SPARK 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).

Let's first merge Rounds2 and Companies – we get out master frame
to merge them we will use company_permalink and permalink for reference
merged data frame has `114942 rows`

By merging we get following columns

```
((114942, 16),
 Index(['company_permalink', 'funding_round_permalink', 'funding_round_type',
        'funding_round_code', 'funded_at', 'raised_amount_usd', 'permalink',
        'name', 'homepage_url', 'category_list', 'status', 'country_code',
        'state_code', 'region', 'city', 'founded_at'],
       dtype='object'))
```

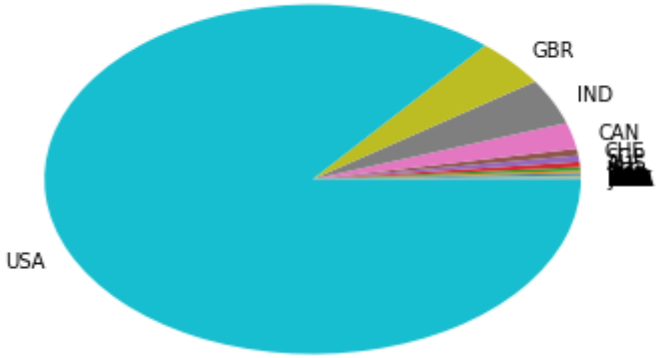
We would drop unnecessary columns from our master frames-

- We are considering country only for location so we can remove region, city, state_code
- We can also remove funding_round_code, funded_at, homepage_url
  Because as per our requirements these columns are not necessary

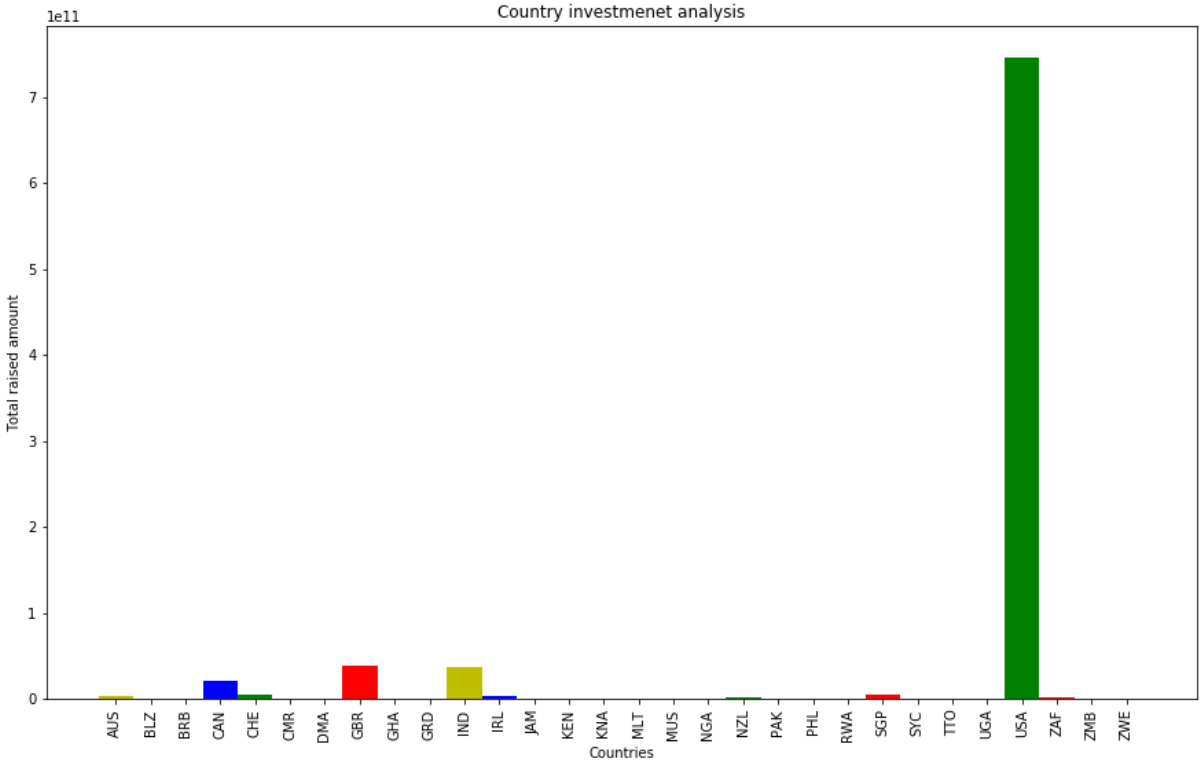
1. **Country analysis:**
   We will only consider English Speaking countries. Below is the list of 30 English speaking countries.

