

PROJECT CASE STUDY PART – 2

CASE STUDY 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.

SOLUTION

First thing what I will do is to find the top five cities where startups have got fundings maximum number of times.

So, first thing is I uploaded the csv file 'startup_funding.csv' to the notebook and then I tried to read the file from the notebook using python as our programming interface for all analysis functions in this case study in a pandas dataframe which here is 'df'.

Then I created a function which will check for the values for the locations of 'New delhi', 'Mumbai', 'Gurgaon', 'Bangalore' and 'Noida'. Also, if there are any values which are in this format "city1/city2" then the function splits the cities with respect to '/'. Now after all if these functions are done then function check if these values are these above cities. If these values are in these cities then these values will be added to the dictionary where these five cities will be keys and the count of number of fundings will be the values for each of these keys.

Now, we will take both of the keys of dictionary and values of dictionary in seprate numpy arrays. Then, we ran a loop and get values for the top five cities and count of fundings these cities got in descending order.

We will get result like this:-

List of locations where most number of times fundings for startups have been observed:

Bangalore - 637

Mumbai - 449

New Delhi - 389

Gurgaon - 241

Noida - 79

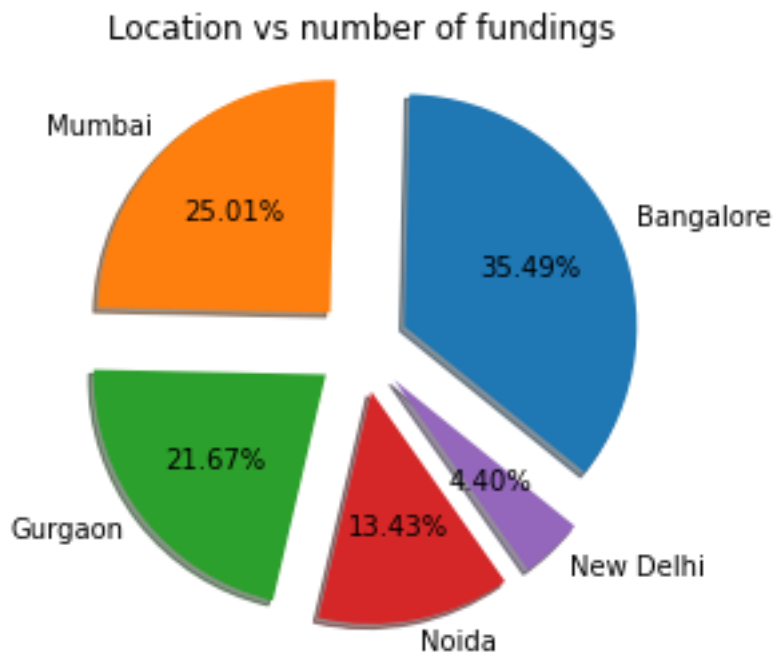
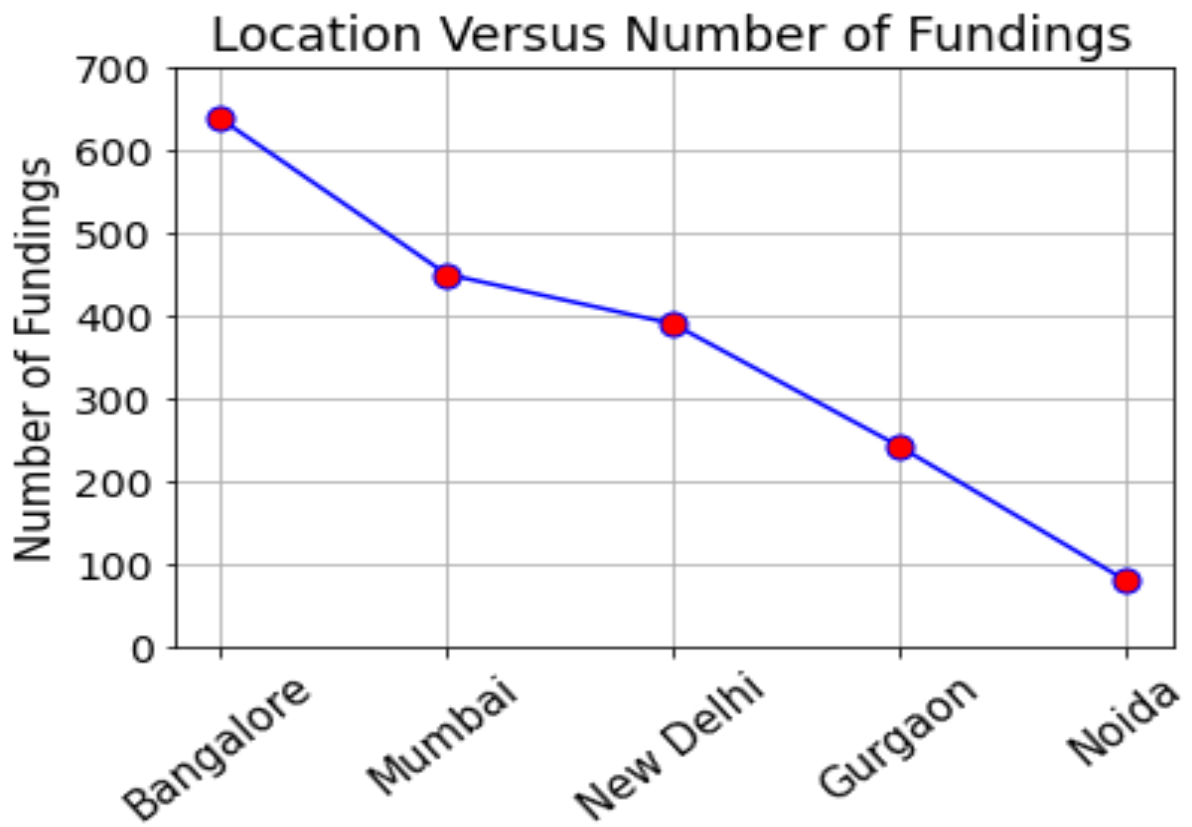
Maximum Funding is done in Bangalore , 637 times.

