Project : Case Study (Part - II) - PDF

Case Study : Questions

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 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.

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.

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.

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.

5. 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 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.

This PDF contains explanation of code in chunks along with the Pie charts

ASK 1

Location where the most number of funding is done means, location where startups has received funding maximum number of times. can choose only between three locations - Bangalore, Mumbai, and NCR.

ANSWER

```
# lets import needed python libraries like pandas, numpy and matplotlib

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

# Now lets make a dataframe of startup csv file

startup =
pd.read_csv(r"C:\Users\vigya\PycharmProjects\Temp1Learn\startup_funding.csv")
df = startup.copy()
```

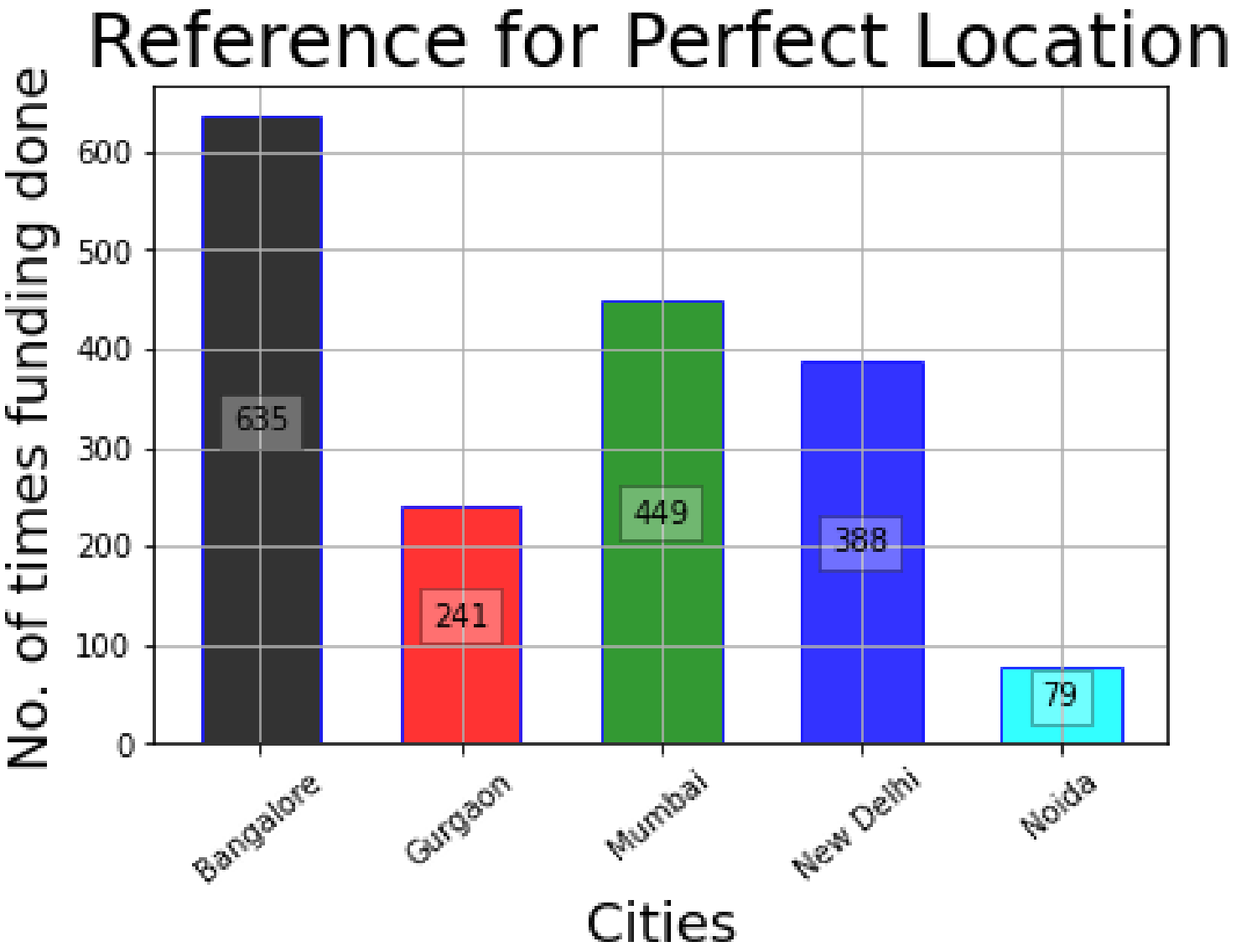
```
# For our task lets make a reference to the column "CityLocation" and drop the NaN values from it

c = df["CityLocation"]
c.dropna(inplace=True)

# Converting the particular column to "c" numpy array and changing all "Delhi" -> "New Delhi" & "bangalore" -> "Bangalore"

c = c.values
c[c == "Delhi"] = "New Delhi"
c[c == "bangalore"] = "Bangalore"

# Traversed through the "c" numpy array and appended normally or after spliting on " / " to a new list "city"
# only if location is 'Bangalore' or 'Mumbai' or 'New Delhi' or 'Noida' or 'Gurgaon'

city = []
for i in range(c.size):
    if " / " in c[i]:
        temp = c[i].split(" / ")
        for j in temp:
            if j == 'Bangalore' or j == 'Mumbai' or j == 'New Delhi' or j == 'Noida' or j == 'Gurgaon':
                city.append(j)
    else:
        if c[i] == 'Bangalore' or c[i] == 'Mumbai' or c[i] == 'New Delhi' or c[i] == 'Noida' or c[i] == 'Gurgaon':
            city.append(c[i])

# Converting list to nparray and finding its unique elements int "loc" and its corresponding counts into "fundtimes"
#also making a copy of counts for use in furthur plotting graph

city = np.array(city)
uncity = np.unique(city, return_counts=True)
loc = uncity[0]
fundtimes = uncity[1]
barfundtimes = uncity[1].copy()

# Taversing through "loc" - unique elements finding max index and fetching city with max time it got investment
# making current fundtime i.e max index as 0 so that furthur on next traversal it is not maximum

for i in range(loc.size):
```

```
    ind = np.argmax(fundtimes)
    print(loc[ind], fundtimes[ind])
    fundtimes[ind] = 0
```

