Q1.) Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, you want to help your friend deciding the location. NCR include Gurgaon, Noida and New Delhi. Find the location where the most number of funding is done. That means, find the location where startups have received funding maximum number of times. Plot the bar graph between location and number of funding. Take city name "Delhi" as "New Delhi". Check the case-sensitiveness of cities also. That means, at some place instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore". For few startups multiple locations are given, one Indian and one Foreign. Consider the startup if any one of the city lies in given locations.

CODE -

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

df = pd.read_csv('startup_funding.csv')

df.loc[df['InvestorsName'].isnull(),'InvestorsName'] = None

cities = ['Bangalore','Mumbai','Gurgaon','Noida','New Delhi']

counts = [0 for i in range(len(cities))]

for i in range(len(cities)):
    counts[i] = len(df[df['CityLocation'].str.contains('^'+cities[i],regex=True,na=False,case=False)])

counts[4] = len(df[df['CityLocation'].str.contains('Delhi',regex=False,na=False,case=False)])


plt.bar(cities,counts)
plt.grid()
plt.xlabel('City')
plt.ylabel('Number of Fundings')
plt.show()
```

Justification of the CODE –

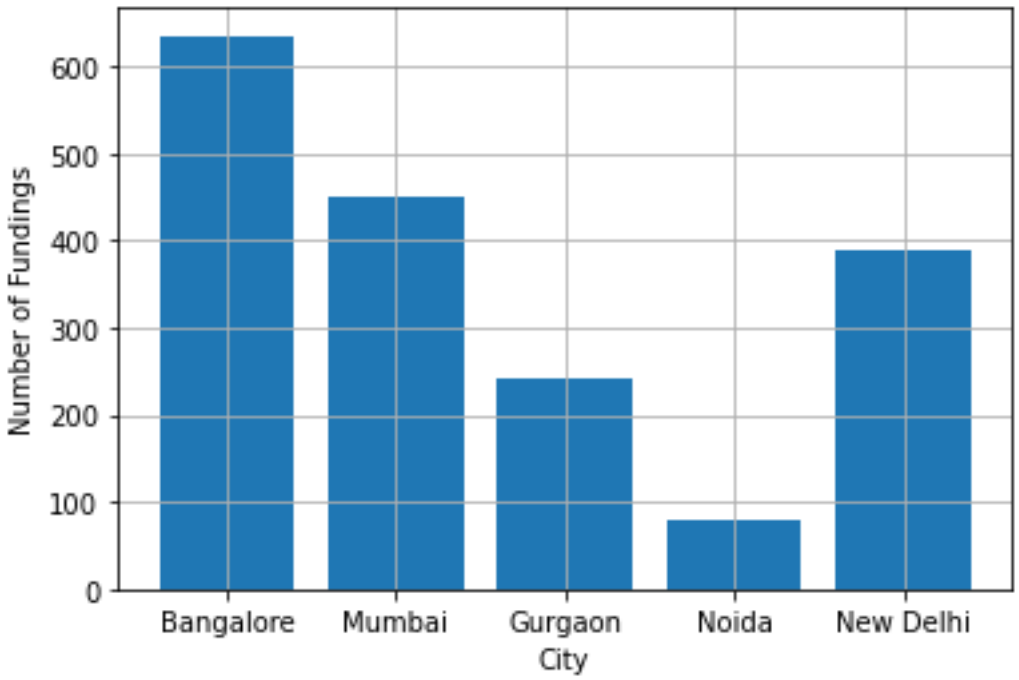
The approach taken towards this code is creating two lists , one with the name of cities and the other with corresponding number of times investments were done in those cities.

For this purpose,

