# Title: Indian Start-up Funding Analysis

**Introduction:**

Indian startups have been making waves in recent years, attracting significant attention and funding from investors both domestically and globally. As the startup ecosystem in India continues to flourish, data analysis becomes crucial in understanding the trends, patterns, and prospects of startup funding. In this article, we delve into the world of Indian startup funding, examining key insights and notable trends.

Data analysis refers to the process of inspecting, cleaning, transforming, and modeling data to discover useful information, draw conclusions, and support decision-making. It involves applying various techniques and methods to uncover patterns, trends, correlations, and insights within a dataset, with the ultimate goal of extracting meaningful and actionable knowledge. Data analysis typically follows a systematic approach that includes several key steps:

1. **Data Collection**: Gathering relevant and reliable data from various sources, which can include structured data from databases, spreadsheets, or surveys, as well as unstructured data such as text documents, images, or social media posts.

2. **Data Cleaning**: Preprocessing and cleaning the data to handle missing values, outliers, inconsistencies, and errors. This step ensures that the dataset is accurate, complete, and suitable for analysis.

3. **Exploratory Data Analysis (EDA):** Conducting an initial exploration of the dataset to gain insights and identify patterns or trends. This involves generating summary statistics, visualizations, and descriptive analysis to understand the distribution, relationships, and characteristics of the data.

4. **Data Modeling and Analysis:** Applying statistical techniques, machine learning algorithms, or other analytical methods to extract meaningful information from the dataset. This step can involve hypothesis testing, regression analysis, clustering, classification, or predictive modeling, depending on the objectives and nature of the data.

5. **Interpretation and Visualization**: Interpreting the results of the data analysis and presenting them in a meaningful and understandable way. Visualizations such as charts, graphs, and dashboards are commonly used to communicate the findings effectively.

6. **Conclusion and Decision Making**: Drawing conclusions based on the analysis and using the insights gained to make informed decisions, solve problems, or address specific research questions or business challenges.

Effective data analysis requires a combination of domain knowledge, statistical expertise, programming skills, and critical thinking. It involves understanding the context and objectives of the analysis, selecting appropriate techniques, and accurately interpreting and communicating the results to stakeholders.

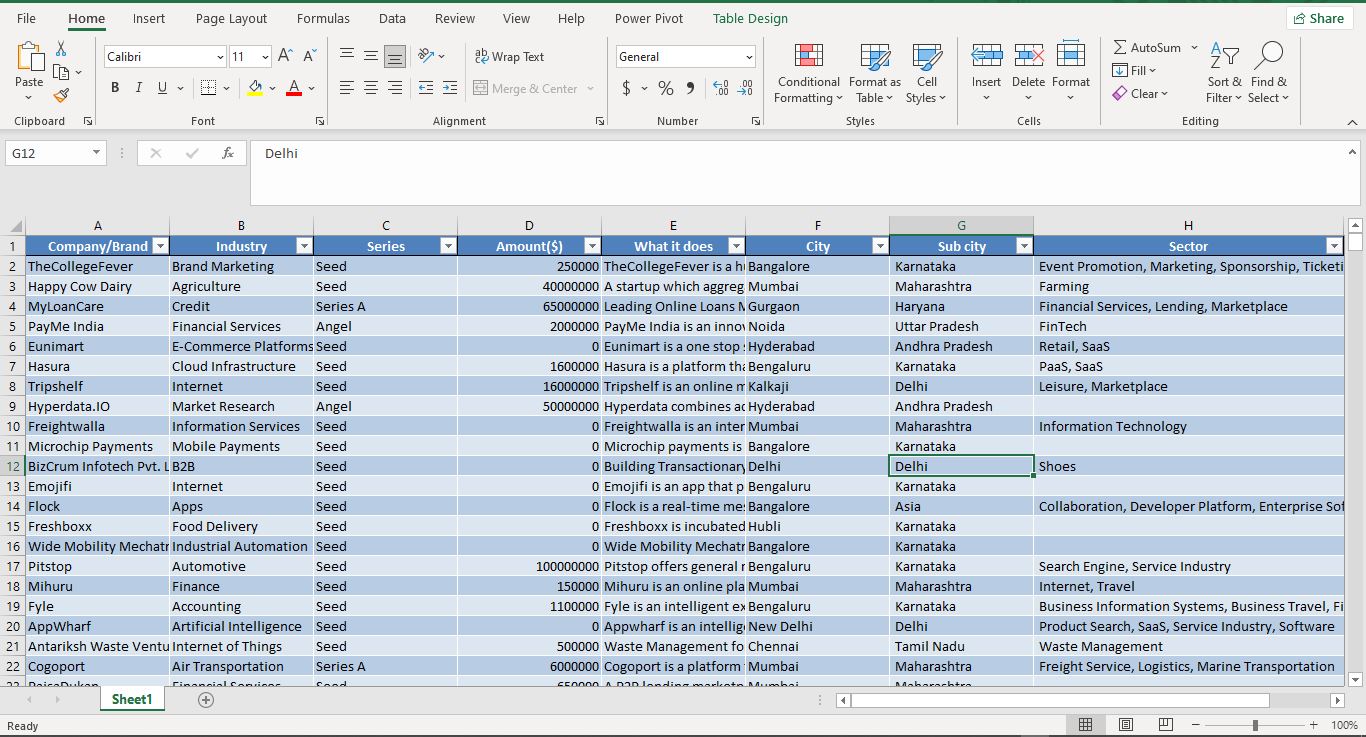
Now, we take a look at the steps of data analysis and how we implemented them into our project. So as part of the project I:

* Asked six questions
* Developed a hypothesis
* Processed the data
* Analyzed the data
* Visualized the data

The goal of this project is to analyze funding received by start-ups in India from 2018 to 2021 and investigate the ecosystem and propose the best course of action as my team is trying to venture into the Indian start-up ecosystem.

**1. Data Collection**: The data was sourced already for me, they were Indian startup funding datasets of 2018, 2019, 2020 and 2021.

The data is shown like below for each year.



The column names and descriptions are as follows:

* Company/Brand: Name of the company/start-up
* Founded: Year start-up was founded
* Sector: Sector of service
* What it does: Description about Company
* Founders: Founders of the Company
* Investor: Investors
* Amount ($): Raised fund
* Stage: Round of funding reached

After viewing the data, six questions were asked to be answered by the analyzing of the data and a hypothesis was developed. The six questions were:

Questions To Be Answered:

• What is the trend in value of Indian start-up funding over the years?

• Has the trend in the number of Indian start-ups increased or decreased over the years?

• Which industries have the highest number of start-ups?

• Which industries have received the highest funding amounts?

• What is the average funding amount for Indian start-ups?

• Are there any correlations between headquarters location and share of funding?

