

# INVESTMENT ASSIGNMENT

## SUBMISSION

**Name:** Kishan

**Batch :** PG Diploma in Machine Learning and AI June-1 2020

**ID :** APFE20N00712

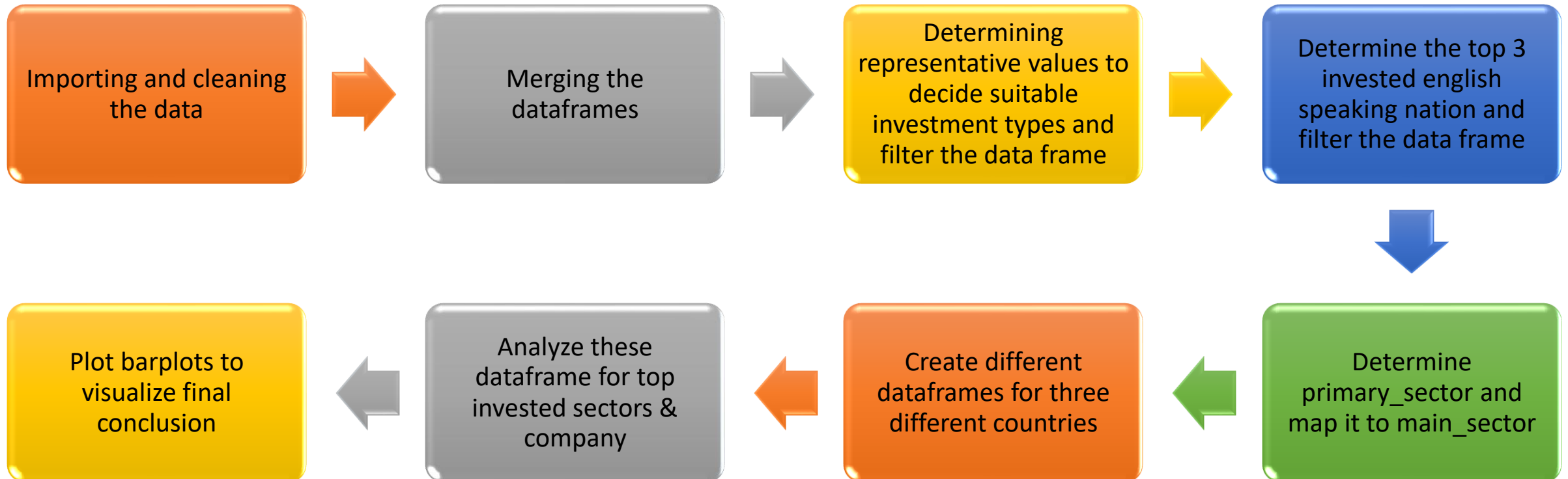
## Abstract

Spark Funds is an Asset management company. Spark Funds wants to make investments in a few companies. The CEO of Spark Funds wants to understand the global trends in investments so that she can take the investment decisions effectively.

There are two minor constraints. The budget is 5-15 million and the CEO only wants to invest in English speaking nation. As Spark funds is an early stage startup investor, it is preferred to invest where others are investing. This analysis main objective is to determine the Investment type, countries and sectors where most of the investment are being done.

For this analysis a sequential approach is followed in which first step is to decide the investment type having the highest investment followed by finding the top 3 English speaking nation with maximum investment amount followed by the various sector in which investment can be done. There is also visual depiction of the data analysis in the form of barplots for a better understanding of the scenario. There are also some assumptions made during the analysis which are mentioned in the code.

Finally it is concluded that which sector, investment type, countries and company will be a better option for Spark Funds



## 1. Import and Cleaning the data:

- ✓ Import companies.csv and rounds2.csv
- ✓ Clean the data using encoding and decoding to remove non ascii characters

## 2. Merging the Dataframe

- ✓ Merging companies and rounds as master\_frame
- ✓ Dropping columns not required for analysis like 'homepage\_url','funding\_round\_permalink','permalink','state\_code' etc.
- ✓ Assumption : Sparks fund will not invest in the closed business entities or sector so the rows with status close is removed.

