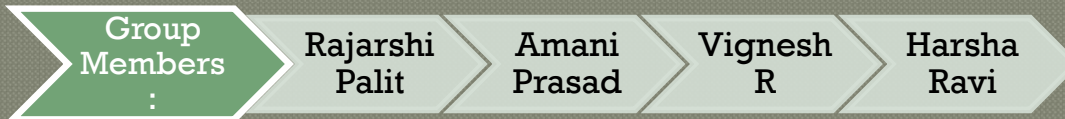# *INVESTMENT CASE STUDY*

SUBMISSION REPORT

- To help Spark Funds, an asset management company, understand the global trends in investments so that it can take investment decisions effectively, while satisfying its constraints.

Group Members :	Rajarshi Palit	Amani Prasad	Vignesh R	Harsha Ravi

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 investors are investing'.

The objective is thus classified into the following sub-goals:

Investment type analysis:

Comparing the typical investment amounts in the venture, seed, angel, private equity etc. so that Spark Funds can choose the type that is best suited for their strategy.
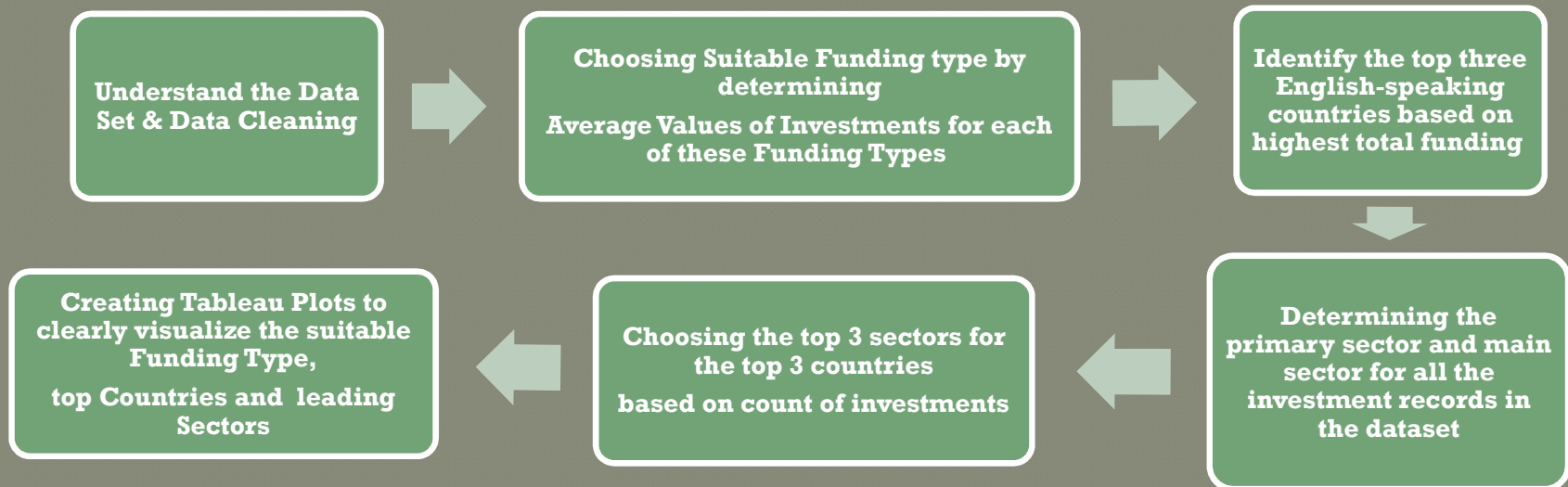
Country analysis:

Identifying the countries which have been the most heavily invested in the past. These will be Spark Funds' favorites as well.

Sector analysis:

Understanding the distribution of investments across the eight main sectors.

Problem Solving Methodology

- The approach for this project has been to divide the entire case study into various checkpoints to meet each of the sub-goals and give more meaning to the analysis. The checkpoints are represented in a sequential flow as below:

Understand the Data Set & Data Cleaning

→

Choosing Suitable Funding type by determining
Average Values of Investments for each of these Funding Types

→

Identify the top three English-speaking countries based on highest total funding

↓

Creating Tableau Plots to clearly visualize the suitable Funding Type,
top Countries and leading Sectors

←

Choosing the top 3 sectors for the top 3 countries
based on count of investments

←

Determining the primary sector and main sector for all the investment records in the dataset

The following data cleansing processes were applied to make the data dependable so that it can provide significant business value by improving Decision Making Process:

Handling Encoding Issues

- Some of the columns from the master_frame dataframe contain special characters due encoding issues. They needed to be handled since deleting the column would have led to important data loss.
- **str.encode('utf-8')** and **str.decode('ascii', 'ignore')** functions were used to first encode the dataset using codec 'utf-8' and finally decoding it using 'ascii' and ignoring errors.

