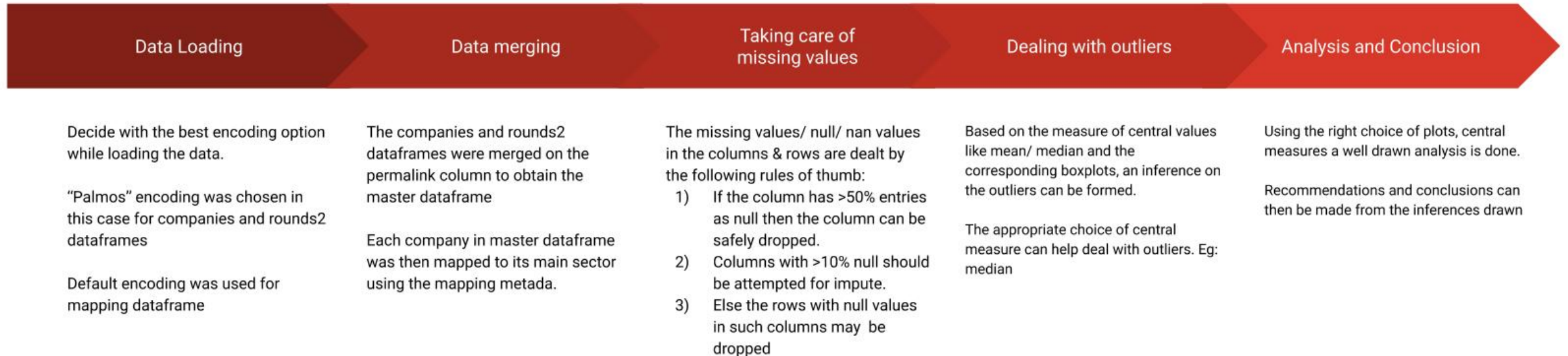# INVESTMENT ASSIGNMENT

# SUBMISSION

Name: Yash Sarawgi

Batch: PGD ML & AI July 2020

# Abstract

- Spark Funds, an asset management company wants to invest in a few companies with the following constraints:
  - 5-15 million USD per round of investment
  - English speaking countries
  - English speaking countries may be considered only where English is an official language
- This presentation uses the CRISP-DM framework for analysis with Exploratory Data Analysis during the process.
- The objective is to help Spark funds decide upon the most valuable investment options for its funds in order to increase its ROI based on:
  - Funding type
  - Country
  - Sector

# Problem Solving Methodology

| Data Loading | Data merging | Taking care of missing values | Dealing with outliers | Analysis and Conclusion |
|---|---|---|---|---|

**Data Loading**

Decide with the best encoding option while loading the data.

"Palmos" encoding was chosen in this case for companies and rounds2 dataframes

Default encoding was used for mapping dataframe

**Data merging**

The companies and rounds2 dataframes were merged on the permalink column to obtain the master dataframe

Each company in master dataframe was then mapped to its main sector using the mapping metada.

**Taking care of missing values**

The missing values/ null/ nan values in the columns & rows are dealt by the following rules of thumb:
1) If the column has >50% entries as null then the column can be safely dropped.
2) Columns with >10% null should be attempted for impute.
3) Else the rows with null values in such columns may be dropped

**Dealing with outliers**

Based on the measure of central values like mean/ median and the corresponding boxplots, an inference on the outliers can be formed.

The appropriate choice of central measure can help deal with outliers. Eg: median

**Analysis and Conclusion**

Using the right choice of plots, central measures a well drawn analysis is done.

