# DATA ANALYSIS AND VISUALIZATION

## UNICORN COMPANIES ANALYSIS

### ❖ Setting the Context: Objectives, Background, and Data Sources

**Objective:** The aim of this project is to gain valuable insights into the startup industry. Specifically, we will focus on identifying the time it takes for startups to become unicorns in their respective industries, as well as which investors are funding these companies and to what extent. Another important aspect of this analysis will be to determine which companies have produced the most unicorns. By examining these key factors, we hope to gain a deeper understanding of the startup ecosystem and provide actionable recommendations to stakeholders in the industry.

**Background:** In recent years, the term "unicorn" has been used to describe a private company with a valuation exceeding $1 billion, yet not listed on any stock exchange. Presently, there are over 1,000 such companies globally, highlighting their increasing significance in the business world. In this project, we will be leveraging a comprehensive dataset containing information on 1,074 unicorn companies from 46 countries and 15 industries. Additionally, the dataset includes data on 1,250 investors, providing us with a detailed view of the funding landscape. By analyzing this data, we aim to gain a better understanding of the factors that contribute to the success of unicorn companies, as well as the characteristics that set them apart from other businesses.

**Data Sources:** The source of our project's dataset is from Kaggle, a platform for data science enthusiasts. The dataset includes information on unicorn companies and is publicly available at the following link: https://www.kaggle.com/datasets/deepcontractor/unicorn-companies-dataset

### ❖ Methodology

For our project's methodology, we will be utilizing a range of software tools to support our analysis. The software tools we will be using are Tableau, Python, R, and MS Excel. These tools will be instrumental in helping us with various aspects of our analysis. Specifically, Tableau will be used primarily for visualization purposes, Python will be leveraged for creating network graphs, and R and MS Excel will be utilized for data cleaning and working with variables. While these software tools will serve as our primary means of analysis, we remain open to incorporating other tools as needed throughout the course of the project.

## ❖ Data Dictionary

| Field | Description | Data Type |
|---|---|---|
| **Company** | Name of the company | String |
| **Valuation** | Company valuation in Billions (B) of dollars | Integer |
| **Year Joined** | The year in which the company reached $1 billion in valuation | Date |
| **Industry** | Industry/Sector in which company is established | String |
| **City** | City the company was founded in | String |
| **Country** | Country the company was founded in | String |
| **Continent** | Continent the company was founded in | String |
| **Year Founded** | Year the company was founded | Date |
| **Funding** | Total amount raised across all funding rounds in Billions (B) or Millions (M) of dollars | Integer |
| **Select Investors** | Top 4 investing firms or individual investors (some have less than 4) | String |
| **Investor Count** | Number of Investor the company/startup has | Integer |
| **Valuation to Funding Ratio** | Ratio of Valuation divided by Funding | Float |

## ❖ Data Cleaning and Preprocessing

As part of the data cleaning and preprocessing steps, we undertook a number of actions to prepare the dataset for analysis. We created two new columns, 'Investor Count' and 'Valuation by Funding Ratio', which will facilitate the creation of interesting visualizations in Tableau. Using R, we identified null values in each column and discovered that the 'funding' and 'investor count' columns had 12 and 55 null values respectively. Additionally, we removed the dollar sign from the 'Valuation' and 'Funding' columns to enable numerical calculations. We also converted the 'Date Joined' and 'Year Founded' columns to year-only format to support visualization efforts. The summary of the dataset before data preprocessing is presented in the image below.

```
   Company            Valuation        Date Joined           Industry
Length:1074        Min.   :  1.000   Min.   :2007-07-02   Length:1074
Class :character   1st Qu.:  1.000   1st Qu.:2019-05-20   Class :character
Mode  :character   Median :  2.000   Median :2021-03-29   Mode  :character
                   Mean   :  3.455   Mean   :2020-05-19
                   3rd Qu.:  3.000   3rd Qu.:2021-09-19
                   Max.   :180.000   Max.   :2022-04-05

     City              Country           Continent          Year Founded
Length:1074        Length:1074        Length:1074        Min.   :1919-01-01
Class :character   Class :character   Class :character   1st Qu.:2011-01-01
Mode  :character   Mode  :character   Mode  :character   Median :2014-01-01
                                                         Mean   :2012-11-23
                                                         3rd Qu.:2016-01-01
                                                         Max.   :2021-01-01

    Funding       Select Investors   Investor Count  Valuation to Funding Ratio
Min.   :  0.0   Length:1074        Min.   : 1.00   Min.   :   0.000
1st Qu.:166.0   Class :character   1st Qu.: 8.00   1st Qu.:   3.402
Median :300.0   Mode  :character   Median :13.00   Median :   5.150
Mean   :338.1                      Mean   :14.51   Mean   :  11.510
3rd Qu.:491.5                      3rd Qu.:19.00   3rd Qu.:   8.148
Max.   :999.0                      Max.   :91.00   Max.   :4000.000
NA's   :12                         NA's   :55
```

