

Spark Fund Investment Analysis

Surender Yadaram

Abstract

- Spark Funds, an asset management company wants to make investments in a few companies. The CEO of Spark Funds wants to understand the global trends in investments so that she can take the investment decisions effectively.
- Spark Funds has two minor constraints for investments:
 - It wants to invest between 5 to 15 million USD per round of investment
 - It wants to invest only in English-speaking countries because of the ease of communication with the companies it would invest in
 - For analysis, consider a country to be English speaking only if English is one of the official languages in that country
- Business objective: The objective is to identify the best sectors, countries, and a suitable investment type for making investments. The overall strategy is to invest where others are investing, implying that the 'best' sectors and countries are the ones 'where most investors are investing'.
- Goals of data analysis: The goals are divided into three sub-goals:
 - Investment type analysis: Compare the typical investment amounts in the venture, seed, angel, private equity etc. so that Spark Funds can choose the type that is best suited for their strategy.
 - Country analysis: Identify the countries which have been the most heavily invested in the past. These will be Spark Funds' favourites as well.
 - Sector analysis: Understand the distribution of investments across the eight main sectors.

Problem solving methodology

Steps followed for doing Investment type, Country and Sector analysis are listed below.

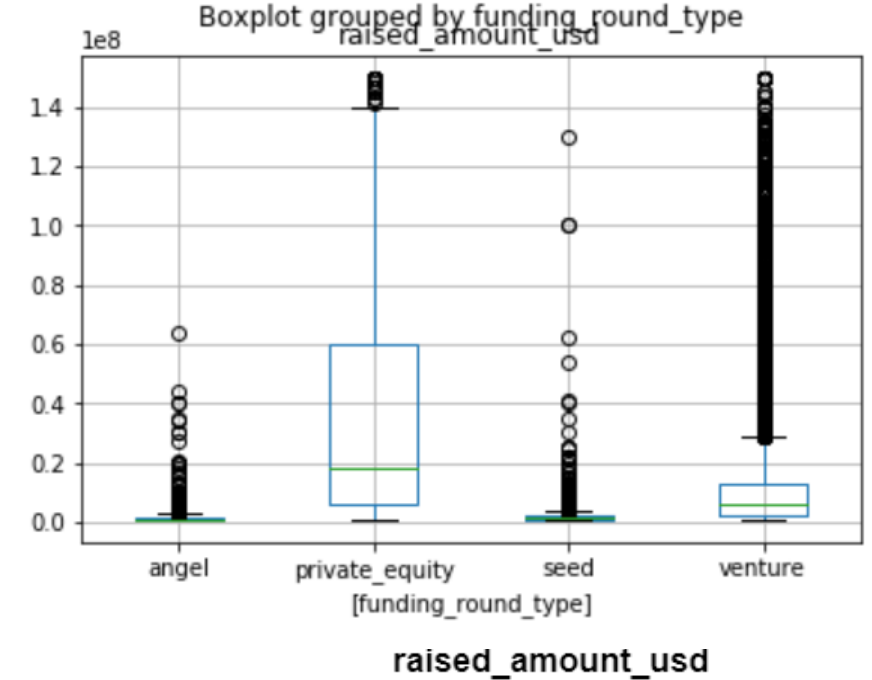
1. Data Cleaning

- Read the companies and rounds2 csv file using pandas library using “engine = python” option
- Use “permalink” column as the unique key for both companies and rounds2 dataframes
- Observe the data for null and junk values using the head() and tail() methods.
- Convert the “permalink” in companies and “company_permalink” to lower case for case consistency
- As data as non ascii characters, use encode() method for encoding the “permalink” as “permalink_encoded” column using “unicode_escape”.
- Find the unique companies using “permalink_encoded” column for both the data frames using unique()
- Merge the data frames rounds2 and companies using left join on the “permalink_encoded” column.
- Filter the data for funding types "seed", "angel", "venture“ and "private_equity“ as Spark Funds are interested only in these funding types.
- As Spark Funds interested in funding 5 to 15 million USD, remove outliers having funding value less than 500k and 150 Million
- Remove the rows having null values in "raised_amount_usd" column in the master_frame dataframe.

