INVESTMENT CASE STUDY

SUBMISSION

Group Name:

1. RANIP HORE
2. AMITH PRADHAN
3. SANCHARI GAUTAM
4. PRATIK NATH

<Abstract>

- Spark Funds, an asset management company wants to make investments in a few companies. Hence, to understand the global trends in investments, real investment data is collected from *crunchbase.com*.
- Spark Funds has two **minor constraints** for investments:
    1. It wants to invest between 5 to 15 million USD per round of investment.
    2. It wants to invest only in English-speaking countries because of the ease of communication with the companies it would invest in.
- **Business objective:**

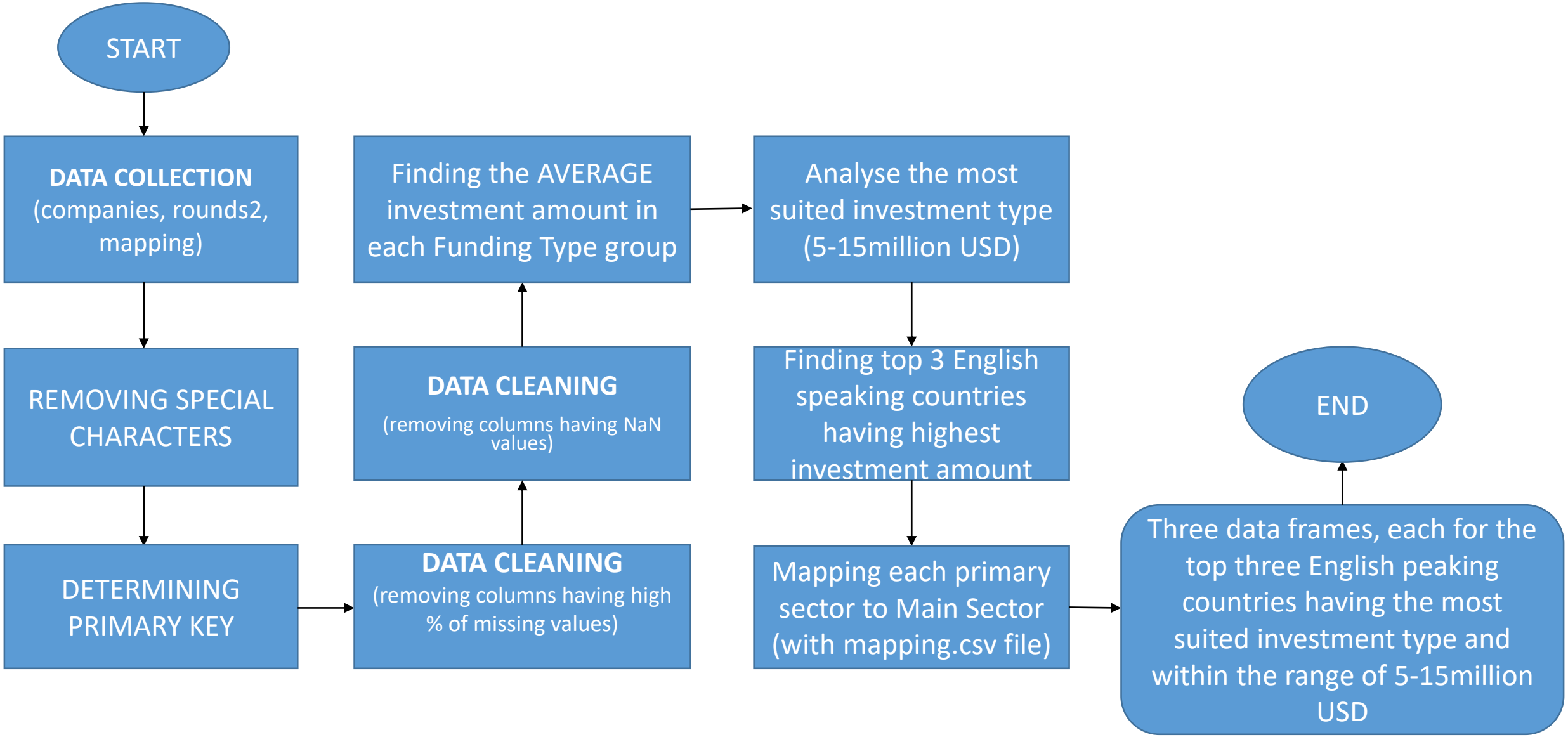
The objective is to identify the best sectors (by sector analysis) , countries (by country analysis), and investment type (by investment analysis) for making investments.

