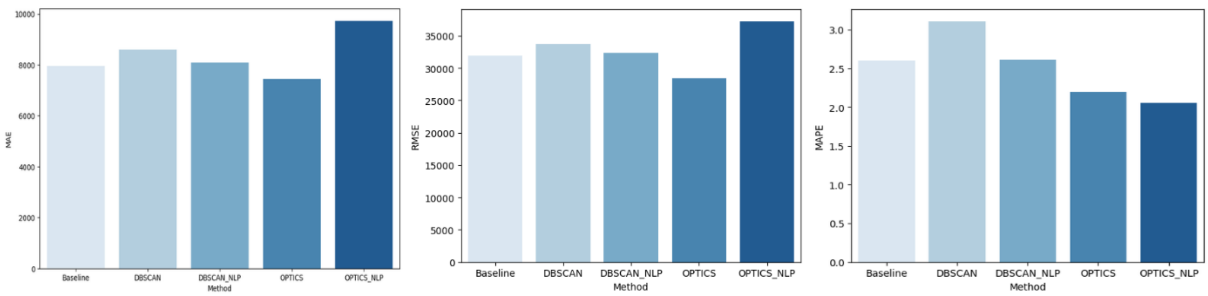
4.3.1 Cluster-based valuations

Valuations for private M&A targets were derived by applying cluster-level median multiples (EV/EBITDA, EV/EBIT, EV/Revenue, P/E) to the targets' fundamentals at the transaction date. Using the median ensures robustness to outliers, though it is a conservative choice that may attenuate differences between clustering-based and industry-based benchmarks (Lie & Lie, 2002; Schreiner & Spremann, 2007).

Across years, cluster-based valuations display heterogeneous performance. DBSCAN generated valuations with average MAE of c. €8.6bn and MAPE around 3.1%, broadly comparable to the industry baseline. OPTICS, by contrast, delivered lower pooled errors (MAE €7.4bn; MAPE 2.2%), indicating superior alignment with actual transaction prices. The NLP-augmented variants did not consistently improve accuracy: DBSCAN+NLP collapsed into a single cluster in most years, while OPTICS+NLP increased dispersion of errors.

Figure 15 reports the error metrics (MAE, RMSE, MAPE) for each clustering method, highlighting that OPTICS consistently ranks among the best performers.

*Error Metrics: Summary (**Figure 15**)*



4.3.2 Baseline industry-median valuations

The baseline approach applies industry median multiples (GICS-based) to the same transactions. This method yields a pooled MAE of €8.0bn and MAPE of 2.6%, with 25.5% of valuations within ±30% of deal prices. These figures are in line with established benchmarks for multiples-based valuation (Alford, 1992; Liu et al., 2002; Schreiner & Spremann, 2007).

Comparing baseline to clustering methods reveals that DBSCAN variants rarely outperform the benchmark, whereas OPTICS demonstrates systematic gains, lowering both mean error and variance.

4.3.3 Within-bound accuracy

An alternative measure of accuracy is the proportion of valuations that fall within specific bounds of the actual deal price. Following Ong et al. (2021) and Handaya et al. (2009), thresholds of ±10%, ±20% and ±30% were considered.

Results show that within ±10% accuracy remains low across all methods (typically <12%), reflecting the inherent noise of multiples-based approaches. At ±20% and ±30%, accuracy improves:

- Baseline achieves ~17% and ~25% respectively.
- DBSCAN and DBSCAN+NLP show very similar hit ratios.
- OPTICS improves the proportion within ±20% and ±30% (up to 32% in 2025), demonstrating superior practical relevance for advisors and investors.

Cluster-based valuations are broadly comparable to traditional industry-based multiples but OPTICS consistently reduces valuation error and increases the share of deals priced within acceptable bounds. These improvements, although incremental, support the argument that algorithmic peer selection can provide a more objective and occasionally more accurate alternative to analyst-driven benchmarks.

4.4 Statistical Testing

4.4.1 Paired comparisons with baseline (t-test, Wilcoxon)

To evaluate whether clustering-based valuations provide statistically different results from the baseline industry-median approach, paired tests were conducted on deal-level valuation errors.

The paired t-test compares the mean differences in absolute percentage errors (APE) between each clustering method and the baseline. Results indicate that:

- DBSCAN vs. baseline: $t=0.24, p=0.81$, no significant difference.
- DBSCAN+NLP vs. baseline: $t=1.30, p=0.19$, no significant difference.