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

- Remove all non English speaking countries from data
- After removing countries our dataset is now has 87002 rows
- 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



-

For further analysis we will take top 10 countries with highest investment
After removing all other countries – we get data of rows – 86668

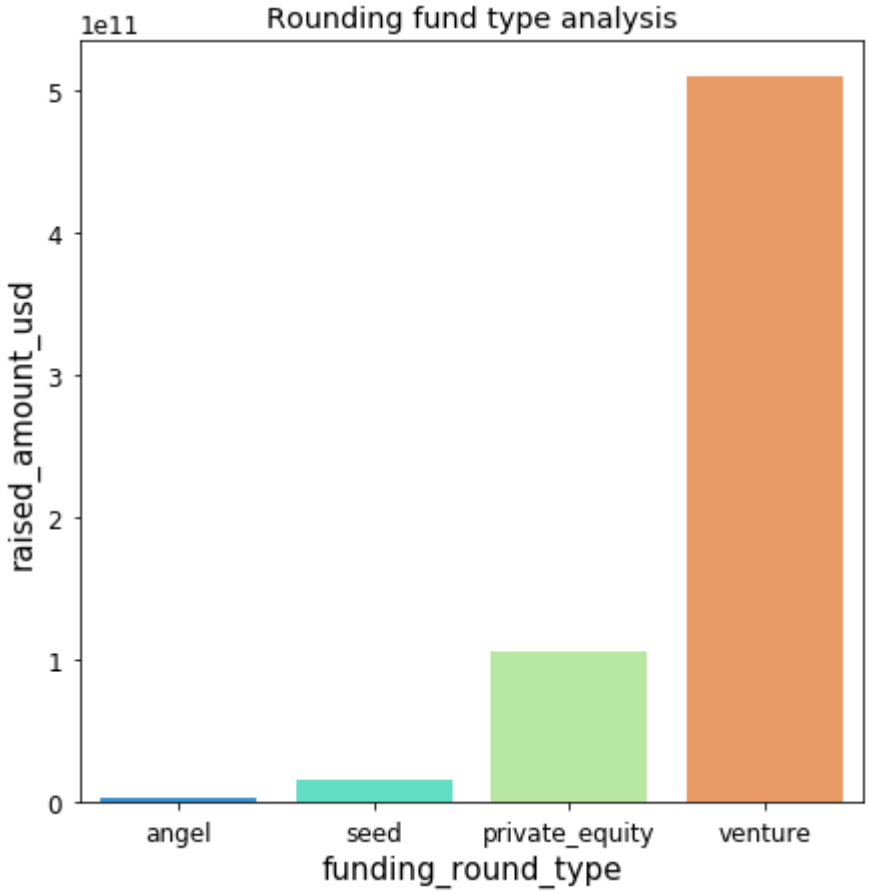
Conclusion: Take top 10 countries with highest investment

2. Funding round type investment

- As per our requirement we are going to consider only 4 types of types
- Angel seed private equity and ventures
- Remove all other types from the data, now we get data of rows 70078

	funding_round_type	raised_amount_usd
0	angel	3.330879e+09
1	seed	1.652551e+10
2	private_equity	1.069869e+11
3	venture	5.103796e+11

