Project: Case Study-II

Answer 1:- In my opinion he start their startup in Bangalore because We find maximum number of funding in Bangalore which is 635.

- In code part first we handle NA value using dropna then we get a new city location without NA.

```
df_start=pd.read_csv('E:\startup_funding.csv',encoding='utf-8')
df_start['CityLocation'].dropna(inplace=True) # To drop the NA values
```

- Then we apply a function which helps to split city to get only Indian cities then we correct cities name using replace function.

```
def separateCity(city):              # Function for finding Indian city
    return city.split('/')[0].strip()
df_start['CityLocation']=df_start['CityLocation'].apply(separateCity)
df_start['CityLocation'].replace("Delhi","New Delhi",inplace=True)   # For replacing city name
df_start['CityLocation'].replace("bangalore","Bangalore",inplace=True) #For rplacing city name
city_number=df_start['CityLocation'].value_counts() # To count the no. of fundings
```

- And then we apply count function to find number of funding in particular cities.

```
city_number=df_start['CityLocation'].value_counts() # To count the no. of fundings
```

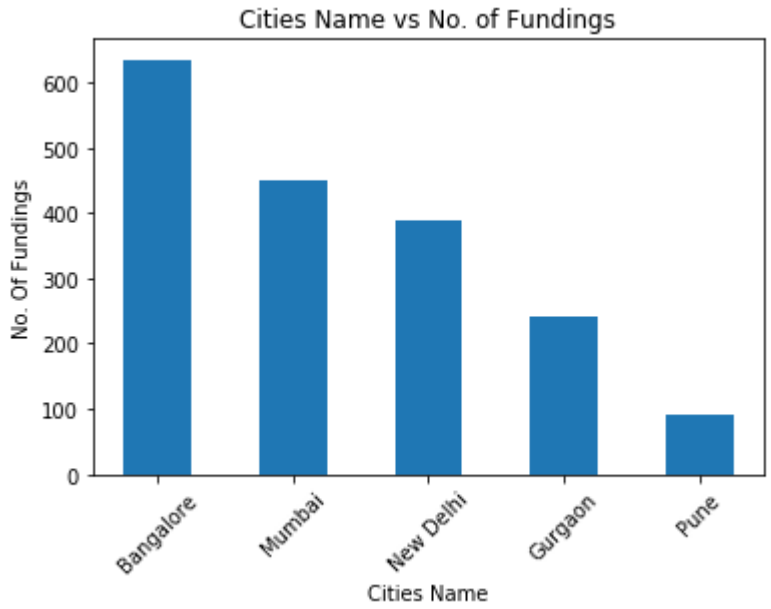
- Then we apply slicing to get top 5 cities having maximum number of funding. And then print top most cities having maximum funding.

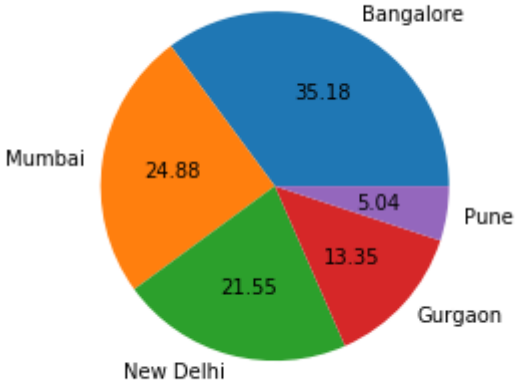
```
city=city_number.index[0:5]
numCity=city_number.values[0:5]
print(city[0],numCity[0])
```

- In next step we plot a bar graph and a pie chart using city name vs Number of funds.

```
#graph plot
plt.bar(city,numCity,width=0.5)

plt.xlabel("Cities Name")
plt.ylabel("No. Of Fundings")
plt.title("Cities Name vs No. of Fundings")
plt.xticks(rotation = 45)
plt.show()
plt.pie(numCity,labels = city,autopct = "%.2f")
plt.show()
```


Cities Name vs No. of Fundings
No. Of Fundings
0
100
200
300
400
500
600
Bangalore
Mumbai
New Delhi
Gurgaon
Pune
Cities Name


Bangalore
35.18
Mumbai
24.88
5.04
Pune
13.35
21.55
Gurgaon
New Delhi

Answer 2:- In my opinion, He will try to invest with Sequoia Capital , Accel partners, Kalaari Capital, SAIF Partners and Indian Angel Network because they are top 5 investor who invested maximum number of time.

