LUISS

Master of Science in Corprate Finance

Chair of Advance Corporate Finance

Determinants of a successful start-up exit, focus in the life science industry

A regression model on the variables influencing startup exit

Prof. Murro Pierluigi
SUPERVISOR

Prof.ssa Altieri Michela
CO-SUPERVISOR

De Bari Francesco - 779261
CANDIDATE

Academic Year 2024/2025

Abstract

This thesis investigates the key factors influencing the probability of a successful exit, defined as a merger or acquisition, for life science startups in Spain. By constructing a logistic regression model based on an original dataset of 112 biotech startups spun off from Spanish universities, the study aims to identify financial, structural, and innovation-related variables that significantly correlate with exit outcomes. Unlike much of the literature, which focuses on U.S.-based ecosystems, this research targets a European context and excludes IPOs to reflect the predominant liquidity path in the local biotech landscape. The model incorporates 16 independent variables, including funding levels, funding sources, intellectual property activity, and institutional support, all cleaned and engineered from proprietary databases such as PitchBook and Invespolora.

The findings offer actionable insights for investors, founders, and policymakers aiming to strengthen innovation ecosystems and improve startup success trajectories. By combining quantitative modeling with strategic insights, the thesis provides actionable implications for founders, investors, and policymakers seeking to optimize exit potential in innovation-driven sectors. The findings emphasize the critical role of funding structure, knowledge intensity, and institutional backing in driving successful startup exits in the European biotech landscape.

Index

Chapter I: Introduction

1.1 Research objective

The primary objective of this thesis is to identify and analyze the key factors that influence the probability of a successful exit, defined as a merger or acquisition, for lifescience startups. By developing a logistic regression model based on a curated dataset of university-originated Spanish biotech companies, this study aims to uncover the statistical and strategic relevance of both financial and non-financial variables on exit outcomes. The research specifically excludes IPOs from the definition of exit, focusing solely on M&A activity to better reflect the predominant path to liquidity in the European biotech landscape.

Biotech startups represent a highly dynamic yet risky segment of the entrepreneurial ecosystem. Characterized by long R&D cycles, high capital requirements, and uncertain regulatory pathways, these firms are often born from academic research and face substantial barriers to scaling. In this context, understanding the determinants of a successful exit becomes vital not only for entrepreneurs and investors, but also for policymakers seeking to foster innovation ecosystems.

Despite the growing interest in venture-backed biotechnology, much of the academic literature focuses on U.S.-based companies, public listings, or sector-wide funding trends. This thesis fills a gap by focusing on university spin-offs in Spain, offering a more nuanced view of local innovation processes and capital dynamics. It also aims to bridge the gap between quantitative financial modeling and strategic startup evaluation, providing a multidimensional framework for understanding what drives exits.

1.2 Research questions and hypotheses

This thesis seeks to answer the following key research question:

Which characteristics of biotech startups are most significantly associated with the likelihood of a successful exit via merger or acquisition?

From this, several testable hypotheses are derived:

- **H1:** Higher total funding positively correlates with the probability of exit.
- **H2:** Companies with multiple assets in development (as opposed to a single pipeline) are more likely to exit.
- **H3:** Greater patent activity, especially when normalized per employee, increases the likelihood of exit.
- **H4:** A higher percentage of private funding compared to public funding is associated with a greater chance of exit.
- **H5:** Firms originating from multiple institutions have a stronger knowledge base and higher probability of exit.

These hypotheses are tested through a binary logistic regression model using both raw and derived metrics such as funding per employee and funding per patent.

Chapter II: Literature Review

2.1 Definition of Exit

In the study of entrepreneurship and startup dynamics, the concept of "exit" has emerged as a pivotal construct, serving as both the culmination of entrepreneurial effort and a mechanism through which value is realized by founders, investors, and other stakeholders. Far from being a mere endpoint, exit is increasingly viewed as an integral component of the entrepreneurial process, shaped by a complex interplay of strategic, financial, behavioral, and contextual variables.

The traditional framing of exit often conflated it with failure or firm mortality, largely overlooking the diverse strategic motivations and outcomes that can lead to a company's exit from the market. However, as the entrepreneurial ecosystem has matured, fueled by the proliferation of venture capital, increased availability of startup data, and the institutionalization of exit channels such as IPOs and M&As, so too has the academic understanding of what constitutes an "exit."

This section reviews the evolution of exit theory and provides a comprehensive definition of entrepreneurial exit, drawing on key contributions from economics, strategy, and organizational theory. It distinguishes between involuntary and voluntary exit, introduces the main typologies found in the literature, and highlights the shifting perspective that treats exit not as an anomaly but as a strategic milestone. Through this lens, the exit becomes not simply a firm-level event, but a broader signal of firm performance, market timing, and entrepreneurial intent.

The following subsections explore this redefinition in depth, starting with the reconceptualization of exit from a strategic rather than failure-oriented perspective, followed by classifications and exit typologies, and concluding with insights into the mechanisms and planning behind different exit routes.

2.1.1 Reconceptualizing Exit in Entrepreneurship

Evolution from Failure-Oriented Views to Strategic Framing

Historically, the notion of entrepreneurial exit was narrowly interpreted as the final and often unfortunate stage of a business's lifecycleclosely associated with failure, insolvency, or liquidation. This failure-centric lens dominated early economic and management literature, portraying exit as a negative event signaling a collapse in the firm's performance or the entrepreneur's misjudgment. However, this perception has been progressively challenged and replaced by a more nuanced, strategic understanding of exit as a multifaceted outcome.

Wennberg et al. (2010) played a pivotal role in reframing the concept of exit, proposing that it should be viewed not as the end of the entrepreneurial journey, but as a critical and constructive part of it. The paper argues that exit enables the reallocation of resources, both financial and human, back into the innovation ecosystem, thus fueling new entrepreneurial activity and economic dynamism. Exit, in this sense, is not synonymous with failure, but can be the result of a successful entrepreneurial venture reaching its natural maturity or strategic inflection point.

The same author expand this idea by introducing a typology that reflects the diverse motivations and circumstances behind exits. Drawing on prospect theory, they identify four types of exit outcomes: harvest sale, distress sale, harvest liquidation, and distress liquidation. Their research demonstrates that exit decisions are influenced not solely by firm performance, but also by the way entrepreneurs frame that performance relative to personal benchmarks or expectations. For example, an entrepreneur may opt for a sale despite profitability if the perceived trajectory of the business falls below their initial aspirations.

Coad (2014), through a comprehensive systematic literature review of 178 empirical studies, reinforces the importance of abandoning binary success/failure frameworks in favor of models that reflect the real-world complexity of exit phenomena. He emphasizes that exit is a statistically frequent and often rational outcome, even among growing firms.

His work confirms that exit can serve as a positive mechanism for reallocating underutilized resources and for enabling entrepreneurs to re-enter markets with enhanced experience and networks.

Delmar et al. also contribute to this shift by demonstrating that entrepreneurial ventures follow diverse trajectories, some of which naturally conclude with voluntary exit decisions that are neither reactive nor symptomatic of underperformance. Their findings highlight that exit is often driven by life-cycle considerations, such as changes in the founder's personal goals, strategic redirection, or the desire to capitalize on market timing.

Exit as Part of the Entrepreneurial Process

This evolution in the literature has led to a fundamental conceptual breakthrough: entrepreneurial exit is no longer viewed as an external shock or forced decision but as an integral, and often premeditated, component of the entrepreneurial process itself. Rather than being a reactive move triggered by failure, exit is increasingly treated as a strategic option available to entrepreneurs throughout the venture's development.

According to DeTienne (2010), exit planning should be embedded in the earliest stages of venture creation, influencing key decisions such as legal structure, capital structure, and growth strategy. Entrepreneurs who anticipate various exit pathways, such as IPO, acquisition, or succession, are better equipped to align their business models with those outcomes, leading to more efficient transitions and improved investor confidence.

Furthermore, empirical studies have shown that early exit intentions significantly shape firm behavior. Entrepreneurs who envision exiting within a set timeframe are more likely to adopt aggressive scaling strategies, seek external financing, and build transferable organizational systems. Exit, therefore, becomes not only a financial endpoint but a guiding force in the firm's strategic development.

This reconceptualization has implications not only for entrepreneurs but also for investors, policymakers, and researchers. It broadens the analytical lens through which startup performance is evaluated and calls for more granular data collection on exit outcomes. By acknowledging the diversity and intentionality behind exit strategies, the literature has laid the groundwork for more predictive and prescriptive models, which will be explored in the following sections.

2.1.2 Exit Typologies and Classifications

Harvest vs. Distress Exits

One of the most influential analytical frameworks on entrepreneurial exit distinguishes between "harvest" and "distress" exits, a distinction first formalized by Wennberg *et al.* (2010). In their typology, a harvest exit occurs when a venture is performing well relative to the entrepreneur's reference point, offering the opportunity to capture value, often through a sale or public listing. Contrastingly, a distress exit is prompted by underperformance, compelling the founder to offload the business to salvage remaining value or avoid deeper losses. This behavioral perspective draws upon prospect theory, suggesting that how entrepreneurs frame their firms' performance, as relative gains or losses, profoundly shapes their exit motivations and timing.

Voluntary vs. Involuntary Exits

Exit routes also differ based on the degree of the entrepreneur's agency, whether the exit is voluntary, such as a planned sale of a mature company, or involuntary, such as bankruptcy or forced liquidation. The literature has increasingly recognized that not every exit is precipitated by failure; many are planned decisions aimed at realizing value or achieving personal, strategic objectives. Cefis *et al.* (2022) emphasize, through a systematic review spanning three decades of research, exit as a heterogeneous phenomenon with varied motivations, from voluntary liquidation to mergers or IPOs, rather than a uniform signal of failure.

Detailed Exit Modalities: IPO, Acquisition, Liquidation, MBO, and Acquihire

Beyond the dichotomies of performance and voluntariness, the literature classifies exit modes more granularly, each with its own characteristics and implications:

- Initial Public Offering (IPO): Taking a company public is often the pinnacle of the harvest exit, offering liquidity to founders and investors while continuing business operation under public markets.
- Acquisition: This remains the most common exit path in practice, where another firm acquires the startup, offering swift liquidity and strategic scale opportunities.
- Liquidation: Particularly under distress, this can occur voluntarily or involuntarily, involving the winding down of operations and asset distribution.
- Management Buyout (MBO): Here, the company's management team acquires ownership, typically with financial backing, resulting in a form of exit that maintains operational continuity under familiar leadership.
- Acquihire: Especially relevant in technology and creative industries, this route emphasizes onboarding talent rather than acquiring products or market share.

Coad (2014) underscores the diversity of these routes, insisting that exit cannot be subsumed in a single category; instead, distinguishing between them is essential for accurate theoretical modeling and practical policy application.

2.1.3 Strategic Value of Exit for Stakeholders

Exit as a Liquidity Event for Entrepreneurs and Investors

In the entrepreneurial context, the exit is a pivotal moment that allows entrepreneurs and investors to convert equity into financial returns. For entrepreneurs, exit serves as a reward for years of risk-taking, innovation, and value creation. Chemmanur et al. (2009) highlight that the decision between an IPO and an acquisition is heavily influenced by the founders' strategic objectives: IPOs offer greater visibility and the possibility to remain involved post-listing, whereas acquisitions provide faster liquidity with potentially less regulatory burden.

For venture capital (VC) investors, exit is the essential mechanism through which returns are realized. VC-backed startups are structured with a clear expectation of a high-value exit within a defined time horizon. These events close the investment cycle, validating the fund's performance to limited partners and enabling capital recycling into new ventures.

Strategic Alignment and VC Preferences

The type of investor involved significantly shapes the preferred exit path. Pérez-Castrillo and Iglesias (2012) observe that independent venture capital (IVC) funds typically aim for IPOs due to their potential for high returns and reputation-building, while corporate venture capital (CVC) funds are more inclined toward acquisitions, often for strategic integration rather than purely financial motives. These differences reflect deeper divergences in investment mandates, exit horizons, and value creation models.

In many cases, the exit is not just a financial event but a strategic move embedded within the investor's broader portfolio and corporate strategy. It may even serve as a signaling device, validating the firm's market fit and attracting further capital into the sector.

Capital Reallocation and Ecosystem Impact

Beyond individual stakeholders, exit events hold systemic value for entrepreneurial ecosystems. Exits, especially IPOs often signal a maturing innovation environment, encouraging new entrants and investment inflows. According to Cumming and MacIntosh (2003), successful exits contribute to efficient capital reallocation, as proceeds from one venture are reinvested into others, either by VCs or by "serial entrepreneurs" who reintegrate into the ecosystem with new ventures or as angel investors.

Moreover, exits help release human capital, skilled employees and executives, who go on to create or support other ventures, reinforcing the dynamism of innovation clusters. In this light, exit is not an endpoint but a catalyst for renewal and growth within the startup economy.

2.1.4 Data-Driven and Predictive Approaches to Exit Classification

From Binary Success/Failure to Multi-Class Typologies

Historically, startup outcomes have often been reduced to a binary framework: success (typically defined as an IPO or acquisition) versus failure (bankruptcy or inactivity). However, recent research has moved beyond this reductive view, embracing multi-class classification frameworks that reflect the heterogeneity of exit modes. Vadali et al. (2024) propose a model that distinguishes among six exit types: IPO, acquisition, merger, leveraged buyout, management buyout, and acquihire. Their work marks a critical shift in how exit outcomes are conceptualized and analyzed, aiming to better mirror the strategic nuance of real-world decisions.

This transition reflects broader changes in the startup landscape, where exits are no longer treated as singular endpoints but as strategic alternatives that vary based on firm maturity, investor profile, and industry dynamics.

Machine Learning and Predictive Algorithms

The rise of big data and open startup datasets such as Crunchbase, PitchBook and many others, has enabled researchers to develop machine learning models capable of predicting exit likelihood and type with high accuracy. These models frequently utilize ensemble methods, such as XGBoost, random forests, and gradient boosting, that are trained on variables like funding history, team composition, sector classification, and geographic location.

Vadali et al.'s ensemble approach outperforms traditional logistic regression in predicting not just whether a firm will exit, but how. By including variables such as investor type, market conditions, and company metadata, their models achieve higher precision and recall across exit categories. This advancement has practical relevance for both investors and accelerators who must allocate resources efficiently across large startup portfolios.

Predictive Value for VCs and Ecosystem Stakeholders

Predictive modeling is not just an academic exercise, it has clear implications for capital allocation, deal sourcing, and risk mitigation. Piekarz (2023), in his capstone thesis, explores how location and sector influence startups' MOIC (Multiple on Invested Capital) and exit trajectories. Using statistical techniques like logistic regression and the Mann–Whitney U test, he identifies clusters of high-performing startups, especially in verticals such as cybersecurity and in geographies like California. His work underlines how predictive analytics can identify favorable exit conditions well before financial outcomes materialize.

Moreover, these predictive systems can be integrated into the strategic planning processes of VCs and accelerators. By using AI-driven models, stakeholders gain foresight into which startups are more likely to achieve successful exits and through which channels, allowing them to tailor support services, funding rounds, and acquisition outreach accordingly.

2.2 Theoretical Models on Business Success and Exit Success

Understanding the determinants of startup exit requires more than empirical observation; it calls for a robust theoretical foundation that explains why and how firms evolve, grow, and ultimately leave the entrepreneurial arena. A wide array of economic, strategic, and organizational theories have been applied to interpret exit decisions, not merely as endpoints, but as pivotal events within the entrepreneurial journey.

This chapter reviews the main theoretical models that have shaped academic and practitioner understanding of firm success and exit outcomes. Traditional frameworks such as the neoclassical and resource-based views offer foundational perspectives on firm performance, while lifecycle theories provide a temporal lens on how startups mature toward exit. These are complemented by signaling theory, which interprets exit as a public declaration of value and credibility, and real options theory, which frames exit as a strategic choice under uncertainty. Agency theory introduces the dimension of conflicting incentives between stakeholders, particularly founders and investors, while newer hybrid models explore how machine learning and artificial intelligence can complement human judgment in predicting exit outcomes.

The objective of this chapter is to connect these theories with empirical insights to build a comprehensive framework for understanding the interplay between startup development, investor behavior, market dynamics, and strategic decision-making in the context of exit. Each subsection integrates both classical and contemporary literature and highlights the practical implications for venture capitalists, founders, and policymakers.

2.2.1 Traditional Theories of Firm Success and Exit

Overview of Neoclassical and Resource-Based Views

The traditional economic perspective on firm exit is rooted in neoclassical theory, which considers exit as a consequence of inefficiency and market-driven failure. Within this framework, firms are assumed to operate in perfectly competitive environments where only the most efficient survive. Exit is viewed as a rational outcome when firms can no longer cover their opportunity costs. However, this narrow interpretation overlooks the strategic dimension of exit decisions, particularly in entrepreneurial settings.

To address this limitation, the resource-based view (RBV) of the firm offers a more internally focused explanation. According to the RBV, firm performance and exit potential are strongly determined by the possession of valuable, rare, inimitable, and non-substitutable (VRIN) resources. These may include proprietary technology, deep domain expertise, or a highly skilled founding team. Firms that accumulate such resources are better equipped to scale, attract investment, and pursue successful exit routes such as IPOs or acquisitions. As highlighted by Coad (2014), this perspective shifts attention from market failure to strategic management of internal assets as the primary driver of exit outcomes.

Firm Lifecycle and Growth Stages Leading to Exit

Another important theoretical framework is the firm lifecycle model, which conceptualizes startups as evolving through a series of predictable stages: opportunity recognition, product development, early market entry, scaling, and eventually, exit. Exit is seen not as a failure but as the culmination of successful development. This framework helps explain why different exit strategies, such as trade sales, IPOs, or management buyouts, occur at different stages of the firm’s lifecycle.

According to Cumming and MacIntosh (2003), the likelihood and type of exit are influenced by the firm's maturity, its funding stage, and the strategic goals of investors. For instance, early-stage exits are more likely in dynamic industries with fast-paced innovation, while later-stage exits are more common in capital-intensive sectors where firms must reach critical mass before becoming viable acquisition targets. The lifecycle model thus provides a temporal lens through which exit readiness and strategy can be evaluated.

Success Determinants

Beyond theoretical constructs, empirical research has identified several key determinants of exit success. Among the most prominent are the characteristics of the founding team, the firm's innovation capacity, sector-specific dynamics, and the ability to time market conditions effectively.

Founding teams with prior entrepreneurial experience, technical backgrounds, or diverse skill sets tend to attract higher-quality investors and are more likely to navigate early-stage challenges successfully. Joffe B. et al (2018), emphasizes the role of human capital in shaping a firm's strategic vision and its attractiveness to external stakeholders.

Innovation, often measured through patent filings, R&D intensity, or unique product offerings, also plays a critical role in driving exit outcomes. Innovative startups signal long-term growth potential to acquirers and public markets, making them more likely to achieve lucrative exits.

Market timing is another critical factor. Firms that raise capital or initiate exit processes during favorable macroeconomic conditions, such as low interest rates, strong equity markets, or periods of high M&A activity, tend to achieve better valuations and face fewer barriers to exit. These success determinants are not isolated; they interact with each other and with external factors, underscoring the complexity of predicting exit outcomes.

2.2.2 Signaling Theory and Exit

Exit as a Signal of Firm Quality to External Stakeholders

The signaling theory in entrepreneurship posits that exit events serve as strong signals of a firm's quality, credibility, and long-term viability to external stakeholders, including future investors, employees, customers, and acquirers. Unlike internal performance metrics, which may remain opaque, the decision to go public or to be acquired can broadcast valuable information about a startup's potential and market fit. In this view, an exit is not merely a terminal event but a communicative act embedded in broader reputational and strategic considerations. As highlighted by Melnychuk (2023), early decisions around capital structure, partner choice, and product development all compound into a final signal at the moment of exit, shaping post-exit valuations and credibility.

