

Name: Shardul Patil

Course: Data Science & Machine Learning Complete Course | 9 Nov '20

Project: Case Study (Part – II)

Note: The file which I have opened in my code in Jupyter had to be downloaded from the coding ninja's module so the path of the file I have put as "C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv". Please in order to check the code put "startup_funding.csv" in open() or pd.read_csv().

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 startup has 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.

Justification:

```
#Q1
import csv
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
li=[]
with open('C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv') as file_obj:
    file_data=csv.DictReader(file_obj,skipinitialspace=True)
    for row in file_data:
        li.append(row['CityLocation'])
```

STEPS:

- First, we open the csv file "startup_funding.csv" and put in file_obj.
- Then, we get the Dictionaries of all rows in form (column name: value) with skipping the initial spaces if there exists in the strings. Therefore, we put all the Dictionaries in the file_data.
- Now we iterate over all the rows in file_data and append all values of "CityLocation" column in list li. Thus we have a list which contains all the cities in it.

```
a=[]
li=np.array(li)
for i in li:
    str=""
    for j in i:
        if j=='/':
            break
        else:
            str=str+j
    a.append(str.strip())
a=np.array(a)
a[a=="Delhi"]="New Delhi"
a[a=="bangalore"]="Bangalore"
a[a=="New Delhi"]="NCR"
a[a=="Gurgaon"]="NCR"
a[a=="Noida"]="NCR"
```

STEPS:

- We make li a numpy array and declare an empty list a. For few startups multiple locations are given, one Indian and one Foreign. Count those startups in Indian startup also. Indian city name is first.
- The For loop iterating over li is basically used to get the Indian cities which come before "/" and are appended in list a. In case if there is no such string with "/", directly the entire string is appended in list a. Now list a contains all the Indian cities and some foreign cities where startup took place independently.
- We make list a numpy array. Have taken City name "Delhi" as "New Delhi". Checked the case-sensitiveness of cities also. That means, at some place instead of "Bangalore", "bangalore" is given. We have converted city name as "Bangalore".
- "NCR" has been considered to represent three locations-"New Delhi", "Gurgaon" and "Noida".

```

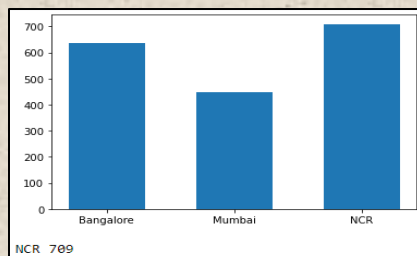
d={}
for i in a:
    if i=="Bangalore":
        d[i]=d.get(i,0)+1
    if i=="Mumbai":
        d[i]=d.get(i,0)+1
    if i=="NCR":
        d[i]=d.get(i,0)+1
city=[]
nooffundings=[]
for i in d:
    city.append(i)
    nooffundings.append(d[i])
plt.bar(city,nooffundings,width=0.6)
plt.show()
val=""
maxval=-1
for i in d:
    if d[i]>maxval:
        maxval=d[i]
        val=i
print(val,maxval)
#NCR is the location where startups has received funding maximum number of times,ie,709 times

```

STEPS:

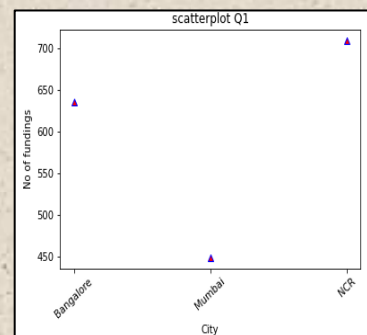
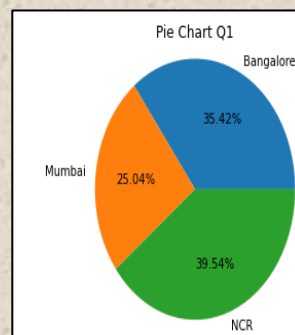
- We create an empty dictionary **d** and iterate over list **a** in order to get frequencies of cities “Bangalore”, “Mumbai” and “NCR”. Dictionary **d** has keys as “Bangalore”, “Mumbai” and “NCR” and values as their number of occurrences in list.
- We create two empty lists **city** and **nooffundings**. On iterating over the dictionary **d**, we append the keys in **city** and values in **nooffundings**.
- Thus a bar graph is being plotted with cities on X axis and number of funding done on Y axis.
- Finally, we iterate over the dictionary **d** to get the highest frequency value along with name of the city and then print

Final Output:



- Thus a bar graph is being plotted with cities on X axis and number of funding done on Y axis.
- Hence it is concluded that “NCR is the location where startups has received funding maximum number of times, i.e., 709 times.

