FUND ANALYSIS

- Objective - Identify best countries, sector and rounding fund type.
- Constraint
  - Investment must be between 5 to 15 million
  - Invest only in English speaking countries
- Strategies
  - Invest where most investors are investing

Goals of data analysis

Goals are divided into three sub-goals:

1) Country analysis
2) Investment type analysis
3) Sector analysis

We will start with Data cleaning and we have filled up null values or if somewhere not necessary then we have also removed data. We imputed raised amount with the help of country and investment round type. We impute null value of Raise amount with average amount of that country (USA) invested investment type (e.g. venture).

We will only consider English Speaking countries.

Below is the list of 30 English speaking countries.

Sr.no	Country Code	Name
1	IND	India
2	USA	USA
3	CAN	Canada
4	GBR	United Kingdom
5	AUS	Australia
6	SGP	Singapore
7	IRL	Ireland
8	NZL	New Zealand
9	CHE	Switzerland
10	NGA	Nigeria
11	ZAF	South Africa
12	MUS	Mauritius
13	KEN	Kenya
14	PHL	Philippines
15	GHA	Ghana
16	UGA	Uganda
17	PAK	Pakistan
18	ZWE	Zimbabwe
19	BRB	Barbados
20	TTO	Trinidad and Tobago
21	CMR	Cameroon
22	MLT	Malta
23	ZMB	Zambia
24	JAM	Jamaica
25	KNA	Saint Kitts and Nevis
26	RWA	Rwanda
27	DMA	Dominica
28	BLZ	Belize
29	GRD	Grenada
30	SYC	Seychelles

1. **Data cleaning**
   - After merging rounds and companies, We have total 114942 rows and 16 columns in our dataset
   - Remove all Non-English-speaking countries
     After this we are left with rows=87002
   - Remove all unnecessary columns
     Now rows =87002 and columns=8
   - Check for null values

     Fig 1.1

```
company_permalink          0
funding_round_type         0
raised_amount_usd      13028
permalink                  0
category_list           1286
status                     0
country_code               0
founded_at             12878
dtype: int64
```

     We can see that we have null entries in raised amount column
     We will impute data in it by replacing null values with mean of country and funding type mean values
   - By looking at the below figures, we can see that permalink is the primary key of company table

     Fig 1.2

```
permalink           0
name                1
homepage_url     5058
category_list    3148
status              0
country_code     6958
state_code       8547
region           8030
city             8028
founded_at      15221
dtype: int64
```

	permalink	name	homepage_url	category_list	status	country_code	state_code	region	city	founded_at
count	66368	66367	61310	63220	66368	59410	57821	58338	58340	51147
unique	66368	66102	61191	27296	4	137	311	1092	5111	3978
top	/Organization/Ownerlistens	Peach	http://www.askforoffer.com	Software	operating	USA	CA	SF Bay Area	San Francisco	01-01-2012
freq	1	4	5	3995	53034	37601	12900	8804	3526	2730

   - We will compare both company permalink and permalink – for this we first need to change everything in lower case

