



2025

STARTUP SUCCESS

**Prediction**

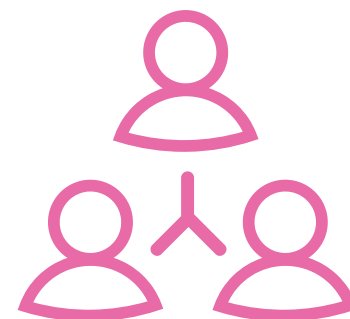
JULIA MOREIRA, GUILHERME SCHWARZ,  
RENAM BIAVATI, MATHEUS FRANCISCO



# MAIN INFO RMAT ION



*Contains info on ~900 U.S. startups (founding, funding, status, etc.) that were founded since 1998 until 2013*



*Key features:*

- *Categorical: is\_technology, status, category\_code, etc.*
- *Numerical: funding\_total\_usd, age\_at\_closing, milestones, etc.*
- *Date: founded\_at, closed\_at, first\_funding\_at, etc.*



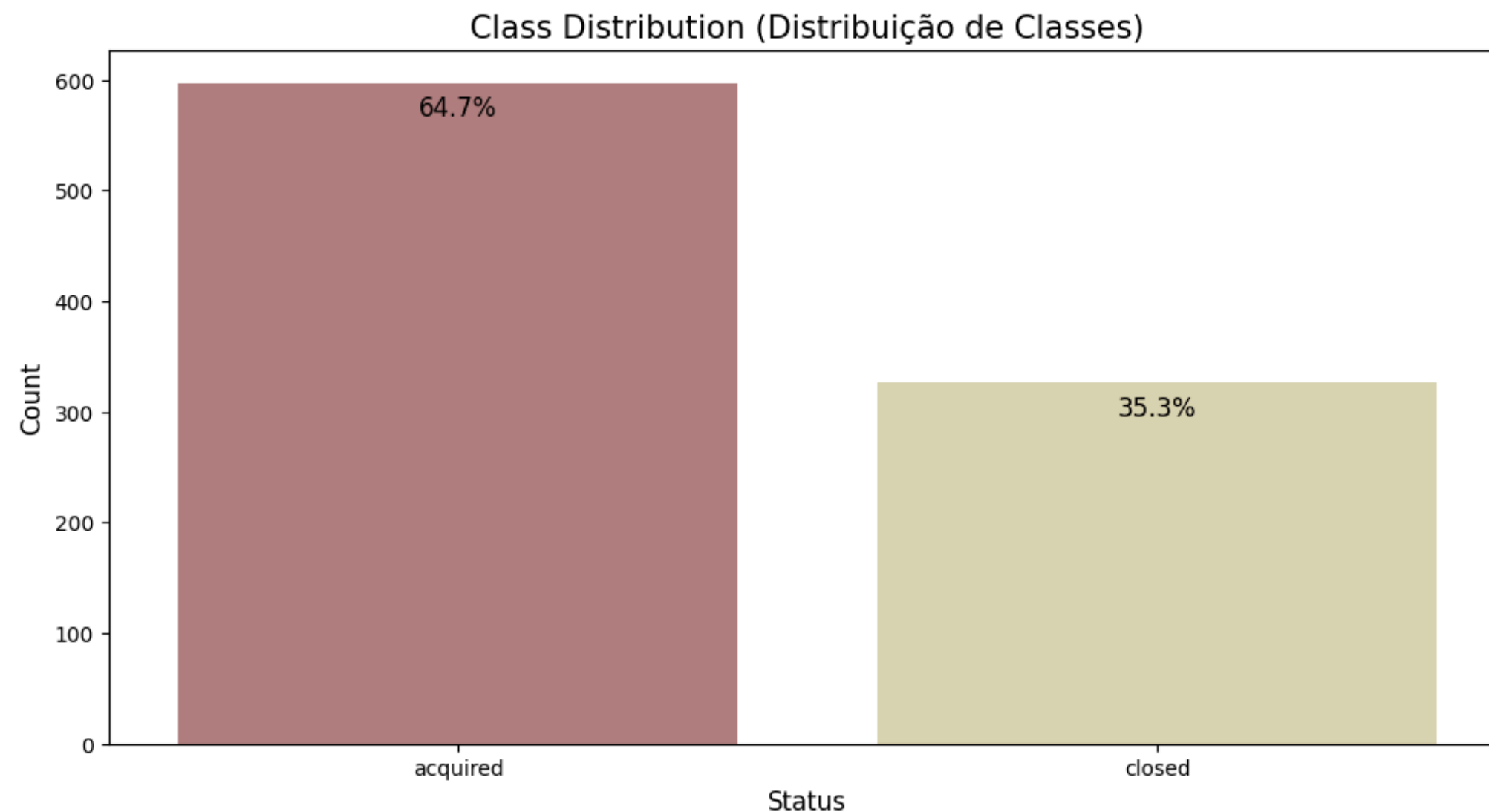
*Labeled dataset: target variable is status (acquired or closed).*

