Unicorn Data Analysis Project

Presented By:
Srilekha

About me

I am Srilekha, a recent graduate with a Bachelor of Technology in Computer Science, specialized in Data Science , from Vaagdevi college of Engineering.

Why Data Science?

I chose Data Science because I honestly enjoy understanding problems and finding answers through data. It feels meaningful when those insights actually help people or improve a business.I'm always excited to learn more and explore new ideas in this fast-growing field.

Unicorn Dataset Analysis

❑ To analyze how **unicorn startups** (companies valued over $1 billion) vary by **industry, country, year founded, and valuation**

❑ The dataset helps understand **global startup growth trends, funding patterns, and innovation hotspots** across different sectors..

❑ The dataset contains **1,070 + rows and 10 columns**

❑ **Key Features Include**: Company Name, Industry, Country, Year Founded, Valuation (in USD billions), Investors, and Employees.

❑ Useful for tasks like **startup valuation prediction, trend analysis, regional investment insights,** and **sector-based comparison.**.

Understanding the Columns

Load the dataset

```
[3]: df=pd.read_csv(r"C:\Users\Asus\Documents\OneDrive\Desktop\Coursera\Labs of EDA All Modules C3\Labs of EDA Module 3\Unicorn_Companies.csv")
df
```

[3]:

	Company	Valuation	Date Joined	Industry	City	Country/Region	Continent	Year Founded	Funding	Select Investors
0	Bytedance	$180B	4/7/17	Artificial intelligence	Beijing	China	Asia	2012	$8B	Sequoia Capital China, SIG Asia Investments, S...
1	SpaceX	$100B	12/1/12	Other	Hawthorne	United States	North America	2002	$7B	Founders Fund, Draper Fisher Jurvetson, Rothen...
2	SHEIN	$100B	7/3/18	E-commerce & direct-to-consumer	Shenzhen	China	Asia	2008	$2B	Tiger Global Management, Sequoia Capital China...
3	Stripe	$95B	1/23/14	Fintech	San Francisco	United States	North America	2010	$2B	Khosla Ventures, LowercaseCapital, capitalG
4	Klarna	$46B	12/12/11	Fintech	Stockholm	Sweden	Europe	2005	$4B	Institutional Venture Partners, Sequoia Capita...
...	...	...	...	...	...	...	...	...	...	...
1069	Zhaogang	$1B	6/29/17	E-commerce & direct-to-consumer	Shanghai	China	Asia	2012	$379M	K2 Ventures, Matrix Partners China, IDG Capital
1070	Zhuan Zhuan	$1B	4/18/17	E-commerce & direct-to-consumer	Beijing	China	Asia	2015	$990M	58.com, Tencent Holdings
1071	Zihaiguo	$1B	5/6/21	Consumer & retail	Chongqing	China	Asia	2018	$80M	Xingwang Investment Management, China Capital ...
1072	Zopa	$1B	10/19/21	Fintech	London	United Kingdom	Europe	2005	$792M	IAG Capital Partners, Augmentum Fintech, North...
1073	Zwift	$1B	9/16/20	E-commerce & direct-to-consumer	Long Beach	United States	North America	2014	$620M	Novator Partners, True, Causeway Media Partners

▪ Handling Missing Data

Missing data in the unicorn dataset occurs when values in certain columns (like City , select Investors) are absent, reducing dataset completeness.

• Columns with the most missing values: City, Select Investors

• Imputing with forward or backward fill methods was avoided, as it could make these columns more biased. Since the number of missing values was very small, they were **dropped from the dataset** to preserve data accuracy.

▪ Handling Duplicates:

Duplicates are repeated records in the dataset that can inflate counts and bias analysis.

• Duplicate records were identified, including the company **"Bolt"** appearing twice due to manual entry errors.

Such duplicates were **removed** to maintain data integrity and ensure each unicorn is represented **only once**..

▪ Fixing Inconsistencies :

Inconsistencies occur when the some inconsistent symbols are present in columns like Valuation.

