Unicorn Boom: Valuation Trends & Emerging Industries in Global Startups, From 2019 - 2021

Anesu David Pasipanodya

All the analysis are based on data in the project and the opinions of the author as of the 22nd of July 2024

<https://github.com/AnesuorDavid/AnesuorDavid.github.io>

Abstract

This project investigates the global landscape of unicorn companies—privately held startups valued at over $1 billion. Leveraging a MySQL database, the project analyses trends in industry valuation, identifies sectors with the highest average company value, and tracks the emergence of new unicorns from 2019 to 2021. By examining valuation patterns and temporal growth, the analysis offers insights into which industries are thriving and where potential investment opportunities may lie.



Contents

[Abstract 2](#_Toc204211941)

[Contents 3](#_Toc204211942)

[Introduction 4](#_Toc204211943)

[Project Overview 5](#_Toc204211944)

[About the dataset 6](#_Toc204211945)

[Data Uploading 7](#_Toc204211946)

[Data Cleaning and Normalisation 9](#_Toc204211947)

[Data Analysis through Aggregation 11](#_Toc204211948)

Introduction

The term *unicorn*—coined to represent the rarity of startups valued at $1 billion or more—has become increasingly common as global venture capital flows surged over the past decade. The rise of platforms such as Robinhood, eToro, and public investment tools has enabled a new wave of amateur investors eager to capitalize on high-growth companies. Simultaneously, market disruptions, technological innovation, and global digitalization have contributed to a record number of unicorns, particularly in the technology and fintech sectors.

This project aims to uncover trends within this phenomenon by analysing a MySQLdatabase of unicorn startups. By focusing on company valuations across industries and tracking their emergence over time, especially between 2019 and 2021, we seek to identify which sectors are experiencing the most growth and which regions and years contributed most to this boom. This exploration not only highlights market dynamics but also provides strategic insight into where the next wave of investment opportunities may emerge.

Project Overview

In this project, we analyze the world’s most valuable startups—unicorn companies, defined as private firms valued at over $1 billion USD. The objective is to understand how these companies are distributed across industries, countries, and time. This analysis will uncover investment trends, growth sectors, and regional startup ecosystems driving global innovation.

The data is first cleaned and structured in MySQL, allowing for filtering, joining, and aggregation across tables. Excel is used for initial exploration and pivot table prototyping. Tableau provides rich visualizations, such as trend lines of unicorn emergence, industry-based heat maps, and valuation dashboards, which communicate insights to both technical and non-technical audiences.

Together, these tools enable a full-cycle data analysis that not only identifies unicorn growth but also reveals which industries and regions are likely to produce the next wave of billion-dollar startups.

**Key Questions to be Answered via Aggregation of the dataset**:

1. Which industries have the highest total and average unicorn valuations, and how concentrated are those valuations among a few companies?
2. How many unicorn companies were created each year between 2019 and 2021, and how does that trend differ across the top 5 industries?
3. Which countries and continents have the most unicorns, and how do their average valuations compare?
4. What is the average time (in years) from founding to unicorn status, and which industries or countries produce faster unicorns?
5. Which cities host the highest number of unicorns, and are these companies clustered in similar industries?
6. Which investors most frequently appear among the top-valued unicorns, and in what industries are they investing?
7. How do valuation-to-funding ratios differ across industries and countries, and what does that say about investment efficiency?
8. Which industries saw the greatest number of new unicorns join the list in 2020 and 2021, and what global factors might have driven this growth?
9. What is the average funding raised per unicorn per country, and how does that relate to the country’s total number of unicorns?
10. Are newer unicorns (post-2020) valued higher or lower than earlier unicorns, and does this trend vary by industry?

About the dataset

The dataset used for this project originates from DataCamp’s Datalabs and focuses on ‘unicorn companies’ —private startups valued at over $1 billion. It is composed of four relational tables, each capturing different dimensions of unicorn data.

In this project the data is organized into four structured tables:

* companies: Includes basic information such as company\_id, company name, city, country, and continent.
* industries: Lists the industry each company operates in, connected via company\_id.
* funding: Contains financial data, including valuation, funding amount, and notable investors.
* dates: Captures date\_joined (when the company reached unicorn status) and year\_founded.

The **companies** table contains general information such as company name, location (city, country, continent), and a unique company\_id serving as the primary key. The **industries** table maps each company\_id to its corresponding industry sector. The **funding** table includes financial details, providing company valuations, total funding raised, and a list of notable investors. The **dates** table records each company’s founding year and the date it officially became a unicorn, enabling time-based analysis. Together, these tables form a comprehensive and analyzable structure ideal for understanding startup growth and investment trends globally.

To prepare the data for analysis, the four CSV files were imported into **MySQL**, where appropriate foreign key relationships and indexing were applied to support structured querying. From there, relevant query outputs and aggregated results were exported into **Microsoft Excel** for basic exploration and formatting. Then **Tableau** was used for interactive data visualization, enabling the creation of dashboards to display metrics such as valuation trends, top industries, funding efficiency, and annual unicorn emergence by geography. Tableau’s drag-and-drop interface allowed for seamless integration with Excel and SQL outputs, helping transform raw unicorn data into meaningful business insights.

Data Uploading

The 4 datasets were uploaded into the schema, where I added the data to the unicorns\_dataset schema.

|  |
| --- |
| The unicorns\_dataset with all the data |

The tables were then checked to see if it was all present.