Here, we printed out the top five cities and number of fundings happened there. Here, we can see that as of our data the maximum number of fundings have been done in bangalore.

To make things more clear and visible and make the analysis easy to understand we created graphs with the help of matplotlib library of python. One is Line graph which has been used to get a line for the top five cities with respect to the number of fundings happened there.

Then, we created a pie chart to show the distribution of percentage of fundings done in these top five cities.

These two charts are shown as below:-



So, now we got the list of all top five cities where maximum fundings were happened and these two graphs It is clear that bangalore has got the maximum number of fundings with '637' and percentage wise '35.49' as compared to other top cities.

Hence, I should suggest my friend to establish his startup in the bangalore city.

CASE STUDY 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.

SOLUTION

Now, we have to find the list of top five investors who have invested maximum number of times in different startups and also more than once in a single startup.

Since we have already read the file through python 'read_csv' function the help of pandas library of python and save it in a pandas dataframe 'df'. now we will concentrate on the solution for this problem.

So, first we will look at the 'InvestorName' column in the 'df' file, If we see there we can see that there are more than one investors for a single startup at a single row.

So, if we want to get the names of all investors individually we will create a function which will split the values in 'InvestorName' column for each row and then we will add those values in the dictionary where the investor name will be the key and the number of times that investor invested in different companies or in the

same company more than once will be stored as the value for that key.

Now we will use apply function for the column 'InvestorName' and will apply the function we created to create a dictionary for containing investor and number of fundings as key – value pair.

Now, as in the last case study we will again make separate numpy arrays for the key and value pairs of this newly created dictionary.

Then, we use 'argsort[::-1]' function to get the indexes in descending order as 'ind' and also, we sliced the 'ind' to ind[0 : 5] to get the top five values from the dictionary.