***Figure 1: Summary of data before cleaning***

Following the data cleaning and preprocessing steps, we present the dataset summary in the accompanying image below. As part of the preprocessing steps, we have separated the Select Investors into different Investor Columns, each separated by a comma. These steps were necessary to ensure the dataset is in a format that is conducive to analysis and helps to mitigate any issues that may arise from data inconsistencies or errors.

```
   Company          Valuation (in Billions)  Year Joined     Industry
Length:1074        Min.   :  1.000         Min.   :2007   Length:1074
Class :character   1st Qu.:  1.000         1st Qu.:2019   Class :character
Mode  :character   Median :  2.000         Median :2021   Mode  :character
                   Mean   :  3.455         Mean   :2020
                   3rd Qu.:  3.000         3rd Qu.:2021
                   Max.   :180.000         Max.   :2022
    City             Country           Continent          Year Founded
Length:1074        Length:1074        Length:1074        Min.   :1919
Class :character   Class :character   Class :character   1st Qu.:2011
Mode  :character   Mode  :character   Mode  :character   Median :2014
                                                         Mean   :2013
                                                         3rd Qu.:2016
                                                         Max.   :2021
Funding (in Billions) Select Investors    Investor 1         Investor 2
Min.   : 0.0000       Length:1074        Length:1074        Length:1074
1st Qu.: 0.2180       Class :character   Class :character   Class :character
Median : 0.3650       Mode  :character   Mode  :character   Mode  :character
Mean   : 0.5511
3rd Qu.: 0.6030
Max.   :14.0000
 Investor 3         Investor 4        Investor Count  Valuation to Funding Ratio
Length:1074        Length:1074        Min.   : 1.00   Min.   :   0.000
Class :character   Class :character   1st Qu.: 8.00   1st Qu.:   3.403
Mode  :character   Mode  :character   Median :13.00   Median :   5.155
                                      Mean   :14.55   Mean   :  11.510
                                      3rd Qu.:19.00   3rd Qu.:   8.151
                                      Max.   :91.00   Max.   :4000.000
```