**QUERY**:

|  |
| --- |
| SELECT \* -- COMPANIES TABLE  FROM companies  LIMIT 5;  SELECT \* -- DATES TABLE  FROM dates  LIMIT 5;  SELECT \* -- FUNDING TABLE  FROM funding  LIMIT 5;  SELECT \* -- INDUSRIES TABLE  FROM industries  LIMIT 5; |

**ANSWER**:

|  |
| --- |
| Screenshots of all the tables |

Data Cleaning and Normalisation

To begin analysing the data effectively, the schema into a usable structure. Since the dataset is distributed across four separate tables — companies, industries, funding, and dates — each table will initially be uploaded into MySQL as its own staging table. These tables share a common column: company\_id, which acts as the **primary key**. This common identifier allows us to establish relationships between the tables, forming a **relational database schema**. After staging, the data will be analysed both by using the aggregating from these individual tables, and consolidated by joining all four tables on the company\_id column. This creates a unified staging table suitable for analysis. This process ensures data integrity and enables smooth interaction with visualization tools like Tableau and Excel for deeper analytical insights.

**QUERY FOR STAGING TABLE DATA**

|  |
| --- |
| -- DATA CLEANING AND-NORMALISATION  -- CREATING STAGING TABLES  CREATE TABLE staging\_companies -- COMPANIES  LIKE companies;  INSERT INTO staging\_companies  SELECT \*  FROM companies;  SELECT \* FROM staging\_companies;  CREATE TABLE staging\_dates -- DATES  LIKE dates;  INSERT INTO staging\_dates  SELECT \*  FROM dates;  SELECT \* FROM staging\_dates;  CREATE TABLE stage\_funding -- FUNDING  LIKE funding;  INSERT INTO stage\_funding  SELECT \*  FROM funding;  SELECT \* FROM stage\_funding;  CREATE TABLE stage\_industry -- INDUSTRY  LIKE industries;  INSERT INTO stage\_industry  SELECT \*  FROM industries;  SELECT \* FROM stage\_industry; |

CREATING A RDB by JOINING TABLES THROUGH SQL

QUERY

|  |
| --- |
| -- CREATING A JOINED TABLE CALLED UNICORN MASTER  CREATE TABLE unicorn\_master  SELECT  c.company\_id,  c.company AS company\_name,  c.city,  c.country,  c.continent,  d.date\_joined,  d.year\_founded,  f.valuation,  f.funding,  i.industry  FROM staging\_companies c  JOIN staging\_dates d ON c.company\_id = d.company\_id  JOIN stage\_funding f ON c.company\_id = f.company\_id  JOIN stage\_industry i ON c.company\_id = i.company\_id;  SELECT \*  FROM unicorn\_master LIMIT 5; |

**ANSWER**:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| company\_id | company\_name | city | country | continent | date\_joined | year\_founded | valuation/$ | funding |
| 189 | Otto Bock HealthCare | Duderstadt | Germany | Europe | 2017-06-24T00:00:00.000 | 1919 | 4billion | 0 |
| 848 | Matrixport |  | Singapore | Asia | 2021-06-01T00:00:00.000 | 2019 | 1billionn | 100million |
| 556 | Cloudinary | Santa Clara | United States | North America | 2022-02-15T00:00:00.000 | 2011 | 2billion | 100million |
| 999 | PLACE | Bellingham | United States | North America | 2021-11-17T00:00:00.000 | 2020 | 1billion | 100million |
| 396 | candy.com | New York | United States | North America | 2021-10-21T00:00:00.000 | 2021 | 2billion | 100million |

\*

Data Analysis through Aggregation

1. *Which industries have the highest total and average unicorn valuations, and how concentrated are those valuations among a few companies?*

To analyse the highest total and the highest average valuations the unicorn\_master table, which is a staging table with an amalgamation of all the data from the 4 tables, will be used to view the data. From here the average and highest total unicorn valuations can be observed and analysed.

**QUERY**:

|  |
| --- |
| -- 1. Which industries have the highest total and average unicorn valuations, and how concentrated are those valuations among a few companies?  SELECT **industry**, SUM(funding) AS **total\_funding**, SUM(valuation) **AS** total\_valuation  FROM unicorn\_master  GROUP BY **industry**  ORDER BY **industry**; |

**ANSWER**:

Order by Average\_Valuation

|  |  |  |
| --- | --- | --- |
| **Industry** | **Total Valuation** | **Average Valuation** |
| Fintech | 836 billion | 3.75 billion |
| E-commerce & direct-to-consumer | 362 billion | 3.38 billion |
| Internet software & services | 595 billion | 2.90 billion |
| Auto & transportation | 81 billion | 2.89 billion |
| Consumer & retail | 68 billion | 2.83 billion |
| Travel | 36 billion | 2.77 billion |
| Hardware | 86 billion | 2.69 billion |
| Health | 198 billion | 2.68 billion |
| Cybersecurity | 129 billion | 2.58 billion |
| Data management & analytics | 98 billion | 2.45 billion |
| Mobile & telecommunications | 89 billion | 2.34 billion |
| Artificial intelligence | 185 billion | 2.26 billion |
| Edtech | 51 billion | 2.04 billion |
| Supply chain, logistics, & delivery | 107 billion | 1.98 billion |
| Other | 108 billion | 1.96 billion |

Order By Total\_Valuation

|  |  |  |
| --- | --- | --- |
| **Industry** | **Total Valuation** | **Average Valuation** |
| Fintech | 836 billion | 3.75 billion |
| Internet software & services | 595 billion | 2.90 billion |
| E-commerce & direct-to-consumer | 362 billion | 3.38 billion |
| Health | 198 billion | 2.68 billion |
| Artificial intelligence | 185 billion | 2.26 billion |
| Cybersecurity | 129 billion | 2.58 billion |
| Other | 108 billion | 1.96 billion |
| Supply chain, logistics, & delivery | 107 billion | 1.98 billion |
| Data management & analytics | 98 billion | 2.45 billion |
| Mobile & telecommunications | 89 billion | 2.34 billion |
| Hardware | 86 billion | 2.69 billion |
| Auto & transportation | 81 billion | 2.89 billion |
| Consumer & retail | 68 billion | 2.83 billion |
| Edtech | 51 billion | 2.04 billion |
| Travel | 36 billion | 2.77 billion |