Checkpoint 1

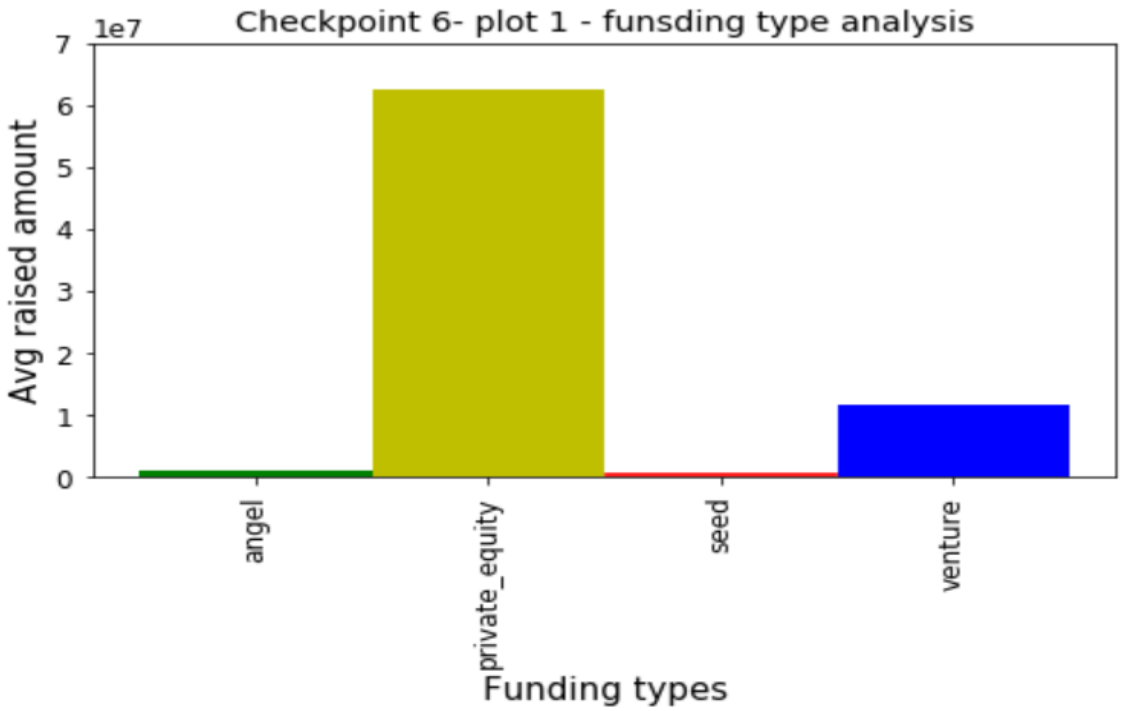
How many **unique companies** are present in **rounds2**?	**90247**
How many **unique companies** are present in **companies**?	**66368**
In the **companies** data frame, which column can be used as the unique key for each company? Write the **name of the column**.	**permalink**
Are there any companies in the rounds2 file which are not present in companies? Answer yes or no: **Y/N**	**N**
Merge the two data frames so that all variables (columns) in the **companies** frame are added to the **rounds2** data frame. Name the merged frame **master_frame**. How many observations are present in master_frame?	**114942**