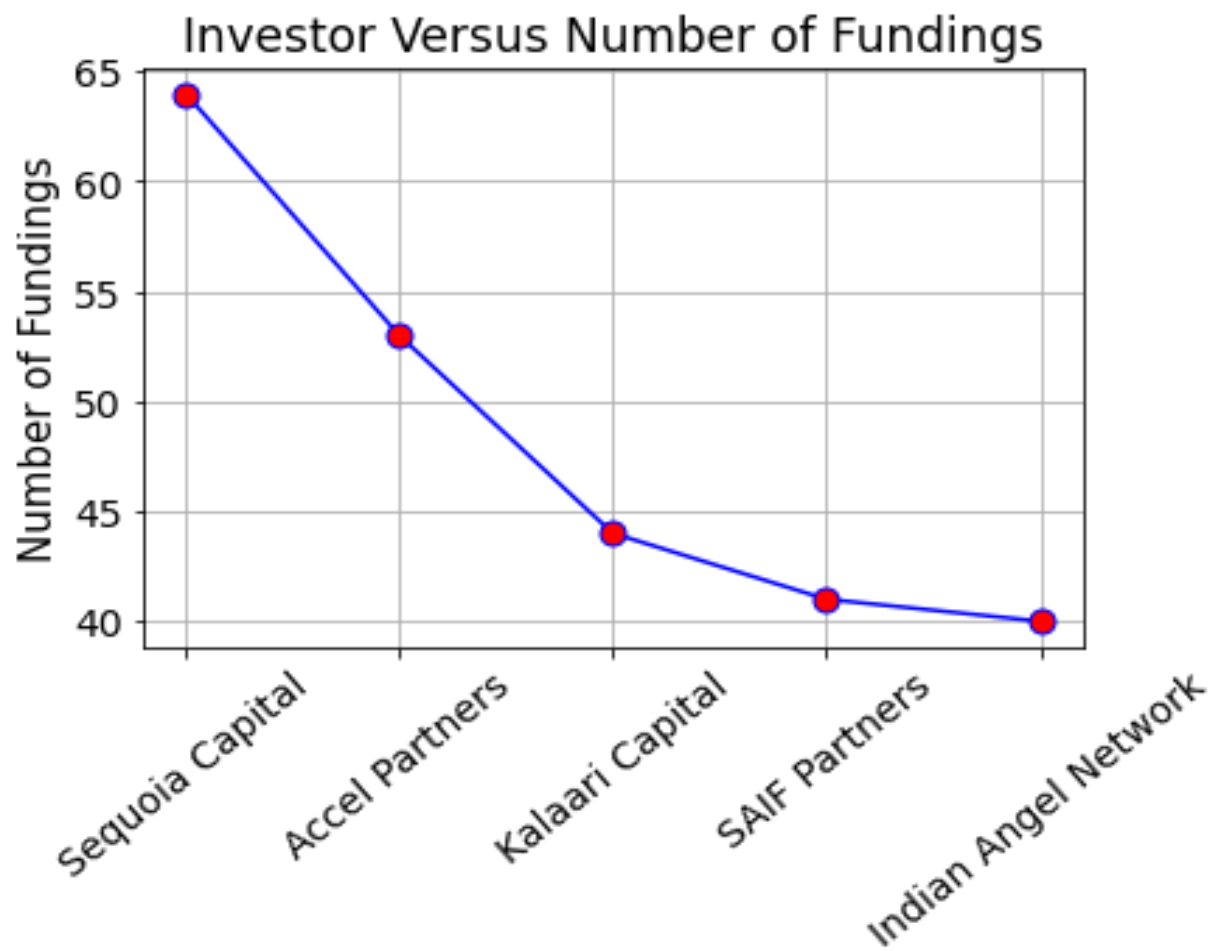
Now, we gave this indexing to our keys and values pairs and ran a loop to get top five investors and number of times they have funded.

We will get this result :-

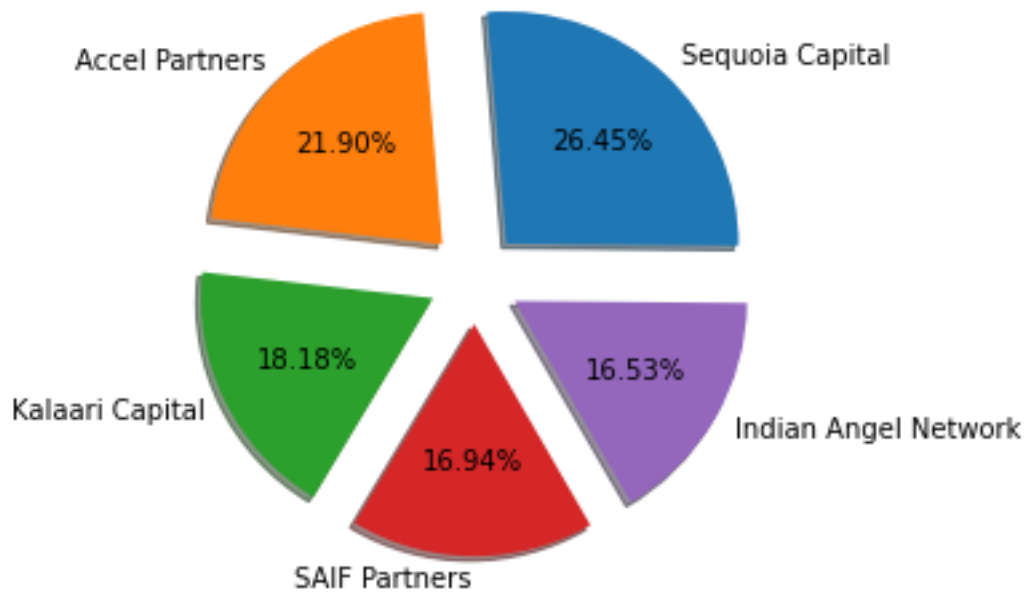
```
Top 5 Investors are:  
Sequoia Capital - 64  
Accel Partners - 53  
Kalaari Capital - 44  
SAIF Partners - 41  
Indian Angel Network - 40
```

So, now we got top five investors at last we will get two graphs for this data. one is line graph and another is pie chart.

