INVESTMENT ASSIGNMENT

# SUBMISSION

Name: Divyansh Sharma

<Abstract>

The Goal of this analysis is to find the best investment verticals which can profit Spark Funds against the investments made in startup's. Our main focus points during this analysis will be the best suited Investment rounds, English speaking countries and the sectors which are experiencing a boom in investments. This research is based on the idea that ' ***If there is a boom in a sector, all the companies of that sector are profited from it*** '.

We have followed Cross-industry standard process for data mining (CRISP-DM) framework for the analysis which is a standard process followed for data mining by experts. Additionally we explored the companies which received the highest funding in the funding round which would suit best for the investment range defined in the problem statement.

During the analysis we have made reasonable assumptions which would help us ignore the data quality issues or defects that exist in the original data. These assumptions are the based on some prior analysis done to assimilate the anomalies present in data.

It is hoped, this investment will benefit Spark Funds Investment Company in its goal to maximize the profit from investments made in startup's in English speaking countries across all the countries. Our suggestions based on the analysis would help the organization grow.

<Analysis>
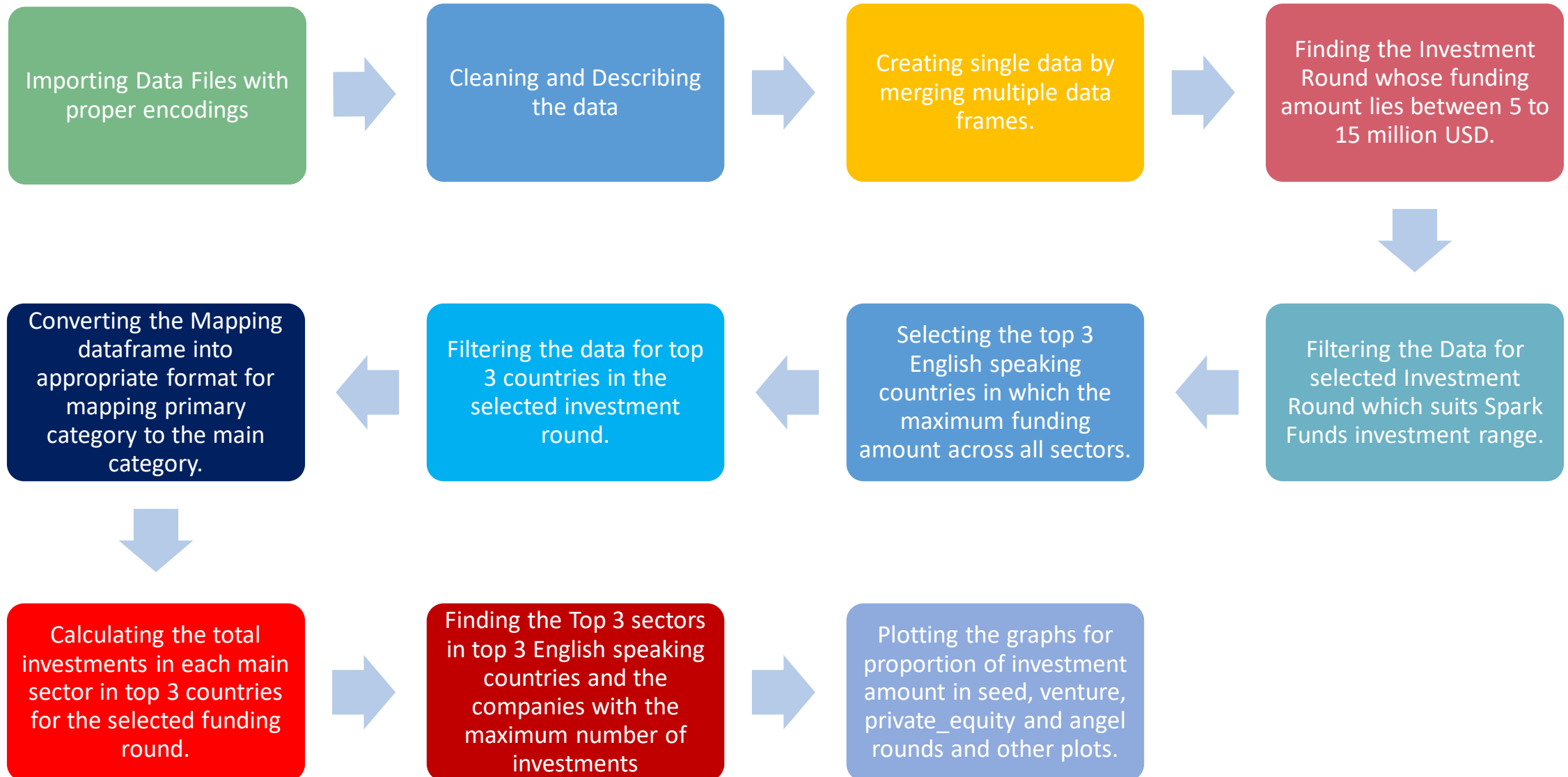
We Received 3 Data Files to perform this analysis and suggest the best verticals for Spark Funds to Invest :

1) **companies.txt** : This file consists of the data of companies like their name, location of registration, company permalink, date of establishment, sectors in which the company operates etc…

2) **Rounds2.csv** : This file consisted of the details of rounds in which the companies registered and the funding details they received in that round.

