****

**Certificate in Junior Python Data Analyst**

**Project Title: Data Analysis on Unicorn Companies**

**Lecturer: Ivan LAU**

**Members: Pang**

**Leung Chi Wa 21**

**Date: 18 July 2025**

Table of Contents

[1. Abstract 4](#_Toc202394131)

[2. Executive Summary 4](#_Toc202394132)

[3. Objectives 5](#_Toc202394133)

[4. Methodology 5](#_Toc202394134)

[5. Expect Outcomes 5](#_Toc202394135)

[6. Exploratory Data Analysis 7](#_Toc202394136)

[6.1 Industry-Based Analysis 7](#_Toc202394137)

[7](#_Toc202394138)

[**Bar Chart Analysis: *Distribution of Companies by Industry*** 8](#_Toc202394139)

[**Key Observations** 8](#_Toc202394140)

[**Enterprise Tech (32.5%)** 8](#_Toc202394141)

[**Financial Services (18.2%)** 8](#_Toc202394142)

[**Industrials (16.2%)** 8](#_Toc202394143)

[**Consumer & Retail (14.7%)** 8](#_Toc202394144)

[**Healthcare & Life Sciences (9.5%)** 8](#_Toc202394145)

[**Media & Entertainment (6.8%)** 9](#_Toc202394146)

[**Insurance (2.0%)** 9](#_Toc202394147)

[**Conclusion** 9](#_Toc202394148)

[6.1.2 Distribution of Equity Funding across Different Industries 10](#_Toc202394149)

[**Left Chart: Total Equity Funding ($B)** 10](#_Toc202394150)

[**Insights:** 10](#_Toc202394151)

[**Right Chart: Mean Equity Funding per Company ($M)** 10](#_Toc202394152)

[**Conclusion** 11](#_Toc202394153)

[6.2 Geographical Analysis:- Top Countries by Valuation 12](#_Toc202394154)

[**Top Countries by Startup Valuation & Company Count** 12](#_Toc202394155)

[**Chart 1: Latest Valuation (in B$)** 12](#_Toc202394156)

[**Chart 2: Number of Companies** 13](#_Toc202394157)

[**Key Insights** 13](#_Toc202394158)

[**United States** 13](#_Toc202394159)

[**China** 13](#_Toc202394160)

[**India** 13](#_Toc202394161)

[**Europe (UK, Germany, France)** 13](#_Toc202394162)

[**Others (e.g., Israel, Brazil, Singapore)** 13](#_Toc202394163)

[**Conclusion** 14](#_Toc202394164)

[6.3 Sector-Based Analysis 15](#_Toc202394165)

[**Chart 1: Total Valuation by Sector** 15](#_Toc202394166)

[**Insights:** 16](#_Toc202394167)

[**Chart 2: Number of Companies by Sector** 16](#_Toc202394168)

[**Insights:** 16](#_Toc202394169)

[**Valuation Efficiency: Average Valuation per Company** 17](#_Toc202394170)

[**Conclusion** 17](#_Toc202394171)

[6.3 Company-Based with Growth Rate Analysis 18](#_Toc202394172)

[**Top Companies by Valuation (2024 vs. 2025)** 18](#_Toc202394173)

[**Growth Rate Highlights (Bottom Section)** 19](#_Toc202394174)

[**Observations:** 19](#_Toc202394175)

[**Conclusion** 19](#_Toc202394176)

[7. Significance 21](#_Toc202394177)

[8. Results and Visualization 21](#_Toc202394178)

[9. Conclusion 21](#_Toc202394179)

[10. Appendix 21](#_Toc202394180)

# 1. Abstract

This project involves performing a comprehensive data analysis of unicorn companies—privately held startups valued at $1 billion or more—using Python, leveraging techniques and concepts learned during our lectures. The main objective is to harness Python’s capabilities to create a diverse set of visualizations, such as charts and graphs, to effectively present key findings and trends related to unicorn companies. These visualizations will highlight critical aspects like industry distribution, geographic concentration, valuation trends, and funding patterns, providing a clear and engaging representation of the data. By applying Python libraries such as pandas for data manipulation, matplotlib and seaborn for plotting, and potentially numpy for numerical computations, the project aims to uncover meaningful insights into the unicorn ecosystem. The resulting graphs will serve as powerful tools for communicating complex data in an accessible format, suitable for presentations, reports, or further analysis. Additionally, the project may explore relationships between variables, such as the correlation between funding and valuation or the growth of unicorns over time, to offer a deeper understanding of the factors driving the success of these high-value startups.

