

# Case Study : Startup Funding

## Scenario :

Funding data of various Indian startups is given in csv format and we need to get some insights from it.

## Imports :

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

Numpy, pandas and matplotlib are imported. Numpy will be used for numerical analysis (if needed), pandas for storing data in a dataframe and then making some sense from it. Matplotlib will be used for plots.

## Import data in a pandas dataframe :

```
startup_funding=pd.read_csv('datasets/startup_funding.csv')
```

## Copying data to df dataframe :

```
df=startup_funding.copy()
```

**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.

Interpretation :

Here we have to find the location among Bangalore, Mumbai, and NCR in which the maximum number of investments are done historically.

Let's see our df :

```
In [3]: df.head()
```

```
Out[3]:
```

	SNo	Date	StartupName	IndustryVertical	SubVertical	CityLocation	InvestorsName	InvestmentType	AmountInUSD	Remarks
0	0	01/08/2017	TouchKin	Technology	Predictive Care Platform	Bangalore	Kae Capital	Private Equity	1,300,000	NaN
1	1	02/08/2017	Ethinos	Technology	Digital Marketing Agency	Mumbai	Triton Investment Advisors	Private Equity	NaN	NaN
2	2	02/08/2017	Leverage Edu	Consumer Internet	Online platform for Higher Education Services	New Delhi	Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...	Seed Funding	NaN	NaN
3	3	02/08/2017	Zepo	Consumer Internet	DIY Ecommerce platform	Mumbai	Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...	Seed Funding	500,000	NaN
4	4	02/08/2017	Click2Clinic	Consumer Internet	healthcare service aggregator	Hyderabad	Narottam Thudi, Shireesh Palle	Seed Funding	850,000	NaN

Basically we just need to know how many times the above mentioned three locations appeared in the CityLocation column.

But there are some differences in name of the cities so let's clean that data first :

```
df = df[df['CityLocation'].notna()] #dropping missing values
```

A function to get 1st city in case multiple are given

```
def separateCity(city):  
    return city.split('/')[0].strip()  
df['CityLocation']=df['CityLocation'].apply(separateCity)
```

Replacing bangalore with Bangalore

```
df['CityLocation'].replace("bangalore","Bangalore",inplace=True)
```

Replacing Delhi with New Delhi

```
df['CityLocation'].replace("Delhi","New Delhi",inplace=True)
```

Now let us get number of times different cities received funding

```
bangalore=(df['CityLocation']=='Bangalore').value_counts()[True]  
new_delhi=(df['CityLocation']=='New Delhi').value_counts()[True]  
gurgaon=(df['CityLocation']=='Gurgaon').value_counts()[True]  
noida=(df['CityLocation']=='Noida').value_counts()[True]
```

**NCR is union of three cities**

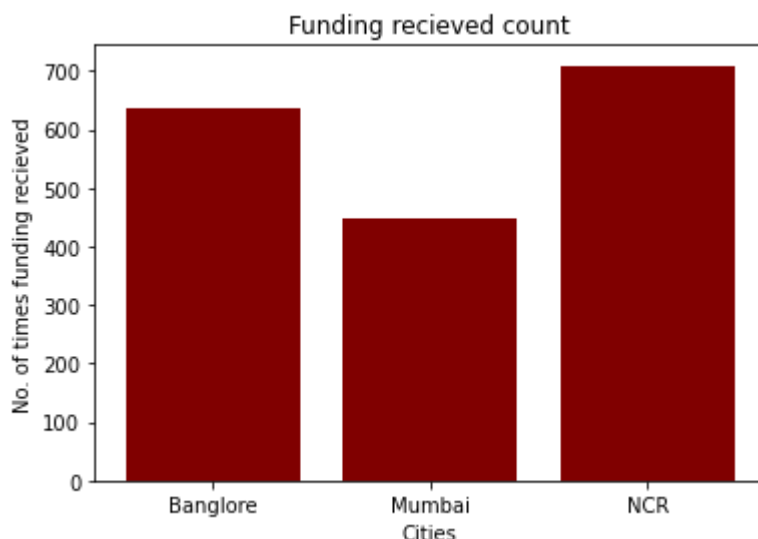
```
ncr=new_delhi+gurgaon+noida  
mumbai=(df['CityLocation']=='Mumbai').value_counts()[True]
```

**So funding count is as follows :**

**Banglore : 635, mumbai : 449, ncr : 709**

**So NCR has received funding maximum number of times that 709 times**

**Bar graph between cities and number of time they have received funding**



**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.**

**Interpretation :**

Here we have to find the investors who have invested the maximum number of times.