- **Strategy:**

To invest where others are investing, implying that the 'best' sectors and countries are the ones 'where most investors are investing'.

- **Outcome:**
    - The most suitable funding type among the four funding types (venture, angel, seed, and private equity)
    - The top three English-speaking countries
    - The most heavily invested main sectors in each of the three countries

<Problem solving methodology>



<Analysis – Data Cleaning Phase>

- The companies and rounds tables are imported with the encoding type ISO-8859-1, having 66368 rows.
- The columns in companies table having null values in companies are computed. It turns out that ‘permalink’ is the column which has zero null values and having 66368 unique values. Hence, ‘permalink’ is considered to be the primary key of companies dataset.
- A similar fashion is applied in rounds2 table. It can be inferred ‘company_permalink’ is the primary key in rounds2 table.
- The companies and rounds2 tables are merged together on the key values and names as ‘master_frame’. After merging, the observation is 114949 rows and 16 columns.
- The columns having higher percentage of missing values are dropped. In master_frame, we find that ‘founded_at’ and ‘funding_round_code’ have 17.85% and 72.91% missing values respectively. So a better decision would be dropping these two columns.
- The column ‘raised_amount_usd’ has 17.39% NaN values. Hence, the rows having NaN values are dropped.
- The ‘master_frame’ now is suitable for analysis. After the data cleaning process, we are left with 94959 rows and 14 columns.

