Project : Case Study

Question 1 answer:

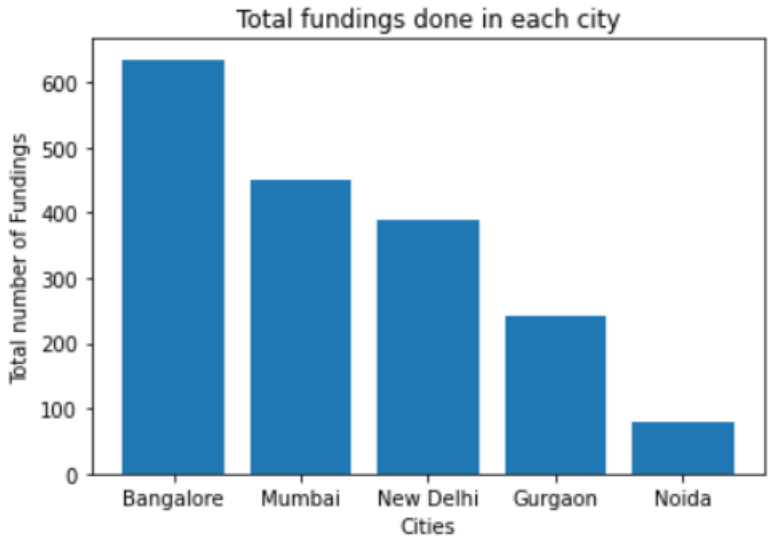
Bangalore 635

Mumbai 449

New Delhi 389

Gurgaon 241

Noida 79



1.Import pandas library as pd and create a dataframe 'df' using (pd.read_csv) function

2.In order to remove all the 'NaN' values in the citylocation colum use dropna function which has two aarguments (subset,inplace). subset is the column in which we want to ramove the 'NaN' values.Inplace==True if we want it change the main dataframe 'df'

3.Some startups are given with two citylocations in order to get the citylocation present in India we use apply function where a function named seperate is passed as argument. We create a function named seperate which takes the citylocation as argument and splits at '/' and returns the city present in india

4. Use replace function for the city column to replace all the citylocations named Delhi with 'New Delhi' and 'bangalore' with 'Bangalore'.

5.Using the value_counts function create a pandas series which gives the frequency i.e no of times each city repeated store it a variable 'data'

6. Create an array 'city' containing the required citylocation as elements.Create another empty array 'num'

7.Acces the requried frequencies from 'data' by iteration through the array 'city'. Append each value into the array 'num'

8.Print the city and the corresponding frequency

9.Plot a bar graph using plt.bar(x,y), add title using plt.title, add label for x and y axes using plt.xlabel() and plt.ylabel()3. Use replace function for the city column to replace all the citylocations named Delhi with 'New Delhi' and 'bangalore' with 'Bangalore'

Question 2 answer :

Sequoia Capital 64

Accel Partners 53

Kalaari Capital 44

SAIF Partners 41

Indian Angel Network 40

1.Import pandas library as pd and create a dataframe 'df' using (pd.read_csv) function

2.In order to remove all the 'NaN' values in the 'InvestorsName' colum use dropna function which has two arguments (subset,inplace). subset is the column in which we want to ramove the 'NaN' values.Inplace==True if we want it change the main dataframe 'df'

3.In some rows there are more than one investor name so we use df['InvestorsName'].apply() function with an argument 'seperate' which is function returns list of all investors of that row.It modifies the 'InvestorsName' colunm

4.We use df.explode() function which takes the column name as argument. It create a new row for each element in the list of the column 'InvestorsName'

5.Using df['InvestorsName'].value_counts() function get all the frequencies of each investor stores in a varible 'data'. Using data.index[:5] create an array 'inv' and using data.values[:5] create an array 'num' which contains investor names and their frequencies.

6.print the investor and his corresponing frequency using for loop. Plot a pie chart using plt.pie(num,labels=inv,autopct='%.2f%%') first argument is the array of numbers which is to be made into percentage and labels is their coresponding names. autopct is used to print the percentage over the sector or partion of pie chart

Question 3 answer:

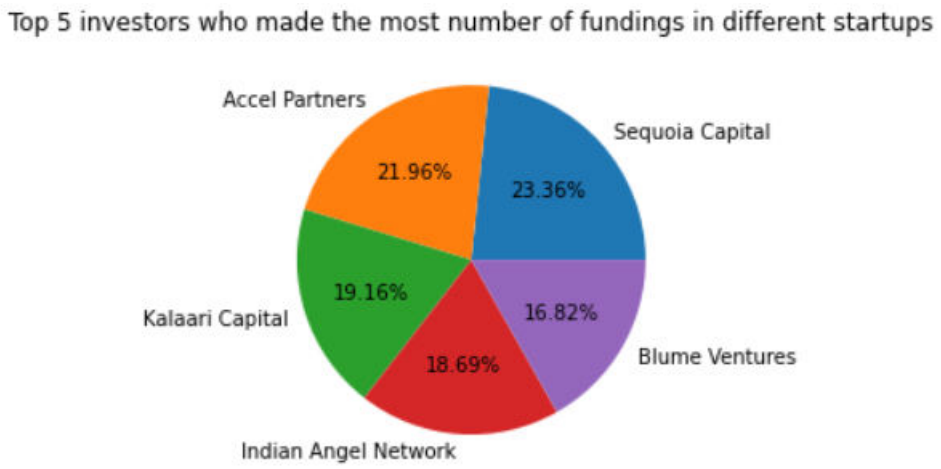
Sequoia Capital 50

Accel Partners 47

Kalaari Capital 41

Indian Angel Network 40

Blume Ventures 36



1.Import pandas library as pd and create a dataframe 'df' using (pd.read_csv) function

2.In order to remove all the 'NaN' values in the 'InvestorsName' colum use dropna function which has two arguments (subset,inplace). subset is the column in which we want to ramove the 'NaN' values.Inplace==True if we want it change the main dataframe 'df'.

3.By using replace function we replace all the error start up names with correct Names

4.In some rows there are more than one investor name so we use df['InvestorsName'].apply() function with an argument 'seperate' which is function returns list of all investors of that row.It modifies the

'InvestorsName' colunm

5.We use df.explode('InvestorsName') function which takes the column name as argument. It create a new row for each element in the list of the column 'InvestorsName'

6.We create different dataframes for each unique investor by using df.groupby('InvestorsName') stored in 'gb'. It takes column names as the argument and creates dataframes of each nique value of that colunm.

7.Created a dictionary 'd' and while iterating interating over the groupbydataframe 'gb' we we create an array 'k' of unique values of startupnames by using inv_info["StartupName"].unique() the length of this 'k' array is key value pair of the company in the dictionary 'd'. (Here len(k) represents the no of unique companies the investor invested)

8.Now using pd.DataFrame(list(d.values()),list(d.keys())) we create a data frame 'dframe' .In orer to keep all the values in descending order we use dframe.sort_values(by=[0],ascending=False). The by argument represent the colums that to be sorted

9.We create two arrays 'inv' and 'num' by using dframe.index[:5] which gives the list of top 5 investors and by iterating over the 'dframe we get their coresponding frequencies that are appended into 'num' array

10.Using for loop we print investor name and corresponding frequency. Plot a pie chart using plt.pie(num,labels=inv,autopct='%.2f%%') first argument is the array of numbers which is to be made into percentage and labels is their coresponding names. autopct is used to print the percentage over the sector or partion of pie chart

Question 4 answer:

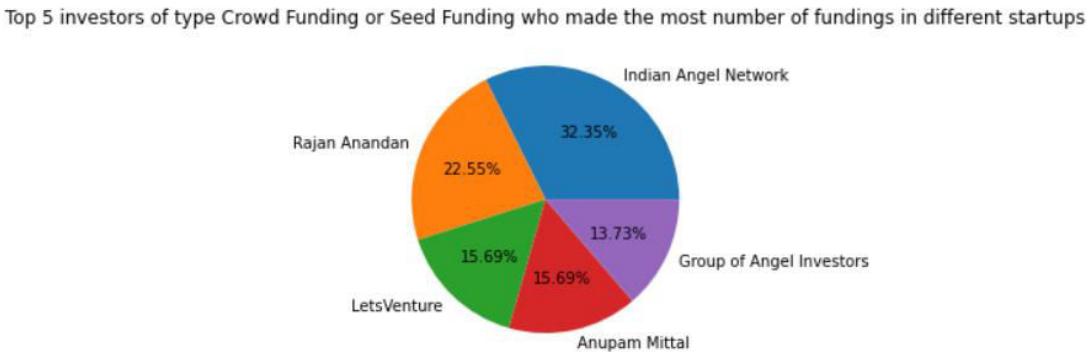
Indian Angel Network 33

Rajan Anandan 23

LetsVenture 16

Anupam Mittal 16

Group of Angel Investors 14



1.Import pandas library as pd and create a dataframe 'df' using (pd.read_csv) function

2.In order to remove all the 'NaN' values in the 'InvestorsName' colum use dropna function which has two arguments (subset,inplace). subset is the column in which we want to ramove the 'NaN' values.Inplace==True if we want it change the main dataframe 'df'.

3.By using replace function we replace all the error start up names with correct Names and all the mistaken investment types by correct names

4. We modify the dataframe 'df' which contains only investors of type Crowd Funding and seed funding

5.In some rows there are more than one investor name so we use df['InvestorsName'].apply() function with an argument 'seperate' which is function returns list of all investors of that row.It modifies the 'InvestorsName' colunm

6.We use df.explode('InvestorsName') function which takes the column name as argument. It creates a new row for each element in the list of the column 'InvestorsName'

7.We create different dataframes for each unique investor by using df.groupby('InvestorsName') stored in 'gb'. It takes column names as the argument and creates dataframes of each nique value of that colunm.

