INVESTMENT CASE STUDY SUBMISSION

Agenda

- Abstract and Business Objective
- How we solved the problem
- High level analysis
- Details analysis
  - ✓ Funding Type
  - ✓ Country
  - ✓ Sector
- Plot 1
  - ✓ Fraction of total investments(globally) and average amount of investment
- Plot 2
  - ✓ Top 9 countries against the total amount of investments of specific funding type
- Plot 3
  - ✓ Top 3 sectors each in specific country
- Conclusion

Abstract

XYZ, an asset management company wants to make investments in a few companies. The CEO of XYZ wants to understand the global trends in investments so that investment decisions can be made effectively.

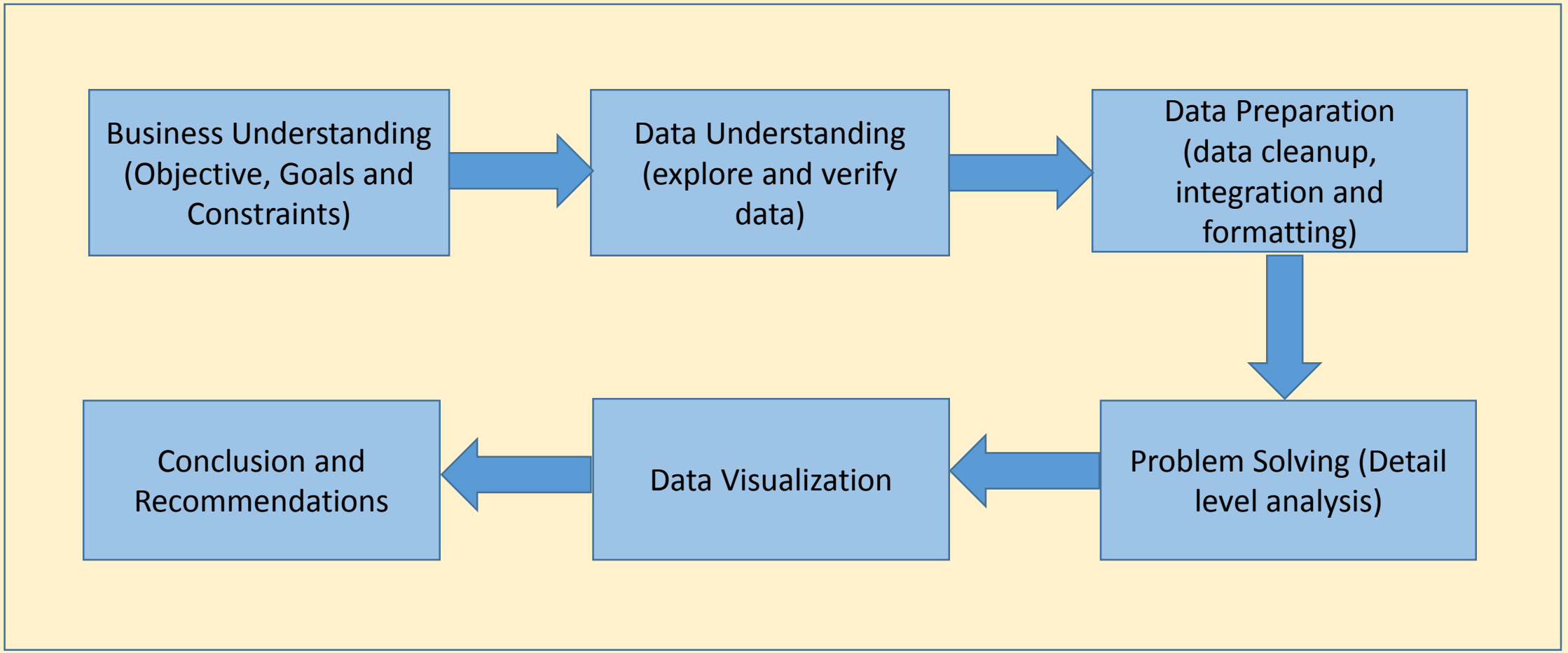
Business Objective

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 investments are happening.

