Justification

Problem 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.

Output 1:

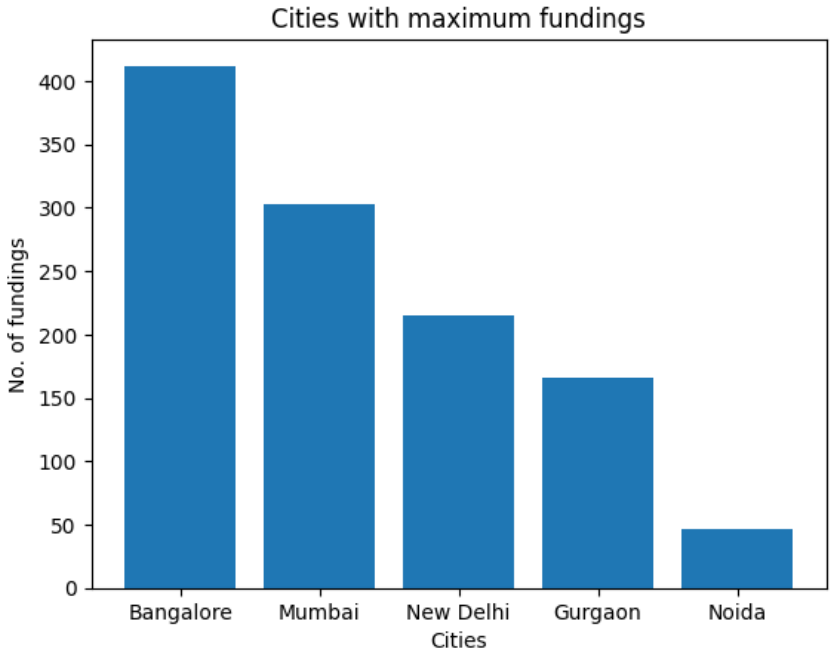
```
Bangalore 412

Mumbai 303

New Delhi 215

Gurgaon 166

Noida 46
```



Explanation: First I dropped the rows having NaN values in CityLocation column in dataset, then I separated the Cities by using a SeperateCity function, then after that corrected the spelling error in bangalore and Delhi (noted in problem). And excluding all the cities except NCR, Bangalore and Mumbai (as mentioned in the problem statement) Next step was to convert the rows of AmountInUSD column to

float then use groupby function to make a column the cleaned columns CityLocation and AmountInUSD, then sorted that in descending order to get the top 5 cities and then print it. Here, I also used a bar graph to represent the top 5 city and their number of fundings which clearly showed that most number of fundings were provided in Bangalore City.

Problem 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

Output 2:

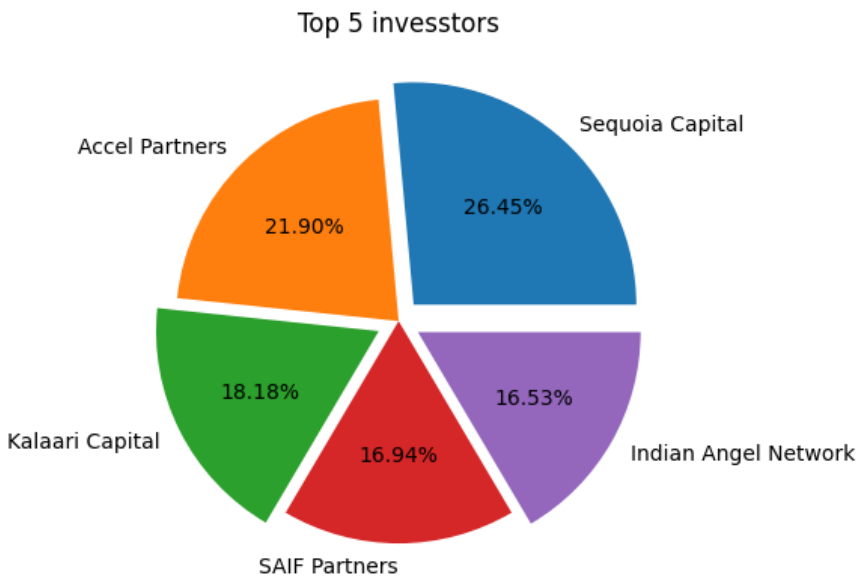
```
Sequoia Capital 64

Accel Partners 53

Kalaari Capital 44

SAIF Partners 41

Indian Angel Network 40
```