These two charts are shown below:-



Investor vs number of fundings for top five investors in percentage



With reference to the list 'Top 5 investors are' and these two graphs it is clear that 'Sequoia Capital' has funded maximum number of times with count as '63' and the percentage as '26.45' largest of all compared to other top five investors whether they are same or different. Also, we can see other top investors in both of these charts.

Hence, I should suggest my friend to contact these top five investors for his funding needs.

CASE STUDY 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.

SOLUTION

For this question if you have understand the question properly, here it says that we have to get top five investors but for different startups. This means we have to get the count of of a startup just once for that particular investor and then after this we will get the top five investors and number of times they funded.

Since, we have already created a pandas dataframe for the csv file 'startup_funding.csv', We will now use two functions 'dropna' to remove all values 'replace' to remove all duplicate values from our dataframe.

Here , we are only just removing duplicate values of important startups like 'Oyo' , 'Ola' , 'Flipkart' , 'Paytm' since we can not remove all duplicate values for all startups.

Now, we will get two lists one for the investors from the 'InvestorsName' column and one for startup name from 'StartupName' column.

Then, we ran a loop to store values in a dictionary as key – value pair where key will be investor and the value will be the count of all individual startups that was funded by that investor. Also here we use a set 's' so that if any startup's name come again then that startup will be counted only once inside that set 's' as we all sets take unique values.

So now we got all unique values for different startups, finally we will save these values as key – value pair where investors is keys and count of startups funded by that investor will be the values

Now, we ran a new loop where we iterate the dictionary with the value 'key' where this 'key' will be the investor's name and then for each key we will take value as the count of all startups that key contains in dictionary 'd'.

At last, we have got keys as investors and values as number of startups a key contains, Now we will make separate lists of keys and value pairs in that dictionary and then use 'argsort[::-1]' to get indexes as 'ind' in descending order and then sliced 'ind' as 'ind[0 :

5]' to get top five values and then we ran a loop to get all top five values in descending order.