Constraints

Investment amount should be between 5 to 15 million USD per round of investment
Investment should only be done in English-speaking countries because of the ease of communication with the companies
XYZ wants to invest where most other investors are investing.
This pattern is often observed among early stage startup investors.

How we solved the problem



High Level Analysis

- Data is given in form of three files
  - A dimension table which contains details about companies – A couple of important attributes worth noting are:
    - Permalink : This is unique id of the company
    - name: This is the name of the Company
    - category_list : Category/categories to which a company belongs
    - country_code : Country code
  - A fact table which contains details about investments, companies where investment is made, amount of investment etc. – A couple of important attributes worth noting are:
    - company_permalink : This is unique id of the company where investment is made
    - raised_amount_usd: Investment made
  - A mapping file which contains details about different categories and associated categories

Assumptions:

- For our analysis, we considered a country to be English speaking only if English is one of the official languages in that country
- We considered a predefined list of countries where English is an official language and considered as part of this analysis
- In the companies data file, there is a variable named as 'category_list'; some of the observations for this variable contains pipe ('|') character. It is determined that left side of this pipe character is nothing but the primary sector

Detail Analysis

- File that contains the companies details are loaded into a data frame viz. ‘companies’
- File that contains the investment details are loaded into a data frame viz. ‘rounds2’
- Next it is a good practice to look out the structure of data to find out for a unique key
  - It is found that in ‘companies’ data frame, 'permalink' is the name which can be used as unique key for the company
  - It is also found that there are no companies in ‘rounds2’ data frame which are not available in companies table
    - ✓ Here we can consider that ‘company_permalink’ in rounds2 refers to the unique key ‘permalink’ of companies table
- File that contains the mapping between categories and main sectors are also loaded in a data frame viz. ‘mapping’
- To meet the objective, this analysis is broken down into several low level analysis
  - ✓ Investment type analysis
  - ✓ Country analysis
  - ✓ Sector analysis

Detail Analysis – Funding Type

- Since we have companies details in one data frame and investment details in another data frame, we chose to merge the data frames to proceed with the further analysis
  - ✓ It is observed that the total observations after merging the two data frames remain the same as total observations in 'rounds2' table
  - ✓ It again turned that each investment has a company associated with it
- From the merged data frame named as 'master_frame', based on the average of investment for each funding type, it was found that 'venture' type is the one who meets the criteria that says "Investment amount should be between 5 to 15 million USD per round of investment"
- We then proceeded considering 'venture' funding type is the one XYZ should look for

Detail Analysis – Country

- Once we found that 'venture' funding type is the one XYZ should concentrate, it is time to find the countries where most investments are happening for 'venture' funding type. This is part of the objective which says "The overall strategy is to invest where others are investing"
- We grouped the countries based on the total investments in each country. There is a constraint that country where XYZ will be investing should be a country where English is official language so that communication can be easy.
  - ✓ It is observed that there is a country which name is empty and is part of top countries where more investments are occurring. We are ignoring this country since there is no way XYZ can invest in country without name
  - ✓ It is also observed that China is part of top countries where more investments are occurring but upon doing a manual check with our predefined country list, we found that in China, English is not official language and hence we ignored this country too
  - ✓ Now we have all the countries in decreasing order of the investments
- We are going to suggest XYZ the top 3 countries where more investments are happening and country has English as an official language. These appear to be "USA", "GBR" and "IND"

