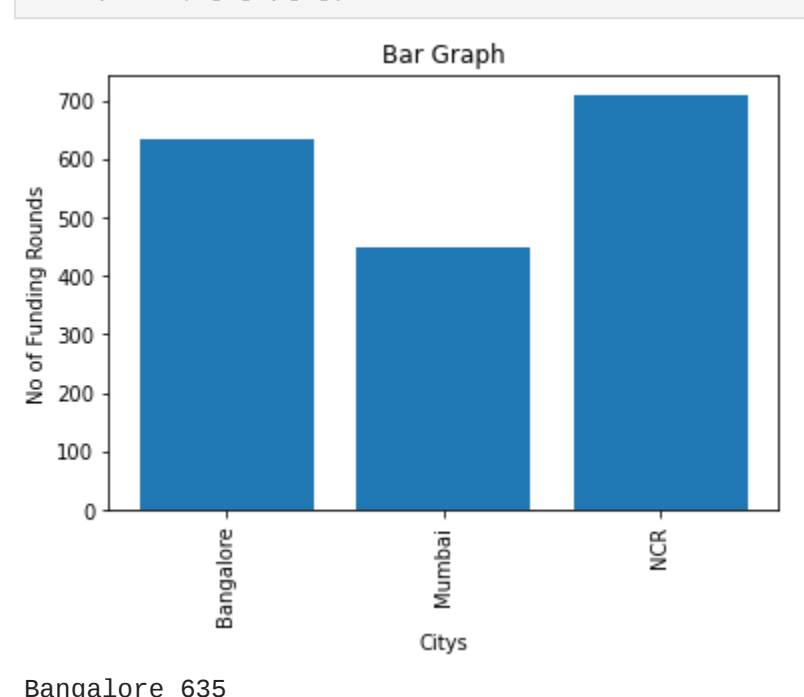
Question 1

In [15]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('startup_funding.csv')
df = data.copy()
df = df.dropna(subset = ["CityLocation"])     ## This will drop NaN values from the CityLocation column


df["CityLocation"] = df["CityLocation"].replace(['Delhi', 'bangalore','Pune / US','Bangalore / SFO','New Delhi / US','Bangalore/ Bangkok','Mumbai / NY','New Delhi/ Houston','Bangal
## Above part does the data cleaning where we correct the name of the cities.



city = df["CityLocation"]          ## Here we assigned the CityLocation column values to variable city.

citys = city.value_counts()[:10]   ## Here we did the value counts of the city where we can identify about the counts
                                   ## of the location where startups received fundings.

loc = citys.index
count = citys.values


##loca = ['Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon', 'Noida']
##counts = [635, 449, 389, 241, 79]




c = ['Bangalore', 'Mumbai', 'NCR']
p = [635, 449, 709]
## Plotting Graph
plt.bar(c, p)
plt.xticks(rotation = 90)
plt.xlabel('Citys')
plt.ylabel('No of Funding Rounds')
plt.title('Bar Graph')

plt.show()