Explanation:

First of all, I dropped rows having NaN values in InvestorsName column of the dataset. Then handled the cases of Undisclosed Investors and made a list for investors name and iterated through the InvestorsName column of the data and then separated the investors who had invested in the same startup and then made a dictionary for it which then was sorted by collections module. After that I plotted a pie chart showing the top 5 investors which clearly showed that Sequoia Capital invested the most number of times.

Problem 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.

Output 3:

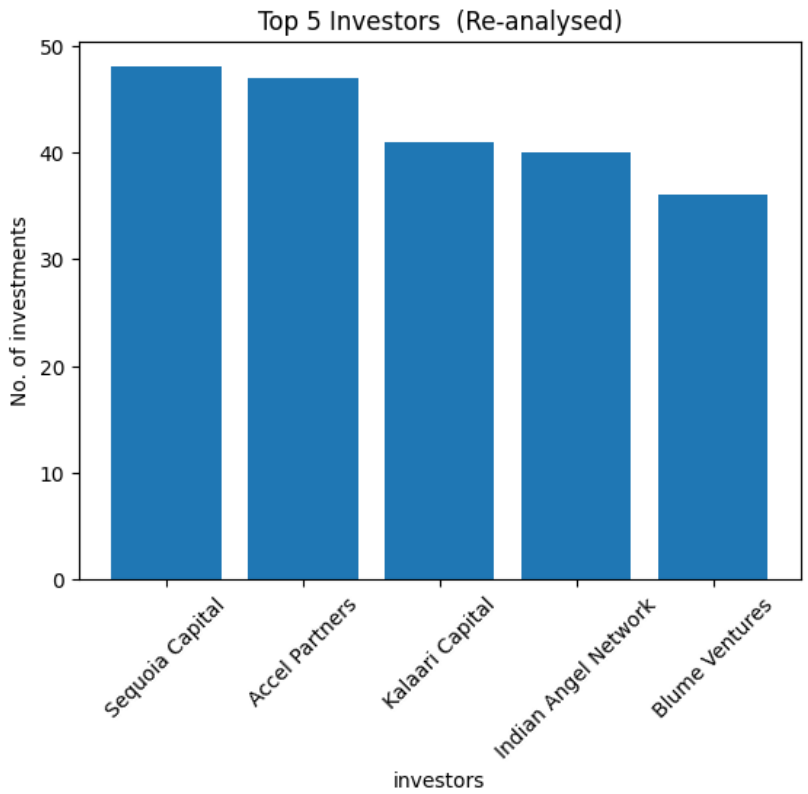
```
Sequoia Capital 48

Accel Partners 47

Kalaari Capital 41

Indian Angel Network 40

Blume Ventures 36
```



Explanation: First of all, I dropped rows having NaN values in InvestorsName column of the given dataset and then corrected the error in the StartupNames Oyo, Ola, Flipkart, and Paytm (as mentioned in the problem) then handled the Undisclosed Investors by dropping them and then made a list of

startups and investors which gave a DataFrame as the output showing the top 5 investors. Also, after that I plotted a bar graph which clearly shows that Sequoia Capital made maximum number of investments and that was 48.

Problem 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.