- They were **standardized** by removing "$" and "B" symbols, trimming spaces, and unifying text formats to ensure **clean and consistent categories** for analysis..
- Cleaned extra spaces, capitalization issues, and irregular values for consistency.
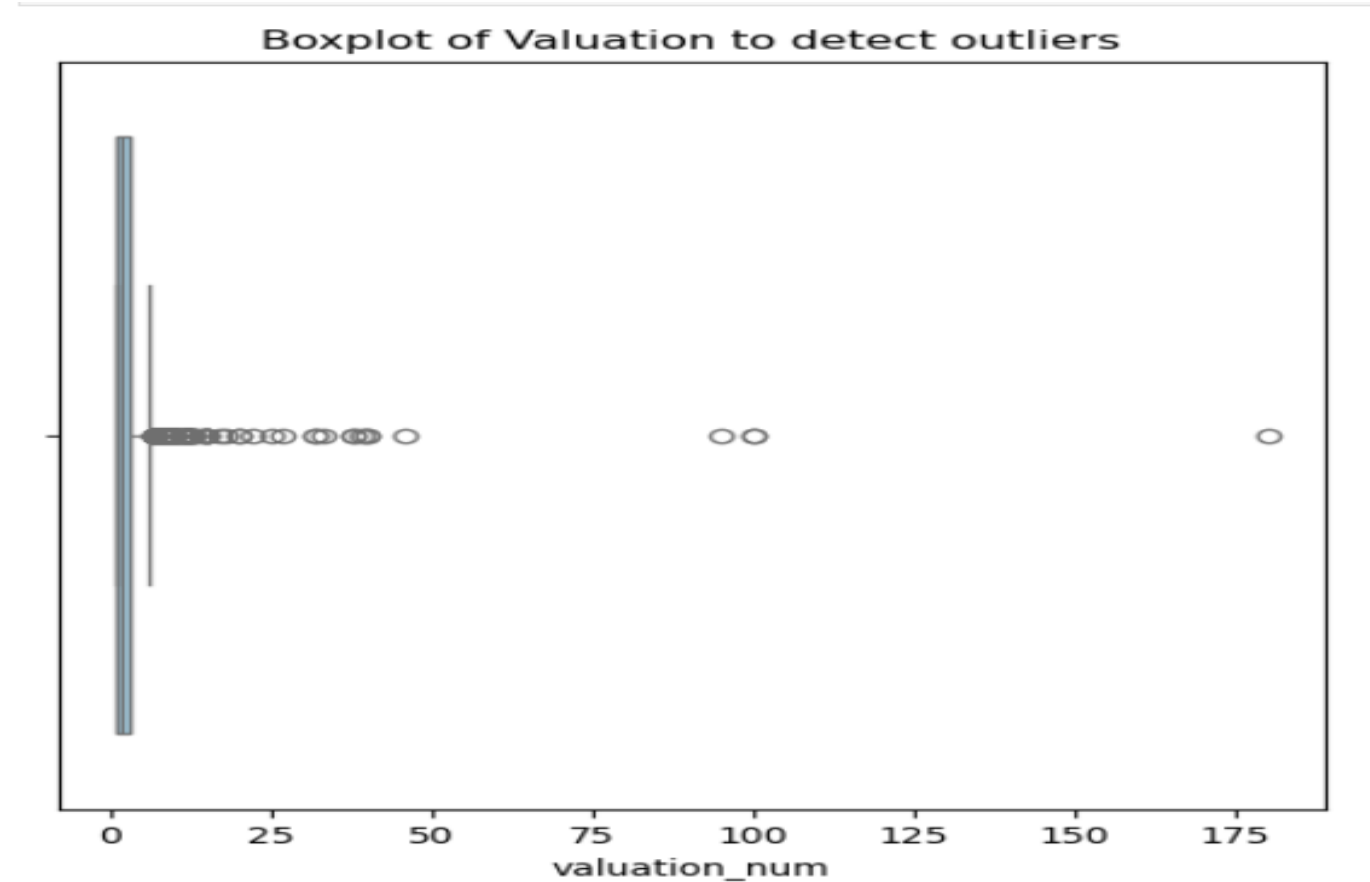
▪ Data Type Conversion:

Data type conversion ensures columns are stored in the correct format for accurate analysis.

- Converted Date Joined to datetime format for proper trend analysis.
- Extracted Year, Month from Year Joined and columns as object type.

.

Handling Outliers

Outliers are data points in the Netflix dataset that deviate significantly from other records and may affect analysis quality.

- The **majority of company valuations** are clustered at the lower end of the scale (close to the left side of the plot).

- The **data distribution is right-skewed**, meaning most unicorns have relatively modest valuations compared to a few ultra-high-value ones.

