

Faculty of Information Science and Technology (FIST)

TDA3121 Data Analytics Fundamentals

Trimester 3 2022/2023

Case Study (40%)

Case Study Title: Startup Insights from 2007 to 2023

Due date: **21st September 2023**

No	Student Name	Student ID	Academic Program	Tutorial Session
1	Ahmad Amir Zakwan bin Ahmad Amran	1221303385	BIA	1B
2	Uithiswary a/p Velan	1191202283	BIA	1B
3	Kiang Min Zhe	1191302919	BIA	1B

MARKING RUBRICS (40%)

TASKS		MARKS	
Report (70%)	Introduction		5
	Methodology		20
	Data Set		10
	Data Analysis and Result		20
	Conclusion		5
	Format		5
	References		5
Presentation (30%)	Clarity		10

	Presentation Slides		10
	Teamwork		10
		Total	/100

*****Marks Indication:**

Component	Excellent	Good	Average	Poor	Very Poor
Introduction (5)	5	4	3	2	1-0
	A well-written executive summary of the project.	An adequate introduction to the project report.	Student has performed moderately work resulting in typical results in the specific task.	Missing section or poorly written introduction to the project report.	Student is not able to achieve the specific task.
Methodology (20)	20-17	16-13	12-8	7-4	3-0
	Systematic and clear explanation of the methodology is provided and related to the project.	Methodology of the project provided but not related to the project.	Student has performed moderately work resulting in typical results in the specific task.	Missing section or irrelevant methodology provided.	Student is not able to achieve the specific task.
Data Set (10)	10-9	8-7	6-5	4-3	2-0
	Comprehensive data set from a reliable source with explanation of the attributes and data file.	A data set from a reliable source but lacks depth and explanation of the attributes; and data file provided.	Student has performed moderately work resulting in typical results in the specific task.	Data set from unknown source; no data file or explanation provided.	Student is not able to achieve the specific task.
Data Analysis and Result (20)	20-17	16-13	12-8	7-4	3-0
	Provide analysis and discuss the results using methods introduced in this course.	Provide analysis but lacks discussion on the findings/results.	Student has performed moderately work resulting in typical results in the specific task.	Only descriptive statistics/charts provided with no analysis or discussion.	Student is not able to achieve the specific task.

Conclusion (5)	5	4	3	2	1-0
	Able to conclude the findings with supporting evidence.	Provide conclusions but lacks supporting evidence.	Student has performed moderately work resulting in typical results in the specific task.	Generally conclude the work done.	Student is not able to achieve the specific task.
Format (5)	5	4	3	2	1-0
	Error-free formatting based on MMU guidelines.	Formatting based on MMU guidelines but have minimal errors.	Student has performed moderately work resulting in typical results in the specific task.	Report is not in appropriate format and have many errors.	Student is not able to achieve the specific task.
References (5)	5	4	3	2	1-0
	Appropriate list of references. Error-free reference formatting based on MMU guidelines.	Reference formatting based on MMU guidelines but have minimal errors.	Student has performed moderately work resulting in typical results in the specific task.	Reference is not in appropriate format and have many errors.	Student is not able to achieve the specific task.
Clarity (10)	10-9	8-7	6-5	4-3	2-0
	The presentation is original, creative, clear and support the project.	The presentation is sufficient to present the main idea of the project.	Student has performed moderately work resulting in typical results in the specific task.	The presentation lacks important information on the idea; further interrogations was required to understand the idea.	Student is not able to achieve the specific task.
Presentation Slides (10)	10-9	8-7	6-5	4-3	2-0
	The presentation was interesting and with required information on the work done.	The presentation includes all Important information on the project.	Student has performed moderately work resulting in typical results in the specific task.	The presentation is not interesting and lacks information.	Student is not able to achieve the specific task.
Teamwork (10)	10-9	8-7	6-5	4-3	2-0
	The presentation shows adequate planning and effective communication and teamwork among members.	The presentation was prepared according to plan with adequate communication and teamwork among members.	Student has performed moderately work resulting in typical results in the specific task.	The presentation was dull and shows lack of planning and poor communication and teamwork among members.	Student is not able to achieve the specific task.

(TO BE FILLED BY STUDENTS)

DECLARATION:

We agree that all members deserve equal marks for this project. We confirm that we have contributed equally to produce an original report in our best effort.

Write down your ID (clearly) and sign (by all members):

Student ID: 1221303385 _____ Signature: *Zakwan* _____

Student ID: 1191202283 _____ Signature: *Ushis*

Student ID: 1191302919 _____ Signature: *Minzhe* _____

Declaration by Group Leader

I hereby declare that all group members' names are correctly included in the above section. I hold a copy of this assignment which I can produce if the original is lost or damaged. I certify that not part of this assignment has been copied from any other student's work or from any other source except where due acknowledgement is made in the assignment/project/etc.

Group Leader's Signature: *Zakwan*

Group Leader's Name: Ahmad Amir Zakwan bin Ahmad Amran

Group Leader's ID: 1221303385

Date: 21 Sept 2023

Group Member's Declaration

(Each group member, including the group leader, must individually fill up and submit this form. This form has to be attached together with the assignment/ project submission.)