Output : -

```
Bangalore 635
Mumbai 449
New Delhi 388
Gurgaon 241
Noida 79
```

```
# As we have recieved the required details we will plot a bar graph

def labels(loc, barfundtimes):
    for i in range(loc.size):
        plt.text(i, barfundtimes[i]//2, barfundtimes[i], ha='center', Bbox=dict(facecolor='white',
alpha=0.3))

plt.bar(loc, barfundtimes, width=0.6, color=['black', 'red', 'green', 'blue', 'cyan'],
edgecolor="blue", linewidth=1, alpha=0.8)
labels(loc, barfundtimes)
style.use('ggplot')
plt.title("Reference for Perfect Location", size=25)
plt.xticks(rotation=40)
plt.xlabel("Cities", size = 18)
plt.ylabel("No. of times funding done", size = 18)
plt.grid()
plt.figure(figsize=(16, 9))
plt.show()
```



Answer to ASK 1

It can be clearly seen in output as well as in graph that Bangalore can be considered as the perfect location as startups have received funding maximum number of times in Bangalore.

ASK 2

Top 5 investors who have invested maximum number of times consider repeat investments in one company also. keeping in mind to ignore undisclosed investors

ANSWER

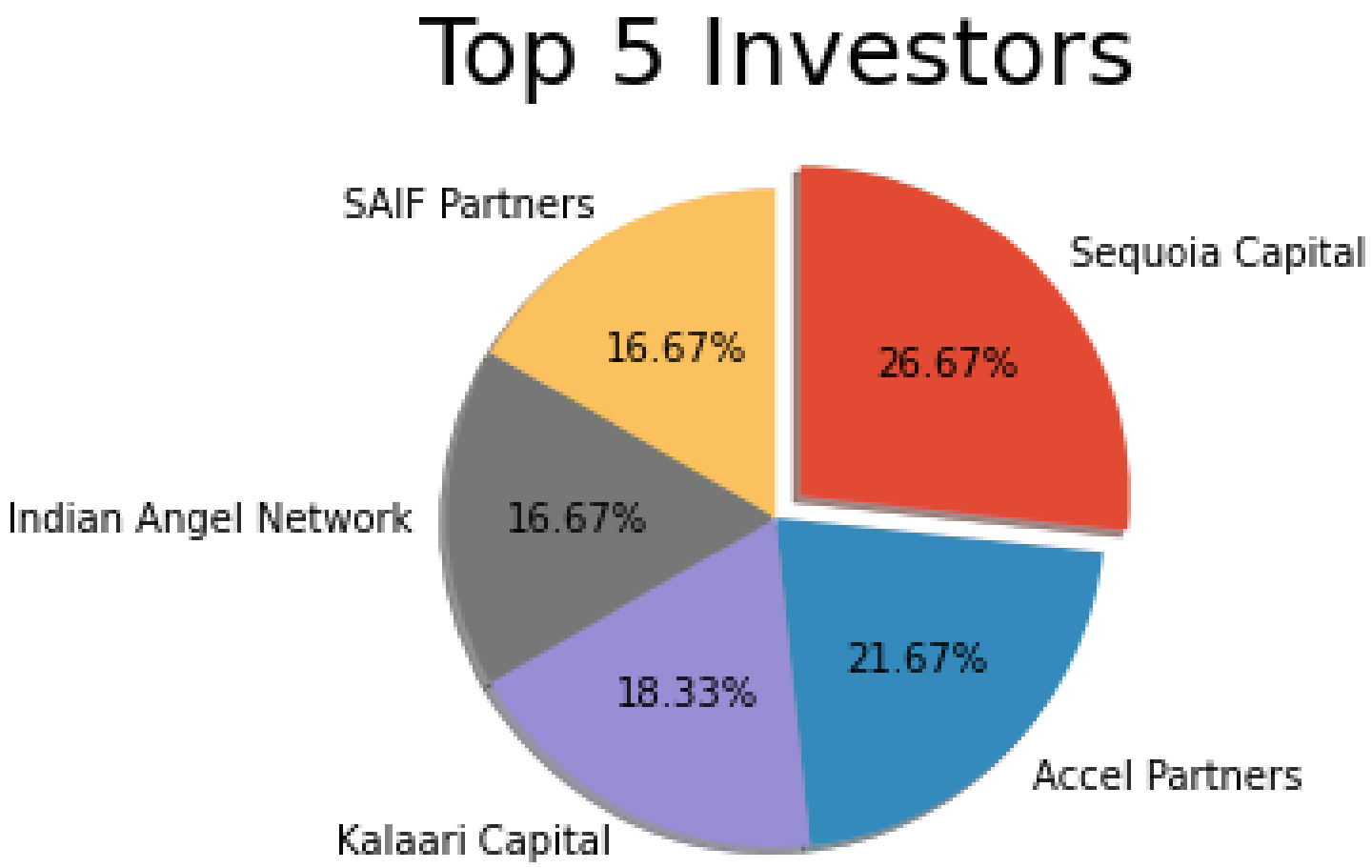
```
# Lets take the column "InvestorsName" and create a reference to it and drop all NaN values from it and create it as a nparray

investor = df["InvestorsName"]
investor.dropna(inplace=True)
investor = investor.values

# "InvestorsName" has ", " -> comma and space seprated values
# We will split on this basis ", " and append them also in f1investor

f1investor = []
for i in range(investor.size):
    if ", " in investor[i]:
        temp = investor[i].split(", ")
        for j in temp:
            f1investor.append(j)
    else:
        f1investor.append(investor[i])

# f1investor also has only comma seprated values now now ","
# we will again split f1investor on the basis of "," and append it in finvestor
# Also we will include conditon of not including if it has Others / empty / undisclosed

finvestor = []