## 3. Determining representative values to decide suitable investment types and filter the data frame:

- ✓ Representative value for four type of investment type is determined i.e; seed, angle, venture and private\_equity
- ✓ Determining the representative value using boxplot and .describe(). There were many outliers hence median will be a better representative value.
- ✓ To determine the average value outliers are removed and mean is calculated.
- ✓ Assumption: The outliers are assumed to be  $> Q3 + 1.5 * IQR$
- ✓ Most suitable type is venture, so master\_frame is filtered to keep data of only venture type.

## Analysis

4. Determine the top 3 invested english speaking nation and filter the data frame :
  - ✓ Create a dataframe top9 containing the list of top 9 heavily invested countries.
  - ✓ Select the top 3 English speaking nation using the Wikipedia list of English speaking nation,
  - ✓ Filter the dataframe to contain the data of only these three countries.
5. Determine primary\_sector and map it to main\_sector:
  - ✓ Create a column primary\_sector as per the given rule.
  - ✓ Import mapping.csv and convert wide format to long format and map primary\_sector to its main\_sector
6. Create different dataframes for three different countries:
  - ✓ Create dataframe D1, D2 and D3 for countries USA, GBR and IND.
  - ✓ Filter for investment limit 5-15 Millions and investment type venture.

## Analysis

### 7. Analyze these dataframe for top invested sectors & company:

- ✓ Use these dataframe to answer all the questions asked. Like top sectors, Most invested company, Top investment count sector
- ✓ The most invested company in USA is Virtustream. GBD is Electric Cloud and IND is FirstCry.com

### 8. Plot barplots to visualize final conclusion:

- ✓ A plot showing the fraction of total investments (globally) in angel, venture, seed, and private equity, and the average amount of investment in each funding type.(Plot 1)
- ✓ A plot showing the top 9 countries against the total amount of investments of funding type Venture. This should make the top 3 countries (USA, GBR, and IND) very clear.(Plot 2)
- ✓ 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).(Plot 3)

