1. **Introduction/Objective**

Funding plays a pivotal role in providing the necessary resources and support for startups to thrive. In this article, we will explore funding received by various startups in India in the period 2018 –2021 through data analysis. The analysis aims to provide insights and recommendations to potential investors that wants to venture into the Indian Start-up Ecosystem and potential insights to secure funding.

1. **Project Structure**

The Cross Industry Standard Process for Data Mining (CRISP-DM) framework was fully adopted in conducting the data analysis process on this project. The sequential steps followed in this were:

1. Business understanding. This stage involved setting the project objectives.
2. Data understanding. What data do we have / need? Is it clean?
3. Data preparation. This entailed performing data cleaning and EDA for use later.
4. Data Visualization and storytelling.
5. Deployment.

**Business Understanding**

**Project overview**

This data analysis project focuses on the funding received by start-ups in India from 2018 to 2021. The objective is to gain insights into the Indian start-up ecosystem and propose the best course of action for our team's venture. By analyzing the data on funding amounts, start-up details, and investor information, we aim to unearth prevailing patterns and gain insights about the opportunities in India's start-up ecosystem to inform decision-making.

**2. Data understanding**

The data used in this project was sampled from different start-up companies in India. It contains funding history for the period 2018 - 2021. The data was obtained from 4 datasets.

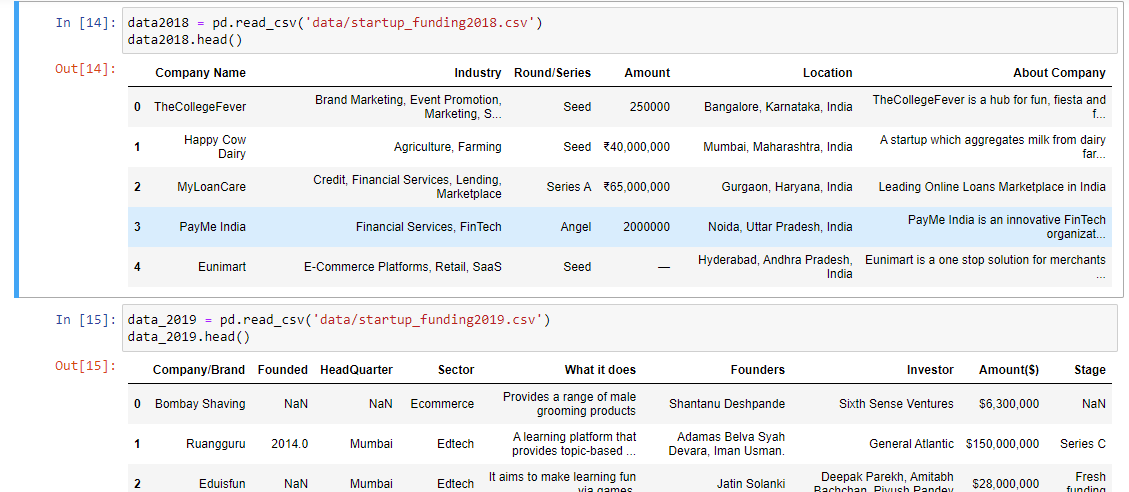
**2.1 Loading packages**

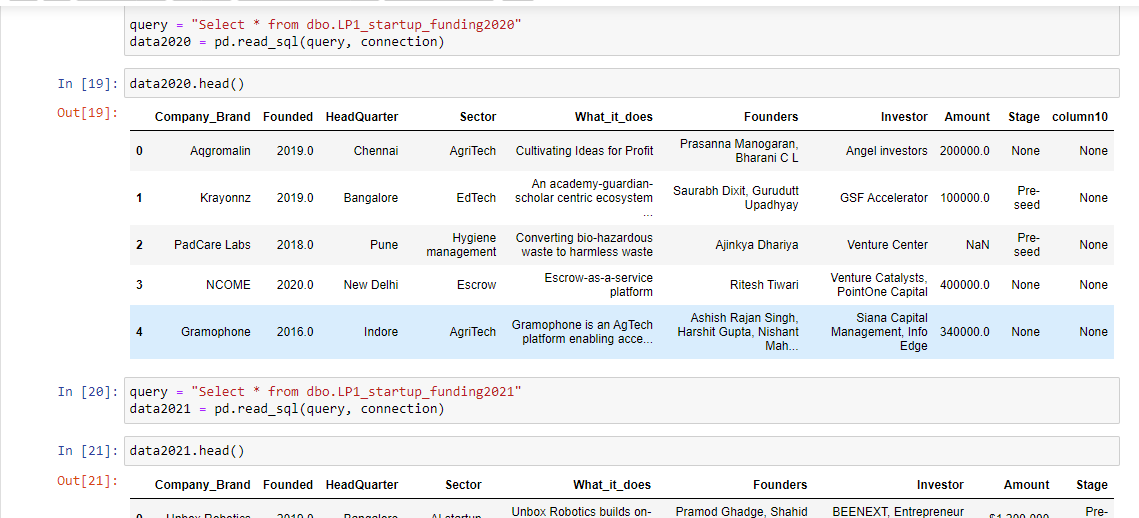
For the data analysis, some Python libraries for data manipulation, data visualization will be loaded as shown below:



**2.2 Loading the data**

To start off we read our data and had a simple preview of each dataset.