for i in range(len(f1investor)):
    if "," in f1investor[i]:
        temp = f1investor[i].split(",")
        for j in temp:
            if "Others" in j or "Undisclosed" in j or j == "":
                continue
            finvestor.append(j)
    else:
        if "Others" in f1investor[i] or "Undisclosed" in f1investor[i] or f1investor[i] == "":
            continue
        finvestor.append(f1investor[i])

# Now we have finvestor containing all seprate values i.e seprate names of investors
# We will get unique investors and their count by :
```

```
uinvestor = np.unique(finvestor, return_counts=True)
inv = uinvestor[0]
count = uinvestor[1]

# For storing data to creat a pie chart we can have seprate empty list

piecount = []
pieinv = []

# Now we will run loop for 5 times and find max index on basis of count each time
# Using that index print investor name and its corresponding count
# Then making that count location 0 , so that next time 2nd heighest will be taken as max
# Also side by side appending elements in empty list we took for creating Pie chart later

for i in range(5):
    ind = np.argmax(count)
    print(inv[ind], count[ind])
    piecount.append(count[ind])
    pieinv.append(inv[ind])
    count[ind] = 0
```

Output :-

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

```
# Plotting pie Chart

plt.pie(piecount, labels=pieinv, autopct="%.2f%%", counterclock=False, startangle=90,
        explode=[0.1, 0, 0, 0, 0], shadow=True)
style.use('ggplot')
plt.title("Top 5 Investors", size=25)
plt.grid()
plt.figure(figsize=(16, 9))
plt.show()
```



Answer to ASK 2

Pie chart clearly specifies the top 5 investors, and among them "Sequoia Capital" has done maximum investment overall.