Group member's name: Ahmad Amir Zakwan bin Ahmad Amran

Student ID: 1221303385

For the purpose of completing this assignment, I have performed the following tasks:

1. I have divided all tasks for my groupmate.
2. I have contributed to all the sections in this report.
3. I have taken part in the presentation.

I hereby declare that I have assessed the final submission and I take full responsibility should there be any inaccuracies, incompleteness, omissions, delays or non-submission.

Group member's signature: Zakwan

Group member's name: Ahmad Amir Zakwan bin Ahmad Amran

Group member's ID: 1221303385

Date: 21.09.23

Group Member's Declaration

(Each group member, including the group leader, must individually fill up and submit this form. This form has to be attached together with the assignment/ project submission.)

Group member's name: Uithiswary a/p Velan _____

Student ID: 1191202283_____

For the purpose of completing this assignment, I have performed the following tasks:

1. I have completed the task assigned for this assignment by the leader. _____
 2. I have contributed to all the sections in this report. _____
 3. I have taken part in the presentation. _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____
- _____

I hereby declare that I have assessed the final submission and I take full responsibility should there be any inaccuracies, incompleteness, omissions, delays or non-submission.

Group member's signature: _____



Group member's name: Uithiswary a/p Velan _____

Group member's ID: 1191202283_____

Date: 21/9/2023_____

Group Member's Declaration

(Each group member, including the group leader, must individually fill up and submit this form. This form has to be attached together with the assignment/ project submission.)

Group member's name: Kiang Min Zhe

Student ID: 1191302919

For the purpose of completing this assignment, I have performed the following tasks:

- [illegible]

I hereby declare that I have assessed the final submission and I take full responsibility should there be any inaccuracies, incompleteness, omissions, delays or non-submission.

Group member's signature: Minzhe

Group member's name: Kiang Min Zhe

Group member's ID: 1191302919

Date: _____

1.0 Introduction

In the ever-evolving landscape of entrepreneurship, startups have consistently been at the forefront of innovation, reshaping industries, and driving economic growth. The period from 2007 to 2023 has been a particularly transformative era for startups, marked by groundbreaking ideas, disruptive technologies, and unprecedented growth. This data analysis project, titled "Startup Insights from 2007 to 2023," delves into a rich dataset containing essential attributes of these pioneering companies, providing a window into the world of startups during this dynamic decade.

A unicorn startup is a privately held company that has achieved a valuation of \$1 billion or more, a remarkable feat that underscores its exceptional growth and potential. These startups, typically in the technology and innovation sectors, often leverage disruptive business models and technological advancements to rapidly expand their market presence. While they remain privately owned, unicorns attract substantial investment from venture capitalists and private equity firms through multiple funding rounds. This investment is based on their growth potential, revenue, and market share, and it fuels their continued expansion and innovation. Examples of well-known unicorn startups include Uber, Airbnb, and SpaceX. However, it's important to note that achieving unicorn status doesn't guarantee long-term success, and these companies still face challenges and uncertainties on their path to profitability and sustainability.

This dataset, which includes critical variables such as company names, valuations (measured in billions of dollars), the date of entry into the startup ecosystem, country and city of operation, industry or sector of operation, and a list of investors, is a valuable repository of information that offers a comprehensive view of the startup landscape during this period. Through rigorous data analysis and exploration, we aim to uncover patterns, trends, and insights that can illuminate the factors contributing to startup

success, the dynamics of startup ecosystems in different regions, and the impact of investor support.

Our primary objective is to answer critical questions that have defined the trajectory of unicorn startups from 2007 to 2023: How did the valuation of startups evolve during this period, and what were the key drivers behind their growth or decline? Did the geographical location of startups influence their success, and were there discernible trends in terms of industry concentration?

As we embark on this data analysis project, we not only seek to uncover the past but also to provide actionable insights for the future. By delving into the nuances of startup data, we aim to offer valuable guidance for entrepreneurs, investors, policymakers, and researchers interested in understanding the dynamics of startup ecosystems.

2.0 Methodology

2.1 Data Collection

The foundation of this case study lies in a carefully curated dataset encompassing startup information from 2007 to 2023. Data for this research was obtained through openly licensed public datasets from the website: [The Complete List Of Unicorn Companies \(cbinsights.com\)](https://www.cbinsights.com/research-unicorn-companies). The dataset acquired from this site consisted of 1235 instances of companies all over the world. Each instance had 7 attributes and the data was collected from 2007 to 2023. The attributes that are used for analysis are described in Table 1.¹

2.2 Data Dictionary

Variables	Description
Company	The name of the startup company
Valuation (\$B)	The valuation of the company is billions of dollars.
Data Joined	The date when the company joined the startup ecosystem.
Country	The country where the company is based.
City	The city where the company is located.
Industry	The industry or sector the company operates in.
Select Investors	The list of select investors who have supported the company.

2.3 Data Analysis Method

2.3.1 Data Cleaning and Preprocessing

Before conducting this analysis, the dataset undergoes thorough cleaning and preprocessing. This involves handling missing data, correcting data types, and ensuring data consistency. Additionally, outliers and anomalies are identified and addressed

¹ <https://www.cbinsights.com/research-unicorn-companies>

appropriately to maintain the integrity of the analysis.

