PROJECT 1: CASE STUDY OF INDIAN STARTUPS (II)

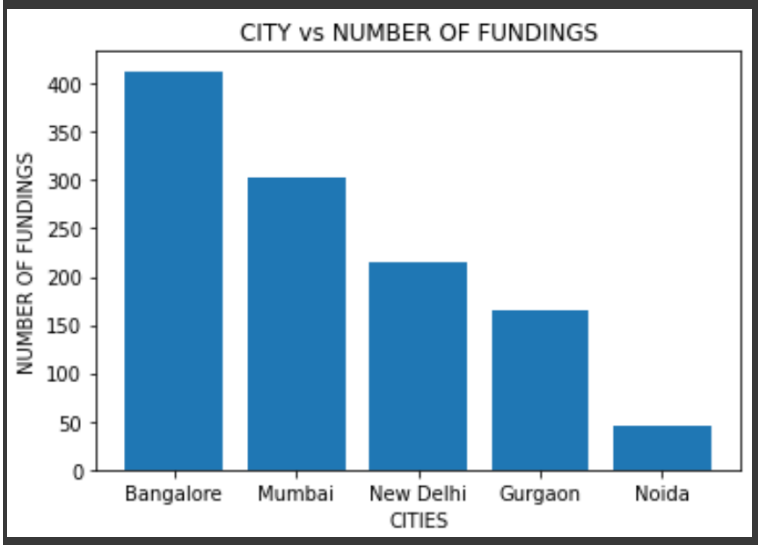
****

Problem 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.

OUTPUT:

```
CITY => NUMBER OF FUNDINGS
--------------------------
1. Bangalore => 413
2. Mumbai => 303
3. New Delhi => 215
4. Gurgaon => 166
5. Noida => 46
```




```
RESULT: BANGALORE is the city where the most number of funding is done.
```

JUSTIFICATION:

STEP 1: We read the dataset from startup_funding.csv file.

STEP 2: We remove the rows where either 'CityLocation' or 'AmountInUSD' is empty because these rows will not be used while counting number of startups that got funding.

STEP 3: As instructed in the problem, we replace 'Delhi' with 'New Delhi', 'bangalore' with 'Bangalore'. Also, we have change 'SFO / Bangalore' and 'Seattle / Bangalore' to 'Bangalore' because it is given in the problem if either of the city is in Target cities then we have to count it.

STEP 4: Since the 'CityLocation' column may contain multiple cities, hence we iterate over the CityLocation column and split the city name separated by '/(slash)' and take the first city.

STEP 5: Now we take only the rows where 'CityLocation' is 'Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon' or 'Noida' because due to financial restriction, my friend can only choose between these cities.

STEP 6: We group the rows together based on CityLocation and count number of fundings for each city.

STEP 7: We print the output. (It can be observed that Bangalore would be the best option.)

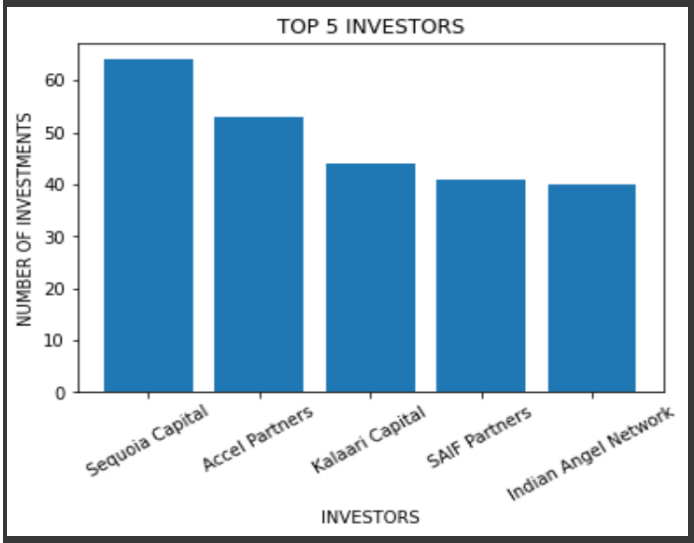
STEP 8: We plot the bar graph for better visualization of the data.

Problem 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.