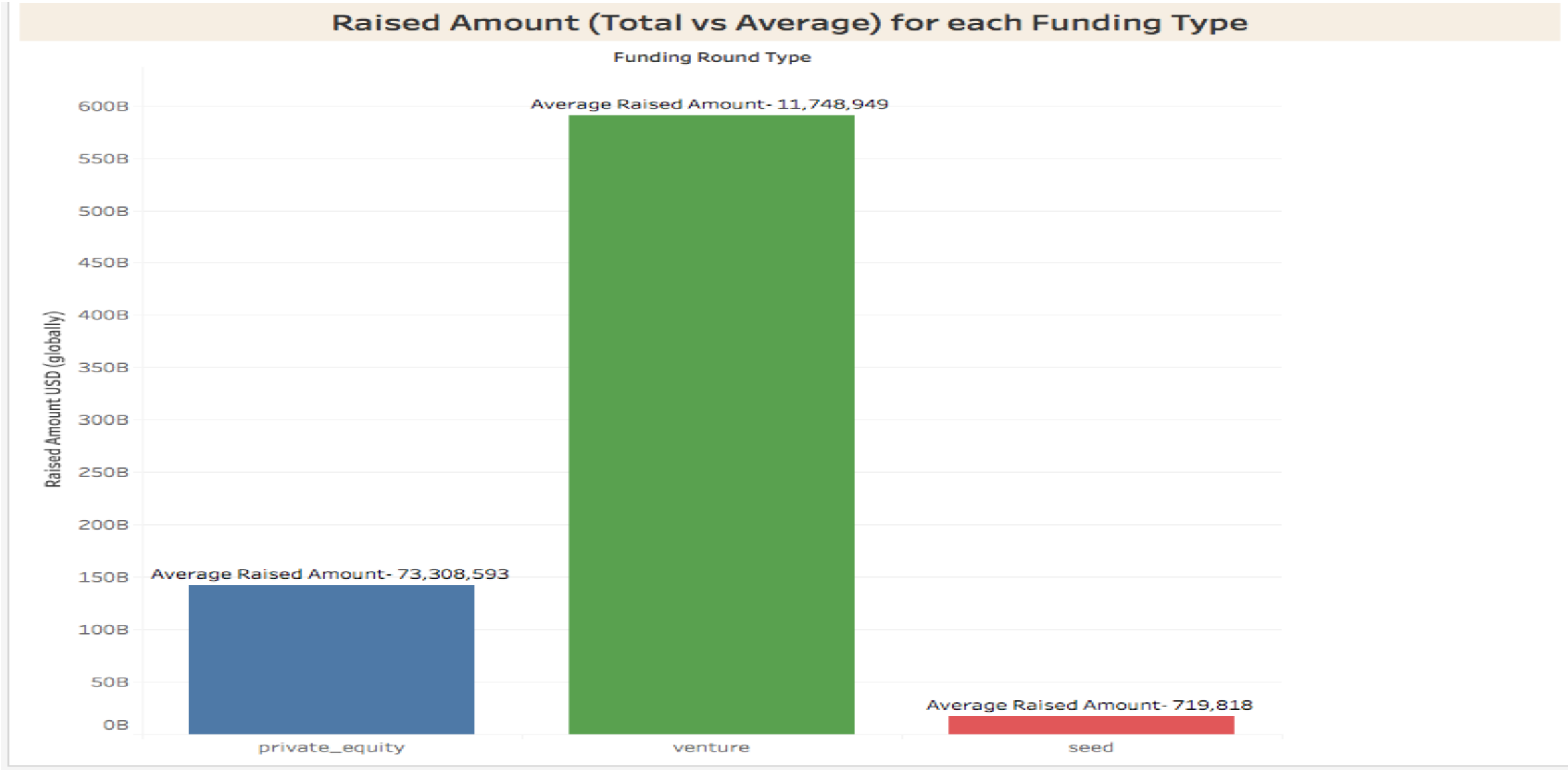
<Analysis – Part 1>

- The average investment in each funding type is determined after grouping the master_frame by ‘funding_type’.
- The most suitable type of investment is found to be ‘Venture’ because it’s average investment lies in the range of 5-15 million dollars.
- The master_frame with only Venture funding type is then grouped by the country_code and the countries with highest total investment is computed. It turns out that USA, Great Britain and India are the three top English speaking countries.
- The mapping dataset provided the name of the eight main sectors along with the names of each sector incategory_list and under which main sector does it fall.
- The row with NaN value in category list has been dropped. Also a trend is observed for the sector names under category_list column having 0 in place of ‘na’. Hence correction has been done to replace 0 with ‘na’ to get the actual sector name.
- The primary_sector column is introduced in master_frame which is the first portion of the ‘category_list’ (before pipe). This primary_sector is mapped to main_sector, which is determined from the mapping dataset.
- A merged_master_frame is computed having all the columns of master_frame along with the main_sector. This frame contains 94959 rows and 16 columns.