- The presence of several extreme outliers indicates **high inequality in valuation** among unicorns.
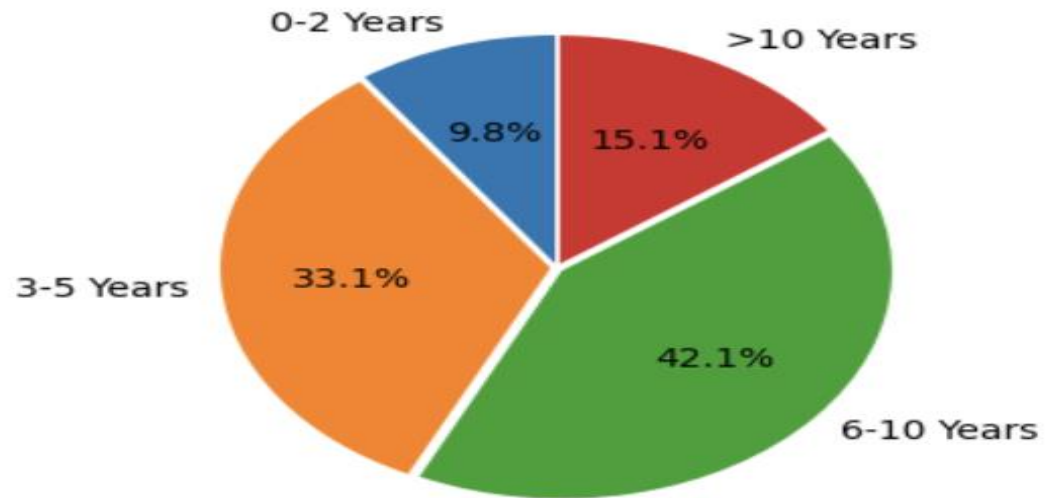

Boxplot of Valuation to detect outliers

■ Univariate Analysis

Univariate analysis examines a single variable at a time to understand its distribution and frequency.