Plots and Graphs:



- I have even made a pie chart though not asked in the question which gives percentage of no of funding done in 3 cities.
- I have even made a scatterplot of 3 cities with their no of funding

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.

Justification:

```
#Q2
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import csv
with open('C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv') as file_obj:
    file_data=csv.DictReader(file_obj,skipinitialspace=True)
    inv=[]
    for row in file_data:
        inv.append(row['InvestorsName'])
```

STEPS:

- First, we open the csv file "startup_funding.csv" and put in file_obj.
- Then, we get the Dictionaries of all rows in form (column name: value) with skipping the initial spaces if there exits in the strings. Therefore, we put all the Dictionaries in the file_data.
- Now we iterate over all the rows in file_data and append all values of "InvestorsName" column in list inv. Thus we have a list which contains all the names of investors in it.

```
a=[]
for i in inv:
    if ',' in i:
        b=i.split(',')
        for j in b:
            a.append(j.strip())
    else:
        a.append(i)
d={}
for i in a:
    d[i]=d.get(i,0)+1
```

STEPS:

- We create an empty list a. In a startup, multiple investors might have invested. So we have to consider each investor for that startup. Thus we split the strings where are separated by "," present in one entire string in a list b. Thus append every string of list b in list a. If there is no "," present in the string just append the entire string in list a.
- We create an empty dictionary d and iterate over list a in order to get frequencies of different investors. Dictionary d thus has keys as names of the investors and values as frequencies of those investors.

```
investors=[]
freq=[]
del d['']
for i in range(5):
    val=''
    maxval=-1
    for i in d:
        if d[i]>maxval:
            maxval=d[i]
            val=i
    investors.append(val)
    freq.append(maxval)
    del d[val]
for i in range(len(freq)):
    print(investors[i],freq[i])
```

STEPS:

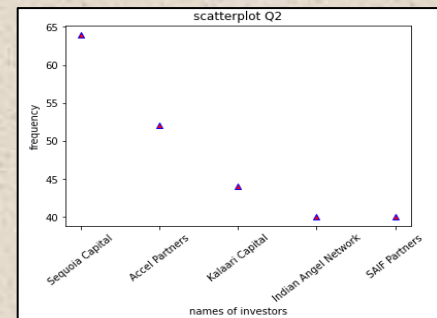
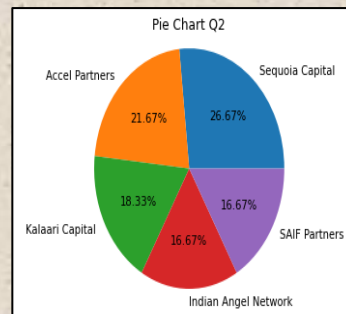
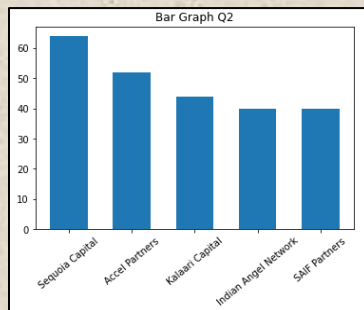
- We create two empty lists investors and freq. Then we delete the part of dictionary which has null values as we have to ignore undisclosed investors.
- Then we iterate over the dictionary d in order to get the top 5 investors along with the number of times they have invested. We append the names of top 5 investors in list investors and their frequencies in list freq.
- Thus we print the output in format (Investor name, number of times he has invested).