2.3.2 Statistical and Analytical Techniques

We used both statistical and analytical techniques to analyze the data we collected in this case study. To gain initial insights from the dataset, we applied the descriptive analysis by finding key statistics such as mean, median and standard deviation. Then we also applied linear regression analysis to evaluate the relationship between variables. With this, we identify the impact of each independent variable on the dependent variable.

2.3.3 Software and library

We have applied some software tools and its libraries to effectively manage and analyze our dataset. We used Microsoft Excel, a general purpose spreadsheet for our initial data cleaning and preprocessing. Excel is an user-friendly interface which allows us to inspect, clean and format our data. For example, in our original dataset there was some duplicate data with punctuation marks and the format of the join date variable also did not follow the standard.

We have used Google Colaboratory which is a cloud-based Jupyter Notebook environment. It served as our collaboration coding environment and provided seamless integration with Python libraries. It is an online platform which allows our team members to work collaboratively in real time, share code and visualizations. Google Colaboratory integrates with Python to run data analysis, create data visualizations and document our methods.

Other than that, we applied the Python programming language and its data science libraries like pandas and NumPy to perform more advanced data manipulation and explanatory analysis. Panda and NumPy libraries were used to do data management tasks such as merging data, filtering observations and aggregating data.

Data visualization is created using Plotly Express, Seaborn and Matplotlib.pyplot which allows us to create informative plots, charts, and graphs that effectively represent key information from our dataset. For predictive modeling, we used scikit-learn (sklearn) as the main machine learning library and sklearn.Route_model module to implement different regression models. We apply this to understand the relationship between variables and make data-driven decisions and predictions.

2.4 Exploratory Data Analysis (EDA):

Exploratory Data Analysis (EDA) is a crucial step in gaining a preliminary understanding of the dataset. We have used EDA to visualize trends in startup valuations over time to identify periods of rapid growth or decline. Other than that, we used it to analyze the geographical distribution of startups to pinpoint regions with a high concentration of startups and explore how valuation varies by location. Lastly, we applied the EDA method to examine the distribution of startups across different industries to determine which sectors have experienced significant valuation growth and which ones have remained relatively stagnant.

2.5 Stakeholders

We would like to present our case study to some stakeholders like startup founders and entrepreneurs, investors and venture capital firms, startup incubators and accelerators, government and economic development agencies and many more. With this, every stakeholder is able to gain the insights of the startup and make reliable decisions.

1. **Startup Founders and Entrepreneurs:** Founders and entrepreneurs in the startup ecosystem are likely to be interested in understanding how the ecosystem thrives. They may use this information to make strategic decisions.
2. **Investors and Venture Capital Firms:** Investors, including venture capitalists and

angel investors, are keen to understand how the timing of their investments correlates with the overall valuation of startups. This analysis can provide insights into investment timing.

3. **Startup Incubators and Accelerators:** Organizations that run startup incubator or accelerator programs may want to assess the impact of their programs over the years. They can use this analysis to track the growth of startups that have participated in their programs.
4. **Government and Economic Development Agencies:** Government agencies and organizations responsible for fostering entrepreneurship and economic development in a region or country may be interested in these results to evaluate the effectiveness of policies and initiatives.
5. **Market Researchers and Analysts:** Professionals in market research and analysis may use this information to identify trends in the startup ecosystem and to provide insights to their clients or organizations.
6. **Business Consultants and Advisors:** Consultants and advisors working with startups may leverage this analysis to provide strategic guidance to their clients regarding the optimal time to enter the market or seek investments.
7. **Academic Researchers and Educators:** Researchers in the fields of entrepreneurship, economics, and business may use the results for academic studies. Educators can incorporate these findings into their courses to illustrate real-world examples.
8. **Media and Journalists:** Media outlets, including tech and business journalists, may find these results valuable for reporting on trends and developments in the startup world.

9. General Public: Individuals with an interest in startups and entrepreneurship, as well as those considering starting their own companies, may find these insights informative and inspirational.

3.0 Data Set

	Company	Valuation (\$B)	Date Joined	Country	City	Industry	Select Investors
0	ByteDance	\$225.00	7/4/2017	China	Beijing	Media & Entertainment	Sequoia Capital China, SIG Asia Investments, S...
1	SpaceX	\$137.00	1/12/2012	United States	Hawthorne	Industrials	Founders Fund, Draper Fisher Jurvetson, Rothen...
2	SHEIN	\$66.00	3/7/2018	Singapore	NaN	Consumer & Retail	Tiger Global Management, Sequoia Capital China...
3	Stripe	\$50.00	23/1/2014	United	San	Financial	Khosla Ventures, LowercaseCapital

Figure 3.1

The dataset contains information about various startup companies and is structured with the following columns according to Figure 3.1:

Company: This column represents the name of each startup company. It serves as a unique identifier for each entity in the dataset.

Valuation (\$B): This column represents the valuation of each startup company in billions of dollars. Valuation is an important financial metric that assesses the estimated worth or market value of a company. It indicates the perceived value of the company's assets, intellectual property, market potential, and other factors.

Date Joined: This column specifies the date when each startup company joined a certain program, network, or platform. This date can be important for analyzing the company's

growth and development over time. It is represented in a date format.

Country: This column indicates the country where each startup company is based or headquartered. It provides geographical information about the company's location.

City: This column specifies the city where each startup company is located within its respective country. It offers more specific location details within a country.