And the hypothesis was:

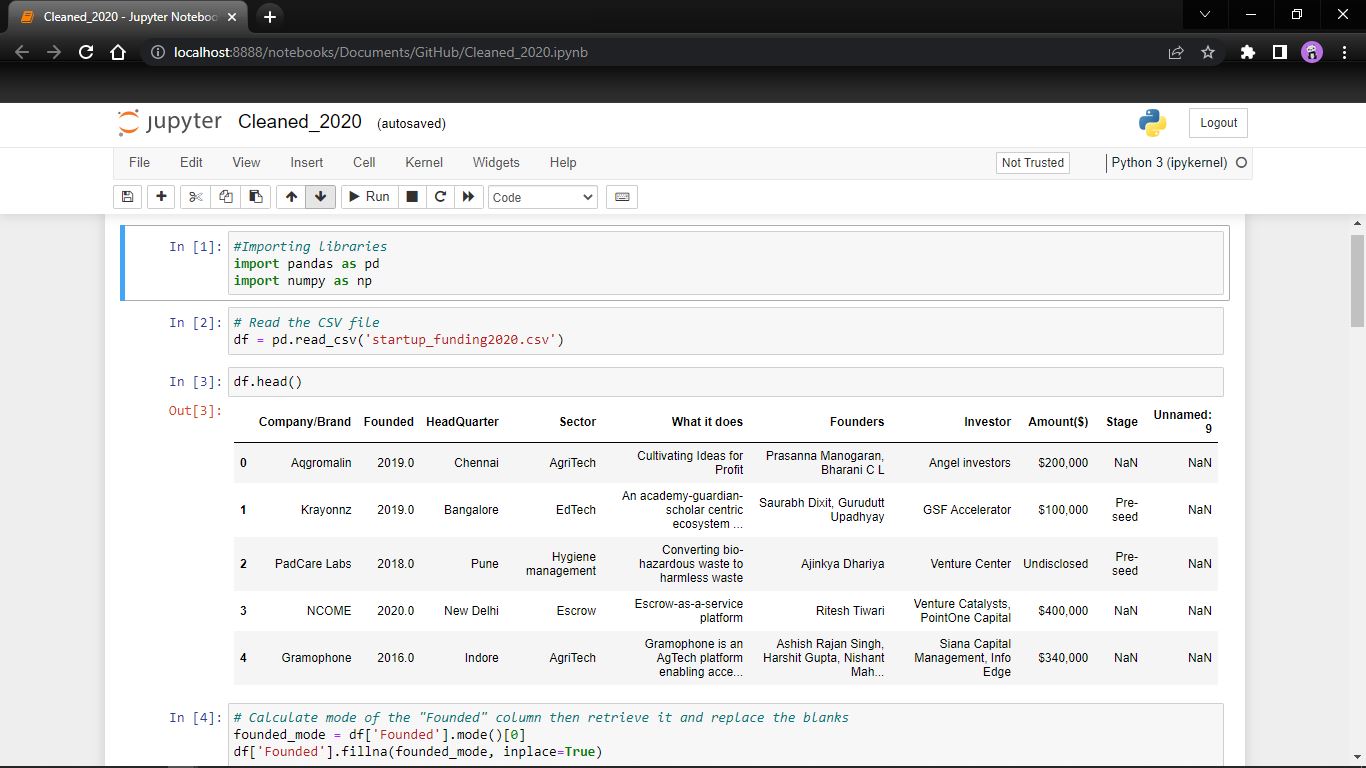
Hypothesis:

Null Hypothesis: There is no significant correlation between City and Amount

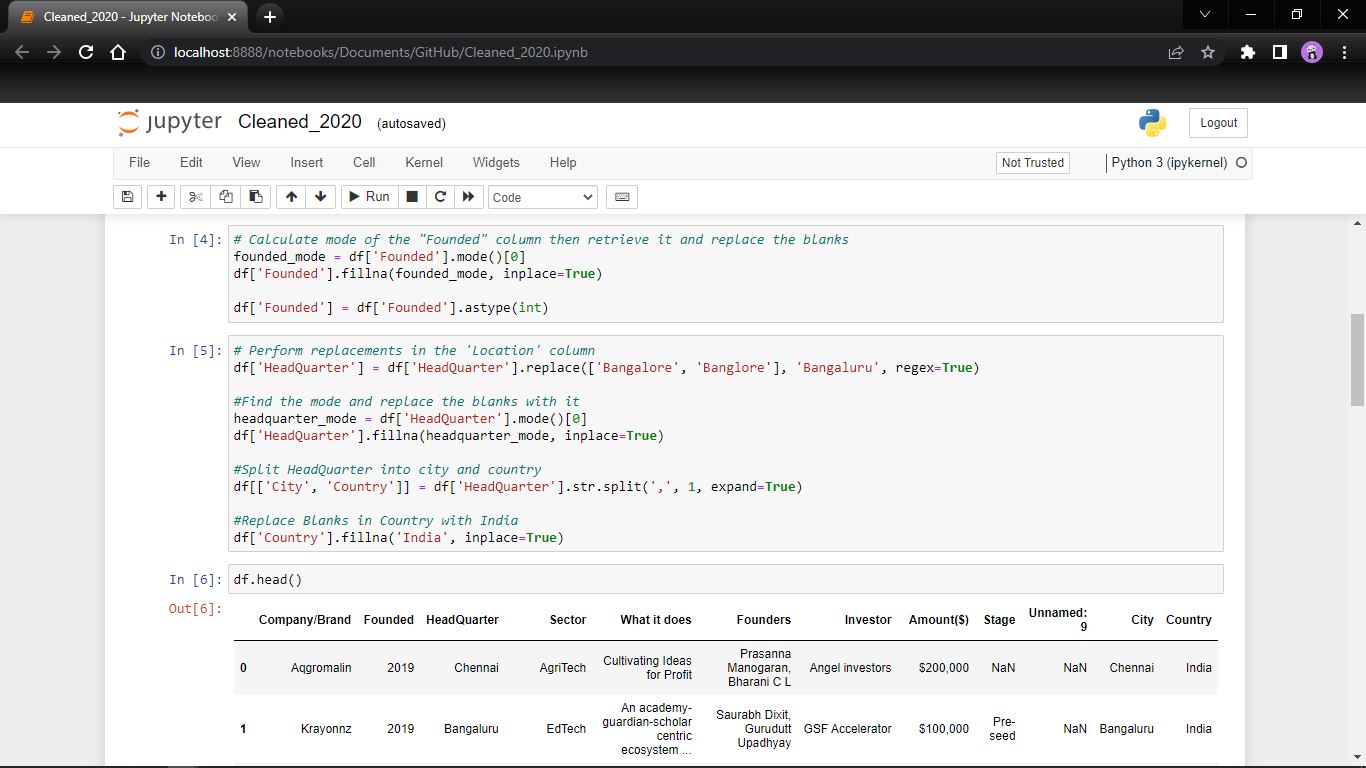
Alternate Hypothesis: There is significant correlation between City and Amount

Then we went on to clean the data.

