

# UNICORN STARTUP ANALYSIS ETL PROJECT

Rathin Patel

## PROBLEM/OBJECTIVE

A common phrase in the investing world is to “follow the money”. Although I’m not sure exactly what that means I like the sound of it. So I thought a great project I could work on to practice and improve my data analysis skills would be to **analyze the top unicorn companies in the world**. Unicorn companies are start-up companies with a value of over \$1 billion. For perspective, Google, Facebook, and Airbnb were once unicorn companies. Thus, I want to get some market insights and see what companies in the future will have that revolutionary impact on the world by “following the money”. I’d like to know which countries are leading in innovation and see where the biggest innovation hubs are in the world as well.

Ever since OpenAI released Chat-GPT 3, the most popular consumer-facing AI application to date, all I’ve been seeing on my social feeds is “AI is going to take over” and “AI will take all our jobs” and while I know a lot of it is just engagement bait, there is definitely some truth to how transformative AI will become to our daily lives and the way we do things. One of my curiosities I’d like to get some insights on upon analyzing the top unicorn companies is how many are in the AI industry, and what industries in general comprise the top unicorn companies. These are the areas that are being funded and where we likely see growth and continuous investment in the future. Best case scenario, I may even find some stock picks or industries to look into for long-term holds.

As I’ve had a decent amount of practice with exploratory data analysis and making visuals with R and Python libraries, I wanted to create an end-to-end project following the **ETL (extract, transform, load)** process. I particularly wanted to learn the ‘extract’ part the most, as when you’re working with data most times in the real world it’s not provided in a CSV by Kaggle or your instructor, rather I would have to learn how to collect the data myself.

**Extract:** Web scraping websites to collect data for relevant information

**Transform:** Transforming the extracted data by storing it in a structured format (SQLite database). Cleaning, structuring, and organizing data to prepare for analysis.

**Load:** Load data into Tableau to create visualizations.

## DATA COLLECTION

Web scraped the table from <https://www.cbinsights.com/research-unicorn-companies> composed of nearly 1200 companies. CBInsights is a reputable source thus I know I’m collecting high-quality data. Data includes Company name, Valuation (in billions), Date joined, country, they are from, city they are in, the industry they are in, and select investors in the company. What I liked about this source is they included the big-name investors in the companies. This additional

variable can be used to get insights into who the big players invested into the companies of the future and funding them.

Web scraped the table

from [https://en.wikipedia.org/wiki/List\\_of\\_unicorn\\_startup\\_companies](https://en.wikipedia.org/wiki/List_of_unicorn_startup_companies) originally for the data source, but wanted a higher-quality data source so I used the one from CBinsights. However, one thing I liked about the Wikipedia table was it included the industry more specifically than CBinsights did. For example, for anything tech related it was classified as “enterprise tech” on CBinsights, whereas Wikipedia for the same company will be more specific, specifying if it was an Artificial Intelligence company, cybersecurity, cryptocurrency etc.

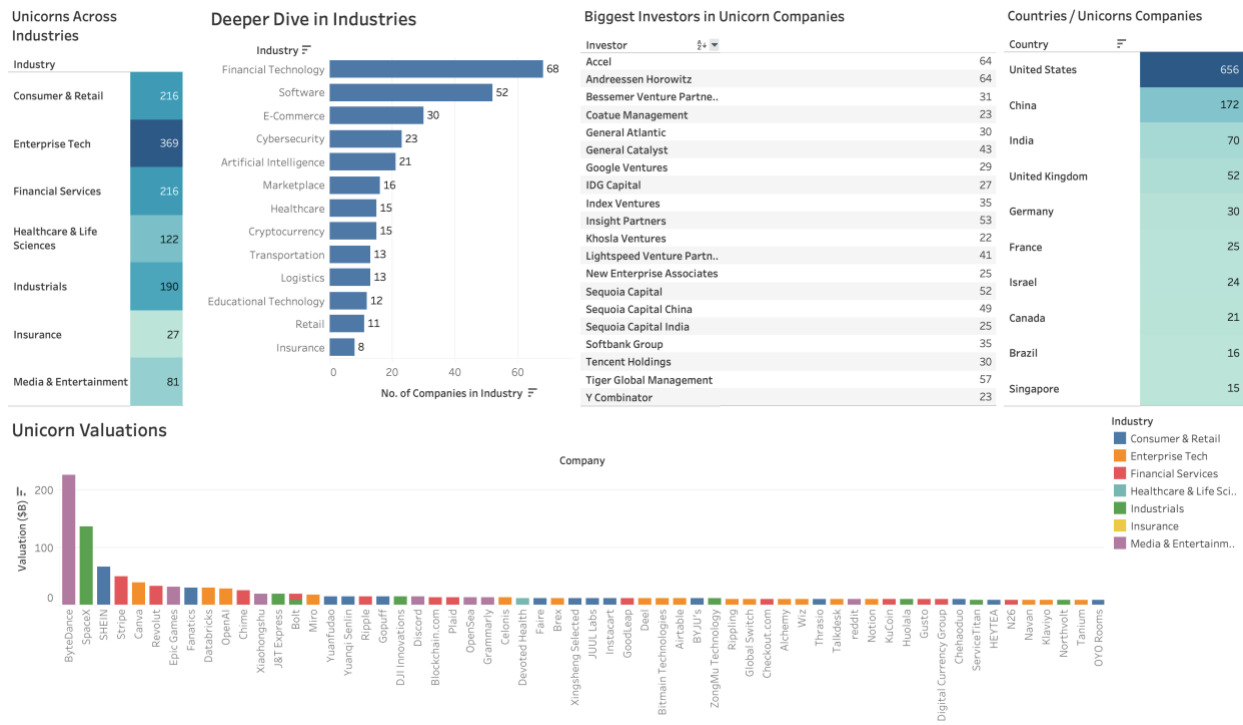
Thus, both datasets were scraped, and stored into SQLite databases to be queried.