for i in range(len(c)):
    print(c[i],p[i])
```




```
Bangalore 635
Mumbai 449
NCR 709
```

Approach

1. First we Opened the File and checked whether NaN values exist in the required column or not.
2. If yes, then we dropped NaN values.
3. Then we did Data Cleaning on the required cities and then we did value counts where we counted the locations where startup received fundings.
4. Then we printed the locations and then plotted the graph as required.

We have found the location where startup received maximum funding rounds and that is Bangalore.

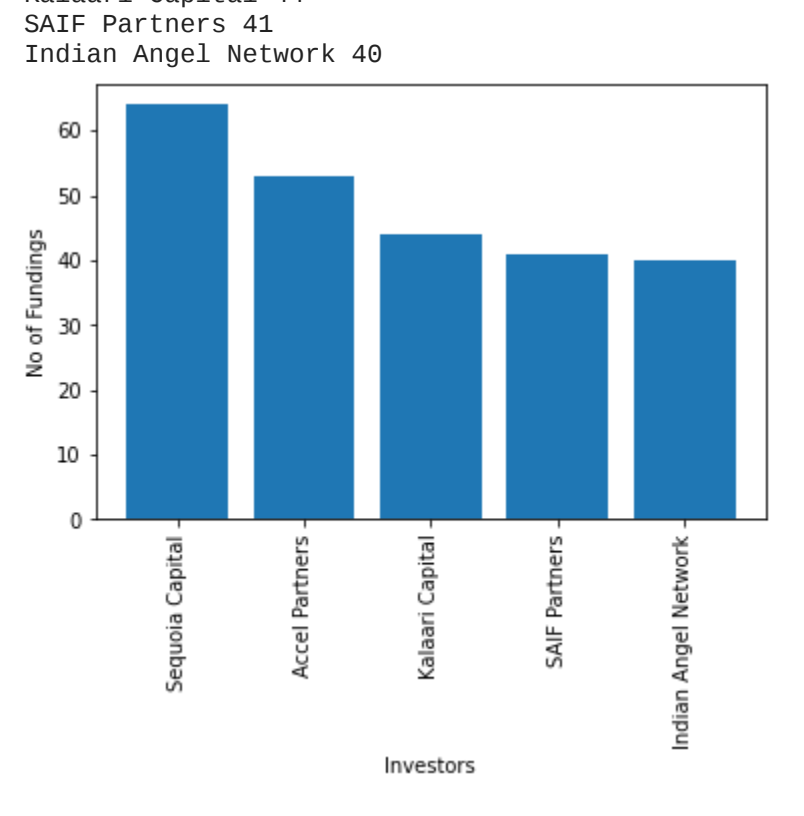
In [ ]:

In [ ]:

Question 2

In [16]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('startup_funding.csv')
df = data.copy()
a = df['InvestorsName'].str.split(',')  ## Here, the data was in the form of complete string and each name was
                                        ## splitted by comma inside the string, so we
                                        ## splitted the data by ',' and assigned it to variable a.
                                        ## So the data becomes in the form of string, and inside it,
                                        ## each string is seprated by comma.


arr = []                               ## Here we created an empty array.

a = list(a)                            ## To work on that column, we converted that value to list.


for i in a:                            ## Now, we Iterated over each elements of list a where we have
                                       ## have values of investors in the form of string.
                                       ## The slight problem I faced in this approach is that while converting
                                       ## a to list, the column 'InvestorsName' was not dropping NaN values.
                                       ## And the type of NaN value was surprisingly of type 'float'.
                                       ## This is maybe because i converted it to array or some other reasons.



    if type(i) != float:               ## So to not include it, we first check if string has multiple investor or
                                       ## single investor, to check that, we first check if type of 'i', i.e.
                                       ## individual element of a is Float or string, if it is not float, it means
                                       ## it is string with multiple investors, So we iterate over it and then we append
                                       ## it to array.

        for j in i:
            arr.append(j)
    else:                              ## If it is of type float, it means it is NaN so we also append it to the array.
        arr.append(i)


df2 = pd.DataFrame(arr)               ## Now, in the array, we have all the names of the investors separated,
                                      ## So we make dataframe of that array and then if we want, we can remove
                                      ## NaN values or we can not.


a = df2[0]
b = a.str.lstrip()                   ## As we created the dataframe, we saw the whitespaces on the first character of string.
                                     ## So here we removed the whitespaces from the first index of string.

c = b.value_counts()[:5]             ## Here we did value counts and here, we can manually discard the NaN count.

d = c.index
e = c.values
for i in range(len(e)):         ## Here we printed the investors name.
    print(d[i], e[i])



## Here we plotted the graph.
plt.bar(d,e)
plt.xticks(rotation = 90)
plt.xlabel("Investors")
plt.ylabel('No of Fundings')
plt.show()
```

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



Approach

1. First i opened the file.
2. Then Here, the data was in the form of complete string and each name was splitted by comma inside the string, so we splitted the data by ',' and assigned it to variable a.
3. Then we created empty array
4. To work on column a, we converted it into list.
5. Now, we Iterated over each elements of list a where we have have values of investors in the form of string. The slight problem I faced in this approach is that while converting a to list, the column 'InvestorsName' was not dropping NaN values. And the type of NaN value was surprisingly of type 'float'. This is maybe because i converted it to array or some other reasons.
6. So to not include it, we first check if string has multiple investor or single investor, to check that, we first check if type of 'i', i.e. individual element of a is Float or string, if it is not float, it means it is string with multiple investors, So we iterate over it and then we append it to array. If it is of type float, it means it is NaN so we also append it to the array.
7. Now, in the array, we have all the names of the investors separated, So we make dataframe of that array and then if we want, we can remove NaN values or we can not. 8. As we created the dataframe, we saw the whitespaces on the first character of string. So we removed the whitespaces from the first index of string.
8. Then we did value counts and now, we can manually discard the NaN count.
9. Then we print the investors name from value counts.