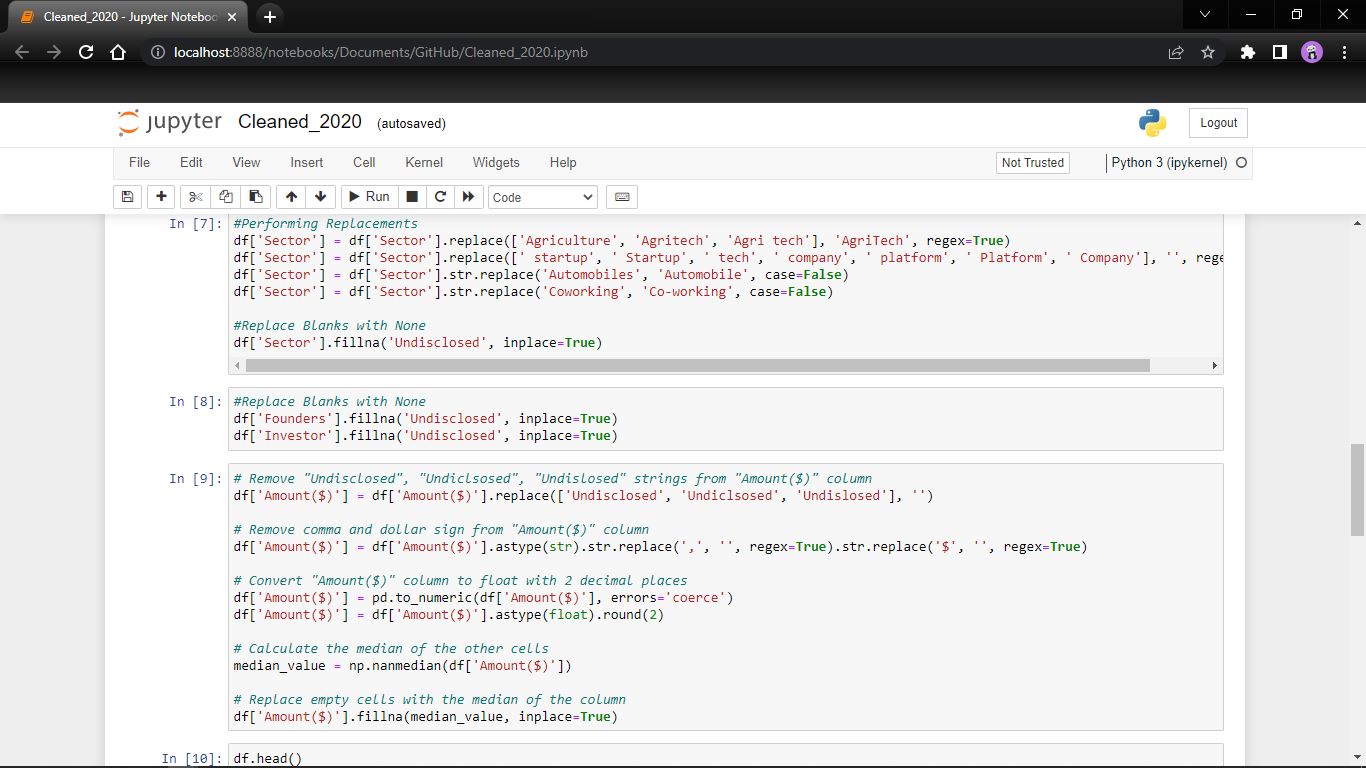
**2. Data Cleaning**: The cleaning of the data was done with jupyter notebook.



Started by importing the libraries, then reading the csv file and showing the first five rows of the dataset. I started from the founded column, it was seen that there were empty cells (missing values) in the founded column so those were replaced with the mode of the founded column, that is the most occurring year.



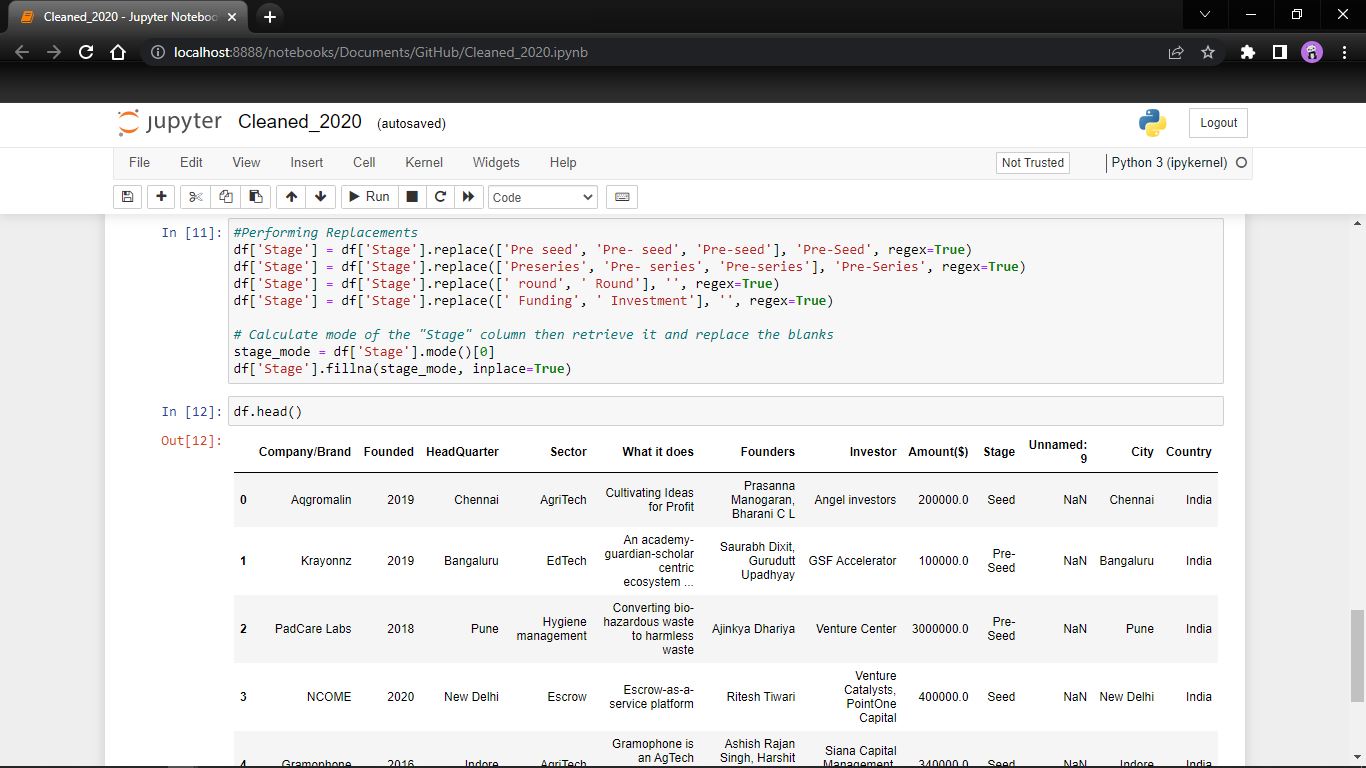
The next column to be cleaned was the headquarter which was split into city and country, it also had empty cells which were replace with the most occurring city, then there were some spelling corrections too done. Duplicates were removed.



Then the Sector column had some missing values, some case issues, uppercase and lowercase which were the same, so I changed them into one and replaced them. Founders and Investors column were only replaced missing values.