*Classification problem dataset.*

# *STATISTICAL*

## *DESCRIPTION*

Number of instances: 923 | Number of features: 49 | Number of classes: 2





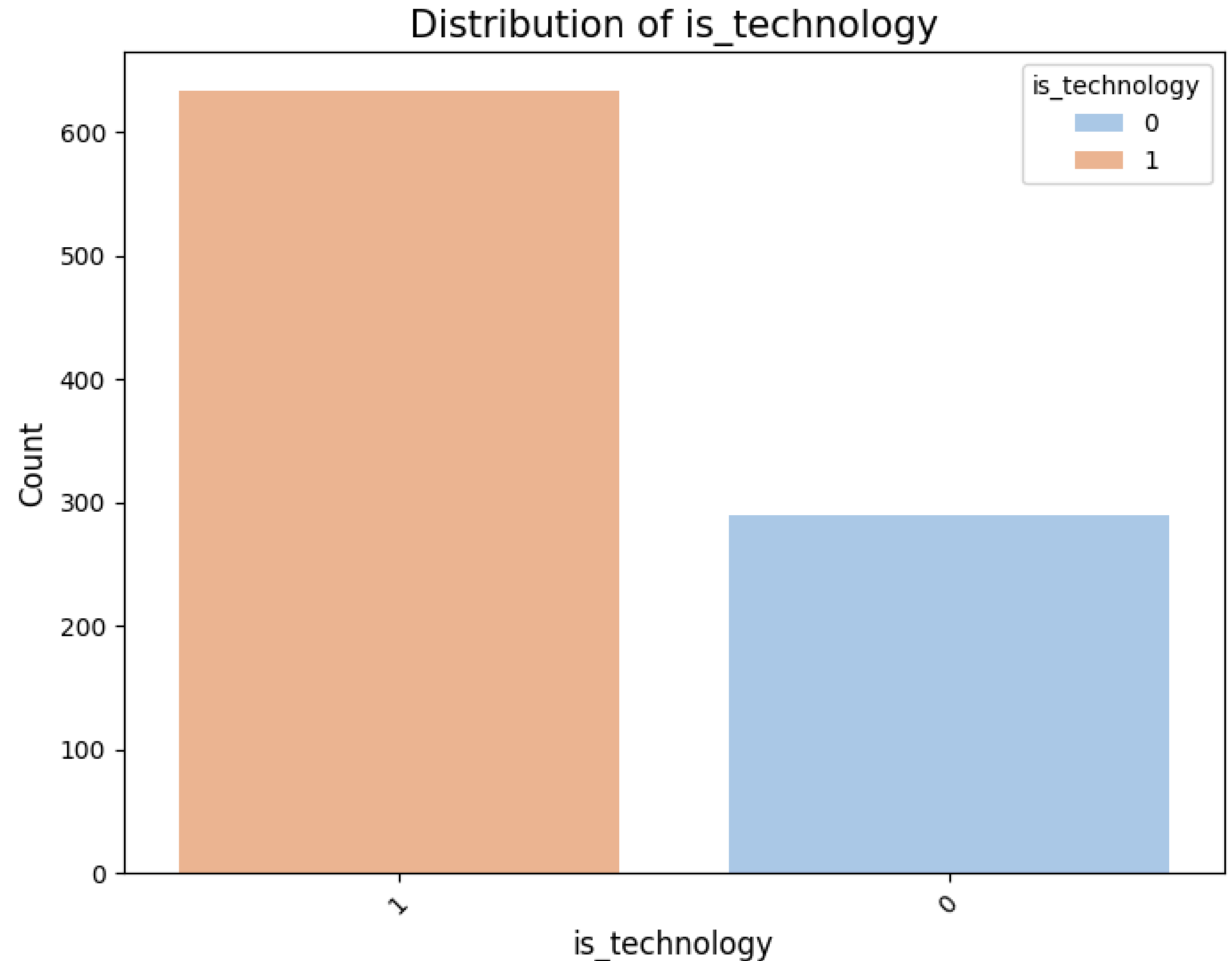
# *UNIVARIATE DATA ANALYSIS*

# ***IS\_TECHNOLOGY COLUMN***

Feature engineering column that indicates if the main activity of an industry is an IT product, like web development and applications.