ASK 3

Top 5 investors who have invested maximum number of times in different companies
errors in startup names - important ones - Ola, Flipkart, Oyo and Paytm.

ANSWER

For this ask we will work upon two columns "InvestorsName" & "StartupName"

```
# For "InvestorsName" we already had a list finvestor containing all seprate values i.e seprate names of investors
# For "StartupName" we will take this column values in nparray

startup = df["StartupName"].values

# Now we will set up conditions for corrections in important startups

startup[startup == "Oyorooms"] = "Oyo"
startup[startup == "Oyo Rooms"] = "Oyo"
startup[startup == "OYO Rooms"] = "Oyo"
startup[startup == "OyoRooms"] = "Oyo"

startup[startup == "Ola Cabs"] = "Ola"
startup[startup == "Olacabs"] = "Ola"

startup[startup == "Flipkart.com"] = "Flipkart"

startup[startup == "Paytm Marketplace"] = "Paytm"

# We have corrected startup column now we will work on investorname column
# We will fill NaN values in "InvestorsName" with some unknown value -> "xx" and then converting it to nparray

investor = df["InvestorsName"]
investor.fillna("xx", inplace=True)
investor = investor.values

# Here size of both "InvestorsName" & "StartupName" is same
# We also have finvestor which is containing all seprate values i.e seprate names of investors
# We will take unique investors name from finvestor into uinvestor

uinvestor = np.unique(finvestor)

# Now we will Traverse over investorName and Split elements firstly on the basis of ", "
# And inplace of appending into other list we will place after spliting at that particular location only

for i in range(investor.size):
    if ", " in investor[i]:
        investor[i] = investor[i].split(", ")

# Now we have list in place of elements which had ", " (comma with space) in them
# But there are certain elements which also had only "," (comma) in them
```

```
# So we will again Traverse over investorName and Split elements secondly on the basis of ","
# And inplace of appending into other list we will place after spliting at that particular location only

for i in range(investor.size):
    if type(investor[i]) is list:
        for j in range(len(investor[i])):
            if "," in investor[i][j]:
                investor[i][j] = investor[i][j].split(",")
# Now we have a list in place of elements which had "," (comma) in them
# That is we may have list inside list in place of elements
# so, investor is having string elements , list having strng elements , list having string and list also

# Lets Now take an empty list "checkinvestor" and a dictonary with empty string of keys startup column

checkinvestor = []
tempdic = {}
for i in startup:
    tempdic[i] = ""

# We will Traverse over Unique Investors (uinvestor) we created from finvestor
# take each its element (inve) and then traverse over just created investor
# if investor has the unique investors name (inve)
# we will check if unique investors name (inve) is in dictonary with correcponding starupname as key
# if yes we will pass this iteration
# if not , we will add the string (unique investors name (inve)) in dictonary with correcponding starupname as key
# And append the empty list "checkinvestor" with the unique investor name(inve)

for inve in uinvestor:
    for i in range(len(investor)):
        if inve in investor[i]:
            if inve in tempdic[startup[i]]:
                continue
            else:
                tempdic[startup[i]] += inve
                checkinvestor.append(inve)


# Now we will will get unique investors from just created list "checkinvestor" with their corresponding counts
```

