



---

# Understanding Tech Startups' Health

Yes — an algorithm to identify unicorns is here.

In 2015, we worked with The New York Times to predict 50 future unicorns (companies that would eventually be valued at \$1 billion or more).

**To date, 24 of them have hit that mark (48%).**

---

- 4 companies were acquired for \$1B+, with the largest being Ele.me, acquired for \$9.5B by Alibaba.
- 8 companies have gone public, all valued at more than \$1B. Chinese used car marketplace Uxin had the highest valuation at IPO (\$2.8B). Most recently, Elastic went public in October at a \$2.5B valuation.
- 12 companies remain private unicorns. Together, they have raised over \$8.0B in total funding.
- Of the non-unicorns (26), 14 went on to raise additional equity funding totaling \$1.4B.

At the risk of sounding immodest, that is pretty good.

And if we were a venture firm, this kind of hit rate would make us legendary.

More recently, in December 2017, we identified the Artificial Intelligence 100 — a ranking of the top 100 private AI companies. In the 11+ months since the AI 100 was unveiled, we've seen:

- 32 of the AI 100 raised additional equity financing totaling \$3.9B. SenseTime raised the most equity financing — \$1.2B since making the AI 100 list.
- 3 were acquired — the largest acquisition went to Flatiron Health, which sold to Roche Holdings for \$1.9B in Feb'18.

## How did we do it?

---

**Our secret weapon has been Mosaic** — an algorithm we've been developing for several years with initial support from the National Science Foundation.

If your work involves picking emerging tech startups in some form — whether as vendors, as partners, as investments, or as customers — it is critical you identify the best companies.

Mosaic is your objective, algorithmic, scalable way to identify these companies.

Below, we share a bit about the Mosaic scoring system, backtesting results, and some recent improvements made to the algorithm.

### **We'll cover:**

- Why we need algorithms to understand startup health
- The signals that Mosaic relies on
- The ideal distribution of Mosaic scores
- Our backtesting results
- What's next
- Our New York Times predictions

## Algorithms to understand startup health

In 2010, we approached the National Science Foundation with the idea that we could use publicly available information and a variety of non-traditional signals to assess the health of private companies.

Having worked at American Express before, our founders had seen the challenges of assessing the health of smaller private companies (aka “thin file” companies.) We believed we could use the vast amounts of unstructured and semi-structured information that is being created to better understand the health of these opaque companies.

The NSF agreed, and in 2010, it gave us \$150,000.

Here is the [CB Insights financing history from CB Insights](#) (yes, very meta.)

The screenshot shows the CB Insights interface for a company's financing history. The table below is extracted from the image:

<input type="checkbox"/>	Date	Round	Amount	Investors	Valuation	Sources
<input type="checkbox"/>	11/9/2015	Series A	\$10M	Pilot Growth Equity	\$63.08M	5
<input type="checkbox"/>	11/26/2013	Grant - III	\$0.5M	National Science Foundation		<a href="#">Deal Terms</a>
<input type="checkbox"/>	8/2/2011	Grant - II	\$0.5M	National Science Foundation		
<input type="checkbox"/>	7/22/2011	Incubator/Accelerator - II	\$0.02M	FinTech Innovation Lab		1
<input type="checkbox"/>	10/2/2010	Grant	\$0.15M	National Science Foundation		
<input type="checkbox"/>	7/22/2009	Incubator/Accelerator		NYU Future Labs		
<input type="checkbox"/>	7/14/2009	In-kind/Services		First Growth Venture Network		

Our initial traction with that first grant from the NSF resulted in 2 additional grants totaling \$1M in 2011 and 2013.

With NSF support, we worked towards a model dubbed Mosaic that would aggregate and synthesize information about these companies from disparate sources and programmatically assess the health of startups.

We believed we could make understanding and identifying the best tech startups less of a crapshoot.

Think of Mosaic as a FICO score for startups.

If we could do this with Mosaic, our belief was that capital, partnerships, talent, time, and attention would flow to the right companies, thereby minimizing misallocations of these resources.