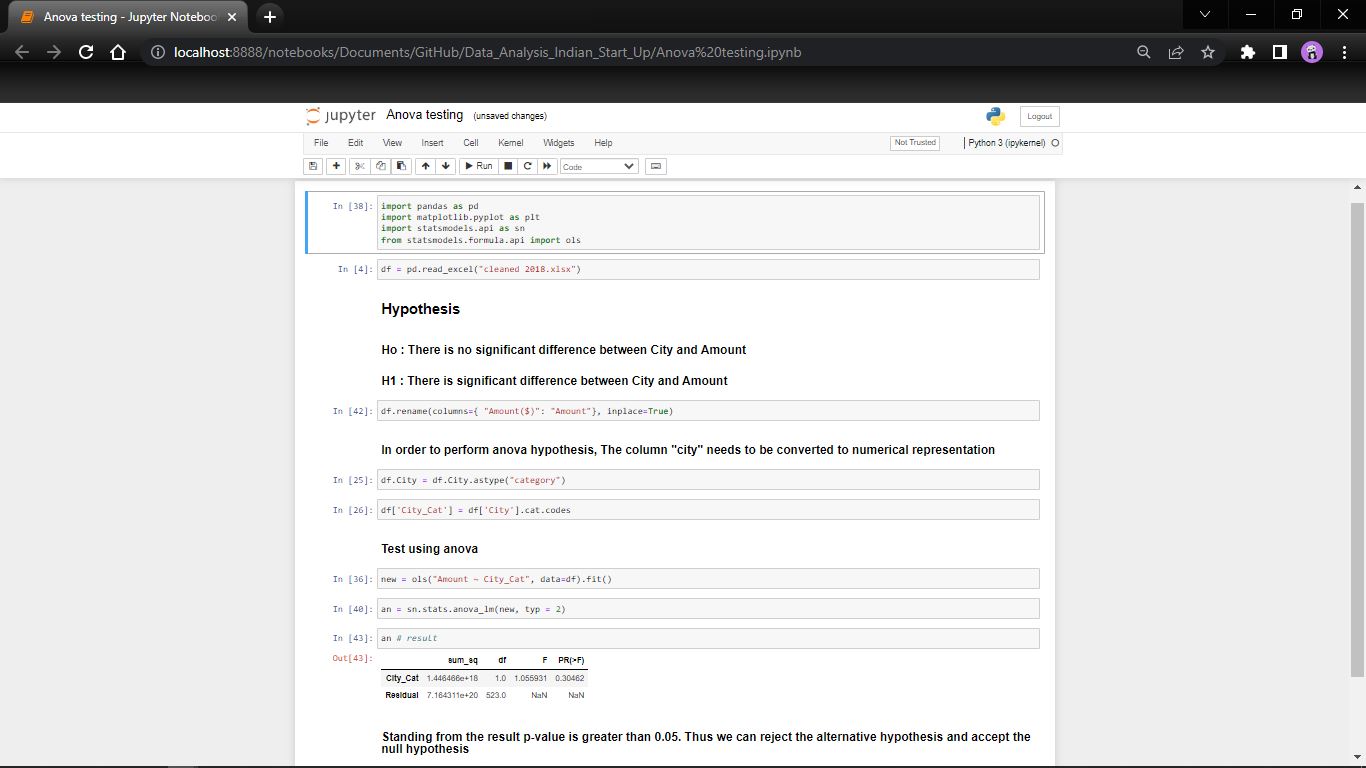
The Amount column was an integer column but had strings of “undisclosed” in them, firstly the “undisclosed” was replaced with blanks then the “$” signs and ”,” comma too were removed before getting the median of total values and it was used to replace them.

The last column was the stage column which had similar naming, missing values and uppercase lowercase issues, similar names and uppercase, lowercase naming were found and replaced so as to be one uniform name, missing values were replaced with the mode value of the column.



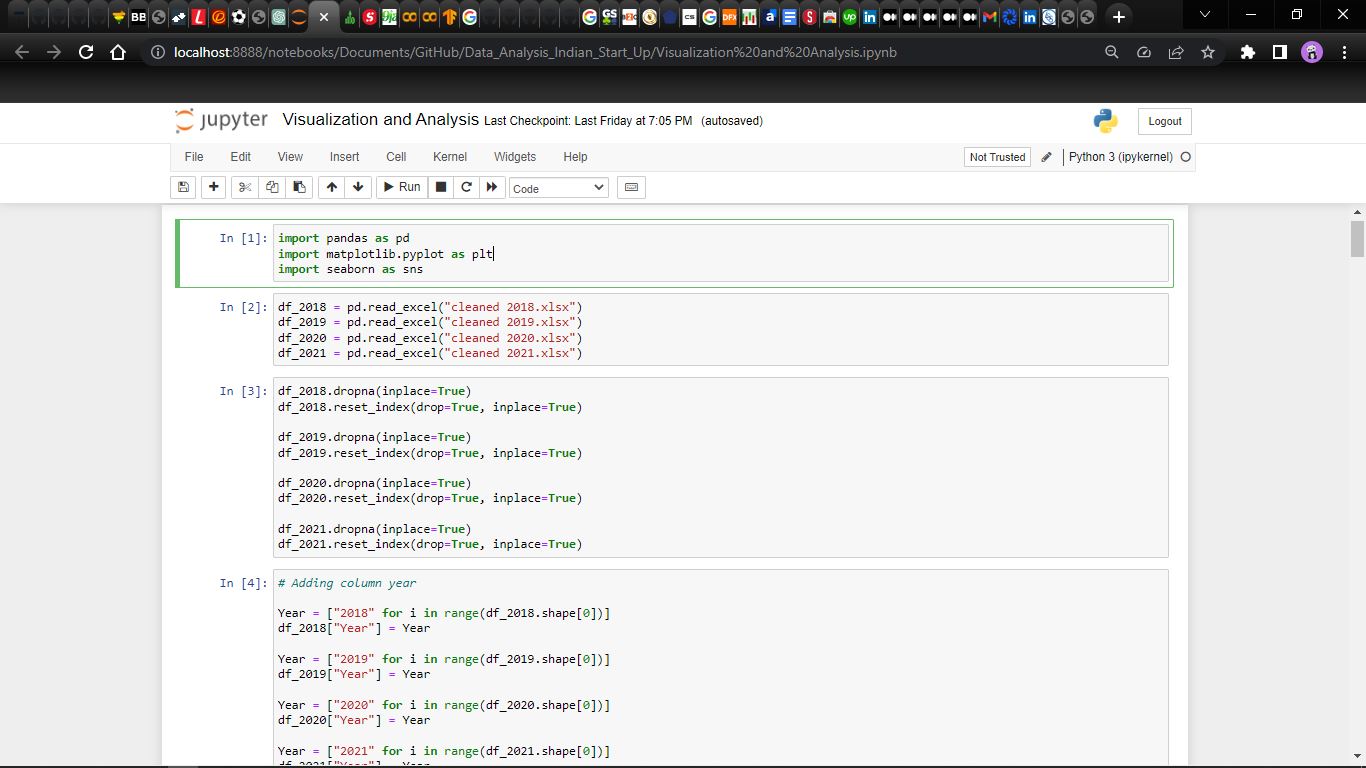
The same process was used for all four datasets, 2018, 2019, 2020 and 2021.

Next was EDA and Analysis.

**3. EDA and Analysis:** The hypothesis testing was done with a null hypothesis and alternative hypothesis. The Null Hypothesis states that there is no significant correlation between City and Amount of Funding while the Alternate Hypothesis states that there is significant correlation between City and Amount of Funding. This was tested out using Anova hypothesis testing.

Thus, it gave us a result greater than 0.05 which means we reject the alternate and accept the null hypothesis.