We found the top 5 investors and they are
Sequoia Capital
Accel Partners
Kalaari Capital
SAIF Partners
Indian Angel Network

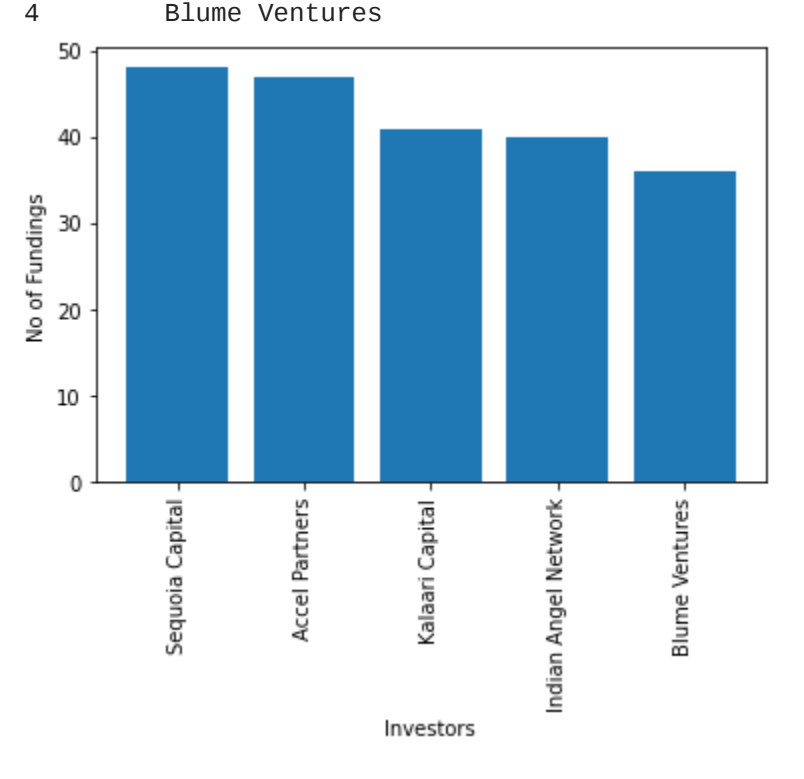
In [ ]:

In [ ]:

Question 3

In [17]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('startup_funding.csv')
df = data.copy()
df["StartupName"].replace("Olacabs","Ola",inplace=True)
df["StartupName"].replace("Ola Cabs","Ola",inplace=True)
df["StartupName"].replace("Flipkart.com","Flipkart",inplace=True)
df["StartupName"].replace("Paytm Marketplace","Paytm",inplace=True)
df["StartupName"].replace("Oyo Rooms","Oyo",inplace=True)
df["StartupName"].replace("Oyorooms","Oyo",inplace=True)
df["StartupName"].replace("OyoRooms","Oyo",inplace=True)
df["StartupName"].replace("OYo Rooms","Oyo",inplace=True)
df["StartupName"].replace("OYO Rooms","Oyo",inplace=True)


# Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
# Dropping NA values

df = df[["StartupName","InvestorsName"]]
df.dropna(inplace=True)

# Creating 2 list for StartupName and InvestorsName

startup_name=list(df["StartupName"].values)
investor_name=list(df["InvestorsName"].values)

# Traversing in the investor_name list , splitting by "," and applying the strip()
i=0
while i<len(investor_name):
    if ',' in investor_name[i] :
        l1 = investor_name[i].split(',')
        j = i
        investor_name[j] = l1[0]
        j = j + 1
        for k in range(1,len(l1)):
            investor_name.insert(j,l1[k].strip())
            startup_name.insert(j,startup_name[i])
            j = j + 1
        i = i + len(l1) - 1
    i = i + 1

# Creating a new list l2 and appending the investor name with their invested startup name
# Creating list with only unique values by set() func.

l2 = []
for i in range(len(startup_name)):
    l2.append(investor_name[i]+"with"+startup_name[i])
l2 = list(set(l2))

# Creating a seperate list for the investor name

l3 = []
for i in l2:
    i = i.split('with')
    l3.append(i[0].strip())

# Creating a new series with list of investor name
# Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name
# Dropping the empty values and printing the top5 investors name with their fundings

df_new = pd.Series(l3)
x = df_new.value_counts().sort_values(ascending=False)
x.drop("",inplace=True)
x = x[0:5]
df = pd.DataFrame(x.index,columns=["Top 5 Investors"])
df["No. of Investment in Different Startup"] = x.values
print(df)