### **So how does Mosaic work?**

## **What signals feed the Mosaic algorithm?**

---

The Mosaic score is comprised of 3 individual models – what we call the 3 M's.

- **Market** – how healthy is the industry the company is in?
- **Money** – what is the financial health of the company?
- **Momentum** – how much traction does the company have?

Each of the M's relies on different signals.

Most attempts we've seen to quantify tech company health have almost exclusively focused on momentum. This is necessary but not sufficient.

Below is a bit on each model (although all the signals utilized are not revealed for obvious reasons).

### **MARKET**

The quality of the market or industry a company competes in is critical. If you are part of a hot industry, that serves as a tailwind to push you along. Conversely, being in an out of favor space means fewer investors, partners, media, and more.

The market model looks at the number of companies in an industry, the financing and exit momentum in the space, and the overall quality and quantity of investors participating in that industry.

### **MONEY**

The money model assesses the financial health of a company, i.e. is it going to run out of money? Our model looks at burn rate, the quality of the investors and syndicate that may be part of the company, its financing position relative to industry peers & competitors, and more.

### **MOMENTUM**

The final model is momentum, where we look at a variety of volume and frequency signals including social media, news/media, sentiment, and partnership & customer momentum.

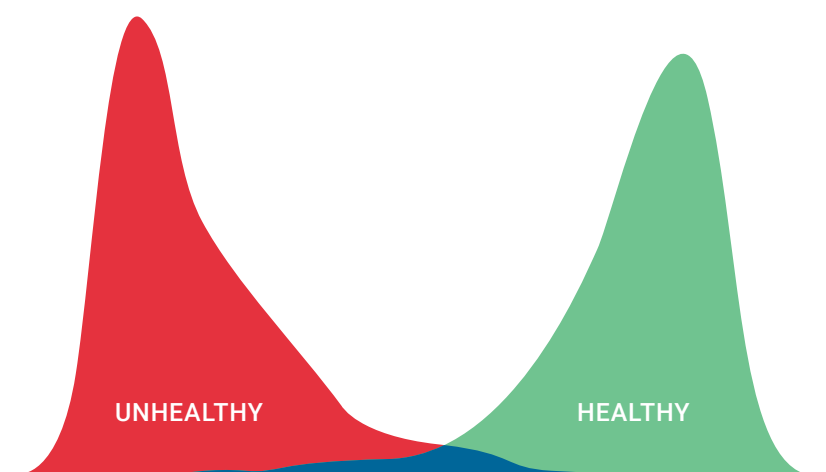
We look at these on an absolute and relative basis vs. peers/industry comparables. The relative piece is critical as it ensures that, for example, enterprise software companies who may get less media attention or who spend less time on social media are not penalized versus consumer-focused tech companies. Each of the 3M models is scored on a scale of 0-1000 and drive the overall Mosaic score (also on 1000 point scale with 1000 being the best score.)

So what would an ideal distribution of Mosaic scores look like, and how are we doing against it?

## The ideal and real distributions of Mosaic scores

---

If Mosaic was perfect, the distribution would look like:



### Ideal Distribution of Mosaic Scores

In this distribution, the Mosaic scores of healthy companies (green) would be weighted towards higher scores and the Mosaic scores of the unhealthy companies (red) would be at the lower end of the range.

However, there are challenges in getting to this theoretical perfect distribution.