- In code part first we add NA in investors name using fillna function and then correct startup names.

```
df.InvestorsName.fillna("",inplace= True)
#correcting names
for i in range(len(df)):
    if df.StartupName[i].find("Flipkart") != -1:
        df.at[i, 'StartupName'] = "Flipkart"
    if df.StartupName[i].find("Ola") != -1:
        df.at[i, 'StartupName'] = "Ola"
    if df.StartupName[i].find("Oyo") != -1:
        df.at[i, 'StartupName'] = "Oyo"
    if df.StartupName[i].find("OYO") != -1:
        df.at[i, 'StartupName'] = "Oyo"
    if df.StartupName[i].find("Paytm") != -1:
        df.at[i, 'StartupName'] = "Paytm"
```

- And then making a dictionary to get the investor name and the count of investments.

```
d = {} #dictionary to get the investorname and the count of investments
for i in range(len(df)):
    if df.InvestorsName[i] != "":
        a = df.InvestorsName[i].split(",")
        for j in a:
            j = j.strip()
            if j in d:
                d[j] = d[j] + 1
            else:
                d[j] = 1
```

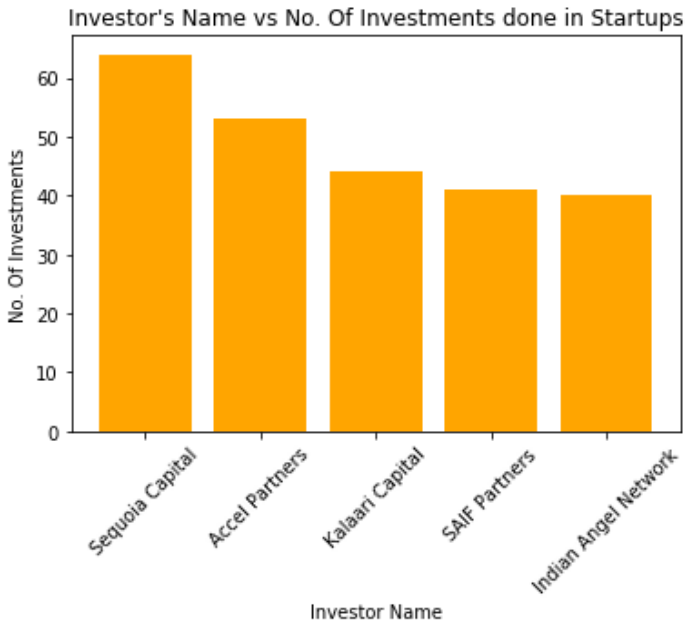
- Then we sort the dictionary to get top 5 investors name and then append their name and count of investments into two different list which helps in to plot the graph.

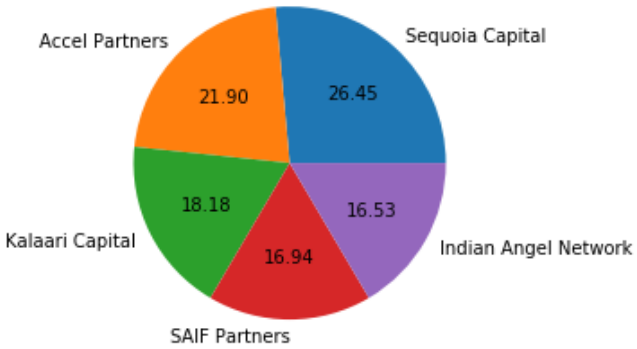
```
sorted_d=sorted_d = sorted(d.items(), key=lambda kv: kv[1],reverse = True)[:5] #will give the sorted list
x1=[]                                                                          # at 0th position and coun
x2=[]
for i in sorted_d:
    print(i[0],i[1])
    x1.append(i[0]) #getting names
    x2.append(i[1]) #getting counts
```

- Then next step to plot a bar graph and a pie chart using investors name vs Number of investments.

```
#graph plot
plt.bar(x1,x2,color = "Orange")

plt.xlabel("Investor Name")
plt.ylabel("No. Of Investments")
plt.title("Investor's Name vs No. Of Investments done in Startups")
plt.xticks(rotation = 45)
plt.show()
plt.pie(x2,labels = x1,autopct = "%.2f")
plt.show()
```





Answer 3:- In my Opinion, he will try to invest with Sequoia Capital, Accel Partners, Kalaari Capital, Indian Angel Network and Blume Venture because they invested in maximum number of times in different companies.

