ANSWER 1

1. The required python libraries were loaded, and the dataset was also loaded into a dataframe. The data was then formatted based on the locations as stated by the question.
2. Afterwards, the locations which had wrong spellings of cities were corrected, and the entire column of the dataframe was stored in a list.
3. The list was then converted to a NumPy array, and only the select locations of 'Bangalore', 'New Delhi', 'Mumbai' and 'Gurgaon' were chosen whose funding was to be considered.
4. A dictionary was made regarding the total investments made to the cities, and we deduced that Bangalore received many fundings for budding start-ups.
5. The location vs funding graph was plotted, and the required customizations on the chart were made to make it look more informative and appealing.

ANSWER 2

1. The required pandas libraries were loaded, and the dataset was also loaded into a dataframe.
2. The null values of the Investor Name column were dropped, and so were the columns that had the data entry as 'Undisclosed'.
3. When there was more than one investor for a particular start-up, they all were separated, and every time they invested in a specific start-up, their investment count increased. And this data was stored in a dictionary, with the keys as the investor names and the number of times they invested as the key values.
4. The keys and values of this dictionary were then converted to a NumPy array separately and based on the key values, the keys were sorted in an ascending order.
5. Then from the end, the last 5 values were chosen i.e., the top 5 investors who had invested the greatest number of times.
6. These investors and the corresponding times they invested was then displayed.

ANSWER 3

1. The required pandas libraries were loaded and using them the dataset was also loaded into a dataframe.
2. The names of the start-ups which were stored all wrong were corrected to ensure the data analysis did not yield different names for the same start-ups.
3. The investor names and the start-ups columns were converted into a NumPy array. And another empty dictionary was created.
4. For every start-up if there are more than 1 investors, then they were separated and corresponding to every start-up all the names of the investors were stored in a list. Hence the start-up names were stored as keys and the values to these keys were all the investors that invested in them.
5. Now this is the most important step. Since many investors invested in the same start-up multiple times, we have to count them only once., For that, the list which was stored as a value to investors were converted to sets. This way the repeated investor names were dropped. After that the length of the set was stored in to a new dictionary as value to the keys that represented the investors.
6. The entire dictionary was sorted the basis of these values and the top 5 investors with the highest number investments was stored in lists separately.
7. These lists were then used to visualize data in the form of a bar graph which was customized to look better.

ANSWER 4

1. The required pandas libraries were loaded and using them the dataset was also loaded into a dataframe.
2. The names of the start-ups which were stored all wrong were corrected to ensure the data analysis did not yield different names for the same start-ups.
3. Now the spelling mistakes in the Investment Type column were fixed and Private Equity and Debt Funding investments were dropped.
4. The investor names and the start-ups columns were converted into a NumPy array. And another empty dictionary was created.
5. For every start-up if there are more than 1 investors, then they were separated and corresponding to every start-up all the names of the investors were stored in a list. Hence the start-up names were stored as keys and the values to these keys were all the investors that invested in them.
6. Now this is the most important step. Since many investors invested in the same start-up multiple times, we have to count them only once., For that, the list which was stored as a value to investors were converted to sets. This way the repeated investor names were dropped. After that the length of the set was stored in to a new dictionary as value to the keys that represented the investors.
7. The entire dictionary was sorted the basis of these values and the top 5 investors with the highest number investments was stored in lists separately.
8. These lists were then used to visualize data in the form of a bar graph which was customized to look better.

ANSWER 5

1. The required pandas libraries were loaded and using them the dataset was also loaded into a dataframe.
2. The names of the start-ups which were stored all wrong were corrected to ensure the data analysis did not yield different names for the same start-ups.
3. Now the spelling mistakes in the Investment Type column were fixed and all rows with investment type other than ‘Private Equity’ were dropped.
4. The investor names and the start-ups columns were converted into a NumPy array. And another empty dictionary was created.
5. For every start-up if there are more than 1 investors, then they were separated and corresponding to every start-up all the names of the investors were stored in a list. Hence the start-up names were stored as keys and the values to these keys were all the investors that invested in them.
6. Now this is the most important step. Since many investors invested in the same start-up multiple times, we have to count them only once., For that, the list which was stored as a value to investors were converted to sets. This way the repeated investor names were dropped. After that the length of the set was stored in to a new dictionary as value to the keys that represented the investors.
7. The entire dictionary was sorted the basis of these values and the top 5 investors with the highest number investments was stored in lists separately.
8. These lists were then used to visualize data in the form of a bar graph which was customized to look better.

ANALYSIS ON THE BASIS OF THE ABOVE 5 QUESTIONS

From Q1 one can clearly see that the ideal Indian city to set up a start-up venture would be Bangalore since the city has seen over 635 investments made in the time from 2015 to 2017.

But which Investor should one approach for an investment opportunity? The most ideal case would be to approach Sequoia Capital which on its own has invested in over 64 start-ups. But one can even approach Accel partners, Kalaari Capital, SAIF partners and Indian Angel Network because these are the top investors in the country. Even though, these investors have invested in multiple start-ups many times, but even if we see the unique number of start-ups, these investors still come out on top.

Once the top investors are finalized which kind of funding to for will narrow down the investors? From the statistics, since the start-up is a beginner venture, the best suited funding would be Seed Funding or Crowd Funding. And one can clearly see that Indian Angel Network was the top most investor in these two categories with investments made in 33 different companies.

Now that an investment has been made by a good investor and a long time has passed since the start-up's launch, it was now time to expand. For that type the best-suited investment type is Private Equity. So, from the data one can understand that Sequoia Capital made 45 investments as a Private Equity Investor, making them a good candidate for asking for an investment for expansion. Other alternatives include the other top 5 investors in the entire country.