**3. Data preparation**

Once we had a strong understanding of our data, we moved onto preparing the data for the analysis. During this stage, the data was cleaned separately for each dataset starting from 2018 up to the 2021 dataset and merged thereafter.

The following issues were handled:

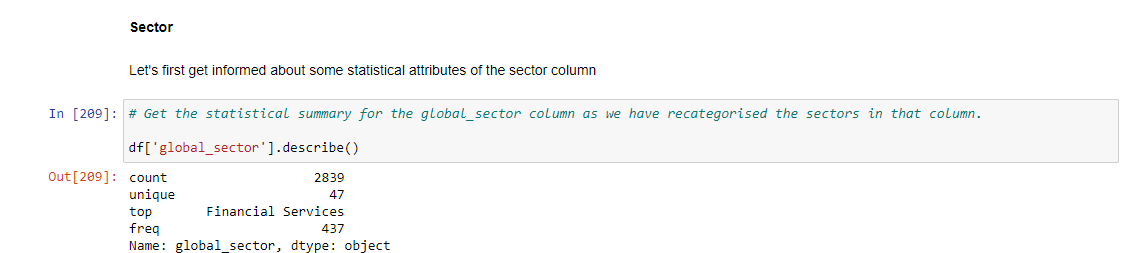
* Detecting and dealing with missing values and duplicates.
* Splitting Location and Industry Columns: The location and industry columns contained multiple values separated by commas. Only the first value was selected as the primary sector.
* Currency Conversion for 2018 Amounts: In the 2018 dataset, the amounts column contained a mix of Indian Rupees (INR) and US Dollars (USD). To standardize the amounts, the Indian Rupee was converted to US Dollars (USD).
* Removing commas and currency signs from the Amount column in all datasets. This allowed the amounts to be properly recognized as numeric values.
* Correcting misplaced/erroneous values especially in the 2021 dataset.
* Dropping an extra column named “*column10*” in the 2020 dataset: This was to align the columns as this column was not present in other datasets.
* Adding “*Funding Year*” column to every dataset which would enable analyzing the funding trends over time.
* Renaming Columns and Concatenation.
* Data type conversions (e.g., numeric data mistakenly encoded as strings)

**4. Exploratory data analysis**

This stage involved conducting an initial exploration of the dataset to gain insights and identify patterns or trends, generating summary statistics, visualizations, and descriptive analysis to understand the distribution, relationships, and characteristics of the data.

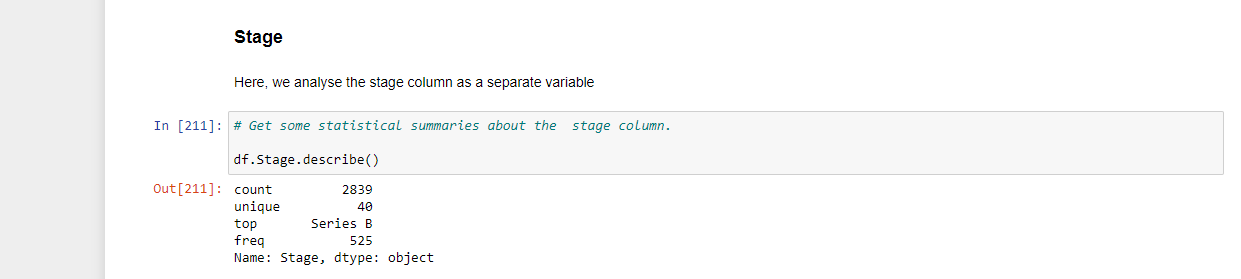
**4.1 Univariate Analysis**

* Univariate analysis is a component of exploratory data analysis (EDA) that focuses on examining and interpreting individual variables in isolation.
* By scrutinizing a variable independently, researchers can uncover valuable information about its properties without considering the influence of other variables.



For the *Sector* column:

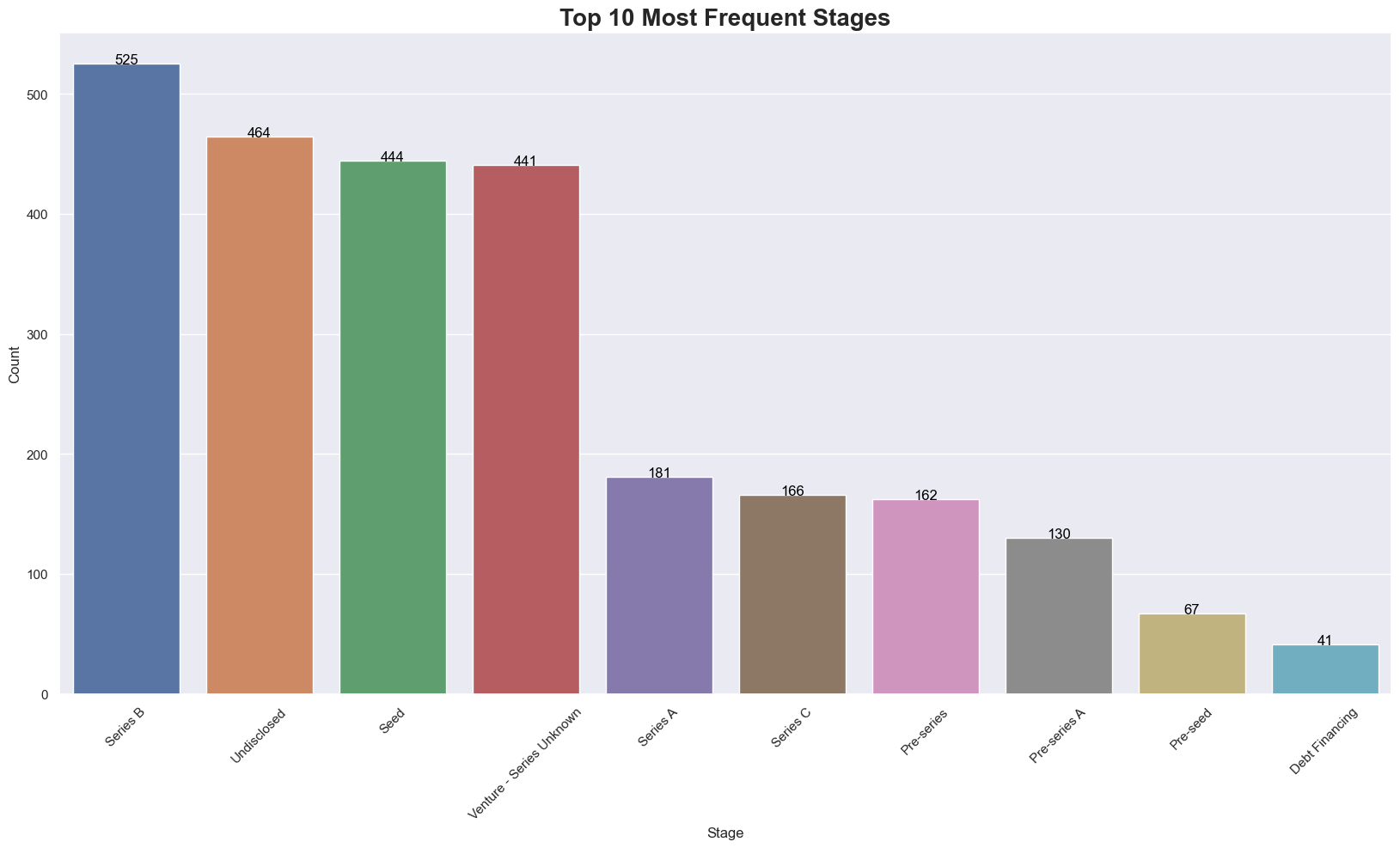
* The recategorized sector column has 2839 rows with 48 unique sectors. '**Financial Services**' occurred most at 361 times.
* This indicates that, the financial services sector is the most dominant in the Indian start-up ecosystem for the period under review.



For the *Stage* column:

* The Series B funding stage was the most common funding stage at which Indian start-ups obtained funding. It dominated with 525 funding deals from 2018 to 2021.

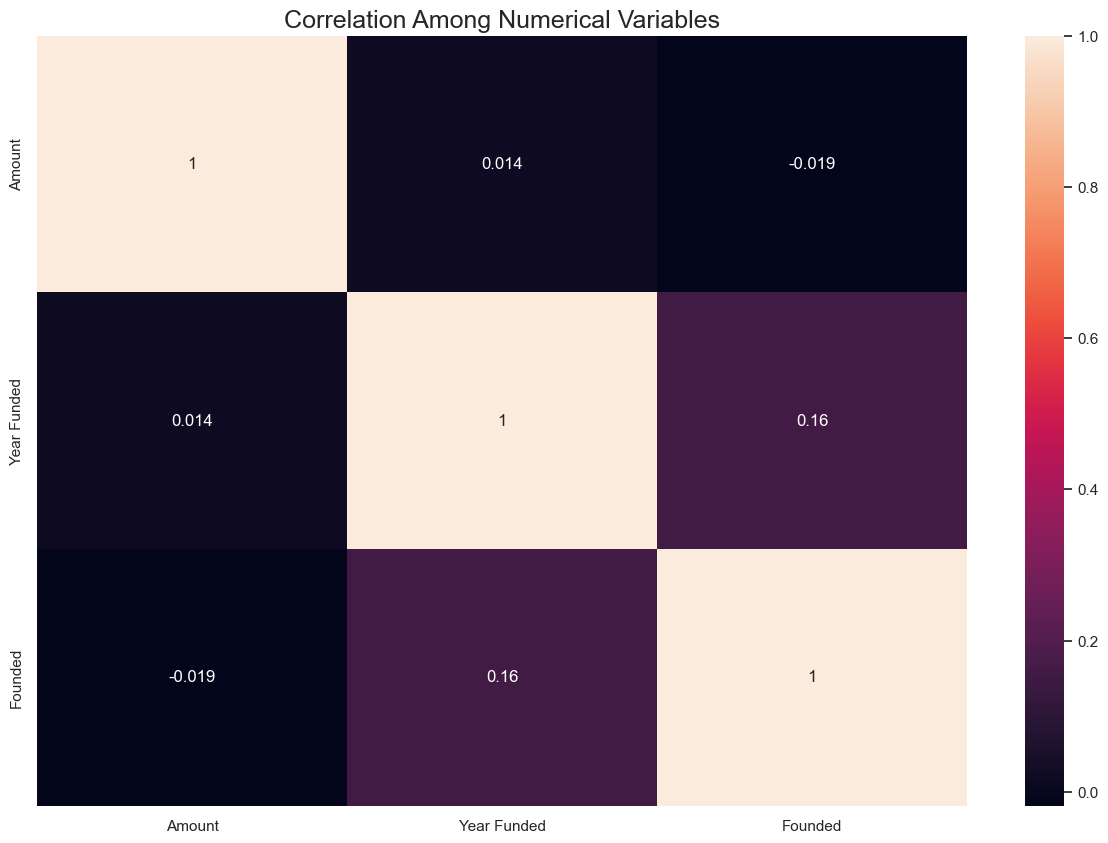
Let us get the visual impressions about these stages and their distribution per counts.



