# INVESTMENT ASSIGNMENT

# SUBMISSION

Name: Jennis Vaishnav

# <Abstract>

The Spark funds want to understand the global market trends. The main motive of this document would be to provide insights from data, which can be used by the CEO of the Spark Funds as in where to invest the money, In which security/funding type, and in which sectors the investment should be made. Some constraints are shared with us. The analysis focuses on where most of the investments are made.

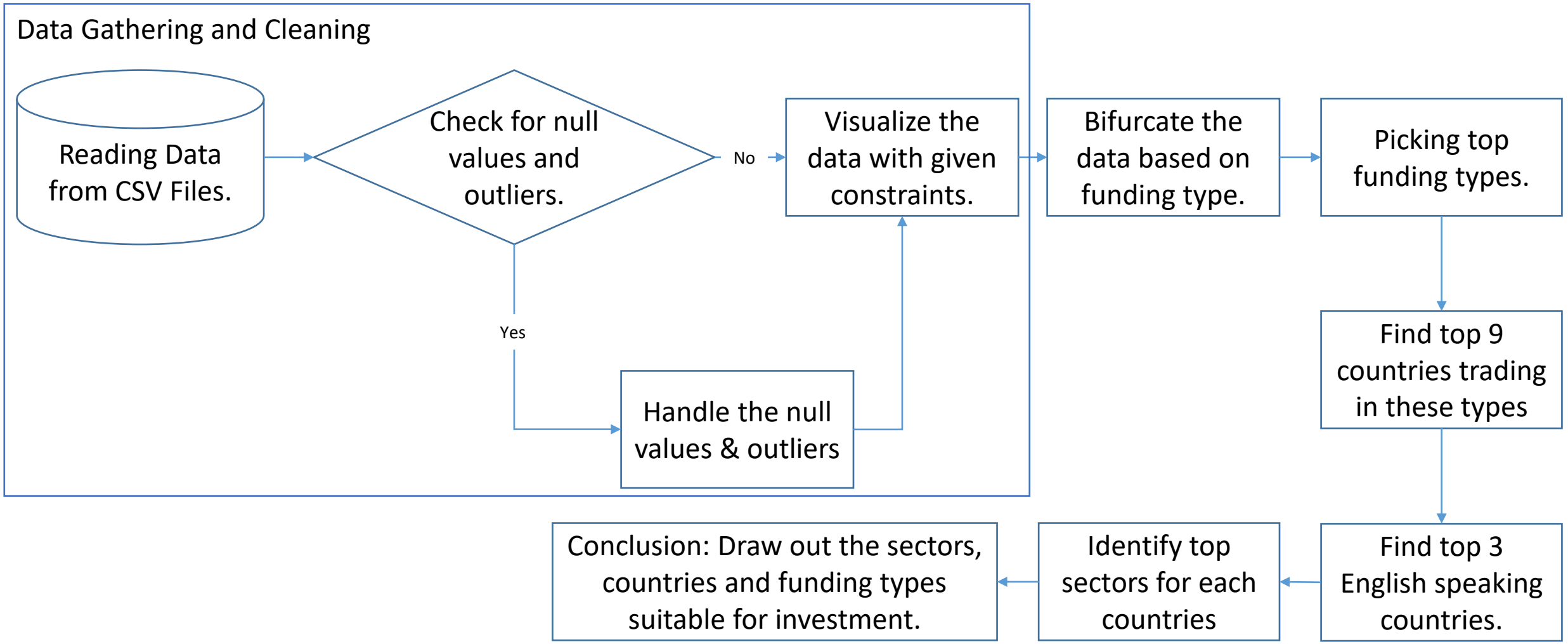
Various visuals have been used to visualize the investments and their trends that are bar graphs, and grouped bar charts. This analysis is not affected by the outliers present in the data. There were certain null entries in the data which were handled before conducting the analysis. The outliers were visualized using the boxplot method, and necessary steps have been taken, in such a way that it won't have a false impact on our analysis.

Now, Throwing some light on the criteria of the analysis, Spark funds wants to invest money from a range between 5 million - 15 million USD. This search/analysis should be only done where English is spoken as the main language as a mode of communication. The entire analysis is divided into a number of checkpoints. Each checkpoint would have some expected results, As in what output the customer will be expecting.

