

Trends in the Startup Ecosystem and how to maximize your Chances of Success from the Beginning according to Data

Introduction:

The Startup Economy – as it is driven by constant innovation – has been subject to change but also to growth in the past decades. As tech driven Startups are relevant employers to us driven, digitally oriented future Alumni's of Ironhack, I decided to dedicate my project to investigating past developments in the Startup ecosystem, correlations of different parameters and prospect future developments.

The following Questions and Hypothesis were targeted in the paper that follows:

General Analysis

- How did the number of companies founded by year develop over time?
- How many companies are there by sector and country?
- How was the growth of number of companies founded by sector in the past years?

Prospecting Growth and Probability

- How is the expected growth for the coming years by sector?
- What is the probability of being acquired by sector?
- What is the probability of being funded and acquired by sector?
- Which sector pays the most for acquisition?

Unicorns:

- How many Unicorns are there by sector and Country?

Hypothesis regarding Correlation:

- Hypothesis 1: There is a positive correlation between the amount of companies founded and the amount of funding by sector.
- Hypothesis 2: There is a positive correlation between the amount of companies founded and closed in a sector.
- Hypothesis 3: There is a correlation between the amount of funding and the probability of being acquired.

The Dataset:

The Data used for the Analysis was drawn from two sources. First was the Crunchbase 2013 Snapshot provided on the Developers Website in form of extensive CSV files and the second was webscrapping form CB Insights' Unicorn Tracker. Both are known as main ressources for the overview of companies and investors.

Assumptions:

- First of all, for the analysis of trends and probability, the data was limited to such gathered after 2000 to keep results unbiased from past situations.
- Second the number of companies founded in a particular sector was assumed to be an indicator for growth in demand and for innovation.
- Third it was assumed that macroeconomic conditions would remain unchanged from those throughout the past 20 years.
- Furthermore, Success needed to be defined as a KPI. See below.

Defining Success

- Success was defined as the companies not being closed.
- Second it was considered to be an indicator of success and future success if a company had received funding over time
- Being acquired or going public is a startup's goal and is therefor also condiered as being successful.

Analysis

1. General Determinations

At first the development of number of companies founded was investigated by total number, industry and country. Please see graphs below.