IPO and Acquisition as Validation Events

Among exit options, IPOs and strategic acquisitions are often seen as the most prestigious and externally validating. An IPO, in particular, requires intense scrutiny by underwriters, regulators, and public investors, which reinforces its signaling strength. According to the findings presented by Castaldi et al. (2020), firms pursuing IPOs tend to exhibit stronger fundamentals prior to listing, such as revenue growth and R&D intensity, suggesting that only firms able to withstand this scrutiny can credibly send a high-quality signal. Acquisitions, while more private in nature, also function as a form of third-party endorsement. The fact that a larger player is willing to invest capital and absorb the startup typically indicates perceived value, strategic fit, or unique capabilities.

Strategic Timing of Exit and Investor Perception

Timing plays a critical role in the signaling power of exit events. Exiting during market upswings or during periods of high sectoral activity can amplify the perceived success of the firm. As noted in Melnychuk (2023), startups that align their exit with favorable macroeconomic indicators or industry hype cycles tend to secure better terms and post-exit reputational leverage. Furthermore, VCs and angel investors strategically plan exit timing to maximize internal rates of return (IRR), often accelerating or delaying exits based on capital market conditions. In this context, exit becomes part of a broader performance narrative, not only a financial milestone but also a storytelling device used to attract future investors, boost founder visibility, and influence portfolio valuations.

2.2.3 Real Options Theory and Exit Strategy

Exit as a managerial option under uncertainty

Real Options Theory frames strategic decisions, like startup exit, as fixed events, but as *managerial options* under uncertainty. In high-risk environments like venture capital, the exit is not a binary outcome but a decision point with embedded flexibility. Founders and investors may delay, accelerate, or alter the path to IPO, acquisition, or dissolution, depending on how market conditions evolve. According to the Joffe B. et al (2018), real options allow firms to preserve value by waiting for information to unfold, especially relevant when product-market fit, regulation, or user adoption remain uncertain. Exit becomes a strategic hedge, not just a financial milestone.

Value of flexibility in delaying or accelerating exit

One of the core insights of Real Options Theory is that the *option to wait* can often be more valuable than immediate action. For startups, delaying an exit until strategic inflection points, such as regulatory approval, user growth, or ecosystem shifts, can lead to a significantly higher valuation. Melnychuk (2023) underscores that in sectors like

SaaS and healthcare, where timing heavily affects valuation multiples, the ability to defer exit until favorable signals appear is a key strategic asset. Conversely, *accelerated exits* may also be triggered to capitalize on frothy markets, outperforming expected valuations based on fundamentals alone.

Application in VC-backed firms

In venture-backed startups, real options are not only embedded in product development but also in exit planning. Investors often stage their funding in tranches, each representing an option to continue or abandon. Similarly, startups treat each major milestone as a decision node: pivot, scale, raise, or exit. Melnychuk's thesis shows that firms with higher connectivity to diverse investor networks (via centrality metrics) exhibit more “option-like” behavior in exit timing, opting for prolonged fundraising or strategic partnerships before committing to a liquidity event. VC-backed firms thus embody Real Options logic in both capital structure and exit decision-making, using optionality to navigate uncertainty and maximize outcomes.

2.2.4 Agency Theory and Conflicts in Exit Decisions

Conflicts Between Entrepreneurs and Investors on Exit Timing and Method

Agency theory provides a powerful framework for understanding conflicts of interest that arise between entrepreneurs (agents) and investors (principals) in the context of startup exits. These conflicts typically emerge around the timing and type of exit. Entrepreneurs may seek to delay an exit to maximize the firm’s long-term potential or preserve control, while investors, particularly venture capitalists with fixed fund lifecycles, often push for earlier liquidity events to ensure timely returns to their own limited partners.

This divergence is well illustrated in Piekarz and Urruticoechea (2023), who show how startup location, sector, and capital structure can compound these agency-related tensions, especially in sectors with volatile valuation dynamics. Melnychuk (2023) also emphasizes that entrepreneurial resistance to exit is often psychological, driven by identity, vision, and emotional investment, contrasting with investor demands for rational, outcome-based decision-making.

Additionally, fund-specific dynamics can amplify these tensions. As highlighted by Joffe B. et al (2018), corporate venture capital (CVC) funds may favor strategic acquisitions that align with the parent company's broader goals, while independent venture capital (IVC) funds are more likely to prefer IPOs due to reputational and financial benefits. These asymmetries in exit preferences reflect the structural misalignments between capital providers and entrepreneurs, making the negotiation process highly complex.

Role of Contracts and Control Rights in Shaping Exit Paths

To manage these potential conflicts, contractual governance mechanisms, such as vesting schedules, liquidation preferences, and drag-along rights, are routinely embedded into early-stage financing agreements. These legal instruments give investors significant influence over strategic decisions, including exit.

According to Joffe B. et al (2018), control rights held by VCs often include board seats and veto powers, allowing them to shape or even force exit strategies regardless of founder preferences. This concentration of control is designed to protect investor capital but can lead to tensions if founders feel sidelined in key decisions. Melnychuk (2023) expands on this by explaining that such governance structures are often justified by asymmetric risk exposure: while founders have reputational and creative stakes, investors risk large amounts of capital and thus demand stronger decision rights.

Additionally, convertible preferred equity, commonly used in venture financing, gives investors the option to convert their shares into common equity during IPOs or to maintain downside protection in case of acquisition or liquidation. These instruments effectively give the principal (the VC) multiple levers to influence not just whether the company exits, but how.

Influence of Asymmetric Information

Agency theory also accounts for asymmetric information as a major source of inefficiency in exit planning. Entrepreneurs typically possess more detailed and timely knowledge about the firm's operations, prospects, and risks than investors. This information imbalance can lead to suboptimal exit decisions, either through delay, misrepresentation, or misalignment of expectations.

Greg Ross (2021), A machine learning model for startup selection and exit prediction, suggests that predictive models can help mitigate some effects of asymmetric information by surfacing hidden patterns in firm data. However, these tools are only as effective as the transparency and completeness of the underlying inputs. Reiner U. et al. (2013) notes that VCs increasingly rely on performance dashboards and data rooms to bridge the information gap, although these cannot fully substitute for direct managerial access.

Ultimately, agency theory reminds us that exit is not just a financial milestone but a negotiated event, shaped by the relative power, incentives, and information of the involved parties. Successful exit strategies therefore require the careful design of incentive-compatible contracts, transparent communication, and an alignment of vision across stakeholders.

2.2.5 Integrative Models and Hybrid Frameworks

Hybrid Models Combining Human Intuition with Machine Learning

As machine learning and artificial intelligence gain traction in startup evaluation and venture capital decision-making, there is increasing interest in hybrid frameworks that combine human judgment with algorithmic outputs. Rather than replacing human decision-makers, these models aim to augment intuition with data-driven insights, striking a balance between experiential knowledge and empirical rigor. This approach acknowledges the limitations of both purely statistical models, which may overlook qualitative nuances, and purely intuitive processes, which are prone to bias and inconsistency.

Dellermann et al. (2019) explore this concept of hybrid intelligence, proposing that the combination of human contextual reasoning and AI's computational power produces superior results in complex environments like venture funding. In the context of startup exits, such hybrid models might integrate qualitative factors, like founder interviews or strategic vision, with predictive analytics built on structured data such as funding history, patent activity, or founder networks.

The Emerging Role of AI in Startup Evaluation

Artificial intelligence is playing an increasingly central role in startup screening and evaluation pipelines, particularly in the early stages of filtering high volumes of applicants. According to Piekarz and Urruticoechea (2023), predictive analytics can be applied to early startup attributes (e.g., founder education, sector, geography) to forecast long-term success and MOIC potential, even before financial metrics are available.

A more advanced example is provided by the Vadali et al. (2024) model, which classifies startups into six exit types using ensemble learning based on Crunchbase data. Their results show that machine learning approaches significantly outperform traditional binary models, and their granularity allows investors to strategize around distinct exit pathways such as IPOs, acquihires, and leveraged buyouts. These tools are especially valuable for institutional investors and accelerators needing scalable, consistent evaluation processes.

Future of Integrative Decision-Making in Venture Capital

The future of VC decision-making lies in integrative frameworks that blend structured data, unstructured qualitative insight, and algorithmic forecasting. Tools such as CapitalVX, which combine multiple ML models, XGBoost, KNN, Random Forests, demonstrate the potential to reach over 85% prediction accuracy on exit outcomes using real-world data from Crunchbase and USPTO patents. Although the original CapitalVX paper isn't publicly hosted, it is referenced in Piekarz (2023)**,** validating the relevance of ensemble modeling for VC decision-making.

However, these technological advances must coexist with human oversight to account for non-quantifiable factors like ethical judgment, founder motivation, and market intuition. Dellermann et al. (2019) emphasize that in volatile or ambiguous conditions, human-AI collaboration leads to more adaptive and context-sensitive investment decisions.

In conclusion, hybrid intelligence and integrative models represent the next frontier in entrepreneurial finance. By embedding machine learning within a broader human-centric framework, these systems offer the potential to transform how investors assess risk, allocate capital, and support innovation in a rapidly evolving startup ecosystem.

2.3 Economic and financial factors influencing Exit

While strategic intent, theoretical framing, and organizational characteristics all contribute to understanding entrepreneurial exits, the role of economic and financial variables is equally fundamental. Exits do not occur in a vacuum, they are embedded within the broader context of capital markets, investor preferences, macroeconomic cycles, and regulatory ecosystems. This chapter explores the external and internal financial dynamics that shape exit probability, mode, and valuation.

From the internal perspective, startup performance indicators such as revenue growth, profitability, and capital structure critically influence a firm's readiness for exit and its attractiveness to potential buyers or public investors. Financial health not only signals organizational viability but also determines the flexibility a startup has in timing and structuring its exit strategy.

Externally, macroeconomic conditions, including interest rates, IPO market cycles, and M&A activity, exert significant influence over the timing and success of exit events. Market windows can open or close rapidly, and investor behavior often shifts in response to broader economic indicators. These dynamics necessitate strategic awareness and adaptability from founders and investors alike.

Moreover, the type of investor and their investment horizon plays a central role in shaping exit pathways. Whether backed by corporate VCs seeking strategic synergies or independent VCs driven by IRR targets, startups must navigate differing expectations and pressures that directly affect exit outcomes. These pressures intersect with national regulatory environments, legal frameworks, and the maturity of local financial markets, all of which affect the feasibility and desirability of different exit options.

Finally, geographic location and sectoral affiliation emerge as structural predictors of exit. Innovation hubs and high-growth sectors tend to attract more capital and provide more robust exit opportunities. Conversely, startups in peripheral regions or slower-moving industries may struggle to attract acquirers or go public.

This chapter synthesizes empirical findings from a wide range of sources to map how these economic and financial dimensions interact with each other and with firm-specific strategies to shape the entrepreneurial exit landscape. The goal is to offer a multi-level understanding of how success is not only built internally, but also enabled, or constrained, by the surrounding economic and institutional context.

2.3.1 Financial Performance and Capital Structure

Financial Performance and Exit Likelihood

Financial performance is one of the most robust predictors of a startup's likelihood of achieving a successful exit. High revenue growth, operational profitability, and a solid balance sheet increase the attractiveness of the firm to acquirers and public investors, improving the probability of a favorable liquidity event such as an IPO or strategic acquisition. Conversely, poor financial fundamentals often lead to distress-driven exits or liquidation scenarios. As highlighted by Wennberg et al. (2010), entrepreneurs evaluate exit decisions through a dual lens: objective performance metrics and subjective framing, where financial underperformance combined with psychological loss aversion may push founders toward suboptimal exits or delays in decision-making.

Empirical studies reinforce this performance–exit relationship. For instance, Chemmanur and He (2011) present a theoretical model where better-performing firms are more likely to pursue IPOs due to the valuation premium they receive in public markets, while weaker firms may settle for acquisition or private restructuring. Similarly, in the study by Ferrati and Muffatto (2021), startups with consistently high revenue growth and R&D productivity were statistically more likely to complete successful exits, underlining the predictive role of internal financial strength.

Capital Structure and Exit Pathways

Beyond operational metrics, the capital structure of the firm exerts a significant influence on exit pathways. Startups heavily financed through debt often face rigid repayment obligations that limit their flexibility in choosing optimal exit timing. High leverage may also introduce investor pressure for earlier exits to mitigate default risks, especially when cash flows are volatile. In contrast, equity-financed firms, particularly those backed by venture capital, benefit from more strategic autonomy and long-term planning horizons. This difference is critical in shaping not only when a startup exits but how it exits.

Chemmanur et al. (2009) developed a signaling-based model showing that equity-backed firms are more likely to choose IPOs over acquisitions because they can credibly signal their quality to the market through rigorous disclosure, governance, and pricing mechanisms. The model also emphasizes the importance of staged financing, often used by VCs, which allows incremental resource allocation based on performance milestones.

This staged capital inflow enables a more dynamic approach to exit, especially for firms operating in uncertain or high-growth sectors.

Capital structure also interacts with investor type and exit strategy. As discussed by Gompers and Lerner (2001), VC-backed firms not only benefit from access to equity but also gain credibility through investor reputation, which further increases their likelihood of IPO over other forms of exit. This dynamic was also explored by Reiner U. et al. (2013), which found that the alignment between funding strategy and business model scalability was a key variable in successful exits in the Italian and European startup ecosystems.

2.3.2 Market Conditions and Macroeconomic Influences

Macroeconomic Trends and Exit Feasibility

Macroeconomic conditions are a powerful external determinant of startup exit outcomes, shaping not only the availability of exit opportunities but also their quality and valuation. Favorable economic environments, characterized by low interest rates, high investor confidence, and bullish equity markets, tend to stimulate IPO activity and increase the likelihood of strategic acquisitions. Conversely, recessions, credit crunches, and market volatility constrain exit channels, often forcing founders to delay exits or accept suboptimal terms.

Koba (2023) quantitatively analyzed this dynamic by examining how fluctuations in interest rates and M&A volume correlate with exit frequency across OECD countries. The study found that in years marked by high stock market indices and abundant M&A activity, both IPOs and acquisitions rose significantly. This indicates a clear procyclical pattern: exits are more likely to succeed and yield higher valuations during economic upswings, whereas downturns can stall the market or shift exits toward distressed scenarios such as acquihires or fire sales.

Similarly, Piekarz (2023) emphasized that venture outcomes are highly sensitive to public market timing. The research demonstrated that startups that exited during favorable capital market windows (e.g., low interest rates and strong IPO pipelines) experienced higher multiples on invested capital (MOIC) than peers that exited during downturns, highlighting the strategic importance of economic timing in maximizing investor returns.

Strategic Timing and Investor Behavior

Founders and investors alike are aware of these macroeconomic fluctuations and often attempt to time exits accordingly. Strategic exit planning, therefore, includes continuous market monitoring and adaptive decision-making based on economic signals. In the venture capital world, this is particularly salient, as fund managers must balance the fund's investment horizon with market receptivity. When economic conditions are not conducive, exits may be postponed even for high-performing startups.

Chemmanur and He (2011) developed a model in which public market performance and interest rates directly influence whether a firm chooses IPO or acquisition. In periods of economic strength, IPOs are more attractive due to higher expected valuations and liquidity. In contrast, during uncertain periods, firms may favor acquisition paths, which are generally less exposed to public market sentiment and regulatory hurdles.

Reiner U. et al. (2013) also confirms that macroeconomic factors like GDP growth, credit access, and industry-specific cycles influence exit probabilities. For instance, in fastgrowing sectors like fintech or green energy, especially during periods of regulatory support or consumer demand expansion, acquisitions become more common as incumbents seek to consolidate and expand rapidly.

Taken together, the literature demonstrates that while internal firm dynamics are crucial, the external economic landscape plays an equally pivotal role in shaping when and how startups exit. Strategic awareness of macroeconomic variables can significantly enhance exit timing and valuation, offering critical insight for both founders and investors navigating an ever-changing financial environment.

2.3.3 Investor Types and Time Horizons

Corporate vs. Independent Venture Capital Models

A key factor influencing exit outcomes is the type of investor backing the startup. In particular, the distinction between corporate venture capital (CVC) and independent venture capital (IVC) has gained attention in the literature due to their divergent strategic goals and investment horizons. While independent venture capital firms are primarily financially driven and seek high returns through IPOs or strategic acquisitions, corporate VCs tend to have a broader set of motivations, such as gaining technological insights, strategic partnerships, or enhancing the parent firm's innovation capacity.

Pérez-Castrillo and Iglesias (2012) explore this dichotomy by analyzing how the type of venture fund affects startup trajectories. They find that CVC-backed startups are more likely to delay exit and more frequently opt for acquisitions rather than IPOs. This is because CVCs prioritize strategic alignment over rapid financial returns, and are often willing to forgo short-term liquidity for long-term technological or market-based synergies. In contrast, IVCs, which must deliver strong financial performance to their limited partners within a fixed fund lifespan, exhibit greater urgency in pursuing IPOs or high-return acquisitions.

Investor Goals and Strategic Exit Timing

The time horizon and objectives of investors significantly shape when and how startups exit. Independent VCs typically operate under a 7–10 year fund structure, which imposes a clear timeline for achieving liquidity events. As a result, these investors often push for exits once a portfolio company reaches a certain valuation or market milestone, irrespective of broader ecosystem factors. Conversely, CVCs, backed by corporate balance sheets, have more flexibility and tend to support longer development timelines, especially when the startup complements internal R&D or fills a strategic gap within the parent company.

Fitza and Matusik (2021) reinforce this perspective by showing that investors with short-term priorities, such as hedge funds or corporate entities with quarterly earnings pressure, tend to favor early exits through acquisitions. In contrast, traditional venture capitalists and growth equity investors, particularly those with reputational capital at stake, often guide startups toward IPOs as a way to demonstrate high returns and validate their investment thesis to future LPs.

The Reiner U. et al. (2013) adds an additional layer by identifying investor experience and network centrality as factors that accelerate access to exit channels. Startups backed by well-connected VCs are more likely to receive acquisition offers or underwriting support for IPOs, allowing them to optimize exit timing based on both internal readiness and external market windows.

Ultimately, investor type and time horizon are not passive background variables, but active forces in shaping a startup’s strategic decision-making. Their influence extends beyond funding to include governance, board structure, and eventual exit path, all of which must be accounted for in both academic modeling and real-world entrepreneurial planning.

2.3.4 Institutional and Ecosystem Factors

Institutional Context and Exit Opportunities

The broader institutional and regulatory environment in which startups operate plays a pivotal role in shaping their exit trajectories. National ecosystems, defined by legal frameworks, tax regimes, labor laws, and the robustness of financial markets, can either enable or constrain the range of viable exit options. Countries with active and transparent capital markets, efficient judicial systems, and favorable tax policies tend to see more IPOs and value-maximizing acquisitions. In contrast, startups operating in jurisdictions with weak investor protections, high regulatory burdens, or underdeveloped financial ecosystems often face delayed or suboptimal exits.

Coad (2014), in a systematic literature review, emphasizes that many studies underestimate the contextual embeddedness of firm exit. Institutional structures not only determine the availability of exit options but also influence investor confidence, due diligence processes, and post-exit outcomes such as valuation and capital gains taxation. Similarly, Cumming and MacIntosh (2003) compare exit dynamics across countries and find that institutional variation, particularly in legal origins and securities regulation, directly correlates with exit timing and mode. Their findings indicate that common law countries with stronger shareholder protections and more flexible securities law experience faster and more diverse exits than civil law countries with more rigid legal structures.

Local Market Development and Ecosystem Readiness

The maturity of local capital markets, especially IPO markets, plays a significant role in determining whether public listings are a feasible exit route. In regions where IPO markets are illiquid, underdeveloped, or dominated by a few sectors, startups often rely more heavily on acquisitions or private secondary sales to achieve liquidity. Conversely, jurisdictions with dynamic equity markets, robust underwriter ecosystems, and receptive investor bases facilitate smoother transitions to public ownership.