3) **Mapping.csv :** This file consists of a mapping list of 600+ categories to 8 main sectors.

During initial analysis we came across a few data quality issues, which are as follows :

• Both the files rounds2.csv and companies.csv contains lot of Unicode characters. These characters were then decoded into ascii format for processing

• Missing values were replaced by 0 for raised_funding_amount variable as there can be companies which did not received any funding in the round they registered for. Also, Missing values for the columns with 'character' values are replaced by blanks ('') as blanks are also used in mapping file.

• Many columns which were not effective in terms of analysis were directly dropped to reduce the data width.

• A Major data Quality Issue was analysed in the mapping file, the names of categories were misspelled as in many instances 'Na / na' was replaced by '0' . We corrected this anomaly before performing further analysis.

<Analysis>

The first analysis was performed to find the best suited investment round for spark funds. During this analysis, the total invested amount and the average funding amount in each funding round was considered to decide the best suited funding round :

- As the range of investment which spark funds want to invest is 5 to 15 million USD.

- There were 4 main Investment rounds which were considered, these are : Venture, Private Equity, Seed and Angel.

- Thus, based on the analysis **Venture** investment round came out to be the best suited round on both the selection criteria's.

- The maximum amount is invested in venture round and the investment range is also suited for spark funds.

In the Second Analysis the top 3 English speaking countries which are the best suited for spark funds were recognized :

- Firstly the records for venture funding round were filtered and the top 9 countries are extracted from the data.

- From the top 9 countries the top 3 English speaking countries are extracted based on the fact that English is their primary language.

- On performing these steps, we found the top 3 countries where the maximum investments are made.

- Based on the analysis, we found that following countries received the highest funding amount across all countries : **USA , GBR , IND** .

- For selecting the Top 3 countries we assumed that the top countries are those which have received the highest total funding across all the countries and companies.

\<Analysis\>

In the third Analysis, following steps were followed :

1) Data for top 3 countries was filtered and the primary category was extracted from the category list based on the assumption that the first category in the list must be considered as primary category.

2) Then this category was mapped to 8 main categories.

3) Further, the top 3 categories for each country were recognized based on the total number investments in each round.

These are the outcomes from this section of analysis :

1) 'Others' category is the top sector in all the 3 countries.

2) Cleantech / Semiconductors is the second most funded main sector in USA and GBR.

3) Social, Analytics and Advertising is the 3rd main sector in USA and GBR while in IND it is 2nd most funded sector.

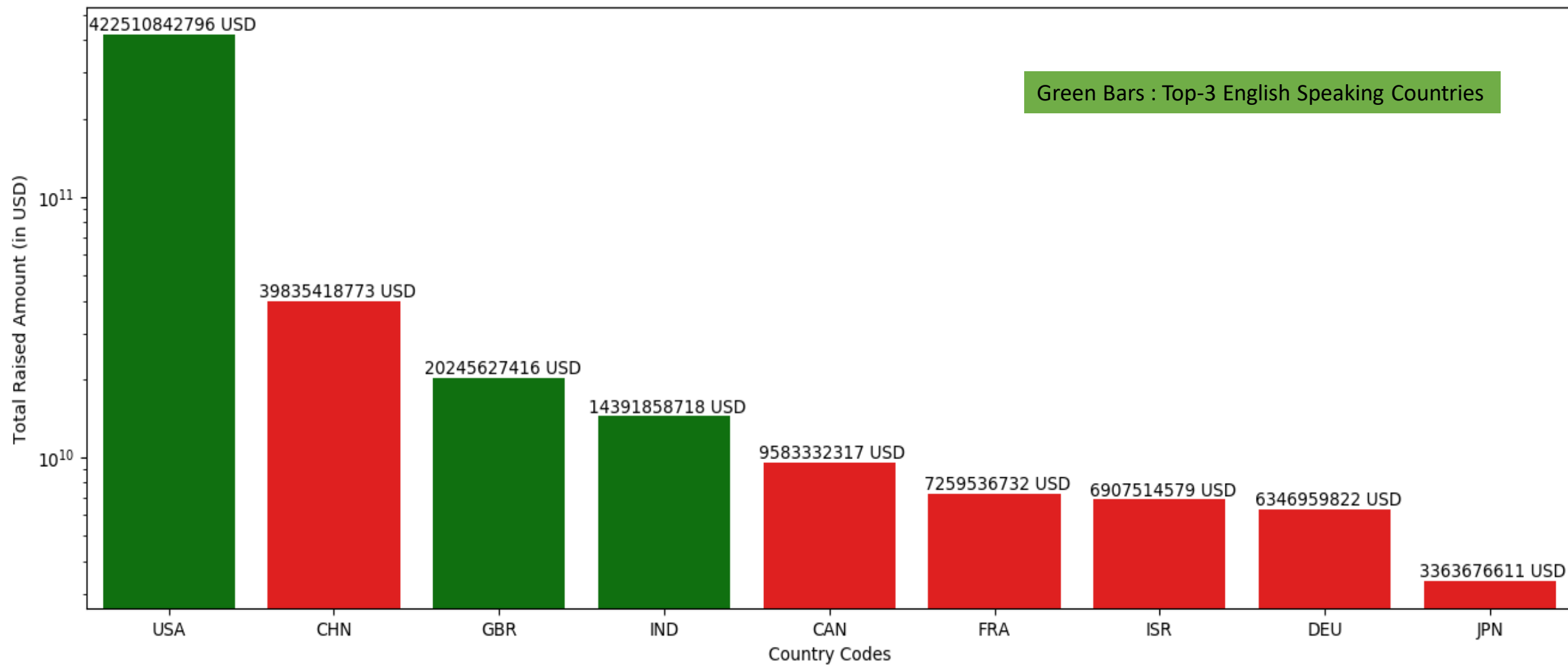4) News, Search and Messaging is the 3rd most funded sector in IND.