Recommendations and conclusions can then be made from the inferences drawn

# Funding Analysis

- Fig 1: There are a large number of outliers in raised amount usd
- Fig 2: The outliers and the large gap between the values of mean and mode for each funding type suggests that median is the right metric to decide the choice of funding and deal with outliers
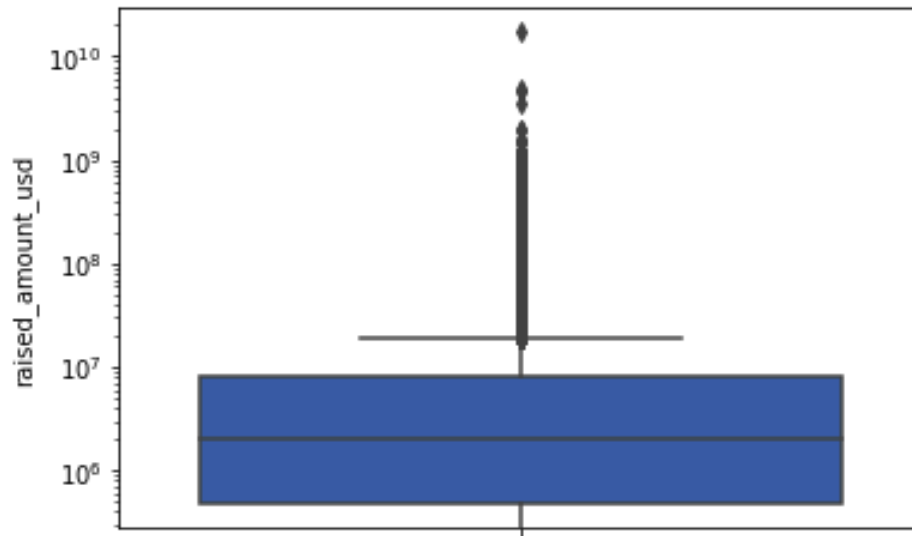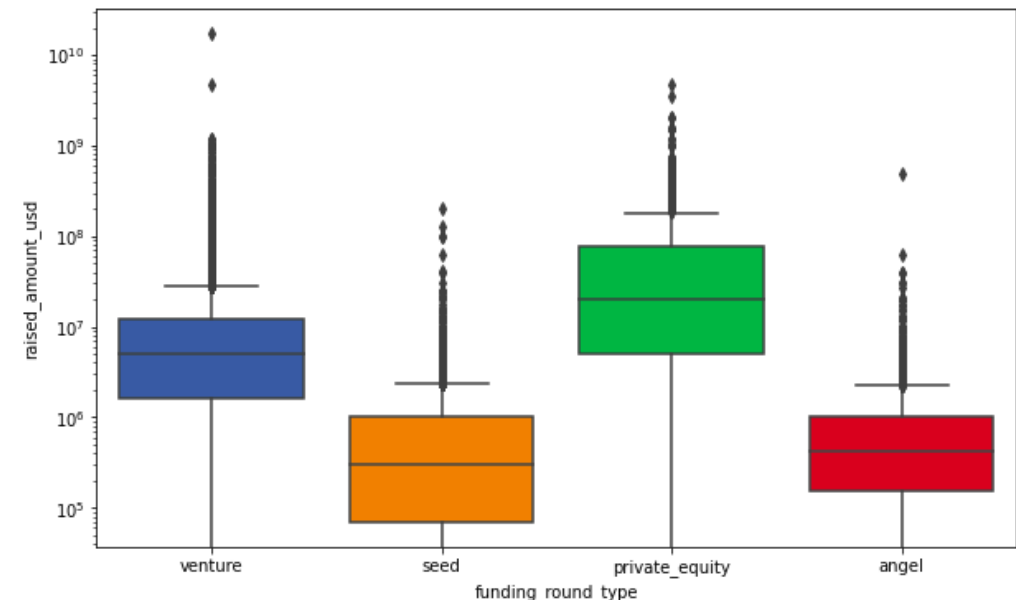


Fig 1: Boxplot of the amount raised in USD



Fig 2: Boxplot of the amount raised in USD for each   category of funding round type

# **Funding Analysis** (Finding the best suitable type)

- Table 1: From this table of categorical (based on funding type) values of raised amount in USD (median); it is clear that the choice of funding for Spark Funds is => **VENTURE** type

  – Constraint: 5-15 million USD per round of investment

  – Angel and Seed have very low median values as per constraint and Private Equity has too high value. Only venture type falls in the range of constraint.