## Extracting name here for plotting graph

x1 = ['Sequoia Capital', 'Accel Partners' ,'Kalaari Capital', 'Indian Angel Network', 'Blume Ventures']
y1 = [48, 47, 41, 40, 36]

plt.bar(x1, y1)
plt.xticks(rotation = 90)
plt.xlabel("Investors")
plt.ylabel('No of Fundings')
plt.show()
```

```
       Top 5 Investors  No. of Investment in Different Startup
0      Sequoia Capital                                      48
1       Accel Partners                                      47
2      Kalaari Capital                                      41
3 Indian Angel Network                                      40
4       Blume Ventures                                      36
```



Approach

1. Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
2. Then we Drop NA values.
3. Then we create 2 list for StartupName and InvestorsName from which we extract the investors name.
4. Traversing in the investor_name list , splitting by "," and applying the strip() so that we can get the correct investors name.
5. Creating a new list l2 and appending the investor name with their invested startup name
6. Creating list with only unique values by set() func. This will create a unique values of investor name for particular startup.
7. Creating a seperate list for the investor name.
8. Creating a new series with list of investor name. Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name Dropping the empty values and printing the top5 investors name with their fundings.
9. Then we extract values for plotting graph.

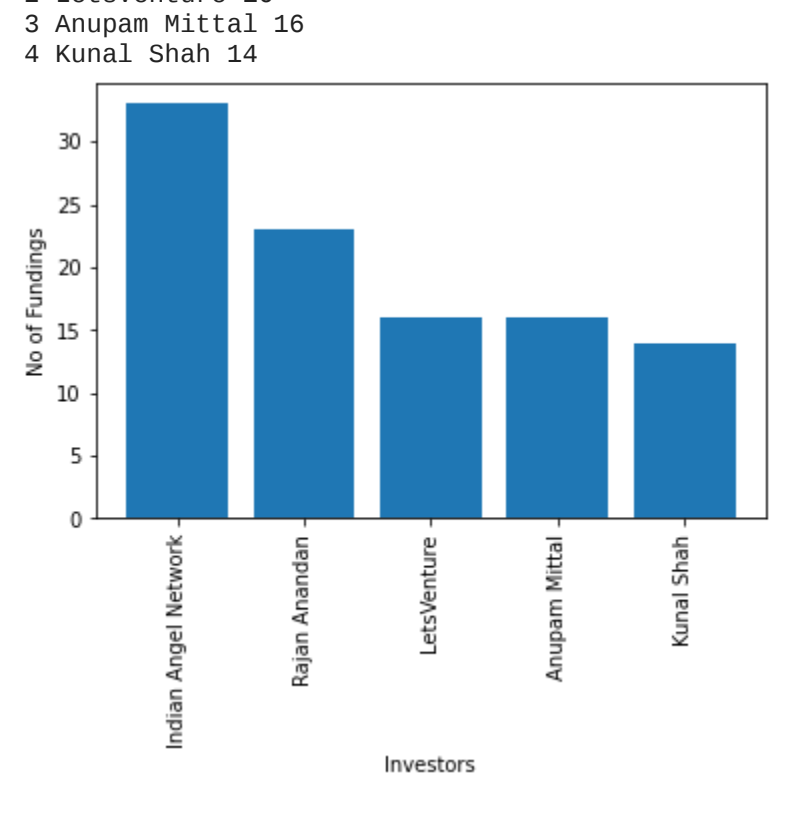
In [ ]:

In [ ]:

Question 4

In [18]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('startup_funding.csv')
df = data.copy()
df["StartupName"].replace("Olacabs","Ola",inplace=True)
df["StartupName"].replace("Ola Cabs","Ola",inplace=True)
df["StartupName"].replace("Flipkart.com","Flipkart",inplace=True)
df["StartupName"].replace("Paytm Marketplace","Paytm",inplace=True)
df["StartupName"].replace("Oyo Rooms","Oyo",inplace=True)
df["StartupName"].replace("Oyorooms","Oyo",inplace=True)
df["StartupName"].replace("OyoRooms","Oyo",inplace=True)
df["StartupName"].replace("OYo Rooms","Oyo",inplace=True)
df["StartupName"].replace("OYO Rooms","Oyo",inplace=True)


# Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
# Dropping NA values

df = df[["StartupName","InvestorsName", 'InvestmentType']]
df.dropna(inplace=True)

## We create dataframe where investmenttype is seed funding and crowd funding.
ab = df[(df["InvestmentType"] == 'Seed Funding') | (df['InvestmentType']== 'Crowd Funding')]


# Creating 2 list for StartupName and InvestorsName

startup_name=list(ab["StartupName"].values)
investor_name=list(ab["InvestorsName"].values)

# Traversing in the investor_name list , splitting by "," and applying the strip()
i=0
while i<len(investor_name):
    if ',' in investor_name[i] :
        l1 = investor_name[i].split(',')
        j = i
        investor_name[j] = l1[0]
        j = j + 1
        for k in range(1,len(l1)):
            investor_name.insert(j,l1[k].strip())
            startup_name.insert(j,startup_name[i])
            j = j + 1
        i = i + len(l1) - 1
    i = i + 1

# Creating a new list l2 and appending the investor name with their invested startup name
# Creating list with only unique values by set() func.

l2 = []
for i in range(len(startup_name)):
    l2.append(investor_name[i]+"with"+startup_name[i])
l2 = list(set(l2))

# Creating a seperate list for the investor name

l3 = []
for i in l2:
    i = i.split('with')
    l3.append(i[0].strip())

# Creating a new series with list of investor name
# Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name
# Dropping the empty values and printing the top5 investors name with their fundings

df_new = pd.Series(l3)
x = df_new.value_counts().sort_values(ascending=False)
x.drop("",inplace=True)
x = x[0:10]
df = pd.DataFrame(x.index,columns=["Top 5 Investors"])
df["No. of Investment in Different Startup"] = x.values




x1 = ['Indian Angel Network', 'Rajan Anandan', 'LetsVenture', 'Anupam Mittal', 'Kunal Shah']
y1 = [33, 23, 16, 16, 14]

print("Top 5 Investors  No. of Investment in Different Startup")
for i in range(len(y1)):
    print(i,x1[i],y1[i])

plt.bar(x1, y1)
plt.xticks(rotation = 90)
plt.xlabel('Investors')
plt.ylabel('No of Fundings')
plt.show()
```

```
Top 5 Investors  No. of Investment in Different Startup
0 Indian Angel Network 33
1 Rajan Anandan 23
2 LetsVenture 16
3 Anupam Mittal 16
4 Kunal Shah 14
```