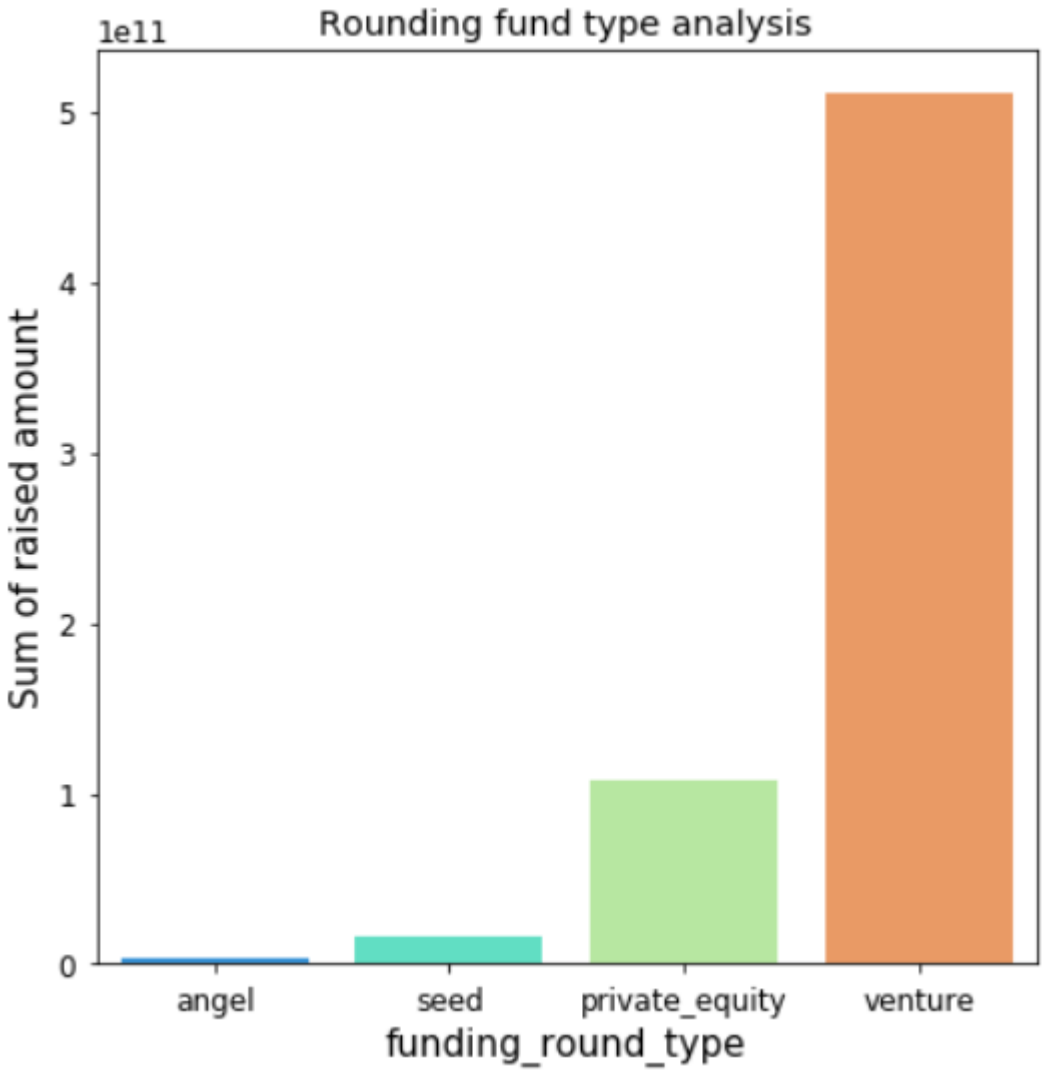
Checkpoint 2

2. Funding type analysis

- Create a master_frame by merging both companies and rounds2 data frames
- As per our requirement we are going to consider only 4 types of funding types Angel seed private equity and ventures
- Check below graph of all 4 types of investment with average of raised amount

- Fig 2.1



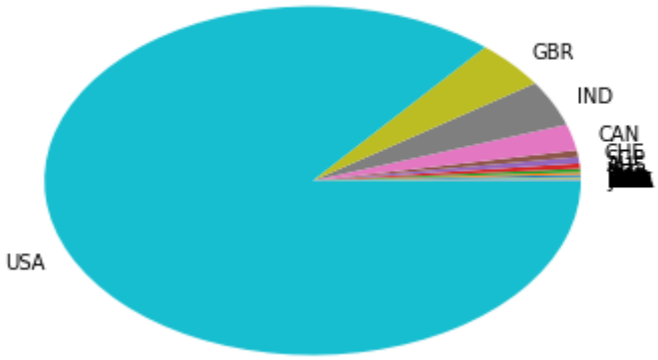


funding_round_type	raised_amount_usd
angel	$850,410.38
private_equity	$62,489,908.00
seed	$820,441.33
venture	$11,549,426.25