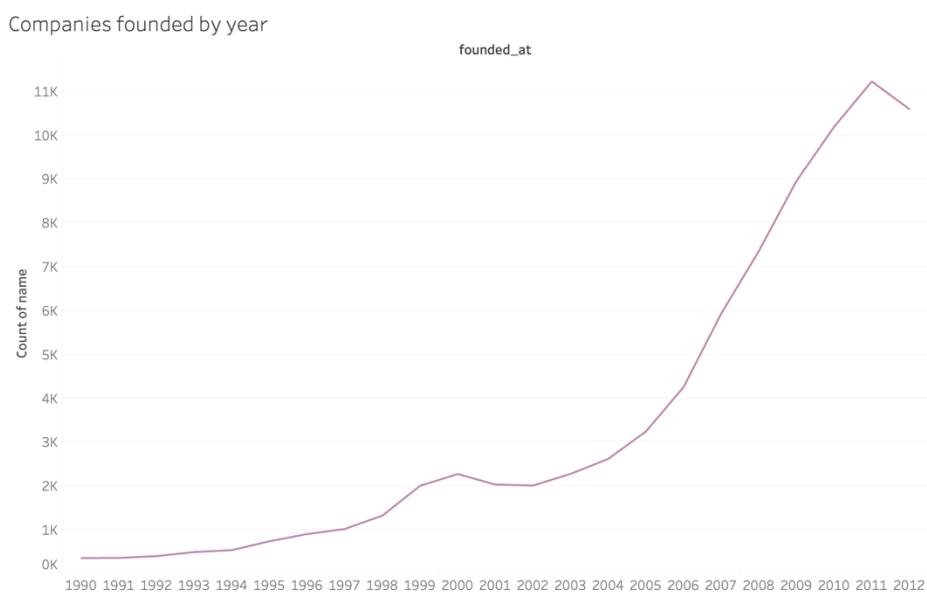


Figure 1: Development of Number in Comps. founded

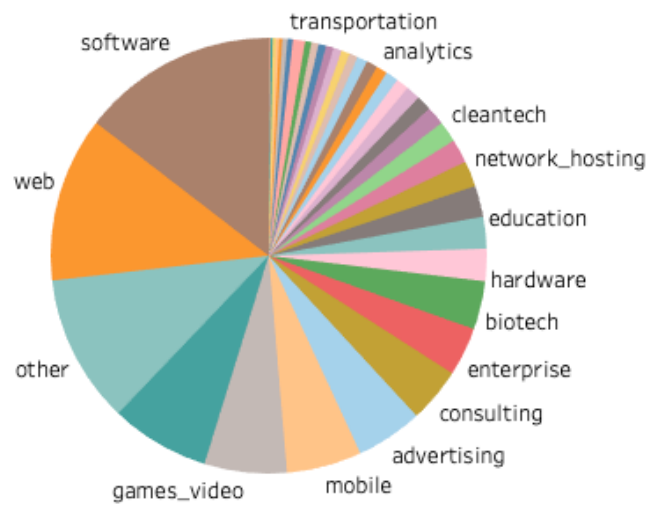


Figure 2: Number of companies by Sector

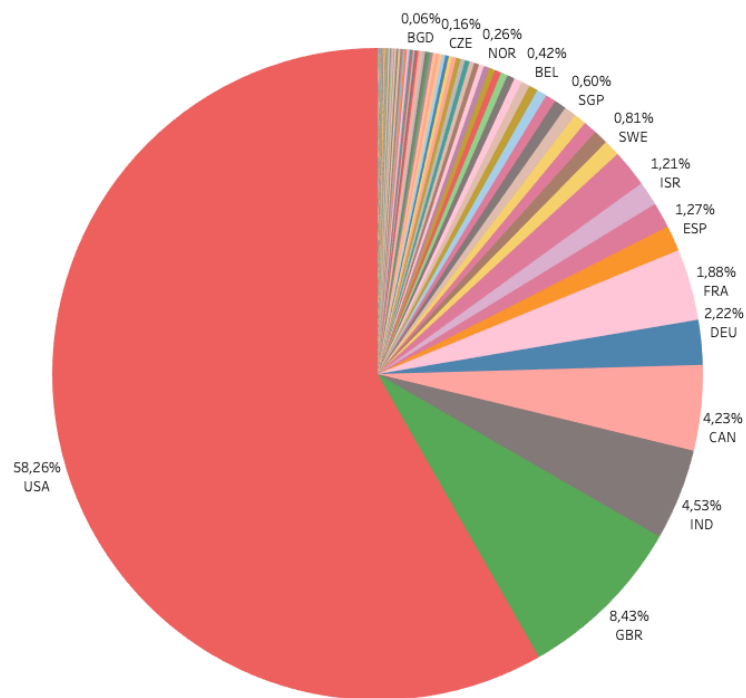


Figure 3: Number of Companies by Country

2. Growth Projections

In order to project growth, 95% confidence intervals for future rates were determined according to rates from the past years alongside the assumption that the general economic conditions would remain the same.

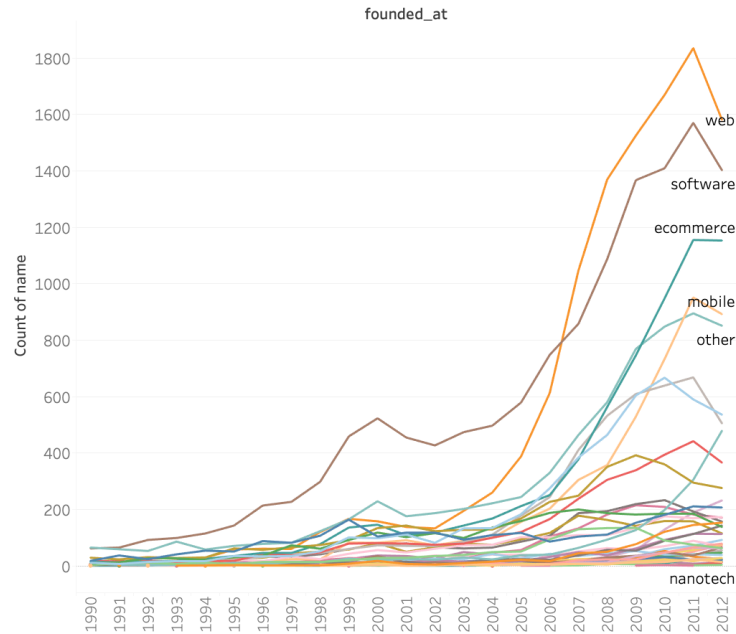


Figure 4: Development of companies by sector over time

The graph above shows the rapid growth of various sectors compared to one another. It is evident which factors have grown the most and which are rising as of end of data.