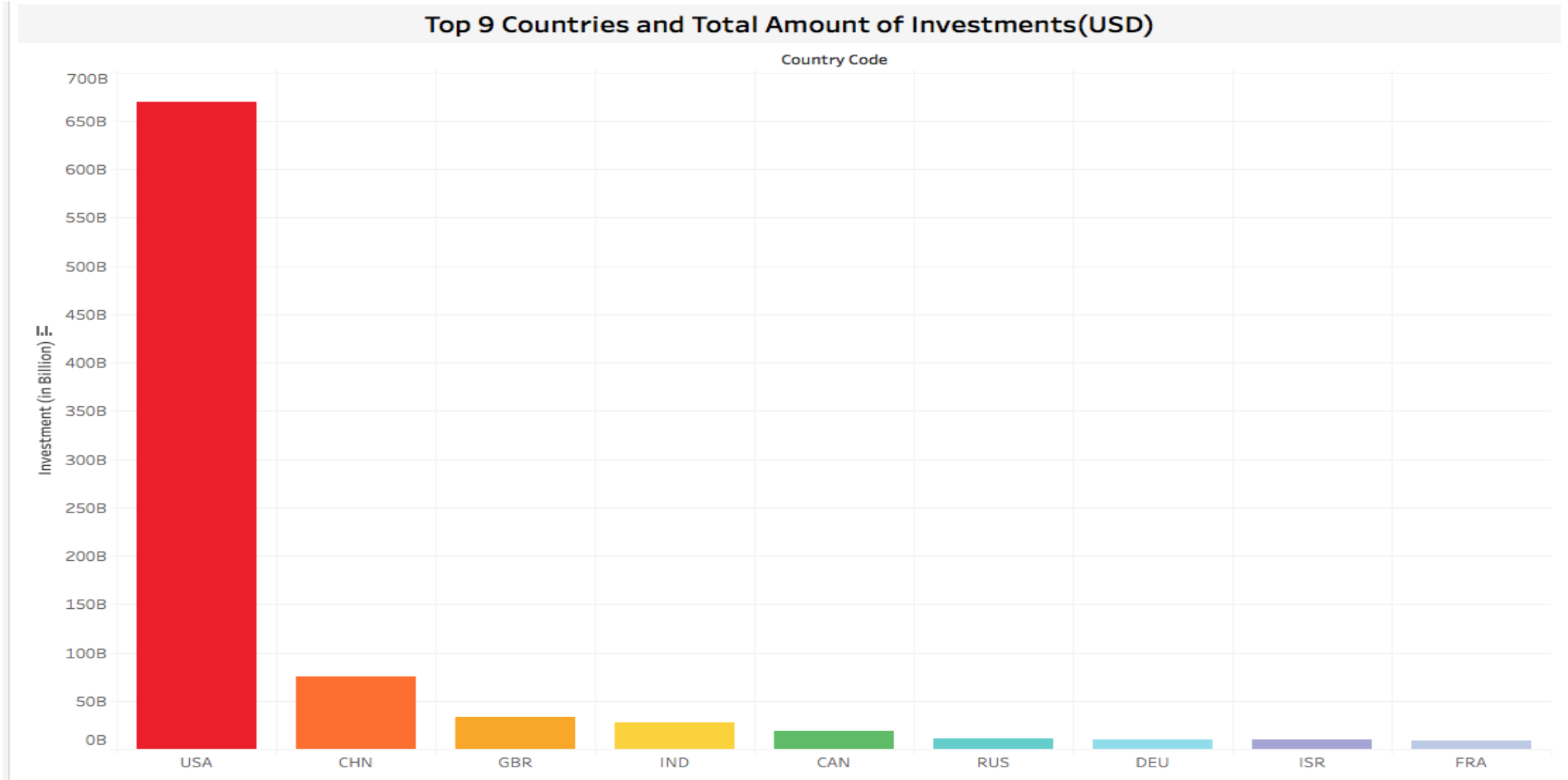
<Analysis – Part 2>

- Three data frames are created for top three English speaking countries with funding type venture and falling within the 5-15 million USD range.
- Two new columns are introduced in each of the data frames. One is ‘total_no_of_investment’ which gives the total number (or count) of investments for each main sector. Other is ‘total_amount_invested’ which gives the total amount invested in each main sector.
- The following analysis is shown in the plots attached thereafter –
  - 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 makes it clear that ‘venture’ is best suited for Spark Funds.
  - A plot showing the top 9 countries against the total amount of investments of funding type ’Venture’. This shows the top 3 countries (USA, Great Britain, and India) very clearly.
  - 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 venture).