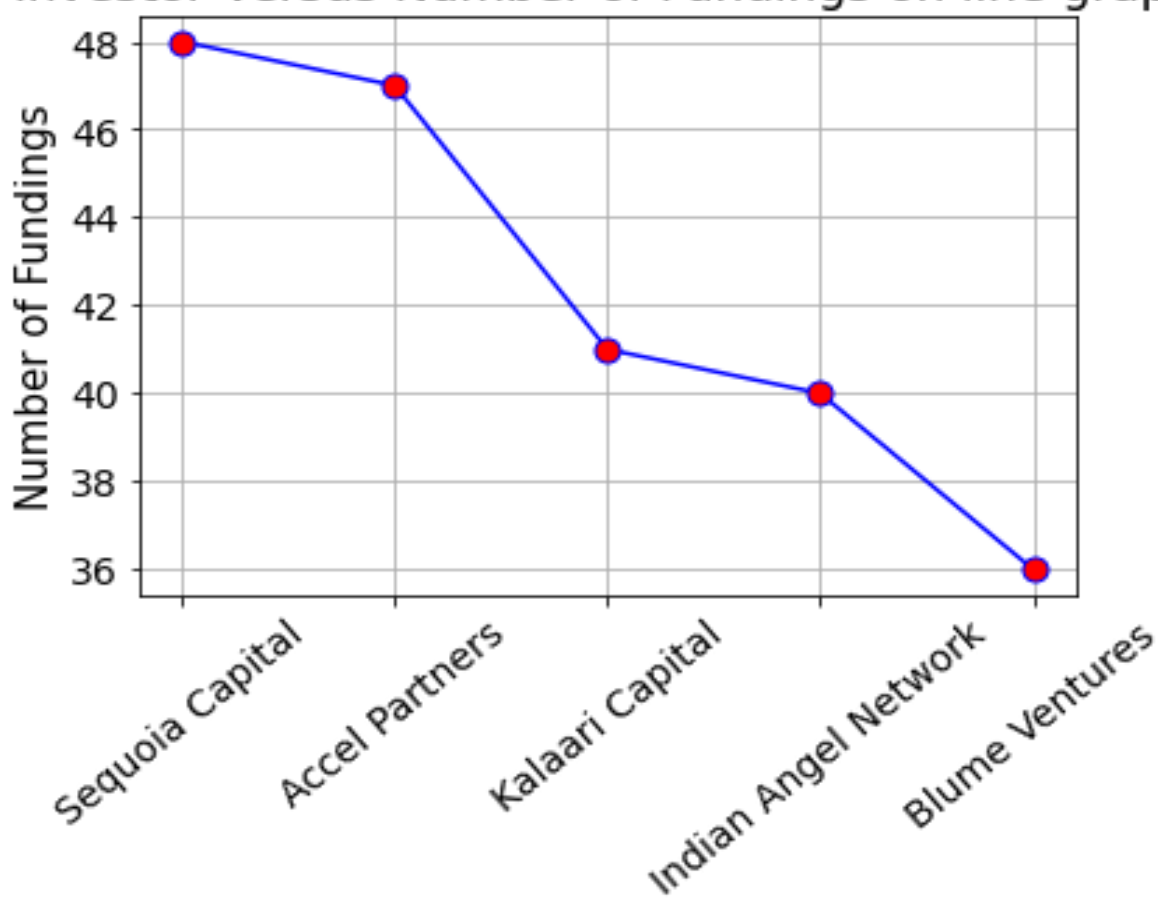
We will get results like this:-

Top 5 Investors are:
Sequoia Capital - 48
Accel Partners - 47
Kalaari Capital - 41
Indian Angel Network - 40
Blume Ventures - 36

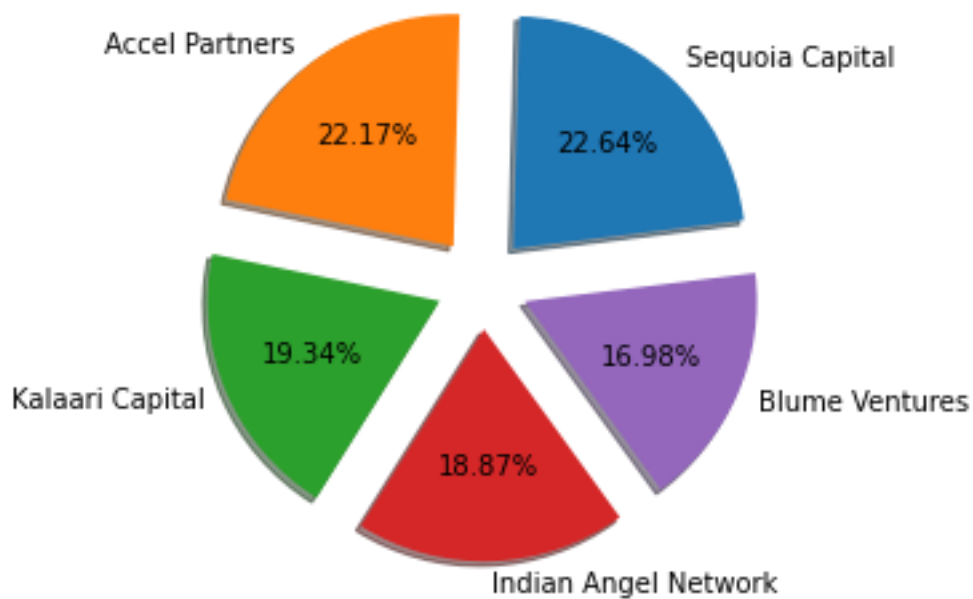
So, now we got top five investors at last we will get two graphs for this data. one is line graph and another is pie chart.

These two charts are shown below:-

Investor Versus Number of Fundings on line graph



Investor vs number of fundings for top five investors in percentage



Finally, we got top five investors who have invested maximum number of times in each startup but only once in a startup from the above results we can conclude that 'Sequoia Capital' is still the largest investor invested 48 times and also largest by percentage wise share of fundings done which here is '22.64%'.

CASE STUDY 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.

SOLUTION

Since, in previous case study we found the top five investors who have invested in a different number of startups taking count of every startup only once funded by that investor. Now, in addition to this we now want to get the investors whom investment type is 'Seed Funding' and 'Crowd Funding'.

Since, we have already created a pandas dataframe for the csv file 'startup_funding.csv', We will now use two functions 'dropna' to remove all values 'replace' to remove all duplicate values from our dataframe.

Here , we are only just removing duplicate values of important startups like 'Oyo' , 'Ola' , 'Flipkart' , 'Paytm' since we can not remove all duplicate values for all startups.

Now, we will provide a condition for the column 'Investment Type' in 'df' where this column will be equal to the 'Seed Funding' and 'Crowd Funding' and then save it again to 'df'.

Now, we will get two lists one for the investors from the 'InvestorsName' column and one for startup name from 'StartupName' column.

Then, we ran a loop to store values in a dictionary as key – value pair where key will be investor and the value will be the count of all individual startups that was funded by that investor. Also here we use a set 's' so that if any startup's name come again then that startup will be counted only once inside that set 's' as we all sets take unique values.

So now we got all unique values for different startups, finally we will save these values as key – value pair where investors is keys and count of startups funded by that investor will be the values

Now, we ran a new loop where we iterate the dictionary with the value 'key' where this 'key' will be the investor's name and then for each key we will take value as the count of all startups that key contains in dictionary 'd'.

At last, we have got keys as investors and values as number of startups a key contains, Now we will make separate lists of keys and value pairs in that dictionary and then use 'argsort[::-1]' to get indexes as 'ind' in descending order and then sliced 'ind' as 'ind[0 : 5]' to get top five values and then we ran a loop to get all top five values in descending order.