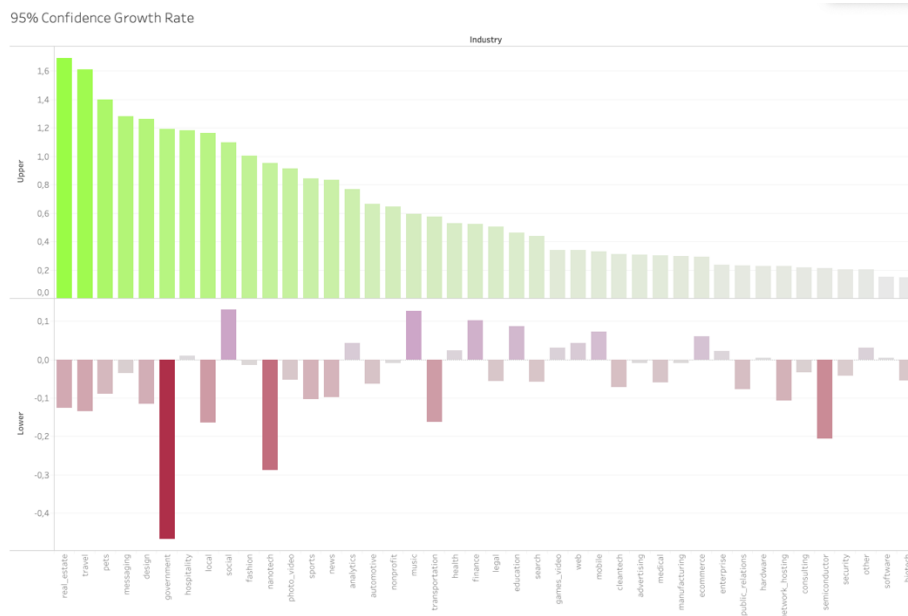


Figure 5: Diagram of possible Growth rates in a 95% CI

The graph above shows the Growth Rates within the 95% Confidence Interval by Sector.

3. Probability of Success

Firstly, the probability of not being closed as a company varies by industry and is evidently not significantly correlated with the number of companies founded. This suggests that there are different success rates in different industries. Hypothesis 2 had to be rejected.

```
In [54]: 1 sns.set(style="white", context="talk")
2 fig, ax = plt.subplots(figsize=(16,10))
3 x = comps_founded_closed_2['ratio_fc']
4 y = comps_founded_closed_2['id']
5 ax = sns.regplot(x, y,
6                 data = comps_founded_closed_2, scatter_kws = {"s": 250},
7                 marker = "+", color = 'r')
8 ax.set(xlabel = "Ratio of Companies closed in Industry (%)", ylabel = "Number of Companies in the Industry")
9 result = stats.linregress(x, y)
10 print("Slope: ", result.slope)
11 print("Intercept: ", result.intercept)
12 print("rvalue: ", result.rvalue)
13 print("pvalue: ", result.pvalue)
14 print("stderr: ", result.stderr)
```

Slope: 32808.10484195256
Intercept: 682.2024870134363
rvalue: 0.2349734639742888
pvalue: 0.13415724135698337
stderr: 21458.501552927584