Detail Analysis – Sector

- Once we have the analysis on the funding type and the top 3 countries, next task is to analyze the top sectors for those 3 countries
- When we look into the 'mapping' data frame, we could easily see that the data is available in wide format. As part of data manipulation we need to convert it into long format
  - ✓ Once we convert the 'mapping' data frame into long format, we will have each category list mapped to a sector.
  - ✓ We observed that one sector named as 'blanks' is also available. This should be considered as part of data manipulation and should not be ignored. It may be of no consequences at this time but retaining may help in further analysis
  - ✓ We also observed that there is a sector named as 'Others'. This sector too should not be avoided as it refers to a particular sector
- As assumed, that in companies data set anything that appears before pipe character ('|') is the name of primary sector, we have split this variable into two separate variables with 'primary sector' and 'sub-sectors' in 'master_frame'. Since '|' is a special character, we need to escape this using "\\"
- Our 'master_frame' data frame currently contains the 'category_list'(renamed as 'primary sector') as one variable, and at the same time the data frame where we have loaded the mapping details(after converting from wide to long) also contains the 'category_list' as one variable
- To do a proper and deep analysis on the sectors, we have merged the mapping file with the 'master_frame' data frame on the basis of primary sector of 'master_frame' data frame and category list of mapping data frame. We named this final data frame as 'fact_table' data frame. We did a inner merge here, which implies that observations in this 'fact_table' will be lesser than 'master_frame' since there are some observations for which there are no match
- We are now interested in finding the top three sectors in each of the three country. So we have made three data frames as 'D1' for USD, 'D2' for GBR and 'D3' for IND

Detail Analysis – Sector contd.

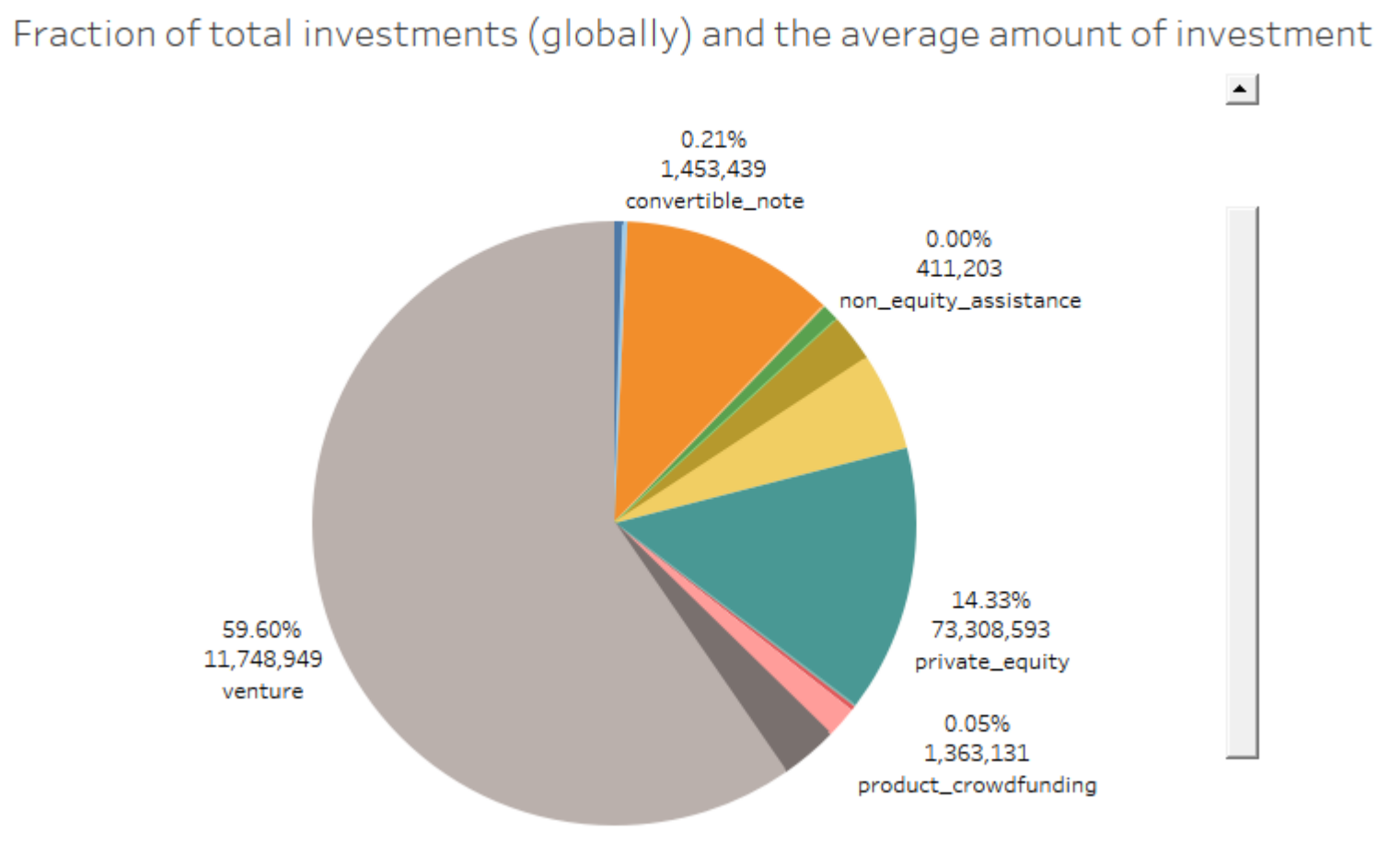
- 'D1', 'D2' and 'D3' data frames are built with below criteria
  - ✓ 'D1' where country code is USD, investment between 5 million and 15 million and investment type is 'venture'
  - ✓ 'D2' where country code is GBR, investment between 5 million and 15 million and investment type is 'venture'
  - ✓ 'D3' where country code is IND, investment between 5 million and 15 million and investment type is 'venture'
- For each of the country, the top 3 sectors are found by aggregating the respective data frame based on main sector against number of investments
- Using the above three data frames, we also calculated the total number and amount of investments in each main sector
- Once we have top 3 sectors for each of the country, we have also drilled down and found out the company where amount of investment is maximum within each sector

