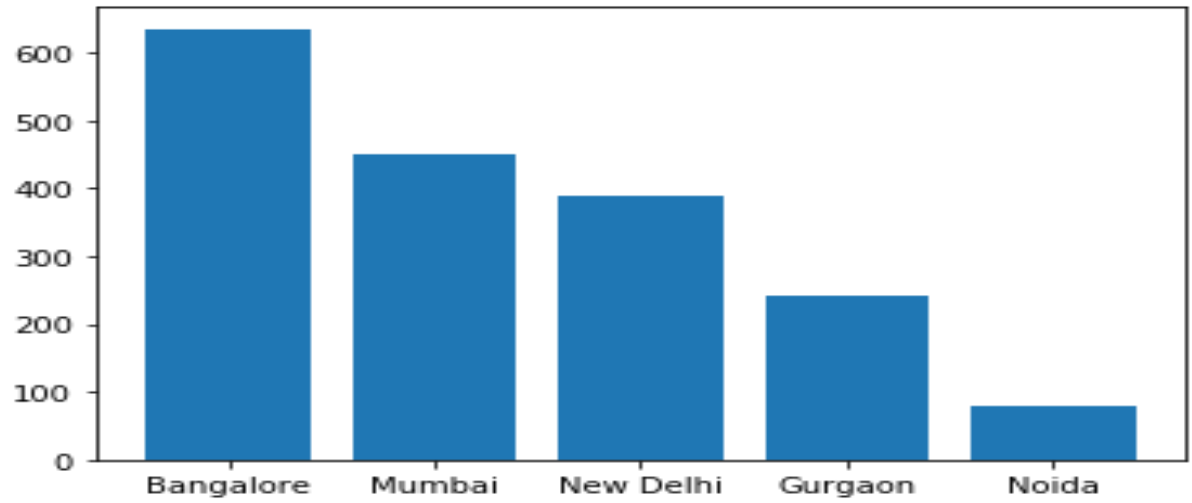
Question 1.
Answer:
As I need the names of the location so I need to make sure that there is no missing value in the CityLocation column, hence I have dropped all the rows wherein missing value is present in CityLocation.
Now second challenge is to handle the case where startup has been established in more than one locations,
now I city names which are based in India, and the question it has been mentioned earlier that, firstcity name
will be based Indian city name, so I have splited the city names with '/' and picked the first name and return
using my city_name function.

There are some cases wherein I need to update city names with proper ones which I have done using pandas
replace method i.e I have update the name of 'Delhi' to 'New Delhi', 'bangalore' to 'Bengalore'. Now I have
proper city names. Now I have used pandas method, value_counts on CityLocation column which returned me the
City Names with the count of startups estabilished there. Now I have made one set of target_locations wherein
my friend wants to establish the startup. Before that I have made two list one for city names and the other one
for counts of startups that is established on that particular city. I did some memory and time optimisation by
modifying the using same lists. After all I have got the number of startups that has been established in the
specified location by my friend and Now I have ploted and got that Bangalore has the highest number of startups
established there. So it would be great if my friend choose to go with location Bangalore.