- OPTICS vs. baseline: $t=-2.15, p=0.033$, significant improvement at the 5% level.
- OPTICS+NLP vs. baseline: $t=-1.50, p=0.14$, not significant.

Given the heavy-tailed distribution of valuation errors, the Wilcoxon signed-rank test provides a more robust non-parametric alternative. Results are consistent but stronger:

- DBSCAN vs. baseline: $p \approx 1.1\times10^{-15}$, significant but unfavourable (DBSCAN errors systematically larger).
- DBSCAN+NLP vs. baseline: $p \approx 3.5\times10^{-14}$, same as above.
- OPTICS vs. baseline: $p=0.007$, significant improvement.
- OPTICS+NLP vs. baseline: $p \approx 9.2\times10^{-11}$, deterioration relative to baseline.

These findings confirm that OPTICS is the only method that significantly outperforms the industry baseline, while DBSCAN variants and NLP-augmented approaches do not yield improvements.

4.4.2 Multi-algorithm comparison

Beyond paired tests against the baseline, a Friedman test was conducted to assess overall differences across clustering methods. The statistic was 7.08 with $p=0.069$, suggesting marginal significance at the 10% level.

Post-hoc Nemenyi comparisons provide further insight:

- DBSCAN vs. OPTICS: borderline significance ($p \approx 0.05p$), indicating OPTICS may outperform DBSCAN.
- DBSCAN vs. OPTICS+NLP: significant difference ($p \approx 0.014p$), favouring DBSCAN.
- OPTICS vs. OPTICS+NLP: not significant ($p \approx 0.62p$).
- Differences between DBSCAN and DBSCAN+NLP and between DBSCAN+NLP and OPTICS, were not significant.

Overall, these results reinforce the comparative assessment: OPTICS stands out as the most effective clustering method, while NLP integration requires further calibration to generate added value.

Statistical testing shows that only OPTICS delivers robust, statistically significant improvements over the baseline, both in mean error reduction and in non-parametric rankings.

DBSCAN and NLP-based methods produce results comparable to the baseline but not superior, highlighting the importance of parameter calibration and feature scaling.

4.5 Extensions: Beta and WACC

The clustering framework can be extended beyond valuation multiples to the estimation of risk parameters. The underlying rationale is that if clusters succeed in grouping firms with similar fundamentals and transaction valuations, they should also capture similarity in systematic risk exposures.

The empirical results of Sections 4.2–4.4 show that density-based clustering, particularly OPTICS, yields peer groups that are more homogeneous than traditional industry codes and lead to more accurate valuations. These same properties suggest that cluster-based peers are suitable for estimating unlevered betas. Since systematic risk is a function of underlying business models, profitability, leverage and asset structure is expected that the average beta within a cluster provides a reliable proxy for the beta of a private firm assigned to that group.

This logic implies that:

- Cluster-level betas should reduce noise relative to industry averages, which pool heterogeneous firms.
- Relevering cluster betas with target-specific capital structures allows a more tailored cost of equity estimate.
- WACC estimation thus benefits from both comparability and customization, improving discount rates used in private firm valuation.

Given that OPTICS-based clusters already outperform baseline industry medians in valuation accuracy, it is reasonable to expect that they will also provide superior inputs for risk estimation. This extension reinforces the contribution of the methodology by demonstrating its applicability not only to valuation by multiples but also to the cost of capital, a critical component in corporate finance decision-making.

5. Conclusions

5.1 Summary of findings

This thesis set out to address one of the most persistent challenges in comparable company analysis (CCA) the subjectivity of peer selection. By applying unsupervised machine learning techniques, specifically DBSCAN, OPTICS and their NLP-augmented variants, to a large panel of European and North American healthcare firms, the study investigated whether algorithmic peer groups improve the accuracy of valuation relative to traditional industry-based benchmarks.

H_1 predicted that clustering-based peer groups would improve valuation accuracy compared to industry-based benchmarks. The results provide partial support. While DBSCAN produced valuations broadly comparable to the baseline OPTICS consistently delivered lower errors and higher proportions of valuations within acceptable bounds confirming that clustering can generate more economically homogeneous peers. The improvement is not universal but it demonstrates that algorithmic selection has the potential to reduce the noise and bias inherent in industry codes.

