Project – Case Study on Indian Startups

Tanmay Sabnis.

Question 1.

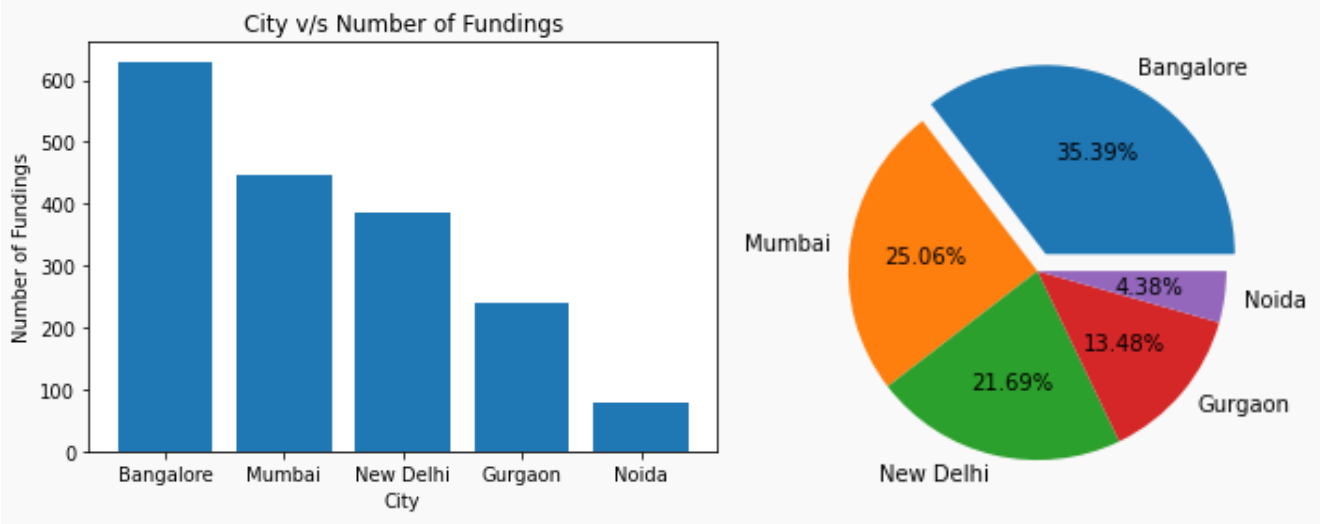
Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, you 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. Sequoia Capital 36

Answer:

The city with maximum number of funding done is Bangalore with around 630 times funding done.

Bangalore occupies 35.39% of the funding market amongst the 5 cities.

Visualizations:



Justification:

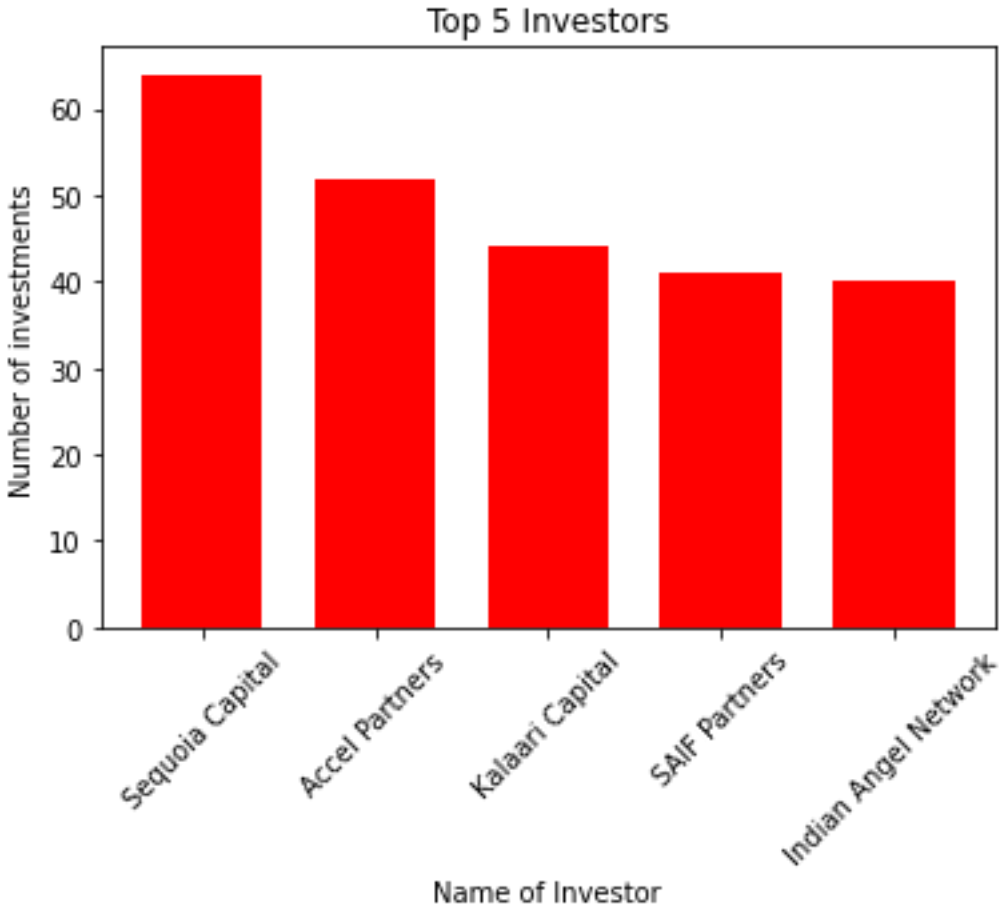
The following logic was used to implement the answer:

1. Importing and Reading the libraries NumPy, Pandas, Matplotlib
2. Dropping all NaN values in 'CityLocation'.
3. Next data cleaning:
    a. Changing Delhi to New Delhi
    b. Splitting the values with 2 cities and retaining only the required city
    c. Capitalising first letter of each city
4. Next extracting insights by using value_counts( )
5. Printing first 5 extracted cities with most funding received

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:



As we can see the top 5 Investors who have invested maximum number of times are:

1. 'Sequoia Capital' - 64
2. 'Accel Partners' - 52
3. 'Kalaari Capital' - 44
4. 'SAIF Partners' - 41
5. 'Indian Angel Network' - 40

Justification:

1. Importing and Reading the libraries NumPy, Pandas, Matplotlib
2. Filtering the data by removing 'undisclosed' investors and appending each investors name in an array
3. Creating a dict and traversing through the array
4. Storing in dictionary, as it will not store duplicate values, our 'value' for key will be count of investments made
5. Storing the keys and values of top 5 investors who invested maximum number of times in 2 NumPy arrays: np_x and np_y
6. Printing first 5 values

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:

The top 5 investors who have invested in maximum number of different start-ups are:

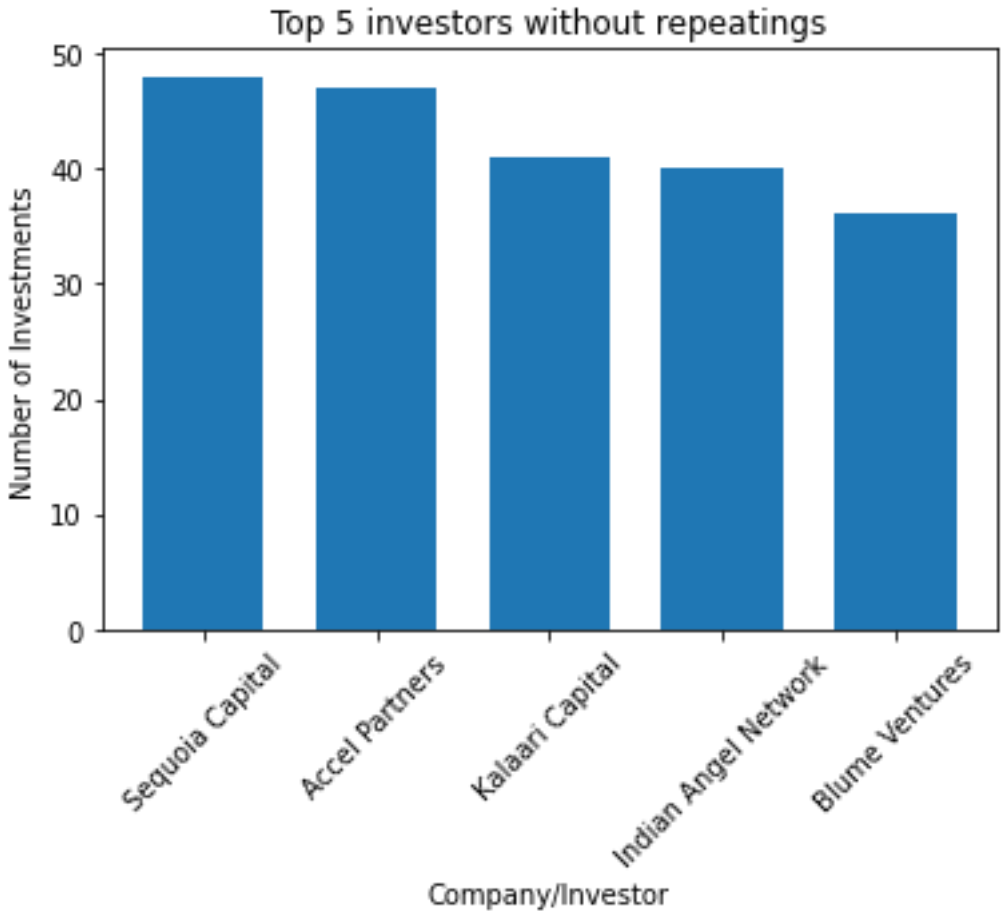
1. Sequoia Capital - 48

2. Accel Partners - 47

3. Kalaari Capital - 41

4. Indian Angel Network - 40

5. Blume Ventures – 36



Justification:

1. Importing and Reading the libraries NumPy, Pandas, Matplotlib.
2. Data Cleaning:
    a. Dropping all NaN values
    b. Clearing spelling mistakes of Ola, Flipkart, etc
3. Creating a Dict
    a. Traverse through all rows
    b. Obtain each value under 'InvestorName'.
        i. It will be either a multi-name string
        ii. Or it will be a single name
    c. For each of the (b) categories:
        i. Obtain split words on '/'
        ii. Check if that name is already in dictionary
            1. If its already in dictionary, it means a set has been created, the add to the set.
            2. If not then create an empty set and add that start-up name as first element of set.
    d. Thus we will get a dictionary with key as distinct names of investors and value as a set of all the investments they have made, it they made investment at 3 places then only 1 will be counted in the set because that's property of a set to retain unique values
4. From this dict, obtain non empty keys and sort them by sorted( ) function
5. Store in different array and print top 5 investors who invested in distinct companies

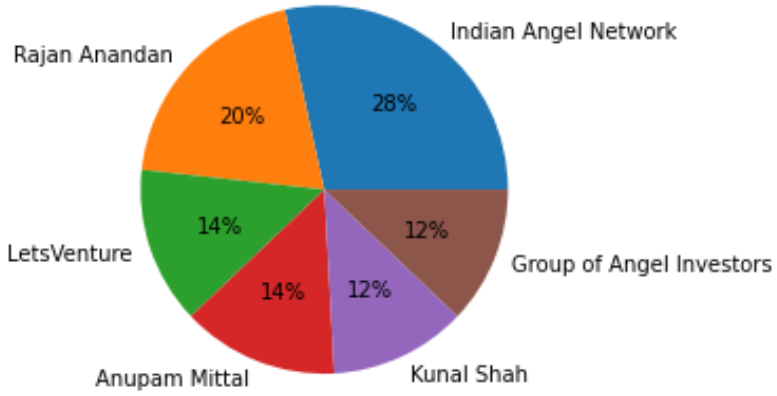
Question 4.

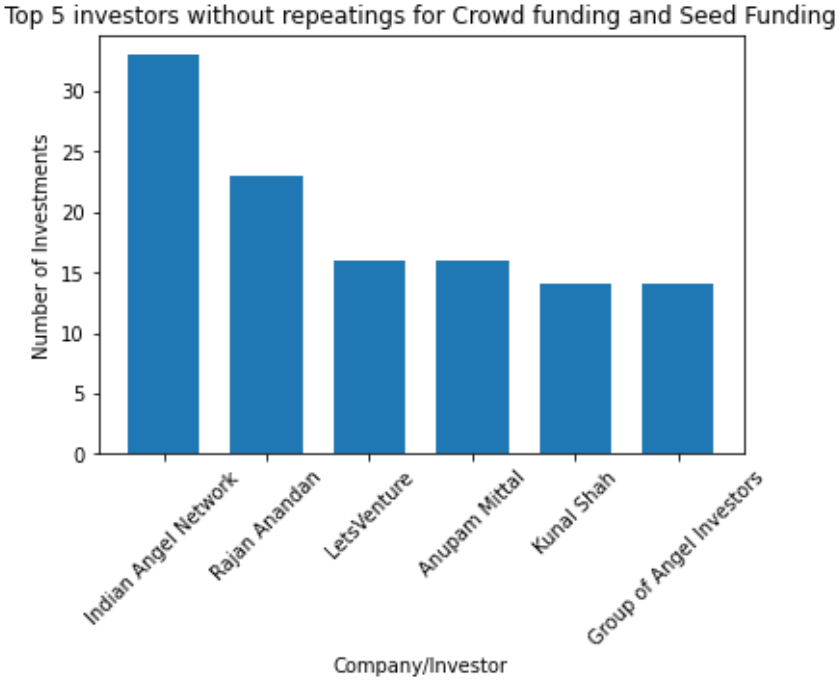
Even after putting so much effort in finding the probable investors, it didn't turn out to be helpful for your friend. So you went to your investor friend to understand the situation better and your investor friend explained to you about the different Investment Types and their features. This new information 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:

Top 5 investors are:

1. Indian Angel Network - 33
2. Rajan Anandan - 23
3. LetsVenture - 16
4. Anupam Mittal - 16
5. Kunal Shah – 14
6. Group of Angel Investors - 14





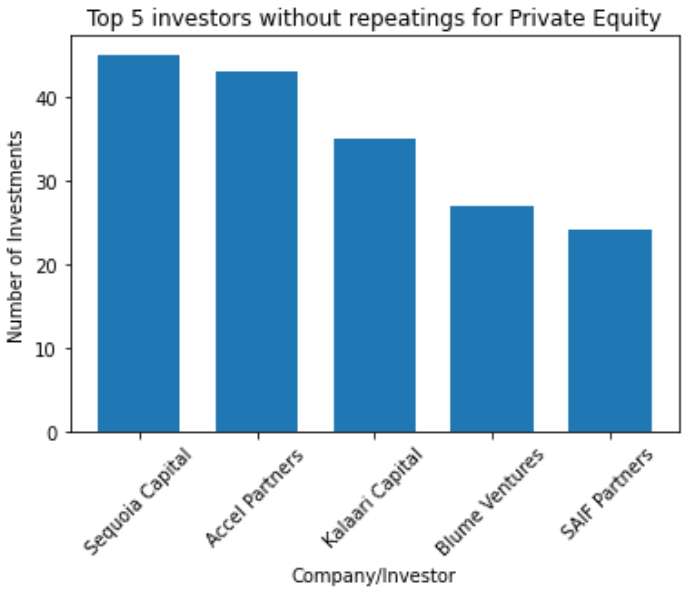
Justification:

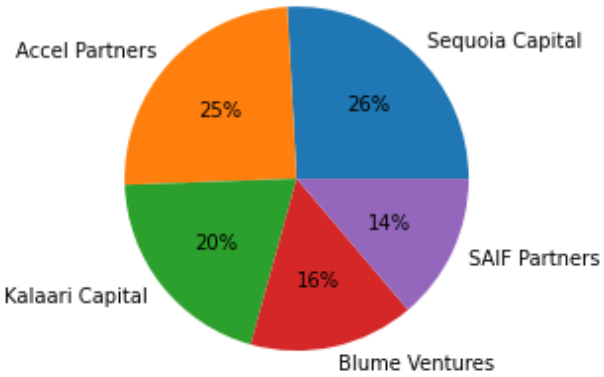
1. Importing and Reading the libraries NumPy, Pandas, Matplotlib.
2. Data Cleaning:
    a. Dropping all NaN values
    b. Retaining dataframe only where investments are of type crowdfunding and seed funding
    c. Clearing spelling mistakes of Ola, Flipkart, etc
3. Creating a Dict
    a. Traverse through all rows
    b. Obtain each value under 'InvestorName'.
        i. It will be either a multi-name string
        ii. Or it will be a single name
    c. For each of the (b) categories:
        i. Obtain split words on '/'
        ii. Check if that name is already in dictionary
            1. If its already in dictionary, it means a set has been created, the add to the set.
            2. If not then create an empty set and add that start-up name as first element of set.
    d. Thus we will get a dictionary with key as distinct names of investors and value as a set of all the investments they have made, it they made investment at 3 places then only 1 will be counted in the set because that's property of a set to retain unique values
4. From this dict, obtain non empty keys and sort them by sorted( ) function
5. Store in different array and print top 5 investors who invested in distinct companies only via crowd funding or 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:





The top 5 Investors for private equity are:

1. Sequoia Capital 45
2. Accel Partners 43
3. Kalaari Capital 35
4. Blume Ventures 27
5. SAIF Partners 24

Justification:

1. Importing and Reading the libraries NumPy, Pandas, Matplotlib.

2. Data Cleaning:

    a. Dropping all NaN values
    b. Retaining only that part of dataframe where 'Private Equity' Investments were made
    c. Clearing spelling mistakes of Ola, Flipkart, etc

3. Creating a Dict

    d. Traverse through all rows
    e. Obtain each value under 'InvestorName'.
        i. It will be either a multi-name string
        ii. Or it will be a single name
    f. For each of the (b) categories:
        i. Obtain split words on '/'
        ii. Check if that name is already in dictionary
            1. If its already in dictionary, it means a set has been created, the add to the set.
            2. If not then create an empty set and add that start-up name as first element of set.
    g. Thus we will get a dictionary with key as distinct names of investors and value as a set of all the investments they have made, it they made investment at 3 places then only 1 will be counted in the set because that's property of a set to retain unique values

4. From this dict, obtain non empty keys and sort them by sorted( ) function

5. Store in different array and print top 5 investors who invested in distinct companies via Private Equity.