# 2. Executive Summary

This project aims to conduct a comprehensive data analysis of unicorn companies—privately held startups valued at $1 billion or more—using Python, leveraging skills acquired through academic coursework. The primary objective is to utilize Python’s robust data analysis and visualization libraries, such as pandas, matplotlib, and seaborn, to generate insightful visualizations that reveal trends, patterns, and relationships within the unicorn ecosystem. The analysis will focus on key metrics, including company valuations, industries, geographic distribution, founding years, and funding amounts, to provide a clear understanding of the factors driving the success of these high-value startups.

# 3. Objectives

1. **Data Exploration and Cleaning**: Import and preprocess a dataset of unicorn companies, addressing missing values, standardizing formats, and ensuring data quality for accurate analysis.
2. **Visualization Development**: Create a variety of graphs, such as bar plots, histograms, box plots, scatter plots, and line charts, to illustrate the distribution of unicorns by industry, country, valuation trends, funding patterns, and growth over time.
3. **Insight Generation**: Identify key trends, such as dominant industries, leading geographic hubs, and correlations between funding and valuation, to inform stakeholders about the unicorn landscape.
4. **Presentation of Findings**: Produce clear, professional visualizations suitable for inclusion in reports, presentations, or strategic discussions, making complex data accessible to diverse audiences.

# 4. Methodology

The project will use a structured approach:

1. **Data Acquisition**: Source a dataset of unicorn companies from Kaggle containing columns like company name, valuation, industry, country, founding year, and total funding.
2. **Data Processing**: Employ pandas to clean and preprocess the data, handling missing values, converting data types, and removing duplicates to ensure reliability.
3. **Exploratory Data Analysis (EDA)**: Conduct initial analysis to summarize key statistics, such as mean valuations, industry counts, and geographic distributions.
4. **Visualization**: Generate graphs using matplotlib and seaborn, including:
   * Bar plots to show the number of unicorns by industry and country.
   * Histograms to depict valuation distributions.
   * Box plots to compare valuations across industries.
   * Scatter plots to explore relationships between funding and valuation.
   * Line charts to track the growth of unicorns over time.
5. **Interpretation**: Summarize findings, highlighting actionable insights, such as emerging industries, key markets, or factors influencing high valuations.

# 5. Expect Outcomes

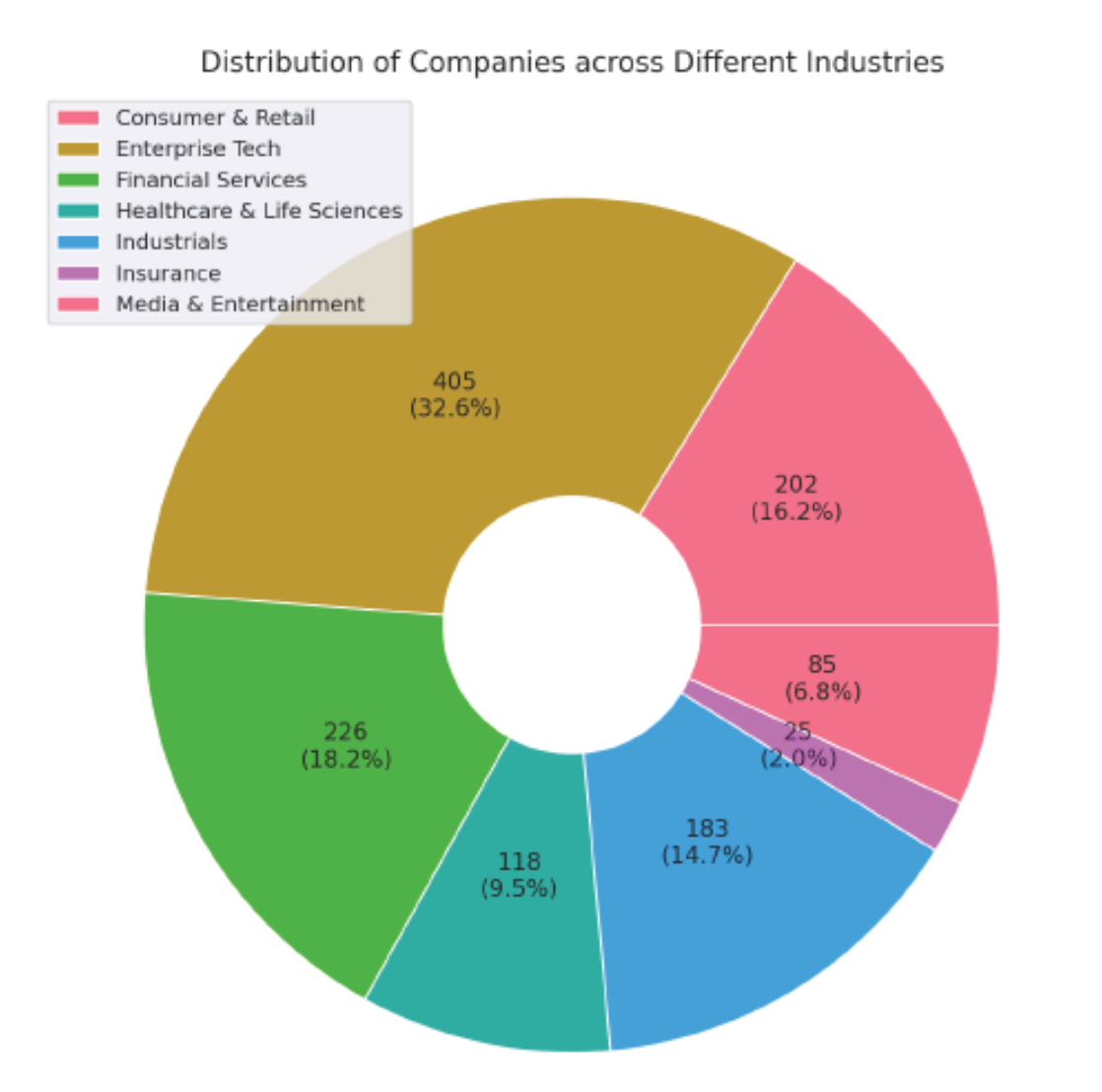
The project will deliver a comprehensive set of visualizations and insights that illuminate the unicorn company landscape. Key deliverables include:

* A cleaned and processed dataset ready for analysis.
* A series of professional-grade graphs saved as JPG files for presentations or reports.
* A summary of findings, identifying dominant industries (e.g., FinTech, AI), leading countries (e.g., USA, China), valuation trends, and potential correlations between funding and valuation.
* Recommendations for stakeholders, such as investors or policymakers, based on observed trends, such as focusing on high-growth sectors or regions.

# 6. Exploratory Data Analysis

# 6.1 Industry-Based Analysis

6.1.1 Distribution of Companies across Different Industries

****

**Bar Chart Analysis: *Distribution of Companies by Industry***

The chart displays the **number of companies** in each industry out of a total of **1,244 companies**. Here's the corrected breakdown:

| **Industry** | **Number of Companies** | **Percentage (%)** |
| --- | --- | --- |
| Enterprise Tech | 405 | 32.5% |
| Financial Services | 226 | 18.2% |
| Industrials | 202 | 16.2% |
| Consumer & Retail | 183 | 14.7% |
| Healthcare & Life Sciences | 118 | 9.5% |
| Media & Entertainment | 85 | 6.8% |
| Insurance | 25 | 2.0% |

**Key Observations**

**Enterprise Tech (32.5%)**

* The largest segment by far.
* Reflects the dominance of software, cloud, AI, and digital infrastructure companies.
* Indicates high scalability and investor interest in tech-driven solutions.

**Financial Services (18.2%)**

* A strong second place.
* Includes fintech, payments, lending, and wealth management platforms.
* Suggests ongoing disruption of traditional finance.

**Industrials (16.2%)**

* A surprisingly large share.
* Likely includes logistics, manufacturing tech, robotics, and supply chain platforms.
* Shows that innovation in physical infrastructure is gaining traction.

**Consumer & Retail (14.7%)**

* Slightly smaller than expected.
* Indicates a more selective or mature market compared to tech and finance.

**Healthcare & Life Sciences (9.5%)**

* A steady presence.
* Reflects the complexity and regulatory intensity of the sector, despite high impact potential.

**Media & Entertainment (6.8%)**

* A modest slice.
* May reflect challenges in monetization, IP, and platform saturation.

