

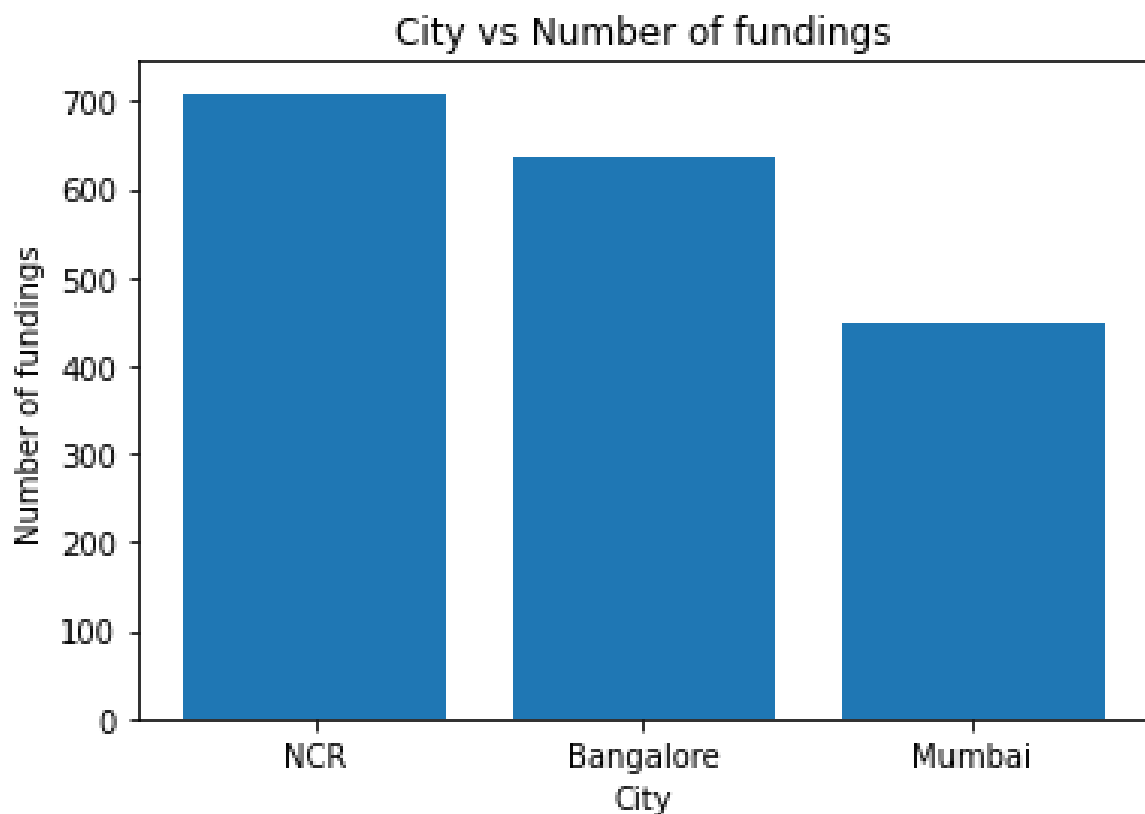
## CASE STUDY - 2

**Q1:** Firstly, I have imported the required modules. After reading data from the CSV file, I have applied data cleaning on the dataset, by removing all the rows where CityLocation has null value using Boolean Indexing. Then I have made a different Data frame City only containing CityLocation. After this I have dropped all the NaN values from that data frame, converted City into an Numpy array and initiate a dictionary to keep a track of the number of entries against NCR, Bangalore and Mumbai. After this I traversed the whole City array. Then for each element applied the necessary changes like consider "Delhi" as "New Delhi", "bangalore" as "Bangalore", meanwhile taking care of start-ups formed in multiple locations. For "Gurgaon", "Noida", "New Delhi", I have taken the city as "NCR". Then I sorted the dictionary in reverse order After this I converted the dictionary to two 1D NumPy array x, y and plotted a bar graph using it.

As a result, we have **NCR** which is the city with the greatest number of fundings i.e., **709**.

