

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 numpy as np**

**df\_start=pd.read\_csv('startup\_funding.csv',encoding='utf-8')**

**df\_start['InvestorsName'].dropna(inplace=True)**

**def createDictionary(array):**

**dictionary={}**

**for i in array:**

**if ',' not in i:**

**if i in dictionary:**

**dictionary[i]=dictionary.get(i)+1**

**else: dictionary[i]=1**

**else:**

**string=i.strip().split(',')**

**for j in string:**

**if j.strip() in dictionary:**

**dictionary[j.strip()]=dictionary.get(j.strip())+1**

**else:**

**dictionary[j.strip()]=1**

**return dictionary**

**dictionary=createDictionary(df\_start['InvestorsName'])**

**dataf=pd.DataFrame(list(dictionary.values()),list(dictionary.keys()))**

**dataf=dataf.sort\_values(by=[0],ascending=False)**

**for i in range(5):**

**print(dataf.index[i],dataf.values[i][0])**

**ANS-Sequoia Capital 64**

**Accel Partners 53**

**Kalaari Capital 44**

**SAIF Partners 41**

**Indian Angel Network 40**

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 numpy as np**

**import csv**

**import matplotlib.pyplot as plt**

**with open('datasets/startup\_funding.csv', encoding='utf8') as file\_obj:**

**file\_data=csv.DictReader(file\_obj, skipinitialspace=True)**

**investors=[]**

**startup=[]**

**for row in file\_data:**

**if not ('Undisclosed' in row['InvestorsName'] or 'undisclosed' in row['InvestorsName']):**

**for i in row['InvestorsName'].split(','):**

**if i.strip()!='':**

**investors.append(i.strip())**

**startup.append(row['StartupName'])**

**for i in range(len(startup)):**

**if 'Ola' in startup[i]:**

**startup[i]='Ola'**

**if 'Flipkart' in startup[i]:**

**startup[i]='Flipkart'**

**if 'Oyo' in startup[i] or 'OYO Rooms' in startup[i]:**

**startup[i]='Oyo'**

**if 'Paytm' in startup[i]:**

**startup[i]='Paytm'**

**np\_startup=np.array(startup)**

**np\_investors=np.array(investors)**

**dic=dict()**

**for i in range(len(np\_startup)):**

**s=set()**

**if np\_investors[i] in dic.keys():**

**dic[np\_investors[i]].add(np\_startup[i])**

**else:**

**s.add(np\_startup[i])**

**dic[np\_investors[i]]=s**

**d=dict()**

**for i in dic.keys():**

**if i in d.keys():**

**d[i]+=len(dic[i])**

**else:**

**d[i]=len(dic[i])**

**x=[]**

**y=[]**

**for i in d.keys():**

**x.append(i)**

**y.append(d[i])**

**np\_x=np.array(x)**

**np\_y=np.array(y)**

**np\_x=np\_x[np.argsort(np\_y)]**

**np\_y=np.sort(np\_y)**

**np\_x=np\_x[len(np\_x)-1:len(np\_x)-1-5:-1]**

**np\_y=np\_y[len(np\_y)-1:len(np\_y)-1-5:-1]**

**plt.bar(np\_x, np\_y, color='green')**

**plt.xlabel('Investors', size=14)**

**plt.ylabel('No. of different companies', size=14)**

**plt.title(' Top 5 investors who have invested maximum number of times in different companies', size=12)**

**plt.xticks(rotation=45)**

**plt.show()**

**for i in range(len(np\_x)):**

**print(np\_x[i], np\_y[i])**

ANSWERS- **Sequoia Capital 48**

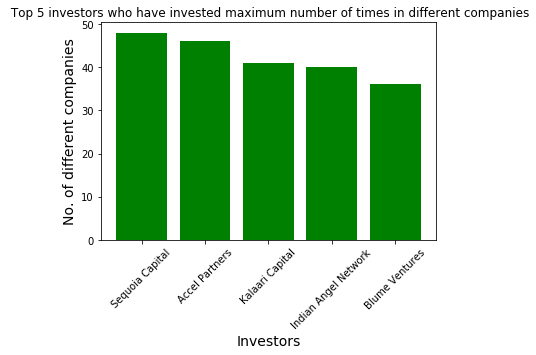
**Accel Partners 46**

**Kalaari Capital 41**

**Indian Angel Network 40**

**Blume Ventures 36**

Graph-



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 numpy as np**

**import csv**

**import matplotlib.pyplot as plt**

**with open('C:\\Users\\PKPAREEK\\Desktop\\New folder (5)\\startup\_funding.csv', encoding='utf8') as file\_obj:**

**file\_data=csv.DictReader(file\_obj, skipinitialspace=True)**

**investors=[]**

**startup=[]**

**for row in file\_data:**

**if row['InvestmentType']=='Seed Funding':**

**if not ('Undisclosed' in row['InvestorsName'] or 'undisclosed' in row['InvestorsName']):**

**for i in row['InvestorsName'].split(','):**

**if i.strip()!