- Most unicorns are created within 6–10 years of their founding
- The histogram shows that very few unicorns were founded before 2000, but their number has grown rapidly after 2010.
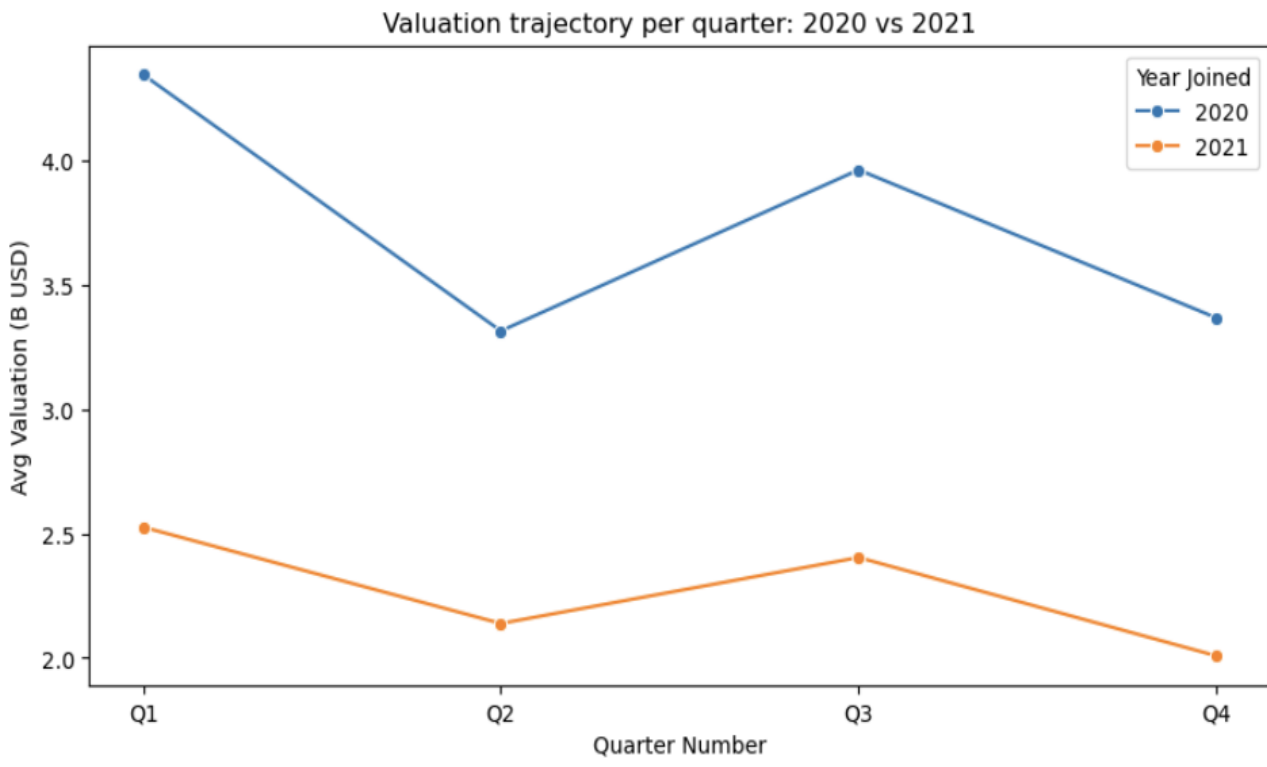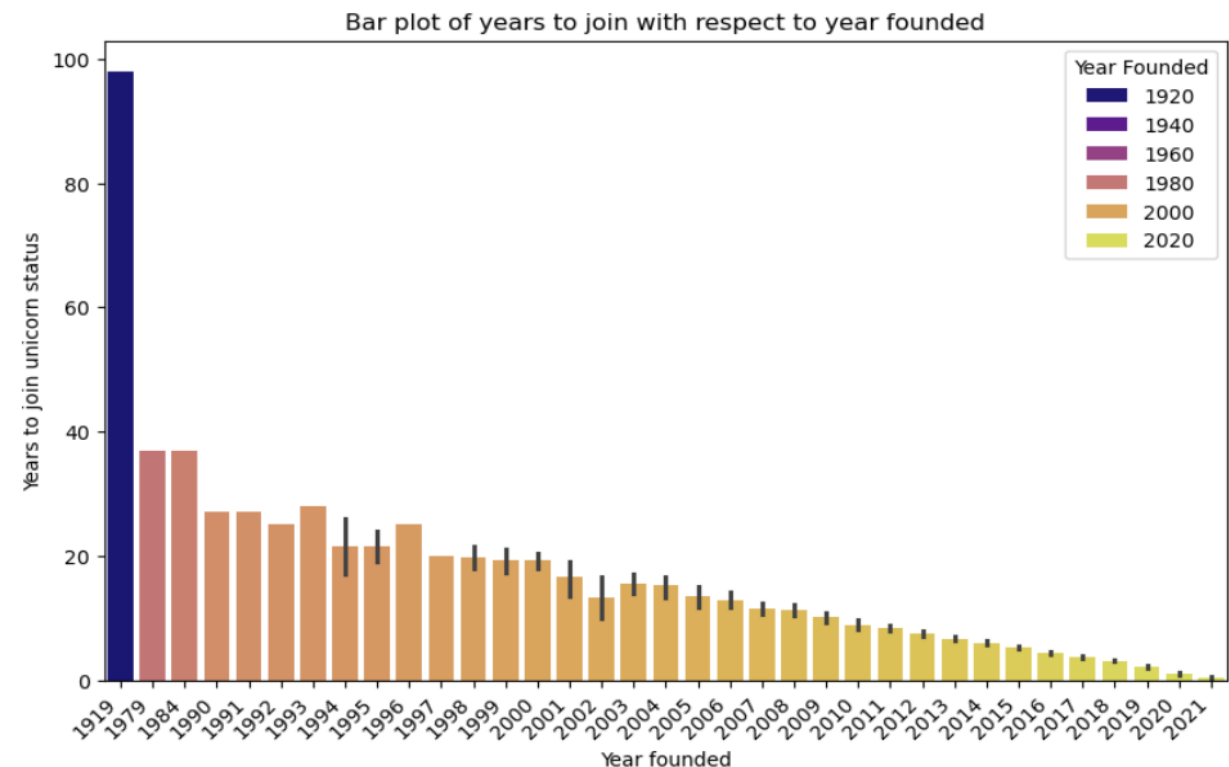