### Output:

NCR

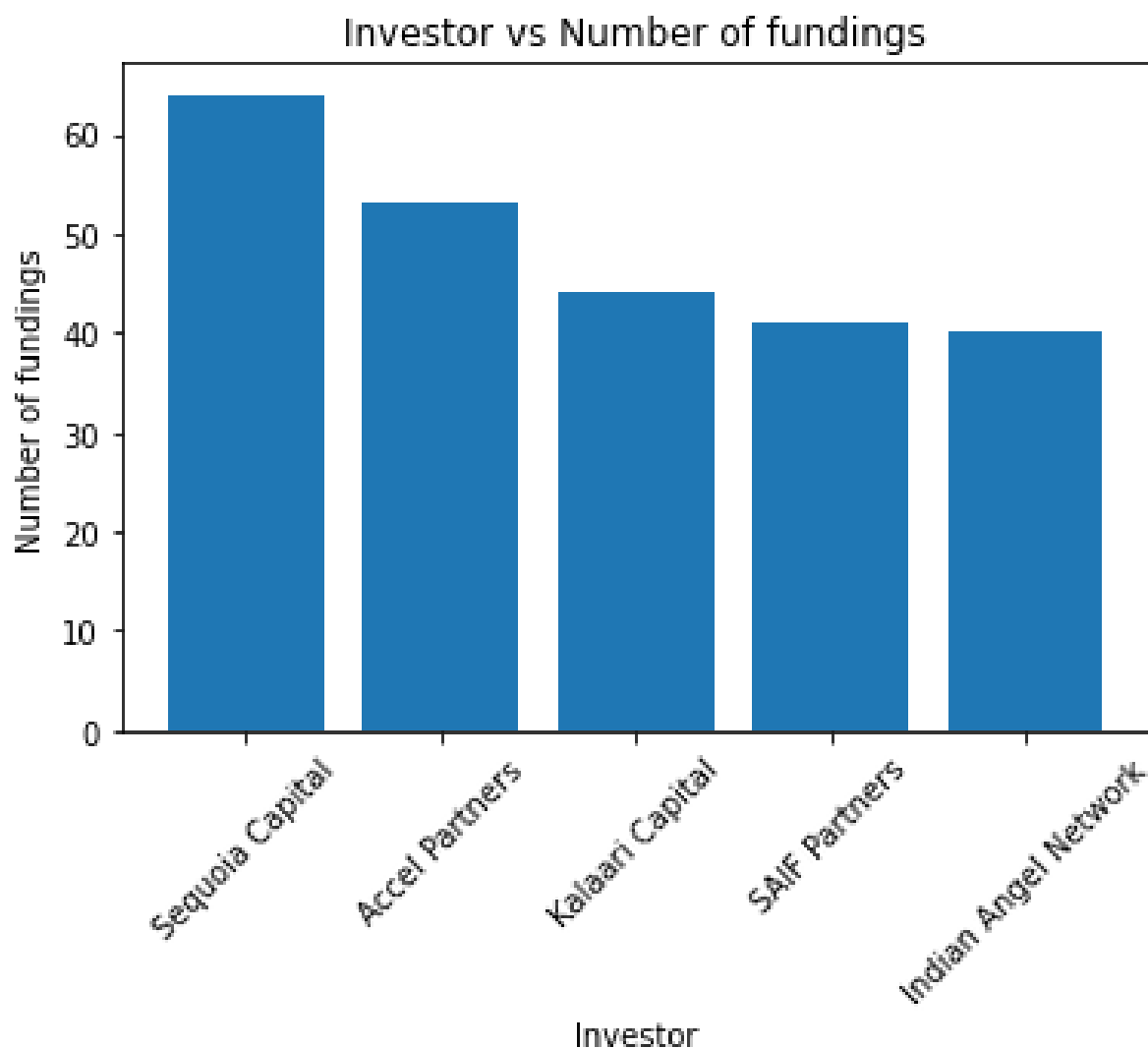


**Q2:** Firstly, I have imported the required modules. After reading data from the CSV file, I have made a different Data frame investor only containing InvestorsName. After this I have dropped all the NaN values from that data frame, converted investor into an Numpy array and initiate a dictionary to keep a track of the investors. After this I traversed the whole City array meanwhile taking care of start-ups funded by multiple investors in the same round. Then I sorted the dictionary in reverse order and convert the dictionary to two 1D NumPy array x, y and plotted a bar graph using it for the top 5 values.