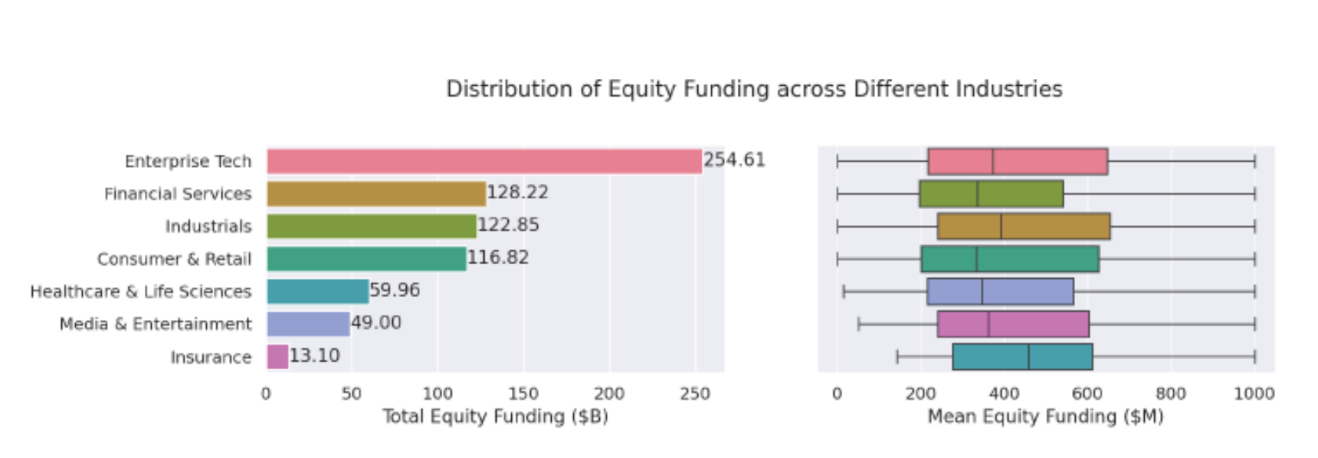
**Insurance (2.0%)**

* The smallest category.
* Suggests high barriers to entry or slower innovation cycles in insurance.

**Conclusion**

* **Enterprise Tech dominates** the startup landscape, accounting for nearly one-third of all companies—underscoring its central role in modern innovation.
* **Financial Services and Industrials** are also robust, reflecting strong momentum in fintech and infrastructure modernization.
* **Consumer & Retail and Healthcare** remain important but are relatively smaller, possibly due to market maturity or regulatory complexity.
* **Media & Entertainment and Insurance** are niche sectors with fewer players, potentially signaling untapped opportunities or structural challenges.

# 6.1.2 Distribution of Equity Funding across Different Industries



**Left Chart: Total Equity Funding ($B)**

This chart shows how much total capital each industry has raised:

| **Industry** | **Total Funding ($B)** |
| --- | --- |
|  |  |
| Enterprise Tech | 254.61 |
| Financial Services | 128.22 |
| Industrials | 122.85 |
| Consumer & Retail | 116.82 |
| Healthcare & Life Sci. | 59.96 |
| Media & Entertainment | 49.00 |
| Insurance | 13.10 |

**Insights:**

* **Enterprise Tech dominates**, raising more than double the next-highest industry.
* **Financial Services, Industrials, and Consumer & Retail** are in a tight second tier, each raising over $100B.
* **Healthcare & Life Sciences** and **Media & Entertainment** are mid-tier.
* **Insurance** lags significantly behind, with just $13.1B in total funding.

**Right Chart: Mean Equity Funding per Company ($M)**

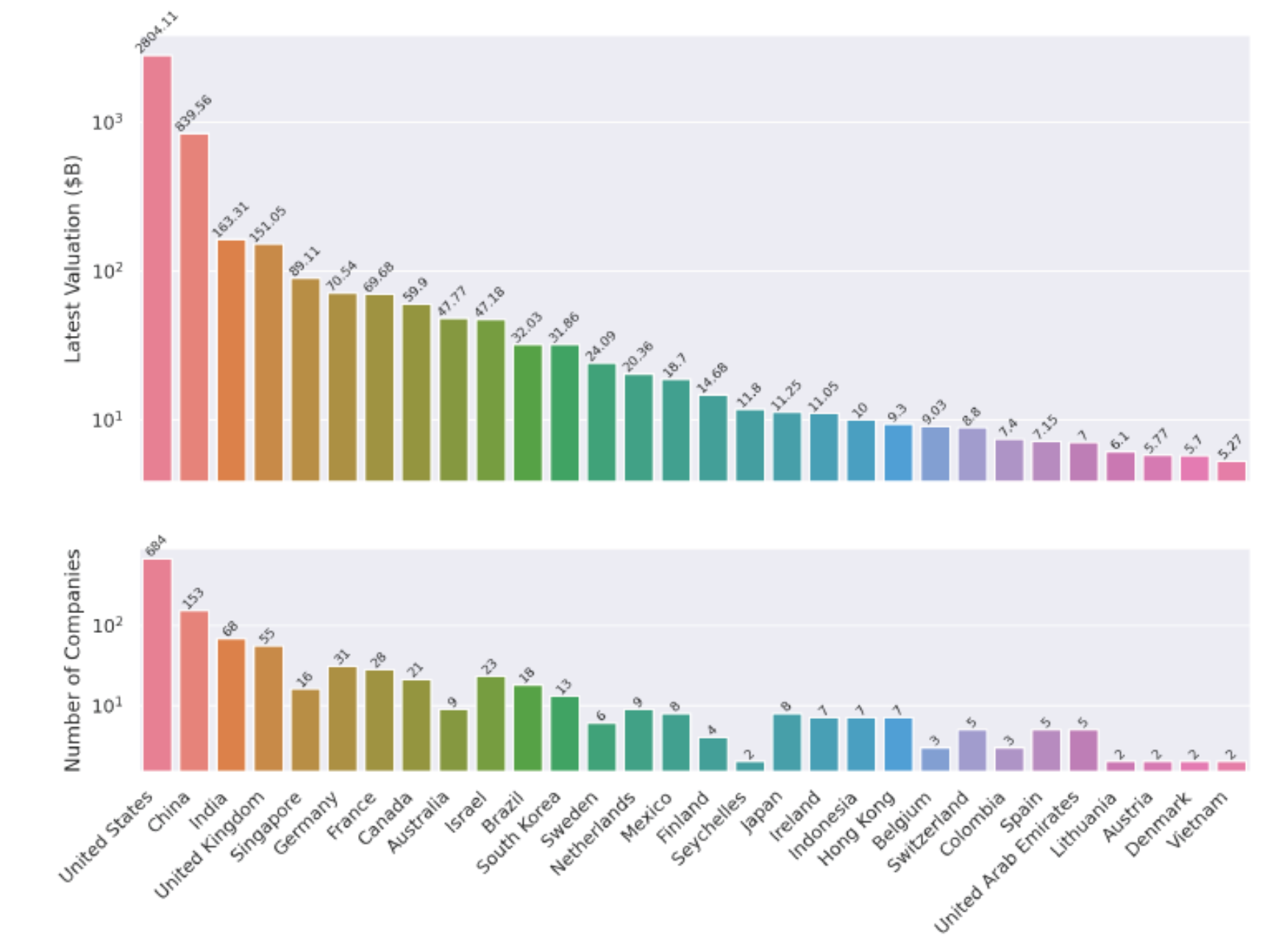
This chart shows the **average funding per company** in each industry.