By just observing the data it’s clear to see Fintech has the highest total valuation, and average valuation, at 836 billion and 3.75 billion respectfully. This shows how the rise of electronic financial services during covid boosted the market and its valuation, especially during the uncertainty of the Covid-19 outbreak and subsequent lockdowns. This is probably why the next 2 largest total valuations both in total and on average are Ecommerce and Internet and software services. This is a clear sign of the rapid adaptation by all the industries to adapt to the changing environment of the time to stay afloat. A telltale sign that this is due to the pandemic is the fact that healthcare is 4th in both valuations, showing the large investment most likely made by governments and private entities in order to combat the spread of Covid 19. This can also be seen in the fact that Travel as an industry had the lowest average valuation. It can simply be attributed to the travel restrictions in place at the time and the low confidence this may have caused in the profut generating capability of the industry at the time. One thing to note is the avg. evaluation of AI companies, as they place number 4 in this category. It is an indicator of the steady growth of this industry, knowing what we do as of the report analysis, where AI companies today are some of the most highly valued in all tech industries.

1. *How many unicorn companies were created each year between 2019 and 2021, and how does that trend differ across the top 5 industries?*

To get this information we have to aggregate the top data from 2019 to 2021, finding the number.

**QUERY**

|  |
| --- |
| SELECT year\_founded, COUNT(company\_name) AS created\_entities  FROM unicorn\_master  WHERE year\_founded IN ('2019', '2020', '2021')  GROUP BY year\_founded  ORDER BY year\_founded; |

**ANSWER**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | year\_founded | created\_entities | | 2019 | 45 | | 2020 | 25 | | 2021 | 11 | |

Then we have to find how many were created in each industry. To do this we have to find the industries with the largest amount unicorn companies.

**QUERY**:

|  |
| --- |
| -- Count per industry  SELECT industry, COUNT(\*) AS total\_unicorn  FROM unicorn\_master  GROUP BY industry  ORDER BY total\_unicorn DESC  LIMIT 5; |

**RESULT:**

|  |  |
| --- | --- |
| industry | total\_unicorn |
| Fintech | 223 |
| Internet software & services | 205 |
| E-commerce & direct-to-consumer | 107 |
| Artificial intelligence | 82 |
| Health | 74 |

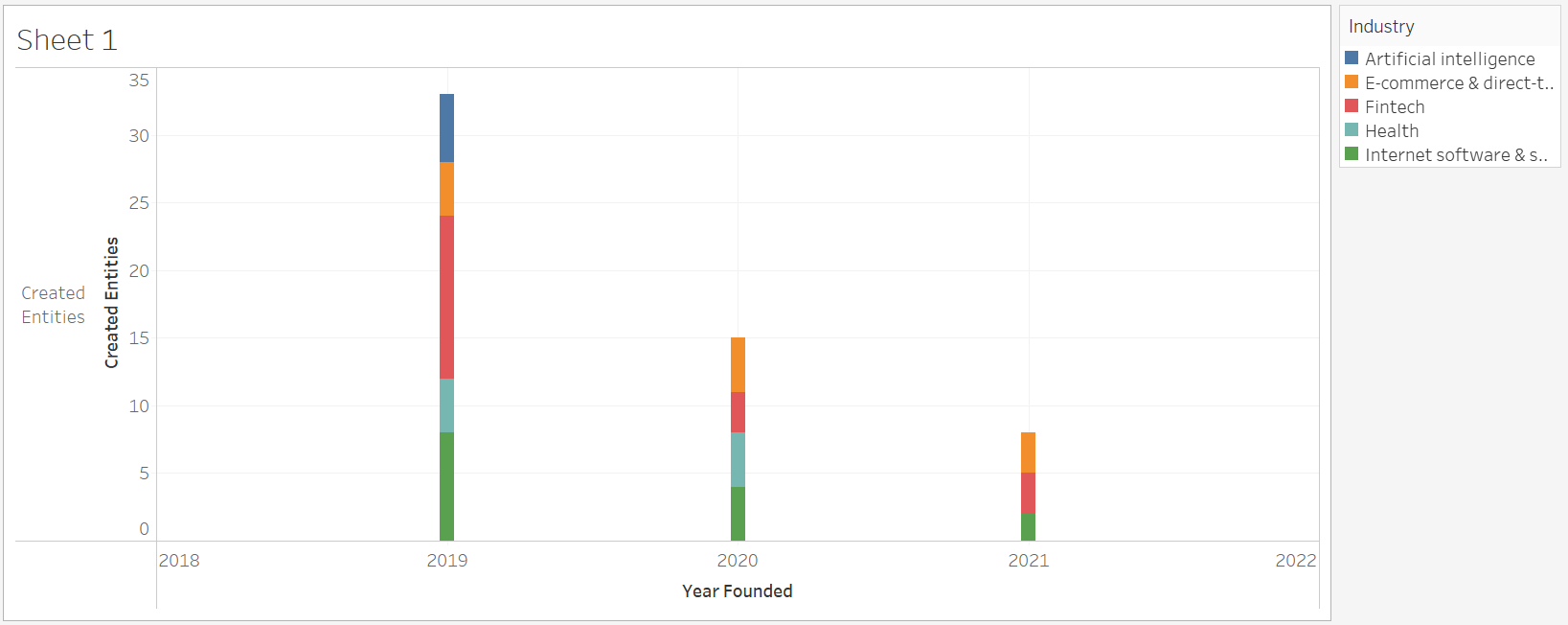
Then we used the name of these industries to query the distribution of the unicorns in each industry per year.