Industry: This column categorizes each startup company into an industry or sector based on its primary area of operation. Industries can vary widely and may include technology, healthcare, finance, e-commerce, and more. This information is valuable for industry-specific analyses.

Select Investors: This column lists the names of select investors or backers who have invested in each startup company. Investors can include individuals, venture capital firms, angel investors, and other entities that provide financial support and resources to help the startup grow.

Overall, this dataset provides comprehensive information about startup companies, including their financial valuation, geographical location, industry focus, key investors, and the date of their entry into becoming a unicorn startup.

4.0 Result

The main aim of this analysis is to understand the ecosystem of a Unicorn Startup that explains what factors make an impact to the success of a startup. As to measure the success of a startup, we use valuation as a dependent variable to track what causes certain startups to bloom well compared to others. Growth rate percentage also can be used as well because a high growth rate may be one of the indicators that the startup is on the right track. We also use descriptive analysis to serve as the foundation for understanding the dataset, revealing trends, relationships, and patterns. It helped generate hypotheses, guided further analyses, and provided valuable insights into the factors influencing unicorn startup success. These insights laid the groundwork for more in-depth investigations and modeling techniques used in the analysis.

4.1 Trend Analysis

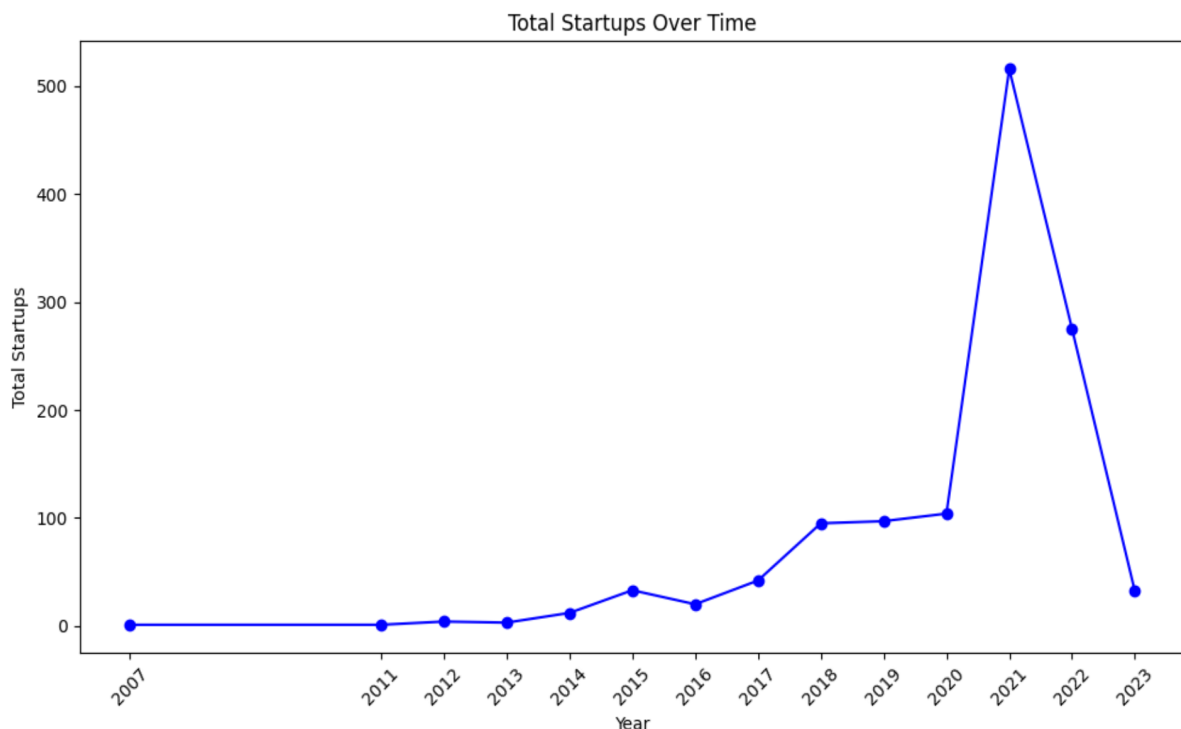


Figure 4.1.1 shows the total startups over time

Figure above shows the line chart graph of how startup growth has changed over the

past decade and in the current year. The total startup in 2007 started with just 1 startup business. Next few years, the startups started to increase slowly and steadily. During the year 2014, there was a notable increase in the startup businesses. As the world advances, the startup business starts to arise. Even Though, during 2016 startups dropped slightly, it started to rise back during 2017. It started to peak from 2020 till 2021 rapidly where the highest startups was 516 in the year 2021. The reason for the increase of startups during that period is covid pandemic. However, it's worth noting that the data also shows a slight decline in the number of startups in 2022 and 2023. This is due to the potential changes in the market.²

