PROJECT 1

# Report

By Mrinal Kalita, NIT Silchar, June 2022

# PROJECT:EXPLANATION AND PLOTS

## Case Study: Questions

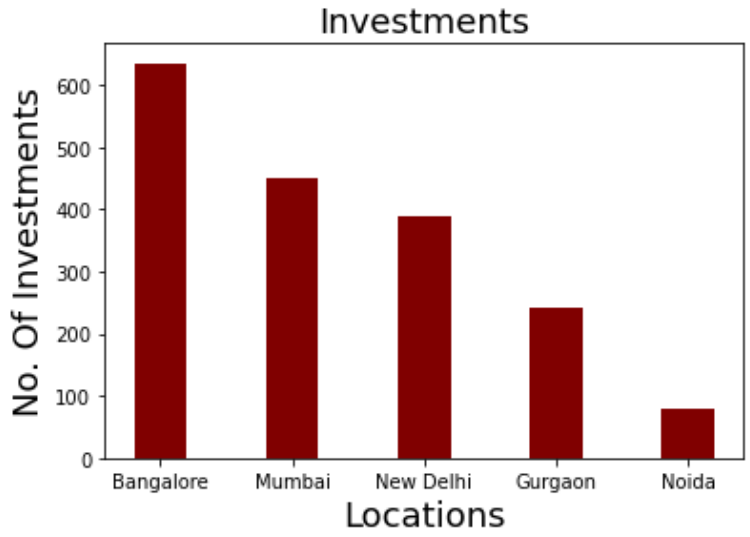
1. Your Friend has developed the Product and he wants to establish the product start-up and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, 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 start-ups 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 start-ups multiple locations are given, one Indian and one Foreign. Consider the start-up if any one of the city lies in given locations.

# Code:

```
#Question 1
import csv
import operator
import matplotlib.pyplot as plt
with open('startup_funding.csv', encoding ='UTF-8') as ob:
    data = csv.DictReader(ob, skipinitialspace=True)
    d={}
    for row in data:
        city=row["CityLocation"]
        if city in d and city!='':
            d[city]+=1
        else:
            d[city]=1
    d={k: v for k, v in sorted(d.items(), key=lambda item: item[1])}
    d = dict(sorted(d.items(), key=operator.itemgetter(1),reverse=True))
    d["Bangalore"]=d["Bangalore"]+d["bangalore"]
    d['New Delhi']=d['New Delhi']+d["Delhi"]



    f=['Bangalore', 'Mumbai', 'New Delhi', 'Gurgaon','Noida']
    #print(f)
    s={}
    for i in f:
        s[i]=d[i]

    for i in s:
        del[d[i]]
    for i in s:
        for j in d:
            if "/" in j:
                if i[1:] in j[1:-1]:
                    s[i]=s[i]+d[j]

    #print(s)
    for k,v in s.items():
        print(k,v)
    Location=list(s.keys())
    No_of_Fundings=list(s.values())
    plt.bar(Location, No_of_Fundings, color ='maroon',width = 0.4)

    plt.xlabel("Locations",fontsize=18)
    plt.ylabel("No. Of Investments",fontsize=18)
    plt.title("Investments",fontsize=18)
    plt.show()
```

# Graph:

```
Bangalore 635
Mumbai 449
New Delhi 389
Gurgaon 241
Noida 79
```




# Explanation:

To solve this particular question I proceeded with the following steps:

- Imported the necessary libraries for data analysis and visualisation:
- Used csv.DictReader to read the csv file in a dictionary format.
- Created a new dictionary ‘d’ and it had  city-location as keys and number of investments as values
- Corrected the irregularities in spellings of “Bangalore” and “New Delhi”.
- Created a new dictionary “s” with required values of city-location and number of investments and ploted the bar graph as per requirements.

# Data Insights:

By analysing the give data I came to a conclusion that the best city to get an investment is Bangalore followed by Mumbai, New Delhi, Gurgaon and Noida (since the client is restricted to the particular regions). Therefore my friend should consider this current city order for seeking investments.

2. Even after trying for so many times, your friend's start-up 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 start-up. Your list will increase the chance of your friend start-up 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 start-up, multiple investors might have invested. So consider each investor for that start-up. Ignore undisclosed investors.