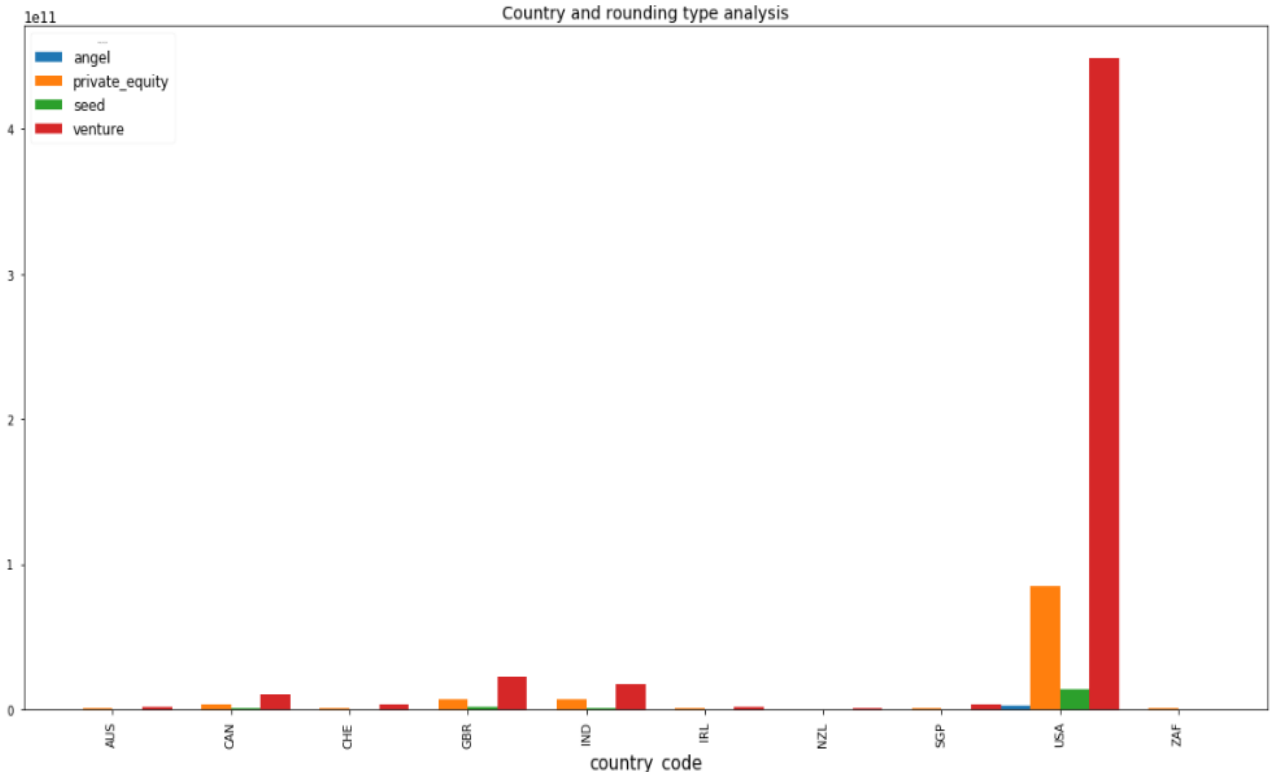
As per the table above venture is the most invested type.
Visually checking



As said, venture seems to have highest amount of investment across all English speaking countries

Let us check country and funding type together

- By looking at the graph we can see that **USA** invests the most in **venture** followed by **private equity**
- We can see that each country has a faire share in venture funding type



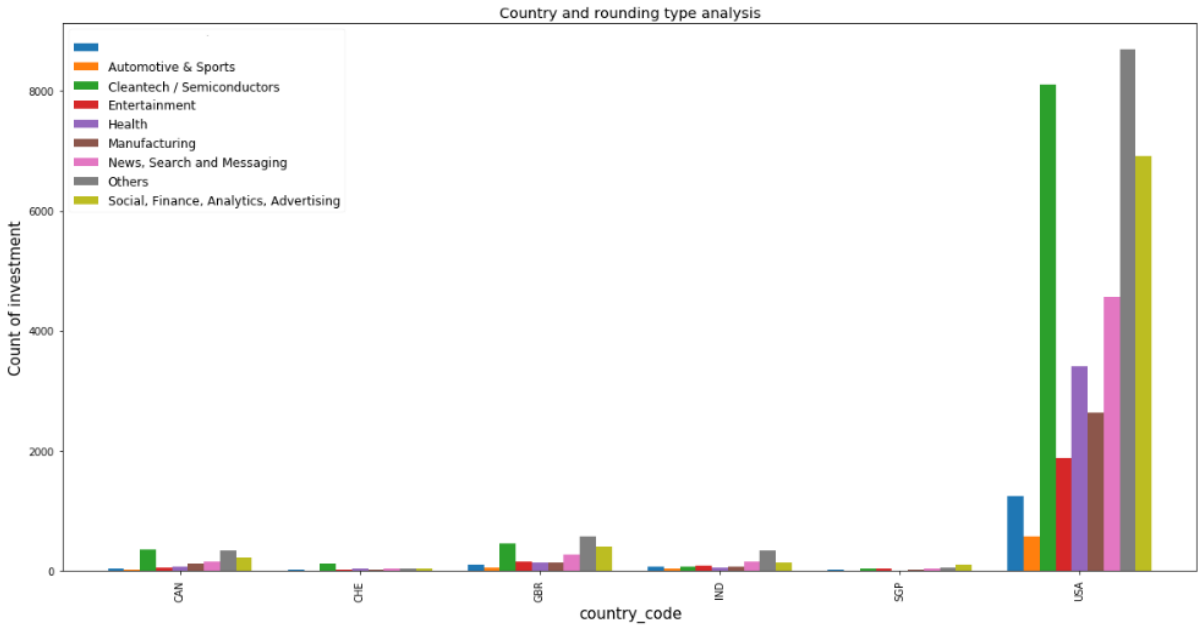
- For our constraint of 5 to 15 million of investment
- We will take mean of each type of fund investment

funding_round_type	raised_amount_usd
angel	854071.66
private_equity	62093388.60
seed	815028.04
venture	11552537.78