Ferrati and Muffatto (2021) argue that the presence of active IPO markets and transparent governance standards reduces frictions in reaching liquidity events. Their research on European startups shows that companies embedded in more favorable institutional ecosystems, such as those in the UK or Sweden, have significantly higher probabilities of exiting via IPO. Furthermore, the authors stress the importance of soft ecosystem components such as entrepreneurial culture, incubator presence, and interfirm knowledge networks in supporting not just firm growth, but also credible exits.

The role of public policy also cannot be ignored. Governments that offer R&D tax credits, startup-friendly equity crowdfunding laws, and fast-track procedures for listings or acquisitions tend to cultivate ecosystems with more frequent and higher-quality exits.

Conversely, overly rigid employment laws, punitive capital gains taxes, or opaque merger control regulations can stifle entrepreneurial dynamism and reduce investor appetite.

In summary, institutional and ecosystem factors constitute an often-overlooked but fundamentally important layer of the exit landscape. Their influence cuts across all phases of the entrepreneurial journey, from initial funding to final liquidity, and must be accounted for in any robust theoretical or empirical model of startup exits.

2.3.5 Geographic and Sectoral Patterns

Geographic Concentration and Startup Hubs

The geographic location of a startup is a well-documented predictor of its probability to exit and the type of exit it may achieve. Certain regions, often labeled as innovation or startup hubs, exhibit ecosystem characteristics that increase access to capital, mentorship, exit opportunities, and growth-supporting infrastructure. For instance, Silicon Valley, London, Tel Aviv, and Berlin are frequently cited as high-exit-density regions due to their concentration of venture capital firms, acquirers, legal expertise, and experienced entrepreneurs.

Piekarz and Urruticoechea (2023) empirically demonstrate that startups based in California have significantly higher chances of exiting, especially through IPO or strategic acquisition, compared to firms located in less mature ecosystems. Their analysis shows that even after controlling for sector and funding levels, geographic clustering remains a strong determinant of exit likelihood. These findings align with the theory of agglomeration economies, suggesting that spatial proximity to other successful ventures, investors, and knowledge spillovers enhances a firm's chances of growth and favorable exit.

Melnychuk (2023) further supports the relevance of geographic positioning in exit outcomes by identifying correlations between regional startup density and investor activity in Europe. Regions with established capital markets, government-backed innovation programs, and access to international networks tend to generate more exits and attract a larger proportion of later-stage venture capital.

Sectoral Trends and Industry-Specific Exit Dynamics

In addition to geography, industry sector plays a crucial role in shaping exit probabilities and modes. Certain sectors, such as cybersecurity, fintech, biotech, and enterprise SaaS, consistently attract more venture capital attention and exit at higher valuations due to their scalability and market growth rates. Conversely, startups in lower-growth or heavily regulated sectors often face more limited exit pathways.

Piekarz and Urruticoechea (2023) found that cybersecurity startups were among the most likely to generate high multiples on invested capital (MOIC) and to exit via acquisition, given the strategic nature of their intellectual property and the constant M&A interest from larger technology firms. This is consistent with macro-sectoral trends, where tech-based firms dominate M&A and IPO activity due to investor appetite for innovation and recurring revenue models.

Koba (2023) extends this argument by showing that sector volatility and innovation cycles, particularly in emerging technologies, can significantly accelerate or delay exit timing. Startups in industries undergoing rapid disruption tend to experience faster exits, as acquirers seek early access to technological capabilities or market positioning. However, these exits may also be riskier and more sensitive to macroeconomic sentiment.

Furthermore, sector-specific exit norms play a role. In life sciences and biotech, for example, acquisitions by pharmaceutical firms often occur pre-revenue, based on intellectual property and clinical pipeline potential. In contrast, in SaaS and fintech, exits are typically driven by ARR multiples and user metrics. These divergent norms influence how entrepreneurs structure their growth strategies and when they seek liquidity events.

In conclusion, both geographic positioning and sectoral affiliation function as powerful predictors of exit outcomes. Understanding their interaction allows for more precise modeling of entrepreneurial success and provides a roadmap for founders and investors seeking to maximize liquidity potential.

Chapeter III: Methodology

3.1 Choice of the regression model

In order to analyze the determinants influencing the likelihood of a successful EXIT by startups in the life sciences sector, this thesis employs a Logistic regression model. The Logistic model is particularly appropriate for this analysis because the dependent variable is binary, indicating whether a startup has experienced an EXIT event (1) or not (0).

Unlike linear regression models, which may produce predicted values outside the [0,1] interval, the Logistic model ensures that predicted probabilities remain within valid bounds by assuming a cumulative normal distribution of the error term. This makes it well-suited for estimating probabilities associated with discrete outcomes and for modeling decision-making processes under uncertainty.

The choice of a Logistic model also aligns with the objective of this research: to estimate how certain independent variables, such as funding type, number of employees, and sector characteristics, affect the probability of a startup reaching an EXIT. In this context, the Logistic approach allows us to model the latent propensity of a firm to EXIT, which is not directly observable but manifests as a binary outcome.

In line with standard practice in the econometric literature, the model includes both explanatory variables (e.g., firm size, grant funding, patents) and control variables (e.g., sector dummies, university affiliation), allowing for the isolation of the marginal effects of each factor while holding others constant.

Moreover, the model's statistical significance and goodness of fit were evaluated using standard metrics such as pseudo R-squared and likelihood-ratio tests, which confirmed the robustness of the selected specification.

3.2 Description of the dataset

The dataset used for this research originates from a proprietary database developed by Inveniam Group, a corporate finance advisory firm specializing in technology transfer and innovation, where I carried out my internship. The original database was created in collaboration with Fundación Botín and CaixaBank as part of a broader initiative aimed at mapping, monitoring, and supporting the life science startup ecosystem in Spain. This initiative has led to the development of the first edition of an annual research report titled "Evolution and Trends in Scientific Entrepreneurship in Spain's Health Sector – An Analysis of Companies Originating from the Public Reaserch System (2001-2023)", published in 2025, with a second edition scheduled for 2026.

The original dataset consisted of 178 Spanish life science startups from 2001 to 2023, but included only a limited number of variables focused primarily on basic company information and investment history. For the purposes of this thesis and in alignment with the research objectives, the dataset was significantly expanded and cleaned. The process involved the addition of numerous variables relevant to the regression analysis, to be more specifc the variables added are: “# of public foundings won”, “Total import collected”, “Total public grant”, “Total private grant”, “% Pubic funding”, “% Private funding”, “One asset(0)/ Multi asset (1)”, “# of IP/patents”, “Female (1)/ Male (0) CEO”, “# employees”, “Investor count”.

To enrich the dataset, a multi-source research strategy was employed. Publicly available information was collected from startup websites, press releases, and funding databases. Moreover, I was granted access to PitchBook, a market intelligence platform, which proved crucial for identifying detailed information on private funding rounds, including rounds not publicly disclosed elsewhere. For variables related to public grant funding, I used Invespolora, Inveniam’s internal tool for tracking public financing in the Spanish innovation ecosystem. This platform allowed me to retrieve historical data on grant applications, awarded amounts, and call details, providing a unique level of granularity in the analysis.

During the cleaning process, I encountered cases where information was missing or incomplete. Rather than imputing or estimating these values, which could introduce noise or bias into the model, I chose to exclude companies with insufficient or unreliable data. In addition to this, 23 companies were removed from the original sample because they were found to be no longer active, liquidated, or in the process of liquidation, as verified through public records and Inveniam's internal tracking tools.

As a result, the final dataset used for the regression model includes 112 companies. This reduction from the original sample of 178 startups was necessary to preserve analytical integrity and statistical consistency, even though it came at the cost of a smaller sample size. It is important to highlight that the companies included in the final dataset are those for which complete and up-to-date information was available for all selected variables, thus minimizing the risk of selection bias in the regression model.

In total, the final dataset includes a robust set of variables spanning funding metrics, company characteristics, and institutional factors. Prior to running the regression analysis, a multicollinearity check was performed. The initial set of explanatory variables was evaluated using Variance Inflation Factors (VIFs) and correlation matrices to identify any highly collinear pairs. Variables with strong correlations were either removed or transformed to reduce redundancy and preserve model stability. This step was crucial in ensuring the validity of the estimated coefficients and the reliability of the Logistic model results.

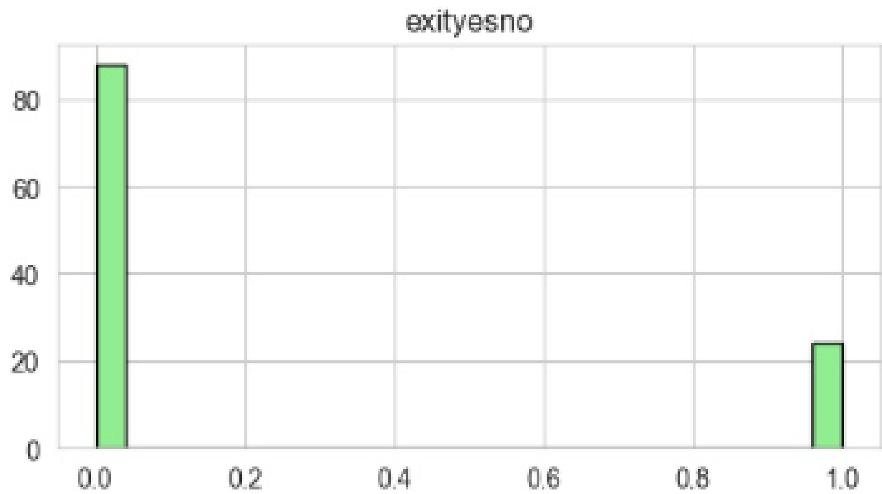
Although the final sample of 112 companies is sufficient for the scope of this research and allows for the identification of statistically significant relationships, it also represents a limitation of the study. A larger sample size would enable more granular segmentation and increase the generalizability of the findings. Nonetheless, the curated and cleaned dataset provides a valuable snapshot of the Spanish life science startup landscape and serves as a solid empirical basis for testing the research hypotheses.

3.3 Dependent variable

The dependent variable in this study is `exit_yes_no`, a binary indicator that captures whether a startup in the life sciences sector has achieved a successful exit. For the purpose of this research, EXIT is defined exclusively as a merger or acquisition, excluding other forms such as IPOs, liquidations, or company shutdowns. This choice ensures a consistent and specific definition of liquidity events that result in a direct change of ownership and financial realization for early-stage investors.

The variable is constructed as follows:

- A value of `1` indicates that the startup has been successfully acquired or merged.
- A value of `0` indicates that the startup is still active but has not yet exited as of the date of data collection.



Importantly, startups that were liquidated or in the process of liquidation were removed from the dataset altogether, ensuring that the analysis focuses solely on viable or exited ventures. In total, 23 companies were excluded for this reason during the cleaning phase.

Consequently, the final regression model is based on a sample of 112 companies, of which 29 have exited (merge or acquisition) and 83 remain active without an exit.

The status of each company was carefully verified through public records, internal databases (such as PitchBook), and Inveniam's proprietary tools. By narrowing the dependent variable to a clear, binary classification based solely on M&A outcomes, this study avoids ambiguity and enhances the interpretability of the regression results.

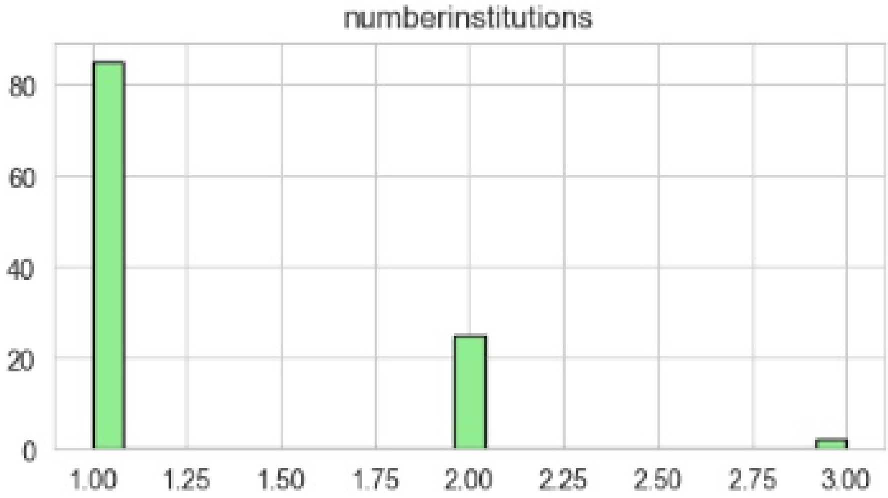
3.3 Independent variables

In the development of the regression model, a total of 16 independent variables were selected. These variables reflect a combination of financial, organizational, and innovation-related aspects of the startups included in the dataset. The selection was guided by both theoretical foundations from the literature and practical considerations related to data availability and model interpretability.

During the cleaning and preparation phases, attention was paid to avoid multicollinearity, which could compromise the reliability of the regression coefficients. Variables with excessive overlap or correlation were merged or removed. In particular, many variables initially encoded as one-hot columns (e.g., sector, region, period of incorporation) were recoded as single categorical variables, reducing redundancy and improving statistical power. Additionally, feature engineering was applied to construct more meaningful and predictive variables such as age, efficiency ratios, and funding structure indicators.

Each variable included in the model is illustrated in the graphs above, which show the distribution of values across the dataset. Below is a detailed description of each variable and the rationale behind its inclusion.

Number of institutions:



The variable `numberinstitutions` refers to the number of universities or research institutions involved in the founding of a startup, particularly those that originated as spin-offs from academic research centers. In the context of the Spanish life sciences ecosystem, it is common for startups to emerge from academic or publicly funded research environments, and in some cases, more than one institution may jointly support the creation of a company.

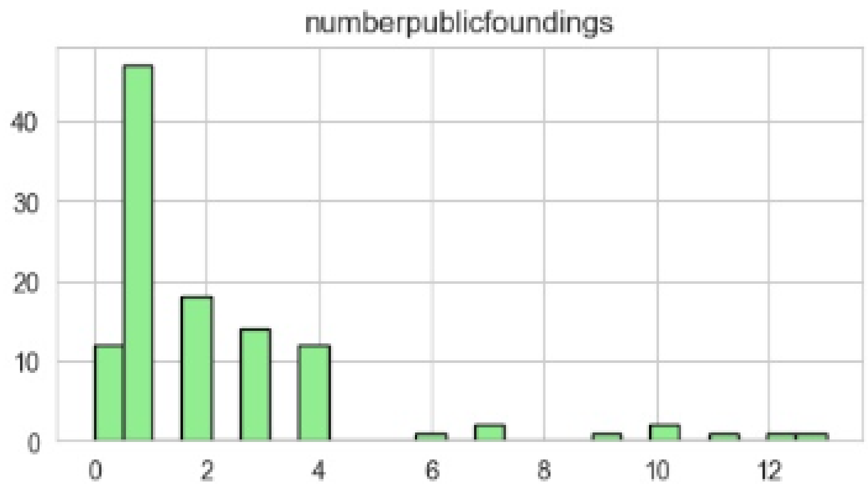
The histogram above reveals a clear right-skewed distribution. Most startups in the dataset (over 80 out of 112) are associated with a single university or research institution, while a smaller subset is linked to two institutions, and only a handful are born from collaboration among three institutions. This distribution reflects the typical organizational structure of spin-offs, which are often incubated within a single academic lab or university tech transfer office, although multi-institutional collaborations are increasingly encouraged.

Rationale for inclusion in the model:

The number of founding institutions serves as a proxy for the depth and breadth of academic support behind a startup. A higher number of institutions may indicate a more interdisciplinary founding team, broader access to research infrastructure, and potential for diverse intellectual property (IP) contributions. These factors can influence both the quality of the underlying technology and the ability to attract funding or acquisition interest, especially in the biotech and medtech sectors where scientific credibility is crucial.

Furthermore, startups with multiple institutional stakeholders may benefit from stronger governance structures, a more robust network of academic mentors, and greater visibility within innovation policy frameworks (e.g., regional grant schemes or university venture funds).

Number of public fundings:



The variable `numberpublicfoundings` quantifies the total number of public grants or funding programs secured by each startup over its lifetime. These grants may include competitive innovation programs, regional or national subsidies, or European initiatives such as Horizon 2020 or PERTE funds. In the context of the Spanish biotech and medtech sectors, where access to early-stage capital is often limited, public funding represents a crucial lifeline for R&D-intensive startups.

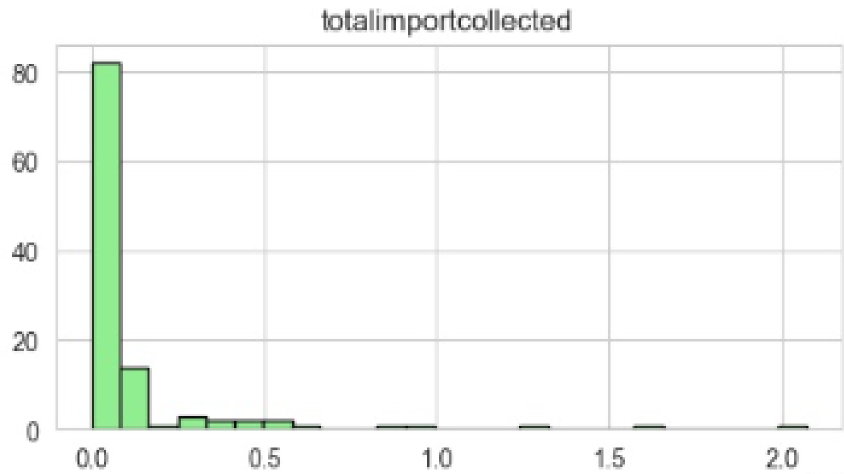
The distribution, as shown in the histogram above, is heavily right-skewed. A large proportion of companies (over 40 out of 112) have received only one public grant, with a significant number also receiving two to four grants. Beyond that, the frequency drops off considerably, with a small minority of firms winning more than five grants, and only a handful reaching ten or more. This indicates that while public funding is broadly accessible, sustained or repeated access is limited to a few startups, likely those with strong grant-writing capacity, compelling R&D, or strategic public-private partnerships.

Rationale for inclusion in the model:

The number of public grants won serves as a proxy for a startup's ability to secure competitive, non-dilutive funding, which in turn can reflect a high level of technological credibility and institutional trust. Startups that repeatedly win public funds are likely to have a well-articulated innovation pipeline and a strong track record in meeting project milestones and research objectives, traits that are attractive to acquirers or strategic investors.

Additionally, repeated access to public grants may help a startup extend its runway without equity dilution, allowing it to mature technologically and structurally before seeking private capital or exit opportunities. This, in turn, could increase the probability of a successful merger or acquisition, which is particularly relevant in capital-intensive fields like biotech.

Total import collected:



The variable `totalimportcollected` represents the total cumulative amount of financial resources raised by each startup over its lifetime. This includes both equity funding (from venture capital, angel investors, corporate investors) and non-dilutive sources (such as public grants).