So we need to find which investors name have appeared maximum number of times in InvestorName column

Let's copy data once again to df and clean it:

```
df.dropna(subset=["InvestorsName"],inplace=True)
invester = df["InvestorsName"].str.split(',')
newarray = [j.strip() for sub in investor for j in sub]
```

**nameseries** contains count of number of times each particular investor invested

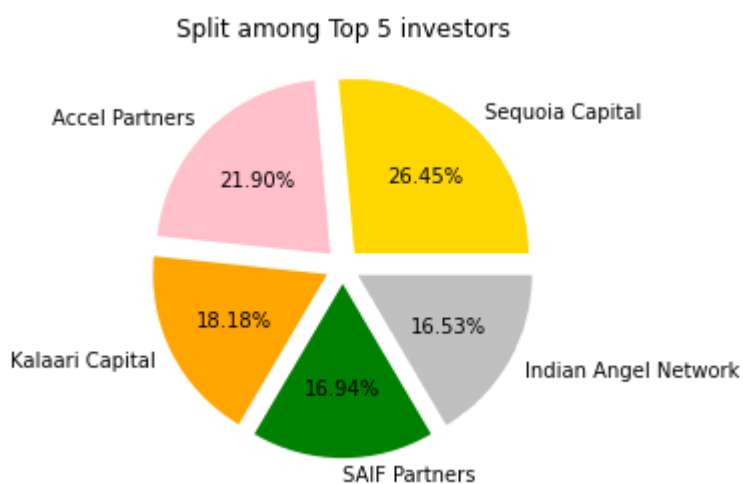
```
nameseries = pd.Series(newarray).value_counts()
```

**top5Investors** contains top 5 investors (in terms of number of times invested)

```
nameseries = pd.Series(newarray).value_counts()
```

**So, Top 5 investors are :**

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



*Question3 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.*

Interpretation :

Here we have to find the investors who have invested a maximum number of times in different startups.

So we need to find which investors name have appeared maximum number of times in InvestorName column but have to consider different investments in same company by same investor only once

**Cleaning StartupName column data for important startups : Ola, Flipkart, Oyo and Paytm :**

```
df = df[df['StartupName'].notna()]
df['StartupName'].replace("Flipkart.com", "Flipkart", inplace=True)
df['StartupName'].replace("Oyo Rooms", "Oyo", inplace=True)
df['StartupName'].replace("Ola cabs", "Ola", inplace=True)
df['StartupName'].replace("Ola Cabs", "Ola", inplace=True)
df['StartupName'].replace("Paytm
Marketplace", "Paytm", inplace=True)
df['StartupName'].replace("Oyorooms", "Oyo", inplace=True)
df['StartupName'].replace("OyoRooms", "Oyo", inplace=True)
df['StartupName'].replace("OYO Rooms", "Oyo", inplace=True)
```

**Cleaning InvestorName column's data**

```
df.dropna(subset=["Indf.dropna(subset=["InvestorsName"], inplace=True), in
place=True)
```

**pairs is a set which contains tuples (investor name, startup in which he invested) so that we don't count an investor investing in the same company at different points of time more than once.**

**top\_investors is an array and will contain investor names**

```
top_investors=[]
pairs=set()
for i in range(len(df)):
```

```

investor = df.iloc[i]["InvestorsName"].split(',')
for j in range(len(investor)):
    investor[j]=investor[j].strip()
    if investor[j]=='':
        continue
    pair=(investor[j],df.iloc[i]["StartupName"])
    if pair in pairs:
        continue
    else:
        pairs.add(pair)
        top_investors.append(investor[j])

nameseries = pd.Series(top_investors).value_counts()

```

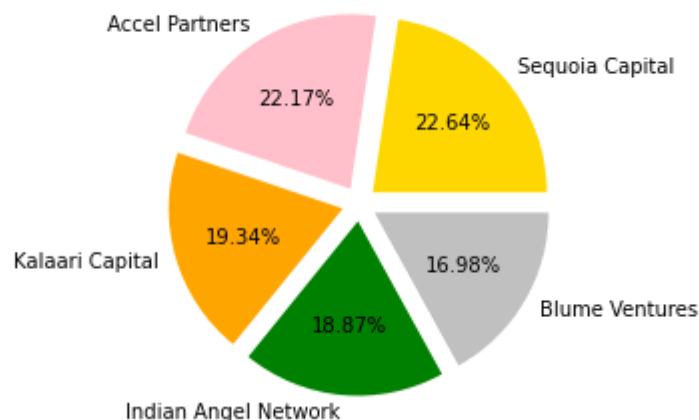
### Extracting top 5 investors from nameseries in top5Investors

```
top5Investors=nameseries[:5]
```

