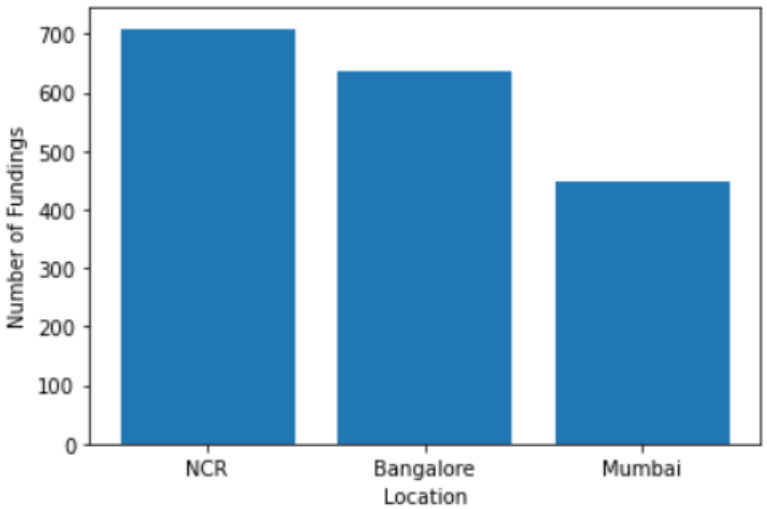
Case Study

Startup Funding

1. Your Friend has developed the Product and he wants to establish the product start-up 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 start-ups 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 start-ups multiple locations are given, one Indian and one Foreign. Consider the start-up if any one of the city lies in given locations.

Answer:




```
Location where most number of fundings are done: NCR
Number of fundings done: 709
```

Interpretation of the Result:

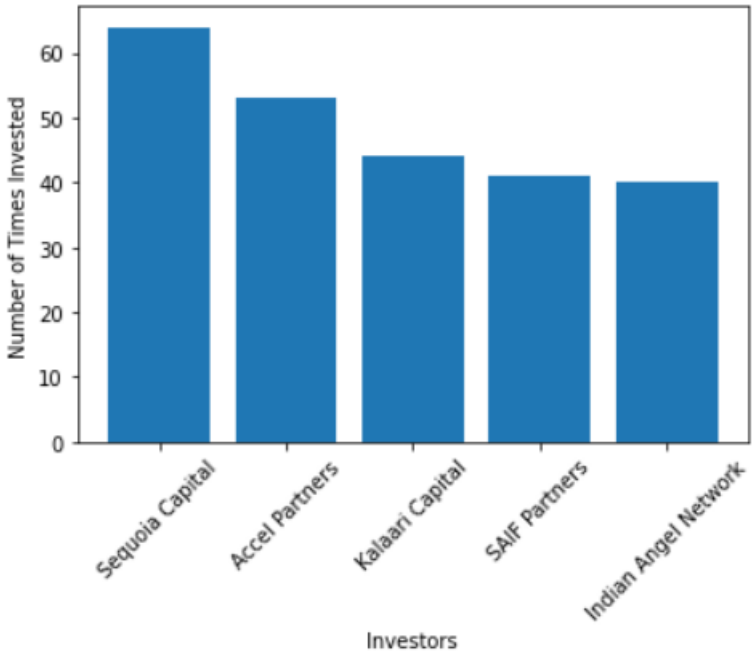
There is a high chance of getting the investment in the NCR region as the data after analysis shows that between 2015-2017, maximum number of fundings were done in NCR region (Gurgaon, Noida and New Delhi).