# Code:

```
#Question 2
import csv
import pandas as pd
import numpy as np
import operator
import matplotlib.pyplot as plt
data = pd.read_csv('startup_funding.csv',encoding = 'utf-8')
data=data.dropna(subset=['InvestorsName'])
data = data[data.InvestorsName != '']
data = data[data.InvestorsName != 'Undisclosed Investors']
data = data[data.InvestorsName != 'undisclosed investor']
data = data[data.InvestorsName != 'Undisclosed investors']
data = data[data.InvestorsName != 'undisclosed investors']
l=[]
for i in data['InvestorsName']:
    if "," in i and i!="":
        x=i.split(',')
        for j in x:
            l.append(j.strip())
#print(l.count("Kalaari Capital"))
s=set(l)

d={}
for i in data["InvestorsName"]:
    if i in d:
        d[i]=d[i]+1
    else:
        d[i]=1

for i in s:
    if i in d:

        d[i]=d[i]+l.count(i)
d={k: v for k, v in sorted(d.items(), key=lambda item: item[1])}
d = dict(sorted(d.items(), key=operator.itemgetter(1),reverse=True))
f=list(d)
f=f[:5]
p={}
for i in f:
    p[i]=d[i]
    print(i,p[i])
```

# Output:

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

# Explanation:

To solve this particular question I proceeded with the following steps:

- Imported the necessary libraries for data analysis and visualisation:
- Used pd.read_csv to read the csv file in a pandas dataframe format.
- Check the irregularities in spellings of “Undisclosed Invsetors” and removed them from the dataframe as per requirement.
- Created a new list ‘l’ and filled it up with the data of investor’s names and checked the strings with comma separates values and filled them accordingly to the main list l.
- Created a set “s” to eliminate the repeated values of list l:
- Created a new dictionary “d” with required values of Investor name and number of investments, sorted them and printed the top 5 investors by using a list “f” and dictionary “p”.

# Data Insights:

By analysing the give data I came to a conclusion that the top investor to seek an investment is “Sequoia Capital” followed by “Accel Partners”, “Kalaari Capital” , “SAIF Partners”, “Indian Angel Network”. Therefore my friend should consider this current order of investors for seeking investments.

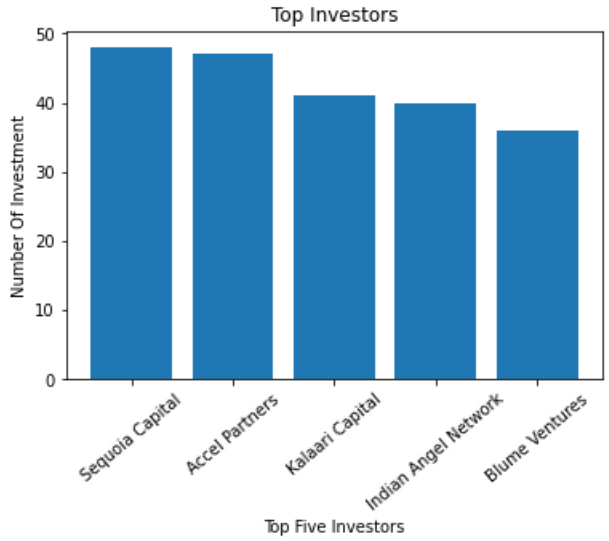
3. After re-analysing the dataset, you found out that some investors have invested in the same start-up at different number of funding rounds. So before finalising the previous list, you want to improvise it by finding the top 5 investors who have invested in different number of start-ups. This list will be more helpful than your previous list in finding the investment for your friend start-up. 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 start-up, count one for that company. There are many errors in start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

# Code:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv('startup_funding.csv',encoding = 'utf-8')
data=data.dropna(subset=['InvestorsName'])
data = data[data.InvestorsName != '']
data = data[data.InvestorsName != 'Undisclosed Investors']
data = data[data.InvestorsName != 'undisclosed investor']
data = data[data.InvestorsName != 'Undisclosed investors']
data = data[data.InvestorsName != 'undisclosed investors']
data['StartupName'].replace('Flipkart.com','Flipkart',inplace = True)
data['StartupName'].replace('Paytm Marketplace','Paytm',inplace = True)
for i in data['StartupName']:
    if "Ola" in i:
        data['StartupName'].replace(i,'Ola',inplace = True)
data['StartupName'].replace('Oyorooms','Oyo',inplace = True)
data['StartupName'].replace('OyoRooms','Oyo',inplace = True)
data['StartupName'].replace('Oyo Rooms','Oyo',inplace = True)
data['StartupName'].replace('OYO Rooms','Oyo',inplace = True)
iv=[]
sn=[]
for k,v in data.iterrows():
    x= v['InvestorsName']
    y = v['StartupName']