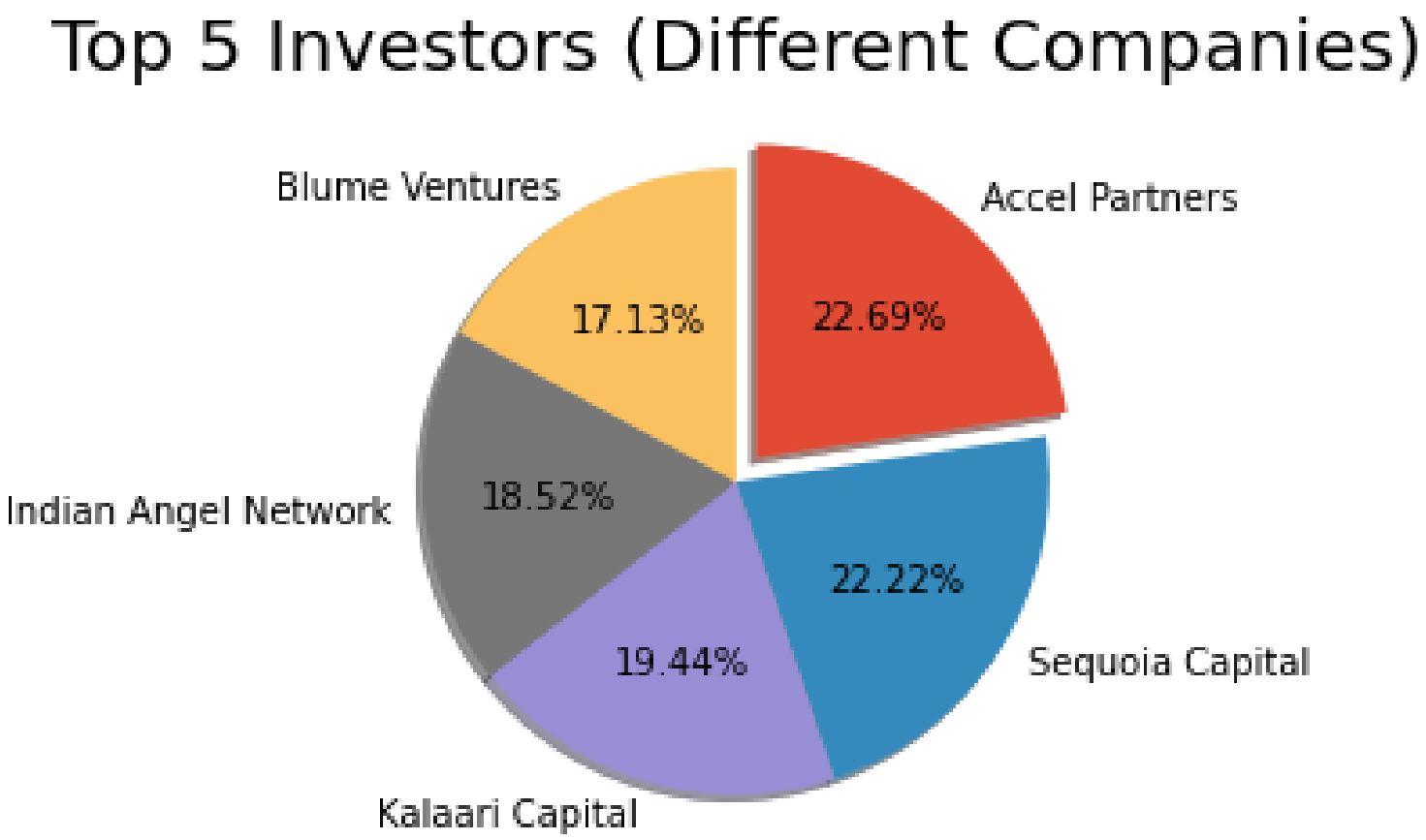
```
checkinvestor = np.array(checkinvestor)
ucheckinvestor = np.unique(checkinvestor, return_counts=True)

# Unique investors name

uin = ucheckinvestor[0]

# Unique investors counts

uco = ucheckinvestor[1]

# For storing data to create a pie chart we can have seprate empty list

pieinv = []
piecount = []

# Now we will run loop for 5 times and find max index on basis of count each time
# Using that index print investor name and its corresponding count
# Then making that count location 0 , so that next time 2nd heighest will be taken as max
# Also side by side appending elements in empty list we took for creating Pie chart later

for i in range(5):
    ind = np.argmax(uco)
    print(uin[ind], uco[ind])
    pieinv.append(uin[ind])
    piecount.append(uco[ind])
    uco[ind] = 0
```

Output:-

```
Accel Partners 49
Sequoia Capital 48
Kalaari Capital 42
Indian Angel Network 40
Blume Ventures 37
```

```
# plotting Pie chart

plt.pie(piecount, labels=pieinv, autopct="%.2f%%", counterclock=False, startangle=90,
        explode=[0.1, 0, 0, 0, 0], shadow=True)
style.use('ggplot')
plt.title("Top 5 Investors (Different Companies)", size=20)
plt.grid()
plt.figure(figsize=(16, 9))
```

```
plt.show()
```



Answer to ASK 3

Pie chart clearly specifies the top 5 investors, and among them "Accel Partners" has done maximum investment in different companies.

ASK 4

Top 5 investors who have invested maximum number of times in different companies
investment type is Crowdfunding or Seed Funding.
Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding"

errors in startup names - important ones - Ola, Flipkart, Oyo and Paytm.

