

Case Study: This dataset has funding information of the Indian startups from January 2015 to August 2017.

Q.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.

In [5]:

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

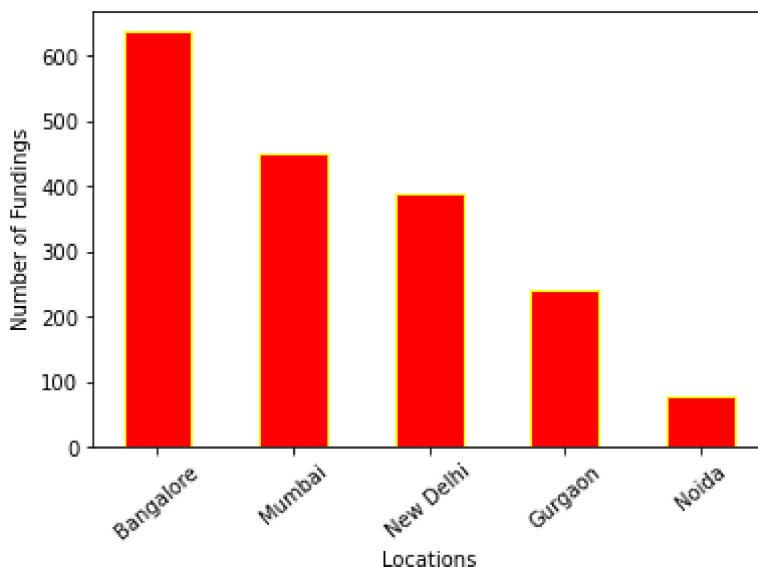
df.dropna(subset = ["CityLocation"],inplace = True)
df["CityLocation"].replace("bangalore","Bangalore",inplace = True)
df["CityLocation"].replace("Delhi","New Delhi",inplace = True)

d1 = {}

for i in df["CityLocation"]:
    if "Bangalore" in i:
        d1["Bangalore"] = d1.get("Bangalore",0) + 1
    elif "Mumbai" in i:
        d1["Mumbai"] = d1.get("Mumbai",0) + 1
    elif "New Delhi" in i:
        d1["New Delhi"] = d1.get("New Delhi",0) + 1
    elif "Noida" in i:
        d1["Noida"] = d1.get("Noida",0) + 1
    elif "Gurgaon" in i:
        d1["Gurgaon"] = d1.get("Gurgaon",0) + 1
cities = sorted(d1, key=d1.get , reverse=True)
fundings = []

for i in cities:
    fundings.append(d1[i])

plt.bar(cities,fundings,width = 0.5,color = "red" , edgecolor = "yellow")
plt.xlabel("Locations")
plt.ylabel("Number of Fundings")
plt.xticks(rotation = 40)
plt.show()
for i in range(5):
    print(cities[i], fundings[i])
print("The location at which the most number of funding is done is " , cities[0])
```



Bangalore 637
Mumbai 449
New Delhi 389
Gurgaon 241

Noida 79

The location at which the most number of funding is done is Bangalore

Q.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.

In [7]:

```
import pandas as pd
import matplotlib.pyplot as plt
startup = pd.read_csv("startup_funding.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)
```

Sequoia Capital
Accel Partners
Kalaari Capital
SAIF Partners
Indian Angel Network

Q.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.

In [4]:

```

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

file_obj = open("startup_funding.csv")
file_data = csv.DictReader(file_obj, skipinitialspace = True)

d = {}

for row in file_data:
    if row["StartupName"] == "Ola Cabs" or row["StartupName"] == "OlaCabs":
        row["StartupName"] = "Ola"
    if row["StartupName"] == "Flipkart.com":
        row["StartupName"] = "Flipkart"
    if row["StartupName"] == "Oyo Rooms" or row["StartupName"] == "OyoRooms" or row["StartupName"] == "Oyo":
        row["StartupName"] = "Oyo"
    if row["StartupName"] == "Paytm Marketplace":
        row["StartupName"] = "Paytm"

    value = row["InvestorsName"].split(",")
    for i in range(len(value)):
        value[i] = value[i].strip()

    for i in value:
        if (row["StartupName"] != None) and i != "" and i != "Undisclosed Investors" and i != "":
            if i in d:
                d[i].add(row["StartupName"])
            else:
                d[i] = set()
                d[i].add(row["StartupName"])

for i in d.keys():
    d[i] = len(d[i])
d1 = sorted(d, key = d.get, reverse = True)

investors = d1[0:5]
num_of_companies = []

top_10_investors = d1[0:10]
number = []