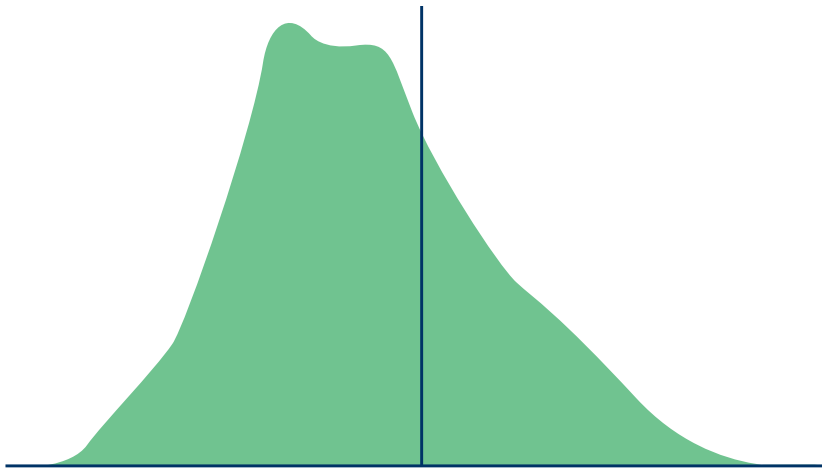
In short, there is immense fog around private tech startup companies. These are inherently opaque organisms. In some instances, they actually use media and other channels to actively obfuscate their true performance.

UNDERSTANDING THE  
HEALTH OF TECH STARTUPS

---

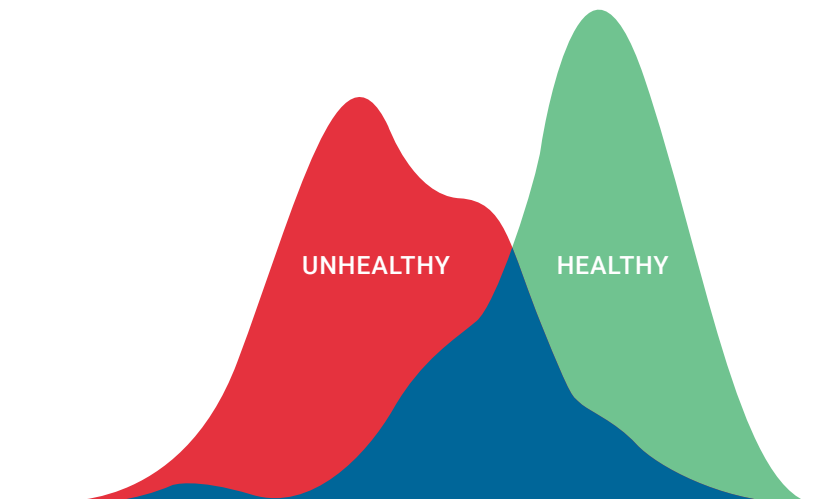
**How does Mosaic do?**

Here is the overall distribution of Mosaic scores. **The median is 420.**



**Overall Distribution of Mosaic Scores**

Here is the Mosaic score distribution for unhealthy vs. healthy companies:



**Actual Distribution of Mosaic Scores**



### How did we backtest Mosaic?

We looked at companies with positive and negative outcomes (described below) and then looked at their Mosaic score 12 months prior.

A positive outcome includes:

- Successful IPO
- Acquisition w/ valuation that is greater than last private valuation
- A \$1B+ private market valuation

A negative outcome includes:

- Bankruptcy / death
- Asset sale
- Acquisition (Talent)

The mean score for companies with positive outcomes was 740. The mean for negative outcomes was 470.

When we analyze companies with a Mosaic score  $\geq 740$ , the outcomes are as follows:

- 83% had a positive outcome
- 17% had a negative outcome

On the flip side, when we analyze companies with a Mosaic score  $\leq 470$ , the outcomes are as follows:

- 97% had a negative outcome
- 3% had a positive outcome

## What's next for Mosaic?

---

We have 4 primary areas of focus for Mosaic going forward.

**More signals:** We've introduced new data to CBI such as patents, earnings transcripts, market sizings, etc., which we believe can be integrated into the 3M models. There are also a host of other data sets that we have on our radar to capture.

**Refinement of signals:** Today, we consider business relationships (customer and partnerships) in our momentum model. With the entity extraction work we've done to extract the names of these partners/customers and track their progress, we believe a model that weighs these business relationships in a more nuanced way (based on size of partner for example) will offer additional precision to Mosaic scores.

**Gathering more data from companies themselves:** To battle the opacity challenge, we have developed and will continue to develop tools that allow companies to paint the most complete picture of themselves for Mosaic. Currently, tens of thousands of companies and investors already update their data via [The Editor](#). We will continue to improve these tools to ensure that companies are putting the most accurate view of their performance in front of investors, customers, partners, and more. (Note: We understand the challenges with company submitted data, i.e. companies don't report their bad news, and have created ways mitigate selection bias.)