* **Enterprise Tech** likely has a high average due to large rounds.
* **Insurance**, despite low total funding, may have a relatively high mean if there are few companies.
* **Consumer & Retail** may have a lower mean, suggesting many companies with smaller rounds.

**Conclusion**

* **Enterprise Tech** is the clear leader in both total and average funding—indicating strong investor confidence and large-scale capital deployment.
* **Financial Services and Industrials** are also major funding recipients, reflecting their importance in infrastructure and fintech innovation.
* **Consumer & Retail** has high total funding but likely lower average funding per company—suggesting a fragmented market with many small players.
* **Insurance** has the lowest total funding but may have a decent average, implying fewer but more capital-intensive startups.
* **Healthcare & Media** sit in the middle—important sectors, but with more modest capital flows.

# 6.2 Geographical Analysis:- Top Countries by Valuation



**Top Countries by Startup Valuation & Company Count**

**Chart 1: Latest Valuation (in B$)**

| **Country** | **Valuation (B$)** |
| --- | --- |
| United States | 2804.11 |
| China | 839.56 |
| India | 163.31 |
| UK | 151.05 |
| Germany | 70.54 |
| France | 69.68 |

* **The U.S. dominates** with nearly **3x** the valuation of China and over **13x** that of India.
* **China** is a strong second, but there's a steep drop-off after that.
* **Europe** (UK, Germany, France) holds a respectable middle tier.

**Chart 2: Number of Companies**

| **Country** | **Companies** |
| --- | --- |
| United States | 684 |
| China | 153 |
| India | 68 |
| UK | 55 |
| Germany | 31 |
| France | 28 |

* The **U.S. again leads** by a wide margin, hosting over **50%** of all companies shown.
* **China and India** follow, reflecting their growing startup ecosystems.
* **European countries** have fewer companies but still contribute significant valuation.

**Key Insights**

**United States**

* Leads in both valuation and company count.
* Suggests not just quantity but also high-value companies (e.g., unicorns, decacorns).

**China**

* Strong second in both metrics.
* High average valuation per company, indicating a mature and well-funded ecosystem.

**India**

* Fewer companies than China but still ranks 3rd in valuation.
* Indicates rising influence in global tech and startup innovation.

**Europe (UK, Germany, France)**

* Moderate number of companies with solid valuations.
* Suggests efficient capital deployment and possibly more late-stage companies.

