**Question 1:**

Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations -  Bangalore, Mumbai, and NCR. As a friend, you want to help your friend deciding the location. NCR include Gurgaon, Noida and New Delhi. Find the location where the most number of funding is done. That means, find the location where startups has received funding maximum number of times. Plot the bar graph between location and number of funding. Take city name "Delhi" as "New Delhi". Check the case-sensitiveness of cities also. That means, at some place instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore". For few startups multiple locations are given, one Indian and one Foreign. Consider the startup if any one of the city lies in given locations**.**

**Answer:**

Afterimporting various library like **pandas**, **matplotlib.pyplot** , **numpy** and choosing the column “CityLocation” from the dataframe I have replaced the wrong spellings of two cities. “Delhi” to “NewDelhi” and “bangalore” to “Bangalore”.

Using **dropna** method I have dropped the rows with “**NaN**” values present and keeping the index same.

Create two **dictionaries** 1. city\_dict for city wise number on fundings and

2. Location\_dict for location wise number of fundings.(as NCR is considered One Location).

**separateCity function:**

* However in the dataframe for CityLocation at some places two locations one Indian and one Foreign and the cities are separate by “/”. So we have to **split** the names by “/”. After **splitting(split())** it becomes a list and after looking at each data in list and **stripping(strip())** it I check for location Mumbai, Bangalore, New Delhi, Gurgaon, Noida .
* If city name is Mumbai Bangalore it is added to both city\_list and location\_list as per it’s name but if city name is New Delhi, Gurgaon, Noida then in location\_list it is added as NCR and in city\_list by the city name.
* Finally data is entered in location\_dict using location\_list and city\_dict using city\_list.

Using the **apply** **method** I have implemented the separateCity function to each row of the entire column of CityLocation.

After **apply method** I have used the **Counter** **method** which work same as a dictionary and counts values for keys for city\_dict and location\_dict.

Now using the most\_common method a list of tuples of highest\_city and highest\_location is created.

Following data city wise and location wise is retrieved from the dataframe.

Table 1 City Wise Number of Funding's

|  |  |
| --- | --- |
| **CITY** | **Number of Funding’s** |
| 1. Bangalore | 637 |
| 1. Mumbai | 449 |
| 1. New Delhi | 389 |
| 1. Gurgaon | 241 |
| 1. Noida | 79 |

**We can visualize the data in a bar graph or pie chart using matplotlib.pyplot.**

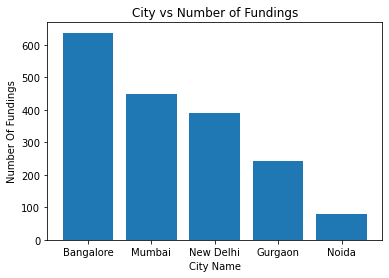
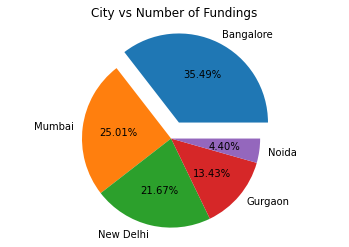


Figure 1: Pie Chart for City vs. Number of funding’s Figure 2: Bar graph for City vs. Number of funding’s

According to the data If **we take city-wise data into consideration** then **Bangalore city which has highest number of funding’s 637** and has received **35.49% of total number of funding’s** then **I would suggest my friend the city of Bangalore for starting his startup.**

However if we take the **NCR location that comprises the city of New Delhi, Gurgaon, Noida** as one location then the data gives us a different outcome.

Table 2 Location Wise Number of Funding's

|  |  |
| --- | --- |
| **Location** | **Number of Funding’s** |
| 1. NCR | 709 |
| 1. Bangalore | 637 |
| 1. Mumbai | 449 |

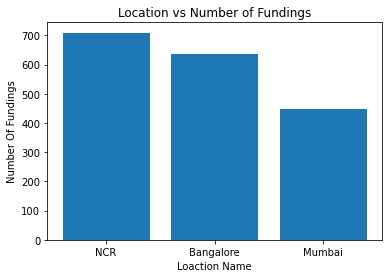
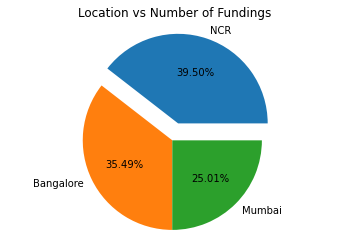


Figure 3: Pie Chart Location vs Number of Funding's Figure 4: Bar Graph Location vs. Number of Funding's

According to the data If we **take location-wise data into consideration** then **NCR location** **which has** **highest number of funding’s 709** and has received **39.50% of total number of funding’s** then **I would suggest my friend the location of NCR for starting his startup.**

***\*\*\*The all mentioned data , graphs and plots are done through the analysis of the file ‘’startup\_funding” attach within this folder . You can check the code documentation in the code folder associated with this pdf as 1 Question Code***.