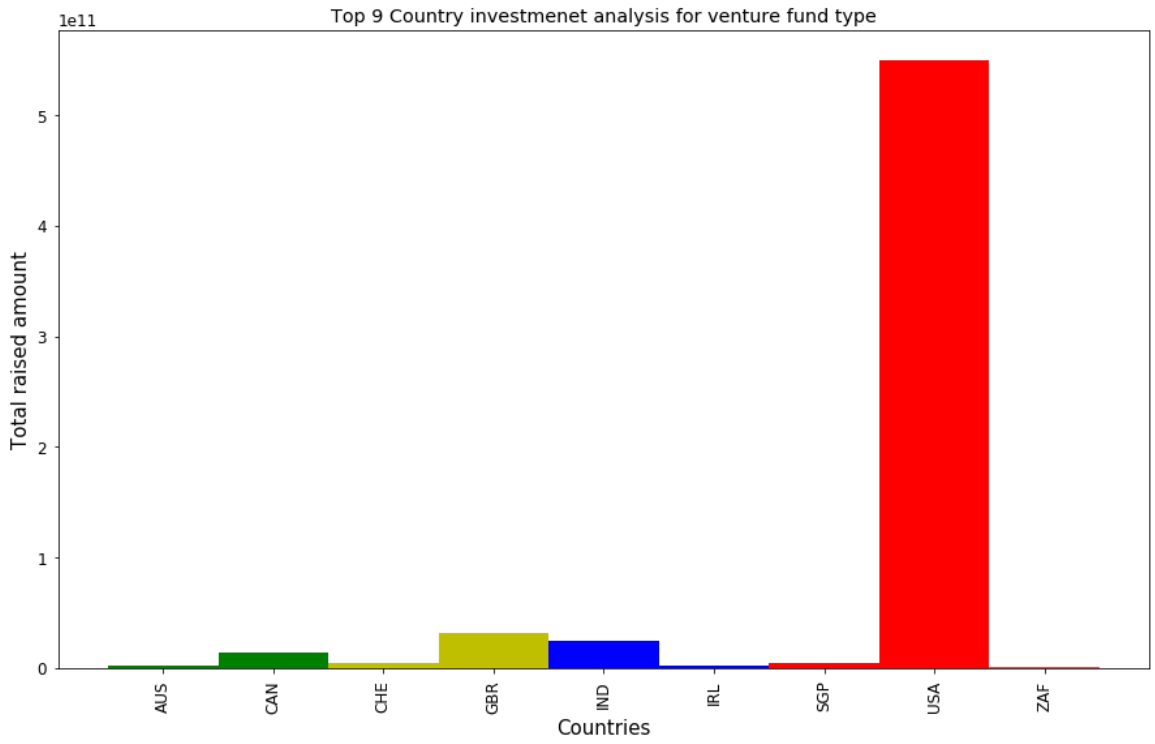
- From above table and figure, as per our requirement 5 to 15 million of investment is valid for **venture** funding type.
- So we are left with 44257 rows

Checkpoint 3

1. **Country analysis:**
   - Remove all non-English speaking countries from data
   - As per the graph below
     We can infer that



   - Pie chart showing distribution of investment, we can see huge difference in USA and other countries
   - USA is followed by GBR,IND CAN,CHE,SGP



