Project Case Study Part II

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.

Solution:

Imported Pandas Numpy and Matplotlib Libraries

Read the CSV file

#We have dropped all the N/A values in the required columns

Created Separate method to get the names of the cities by the split indicator "/"

Applied Separate method on the City Location column to apply this on all the row values

Corrected city names of “Delhi” and “Bangalore” as mentioned in the problem statement

#Removed Undisclosed and unknown values from the 'Amount in USD' column

#Grouped 'City Locations' and 'Amount in USD’ columns to get the no.of times each city got funded for Startups

#Sorted the result values from above step in descending order to get the top 5 city locations where the most no.of funding is done.

Created 2 lists for storing the index and values of the DataFrame which helped in plotting the bar Graph

#Checking for the Eligible Cities with the index data of the DataFrame and if the requirement matches, I have appended the values and index data to the previously created lists

Used a counter variable to check for the Exact Top 5 Cities.

Plotted Bar Graph for the detailed analysis

———

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.

Solution:

Imported Pandas Numpy and Matplotlib Libraries

Read the CSV file

#We have dropped all the N/A values in the required columns

#ignoring the values of investors whose name is given as "Undisclosed Investors"

Seperated Investor Names by split indicator "," and added these Investor Names to a List

copied the list values to Dictionary

#Used Counter method to count the number of occurence of investor names

Used most_common method to find the most repeated investor names

#Created 2 lists for storing the index and values of the DataFrame which helped in plotting the Pie chart

Used a counter variable to check for the Exact Top 5 Investors.

Plotted Pie Chart for the detailed analysis

——

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.

Solution:

#imported Pandas, Numpy and Matplotlib libraries

#Read CSV file

#Corrected names of some important startups like "Ola", "Oyo", "Flipkart" and "Paytm" by replace method.

#Removed Undisclosed and unknown Investors from the "Investors Name" Column

#Separated names of Investors based on "," split criteria

#created 2 lists for adding the values of separated investor names and the corresponding startups funded.

#Created a new DataFrame by the values of above created lists as columns.

fetched unique values by grouping Investor Names and Startup Names

Sorted the grouped values in descending order to get the top 5 investors who have invested maximum number of times in different companies.

#Created 2 lists for storing the index and values of the DataFrame which helped in plotting the Bar Graph

Used a counter variable to check for the Exact Top 5 Investors.

#plotted bar graph for the detailed analysis

——

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.

Solution:

#imported Pandas, Numpy and Matplotlib libraries

#Read CSV file

#Corrected names of some important startups like “Ola”, “Oyo”, “Flipkart” and ”Paytm” by replace method.

#Corrected names of some important Investment Types like Private Equity, Seed Funding, Debt Funding and Crowdfunding

#Removed Undisclosed and unknown Investors from the “Investors Name” Column

#Separated names of Investors based on "," split criteria

Created 2 lists for adding the values of separated investor names and the corresponding startups names which has the Investment type as either Seed Funding or Crowd Funding.

#Created a new DataFrame by the values of above created lists as columns.

#Grouped ‘Investors Names' and ‘Startup Names’ columns to get the no.of times each Investor has funded for Startups

Sorted the grouped values in descending order to get the top 5 investors who have invested in Startups of Investment type either Crowdfunding or Seed Funding.

#Created 2 lists for storing the index and values of the DataFrame which helped in plotting the Bar Graph

Used a counter variable to check for the Exact Top 5 Investors.

#plotted bar graph for the detailed analysis

——

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.

Solution:

#imported Pandas, Numpy and Matplotlib libraries

#Read CSV file

#Corrected names of some important startups like “Ola”, “Oyo”, “Flipkart” and ”Paytm” by replace method.

#Corrected names of some important Investment Types like Private Equity, Seed Funding, Debt Funding and Crowdfunding

#Removed Undisclosed and unknown Investors from the “Investors Name” Column

#Separated names of Investors based on "," split criteria

Created 2 lists for adding the values of separated investor names and the corresponding startups names which has the Investment type as Private Equity

#Created a new DataFrame by the values of above created lists as columns.

#Grouped ‘Investors Names' and ‘Startup Names’ columns to get the no.of unique values

Sorted the grouped values in descending order to get the top 5 investors who have invested in Startups of Investment type Private Equity

#Created 2 lists for storing the index and values of the DataFrame which helped in plotting the Bar Graph

Used a counter variable to check for the Exact Top 5 Investors.

#plotted bar graph for the detailed analysis