Justification:

1. First NAN values were dropped from CityLocation column as they would not contribute to the result. The pandas.series generated was assigned to the cities variable. Every entry of a city means that a start-up received funding in that city. Thus value counts of cities/locations of this column would give us our answer.
2. Then a SeparateCity() function was defined to separate cities from every entry of the CityLocation column, on a '/' and capture the city at 0^{th} index which would be an Indian city for every entry and then strip of any whitespace.
3. The above function was then applied to every entry of the series.
4. Data cleaning was then done according to the question. New Delhi, Gurgaon and Noida were then replaced by NCR as that was to be considered as per the question.
5. Value counts function was used which gave the count of the cities/locations in the column in decreasing order as needed. Top three locations were NCR, Bangalore and Mumbai.
6. Index and values were assigned to variables x and y accordingly and the bar graph was plotted.
7. The final answer was then printed by accessing the first entry of each of the variables x and y.

2. Even after trying for so many times, your friend's start-up 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 start-up. Your list will increase the chance of your friend start-up 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 start-up, multiple investors might have invested. So consider each investor for that start-up. Ignore undisclosed investors.

Answer:




```
Top investors in decreasing order of number of times invested are:
Sequoia Capital
Accel Partners
Kalaari Capital
SAIF Partners
Indian Angel Network
```

Interpretation of the Result:

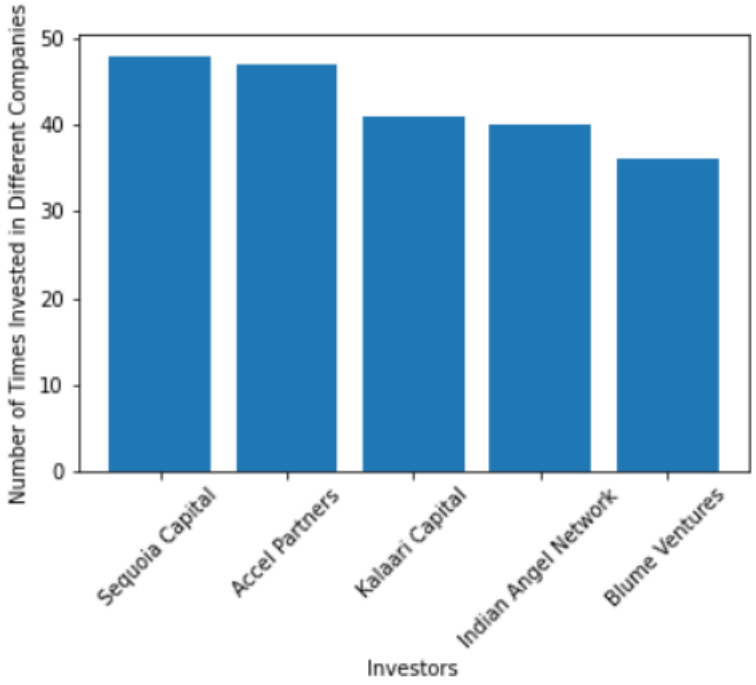
The above investors can be contacted for investment and there is a higher chance of getting an investment from them than just contacting randomly as the above investors have invested maximum number of times in start-ups from 2015-2017.

Justification:

1. First the InvestorsName column was used to create a new dataframe and this was assigned to a variable. NAN values were dropped as they would not contribute to the result. An investor name in the column means that the investor invested in a start-up. Thus value counts of investors would give us our answer.
2. A function SeparateInvestors_RepetitionsAllowed() was defined to count each investor in a company individually. The investor string is passed to the function and investors are split using ‘,’. The investors are then stripped off any whitespace and appended to a list and empty strings are skipped.
3. The above function is then applied to every entry of the InvestorsName column this would add lists of investors in place of strings in the InvestorsName column. To extract each and every investor from the list and add them as new rows below the current row explode() has been used. This would result in every entry now being of a single investor. As repeat investments in the same company were to be considered, investors were not grouped according to the companies.
4. The investors can then be counted using value counts. The top 5 investors in decreasing order were assigned to the investors variable and their counts were assigned to values variable accordingly.
5. Bar graph was plotted between investors and number of times invested.
6. Final answer was printed.