Question 2:

Even after trying for so many times, your friend’s startup could not find the investment. So you decided to take this matter in your hand and try to find the list of investors who probably can invest in your friend’s startup. Your list will increase the chance of your friend startup getting some initial investment by contacting these investors. Find the top 5 investors who have invested maximum number of times (consider repeat investments in one company also). In a startup, multiple investors might have invested. So consider each investor for that startup. Ignore undisclosed investors.

**Answer:**

Afterimporting various library like **pandas**, **matplotlib.pyplot** , **numpy** and choosing the column “InvestorsName” from the dataframe I have replaced the wrong spelling of “Undisclosed investors” to “Undisclosed Investors” .

Using **dropna** method I have dropped the rows with “**NaN**” values present and keeping the index same.

Create one **dictionary** **investor\_dict** for investor wise number on funding’s.

**separateInvestor function:**

* However in the dataframe for **InvestorsName** at some places multiple investors name is given and the investors are separated by “,”. So we have to **split** the names by “,”. After **splitting(split())** it becomes a list and after looking at each data in list and **stripping(strip())** it I check if investor name

Is not “” and “Undisclosed Investors”.

* **If not “” and “Undisclosed Investors”** then it is appended to name\_li.
* Finally data is entered in **investor\_dict** using **location\_list** .

Using the **apply** **method** I have implemented the **separateInvestor** function to each row of the entire column of InvestorsName.

After **apply method** I have used the **Counter** **method** which work same as a dictionary and counts values for keys for **investor\_dict**.

Now using the most\_common method a list of tuples of highest\_investor is created.

Following data investor wise is retrieved from the dataframe.

Table 3 Investor wise Number of Funding's Top 5

|  |  |
| --- | --- |
| **Investor Name** | **Number of Funding’s** |
| 1. Sequoia Capital | 64 |
| 1. Accel Partners | 53 |
| 1. Kalaari Capital | 44 |
| 1. SAIF Partners | 41 |
| 1. Indian Angel Network | 40 |

**We can visualize the data in a bar graph or pie chart using matplotlib.pyplot.**

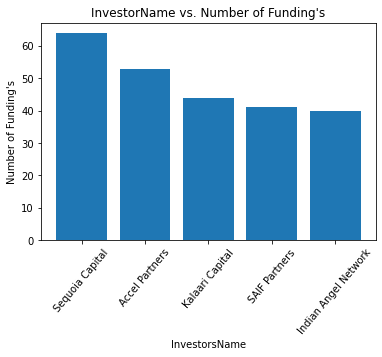
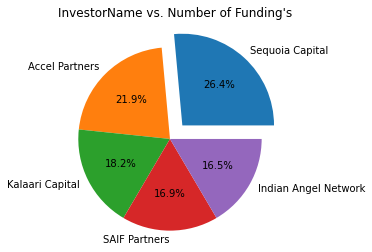
****

Figure 5: Bar Graph Investor vs. Number of Funding's Figure 6: Pie Chart Investor Name vs. Number of funding's

According to the data If **we take Investor-wise data into consideration** then **Sequoia Capital has made highest number of funding’s 64** which is **26.4% of total number of funding’s of TOP 5 Investors ignoring Undisclosed Investors.**

However if we talk about the top 10 Investors then the top 5 investors have done a significant amount of funding’s.

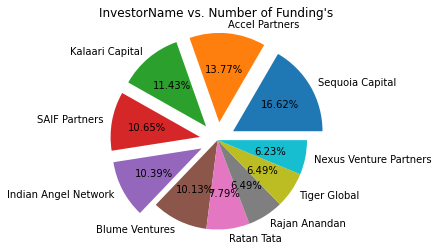


Figure 7: Top 10 Investors with their share in Number of funding's

From the pie chart we can also see that the top 5 investors share 62.86% share among the top 10 investors in number of investments invested(ignoring Undisclosed Investors) where **Sequoia capital share the maximum with 16.62% and Indian Angel Network be the least with 10.39% among the top 5.**

***\*\*\*The all mentioned data , graphs and plots are done through the analysis of the file ‘’startup\_funding” attach within this folder . You can check the code documentation in the code folder associated with this pdf as 2 Question Code***.