Analysis

2. Funding Type Analysis.

- Once the data clean up is done for "raised_amount_usd" column by removing nulls, perform the summary statistics using describe()
- The most representative value of the investment amount is found to be Average.
- Data imputing is considered as we have almost 80% of data for analysis after data clean up



	count	mean	std	min	25%	50%	75%	max
funding_round_type								
angel	2229.0	1.646911e+06	3.006388e+06	500000.0	653426.0	1000000.0	1633097.00	63590263.0
private_equity	1610.0	3.598456e+07	3.902047e+07	500000.0	5968334.5	18382000.0	59755068.25	150000000.0
seed	9673.0	1.546603e+06	2.595150e+06	500000.0	747936.0	1100000.0	1880000.00	130000000.0
venture	45216.0	1.094656e+07	1.524231e+07	500000.0	2400000.0	5817119.0	13000000.00	150000000.0

The suitable funding type is “venture” type funding

Analysis

3. Country Type Analysis

- Remove the rows having null values for “country_code” in the master_frame
- Filter the data for the funding type “venture”
- Group by the resultant dataframe by the “country_code” and do the sum of the “raised_amount_usd”.
- Sort the total amount in the decending order to get the top 9 countries based on the total investment.

	index	country_code	raised_amount_usd
0	91	USA	3.569427e+11
1	15	CHN	2.590326e+10
2	28	GBR	1.829193e+10
3	37	IND	1.079229e+10
4	12	CAN	9.386889e+09
5	40	ISR	6.903039e+09
6	27	FRA	6.524318e+09
7	20	DEU	6.172076e+09
8	43	JPN	2.888074e+09

The top three English-speaking countries that received highest funding are USA, UK and India

Analysis

4. Sector Analysis 1

- Remove the rows having null values for “category_list” column in master_frame
- Extract the “primary_sector” from “category_list” column in master_frame
- Read the mapping csv file and remove null values for “category_list” using pandas.
- Correct the spelling mistake of 0 instead of ‘na’ for “category_list” in mapping data frame and convert it to lower case for case consistency
- Write a method to map the “primary_sector” in master_frame to “category_list” in mapping data frame by creating a main_sector column.
- Remove the rows in the master_frame for which primary sector is not mapped to main sector

5. Sector Analysis 2

To create three separate data frames D1, D2 and D3 for each of the three countries containing the observations of funding type FT falling within the 5-15 million USD range, follow below steps

- Filter the master_frame for the country, funding type venture and raised amount in 5 to 15 million category
- Find the total count of investment and total amount for each country
- Group the country specific investments by main_sector and count the investments by sorting in descending order to get the top 3 sectors.
- Get the company received highest funding in the top 2 sectors for the top 3 English speaking countries

