

PROJECT: CASE STUDY

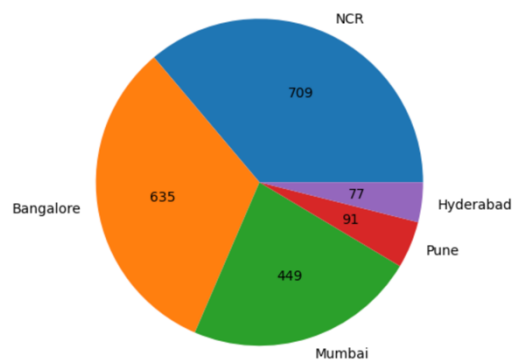
Starting with importing all the necessary libraries that might be useful such as csv, numpy, pandas and matplotlib. Next, using pandas importing the file into a dataframe named df, which will be used as the parent df and copies will be created for each problem so that the original df does not change.

Q1 - 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.

First things first, dropping the nan values in accordance to CityLocation column. Then for few startups multiple locations are given, to update that I have separate the multiple location for startups using separateCity function that I have created. Then, I have corrected the names of cities and clubbed delhi, noida and gurgaon as NCR. Then I have used value_counts() to check the city with most number of fundings.

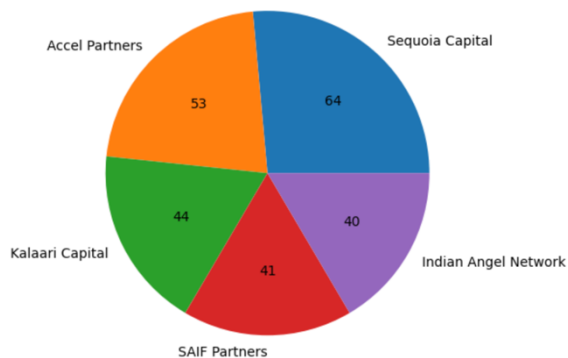
Here are my findings:

My Friend should consider establishing the product startup in **NCR location** as it has the most number of startup fundings. NCR has the most fundings (709) followed by Bangalore (635) and Mumbai (449).



Q2 - 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.

First things first, dropping the nan values in accordance to InvestorsName column. Then I have separated the multiple investors in one startup and stored in a list as we need to find the investors with most number of investments. Also, I have ignored the Undisclosed investors. Then with the help of a dictionary, I have counted the frequency of each investor and stored it. Then I have sorted the dictionary in descending order.



Here are my findings:

These are the top 5 investors with the following number of investments.

Sequoia Capital 64

Accel Partners 53

Kalaari Capital 44

SAIF Partners 41

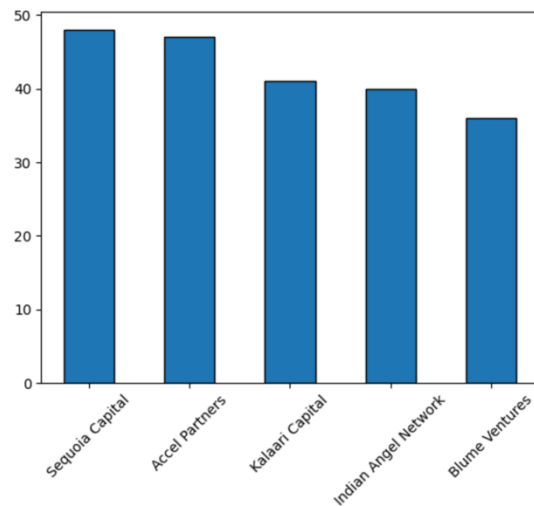
Indian Angel Network 40

Q3 - 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.

First things first, dropping the nan values in accordance to InvestorsName and StartupName column. Then I have checked the duplicate names of Startups like Ola, Flipkart, Oyo and Paytm and updated them. Also, I have ignored the Undisclosed investors. Then I have separated the multiple investors in one startup and updated it in the df. Now, as we need to find investors with max investments in different companies, I have created two lists where every index will correspond to investorName and startupName.

Eg: [InvestorA, InvestorB, InvestorC ...] [Startup1, Startup1, Startup2 ...]

After the lists have been created, I have updated the dataframe data with the lists. Then I have grouped the data using `nunique()` which returns dataframe with counts of unique elements in each position. At last I have sorted the dataframe in descending order.



Here are my findings:

These are the top 5 investors with max investments in different companies along with the following number of investments.

Sequoia Capital 48
 Accel Partners 47
 Kalaari Capital 41
 Indian Angel Network 40
 Blume Ventures 36

Q4 - 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.

First things first, dropping the null values in accordance to InvestorsName, StartupName, and InvestmentType. Then I have checked the duplicate names of Startups like Ola, Flipkart, Oyo, Paytm, Private Equity, Seed Funding, and Crowd Funding and updated them. Also, I have ignored the Undisclosed investors. Now, as we need to get the investments of investment type Seed Funding or Crowd Funding, I have updated the df with only Crowd Funding and Seed Funding investments. Then I have separated the multiple investors in one startup and updated it in the df. Now, as we need to find investors with max investments in different companies, I have created two lists where every index will correspond to investorName and startupName.

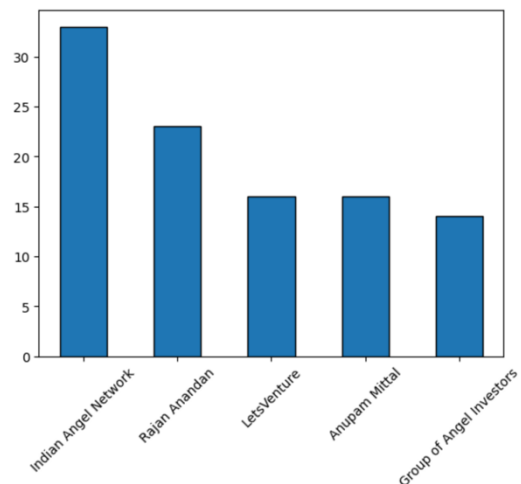
Eg: [InvestorA, InvestorB, InvestorC ...] [Startup1, Startup1, Startup2 ...]

After the lists have been created , I have updated the dataframe data with the lists. Then I have grouped the data using `nunique()` which returns dataframe with counts of unique elements in each position. At last I have sorted the dataframe in descending order.

Here are my findings:

These are the Top 5 investors who have invested in a different number of startups and their investment type is Crowdfunding or Seed Funding.

Indian Angel Network 33
Rajan Anandan 23
LetsVenture 16
Anupam Mittal 16
Group of Angel Investors 14



Q5 - 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.

First things first, dropping the null values in accordance to InvestorsName, StartupName, and InvestmentType. Then I have checked the duplicate names of Startups like Ola, Flipkart, Oyo, Paytm, Private Equity, Seed Funding, and Crowd Funding and updated them. Also, I have ignored the Undisclosed investors. Now, as we need to get the investments of investment type Seed Funding or Crowd Funding, I have updated the df with only Private Equity investments. Then I have separated the multiple investors in one startup and updated it in the df. Now, as we need to find investors with max investments in different companies, I have created two lists where every index will correspond to investorName and startupName.

Eg: [InvestorA, InvestorB, InvestorC ...] [Startup1, Startup1, Startup2 ...]

After the lists have been created , I have updated the dataframe data with the lists. Then I have grouped the data using `nunique()` which returns dataframe with counts of unique elements in each position. At last I have sorted the dataframe in descending order.

Here are my findings:

Top 5 investors who have invested in a different number of startups and their investment type is Private Equity.

Sequoia Capital 45
Accel Partners 43
Kalaari Capital 35
Blume Ventures 27
SAIF Partners 24