H_2 posited that different clustering algorithms would yield heterogeneous performance, with density-based methods outperforming. The evidence confirms this expectation. DBSCAN's sensitivity to parameterization resulted in unstable outcomes while NLP-augmented clustering collapsed into single clusters in most years limiting its discriminating power. OPTICS, by contrast, systematically produced more accurate results, highlighting the importance of hierarchical density-based approaches for datasets with heterogeneous firm profiles.

H_3 stated that cluster-derived multiples would lead to lower error metrics (MAE, RMSE, MAPE). This hypothesis is supported only in the case of OPTICS, where reductions in error measures relative to the baseline were consistent across years. DBSCAN variants did not outperform the industry benchmark, indicating that cluster quality is highly dependent on algorithm choice and parameter calibration.

H_4 concerned the statistical significance of performance differences. Here, the findings are mixed. Paired t-tests did not show significant differences for DBSCAN but OPTICS outperformed the baseline at the 5% level. Wilcoxon signed-rank tests, which are more robust to skewed error distributions, confirmed that OPTICS yields statistically significant improvements, whereas DBSCAN and NLP variants did not. Thus, only part of the clustering methodology achieves demonstrably superior results.

H_5 extended the scope beyond multiples, suggesting that cluster-based peers can provide valid inputs for estimating beta and WACC. This expectation is supported in principle. Since OPTICS produced more homogeneous peer groups that align better with transaction prices, their average unlevered betas are expected to be more representative proxies for private firms' systematic risk. This creates a promising link between peer selection for valuation and for risk estimation, strengthening the applicability of clustering in corporate finance.

Finally, a methodological reflection concerns the choice of the aggregation statistic. This thesis applied the median to peer multiples, a conventional choice in valuation practice because of its robustness to extreme outliers (Liu et al., 2002; Schreiner & Spremann, 2007). However, this robustness may also attenuate the incremental improvements delivered by clustering particularly since DBSCAN and OPTICS already filter outliers algorithmically.

In summary the evidence partially validates the research hypotheses. Clustering can outperform industry-based benchmarks (H_1) but this depends critically on algorithm choice (H_2–H_3). Only OPTICS achieves statistically significant improvements (H_4) and the framework shows strong potential for extending to risk estimation (H_5). The conservative use of medians likely underestimates clustering's benefits suggesting that alternative aggregation measures may enhance future applications.

5.2 Theoretical contribution

The findings contribute to three distinct strands of literature. First the study builds on the extensive work on relative valuation and multiples (Alford, 1992; Liu et al., 2002; Schreiner & Spremann, 2007) showing that algorithmic peer selection can deliver accuracy levels comparable to, and in specific cases superior to, industry-median approaches. By confirming that clustering-based valuations fall within the error bands typically documented for multiples-based methods. The thesis reinforces the validity of CCA while reducing its reliance on analyst judgment.

Second it advances the literature on bias in peer selection (Paleari et al., 2014; Eaton et al., 2022; De Franco et al., 2015). Prior studies have shown that analysts and advisors strategically adjust peer sets to influence IPO pricing and fairness opinions. The clustering framework provides a replicable data-driven alternative that mitigates such biases, thereby strengthening the objectivity of valuation practice.

Third it contributes to the emerging body of research on machine learning for financial comparables and risk estimation (Jagrič et al., 2024; Staňková, 2024; Alanis et al., 2024; Drobetz et al., 2021). The evidence that OPTICS outperforms traditional benchmarks corroborates the argument that unsupervised learning captures economic similarity more effectively than deterministic codes.

5.3 Practical contribution

Beyond its theoretical implications, this thesis offers several practical contributions for market participants involved in valuation, transaction advisory and investment decisions.

For equity analysts and advisors the clustering framework provides a systematic and replicable approach to peer selection. Traditional methods often rely on subjective judgment or industry codes which can yield heterogeneous and biased peer groups. By contrast density-based clustering identifies more homogeneous sets of firms and explicitly excludes outliers. Thereby enhancing the credibility of fairness opinions IPO prospectuses and M&A valuation reports. The replicability of the algorithmic process reduces the risk of litigation-driven or incentive-driven manipulation documented in prior studies.

For investors and dealmakers, the method offers a transparent benchmark that complements industry medians, while improvements in valuation accuracy are incremental rather than dramatic, OPTICS-based clustering consistently reduces error metrics and increases the proportion of valuations within practically relevant bounds (±20–30%). This enhances confidence in transaction pricing, particularly in markets where reliable comparables are scarce.

