PROJECT: CASE STUDY ON INDIAN STARTUPS

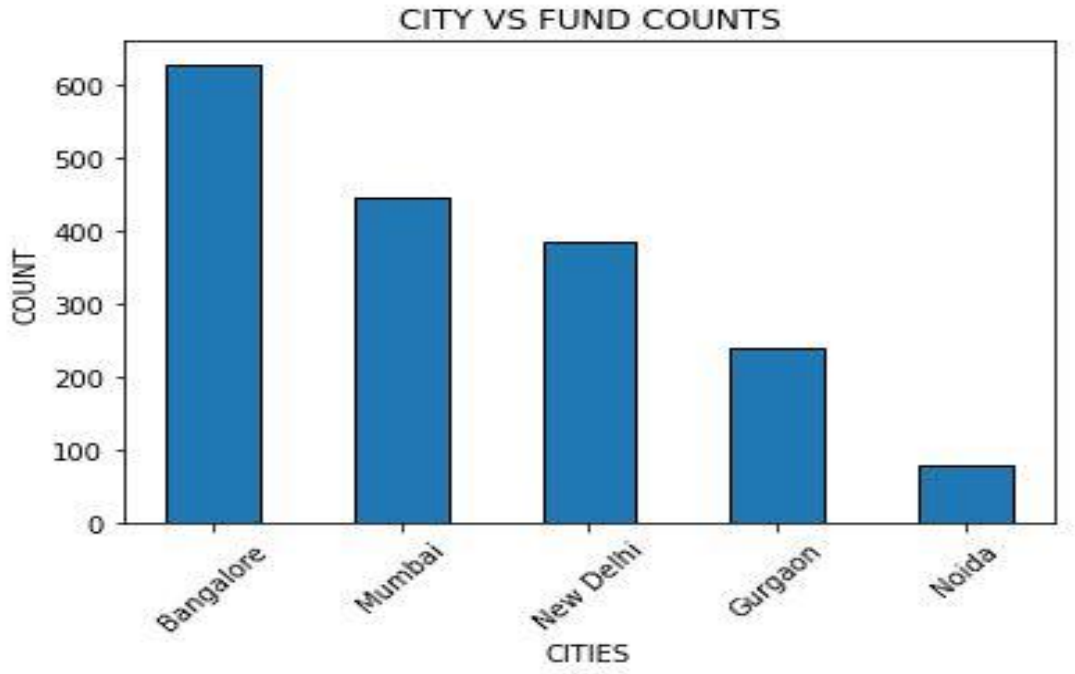
1. Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, 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 pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv('startupfundings.csv')
df.dropna(subset=['CityLocation'])
df['CityLocation'].replace("Delhi","New Delhi",inplace=True)
df['CityLocation'].replace("bangalore","Bangalore",inplace=True)
df=df[(df.CityLocation=='Bangalore') | (df.CityLocation == 'New Delhi') | (df.CityLocation == 'Noida') |
(df.CityLocation == 'Gurgaon') | (df.CityLocation == 'Mumbai')]

d={}
for i in df['CityLocation']:
    if i in d:
        d[i]+=1
    else:
        d[i]=1

x=list(d.keys())
y=list(d.values())
print(x)
print(y)

plt.bar(x,y,width=0.5,edgecolor='black')
plt.xticks(rotation=45)
plt.title('CITY VS FUND COUNTS')
plt.xlabel('CITIES')
plt.ylabel('COUNT')
plt.show()
```

OUTPUT:

```
['Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon', 'Noida']
[628, 446, 385, 240, 78]
```



EXPLANATION AND ANALYSIS OF THE PROBLEM:

In this problem we were supposed to find out the number of fundings received in the cities of BANGALORE, MUMBAI and DELHI NCR which consists of three cities i.e. NEW DELHI, GURGAON and NOIDA.
First of all we need to drop the null values from the CityLocation column of the dataframe. Then I have replaced all the inappropriate casing of the values to the correct casing i.e. the first letter of each word in the CityLocation is column is made capital. After taking care of the casing I filtered out the data that I actually need i.e. the dataframe with only CityLocation as BANGALORE, MUMBAI, NEW DELHI, GURGAON, and NOIDA. After this, I created a dictionary and appended the cities from the dataframe as keys and initially set their value to zero. The values in the dictionary were incremented by one whenever a city was iterated again.
Then for plotting the graph I explicitly took other lists for the keys and their corresponding values in the dictionary. Used the matplotlib module to plot the graph.

INTERPRETATION OF RESULT:
From the bar graph we can clearly see that the city BANGALORE has received the highest number of fundings followed by MUMBAI, NEW DELHI, GURGAON and the least number of fundings was received by NOIDA.

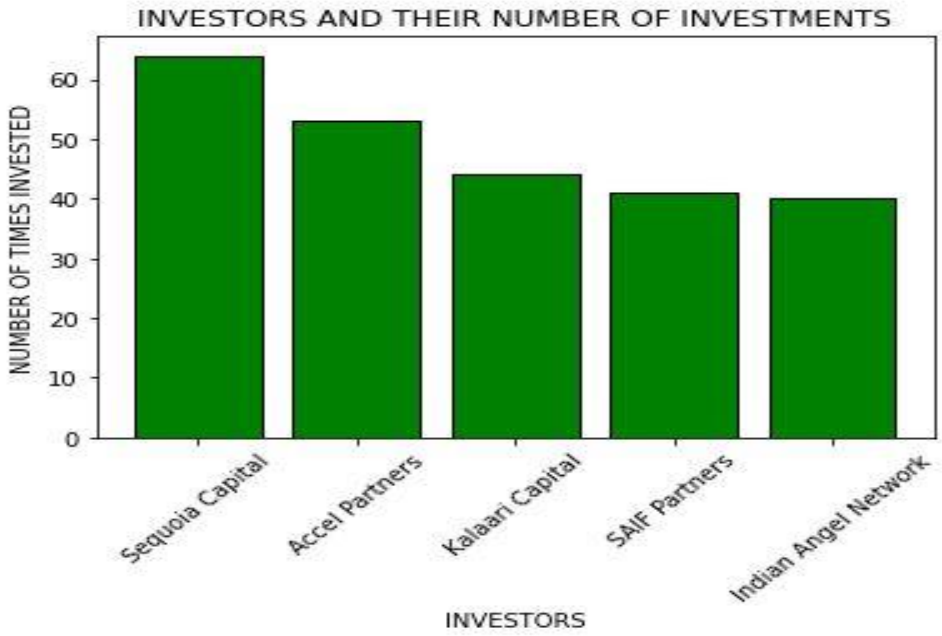
2. Even after trying for so many times, your friend's startup could not find the investment. So you decided to take this matter in your hand and try to find the list of investors who probably can invest in your friend's startup. Your list will increase the chance of your friend startup getting some initial 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 pandas as pd
import matplotlib.pyplot as plt
startup = pd.read_csv("startupfundings.csv")
df = startup.copy()

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

d = {}
for i in df["InvestorsName"].values:
    if "," in i:
        for j in i.strip().split(','):
            d[j.strip()] = d.get(j.strip(),0) + 1
    else:
        d[i.strip()] = d.get(i.strip(),0) + 1



d1 = sorted(d, key=d.get , reverse=True)[0:5]
for i in d1:
    print(i)
d={k: v for k, v in sorted(d.items(), key=lambda item: item[1], reverse=True)}

x=list(d.keys())[0:5]
y=list(d.values())[0:5]

plt.bar(x,y, edgecolor='black', color='green')
plt.xticks(rotation=45)
plt.title('INVESTORS AND THEIR NUMBER OF INVESTMENTS')
plt.xlabel('INVESTORS')
plt.ylabel('NUMBER OF TIMES INVESTED')
plt.show()
```

OUTPUT:

```
Sequoia Capital
Accel Partners
Kalaari Capital
SAIF Partners
Indian Angel Network
```



EXPLANATION AND ANALYSIS OF THE PROBLEM:

In this problem we found out the top 5 investors who have invested maximum number of times. We included the repeated investments by the investors.