These come out to be: **Sequoia Capital 64, Accel Partners 53, Kalaari Capital 44, SAIF Partners 41, Indian Angel Network 40.**

### Output:

```
Sequoia Capital 64
Accel Partners 53
Kalaari Capital 44
SAIF Partners 41
Indian Angel Network 40
```

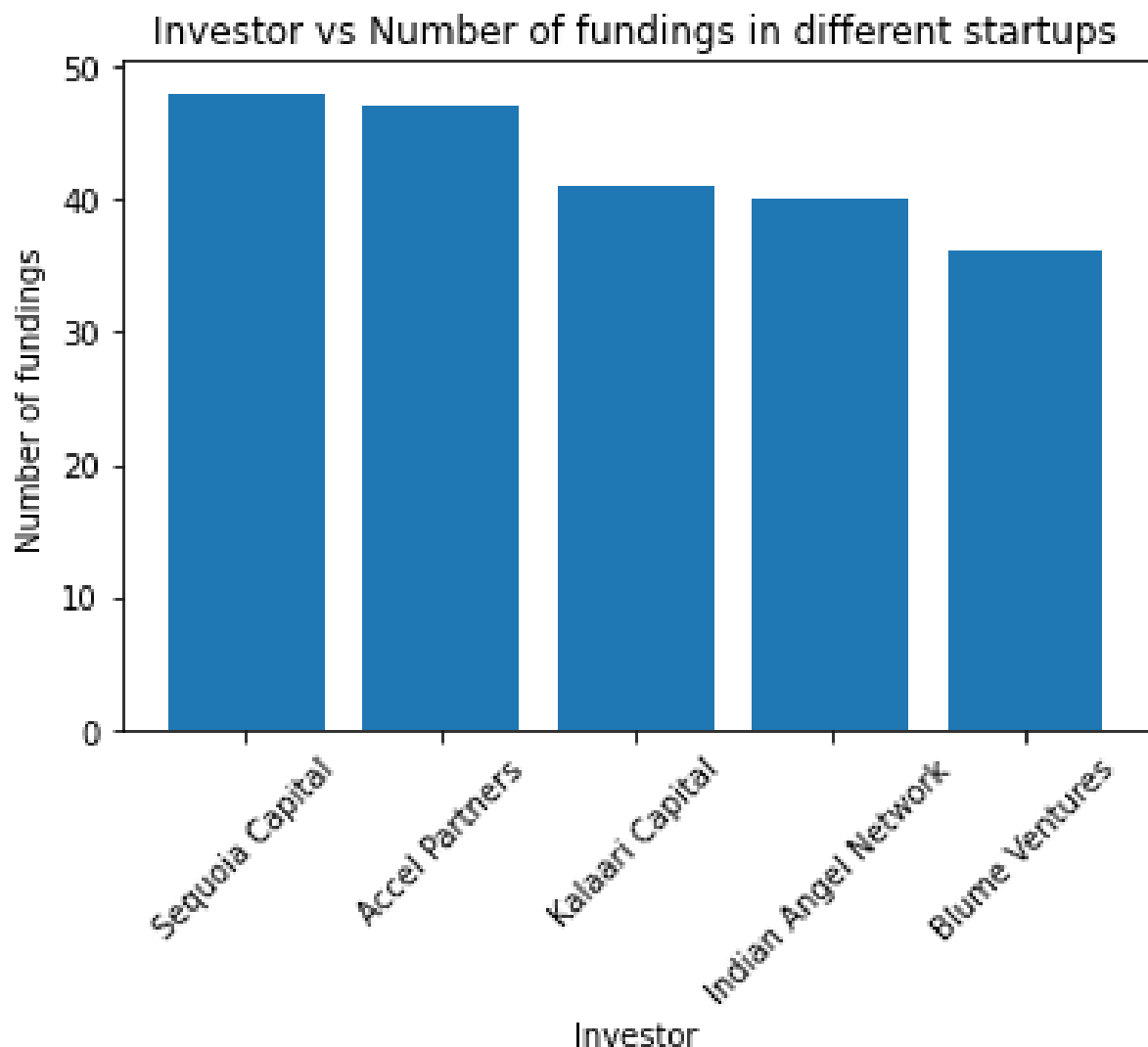


**Q3:** Firstly, I have imported the required modules. After reading data from the CSV file, I have applied data cleaning on the dataset, by removing all the rows where InvestorsName has null value using Boolean Indexing, after this I have reset the indexes, and initiated a dictionary. Then I traverse the whole data frame and keep track of the investors and the name of the start-up in each iteration. After this I clean the name of start-ups and take care of all the multiple investors in each round of funding. For every investor in the dictionary, I maintain a set of start-ups the investor has invested in. After this I alter the value corresponding to each key with the length of the set respective to each key. Then I sorted the dictionary in reverse order and convert the dictionary to two 1D NumPy array x, y and plotted a bar graph using it for the top 5 values.