Unicorn Creation by Time Frame

Year Founded histogram

Bivariate Analysis

Bivariate analysis studies the relationship between two variables to identify associations, trends, or dependencies.
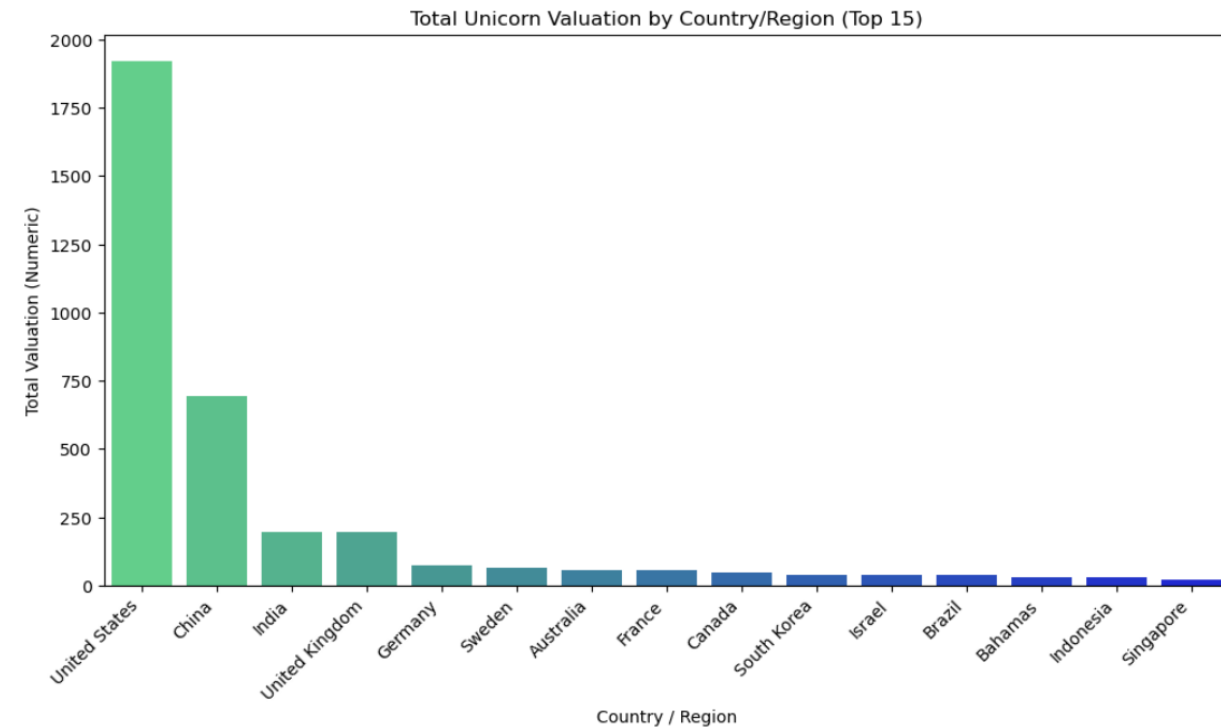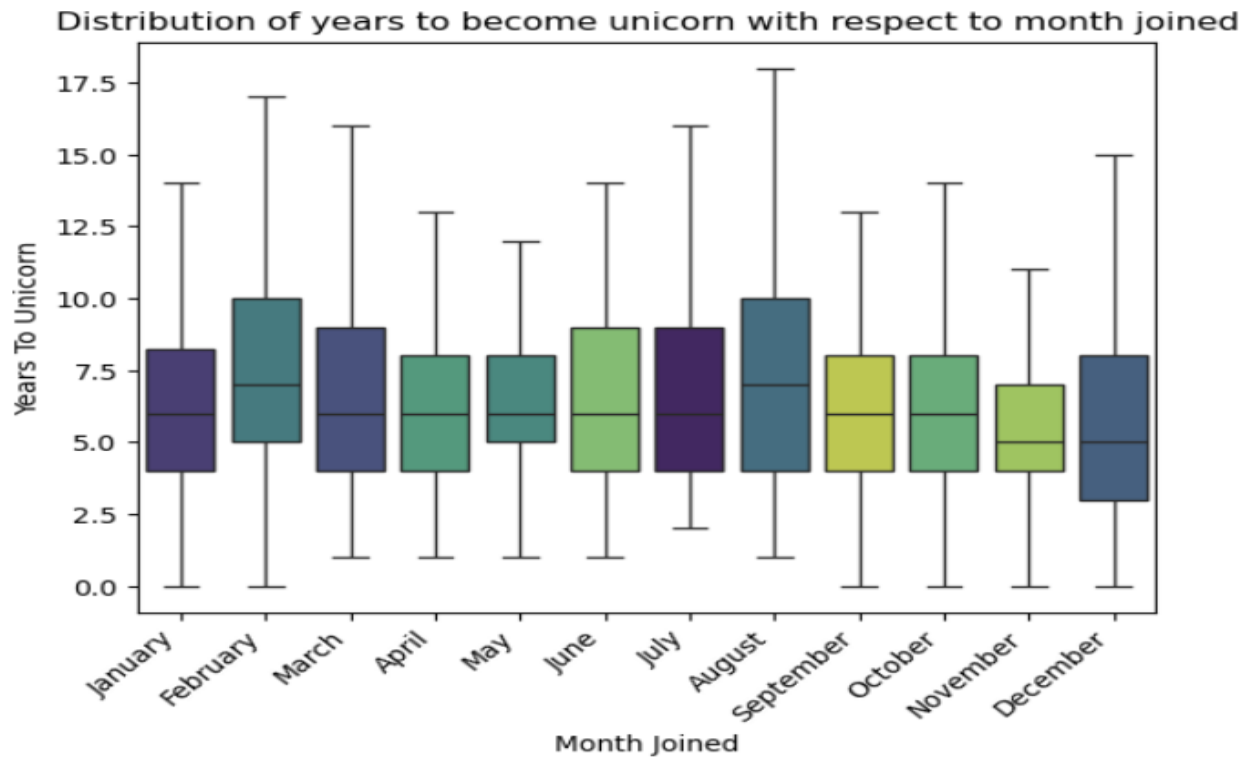
1. Numerical vs Numerical:

- Line chart comparing average valuations (in billions USD) by quarter for companies that joined in 2020 vs 2021.
- Bar plot showing how long companies took to reach unicorn status based on their founding year..
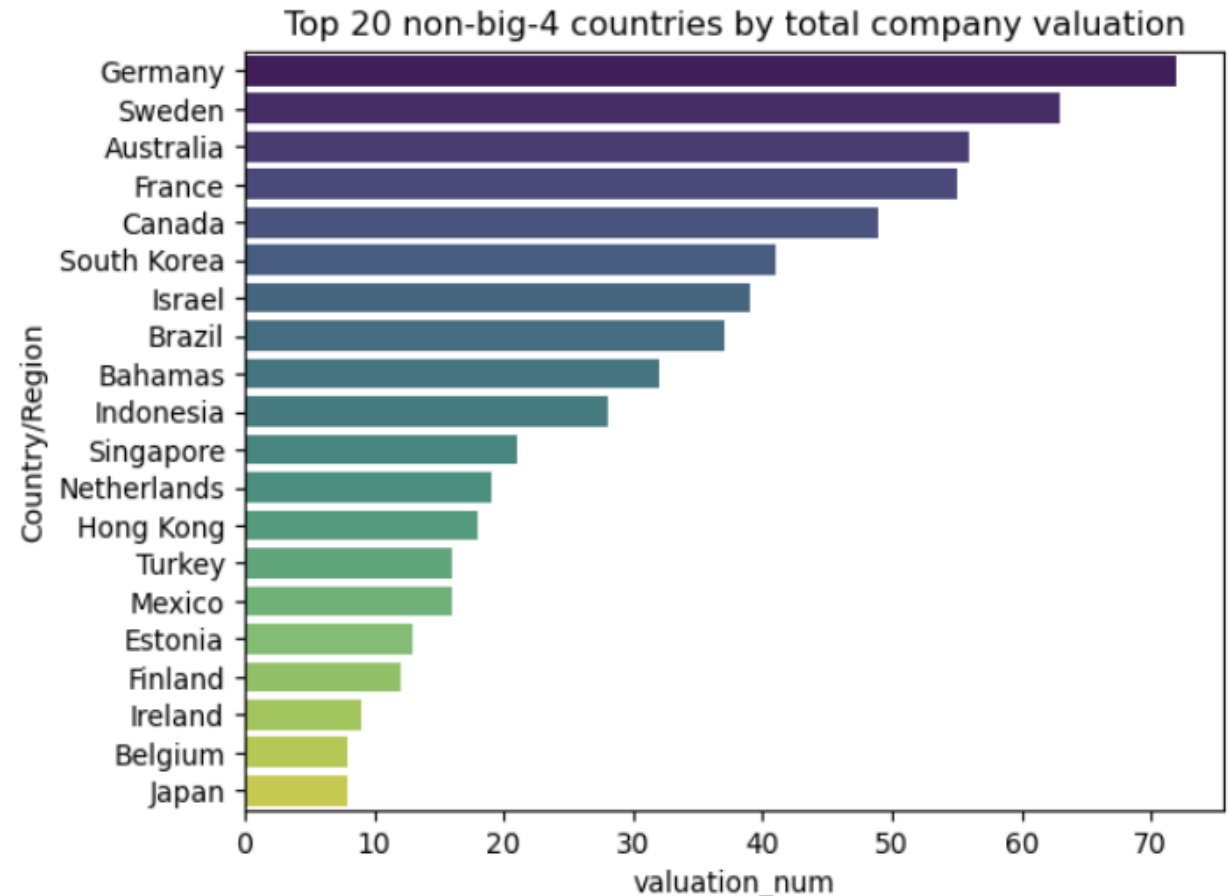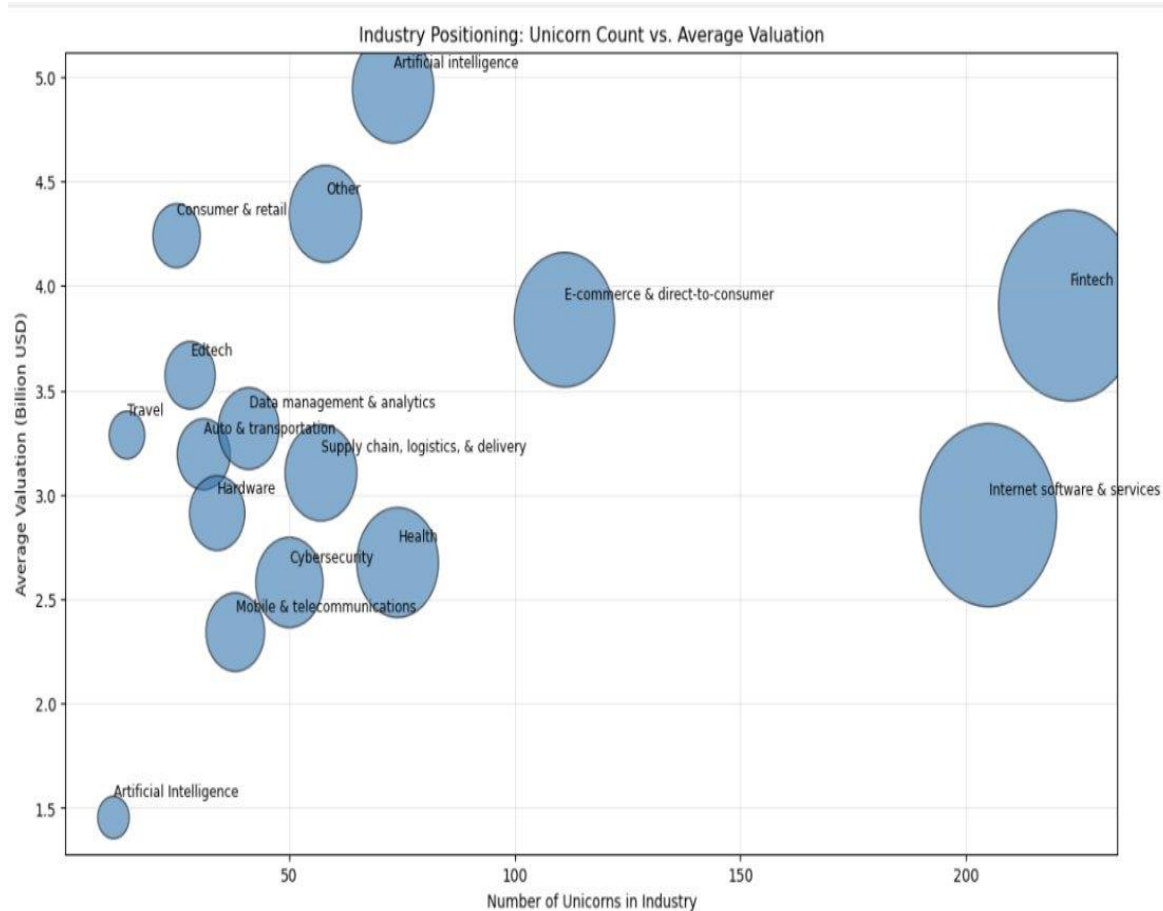
2. Numerical vs Categorical:

- Startups that joined in **October to December** reached unicorn status a bit **faster on average**, compared to those in earlier months.
- The **United States clearly leads** with the highest total unicorn valuation, far ahead of any other country.

Distribution of years to become unicorn with respect to month joined

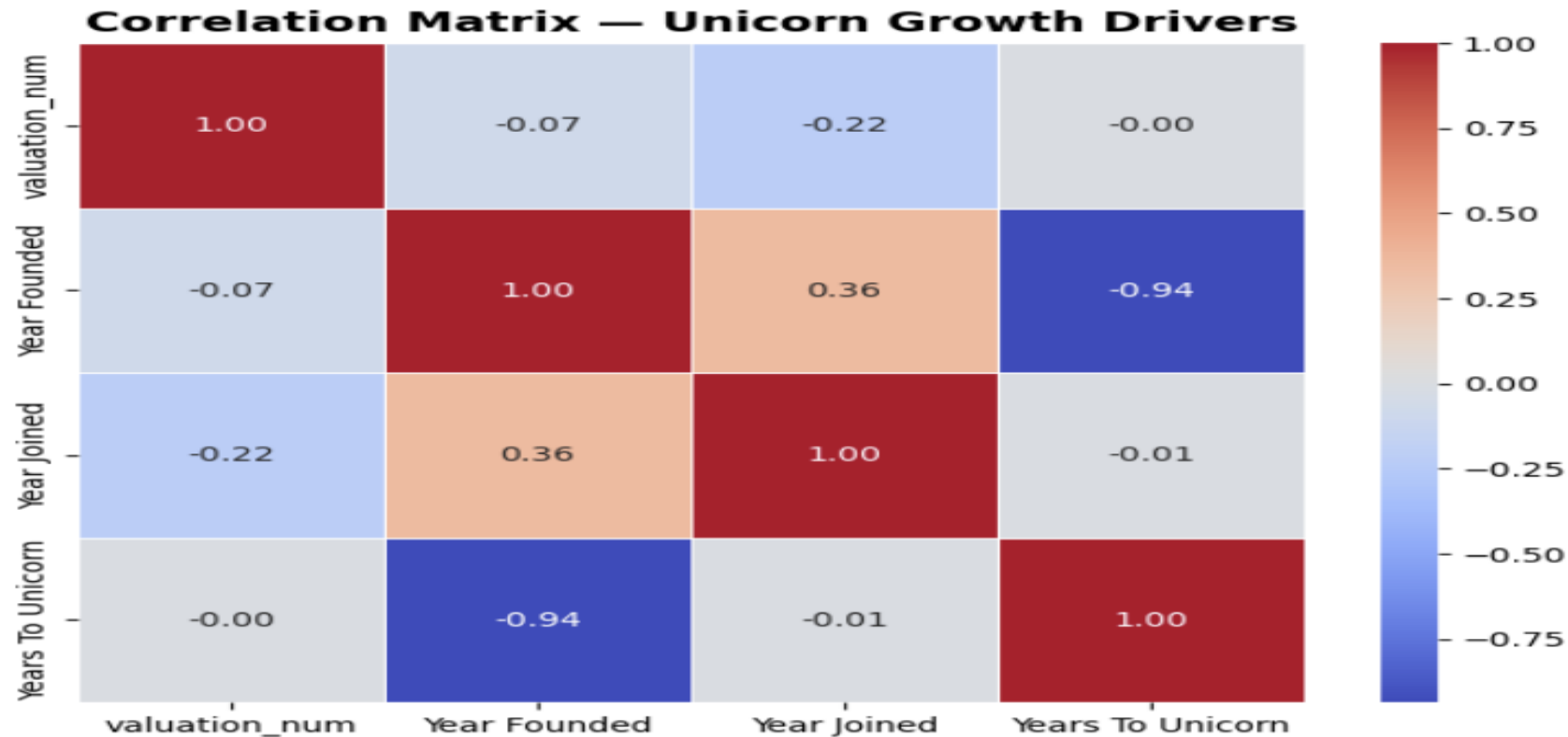Total Unicorn Valuation by Country/Region (Top 15)