-

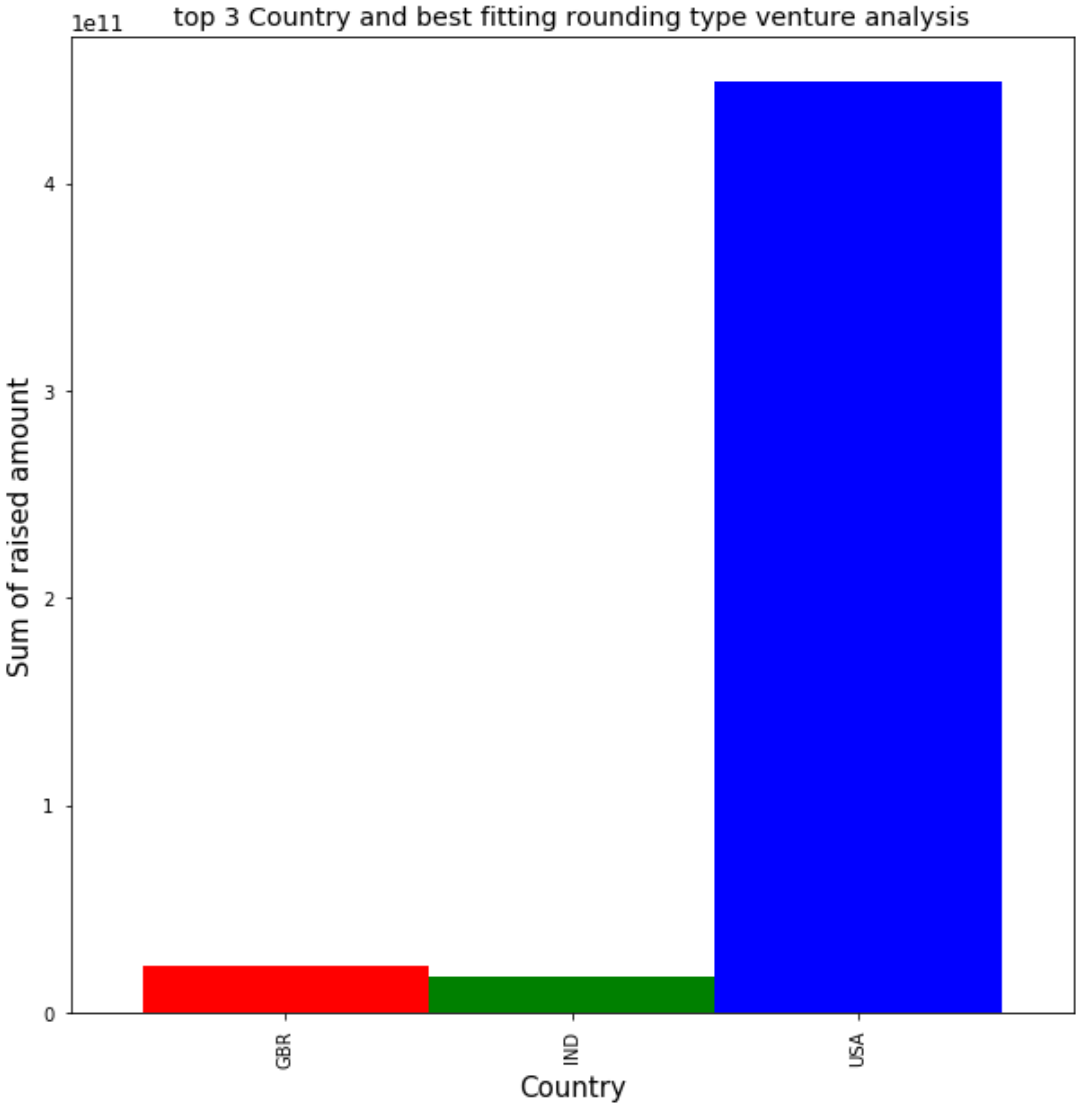
country_code	raised_amount_usd
USA	$549,568,807,428.47
GBR	$31,389,813,215.13
IND	$25,209,944,767.39
CAN	$14,388,313,008.34
SGP	$4,984,684,956.51
CHE	$4,479,831,378.53
IRL	$2,842,329,420.59
AUS	$2,418,962,158.57
ZAF	$1,284,545,840.53

For further analysis we will take top 3 countries with highest investment

Conclusion: Take top 3 countries with highest investment

1	Top English-speaking country	USA
2	Second English-speaking country	United Kingdom
3	Third English-speaking country	India

Now we are taking only 3 countries and invest type venture so we have 41667 row left in data to analyse.


top 3 Country and best fitting rounding type venture analysis
1e11
Sum of raised amount
0
1
2
3
4
GBR
IND
USA
Country