```
counts[i] = len(df[df['CityLocation'].str.contains('^'+cities[i],regex=True,na=False,case=False)])
```

here , the rows where city name is or contains the given city names is being extracted and their length, i.e count is being recorded in the count array. Case = False takes care of the different case letters present in the city name. regex = true and ^cities[i] make sure that only entries where the name starts with Indian cities are included for entries where one Indian and one foreign city is present. So, Dallas/Hyderabad won't be included but Hyderabad/Dallas would be included.
Thereafter a bar graph is plotted between the cities and the number of fundings made in that city.

Plots and Final Suggestions –



For the given cities, The maximum number of fundings have been done in Bangalore(around 650) as compared to the other cities.

Q2.) Even after trying for so many times, your friend's startup could not find the investment. So you decided to take this matter in your hand and try to find the list of investors who probably can invest in your friend's startup. Your list will increase the chance of your friend startup getting some initial investment by contacting these investors. Find the top 5 investors who have invested maximum number of times (consider repeat investments in one company also). In a startup, multiple investors might have invested. So consider each investor for that startup. Ignore undisclosed investors.

CODE –

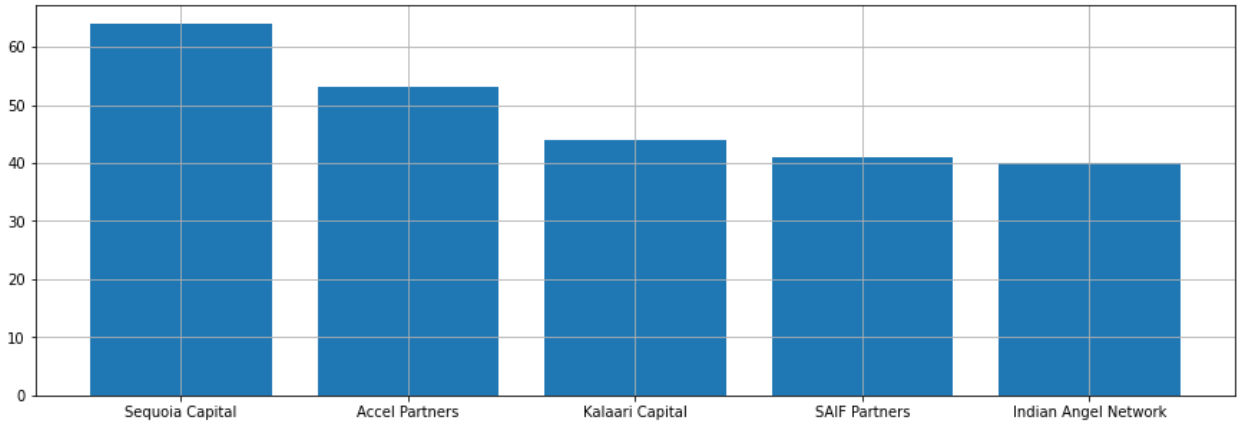
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

df = pd.read_csv('startup_funding.csv')
df.loc[df['InvestorsName'].isnull(),'InvestorsName'] = None
dict = {}

for i in df['InvestorsName'].str.split(',').values:
   if(i is not None):
      for k in i:
         if(k is not None):
            dict[k.strip()] = dict.get(k.strip(),0)+1

listx = []
key=''
value=0
for i in dict:
   listx.append([dict[i],i])
listx.sort(reverse=True)

for i in range(5):
   inv_name[i] = listx[i][1]
   inv_num[i] = listx[i][0]

plt.bar(inv_name,inv_num)
plt.grid()
plt.rcParams['figure.figsize']=(12,5)
plt.show()
```

Justification for CODE –

In this code, to get the number of times an investor has invested in startups, a dictionary is being maintained. Every key of the dictionary is the investor name and its value is the number of times that investor has invested. For this, the InvestorName column is being traversed. Every entry is being treated as a string. To access each investor, the string is being split using comma(‘,’) as a separator. Thereafter if the investor doesn’t already exist in the dict, it is being initialized with a value of 1. Otherwise it is being incremented by 1.

Finally the dict is being transformed into a list of lists with the first entry being the number of investments and second being the name of investor. This list is then sorted in descending order and the top 5 values are procured.

Plots and Final Suggestions –



EXPLANATION – According to the data, the top 5 investors are-

1. Sequoia Capital (64)
2. Accel Partners (53)
3. Kalaari Capital (44)
4. SAIF Partners (41)
5. Indian Angel Network (40)

Q3.) After re-analysing the dataset you found out that some investors have invested in the same startup at different number of funding rounds. So before finalising the previous list, you want to improvise it by finding the top 5 investors who have invested in different number of startups. This list will be more helpful than your previous list in finding the investment for your friend startup. Find the top 5 investors who have invested maximum number of times in different companies. That means, if one investor has invested multiple times in one startup, count one for that company. There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