- **Fintech and Internet software** are the top industries with the **most unicorns and biggest total valuations**.

- Germany has the **highest total unicorn valuation**, indicating a strong and mature startup ecosystem in Europe.

Industry Positioning: Unicorn Count vs. Average Valuation

Top 20 non-big-4 countries by total company valuation

Multivariate Analysis

Multivariate analysis explores relationships among three or more variables simultaneously.

- There's a **strong negative correlation (-0.94)** between **Year Founded** and **Years to Unicorn**, meaning companies founded more recently tend to achieve unicorn status in fewer years.

- The **valuation_num** shows **weak correlations** with all other variables (ranging from -0.22 to -0.07)

Correlation Matrix — Unicorn Growth Drivers

Geographical Distribution of Unicorn Valuations (Non–Big Four)

- The graph is a **world map visualization** showing **total company valuations by country**, excluding the Big Four (USA, China, India, and the UK).

- Countries like **Germany, Sweden, Australia, and France** have the **highest total unicorn valuations** among non–Big Four Countries & **South Korea, Israel, and Brazil** also show strong unicorn growth in their regions.

- The **global startup ecosystem** is expanding, with emerging countries playing a growing role in unicorn creation.

Total company valuations by country (non-big-four)

Country/Region
- Germany
- Sweden
- Australia
- France
- Canada
- South Korea
- Israel
- Brazil

Insights

➢ Startups founded in recent years are reaching unicorn status much faster, showing accelerated global startup growth.

➢ There is a strong negative correlation between founding year and years to unicorn, highlighting shorter growth cycles for modern startups.

➢ Germany leads among non–Big 4 nations, followed by Sweden and Australia, showcasing Europe's strong innovation base.

➢ Fintech, E-commerce, and Artificial Intelligence dominate the unicorn landscape, driven by digital adoption and investor interest.

➢ Most unicorns were founded after 2010, reflecting the surge in tech-driven and venture-backed startups in the last decade.

➢ Cities like San Francisco, Beijing, and London remain top unicorn hubs, while new cities are emerging as secondary innovation centers.

Conclusion

❖ The global startup ecosystem is evolving rapidly, with newer companies achieving unicorn status faster than ever before

❖ Innovation, digital transformation, and investor confidence are the key forces driving high valuations across industries..

❖ While the Big 4 countries continue to dominate, emerging markets are quickly gaining ground, creating a more balanced global startup landscape.

❖ Fintech, E-commerce, and AI remain the leading sectors, shaping the next wave of billion-dollar companies.

❖ Overall, the data highlights a shift toward faster growth, broader geographic diversity, and technology-driven success in the unicorn economy..

Future Scope:

1. We can develop an interactive **Unicorn Insights Dashboard** to track company valuations, industries, and country-wise growth in real time.

2. We can build a **Unicorn Trend Predictor** to forecast future high-growth sectors and emerging startup hubs using historical data.

3. We can create a **Funding Pattern Analyzer** to identify investor behavior, funding rounds, and their impact on valuation speed..

4. We can design a **Regional Ecosystem Tracker** to highlight underrepresented countries showing potential for unicorn emergence.

5. We can Implement a **Sector Intelligence Model** to analyze which industries are likely to dominate the next wave of unicorns..

6. We can Integrate all analytics into a unified **Global Startup Intelligence Platform** for investors, policymakers, and entrepreneurs

THANK YOU

INNOMATICS
RESEARCH LABS