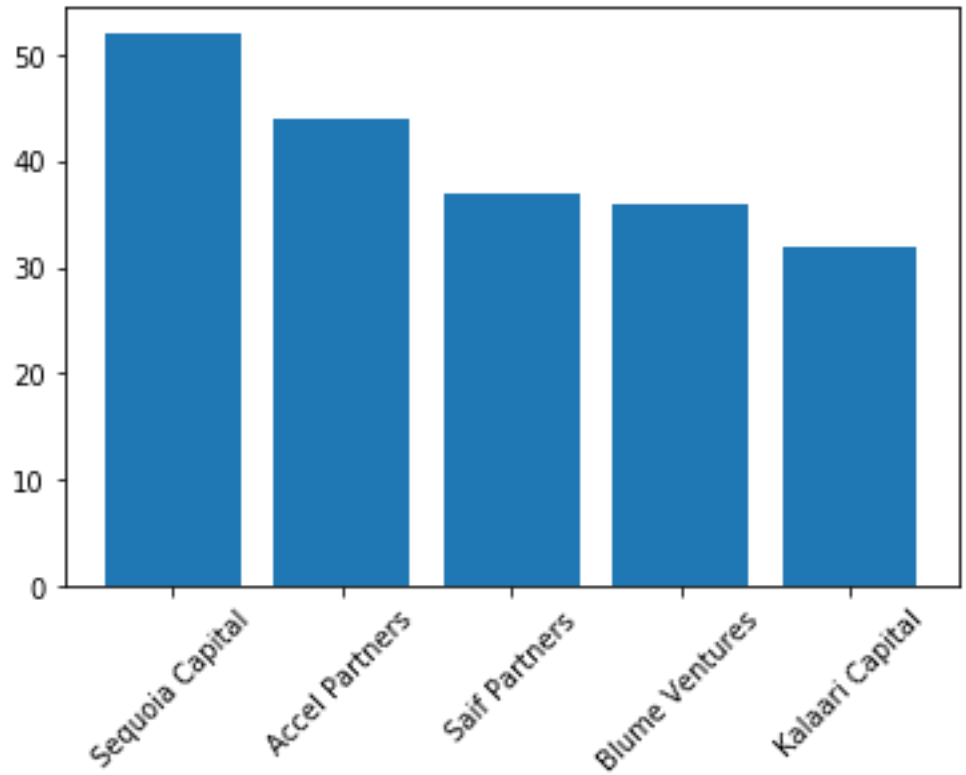


Question 1 Plot

Question 2.
Answer:
Now I know the locations wherein my friend wants to establish the startup, so I can eliminate
all other cities. In order to do that first I have followed the same process of updating the names and
all. Once that has been done, I have used one function to make all other city as '' so that I can eliminate
using this condition. So my funciton filter_city takes one argument and argument is checked whether it
is present in the set target_locations wherein I have mentioned specified city names. If name present then
do nothing other than make ''. In the next step I have picked all the indices where CityLocation is ''
after that I have used boolean indexing on the data_frame and all other cities got filtered out.

In second part, I have made one investors dictionary which will help me get the count of startup
investment. After following the first process, now I am making sure that I have no missing value in
InvestorNames by dropping all rows wherein InvestorName(s) is/are missing. In InvestorNames column there could
be multiple names so to update accordingly using investor_names method, this method takes one argument of investornames
and split on the basis of ',' so I will get all the names which I have stored in names and now I have iterator over
all the investor names and updated the occurence of investorname in the investors dictionary. I have handle one case
that is after spliting we might get empty string so name should not be empty.

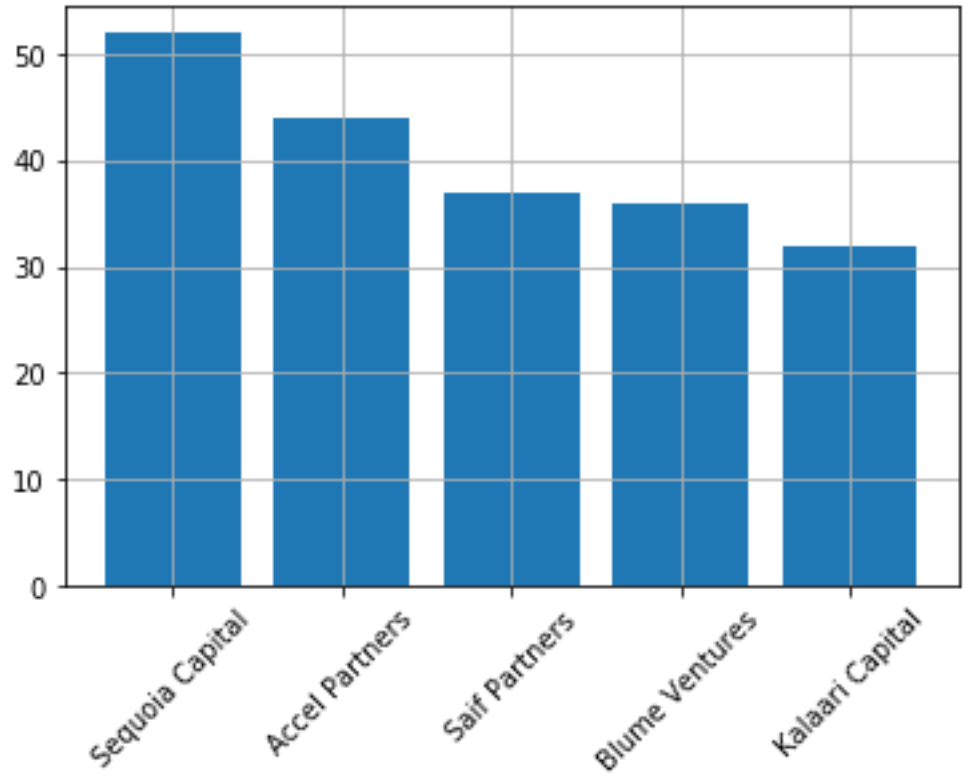
Last part of the problem boil down to simply making one list of list where the first element of the inner
list will be the count and second will be the name of the investor. After that I have sorted the list of list based
on my first column and from there I have ignored the first one as it is undisclosed investors and I put names and
values in a different lists so I will need to pass two argument in bar graph function. From the bar graph we can see
that Sequoia Capital has the highest number of investment in the startups. So if my friend approaches Sequoia Capital
then it will give him the maximum chance of getting investment.