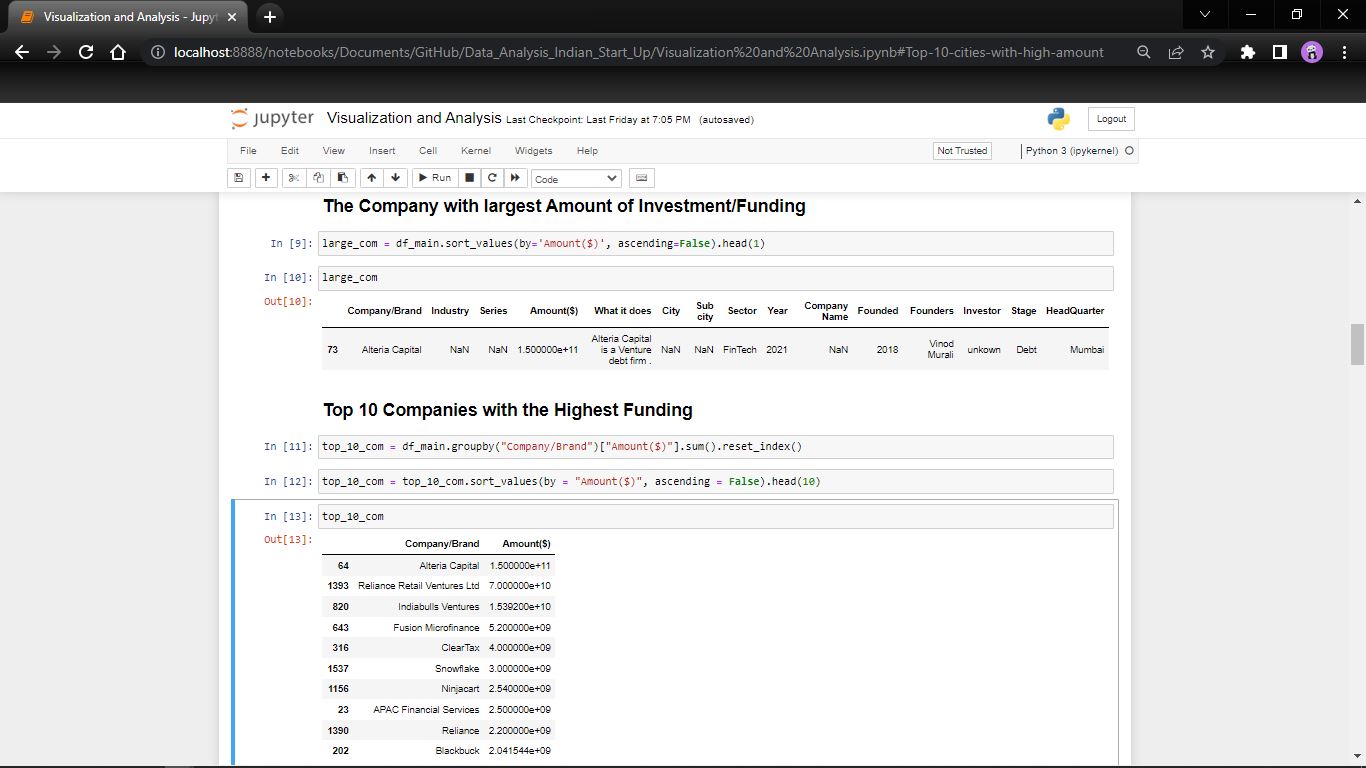
Then came the analysis part of the project, this was also done in jupyter notebook with the help of some visualizations so seaborn was imported as sns and matplotlib pyplot as plt, before we imported all the four datasets separately for merging, the year column was added to the merged dataset, the dataset was also checked for null values.



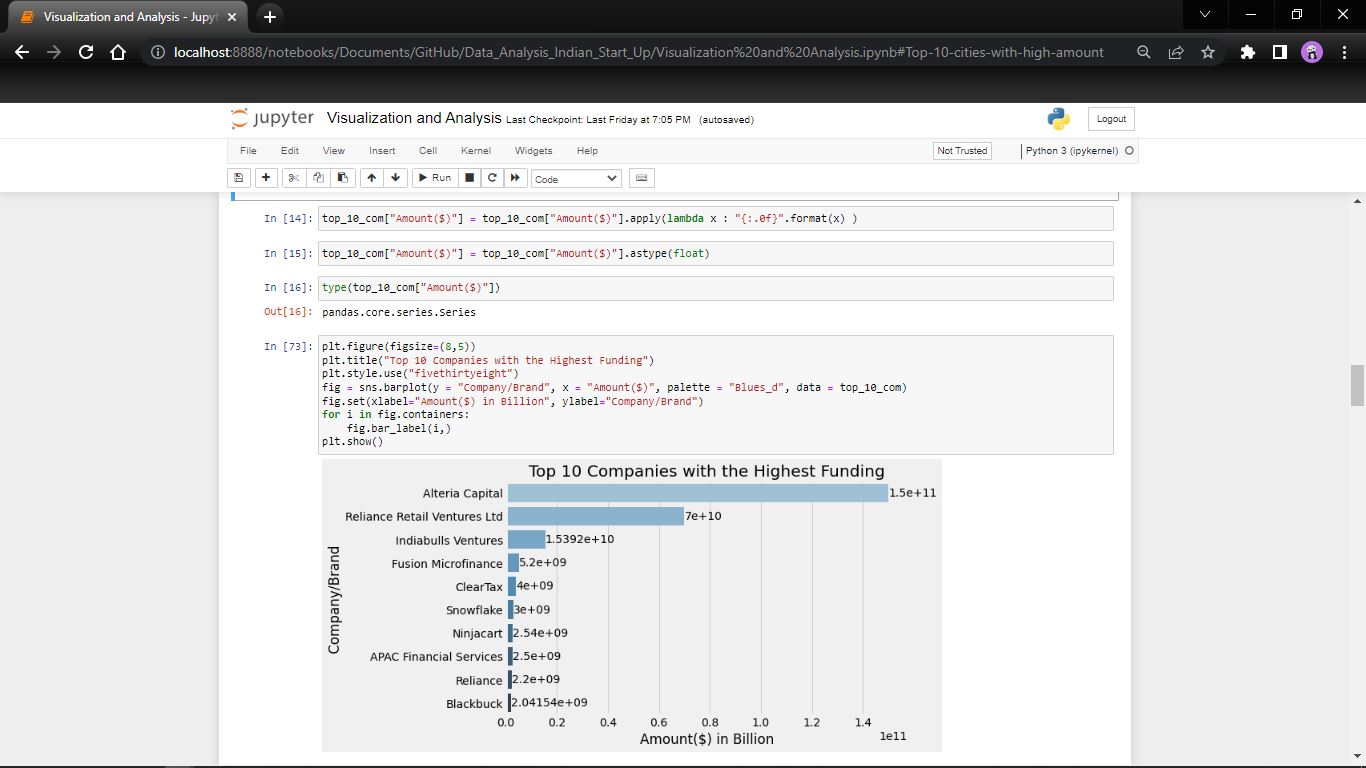
After merging, analysis and visualization started.

Firstly, I checked for the company with the largest amount of investment and then the top 10 companies with highest Amount of funding.

This shows **Alteria Capital** as the company with the largest investment with a funding of $150,000,000,000 which is a venture debt firm, it is in the FinTech Sector and was founded in 2018 by Vinod Murali with their headquarters in Mumbai. It is currently in Debt stage.