* The diagram above depicts the fact that most funded stage in the Indian start-up funding ecosystem is the **Series B stage**. 361 successful funding was obtained at this stage throughout the period.
* However, the number of start-ups in the said ecosystem whose funding stages were not disclosed are just 61 short of that of the top funded stage.
* Also, the third most stage at which funding was obtained among these said start-ups is the **Seed** stage. Funding was obtained 444 times at this stage from 2018 through to 2021.
* Meanwhile at the venture equity stage whose series are unknown, 441 funding deals were obtained then a drastic drop in number of deals occurred with Series A, Series C, Pre-series, Pre-seed and the last for the top ten have been at the debt financing stage.

**4.2 Multivariate analysis**

This form of exploratory data analysis seeks to find relationships among numeric variables. A correlation heatmap is used in this regard.

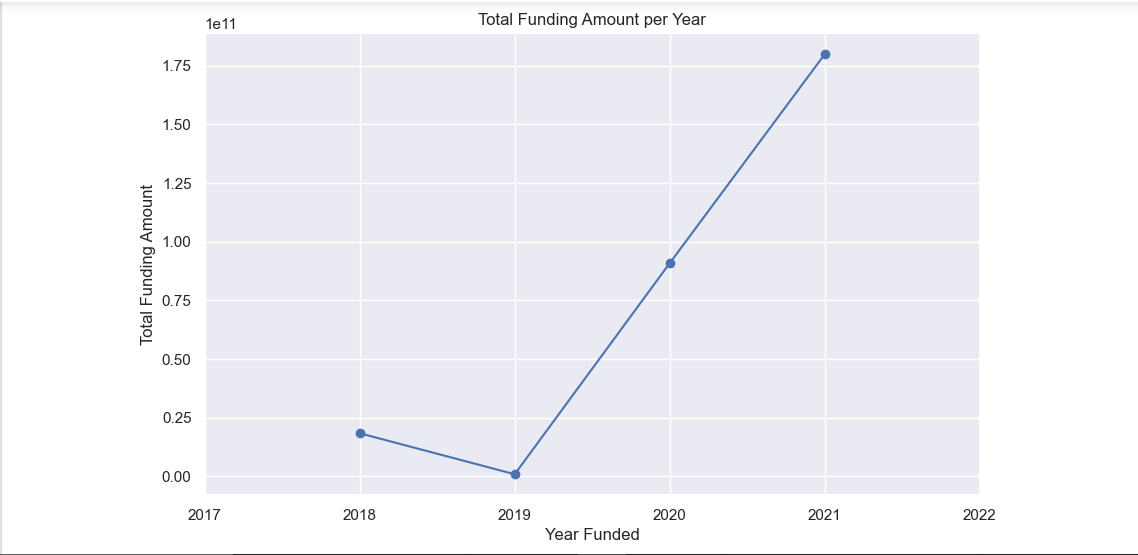


* From our correlation heatmap, there seems to be a very weak positive relationship between the year funded and funding amount received.
* For founding years of companies and funding obtained, there exists a weak negative relationship.
* This implies that the amount funded does not depend on the year in which it was received. Also, it does not rely on the age of such companies.

### **To guide our analysis, we formulated 5 SMART questions:**

1. **What is the overall trend in funding received by startups in India from 2018-**2021?

