

First Question

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.

Explanation

1)

In this question, we have to find the perfect location to establish the product startup where getting the investment has a high chance among the Bangalore, Mumbai, **and** NCR(New Delhi, Gurgaon, Noida).

First of all, CityLocation can be separately stored **in** dff (pandas series object) from the dataframe.

Now, remove all the NaN value from the dff.

As dff can contain values like: ("Mumbai/Maharashtra") so split it using `split("/")` **and return** the string which **is** at "0th" index after striping using (`strip()` method so that it can vanish the leading space).

Now **for** each city count the number of investement had done **in** that city store it **in** cityNumber.

for NCR=number of total investement **is** of New Delhi,Gurgaon **and** Noida

then plot the graph such that, Number of Funding against city (NCR,Mumbai, Bangalore)

From the garph it **is** clear that, the number of Funding **in** NCR **is** maximum.

Code

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

```

df=pd.read_csv("onlyfiles/startup_funding.csv",encoding="utf-8")
dff=df['CityLocation'].dropna()

def seprateCity(city):
    return city.split('/')[0].strip().title()

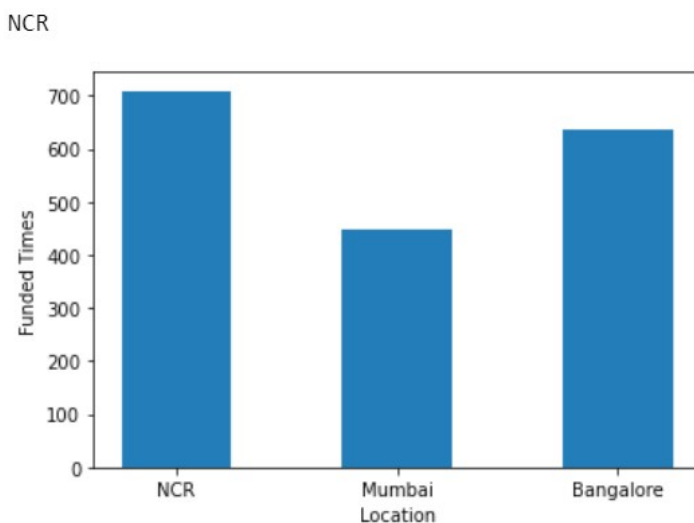
dff=dff.apply(seprateCity)
dff.replace("Delhi","New Delhi",inplace=True)
dff.replace("banglore","Bangalore",inplace=True)
cityNumber=dff.value_counts()

NCR=cityNumber['New Delhi']+cityNumber['Gurgaon']+cityNumber['Noida']
Bangalore=cityNumber['Bangalore']
Mumbai=cityNumber['Mumbai']

location=["NCR","Mumbai","Bangalore"]
FundedTimes=[NCR,Mumbai,Bangalore]
print("NCR")
plt.bar(location,FundedTimes,width=0.5)
plt.xlabel("Location")
plt.ylabel("Funded Times")
plt.show()

```

Output:



Second Question

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.

Explanation

In this, we have to find the top 5 investor who invested maximum number of times in startup.

So, first of all store the "InvestorName" separately into investor(vairable)

Now, drop the NaN value from investor.

Then, for each investor count the number of times they investement in startup for that we created a function InvestorDictionary which return dictionary such that it's key store name of investor and it's value store number of times they have invested.

Then, store the dictionary into dataf(dataframe) and sort in descending order.

Print first 5 investor name they are the top 5 investor who invested maximum number of times.

Code

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

df=pd.read_csv("onlyfiles/startup_funding.csv",encoding="utf-8")
investor=df['InvestorsName'].dropna()

def InvestorDictionary(array):
    Investor=dict()
    for i in array:
        if ',' not in i:
            if i in Investor:
                Investor[i]=Investor.get(i)+1
            else:
                Investor[i]=1
        else:
            string=i.strip().split(',')
            for j in string:
                if j.strip() in Investor:
                    Investor[j.strip()]=Investor[j.strip()]+1
                else:
                    Investor[j.strip()]=1
    return Investor
Investor=InvestorDictionary(investor)

dataf=pd.DataFrame(list(Investor.values()),list(Investor.keys()))
dataf=dataf.sort_values(by=[0],ascending=False)[:5]
investFrequency=list()
investorName=list()
```

```

for i in range(5):
    ans=int(dataf.values[i])
    investFrequency.append(ans)
    investorName.append(dataf.index[i])
    print(dataf.index[i],ans)
plt.xticks(rotation=65)
plt.bar(investorName,investFrequency)
plt.show()