- In code part first we fill NA in investors name using fillna function and then change startup name according to spelling correct.

```
df.InvestorsName.fillna("",inplace= True)

for i in range(len(df)):
    if df.StartupName[i].find("Flipkart") != -1:
        df.at[i, 'StartupName'] = "Flipkart"
    if df.StartupName[i].find("Ola") != -1:
        df.at[i, 'StartupName'] = "Ola"
    if df.StartupName[i].find("Oyo") != -1:
        df.at[i, 'StartupName'] = "Oyo"
    if df.StartupName[i].find("OYO") != -1:
        df.at[i, 'StartupName'] = "Oyo"
    if df.StartupName[i].find("Paytm") != -1:
        df.at[i, 'StartupName'] = "Paytm"
```

- Then we make dictionary for counting number of investment round as per the startup name and the investor name. Then we find investor name present in the in dictionary and split on basis of coma then append it into a list which is list of investors.

```
d = {} #dictionary for counting no. of investment round as per the startup name and the investor name
for i in range(len(df)):
    if df.InvestorsName[i] != "":
        a = df.InvestorsName[i].split(",")
        for j in a:
            j = j.strip()
            if (df.StartupName[i],j) in d:
                d[(df.StartupName[i],j)] = d[(df.StartupName[i],j)] + 1
            else:
                d[(df.StartupName[i],j)] = 1
```

- Make a list of investors and append the dictionary value in it to get investors.

```
x = [] #List of investors
for i in d:
    x.append(i[1])
```

- Then we make another dictionary for finding investors without repetitive investment done by the investor.

```
dic = {} # for finding investors without repetitive invesment done by the investor
for i in x:
    if i in dic:
        dic[i] = dic[i] + 1
    else:
        dic[i] = 1
dic[""] = 0
```

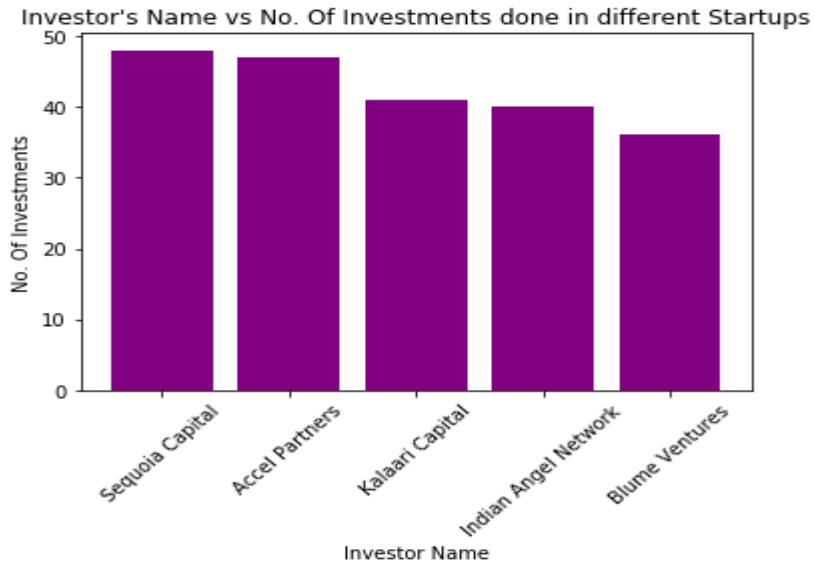
- Then to find the top 5 investor name we apply sort function in descending order.

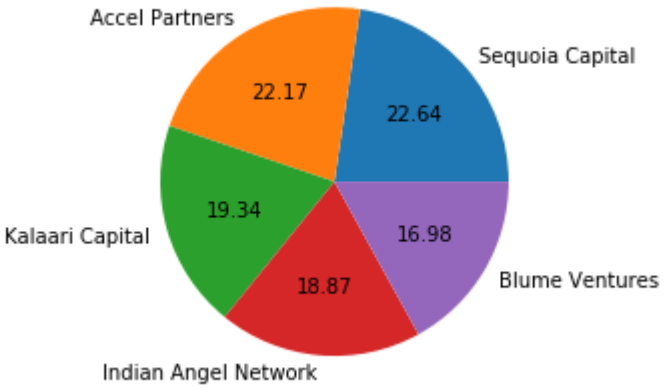
```
sorted_d = sorted(dic.items(), key=lambda kv: kv[1],reverse = True) # sorted list in descen
for i in sorted_d[:5]:
    print(i[0],i[1])
    x1.append(i[0])
    x2.append(i[1])
```

- Finally, plot a bar graph and a pie graph on the basis of investor name and Number of investment.

```
#graph plot
plt.bar(x1,x2,color = "purple")

plt.xlabel("Investor Name")
plt.ylabel("No. Of Investments")
plt.title("Investor's Name vs No. Of Investments done in different Startups")
plt.xticks(rotation = 45)
plt.show()
plt.pie(x2,labels = x1,autopct = "%.2f")
plt.show()
```




Accel Partners
22.17
Sequoia Capital
22.64
Kalaari Capital
19.34
16.98
Blume Ventures
18.87
Indian Angel Network

Answer 4:- As we go my investor friend to understand the situation better it will explain about investment type and features so that I conclude that the top 5 investors Indian Angel Network, Rajan Anandan, LetsVenture, Anupam Mittal,Kunal Shah are best for investment because their investment types are Seed Funding and Crowd Funding. Which is best type for an early stage startup.