Table 1: Average (Median) values of raised amout in USD based on funding round type categories

| Funding Round Type | Average (Median) values |
|---|---|
| Angel | 414,906 |
| Private Equity | 20,000,000 |
| Seed | 300,000 |
| **Venture** | **5,000,000** |

# Country Analysis

- Table 2: The table clearly shows the top 9 countries based on the total value of investment. Top 3 countries (code) of interest for Spark Funds are : **USA**, **GBR** and **IND**

  - Constraint: Only English speaking countries

  - China (CHN) though ranked 2 doesn't fit in the constraint as English is not the official language there

Table 2: Top 9 countries based on the total amount of investment

| | country_code | raised_amount_usd |
|---|---|---|
| 0 | USA | 420068029342.0000 |
| 1 | CHN | 39338918773.0000 |
| 2 | GBR | 20072813004.0000 |
| 3 | IND | 14261508718.0000 |
| 4 | CAN | 9482217668.0000 |
| 5 | FRA | 7226851352.0000 |
| 6 | ISR | 6854350477.0000 |
| 7 | DEU | 6306921981.0000 |
| 8 | JPN | 3167647127.0000 |
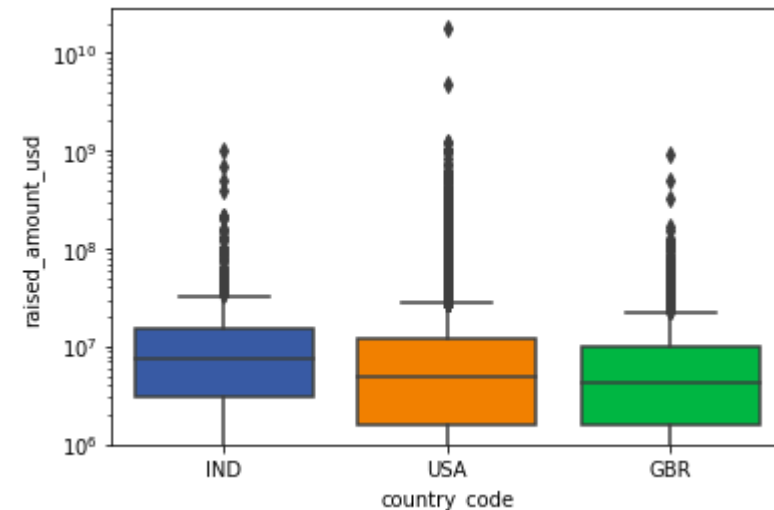
Fig 3: Boxplot of raised amount in USD based on countries

# Sector Analysis

- There are 8 main sectors that each company can be classified into.

- Table 3: Summaries the sector-wise analysis of each country based on the number of investments and total amount of investment in each.

  - Based on the number of investments (circled):

    - The top 3 sectors in USA:
      - Others
      - Social, Finance, Analytics and Advertising
      - Cleantech / Semiconductors

    - The top 3 sectors in GBR:
      - Others
      - Social, Finance, Analytics and Advertising
      - Cleantech / Semiconductors

    - The top 3 sectors in India:
      - Others
      - Social, Finance, Analytics and Advertising
      - News, Search and Advertising