8.Created a dictionary 'd' and while iterating interating over the groupbydataframes 'gb' we we create an array 'k' of unique values of startupnames by using inv_info["StartupName"].unique() the length of this 'k' array is key value pair of the company in the dictionary 'd'. (Here len(k) represents the no of unique companies the investor invested)

9.Now using pd.DataFrame(list(d.values()),list(d.keys())) we create a data frame 'dframe' .In orer to keep all the values in descending order we use dframe.sort_values(by=[0],ascending=False). The by argument represent the colums that to be sorted

10. We remove the 'Undisclousure Investors' present in the top 5 as we have to ignore them and also we must ignore we modified the NaN as "" (empty string) in the seperate function we have to remove "" using dframe.drop("",inplace==True)

11.We create two arrays 'inv' and 'num' by using dframe.index[:5] which gives the list of top 5 investors and by iterating over the 'dframe we get their coresponding frequencies that are appended into 'num' array

12.Using for loop we prin tinvestor name and corresponding frequency. Plot a pie chart using plt.pie(num,labels=inv,autopct='%.2f%%') first argument is the array of numbers which is to be made into percentage and labels is their coresponding names. autopct is used to print the percentage over the sector or partion of pie chart

Question 5 answer:

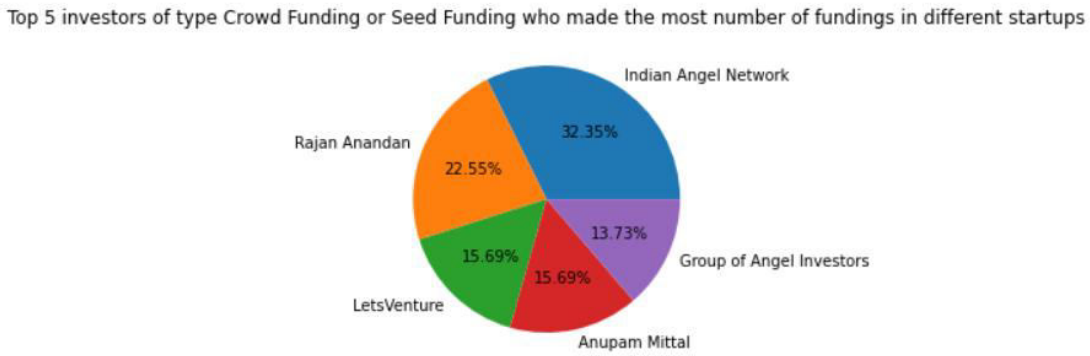
Indian Angel Network 33

Rajan Anandan 23

LetsVenture 16

Anupam Mittal 16

Group of Angel Investors 14



1.Import pandas library as pd and create a dataframe 'df' using (pd.read_csv) function

2.In order to remove all the 'NaN' values in the 'InvestorsName' colum use dropna function which has two arguments (subset,inplace). subset is the column in which we want to ramove the 'NaN' values.Inplace==True if we want it change the main dataframe 'df'.

3.By using replace function we replace all the error start up names with correct Names and all the mistaken investment types by correct names

4. We modify the dataframe 'df' which contains only investors of type Private Equity

5.In some rows there are more than one investor name so we use df['InvestorsName'].apply() function with an argument 'seperate' which is function returns list of all investors of that row.It modifies the 'InvestorsName' colunm

6.We use df.explode('InvestorsName') function which takes the column name as argument. It creates a new row for each element in the list of the column 'InvestorsName'

7.We create different dataframes for each unique investor by using df.groupby('InvestorsName') stored in 'gb'. It takes column names as the argument and creates dataframes of each nique value of that colunm.

8.Created a dictionary 'd' and while iterating interating over the groupbydataframes 'gb' we we create an array 'k' of unique values of startupnames by using inv_info["StartupName"].unique() the length of this 'k' array is key value pair of the company in the dictionary 'd'. (Here len(k) represents the no of unique companies the investor invested)

9.Now using pd.DataFrame(list(d.values()),list(d.keys())) we create a data frame 'dframe' .In orer to keep all the values in descending order we use dframe.sort_values(by=[0],ascending=False). The by argument represent the colums that to be sorted

10. We remove the 'Undisclousure Investors' present in the top 5 as we have to ignore them and also we must ignore we modified the NaN as "" (empty string) in the seperate function we have to remove "" using dframe.drop("",inplace==True)

11.We create two arrays 'inv' and 'num' by using dframe.index[:5] which gives the list of top 5 investors and by iterating over the 'dframe we get their coresponding frequencies that are appended into 'num' array

12.Using for loop we prin tinvestor name and corresponding frequency. Plot a pie chart using plt.pie(num,labels=inv,autopct='%.2f%%') first argument is the array of numbers which is to be made into percentage and labels is their coresponding names. autopct is used to print the percentage over the sector or partion of pie chart