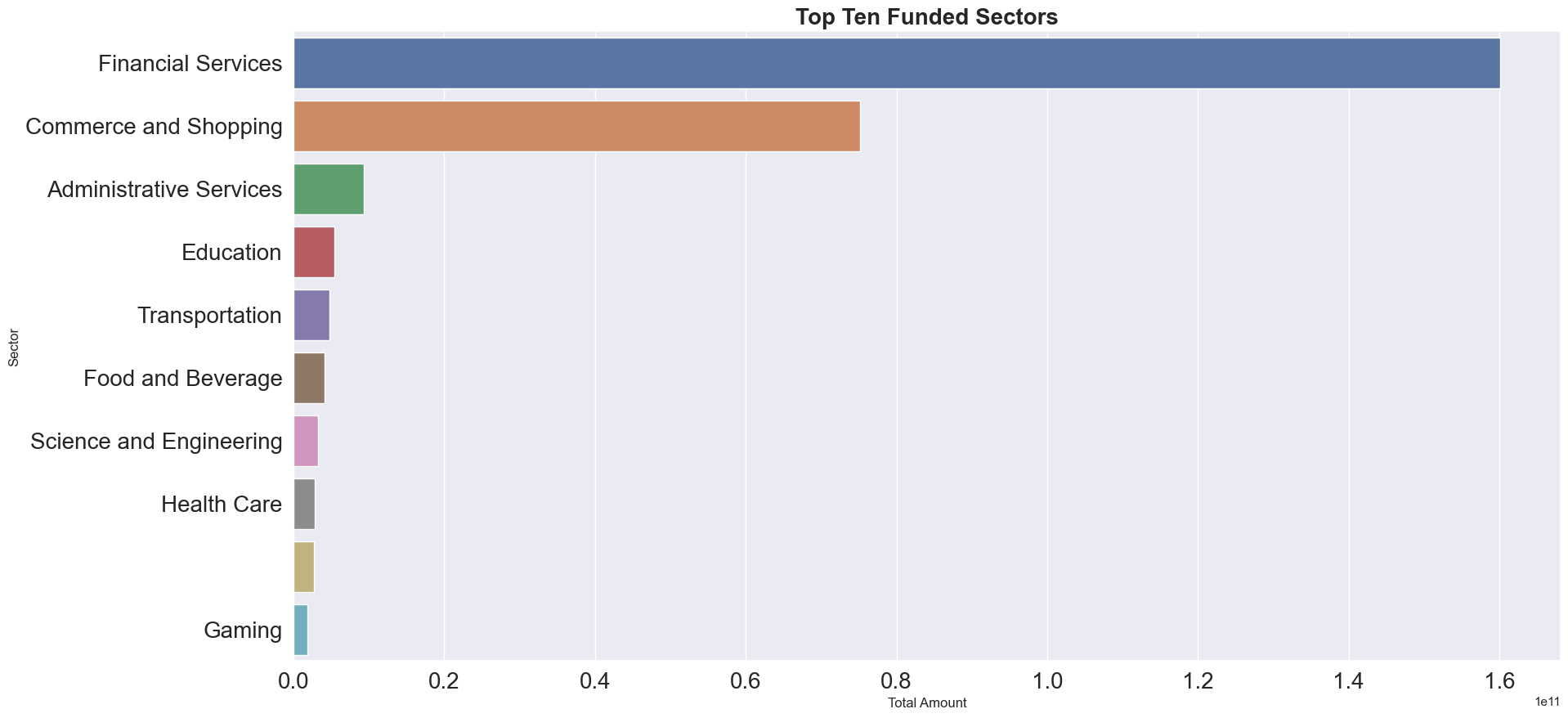
To address this question, we used the ‘groupby’ function in python to group the data according to ‘Year Funded’ and aggregate ‘Amount’ by sum. The resulting visual is illustrated here:



* For the period 2018 -2019, there was a downward trend in the amount of money funded into startups.
* From 2019, the money funded has steadily increased.
* **Question 2**

#### **Which industries or sectors have received the highest funding during this period?**

We grouped the data by the ‘Sector’ column and got the sum of the amounts. The diagram is presented below:

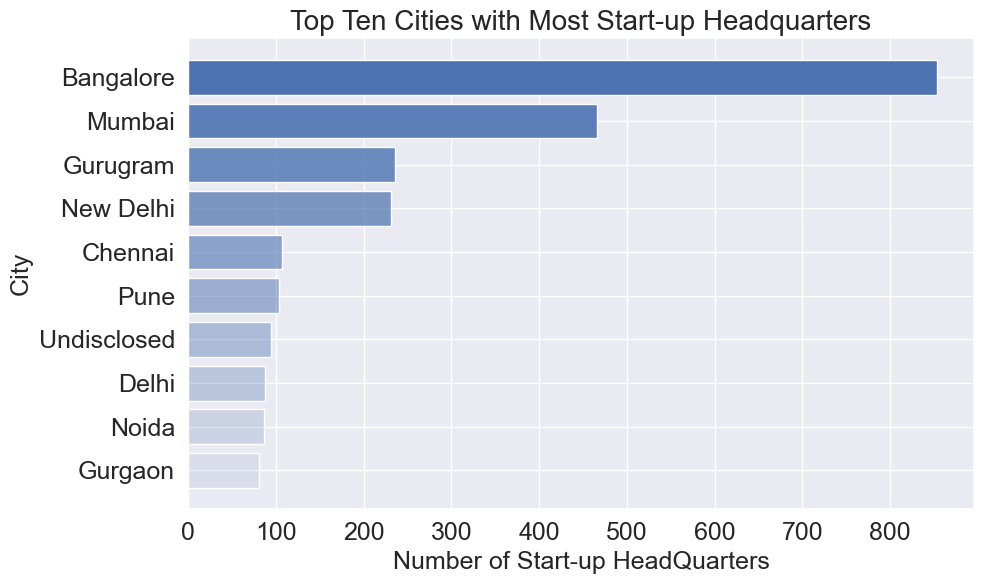


From the graph above, ***Financial Services*** is the highest funded sector, which is then followed by ***Commerce and Shopping*** then ***Administrative Services***. Education, Transport, Food and Beverage are also among the top ten most funded sectors.