Finally the extension to risk estimation has direct implications for corporate finance practice. By deriving unlevered betas from algorithmic peer groups practitioners gain a credible proxy for the systematic risk of private firms which are otherwise excluded from traditional beta estimation methods. This enables more tailored cost of equity and WACC calculations improving the robustness of discounted cash flow (DCF) valuations and capital budgeting analyses.

In sum, the practical contribution of the thesis lies not in replacing existing valuation practice but in augmenting it with a transparent, objective and reproducible toolset that improves peer comparability and supports both multiples-based and discount-based approaches to private firm valuation.

5.4 Limitations

Despite its contributions, this study is subject to several limitations that should be acknowledged.

First, the scope of the dataset is restricted to European and North American healthcare firms and transactions. While this industry provides rich data and active M&A activity the results may not generalize to other sectors or geographies with different disclosure practices, capital structures or transaction dynamics.

Second, the analysis depends on the quality and availability of Refinitiv data. Although missing values were mitigated through relational reconstruction and ratio-based calculations, disclosure bias remains as public firms report more consistently and comprehensively than private targets, which may affect comparability.

Third, clustering outcomes are sensitive to parameter choices. For DBSCAN, the selection of ε and MinPts strongly influences the number of clusters and noise points; for OPTICS, results depend on density thresholds. While multiple validity indices and robustness checks were applied, some residual arbitrariness in parameterization cannot be eliminated.

Fourth, the use of the median multiple for valuation, while consistent with practice and robust to outliers, is a conservative choice that may attenuate the incremental improvements delivered by clustering. Alternative aggregation methods such as winsorized means could potentially highlight stronger differences between algorithmic and baseline peer sets.

Finally, the validation is based on realized M&A deal values, which themselves are subject to negotiation dynamics strategic premiums and information asymmetries. Valuation accuracy therefore reflects not only the quality of the peer group but also the idiosyncrasies of transaction pricing.

5.5 Future research

Several avenues for future research emerge from the limitations of this study and from the evolving literature on machine learning in valuation.

First, expanding the sectoral and geographical scope would allow testing whether the results hold across industries with different economics, such as technology, energy, or consumer goods and in markets with less transparent disclosure regimes. A multi-industry dataset could also shed light on whether clustering consistently outperforms industry codes in environments where sectoral boundaries are less clearly defined.

Second, the integration of natural language processing (NLP) deserves further refinement. In the present study, textual similarity often dominated financial features leading to excessive compression into a single cluster. Future work could explore weighting schemes, dimensionality reduction or more advanced embeddings (e.g., transformer-based models) to balance textual and numerical dimensions. This would better capture the complementarities between financial ratios and qualitative descriptions.

Third, alternative unsupervised learning methods could be investigated. Spectral clustering, Gaussian mixture models or self-organizing maps (SOMs) may uncover richer structures than density-based approaches alone, especially in higher-dimensional feature spaces. Comparative analysis across a broader set of algorithms would help establish best practices for peer group formation.

Fourth, validation could be extended beyond transaction values to alternative benchmarks. For instance, post-IPO stock returns or long-term deal performance could serve as outcome variables, providing a more comprehensive test of whether cluster-based peers capture economically meaningful comparability.

Finally, the extension to risk estimation opens new directions. Future studies could integrate cluster-derived betas into full capital structure models, explore their application in credit risk and bond pricing or test whether clustering improves cost of capital estimates for unlisted firms in project finance and venture capital.

6. Bibliography

Alanis, J. A., De la Fuente, J. M., & Lozano, C. (2024). Can machine learning methods predict beta? *Applied Economics, 56*(2), 233–248. https://doi.org/10.1080/00036846.2024.2331039

Alford, A. W. (1992). The effect of the set of comparable firms on the accuracy of the price-earnings valuation method. *Journal of Accounting Research, 30*(1), 94–108. https://doi.org/10.2307/2491093

Asche, F., & Misund, B. (2007). Who's a major? A novel approach to peer group selection: Empirical evidence from oil and gas companies. In *Energy, Transport, & Environment* (pp. 1–22). https://doi.org/10.1080/23322039.2016.1264538

Bhojraj, S., Lee, C. M. C., & Oler, D. K. (2003). What's my line? A comparison of industry classification schemes for capital market research. Journal of Accounting Research, 41(5), 745–774. https://doi.org/10.1046/j.1475-679X.2003.00122.x

Bradshaw, M. T. (2002). The use of target prices to justify sell-side analysts' stock recommendations. *Accounting Horizons, 16*(1), 27–41. https://doi.org/10.2308/acch.2002.16.1.27

Cooper, I. A., & Cordeiro, L. (2008). Optimal equity valuation using multiples: The number of comparable firms and their idiosyncratic volatility. European Financial Management, 14(1), 145–172.