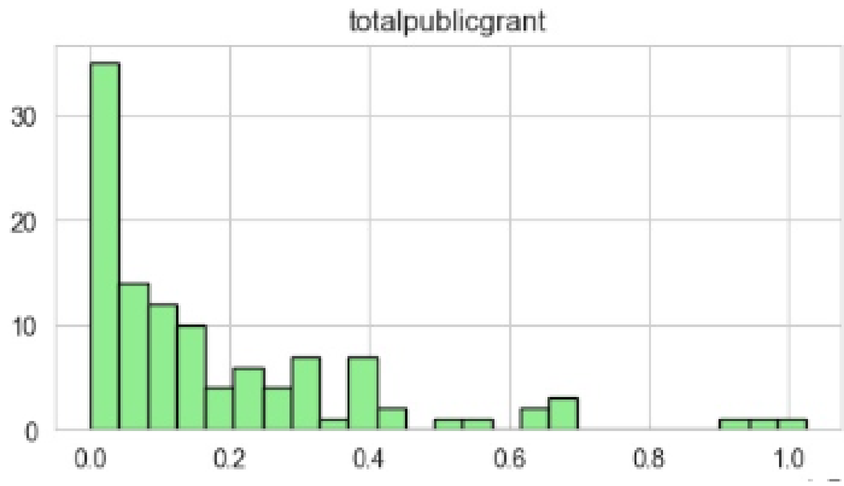
As observed in the histogram, the distribution is extremely right-skewed, with a majority of startups raising relatively small amounts (typically under €5 million), and a few outliers raising significantly higher figures (over €20 million). This power-law-like distribution is typical of early-stage markets like biotech and medtech, where capital is concentrated in a small number of high-potential companies, often backed by institutional investors.

Rationale for inclusion in the model:

The rationale behind including this variable is based on the assumption that the total capital raised is not just a proxy for startup size, but also for investor trust, scalability potential, and maturity. In the life sciences, especially, raising large sums of capital is crucial to support long and costly development cycles (e.g., clinical trials, regulatory certifications, IP protections). Therefore, a startup that has been able to raise substantial resources over time may be closer to commercial viability or technological validation, both of which increase its attractiveness to acquirers.

Furthermore, investors generally perform rigorous due diligence before allocating significant amounts of money, thus, a high value of `totalimportcollected` can indirectly reflect a startup's credibility, governance quality, and strategic roadmap. However, high funding alone is not always a positive indicator; it may also point to capital-intensive models that are not yet profitable. This duality makes the variable especially interesting from a statistical perspective, as it allows the regression model to quantify whether financial abundance alone correlates positively or negatively with eixt probability.

Total public grant:



The variable `totalpublicgrant` captures the total monetary value of public grants awarded to each startup throughout its operational life. It reflects the aggregate amount of non-dilutive public funding the startup has received, typically from national agencies, regional programs, EU funds, or innovation-based competitions. This figure is expressed in euros and was retrieved from Inveniam's proprietary database *Invespolora*, where detailed public funding histories (including calls submitted and amounts awarded) are archived.

The histogram shows a right-skewed distribution, with the vast majority of companies receiving grants below €200,000. A long tail extends toward €1 million, but only a few firms achieved such high levels of public financing. Notably, there is a more even spread in the mid-range compared to previous variables (like total import collected), indicating greater heterogeneity in public grant support among startups.

This visual distribution underlines a key characteristic of public grant allocation in the lifescience space: while many startups receive small seed-stage or innovation grants, only a few consistently secure large-scale public funds, often through consortia participation, European Horizon programs, or advanced-stage clinical subsidies.

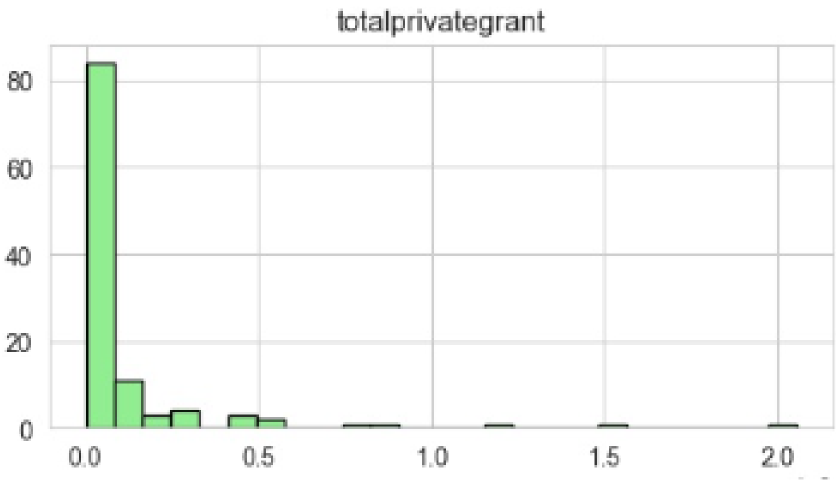
Rationale for inclusion in the model:

From a strategic standpoint, `totalpublicgrant` plays a multidimensional role in the regression model.

1. **Indicator of Institutional Support**: A higher grant total may signal that a startup is recognized by public institutions as having high potential or strategic relevance. Public funding bodies typically require in-depth evaluation of business plans and R&D milestones. Thus, this variable indirectly captures external validation and alignment with policy priorities (e.g., digital health, rare diseases, or medtech devices).
2. **Non-Dilutive Capital as Growth Driver**: Public funding allows startups to finance core research and development without diluting equity ownership. In industries like life sciences, where long product development timelines and regulatory hurdles exist, this can make a critical difference. Startups able to secure substantial public grants are better positioned to retain control, attract private investors later, or bootstrap their way to acquisition-readiness.
3. **Risk Mitigation for Investors**: High levels of public grants may also de-risk the company from an investor perspective, as it implies that certain technological validations or certifications are already funded or underway. This improves the likelihood of M&A interest.
4. **Potential Bias & Interpretation Limits**: However, the relationship is not always linear. Public grants often come with rigid milestones, spending limitations, and

may delay market responsiveness. Some well-funded companies may become public grant dependent without market traction. Therefore, this variable must be interpreted alongside efficiency indicators such as `funding_per_employee` and `funding_per_patent`.

Total private grant:



The variable `totalprivategrant` represents the total cumulative amount of private funding secured by the startup over its lifetime, expressed in euros. This includes equity investments, convertible notes, venture capital funding, and other forms of private financing not classified as public subsidies or grants. The data was retrieved primarily through Inveniam's internal Pitchbook terminal, allowing access to precise investment figures and funding round details that are often unavailable through public databases or startup websites.

The histogram displays a highly right-skewed distribution, with a significant concentration of companies clustered around the €0–€100,000 range. A long tail extends to amounts exceeding €2 million, though these are rare. This graph mirrors the typical funding structure of life science startups, where only a small subset manages to raise substantial private capital.

This distribution confirms that most companies operate in capital-constrained environments, possibly due to early-stage status, risk aversion of VCs in Spain, or insufficient traction. However, a few firms appear to have successfully secured multiple funding rounds or attracted institutional investors with larger tickets.

Rationale for inclusion in the model:

The strategic inclusion of `totalprivategrant` in the regression model is critical for several reasons:

1. **Proxy for Market Validation**

   Private capital requires passing rigorous due diligence processes. A higher total private funding figure reflects investor confidence, commercial potential, and perceived scalability of the startup's technology or product. In this sense, `totalprivategrant` serves as a proxy for private market validation, a crucial success driver in predicting exit outcomes.

2. **Fuel for Growth and Exit Readiness**

   Private funding often finances growth acceleration, including hiring, clinical trials, certifications, and internationalization, all key to making a startup attractive for acquisition. Therefore, this variable provides insight into whether the startup had sufficient resources to reach exit-relevant milestones.

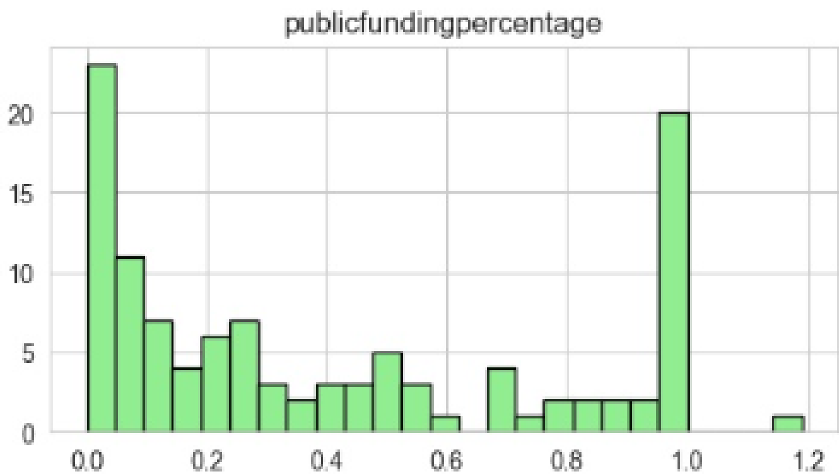
3. **Complement to Public Grant Data**

   When analyzed in conjunction with `totalpublicgrant`, this variable allows the model to explore funding mix dynamics. Some firms rely heavily on public funds (often tied to R&D), while others exhibit greater reliance on private venture ecosystems. This comparison supports richer interpretations on how different financing strategies affect the probability of M&A.

4. **Control for Size and Lifecycle**

   Funding levels also correlate with company maturity, and their inclusion helps control for life-cycle effects. While `company_age_years` captures temporal maturity, `totalprivategrant` captures financial maturity—an essential factor in signaling exit readiness.

Public funding percentage:



The variable `publicfundingpercentage` represents the share of public funding received by each startup relative to the total funds raised, meaning the sum of public grants and private investments.

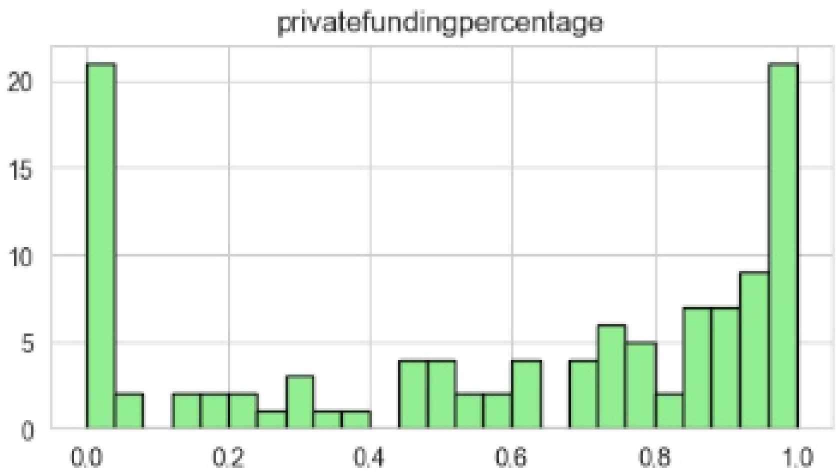
This metric offers a direct measure of the degree to which a startup relies on public funding, such as European programs (e.g., Horizon 2020, EIC), national grants (CDTI), or regional support schemes.

The graph shows a highly polarized distribution: on one side, many companies are concentrated around zero, indicating they received little to no public funding; on the other side, there is a spike around the value of 1, representing companies entirely funded by public grants with no private investment. This distribution clearly reveals two distinct strategic approaches within the Spanish startup ecosystem: startups that rely solely on public grants, often academic spin-offs, and others that are entirely funded through private capital or a mixed structure that leans heavily one way.

Rationale for inclusion in the model:

The inclusion of this variable in the model is driven by the need to understand how the funding structure impacts the likelihood of exit success, particularly in the biotech sector, where public capital often plays a key role in early-stage development. This metric captures strategic differences between startups that are highly dependent on public funds, typically still in early technological or preclinical phases, and those that have attracted private investors, signaling a greater level of market confidence in the company's commercial potential. Moreover, the public funding percentage is indicative of a startup's ability (or inability) to position itself on a market-oriented growth trajectory, and therefore closer to M&A scenarios. Given the nature of the sample, which includes a significant number of academic spin-offs, this variable serves as a bridge between the institutional origins of a company and its capacity to attract venture capital. Lastly, considering regional disparities in grant availability and distribution rates, this metric allows the model to control for potential policy-based distortions, offering insight into how local ecosystems may influence fundraising dynamics.

Private funding percentage:



The variable `privatefundingpercentage` represents the share of private investment received by each startup relative to the total amount of capital raised.

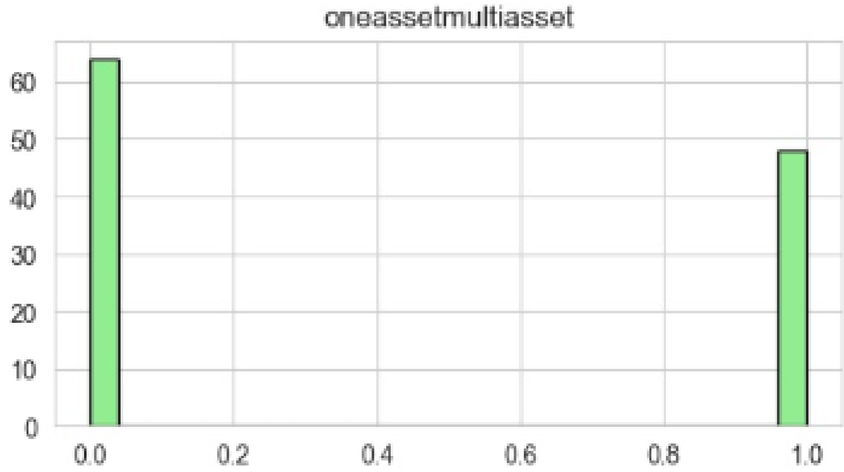
This metric complements its public counterpart and captures the degree of reliance on private investors, including venture capital funds, business angels, corporate VCs, or private equity players.

The distribution of this variable reveals a bimodal pattern, with peaks around 0 and 1. This indicates that many startups are either almost entirely publicly funded (close to 0) or almost exclusively privately funded (close to 1). A significant number of companies lie along the spectrum in between, indicating mixed funding strategies. The tailing spread suggests a wide variety of capital structure approaches across the sample.

Rationale for inclusion in the model:

The inclusion of `privatefundingpercentage` is essential to differentiate startups that successfully attracted private capital, a key milestone in scaling and commercial validation, from those that relied primarily on public mechanisms. From a strategic standpoint, private capital injection often reflects strong investor confidence, better governance practices, and a market-oriented business model. Moreover, in the biotech and medtech sectors, where development cycles are long and capital-intensive, access to private funding may be crucial for survival beyond the R&D phase. The presence of private investors is frequently associated with stronger networks, clearer exit planning, and readiness for M&A activity, especially in later stages. This variable, therefore, functions as a proxy for market traction and investor attractiveness, two key aspects likely to influence the probability of exit. It also serves to control for possible capital structure biases, especially when comparing university spin-offs to VC-backed startups. Combined with the public funding percentage, it enables the model to better understand how different financing compositions may affect a company's strategic trajectory and exit potential.

One asset multi asset:



The variable `oneassetmultiasset` is a binary indicator that identifies whether a company is developing a single asset (coded as 0) or multiple assets/pipelines simultaneously (coded as 1). It reflects the scope and diversification strategy of the startup's research and development efforts.

The distribution is split between two distinct bars at 0 and 1, confirming the binary nature of this variable. A slightly higher number of companies (about 60) focus on a single asset, while the remaining ~50 companies have diversified into multiple assets or therapeutic targets. This dichotomy highlights two prevailing strategic approaches in the life sciences startup ecosystem: deep focus on one high-potential product versus portfolio diversification.

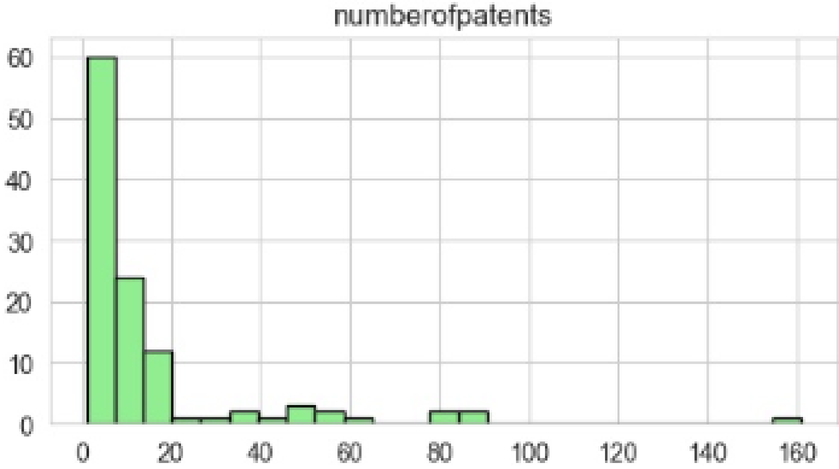
Rationale for inclusion in the model:

The inclusion of `oneassetmultiasset` in the regression is rooted in its ability to capture strategic differences in risk management and scalability potential. Startups with a single-asset focus often adopt a “go big or go home” strategy, concentrating all efforts and funding on the success of one pipeline. While this can lead to high returns in case of breakthrough success, it also implies higher concentration risk, making them more vulnerable to clinical failure, regulatory hurdles, or market rejection.

On the other hand, startups with multiple assets tend to follow a portfolio approach, which mitigates risk by diversifying across different therapeutic areas or technologies. These companies might appeal more to investors or acquirers due to greater optionality, potential for platform technologies, or simply a wider range of revenue streams. From an M&A perspective, a diversified pipeline may signal strategic flexibility and longer-term value, making the firm a more attractive acquisition target.

Therefore, this variable adds depth to the model by controlling for R&D scope and strategic orientation, which are critical dimensions in determining a company’s likelihood to exit. It also helps differentiate between capital allocation strategies and their outcomes in the competitive and resource-constrained environment of biotech innovation.

Number of patents:



The variable `numberofpatents` represents the total number of patents filed by each company in the dataset. It is a numerical and discrete variable, directly reflecting the level of intellectual property (IP) development, which is especially critical in the life sciences and biotech sectors.

The histogram displays a right-skewed distribution, where the majority of companies hold between 0 and 20 patents, with a strong peak in the 0–10 range. A smaller number of companies hold larger patent portfolios, reaching up to more than 160 patents, though these cases are rare and should be considered outliers. This long-tail distribution is typical in R&D-intensive sectors where a few highly capitalized firms or platform-based startups accumulate a large volume of IP, while the majority operate with smaller, focused patent portfolios.

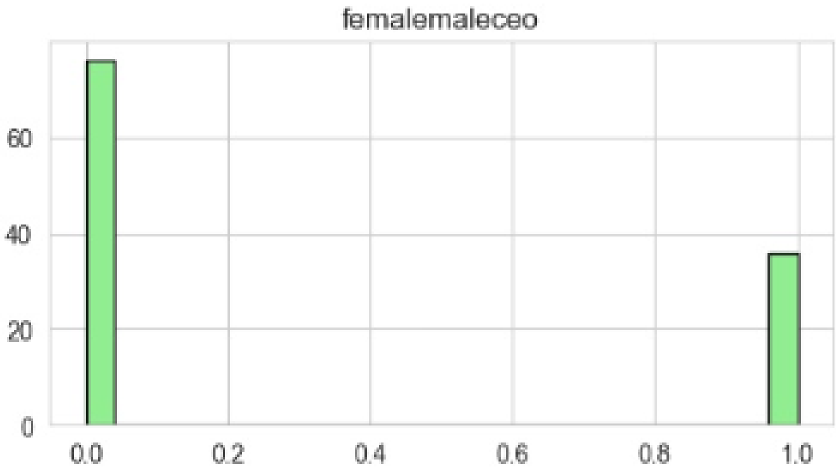
Rationale for inclusion in the model:

In biotechnology and medical technology sectors, intellectual property is a core asset. Patents provide a legal framework for protecting innovations, securing market exclusivity, and enhancing a startup's valuation. The number of patents is not just a count; it is a proxy for technological maturity, innovation capability, and competitive advantage.

From a regression modeling standpoint, including this variable enables the analysis to capture the extent of R&D output and its correlation with a company's ability to exit. Companies with a larger number of patents may have stronger licensing opportunities, acquisition appeal, or the potential for commercial success, making them more likely to achieve a successful exit via acquisition. Moreover, patents are often a key metric that institutional investors and pharmaceutical acquirers evaluate when assessing a target's potential.

It is also worth noting that while the raw number is important, in isolation it may not reflect efficiency or quality. That is why complementary variables such as `patents_per_employee` are also introduced in the model to capture relative innovation efficiency, but the absolute patent count remains a fundamental indicator of technological substance.