CODE -

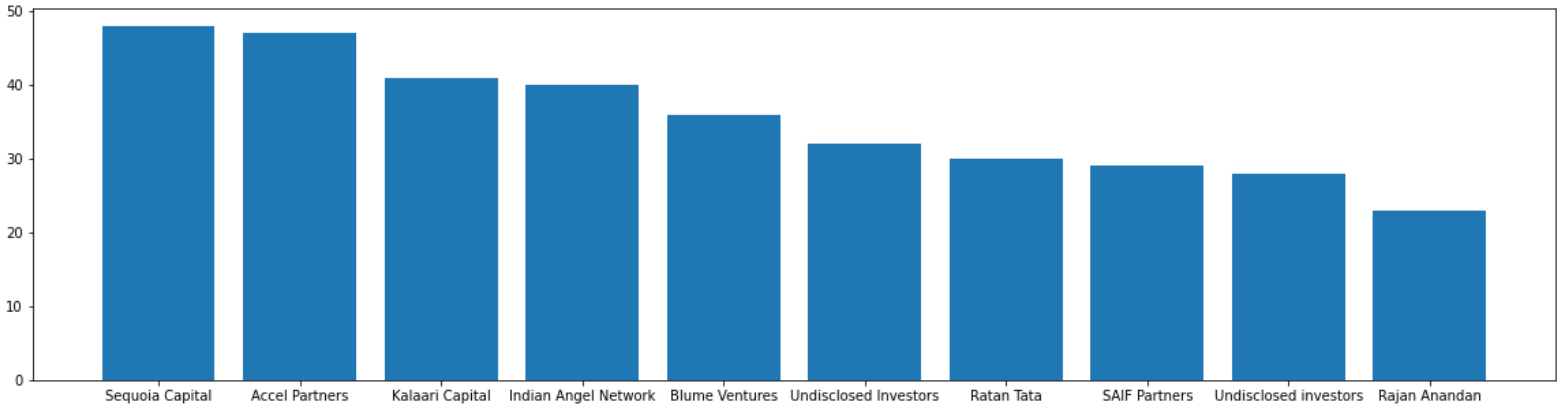
```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

df = pd.read_csv('startup_funding.csv')

df.loc[df['InvestorsName'].isnull(),'InvestorsName'] = None

df['StartupName'] = df['StartupName'].str.replace('Olacabs','Ola')
df['StartupName'] = df['StartupName'].str.replace('Ola Cabs','Ola')
df['StartupName'] = df['StartupName'].str.replace('Oyo Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('OyoRooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Oyorooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('OYO Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Flipkart.com','Flipkart')
df['StartupName'] = df['StartupName'].str.replace('Oyo Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Paytm Marketplace','Paytm')


dicti = {}

for (i,j) in zip(df['InvestorsName'].str.split(',').values,df['StartupName'].values):
    if(i is not None):
        startupname = j.strip()
        for k in i:
            if(k is not None):
                temp_list = dicti.get(k.strip())
                if(temp_list is None):
```

```
            temp_list=[]
        temp_list.append(startupname)
        dicti[k.strip()]=temp_list

for i in dicti:
  dicti[i]=list(set(dicti[i]))

listi = [[k,v] for k,v in dicti.items()]

for i in range(len(listi)-1):
  for j in range(i+1,len(listi)):
    if(len(listi[i][1])<len(listi[j][1])):
      listi[i],listi[j]=listi[j],listi[i]
investor=[]
investments=[]

for i in listi:
  if(i[0] != ''):
    investor.append(i[0])
    investments.append(len(i[1]))

x = investor[:10]
y = investments[:10]

plt.rcParams["figure.figsize"] = (20,5)
plt.bar(x,y)
plt.show()
```

Justification of CODE -

This code is similar to the previous code except, that instead of storing the number of times each investor has invested, we're storing a list of all the startups they've invested in. Then, in order to get unique values, we're converting the list into a set. The length of the set would thus give us the number of unique startups the investor has invested in.

Plots and Results -



EXPLANATION – According to the data, the top 5 investors investing in different startups are-

1. Sequoia Capital (48)
2. Accel Partners (47)
3. Kalaari Capital (41)
4. Indian Angel Network (40)
5. Blume Ventures(36)

Q4.) Even after putting so much effort in finding the probable investors, it didn't turn out to be helpful for your friend. So you went to your investor friend to understand the situation better and your investor friend explained to you about the different Investment Types and their features. This new information will be helpful in finding the right investor. Since your friend startup is at an early stage startup, the best-suited investment type would be - Seed Funding and Crowdfunding. Find the top 5 investors who have invested in a different number of startups and their investment type is Crowdfunding or Seed Funding. Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column. There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