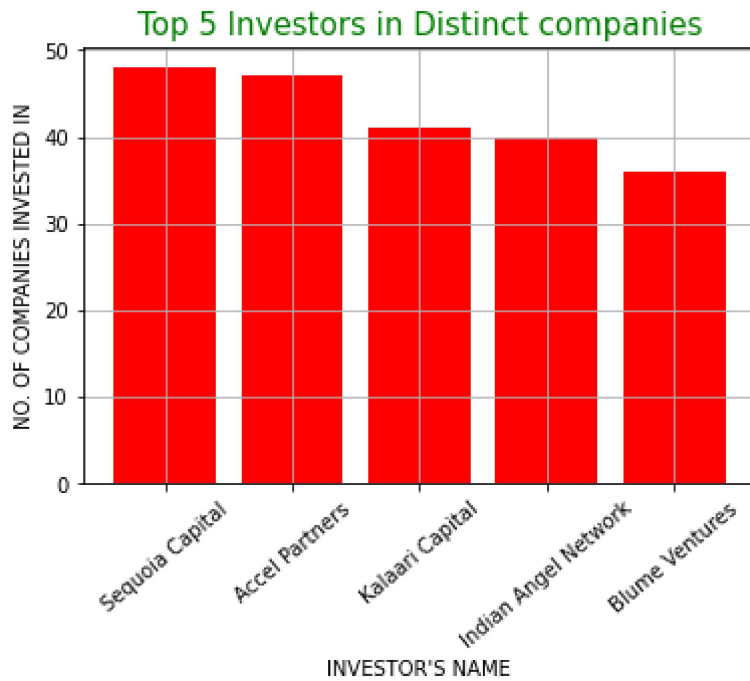
for i in top_10_investors:
    number.append(d[i])

for i in investors:
    num_of_companies.append(d[i])

plt.bar(investors, num_of_companies, color = 'red')
plt.title("Top 5 Investors in Distinct companies", fontsize = 15, color = 'green')
plt.xlabel("INVESTOR'S NAME", fontsize = 10)
plt.ylabel("NO. OF COMPANIES INVESTED IN", fontsize = 10)
plt.xticks(rotation = 40)
plt.grid()
plt.show()

for i in investors:
    print(i, d[i])

```



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

Q.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.

In [7]:

```

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

file_obj = open("startup_funding.csv")
file_data = csv.DictReader(file_obj, skipinitialspace = True)

d = {}

for row in file_data:
    if row["StartupName"] == "Ola Cabs" or row["StartupName"] == "OlaCabs":
        row["StartupName"] = "Ola"
    if row["StartupName"] == "Flipkart.com":
        row["StartupName"] = "Flipkart"
    if row["StartupName"] == "Oyo Rooms" or row["StartupName"] == "OyoRooms" or row["StartupName"] == "Oyo":
        row["StartupName"] = "Oyo"
    if row["StartupName"] == "Paytm Marketplace":
        row["StartupName"] = "Paytm"

    if row["InvestmentType"] == "SeedFunding":
        row["InvestmentType"] = "Seed Funding"
    if row["InvestmentType"] == "Crowd funding":
        row["InvestmentType"] = "Crowd Funding"

    value = row["InvestorsName"].split(",")
    for i in range(len(value)):
        value[i] = value[i].strip()
    for i in value:
        if (row["StartupName"] != None) and i != "" and i != "Undisclosed Investors" and i != "Crowd Funding":
            if i in d:
                d[i].add(row["StartupName"])
            else:
                d[i] = set()
                d[i].add(row["StartupName"])

for i in d.keys():
    d[i] = len(d[i])

d1 = sorted(d, key = d.get, reverse = True)

investors = d1[0:5]
num_of_companies = []

top_10_investors = d1[0:10]
number = []

for i in top_10_investors:
    number.append(d[i])

for i in investors:
    num_of_companies.append(d[i])

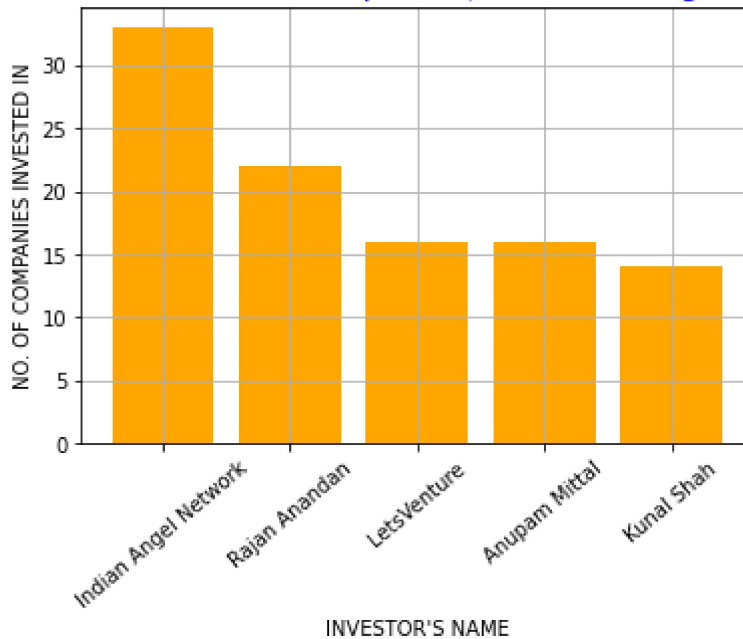
plt.bar(investors, num_of_companies, color = 'orange')
plt.title("Top 5 Investors in Distinct companies(Crowd Funding, Seed Funding)", fontsize = 12)
plt.xlabel("INVESTOR'S NAME", fontsize = 10)
plt.ylabel("NO. OF COMPANIES INVESTED IN", fontsize = 10)
plt.xticks(rotation = 40)

```