ANSWER

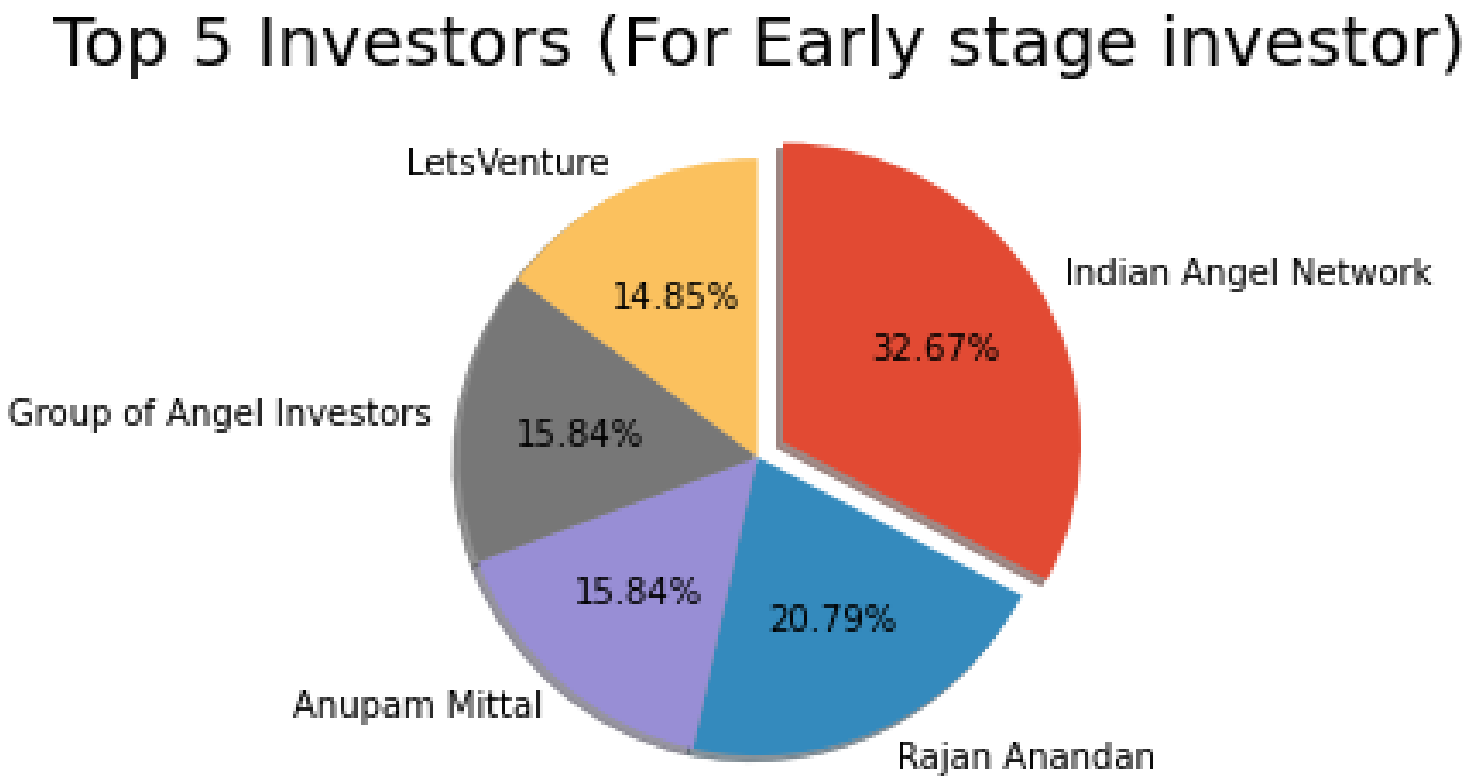
```
# For this ask we will work upon three columns "InvestorsName" & "StartupName" & "InvestmentType"
# For "InvestorsName" we already had a list finvestor containing all seprate values i.e seprate names of investors
# and we made a unique invetors name from finster -> uinvestor
# We also have investor with string elements , list having strng elements , list having string and list also
# We have startup with all the corrections made in startupname
# We just need to get investmentType done.

# Taking "InvestmentType" column and filling all NaN with temporary value - > "xx"
# And then converting it to nparray

df["InvestmentType"].fillna("xx", inplace=True)
investmenttype = df["InvestmentType"].values

# Correct spelling of investment types

investmenttype[investmenttype == "Crowd funding"] = 'Crowd Funding'
investmenttype[investmenttype == 'SeedFunding'] = 'Seed Funding'
investmenttype[investmenttype == 'PrivateEquity'] = 'Private Equity'

# Lets Now take an empty list "checkinvestor" and a dictonary with empty string of keys startup column

checkinvestor = []
tempdic = {}
for i in startup:
    tempdic[i] = ""

# We will Traverse over Unique Investors (uinvestor) we created from finvestor
# take each its element (inve) and then traverse over just created investor
# if investor has the unique investors name (inve)
# we will check if unique investors name (inve) is in dictonary with correcponding starupname as key
# if yes we will pass this iteration
# if not , we will check if investmenttype is "Crowd Funding" or 'Seed Funding'
```

```
# we will add the string (unique investors name (inve)) in dictonary with correcponding starupname as key
# And append the empty list "checkinvestor" with the unique investor name(inve)

for inve in uinvestor:
    for i in range(len(investor)):
        if inve in investor[i]:
            if inve in tempdic[startup[i]]:
                continue
            else:
                if investmenttype[i] == "Crowd Funding" or investmenttype[i] == 'Seed Funding':
                    tempdic[startup[i]] += inve
                    checkinvestor.append(inve)

# Now we will will get unique investors from just created list "checkinvestor" with their corresponding counts

checkinvestor = np.array(checkinvestor)
ucheckinvestor = np.unique(checkinvestor, return_counts=True)

# Unique investors name

uin = ucheckinvestor[0]

# Unique investors counts

uco = ucheckinvestor[1]

# For storing data to create a pie chart we can have seprate empty list

pieinv = []
piecount = []

# Now we will run loop for 5 times and find max index on basis of count each time
# Using that index print investor name and its corresponding count
# Then making that count location 0 , so that next time 2nd heighest will be taken as max
# Also side by side appending elements in empty list we took for creating Pie chart later

for i in range(5):
    ind = np.argmax(uco)
    print(uin[ind], uco[ind])
    pieinv.append(uin[ind])
    piecount.append(uco[ind])
    uco[ind] = 0
```

