

Case Study 2

June 9, 2020

1 Case Study

1.0.1 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.

1.0.2 Importing the file

```
[77]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
startup = pd.read_csv("startup_funding.csv")
startup.head()
```

```
[77]: SNo      Date      StartupName  IndustryVertical \
0      0  01/08/2017      TouchKin           Technology
1      1  02/08/2017      Ethinos           Technology
2      2  02/08/2017  Leverage Edu  Consumer Internet
3      3  02/08/2017        Zepo  Consumer Internet
4      4  02/08/2017  Click2Clinic  Consumer Internet

                                SubVertical CityLocation \
0                                Predictive Care Platform  Bangalore
1                                Digital Marketing Agency    Mumbai
2  Online platform for Higher Education Services  New Delhi
3                                DIY Ecommerce platform    Mumbai
4                                healthcare service aggregator  Hyderabad
```

	InvestorsName	InvestmentType	\
0	Kae Capital	Private Equity	
1	Triton Investment Advisors	Private Equity	
2	Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...	Seed Funding	
3	Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...	Seed Funding	
4	Narottam Thudi, Shireesh Palle	Seed Funding	

	AmountInUSD	Remarks
0	1,300,000	NaN
1	NaN	NaN
2	NaN	NaN
3	500,000	NaN
4	850,000	NaN

1.0.3 Top 10 Indian cities to get startup funding

```
[78]: startup["CityLocation"].dropna(inplace=True)
def split(cityname):
    return cityname.split('/')[0].strip()

startup["CityLocation"] = startup["CityLocation"].apply(split)
startup['CityLocation'].replace('Delhi', 'New Delhi', inplace = True)
startup['CityLocation'].replace('bangalore', 'Bangalore', inplace = True)
startup["CityLocation"].value_counts().head(10)
```

```
[78]: Bangalore    635
      Mumbai      449
      New Delhi   389
      Gurgaon     241
      Pune        91
      Noida       79
      Hyderabad   77
      Chennai     67
      Ahmedabad   35
      Jaipur      25
      Name: CityLocation, dtype: int64
```

These are the top 10 cities that he can choose for startup with no of funding each cities has.

1.0.4 No of funding done in preferred cities

```
[79]: preferred = ["Bangalore", "Mumbai", "NCR"]
      dic = {}
      for i in preferred:
          if i == "NCR":
```

```

        value = ((startup["CityLocation"] == "New Delhi") |
↳(startup["CityLocation"] == "Gurgaon") | (startup["CityLocation"] ==
↳"Noida")).sum()
        dic[i] = dic.get(i, 0) + value
    else:
        value = (startup["CityLocation"] == i).sum()
        dic[i] = dic.get(i, 0) + value
print(dic)

```

```
{'Bangalore': 635, 'Mumbai': 449, 'NCR': 709}
```

1.0.5 City with maximum no of startup funding

```
[88]: import operator
result = sorted(dic.items(), key = operator.itemgetter(1), reverse = True)[0]
print(result)