Firstly we dropped all the null values in the InvestorsName column of the dataframe. After doing this, we made a dictionary and iterated through the InvestorsName column and pushed the names in the dictionary as keys and set their initial values as one. Every time a the InvestorsName was encountered their respective values were incremented by one. If we encountered a column with multiple values in it, we first separated the values in the cell by comma and made a list of the names in the cell. Then we iterated through this list and did the same as we were doing for the column initially. Now that we have our investors names as keys and their number of investments as values we proceeded to sort the dictionary based on the values. After this, the top 5 keys and values were pushed to other lists x and y respectively and these lists were used to plot the graph using matplotlib module.

INTERPRETATION OF RESULT:

If we look at the graph we can see that Sequoia Capital has made the most number of investments followed by Accel Partners, Kalaari Capital and SAIF Partners. Indian Angel Network stood at 5^{th} position in number of fundings.

3. 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.

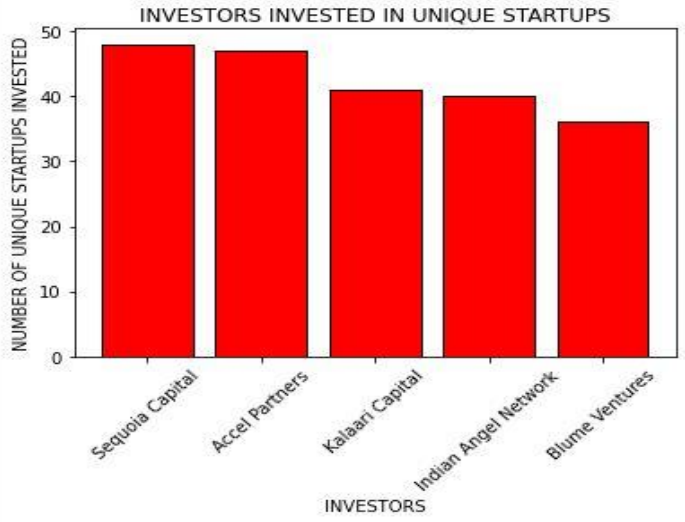
CODE:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
startup = pd.read_csv("startupfundings.csv")
df = startup.copy()

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

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("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)
df["StartupName"].replace("Paytm Marketplace","Paytm",inplace = True)
d = {}
for i in df.index:
    e = df["InvestorsName"][i].strip()
    if "," in e:
        for j in e.strip().split(','):
            if j.strip() in d:
                d[j.strip()].add(df["StartupName"][i].strip())
            else:
                s = set()
                d[j.strip()] = s
                d[j.strip()].add(df["StartupName"][i].strip())
    else:
        a = e.strip()
        if a in d:
            d[a].add(df["StartupName"][i].strip())
        else:
            s = set()
            d[a] = s
            d[a].add(df["StartupName"][i].strip())

d1 = {}
for i in d:
    if i == "":
        continue
    d1[i] = len(d[i])

d2 = sorted(d1, key=d1.get , reverse=True)[0:5]
for i in d2:
```

```
    print(i)

d3={}
for i in d:
    d3[i]=len(d[i])

d3={k: v for k, v in sorted(d3.items(), key=lambda item: item[1],reverse=True)}
del d3['']


x=list(d3.keys())[0:5]
y=list(d3.values())[0:5]

plt.bar(x,y, edgecolor='black', color='red')
plt.xticks(rotation=45)
plt.title('INVESTORS INVESTED IN UNIQUE STARTUPS')
plt.xlabel('INVESTORS')
plt.ylabel('NUMBER OF UNIQUE STARTUPS INVESTED')
plt.show()
```

OUTPUT:

```
Sequoia Capital
Accel Partners
Kalaari Capital
Indian Angel Network
Blume Ventures
```



EXPLANATION AND ANALYSIS OF THE PROBLEM:

In this problem we were supposed to find the top 5 investors who have invested maximum number of times in different companies. We started by dropping the null values in the StartupName and InvestorsName column. After doing this we corrected the names of some prominent startups in our dataframe such as ola, oyo, paytm and flipkart. Then we created a dictionary in which we kept the InvestorsName as keys and the startupnames they have funded as values. We used sets as the dictionary values because it eliminates the duplicacy by keeping the startup name only once if it gets repeated for the same InvestorsName. Then we created dictionary d1 which containes the investors name and their unique fund count in a startup by calculating the length of the set for each investor in previous dictionary. Then we created dictionary d2 which contains the key value pairs of d1 in descending order since we wanted to find the top 5 investors only. Then we created two lists x and y to keep the keys and values of top 5 items in dictionary d2. Then we used the matplotlib module to plot a bar graph.

INTERPRETATION OF RESULT:

From the graph we can see that Sequoia Capital has invested the most number of time for different startups followed by Accel Partners, Kalaari Capital, Indian Angel network and lastly Blume Ventu res.

4. 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:

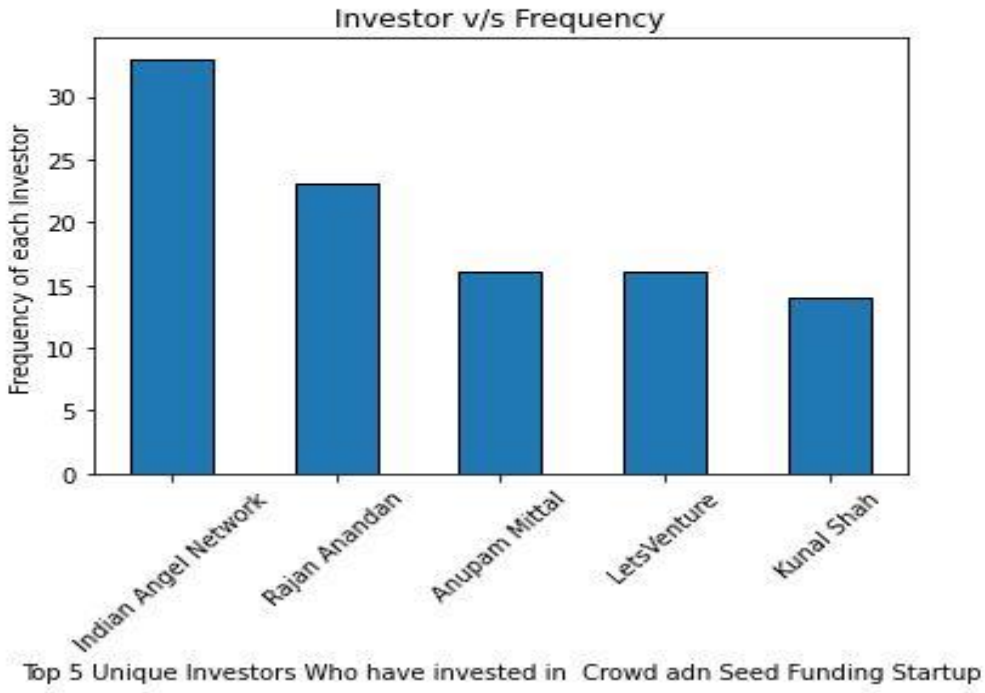
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('startupfundings.csv' , encoding ='utf-8')
df


df['StartupName'] = df['StartupName'].str.lower()
df['StartupName'].replace(['flipkart.com','flipkart'],'flipkart',inplace=True)
df['StartupName'].replace(['oyo rooms','oyorooms','oyo'],'oyo',inplace=True)
df['StartupName'].replace(['ola','ola cabs'],'ola',inplace=True)
df['StartupName'].replace(['paytm marketplace','paytm'],'paytm',inplace=True)


invest_type = df['InvestmentType']
invest_type = invest_type.unique()
invest_type


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


arr = df[['InvestorsName','InvestmentType','StartupName']]
arr1= arr[(arr['InvestmentType'] == 'Seed Funding') | (arr['InvestmentType'] == 'Crowd Funding')]
arr1


arr1.drop('InvestmentType',inplace=True,axis = 1)# dropping null values from data Frame and dropping column
investment type
arr1
```

```
arr2 = arr1['InvestorsName'].str.split(',',expand=True)
arr2.columns =['c0','c1','c2','c3','c4','c5','c6','c7','c8','c9']
arr2['StartupName'] = arr1['StartupName']# adding startup column in the data Frame
arr2


#getting unique startupname values from the data frame
arr3 = arr2['StartupName'].unique()
arr3


df1 = arr2.groupby('StartupName')#grouping by according to startup name


# Creating List
l = []
for key in range(len(arr3)):
    df2 = df1.get_group(arr3[key])
    s = set()
    for i in range(len(df2)):
        for j in range(len(df2.columns)):
            if(type(df2.iloc[i,j])==str ):
                s.add(df2.iloc[i,j].strip())# adding elemnets to set
                t = list(s)# Converting set to list
            else:
                continue
    l.append(t)    #appending each list into another list list l
l


d = {}
for i in range(len(l)):
    for j in range(len(l[i])):
        temp = l[i][j]
        d[temp] = d.get(temp,0)+1


t = {}# taking another dictionary havng no null values from dictionary d
for i in d:
    if(i !=''): # removing null string
        t[i]=d[i]
del t['Undisclosed Investors']# removing Undisclosed Investors
del t['Undisclosed investors']


from collections import Counter
dic = dict(Counter(t).most_common(5))
x = []
y= []
for i in dic:
    x.append(i) #Getting keys from dictionary
```

```
    y.append(dic[i])# Getting Freq of the values
    print(i,dic[i])


plt.bar(x,y,edgecolor ='black',width=0.5)
plt.xlabel('Top 5 Unique Investors Who have invested in  Crowd adn Seed Funding Startup')
plt.ylabel('Frequency of each Investor')
plt.xticks(rotation =45)
plt.title('Investor v/s Frequency')
plt.show()
```