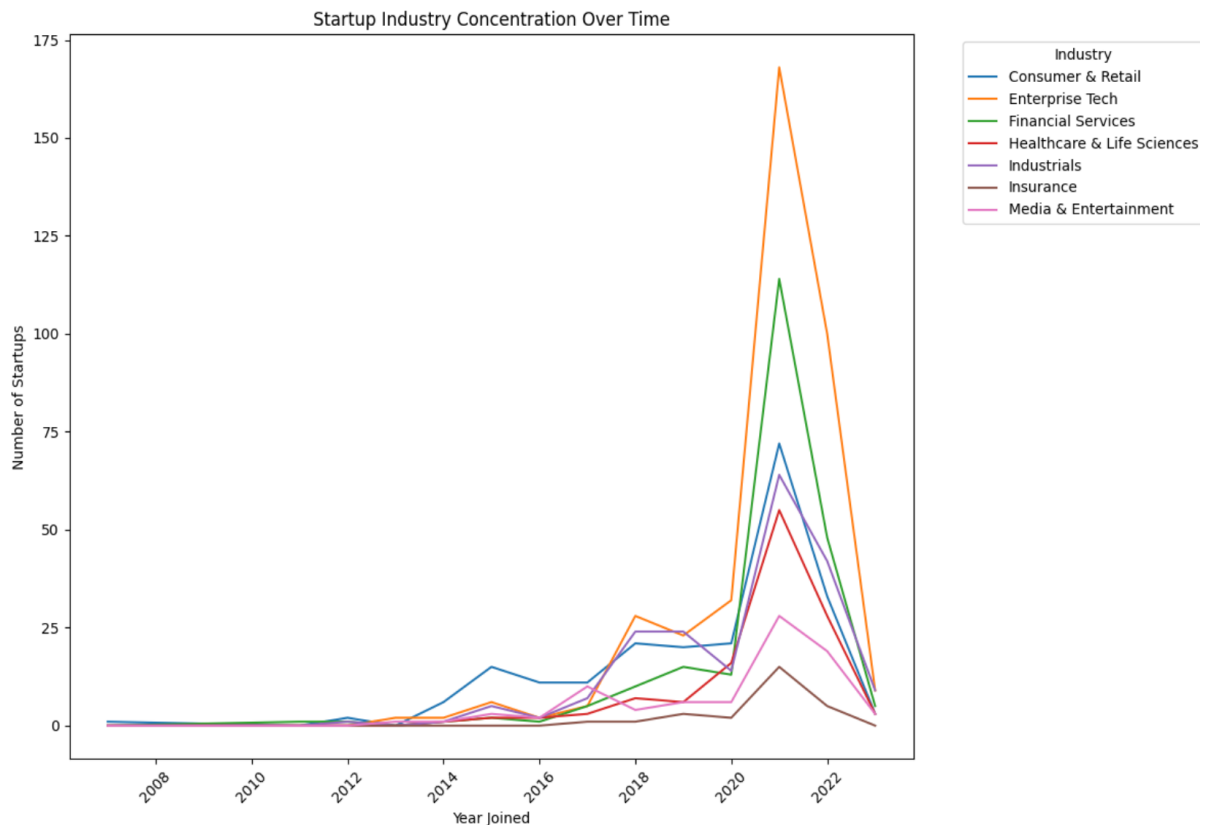


Figure 4.1.2 shows line chart for startup industry concentration over time

From the figure 4.1.2, we can observe the exponential growth of startups across various

² <https://www.cbinsights.com/research/unicorn-company-slowdown/>

industries. Based on the Line chart, the most substantial surge is the “Enterprise Tech” industry where it rose to dominate the startup businesses in 2021 and 2022. This remarkable growth is an indication of the increasing importance of technology in today’s business world. There are few enterprise tech startups that have grown famous like OpenAI and Canva. “Consumer & Retail” startups have steadily grown over the years. It maintained modest growth in the startup industry in the early years and started to have significant growth around 2021. This shows the shifts of consumer behavior and preferences to adapt based on the current trend. For example, there are some startups in the Consumer & Retail industry such as SHEIN and Fanatics. Though many industries peaked rapidly during the year 2020, the Insurance industry saw a slight uptick in 2021.

4.2 Growth Rate against Industry

The analysis of growth rates against different industries provides a valuable perspective on the performance and dynamics of various sectors within the economy. By examining how industries have grown over a specific period, we can gain insights into which sectors are experiencing rapid expansion and which ones are evolving at a more moderate pace.