**QUERY**:

|  |
| --- |
| SELECT year\_founded, industry, COUNT(company\_name) AS created\_entities  FROM unicorn\_master  WHERE year\_founded IN ('2019', '2020', '2021')  AND industry IN (  'Fintech',  'Internet software & services',  'E-commerce & direct-to-consumer',  'Artificial intelligence',  'Health'  )  GROUP BY year\_founded, industry  ORDER BY industry, year\_founded; |

**ANSWER**:

|  |  |  |
| --- | --- | --- |
| year\_founded | industry | created\_entities |
| 2019 | Artificial intelligence | 5 |
| 2019 | E-commerce & direct-to-consumer | 4 |
| 2020 | E-commerce & direct-to-consumer | 4 |
| 2021 | E-commerce & direct-to-consumer | 3 |
| 2019 | Fintech | 12 |
| 2020 | Fintech | 3 |
| 2021 | Fintech | 3 |
| 2019 | Health | 4 |
| 2020 | Health | 4 |
| 2019 | Internet software & services | 8 |
| 2020 | Internet software & services | 4 |
| 2021 | Internet software & services | 2 |

 This table was pasted in an excel file named ‘Trends of businesses per year, which was uploaded in Tableau visualised as a histogram.

The histogram has the year of the unicorns founding on the X-axis as ‘Year Founding’. On the Y-axis is the count of created entities The colour legend is outlined on the far right, displaying Artificial Intelligence as blue, E-commerce and & direct-to-consumer in amber, Fintech in Red Health in turquoise, and Internet software & services in green.

Upon analysis it is clear to see that a majority of Fintech entities (red) were created in 2019, with 12 being created in that year. In 2020 there was a stark decrease in created Fintech entities, with 4 being created during the first 2 pandemic years. The pandemic seems to have been a deciding factor in the decrease of unicorns in newly created entities. In 2019 a large amount of Internet software & services(green) created with 8 unicorn entities being created this year reflecting a large interest in this industry as a prospective investment venture. As the pandemic began, newly created unicorn companies of this industry steadily decreased in the following 2 years with it halving twice in the 2020 and 2021 respectively. One industry that had consistent growth in terms of newly created unicorn industries over the 3 year period was the E-commerce & direct-to-consumer (amber). This is shown with the fact that 4 were created in 2019, 4 in 2020, and 3 in 2021 showing that this industry had consistent growth during the pandemic. Health (turquoise) was another type of industry that grew from 2019 to 2020 however no new ones were created in 2021. This maybe as a result of the pandemic, as established health companies may have been able to use existing infrastructure to benefit from the situation. However, due to difficulties as a result of restrictions it may have been tricky to establish new companies in that time period. The last but most peculiar industry to analyse is the AI related data (blue) as it had a large growth in 2019, but as a result of changes in priorities, it dropped to 0 new newly created unicorn industries in 2020 and 2021 respectively.

1. *Which countries have the most unicorns, and how do their average valuations compare?*

In order to analyse the countries with the most unicorns, as well as comparing their average evaluations we have to aggregate the countries with the highest number of unicorns, then use this list to compare their average valuations

**QUERY**:

|  |
| --- |
| SELECT country, COUNT(company\_name) as No\_of\_unicorns  FROM unicorn\_master  GROUP BY country  ORDER BY No\_of\_unicorns DESC  LIMIT 5; |

**ANSWER**:

|  |  |
| --- | --- |
| country | No\_of\_unicorns |
| United States | 553 |
| China | 166 |
| India | 61 |
| United Kingdom | 42 |
| Germany | 26 |

From this list we can compare average evaluations

**QUERY**

|  |
| --- |
| -- Highest average valuation  SELECT country, avg(valuation) AS avg\_val  FROM unicorn\_master  WHERE country IN ('United States', 'China' ,'India', 'United Kingdom' , 'Germany')  group by country  ORDER BY avg\_val; |

**ANSWER**

|  |  |
| --- | --- |
| **Country** | **Average Valuation (in Words)** |
| United Kingdom | 4.38 billion USD |
| United States | 2.95 billion USD |
| Germany | 2.77 billion USD |
| China | 2.75 billion USD |
| India | 2.39 billion USD |

* Here we can observe that America has by far the largest number of unicorns with near 3 times as much as China the second highest number. This is as of 2019, and is due to its commonly known reputation of being a good place to invest. It also has the 2nd highest average valuation.
* China has the second highest number of unicorn entities however it has the 4th highest avg\_val. This maybe due o the still unpredictable policy nature of the government and investors reluctance to put large sums of invest in the country relative to its counterparts
* India has the 3rd highest number of unicorns, however the lowest avg\_val. This can be easily explained as being due to india being a developing country with smaller relative valuations but new emerging companies sprouting from new industries.
* The United Kingdom has the 2nd lowest number of unicorn industries however the highest avg\_val as a lot of legacy companies have been there and built trust over time enough for people to invest in them.
* Germany rounds out the top five with 26 unicorn companies and an average valuation of approximately 2.77 billion USD. While its unicorn count is modest compared to the top three countries, its average valuation places it third overall. This suggests that although Germany produces fewer unicorns, the ones it does produce tend to be relatively well-capitalized and stable. This could reflect a business environment that emphasizes sustainable growth, strong regulatory standards, and high barriers to entry—resulting in fewer, but more mature and higher-valued ventures making it to unicorn status

1. *What is the average time (in years) from founding to unicorn status, and which industries or countries produce faster unicorns?*

To analyze the average time (in years) it takes for a company to reach unicorn status from its founding—and identify which industries or countries tend to produce unicorns faster—we must calculate the time difference between the year\_founded and the year of date\_joined for each company. This result is then aggregated by country or industry to determine averages.