**Question 3:**

After re-analysing the dataset you found out that some investors have invested in the same startup at different number of funding rounds. So before finalising the previous list, you want to improvise it by finding the top 5 investors who have invested in different number of startups. This list will be more helpful than your previous list in finding the investment for your friend startup. Find the top 5 investors who have invested maximum number of times in different companies. That means, if one investor has invested multiple times in one startup, count one for that company. There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

**Answer:**

In the previous list we count the number of investments by investors in same startup multiple times but now we will count them only once as the reason that investor invested multiple times in a startup is because they get profit but since our startup is just started so it needs investment first time. So we will count the number of investments made by investors in different startups

Afterimporting various library like **pandas**, **matplotlib.pyplot** , **numpy** and choosing the columns “**StartupName**” and “**InvestorsName**” from the dataframe I have replaced the wrong spelling of **“Ola”, “Flipkart”, “Oyo” and “Paytm”** in “StartupName” column and “Undisclosed Investors” in “InvestorsName” column.

Using **dropna** method I have dropped the rows with “**NaN**” values present and keeping the index same.

Create one **dictionary** **investor\_dict**  for investor wise number on funding’s.

* Now using the value\_counts function I have counted the values of different startups and have a list of StartupName using names present at the index.
* As now we have to consider multiple investments by same investor in a company as one we have to use the StartupName to get to the each investor of that Startup.
* Finding the index where StartupName matches and seeimg the investors at that location using **.loc[index]** we get the all the investors name.
* Now one by one the names are edited using split() and strip() and checked for not “” and not “Undisclosed Investor” and then are append in a list name\_li.
* Now to get all the **unique names** in the name\_li we make it a **numpy array** and apply **the unique function to the numpy array** and hence we get a numpy array of all **unique investor who invested in the startup**.
* **Now make the dictionary investor\_dict using the unique numpy array.**

Use **Counter** **method** which work same as a dictionary and counts values for keys for **investor\_dict**.

Now using the most\_common method a list of tuples of highest\_investor is created.

Following data investor wise is retrieved from the dataframe.

Table 4: Top 5 Investor wise Number of Funding's In Different Startup.

|  |  |
| --- | --- |
| **Investor Name** | **Number of Funding’s** |
| 1. Sequoia Capital | 48 |
| 1. Accel Partners | 47 |
| 1. Kalaari Capital | 41 |
| 1. Indian Angel Network | 40 |
| 1. Blume Ventures | 36 |

**We can visualize the data in a bar graph or pie chart using matplotlib.pyplot.**

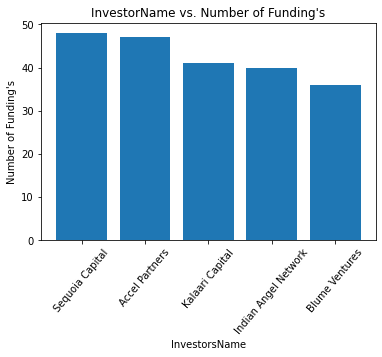
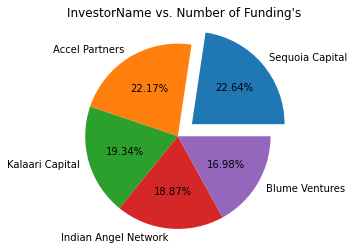


Figure 8: Pie Chart Investor Name vs. Number of funding's Figure 9: Bar Graph Investor vs. Number of Funding's

The bar graph shows that there is not much difference in number of investments made by the top 2-Sequoia Capital and Accel Partners but after that there is difference.

If we look at the percentage share of top 5 among themselves in the pie chart then we will find that all share a considerable amount. Sequoia Capital has 22.64% and Accel Partners has 22.17% of the total funding’s in top 5 investors.

However if we talk about the top 10 Investors then the top 5 investors have done a significant amount of funding’s.

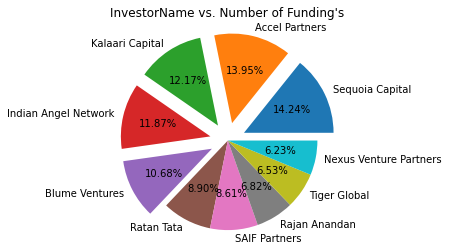


Figure 10: Top 10 Investors with their share in Number of funding's

From the pie chart we can also see that the top 5 investors share 62.91% share among the top 10 investors in number of investments invested(ignoring Undisclosed Investors) where **Sequoia capital share the maximum with 14.24% and Blume Ventures be the least with 10.68% among the top 5.**

***\*\*\*The all mentioned data , graphs and plots are done through the analysis of the file ‘’startup\_funding” attach within this folder . You can check the code documentation in the code folder associated with this pdf as 3 Question Code***.