We will be comparing the trends of typical funding types that are seed, angel, private equity, and venture. Past performance of these funding types would be analyzed. From these analyzed data visualization we will find which funding type is heavily invested. In fact, heavily invested funding type would be considered as favorites of Spark funds. There weren't any major technical challenge faced during the study except a few. The first one was the type of encoding to be used at the time of reading the file. The second challenge was to reframe the entire data frame. Standard and robust methods were used to solve those challenges.

From a safety perspective, This analysis shall also be compared and analyzed further with the current situation of the market. That is the current state of the market should also be considered before investing.

# <Problem solving methodology>



# <Analysis>

1. First we imported the files and libraries necessary for performing the analysis operation.
2. Using palmos encoding both the files were read.
3. Before find the unique number of rows the rows were converted to upper case.
4. “Permalink” was identified as unique key for each company.
5. Then we checked if any company in round2 csv files was absent in company, but the result turned out to be:- All the files were present in the file.
6. Later both the dataset were merged by left join by “company_permalink” and “permalink” as they are the primary keys in the datasets.
7. Then we filtered the data on based of funding type as client wanted:
    i. Venture
    ii. Angel
    iii. Seed
    iv. Private Equity

# <Analysis>

8. The rows where were null or of no use where filtered out. And we checked the percentage of rows that were null, Which turned out to be 14%. Around 14% of the rows were null. They were dropped it for later use to avoid false visualization.

9. Representative of each funding type was to be decided, So to represent the funding types we had a few options to represent, i.e:

   i. Mean

   ii. Median

A look on boxplot of the funds that are raised:

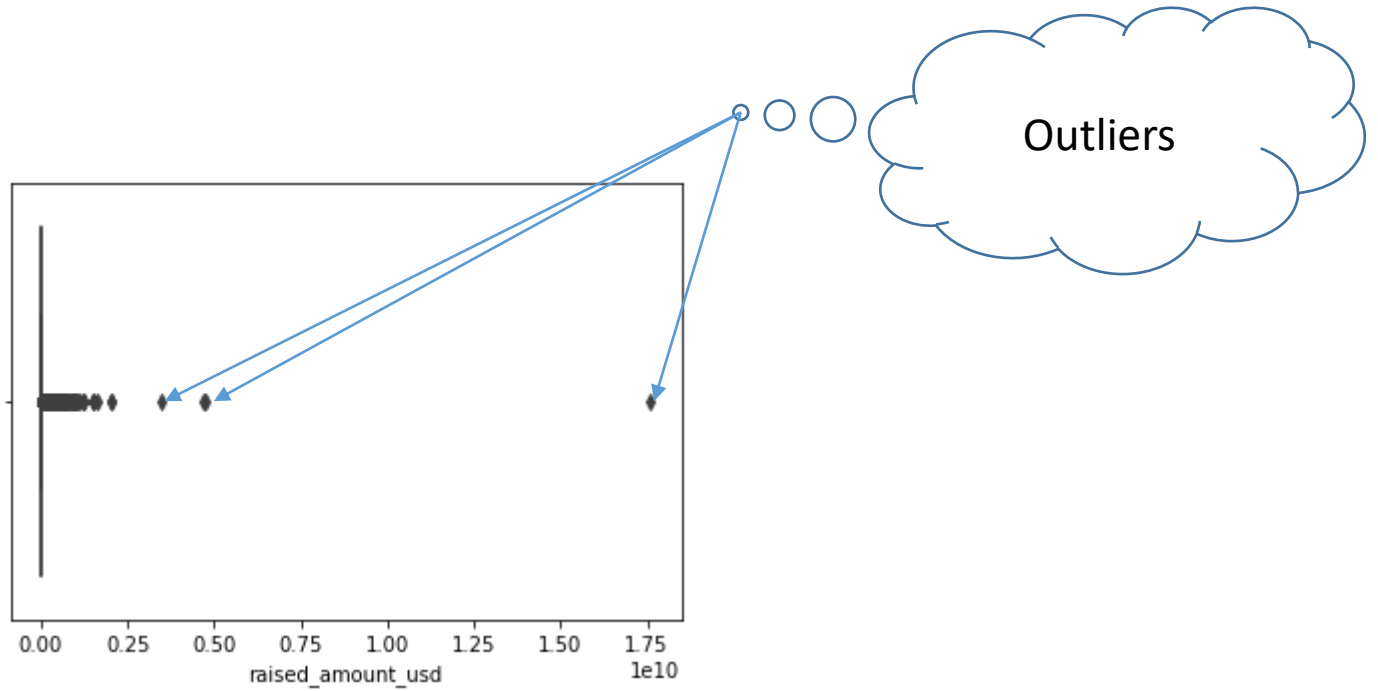



Figure:1.0

# <Analysis>

This are not the only outliers though, So having a look at outliers, we can be sure of one thing that we cannot take mean values to represent the funding type column. So using use median to analyse/represent the funding type column.