Final Output:

Sequoia Capital 64
Accel Partners 52
Kalaari Capital 44
Indian Angel Network 40
SAIF Partners 40

- Thus we have printed the output in format (Investor name, number of times he has invested).
- Thus we have the top 5 investors who have invested maximum number of times (consider repeat investments in one company also).
1. Sequoia Capital 2. Accel Partners 3. Kalaari Capital 4. Indian Angel Network 5. SAIF Partners

Plots and Graphs:



- I have made a bar graph though not asked in the question which gives frequency of top 5 investors.
- I have even made a scatterplot of top 5 investors with their frequencies.
- I have even made a pie chart though not asked in the question which gives percentage of investment done by top 5 investors.

Q3). After re-analyzing the dataset you found out that some investors have invested in the same startup at different number of funding rounds. So before finalizing 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.

Justification:

```
#Q3
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import csv
with open('C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv') as file_obj:
    file_data=csv.DictReader(file_obj,skipinitialspace=True)
    file_list=list(file_data)
```

STEPS:

- First, we open the csv file "startup_funding.csv" and put in file_obj.
- Then, we get the Dictionaries of all rows in form (column name: value) with skipping the initial spaces if there exists in the strings. Therefore, we put all the Dictionaries in the file_data.
- Thus we form a list of all dictionaries of file_data named as file_list.


```

inv=[]
company=[]
for row in file_list:
    if row['StartupName']=="Flipkart.com":
        row['StartupName']="Flipkart"
    if row['StartupName']=="Oyo Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OYO Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Oyorooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OyoRooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Ola Cabs":
        row['StartupName']="Ola"
    if row['StartupName']=="OlaCabs":
        row['StartupName']="Ola"
    if row['StartupName']=="Paytm Marketplace":
        row['StartupName']="Paytm"
    i=row['InvestorsName']
    if ',' in i:
        b=i.split(',')
        for j in range(len(b)):
            company.append(row['StartupName'])
        for j in b:
            inv.append(j.strip())
    else:
        inv.append(i)
        company.append(row['StartupName'])

```

STEPS:

- We create a two empty lists inv and company.
- There are many errors in startup names. Ignore correcting all; just handle the important ones - Ola, Flipkart, Oyo and Paytm. Thus we have converted all the related Startup names to the important ones on iterating through row dictionaries in file_list.
- Now we take investors name in i variable. If it contains many names separated by ",", we split all the names and put into list b which again iterated and the names are appended in list inv and thus we add Startup name corresponding to all the investor names in list company times equal to length of list b .
- If there is no ",", in i, just append investor name in list inv and append Startup name in list b once.

```

d={}
for i in range(len(inv)):
    if company[i]!='':
        if inv[i] in d:
            d[inv[i]].add(company[i])
        else:
            d[inv[i]]=set()
            d[inv[i]].add(company[i])
for i in d:
    k=len(d[i])
    d[i]=k

```

STEPS:

- We create an empty dictionary d. On iterating through both the lists inv and company, we keys as names of investors and values as set of all different companies where they have invested ignoring the null values. We have used a set here so as to get different companies of that particular investor and avoid repetition.
- Again we iterate on dictionary d, where we change the values of d equal to length of set as we need number of different companies where they have invested and are not concerned about the names at all. Finally we have a dictionary d with keys as investor names and values as number of different companies they have invested in.

```

investors=[]
freq=[]
del d['']
for i in range(5):
    val=''
    maxval=-1
    for i in d:
        if d[i]>maxval:
            maxval=d[i]
            val=i
    investors.append(val)
    freq.append(maxval)
    del d[val]
for i in range(len(freq)):
    print(investors[i],freq[i])

```

STEPS:

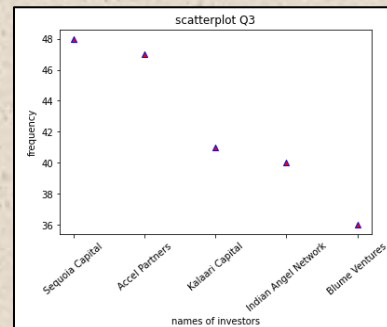
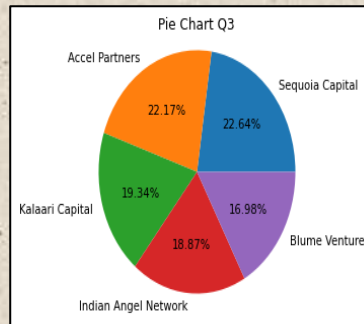
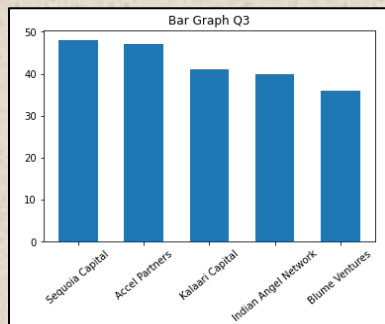
- We create two empty lists investors and freq. Then we delete the part of dictionary which has null values as we have to ignore undisclosed investors.
- Then we iterate over the dictionary d in order to get the top 5 investors along with the number of times they have invested. We append the names of top 5 investors in list investors and their frequencies in list freq.
- Thus we print the output in format (Investor name, number of times he has invested).