Table 3: Sector wise analysis of each country on number of investments (count) and total amount of investments (sum)

| country_code | main_sector | count | sum |
|---|---|---|---|
| GBR | Automotive & Sports | 16 | 167051565.0000 |
| | Cleantech / Semiconductors | 130 | 1163990056.0000 |
| | Entertainment | 56 | 482784687.0000 |
| | Health | 24 | 214537510.0000 |
| | Manufacturing | 42 | 361940335.0000 |
| | News, Search and Messaging | 73 | 615746235.0000 |
| | Others | 147 | 1283624289.0000 |
| | Social, Finance, Analytics, Advertising | 133 | 1089404014.0000 |
| IND | Automotive & Sports | 13 | 136900000.0000 |
| | Cleantech / Semiconductors | 20 | 165380000.0000 |
| | Entertainment | 33 | 280830000.0000 |
| | Health | 19 | 167740000.0000 |
| | Manufacturing | 21 | 200900000.0000 |
| | News, Search and Messaging | 52 | 433834545.0000 |
| | Others | 110 | 1013409507.0000 |
| | Social, Finance, Analytics, Advertising | 60 | 550549550.0000 |
| USA | Automotive & Sports | 167 | 1454104361.0000 |
| | Cleantech / Semiconductors | 2350 | 21633430822.0000 |
| | Entertainment | 591 | 5099197982.0000 |
| | Health | 909 | 8211859357.0000 |
| | Manufacturing | 799 | 7258553378.0000 |
| | News, Search and Messaging | 1583 | 13971567428.0000 |
| | Others | 2950 | 26321007002.0000 |
| | Social, Finance, Analytics, Advertising | 2714 | 23807376964.0000 |

# Sector Analysis (Companies of Interest)

- Table 4: It lists the companies with the highest amount of investment in the top 2 sectors (based on number of investment) of the top 3 countries.

Table 4: Companies with highest investment in the top 2
sectors (based on number of investment) of top 3 countries

| Sector (down) & Country (right) | C1 (USA) | C2 (GBR) | C3 (IND) |
|---|---|---|---|
| Others | Virtustream | Electric Cloud | FirstCry.com |
| Social, Finance, Analytics and Advertising | SST Inc. (Formerly ShotSpotter) | Celltick Technologies | Manthan Systems |