**Extending to other industries:** We believe there are other sectors such as consumer goods & services that Mosaic can be extended to. In the medium-term, we believe a Mosaic score for biotech/pharma would be very compelling.

We're excited to extend the work we've done with Mosaic to demystify the health of emerging tech companies even further.

UNDERSTANDING THE  
HEALTH OF TECH STARTUPS

OUR NEW YORK TIMES 50 FUTURE UNICORNS

As mentioned, we made predictions in 2015 about 50 future unicorns.  
24 have gone on to become unicorns.

Here are the 24 unicorns we predicted in 2015 with their current valuations.



PRIVATE UNICORNS

coinbase

Valued at \$8B

OYO

Valued at \$5B

DOORDASH

Valued at \$4B

deliveroo

Valued at \$2B

GUSTO

Valued at \$2B

AVANT

Valued at \$1.9B

SQUARESPACE

Valued at \$1.7B

freshworks

Valued at \$1.5B

INTERCOM

Valued at \$1.3B

POSTMATES

Valued at \$1.2B

Thumbtack

Valued at \$1.2B

17ZUOYE

一起

Valued at \$1B

ACQUIRED

ELE.ME

饿了么

Acquired by  
Alibaba for \$9.5B

flatiron

Acquired by  
Roche Holding for \$2.1B

DOLLAR SHAVE CLUB

Acquired by  
Unilever for \$1B

PillPack

Acquired by  
Amazon for \$1B

IPO

优信  
UXIN GROUP

Jun'18 IPO  
\$2.8B valuation

elastic

Oct'18 IPO  
\$2.5B valuation

HelloFRESH

Nov'17 IPO  
\$2B valuation

Upwork™

Oct'18 IPO  
\$1.6B valuation

okta

Apr'17 IPO  
\$1.5B valuation

SONOS

Aug'18 IPO,  
\$1.5B valuation

zuora

Apr'18 IPO  
\$1.4B valuation

STITCH FIX

Nov'17 IPO  
\$1.4B valuation

## UNDERSTANDING THE HEALTH OF TECH STARTUPS

The remaining 26 companies are listed below.

### **Betterment**

\$712M valuation  
raised \$170M since  
NYT-CBI unicorn prediction

### **Collective Health**

raised \$191M since  
NYT-CBI unicorn prediction

### **DigitalOcean**

raised \$130M since  
NYT-CBI unicorn prediction

### **eDaijia**

### **Greenhouse Software**

raised \$85M since  
NYT-CBI unicorn prediction

### **Grofers**

raised \$181M since  
NYT-CBI unicorn prediction

### **Harry's Razor Company**

raised \$112M since  
NYT-CBI unicorn prediction

### **HotelTonight**

raised \$37M since  
NYT-CBI unicorn prediction

### **Mixpanel**

### **One Medical Group**

raised \$285M since  
NYT-CBI unicorn prediction

### **Optimizely**

raised \$58M since  
NYT-CBI unicorn prediction

### **Planet Labs**

### **Privia Health**

### **Raise Marketplace**

raised \$60M since  
NYT-CBI unicorn prediction

### **Skyhigh Networks**

acquired by McAfee for \$720M

### **Stack Exchange**

### **Taboola**

raised a total of \$160M

### **Takealot Online**

majority stake by  
Naspers in April 2017

### **Wealthfront**

valuation \$500M  
raised \$75M in Jan 2018

*We also made a handful of bad predictions consisting of companies that died, have underperformed or which had major valuation down rounds.*

*These include 3D Robotics, Airware, Beepi, Cyanogen, Munchery, Practice Fusion, and Rethink Robotics.*



**If your job involves picking startups  
– whether as partners, investments,  
vendors or acquisitions – it’s time  
to tilt the odds in your favor.**

---

[CLICK HERE TO SIGN UP FOR FREE](#)