OUTPUT:

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



EXPLANATION AND ANALYSIS OF THE PROBLEM:

In this problem we had to find out the top 5 investors who have made investments in CROWD FUNDING and SEED FUNDING type of investments. We were not allowed to count duplicate investors for any startups. We started by dropping the null values in the StartupName and InvestorsName column. After doing this we corrected the names of some prominent startups in our dataframe such as ola, oyo, paytm and flipkart. Then we filtered the dataframe to only those rows which contained Crowd Funding or Seed Funding in their INVESTMENTTYPE column. Then we created a dictionary in which we kept the InvestorsName as keys and the startupnames they have funded as values. We used sets as the dictionary values because it eliminates the duplicacy by keeping the startup name only once if it gets repeated for the same InvestorsName. Then we created dictionary d1 which containes the investors name and their unique fund count in a startup by calculating the length of the set for each investor in previous dictionary. Then we created dictionary d2 which contains the key value pairs of d1 in descending order since we wanted to find the top 5 investors only. Then we created two lists x and y to keep the keys and values of top 5 items in dictionary d2. Then we used the matplotlib module to plot a bar graph.

INTERPRETATION OF RESULT:

From the above graph we can see that INDIAN ANGEL NETWORK has topped our list, followed by RAJAN ANANDAN, ANUPAM MITTAL, LETSVENTURE, and KUNAL SHAH in the respective orders.

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

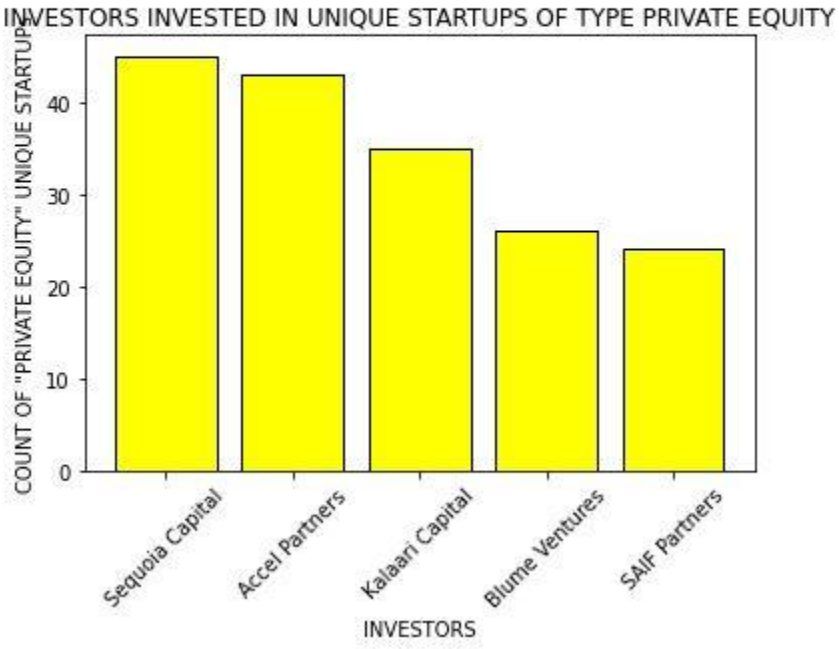
CODE:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
startup = pd.read_csv("startupfundings.csv")
df = startup.copy()