OUTPUT:

```
INVESTOR => NUMBER OF INVESTMENTS
--------------------------------
1. Sequoia Capital => 64
2. Accel Partners => 53
3. Kalaari Capital => 44
4. SAIF Partners => 41
5. Indian Angel Network => 40
```



JUSTIFICATION:

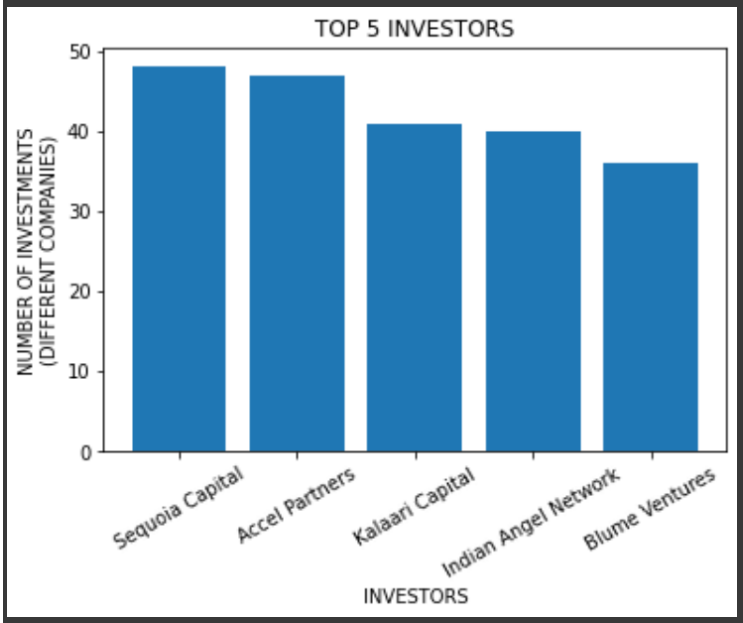
STEP 1: We read the dataset from startup_funding.csv file.
STEP 2: We remove the rows where either 'InvestorsName' is either empty or undisclosed.
STEP 3: Since the 'InvestorsName' column contains multiple names, hence we iterate over the 'InvestorsName' column and split the names separated by ',(comma)' and store a list of all names in a new column 'Investors'.
STEP 4: Now, we iterate over all the rows and all investors in each row, and count number of investments made by all investors.
STEP 5: We sort the result in descending order and take top 5 investors.
STEP 6: We print the output.
STEP 7: We plot the bar graph for better visualization of the data.

Problem 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.

OUTPUT:

```
INVESTOR => NUMBER OF INVESTMENTS
---------------------------------
1. Sequoia Capital => 48
2. Accel Partners => 47
3. Kalaari Capital => 41
4. Indian Angel Network => 40
5. Blume Ventures => 36
```



JUSTIFICATION:

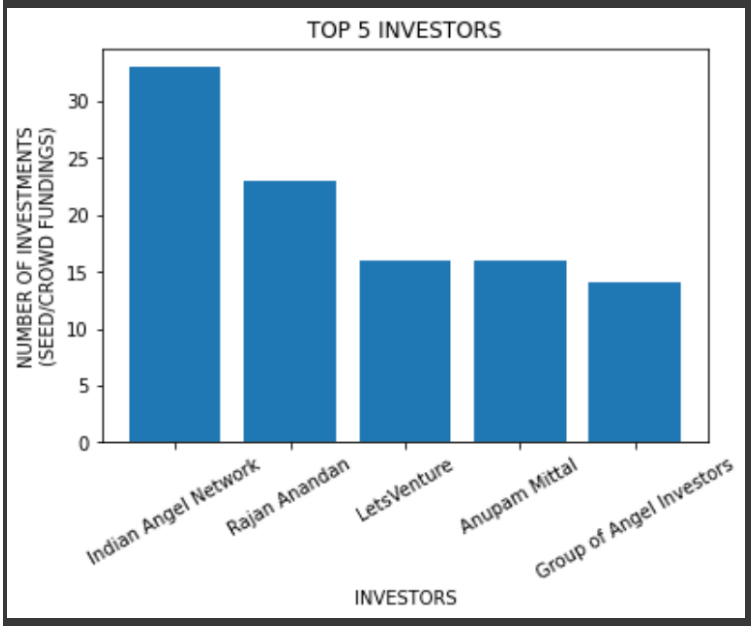
STEP 1: We read the dataset from startup_funding.csv file.
STEP 2: We remove the rows where either 'InvestorsName' is either empty or undisclosed.
STEP 3: We correct all the names related to 'Flipkart', 'Ola', 'Paytm' and 'Oyo'.
STEP 4: Since the 'InvestorsName' column contains multiple names, hence we iterate over the 'InvestorsName' column and split the names separated by ',(comma)' and store a list of all names in a new column 'Investors'.
STEP 5: Now, we iterate over all the rows and all investors in each row, and make a list of startups in which each investor has invested. Then we take the length of unique elements in each list and that would give us the number of different companies in which an investor has invested.
STEP 6: We sort the result in descending order and take top 5 investors.
STEP 7: We print the output.
STEP 8: We plot the bar graph for better visualization of the data.

Problem 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.

OUTPUT:

```
INVESTOR => NUMBER OF INVESTMENTS
---------------------------------
1. Indian Angel Network => 33
2. Rajan Anandan => 23
3. LetsVenture => 16
4. Anupam Mittal => 16
5. Group of Angel Investors => 14
```



JUSTIFICATION:

STEP 1: We read the dataset from startup_funding.csv file.