='':**

**investors.append(i.strip())**

**startup.append(row['StartupName'])**

**for i in range(len(startup)):**

**if 'Ola' in startup[i]:**

**startup[i]='Ola'**

**if 'Flipkart' in startup[i]:**

**startup[i]='Flipkart'**

**if 'Oyo' in startup[i] or 'OYO Rooms' in startup[i]:**

**startup[i]='Oyo'**

**if 'Paytm' in startup[i]:**

**startup[i]='Paytm'**

**np\_startup=np.array(startup)**

**np\_investors=np.array(investors)**

**dic=dict()**

**for i in range(len(np\_startup)):**

**s=set()**

**if np\_investors[i] in dic.keys():**

**dic[np\_investors[i]].add(np\_startup[i])**

**else:**

**s.add(np\_startup[i])**

**dic[np\_investors[i]]=s**

**d=dict()**

**for i in dic.keys():**

**if i in d.keys():**

**d[i]+=len(dic[i])**

**else:**

**d[i]=len(dic[i])**

**x=[]**

**y=[]**

**for i in d.keys():**

**x.append(i)**

**y.append(d[i])**

**np\_x=np.array(x)**

**np\_y=np.array(y)**

**np\_x=np\_x[np.argsort(np\_y)]**

**np\_y=np.sort(np\_y)**

**np\_x=np\_x[len(np\_x)-1:len(np\_x)-1-5:-1]**

**np\_y=np\_y[len(np\_y)-1:len(np\_y)-1-5:-1]**

**plt.bar(np\_x, np\_y, color='green')**

**plt.xlabel('Investors', size=14)**

**plt.ylabel('No. of different companies', size=14)**

**plt.title(' Top 5 investors who have invested maximum number of times in different companies', size=12)**

**plt.xticks(rotation=45)**

**plt.show()**

**for i in range(len(np\_x)):**

**print(np\_x[i], np\_y[i])**

**ANS.**

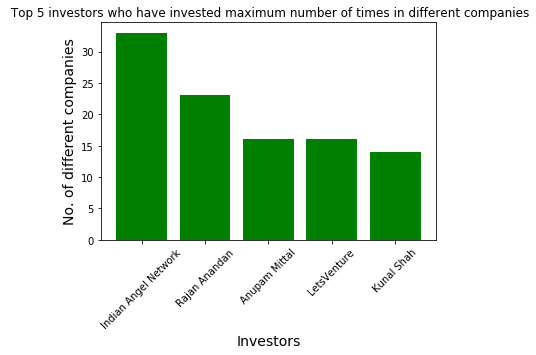
***Indian Angel Network 33***

***Rajan Anandan 23***

***Anupam Mittal 16***

***LetsVenture 16***

***Kunal Shah 14***

Graph-

1. 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.

**import numpy as np**

**import csv**

**import matplotlib.pyplot as plt**

**with open('C:\\Users\\PKPAREEK\\Desktop\\New folder (5)\\startup\_funding.csv', encoding='utf8') as file\_obj:**

**file\_data=csv.DictReader(file\_obj, skipinitialspace=True)**

**investors=[]**

**startup=[]**

**for row in file\_data:**

**if row['InvestmentType']=='Private Equity':**

**if not ('Undisclosed' in row['InvestorsName'] or 'undisclosed' in row['InvestorsName']):**

**for i in row['InvestorsName'].split(','):**

**if i.strip()!='':**

**investors.append(i.strip())**

**startup.append(row['StartupName'])**

**for i in range(len(startup)):**

**if 'Ola' in startup[i]:**

**startup[i]='Ola'**

**if 'Flipkart' in startup[i]:**

**startup[i]='Flipkart'**

**if 'Oyo' in startup[i] or 'OYO Rooms' in startup[i]:**

**startup[i]='Oyo'**

**if 'Paytm' in startup[i]:**

**startup[i]='Paytm'**

**np\_startup=np.array(startup)**

**np\_investors=np.array(investors)**

**dic=dict()**

**for i in range(len(np\_startup)):**

**s=set()**

**if np\_investors[i] in dic.keys():**

**dic[np\_investors[i]].add(np\_startup[i])**

**else:**

**s.add(np\_startup[i])**

**dic[np\_investors[i]]=s**

**d=dict()**

**for i in dic.keys():**

**if i in d.keys():**

**d[i]+=len(dic[i])**

**else:**

**d[i]=len(dic[i])**

**x=[]**

**y=[]**

**for i in d.keys():**

**x.append(i)**

**y.append(d[i])**

**np\_x=np.array(x)**

**np\_y=np.array(y)**

**np\_x=np\_x[np.argsort(np\_y)]**

**np\_y=np.sort(np\_y)**

**np\_x=np\_x[len(np\_x)-1:len(np\_x)-1-5:-1]**

**np\_y=np\_y[len(np\_y)-1:len(np\_y)-1-5:-1]**

**plt.bar(np\_x, np\_y, color='green')**

**plt.xlabel('Investors', size=14)**

**plt.ylabel('No. of different companies', size=14)**

**plt.title(' Top 5 investors who have invested maximum number of times in different companies', size=12)**

**plt.xticks(rotation=45)**

**plt.show()**

**for i in range(len(np\_x)):**

**print(np\_x[i], np\_y[i])**

**ANS. Sequoia Capital 45**

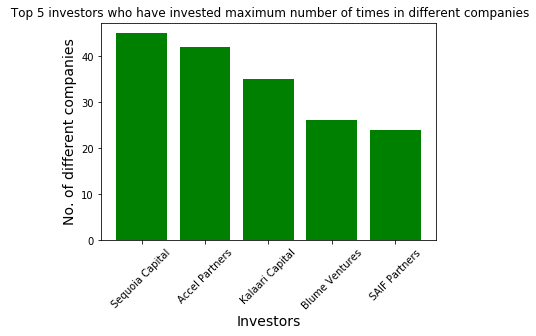
**Accel Partners 42**

**+**

**Blume Ventures 26**

**SAIF Partners 24**

**Graph-**

****