Case Study Project Documentation

Ans 1: The data frame was made using pandas library in python and then the following tasks were carried out.

1. Handling the Nan of City Location
2. Separating cities based on '/' and keeping the first name only.
3. Replacing the duplicate and incorrect names
4. Counting the number of StartUp's in different cities
5. Printing the data
6. Plotting the Bar Graph.

The Answer came out to be :

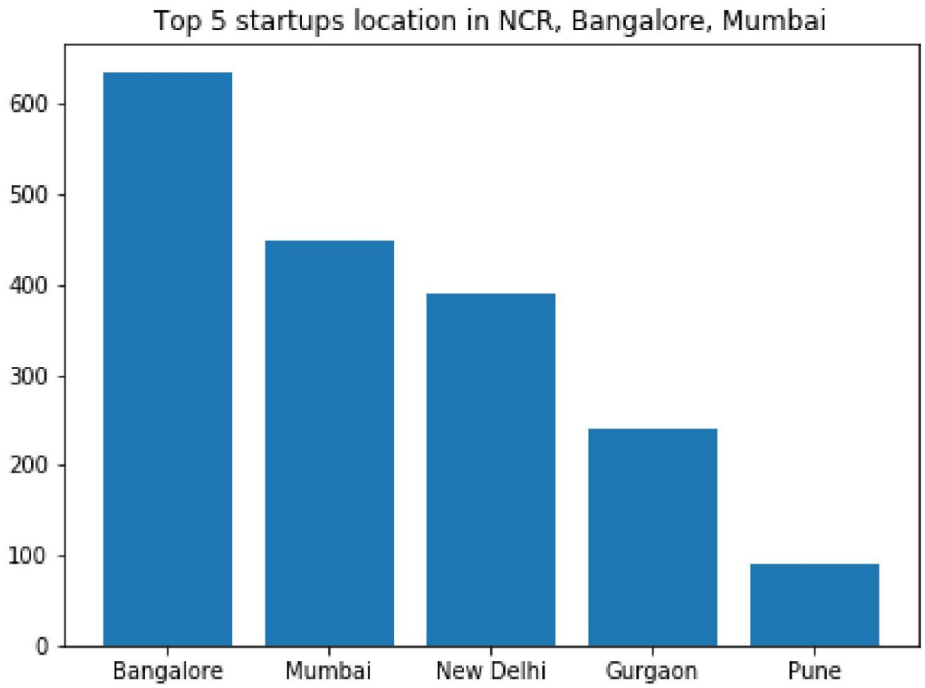
Bangalore 635

Mumbai 449

New Delhi 389

Gurgaon 241

Pune 91



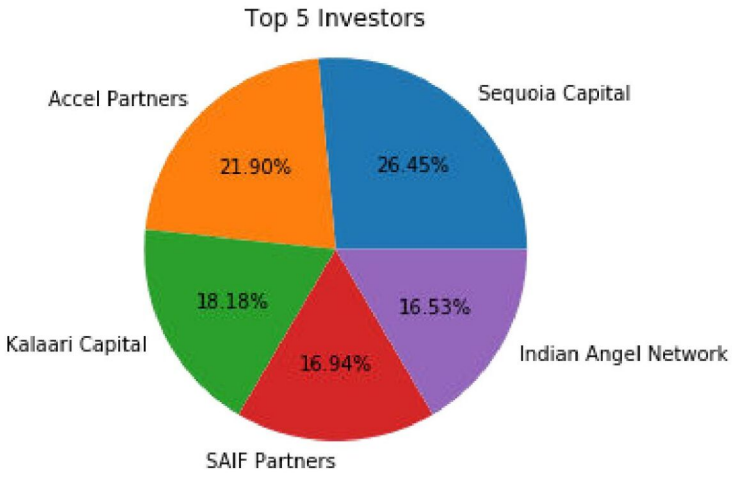
Ans 2: The data frame was made using pandas library in python and then the following tasks were carried out.

1. Handling the Nan values in investors Names.
2. Dictionary function for counting the number of funding's for each investor.
3. Array of investors names is passed.
4. In some records, there are multiple Investors names separated by comma.
5. if the single name is present and increased the counter.

6. If the multiple names are present then split by the comma and create an array of names. Extract the name, put into the dictionary and increase the counter.
7. Applying the function to the column.
8. Creating the data frame from keys and values of the dictionary.
9. Sorting the data frame based on the number of investment in reverse order.

The Answer came out to be :

Sequoia Capital	**64**
Accel Partners	**53**
Kalahari Capital	**44**
SAIF Partners	**41**
Indian Angel Network	**40**

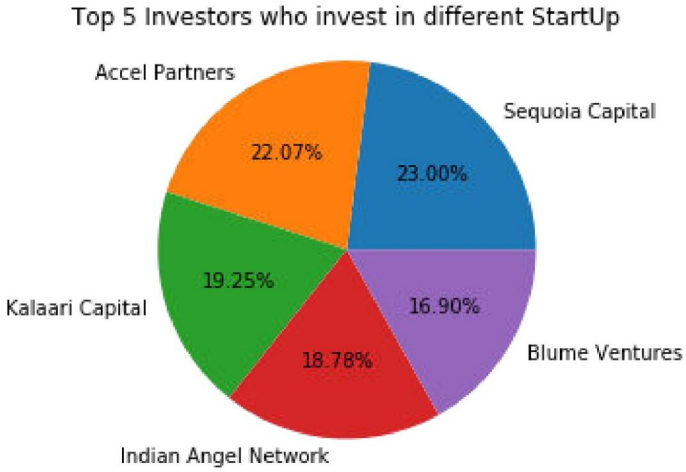


Ans 3: The data frame was made using pandas library in python and then the following tasks were carried out.