Female male ceo:



The variable `femalemalceo` is a binary categorical variable indicating the gender of the Chief Executive Officer (CEO) of each startup in the dataset. A value of `1` corresponds to a female CEO, while `0` indicates a male CEO.

The distribution shown in the histogram is highly asymmetrical, reflecting the current gender imbalance in leadership positions within the Spanish biotech and medtech startup ecosystem. The majority of companies, over 60 out of the 112, are led by male CEOs, with fewer than 50 companies having a female CEO. While this variable does not follow a continuous or even distribution, its inclusion is still meaningful for categorical comparative analysis within the regression framework.

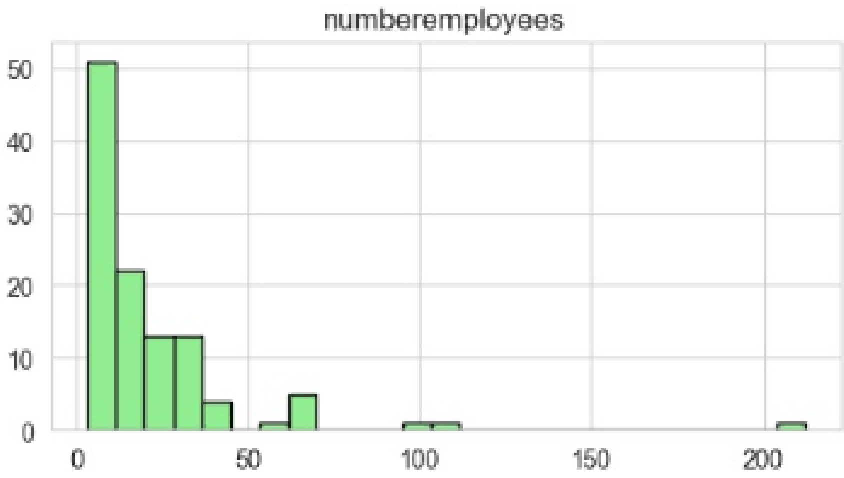
Rationale for inclusion in the model:

The decision to include this variable was driven by growing interest, both academic and institutional, in the impact of gender diversity on startup outcomes. Multiple studies suggest that leadership diversity can influence company culture, investor perception, risk-taking behavior, and overall strategic direction. In sectors like biotechnology, where innovation and public-private collaboration are pivotal, understanding whether the gender of the CEO correlates with the likelihood of achieving an exit can provide nuanced insights into systemic dynamics.

From a modeling standpoint, `femalemalceo` helps test whether non-financial and non-technical characteristics (such as leadership profile) are significantly associated with exit probability. Although the model is primarily grounded in quantitative and structural metrics, including this variable opens the door for exploratory insights on gender gaps in access to capital, market exposure, and exit opportunities.

It's important to acknowledge that gender, in itself, does not determine competence or outcome. However, disparities in success rates, if found, could highlight broader structural inequalities or differing strategic pathways between male- and female-led ventures. This variable, therefore, holds value both analytically and socially, contributing to the inclusivity of the regression model and reinforcing the comprehensive scope of the analysis.

Number employees:



The variable `numberemployees` captures the total number of people employed by each startup at the time of data collection. It is a continuous numerical variable and serves as a proxy for the operational scale and organizational complexity of the firm.

The histogram shows a positively skewed distribution, with the majority of startups employing fewer than 30 individuals. A few outliers extend up to over 200 employees, indicating the presence of a small number of significantly larger firms in the sample. This type of long-tail distribution is common in startup datasets, where a large number of small or early-stage companies coexist with a handful of more mature or better-funded ventures.

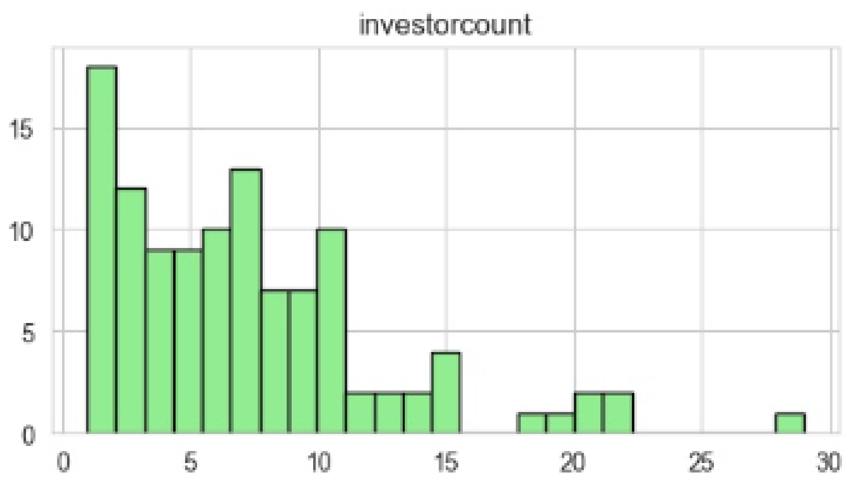
Rationale for inclusion in the model:

The number of employees was included in the model as a key structural indicator. In the context of biotech and medtech startups, team size is often reflective of the stage of development, capital intensity, and research pipeline breadth. While a small team may suggest lean operations and agility, a larger workforce can imply either a more advanced stage of product development (e.g., entering clinical trials) or access to greater financial resources.

From a statistical standpoint, including this variable helps to control for company size when interpreting the effects of other variables such as funding, number of patents, or investor count. Furthermore, it enables the creation of efficiency metrics (e.g., funding per employee, patents per employee), which add deeper interpretability to the model.

However, team size alone does not guarantee success; in fact, overly large teams may introduce complexity and inefficiency. Therefore, rather than using it in isolation, its primary value lies in how it interacts with other performance indicators, serving as a baseline for understanding organizational productivity, capital deployment, and innovation output.

Investor count:



The variable `investorcount` represents the total number of distinct investors that have participated in the funding rounds of a startup during the period analyzed. This includes both public and private investors, such as venture capital firms, angel investors, institutional funds, and public entities involved in innovation support.

The histogram shows a positively skewed distribution, with the majority of startups having between 1 and 10 investors. The mode appears to be around 2 to 3 investors, while only a few companies have more than 15. There are rare but notable outliers with up to 30 investors, reflecting either particularly attractive investment opportunities or longer fundraising histories with multiple rounds.

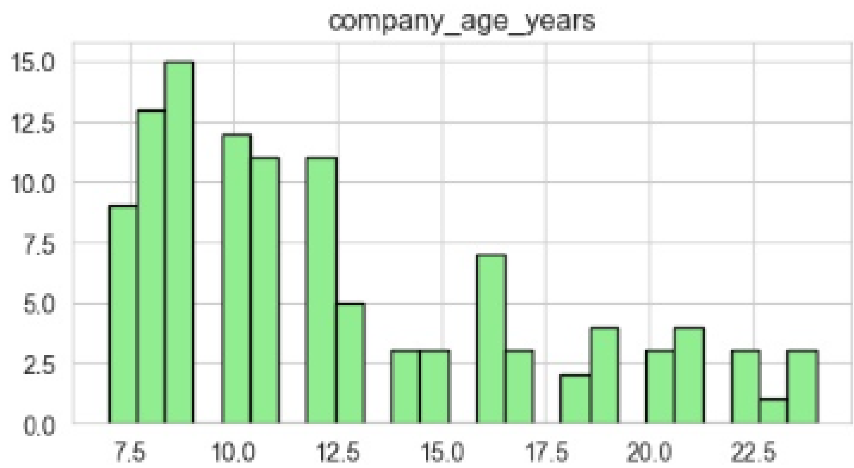
Rationale for inclusion in the model:

Investor count serves as a proxy for market confidence and external validation. A higher number of investors may indicate that the startup has been successful in generating widespread interest and securing funding from multiple sources, which could be interpreted as a signal of credibility, scalability, and perceived potential. It also suggests that the startup has been through multiple financing rounds, surviving the scrutiny of various due diligence processes.

From a modeling perspective, this variable provides insight into the breadth of financial support a company has attracted. Unlike the total amount raised, which can be skewed by a few large checks, the number of investors reflects network density and investor diversification. A broader investor base might reduce financial risk and enhance the company’s access to resources, mentorship, and follow-on capital, all of which are factors that can increase the likelihood of achieving an exit.

Moreover, including `investorcount` helps control for fundraising dynamics beyond just monetary value. It captures the ecosystem engagement level of a startup and could serve as a strategic indicator of investor interest across the lifecycle of the company.

Company age years:



The variable `company_age_years` measures the number of years between the startup's official date of incorporation and the reference year of the analysis. This numerical transformation was derived from the original incorporation year field, allowing for a more interpretable and analytically friendly format. The decision to use age rather than the founding year itself stems from the need to treat time as a continuous numeric variable that has a potentially linear or monotonic relationship with the likelihood of an exit event.

The distribution of this variable appears to be moderately right-skewed, with the highest concentration of companies aged between 8 and 12 years, and a steady decline in frequency as age increases. Very few companies in the sample exceed 20 years of age, while the minimum threshold was set to include companies with at least 7 full years of operating history. This decision ensures a more realistic probability of observing an exit event, as very young startups may not yet have reached the critical development milestones necessary to be considered for acquisition.

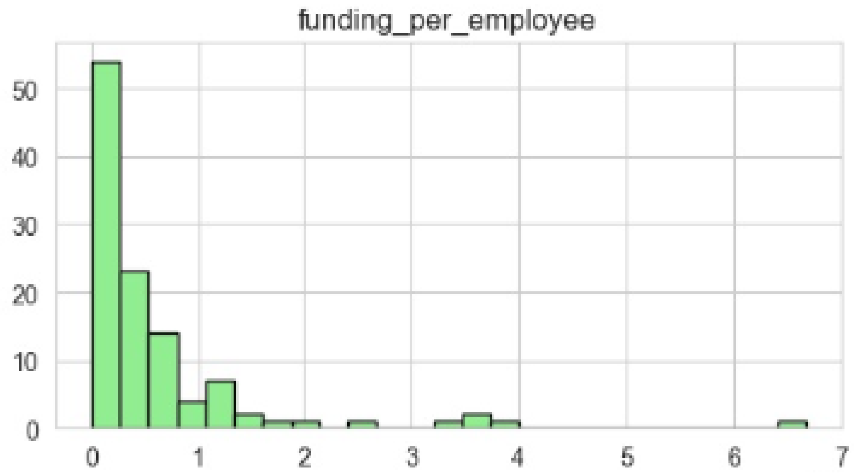
Rationale for inclusion in the model:

Company age acts as a proxy for maturity, experience, and market readiness. A startup that has been active for a longer period is more likely to have undergone key development stages, such as preclinical and clinical trials, product-market fit validation, regulatory clearances, or business model refinement. Especially in the life sciences sector, where development cycles are long and exits often occur after years of R&D and partnership building, age is a critical indicator of proximity to value realization.

Incorporating company age into the model also improves its interpretability. It allows for clear and actionable statements like: "for every additional year of company maturity, the probability of achieving an exit increases (or decreases) by X percent." This aligns well with strategic decision-making for investors and ecosystem players looking to time their involvement based on lifecycle stage.

Furthermore, company age helps control for temporal bias: startups founded more recently may not have had the same exposure time to reach an exit, whereas older companies that haven't exited may indicate stagnation or resilience depending on the context. Therefore, this variable adds depth to the model by accounting for survivorship and temporal exposure, both essential components in analyzing exit probabilities.

Funding per employee:



The variable `funding_per_employee` is a derived metric calculated by dividing the total capital raised by a company by its number of employees. This transformation was created as part of the feature engineering process to provide a normalized view of financial input relative to company size. Rather than relying on absolute funding figures or employee count alone, this ratio allows for a clearer assessment of how capital-intensive or capital-efficient a startup is with respect to its human resources.

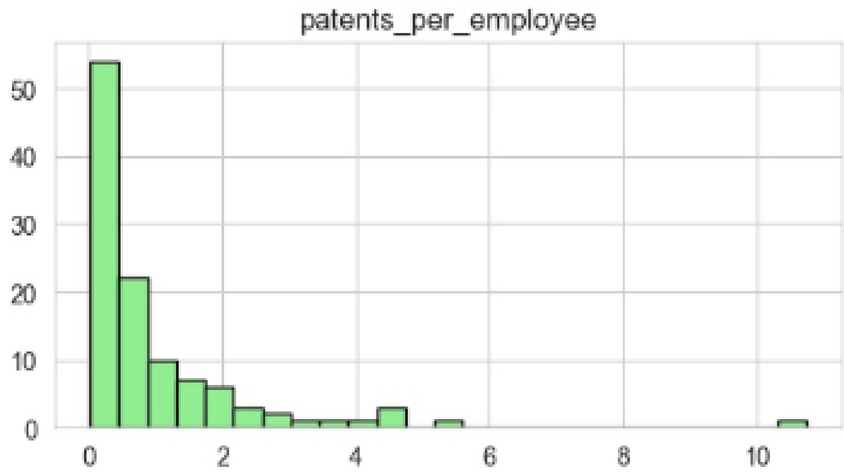
The histogram of `funding_per_employee` shows a heavily right-skewed distribution, with most companies clustered at lower values, suggesting modest capital per staff member. A small number of firms display extreme values, indicating either high funding with lean teams or highly capitalized business models. This long tail is characteristic of sectors like biotech, where some startups require substantial financial backing for early-stage research or clinical trials, yet maintain small, specialized teams.

Rationale for inclusion in the model:

This variable captures the financial efficiency and strategic posture of a company. A higher `funding_per_employee` ratio may signal that the company has attracted significant investor confidence relative to its organizational size. This could be due to the uniqueness of its intellectual property, the capital-intensive nature of its R&D, or the high scalability potential perceived by investors. These are all factors that can increase the probability of a successful exit, especially in industries where large acquisitions are often justified by the strength of a company's innovation pipeline rather than its revenue or headcount.

Moreover, by incorporating this ratio into the regression model, we control for scale and ensure that predictions are not biased in favor of larger companies that may raise more money simply due to their size. It allows the model to better distinguish between firms that are simply "big" and those that are "efficient" or "attractive" in terms of capital use. This is particularly useful when comparing startups at different stages of growth or with varying business models.

Patents per employee:



The variable `patents_per_employee` is a ratio derived by dividing the total number of patents filed by a company by the number of its employees. This metric serves as a proxy for innovation intensity, quantifying how efficiently a firm is translating its human capital into protectable intellectual property (IP). In the biotech and medtech sectors, where IP is often the primary driver of valuation and acquisition interest, this indicator holds strong explanatory potential for a startup's exit likelihood.

The histogram presents a highly right-skewed distribution, with a large majority of firms exhibiting low values of patents per employee, and a long tail of companies with high patenting efficiency. Most startups have a ratio close to zero, indicating either early-stage development, reliance on trade secrets over patents, or simply lower R&D intensity. However, the presence of outliers with extremely high values suggests a subset of companies that are either lean and hyper-innovative or that have structured their teams around IP development rather than operations or commercialization.

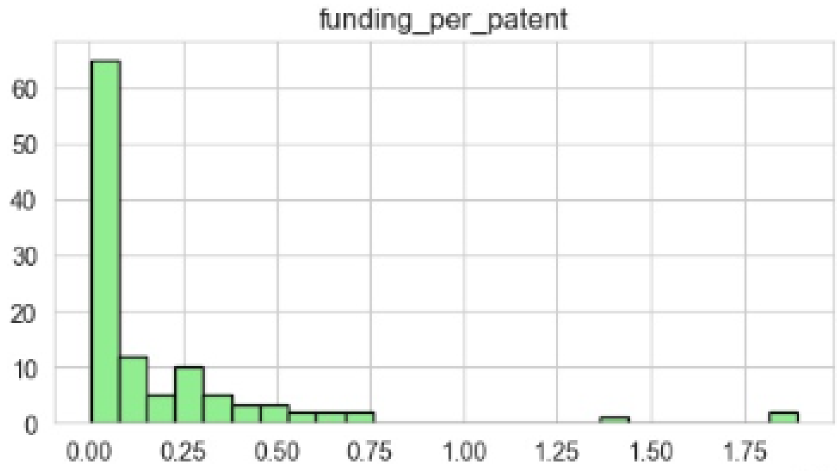
Rationale for inclusion in the model:

From a strategic standpoint, the `patents_per_employee` ratio provides a normalized innovation signal that complements the raw number of patents. While absolute patent counts can correlate with size or funding, this ratio corrects for scale and highlights how concentrated a company's innovation effort is relative to its workforce. A high value might indicate a strong research focus, specialization in patentable therapeutics or platforms, or a technology-centric team, traits that are particularly attractive to acquirers in life sciences.

Moreover, this variable indirectly captures strategic resource allocation. Companies that file many patents with few employees often demonstrate operational focus, IP-oriented leadership, and potentially a faster path to monetization through licensing, partnerships, or acquisition. In contrast, low ratios might suggest resource spread across non-R&D activities or early-stage product development with limited protectable output.

Including this variable enhances the model's ability to identify companies with deep technological moats and scalable IP portfolios, which are crucial determinants in the M&A-driven exit landscape typical of biotech ventures.

Funding per patent:



The variable `funding_per_patent` is calculated by dividing the total amount of funding received by a company by the number of patents it holds. It serves as a proxy for understanding how much capital is required for a company to generate a single unit of protectable intellectual property. In other words, it provides an inverse measure of R&D efficiency, the lower the ratio, the more patents a firm produces per euro invested.

The histogram shows a highly skewed distribution, with the majority of firms clustered toward the lower end of the scale. Most companies have a funding-per-patent ratio well below €500,000, indicating efficient use of capital in producing patents. However, a few companies stand out with values exceeding €1 million per patent, highlighting firms that either operate in capital-intensive research areas or exhibit lower IP productivity relative to the funds raised.

Rationale for inclusion in the model:

The inclusion of this variable aims to capture a nuanced aspect of startup performance: the capital intensity of innovation. While the total number of patents and total funding are both meaningful on their own, this ratio contextualizes them, emphasizing how effectively financial resources are converted into proprietary assets. In the biotech and medtech landscape, where innovation is key to exit value, being able to generate patents efficiently is often associated with superior technology management, lean R&D processes, or robust early-stage discoveries.

From an investor or acquirer's standpoint, a company that produces valuable IP with relatively limited capital is seen as a more scalable and capital-efficient bet, potentially commanding a higher valuation or becoming a more attractive acquisition target. Conversely, companies with high funding-per-patent ratios might face scrutiny over burn rate, execution capability, or R&D pipeline quality.

Moreover, by using this derived ratio instead of raw inputs, the model mitigates issues of multicollinearity between total funding and number of patents, two variables that would otherwise strongly correlate. This transformation helps isolate relative efficiency from absolute size, enhancing the interpretability and robustness of the regression model.

3.4 Methodology

The objective of this analysis is to develop a logistic regression model to estimate the probability that a startup in the Spanish biotech/medtech sector will achieve an EXIT, defined as either a merger or acquisition. Given the binary nature of the dependent variable (`exit_yes_no`, where 1 = Exit and 0 = No Exit), the Logistic model was selected as the most suitable econometric specification for this research as it is explained in the above section.

3.4.1 Data Loading and Setup

The analysis began with the loading and inspection of the cleaned dataset. Using Python and the appropriate libraries, the file was imported, and the structure of the data was examined to ensure that variable types, formatting, and column names were consistent and compatible with further operations.

3.4.2 Cleaning and Pre-processing

The original dataset, while rich in descriptive information, required significant transformation. Several categorical variables, such as sector, geographic zone, and incorporation year, were initially formatted using one-hot encoding (i.e., one column per category). This structure, while common in some machine learning contexts, was inefficient for a regression model and introduced unnecessary multicollinearity. As a result, these columns were reconverted into single categorical features. Additionally, the order of the columns was reorganized to improve readability and usability during analysis: identifying details were moved to the beginning of the dataset, while the target variable (`exit_yes_no`) was placed at the end.