Output 4:

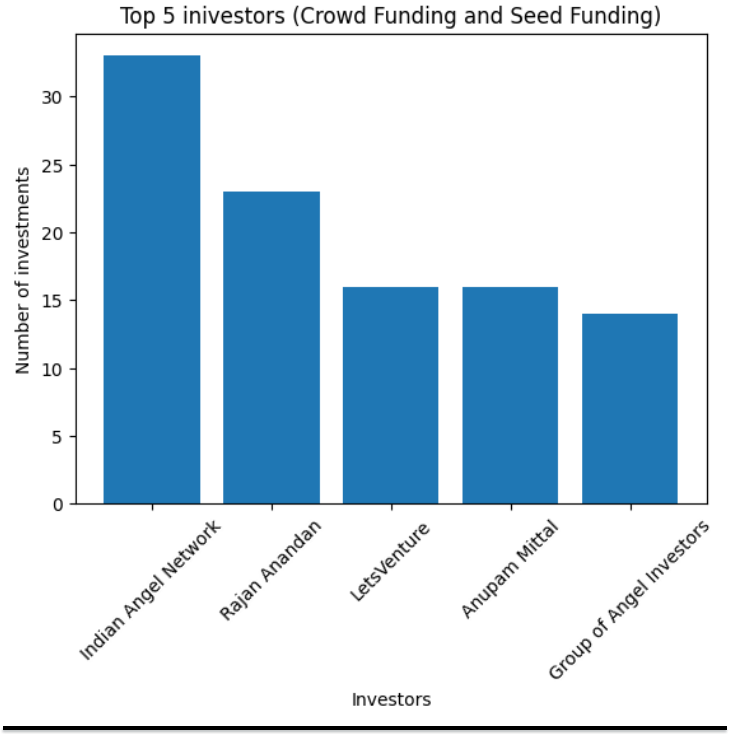
```
Indian Angel Network 33

Rajan Anandan 23

LetsVenture 16

Anupam Mittal 16

Group of Angel Investors 14
```



Explanation: First I dropped the rows having NaN value in InevstorsName and StartupName columns of the given dataset. Then I corrected the errors in InvestmentType, StartupName and InvestorName columns. Next, I set the data only to Crowd Funding and Seed Funding investment type, After that, I made a investor and startup list by which I made a DataFrame and sorted it to find the top 5 investors.

Also, I plotted a bar graph which clearly shows that Indian Angel Network made maximum fundings of Crowd Funding or Seed Funding type.

Problem 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.

Output 5:

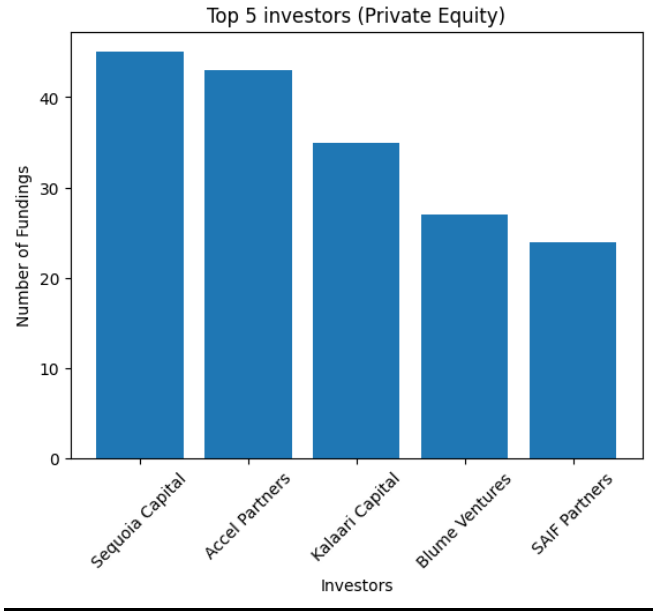
```
Sequoia Capital 45

Accel Partners 43

Kalaari Capital 35

Blume Ventures 27

SAIF Partners 24
```



Explanation: First I dropped the rows having NaN value in InevstorsName and StartupName columns of the given dataset. Then I corrected the errors in InvestmentType, StartupName and InvestorName columns. And then I set the data only to Private Equity investment type. After that, I made a investor and startup list by which I made a DataFrame and sorted it to find the top 5 investors.

Also, I plotted a bar graph which clearly shows that Sequoia Capital made maximum fundings of Private Equity type.