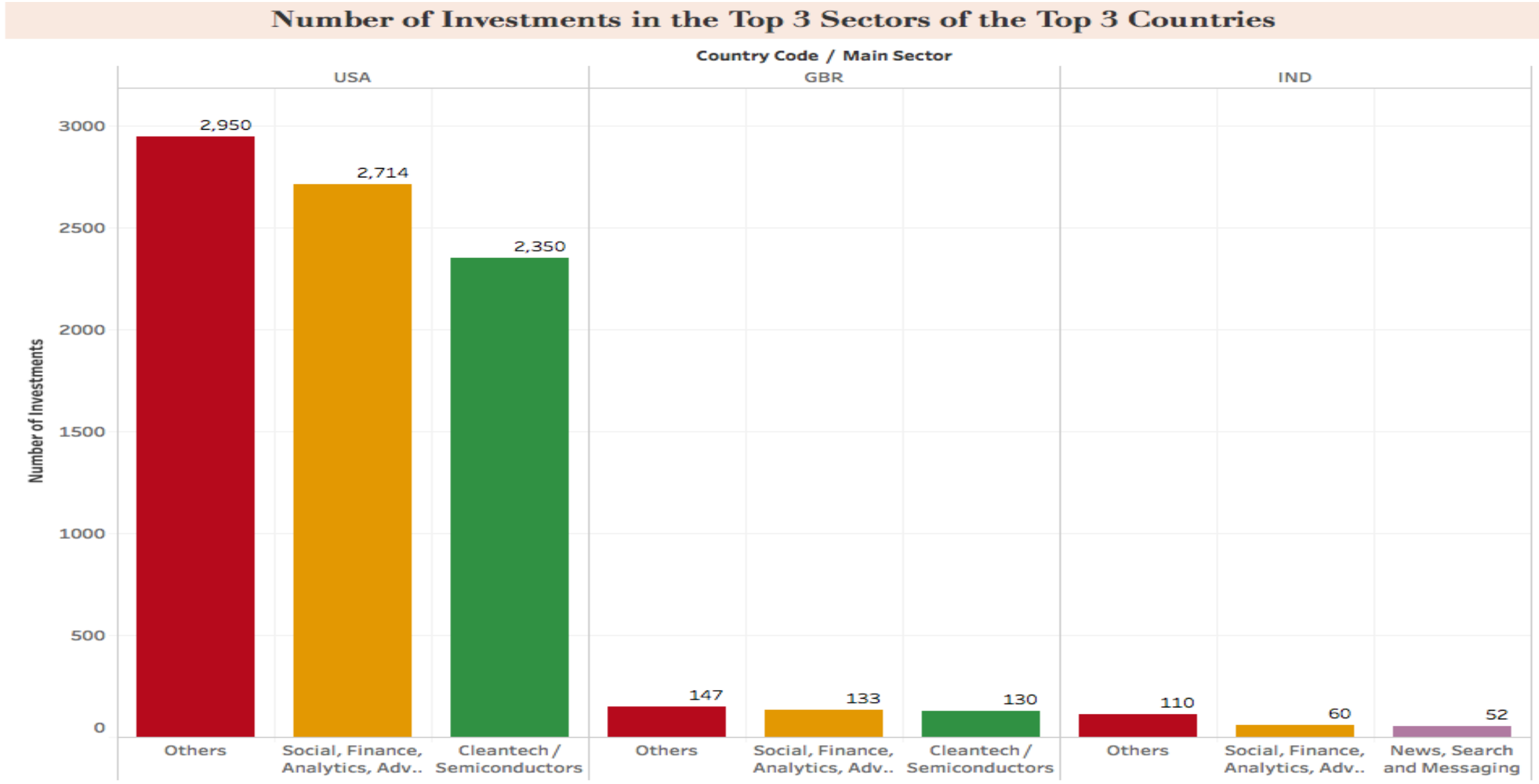
<Results – PLOT 1>



<Results – PLOT 2>



<Results – PLOT 3>



<Conclusions>

The Spark Funds Company should invest in the countries of USA, Great Britain and India within their investment budget of 5-15 million dollars. The funding type chosen should be Venture type as it serves all the criterion.

According to the country it decides to invest on , it should invest on any one of the top three sectors as provided below :

USA : Others, ‘Social, Finance, Analytics, Advertising’ and ‘Cleantech / Semiconductors’.

GBR : Others, ‘Social, Finance, Analytics, Advertising’ and ‘Cleantech / Semiconductors’.

IND : Others, ‘Social, Finance, Analytics, Advertising’ and ‘News, Search and Messaging’.