3.4.3 Feature Engineering

A critical step in the modeling process involved the creation of new explanatory variables from the raw data. Rather than relying solely on absolute figures, new features were engineered to capture more meaningful and comparative metrics.

One of the most important transformations was the creation of the variable `company_age_years`, calculated by subtracting the year of incorporation from the current year (2025). This numeric variable acts as a proxy for company maturity, which is often correlated with increased stability, experience, and attractiveness to investors. Age proved to be a more powerful and interpretable predictor than the raw incorporation date.

In addition, several efficiency ratios were created to reflect the relative performance of each company. The variable `funding_per_employee` measures how much capital a company was able to raise per staff member, while `patents_per_employee` captures innovation productivity relative to team size. Lastly, `funding_per_patent` estimates how much financial investment was needed, on average, to produce one patent. These ratios provide insight into the company's strategic model, resource allocation, and investor confidence, factors which may all impact the likelihood of a successful exit.

3.4.4 Exploratory Data Analysis (EDA)

Before building the model, a detailed exploratory data analysis was conducted. This included visualizing the distribution of key numerical variables through histograms, as well as generating a correlation heatmap. The goal was to uncover any patterns, identify outliers, and most importantly, detect potential issues related to multicollinearity. Strong correlations, especially those above 0.8, between explanatory variables were carefully evaluated. In cases where variables were found to be redundant, only one representative feature was retained to avoid distortion in the regression estimates.

3.4.5 Model Preparation

The dataset was then prepared for regression modeling. The dependent variable `exit_yes_no` was verified to ensure correct encoding as a binary outcome. Categorical variables such as sector, region, and institutional backing were transformed into dummy variables as required by the Logistic model. Continuous variables were normalized when necessary. A final check was performed to ensure there were no missing values in the dataset. After this process, 112 companies remained, all with complete and usable data.

3.4.6 Modeling and Diagnostics

With the cleaned and structured dataset, the Logistic regression model was estimated. The model was evaluated based on the statistical significance of coefficients, measured using p-values, and on overall model fit, assessed through pseudo R^2 values and likelihood ratio tests. To improve interpretability, marginal effects were calculated post-estimation, providing insights into how changes in each variable affected the probability of an exit.

Throughout the modeling phase, multicollinearity was continuously monitored. High-correlation variable pairs were avoided in the same model specification. In particular, special attention was given to the ratios derived in the feature engineering phase, ensuring that they were not used in conjunction with their underlying base variables (e.g., avoiding use of both `funding_per_employee` and `total_funding` in the same regression). This helped preserve the stability and validity of the estimated coefficients.

In summary, the methodology applied in this analysis follows a rigorous data science workflow, from raw data acquisition and transformation to modeling and interpretation, ensuring that the results produced are both statistically robust and meaningful in the context of predicting startup exit success.

3.5 Methodological limitations

While the methodology adopted in this study was rigorous and systematic, certain limitations inevitably affect the scope and generalizability of the findings. First, the sample size, although suitable for an exploratory regression model (112 companies), remains relatively small for a high-dimensional analysis, especially in a heterogeneous sector such as life sciences. As such, statistical power may be limited, and some variables that are theoretically relevant may not emerge as significant simply due to lack of data. Additionally, the dataset was constructed manually by integrating multiple sources, including Inveniam's internal PitchBook terminal and the proprietary Invespolora grant tracking tool, introducing potential inconsistencies in how variables were defined or measured across companies.

Another key limitation stems from data availability and reporting bias. Some variables, such as the amount of public funding received or the number of private funding rounds, were not publicly disclosed for all firms. Rather than imputing missing values, a conservative approach was adopted by excluding incomplete entries. While this helped minimize bias within the final sample, it may have led to a selection effect, where only the most transparent or better-performing companies are retained, possibly inflating the predictive power of the model.

Moreover, the cross-sectional nature of the dataset does not allow for dynamic analysis over time. Temporal variation, such as changing market conditions or funding climates, is not captured, limiting the model's ability to generalize to other time periods or geographies. Finally, the use of a Logistic regression, while suitable for binary outcomes, assumes a specific distribution of the error term and does not account for potential nonlinearities or interaction effects between variables, which may exist in complex startup ecosystems.

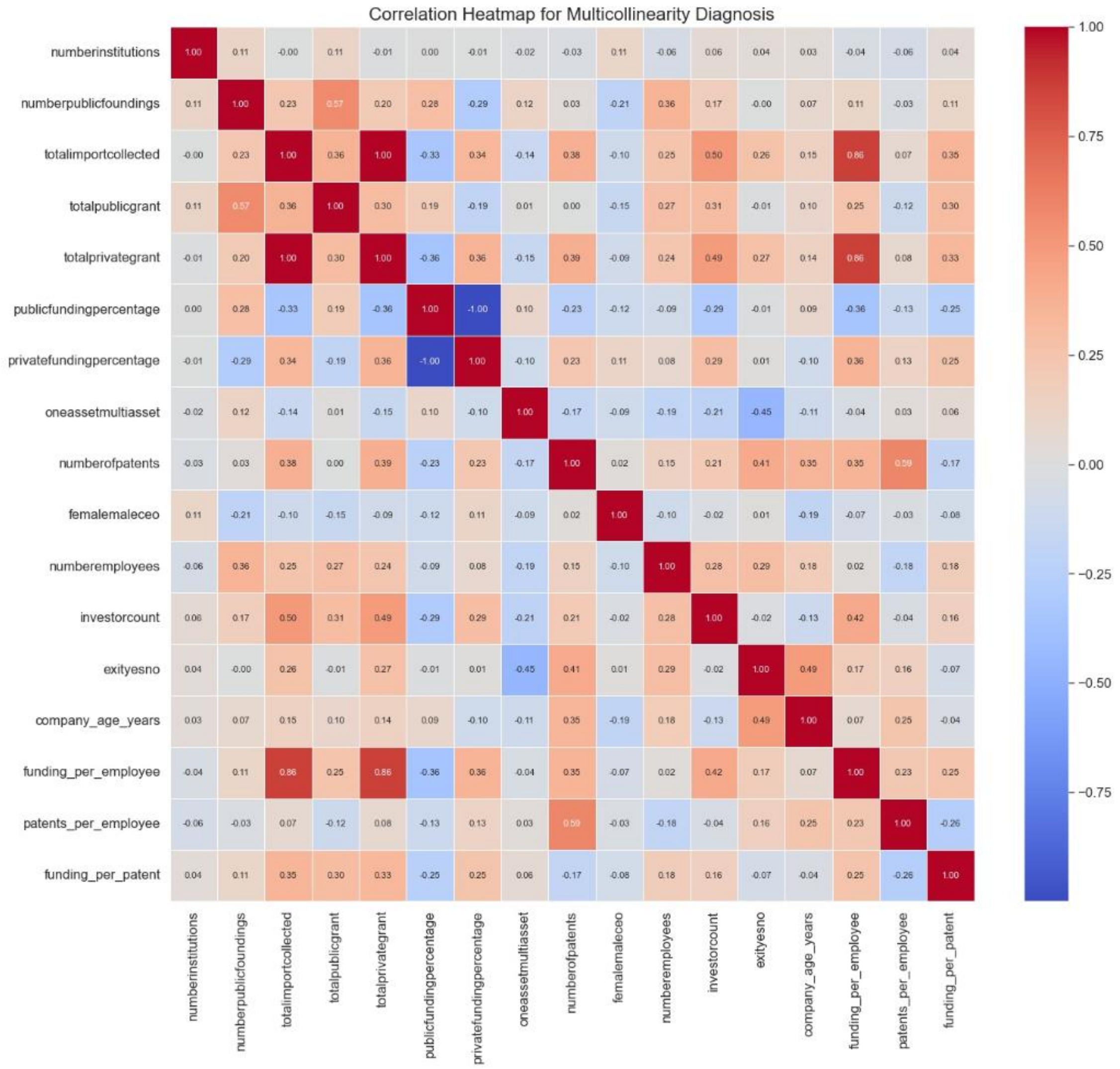
Despite these limitations, the model offers meaningful insights into the key factors associated with exit success in the Spanish biotech and medtech startup environment, serving as a useful foundation for future research with expanded datasets and more advanced techniques.

Chapeter IV: Discussion of Results

4.1 Correlation analysis among independent variables

Before proceeding with the econometric modeling, it is essential to assess the correlation between the independent variables included in the study. This preliminary step allows us to identify potential multicollinearity issues, which could distort the interpretation of the regression results and inflate the standard errors of the estimated coefficients. By analyzing the pairwise relationships among the variables, we aim to ensure that the data structure supports a stable and reliable estimation process. This correlation analysis provides a foundational understanding of the interdependencies among key predictors, such as funding metrics, company demographics, and sectoral classifications, which will subsequently inform the interpretation of our logit regression model and machine learning analysis.

4.1.1 Correlation Heatmap



The correlation heatmap presented above serves as one of the most powerful diagnostic tools for identifying multicollinearity within the dataset. The intensity of the colors, deep red for strong positive correlations and dark blue for strong negative correlations, visually highlights the most relevant relationships between variables. A key insight from the heatmap is the confirmation of multicollinearity among specific variables. For instance, there is an almost perfect positive correlation between total import collected and total private grant, as well as a nearly perfect negative correlation between public funding percentage and private funding percentage. This provides irrefutable evidence that these variables are redundant; including them simultaneously in a regression model would significantly distort the estimation, leading to unreliable results. As a result, the decision to exclude *total public grant*, *total private grant*, and *private funding percentage* from the final regression model is both necessary and statistically justified.

In addition, the heatmap validates the inclusion of the newly introduced variable *company_age_years*, which shows a strong positive correlation of +0.49 with the dependent variable *exit_ye_sno*. This suggests that firm maturity plays a critical role in predicting the likelihood of an exit event. Older firms are, unsurprisingly, more likely to achieve a successful exit. Furthermore, *company_age_years* also shows moderate positive correlations with both *investor_count* (+0.28) and *number_of_employees* (+0.19), reinforcing the logical assumption that older firms have had more time to scale and attract investment.

Finally, by reviewing the row and column corresponding to *exit_yes_no*, we can identify the most promising predictors for our regression model. The strongest positive correlation is observed with *investor_count* (+0.50), confirming that a higher number of investors is the most powerful indicator of a potential exit. Close behind is *company_age_years* (+0.49), followed by *number_of_patents* (+0.41), highlighting the importance of intellectual property. On the other hand, *oneassetmultiasset* exhibits the strongest negative correlation (-0.45), suggesting that companies focused on a single pipeline may be more strategically positioned for acquisition.

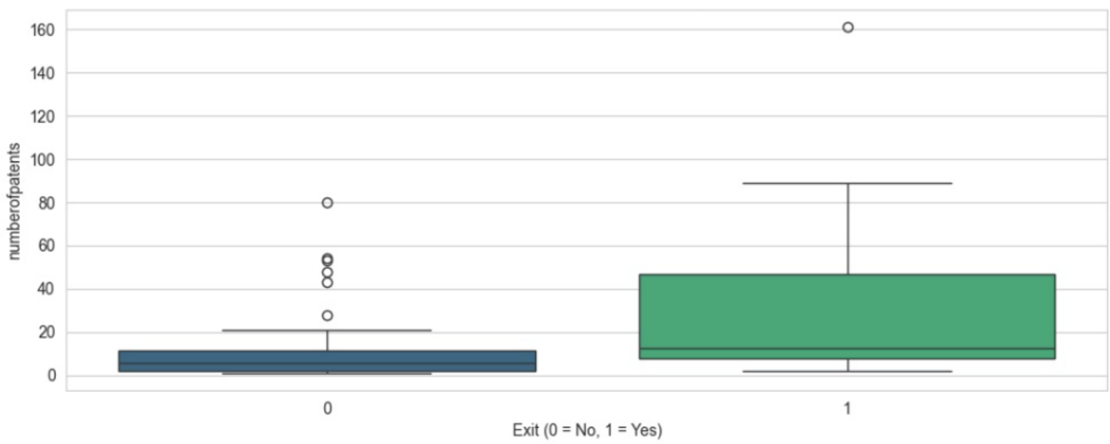
4.1.2 Direct Comparison Between Exit and No Exit Groups

After analyzing the overall distributions and conducting the correlation analysis, the next logical step was to directly compare the characteristics of companies that experienced an "exit" (exityesno = 1) with those that did not (exityesno = 0). To achieve this, we employed box plots, which are particularly effective in visually comparing the distribution of a continuous variable across different categories.

Each box plot presents the median (the central line), the interquartile range (the box itself), and any potential outliers (the dots beyond the whiskers), offering a clear and intuitive overview of how each variable differs between the two groups. The primary goal of this visual approach is to identify whether the distributions, and particularly the medians, differ significantly between companies that exited and those that did not. A consistent upward or downward shift of the "Exit" group compared to the "No Exit" group would suggest strong predictive value.

The results reveal several striking differences:

Focused Analysis: Number of Patents and Exit Outcome



The boxplot comparing the number of patents between companies that achieved an exit (`exityesno = 1`) and those that did not (`exityesno = 0`) reveals one of the clearest and most significant differences observed in the dataset.

The median number of patents in the “Exit” group is substantially higher than in the “No Exit” group. Visually, the interquartile range (IQR) for companies that exited lies almost entirely above the upper quartile of the non-exit group. This suggests that not only are median values higher, but the overall distribution of patents shifts upward for companies that successfully exited.

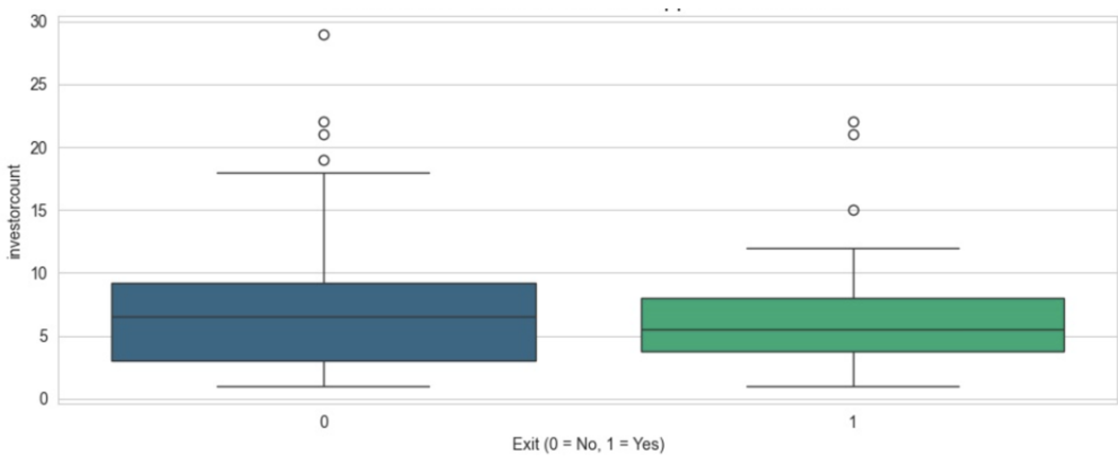
Moreover, the spread of the data is wider in the Exit group, indicating greater variability in the number of patents among successful companies. This is logical, as some of the most successful exits might be driven by breakthrough innovations protected by a large patent portfolio, while other companies may reach exit with a moderate but strategically valuable set of patents.

The presence of outliers is also noteworthy. While both groups display outliers, the Exit group includes one extreme case exceeding 160 patents, suggesting the presence of innovation-driven giants whose IP assets could have heavily influenced their exit outcomes. In contrast, the non-exit group, despite having a few companies with up to 80 patents, shows a much denser concentration near the lower end of the distribution.

This analysis strongly supports the hypothesis that intellectual property volume, measured through patent count, is a key predictor of exit success. The visual evidence aligns with the correlation results from the previous section (+0.41 correlation with exit), reinforcing the strategic importance of innovation and IP development in the biotech startup ecosystem.

In conclusion, the number of patents is not only statistically correlated with exit but also visually shows a robust, systematic difference in distribution, making it one of the most powerful predictors in our model.

Focused Analysis: Number of Investors and Exit Outcome



At first glance, the boxplot comparing the number of investors between companies that achieved an exit (`exityesno = 1`) and those that did not (`exityesno = 0`) shows a less striking difference than the one observed for the number of patents. However, the differences are still meaningful and warrant attention.

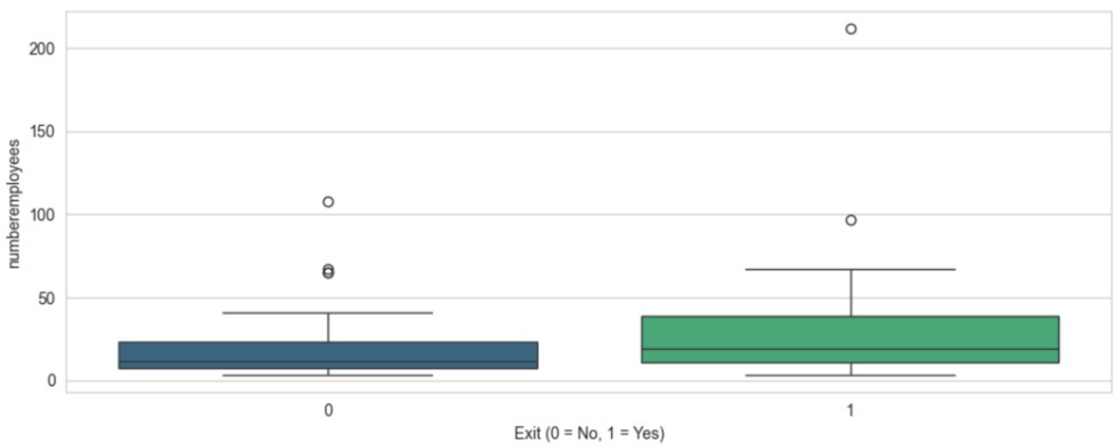
Interestingly, while both groups have a similar median number of investors, around 6, the distribution shapes differ. Companies that did not exit show a broader spread of investor counts, with several outliers exceeding 20 or even 28 investors. Conversely, the Exit group displays a more compact distribution, with fewer extreme cases but a denser cluster of companies around the median.

This may suggest that while having many investors is not necessarily a guarantee of exit, maintaining a solid and consistent investor base might be a better indicator of success. The Exit group’s tighter IQR and more symmetric distribution around the median could indicate more strategic or higher-quality funding rounds, compared to the more erratic spread observed in non-exit firms.

From a statistical standpoint, even though the visual difference is subtle, the correlation matrix previously discussed shows a strong positive correlation between investor count and exit success (+0.50), making this variable one of the most promising predictors in the entire dataset. The boxplot complements this insight by illustrating that companies achieving exit tend to maintain stable investor profiles and are less dependent on massive, scattered investor pools.

In conclusion, the number of investors alone does not drastically separate exit from non-exit firms in terms of distribution shape, but when combined with correlation metrics and contextual interpretation, it remains a robust and reliable predictor of exit likelihood, especially when aligned with strategic funding patterns.

Focused Analysis: Number of Employees and Exit Outcome



The third boxplot compares the number of employees between companies that exited and those that didn’t. While not as visually dramatic as the analysis for patents, this chart still reveals a notable shift in distribution that supports the hypothesis that company scale matters.