**QUERY**:

|  |
| --- |
| SELECT country,  ROUND(AVG(YEAR(date\_joined) - year\_founded), 2) AS avg\_years  FROM unicorn\_master  GROUP BY country  ORDER BY avg\_years DESC; |

**ANSWER**:

|  |  |
| --- | --- |
| country | avg\_years |
| Czech Republic | 2 |
| Italy | 3 |
| Bahamas | 3 |
| Senegal | 3 |
| Nigeria | 3 |
| Colombia | 3.5 |
| Philippines | 4 |
| Argentina | 4 |
| Mexico | 5 |
| Vietnam | 5 |
| Chile | 5 |
| Indonesia | 5.4 |
| Austria | 5.5 |
| Hong Kong | 5.67 |
| China | 5.85 |
| Malaysia | 6 |
| Denmark | 6 |
| Turkey | 6 |
| Estonia | 6 |
| Thailand | 6 |
| Japan | 6.6 |
| Sweden | 6.75 |
| United States | 6.78 |

Companies with the fastest rate of acquiring unicorn status

**QUERY**:

|  |
| --- |
| SELECT industry,  ROUND(AVG(YEAR(date\_joined) - year\_founded), 2) AS avg\_years  FROM unicorn\_master  GROUP BY industry  ORDER BY avg\_years ASC; |

**ANSWER**:

|  |  |
| --- | --- |
| industry | avg\_years |
| Auto & transportation | 5.21 |
| Hardware | 5.63 |
| Artificial intelligence | 5.94 |
| Mobile & telecommunications | 6.32 |
| E-commerce & direct-to-consumer | 6.4 |
| Fintech | 6.55 |
| Travel | 6.62 |
| Cybersecurity | 6.8 |
| Supply chain, logistics, & delivery | 7.09 |
| Other | 7.47 |
| Internet software & services | 7.87 |
| Edtech | 7.92 |
| Data management & analytics | 8.13 |
| Health | 8.19 |
| Consumer & retail | 8.33 |

* The average time it takes for a company to reach unicorn status from its founding reveals important insights about national startup ecosystems and industry-level acceleration. By calculating the difference between the year each company was founded and the year it became a unicorn, we can assess how quickly high-value companies are emerging across various contexts.
* Among countries, the Czech Republic stands out with the shortest average time to unicorn status, at just 2 years. This is followed by Italy, the Bahamas, Senegal, and Nigeria, all averaging around 3 years. These unusually fast transitions suggest targeted investment environments or exceptional scalability within individual success stories, rather than broad industry trends.
* In contrast, more established ecosystems such as China and the United States exhibit longer average times of 5.85 years and 6.78 years respectively. This could be due to higher regulatory standards, larger markets requiring more time to penetrate, or more rigorous funding rounds before reaching billion-dollar valuations.
* Industry-wise, Auto & transportation companies achieve unicorn status the fastest, with an average of 5.21 years from founding. This rapid ascent may be driven by intense innovation and investment in electric vehicles, autonomous transport, and mobility tech. Close behind is Hardware at 5.63 years, reflecting a fast-moving sector where clear product-market fit and investor backing can lead to quick valuations.
* Artificial Intelligence also shows a relatively short average time to unicorn status at 5.94 years, likely due to the high demand for automation and AI-driven solutions in multiple sectors.
* On the other end of the spectrum, industries such as Consumer & Retail (8.33 years), Health (8.19 years), and Data Management & Analytics (8.13 years) have the longest average timeframes. These sectors often require extensive research and development, regulatory approvals, or large-scale infrastructure before achieving high valuation thresholds, which can slow down the path to unicorn status.
* Overall, the analysis shows that the speed to unicorn status is influenced both by country-specific factors—such as policy environments, market maturity, and investor culture—and by the nature of the industry itself, particularly the capital intensity and innovation cycles required.

1. *Which cities host the highest number of unicorns, and are these companies clustered in similar industries?*

To identify the cities hosting the highest number of unicorn companies, we need to group the data by city and count the number of unicorns headquartered in each. This helps us understand global startup hotspots and where unicorn formation is most concentrated.

This query returns the top 10 cities by unicorn count, allowing further analysis on potential industry clustering patterns in the next set.

**QUERY**

|  |
| --- |
| SELECT city, COUNT(company\_name) AS no\_of\_unicorns  FROM unicorn\_master  GROUP BY city  ORDER BY no\_of\_unicorns DESC  LIMIT 10; |

**ANSWER**:

|  |  |
| --- | --- |
| city | no\_of\_unicorns |
| San Francisco | 149 |
| New York | 103 |
| Beijing | 59 |
| Shanghai | 43 |
| London | 33 |
| Bengaluru | 26 |
| Shenzhen | 19 |
| Paris | 19 |
| Palo Alto | 18 |
| Berlin | 17 |

Then this subquery is created to show the distribution of the unicorns in the industries

**QUERY**:

|  |
| --- |
| SELECT city, industry, COUNT(company\_name) AS unicorn\_count  FROM unicorn\_master  WHERE city IN (  SELECT city  FROM unicorn\_master  GROUP BY city  ORDER BY COUNT(company\_name) DESC  LIMIT 10  )  GROUP BY city, industry  ORDER BY city, unicorn\_count DESC; |

**ANSWER (SEE city vs industry clustering.xsl)**

The data was exported to excel via copy and paste. It was imported to Tableau as an excel file where it was displaced as a heat highlight table to more effectively analyse the different types of the data.



