Dataset Overview

Dataset Overview

This dataset, sourced from Luc Wynant's research, captures 260 private equity-backed buyout transactions with 187 detailed columns. Each row represents a distinct investment deal, encompassing data about fund structure, company profiles, financial terms, governance mechanisms, and exit strategies. The dataset spans 25 years, from 1999 to 2024, making it a valuable longitudinal resource for academic and industry-based financial analysis. It offers researchers a rich context to study trends in ownership, control rights, and performance outcomes.

Data Issues Observed

1. Missing Data

Over **5,000 cells** in the dataset contain missing values. These are spread across both numerical and categorical columns. Their presence could bias the statistical interpretations if left untreated. Addressing missingness is **essential** to ensure model reliability and to prevent downstream errors in machine learning or econometric modeling. A total of **187 columns** were identified as having at least one missing value. These missing values appear both systematically (entire columns or sections) and randomly, requiring tailored imputation strategies**.** **Missing Value Summary**

2. Error Codes Identified

The dataset uses placeholder error codes like **-9999**, **-8888**, and **-7777** . These codes were manually or systematically entered during data entry stages. They must be reclassified as **NaN** (missing) values for compatibility with analytical tools and to maintain semantic clarity. A total of **124 columns** contain at least one such placeholder error code. These errors can severely affect numeric calculations, visualizations, and summary statistics. **Error value Summary**

3. Inconsistent Data Types

Several variables intended to be **numeric** or **date-based** were stored as strings or object types. This included year columns that could not be parsed as integers and binary flags not correctly typed as booleans or categories. Type inconsistency can severely limit filtering, sorting, and statistical operations, especially in **time-based analyses**, and can cause type mismatch errors during computation.

Data Cleaning & Preparation

1. Removal of Empty Structures

All rows and columns containing only missing or error-coded entries were removed (NonsubordinatedBondLoansNi). Additionally, unnamed columns (often artifacts of Excel formatting) were dropped. This step **streamlined the dataset**, improved clarity, and ensured each variable retained some informational value for analysis. It also reduced memory usage and enhanced processing speed.

2. Data Type Normalization
Each column's data type was corrected to match its intended use. Float-type columns with whole numbers were converted to integers to save memory and improve readability. Binary and categorical fields were explicitly cast as **category** types, enabling more efficient **grouping**, **filtering**, and **visualization** during exploratory analysis.

3. Error Code Replacement
All instances of **-9999**, **-8888**, and **-7777** were replaced with NaN to standardize missing value representation. This step allowed for **consistent treatment** during imputation and helped prevent **erroneous calculations** caused by misinterpreting error codes as real numbers.

4. Filling in Missing Values
Numeric columns had their missing values replaced with the **median**, a robust measure that is resistant to extreme outliers. Categorical columns used the **most frequently occurring value (mode)**, preserving the general distribution and minimizing distortion of categories. In special cases, domain knowledge was used to infer missing entries where appropriate.

Exploratory Data Analysis (EDA)

1. Deal Summary
A total of **247 completed transactions** were analyzed, covering a range of firm sizes, years, and regions. Notably, **37.2%** of deals resulted in successful exits, indicating substantial variation in exit outcomes. This variability is useful for further correlation with governance or financial variables.

2. Management Equity Ownership

a) Ownership internal stakeholders

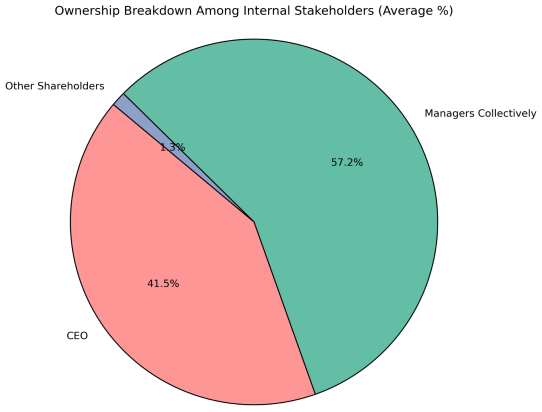


Image 1: Ownership internal stakeholders

- **The image 1** illustrates how ownership is generally divided among key internal stakeholders in the deals we analyzed. On average, the **CEO holds about 41.5% of the shares**, the **management team together owns 57.2%**, and only a small portion—**1.3%**—is owned by other shareholders.

- To find these numbers, we summed up the ownership percentages of each group across all the deals in our dataset. Then, we calculated the average shares for each group and adjusted the totals so they add up to 100%. This approach gives a clear and consistent view of how ownership is typically split among the company's internal leadership.