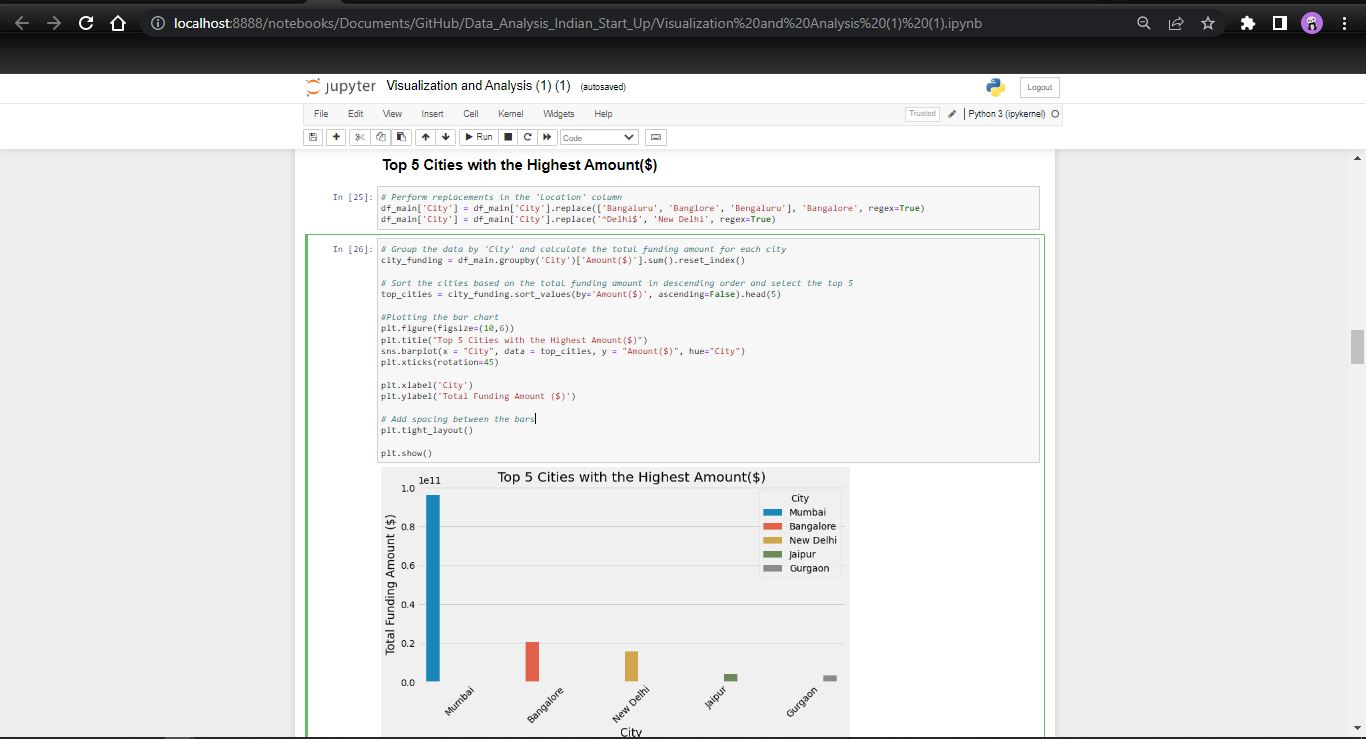


This shows the top 10 companies with the highest funding amount, Alteria Capital coming in first with $150,000,000,000, then Reliance Retail Ventures Ltd with $70,000,000,000 and then Indiabulls Ventures with $15,392,000,000 in third, they are all located in Mumbai.

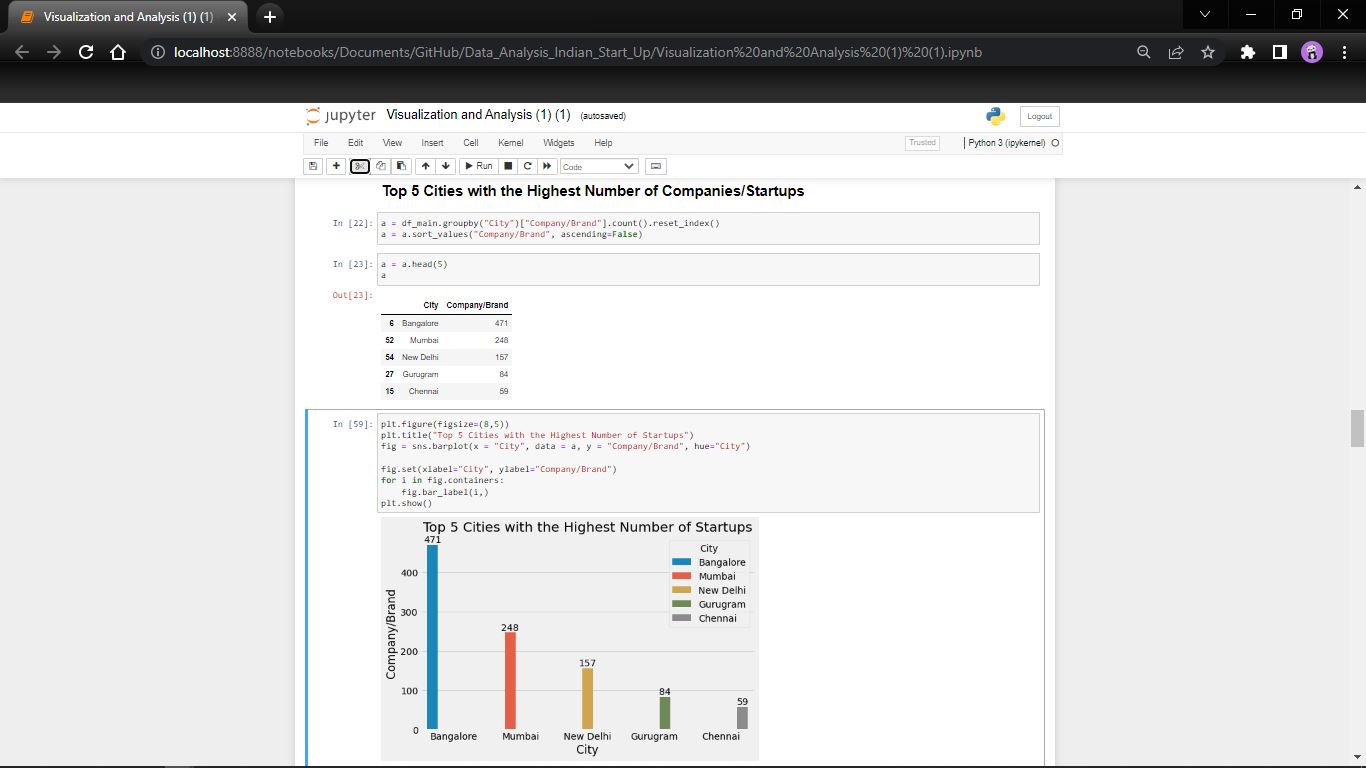


I also checked for the cities with the highest amount of funding and also the cities with the highest number of companies.