### **Question 3**

#### **What is the Distribution of Start-ups Across the Cities in India?**

Here, we plotted the population of start-ups headquarters per city as shown in the diagram below.

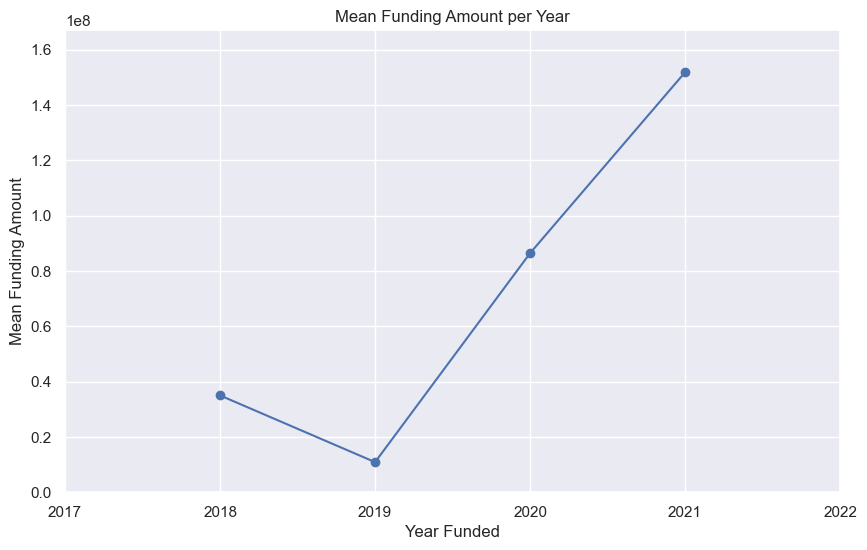


Most startups have their headquarters located in ***Bengaluru (Bangalore),*** which is in the Indian State of Karnataka followed by ***Mumbai*** then ***Gurugram*** and ***New Delhi***.

### **Question 4**

#### **What is the average funding amount received by start-ups in India during this period?**

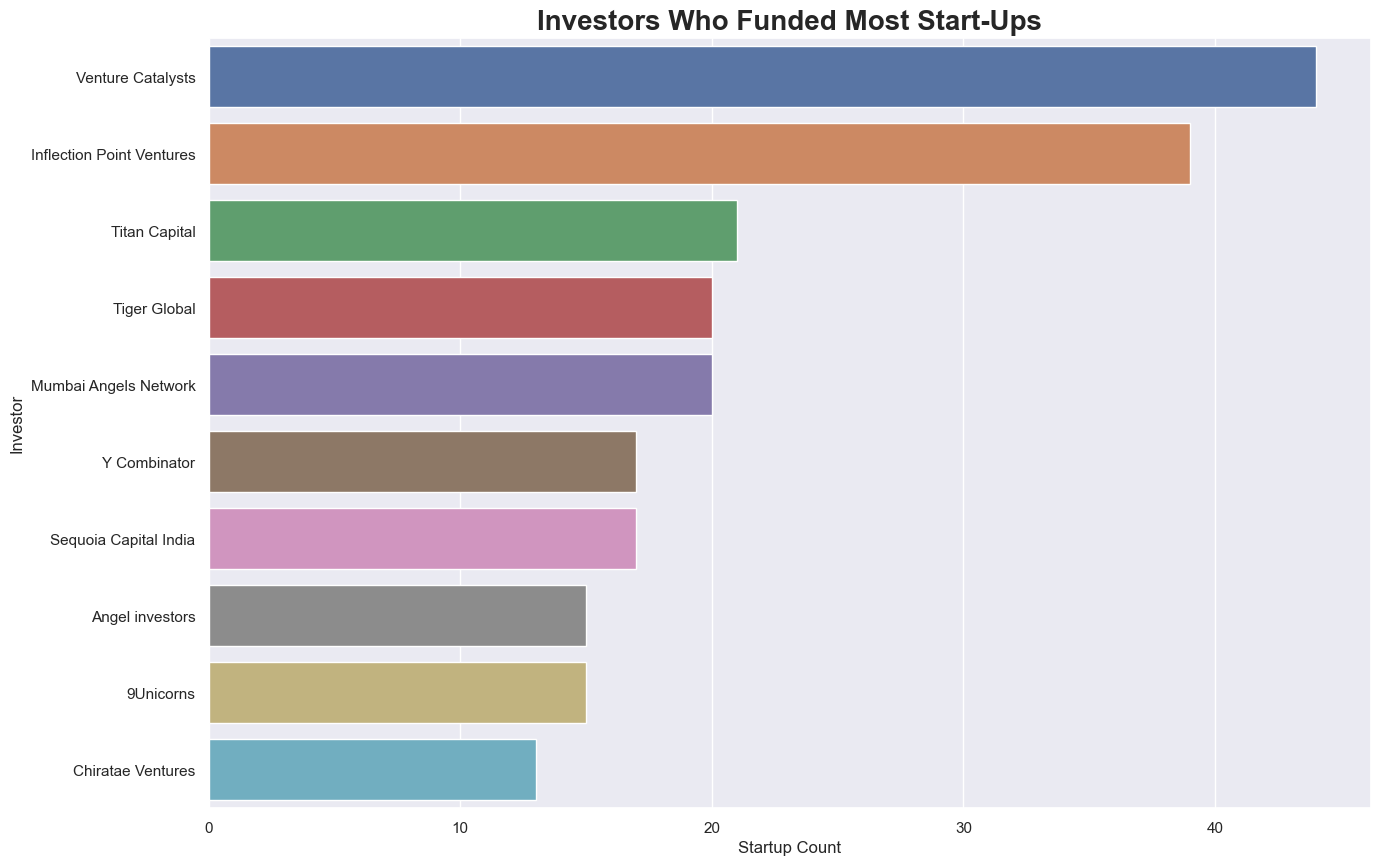
To illustrate the trend of average funding received by start-ups over the period of 2018 to 2021, we grouped the data by ‘Year Funded’ and aggregated the average (mean) of ‘Amount’.