# Results

Plot 1

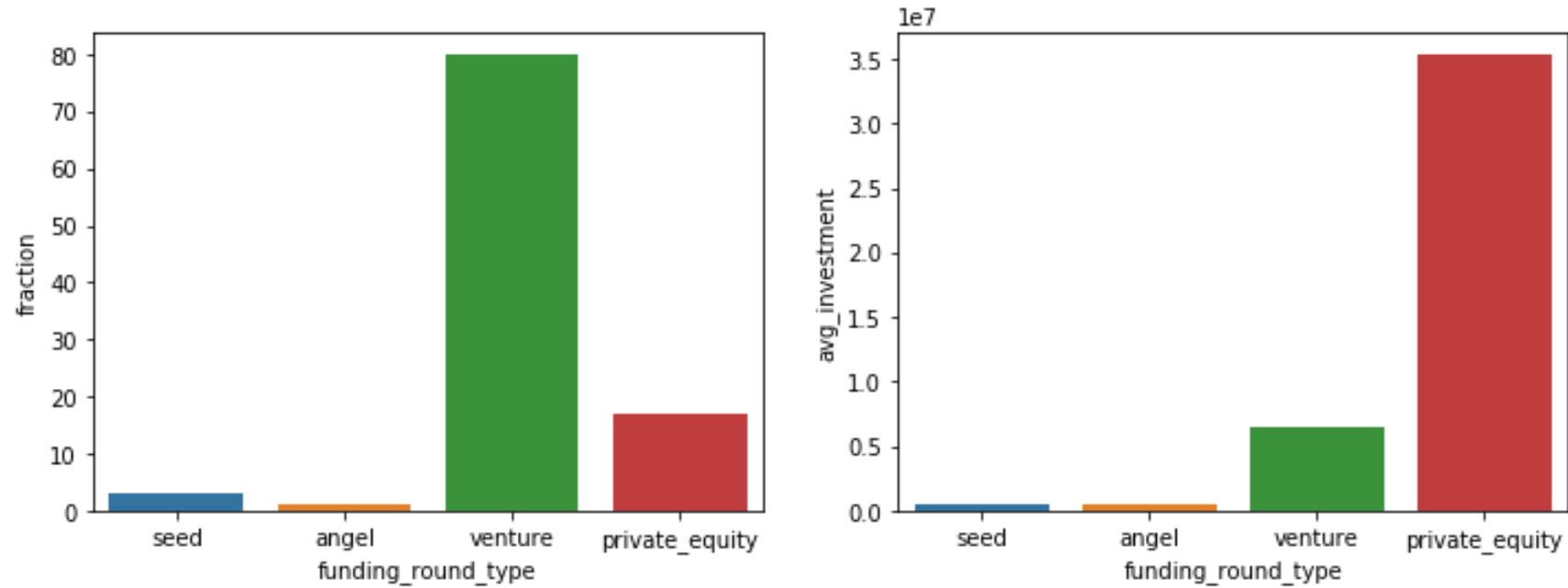


Fig 1: Barplot showing fraction of investment amount of funding type and average funding values