**Question 4:**

Even after putting so much effort in finding the probable investors, it didn't turn out to be helpful for your friend. So you went to your investor friend to understand the situation better and your investor friend explained to you about the different Investment Types and their features. This new information will be helpful in finding the right investor. Since your friend startup is at an early stage startup, the best-suited investment type would be - Seed Funding and Crowdfunding. Find the top 5 investors who have invested in a different number of startups and their investment type is Crowdfunding or Seed Funding. Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column. There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

**Answer:**

In the above list we find the top 5 investors who invested maximum number of times in different startups considering all types of investment . But our friend startup is at an early stage , so it will need only seed or crowd funding .

Afterimporting various library like **pandas**, **matplotlib.pyplot** , **numpy** and choosing the 3 columns “**StartupName**” ,“**InvestorsName**” and “**InvestmentType**” from the dataframe I have replaced the wrong spelling of **“Ola”, “Flipkart”, “Oyo” and “Paytm”** in “StartupName” column and “**Undisclosed Investors**” in “InvestorsName” column and “**Seed Funding**”, “**Private Equity**” and “**Crowd Funding**” in “InvestmentType” column.

Using **dropna** method I have dropped the rows with “**NaN**” values present and keeping the index same.

**I drop the index in dataframe where** **dataframe[“InvestmentType”] is equal to “Private Equity” or “Debt Funding” to only focus on our dataframe for “Seed Funding” and “Crowd Funding”.**

Create one **dictionary** **investor\_dict**  for investor wise number on funding’s.

* Now using the value\_counts function I have counted the values of different startups and have a list of StartupName using names present at the index.
* As now we have to consider multiple investments by same investor in a company as one we have to use the StartupName to get to the each investor of that Startup.
* Finding the index where StartupName matches and seeimg the investors at that location using **.loc[index]** we get the all the investors name.
* Now one by one the names are edited using split() and strip() and checked for not “” and not “Undisclosed Investor” and then are append in a list name\_li.
* Now to get all the **unique names** in the name\_li we make it a **numpy array** and apply **the unique function to the numpy array** and hence we get a numpy array of all **unique investor who invested in the startup**.
* **Now make the dictionary investor\_dict using the unique numpy array.**

Use **Counter** **method** which work same as a dictionary and counts values for keys for **investor\_dict**.

Now using the most\_common method a list of tuples of highest\_investor is created.

Following data investor wise is retrieved from the dataframe.

Table 5: Top 5 Investor wise Number of Funding's in terms of Seed Funding and Crowd Funding

|  |  |
| --- | --- |
| **Investor Name** | **Number of Funding’s** |
| 1. Indian Angel Network | 33 |
| 1. Ranjan Anandan | 23 |
| 1. LetsVenture | 16 |
| 1. Anupam Mittal | 16 |
| 1. Group of Angel Investors | 14 |

**We can visualize the data in a bar graph or pie chart using matplotlib.pyplot.**

***\*\*\*The all mentioned data , graphs and plots are done through the analysis of the file ‘’startup\_funding” attach within this folder . You can check the code documentation in the code folder associated with this pdf as 4 Question Code***

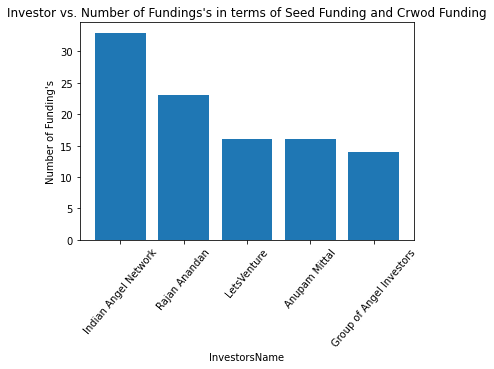


Figure 11: Bar Graph Investor vs. Number of Funding's for Seed Funding and Crowd Funding

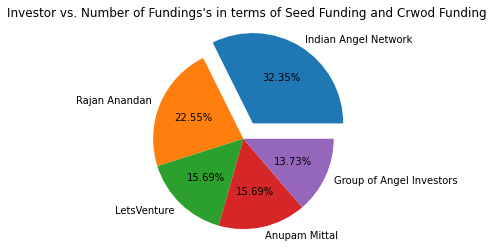


Figure 12: Pie Chart Investor vs. Number of Funding's for Seed Funding and Crowd Funding

It is clear from the bar graph also that **Indian Angel Network** do the most number of investments(**33**) in different startups of type seed funding and crowd funding. **Anupam Mittal and LetsVenture** invested same number of times(**16**).