## **EXPLORATORY DATA ANALYSIS**

SQL Queries were used to solve the following questions:

- The top 10 countries with the most unicorn companies
- Number of companies that were in the 7 distinct industries from the CB insights scrape
- Number of companies in the more specific industries from the Wikipedia scrape (deeper dive into industries)
- Top 50 most common investors found in the ‘select investors’ column in the CB insights table, and how many unicorn companies they are invested in
- Valuations of all the companies (more for data vis purposes)

# DATA VISUALIZATION



(interactive version on tableau public)

[https://public.tableau.com/app/profile/rathin.patel7363/viz/Unicorn\\_Startup\\_Analysis/Dashboard1?publish=yes](https://public.tableau.com/app/profile/rathin.patel7363/viz/Unicorn_Startup_Analysis/Dashboard1?publish=yes)

## DATA ANALYSIS/DISCUSSION

Upon collecting my data, storing it, exploring it with SQL queries, and finally making visualizations with my queries I learned a decent amount about the landscape of the unicorn companies in the world currently.

While some things did slightly surprise me (not as many unicorns focusing on AI as I thought), a lot of things didn't. I wasn't that surprised ultimately because the industries with more unicorns (financial technology, software, e-commerce, and cybersecurity) have been the biggest industries in the last couple of decades and they will still be very important areas whether AI has a revolutionary impact on them or not. AI-focused unicorn companies may not be as high as in some other industries as AI technology is complex and constantly evolving and achieving market traction and demonstrating long-term value can take time. A lot of investment is required. Data availability and quality is also a huge inhibitor to AI growth as getting access to the right data can be challenging, especially in industries with limited digitization or privacy concerns. Startups in the AI space also need to navigate regulatory and ethical challenges while developing their technology and business models. Additionally, we can't forget about the competition from

established players, as they are likely doing some of the most important work in AI right now due to their resources and talent, like Meta, Google, Amazon, and Microsoft.

I wasn't surprised the United States has the most unicorns by a staggering lead, and that the second and third countries with the most unicorns are behind by a lot (China and India).

As for learning about the biggest investors in these unicorn companies, I will have to study some of their holdings. I can "follow the money" and hopefully get some insights into where I should consider investing. At the minimum, I know where some of the most resourceful and highly capable investment firms are placing their bets.

One thing I did find interesting about the valuations, specifically the top companies, is how simple the problems they are solving and what they're focused on. This made me remember something I learned in the past which is working on basic ideas/human experiences but doing them exceptionally well. For example, ByteDance is in the business of social media/entertainment (a very simple use case of entertaining people, connecting people) but they are the highest valued unicorn. Whereas I would have thought the top 10 would have comprised of more companies in AI, space exploration, or some of the other areas of interest we don't have a lot of advancement in (funnily SpaceX, a space exploration company, is number 2 on the list). Similarly, SHEIN (number 3) is just a consumer and retail platform for people to purchase clothes, again a very simple use case, but done exceptionally well and with a refreshing business model. As well as Stripe (a relatively simple use case for handling financial transactions) but done really well. Thus, I'd have imagined the top 10 unicorn companies were more similar to SpaceX and other areas of interest we don't have a lot of advancement in, but rather its companies working on simpler use cases on top of utilizing some of the bigger technological breakthroughs of the last couple of decades such as cloud computing, how fast we can move data, and the rise of e-commerce.