NOTE:

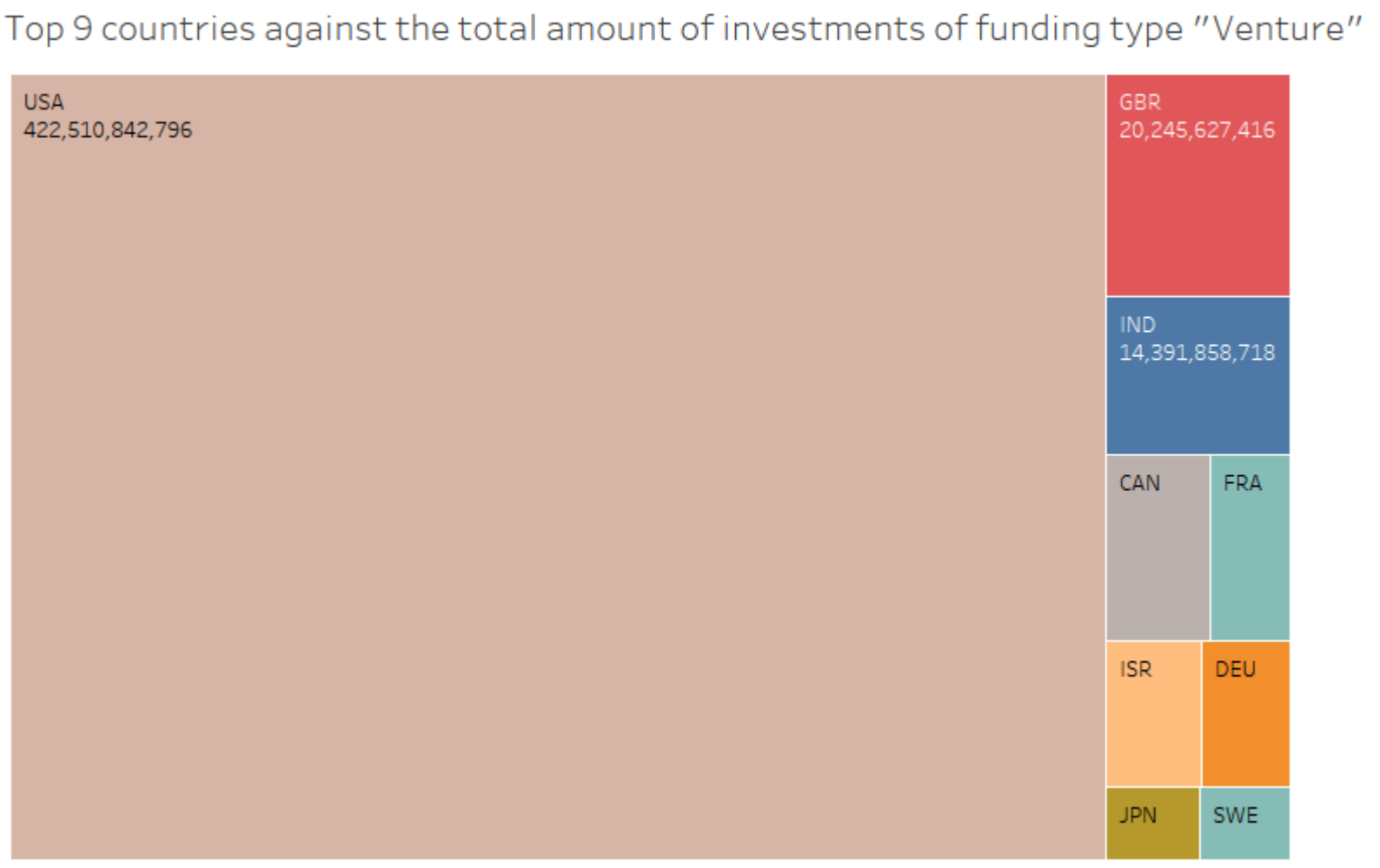
To plot our findings in tableau, we have exported the data from R solution to csv as two files:

1. First file which contains merged details between companies and rounds2 dataset. This will help us in plotting data for "Fraction of total investments (globally) and the average amount of investment" and "Top 9 countries against the total amount of investments of funding type "Venture""
2. Second file which contains the above file but also merged with mapping file(after conversion from wide to long). This will help us in finding " Top 3 sectors in the top 3 countries"