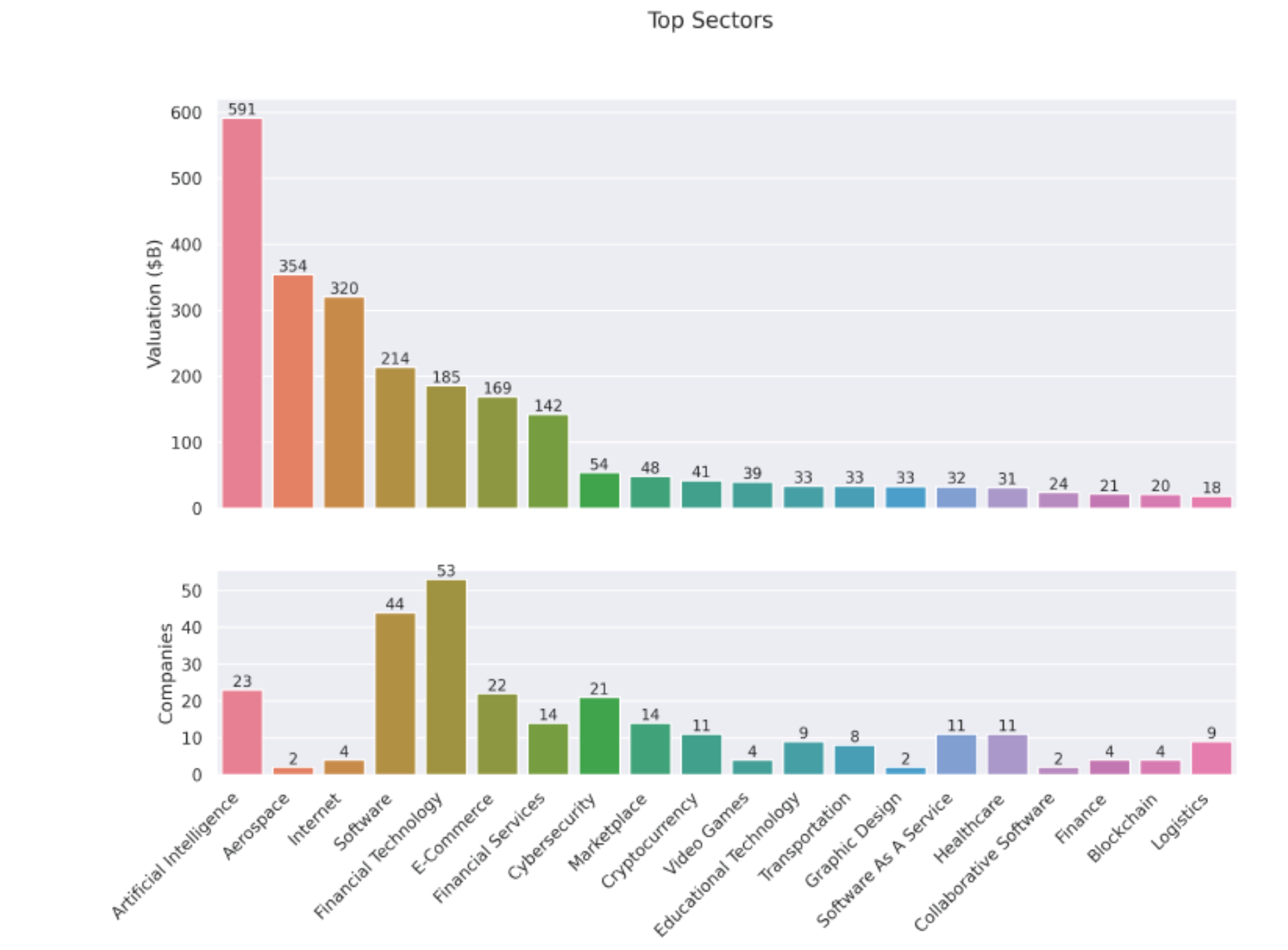
**Others (e.g., Israel, Brazil, Singapore)**

* Smaller ecosystems but still producing high-value companies.
* May represent emerging innovation hubs.

**Conclusion**

* The **U.S. is the undisputed leader** in both startup volume and valuation.
* **China and India** are rapidly scaling, with China showing higher capital concentration.
* **Europe** maintains a strong presence with fewer but valuable companies.
* **Emerging markets** like Israel, Brazil, and Southeast Asia are punching above their weight in valuation per company.

# 6.3 Sector-Based Analysis



1. **Chart 1: Total Valuation ($B)**
2. **Chart 2: Number of Companies**

These charts help us understand which sectors are attracting the most capital and where the highest concentration of startups lies.

**Chart 1: Total Valuation by Sector**

This chart ranks sectors by the **cumulative valuation** of companies within each.

| **Rank** | **Sector** | **Valuation ($B)** |
| --- | --- | --- |
| 1 | Artificial Intelligence | 591 |
| 2 | Aerospace | 354 |
| 3 | Internet | 320 |
| 4 | Software | 214 |
| 5 | Financial Technology | 185 |
| 6 | E-Commerce | 169 |
| 7 | Financial Services | 109 |
| 8 | Cybersecurity | 142 |
| 9 | Marketplace | 54 |
| 10 | Cryptocurrency | 41 |

**Insights:**

* **Artificial Intelligence** leads by a wide margin, reflecting the explosive growth and investor enthusiasm in generative AI, machine learning, and automation.
* **Aerospace** and **Internet** follow, likely driven by a few high-value players (e.g., SpaceX, Starlink, or major platform companies).
* **Software, Financial Technology, and E-Commerce** show strong valuations, indicating their strategic importance in a post-pandemic, digitally connected world.