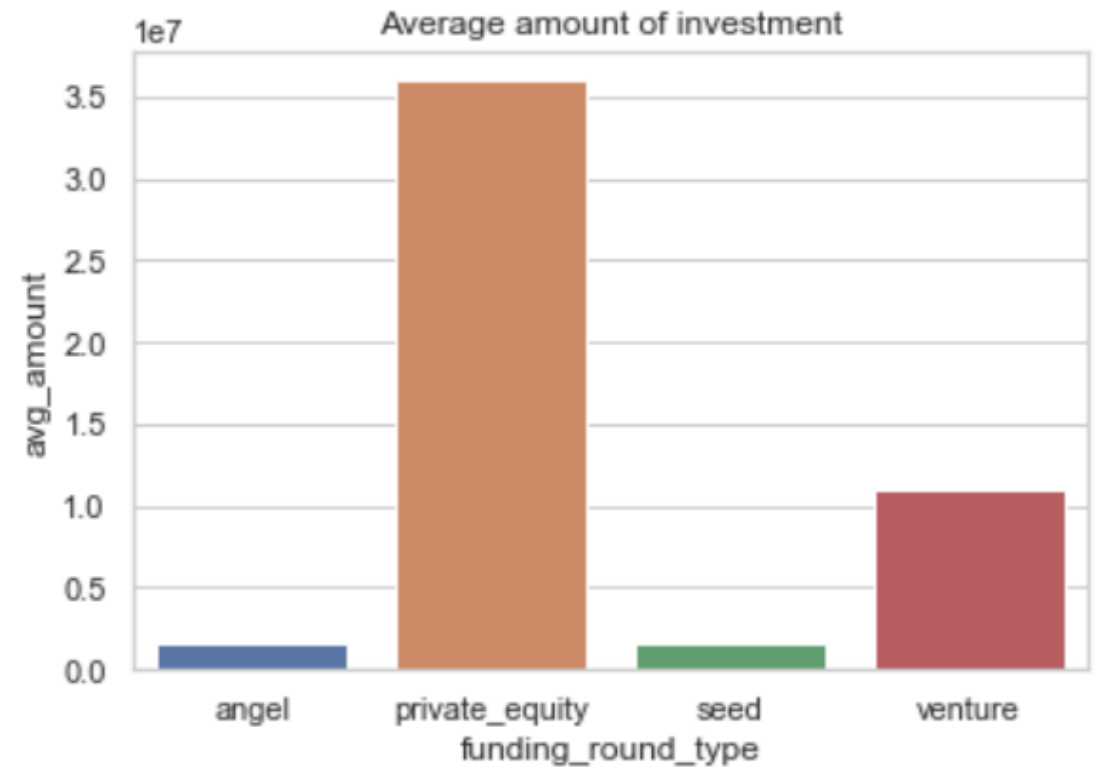
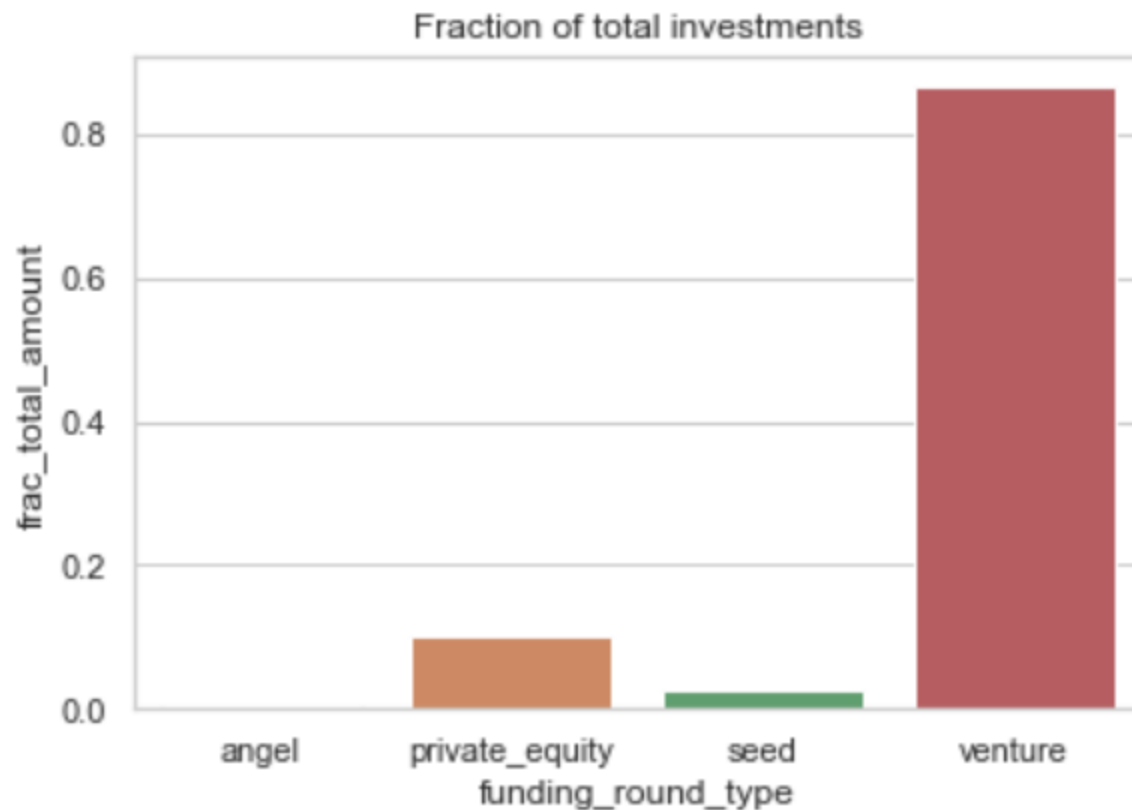
Analysis

5. Sector Analysis 2

Sector-wise Investment Analysis				
Sl.no	Questions	C1	C2	C3
1	Total number of Investments (count)	12063	621	328
2	Total amount of investment (USD)	1.07757E+11	5379078691	2949543602
3	Top Sector name (no. of investment-wise)	Others	Others	Others
4	Second Sector name (no. of investment-wise)	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising
5	Third Sector name (no. of investment-wise)	Cleantech / Semiconductors	Cleantech / Semiconductors	News, Search and Messaging
6	Number of investments in top sector (3)	2950	147	110
7	Number of investments in second sector (4)	2714	133	60
8	Number of investments in third sector (5)	2350	130	52
9	For point 3 (top sector count-wise), which company received the highest investment?	Virtustream	Electric Cloud	FirstCry.com
10	For point 4 (second best sector count-wise), which company received the highest investment?	SST Inc. (Formerly ShotSpotter)	Celltick Technologies	Manthan Systems