```

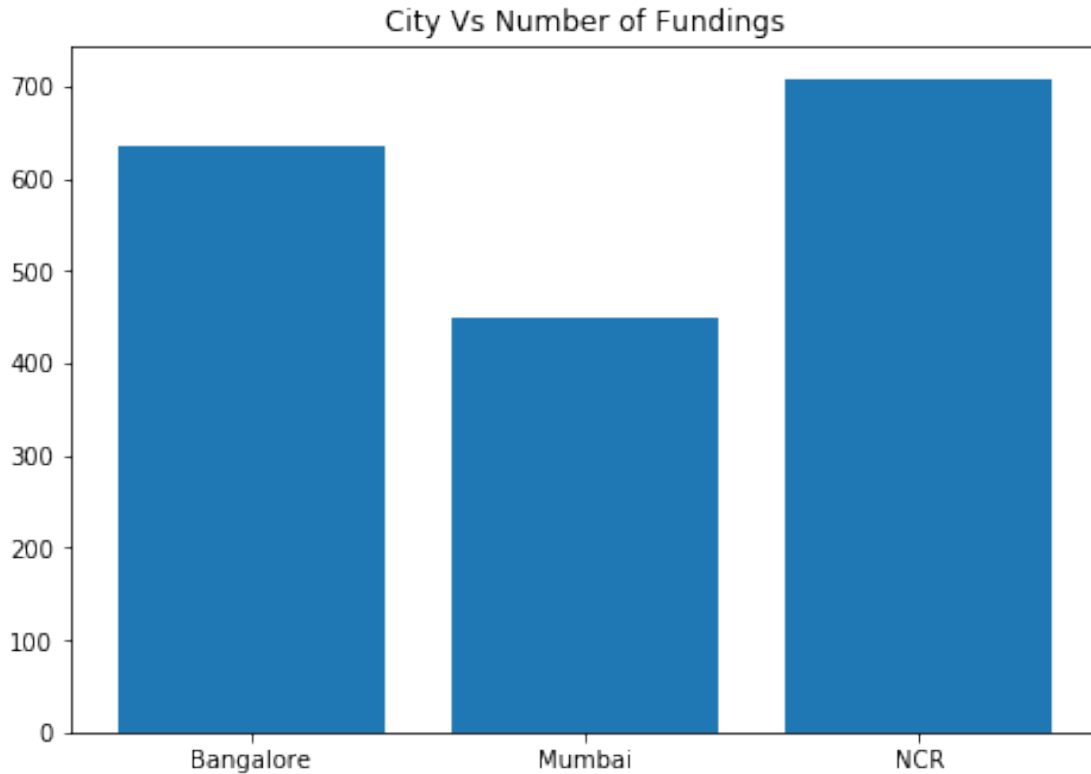
```
('NCR', 709)
```

From this result it can be seen that best location for him to locate his startup will be NCR as it has highest no of fundings.

1.0.6 Plotting the graph

```
[90]: fig = plt.figure()
x = fig.add_axes([1,0,1,1])
x.set_title('City Vs Number of Fundings')
x.bar(dic.keys(), dic.values())
plt.show()

```



1.0.7 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.

1.0.8 Importing the file

```
[118]: startup1 = pd.read_csv("startup_funding.csv")
```

1.0.9 Correction in Startup Names

```
[119]: startup1["StartupName"].replace("Ola Cabs", "Ola", inplace = True)
startup1["StartupName"].replace("OlaCabs", "Ola", inplace = True)
startup1["StartupName"].replace("Paytm Marketplace", "Paytm", inplace = True)
startup1["StartupName"].replace("Flipkart.com", "Flipkart", inplace = True)
startup1["StartupName"].replace("Oyo Rooms", "Oyo", inplace = True)
startup1["StartupName"].replace("Oyorooms", "Oyo", inplace = True)
startup1["StartupName"].replace("OyoRooms", "Oyo", inplace = True)
```

```
startup1["StartupName"].replace("OYO Rooms", "Oyo", inplace = True)
startup1["AmountInUSD"].fillna("0", inplace = True)
```

1.0.10 Converging into numeric values

```
[120]: def num_1(value):
        if value != "":
            return int(float(value.replace(',', '')))
startup1["AmountInUSD"] = startup1["AmountInUSD"].apply(num_1)
startup1["InvestorsName"].dropna(inplace = True)

dic = {}
def name(val):
    arr = val.split(",")
    for i in arr:
        key = i.strip()
        dic[key] = dic.get(key, 0) + 1

startup1["InvestorsName"].apply(name)
name = []
value = []
for key in dic:
    name.append(key)
    value.append(dic[key])
```

1.0.11 Making array and printing result

```
[121]: np_values = np.array(value)
        np_names = np.array(name)

        X = np_values.argsort()[::-1][:5]
        for i in X:
            print(np_names[i], np_values[i])
```

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

So these are top 5 investors that have invested maximum no. of times and by contacting them first will help him to get investment for his startup.

1.0.12 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.