Question 2 Plot

Question 3.
Answer:
I have followed the answer of question Q.1 and Q.2, herein I just need to twick the case, i.e I need to store all unique startups names wherein investor invested. Before that I have made sure
all the mentioned startup names are modified accordingly using updateStartupName. Previously I was
storing the occurence of Investor but here I will be storing the startupname wherein Investor invested.
So my investors dictory will contain all the investors and their investments in differnt startup.
Here I am storing the startup names in set so if startup name is already present so I don't need to
count it one more time. So investors will store like this
{'InvestorName':{'Startup1','Startup2'}}. In
order to make the above logic I will need to combine startupname and investornames so I did that by
creating one more column with name startup_investor. Once it has been done, I have applied one function
with name startup_investors wherein I have implemented the above logic.

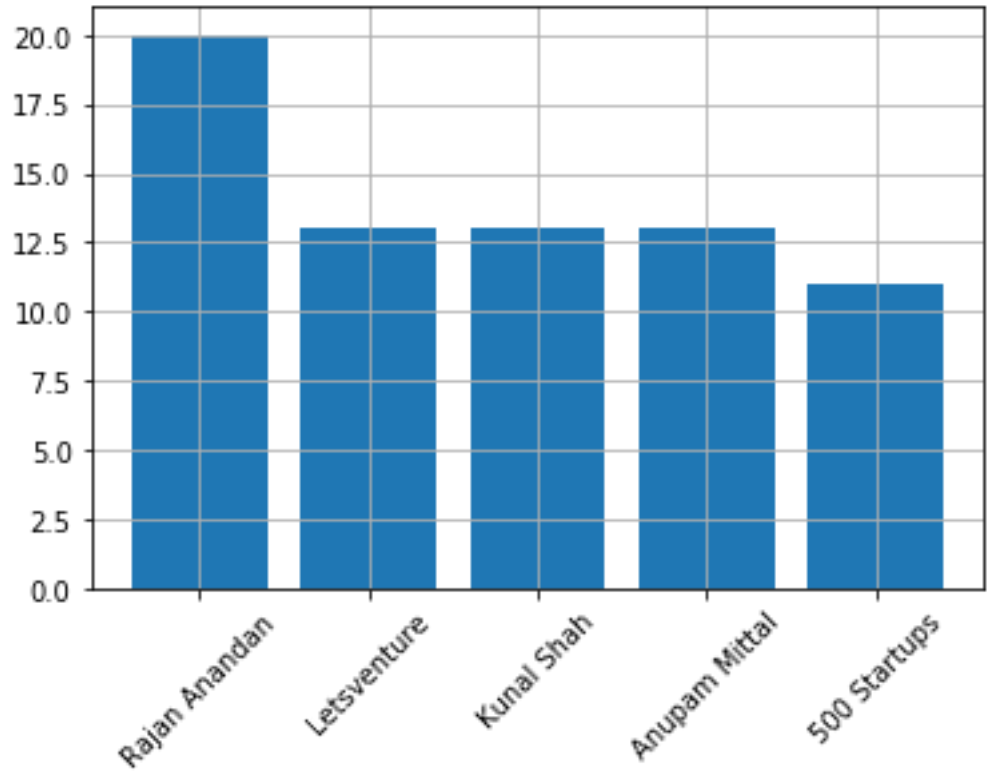
In the next step what I have done, I have investors and all the startup associated with them, so
I have made one list of list wherein I have store count of different startup and the investorname. After
that I have sorted the list of list with first column and picked top five investors and ignore undisclosed
investor. At last I have made two different list just to plot the bar graph. From the bar graph if we see
still Sequoia Capital has invested maximum number of times in different startups, so It will be good choice
to approach Sequoia Capital.