3. After re-analysing the dataset you found out that some investors have invested in the same start-up 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 start-ups. This list will be more helpful than your previous list in finding the investment for your friend start-up. 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 start-up, count one for that company. There are many errors in start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

Answer:




```
Top investors in decreasing order of number of times invested in differ
ent companies are:
Sequoia Capital
Accel Partners
Kalaari Capital
Indian Angel Network
Blume Ventures
```

Interpretation of the Result:

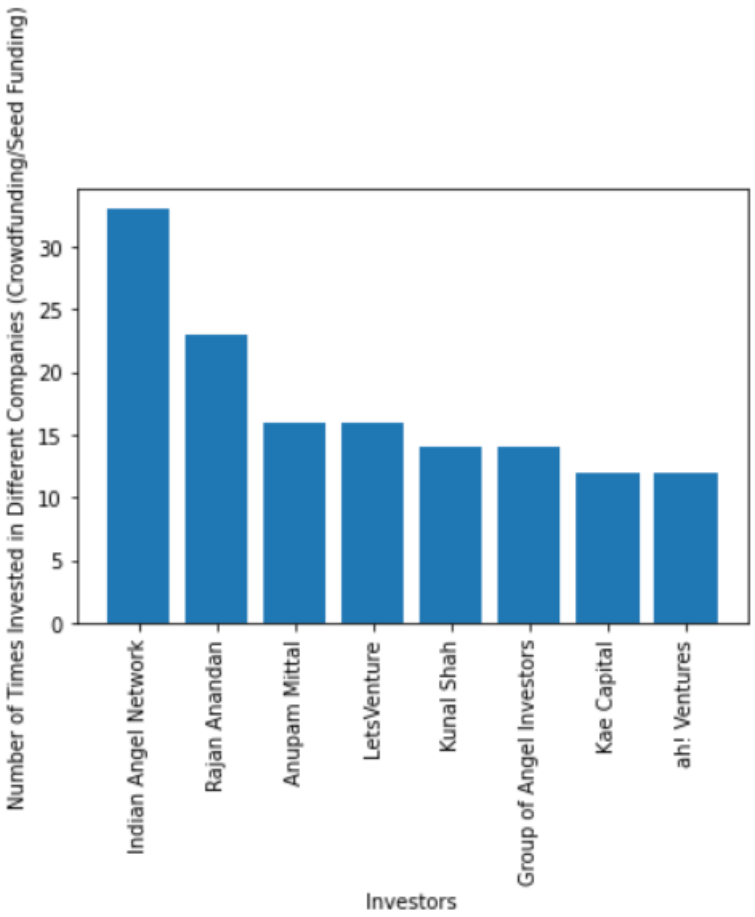
The above investors can be contacted for investment and there is a higher chance of getting an investment from them than just contacting randomly as the above investors have invested maximum number of times in different start-ups from 2015-2017. Investments by the same investor in the same start-up in different funding rounds was counted as one. This list is more exclusive and useful than the previous one as redundant investments were not included.

Justification:

1. A new DataFrame was created using columns StartupName and InvestorsName and assigned to a variable. NAN values were dropped as they would not contribute to the result. An investor name in the column means that the investor invested in a start-up. Thus value counts of investors who invested in different start-ups would give us our answer.
2. Data Cleaning needed to be done for start-up names. Thus all start-up names were viewed using value counts. A simple ctrl+f (find) search was done for wrong/misspelled start-up names. The wrong start-up names were then replaced with the correct ones.
3. Now as multiple investments by the same investor in a company were to be counted as one, I grouped the investors according to the company. I grouped by the StartupName and analogous to summing remaining numeric values group-wise, I joined the strings of investors group-wise by applying join to InvestorsName column. I then reset the index.
4. Then I defined a function SeparateInvestors_RepetitionsNotAllowed() that takes in the investors string. The function then declares a set and I split the investors by a ‘,’ and added non empty strings of individual investors after stripping whitespace to the set. Now if any investor invested multiple times he would appear multiple times in the investor string passed to the function. But as the investors are being added to a set, the set would contain only unique values and investors who invested multiple times in the same company would appear only once, thus they would consequently be counted only once.
5. I then applied this function to every entry of the InvestorsName column which would replace every entry with a set of unique investors which I then exploded as done in the previous question.
6. I then reset the index.
7. I then used the value_counts() function to assign the top five investors and their values (no. of times invested in different companies) to variables.
8. I the plotted a graph between Investors and Number of Times Invested in Different Companies.
9. I then printed the answer.