***Figure 2: Summary of data after cleaning***

## ❖ Data Analysis and Results

1. TOP 3 INDUSTRY WITH HIGHEST FUNDING AND VALUATION

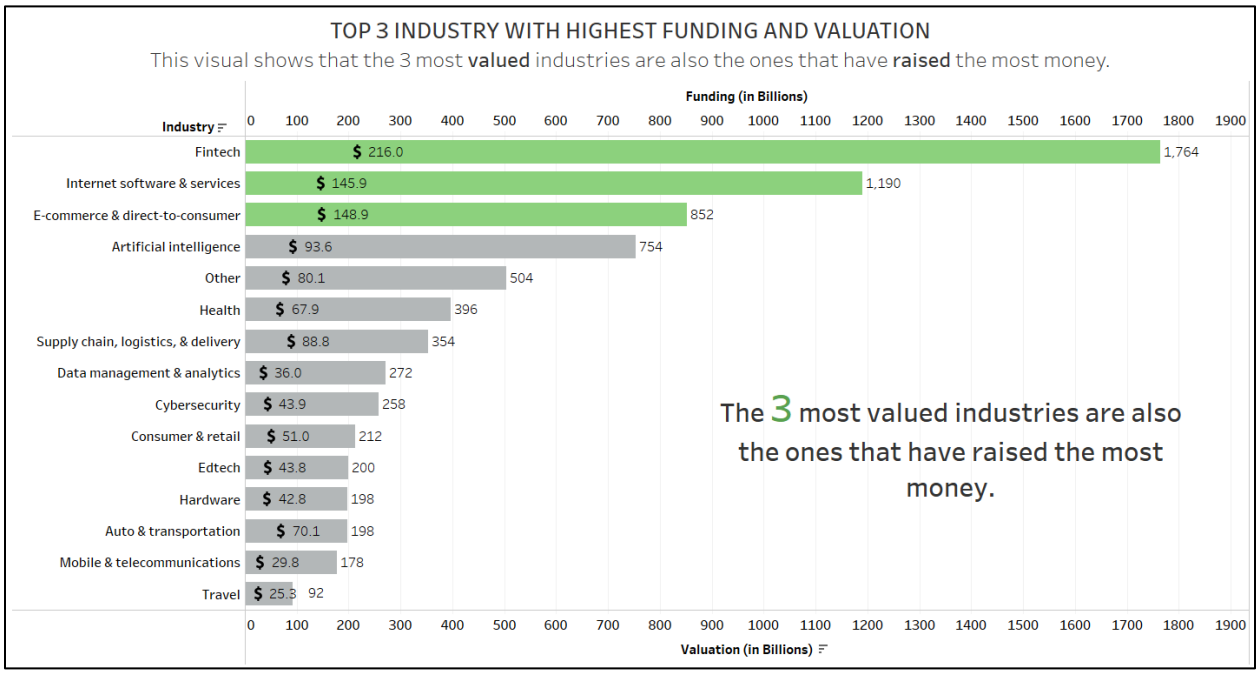


***Figure 3: Top 3 industry with highest funding and valuation***

The visualization presented highlights the three industries that have raised the most funding and achieved the highest valuations. Based on the data, it is clear that the Fintech, Internet Software and Services, and E-commerce & Direct-to-Consumer industries have emerged as the dominant players in the startup landscape. These industries have raised a significant amount of funding and have been able to achieve impressive valuations as a result. For instance, the Fintech industry has raised around $216 billion in funding, with the total valuation of all companies reaching $1764 billion. Likewise, the Internet Software and Services and E-commerce industries have raised approximately $146 billion and $149 billion, respectively, with a combined valuation of $1190 billion and $852 billion.

Interestingly, the three most valued industries are also the ones that have been able to raise the most money. This indicates that there is a clear correlation between the ability to raise funding and achieving a high valuation. By examining these trends, we can identify the industries that are most attractive to investors and provide insights into the factors that contribute to their success.

To explore this further, we will be examining an additional plot.

## 2. TOP 10 COMPANIES BY INDUSTRY

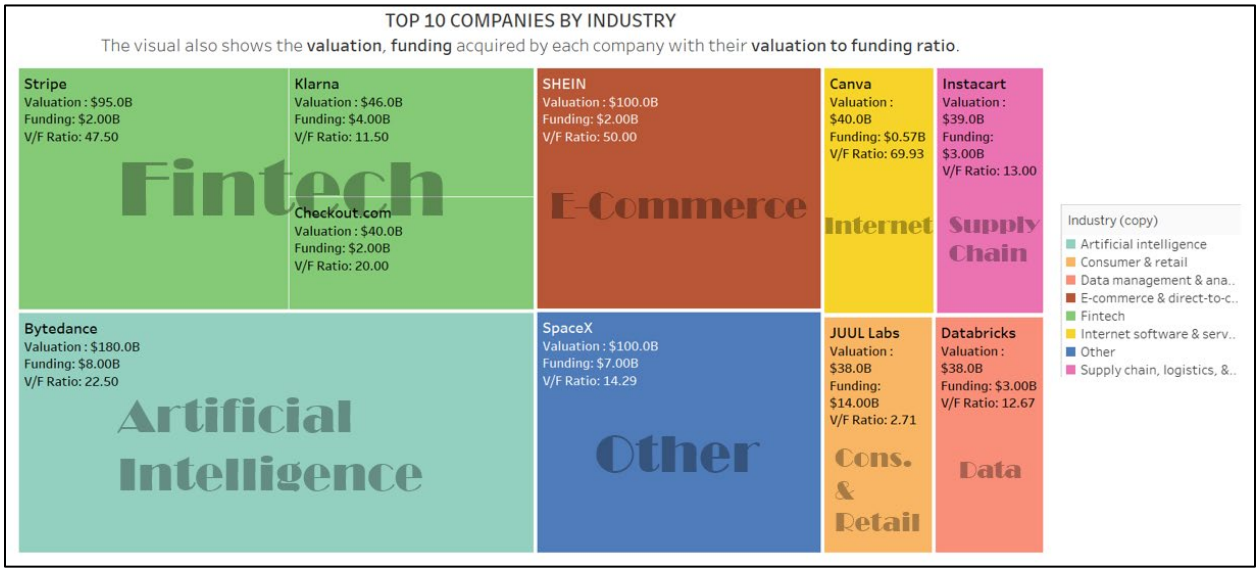


***Figure 4: Top 10 companies by industry***

The presented visualization provides a detailed look at the companies that are dominating their respective industries, along with their valuations, funding, and valuation-to-funding ratios. By examining this information, we can identify the companies that are most successful in their respective industries and gain insights into the factors that contribute to their success.

Based on the data, it is clear that Stripe, Klarna, and Checkout.com are the leading companies in the Fintech industry. Similarly, Shein is the dominant player in the E-commerce and Direct-to-Consumer industry, while Canva is the clear leader in the Internet Software and Services industry. Instacart has emerged as the leading company in the Supply Chain, Logistic and Delivery industry, while Bytedance dominates the Artificial Intelligence industry. SpaceX is the dominant player in the Other industry, while JUUL Labs is the leading company in the Consumer & Retail industry. Finally, Databricks has emerged as the leading player in the Data Management and Analytics industry.

Interestingly, Canva has the highest Valuation-to-Funding Ratio, indicating that it is the most efficient at converting funding into valuation. This is followed by Shein, Stripe, and Bytedance, indicating that these companies are also highly effective at using funding to increase their valuations.

Now that we have identified which companies are leading in their industries, let's explore how long it takes to achieve unicorn status.