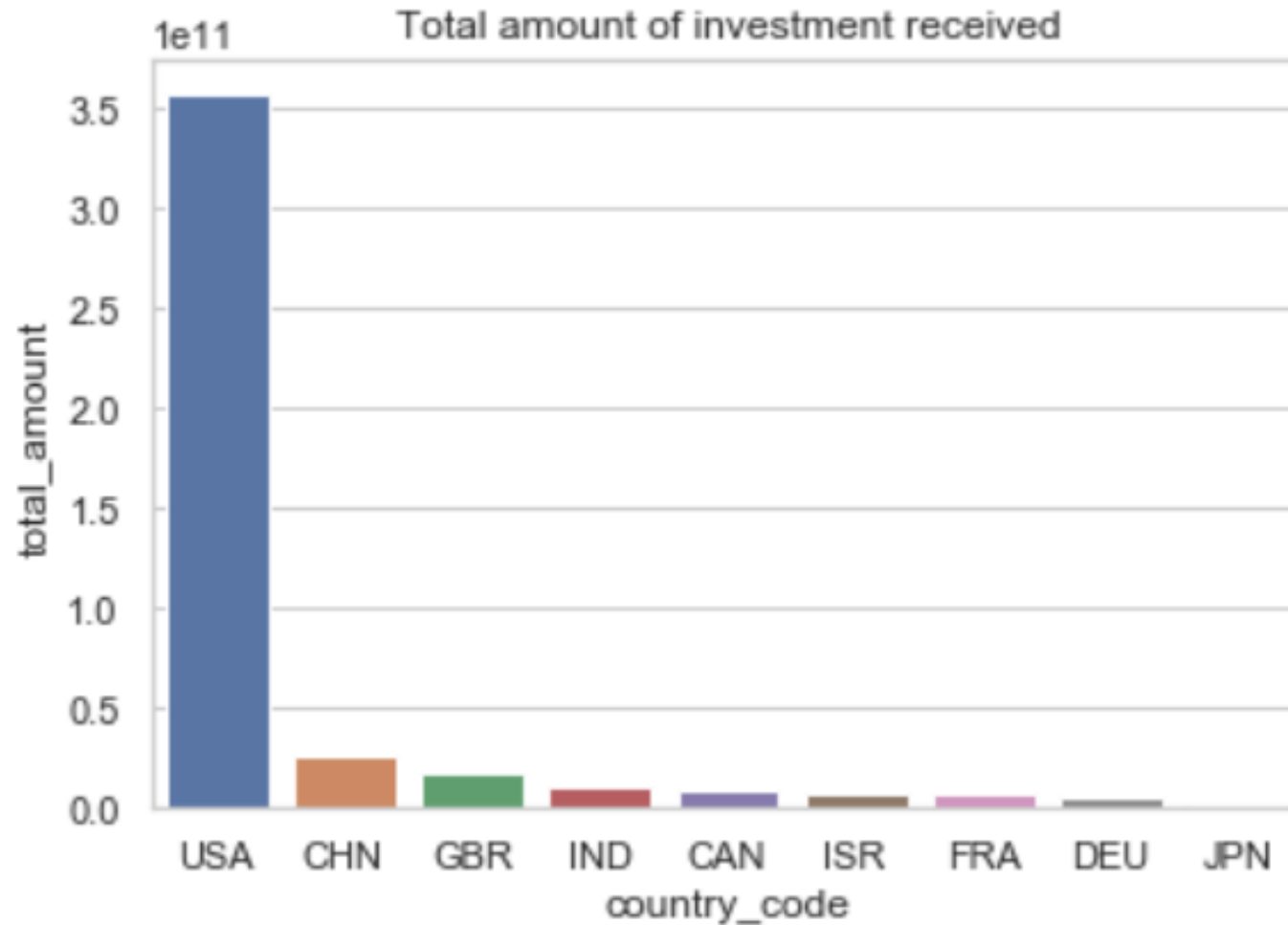
6. Plot Results

Plot 1. A plot showing the fraction of total investments (globally) in angel, venture, seed, and private equity, and the average amount of investment in each funding type.