These come out to be: **Sequoia Capital 48, Accel Partners 47, Kalaari Capital 41, Indian Angel Network 40, Blume Ventures 36.**

**Output:**

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



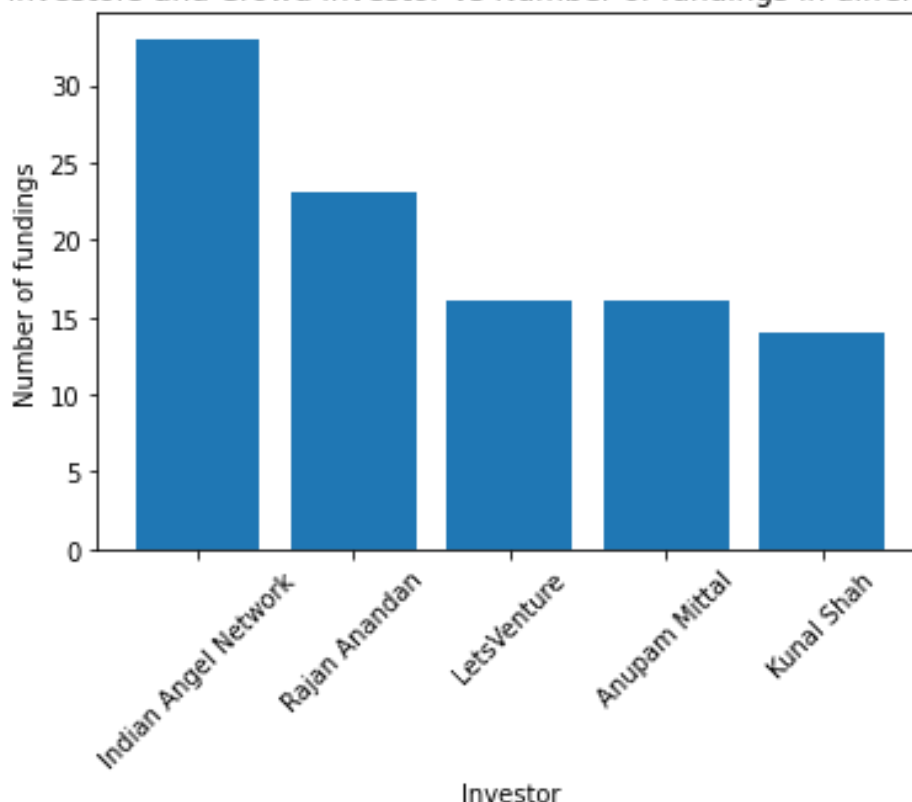
**Q4:** Firstly, I have imported the required modules. After reading data from the CSV file, I have applied data cleaning on the dataset, by providing consistency to the parameter InvestmentType, after I have kept only those records in the dataframe where InvestmentType is either “Seed Funding” or “Crowd Funding” and initiated a dictionary. After this I dropped the null values and reset the indexes. Then I traverse the whole data frame and keep track of the investors and the name of the start-up in each iteration. After this I clean the name of start-ups and take care of all the multiple investors in each round of funding. For every investor in the dictionary, I maintain a set of start-ups the investor has invested in. After this I alter the value corresponding to each key with the length of the set respective to each key. Then I manipulated the values of “Undisclosed Investors” and “Undisclosed investors” to 0. Then I sorted the dictionary in reverse order and convert the dictionary to two 1D NumPy array x, y and plotted a bar graph using it for the top 5 values.

These come out to be: **Indian Angel Network 33, Rajan Anandan 23, LetsVenture 16, Anupam Mittal 16, Kunal Shah 14.**

#### Output:

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

**Seed Investors and Crowd Investor vs Number of fundings in different startups**



**Q5:** Firstly, I have imported the required modules. After reading data from the CSV file, I have applied data cleaning on the dataset, by providing consistency to the parameter InvestmentType, after I have kept only those records in the dataframe where InvestmentType is "Private Equity" and initiated a dictionary. After this I dropped the null values and reset the indexes. Then I traverse the whole data frame and keep track of the investors and the name of the start-up in each iteration. After this I clean the name of start-ups and take care of all the multiple investors in each round of funding. For every investor in the dictionary, I maintain a set of start-ups the investor has invested in. After this I alter the value corresponding to each key with the length of the set respective to each key. Then I manipulated the values of "Undisclosed Investors" and "Undisclosed investors" to 0. Then I sorted the dictionary in reverse order and convert the dictionary to two 1D NumPy array x, y and plotted a bar graph using it for the top 5 values.

These come out to be: **Sequoia Capital 45, Accel Partners 43, Kalaari Capital 35, Blume Ventures 27, SAIF Partners 24.**

### Output

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

Private Equity Investor vs Number of fundings in different startups