# Results: Plot 1

Plot 1.1: Total amount raised for each funding type as a fraction of the total investments (globally)
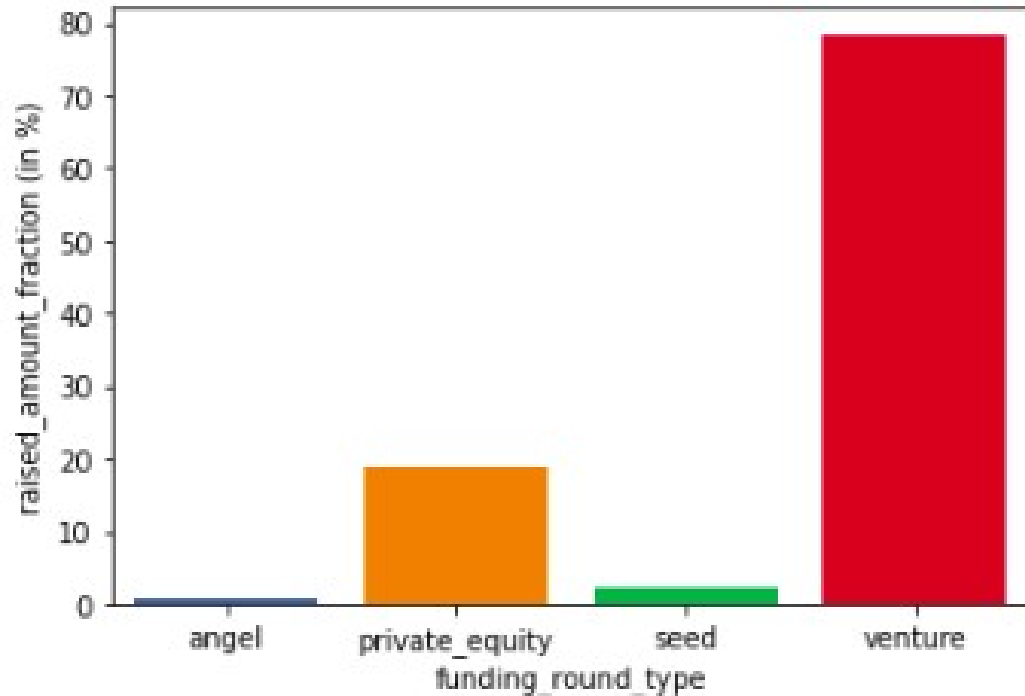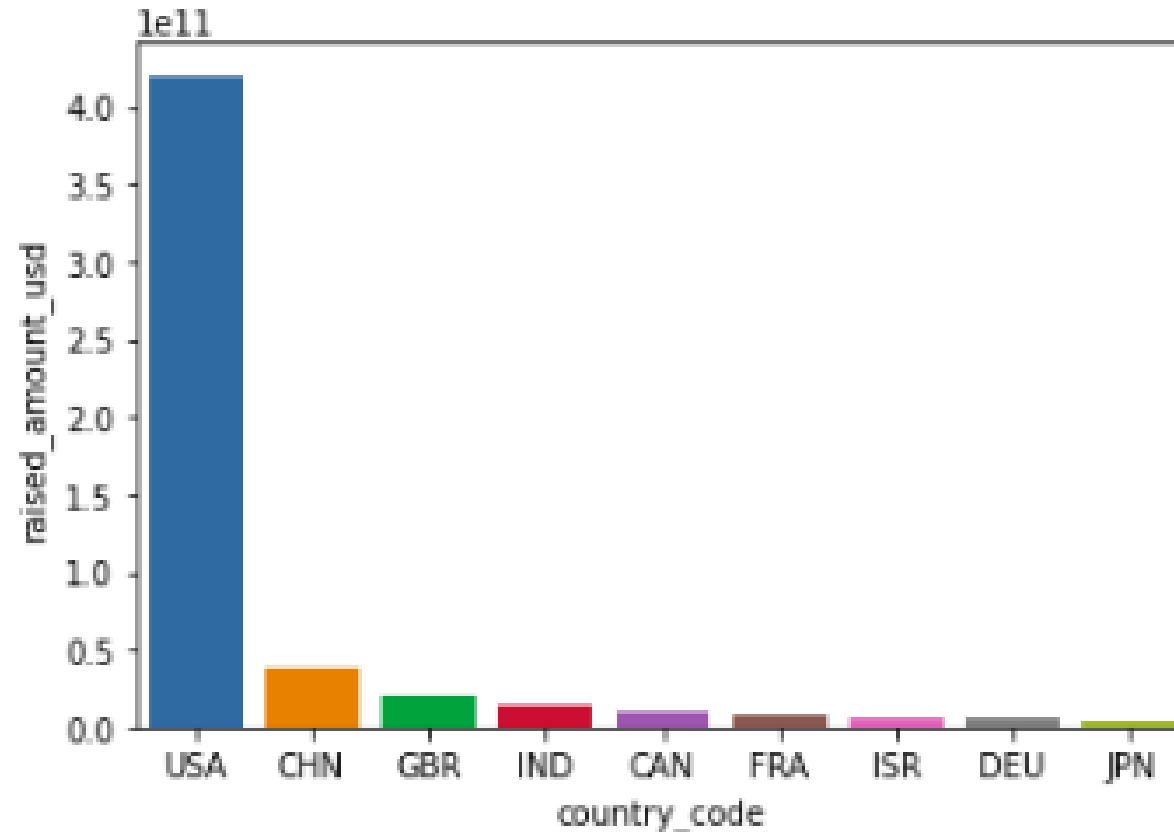
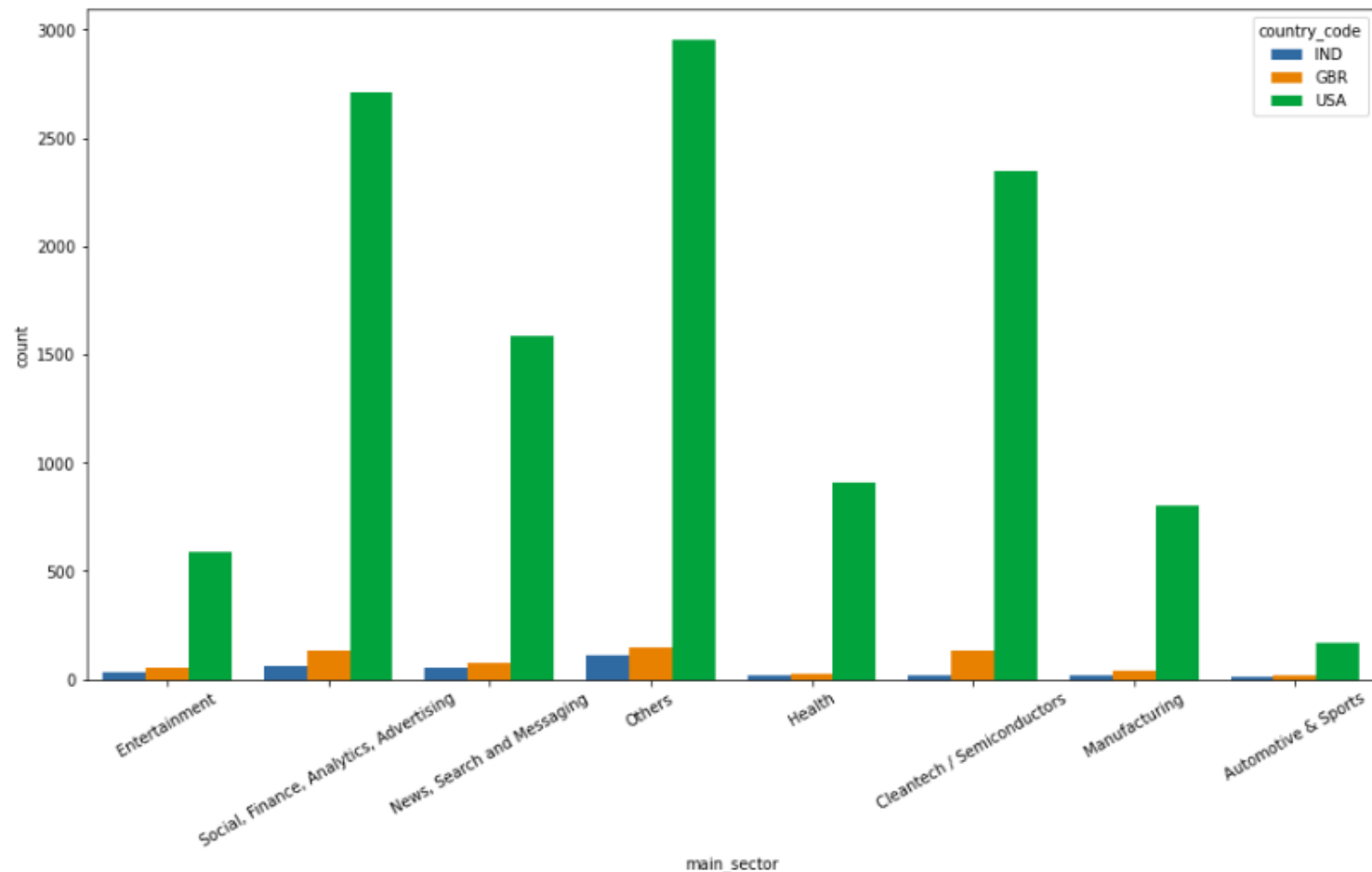Plot 1.2: Median of