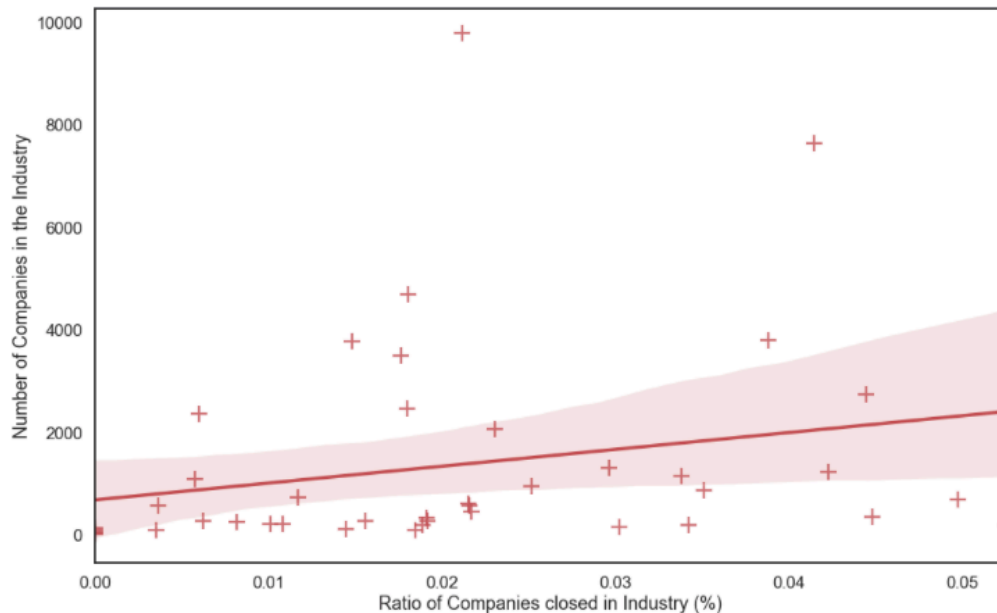


Figure 6: Ratio of Companies closed vs. Number of companies founded

Secondly, the correlation between the number of companies founded in an industry and the amount of funding, was proven to be significantly negative by p-value, which means that the more companies are founded in an industry, the harder it gets to receive funding. Therefore Hypothesis 1 had to be rejected as well.

```
In [30]: 1 sns.set(style="white", context="talk")
2 fig, ax = plt.subplots(figsize=(16,10))
3 x = funding_probability_industry['percent_funded']
4 y = funding_probability_industry['number']
5 ax = sns.regplot(x, y,
6                 data = funding_probability_industry, scatter_kws = {"s": 250},
7                 marker = "o", color = 'b')
8 ax.set(xlabel = "Percentage of Companies funded in an industry (%)", ylabel = "Number of Companies in the Indus
9 result = stats.linregress(x, y)
10 print("Slope: ", result.slope)
11 print("Intercept: ", result.intercept)
12 print("rvalue: ", result.rvalue)
13 print("pvalue: ", result.pvalue)
14 print("stderr: ", result.stderr)

Slope: -6593.698635273976
Intercept: 4855.2986355256035
rvalue: -0.3126823887315767
pvalue: 0.043786709837793286
stderr: 3167.0450601021307
```