Following the same pattern exhibited by total funding, there was a downward trajectory in the average money funded in startups but from 2019 there was a steady increase.

### **Question 5**

#### **Which Investors Funded Most Start-Ups Over the Period?**

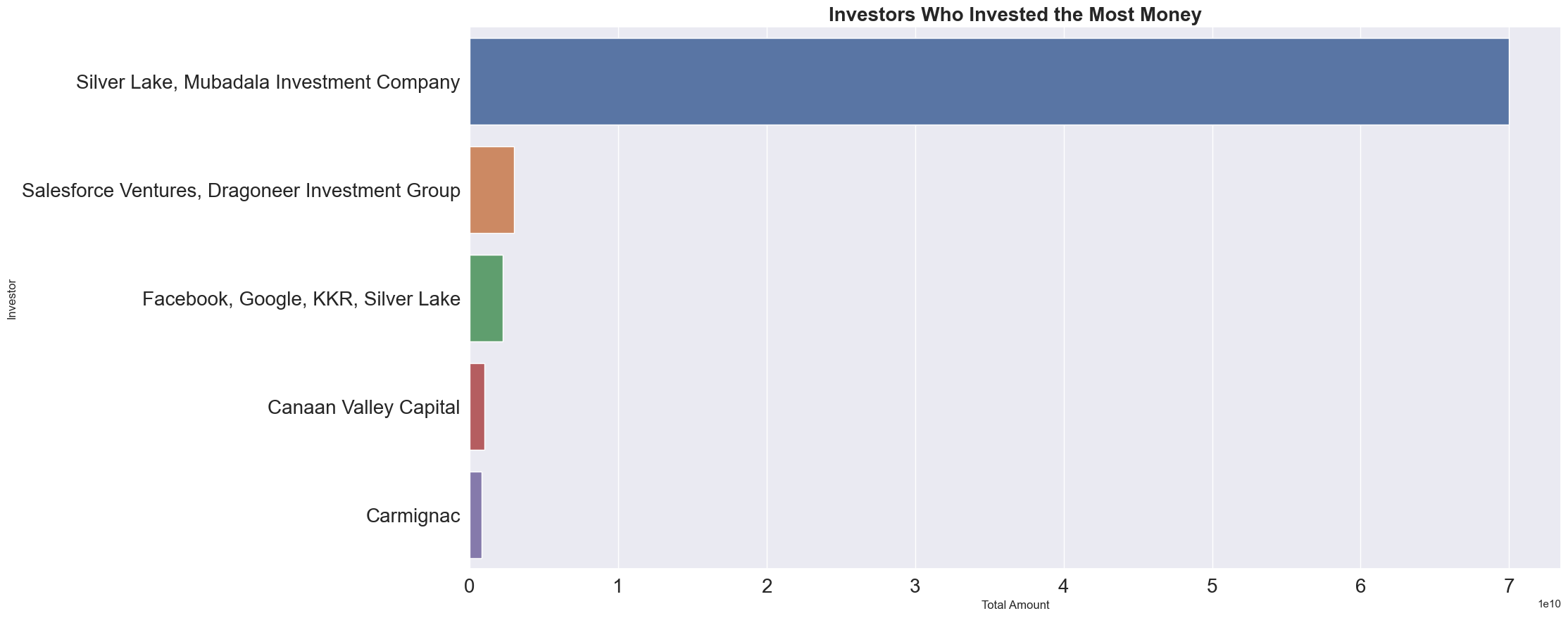


For most startups, their investors were either Unknown or Undisclosed. We assume it makes sense to expurgate such information from our analysis since it will not be of much benefit to our stakeholders.