Question 3 Plot

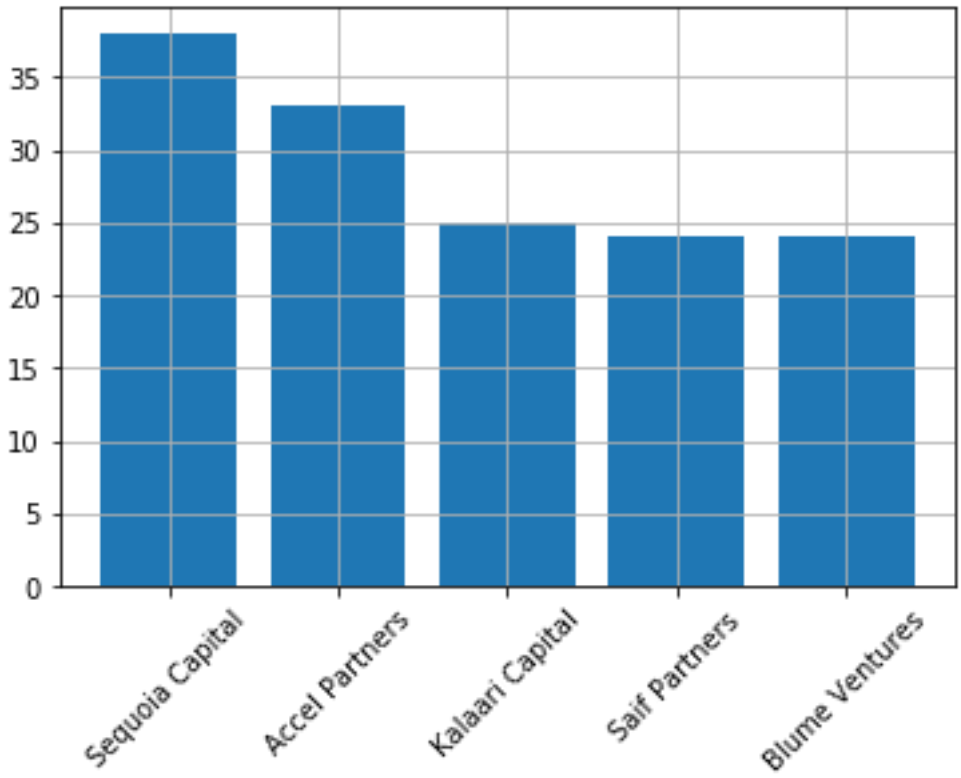
Question 4.
Answer:
Following the previous steps and here I just need to filter out the InvestmentType which are other than
'Seed Funding and Crowd Funding'. To make sure I don't miss out any first I have correct spelling of all the
InvestmentType by using updateType. So once I get all the proper InvestmentType I can apply the filter so i did
that using boolean indexing, I have taken only those rows wherein InvestmentType is Seed Fuding or Crowd Funding.

Once that is done, now I have followed the previous step which I had followed in Q.3 to figure out all the different
startup wherein investor has invested. And followed the process till I get the top five Investors. After ploting
the bar graph I have found that Rajan Anandan did the highest number of investment as Seed funding or crowd funding so
it will be better to approach Rajan Anandan.



Question 4 Plot

Question 5.
Answer:
This question is almost similar to Q4 but here I need to filter out all the InvestmentType other than
'Private Equity'. I did all the process which I have followed in Q4. just I have alter the condition of filtering out
the InvestmentType. After doing all, if I plot a bar graph then we can notice that Sequoia Capital has maximum number
of investment as investment type='Private Equity' in different startup. So it is better to contact Sequoia Capital.



Question 5 Plot