## 3. AVERAGE TIME TO BECOME A UNICORN

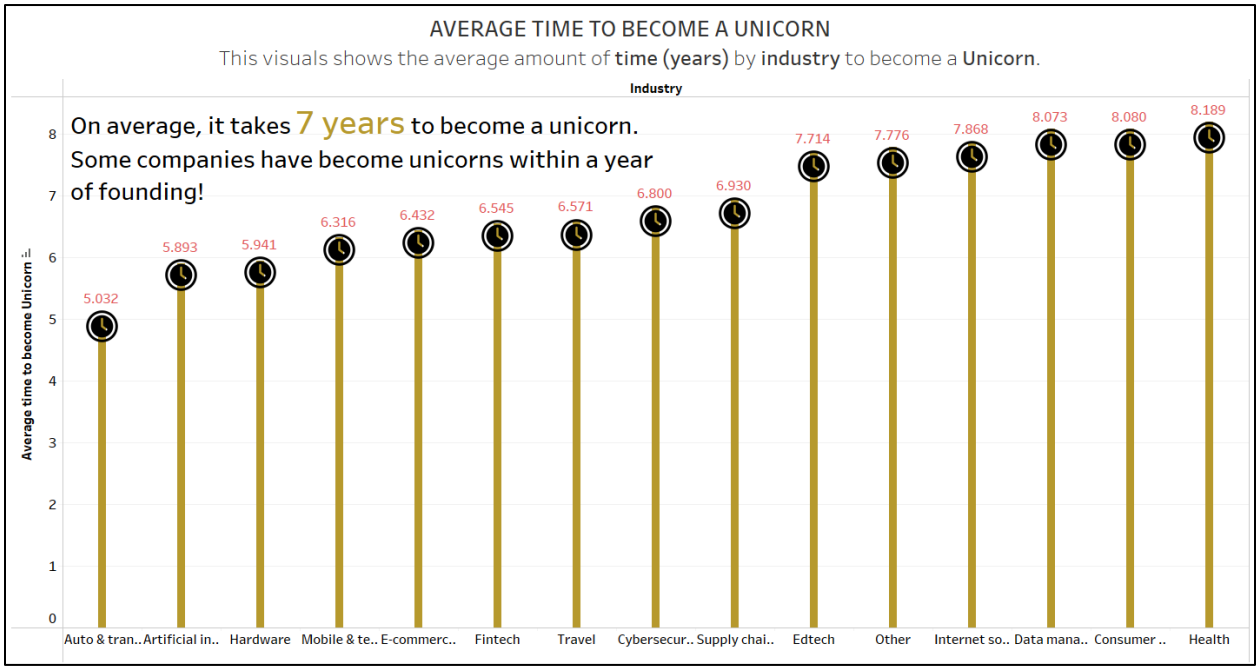


***Figure 5: Average time to become a Unicorn***

This visualization displays the average time it takes for companies to reach unicorn status in different industries. The Auto and Transportation industry has the lowest average time of just 5 years, indicating that it takes relatively little time to reach unicorn status in this industry. On the other hand, the health industry has the highest average time of 8.189 years.

In some technology fields such as Artificial Intelligence and Cyber Security, it takes around 5.8 and 6.8 years on average, respectively, to become a unicorn.

Interestingly, although the average time it takes for a company to become a unicorn is 7 years, there are some exceptional cases where companies have achieved unicorn status within just one year of being founded.

Next, we will examine the top investors and their investments in different industries.