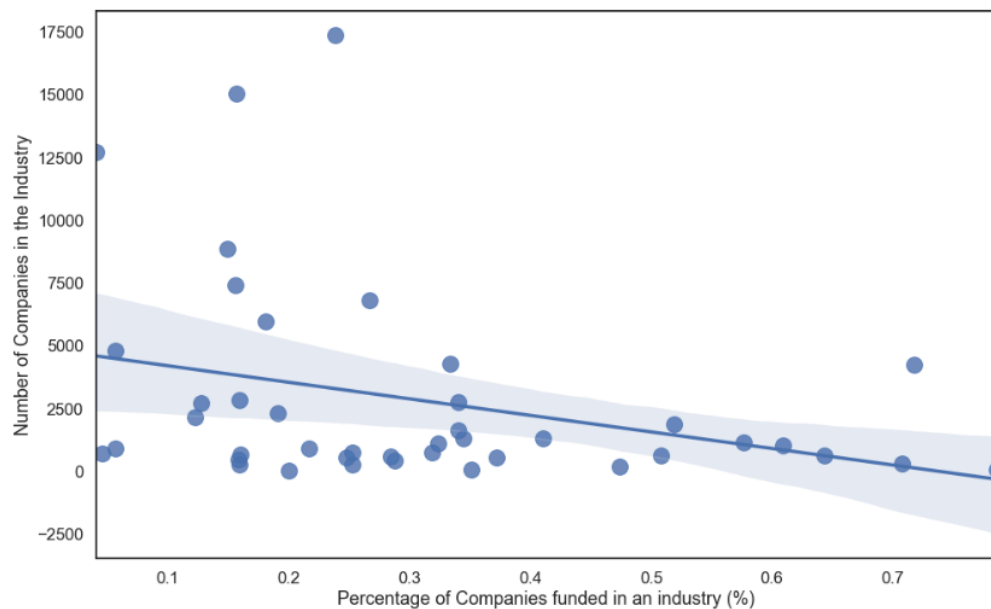
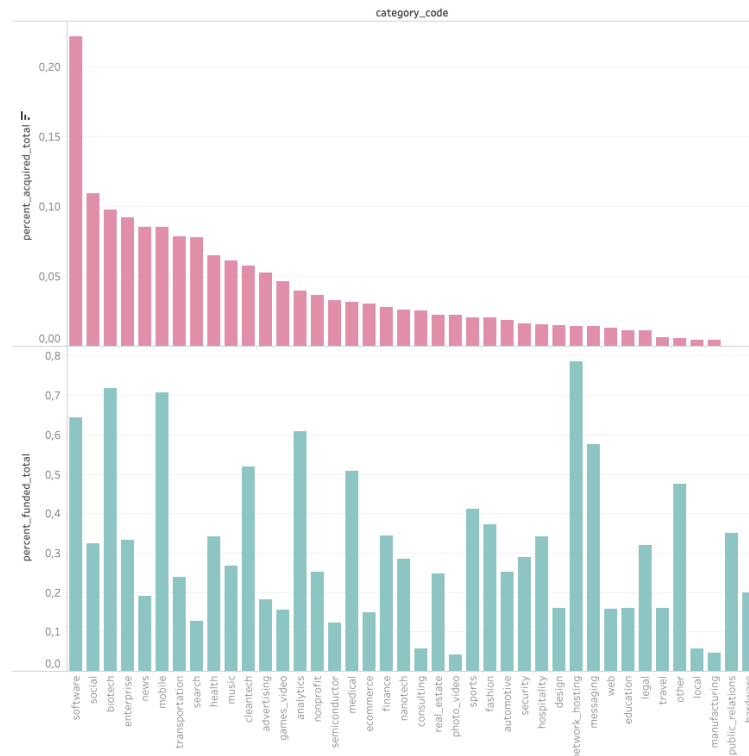


Figure 7: Number of Companies founded vs Funded

Third, Hypothesis 3 stating that the amount of funding a company receives raises its probability of being acquired had to be rejected due to an insignificant correlation according to the linear regression performed.

Probability of being Acquired by Sector



Slope: 554174783.0721252
Intercept: 991433602.7657511
rvalue: 0.295620385276551
pvalue: 0.06402363322474158
stderr: 290511185.48998517

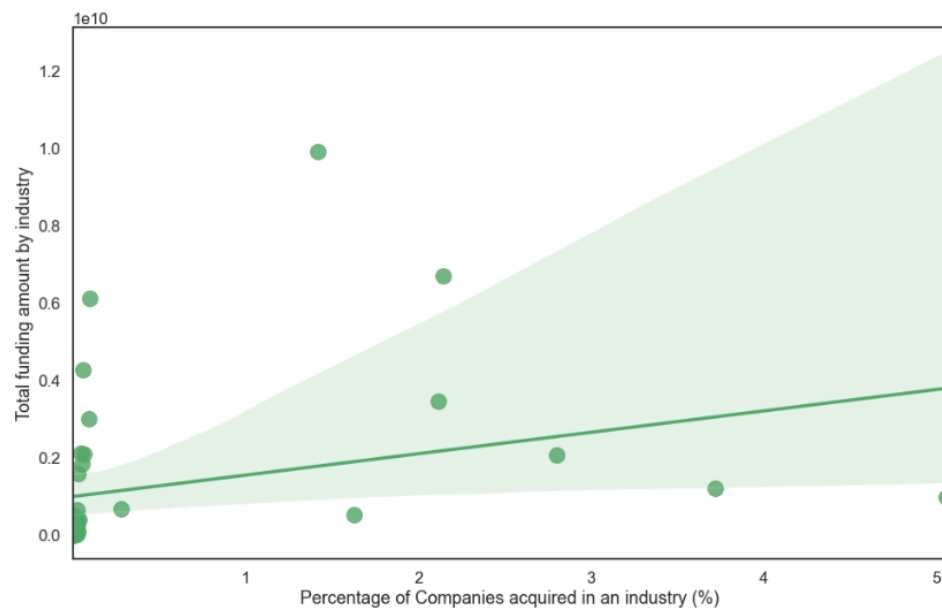


Figure 8: Total funding amount by industry vs. % of companies acquired

4. Unicorns:

Lastly the List of Unicorns was investigated by country and industry.