- In coding part first we correct the spelling of investment types using replace function and then fill NA to handling na in funding column using fillna function.

```
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)
df['AmountInUSD'].fillna('0',inplace = True) # nan handling of funding column
```

- Then make two dictionary fromkeys which replace in startup names using replace fun ction and then for unknown investor names replaced by '0' using fillna function.

```
x = dict.fromkeys(['OYO Rooms', 'Oyo Rooms','OyoRooms','Oyorooms'], 'Oyo')
y = dict.fromkeys(['Ola Cabs','Olacabs'],'Ola')
df['StartupName']=df['StartupName'].replace(x)
df['StartupName']=df['StartupName'].replace(y)
df['StartupName']=df['StartupName'].replace({"Flipkart.com":"Flipkart"})
df['StartupName']=df['StartupName'].replace({'Paytm Marketplace':'Paytm'})
df['InvestorsName'].fillna('0',inplace = True) # unknown investor names replaced by '0'
```

- And then make a dictionary which uses investors as key for which which value is an other dictionary with startup names as keys and number of investments made in that startup as value.

  For ex.- 'Sequoia Capital': {'Ola':5, 'Flipkart':2} could be a key value pair in d(so Sequ oia would have invested 5 times in Ola and twice in Flipkart)

  After this trailing comma caused ' ' to appear as investor name so it had to be removed using rstrip function and then apply a condition for ignoring undisclosed investors.

```
d = {}

for _,row in df.iterrows():
    x = row['InvestorsName'].strip().rstrip(',')
    itype = row['InvestmentType']
    if itype=='Seed Funding' or itype=='Crowd Funding':
        startup = row['StartupName'].strip()
        names = x.split(',')
        for name in names:
            if 'undisclosed' in name.lower(): #ignoring undisclosed investors
                break
            else:
                name = name.strip()
                if name in d:
                    if startup in d[name]:
                        d[name][startup]+=1
                    else:
                        d[name][startup]=1
                else:
                    d[name]={}
                    d[name][startup]=1
```

- Then we make another dictionary to counts number of startups an investor has invested in, as is required by question.
  For ex.- 'Sequoia Capital': 2 could be a key value pair in d2 corresponding to example key-value pair in d i.e. Sequoia Capital invested in 2 startups (ola and flipkart)

  And then we apply sorted function to sort the dictionary in descending order of count and finally print the investor and their count.

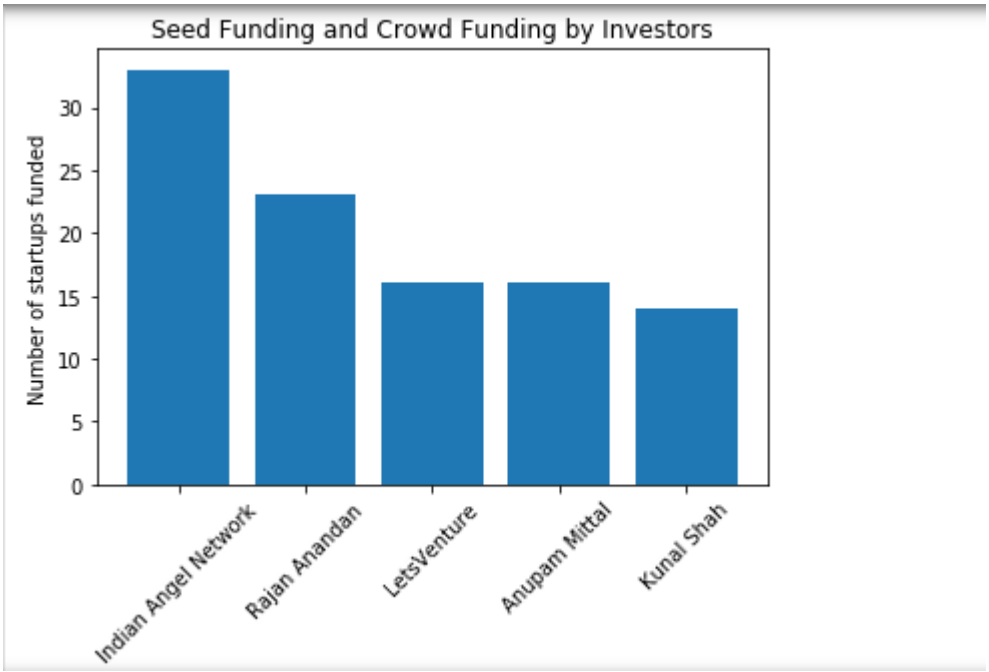
```
for investor in d:
    d2[investor] = len(d[investor])

d2 = sorted(d2.items(), key=lambda x: x[1], reverse=True)[:5]
# sorted in descending order of count

investors = [t[0] for t in d2]
counts = [t[1] for t in d2]
for i in range(len(d2)):
    print(investors[i],counts[i])
```