Final Output:

```
Sequoia Capital 48
Accel Partners 47
Kalaari Capital 41
Indian Angel Network 40
Blume Ventures 36
```

- Thus we have printed the output in format (Investor name, number of times he has invested).
- Thus we have the top 5 investors who have invested maximum number of times in different companies.
1. Sequoia Capital 2. Accel Partners 3. Kalaari Capital 4. Indian Angel Network 5. Blume Ventures

Plots and Graphs:



- I have made a bar graph though not asked in the question which gives frequency of top 5 investors.
- I have even made a scatterplot of top 5 investors with their frequencies.
- I have even made a pie chart though not asked in the question which gives percentage of investment done by top 5 investors.

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 is - "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.

Justification:

```
#Q4
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import csv
with open('C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv') as file_obj:
    file_data=csv.DictReader(file_obj,skipinitialspace=True)
    file_list=list(file_data)
```

STEPS:

- First, we open the csv file "startup_funding.csv" and put in file_obj.
- Then, we get the Dictionaries of all rows in form (column name: value) with skipping the initial spaces if there exists in the strings. Therefore, we put all the Dictionaries in the file_data.
- Thus we form a list of all dictionaries of file_data named as file_list.

```

inv=[]
company=[]
for row in file_list:
    if row['StartupName']=="Flipkart.com":
        row['StartupName']="Flipkart"
    if row['StartupName']=="Oyo Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OYO Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Oyornrooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OyoRooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Ola Cabs":
        row['StartupName']="Ola"
    if row['StartupName']=="OlaCabs":
        row['StartupName']="Ola"
    if row['StartupName']=="Paytm Marketplace":
        row['StartupName']="Paytm"
    if row['InvestmentType']=="Crowd funding":
        row['InvestmentType']="Crowd Funding"
    if row['InvestmentType']=="SeedFunding":
        row['InvestmentType']="Seed Funding"
    if (row['InvestmentType']=="Seed Funding") or (row['InvestmentType']=="Crowd funding"):
        i=row['InvestorsName']
        if ',' in i:
            h=i.split(',')
            for j in range(len(h)):
                company.append(row['StartupName'])
            for j in h:
                inv.append(j.strip())
        else:
            inv.append(i)
            company.append(row['StartupName'])

```

STEPS:

- We create a two empty lists inv and company.
- There are many errors in startup names. Ignore correcting all; just handle the important ones - Ola, Flipkart, Oyo and Paytm. Thus we have converted all the related Startup names to the important ones on iterating through row dictionaries in file_list. Correct spelling of investment types is - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. But since we concerned more about "Seed Funding" and "Crowd Funding", we have corrected only their names.
- If the InvestmentType is "Seed Funding" or "Crowd Funding", then we take investors name in i variable. If it contains many names separated by ",", we split all the names and put into list b which again iterated and the names are appended in list inv and thus we add Startup name corresponding to all the investor names in list company times equal to length of list b.
- If there is no "," in i, just append investor name in list inv and append Startup name in list b once.

```

d={}
for i in range(len(inv)):
    if company[i]!='':
        if inv[i] in d:
            d[inv[i]].add(company[i])
        else:
            d[inv[i]]=set()
            d[inv[i]].add(company[i])
for i in d:
    k=len(d[i])
    d[i]=k

```

STEPS:

- We create an empty dictionary d. On iterating through both the lists inv and company, we keys as names of investors and values as set of all different companies where they have invested ignoring the null values. We have used a set here so as to get different companies of that particular investor and avoid repetition.
- Again we iterate on dictionary d, where we change the values of d equal to length of set as we need number of different companies where they have invested and are not concerned about the names at all. Finally we have a dictionary d with keys as investor names and values as number of different companies they have invested in.

```

investors=[]
freq=[]
del d['']
for i in range(5):
    val=''
    maxval=-1
    for i in d:
        if d[i]>maxval:
            maxval=d[i]
            val=i
    investors.append(val)
    freq.append(maxval)
    del d[val]
for i in range(len(freq)):
    print(investors[i],freq[i])

```

STEPS:

- We create two empty lists investors and freq. Then we delete the part of dictionary which has null values as we have to ignore undisclosed investors.
- Then we iterate over the dictionary d in order to get the top 5 investors along with the number of times they have invested. We append the names of top 5 investors in list investors and their frequencies in list freq.
- Thus we print the output in format (Investor name, number of times he has invested).

Final Output:

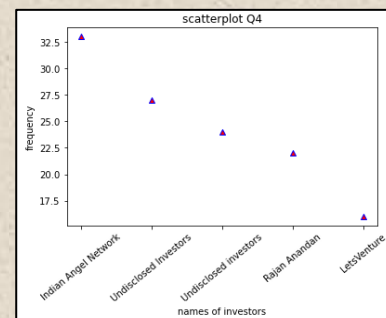
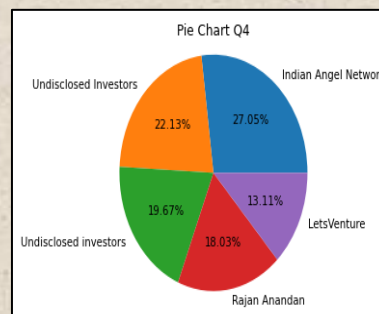
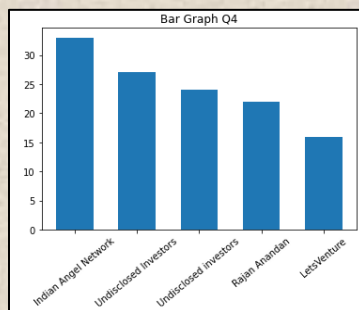
```

Indian Angel Network 33
Undisclosed Investors 27
Undisclosed investors 24
Rajan Anandan 22
LetsVenture 16

```

- Thus we have printed the output in format (Investor name, number of times he has invested).
- Thus we have the top 5 investors who have invested maximum number of times in different companies in investment type as “Crowd Funding” or “Seed Funding”.
1. Indian Angel Network 2. Undisclosed Investors 3. Undisclosed investors 4. Rajan Anandan 5. LetsVenture

Plots and Graphs:



- I have made a bar graph though not asked in the question which gives frequency of top 5 investors.
- I have even made a scatterplot of top 5 investors with their frequencies.
- I have even made a pie chart though not asked in the question which gives percentage of investment done by top 5 investors.

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 savior to help your friend and want to create a list of probable new new investors. Before moving forward you remember your investor friend advice that finding the investors by analyzing 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 is - "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.

Justification:

```
#Q5
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import csv
with open('C:/Users/Administrator.10000868-WOR-LP/Desktop/coding ninjas program/startup_funding.csv') as file_obj:
    file_data=csv.DictReader(file_obj,skipinitialspace=True)
    file_list=list(file_data)
```

STEPS:

- First, we open the csv file "startup_funding.csv" and put in file_obj.
- Then, we get the Dictionaries of all rows in form (column name: value) with skipping the initial spaces if there exists in the strings. Therefore, we put all the Dictionaries in the file_data.
- Thus we form a list of all dictionaries of file_data named as file_list.

```
inv=[]
company=[]
for row in file_list:
    if row['StartupName']=="flipkart.com":
        row['StartupName']="Flipkart"
    if row['StartupName']=="Oyo Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OYO Rooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Oyorooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="OyoRooms":
        row['StartupName']="Oyo"
    if row['StartupName']=="Ola Cabs":
        row['StartupName']="Ola"
    if row['StartupName']=="OlaCabs":
        row['StartupName']="Ola"
    if row['StartupName']=="Paytm Marketplace":
        row['StartupName']="Paytm"
    if row['InvestmentType']=="PrivateEquity":
        row['InvestmentType']="Private Equity"
    if row['InvestmentType']=="Private Equity":
        i=row['InvestorsName']
        if ',' in i:
            b=i.split(',')
            for j in range(len(b)):
                company.append(row['StartupName'])
                for j in b:
                    inv.append(j.strip())
        else:
            inv.append(i)
            company.append(row['StartupName'])
```