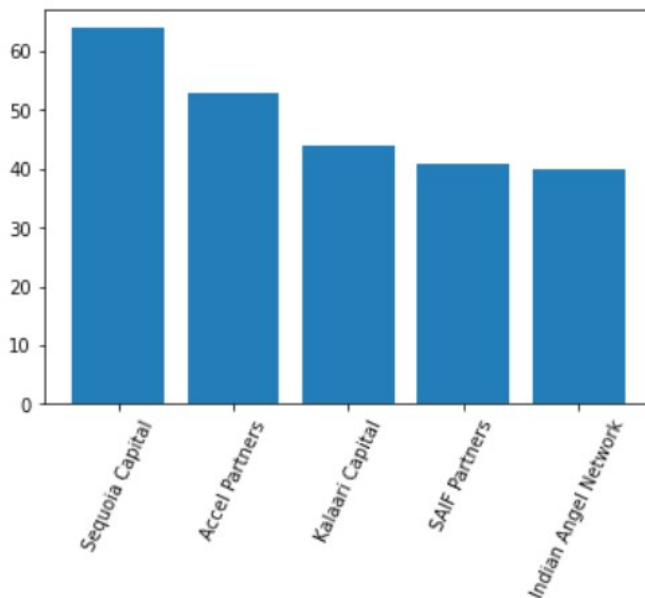
```

Output:

```

Sequoia Capital 64
Accel Partners 53
Kalaari Capital 44
SAIF Partners 41
Indian Angel Network 40

```



Third Question

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 improve 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.

Explanation

In this, we have to find the top 5 investor who invested maximum number of times in different startup.

So, first of all drop the NULL values from the InvestorsName and StartupName.

Then, correct spelling of top startup.

Now, for each investor store the startup in which they invested take care that if ones they invested in startup then the same investor can't invest in that startup as we have to find the maximum number of investement in different startup.

So for that we can use "defaultdict" from the collections

We store the value such that:

```
dict['InvestorName']={startup1,startup2,startup3,...}
```

so, now we can give that in how much different startup the investor invested by giving the size of dict['InvestorName'] for investor.

Hence, store the investor name key and size of key(variable) in val(variable)

then, form a dataframe using key and val
sort it in descending order

Then, print the first 5 investor, they are the top 5 investor who invested in different startup.

Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

df=pd.read_csv("onlyfiles/startup_funding.csv",encoding="utf-8")
df['InvestorsName'].dropna(inplace=True)
df['StartupName'].dropna(inplace=True)
df['StartupName'].replace("Olacabs","Ola",inplace=True)
df['StartupName'].replace("Ola Cabs","Ola",inplace=True)
df['StartupName'].replace("Flipkart.com","Flipkart",inplace=True)
df['StartupName'].replace("Paytm Marketplace","Paytm",inplace=True)
df['StartupName'].replace("Oyo Rooms","Oyo",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.reset_index(inplace=True)

def InvestorDictionay():
    Investor=defaultdict(list)
    for i in range(0,len(df)):
        startupName=df['StartupName'][i]
```

```

        investor=str(df['InvestorsName'][i]).strip().split(',')
        for j in range(len(investor)):
            string=investor[j].strip()
            if(len(string)==0 or string=="Undisclosed Investors" or
string=="Undisclosed investors"):
                continue
            if string not in Investor:
                Investor[string].append(startupName)
            elif startupName not in Investor[string]:
                Investor[string].append(startupName)
        return Investor

Investor=InvestorDictionay()
key=Investor.keys()
vla=Investor.values()
Invest=dict()
for i in key:
    Invest[i]=len(Investor[i])
dataf=pd.DataFrame(list(Invest.values()),list(Invest.keys()))
dataf=dataf.sort_values(by=[0],ascending=False)[:5]
investFrequency=list()
investorName=list()
for i in range(5):
    ans=int(dataf.values[i])
    investFrequency.append(ans)
    investorName.append(dataf.index[i])
    print(dataf.index[i],ans)
plt.xticks(rotation=65)
plt.bar(investorName,investFrequency)
plt.show()

```

Output:

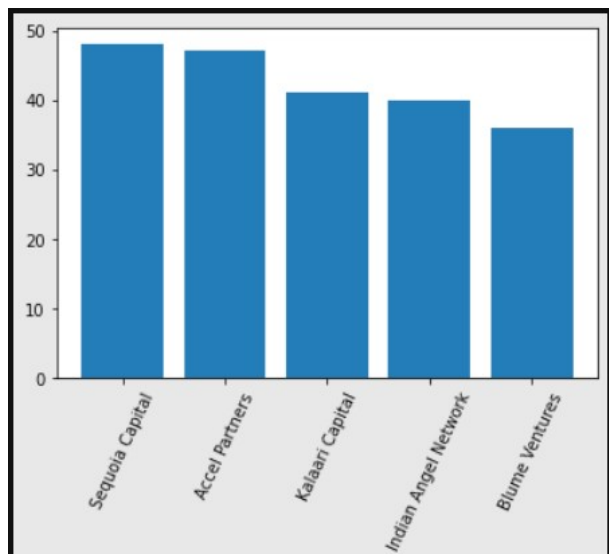
Sequoia Capital 48

Accel Partners 47

Kalaari Capital 41

Indian Angel Network 40

Blume Ventures 36



Fourth Question

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.

Explanation

In this, we have to find the top 5 investor who invested maximum number of times **in** different startup having investment type Seed Funding **or** Crowdfunding.
So, first of all drop the NULL values form the InvestorsName ,

StartupName and InvestmentType.

Then, correct spelling of top startup and the investment types.

Now, for each investor store the startup in which they invested take care that if ones they invested in startup then the same investor can't investe in that startup as we have to find the maximum number of investement in different startup.

So for that we can use "defaultdict" from the collections

We store the value such that:

```
dict['InvestorName']={startup1,startup2,startup3,...}
```

so, now we can give that in how much different startup the investor invested by giving the size of dict['InvestorName'] for investor.

If the investement type is "Debt Funding" or "Private Equity" then don't store it in dict.

Hence, store the investor name key and size of key(vairable) in val(variable)

then, form a dataframe using key and val
sort it in descending order

Then, print the first 5 investor, they are the top 5 investor who invested in different startup.

Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

df=pd.read_csv("onlyfiles/startup_funding.csv",encoding="utf-8")

df['InvestorsName'].dropna(inplace=True)
df['StartupName'].dropna(inplace=True)
df['InvestmentType'].dropna(inplace=True)
df['StartupName'].replace("Olacabs","Ola",inplace=True)
df['StartupName'].replace("Ola Cabs","Ola",inplace=True)
df['StartupName'].replace("Flipkart.com","Flipkart",inplace=True)
df['StartupName'].replace("Paytm Marketplace","Paytm",inplace=True)
df['StartupName'].replace("Oyo Rooms","Oyo",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['InvestmentType'].replace("SeedFunding","Seed
Funding",inplace=True)
df['InvestmentType'].replace("PrivateEquity","Private
Equity",inplace=True)
df['InvestmentType'].replace("Crowd funding","Crowd
Funding",inplace=True)
```



```

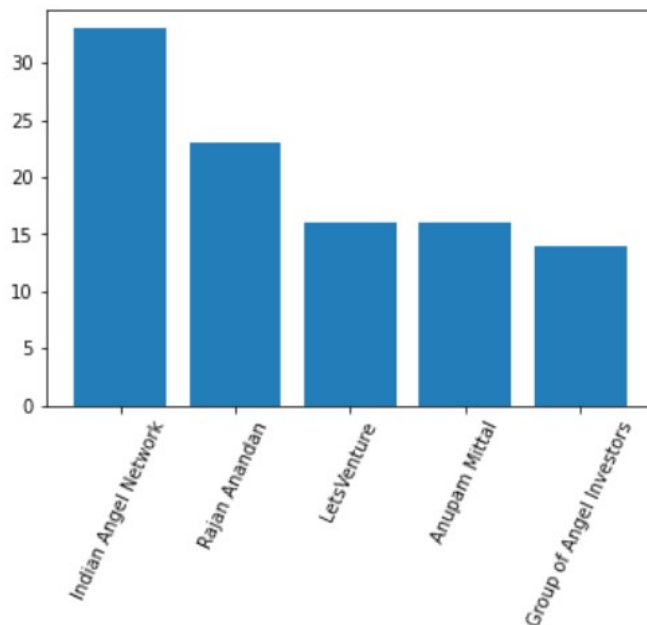
df.reset_index(inplace=True)
def InvestorDictionay():
    Investor=defaultdict(list)
    for i in range(0,len(df)):
        investmentType=df['InvestmentType'][i]
        if (investmentType=="Debt Funding" or investmentType=="Private
Equity"):
            continue
        startupName=df['StartupName'][i]
        investor=str(df['InvestorsName'][i]).strip().split(',')
        for j in range(len(investor)):
            string=investor[j].strip()
            if(len(string)==0 or string=="Undisclosed Investors" or
string=="Undisclosed investors"):
                continue
            if string not in Investor:
                Investor[string].append(startupName)
            elif startupName not in Investor[string]:
                Investor[string].append(startupName)
    return Investor