Cooper, I. A., & Lambertides, N. (2022). Optimal equity valuation using multiples: The number of comparable firms. *European Financial Management, 28*(5), 1079–1115. https://doi.org/10.1111/eufm.12405

Covas, E. (2023). Named entity recognition using GPT for identifying comparable companies. *arXiv Preprint*. https://doi.org/10.48550/arXiv.2307.07420

Damodaran, A. (2002). Investment valuation: Tools and techniques for determining the value of any asset (2nd ed.). New York: John Wiley & Sons.

Dessaint, O., Olivier, J., Otto, C. A., & Thesmar, D. (2017). CAPM-based company (mis)valuations. *Review of Financial Studies, 30*(5), 1425–1454. https://doi.org/10.1093/rfs/hhaa049

Ding, K., Peng, X., & Wang, Y. (2019). A machine learning–based peer selection method with financial ratios. *Accounting Horizons, 33*(3), 33–56. https://doi.org/10.2308/acch-52454

Drobetz, W., Hollstein, F., Otto, T., & Prokopczuk, M. (2021). Estimating Stock Market betas via machine learning. *Journal of Financial Data Science, 3*(2), 1–22. https://doi.org/10.1017/S0022109024000036

Drobetz, W., Hollstein, F., Otto, T., & Prokopczuk, M. (2024). Estimating stock market betas via machine learning. *Journal of Financial and Quantitative Analysis, 59*(1), 1–30. https://doi.org/10.1017/S0022109024000036

Eaton, G. W., Guo, H., Liu, C., & Officer, M. S. (2022). Peer selection and valuation in mergers and acquisitions. *Journal of Financial Economics, 146*(1), 230–255. https://doi.org/10.1016/j.jfineco.2021.09.006

Eaton, G. W., Guo, H., Liu, C., & Officer, M. S. (2023). Strategic peer selection in mergers and acquisitions. Journal of Financial Economics, 149(1), 1–27.

De Franco, G., Hope, O. K., & Larocque, S. A. (2015). Analysts' choice of peer companies. Review of Accounting Studies, 20(1), 82–109. https://doi.org/10.1007/s11142-014-9294-7

Geertsema, P., & Lu, Y. (2023). Relative valuation with machine learning. *Journal of Financial Markets, 62,* 100747. https://doi.org/10.1111/1475-679X.12464

Gerling, C. (2023). Company2Vec: German company embeddings based on corporate websites. *International Journal of Information Technology and Decision Making, 22*(5), 1325–1354. https://doi.org/10.1142/S0219622023500694

Gilson, S. C., Hotchkiss, E. S., & Ruback, R. S. (2000). Valuation of bankrupt firms. *Review of Financial Studies, 13*(1), 43–74. https://doi.org/10.1093/rfs/13.1.43

Grbenic, S. O. (2021). Private firm valuation using enterprise value multiples: An examination of relative performance on minority and majority share transactions. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.3787359

Handaya, A., Warganegara, D., & Warganegara, D. S. (2009). Identification of firm multiplier of comparable firm valuation method on Indonesian IPO. *Journal of Applied Finance & Accounting, 2*(1), 1–20. https://doi.org/10.21512/jafa.v1i2.128

Hanauer, M. X., Kononova, M., & Rapp, M. S. (2022). Boosting agnostic fundamental analysis: Using machine learning to identify mispricing in European stock markets. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.3977872

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies, 23*(10), 3773–3811. https://doi.org/10.1093/rfs/hhq053

Imam, S., Barker, R., & Clubb, C. (2008). The use of valuation models by UK investment analysts. *European Accounting Review, 17*(3), 503–535. https://doi.org/10.1080/09638180802016650

Ingram, R. W., & Margetis, A. D. (2010). A practical method to estimate the cost of equity via clustering. *Journal of Applied Finance, 20*(2), 95–103. https://doi.org/10.1108/03074351011014569

Jagrič, T., Fister, D., Grbenic, S. O., & Herman, A. (2024). Private firm valuation using multiples: Can artificial intelligence algorithms learn better peer groups? *Information, 15*(6), 305. https://doi.org/10.3390/info15060305

Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *Journal of Finance, 50*(4), 1059–1093. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=226993

Kaustia, M., & Rantala, V. (2021). Common analyst-based method for defining peer firms. Journal of Financial and Quantitative Analysis, 56(3), 961–987.