STEPS:

- We create a two empty lists inv and company.
- There are many errors in startup names. Ignore correcting all; just handle the important ones - Ola, Flipkart, Oyo and Paytm. Thus we have converted all the related Startup names to the important ones on iterating through row dictionaries in file_list. Correct spelling of investment types is - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. But since we concerned more about "Private Equity", we have corrected only their names.
- If the InvestmentType is "Private Equity", then we take investors name in i variable. If it contains many names separated by ",", we split all the names and put into list b which again iterated and the names are appended in list inv and thus we add Startup name corresponding to all the investor names in list company times equal to length of list b.
- If there is no "," in i, just append investor name in list inv and append Startup name in list b once.

```

d={}
for i in range(len(inv)):
    if company[i]!='':
        if inv[i] in d:
            d[inv[i]].add(company[i])
        else:
            d[inv[i]]=set()
            d[inv[i]].add(company[i])
for i in d:
    k=len(d[i])
    d[i]=k

```

STEPS:

- We create an empty dictionary d. On iterating through both the lists inv and company, we keys as names of investors and values as set of all different companies where they have invested ignoring the null values. We have used a set here so as to get different companies of that particular investor and avoid repetition.
- Again we iterate on dictionary d, where we change the values of d equal to length of set as we need number of different companies where they have invested and are not concerned about the names at all. Finally we have a dictionary d with keys as investor names and values as number of different companies they have invested in.

```

investors=[]
freq=[]
del d['']
for i in range(5):
    val=''
    maxval=-1
    for i in d:
        if d[i]>maxval:
            maxval=d[i]
            val=i
    investors.append(val)
    freq.append(maxval)
    del d[val]
for i in range(len(freq)):
    print(investors[i],freq[i])

```

STEPS:

- We create two empty lists investors and freq. Then we delete the part of dictionary which has null values as we have to ignore undisclosed investors.
- Then we iterate over the dictionary d in order to get the top 5 investors along with the number of times they have invested. We append the names of top 5 investors in list investors and their frequencies in list freq.
- Thus we print the output in format (Investor name, number of times he has invested).

Final Output:

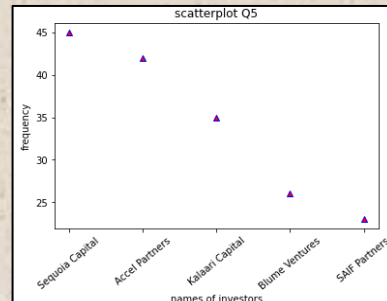
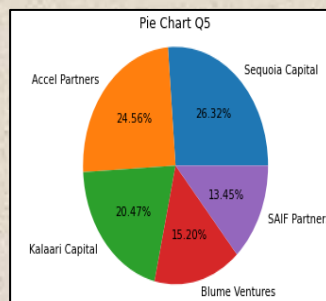
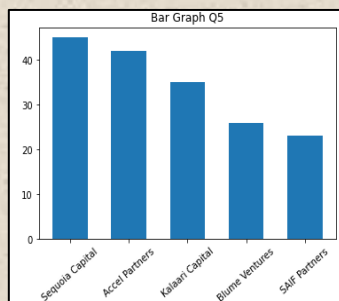
```

Sequoia Capital 45
Accel Partners 42
Kalaari Capital 35
Blume Ventures 26
SAIF Partners 23

```

- Thus we have printed the output in format (Investor name, number of times he has invested).
- Thus we have the top 5 investors who have invested maximum number of times in different companies in investment type as "Private Equity".
1. Sequoia Capital 2.Accel Partners 3.Kalaari Capital 4.Blume Ventures 5.SAIF Partners

Plots and Graphs:



- I have made a bar graph though not asked in the question which gives frequency of top 5 investors.
- I have even made a scatterplot of top 5 investors with their frequencies.
- I have even made a pie chart though not asked in the question which gives percentage of investment done by top 5 investors.