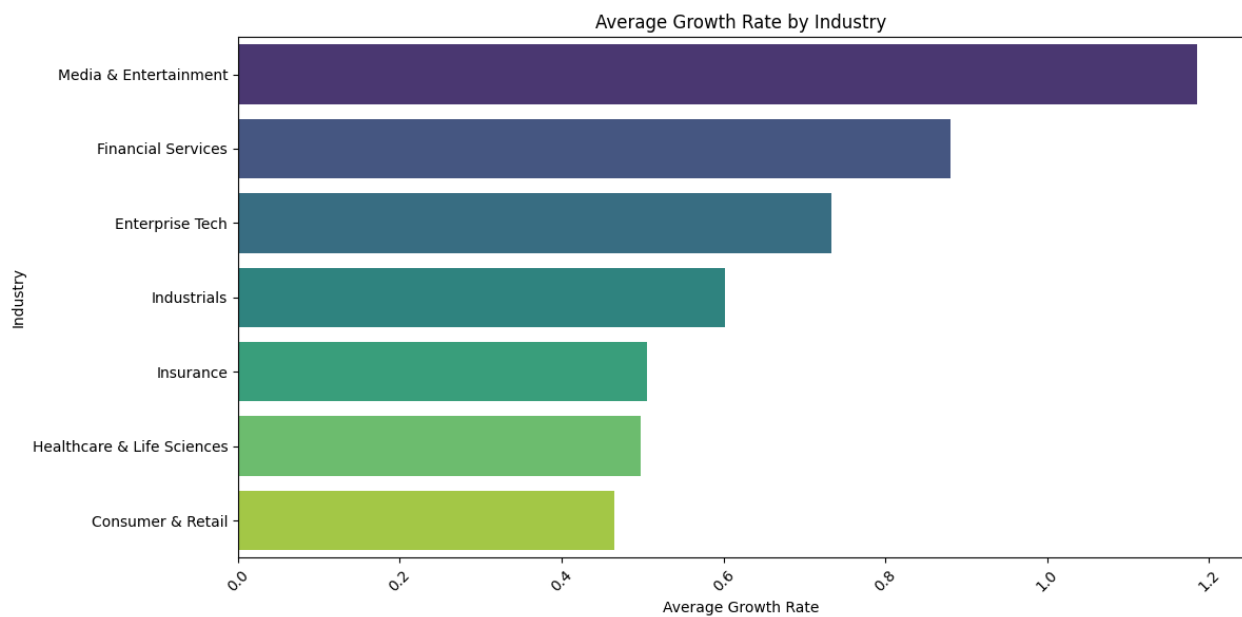


Figure 4.2.1 shows histogram for Average Growth Rate by Industry (per year)

Based on Figure 4.2.1, The Media & Entertainment industry has the highest average growth rate among all the industries. This industry likely experienced significant growth, possibly driven by the adoption of new technologies, increased demand for digital content, or other factors contributing to its expansion. The Financial Services sector has the second-highest average growth rate. The growth may be due to innovations in fintech, expanding customer bases, or strategic investments in technology and services. The Consumer & Retail industry has the lowest average growth rate among the listed industries. While growth in this sector may be relatively slower, it could be influenced by changing consumer preferences, competition, or economic conditions affecting retail businesses. These observations provide insights into which industries have experienced varying degrees of growth, which can be valuable for investors, policymakers, and businesses looking to make informed decisions regarding investments, partnerships, or market entry strategies.

4.3 Date Join against Valuation

Exploring the relationship between a startup's date of joining and its valuation through a linear regression model offers valuable insights into the dynamics of business growth. This analysis aims to quantify the impact of a company's age, measured from its inception, on its financial success. By applying linear regression techniques, we can discern whether there is a statistically significant correlation between the two variables. The resulting regression equation and associated coefficients provide a quantitative understanding of how the timing of a startup's entry into the market influences its valuation.

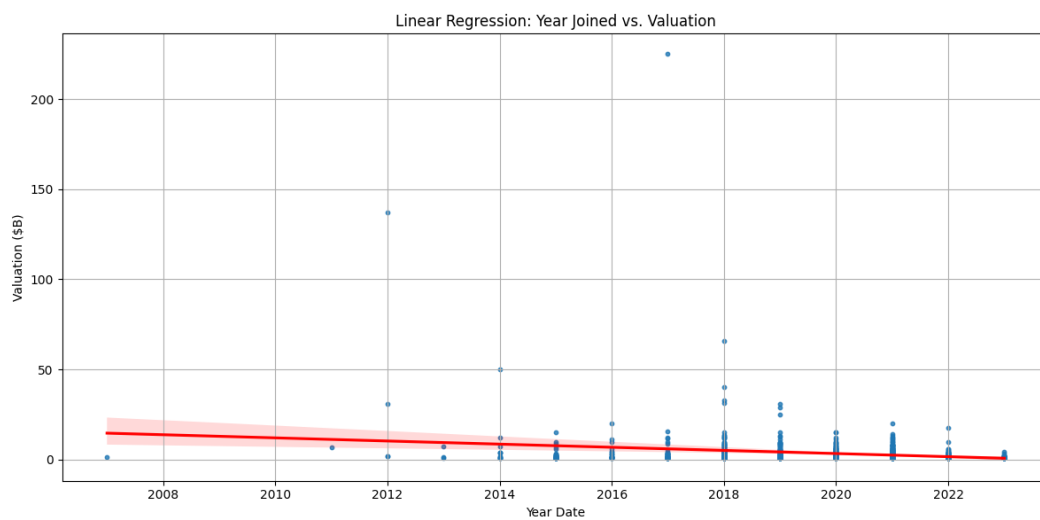


Figure 4.3.1 Linear Regression : Year Joined against Valuation (\$B)

OLS Regression Results						
Dep. Variable:	Valuation	R-squared:	0.042			
Model:	OLS	Adj. R-squared:	0.042			
Method:	Least Squares	F-statistic:	54.49			
Date:	Mon, 18 Sep 2023	Prob (F-statistic):	2.86e-13			
Time:	20:41:39	Log-Likelihood:	-4355.5			
No. Observations:	1235	AIC:	8715.			
Df Residuals:	1233	BIC:	8725.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1760.3879	238.058	7.395	0.000	1293.345	2227.431
JoinYear	-0.8698	0.118	-7.382	0.000	-1.101	-0.639
Omnibus:	2711.415	Durbin-Watson:	0.213			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10606910.597			
Skew:	18.866	Prob(JB):	0.00			
Kurtosis:	455.441	Cond. No.	2.05e+06			

Figure 4.3.2 OLS Regression Result

Based on Figure 4.3.2, R-squared measures the proportion of the variance in the dependent variable (Valuation) that can be explained by the independent variable (JoinYear). In this case, the model explains only 4.2% of the variance, suggesting a weak relationship.