Approach

1. Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
2. Then we Drop NA values. Then We create dataframe where investmenttype is seed funding and crowd funding.

1. Then we create 2 list for StartupName and InvestorsName from which we extract the investors name.
2. Traversing in the investor_name list , splitting by "," and applying the strip() so that we can get the correct investors name.
3. Creating a new list l2 and appending the investor name with their invested startup name
4. Creating list with only unique values by set() func. This will create a unique values of investor name for particular startup.
5. Creating a seperate list for the investor name.
6. Creating a new series with list of investor name. Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name Dropping the empty values and printing the top5 investors name with their fundings.
7. Then we extract values for plotting graph.

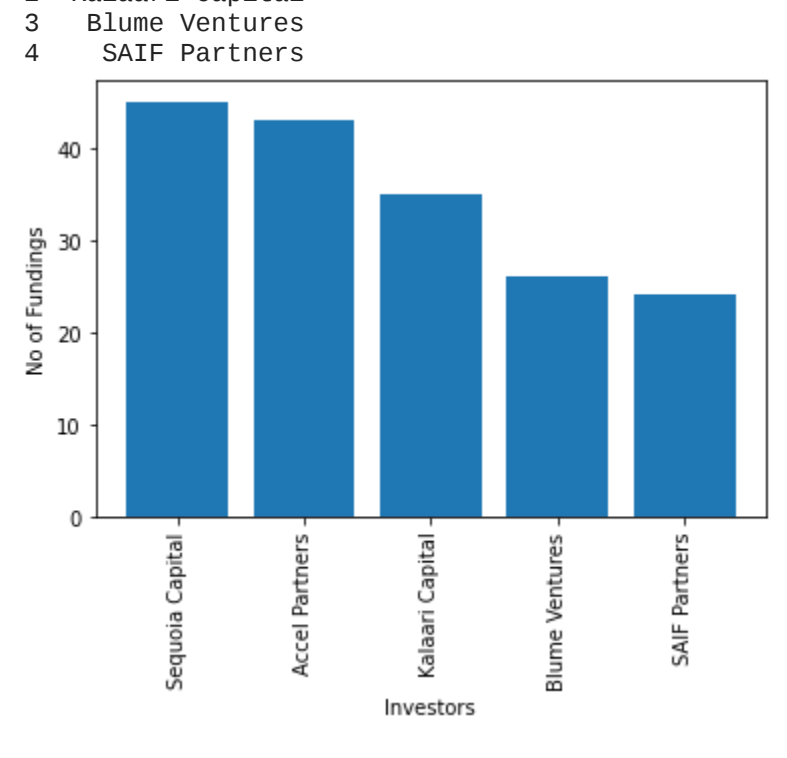
In [ ]:

In [ ]:

Question 5

In [19]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('startup_funding.csv')
df = data.copy()
df["StartupName"].replace("Olacabs","Ola",inplace=True)
df["StartupName"].replace("Ola Cabs","Ola",inplace=True)
df["StartupName"].replace("Flipkart.com","Flipkart",inplace=True)
df["StartupName"].replace("Paytm Marketplace","Paytm",inplace=True)
df["StartupName"].replace("Oyo Rooms","Oyo",inplace=True)
df["StartupName"].replace("Oyorooms","Oyo",inplace=True)
df["StartupName"].replace("OyoRooms","Oyo",inplace=True)
df["StartupName"].replace("OYo Rooms","Oyo",inplace=True)
df["StartupName"].replace("OYO Rooms","Oyo",inplace=True)


# Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
# Dropping NA values

df = df[["StartupName","InvestorsName", 'InvestmentType']]
df.dropna(inplace=True)
ab = df[df["InvestmentType"] == 'Private Equity']


# Creating 2 list for StartupName and InvestorsName

startup_name=list(ab["StartupName"].values)
investor_name=list(ab["InvestorsName"].values)

# Traversing in the investor_name list , splitting by "," and applying the strip()
i=0
while i<len(investor_name):
    if ',' in investor_name[i] :
        l1 = investor_name[i].split(',')
        j = i
        investor_name[j] = l1[0]
        j = j + 1
        for k in range(1,len(l1)):
            investor_name.insert(j,l1[k].strip())
            startup_name.insert(j,startup_name[i])
            j = j + 1
        i = i + len(l1) - 1
    i = i + 1

# Creating a new list l2 and appending the investor name with their invested startup name
# Creating list with only unique values by set() func.

l2 = []
for i in range(len(startup_name)):
    l2.append(investor_name[i]+"with"+startup_name[i])
l2 = list(set(l2))

# Creating a seperate list for the investor name

l3 = []
for i in l2:
    i = i.split('with')
    l3.append(i[0].strip())

# Creating a new series with list of investor name
# Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name
# Dropping the empty values and printing the top5 investors name with their fundings

df_new = pd.Series(l3)
x = df_new.value_counts().sort_values(ascending=False)
x.drop("",inplace=True)
x = x[0:5]
df = pd.DataFrame(x.index,columns=["Top 5 Investors"])
df["No. of Investment in Different Startup"] = x.values
print(df)

## Extracting the names for plotting graph manually

x1 = ['Sequoia Capital', 'Accel Partners', 'Kalaari Capital', 'Blume Ventures', 'SAIF Partners']
y1 = [45, 43, 35, 26, 24]



plt.bar(x1, y1)
plt.xticks(rotation = 90)
plt.xlabel('Investors')
plt.ylabel('No of Fundings')
plt.show()
```

```
   Top 5 Investors  No. of Investment in Different Startup
0  Sequoia Capital                                      45
1   Accel Partners                                      43
2  Kalaari Capital                                      35
3   Blume Ventures                                      26
4    SAIF Partners                                      24
```



Approach

1. Creating data frame consisting of only 2 columns "StartupName" and "InvestorsName"
2. Then we Drop NA values. Then We create dataframe where investmenttype is Private Equity.

1. Then we create 2 list for StartupName and InvestorsName from which we extract the investors name.
2. Traversing in the investor_name list , splitting by "," and applying the strip() so that we can get the correct investors name.
3. Creating a new list l2 and appending the investor name with their invested startup name
4. Creating list with only unique values by set() func. This will create a unique values of investor name for particular startup.
5. Creating a seperate list for the investor name.
6. Creating a new series with list of investor name. Value_counts().sort_values(ascending=False) gives the count of no. of fundings with their investor name Dropping the empty values and printing the top5 investors name with their fundings.
7. Then we extract values for plotting graph.