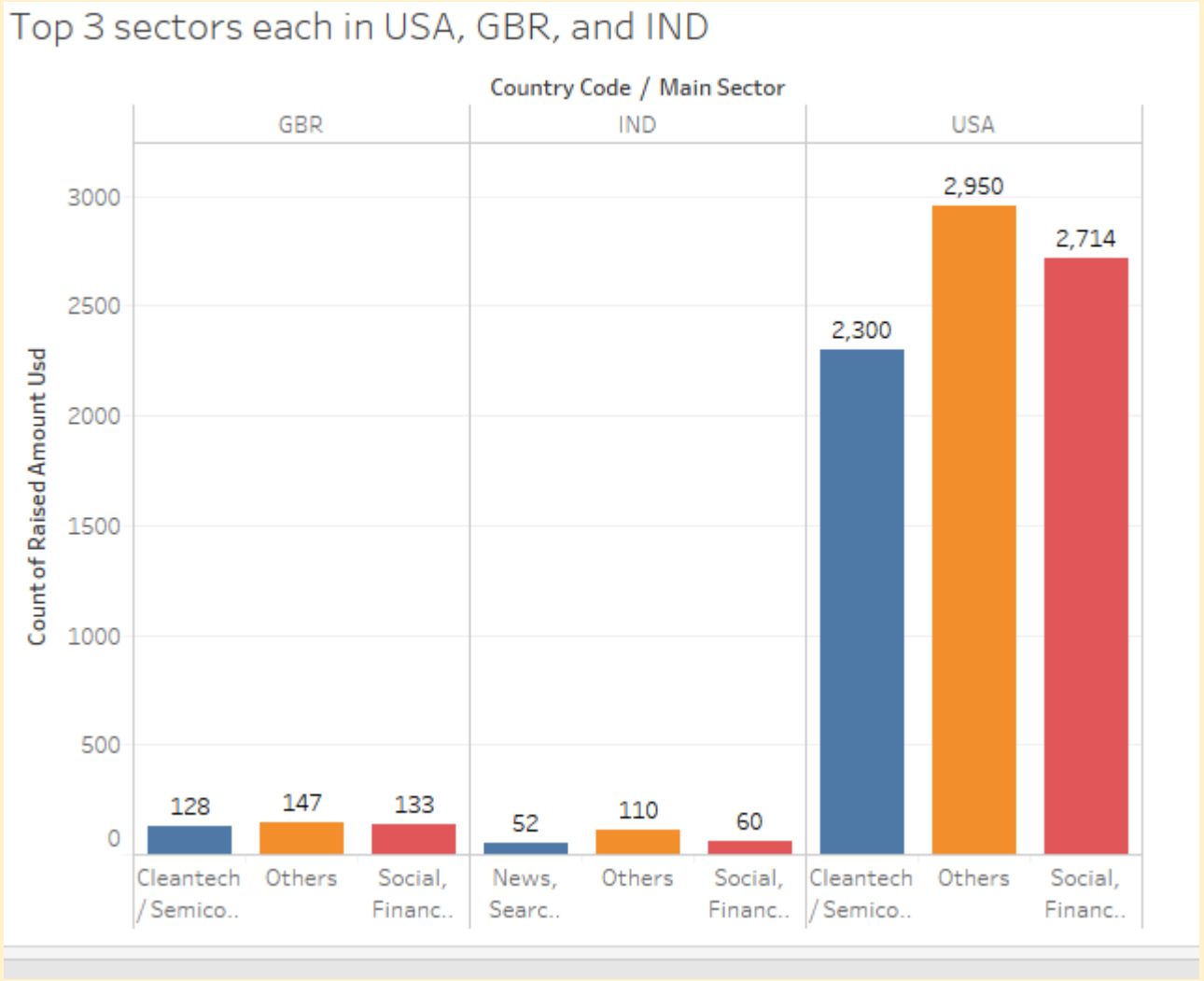
Plot 1


Fraction of total investments (globally) and the average amount of investment
0.21%
1,453,439
convertible_note
0.00%
411,203
non_equity_assistance
14.33%
73,308,593
private_equity
0.05%
1,363,131
product_crowdfunding
59.60%
11,748,949
venture

Plot 2


Top 9 countries against the total amount of investments of funding type "Venture"
USA
422,510,842,796
GBR
20,245,627,416
IND
14,391,858,718
CAN
FRA
ISR
DEU
JPN
SWE

Plot 3


Top 3 sectors each in USA, GBR, and IND
Country Code / Main Sector
GBR
IND
USA
Count of Raised Amount Usd
3000
2500
2000
1500
1000
500
0
128
147
133
52
110
60
2,300
2,950
2,714
Cleantech / Semico..
Others
Social, Financ..
News, Searc..
Others
Social, Financ..
Cleantech / Semico..
Others
Social, Financ..

Conclusions

1. XYZ should investment in the investment type “Venture”

2. XYZ should invest in following three countries
    1. USA – United States of America
    2. GBR – Great Britain
    3. IND – India

3. XYZ should target the following main sectors in the order specified
    1. Others
    2. Social, Finance, Analytics, Advertising
    3. Cleantech / Semiconductors
    4. News, Search and Messaging