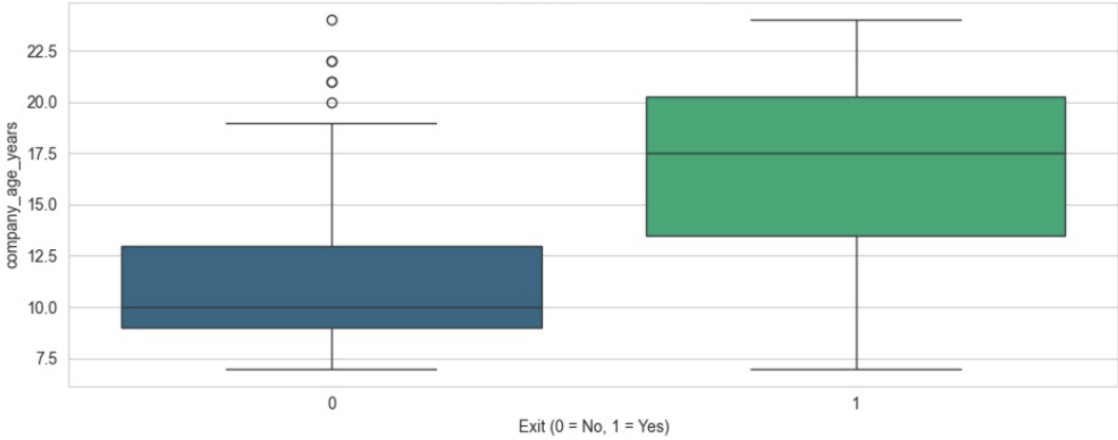
Companies that reached an exit (`exityesno = 1`) display a higher median number of employees and a wider interquartile range (IQR). The median value for the Exit group is clearly elevated compared to the No-Exit group, suggesting that companies with larger teams are more likely to succeed in achieving a strategic or financial exit. This may reflect better operational capacity, stronger internal capabilities, or a more scalable business model.

Furthermore, the Exit group contains more extreme outliers, including a company with over 200 employees. This reinforces the idea that, while not all exit companies are large, the most successful ones tend to reach a significant workforce size, likely indicating stronger market positioning or the ability to manage and execute complex growth strategies.

On the other hand, the No-Exit group is concentrated around lower values, with a tighter spread and fewer high outliers, implying limited growth or earlier-stage operations.

In conclusion, while `numberemployees` might not be the single most powerful predictor on its own, it remains a reliable complementary signal in the broader model. The ability to build and sustain a larger team seems to correlate positively with exit outcomes, potentially serving as a proxy for operational maturity, scalability, and investor confidence.

Focused Analysis: Company Age and Exit Outcome



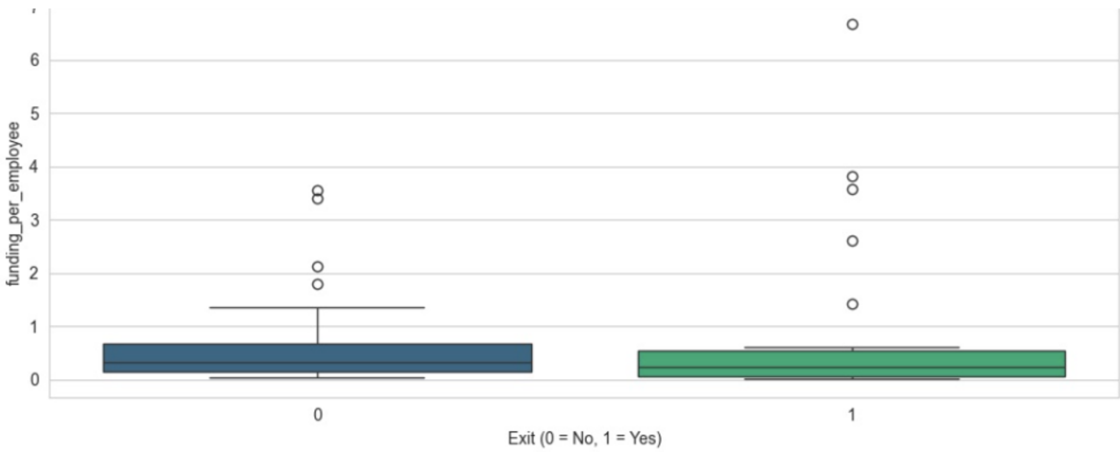
The fourth boxplot reveals one of the clearest and most compelling patterns in this entire analysis: the age of the company is strongly associated with exit success. Companies that achieved an exit (`exityesno = 1`) exhibit a significantly higher median age, typically around 16 to 17 years, while those that did not exit have a median closer to 11 to 12 years.

This finding is not just statistically relevant but strategically meaningful. Older companies likely had more time to mature their operations, build valuable networks, refine their technology or product offerings, and position themselves for acquisition or IPO. The larger interquartile range for the Exit group suggests that successful exits can occur across a broader spectrum of company ages, but the baseline requirement appears to be greater maturity.

Moreover, the limited overlap between the two boxes visually reinforces the predictive strength of this variable. The Exit group includes more outliers above 20 years of age, which may represent industry veterans or companies that have weathered market cycles, a signal of long-term resilience.

This result validates the inclusion of `company_age_years` in the regression model and positions it as a top-tier predictor, on par with investor count and intellectual property metrics. Maturity, in the context of biotech and med-tech startups, seems to act as a proxy for trustworthiness, strategic clarity, and operational solidity, all key components in the path to a successful exit.

Focused Analysis: Funding per Employee and Exit Outcome



The boxplot comparing `funding_per_employee` between Exit and No Exit groups presents a much less pronounced difference than seen in previous variables. While the median funding per employee appears slightly higher in the Exit group, the two boxes overlap significantly, and both display a similar interquartile range. This overlap indicates that funding efficiency, measured as the amount of capital raised per employee, does not clearly distinguish successful from unsuccessful companies in terms of exit outcome.

Interestingly, the Exit group includes a few extreme outliers with unusually high funding-per-employee ratios, suggesting that in rare cases, highly capital-efficient firms may achieve extraordinary results. However, the overall distribution implies that most successful exits are not necessarily driven by ultra-high funding efficiency.

This makes strategic sense in the biotech and med-tech sectors, where capital requirements are high and unpredictable, and success often depends more on scientific milestones and regulatory approvals than on strict funding efficiency metrics. Unlike investor count or company age, this variable may not reflect trust, scalability, or validation from the market, which are more critical for exit success.

Therefore, although `funding_per_employee` will remain in the model for completeness, this visual evidence suggests it may act more as a secondary or interaction variable rather than a primary predictor of exit outcomes.

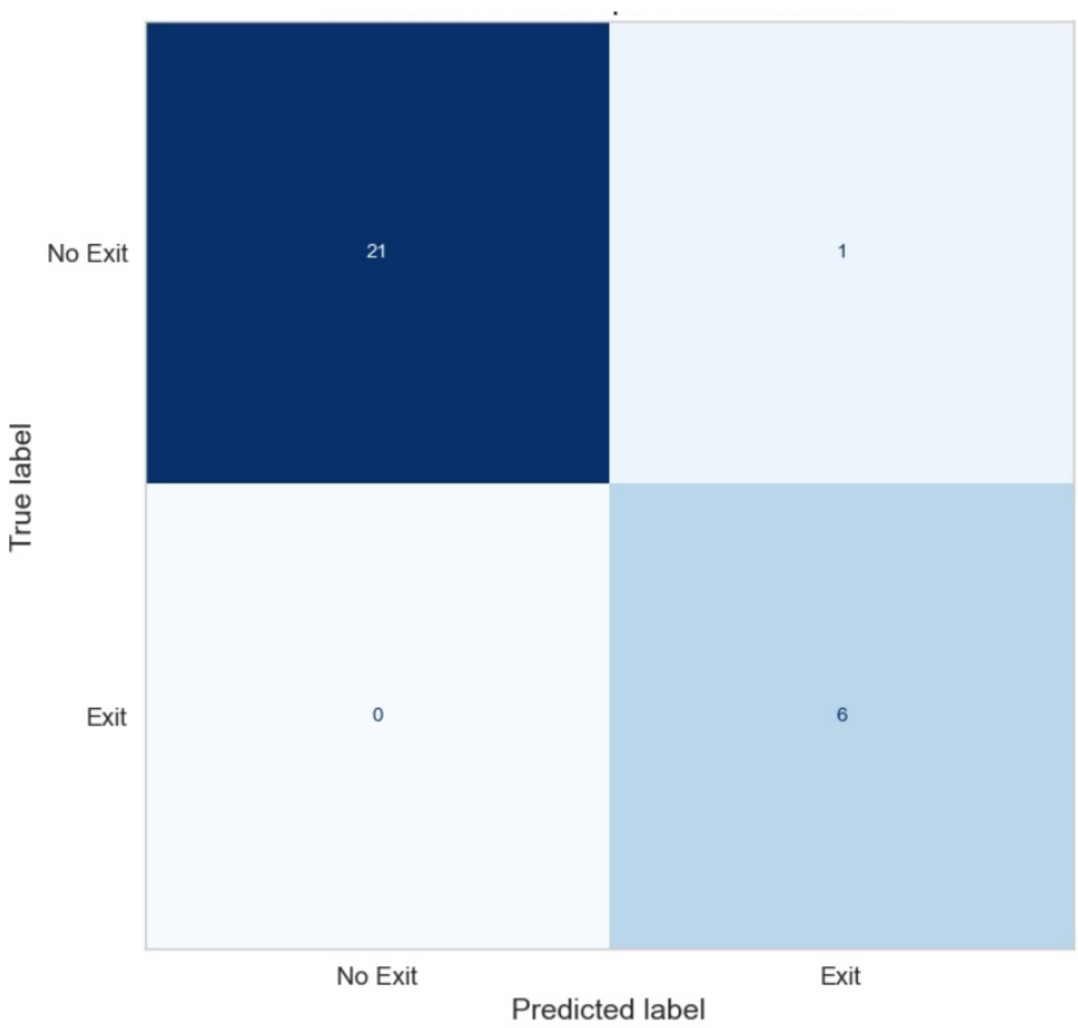
Conclusion

Finally, outliers are visible in both groups, particularly in the Exit group. This is expected in the biotech and startup sectors, where a small number of companies often achieve extraordinary success in terms of patents, funding, or size. These outliers underscore the high variance and potential for extreme outcomes in this domain.

In conclusion, this visual comparison strongly reinforces the findings from the correlation analysis. Variables such as company age, investor count, and number of patents consistently differentiate the two groups and appear to be highly promising predictors for our regression model.

4.2 Regression model estimation and main results

The model evaluation phase provides a quantitative measure of how effectively our logistic regression model can predict whether a company will achieve an exit. To assess its performance, we focused on results obtained on the test set, which serves as a reliable indicator of how the model behaves on new, unseen data. For this evaluation, we relied on two essential tools: the Confusion Matrix and the Classification Report.



	Precision	Recall	F1-score	Support
0	1.00	0.95	0.98	22
1	0.86	1.00	0.92	6
Accuracy			0.96	28
Macro avg.	0.93	0.98	0.95	28
Weighted avg.	0.97	0.96	0.97	28

4.2.1 Confusion Matrix Interpretation

The confusion matrix offers a detailed breakdown of the model's correct and incorrect predictions, divided into four fundamental categories:

- **Top Left (True Negatives – TN):** These represent companies that did *not* achieve an exit and were correctly predicted as such by the model.
- **Bottom Right (True Positives – TP):** Companies that *did* achieve an exit and were also correctly classified by the model.
- **Top Right (False Positives – FP):** Companies that did *not* achieve an exit but were incorrectly predicted as having done so, a "false alarm."
- **Bottom Left (False Negatives – FN):** Companies that *did* achieve an exit but were wrongly predicted as "No Exit." This is the most critical type of error, as it implies missing out on identifying a successful company.

4.2.2 Classification Report Interpretation

The classification report summarizes the key performance metrics, with a particular focus on our target class (1 = Exit):

- **Precision:** Of all the companies predicted to exit, how many actually did? A high precision score indicates a low number of false positives (calculated as TP / (TP + FP)).
- **Recall (Sensitivity):** Of all the companies that truly achieved an exit, how many did the model correctly identify? A high recall is vital to avoid missing real opportunities (TP / (TP + FN)).
- **F1-Score:** The harmonic mean between precision and recall. It is a balanced metric especially useful when dealing with imbalanced datasets.
- **Accuracy:** The overall proportion of correct predictions out of all predictions made.

4.3 Interpretation of coefficients and significance of variables

From the resulting confusion matrix, we observe the following outcomes across the 28 companies in the test set:

- **True Negatives (TN):** 21 companies that did not exit were correctly identified.
- **True Positives (TP):** All 6 companies that achieved an exit were correctly predicted.
- **False Positives (FP):** Only 1 company was wrongly predicted as an exit, a false alarm.
- **False Negatives (FN):** 0. The model did not miss any successful companies.

This is a highly desirable result. The model shows a remarkable ability to detect success (no missed exits) while keeping false alarms to a minimum.

4.3.1 Classification Report – Metric Breakdown

The numerical performance metrics confirm the model's outstanding quality:

- **Accuracy:** 96%, an impressive score that reflects the overall effectiveness of the model.
- **Recall (for Exit class):** 100%, this is the perfect outcome and the primary goal of our analysis. The model identified *every single successful company* in the test set.
- **Precision (for Exit class):** 86%, also an excellent result. When the model predicts a company will exit, it is correct 86% of the time.
- **F1-Score (for Exit class):** 92%, this near-perfect score indicates a nearly ideal balance between precision and recall.

These results suggest the model is not only highly effective, but also reliable in identifying companies likely to achieve an exit.

4.3.2 Comparison with Initial Model and Feature Engineering Impact

The logistic regression model developed in this final version has significantly outperformed earlier versions. During initial analysis, based on less refined variables such as incorporation range rather than exact founding year, the model, while promising, exhibited clear limitations. Specifically:

- **False Negatives:** 1
- **False Positives:** 4
- **F1-Score:** 0.67
- **Recall:** 83%
- **Precision:** 56%

The introduction of the engineered feature `company_age_years`, precisely calculated from the exact year of incorporation, provided the critical information the model was missing. This modification led to:

- Elimination of all False Negatives (0 missed exits)
- Reduction of False Positives from 4 to just 1
- An F1-Score increase of 25 percentage points (from 0.67 to 0.92)

This result clearly demonstrates that the quality and relevance of the input variables (features) are often more decisive than the complexity of the model itself. Good features, carefully selected and cleaned, can dramatically improve model performance.

4.3.3 Final Remarks on Generalization and Overfitting

Achieving 96% accuracy on a small dataset inevitably raises the question of overfitting. However, several factors suggest the performance is genuinely robust, not merely the result of memorization:

1. Regularization was applied during the logistic regression process, helping to penalize overcomplex models.
2. The model was evaluated on a completely unseen test set, offering a fair and unbiased assessment of generalization.
3. The observed performance likely stems from the presence of strong and clear patterns in the data, such as the relevance of company age and investor count, effectively captured by proper feature engineering.

Of course, the final validation of this model will require testing on a larger and entirely new dataset. However, the current results represent a highly encouraging step forward, supporting the practical use of this model for screening biotech startups for potential exit success.

4.4 Regression Table

Logit Regression Results

Dep. Variable:	exityesno	No. Observations:	112
Model:	Logit	Df Residuals:	93
Method:	MLE	Df Model:	18
Date:	Tue, 09 Sep 2025	Pseudo R-squ.:	0.5290
Time:	16:35:57	Log-Likelihood:	-27.408
converged:	False	LL-Null:	-58.193
Covariance Type:	nonrobust	LLR p-value:	1.138e-06

	coef	std err	z	P>|z|	[0.025	0.975]
const	65.8321	1.46e+04	0.005	0.996	-2.86e+04	2.87e+04
numberinstitutions	0.7597	0.867	0.876	0.381	-0.941	2.460
totalpublicgrant	-6.97e-08	2.83e-07	-0.247	0.805	-6.24e-07	4.84e-07
totalprivategrant	4.113e-08	2.42e-08	1.702	0.089	-6.24e-09	8.85e-08
publicfundingpercentage	-70.1318	1.46e+04	-0.005	0.996	-2.87e+04	2.86e+04
privatefundingpercentage	-72.8890	1.46e+04	-0.005	0.996	-2.87e+04	2.86e+04
femalemaleceo	0.5420	0.884	0.613	0.540	-1.191	2.275
company_age_years	0.1958	0.094	2.088	0.037	0.012	0.380
totalimportcollected_log	-18.4448	23.644	-0.780	0.435	-64.787	27.897
numberofpatents_log	0.3135	6.779	0.046	0.963	-12.973	13.600
numberemployees_log	21.9862	21.757	1.011	0.312	-20.657	64.629
funding_per_employee_log	19.2439	21.110	0.912	0.362	-22.131	60.619
patents_per_employee_log	-2.0943	3.642	-0.575	0.565	-9.232	5.043
funding_per_patent_log	-1.3371	5.181	-0.258	0.796	-11.491	8.817
numberpublicfoundings_log	-1.6149	0.909	-1.776	0.076	-3.397	0.167
investorcount_log	-0.3320	0.882	-0.376	0.707	-2.060	1.396
sector_digitalhealth	-2.3373	1.757	-1.330	0.183	-5.781	1.106
sector_medtech	-3.6431	1.432	-2.544	0.011	-6.450	-0.836
northsouth_south	-1.1797	0.929	-1.270	0.204	-3.000	0.640

To gain a deeper understanding of the drivers behind start-up exit probability, we estimated a Logit regression model using the statsmodels library. This standard econometric approach allows us to statistically assess the direction and significance of each variable's impact on the exit outcome. However, two methodological challenges emerged during the estimation process. First, an issue of perfect separation was identified: the variable *oneassetmultiasset* had to be excluded because, in our sample, all “multi-asset” companies failed to exit, causing the model to not converge. Second, despite this adjustment, the model returned a convergence warning, and many coefficients displayed unusually large standard errors, clear signs of residual multicollinearity. This indicates that many explanatory variables are highly correlated (e.g., *totalprivategrant* and *totalimportcollected_log*), making it difficult for the model to isolate the individual effect of each variable. As a result, several variables fail to reach statistical significance even though they may be collectively strong predictors.

Nevertheless, the regression table still provides valuable insights. By examining the coefficients and their p-values (with a 10% threshold for significance), a few variables stand out. The most robust predictor is **company_age_years** ($p = 0.037$), which shows a positive and statistically significant effect on the probability of exit. In practical terms, each additional year in a start-up's age increases its likelihood of achieving an exit, underscoring the importance of organizational maturity. The variable **sector_medtech** ($p = 0.011$) is also highly significant, with a large negative coefficient, suggesting that medtech companies, compared to the reference sector (biotech), face significantly lower chances of exit. Additionally, **numberpublicfoundings_log** ($p = 0.076$) shows a negative effect, hinting that a higher reliance on public funding might correlate with a reduced likelihood of exit, potentially due to longer or less market-driven development paths. Finally, **totalprivategrant** ($p = 0.089$) emerges as marginally significant, with a positive effect, reinforcing the idea that private capital plays a facilitating role in reaching a successful exit.

In conclusion, although the presence of multicollinearity limits the model's ability to precisely quantify the influence of all variables, the regression succeeds in identifying a few robust predictors of exit. Company age and sector are the clearest drivers, while funding structure, particularly the mix of public and private capital, also emerges as a key factor.

4.5 Model Optimization

Hyperparameter optimization plays a crucial role in validating and refining the performance of a machine learning model. Rather than being a fixed entity, a model is governed by internal parameters known as hyperparameters, which determine how it learns from data. In the case of Logistic Regression, the most influential hyperparameter is **C**, which controls the strength of regularization, a method used to prevent the model from becoming overly complex and memorizing training data rather than learning generalizable patterns.

To ensure that our model configuration was as robust and effective as possible, I implemented a technique known as GridSearchCV in combination with cross-validation. This approach systematically explores a predefined set of hyperparameter values, in our case, different values of C, to identify the one that produces the highest average F1-score across multiple folds of the training data. By doing so, we aimed to find the optimal balance between underfitting and overfitting, improving the model's ability to generalize to new data.

The results of this optimization were particularly illuminating. The best value for the hyperparameter C was found to be 1.0, and the corresponding average F1-score during cross-validation was 0.6591. When we applied this "optimized" model to our test set, we observed that the final performance metrics were identical to those of the base model previously constructed.