10. Now comes the most difficult part of the project. We were suppose to reshape the data frame. In laymen language we would be transforming the rows in columns and columns to the rows. We came across various methods such as stack, unstack, and melt method. Out of which melt method was opted.

11. Now we were in a position to start analysis of sector country wise. Before performing the analysis the data frame was filtered out with the budget constraint of the client. Then we found out top performing sectors for each country. Which made out picture quite clear.

12. Later we also found the amount of money raised by the top 3 countries in which Spark Funds would be interested in.

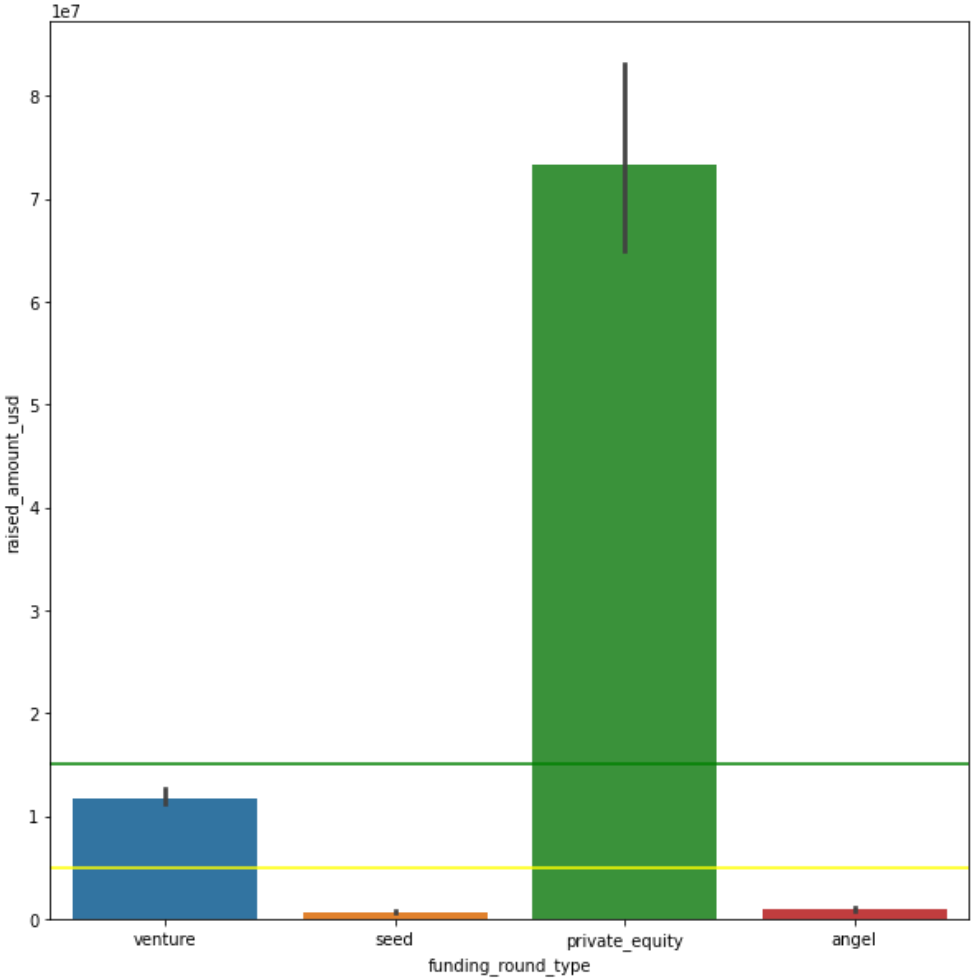
13. Out of the total investments made, we filtered out the sector which was highest among the amount raised.