Amount raised in USD for each funding type
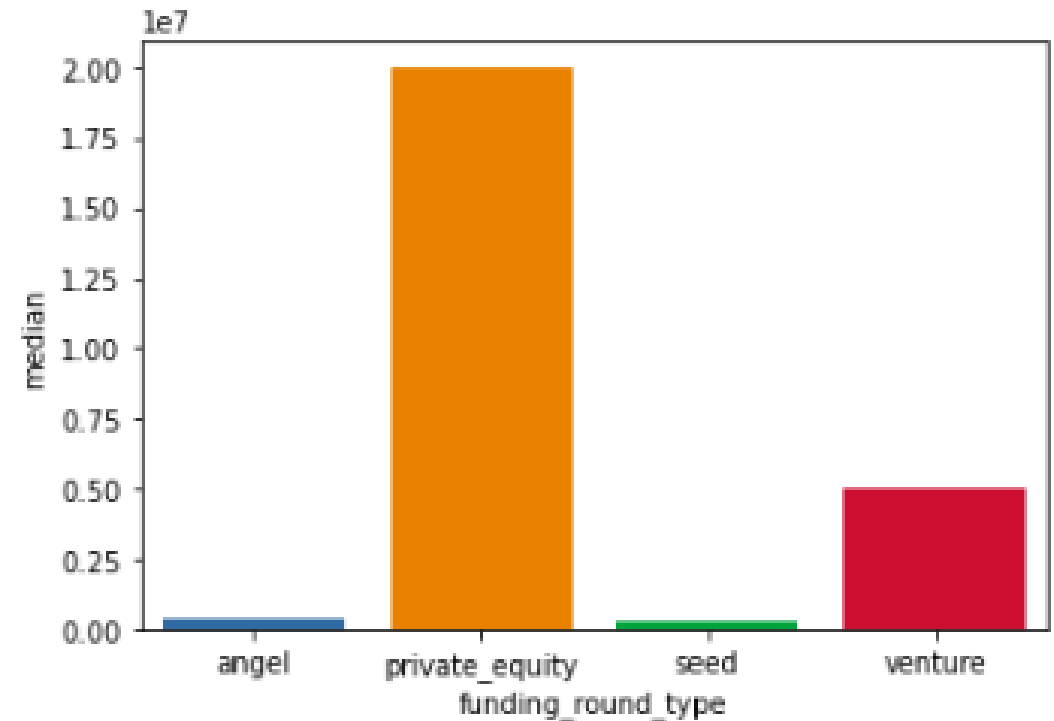
# Results: Plot 2 (for venture type funding)