- At last we plot a bar graph between investor and Number of startups funded.

```
plt.bar(investors,counts)
plt.xticks(rotation = 45, horizontalalignment = 'center')
plt.title('Seed Funding and Crowd Funding by Investors')
plt.xlabel('Investor')
plt.ylabel('Number of startups funded')
plt.show()
```



Answer 5:- As we go my investor friend to understand the situation better it will explain abou t investment type and features, now I conclude that if he wants to expand their startup it will i nvest with these top 5 investors are Sequoia Capital,Accel Partners, Kalaari Capital, Blume V entures, SAIF Partners because their investment type is Private Equity.

- In coding part first we correct the spelling of investment types using replace fun ction and then fill NA to handling na in funding column using fillna function.

```
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)
df['AmountInUSD'].fillna('0',inplace = True) # nan handling of funding column
```

- Then make two dictionary fromkeys which replace in startup names using replace fun ction and then for unknown investor names replaced by '0' using fillna function.

```
x = dict.fromkeys(['OYO Rooms', 'Oyo Rooms','OyoRooms','Oyorooms'], 'Oyo')
y = dict.fromkeys(['Ola Cabs','Olacabs'],'Ola')
df['StartupName']=df['StartupName'].replace(x)
df['StartupName']=df['StartupName'].replace(y)
df['StartupName']=df['StartupName'].replace({"Flipkart.com":"Flipkart"})
df['StartupName']=df['StartupName'].replace({'Paytm Marketplace':'Paytm'})
df['InvestorsName'].fillna('0',inplace = True) # unknown investor names replaced by '0'
```

- And then make a dictionary which uses investors as key for which which value is anot her dictionary with startup names as keys and number of investments made in that startup as value.

  For ex.- 'Sequoia Capital': {'Ola':5, 'Flipkart':2} could be a key value pair in d(so Sequ oia would have invested 5 times in Ola and twice in Flipkart)

  After this trailing comma caused ' ' to appear as investor name so it had to be removed using rstrip function and then apply a condition for ignoring undisclosed investors.

```
d = {}

for _,row in df.iterrows():
    x = row['InvestorsName'].strip().rstrip(',')
    itype = row['InvestmentType']
    if itype=='Seed Funding' or itype=='Crowd Funding':
        startup = row['StartupName'].strip()
        names = x.split(',')
        for name in names:
            if 'undisclosed' in name.lower(): #ignoring undisclosed investors
                break
            else:
                name = name.strip()
                if name in d:
                    if startup in d[name]:
                        d[name][startup]+=1
                    else:
                        d[name][startup]=1
                else:
                    d[name]={}
                    d[name][startup]=1
```

- Then we make another dictionary to counts number of startups an investor has investedin, as is required by question.
  For ex.- 'Sequoia Capital': 2 could be a key value pair in d2 corresponding to example key-value pair in d i.e. Sequoia Capital invested in 2 startups (ola and flipkart)

  And then we apply sorted function to sort the dictionary in descending order of count and finally print the investor and their count.

```
for investor in d:
    d2[investor] = len(d[investor])

d2 = sorted(d2.items(), key=lambda x: x[1], reverse=True)[:5]
# sorted in descending order of count

investors = [t[0] for t in d2]
counts = [t[1] for t in d2]
for i in range(len(d2)):
    print(investors[i],counts[i])
```

- At last we plot a bar graph between investor and Number of startups funded.

```
plt.bar(investors,counts)
plt.xticks(rotation = 45, horizontalalignment = 'center')
plt.title('Seed Funding and Crowd Funding by Investors')
plt.xlabel('Investor')
plt.ylabel('Number of startups funded')
plt.show()
```