df.loc[df.InvestmentType == 'Crowd funding'] = 'Crowd Funding'
df.loc[df.InvestmentType == 'PrivateEquity'] = 'Private Equity'
df.loc[df.InvestmentType == 'SeedFunding'] = 'Seed Funding'

df.InvestmentType.loc[df.InvestmentType == 'Crowd funding'] = 'Crowd Funding'
df.InvestmentType.loc[df.InvestmentType == 'PrivateEquity'] = 'Private Equity'
df.InvestmentType.loc[df.InvestmentType == 'SeedFunding'] = 'Seed Funding'

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

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

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


d = {}
for i in df.index:
    e = df["InvestorsName"][i].strip()
    if "," in e:
        for j in e.strip().split(','):
            if j.strip() in d:
                d[j.strip()].add(df["StartupName"][i].strip())
            else:
```

```
            s = set()
            d[j.strip()] = s
            d[j.strip()].add(df["StartupName"][i].strip())
    else:
        a = e.strip()
        if a in d:
            d[a].add(df["StartupName"][i].strip())
        else:
            s = set()
            d[a] = s
            d[a].add(df["StartupName"][i].strip())

d1 = {}
for i in d:
    if i == "":
        continue
    d1[i] = len(d[i])

d2 = sorted(d1, key=d1.get , reverse=True)[0:5]
for i in d2:
    print(i)


d3={}
for i in d:
    d3[i]=len(d[i])

d3={k: v for k, v in sorted(d3.items(), key=lambda item: item[1],reverse=True)}
del d3['']


x=list(d3.keys())[0:5]
y=list(d3.values())[0:5]

plt.bar(x,y, edgecolor='black', color='yellow')
plt.xticks(rotation=45)
plt.title('INVESTORS INVESTED IN UNIQUE STARTUPS OF TYPE PRIVATE EQUITY')
plt.xlabel('INVESTORS')
plt.ylabel('COUNT OF "PRIVATE EQUITY" UNIQUE STARTUPS')
plt.show()
```

OUTPUT:

```
Sequoia Capital
Accel Partners
Kalaari Capital
Blume Ventures
SAIF Partners
```



EXPLANATION AND ANALYSIS OF THE PROBLEM:

In the 5th question, we were asked to Find the top 5 investors who have invested in a different number of startups and their investment type is Private Equity. Firstly we dropped all the null values from the dataframe columns StartupName and InvestmentType. Next we need to correct the InvestmentName column values as per the given correct values. For this we can print the unique values in that column and know what all names are to be changed. Then, we need to correct the StartupNames of some of the prominent Startups like OYO, OLA, PAYTM, and FLIPKART. Then we created a dictionary d that keeps all the InvestorsNames as keys and the startups they have funded in as a value which is in form of set. And as we know that sets will avoid duplicacy in them, the values of each InvestorsName will be a series of unique startupnames they have funded in. Next we create a dictionary d1 using d which contains same keys as before in d but the values are changed as length of the set in the dictionary d. So, now the dictionary d1 contains investorsnames as keys and the number of unique startups they have funded in as values. Next d2 is created which is the sorted version of d1 dictionary in descending order. Next two lists x and y are made that will keep the values of top 5 keys and values respectively in dictionary d2 . These two lists are used to create a bar graph that represents the top 5 investornames that have invested in PRIVATE EQUITY investment type of startups.

INTERPRETATION OF RESULT:

From the above graph we can interpret that Sequoia Capital has invested the most in PRIVATE EQUITY type investments followed by Accel Partners, Kalaari Capital, Blume Ventures and SAIF Partners respectively in that order.