Project: Case Study-II

Justification

Part-1:

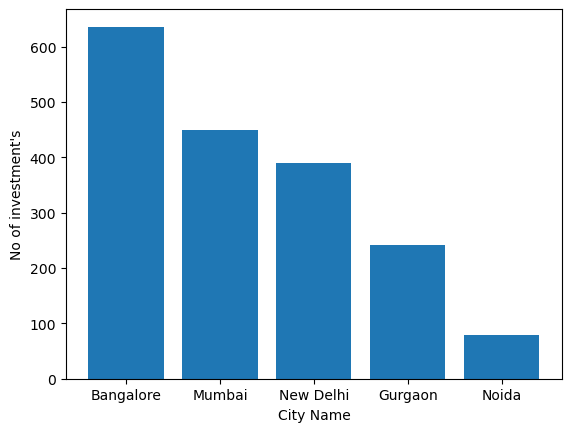
Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, you 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.

Explanation:

1. Imported required libraries namely NumPy, pandas, matplotlib.
2. Stored the data of file ‘Startupfunding.csv’ in a variable named ‘data’.
3. Dropped the rows containing NaN values in column ‘CityLocation’
4. Corrected the name such as ‘Delhi’, ‘bangalore’ in column ‘CityLocation’.
5. Iterating through the values of ‘CityLocation’ and splitting the locations in case of multiple location given.
6. Storing the splitted values as key in a dictionary ‘d’ and their number of occurrences as values only if city names are ‘New Delhi’, ‘Mumbai’, ‘Bangalore’, ‘Gurgaon’, ‘Noida’.
7. Creating a DataFrame named ‘counter’ of dictionary ‘d’ using keys as indexes and key-values as values corresponding to the index.
8. Then sorting the values present at column index [0] of the counter.
9. Since the city with largest no of investments is at top so printing the 0’th value of counter index & counter values.
10. Also plotting the bar graph (No of investments vs City) using matplotlib library.

Output:

Bangalore 636



Part-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.

Explanation:

1. Storing the data of file ‘Startupfunding.csv’ in a variable named ‘data’.
2. Dropping the rows containing NaN values in column ‘InvestorsName’.
3. Iterating through the values of ‘InvestorsName’ and splitting the Investors name in case of multiple investors are given.
4. Storing the splitted values in a list named ‘investor’.
5. Converting the list ‘investor’ to a numpy array.
6. Then applying np.unique() function to get unique values and their count respectively stored in list ‘unique’ and list ‘counts’.
7. Storing the indexes of the sorted values in ‘counts’ in ‘sortindex’ by applying np.argsort(counts).
8. Iterating through last 5 values of the ‘sortindex’ top get top 5 indexes having maximum count.
9. Printing the values at their respective index in ‘unique’ and ‘counts’
10. Also, plotting a pie chart with the percentage split of these top 5 investors.

Output:

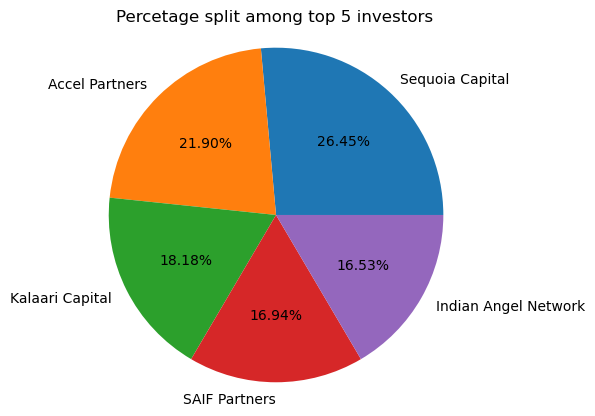
Sequoia Capital 64

Accel Partners 53

Kalaari Capital 44

SAIF Partners 41

Indian Angel Network 40



Part-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.

Explanation:

1. Storing a copy of data of file ‘Startupfunding.csv’ in a variable named ‘df’.
2. Dropping the rows containing NaN values in columns ‘StartupName’ & ‘InvestorsName’.
3. Correcting names of some startups like Flipkart, Ola, Oyo and Paytm.
4. Iterating through all indexes of df and splitting values of that index in column ‘InvestorsName’ in case of multiple names given.
5. Storing all investors name as key in a dictionary ‘d’ and a set containing name of companies that investor invested as key-values.
6. Making another dictionary ‘d1’ to store the investors name as key and the no of companies each of them invested as key-values.
7. A list ‘d2’ contains the name of top 5 investors sorted according to no of companies they invested in descending order.
8. Printing the name of top 5 investors with the no of companies they invested in.
9. Also, plotting pie chart with the percentage split of these top 5 companies.