Output :-

```
Indian Angel Network 33
Rajan Anandan 21
Anupam Mittal 16
Group of Angel Investors 16
LetsVenture 15
```

```
# Plotting Pie Chart

plt.pie(piecount, labels=pieinv, autopct="%.2f%%", counterclock=False, startangle=90,
        explode=[0.1, 0, 0, 0, 0], shadow=True)
style.use('ggplot')
plt.title("Top 5 Investors (For Early stage investor)", size=20)
plt.grid()
plt.figure(figsize=(16, 9))
plt.show()
```



Answer to ASK 4

Pie chart clearly specifies the top 5 investors, and among them "Indian Angel Network" has done maximum investment in different

companies
and investment type is "Crowd Funding" or 'Seed Funding' for early stage

ASK 5

Top 5 investors who have invested maximum number of times in different companies
investment type is Private Equity
Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding"
errors in startup names - important ones - Ola, Flipkart, Oyo and Paytm.

ANSWER

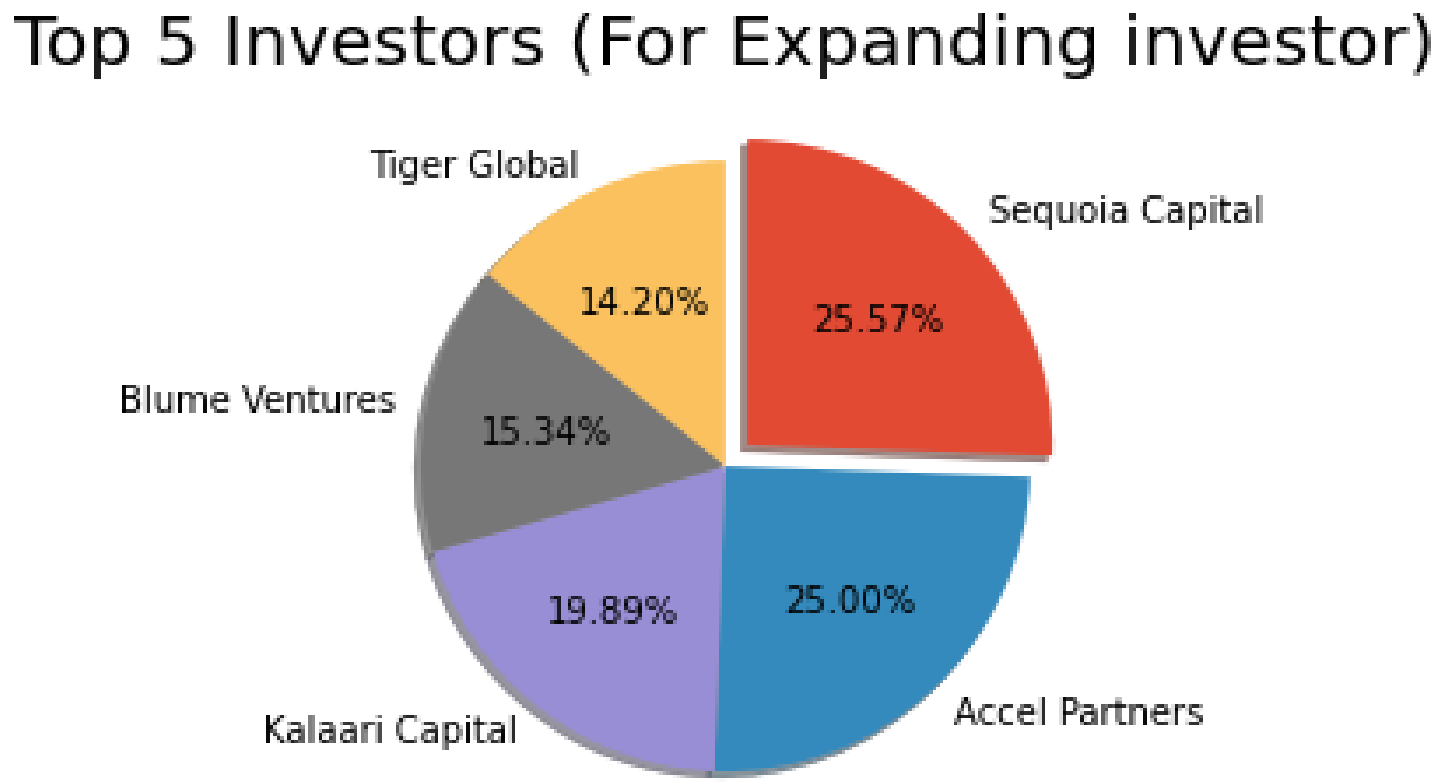
```
# For this ask we will work upon three columns "InvestorsName" & "StartupName" & "InvestmentType"
# For "InvestorsName" we already had a list finvestor containing all seprate values i.e seprate names of investors
# and we made a unique investors name from finvestor -> uinvestor
# We also have investor with string elements , list having strng elements , list having string and list also
# We have startup with all the corrections made in startupname
# We also have investmentType done.

# Lets Now take an empty list "checkinvestor" and a dictonary with empty string of keys startup column

checkinvestor = []
tempdic = {}
for i in startup:
    tempdic[i] = ""

# We will Traverse over Unique Investors (uinvestor) we created from finvestor
# take each its element (inve) and then traverse over just created investor
# if investor has the unique investors name (inve)
# we will check if unique investors name (inve) is in dictonary with correcponding starupname as key
```

```