4. TOP 10 INVESTORS AND THEIR RANKING BY INVESTMENT

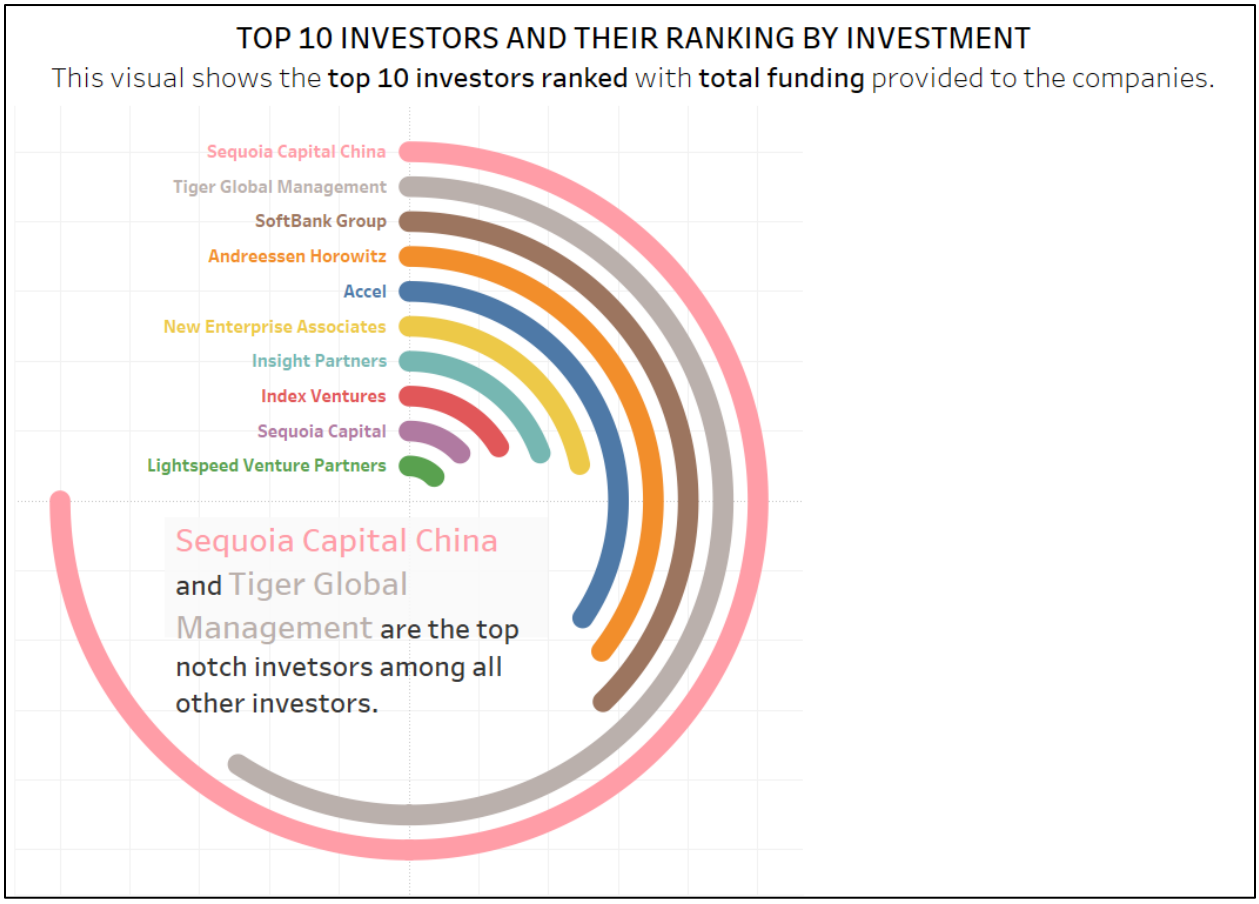


***Figure 6: Top 10 investors and their ranking by investment***

The presented radial chart displays the top 10 investors ranked by their total funding provided to the companies. Sequoia Capital China and Tiger Global Management occupy the top two positions among the 10 investors listed. It is noticeable that Andreessen Horowitz, Accel, and New Enterprise Associates have invested similar amounts in the companies. Conversely, Lightspeed Venture Partners have made the least investment among the top 10 investors.

Now that we know which investors are leading the pack, let's explore the industries they are investing in and the amount they are providing in funding.