Investor=InvestorDictionay()
key=Investor.keys()
vla=Investor.values()
Invest=dict()
for i in key:
    Invest[i]=len(Investor[i])
dataf=pd.DataFrame(list(Invest.values()),list(Invest.keys()))
dataf=dataf.sort_values(by=[0],ascending=False)[:5]
investFrequency=list()
investorName=list()
for i in range(5):
    ans=int(dataf.values[i])
    investFrequency.append(ans)
    investorName.append(dataf.index[i])
    print(dataf.index[i],ans)
plt.xticks(rotation=65)
plt.bar(investorName,investFrequency)
plt.show()

```

Output:

```
Indian Angel Network 33
Rajan Anandan 23
LetsVenture 16
Anupam Mittal 16
Group of Angel Investors 14
```



Fifth Question

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.

Explanation

In this, we have to find the top 5 investor who invested maximum number of times **in** different startup having investment type **"Private Equity"**.

So, first of all drop the NULL values form the InvestorsName, StartupName **and** InvestmentType.

Then, correct spelling of top startup **and** InvestmentType.
Now, **for** each investor store the startup **in** which they invested take care that **if** ones they invested **in** startup then the same investor can't investe in that startup as we have to find the maximum number of investement in different startup.

So for that we can use "defaultdict" from the collections
We store the value such that:
 dict['InvestorName']={startup1,startup2,startup3,...}
so, now we can give that in how much different startup the investor invested by giving the size of dict['InvestorName'] for investor.
If the investement type is "Debt Funding" or "Seed Funding" or "Crowd Funding" then don't store it **in** dict.

Hence, store the investor name key **and** size of key(vairable) **in** val(variable)
then, form a dataframe using key **and** val
sort it **in** descending order

Then, print the first 5 investor, they are the top 5 investor who invested **in** different startup.

Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from collections import defaultdict

df=pd.read_csv("onlyfiles/startup_funding.csv",encoding="utf-8")

df['InvestorsName'].dropna(inplace=True)
df['StartupName'].dropna(inplace=True)
df['InvestmentType'].dropna(inplace=True)
df['StartupName'].replace("Olacabs","Ola",inplace=True)
df['StartupName'].replace("Ola Cabs","Ola",inplace=True)
df['StartupName'].replace("Flipkart.com","Flipkart",inplace=True)
df['StartupName'].replace("Paytm Marketplace","Paytm",inplace=True)
df['StartupName'].replace("Oyo Rooms","Oyo",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['InvestmentType'].replace("SeedFunding","Seed
Funding",inplace=True)
df['InvestmentType'].replace("PrivateEquity","Private
Equity",inplace=True)
df['InvestmentType'].replace("Crowd funding","Crowd
Funding",inplace=True)
```

```

df.reset_index(inplace=True)
def InvestorDictionary():
    Investor=defaultdict(list)
    for i in range(0,len(df)):
        investmentType=df['InvestmentType'][i]
        if (investmentType=="Crowd Funding" or investmentType=="Seed
Funding" or investmentType=="Debt Funding"):
            continue
        startupName=df['StartupName'][i]
        investor=str(df['InvestorsName'][i]).strip().split(',')
        for j in range(len(investor)):
            string=investor[j].strip()
            if(len(string)==0 or string=="Undisclosed Investors" or
string=="Undisclosed investors"):
                continue
            if string not in Investor:
                Investor[string].append(startupName)
            elif startupName not in Investor[string]:
                Investor[string].append(startupName)
    return Investor

```

```

Investor=InvestorDictionary()
key=Investor.keys()
vla=Investor.values()
Invest=dict()
for i in key:
    Invest[i]=len(Investor[i])
dataf=pd.DataFrame(list(Invest.values()),list(Invest.keys()))
dataf=dataf.sort_values(by=[0],ascending=False)[:5]
investFrequency=list()
investorName=list()
for i in range(5):
    ans=int(dataf.values[i])
    investFrequency.append(ans)
    investorName.append(dataf.index[i])
    print(dataf.index[i],ans)
plt.xticks(rotation=65)
plt.bar(investorName,investFrequency)
plt.show()

```

Output:

Sequoia Capital 45
Accel Partners 43
Kalaari Capital 35
Blume Ventures 27
SAIF Partners 24