-
- So, we can see that **venture** falls under our favourable constraint – it has **less than 15million of investment** and is followed by many countries for investment option
- From graph we can also see that counties **except USA, GBR, IND, CAN, CHE, SGP** has least or no investment in venture, so we will remove those from our data
- After removing countries and all funding type except venture, now we are left with 43959 rows

3. Sector analysis

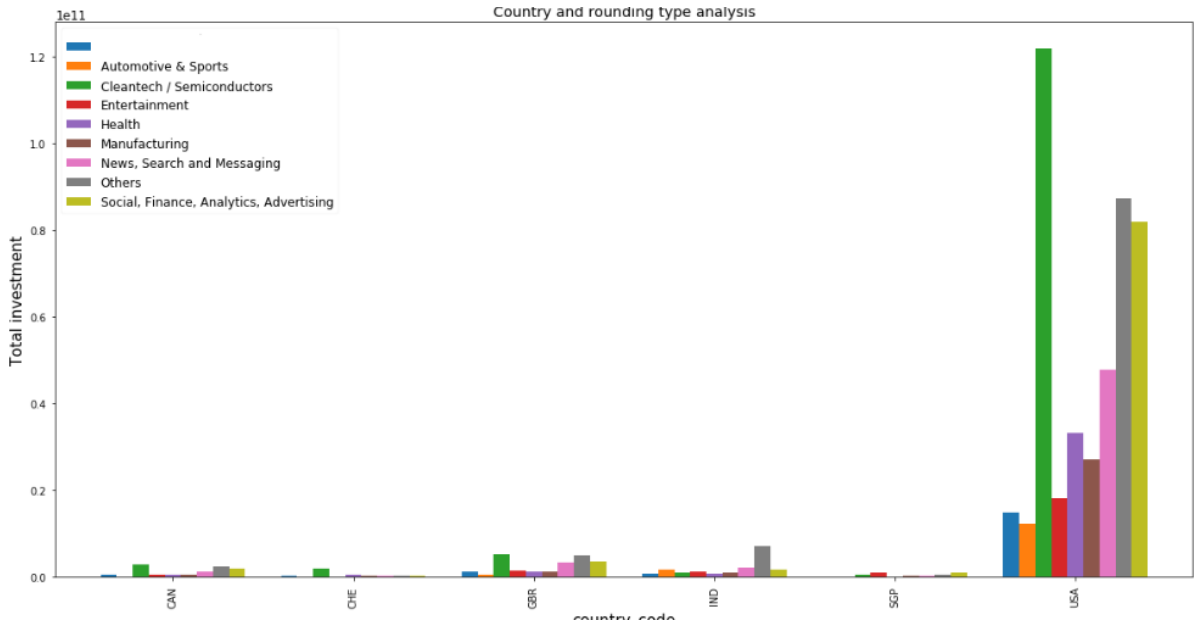
- We had sector data in mapping file
- First, we need to merge sector with the already merged file
- After merging we get following graph
-



- We can see that from all the 8 sectors
- USA invests more no of times in Others,Cleantech/Semiconductors,Social/Finance/Analytics/Advertising
- GBR invests mostly in Others,Cleantech/Semiconductors,Social/Analytics/Advertising are the most occurred investments
- IND invests mostly in Others and News
- CAN invest mostly in Cleantech/Semiconductor and Others are

The above visual was related to no of investment made

Now let us check **amount spent on each sector** by each **country**



From the graph it is seen that

- **USA** invests more amount in Cleantech/Semiconductors, Others,Social/Finance/Analytics/Advertising
- **United Kindom** invests most amount Others,Cleantech/Semiconductors,Social/Analytics/Advertising
- **India** invests most in Others
- **Canada** invests most amount in Cleantech/Semiconductor and Others

Conclusion:

- As per analysis and requirement
- Spark fund needs to look into following points before investment
    - **Countries USA, GBR, IND, CAN, CHE and SGP** 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.