There are a few assumptions made during this part of analysis :

1) Nan in Mapping is replaced with ('') blanks.

2) Primary categories which are not present in mapping file are mapped to Nan as there were only 79 such instances in a this big data sample thus not making a big difference in the overall analysis.
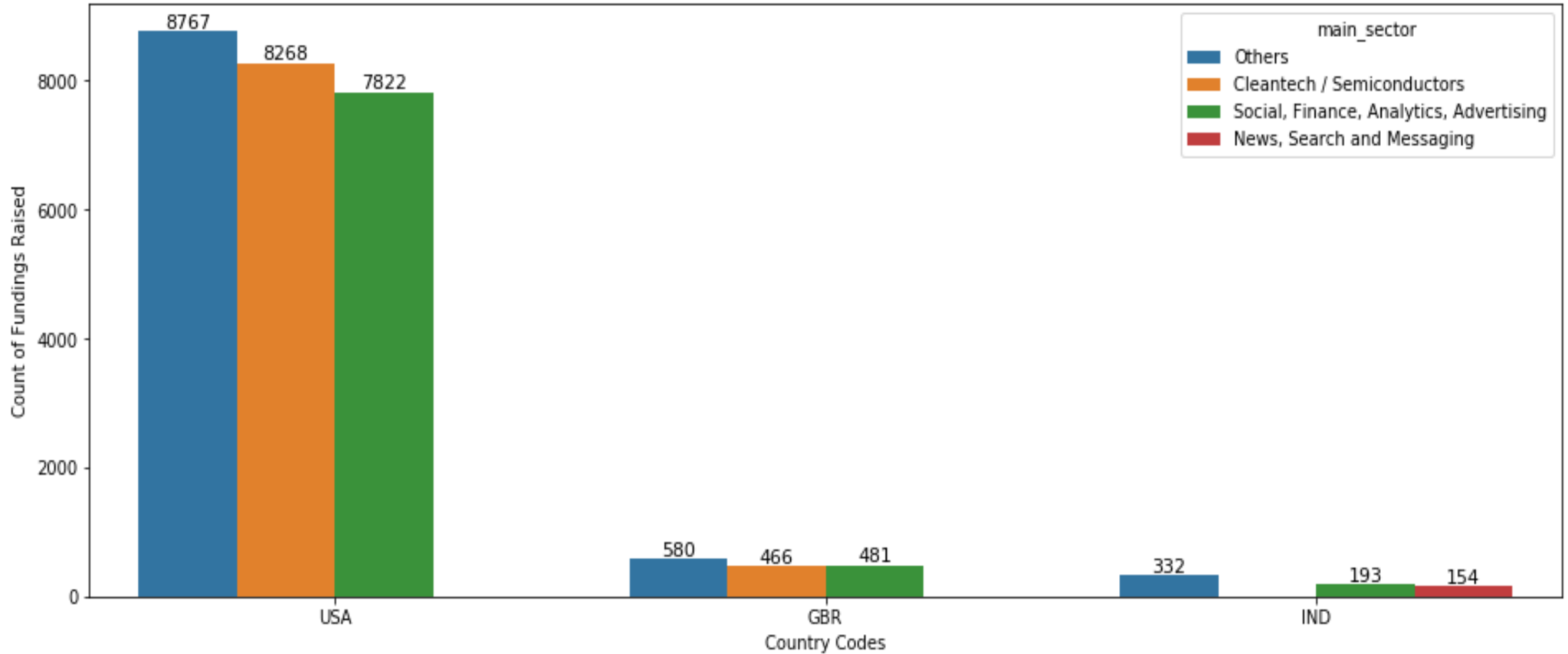
<Results>