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 start-up is at an early stage start-up, the best-suited investment type would be - Seed Funding and Crowdfunding. Find the top 5 investors who have invested in a different number of start-ups 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 start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

Answer:



```
Top investors in decreasing order of number of times invested in differ
ent companies where investment type was Crowd Funding or Seed Funding a
re:
Indian Angel Network
Rajan Anandan
LetsVenture
Anupam Mittal
Kunal Shah
Group of Angel Investors
ah! Ventures
Kae Capital
```

Interpretation of the Result:

The above investors can be contacted for investment and there is a higher chance of getting an investment from them than just contacting the top 5 investors as the above investors have invested maximum number of times in different start-ups from 2015-2017 and their investment type was Crowdfunding or Seed Funding. This would be really helpful for my friend to find the right investor as his start-up is in an early stage and the above two funding types would be best suited. Investments by the same investor in the same start-up in different funding rounds was counted as one as required by the question. This list is more exclusive and useful than the previous one as redundant investments were not included and only specific useful investment types were considered.

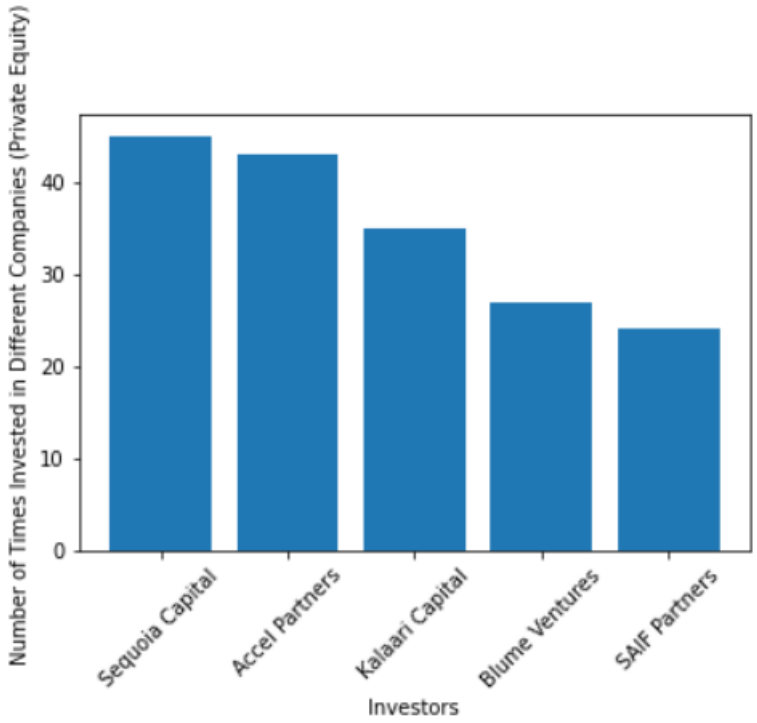
Justification:

1. Data Cleaning was needed for spelling errors in investment types. Thus the investment types were viewed using the value_counts() function and wrong spellings were replaced accordingly.
2. Now only entries where the InvestmentType was Crowdfunding or Seed Funding were considered. This was done by creating a Boolean array using the bitwise OR operator which ORed every entry of the two Boolean arrays generated. The data generated was then assigned to a variable data1.
3. A new DataFrame was created using columns StartupName and InvestorsName and assigned to a variable. NAN values were dropped as they would not contribute to the result. An investor name in the column means that the investor invested in a start-up. Thus value counts of investors who invested in different start-ups would give us our answer.
4. Data Cleaning needed to be done for start-up names. Thus all start-up names were viewed using value counts (done in previous question). A simple ctrl+f (find) search was done for wrong/misspelled start-up names. The wrong start-up names were then replaced with the correct ones.
5. Now as multiple investments by the same investor in a company were to be counted as one, I grouped the investors according to the company. I grouped by the StartupName and analogous to summing remaining numeric values group-wise, I joined the strings of investors group-wise by applying join to InvestorsName column. I then reset the index.