Frequency Distribution of Tech Companies:

- Tech (1): 68.6%
- Non-Tech (0): 31.4%

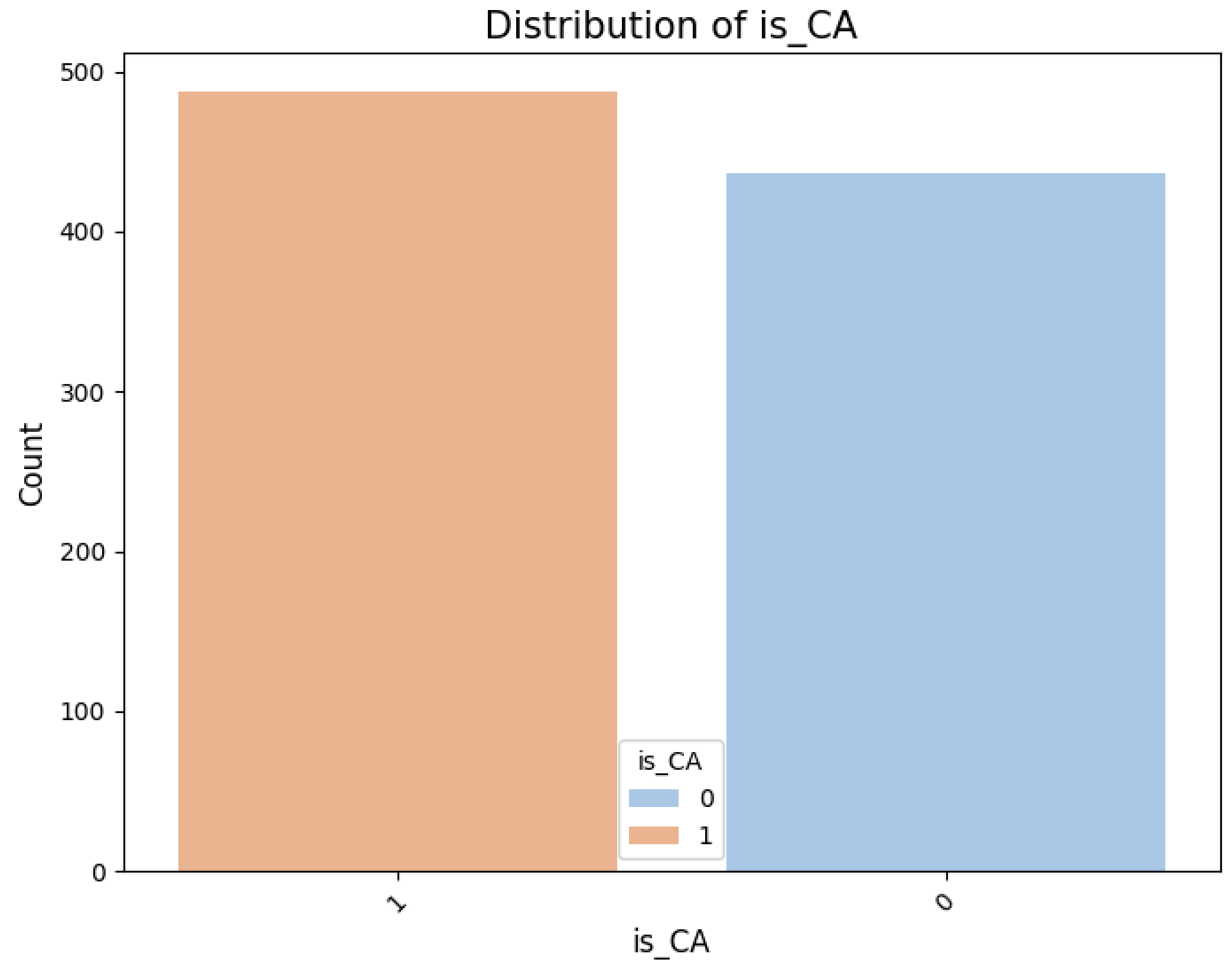


# *IS\_CA COLUMN*

Define if the startup was founded in California or not.

Frequency Distribution of Companies in California:

- Based in CA (1): 52.8%
- Not in CA (0): 47.2%

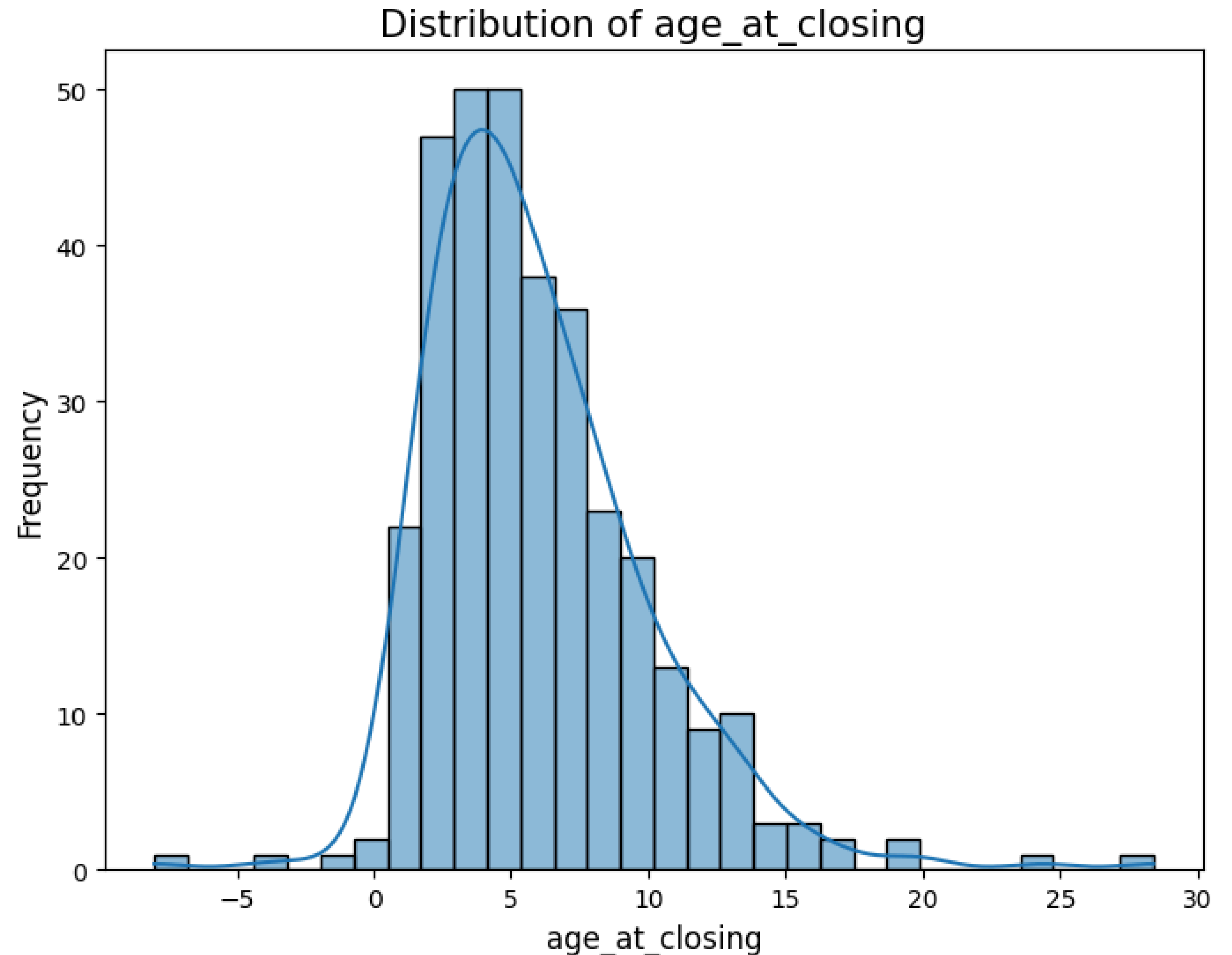


# ***AGE\_AT\_CLOSING COLUMN***

Data about how many years the company had when it was closed.

Statistics for Age at Closing:

- Mean: 5.96 years
- Std. Deviation: 4.04
- Skewness: 1.29 → Right-skewed
- Kurtosis: 4.15 → Leptokurtic (heavy tails)