Checkpoint 4 & 5

Sector analysis 1 & 2

- We had sector data in mapping file
- First, we need to merge sector with the already merged master_frame dataset
- Below is the merged data frame lookup after adding primary sector in dataset

mpany_permalink	funding_round_type	raised_amount_usd	permalink	category_list	status	country_code	founded_at	primary_sector	main_sector
/organization/-fame	venture	1.000000e+07	/organization/-fame	Media	operating	IND	NaN	Entertainment	Entertainment
ganization/-qounter	venture	1.169127e+07	/organization/-qounter	Application Platforms|Real Time|Social Network...	operating	USA	04-09-2014	News, Search and Messaging	News, Search and Messaging,Social, Finance, An...
/organization/004-technologies	venture	1.169127e+07	/organization/004-technologies	Software	operating	USA	01-01-2010	Others	Others
organization/0xdata	venture	2.000000e+07	/organization/0xdata	Analytics	operating	USA	01-01-2011	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising
organization/0xdata	venture	1.700000e+06	/organization/0xdata	Analytics	operating	USA	01-01-2011	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising

Now we will divide our data in to 3 new country wise data frame.

```
Rows of D1_USA 38034 Columns of D1_USA 10 Total amount raised by USA= 444514477010.36646
Rows of D2_GBR 2265 Columns of D2_GBR 10 Total amount raised by GBR= 22279635651.778175
Rows of D3_IND 984 Columns of D3_IND 10 Total amount raised by IND= 17143373631.191803
```

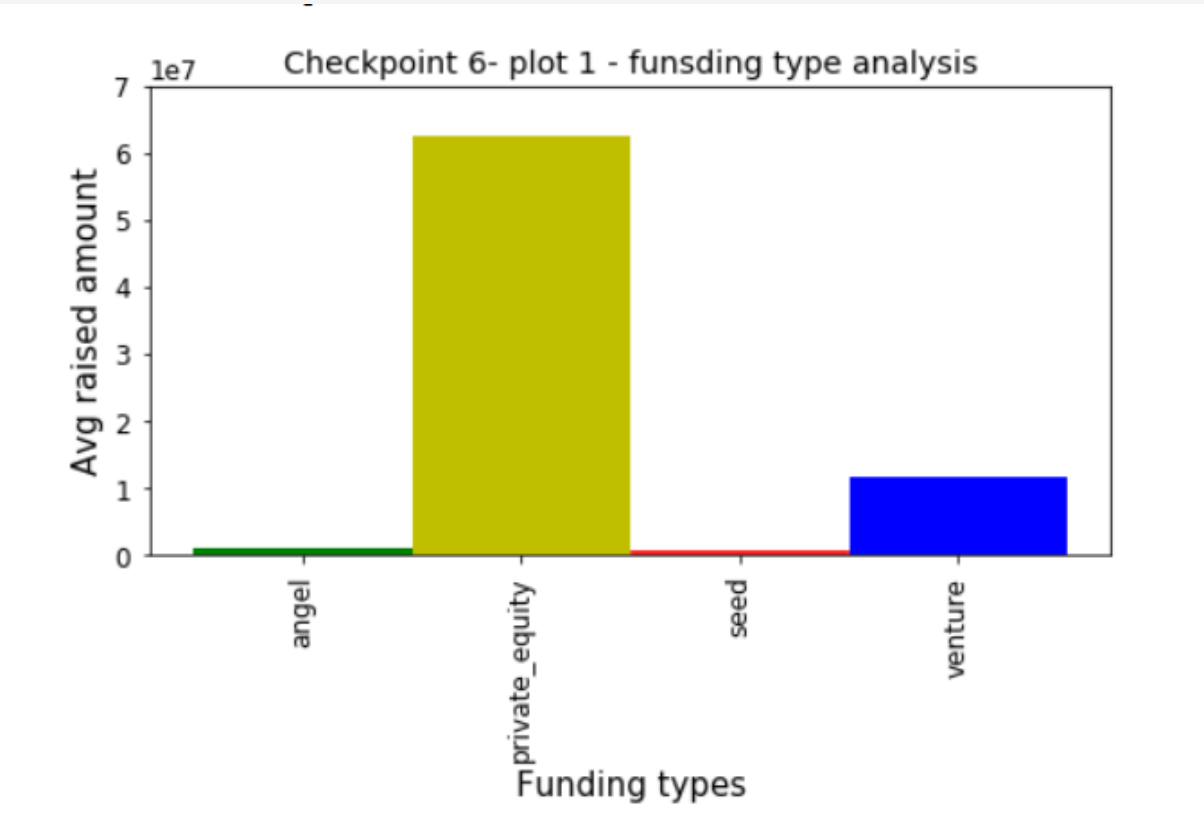
By analysing d1,d2,d3 We got following results.