14. Lastly, The top sectors, top 9 countries and many other observations were visualised which will be acknowledged in the upcoming slides.

# <Results>

Plot 1

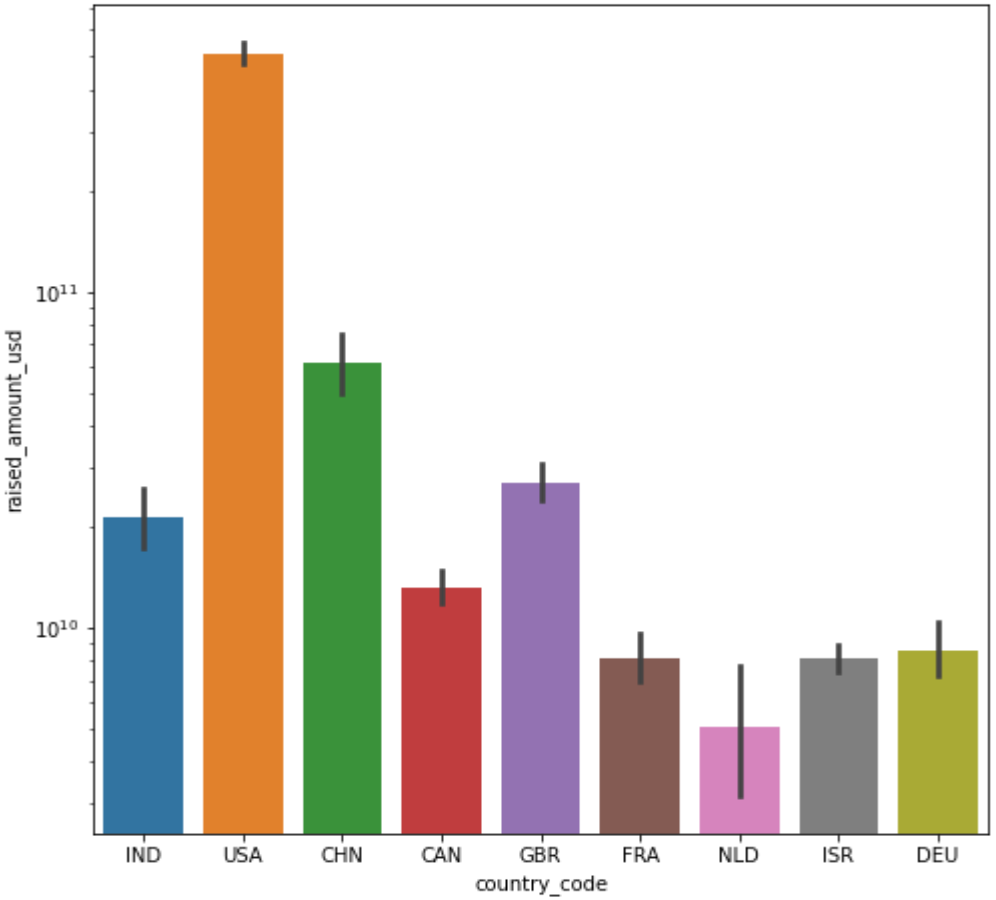
- The top notch here is clearly of private equity, but surprisingly it won't be considered for investment as it is out of the range. Our objective was to find funding type between 5M to 15M USD.
- Those two lines green and yellow, represents the upper limit of the investment and lower limit of the investment respectively.
- Venture becomes our funding type of interest as it is within the range and has in raised amount greater than other two types, i.e, seed and angel.
- The bar graphs is plotted with raised amount in USD on the Y-axis, and funding type on the X-axis.