The JoinYear coefficient which is -0.8698 represents the estimated change in Valuation for a one-unit increase in JoinYear. In this case, it suggests that, on average, as the JoinYear increases by one year, Valuation decreases by approximately 0.8698 units.³

So as a conclusion, this linear regression model suggests that there is a statistically significant, albeit weak, relationship between a startup's JoinYear and its Valuation. However, the low R-squared value indicates that JoinYear alone does not explain much of the variation in Valuation, and other factors may play a more substantial role.

³<https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/>

4.4 Geographical factor

“Does Geographical presence have impacts on startups?”

Based on this big question, we will be answering if geographical presence has impacts or not on startups. In this part, we will analyze the dataset, give reasons and examples on geographical factors.

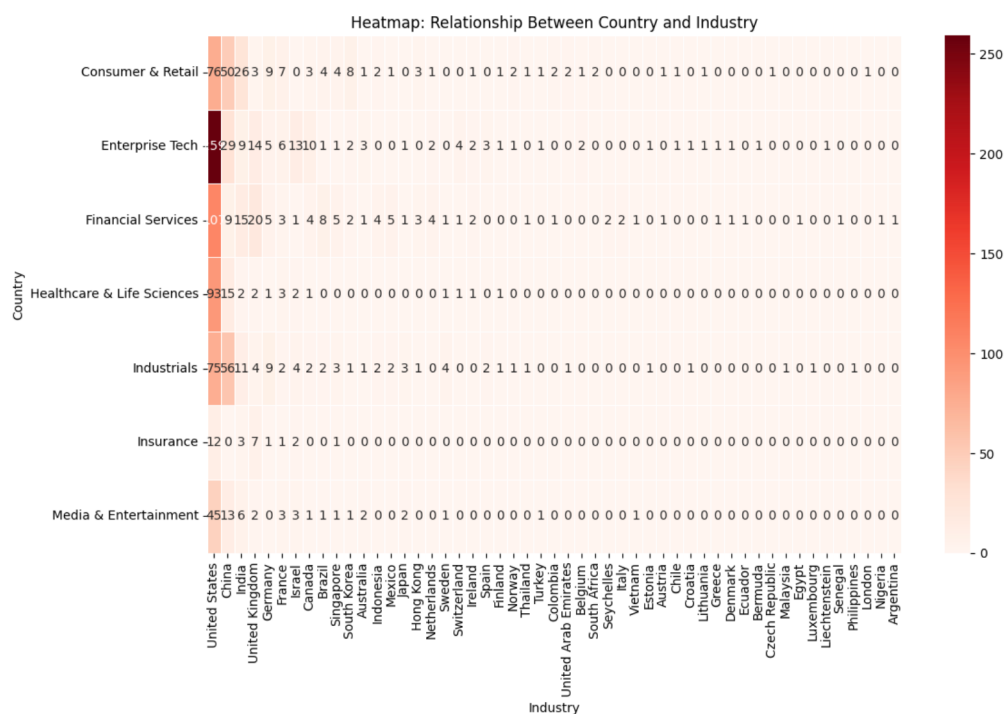


Figure 4.4.1 shows the heatmap of the relationship between country and industry.

This figure shows the relationship between country and industry. From this heatmap, we know that the United States has the most startups in the enterprise tech industry. The US has a significant number of startups in all other industries. Other than that, China is the second country which has the number of startups in industries. From this heatmap, we know that small countries have less startups than big countries like the US, China and India.

