Project : Case Study (Part- II) – Justification Summary

1. Location with most investments

Location is stored in CityLocation column and below are corrections needed before applying any logic to it.

df['CityLocation'].replace('Delhi','NCR',inplace=True)
df['CityLocation'].replace('New Delhi','NCR',inplace=True)
df['CityLocation'].replace('Gurgaon','NCR',inplace=True)
df['CityLocation'].replace('Noida','NCR',inplace=True)
df['CityLocation'].replace('bangalore','Bangalore',inplace=True)

CityLocation cities are stored in a dictionary by iterating over data frame values for not null CityLocations.

Sorted the dictionary by key value and below is the graph generated for NCR, Bangalore and Mumbai area as requested in Question.

Ans : So based on the graph **NCR** is perfect location for setting up a startup.
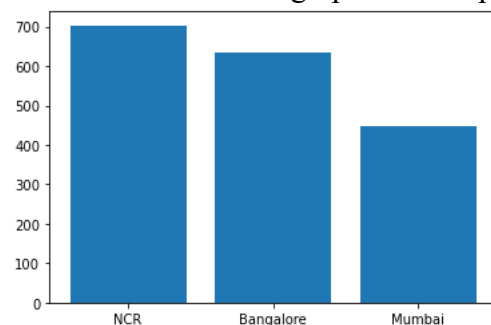

```
NCR 703
Bangalore 635
Mumbai 449
```

2. Top 5 Investors

Get to top 5 investors, first NaN are dropped by dropna function for InvestorsName column, then InvestorsName are splitted with ',' and stored in investor list.

The list is further analyzed by Pandas Series method to get top 5 Investors by count in multiple startup keeping repeat startup count as different.

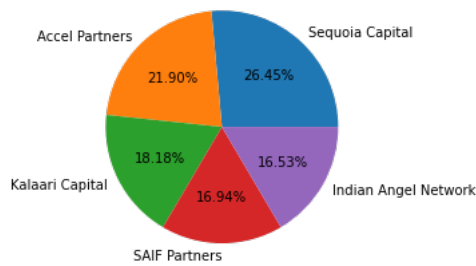Ans: Below are the top 5 investors and pie chart of their % of distribution

```
Sequoia Capital 64

Accel Partners 53

Kalaari Capital 44

SAIF Partners 41

Indian Angel Network 40
```

3. Top 5 investors who have invested in different number of startups

To get the Top 5 investors, keep multiple investments in same company investment as single, I have used set to store list of Startups against City so that it auto removes duplicates.

Below are the data corrections done as per question description:

df['StartupName'].replace('Flipkart.com','Flipkart',inplace=True)

df['StartupName'].replace('Ola Cabs','Ola',inplace=True)

df['StartupName'].replace('Olacabs','Ola',inplace=True)

df['StartupName'].replace('Oyo Rooms','Oyo',inplace=True)

df['StartupName'].replace('Paytm Marketplace','Paytm',inplace=True)

Undisclosed Investors are skipped by if logic so that it's not counted as per question description.

Ans:

```
[50, 47, 41, 40, 36]
['Sequoia Capital', 'Accel Partners', 'Kalaari Capital', 'Indian Angel Netw
ork', 'Blume Ventures']
```
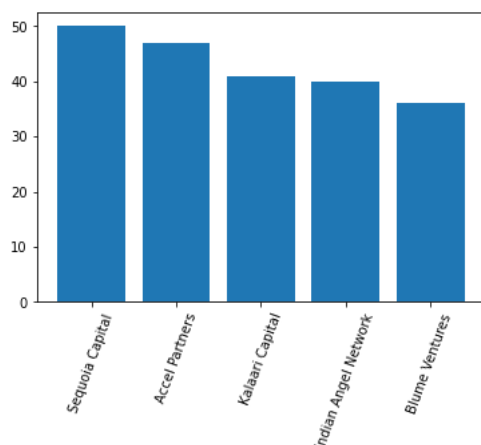

4. top 5 investors who have invested in a different number of startups and their investment type is Crowdfunding or Seed Funding

To get Top 5 investors in Crowd Funding and Seed Funding, first data corrections are done:

df['InvestmentType'].replace('SeedFunding','Seed Funding',inplace=True)

df['InvestmentType'].replace('Crowd funding','Crowd Funding',inplace=True)

df['InvestmentType'].replace('PrivateEquity','Private Equity',inplace=True)

NaN are dropped by dropna method:

df.dropna(subset=["InvestorsName"],inplace=True)
df.dropna(subset=["InvestmentType"],inplace=True)
and data frame is iterated only for Seed Funding and Crowd Funding by data frame filter.
Rest of the processing logic is same as problem3 by storing list of companies in set to take care of duplicates.
Ans:

```
[33, 23, 16, 16, 14]
['Indian Angel Network', 'Rajan Anandan', 'LetsVenture', 'Anupam Mittal', '
Kunal Shah']
```
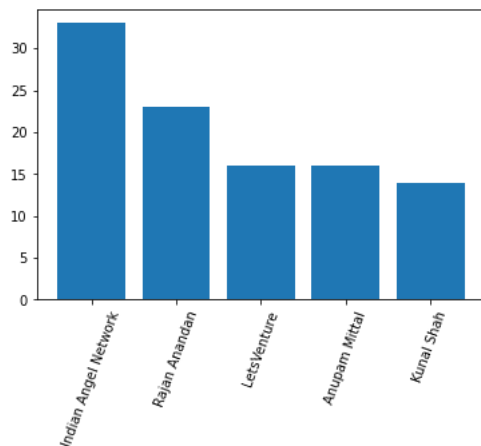

5. Find the top 5 investors who have invested in a different number of startups and their investment type is Private Equity

It's almost same as problem 4 except Investment Type is filtered only for Private Equity to find out the top5 Investors.

Ans:

```
[47, 43, 35, 27, 24]
['Sequoia Capital', 'Accel Partners', 'Kalaari Capital', 'Blume Ventures',
'SAIF Partners']
```