## 5. TOP 10 INVESTORS WITH THEIR FUNDING BY INDUSTRY

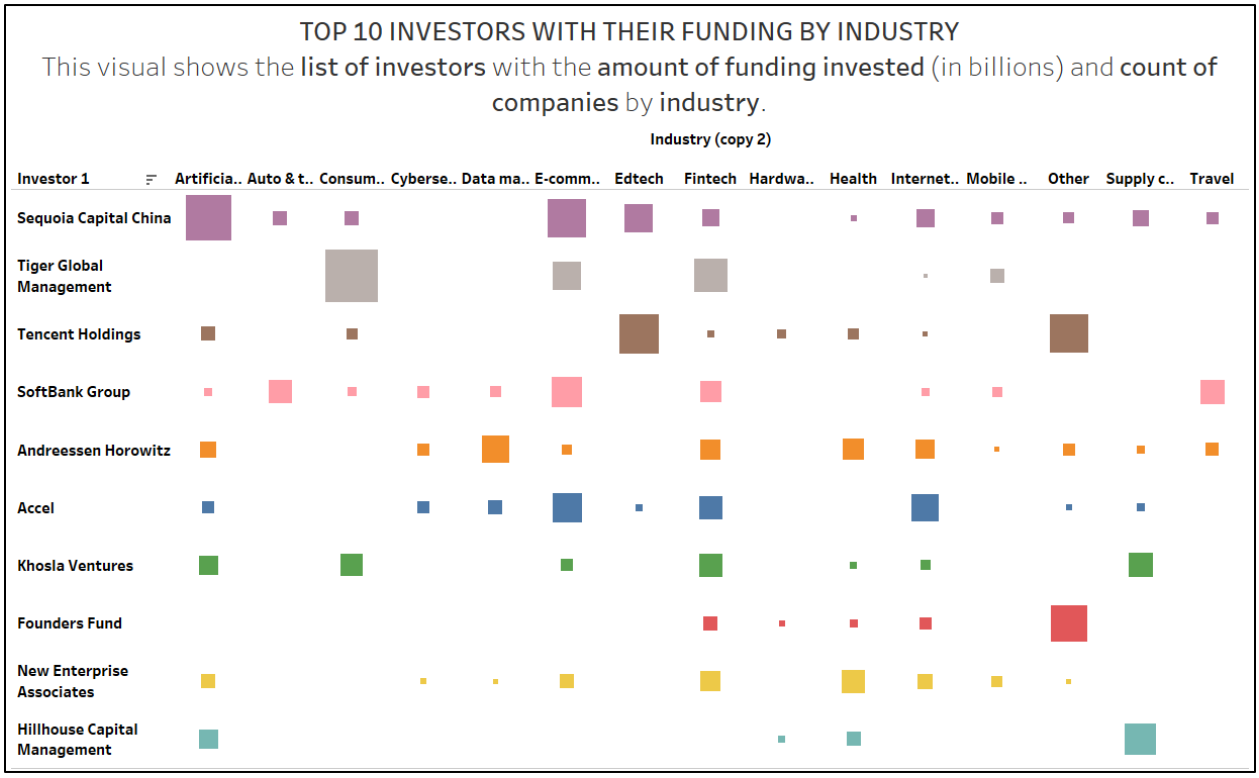


***Figure 7: Top 10 investors with their funding by industry***

This visualization displays the investment patterns of top investors across different industries. The x-axis shows the industry categories while the y-axis represents the investors. The size of the square represents the amount of funding invested by the respective investors in a particular industry.

For example, Sequoia Capital China, which is the leading investor, has made significant investments in the Artificial Intelligence and E-commerce & direct-to-consumer industries. Similarly, Tiger Global Management has focused mainly on the Consumer & Retail and Fintech sectors. Interestingly, the auto & transportation and travel industries are receiving comparatively less investment from top investors. Conversely, the investors are primarily focusing on the burgeoning and dynamic technology industries such as Artificial Intelligence, E-commerce & direct-to-consumer, Fintech, and Internet Software & Services.

Overall, this visualization highlights the investment trends and preferences of top investors across various industries.