Plot 2: Top 9 countries with the total amount of investment made in each country

(English speaking countries of interest are: USA, GBR & IND)

# Results: Plot 3 (Plot 3.1) (for venture type funding)

Plot 3.1 The total number of investments made in each sector for the top 3 countries

# Results: Plot 3 (Plot 3.2) (for venture type funding)

Plot 3.2 The total number of investments made in each country across various sectors

(This clearly depicts the top 3 sectors (☆marked) in each country. Hence, better insights than representation in Plot 3.1)
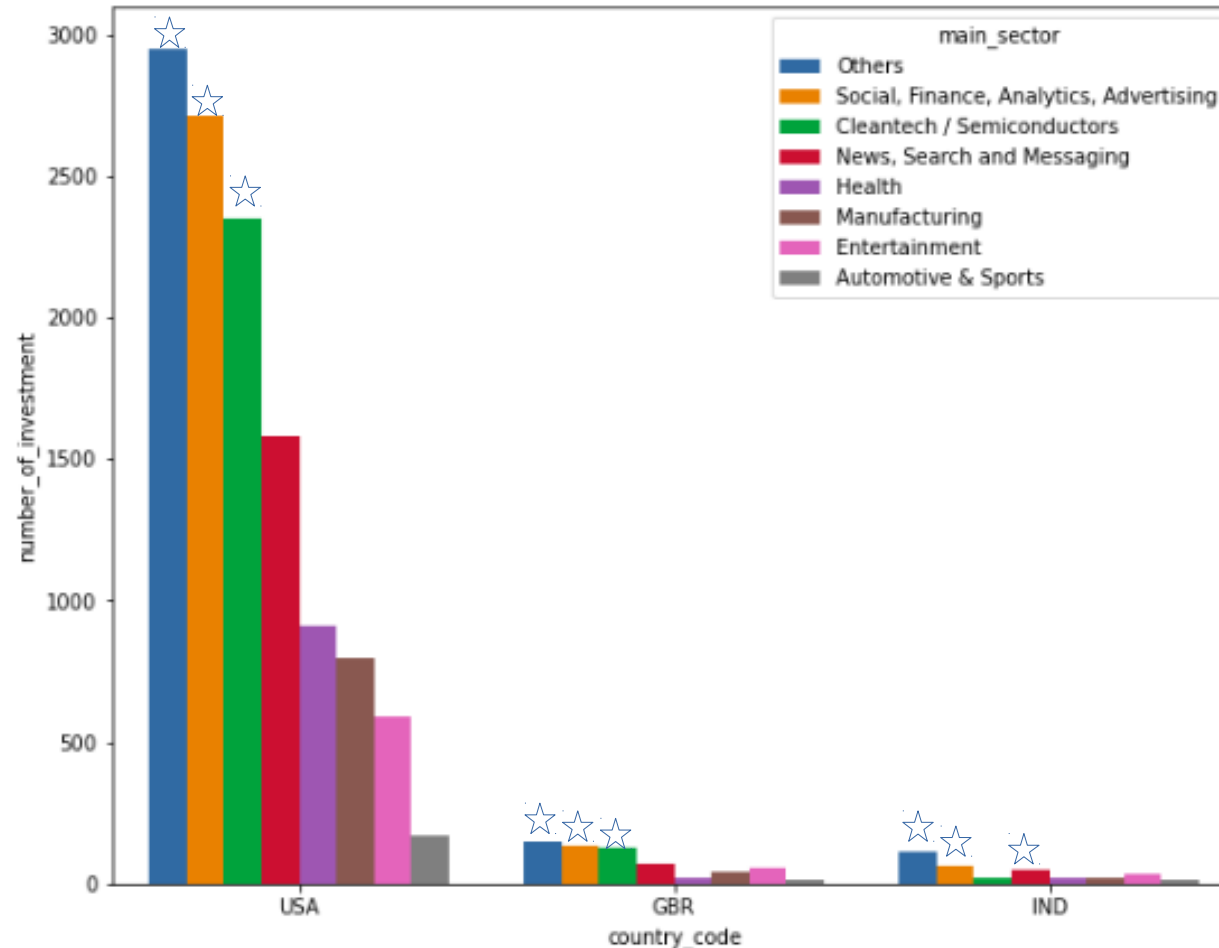
# Inferences & Recommendations

- From Plot1.1 and Plot 1.2 it is seen that largest fraction of the investments are made in the venture type of funding; also the median value of venture type falls in the range of 5-15 million USD as per constraint, making venture type the best option of investment.

- From Plot 2 based on the number of investments, the countries that Spark Funds should focus for investment may be United States (USA), United kingdom (GBR) or India (IND).

- From plot 3.2 the best sectors for investment in each country have been traced out.

- For highest ROI, based on the analysis and results, Table 4 lists a few companies in the top 2 sectors (same across countries, proving their relevance as a choice). It would be strongly recommended for Spark Funds to create a diverse portfolio by choosing companies across countries and sectors as shown in Table 4.

# Conclusions

- Based on the analysis, results, inferences and recommendations, the following conclusions have been drawn for Spark Funds:

  - Funding type that is most suitable : Venture type

  - Countries that are a safe choice to choose companies from: USA, GBR, IND

  - Sectors that have proven across 3 countries:
    - Others
    - Social, Finance, Analytics and Advertising

  - For further investments "Cleantech / Semiconductors" in USA and GBR; "News, Search and Messaging" in IND can be choice of sectors where Spark Funds can look for investing

  - A few companies with highest amount of investment in the top 2 sectors are:
    - Virtustream
    - Electric Cloud
    - FirstCry.com
    - SST Inc.
    - Celltick Technologies
    - Manthan Systems