Dropping Null values in Rows & Columns

- The **null values in some columns** that aren't required for analysis have been dropped from the master_dataframe
- High Percentage of Null values in columns like 'raised_amount_usd', 'country_code' and 'category_list' doesnt make any sense or help in analysis. Hence **all rows containing null values** for these columns were dropped from master_frame.

Handling anomalies in Main Sector values

- There are some sectors in the category_list column of mapping database where the string 'na' is replaced by '0'. E.g. '**A0lytics**'. Contrarily, there are instances where '0' shouldn't be replaced by 'na'. E.g. '**Enterprise 2.0**'.
- Used **re.compile** to generate a pattern to replace the '0' character with 'na' where ever needed

The various funding types such as seed, venture, angel, etc. depend on the type of the company (startup, corporate, etc.), its stage (early stage startup, funded startup, etc.), the amount of funding (a few million USD to a billion USD), and so on. The aim of this analysis is to choose one among the four popular investment types for each of Spark Funds' potential investments.

Calculating the average investment amount for each among the four popular funding types (venture, angel, seed, and private equity)

- Grouping the dataset by 'funding_type' column, the mean funding amount in USD was determined for the four funding types - venture, angel, seed, and private equity.
- Comparing the results with Spark Funds' funding constraint (investing between **5 to 15 million USD** per round of investment), **'Venture'** was found to be the most suitable funding type.

Calculating the fraction of total investments (globally) in venture, seed and private equity

- Grouping the dataset by 'funding_type' column, the fractional count of investments was determined for the funding types - venture, seed, and private equity.
- Comparing the results with Spark Funds' funding constraint (**wanting to invest where most other investors are investing**), **'Venture'** was found to be the most suitable funding type.