6. PROPORTION OF VALUATION BY INDUSTRY FOR EACH CONTINENT

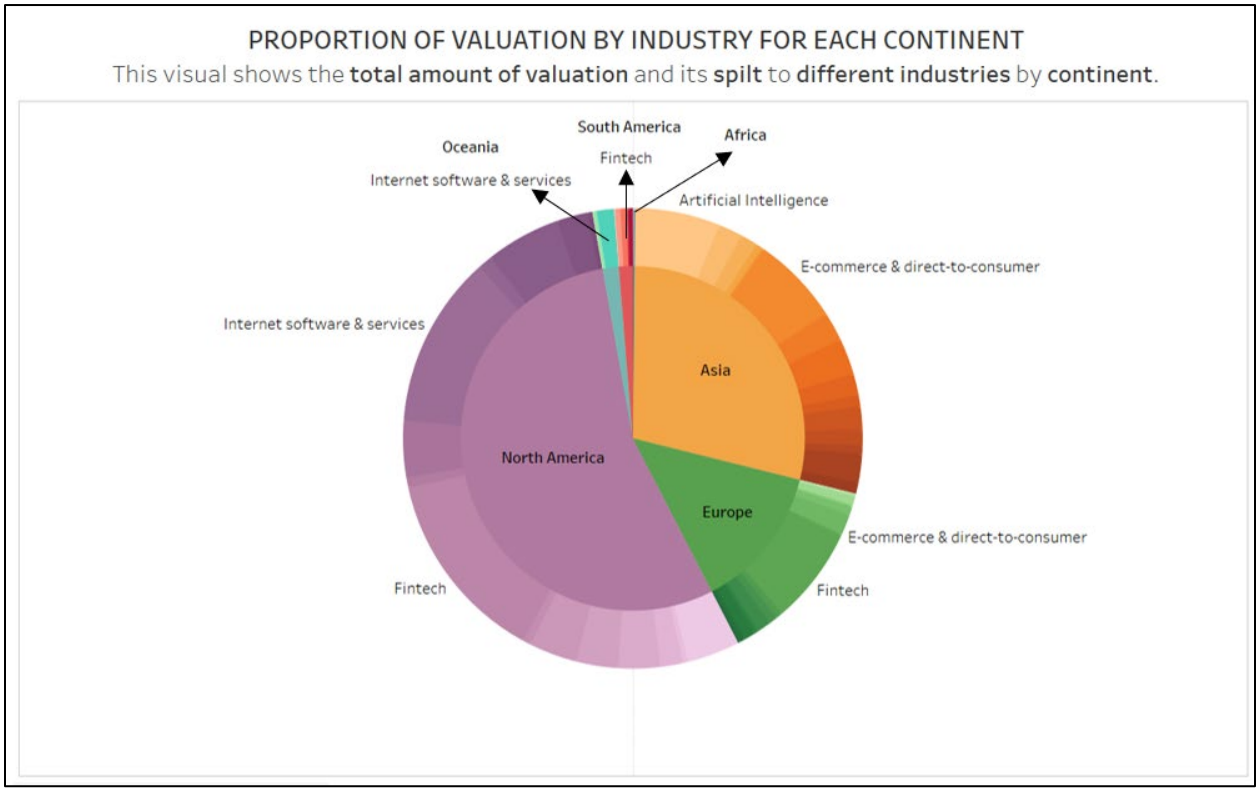


***Figure 8: Proportion of valuation by industry for each continent***

The given visualization depicts the continents with the highest-valued industries across the world. North America has the highest number of unicorn companies, which make up around 50% of the total companies. The valuation of these companies is dominated by two industries, Fintech and Internet Software & Services. On the other hand, Asia is the second highest continent where E-commerce & direct-to-consumer and Artificial Intelligence industries dominate the market. The third highest continent, Europe, is also dominated by Fintech and E-commerce & direct-to-consumer industries. The remaining continents, including Oceania, South America, and Africa, have fewer unicorn companies than the other continents. Oceania's unicorn companies are mainly in the Internet Software and Service industry, while South America's unicorn companies are mostly in the Fintech industry.

To summarize, North America leads the world in terms of the number of unicorn companies, and its valuation is primarily driven by Fintech and Internet Software & Services. Asia and Europe follow, with their valuations being driven by E-commerce & direct-to-consumer, Artificial Intelligence, Fintech, and E-commerce & direct-to-consumer industries. The other continents have fewer unicorn companies, with Oceania being mainly dominated by the Internet Software and Service industry, while South America is primarily dominated by Fintech.

## 7. WORLD DISTRIBUTION OF UNICORN ORIGIN

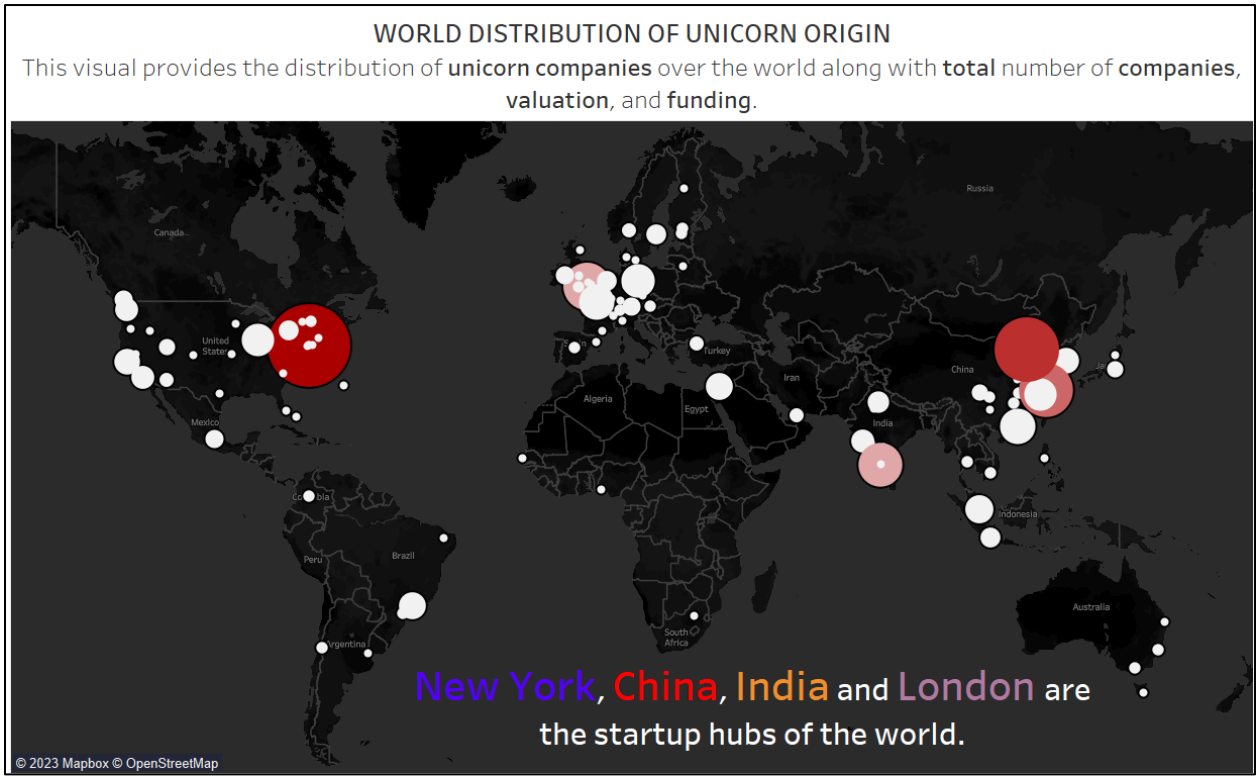


***Figure 9: World distribution of Unicorn origin***

The visualization displays the locations with the highest number of unicorn companies. The size of each bubble represents the number of companies originating from that specific area/city/state. New York in North America has the highest number of unicorn companies with 206, followed by Beijing and Shanghai in China, London in Europe, and Bengaluru in India. These areas can be considered as startup hubs of the world due to the high number of successful unicorn companies originating from there.