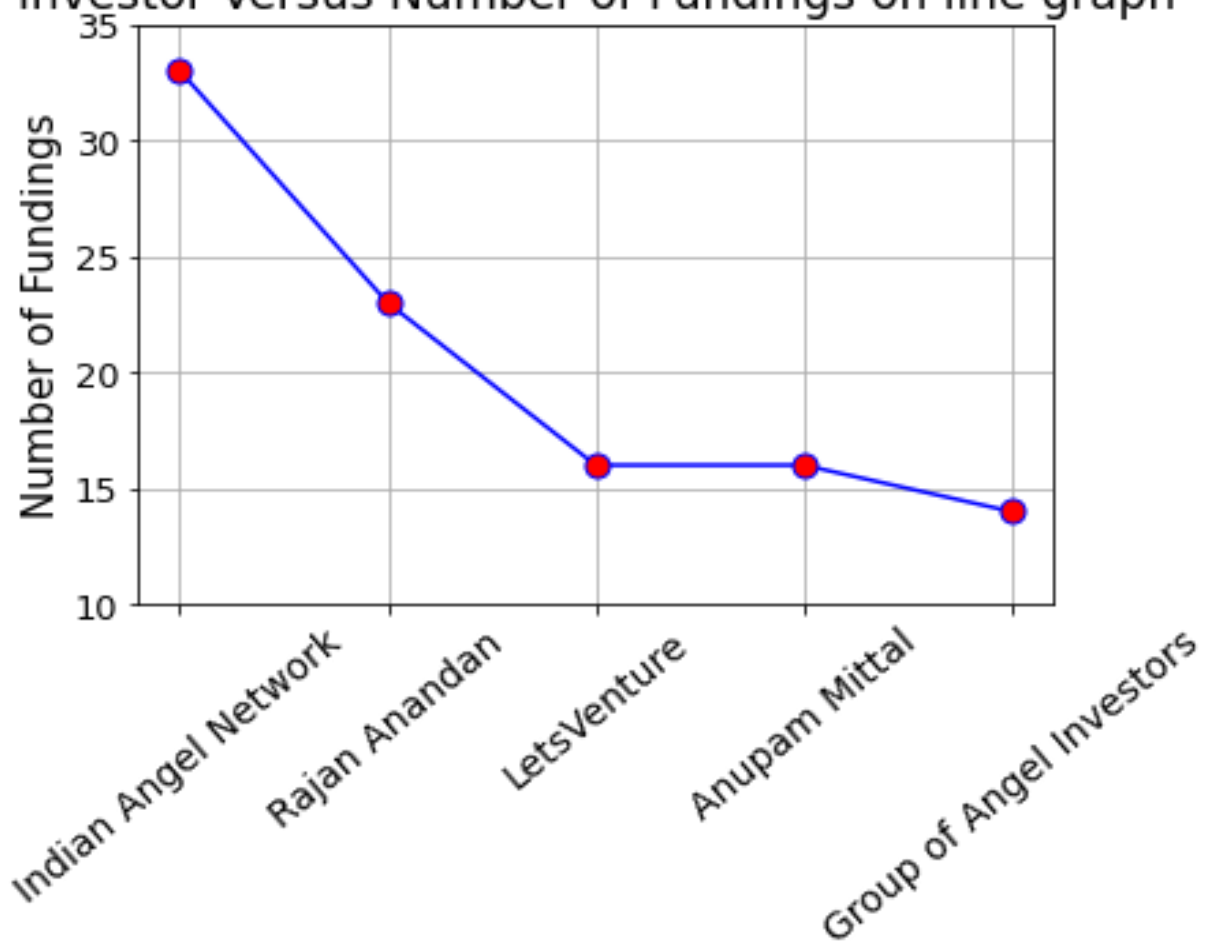
We will get results like this:-

```
Top 5 Investors for Investment type- Seed
Funding and Crowd Funding are:
Indian Angel Network - 33
Rajan Anandan - 23
LetsVenture - 16
Anupam Mittal - 16
Group of Angel Investors - 14
```