6. I then applied the SeparateInvestors_RepetitionsNotAllowed() function to every entry of the InvestorsName column which would replace every entry with a set of unique investors which I then exploded as done in the previous question.
7. I then reset the index.
8. I then picked up only those investor names where investor names were NOT undisclosed as undisclosed ones would not provide me any useful information. I did this using the bitwise and operation.
9. I then used the value_counts() function to view the top five investors. As Anupam Mittal and LetsVenture were tied for third position; Kunal Shah and Group of Angel Investors were tied for fourth position and Kae Capital and ah! Ventures were tied for fifth position I had to include all of them in the answer.
10. I then assigned the top five investors and their values (no. of times invested in different companies) to variables.
11. I the plotted a graph between Investors and Number of Times Invested in Different Companies (Crowdfunding/Seed Funding).
12. I then printed the answer.

5. Due to your immense help, your friend start-up successfully got seed funding and it is on the operational mode. Now your friend wants to expand his start-up and he is looking for new investors for his start-up. 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 start-up 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 start-ups 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 start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

Answer:




```
Top investors in decreasing order of number of times invested in differ
ent companies where investment type was Private Equity are:
Sequoia Capital
Accel Partners
Kalaari Capital
Blume Ventures
SAIF Partners
```

Interpretation of the Result:

As my friend's start-up successfully got seed funding and is now in the operational mode further expansion would require new investors. Now as start-up is in growth stage so the best-suited investment type is Private Equity. Thus the above list of top 5 investors can be contacted for investment and there is a higher chance of getting an investment from them than just contacting the top 5 investors from previous question as the above investors have invested maximum number of times in different start-ups from 2015-2017 and their investment type is Private Equity. This would be really helpful for my friend to find the right investor as his start-up is now in a growth stage. Investments by the same investor in the same start-up in different funding rounds was counted as one as required by the question. This list is more exclusive and useful than the previous one as redundant investments were not included and only specific useful investment types were considered.

Justification:

1. Data Cleaning for spelling errors in investment types was already done in the last question.
2. Now only entries where the InvestmentType was Private Equity were considered. This was done by creating a Boolean array. The data generated was then assigned to a variable data2.
3. A new DataFrame was created using columns StartupName and InvestorsName and assigned to a variable. NAN values were dropped as they would not contribute to the result. An investor name in the column means that the investor invested in a start-up. Thus value counts of investors who invested in different start-ups would give us our answer.
4. Data Cleaning needed to be done for start-up names. Thus all start-up names were viewed using value counts (done in third question). A simple ctrl+f (find) search was done for wrong/misspelled start-up names. The wrong start-up names were then replaced with the correct ones.
5. Now as multiple investments by the same investor in a company were to be counted as one, I grouped the investors according to the company. I grouped by the StartupName and analogous to summing remaining numeric values group-wise, I joined the strings of investors group-wise by applying join to InvestorsName column. I then reset the index.
6. I then applied the SeparateInvestors_RepetitionsNotAllowed() function to every entry of the InvestorsName column which would replace every entry with a set of unique investors which I then exploded as done in the previous question.
7. I then reset the index.
8. I then picked up only those investor names where investor names were NOT undisclosed as undisclosed ones would not provide me any useful information. I did this using the bitwise and operation.
9. I then used the value_counts() function to view the top five investors.
10. I then assigned the top five investors and their values (no. of times invested in different companies) to variables.
11. I the plotted a graph between Investors and Number of Times Invested in Different Companies (Private Equity).
12. I then printed the answer.