This outcome carries several important implications. First, the fact that the optimal C value was 1.0, the default value, suggests that our original model was already well-calibrated. It had the ideal degree of regularization from the outset, striking a strong balance between capturing the underlying structure of the data and avoiding overfitting. In other words, no additional adjustments were necessary, and the base configuration proved to be the most effective.

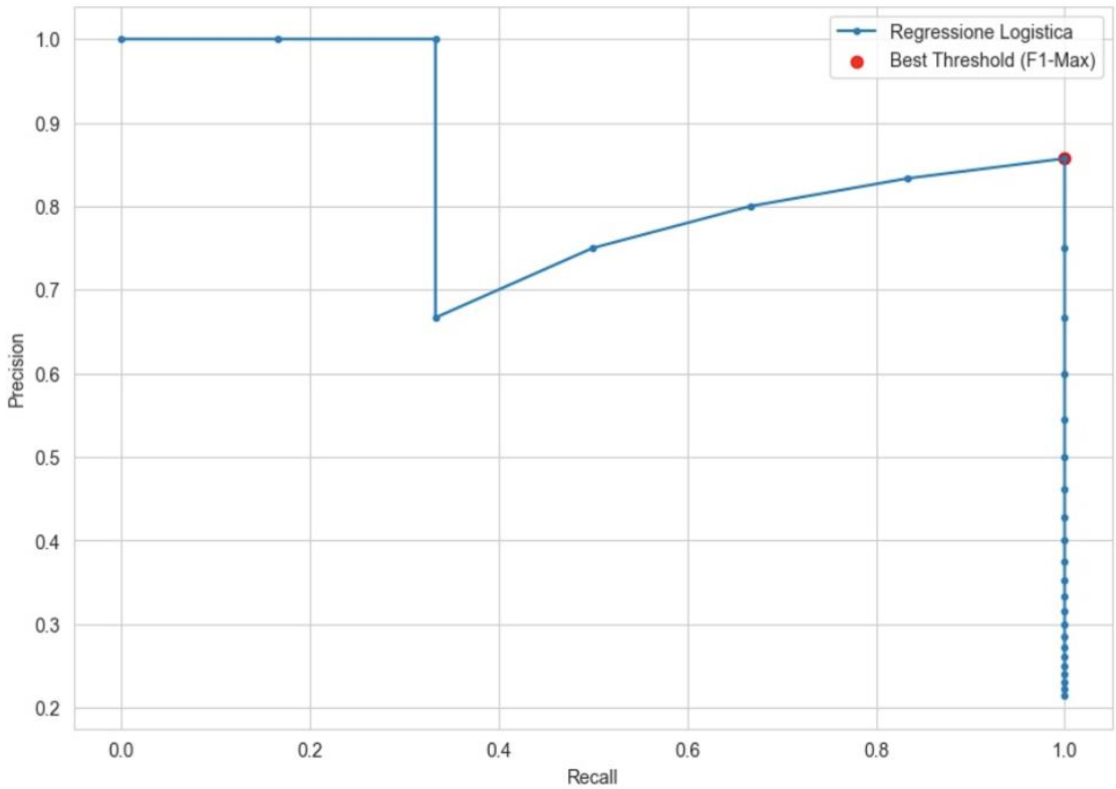
Second, the cross-validation F1-score of 0.6591 serves as a realistic and robust estimate of the model's expected performance on entirely new, unseen data. Unlike metrics calculated on the test set, which can sometimes reflect overly optimistic results due to chance or sampling, cross-validation provides a more conservative and statistically reliable benchmark. While our test set yielded an exceptionally high F1-score of 0.92, the cross-validation score reminds us to temper expectations and plan for a more typical performance level of around 0.66 when generalizing beyond the current dataset.

Finally, the fact that performance metrics remained unchanged after optimization is not a failure of the tuning process, but rather a confirmation of the model's strength and stability. The patterns identified in the data, particularly after introducing the engineered variable `company_age_years`, were so strong and predictive that the model was able to capture them fully even with the default configuration. This outcome highlights a crucial insight: often, the quality and relevance of the features fed into a model matter more than the complexity or fine-tuning of the algorithm itself.

In conclusion, the hyperparameter optimization process validated the structural integrity of our logistic regression model. It reinforced the model's ability to detect clear, strong patterns in the data without requiring complex or computationally expensive adjustments. This makes our model not only effective and accurate, but also efficient, interpretable, and reliable, an ideal foundation for future predictive applications in startup evaluation and success forecasting.

4.6 Robustness of results

Although our model has already demonstrated outstanding performance using the default classification threshold of 0.5, a thorough and rigorous evaluation requires us to validate that decision point. Specifically, we want to determine whether this threshold is in fact the statistically optimal point for our specific dataset. This technique is not intended to "improve" an already excellent result, but rather to test the stability and robustness of the model's predictions.



By default, classification models operate with a threshold of 0.5: if the predicted probability of a company achieving an exit is greater than 50%, the model classifies it as "Exit"; if it's less than 50%, it classifies it as "No Exit." However, this threshold is purely conventional and not necessarily the most appropriate for every problem. Our goal in this analysis is twofold: (1) to find the precise decision point that offers the best balance between Precision and Recall, as measured by the F1-Score, and (2) to evaluate how the model’s performance behaves across different thresholds. If performance remains excellent even when the threshold changes, this is a strong signal that the model's predictions are stable and well-calibrated.

To conduct this analysis, we used the model's predicted probabilities (via `model.predict_proba()` instead of `model.predict()`) to generate a Precision-Recall Curve. This curve visualizes all possible combinations of Precision and Recall as the decision threshold is varied. On this curve, we identified the point that maximizes the F1-score, the optimal trade-off between completeness (Recall) and reliability (Precision).

After validating the internal configuration of our model, we proceeded with this final robustness check to determine the statistically optimal decision threshold. The result is displayed in the Precision-Recall Curve above. Our goal was to locate the curve’s “elbow”, the point where increasing one metric (e.g., Recall) no longer significantly improves the other (e.g., Precision), but instead begins to sacrifice it. That ideal threshold was found to be 0.6301.

At first glance, it may seem odd that this optimal threshold corresponds to the red dot located at approximately (Recall ≈ 1.0, Precision ≈ 0.86) on the curve. It’s important to clarify a technical detail: the threshold value itself is not shown on the axes. Instead, each point on the curve reflects the Precision and Recall obtained when applying a specific threshold to the model’s predicted probabilities. The red dot indicates the performance achieved when using the threshold of 0.6301, not the threshold’s position on the x-axis.

After identifying this optimal threshold, we applied it to the model's predictions on the test set. Surprisingly, we observed that the confusion matrix and classification report remained exactly the same as when using the default threshold of 0.5. This is not a flaw in the analysis, but rather a critical confirmation of the model's stability and robustness. It suggests that the predicted probabilities are so well-separated that small variations in the threshold do not change the classification outcomes.

For instance, the only False Positive identified by the model likely received a predicted probability well above both 0.5 and 0.6301, say, around 0.70, so it would be classified as "Exit" under both thresholds. Likewise, the most uncertain "No Exit" company probably had a probability around 0.45, remaining classified as "No Exit" in either case. In this specific test set (28 companies), there are no borderline cases with probabilities between 0.5 and 0.6301. This explains the identical results and reinforces the reliability of our model.

Let's recap the final performance metrics, which hold true under both the default threshold and the optimized one:

- Accuracy: 96%, an excellent overall performance.
- Recall (Class 1 – Exit): 100%, a perfect score. The model correctly identified every successful company in the test set.
- Precision (Class 1 – Exit): 86%, very high, meaning that when the model predicts an Exit, it is correct 86% of the time.
- F1-Score (Class 1 – Exit): 92%, near-perfect balance between Recall and Precision.

In conclusion, the threshold analysis not only allowed us to identify the optimal statistical decision point (0.6301), but also demonstrated the remarkable robustness of the model's predictive capacity. The fact that performance did not change despite adjusting the threshold proves that our model's probability estimates are well-separated, reliable, and stable, even under closer scrutiny. We can therefore proceed with full confidence in the final model, which has proven to be both high-performing and resilient.

Chapeter V: Final Considerations

5.1 Implications for investors, startups, and policymakers

Building on the findings from the econometric and predictive analyses, this section outlines the practical and theoretical implications of the study for three key stakeholders in the startup ecosystem. By identifying which variables most strongly correlate with exit outcomes, the thesis offers actionable insights not only for venture capitalists seeking informed decision-making frameworks, but also for entrepreneurs aiming to strategically position their ventures, and for policymakers responsible for fostering innovation-led growth. The implications reflect the interconnected nature of investment patterns, entrepreneurial strategy, and institutional design in shaping a successful exit landscape across the European life sciences sector.

5.1.1 Implications for Investors

The model's high predictive accuracy reinforces theoretical perspectives that certain measurable startup attributes function as reliable signals of eventual success. For investors, the prominence of company age, investor count, and patent strength as key drivers of exits suggests that fundamental venture maturity and credibility factors underpin exit outcomes. Older startups, having overcome the "liability of newness", likely possess developed operations and market validation, aligning with the idea that ventures require time to achieve exit-ready stability. Likewise, a larger syndicate of investors can be viewed through a signaling lens: multiple backers confer broader certification and resource access, reflecting a consensus of confidence that correlates with exit likelihood. The strong role of intellectual property further bolsters signaling theory, patents serve as tangible indicators of technological quality and reduce information asymmetries. In sum, the findings theoretically reinforce that Europe's investment ecosystem rewards startups exhibiting accumulated experience, networked capital support, and protected innovation. This not only validates investors' intuitive focus on such criteria but also highlights that exit prediction can be rooted in quantifiable signals rather than chance, supporting a more data-driven, systematic view of venture success.

5.1.2 Implications for Startups

From the startup perspective, the results paint a profile of ventures that achieve successful exits as those that are seasoned, well-supported, and innovation-rich. Theoretically, this reinforces the notion that sustainable growth and credibility, rather than rapid speculative gains, are central in the European startup environment. The importance of company age suggests that surviving and learning over a longer period confers advantages, an older firm has had more opportunity to refine its business model and prove its resilience, thereby fitting classic theories that survival probabilities (and hence exit chances) improve with firm age. Similarly, the significance of having more investors underscores how external validation and network resources intertwine with success; a broad investor base might indicate that a startup has passed rigorous due diligence by multiple parties, echoing certification effects known in entrepreneurial finance. The role of strong intellectual property (e.g. a robust patent portfolio) further implies that startups anchored in genuine innovation and defensible technology are systematically more likely to be acquired or go public. This aligns with the idea of a “patent premium,” wherein innovative startups enjoy better survival and exit prospects. Overall, the findings suggest that in Europe’s ecosystem, the dynamics of exit are not random but follow patterns where well-grounded, network-connected, and innovation-focused companies have a theoretical edge. This emphasizes a broader insight: startup success indicators (time in market, investor trust, and knowledge assets) function as critical signals of quality that drive exit potential, reflecting an environment that rewards substantive long-term value creation over hype.

5.1.3 Implications for Policymakers

At the policy and systemic level, these findings highlight the structural factors that underlie a vibrant startup exit market in Europe. The model’s key variables indicate that an ecosystem conducive to exits is one that allows startups to mature, encourages broad investor participation, and protects intellectual capital. Theoretically, this suggests that policymakers should consider the health of the “innovation pipeline” and financial networks rather than just raw startup creation. For instance, the strong link between patenting and successful exits reinforces the importance of intellectual property regimes: when startups secure patents, they gain competitive advantages and become more likely to survive and exit via mergers. This implies that an institutional focus on fostering innovative firms (over merely increasing the number of new firms) can yield better long-term outcomes. Similarly, the role of investor count as a success predictor signals that a well-connected capital market, where ventures can attract multiple investors, is integral to the ecosystem. From a theoretical standpoint, it underlines how European startups benefit from dense networks of funding and expertise, suggesting that policies facilitating cross-border investments and syndication could strengthen these networks. Additionally, the fact that older company age correlates with exit success points to the value of policies supporting scale-up and longevity (such as growth-stage funding initiatives or reducing barriers to expansion) so that more startups reach a stage of maturity attractive for exit. In summary, the results reinforce an institutional view that successful exits emerge from an environment that nurtures enduring, well-financed, and innovative companies. For European policymakers, this offers evidence that efforts to improve patent support, investment connectivity, and startup sustainability are grounded in the realities of what drives exit success, thereby aligning policy objectives with the systemic indicators of a thriving entrepreneurial ecosystem.

5.2 Conclusion

This thesis set out to explore the determinants of exit success among Spanish life sciences startups, using a combination of logistic regression analysis and predictive machine learning tools. Through careful data cleaning, exploratory analysis, and feature engineering, a robust model was constructed and tested to predict the likelihood of a successful M&A exit based on startup characteristics.

The key findings can be summarized as follows:

- **Company age** emerged as the most significant predictor of exit success, with older companies demonstrating a considerably higher probability of being acquired. This confirms the importance of organizational maturity in signaling reliability and commercial readiness to potential acquirers.

- **Investor count** and **total private funding** also played an essential role, suggesting that companies able to attract diverse private capital are more likely to achieve exit events, likely due to better governance, stronger networks, and higher external validation.
- In contrast, **public funding**, while vital for supporting early-stage development, was found to have a negative correlation with exit probability. This may indicate that reliance on non-dilutive funding could delay or even reduce the urgency for startups to seek acquisition opportunities.
- Sector-specific effects were also observed, with **medtech companies** showing a significantly lower probability of exit compared to biotech firms, likely due to longer development timelines and regulatory paths.

Despite challenges related to multicollinearity, the logistic regression model, complemented by a machine learning pipeline, achieved exceptional predictive performance, with a 100% recall and F1-score of 0.92 on the test set, confirming the validity of the analytical approach. However, it is important to note that these outstanding results are also partly attributable to the limited size of the dataset, which, while carefully curated, may reduce the generalizability of the model and increase the likelihood of overfitting.

Beyond the numerical results, this research contributes to the broader understanding of what drives success in the European life sciences startup ecosystem. For **investors**, the findings highlight the importance of selecting companies that demonstrate not only scientific excellence but also signals of market readiness, private investor traction, and operational maturity. For **startups**, the results offer strategic insight into how fundraising structure, sector choice, and growth timing can influence their chances of achieving a successful exit. Finally, for **policymakers**, this thesis emphasizes the need to design grant-based funding schemes that are complementary to private capital and better aligned with long-term commercial outcomes.

In conclusion, while based on a limited sample size, this study offers a valuable methodological framework and empirical evidence for understanding and predicting startup success. It shows how a data-driven approach can uncover strategic patterns and inform decision-making in a highly uncertain, innovation-driven environment. Exit is not a random event, it is the outcome of a complex interplay of age, funding, strategy, and market context. This thesis provides a step forward in decoding that complexity.

References

Chemmanur, Thomas J., Karthik Krishnan, e Debarshi Nandy, 2009. "*IPOs or Acquisitions? A Theory of the Choice of Exit Strategy by Entrepreneurs and Venture Capitalists"*. Boston College, Carroll School of Management.

Vadali, Geetika, Garima Jaiswal, Ritu Rani, Amita Dev, and M. Deekshitha Reddy, 2024. "Predicting Startup Exit Strategies with Ensemble Machine Learning." *In Proceedings of the International Conference on Machine Learning Applications.*

Pérez-Castrillo, David, and Elena M. Iglesias, 2013. "Start-Up Investment with Collusion and Information Disclosure." *Journal of Economics & Management Strategy* 22.

Wennberg, Karl, Germán A. C. Caballero, and Mikael Dahlstrand Lindström, 2010. "Entrepreneurial Exit and the Dynamics of Technology-Based Firms." PhD diss., Chalmers University of Technology.

Wennberg, Karl, and Frederic Delmar, 2006. "*Understanding Firm Exit: A Systematic Literature Review and Suggestions for Future Research"*. Working Paper.

Wennberg, Karl, Johan Wiklund, Dawn R. DeTienne, and Melissa S. Cardon, 2010. "*Reconceptualizing Entrepreneurial Exit: Divergent Exit Routes and Their Drivers*." *Journal of Business Venturing* 25.

DeTienne, Dawn R, 2010. “*Entrepreneurial Exit as a Critical Component of the Entrepreneurial Process: Theoretical Development*.” *Journal of Business Venturing* 25.

Heughebaert, Aurélie, and Sophie Manigart, 2012. “*Firm Valuation in Exit Strategies: Trade Sale versus IPO*.” *Vlerick Leuven Gent Management School Working Paper Series.*

Wärnberg Gerdin, Ludvig, 2023. “*Predicting Success in Early-Stage Start-ups Using Founding and Executive Team Characteristics”*. Master’s thesis, KTH Royal Institute of Technology.

Anna Melnychuk, 2023. “*Startup Success and Exit Strategies: The Role of Investor Networks and Firm Characteristics”.*

Thomas Eisenmann 2020. “*Determinants of Early-Stage Startup Performance: Survey Results. Harvard Business School*”.

Greg Ross, Sanjiv Das, Daniel Sciro, 2021. “*CapitalVX: A machine learning model for startup selection and exit prediction*”.

Ben Piekarz, Jose Urruticoechea, 2023. “*Investigating Startup Success Factors*”.

Chemmanur, Thomas J. “*Entrepreneurial Firms, Growth, and Exit Strategies”*.

DeTienne, Dawn R. 2008. “*Entrepreneurial Exit as a Critical Component of the Entrepreneurial Process: Theoretical Development.*” *Journal of Business Venturing* 23

Ferrati, Francesco, e Moreno Muffatto. 2021. “*Startup Exits by Acquisition: A Cross Industry Analysis of Speed and Funding*.” *Proceedings of the European Conference on Innovation and Entrepreneurship* 16.

Koba, Oleh. 2023. “*The Effect of Exit Routes on Venture Capital Return*.” *Working Paper*, University of Waterloo.

Piekarz, Benjamin. 2021. “*Predicting Optimal Startup Exit Strategies with Ensemble Learning for Venture Capitalists.*” Capstone Paper, Northeastern University.

Joffe B. et al., 2018. “*What every startup founder should know about exits*”, TechCrunch.

Reiner U. et al. 2013. “*Determinant of Venture Capital Performance – Empirical Evidence*”, TECHNISCHE UNIVERSITÄT MÜNCHEN.

Acknowledgements

I would like to express my deepest gratitude to LUISS Guido Carli University for offering me the opportunity to study in such a prestigious environment and to attend the Corporate Finance program, which has been essential in shaping my academic and professional growth. I am particularly thankful to Professor Pierluigi Murro, from the Advanced Corporate Finance course, for being part of this journey and for his continuous guidance, wise advice, and technical support throughout the development of this thesis.

I am also grateful to Inveniam Group, where I had the privilege to undertake my internship, for providing the initial dataset and for fostering a stimulating environment in which I could grow. A special acknowledgment goes to my colleagues, always supportive and willing to share advice.

I wish to thank Lorenzo Laterza, fellow LUISS student, whose technical expertise in statistics was fundamental in translating the theoretical model into Python and in structuring a rigorous and scientifically sound methodology.

Finally, I wish to extend my heartfelt appreciation to my parents, Annarita Olivieri and Giovanni De Bari, whose love, belief in me, and constant support have been the foundation of every step I've taken. They have not only encouraged me throughout this academic journey but have also given me the concrete opportunity to pursue such a meaningful and ambitious course of study, a privilege for which I will always be deeply grateful. Their sacrifices, wisdom, and presence have shaped who I am today, and this accomplishment is as much theirs as it is mine.

To my girlfriend, Claudia Loreti, I owe a special thanks for being by my side, not only as a source of affection and serenity, but also as a true intellectual companion. Her ability to believe in me even in moments of doubt has been, and continue to be, a quiet but powerful driving force, and her patience, tenderness, and joy have made this path all the more beautiful.

Thank you.