b) **Private Equity Investors**

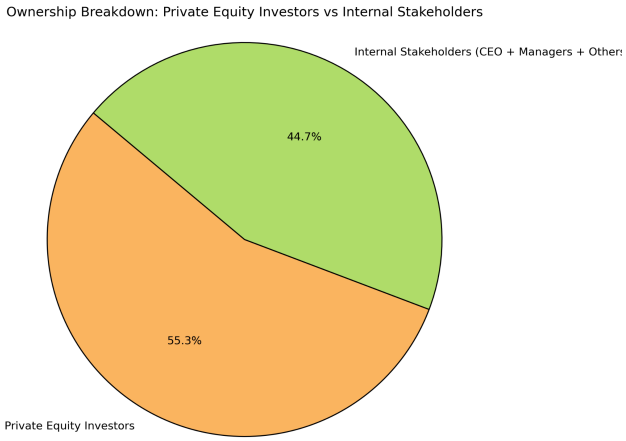


Image 2: Private Equity Investors vs Internal Stakeholdersl

- **Image 2** illustrates the ownership distribution between **Private Equity Investors** and **Internal Stakeholders** such as the CEO, managers, and other internal team members. According to the chart, **Private Equity Investors hold 55.3%** of the company, while **Internal Stakeholders hold 44.7%**. Although this suggests a relatively balanced share, it is important to understand that in most real-world cases, **outside investors typically own the majority of the company**, and **CEOs or management teams usually hold much smaller stakes**. Therefore, while this chart is useful to understand the internal distribution of ownership within the company, it does **not present the complete picture** of total ownership. Additional details about other external investors would be necessary to fully understand who controls the company.

3. Board Structure

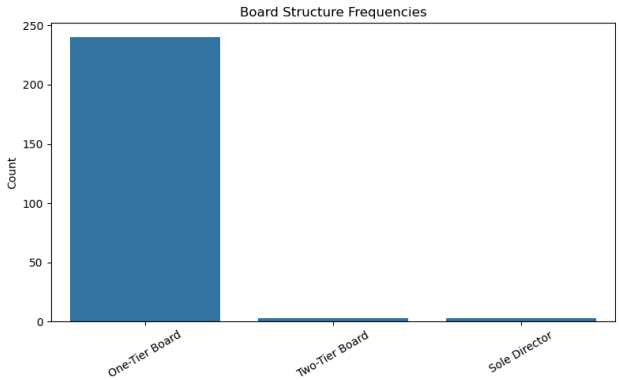


Image 3 : Board structure frequencies

- From **image 3** it's clear that **the vast majority of deals adopt a one-tier board structure**, with around **240 occurrences**, which reflects approximately **98%** of the dataset. This structure combines executive and non-executive roles into a single board, emphasizing a **unitary**

governance model. In contrast, **only a few deals (around 3 each)** feature a **two-tier board**—which separates supervisory and management responsibilities—or a **sole director** setup. This overwhelming preference for one-tier boards highlights the **dominance of unified board structures** in the observed dataset.

4. Control Clauses in Deals

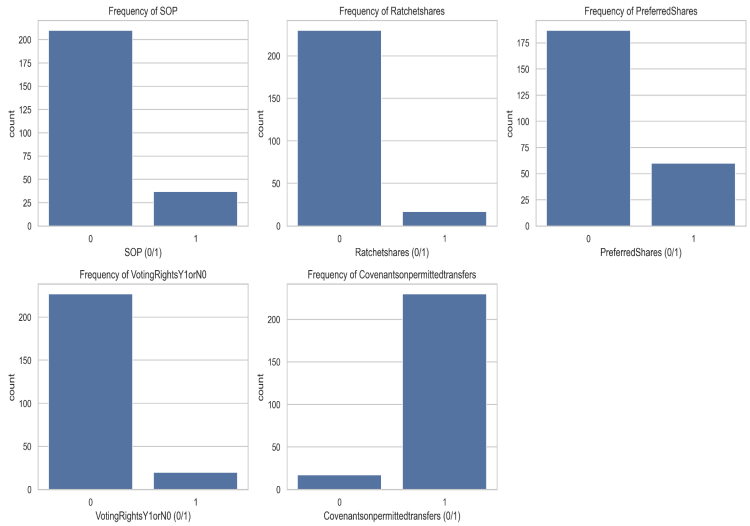


Image 4: Control clauses

- The image 4 presents five bar charts, each showing how often specific contractual clauses appear in private equity (PE) buyout deals. These clauses are measured using binary indicators from the dataset, where **0 means the clause is not present** and **1 means the clause is present**. The **x-axis** of each chart represents these binary values (0 or 1), while the **y-axis** shows the **number of contracts** (count) that fall under each category.

- The first chart, ***Frequency of SOP*** *(Stock Option Plans)*, shows that most contracts do not include SOPs, with over 200 contracts marked as 0 and only a small number (around 37) marked as 1. This suggests that SOPs are relatively uncommon in the sample. The second chart, *Frequency of* ***Ratchetshares***, similarly shows a low adoption rate: very few contracts use ratchet share mechanisms. The third chart, *Frequency of* ***PreferredShares***, indicates that while still less common than absence, preferred shares are present in about 60 contracts, showing moderate usage.

- The fourth chart, *Frequency of* ***VotingRightsY1orN0***, also displays a strong preference for not including special voting rights, with most contracts recorded as 0. Lastly, the *Frequency of* ***Covenantsonpermittedtransfers*** chart is the opposite: nearly all contracts include covenants restricting permitted transfers. Together, these graphs provide a clear picture of how often key control mechanisms appear in real-world PE contracts

adsfadfadfadfasdf