1. Handling the Nan of StartUp's Name
2. Replacing the duplicate and incorrect names in StartUp Name Column.
3. Drop the "disclosed Investors" from the Dataframe.
4. Select the StartupName and InvestorsName and zip to both column, covert into the "data" List.
5. Dictionary function for counting the number of unique startup for each investor.
6. In some records, there are multiple Investors names separated by comma.
7. If the single name is present and add the Startup Name into the List.
8. If the multiple names are present then split by the comma and create an array of names. Extract the name, put into the dictionary and the Startup Name into the List.
9. Again revisit the whole dictionary, for every key (InvestorsName) count the length of StartUp List and put the length of the list into the key Value.
10. Sorting the data frame based on the number of investments in reverse order.

The Answer came out to be :

Sequoia Capital:	**49**
Accel Partners:	**47**
Kalaari Capital:	**41**
Indian Angel Network:	**40**
Blume Ventures:	**36**



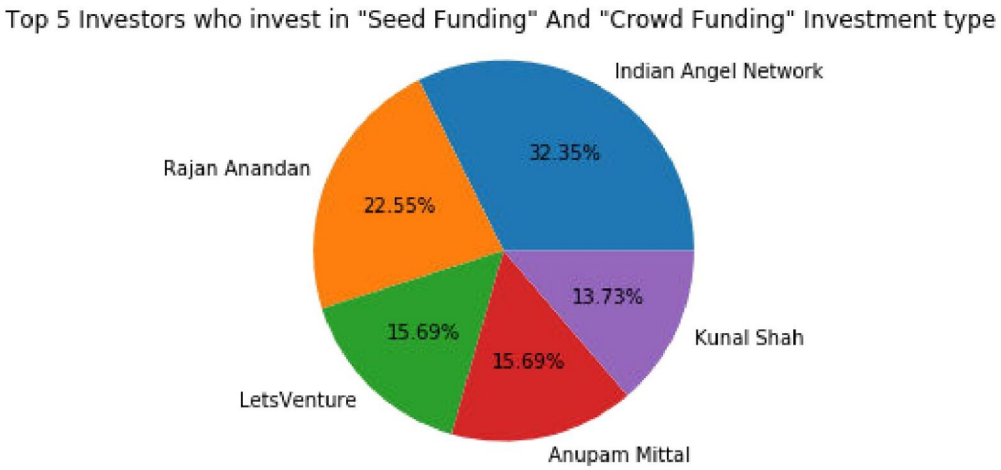
Ans 4: The data frame was made using pandas library in python and then the following tasks were carried out.

1. Handling the Nan of StartUp's Name
2. Replacing the duplicate and incorrect names in StartUp Name Column.
3. Handling the Nan of Investment Type
4. Replacing the duplicate and incorrect names in InvestmentType Column.
5. Drop the "disclosed Investors" from the Dataframe.
6. Select the Dataframe where investment Type is "Seed Funding" and "Crowd Funding".
7. Select the StartupName and InvestorsName and zip to both column, covert into the "data" List.
8. Dictionary function for counting the number of unique startup for each investor.
9. In some records, there are multiple Investors names separated by comma.
10. If the single name is present and add the Startup Name into the List.
11. If the multiple names are present then split by the comma and create an array of names. Extract the name, put into the dictionary and the Startup Name into the List.
12. Again revisit the whole dictionary, for every key (InvestorsName) count the length of StartUp List and put the length of the list into the key Value.

13. Sorting the data frame based on the number of investment in reverse order.

The Answer came out to be :

Indian Angel Network:	**33**
Rajan Anandan:	**23**
LetsVenture:	**16**
Anupam Mittal:	**16**
Kunal Shah:	**14**



Ans 5: The data frame was made using pandas library in python and then the following tasks were carried out.

1. Handling the Nan of StartUp's Name
2. Replacing the duplicate and incorrect names in StartUp Name Column.
3. Handling the Nan of Investment Type
4. Replacing the duplicate and incorrect names in InvestmentType Column.
5. Drop the "disclosed Investors" from the Dataframe.
6. Select the Dataframe where investment Type is "Private Equity".
7. Select the StartupName and InvestorsName and zip to both column, covert into the "data" List.
8. Dictionary function for counting the number of unique startup for each investor.
9. In some records, there are multiple Investors names separated by comma.
10. If the single name is present and add the Startup Name into the List.
11. If the multiple names are present then split by the comma and create an array of names. Extract the name, put into the dictionary and the Startup Name into the List.
12. Again revisit the whole dictionary, for every key (InvestorsName) count the length of StartUp List and put the length of the list into the key Value.
13. Sorting the data frame based on the number of investment in reverse order.

The Answer came out to be :

Sequoia Capital: 46

Accel Partners: 43

Kalaari Capital: 35

Blume Ventures: 27

SAIF Partners: 24