1.0.13 Importing file and correction in startup names

```
[96]: startup2 = pd.read_csv("startup_funding.csv")

startup2["StartupName"].replace("Ola Cabs", "Ola", inplace = True)
startup2["StartupName"].replace("OlaCabs", "Ola", inplace = True)
startup2["StartupName"].replace("Paytm Marketplace", "Paytm", inplace = True)
startup2["StartupName"].replace("Flipkart.com", "Flipkart", inplace = True)
startup2["StartupName"].replace("Oyo Rooms", "Oyo", inplace = True)
startup2["StartupName"].replace("Oyorooms", "Oyo", inplace = True)
startup2["StartupName"].replace("OyoRooms", "Oyo", inplace = True)
startup2["StartupName"].replace("OYO Rooms", "Oyo", inplace = True)
startup2["AmountInUSD"].fillna("0", inplace = True)
startup2.dropna(subset = ["StartupName", "InvestorsName"], inplace = True)
```

1.0.14 Creating dictionary to hold the keys as investor name and values as a set of all the startups invested in.

```
[99]: d = {}
for i in startup2.index:
    investorName = startup2["InvestorsName"][i].strip()
    if "," in investorName:
        arrName = investorName.strip().split(",")
        for j in arrName:
            if j.strip() in d:
                d[j.strip()].add(startup2["StartupName"][i].strip())
            else:
                s = set()
                d[j.strip()] = s
                d[j.strip()].add(startup2["StartupName"][i].strip())
    else:
        stripped = investorName.strip()
        if stripped in d:
            d[stripped].add((startup2["StartupName"][i].strip()))
        else:
            s = set()
```

```
d[stripped] = s
d[stripped].add(startup2["StartupName"][i].strip())
```

1.0.15 Another dictionary to keep a count of the number of startups each investor invested in and then printing it

```
[100]: d1 = {}
for i in d:
    if i == "":
        continue
    d1[i] = len(d[i])

d2 = sorted(d1, key = d1.get, reverse = True)
for i in d2[:5]:
    print(i, d1[i])
```

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

So these are the top 5 investor names who have invested maximum no. of times in different startup's. For him first priority would be to contact Sequoia Capital for his startup.

1.0.16 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.

```
[102]: startup3 = pd.read_csv("startup_funding.csv")

startup3["AmountInUSD"].fillna("0", inplace = True)
def num_1(value):
    if value != "":
        return int(float(value.replace(',','')))

```

```

startup3["AmountInUSD"] = startup3["AmountInUSD"].apply(num_1)

# Correcting the investment typo errors
startup3["InvestmentType"].replace('SeedFunding', 'Seed Funding', inplace =
↳ True)
startup3["InvestmentType"].replace('PrivateEquity', 'Private Equity', inplace =
↳ True)
startup3["InvestmentType"].replace('Crowd funding', 'Crowd Funding', inplace =
↳ True)

# Correcting the Startup names typo errors
startup3["StartupName"].replace("Ola Cabs", "Ola", inplace = True)
startup3["StartupName"].replace("OlaCabs", "Ola", inplace = True)
startup3["StartupName"].replace("Paytm Marketplace", "Paytm", inplace = True)
startup3["StartupName"].replace("Flipkart.com", "Flipkart", inplace = True)
startup3["StartupName"].replace("Oyo Rooms", "Oyo", inplace = True)
startup3["StartupName"].replace("Oyorooms", "Oyo", inplace = True)
startup3["StartupName"].replace("OyoRooms", "Oyo", inplace = True)
startup3["StartupName"].replace("OYO Rooms", "Oyo", inplace = True)
startup3["AmountInUSD"].fillna("0", inplace = True)

# TAKING CARE OF THE NA VALUES:
startup3.dropna(subset =
↳ ["StartupName", "InvestorsName", "InvestmentType"], inplace = True)
startup3_copy = startup3.copy()

# Limiting only to seed funding and crowd funding
Condition = (startup3_copy["InvestmentType"] == "Crowdfunding") |
↳ (startup3_copy["InvestmentType"] == 'Seed Funding')
startup3_copy = startup3_copy[Condition]