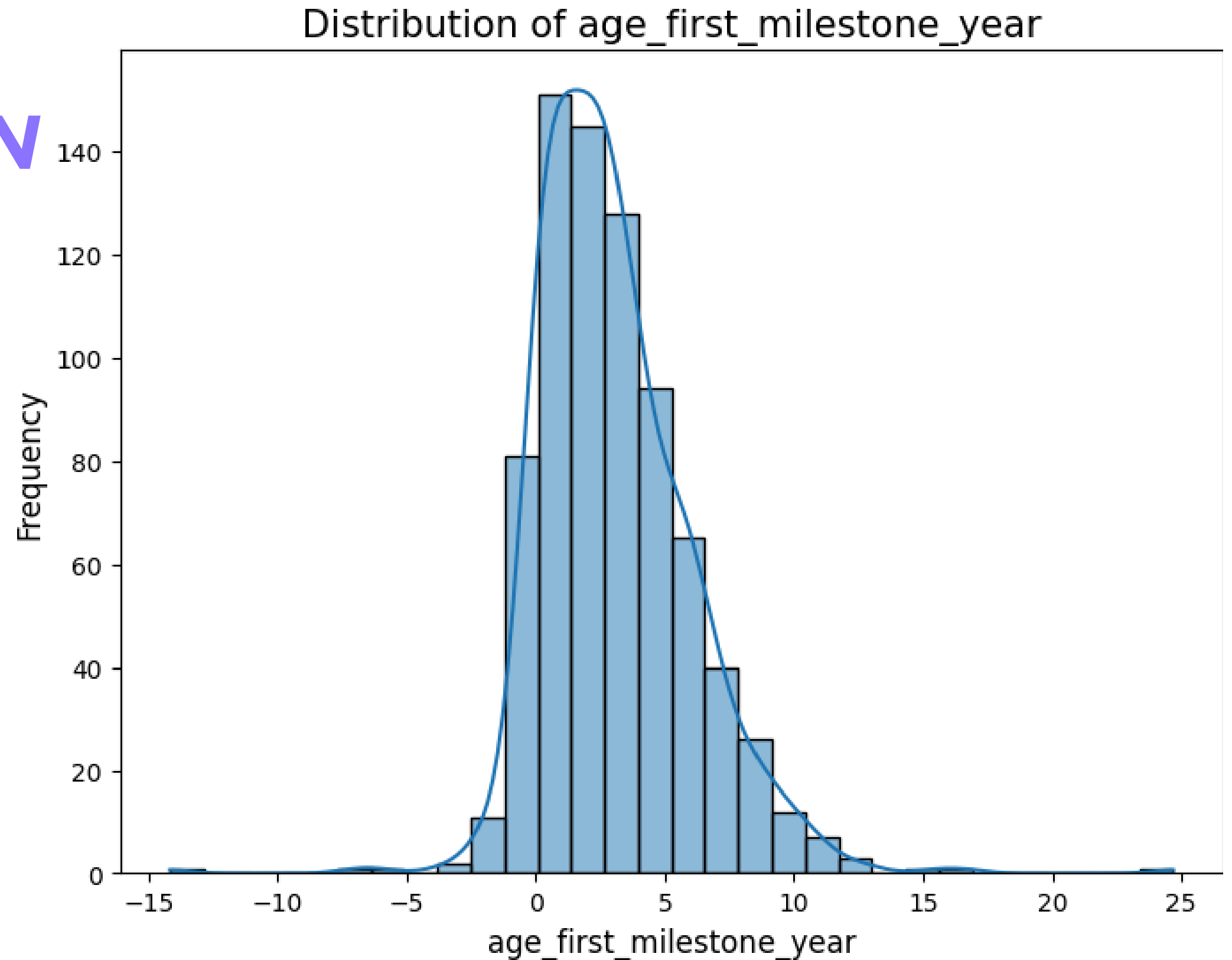


# ***AGE\_FIRST\_MILEST ONE\_YEAR COLUMN***

How many years old was the startup when it achieved its first milestone

Statistics for Age at First Milestone:

- Mean: 3.06 years
- Std. Deviation: 2.98
- Skewness: 0.94 → Moderately right-skewed
- Kurtosis: 5.60 → Leptokurtic (heavy tails)