# if yes we will pass this iteration
# if not , we will check if investmenttype is "Private Equity"
# we will add the string (unique investors name (inve)) in dictonary with correcponding starupname as key
# And append the empty list "checkinvestor" with the unique investor name(inve)

for inve in uinvestor:
    for i in range(len(investor)):
        if inve in investor[i]:
            if inve in tempdic[startup[i]]:
                continue
            else:
                if investmenttype[i] == "Private Equity":
                    tempdic[startup[i]] += inve
                    checkinvestor.append(inve)

# Now we will will get unique investors from just created list "checkinvestor" with their corresponding counts

checkinvestor = np.array(checkinvestor)
ucheckinvestor = np.unique(checkinvestor, return_counts=True)

# Unique investors name

uin = ucheckinvestor[0]

# Unique investors counts

uco = ucheckinvestor[1]

# For storing data to create a pie chart we can have seprate empty list

pieinv = []
piecount = []

# Now we will run loop for 5 times and find max index on basis of count each time
# Using that index print investor name and its corresponding count
# Then making that count location 0 , so that next time 2nd heighest will be taken as max
# Also side by side appending elements in empty list we took for creating Pie chart later

for i in range(5):
    ind = np.argmax(uco)
    print(uin[ind], uco[ind])
    pieinv.append(uin[ind])
    piecount.append(uco[ind])
    uco[ind] = 0
```

Output :-

```
Sequoia Capital 45
Accel Partners 44
Kalaari Capital 35
Blume Ventures 27
Tiger Global 25
```

```
# Plotting Pie Chart

plt.pie(piecount, labels=pieinv, autopct="%.2f%%", counterclock=False, startangle=90,
        explode=[0.1, 0, 0, 0, 0], shadow=True)
style.use('ggplot')
plt.title("Top 5 Investors (For Expanding investor)", size=20)
plt.grid()
plt.figure(figsize=(16, 9))
plt.show()
```



Answer to ASK 5

Pie chart clearly specifies the top 5 investors, and among them "Sequoia Capital" has done maximum investment in different companies
and investment type is "Private Equity" for expanding startup stage