**Chart 2: Number of Companies by Sector**

This chart shows how many companies are active in each sector.

| **Rank** | **Sector** | **Number of Companies** |
| --- | --- | --- |
| 1 | Financial Technology | 53 |
| 2 | Software | 44 |
| 3 | Artificial Intelligence | 23 |
| 4 | E-Commerce | 22 |
| 5 | Cybersecurity | 21 |
| 6 | Financial Services | 14 |
| 7 | Marketplace | 14 |
| 8 | Cryptocurrency | 11 |
| 9 | Internet | 4 |
| 10 | Aerospace | 2 |

**Insights:**

* **Fintech and Software** dominate in company count, suggesting lower barriers to entry and broad applicability.
* **AI**, despite being #1 in valuation, has fewer companies—indicating high capital concentration per firm.
* **Aerospace and Internet** have very few companies but extremely high valuations, likely skewed by a few mega-unicorns.

**Valuation Efficiency: Average Valuation per Company**

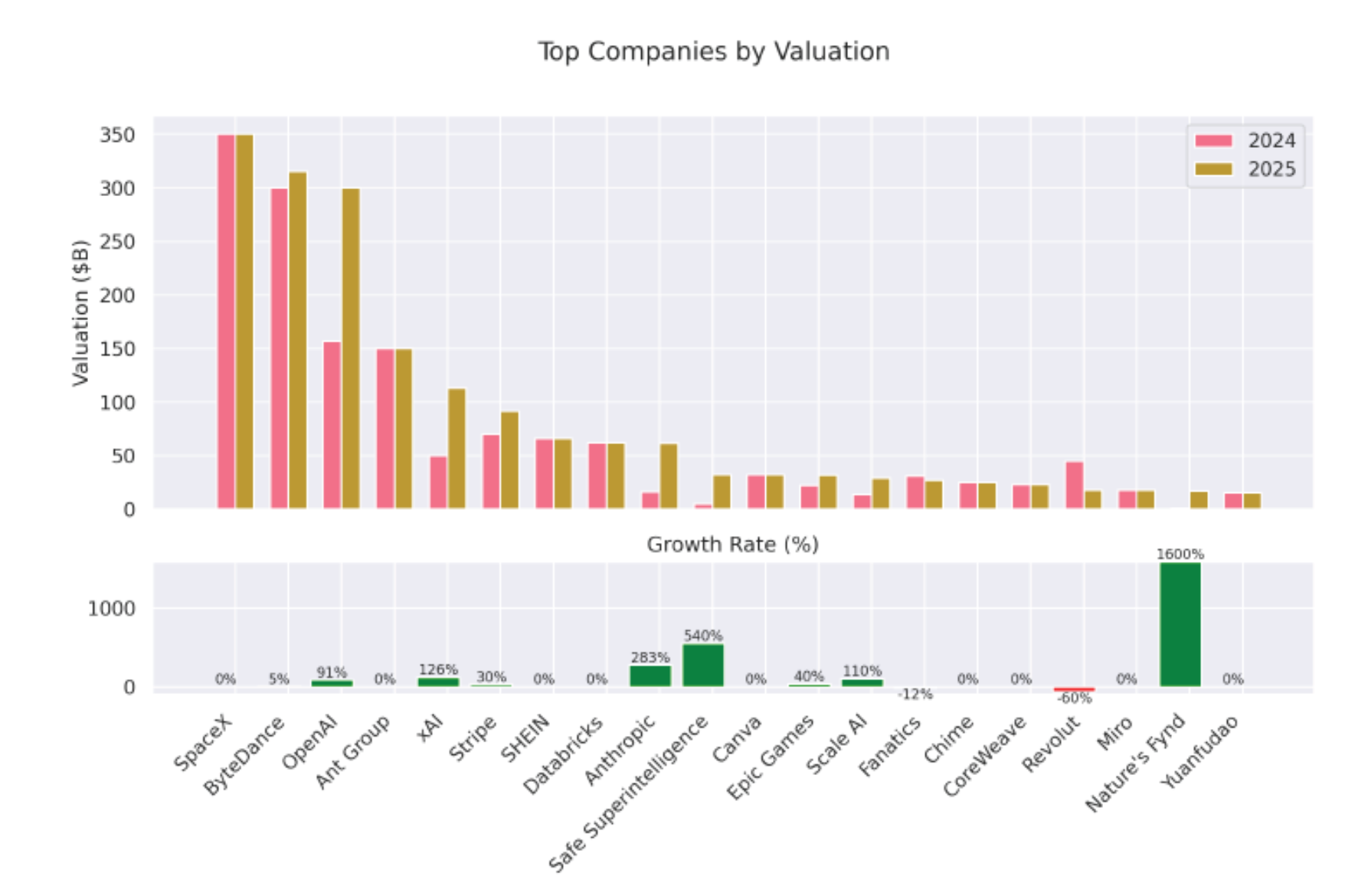
Let’s calculate how much valuation each company commands on average:

| **Sector** | **Valuation ($B)** | **Companies** | **Avg. Valuation/Company ($B)** |
| --- | --- | --- | --- |
| Aerospace | 354 | 2 | **177.0** |
| Internet | 320 | 4 | **80.0** |
| Artificial Intelligence | 591 | 23 | **25.7** |
| Health | 167 | 20 | 8.4 |
| Biotechnology | 127 | 19 | 6.7 |
| Cybersecurity | 139 | 21 | 6.6 |
| Transportation | 132 | 17 | 7.8 |
| Software | 214 | 44 | 4.9 |
| E-Commerce | 157 | 22 | 7.1 |
| Financial Technology | 185 | 53 | **3.5** |

**Conclusion**

* **Artificial Intelligence is the most valuable sector overall**, with a strong average valuation per company—showing both depth and scale.
* **Aerospace and Internet** are outliers: few companies, but massive valuations—likely driven by a handful of dominant players.
* **Fintech and Software** have the most companies but the lowest average valuation per firm, suggesting high competition and market saturation.
* **Health, Cybersecurity, and Biotech** strike a balance between company count and valuation, indicating steady investor confidence.

# 6.3 Company-Based with Growth Rate Analysis



**Top Companies by Valuation (2024 vs. 2025)**

The upper section of the chart compares company valuations in billions of dollars:

* **Top 5 by 2025 Valuation**:
  1. **SpaceX** – remains the highest-valued company, unchanged from 2024.
  2. **ByteDance** – slight increase (~5% growth).
  3. **OpenAI** – major jump, reflecting a 91% increase.
  4. **Ant Group** – valuation remains flat.
  5. **XAI** – huge leap, up 126%.
* **Other notable players**:
  1. **Anthropic** and **Safe Superintelligence** show explosive growth.
  2. **Stripe**, **Epic Games**, and **Scale AI** also post solid gains.
  3. **Fanatics** and **Revolut** decline in valuation, with Revolut dropping sharply.

**Growth Rate Highlights (Bottom Section)**

| **Company** | **Growth Rate (%)** |
| --- | --- |
| Nature’s Fynd | **1600%** |
| Safe Superintelligence | 540% |
| Anthropic | 283% |
| Scale AI | 110% |
| XAI | 126% |
| OpenAI | 91% |
| Stripe | 30% |
| Epic Games | 40% |
| Fanatics | -12% |
| Miro | -60% |
| Others | 0% |

**Observations:**

* **AI companies dominate growth**: OpenAI, Anthropic, XAI, Scale AI, and Safe Superintelligence all post triple-digit growth.
* **Nature’s Fynd** (a biotech/alt-protein company) is the outlier with a staggering 1600% growth—possibly due to a new funding round or breakthrough.
* **Fanatics** and **Miro** are the only companies with negative growth, suggesting valuation corrections or strategic setbacks.
* Several companies (e.g., SpaceX, SHEIN, Canva, Chime) show **0% growth**, indicating stable but stagnant valuations.

**Conclusion**

* **AI is the clear growth engine** in 2025, with multiple companies seeing valuations more than double.
* **Nature’s Fynd** stands out as a breakout success, likely due to a major funding event or market shift.
* **Stability at the top**: SpaceX and ByteDance remain dominant but show little to no growth.
* **Caution flags**: Miro’s -60% drop and Fanatics’ decline suggest market corrections or internal challenges.

**Remarks:**

* 1. Nature’s Fynd自然之源是一家利用黃石國家公園的微生物生產蛋白質的食品公司，創造了肉類和乳製品替代品 這種名為「Fusarium strain flavolapis」的微生物，發酵後會生成一種富含蛋白質的物質，名為「Fy」。 Nature 's Fynd 利用 Fy 開發了諸如素食早餐餅和無奶優格等產品。
  2. Safe Superintelligence Inc. 或 SSI Inc. 是一家美國人工智慧公司，由 Ilya Sutskever（OpenAI 前首席科學家）、Daniel Gross（蘋果 AI 前負責人）和 Daniel Levy（投資者兼 AI 研究員）創立。

# 7. Significance

This project demonstrates proficiency in Python-based data analysis and visualization, showcasing the ability to transform raw data into actionable insights. By analyzing unicorn companies, the project contributes to understanding the dynamics of high-growth startups, which are critical to innovation and economic development. The visualizations will serve as powerful tools for communicating findings to stakeholders, supporting strategic decision-making in investment, entrepreneurship, or policy development.

# 8. Results and Visualization

# 9. Conclusion

# 10. Appendix

**The End**