    i = str(x)
    l = i.split(',')

    for j in l:
        if j != '':
            j = j.strip()
            iv.append(j)
            sn.append(y)

d = pd.DataFrame({'InvestorsName' : iv,'StartupName' : sn})
d = d.groupby('InvestorsName')['StartupName'].nunique()
d = d.sort_values(ascending = False)

n = []
p = []
for i in range(5):
    print(d.index[i],d.values[i])
    n.append(d.index[i])
    p.append(d.values[i])

plt.bar(n,p)
plt.xticks(rotation=40)
plt.xlabel('Top Five Investors')
plt.ylabel('Number Of Investment')
plt.title('Top Investors')
plt.show()
```

# Graph:

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




# Explanation:

To solve this particular question I proceeded with the following steps:

- Imported the necessary libraries for data analysis and visualisation:
- Used pd.read_csv to read the csv file in a pandas dataframe format.
- Checked the irregularities in spellings of “Undisclosed Invsetors” and removed them from the dataframe as per requirement.
- Corrected the irregularities in spellings of “Ola”, “Flipkart’’, “Oyo” and “Paytm”.
- Created two new lists “nv” and “sn” to store the data of investor name and starup name respectively.
- Created a new dataframe “d” with required values of Investor name and startup name, filtered out the unique values using the fuction nunique() and then sorted the values in desending order.
- Appended the required values to two lists “n” and “p” and ploted the bar graph as per requirements using matplotlib functions.

# Data Insights:

By analysing the give data I came to a conclusion that the top investor to seek an investment is “Sequoia Capital” followed by “Accel Partners”, “Kalaari Capital” , “Indian Angel Network” and “Blume Ventures”. Therefore my friend should consider this current order of investors for seeking investments.

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 start-up is at an early-stage start-up, the best-suited investment type would be - Seed Funding and Crowdfunding. Find the top 5 investors who have invested in a different number of start-ups and their investment type is Crowdfunding or Seed Funding. Correct spelling of investment types is - "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 start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

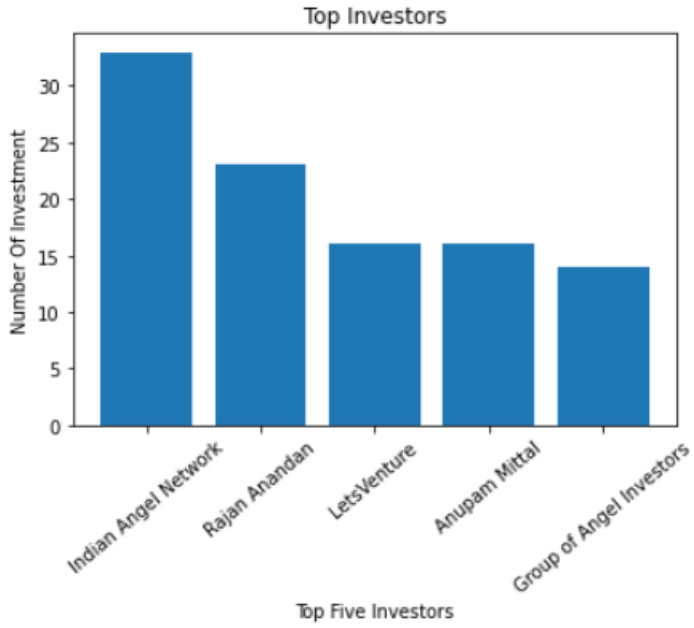
# Code:

```
#Question 4
import csv
import pandas as pd
import numpy as np
import operator
import matplotlib.pyplot as plt
data = pd.read_csv('startup_funding.csv',encoding = 'utf-8')
data=data.dropna(subset=['InvestorsName'])

data['InvestmentType'].replace('PrivateEquity','Private Equity',inplace=True)
data['InvestmentType'].replace('Crowd funding','Crowd Funding',inplace=True)
data['InvestmentType'].replace('SeedFunding','Seed Funding',inplace=True)

data = data[data.InvestorsName != '']
data = data[data.InvestorsName != 'Undisclosed Investors']
data = data[data.InvestorsName != 'undisclosed investor']
data = data[data.InvestorsName != 'Undisclosed investors']
data = data[data.InvestorsName != 'undisclosed investors']

data['StartupName'].replace('Flipkart.com','Flipkart',inplace = True)
data['StartupName'].replace('Paytm Marketplace','Paytm',inplace = True)
for i in data['StartupName']:
    if "Ola" in i:
        data['StartupName'].replace(i,'Ola',inplace = True)
data['StartupName'].replace('Oyorooms','Oyo',inplace = True)
data['StartupName'].replace('OyoRooms','Oyo',inplace = True)
data['StartupName'].replace('Oyo Rooms','Oyo',inplace = True)
data['StartupName'].replace('OYO Rooms','Oyo',inplace = True)

data = data[(data.InvestmentType =='Seed Funding') | (data.InvestmentType == 'Crowd Funding')]

iv=[]
sn=[]
```

```
for k,v in data.iterrows():
    x= v['InvestorsName']
    y = v['StartupName']

    i = str(x)
    l = i.split(',')

    for j in l:
        if j != '':
            j = j.strip()
            iv.append(j)
            sn.append(y)

d = pd.DataFrame({'InvestorsName' : iv,'StartupName' : sn})
d = d.groupby('InvestorsName')['StartupName'].nunique()
d = d.sort_values(ascending = False)

n = []
p = []
for i in range(5):
    print(d.index[i],d.values[i])
    n.append(d.index[i])
    p.append(d.values[i])

plt.bar(n,p)
plt.xlabel('Top Five Investors')
plt.ylabel('Number Of Investment')
plt.title('Top Investors')
plt.xticks(rotation=40)
plt.show()
```

# Graph:

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




# Explanation:

To solve this particular question I proceeded with the following steps:

- Imported the necessary libraries for data analysis and visualisation:
- Used pd.read_csv to read the csv file in a pandas dataframe format.
- Checked the irregularities in spellings of “Undisclosed Invsetors” and removed them from the dataframe as per requirement.
- Corrected the irregularities in spellings of “Ola”, “Flipkart’’, “Oyo”, “Paytm”, “Private Equity”, “Crowd Funding”, “Seed Funding”.
- Filtered the data frame by only picking “Crowd Funding” or “Seed Funding” as investment types.
- Created two new lists “nv” and “sn” to store the data of investor name and starup name respectively.
- Created a new dataframe “d” with required values of Investor name and startup name, filtered out the unique values using the fuction nunique() and then sorted the values in desending order.
- Appended the required  values to two lists “n” and “p” and ploted the bar graph as per requirements using matplotlib functions.

# Data Insights:

By analysing the give data I came to a conclusion that the top investor to seek an investment in Crowdfunding or Seed Funding is “Indian Angel Network” followed by “Ranjan Anandan”, “LetsVentures”, “Anupam Mittal” and “Group of Angel Investors” . Therefore my friend should consider this current order of investors for seeking investments.

5. Due to your immense help, your friend start-up successfully got seed funding and it is on the operational mode. Now your friend wants to expand his start-up and he is looking for new investors for his start-up. Now you again come as a saviour to help your friend and want to create a list of probable new investors. Before moving forward, you remember your investor friend advice that finding the investors by analysing the investment type. Since your friend start-up 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 start-ups and their investment type is Private Equity. Correct spelling of investment types is - "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 start-up names. Ignore correcting all, just handle the important ones - Ola, Flipkart, Oyo and Paytm.

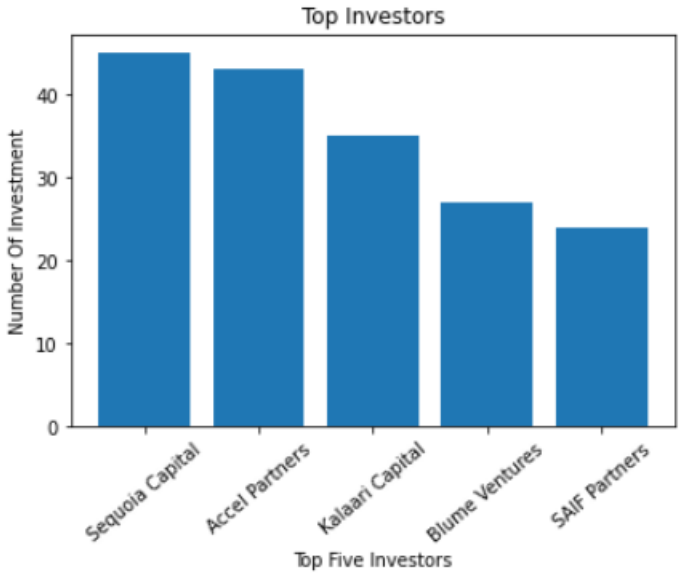
# Code:

```
#Question 5
import csv
import pandas as pd
import numpy as np
import operator
import matplotlib.pyplot as plt
data = pd.read_csv('startup_funding.csv',encoding = 'utf-8')
data=data.dropna(subset=['InvestorsName'])

data['InvestmentType'].replace('PrivateEquity','Private Equity',inplace=True)
data['InvestmentType'].replace('Crowd funding','Crowd Funding',inplace=True)
data['InvestmentType'].replace('SeedFunding','Seed Funding',inplace=True)

data = data[data.InvestorsName != '']
data = data[data.InvestorsName != 'Undisclosed Investors']
data = data[data.InvestorsName != 'undisclosed investor']
data = data[data.InvestorsName != 'Undisclosed investors']
data = data[data.InvestorsName != 'undisclosed investors']

data['StartupName'].replace('Flipkart.com','Flipkart',inplace = True)
data['StartupName'].replace('Paytm Marketplace','Paytm',inplace = True)
for i in data['StartupName']:
    if "Ola" in i:
        data['StartupName'].replace(i,'Ola',inplace = True)
data['StartupName'].replace('Oyorooms','Oyo',inplace = True)
data['StartupName'].replace('OyoRooms','Oyo',inplace = True)
data['StartupName'].replace('Oyo Rooms','Oyo',inplace = True)
data['StartupName'].replace('OYO Rooms','Oyo',inplace = True)

data = data[(data.InvestmentType =='Private Equity')]

iv=[]
sn=[]
```

```
for k,v in data.iterrows():
    x= v['InvestorsName']
    y = v['StartupName']

    i = str(x)
    l = i.split(',')

    for j in l:
        if j != '':
            j = j.strip()
            iv.append(j)
            sn.append(y)

d = pd.DataFrame({'InvestorsName' : iv,'StartupName' : sn})
d = d.groupby('InvestorsName')['StartupName'].nunique()
d = d.sort_values(ascending = False)

n = []
p = []
for i in range(5):
    print(d.index[i],d.values[i])
    n.append(d.index[i])
    p.append(d.values[i])

plt.bar(n,p)
plt.xlabel('Top Five Investors')
plt.ylabel('Number Of Investment')
plt.title('Top Investors')
plt.xticks(rotation=40)
plt.show()
```

# Graph:

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




# Explanation:

To solve this particular question I proceeded with the following steps:

- Imported the necessary libraries for data analysis and visualisation:
- Used pd.read_csv to read the csv file in a pandas dataframe format.
- Checked the irregularities in spellings of “Undisclosed Invsetors” and removed them from the dataframe as per requirement.
- Corrected the irregularities in spellings of “Ola”, “Flipkart’’, “Oyo”, “Paytm”, “Private Equity”, “Crowd Funding”, “Seed Funding”.
- Filtered the data frame by only picking “Private Equity” as investment types.
- Created two new lists “nv” and “sn” to store the data of investor name and starup name respectively.
- Created a new dataframe “d” with required values of Investor name and startup name, filtered out the unique values using the fuction nunique() and then sorted the values in desending order.
- Appended the required values to two lists “n” and “p” and ploted the bar graph as per requirements using matplotlib functions.

# Data Insights:

By analysing the give data I came to a conclusion that the top investor to seek an investment in Private Equity is “Sequoia Capital” followed by “Accel Partners”, “Kalaari Capital”, “Blume Ventures” and “SAIF Partners” . Therefore my friend should consider this current order of investors for seeking investments.