CODE –

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

df = pd.read_csv('startup_funding.csv')

df['AmountInUSD'] = df['AmountInUSD'].str.replace(',','')

df['AmountInUSD'] = pd.to_numeric(df['AmountInUSD'])

df.loc[df['AmountInUSD'].isnull() , 'AmountInUSD'] = 0

df.loc[df['InvestorsName'].isnull(),'InvestorsName'] = None

df['StartupName'] = df['StartupName'].str.replace('Olacabs','Ola')
df['StartupName'] = df['StartupName'].str.replace('Ola Cabs','Ola')
df['StartupName'] = df['StartupName'].str.replace('Oyo Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('OyoRooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Oyorooms','Oyo')
```

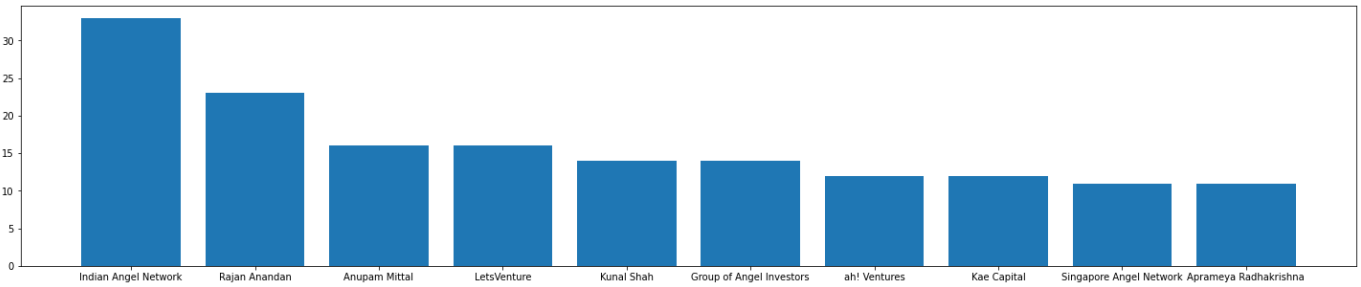
```
df['StartupName'] = df['StartupName'].str.replace('OYO Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Flipkart.com','Flipkart')
df['StartupName'] = df['StartupName'].str.replace('Paytm Marketplace','Paytm')

df['InvestmentType'] = df['InvestmentType'].str.replace('PrivateEquity','Private
Equity')
df['InvestmentType'] = df['InvestmentType'].str.replace('SeedFunding','Seed
Funding')
df['InvestmentType'] = df['InvestmentType'].str.replace('Crowd funding','Crowd
Funding')

df_copy = df[(df['InvestmentType']=='Seed
Funding')|(df['InvestmentType']=='Crowd Funding')]


dicti = {}

for (i,j) in
zip(df_copy['InvestorsName'].str.split(',').values,df_copy['StartupName'].values
):
    if(i is not None):
        startupname = j.strip()
        for k in i:
            if(k is not None):
                temp_list = dicti.get(k.strip())
                if(temp_list is None):
                    temp_list=[]
                temp_list.append(startupname)
                dicti[k.strip()]=temp_list

for i in dicti:
    dicti[i]=list(set(dicti[i]))

listi = [[k,v] for k,v in dicti.items()]

for i in range(len(listi)-1):
    for j in range(i+1,len(listi)):
        if(len(listi[i][1])<len(listi[j][1])):
```

```
            listi[i],listi[j]=listi[j],listi[i]
investor=[]
investments=[]

for i in listi:
    if(i[0] != '' and i[0]!='Undisclosed Investors' and i[0]!='Undisclosed
investors'):
        investor.append(i[0])
        investments.append(len(i[1]))

x = investor[:10]
y = investments[:10]

plt.rcParams["figure.figsize"] = (20,5)
plt.bar(x,y)
plt.show()
```

Justification of CODE -

This code is the same as the previous code except that we've created a copy of the original dataframe by only including the investments where investment type is either Seed Funding or Crowd Funding. Thereafter the same process of extracting the unique startups from the dataframe is followed.

Plots and Results –



EXPLANATION – According to the data, the top 5 investors investing in different startups having investment type as seed funding or crowd funding are-

1. Indian Angel Network (33)
2. Rajan Anandan(23)
3. Anupam Mittal(16)
4. LetsVenture(16)
5. Kunal Shah(14)

Q5.) Due to your immense help, your friend startup successfully got seed funding and it is on the operational mode. Now your friend wants to expand his startup and he is looking for new investors for his startup. Now you again come as a saviour to help your friend and want to create a list of probable new investors. Before moving forward you remember your investor friend advice that finding the investors by analysing the investment type. Since your friend startup is not in early phase it is in growth stage so the best-suited investment type is Private Equity. Find the top 5 investors who have invested in a different number of startups and their investment type is Private Equity. Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column.There are many errors in startup names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm..

CODE –

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np

df = pd.read_csv('startup_funding.csv')

df['AmountInUSD'] = df['AmountInUSD'].str.replace(',','')

df['AmountInUSD'] = pd.to_numeric(df['AmountInUSD'])

df.loc[df['AmountInUSD'].isnull() , 'AmountInUSD'] = 0

df.loc[df['InvestorsName'].isnull(),'InvestorsName'] = None

df['StartupName'] = df['StartupName'].str.replace('Olacabs','Ola')
df['StartupName'] = df['StartupName'].str.replace('Ola Cabs','Ola')
```