* The heat map above has city on the Y axis and industry at the X axis. Upon analysis it is easy to come up with insights about the unicorn landscape. It shows how the above cities have evolved as specialized centres for different industries Beijing is the top unicorn hub, with a diverse range of industries represented. It leads in E-commerce & direct-to-consumer (10), Mobile & telecommunications (9), and Artificial Intelligence (9). Edtech also holds a strong presence with 6 unicorns, followed by Internet software & services (5). Sectors like Fintech, Hardware, and Consumer & retail each have 4, indicating a well-rounded tech ecosystem with strong government and infrastructure support.
* Bengaluru ranks next, dominated by Internet software & services (9) and Fintech (6). E-commerce (3) and Edtech (2) follow closely. Smaller counts in Health, Auto & transportation, and Mobile suggest emerging diversification but a clear lean towards SaaS and financial platforms.
* Berlin’s unicorns are most prominent in Fintech and E-commerce (4 each), with Travel (3) and Internet software (2) close behind. This spread indicates a consumer-focused digital market, with moderate representation across tech verticals.
* London’s unicorn scene is heavily skewed towards Fintech (24), with all other industries trailing far behind. Its position as a global financial centre likely drives this trend, while smaller counts in AI, Mobile, and E-commerce reflect secondary activity.
* New York shows the most balanced mix. Fintech dominates (33), but Internet software (20), Health (14), and Cybersecurity (9) also have strong counts. This diversity reflects a broad, well-funded startup ecosystem with cross-sector appeal.
* Palo Alto is smaller in volume but reflects deep tech roots. Internet software (5), AI (4), and Data analytics (3) are the top sectors. Other industries like Cybersecurity, Fintech, and Hardware appear once each, showing focused innovation in niche verticals.

1. *Which investors most frequently appear among the top-valued unicorns, and in what industries are they investing?*

To add investor data into the unified unicorn\_master table, the select\_investors column from the original funding dataset needs to be integrated. Since unicorn\_master is already a merged table combining key information from all the staging tables, the process involves two main steps: first, adding a new column to hold the investor data, and second, populating that column by matching records based on a shared key — in this case, the company\_id field which uniquely identifies each unicorn entity across all tables.

**QUERY**:

|  |
| --- |
| SELECT \*  FROM unicorn\_master;  ALTER TABLE unicorn\_master  ADD COLUMN select\_investors TEXT;  UPDATE unicorn\_master um  JOIN funding f ON um.company\_id = f.company\_id  SET um.select\_investors = f.select\_investors;  SELECT select\_investors  FROM unicorn\_master  LIMIT 3; |

**RESULT**:

|  |
| --- |
| select\_investors |
| EQT Partners |
| Dragonfly Captial, Qiming Venture Partners, DST Global |
| Blackstone, Bessemer Venture Partners |

Then from here we have to define top valued unicorns. We’ll do this by making a temporary tables

**SQL QUERY:**

|  |
| --- |
| -- Adjust the LIMIT or percentage threshold as needed  CREATE TEMPORARY TABLE top\_valued\_unicorns AS  SELECT \*  FROM unicorn\_master  ORDER BY valuation DESC  LIMIT 100; -- Top 100 unicorns |

From here we can query how many times the investor comes up in the list

**SQL QUERY**

|  |
| --- |
| SELECT select\_investors, count(select\_investors) as No\_of\_investments  FROM unicorn\_master  GROUP BY select\_investors DESC LIMIT 5; |

**ANSWER**

|  |  |
| --- | --- |
| select\_investors | No\_of\_investments |
| Sequoia Capital | 3 |
| Qualcomm Ventures, Accel, Canaan Partners | 2 |
| Speedinvest, Valar Ventures, Uniqa Ventures | 2 |
| Insight Partners, Sequoia Capital, Index Ventures | 2 |
| General Atlantic | 2 |
| Greylock Partners, Google Ventures, BlackRock | 2 |
| Undisclosed | 2 |
| Tiger Global Management, Sequoia Capital India, Ribbit Capital | 2 |
| SoftBank Group | 2 |
| Sequoia Capital China, Qiming Venture Partners, Tencent Holdings | 2 |
| Shunwei Capital Partners, China Media Group, Guangzhou Huiyin Aofeng Equity Investment Fund | 2 |
| Pelion Venture Partners, Foundation Capital, Thoma Bravo | 2 |
| Global Founders Capital, Shea Ventures, Greycroft | 2 |
| Two Sigma Ventures, Flint Capital, Commerce Ventures | 2 |

Sequoia Capital stands out as the most frequent investor among top-valued unicorns, appearing in three separate investment cases, indicating its strong presence in high-growth ventures. Several other investor groups, such as SoftBank Group and General Atlantic, appear twice, often as part of syndicates. This highlights a trend where top unicorns attract backing from major firms through collaborative funding. The presence of diverse investor combinations also suggests cross-sector confidence and strategic pooling of capital to support startups with significant market potential.

1. *Which industries have the highest average valuation-to-funding ratio, and what does this suggest about return on investment potential?*

To do this we’ll have to first calculate the ratios of each industries total funding vs their valuation and then compare them from country to country

Ratios:

**QUERY**

|  |
| --- |
| SELECT  industry,  ROUND(AVG(valuation / funding), 2) AS avg\_valuation\_funding\_ratio -- Average Ratio  FROM  unicorn\_master  WHERE  funding IS NOT NULL AND funding != 0 -- Where it is not 0 or null  GROUP BY  industry  ORDER BY  avg\_valuation\_funding\_ratio DESC; |

To successfully analyse the data the top 5 ratios can be analysed

Top 5:

|  |  |
| --- | --- |
| industry | avg\_valuation\_funding\_ratio |
| Internet software & services | 28.74 |
| Other | 9.53 |
| Mobile & telecommunications | 9.39 |
| Fintech | 9.14 |
| Hardware | 8.44 |

Here, one can clearly see that Internet Software & Services stands out as the most profit-yielding industry, boasting a return on investment (ROI) ratio of 28.74 - over four times higher than the next most efficient sectors. This significantly higher valuation-to-funding ratio highlights its dominance in capital efficiency and investor appeal. Following that are industries grouped under “Other,” as well as Mobile & Telecommunications and Fintech, each with ROI ratios hovering around 8, suggesting solid but comparatively modest returns relative to capital invested.