```
plt.grid()
plt.show()

for i in investors:
    print(i,d[i])
```

Top 5 Investors in Distinct companies(Crowd Funding, Seed Funding)



Indian Angel Network 33
Rajan Anandan 22
LetsVenture 16
Anupam Mittal 16
Kunal Shah 14

Q.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.

In [9]:

```

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

file_obj = open("startup_funding.csv")
file_data = csv.DictReader(file_obj, skipinitialspace = True)

d = {}

for row in file_data:
    if row["StartupName"] == "Ola Cabs" or row["StartupName"] == "OlaCabs":
        row["StartupName"] = "Ola"
    if row["StartupName"] == "Flipkart.com":
        row["StartupName"] = "Flipkart"
    if row["StartupName"] == "Oyo Rooms" or row["StartupName"] == "OyoRooms" or row["StartupName"] == "Oyo":
        row["StartupName"] = "Oyo"
    if row["StartupName"] == "Paytm Marketplace":
        row["StartupName"] = "Paytm"

    if row["InvestmentType"] == "PrivateEquity":
        row["InvestmentType"] = "Private Equity"

    value = row["InvestorsName"].split(",")
    for i in range(len(value)):
        value[i] = value[i].strip()
    for i in value:
        if (row["StartupName"] != None) and i != "" and i != "Undisclosed Investors" and i != "":
            if i in d:
                d[i].add(row["StartupName"])
            else:
                d[i] = set()
                d[i].add(row["StartupName"])

for i in d.keys():
    d[i] = len(d[i])

d1 = sorted(d, key = d.get, reverse = True)

investors = d1[0:5]
num_of_companies = []

top_10_investors = d1[0:10]
number = []

for i in top_10_investors:
    number.append(d[i])

for i in investors:
    num_of_companies.append(d[i])

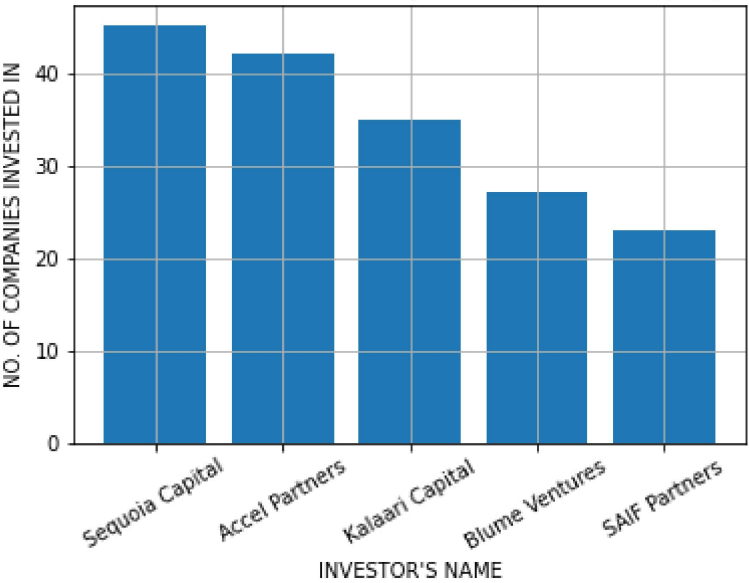
plt.bar(investors, num_of_companies)
plt.title("Top 5 Investors in Distinct companies(Private Equity Funding)", fontsize = 15)
plt.xlabel("INVESTOR'S NAME", fontsize = 10)
plt.ylabel("NO. OF COMPANIES INVESTED IN", fontsize = 10)
plt.xticks(rotation = 30)
plt.grid()
plt.show()

```



```
for i in investors:  
    print(i,d[i])
```

Top 5 Investors in Distinct companies(Private Equity Funding)



Sequoia Capital 45
Accel Partners 42
Kalaari Capital 35
Blume Ventures 27
SAIF Partners 23