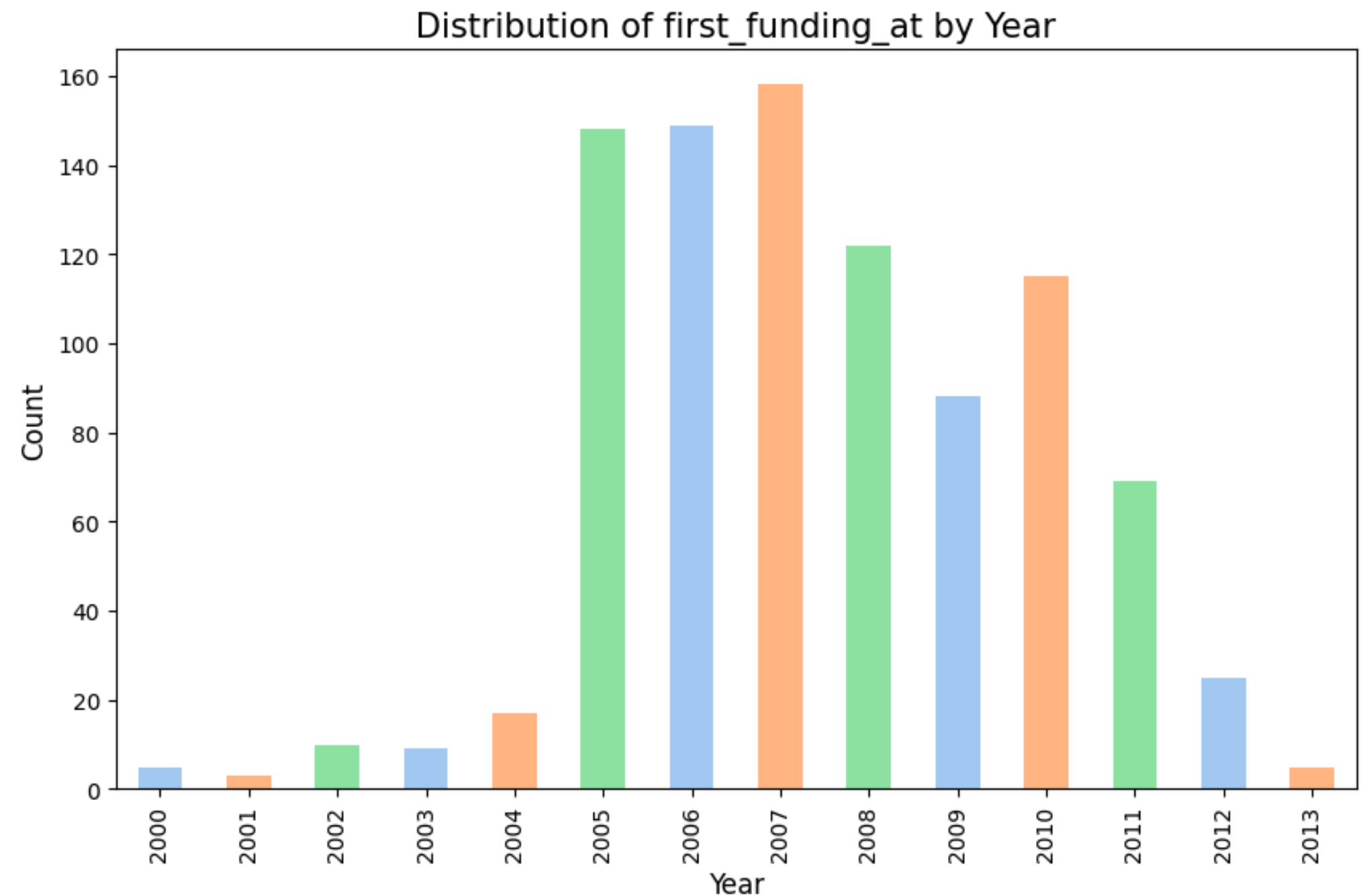
# ***FIRST\_FUNDING\_AT\_BY\_YEAR***

## ***COLUMN***

Show how many startups received their first funding by year.

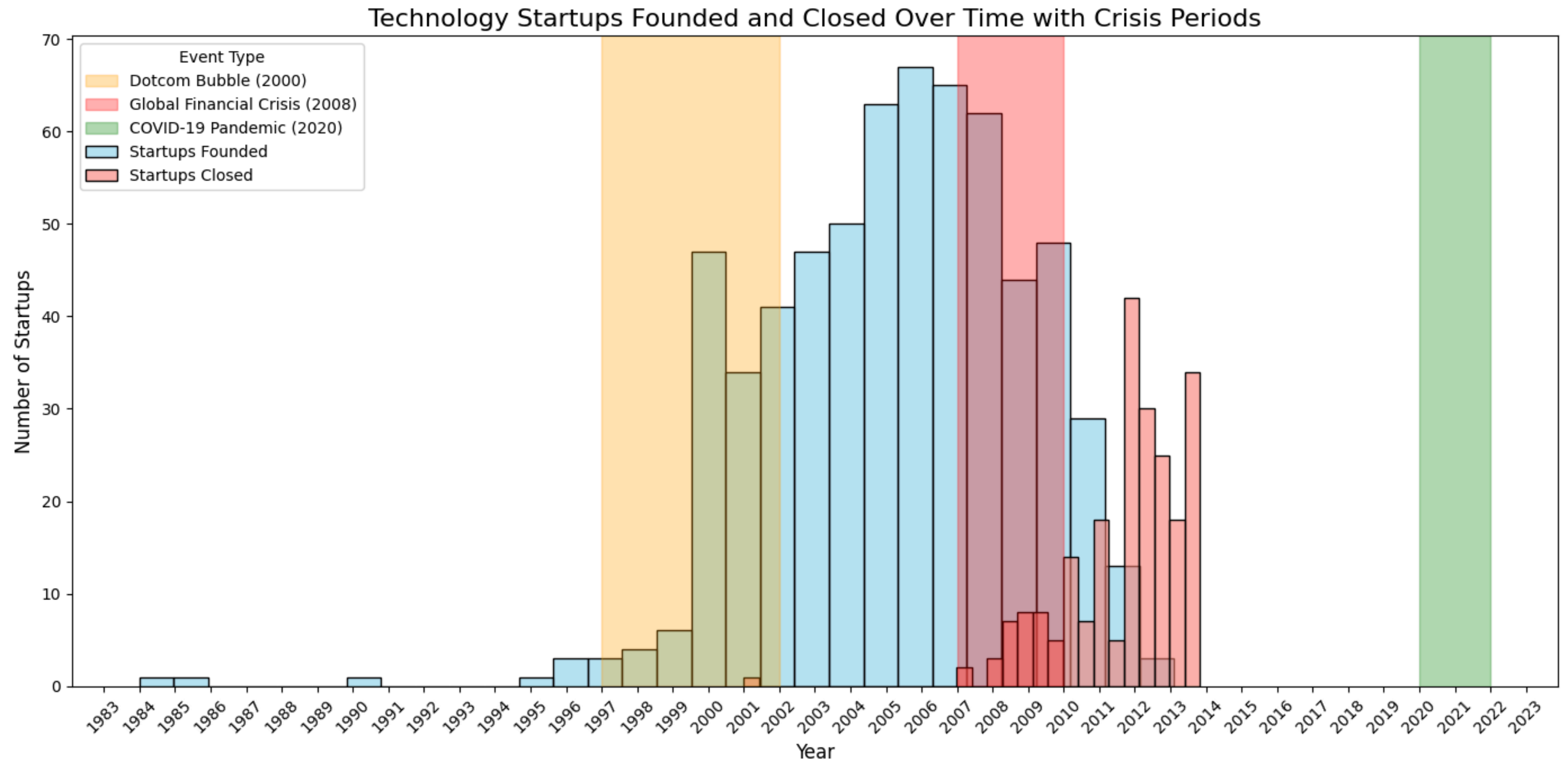
Frequency Distribution of Last Funding Year:

- Peaks in 2010 (175), 2011 (131), and 2008 (127)
- Lowest in 2001 (3) and 2004 (5)
- Noticeable drop after 2012, with only 30 in 2013



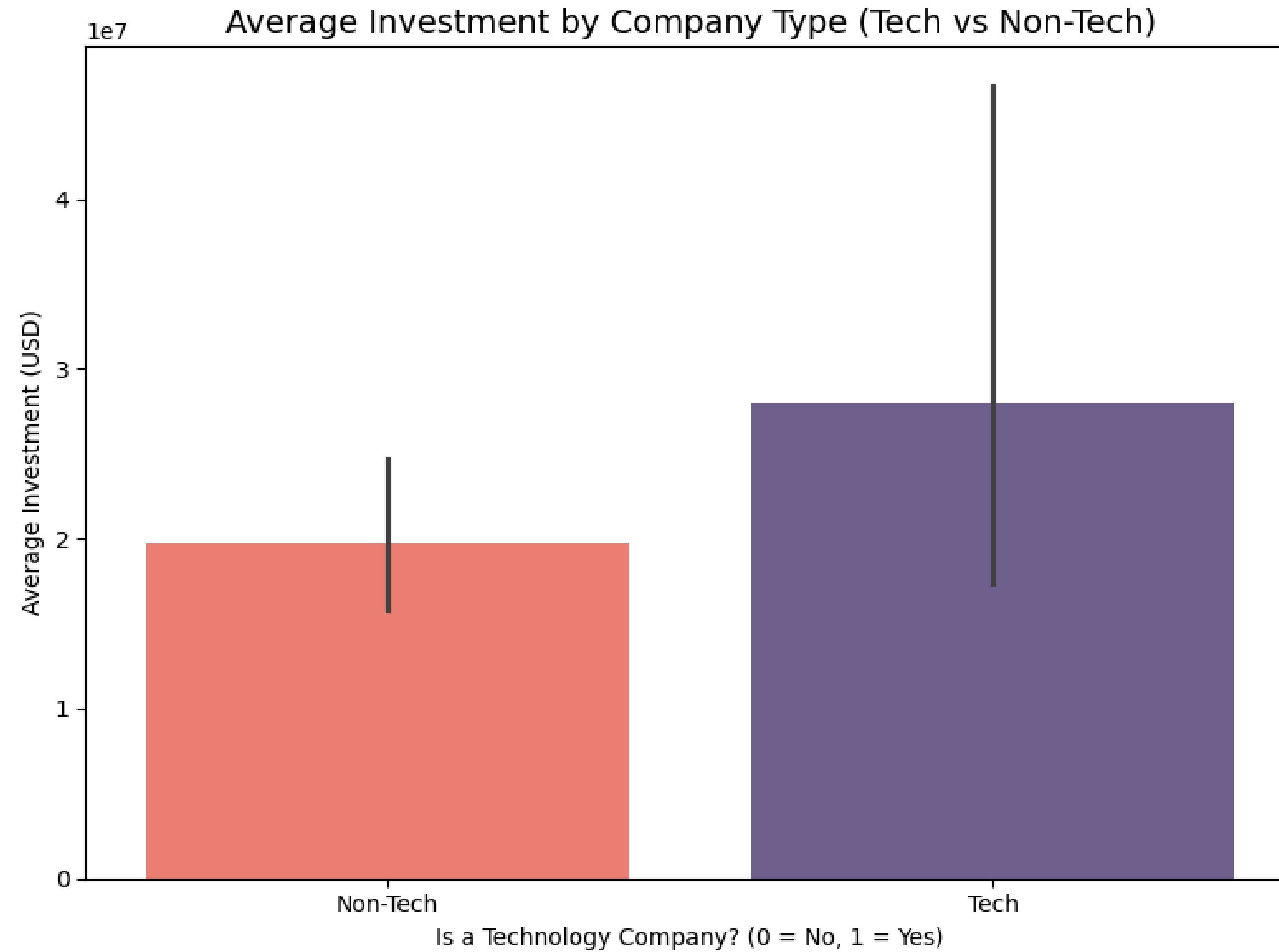


# ***MULTIVARIATE DATA ANALYSIS***

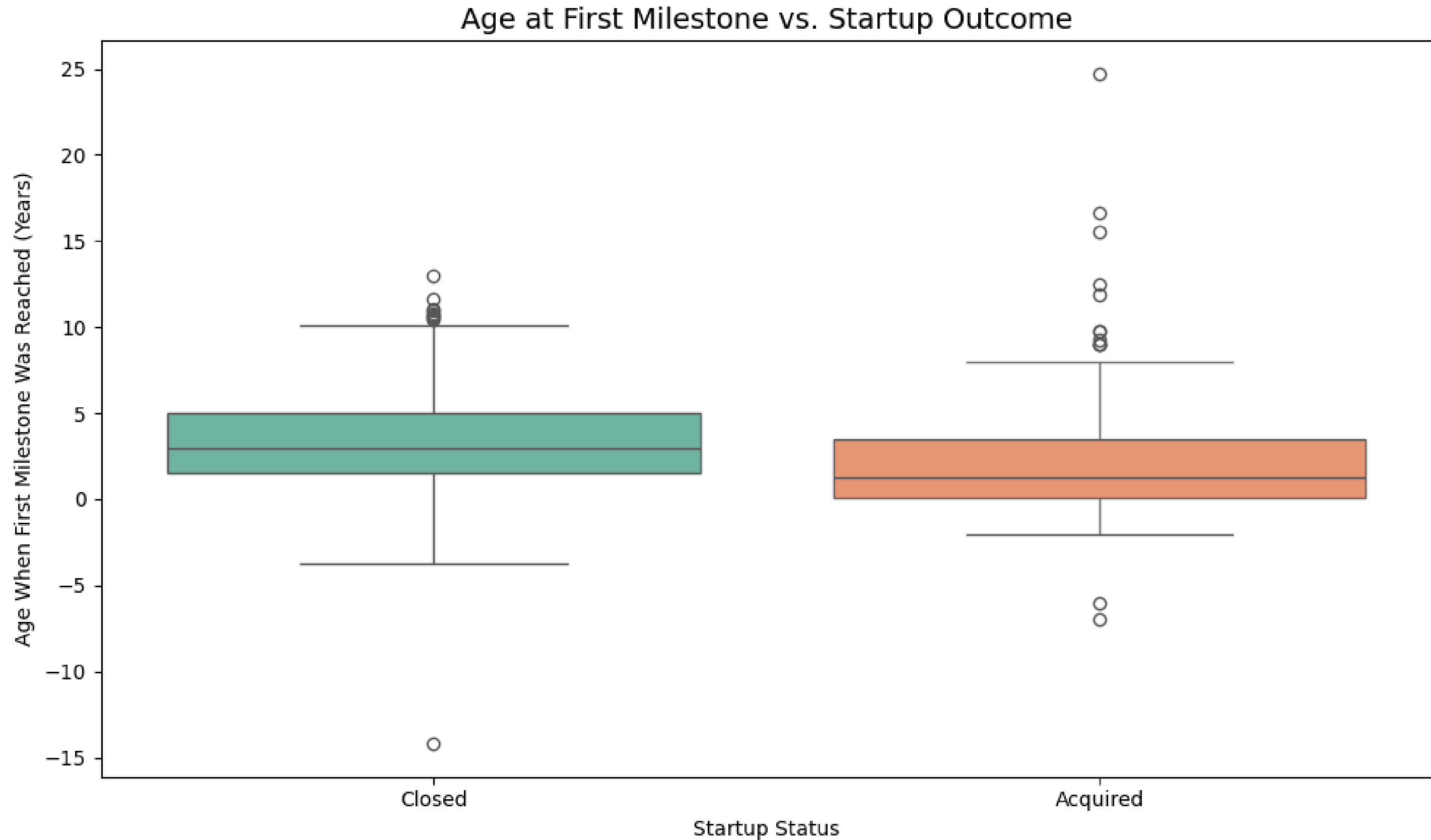


**FINDINGS:** 2008 CRISIS LED TO THE HIGHEST POST-CRISIS SHUTDOWNS, INCLUDING TECH STARTUPS.

**NOTE:** DATASET LACKS DATA AFTER 2013, LIMITING ANALYSIS OF RECENT CRISES LIKE COVID-19.



**FINDINGS:** TECH COMPANIES RECEIVED HIGHER AVERAGE FUNDING THAN THE OTHER COMPANIES



***FINDINGS:*** NO CLEAR CORRELATION FOUND—BOTH GROUPS SHOWED SIMILAR DISTRIBUTIONS. EARLY MILESTONES DON'T STRONGLY PREDICT SUCCESS.