STEP 2: We remove the rows where either 'InvestorsName' is either empty or undisclosed.

STEP 3: We make corrections in the names of 'InvestmentType' by replacing 'Crowd funding', 'PrivateEquity', 'SeedFunding' with 'Crowd Funding', 'Private Equity' and 'Seed Funding' respectively.

STEP 4: We correct all the names related to 'Flipkart', 'Ola', 'Paytm' and 'Oyo'.

STEP 5: We only take those rows where 'InvestmentType' is either 'Seed Funding' or 'Crowd Funding'.

STEP 6: Since the 'InvestorsName' column contains multiple names, hence we iterate over the 'InvestorsName' column and split the names separated by ',(comma)' and store a list of all names in a new column 'Investors'.

STEP 7: Now, we iterate over all the rows and all investors in each row, and make a list of startups in which each investor has invested. Then we take the length of unique elements in each list and that would give us the number of different companies in which an investor has invested with InvestmentType either 'Seed Funding' or 'Crowd Funding'.

STEP 8: We sort the result in descending order and take top 5 investors. (NOTE: Here we are getting two names 'Group of Angel Investors' and 'Kunal Shah' with same frequency i.e.14 at position 5th and 6th. So we are taking the first one in top 5. Other can also be taken.)

STEP 9: We print the output.

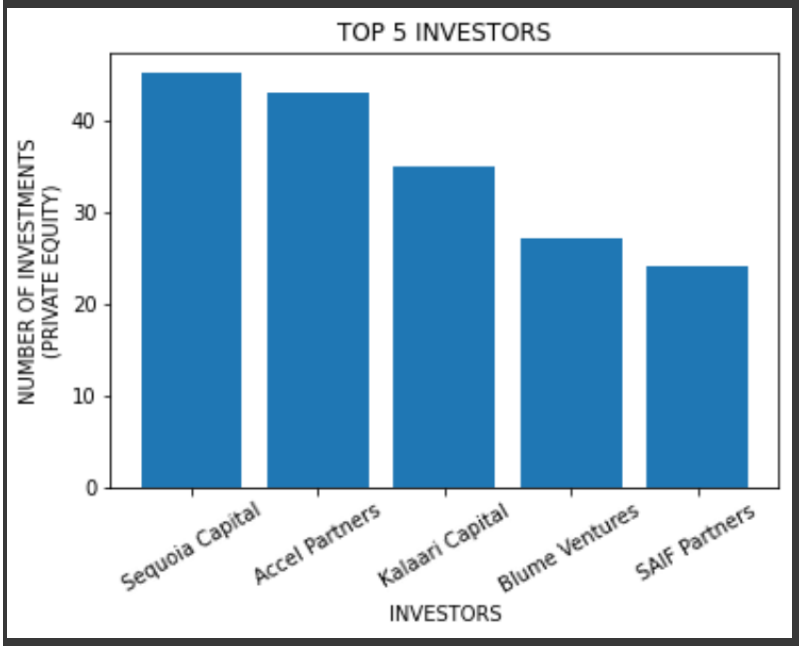
STEP 10: We plot the bar graph for better visualization of the data.

Problem 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.

OUTPUT:

```
INVESTOR => NUMBER OF INVESTMENTS
---------------------------------
1. Sequoia Capital => 45
2. Accel Partners => 43
3. Kalaari Capital => 35
4. Blume Ventures => 27
5. SAIF Partners => 24
```



JUSTIFICATION:

STEP 1: We read the dataset from startup_funding.csv file.

STEP 2: We remove the rows where either 'InvestorsName' is either empty or undisclosed.

STEP 3: We make corrections in the names of 'InvestmentType' by replacing 'Crowd funding', 'PrivateEquity', 'SeedFunding' with 'Crowd Funding', 'Private Equity' and 'Seed Funding' respectively.

STEP 4: We correct all the names related to 'Flipkart', 'Ola', 'Paytm' and 'Oyo'.

STEP 5: We only take those rows where 'InvestmentType' is 'Private Equity'.

STEP 6: Since the 'InvestorsName' column contains multiple names, hence we iterate over the 'InvestorsName' column and split the names separated by ',(comma)' and store a list of all names in a new column 'Investors'.

STEP 7: Now, we iterate over all the rows and all investors in each row, and make a list of startups in which each investor has invested. Then we take the length of unique elements in each list and that would give us the number of different companies in which an investor has invested with InvestmentType 'Private Equity'.

STEP 8: We sort the result in descending order and take top 5 investors.

STEP 9: We print the output.

STEP 10: We plot the bar graph for better visualization of the data.