# Results

Plot 2

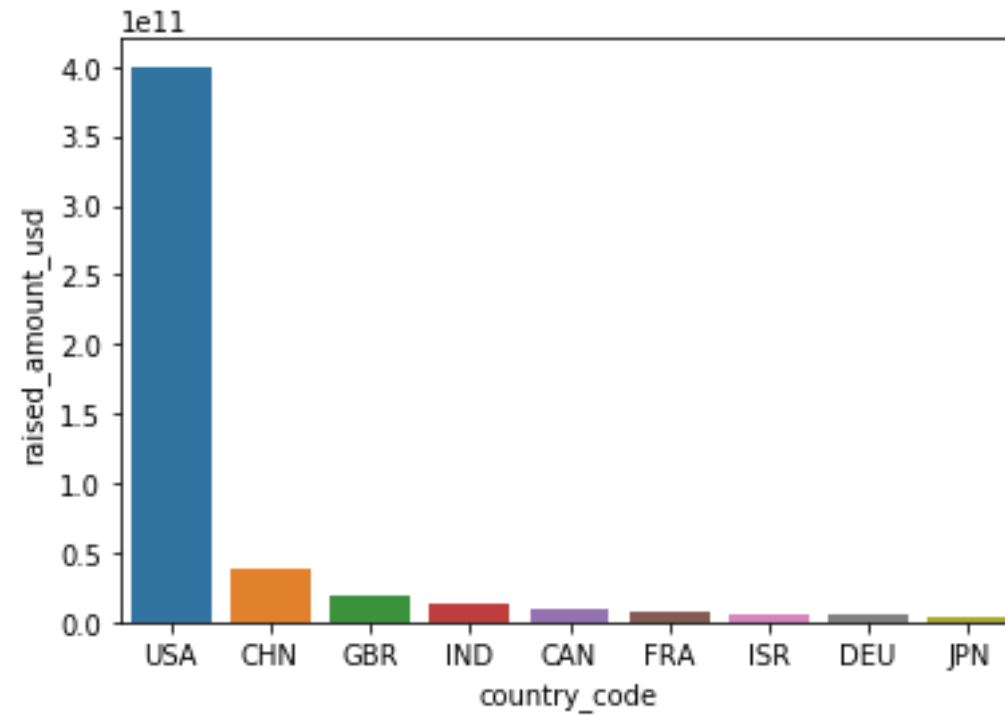


Fig 2: Barplot showing amount of US Dollars raised in various countries



## Results

Plot 3

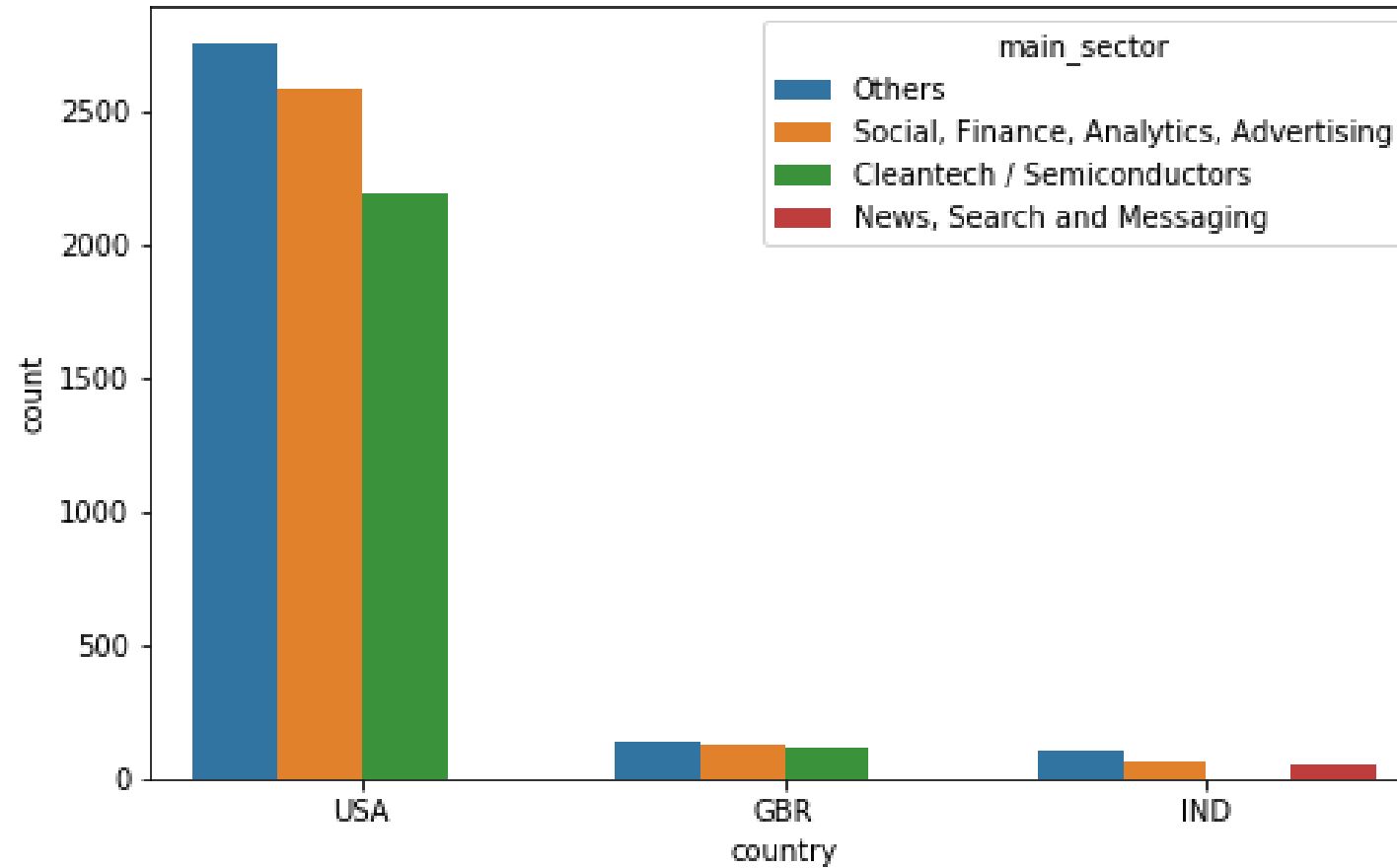


Fig 3: Barplot showing Investment count in top 3 sectors in USA, GBR and IND

## Conclusions

- The most suitable investment type having the maximum fraction of investment and average value lying in 5-15 Millions USD is **Venture**. (Plot 1)
- Top 3 most invested English speaking nation is **USA(United States)**, **GBR(United Kingdom)** and **IND(India)**. (Plot 2)
- The top 3 investment sectors of USA, GBR and IND are :  
USA :- Others; Social, Finance, Analytics, Advertising; Cleantech / Semiconductors  
GBR :- Others; Social, Finance, Analytics, Advertising; Cleantech / Semiconductors  
IND :- Others; Social, Finance, Analytics, Advertising; News, Search and Messaging

**Sparks Fund can use this analysis to plan their investment.**