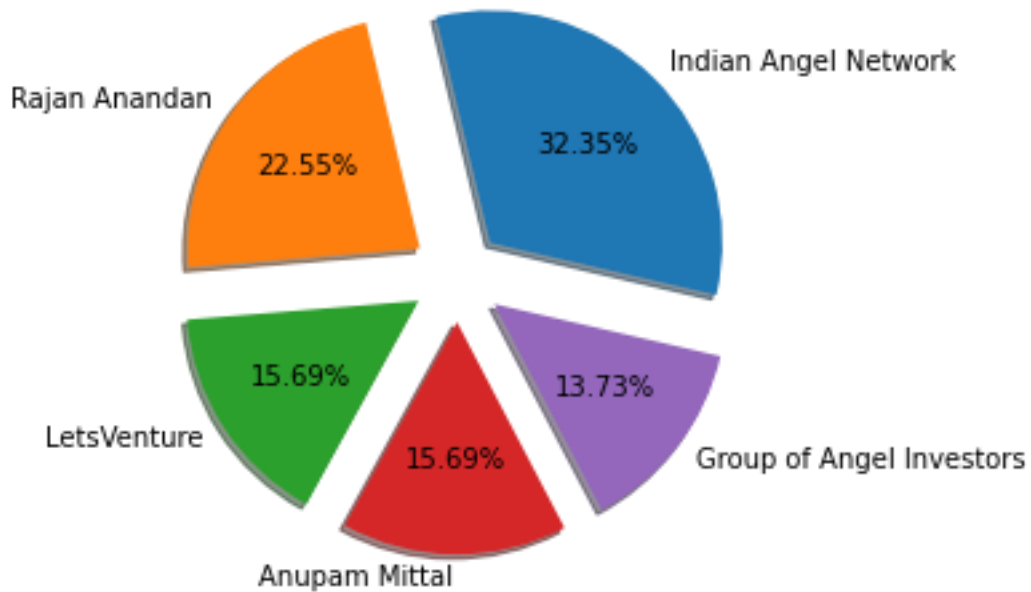
So, now we got top five investors whom investment type is 'Seed Funding' and 'Crowd Funding'. At last we will get two graphs for this data. one is line graph and another is pie chart

These two charts are shown below:-

Investor Versus Number of Fundings on line graph



Investor vs number of fundings for top five investors in percentage



Finally, we got top five investors who have invested maximum number of times in each startup but only once in a startup where investment type is 'Seed Funding' and 'Crowd Funding' and from the above results we can conclude that as for above condition 'Indian angel network' is the largest investor invested 48 times and also largest by percentage wise share of fundings done which here is '32.35%'.

CASE STUDY 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.

SOLUTION

Now, my friend's startup has got funding through investment type of 'Seed Funding' and now it is on operational mode so now we want to get another Funding for my friend which is 'Private Equity' because my friend's startup is in the growth stage.

Hence, now we want to find top five investors whom invested in different number of startups with each startup just once and here the investment type is 'Private Equity'.

So, the process is just similar to the case study as done before this one,

Since, we have already created a pandas dataframe for the csv file 'startup_funding.csv', We will now use two functions 'dropna' to remove all values 'replace' to remove all duplicate values from our dataframe.

Here , we are only just removing duplicate values of important startups like 'Oyo' , 'Ola' , 'Flipkart' , 'Paytm' since we can not remove all duplicate values for all startups.

Now, we will provide a condition for the column 'Investment Type' in 'df' where this column will be equal to the 'Seed Funding' and 'Crowd Funding' and then save it again to 'df'.

Now, we will get two lists one for the investors from the 'InvestorsName' column and one for startup name from 'StartupName' column.

Then, we ran a loop to store values in a dictionary as key – value pair where key will be investor and the value will be the count of all individual startups that was funded by that investor. Also here we use a set 's' so that if any startup's name come again then that startup will be counted only once inside that set 's' as we all sets take unique values.

So now we got all unique values for different startups, finally we will save these values as key – value pair where investors is keys and count of startups funded by that investor will be the values

Now, we ran a new loop where we iterate the dictionary with the value 'key' where this 'key' will be the investor's name and then for each key we will take value as the count of all startups that key contains in dictionary 'd'.

At last, we have got keys as investors and values as number of startups a key contains, Now we will make separate lists of keys and value pairs in that dictionary and then use 'argsort[::-1]' to get indexes as 'ind' in descending order and then sliced 'ind' as 'ind[0 : 5]' to get top five values and then we ran a loop to get all top five values in descending order.

We will get results like this:-

Top 5 Investors for Investment type- Private Equity are:

Sequoia Capital - 45

Accel Partners - 43

Kalaari Capital - 35

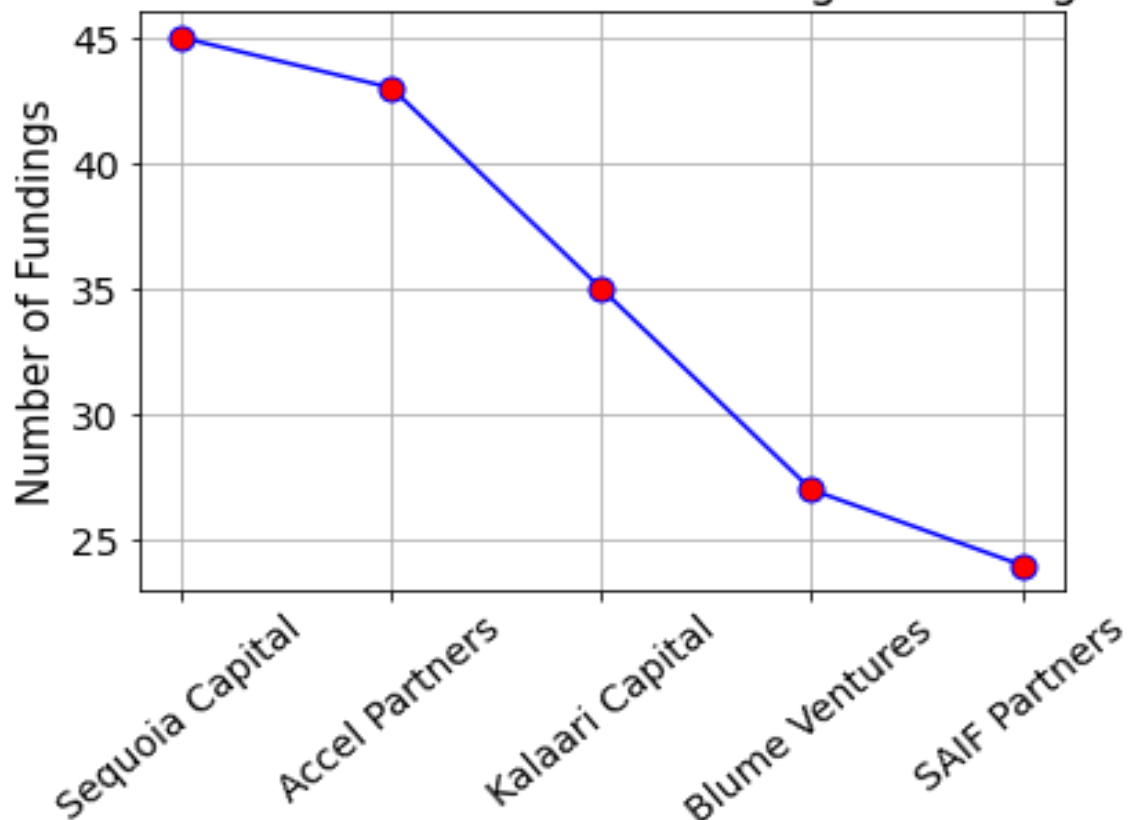
Blume Ventures - 27

SAIF Partners - 24

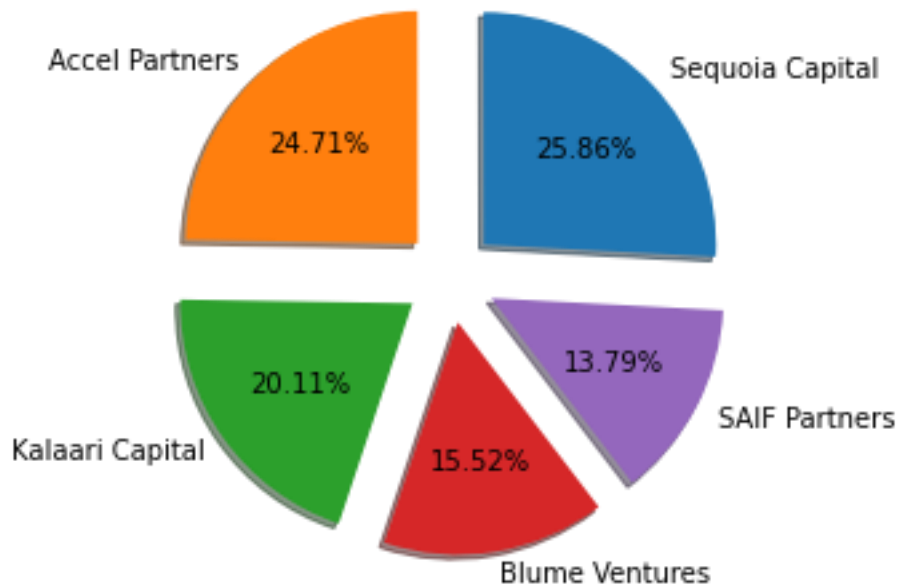
So, now we got top five investors whom investment type is 'Private Equity'. At last we will get two graphs for this data. one is line graph and another is pie chart

These two charts are shown below:-

Investor Versus Number of Fundings on line graph



Investor vs number of fundings for top five investors in percentage



Finally, we got top five investors who have invested maximum number of times in each startup but only once in a startup where investment type is 'Private Equity' and from the above results we can conclude that as for above condition 'Indian angel network' is the largest investor invested 48 times and also largest by percentage wise share of fundings done which here is '32.35%'.