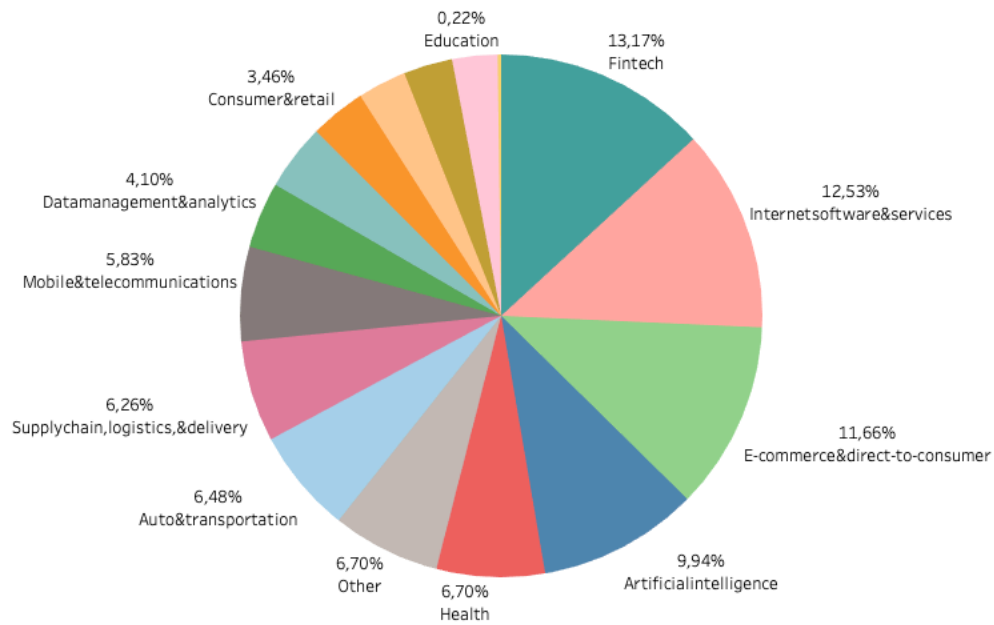


Figure 9: Unicorns by Sector

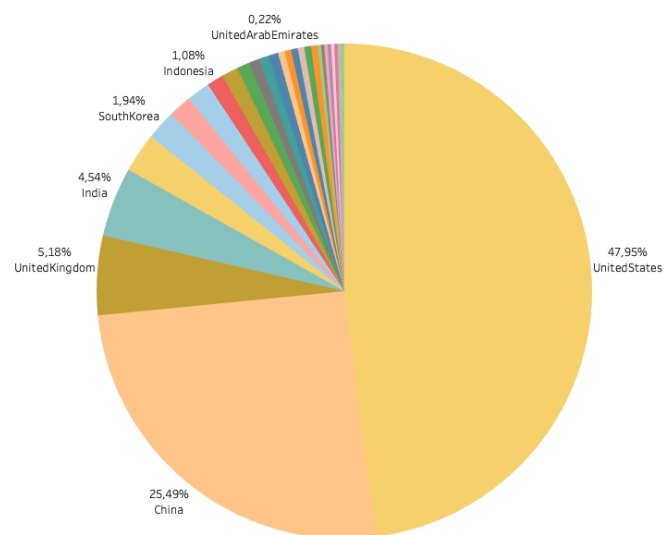


Figure 10: Unicorns by Country

Summary of the Main Insights:

- Most companies were founded in the US in Software, a trend that seems to continue.
- Sectors with the highest possible growth rates will be real estate, travel and pets.
- There is no positive correlation between the amount of companies founded and the amount of funding by sector.
- There is no positive correlation between the amount of companies founded and closed in a sector.
- There is no positive correlation between the amount of funding and the probability of being acquired.
- Most Unicorns are founded in the USA (47,95%), second is China (25,49%), followed by the UK with 5,18%.
- Most Unicorns operate in the sectors: Fintech, Ecommerce, Software and Artificial Intelligence.
- Success can be assumed most likely by founding a software / enterprise software company in the United States.

Sources:

- Crunchbase 2013 Snapshot: [Link](#)
- CB Insights Unicorn Tracker: [Link](#)