<Results>

<Conclusions>
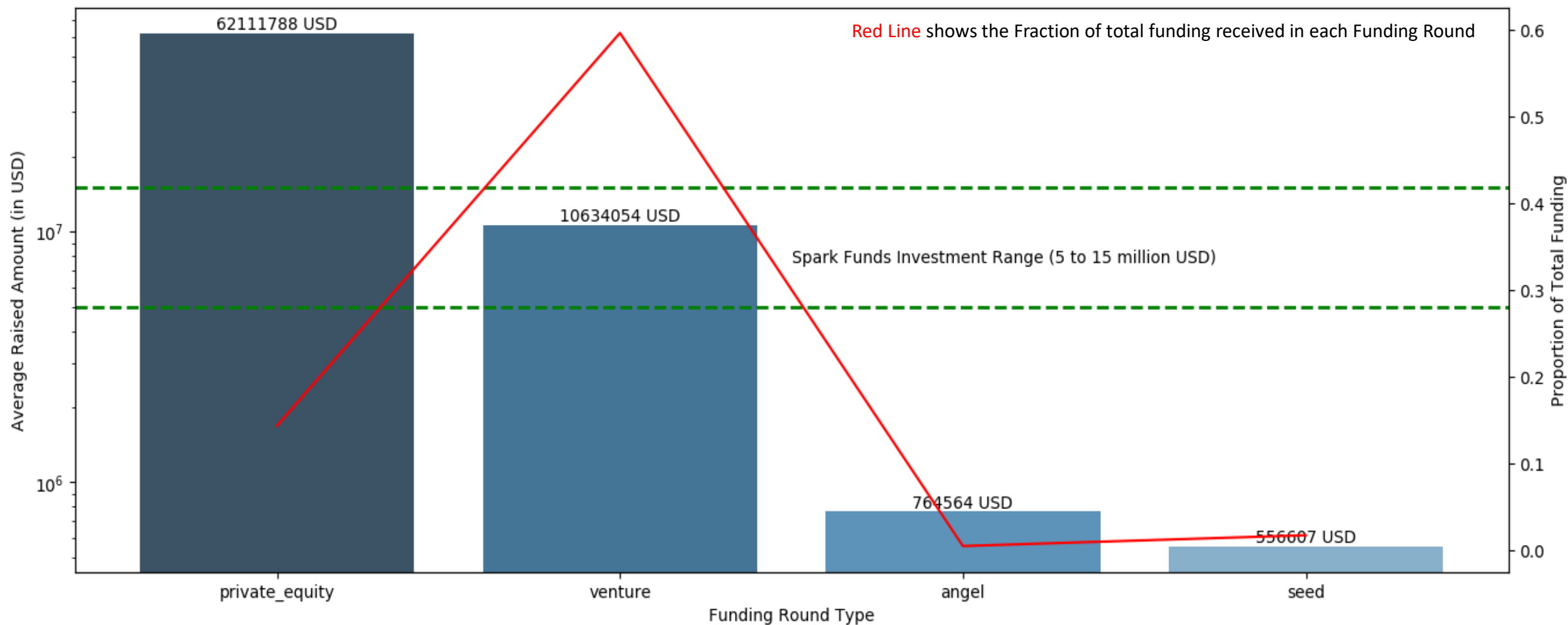
Based on the analysis we can conclude :

- **Venture Round** is the best suited funding round for Spark Funds.

- **USA**, **GBR** and **IND** are the most suited Countries to Invest in as these are the top 3 English speaking countries where maximum amount of investments are being made.

- 'Others' Sector is the most invested sector in the top 3 countries thus Spark Funds can also invest in this sector in any of the 3 countries.

- **Cleantech / Semiconductors** is the second most funded main sector in USA and GBR.

- **Social, Analytics and Advertising** is the 3rd main sector in USA and GBR while in IND it is 2nd most funded sector.

- **News, Search and Messaging** is the 3rd most funded sector in IND.

- Thus, Spark Funds can invest in these sectors for better returns.

- We suggest Spark Funds to invest in the following 6 companies for best returns :
  - **/Organization/Social-Finance**
  - **/Organization/Freescale**
  - **/Organization/Oneweb**
  - **/Organization/Powa-Technologies**
  - **/Organization/Flipkart**
  - **/Organization/Shopclues-Com**