The leading investor that had invested in most startups was therefore Venture Catalysts, which was closely followed by Inflection Point Ventures. Thereafter came Titan Capital, Tiger Global and Mumbai Angels Network.

### **Question 6**

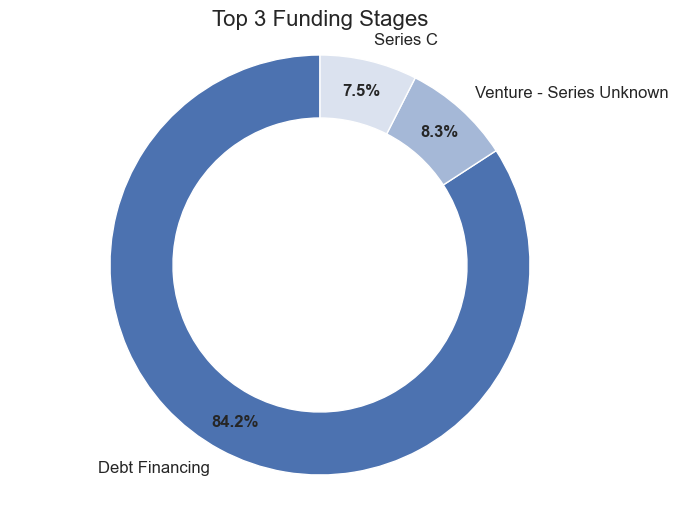
#### **Which Investors Funded the Most Money on Average Over the Period?**



Silver Lake and Mubadala Investment Company are the leading investors who funded the highest average amount into startups.

### **Question 7**

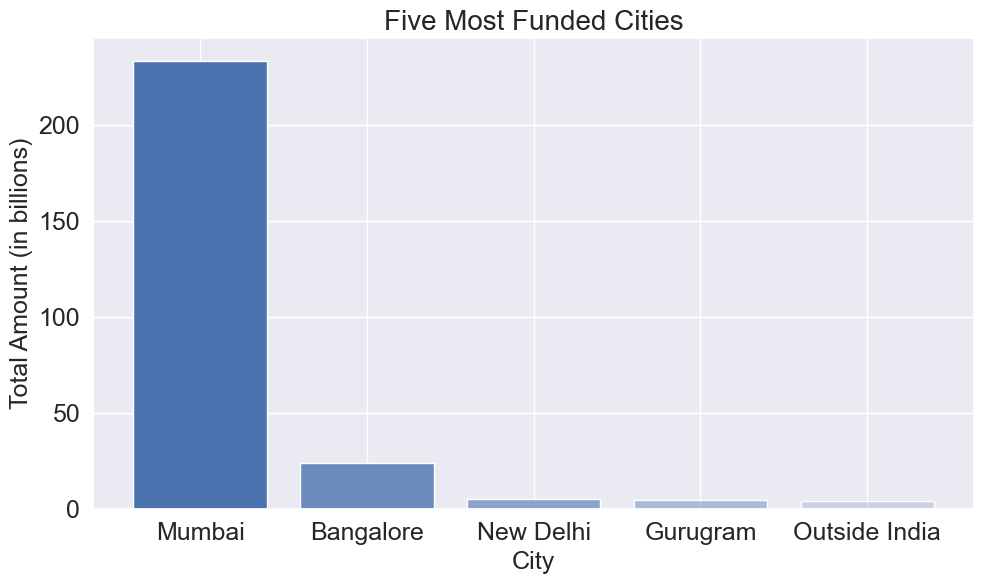
#### **At What Stages were Most Start-ups Funded?**



Dept financing recorded the highest funding in the Indian start-up ecosystem with a whooping ***84.2*** ***percent*** share among three top stages of funding over the period of 2018 to 2021. This was followed by venture equity whose series was not known then Series C. These were just 8.3% and 7.5% respectively.

### **Question 8**

#### **Which 5 Cities Recorded Most Funding?**



According to the analysis, the most highly funded start-ups are in Mumbai as it received over 233 billion dollars over the period. The next was Bangalore, which had a little over 24 billion dollars. New Delhi then came with about 5 billion dollars with just minute gaps from Gurugram and those located outside India.

### **Conclusion**

We discovered through our analysis that funding dropped from 2018 to 2019 by over 17 billion dollars before rising steadily through to 2021. We therefore fail to reject the null hypothesis which states that the funding received by start-ups in India has not demonstrated consistent upward trajectory over the years.

We also found out that there are significant disparities in funding received during various stages. The debt financing stage received the highest funding over the period with 150.7 billion. Venture capital with unknown series, however, came second with 14.9 billion and Series C with 13.4 billion.

**Recommendations**

* Situation analysis for a venture into the Indian ecosystem regarding funding of start-ups is particularly important as it is highly likely to receive funding as a start-up when headquartered at Mumbai, Bangalore, New Delhi and Gurugram respectively.
* Alternative financing sources must be implored to support a start-up since most of the start-ups failed to receive highly significant funding at their initial stages. The earliest highly funded was at the Series C stage when it is expected to be a market fit.
* Considering the sector to venture into as far as funding for a start-up in India is concerned is vital as it is more likely to receive funding when at the financial services, commerce and shopping, administrative and education sectors, respectively.