Output:

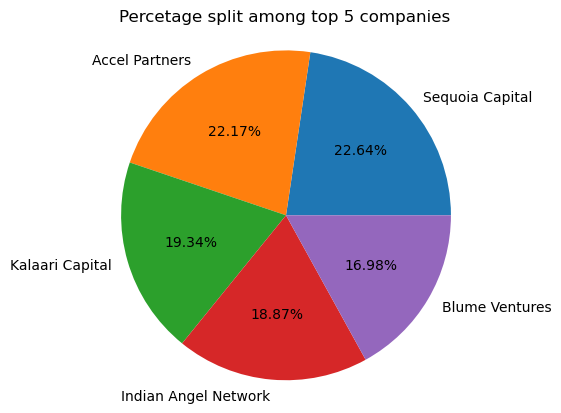
Sequoia Capital 48

Accel Partners 47

Kalaari Capital 41

Indian Angel Network 40

Blume Ventures 36



Part-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.

Explanation:

1. Storing a copy of data of file ‘Startupfunding.csv’ in a variable named ‘data’.
2. Dropping the rows containing NaN values in columns ‘InvestorsName’ , ‘InvestmentType’ ,’StartupName’.
3. Correcting names of some startup names like Flipkart, Ola, Oyo and Paytm and some investment type names ‘Seed Funding’ and ‘Crowd Funding’.
4. Removing all rows containing investment type other than ‘Seed Funding’ and ‘Crowd Funding’.
5. Iterating through all indexes of data and splitting values of that index in column ‘InvestorsName’ in case of multiple names given.
6. Storing all investors name as key in a dictionary ‘d’ and a set containing name of companies that investor invested as key-values.
7. Making another dictionary ‘d1’ to store the investors name as key and the no of companies each of them invested as key-values.
8. A list ‘d2’ contains the name of top 5 investors sorted according to no of companies they invested in descending order.
9. Printing the name of top 5 investors with the no of companies they invested in.
10. Also plotting the bar graph (No of investments vs Investor name) using matplotlib library.

Output:

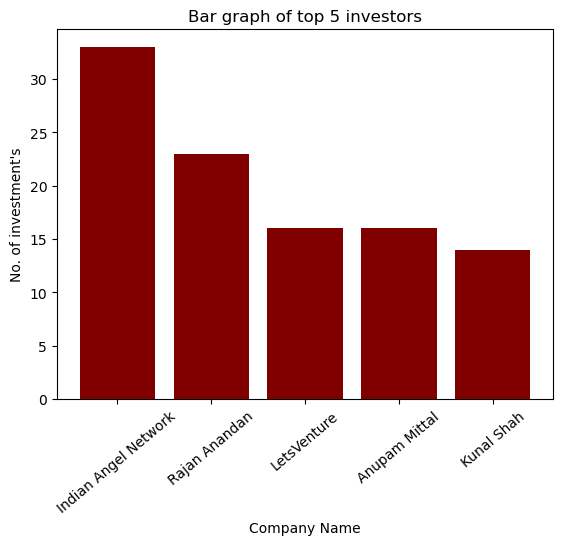
Indian Angel Network 33

Rajan Anandan 23

LetsVenture 16

Anupam Mittal 16

Kunal Shah 14



Part-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.

Explanation:

1. Storing a copy of data of file ‘Startupfunding.csv’ in a variable named ‘data’.
2. Dropping the rows containing NaN values in columns ‘InvestorsName’ , ‘InvestmentType’ ,’StartupName’.
3. Correcting names of some startup names like Flipkart, Ola, Oyo and Paytm and some investment type name ‘Private Equity’.
4. Removing all rows containing investment type other than ‘Private Equity’.
5. Iterating through all indexes of data and splitting values of that index in column ‘InvestorsName’ in case of multiple names given.
6. Storing all investors name as key in a dictionary ‘d’ and a set containing name of companies that investor invested as key-values.
7. Making another dictionary ‘d1’ to store the investors name as key and the no of companies each of them invested as key-values.
8. A list ‘d2’ contains the name of top 5 investors sorted according to no of companies they invested in descending order.
9. Printing the name of top 5 investors with the no of companies they invested in.
10. Also plotting the bar graph (No of investments vs Investor name) using matplotlib library.

Output:

Sequoia Capital 46

Accel Partners 43

Kalaari Capital 35

Blume Ventures 27

SAIF Partners 24