1. *Which industries saw the greatest number of new unicorns join the list in 2020 and 2021, and what global factors might have driven this growth?*

The industry that had the greatest number of new unicorns is the one with the most number of industries with founding dates in 2020 and 2021.

**QUERY:**

|  |
| --- |
| SELECT industry, COUNT(company\_name) as new\_unicorn  FROM unicorn\_master  WHERE year\_founded IN (2019, 2020)  GROUP BY industry  ORDER BY new\_unicorn DESC |

**ANSWER:**

|  |  |
| --- | --- |
| industry | new\_unicorn |
| Fintech | 15 |
| Internet software & services | 12 |
| Cybersecurity | 8 |
| E-commerce & direct-to-consumer | 8 |
| Health | 8 |
| Artificial intelligence | 5 |
| Other | 5 |
| Consumer & retail | 2 |
| Mobile & telecommunications | 2 |
| Data management & analytics | 2 |
| Hardware | 1 |
| Supply chain, logistics, & delivery | 1 |
| Auto & transportation | 1 |

Between 2020 and 2021, the rise in new unicorns was heavily influenced by the global pandemic and the increasing reliance on digital infrastructure. Fintech emerged as the top-performing industry, with 15 new unicorns, as consumers and businesses worldwide shifted toward online payments, digital wallets, and decentralized finance solutions. Internet software & services followed with 12, reflecting the boom in remote work tools, SaaS platforms, and cloud-based collaboration. Cybersecurity, Health, and E-commerce all tied with 8 new entries each, showing the growing need for secure digital environments, accessible healthcare tech, and online consumer services during lockdowns. Artificial Intelligence also saw growth with 5 unicorns, benefiting from increased demand for automation, smart analytics, and virtual assistance. Other tech-focused industries gained traction as well, showcasing the resilience and adaptability of tech during uncertainty. Overall, the pandemic acted as a catalyst for rapid innovation, accelerating trends that were already on the rise and reshaping the global startup landscape.

1. *What is the average funding raised per unicorn per country, and how does that relate to the country’s total number of unicorns?*

To analyze the average funding raised per unicorn per country and its relationship with the total number of unicorns, two key metrics must be calculated: the total number of unicorns per country, and the average funding per unicorn in that country. This helps determine whether countries with more unicorns are also providing more funding per startup—or if fewer, more heavily funded startups are skewing the averages.

SQL QUERY:

|  |
| --- |
| SELECT country, COUNT(company\_name) AS total\_unicorns -- Total unicorns per country  FROM unicorn\_master  GROUP BY country  ORDER BY total\_unicorns DESC;  SELECT country, ROUND(AVG(funding), 2) AS avg\_funding -- Average FUnding  FROM unicorn\_master  GROUP BY country  ORDER BY avg\_funding DESC;  SELECT country, -- United queries  COUNT(company\_name) AS total\_unicorns,  ROUND(AVG(funding), 2) AS avg\_funding  FROM unicorn\_master  GROUP BY country  ORDER BY avg\_funding DESC; |

The results produce 46 rows with varying results. From the data, it’s evident that smaller or emerging markets like the Bahamas, Colombia, and Turkey top the list in terms of average funding per unicorn, each receiving well over $800 million to $2 billion per company. This often reflects a concentration of large capital into very few standout companies in those regions.

In contrast, larger ecosystems like the United States, China, and India, despite having the highest number of unicorns, show lower average funding per unicorn. For example:

* United States: 553 unicorns at an average of $421.6 million
* China: 166 unicorns at $490.6 million
* India: 61 unicorns at $519 million

This suggests that while these countries are powerhouses for startup creation, the capital is spread across a larger number of companies. Countries like the United Kingdom, Germany, and France fall in the mid-range, showing balanced startup ecosystems with solid average funding and a fair number of unicorns.

1. *Are newer unicorns (post-2020) valued higher or lower than earlier unicorns, and does this trend vary by industry*

To do this average unicorn valuation of companies pre 2020 would then be compared to that post 2020. Then it can be seen in the same way from industry to industry.

Pre-2020

**QUERY**:

|  |
| --- |
| SELECT AVG(valuation) AS pre\_2020\_avg\_val  FROM unicorn\_master  WHERE year\_founded < 2020; |

**ANSWER**:

|  |
| --- |
| pre\_2020\_avg\_val |
| 2.92 billion USD |

Post-2020

**QUERY**

|  |
| --- |
| SELECT AVG(valuation) AS avg\_val\_from\_2020  FROM unicorn\_master  WHERE year\_founded >= 2020 |

|  |
| --- |
| avg\_val\_from\_2020 |
| 1.83 billion |

From this we can easily see that unicorn companies are valued higher post 2020 than pre-2020. Its clearly as a result of the after effects of the Corona Virus pandemic, which caused a general slowdown in all industries. But we can still gather further insights by analysing data from industry to industry.