Sr.no	Question	USA	United Kingdom	India
1	total no of investments	38034	2265	984
2	total amount of investment	$444,514,477,010.37	$22,279,635,651.78	$17,143,373,631.19
3	top sector(based on count of investments)	8693(others)	571(others)	329(others)
4	Second-best sector (based on count of investments)	8113(cleantech/semiconductors)	456(cleantech/semiconductors)	154(News, Search and Messaging)
5	Third-best sector (based on count of investments)	6916(Social, Finance, Analytics, Advertising)	405(Social, Finance, Analytics, Advertising)	130(Social, Finance, Analytics, Advertising)
6	Number of investments in the top sector (refer to point 3)	8693	571	329
7	Number of investments in the second-best sector (refer to point 4)	8113	456	154
8	Number of investments in the third-best sector (refer to point 5)	6916	405	130
9	For the top sector count-wise (point 3), which company received the highest investment?	sofi	one web	flipkart
10	For the second-best sector count-wise (point 4), which company received the highest	snapchat	farfetch	snapdeal

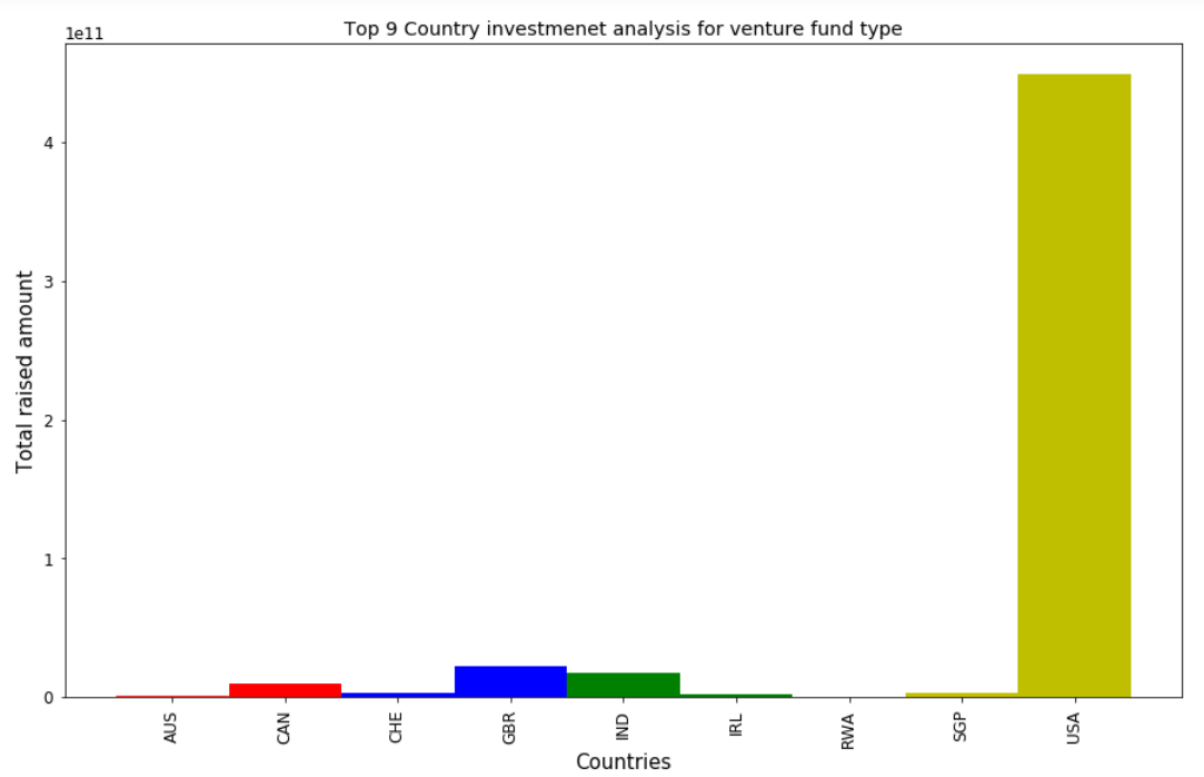
Checkpoint 6 Plots