# <Results>

Plot 2

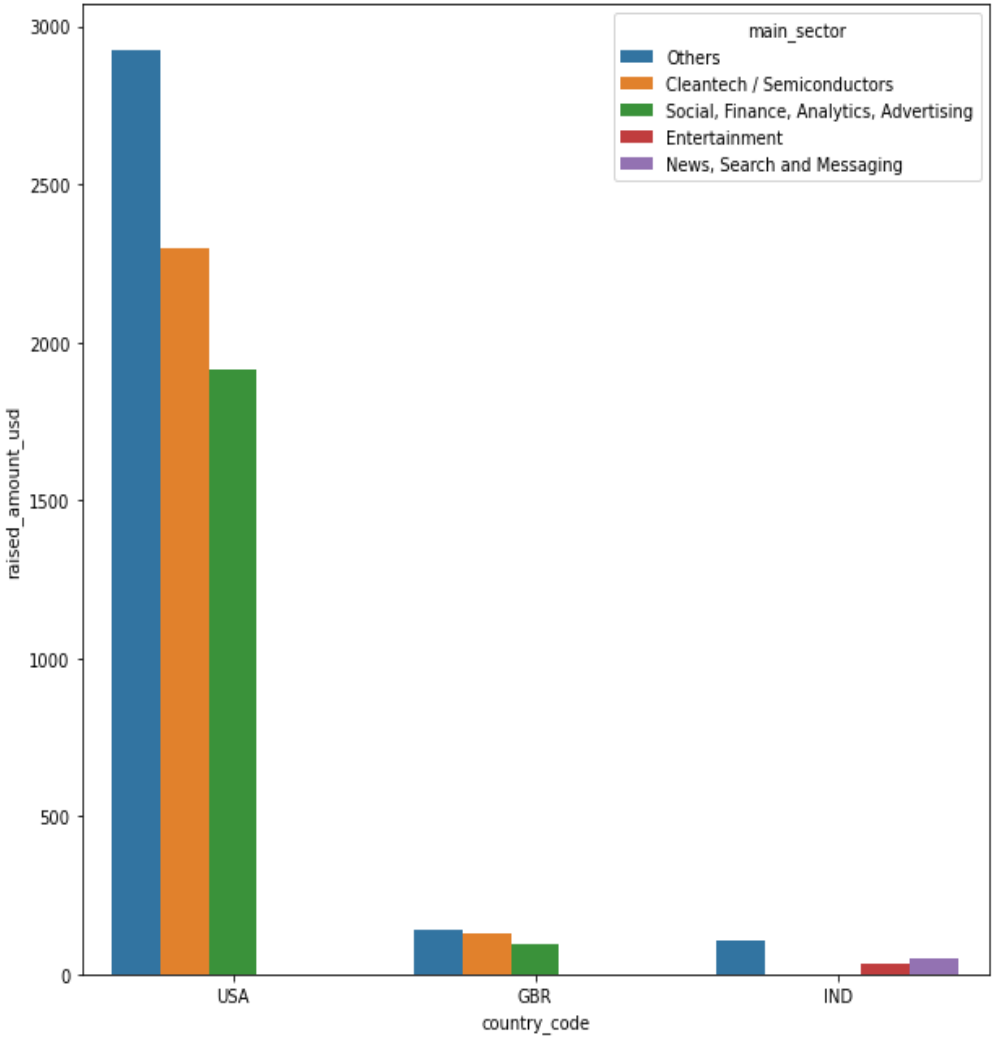
- On the X-axis there are countries, and on the Y-axis there is the amount of USD raised by the countries. Here we have taken log on the Y-axis to enlarge the small differences.
- Clearly, USA has raised highest amount of USD. Followed by China, GBR, India, Canada and so on.
- Here we will omit China in our analysis, as spark funds only wants to invest in English speaking countries.
- After removing China the top order get's slightly different, Still USA would remain on top, followed by GBR and India.



# <Results>

Plot 3

- Here on the X-axis countries are represented. And on the Y-axis amount raised in USD.
- The thing to note here is we clubbed the top sectors of each country in the bar plot.
- One can easily identify the top performing sectors of each country and find insights accordingly.
- Others is top performing sector in all the countries if we observe carefully.
- As we've already seen that USA is having highest amount of USD raised, we can now further see the top performing sectors, which has generated a certain amount of USD.



# <Conclusions>

Based on the above slides we can have an clear analysis as in which funding type to be invested, and also identify the top performing sectors. This analysis can be easily visualised using a tabular structure, as follows:

| Country | Main sector | Number of investments in top sector |
|---|---|---|
| USA | Others | 64300000 |
| | Cleantech / Semiconductors | 75300000 |
| | Social, Finance, Analytics, Advertising | 61800000 |
| GBR | Others | 37000000 |
| | Cleantech / Semiconductors | 35600000 |
| | Social, Finance, Analytics, Advertising | 37500000 |
| IND | Others | 39000000 |
| | News, Search and Messaging | 33000000 |
| | Entertainment | 21000000 |