Unicorn companies are privately-owned startups that are valued at over $1 billion. The large number of unicorn companies in these areas suggests that they provide a favorable environment for startups to grow and succeed. It can be inferred that these areas have well-developed infrastructure, support systems, and a large pool of talented individuals.

This visualization is useful for understanding the concentration of successful startups in specific areas and can aid entrepreneurs in deciding where to establish their own businesses.

8. INVESTOR – COMPANY NETWORK GRAPH

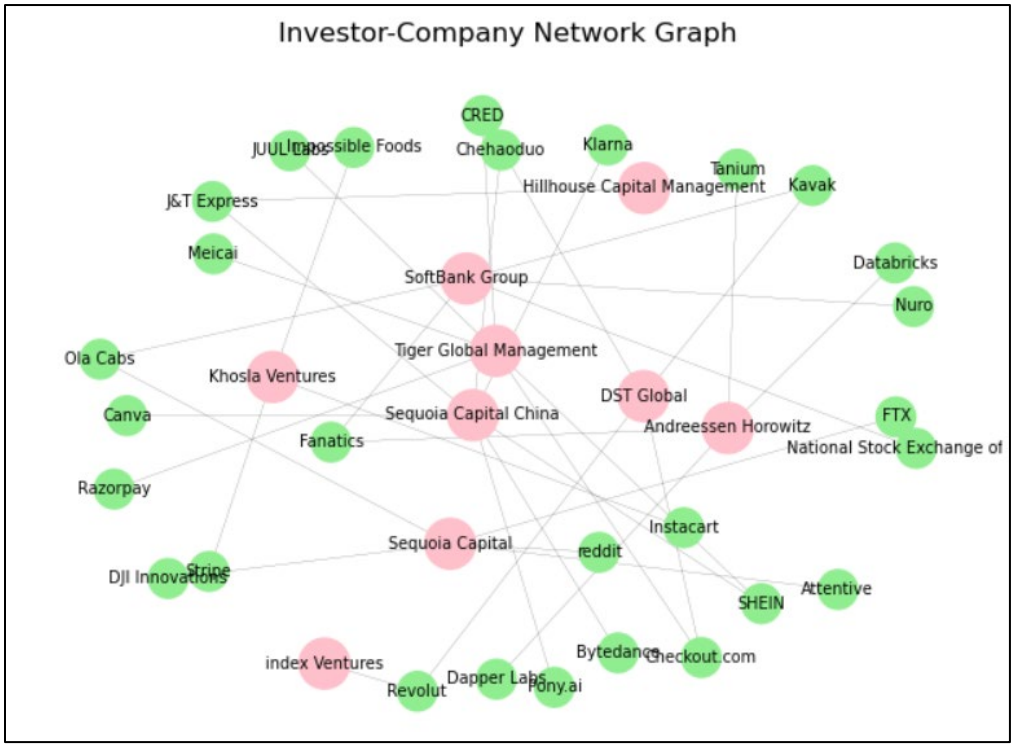


***Figure 10: Investor-Company network graph***

This visualization is a graph created using the Python programming language and the NetworkX library. The nodes in the graph represent investors, and the green nodes represent the companies they have invested in. The graph helps us understand which investors are investing in which companies.

We can observe that Sequoia Capital China, Tiger Global Management, and SoftBank Group have invested in the most unicorn companies, such as Bytedance, Instacart, Medicai, J&T Express, Ola Cabs, and Fanatics, among others. This visualization is useful for identifying the top investors in the startup world and the companies they have invested in.

It can help entrepreneurs and startup founders determine which investors to approach for funding, based on their investment history and the companies they have supported in the past.

## ❖ Conclusion

→ The analysis has revealed that Fintech, Internet Software and Services, and E-commerce & Direct-to-Consumer industries have raised the most funding and achieved the highest valuations.
→ We also identified the leading companies in each industry, such as Stripe, Klarna, and Checkout.com in Fintech, Shein in E-commerce, and Canva in Internet Software and Services.
→ Additionally, we explored the average time it takes for companies to reach unicorn status in different industries and the top investors providing funding to these companies.
→ We highlighted the dominant players in the startup landscape across various industries.
→ We found that top investors primarily focus on technology-driven industries such as Fintech, E-commerce & direct-to-consumer, Artificial Intelligence, and Internet Software & Services.
→ We highlighted the dominance of North America, Asia, and Europe in terms of the number of unicorn companies and their valuation.
→ We showcased the startup hubs of the world, with New York, Beijing, Shanghai, London, and Bengaluru having the highest number of unicorn companies.
→ We also provided insights into the top investors and the companies they have invested in, which can aid entrepreneurs in securing funding for their own startups.