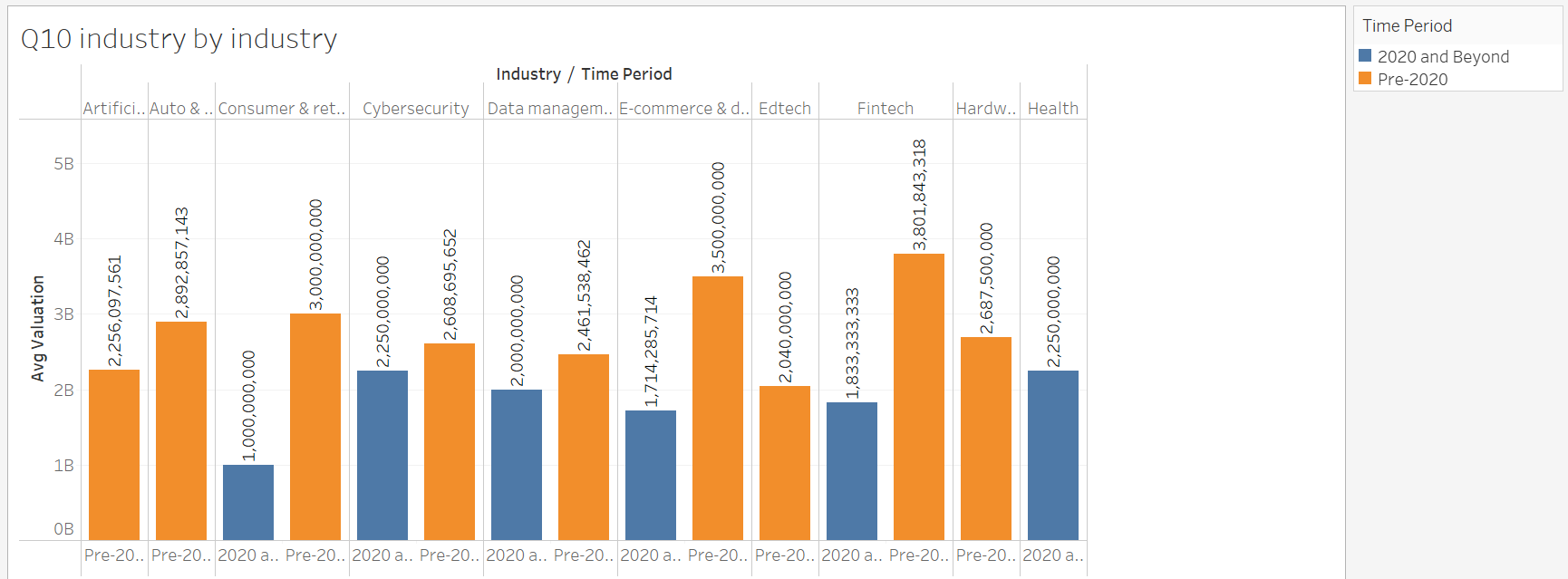
Pre-2020 AND From 2020

**QUERY**

|  |
| --- |
| SELECT  industry,  CASE  WHEN year\_founded < 2020 THEN 'Pre-2020'  ELSE '2020 and Beyond'  END AS time\_period,  AVG(valuation) AS avg\_valuation  FROM unicorn\_master  GROUP BY industry, time\_period  ORDER BY industry, time\_period; |

**ANSWER**:

|  |  |  |
| --- | --- | --- |
| industry | time\_period | avg\_valuation |
| Artificial intelligence | Pre-2020 | **2.26 billion** |
| Auto & transportation | Pre-2020 | **2.89 billion** |
| Consumer & retail | Pre-2020 | **3.00 billion** |
| Cybersecurity | Pre-2020 | **2.61 billion** |
| Data management & analytics | Pre-2020 | **2.46 billion** |
| E-commerce & direct-to-consumer | Pre-2020 | **3.50 billion** |
| Edtech | Pre-2020 | **2.04 billion** |
| Fintech | Pre-2020 | **3.80 billion** |
| Hardware | Pre-2020 | **2.69 billion** |
| Health | Pre-2020 | **2.70 billion** |
| Internet software & services | Pre-2020 | **2.95 billion** |
| Mobile & telecommunications | Pre-2020 | **2.33 billion** |
| Other | Pre-2020 | **1.94 billion** |
| Supply chain, logistics, & delivery | Pre-2020 | **1.98 billion** |
| Travel | Pre-2020 | **2.26 billion** |
| Consumer & retail | 2020 and Beyond | **2.89 billion** |
| Cybersecurity | 2020 and Beyond | **3.00 billion** |
| Data management & analytics | 2020 and Beyond | **2.61 billion** |
| E-commerce & direct-to-consumer | 2020 and Beyond | **2.46 billion** |
| Fintech | 2020 and Beyond | **3.50 billion** |
| Health | 2020 and Beyond | **2.04 billion** |
| Internet software & services | 2020 and Beyond | **3.80 billion** |
| Mobile & telecommunications | 2020 and Beyond | **2.69 billion** |

I visualised this in Tableau to get better insights.

In this bar chart Average valuation in billions is plotted against Industry/Time Period. Unicorn valuations reveal a shift across industries before and after 2020. Pre-2020, industries like Fintech ($3.80B), E-commerce ($3.50B), and Consumer & Retail ($3.00B) led in average valuations, reflecting sustained investor confidence. Post-2020, many industries saw declines—Consumer & Retail dropped to $1.00B—while others like Mobile & Telecom rose to $2.50B. Cybersecurity and Health maintained strong valuations across both periods (~$2.25B), showing resilience. The data suggests that while traditional sectors cooled post-2020, others adapted or gained relevance amid global disruptions, revealing shifting investment trends and evolving market confidence in newer tech-driven sectors.

Acknowledgements

Datacamp Labs for the data

<https://www.datacamp.com/datalab>