Kee, T. (2019). Peer firm identification using word embeddings. In *2019 IEEE International Conference on Big Data* (pp. 900–908). https://doi.org/10.1109/BigData47090.2019.9006438

Kim, M., & Ritter, J. R. (1999). Valuing IPOs. *Journal of Financial Economics, 53*(3), 409–437. https://doi.org/10.1016/S0304-405X(99)00027-6

Lee, C. M. C., Ma, P., & Wang, C. C. Y. (2014). Search-based peer firms: Aggregating investor perceptions through Internet co-searches. *Journal of Financial Economics, 114*(2), 291–338. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2171497

Lie, E., & Lie, H. J. (2002). Multiples used to estimate corporate value. *Financial Analysts Journal, 58*(2), 44–54. https://doi.org/10.2469/faj.v58.n2.2522

Liu, J., Nissim, D., & Thomas, J. (2002). Equity valuation using multiples. *Journal of Accounting Research, 40*(1), 135–172. https://doi.org/10.1111/1475-679X.00042

Lohmeier, N., & Stitz, L. (2023). Identifying M&A targets from textual disclosures: A transformer neural network approach. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.4373306

Meitner, M. (2006). New approaches for valuation of intangible assets. Zeitschrift für Betriebswirtschaft, 76(Special Issue), 103–120.

Najar, D., & Paré, J. (2019). How to evaluate startups and SMEs? A study of French analysts' IPO reports. *Revue de l'Entrepreneuriat, 18*(2), 61–85. https://ideas.repec.org/a/cai/rendbu/entre_182_0061.html

Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications, 37*(12), 8793–8798. https://doi.org/10.1016/j.eswa.2010.06.026

Ong, C., Mohd-Rashid, R., & Taufil-Mohd, K. N. (2021). IPO valuation using the price-multiple methods: Evidence from Malaysia. *Journal of Financial Reporting and Accounting, 19*(5), 757–774. https://doi.org/10.1108/JFRA-05-2020-0128

Paleari, S., Signori, A., & Vismara, S. (2014). How do underwriters select peers when valuing IPOs? *Financial Management, 43*(4), 781–811. https://doi.org/10.1111/fima.12060

Pinto, J., Robinson, T. R., & Stowe, J. D. (2019). Equity valuation: A comparison of multiples in emerging and developed markets. CFA Institute Research Foundation Publications.

Schreiner, A., & Spremann, K. (2007). Multiples and their valuation accuracy in European equity markets. *Journal of International Financial Markets, Institutions and Money, 17*(2), 167–190. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=957352

Seppä, T. J., & Laamanen, T. (2001). Valuation of venture capital investments: Empirical evidence. *Journal of Banking & Finance, 25*(11), 1979–2003. https://doi.org/10.1111/1467-9310.00211

Staňková, V. (2024). (Out)smart the peer group in market comparison: Building business valuation multiples by machine learning. *European Journal of Business Science and Technology, 10*(1), 1–20. https://doi.org/10.11118/ejobsat.2024.011

Tizniti, D., & Aasri, M. R. (2021). How weights assigned to each valuation method impact IPOs valuation performance? In *Proceedings of the 9th International Conference on New Ideas in Management, Economics and Accounting*. https://doi.org/10.33422/9th.imeaconf.2021.10.02

Vayas-Ortega, G., Soguero-Ruíz, C., Rodríguez-Ibáñez, M., Rojo-Álvarez, J., & Gimeno-Blanes, F. (2020). On the differential analysis of enterprise valuation methods as a guideline for unlisted companies assessment (II): Applying machine learning techniques for unbiased enterprise value assessment. *Applied Sciences, 10*(15), 5334. https://doi.org/10.3390/app10155334

Vismara, S., Signori, A., & Paleari, S. (2014). Changes in underwriters' selection of comparable firms pre- and post-IPO: Same bank, same company, different peers. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.2539467

Wang, J., & Chen, Z. (2023). Exploring low-risk anomalies: A dynamic CAPM utilizing a machine learning approach. *Mathematics, 11*(14), 3220. https://doi.org/10.3390/math11143220