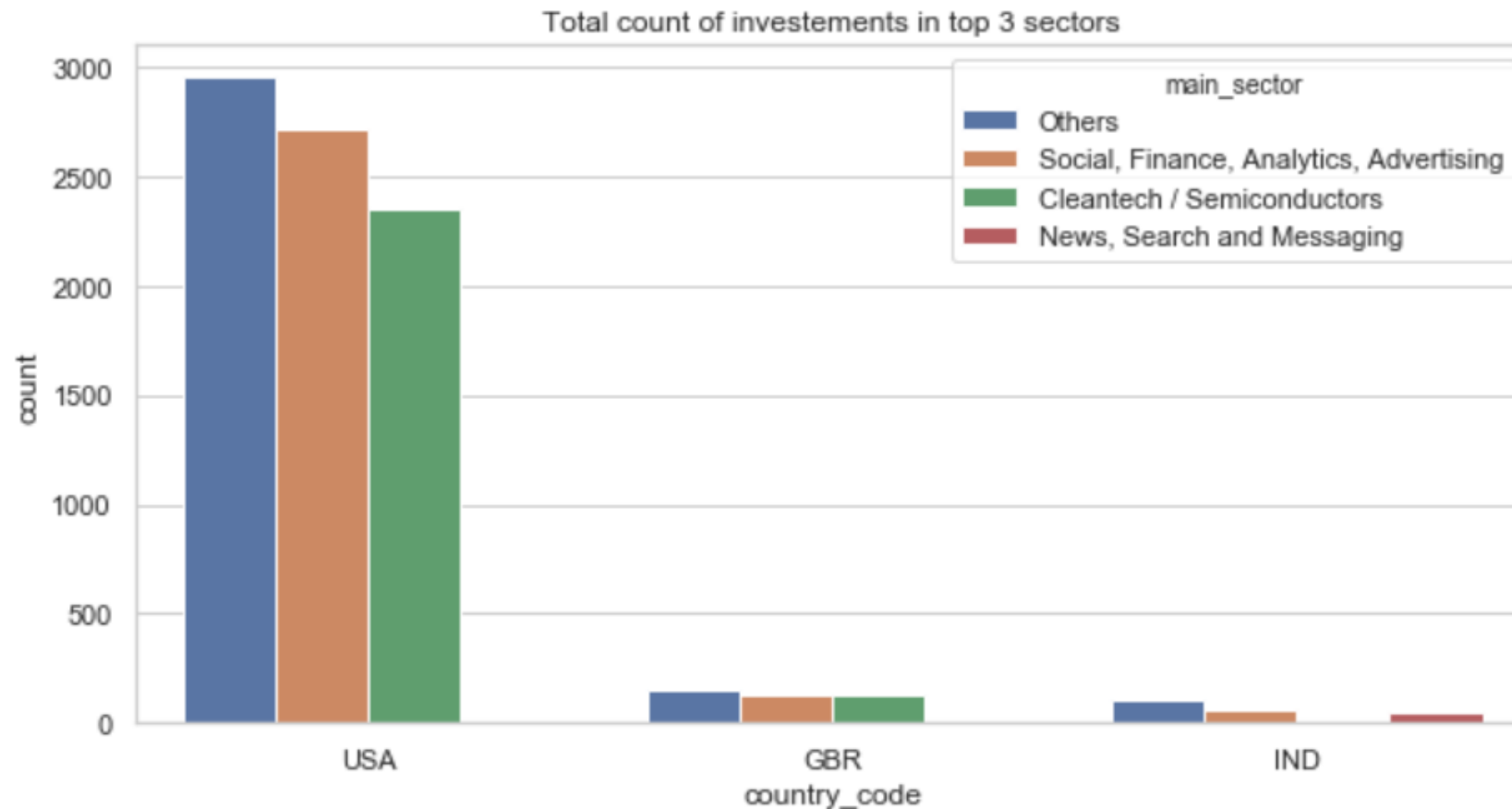
6. Plot Results

Plot 2. A plot showing the top 9 countries against the total amount of investments of funding type FT(venture type funding).



6. Plot Results

Plot 3. A plot showing the number of investments in the top 3 sectors of the top 3 countries on one chart (for the chosen venture type funding).



Conclusions

- ✓ Data cleanup is done the given data for doing the Investment analysis
- ✓ The most representative value of the investment amount is found to be Average in funding type analysis.
- ✓ The suitable funding type for SparkFunds in 5 to 15 million category is “venture” type funding
- ✓ The top three English-speaking countries that received highest venture funding are USA, UK and India
- ✓ The top3 sectors in USA and UK are Others, Social, Finance, Analytics, Advertising and Cleantech / Semiconductors
- ✓ The top3 sectors in India are Others, Social, Finance, Analytics, Advertising and News, Search and Messaging