1. A plot showing the fraction of total investments (globally) in venture, seed, and private equity, and the average amount of investment in each funding type. This chart should make it clear that a certain funding type (FT) is best suited for Spark Funds.

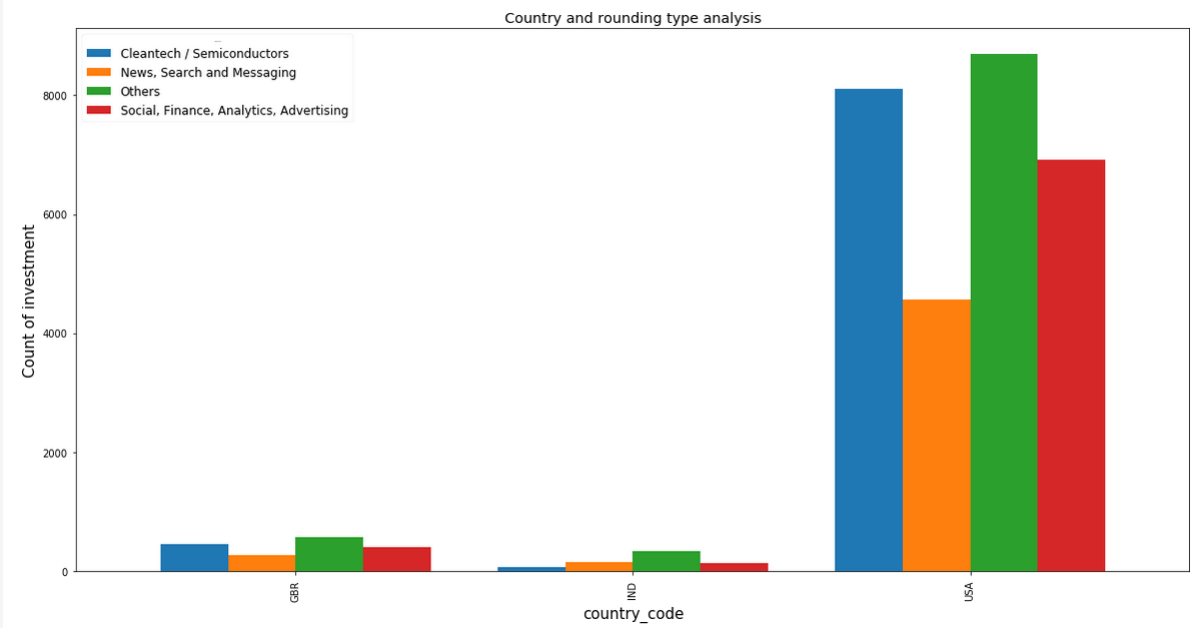
This is global chart but English-speaking countries.



2. A plot showing the top 9 countries against the total amount of investments of funding type FT. This should make the top 3 countries (Country 1, Country 2, and Country 3) very clear.



3. A plot showing the number of investments in the **top 3 sectors** of the **top 3 countries** on one chart (for the chosen investment type FT).



Conclusion:

- As per analysis and requirement
- Spark fund needs to look into following points before investment
    - Countries USA, GBR, IND need to be considered for investment
    - Fund type that falls under 5 to 15 million is venture, so venture should be considered for investing funds
    - Sectors which are mostly considered in investment are Others,Cleantech/Semiconductor, Social/Finanace/Analytics/Advertisinsg and News/Search and messaging.