Funding Type Analysis: Visualization
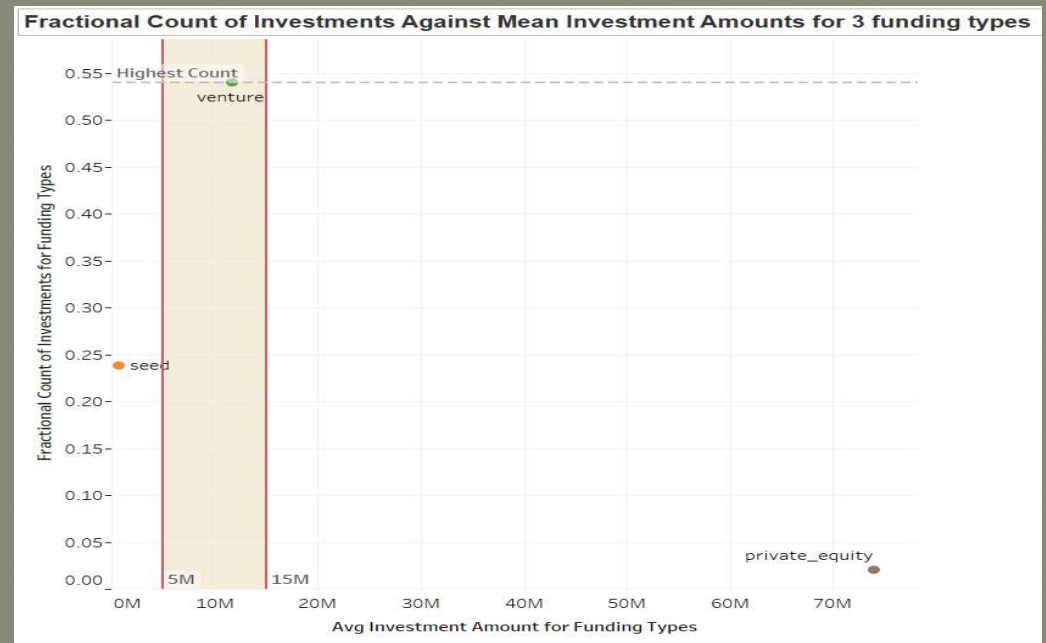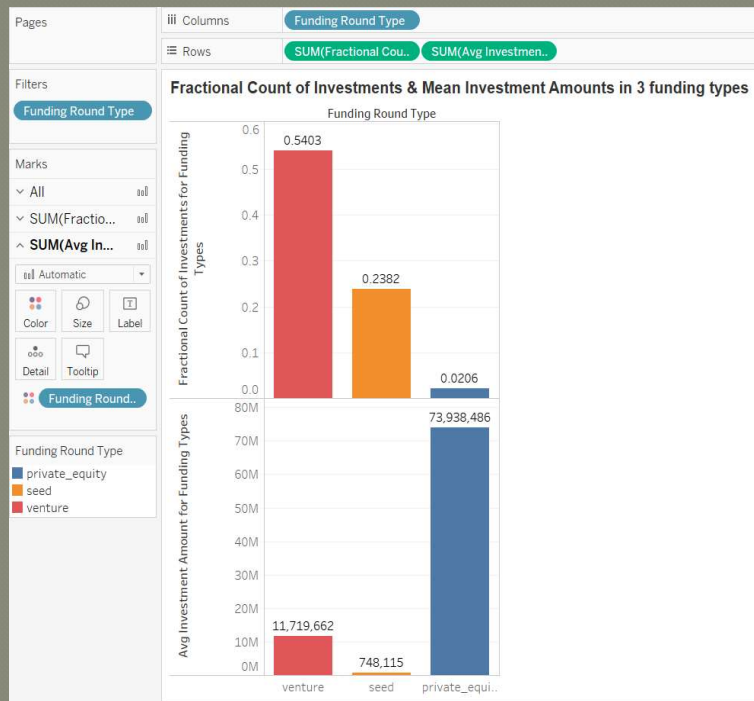
- These 2 plots were created in Tableau.
- The plots were created for only 3 funding_types – venture, seed and private_equity.
- The fractional investment counts and Average investment amounts were created as calculated fields using FIXED functions as in the pictures beside and plotted as rows against Funding Round Type.
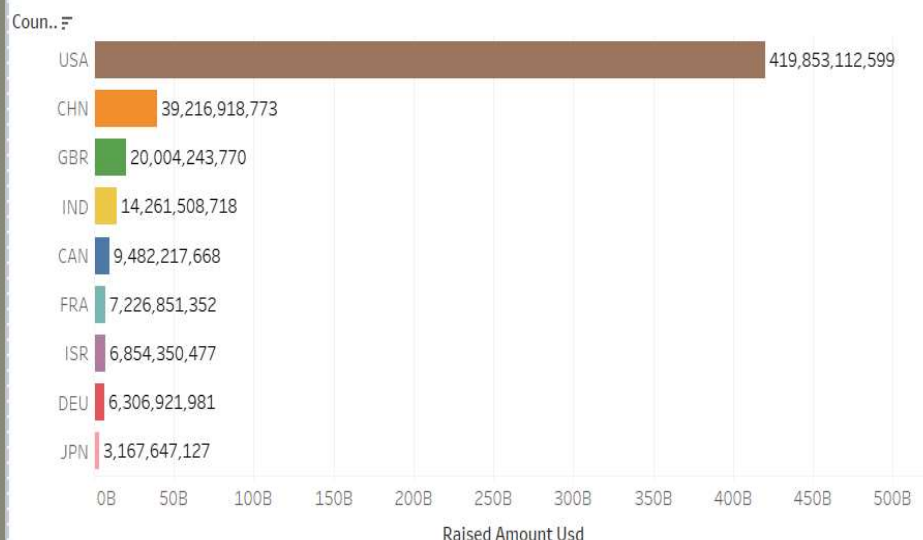
The aim of this analysis is to identify countries with the highest amount of funding for the chosen investment type - 'Venture'. This is a part of its broader strategy to invest where most investments are occurring.

Tableau plot was created to visually identify the top 9 countries receiving highest total funding for funding type – 'Venture'

Finding top 9 countries with highest funding (for investment type 'Venture') and identifying the top 3 English speaking countries from the list.

- Grouping the dataset by 'country_code' column, the sum of funding amount in USD was determined for the **top 9** countries.
- Comparing the results with Spark Funds' funding constraint (investing only in English-speaking countries because of the ease of communication with them), **'USA', 'GBR'** & **'IND'** were found to be the top 3 suitable countries. (**'CHN'** isn't an english-speaking country.)

Top 9 countries based on highest total amount of investments for funding type 'venture'

Coun.. ⌐

Country	Raised Amount Usd
USA	419,853,112,599
CHN	39,216,918,773
GBR	20,004,243,770
IND	14,261,508,718
CAN	9,482,217,668
FRA	7,226,851,352
ISR	6,854,350,477
DEU	6,306,921,981
JPN	3,167,647,127

Raised Amount Usd

The aim of this analysis is to identify the most profitable sectors by understanding the distribution of investments across the eight main sectors.

Identifying the 'primary sector' and 'main sector' of each investment record from the master_frame dataframe

- Extracted the **'primary sector' of each investment** from the category_list column(1st category) in master_frame dataframe.
- The mapping file was transposed to read the column header (main sector name) based on row values (where value=1) for each of the category list values in each row.
- Imported the transformed mapping file into a dataframe and map each primary sector to one of the eight **main sectors**.