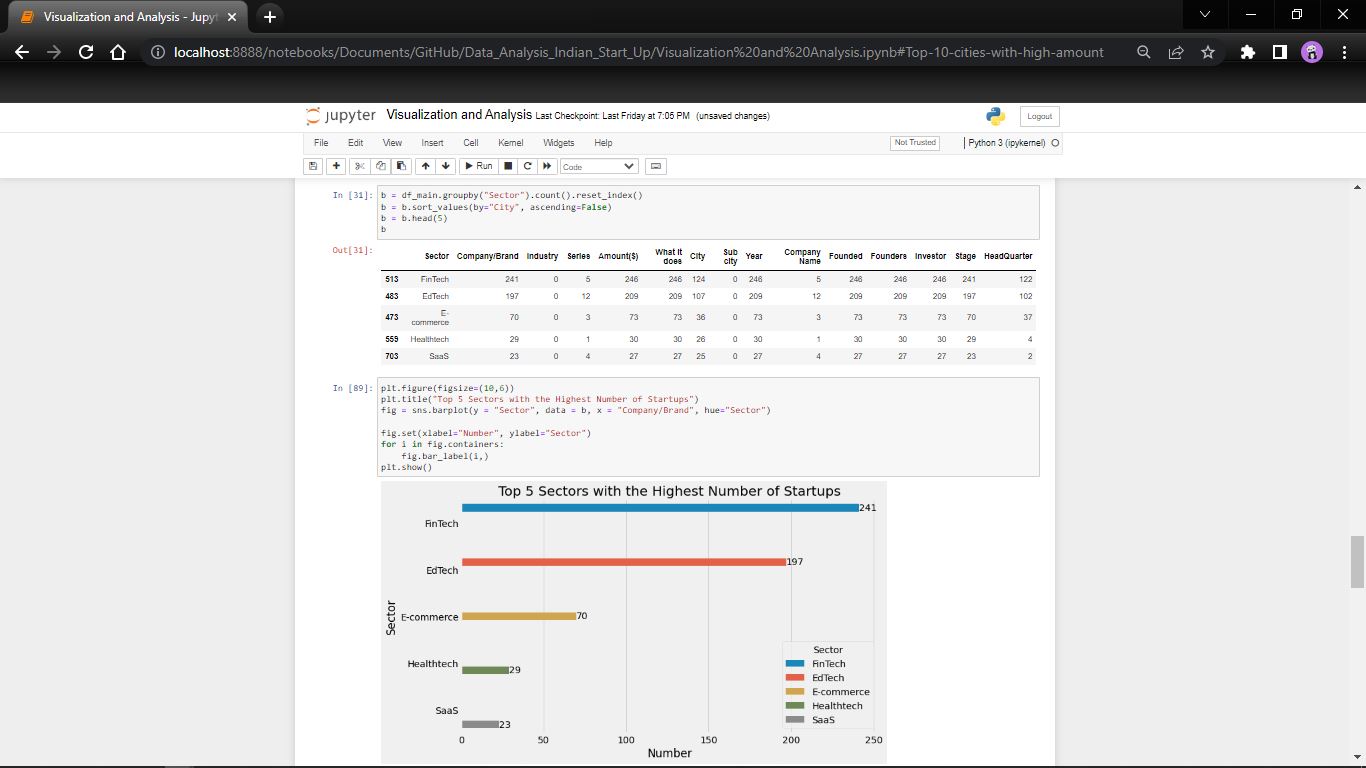
This shows the Cities with the highest amount of funding, Number 1 is Mumbai by a really large margin followed by Bangalore and then New Delhi.



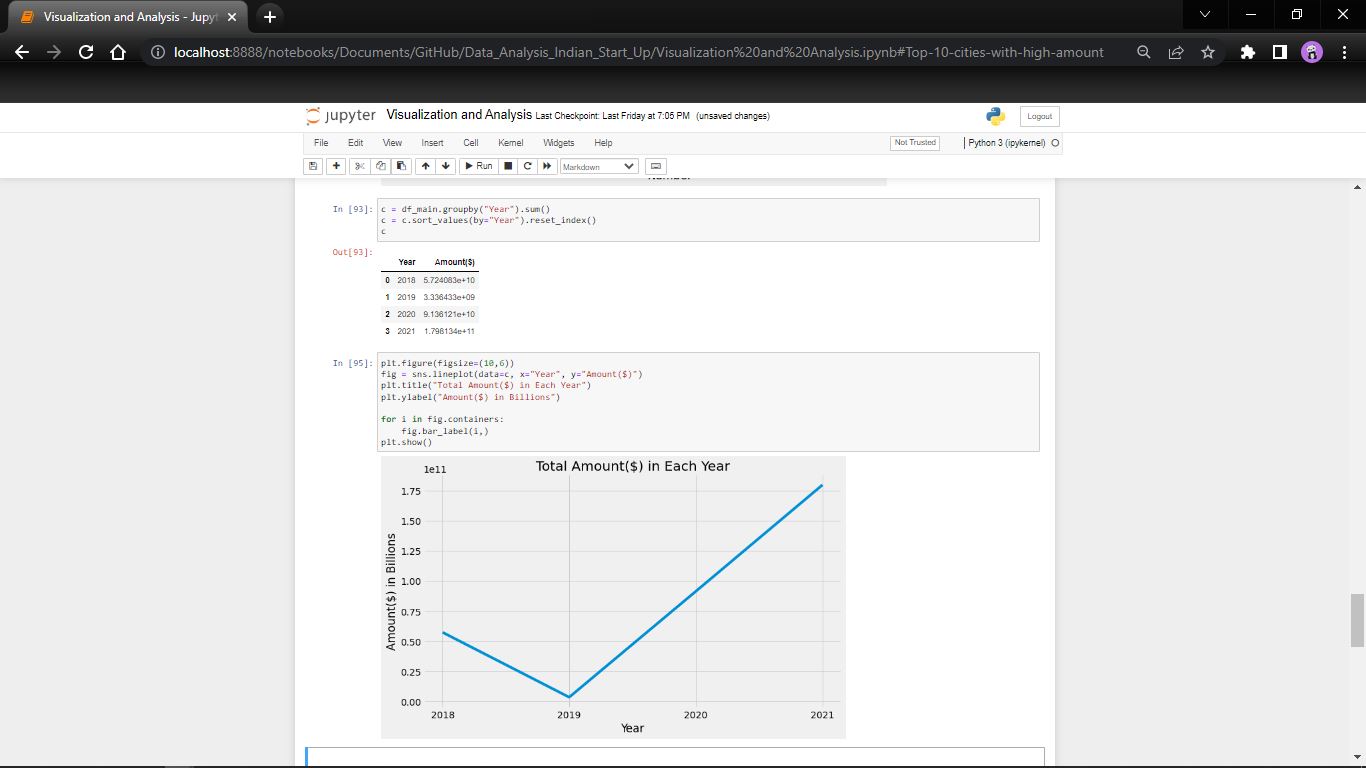
This is the visualization for the top 5 cities with the highest number of startups, Bangalore being outstandingly higher than others with 471 startups and then followed by Mumbai with 248 startups then New Delhi with 157 startups.