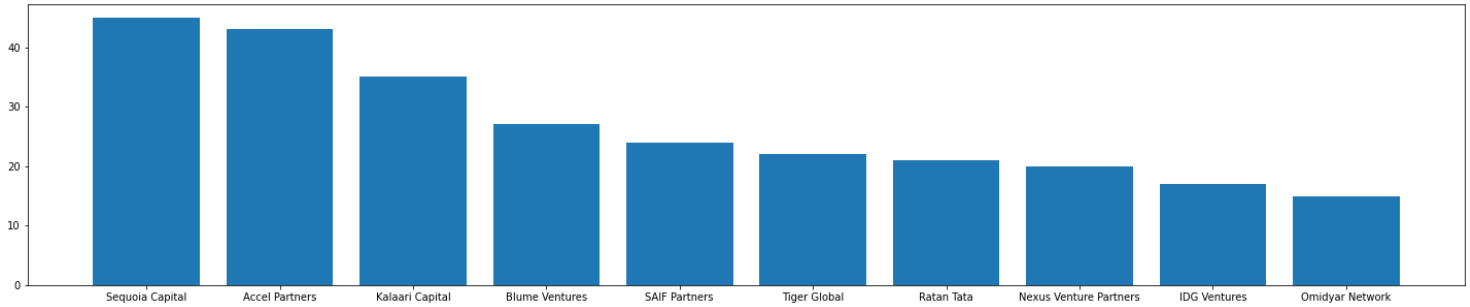
```
df['StartupName'] = df['StartupName'].str.replace('Oyo Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('OyoRooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Oyorooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('OYO Rooms','Oyo')
df['StartupName'] = df['StartupName'].str.replace('Flipkart.com','Flipkart')
df['StartupName'] = df['StartupName'].str.replace('Paytm Marketplace','Paytm')

df['InvestmentType'] = df['InvestmentType'].str.replace('PrivateEquity','Private
Equity')
df['InvestmentType'] = df['InvestmentType'].str.replace('SeedFunding','Seed
Funding')
df['InvestmentType'] = df['InvestmentType'].str.replace('Crowd funding','Crowd
Funding')

df_copy = df[df['InvestmentType']=='Private Equity']


dicti = {}

for (i,j) in
zip(df_copy['InvestorsName'].str.split(',').values,df_copy['StartupName'].values
):
    if(i is not None):
        startupname = j.strip()
        for k in i:
            if(k is not None):
                temp_list = dicti.get(k.strip())
                if(temp_list is None):
                    temp_list=[]
                temp_list.append(startupname)
                dicti[k.strip()]=temp_list

for i in dicti:
    dicti[i]=list(set(dicti[i]))

listi = [[k,v] for k,v in dicti.items()]

for i in range(len(listi)-1):
```

```
    for j in range(i+1,len(listi)):
        if(len(listi[i][1])<len(listi[j][1])):
            listi[i],listi[j]=listi[j],listi[i]
investor=[]
investments=[]

for i in listi:
    if(i[0] != '' and i[0]!='Undisclosed Investors' and i[0]!='Undisclosed
investors'):
        investor.append(i[0])
        investments.append(len(i[1]))

x = investor[:10]
y = investments[:10]

plt.rcParams["figure.figsize"] = (20,5)
plt.bar(x,y)
plt.show()
```

Justification of the Code –

This code is the same as the previous code except that we've created a copy of the original dataframe by only including the investments where investment type is Private Equity. Thereafter the same process of extracting the unique startups from the dataframe is followed.

Plots and Results-



EXPLANATION – According to the data, the top 5 investors investing in different startups having investment type as Private Equity are-

1.) Sequoia Capital(45)
2.) Accel Partners(43)
3.) Kalaari Capital(35)
4.) Blume Ventures(27)
5.) SAIF Partners(24)