```

```

[103]: d = {}
for i in startup3_copy.index:
    investorName = startup3["InvestorsName"][i].strip()
    if "," in investorName:
        arrName = investorName.strip().split(",")
        for j in arrName:
            if j.strip() in d:
                d[j.strip()].add(startup3["StartupName"][i].strip())
            else:
                s = set()
                d[j.strip()] = s
                d[j.strip()].add(startup3["StartupName"][i].strip())
    else:
        stripped = investorName.strip()
        if stripped in d:

```



```

        d[stripped].add((startup3["StartupName"][i].strip()))
    else:
        s = set()
        d[stripped] = s
        d[stripped].add(startup3["StartupName"][i].strip())

d1 = {}
for i in d:
    if i == "":
        continue
    d1[i] = len(d[i])

del d1["Undisclosed Investors"]
del d1["Undisclosed investors"]

d2 = sorted(d1, key = d1.get, reverse = True)

for i in d2[:5]:
    print(i, d1[i])

```

Indian Angel Network 33
 Rajan Anandan 23
 LetsVenture 16
 Anupam Mittal 16
 Kunal Shah 14

So these are the top 5 investors who have invested in different startup's and their investment type is what is suited by him i.e. Crowd Funding or Seed Funding

1.0.17 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.

```
[104]: startup4 = pd.read_csv("startup_funding.csv")

startup4["AmountInUSD"].fillna("0", inplace = True)
def num_1(value):
    if value != "":
        return int(float(value.replace(',','')))

startup4["AmountInUSD"] = startup4["AmountInUSD"].apply(num_1)

# Correcting the investment typo errors
startup4["InvestmentType"].replace('SeedFunding', 'Seed Funding', inplace = True)
startup4["InvestmentType"].replace('PrivateEquity', 'Private Equity', inplace = True)
startup4["InvestmentType"].replace('Crowd funding', 'Crowd Funding', inplace = True)

# Correcting the Startup names typo errors
startup4["StartupName"].replace("Ola Cabs", "Ola", inplace = True)
startup4["StartupName"].replace("OlaCabs", "Ola", inplace = True)
startup4["StartupName"].replace("Paytm Marketplace", "Paytm", inplace = True)
startup4["StartupName"].replace("Flipkart.com", "Flipkart", inplace = True)
startup4["StartupName"].replace("Oyo Rooms", "Oyo", inplace = True)
startup4["StartupName"].replace("Oyorooms", "Oyo", inplace = True)
startup4["StartupName"].replace("OyoRooms", "Oyo", inplace = True)
startup4["StartupName"].replace("OYO Rooms", "Oyo", inplace = True)
startup4["AmountInUSD"].fillna("0", inplace = True)

# TAKING CARE OF THE NA VALUES:
startup4.dropna(subset = ["StartupName", "InvestorsName", "InvestmentType"], inplace = True)
startup4_copy= startup3.copy()
```

```
#Limiting only to private equity
Condition = (startup4_copy["InvestmentType"] == "Private Equity")
startup4_copy = startup4_copy[Condition]
```

```
[105]: d = {}
for i in startup4_copy.index:
    investorName = startup4["InvestorsName"][i].strip()
    if "," in investorName:
        arrName = investorName.strip().split(",")
        for j in arrName:
            if j.strip() in d:
                d[j.strip()].add(startup4["StartupName"][i].strip())
            else:
                s = set()
                d[j.strip()] = s
                d[j.strip()].add(startup4["StartupName"][i].strip())

    else:
        stripped = investorName.strip()
        if stripped in d:
            d[stripped].add((startup4["StartupName"][i].strip()))
        else:
            s = set()
            d[stripped] = s
            d[stripped].add(startup4["StartupName"][i].strip())

d1 = {}
for i in d:
    if i == "":
        continue
    d1[i] = len(d[i])

del d1["Undisclosed Investors"]
del d1["Undisclosed investors"]

d2 = sorted(d1, key = d1.get, reverse = True)

for i in d2[:5]:
    print(i, d1[i])
```

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

Since his startup got invested now he want to expand his startup and need top 5 investors who can do investment as Private Equity.