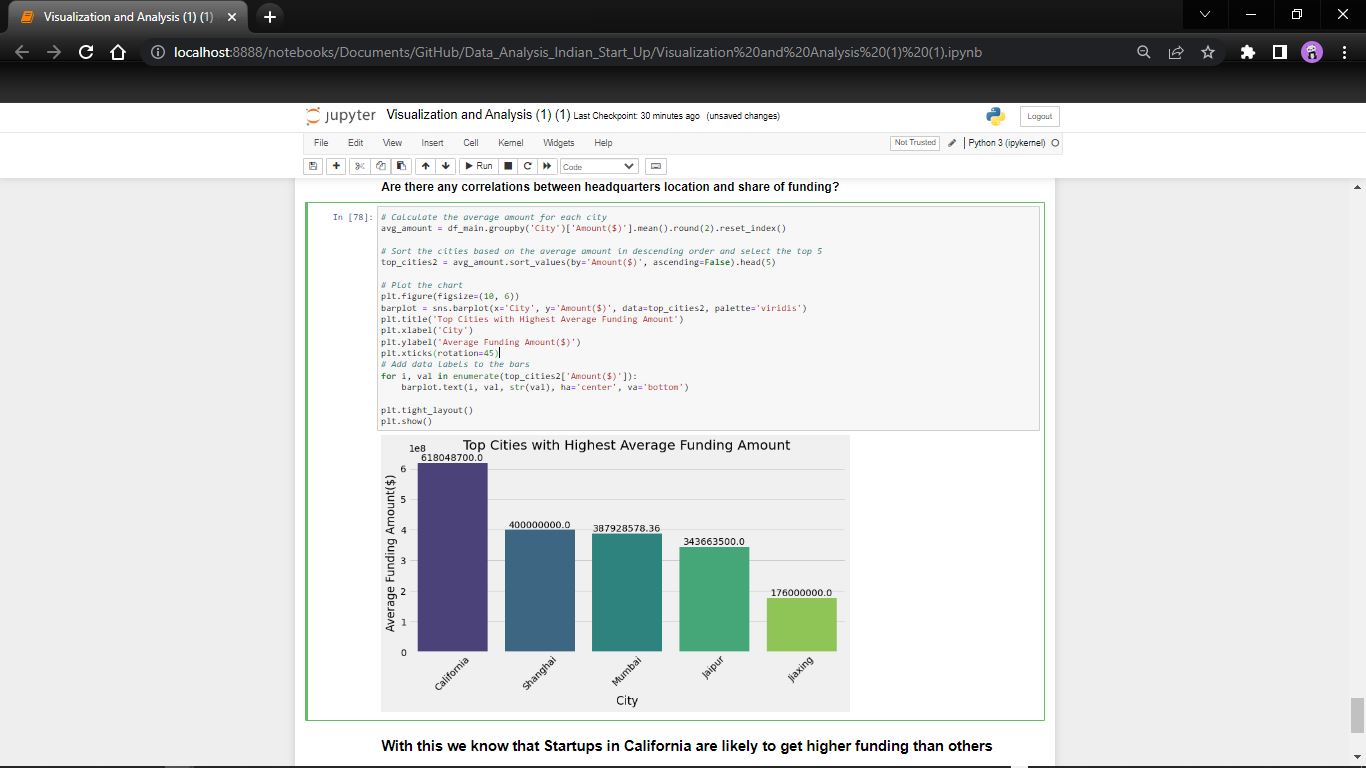
Next, was getting the top 5 Sectors with the Highest Number of Startups, FinTech had the highest number of startups with 241 Startups, next was EdTech with 197 Startups and thirdly, E-commerce with 70 Startups, as shown in the image below.



I also got the total amount in each year and it is seen that in general it keeps moving up except the dip in 2019.



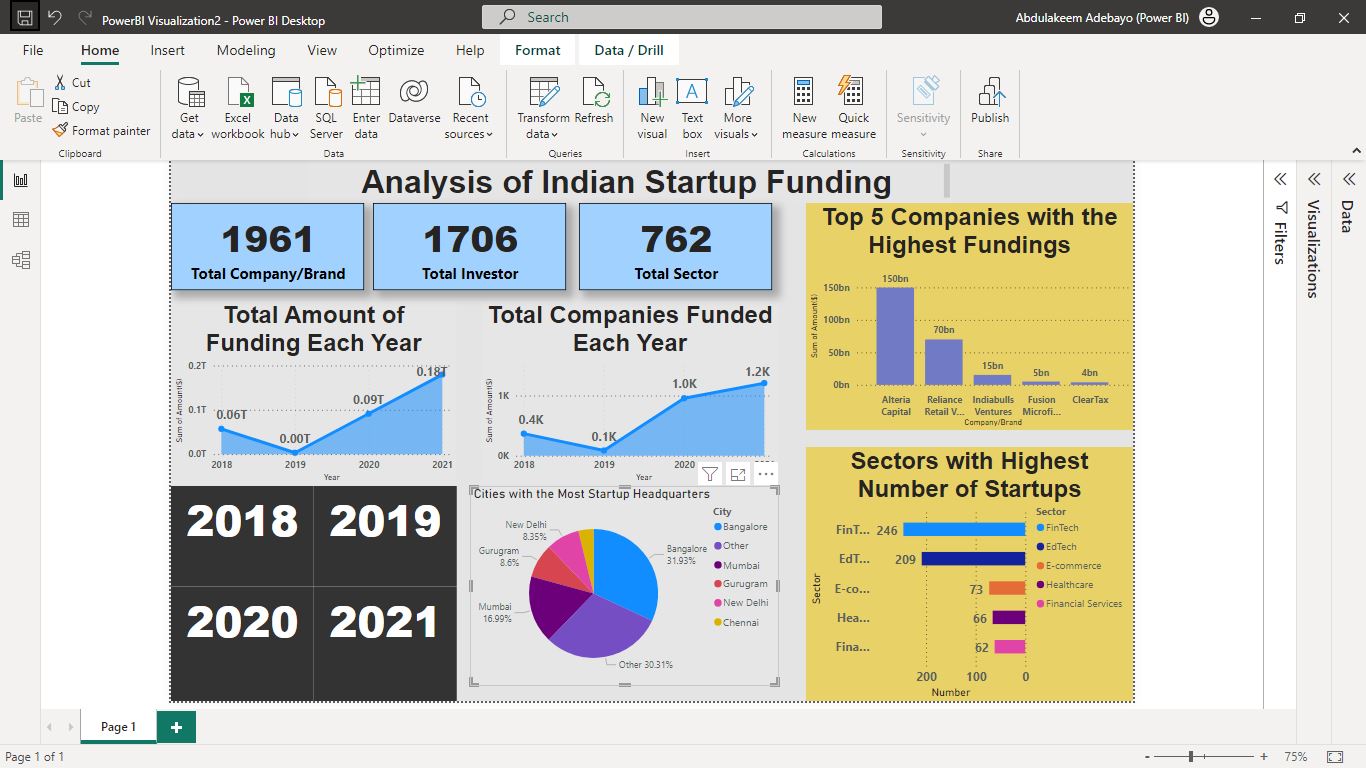
Lastly, I wanted to see the average funding of each city, which would give me a better view of how well each city is being funded on average not the highest or the lowest.



And California had the highest average by a margin, Shanghai was next before the favored Mumbai came in 3rd.

After analyzing and answering the 6 questions, next was visualization using PowerBI.

**5. Visualization**: The visualization depicts a total of 1,961 companies, a total of 1,706 investors and a total of 762 sectors and the four years (2018 – 2021). The Total amount of funding each year and the total companies funded each year is directly proportion which means there is an increasing growth of startups each year. The top 3 companies with the highest fundings are located in Mumbai, while the sector with the most startups is FinTech and the city is Bangalore.



**6. Conclusion**: In conclusion, the best course of action to take venturing into the Indian start-up ecosystem is venture into a Fintech company with its headquarters in Mumbai. Mumbai as shown in the analysis plays a significant role in making startups successful, it has the highest top 3 funded companies, it is the 3rd most averagely funded city and also popular among startups. FinTech has the highest funded company and also the most popular startup.