Creating 3 dataframes for each of the top 3 countries for funding type 'Venture' and investment range of 5-15 M USD and determining the best sectors in each

- Created 3 dataframes -- **'D1', 'D2'** & **'D3'** -- corresponding to the top 3 countries **'USA', 'GBR'** & **'IND'**, by filtering the master_frame based on respective country_code, funding amount limit – 5 to 15 mil USD and funding type 'venture'.
- Using the dataframes, determined the **top 3 sectors** and the amount of investments in each of them for all the 3 countries.
- Identified the **country receiving highest** investment amount in the top 2 sectors of each of the top 3 countries.

Findings	USA	GBR	IND
Top Sector name	Others	Others	Others
Second Sector name	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising	Social, Finance, Analytics, Advertising
Third Sector name	Cleantech / Semiconductors	Cleantech / Semiconductors	News, Search and Messaging

Results: Top 3 sectors in the top 3 countries
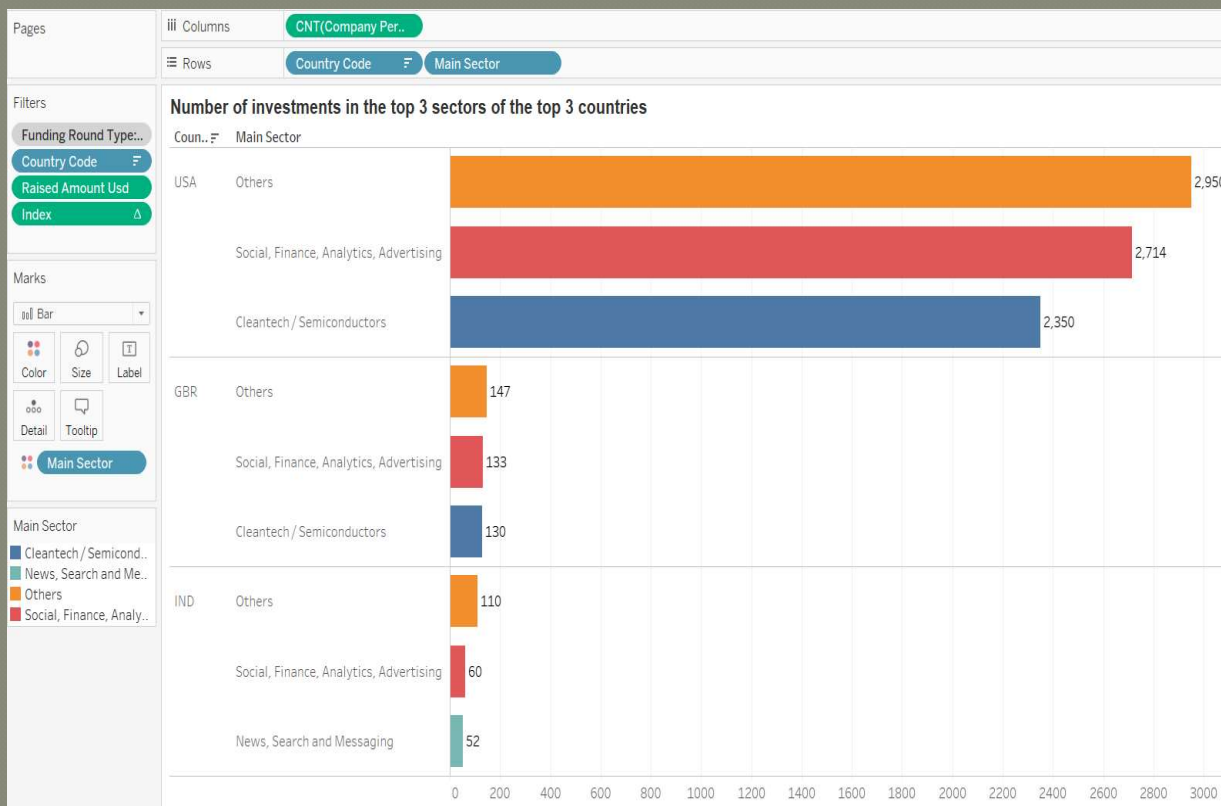
• This plot was created in Tableau.

• The plot shows the Top 3 Sectors in each of the Top 3 Countries and their count of investments

• Following filters have been used to extract required data:

1. **Funding Round Type**(with apply to context option) = '**Venture**'
2. **Country Code** in '**USA**', '**GBR**' & '**IND**'
3. **Raised Amount USD** between 5 to 15 mil USD &
4. Calculated Field with **Index()** function to select top 3 sector in each of top 3 countries.