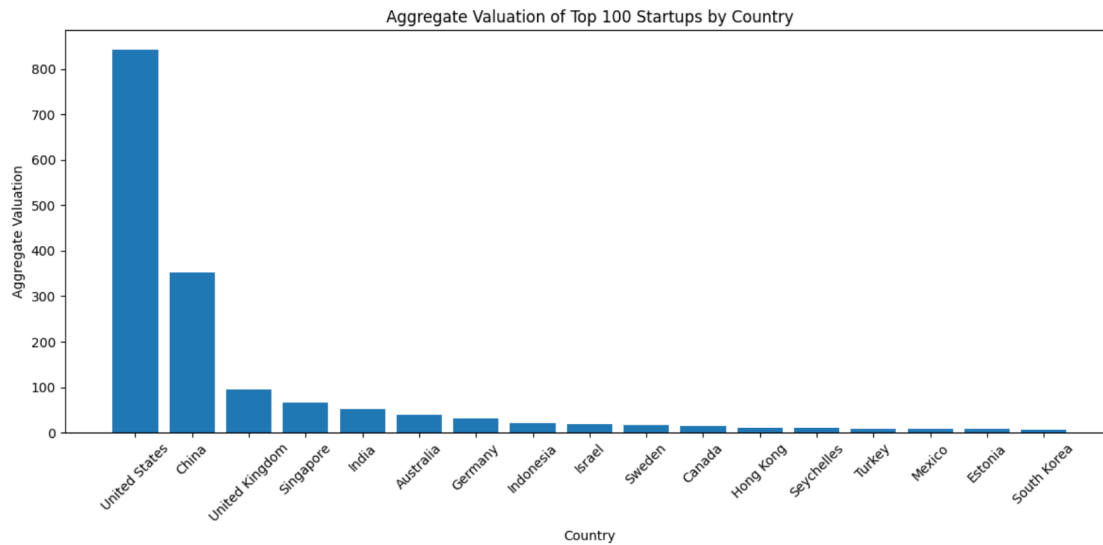


Figure 4.3.3 displays the aggregate valuation of top 100 startups by country

The bar chart above presents a clear visualization of top 100 startups by country. The United States stands out with the highest aggregate valuation, \$843.4 billion. This shows the position of the US as a global platform for startups. Secondly, China is the country that follows after the US in the top 100 startups with a \$352.18 billion valuation. It reflects that China is having rapid growth and expanding in the startup ecosystem. Other than these 2 countries, other countries relatively have lower valuation because of the limited market size, investment climate, access to resources, and regulatory environment.

Based on the analysis, we know geographical presences have impacts on startups. One of the reasons is Gross Domestic Product (GDP) which plays an important role in startup businesses. This is because countries with high GDPs often have larger consumer markets and higher disposable income which will increase sales and its growth. For example, the United States has a gross domestic product (GDP) of \$20.89 trillion and

14.72 trillion for China.⁴ It also shows that the United States and China are the most contributing countries on the top 100 startups bar chart. With this, we can conclude that Gdp is one of the geographical factors that impacts startups.

Other than that, the other reason geographical presence impacts on startups is government policies, regulations and incentives. What are government policies, regulations and incentives? Government policies are guidelines implemented by the government to address specific issues, regulations are specific rules and laws established by a government to govern the behavior of businesses and industries, and incentives are policies conducted by the government to encourage businesses. There are few examples of it which are Intellectual Property Protection for the United States and “Made in China 2025” plans for China. The US has strong intellectual property protection laws which enable the startups in Silicon Valley to protect their innovative products and ideas. Thus, we can wrap that government policies, regulations and incentives are the geographical factor that impacts startup business.

The last reason we say geographical presences have impacts on startups are Economic Ecosystem and Populations. The definition of Economic Ecosystem is a dynamic and stable business location. Silicon Valley is an example of an economic ecosystem which is a successful startup ecosystem. It is a place for the concentration of technology companies and venture capital. Population is also the reason that contributes to startups. China has a massive population of 1.4118 billion people who directly or indirectly impact the startup businesses. With over a billion potential investors and consumers, China’s startups have a way to a huge domestic market. For example, companies like Alibaba and Tencent used this big market to become global tech giants.

⁴ <https://www.investopedia.com/insights/worlds-top-economies/>

With this, we can assume that Economic ecosystem and population have impacts on startups.

Thus, we answer the big question that geographical presence does have impacts on startups.

5.0 Conclusion

Our analysis has provided valuable insights into the factors that impact unicorn startup success. We have explored various aspects of the dataset, including the relationship between factors like the date of joining, industry, geography, and valuation. While our analysis has shed light on some important trends and correlations, it's essential to acknowledge that there are numerous other factors influencing a startup's success that are not present in our dataset.

Our findings indicate that factors such as the industry in which a startup operates and the geography can play a role in determining a company's valuation and growth rate. However, the complex and multifaceted nature of startup success means that our analysis only scratches the surface.

Many critical factors, such as the startup's leadership team, product innovation, market dynamics, competition, and economic conditions, are not explicitly captured in our dataset. These unaccounted-for variables can significantly impact a startup's trajectory and ultimate success.

Therefore, while our analysis provides valuable insights and highlights certain influential factors, it is crucial to approach startup analysis with a holistic perspective. Startups operate in dynamic and ever-changing environments, and their success depends on a combination of both observable and latent factors.

In future analyses, it would be beneficial to incorporate a more comprehensive dataset that includes a broader range of variables to gain a deeper understanding of the intricate factors that contribute to startup success. Such a holistic approach will provide more accurate and actionable insights for entrepreneurs, investors, and stakeholders in the

startup ecosystem.

In conclusion by doing this project, the utilization of Python as a tool for data analysis emerges as a remarkably potent and precise approach. Before embarking on the journey of data analysis, it is imperative to define our goals and priorities meticulously to ensure that we are not wasting our efforts on irrelevant or inconsequential data points. This project has illuminated a crucial insight which is that the creation of tables and graphs is no longer the exclusive domain of conventional software like Microsoft Excel or other dedicated business analysis tools; coding offers us a powerful alternative. Furthermore, as we traversed this project's terrain, we found ourselves delving into a diverse array of data analysis methods, expanding our knowledge and capabilities in this intricate field. This journey not only reinforced the efficacy of Python but also broadened our analytical horizons, equipping us with a versatile skill set for future endeavors.

6.0 References

1. <https://www.cbinsights.com/research-unicorn-companies>
2. <https://www.cbinsights.com/research/unicorn-company-slowdown/>
3. [What is Linear Regression? - Statistics Solutions](#)
4. <https://www.investopedia.com/insights/worlds-top-economies/>