One more thing to be noted is that **crowd funding occurs only two times** which are done by Indian Angel Network and Rajan Anandan. The majority is done of type Seed Funding.

**Question 5:**

Due to your immense help, your friend startup successfully got seed funding and it is on the operational mode. Now your friend wants to expand his startup and he is looking for new investors for his startup. Now you again come as a saviour to help your friend and want to create a list of probable new new investors. Before moving forward you remember your investor friend advice that finding the investors by analysing the investment type. Since your friend startup is not in early phase it is in growth stage so the best-suited investment type is Private Equity. Find the top 5 investors who have invested in a different number of startups and their investment type is Private Equity. Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column. There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

**Answer:**

Now our friend startup is successfully established and now we want to expand it. Our friend startup is in growth stage so the best suited investment type is Private Equity. So we will filter the list obtained in third question to the investors list who invested maximum number of times in different startups of investment type Private Equity.

Afterimporting various library like **pandas**, **matplotlib.pyplot** , **numpy** and choosing the 3 columns “**StartupName**” ,“**InvestorsName**” and “**InvestmentType**” from the dataframe I have replaced the wrong spelling of **“Ola”, “Flipkart”, “Oyo” and “Paytm”** in “StartupName” column and “**Undisclosed Investors**” in “InvestorsName” column and “**Seed Funding**”, “**Private Equity**” and “**Crowd Funding**” in “InvestmentType” column.

Using **dropna** method I have dropped the rows with “**NaN**” values present and keeping the index same.

**I drop the index in dataframe where** **dataframe[“InvestmentType”] is equal to “Seed Funding” or “Debt Funding” and “Crowd Funding” to only focus on our dataframe for “Private Equity” .**

Create one **dictionary** **investor\_dict**  for investor wise number on funding’s.

* Now using the value\_counts function I have counted the values of different startups and have a list of StartupName using names present at the index.
* As now we have to consider multiple investments by same investor in a company as one we have to use the StartupName to get to the each investor of that Startup.
* Finding the index where StartupName matches and seeing the investors at that location using **.loc[index]** we get the all the investors name.
* Now one by one the names are edited using split() and strip() and checked for not “” and not “Undisclosed Investor” and then are append in a list name\_li.
* Now to get all the **unique names** in the name\_li we make it a **numpy array** and apply **the unique function to the numpy array** and hence we get a numpy array of all **unique investor who invested in the startup**.
* **Now make the dictionary investor\_dict using the unique numpy array.**

Use **Counter** **method** which work same as a dictionary and counts values for keys for **investor\_dict**.

Now using the most\_common method a list of tuples of highest\_investor is created.

Following data investor wise is retrieved from the dataframe.

Table 6:Top 5 Investor wise Number of Funding's where Investment Type is Private Equity

|  |  |
| --- | --- |
| **Investor Name** | **Number of Funding’s** |
| 1. Sequoia Capital | 45 |
| 1. Accel Partners | 43 |
| 1. Kalaari Capital | 35 |
| 1. Blume Ventures | 27 |
| 1. SAIF Partners | 24 |

**We can visualize the data in a bar graph or pie chart using matplotlib.pyplot.**

***\*\*\*The all mentioned data , graphs and plots are done through the analysis of the file ‘’startup\_funding” attach within this folder . You can check the code documentation in the code folder associated with this pdf as 5 Question Code***

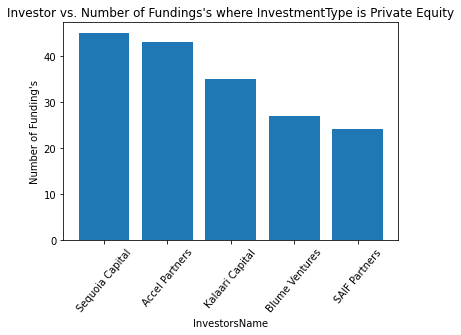
****

Figure 13: Bar Graph Investor vs. Number of Funding' in terms of Private Equity

It is clear from the bar graph also that **Sequoia Capital** do the most number of investments(**45**) in different startups of InvestmentType Private Equity.

In total Seed Funding is done **1301** times, Private Equity is done **1067** times Crowd Funding is done **2** times and Debt Funding is done **1** time.

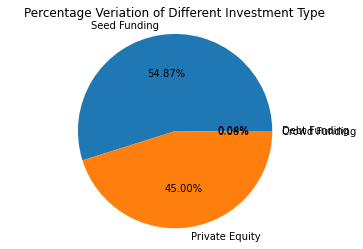


Figure 14: Investment Type Distribution