So, Top 5 investors are :

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

Split among Top 5 investors with each startup they invested in is counted as single investment



*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.*

## Correcting investment types (the ones we need for current question)

```
df['InvestmentType'].replace("SeedFunding","Seed Funding",inplace=True)
df['InvestmentType'].replace("Crowd funding","Crowd
Funding",inplace=True)
df = df[df['InvestmentType'].notna()]
df.dropna(subset=["InvestorsName"],inplace=True)
```

**topInvestors** is a dictionary containing count of number of times an investor invested with InvestmentType as Seed Funding or Crowdfunding

```
topInvestors={}
for i in range(len(df)):
    if df.iloc[i]['InvestmentType']=='Seed Funding' or
df.iloc[i]['InvestmentType']=='Crowd Funding':
        investor = df.iloc[i]["InvestorsName"].split(',')
        for j in range(len(investor)):
            investor[j]=investor[j].strip()
            if investor[j]==' ' or investor[j]=='Undisclosed investors'
or investor[j]=='Undisclosed Investors':
                continue
            if investor[j] in topInvestors:
                topInvestors[investor[j]]+=1
            else:
                topInvestors[investor[j]]=1
```

**importing itemgetter to sort dictionary topInvestors and get top 5 investors in res**

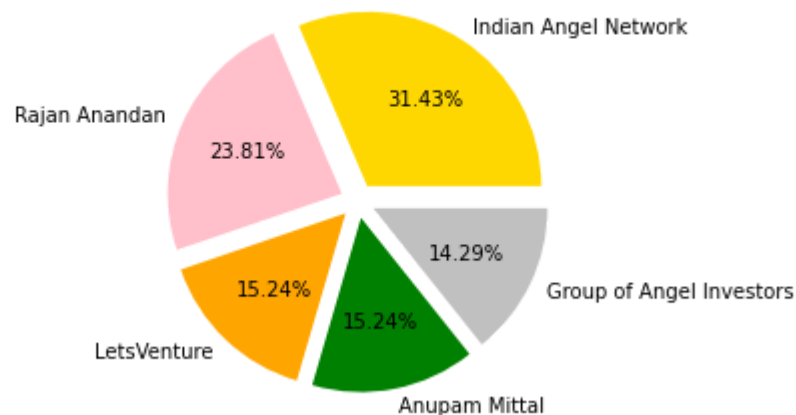
```
from operator import itemgetter
res = dict(sorted(topInvestors.items(), key = itemgetter(1), reverse = True)[:5])
for i in res:
    print(i,res[i])
```

**So top 5 investors with their investment type as Seed Funding or Crowd**

**Funding are :**

**Indian Angel Network, Rajan Anandan, LetsVenture, Anupam Mittal, Group of Angel Investors.**

Distribution among top 5 investors with their investment type as Seed Funding or Crowd Funding



**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.

Copying data to df:

```
df=startup_funding.copy()
```

**Correcting investment types (the ones we need for current question)**

```
df['InvestmentType'].replace("PrivateEquity","Private  
Equity",inplace=True)  
df = df[df['InvestmentType'].notna()]  
df.dropna(subset=["InvestorsName"],inplace=True)
```

**topInvestors** is a dictionary containing count of number of times an investor invested with InvestmentType as Private Equity

```
topInvestors={}  
for i in range(len(df)):
```



```

if df.iloc[i]['InvestmentType']=='Private Equity':
    investor = df.iloc[i]["InvestorsName"].split(',')
    for j in range(len(investor)):
        investor[j]=investor[j].strip()
        if investor[j]==' ' or investor[j]=='Undisclosed investors'
or investor[j]=='Undisclosed Investors':
            continue
        if investor[j] in topInvestors:
            topInvestors[investor[j]]+=1
        else:
            topInvestors[investor[j]]=1

res = dict(sorted(topInvestors.items(), key = itemgetter(1), reverse =
True)[:5])
for i in res:
    print(i,res[i])

```

**So top 5 investors with their investment type as Private Equity are : Sequoia Capital, Accel Partners, Kalaari Capital, SAIF Partners, Blume Ventures.**

Distribution among top 5 investors with their investment type as Private Equity

