

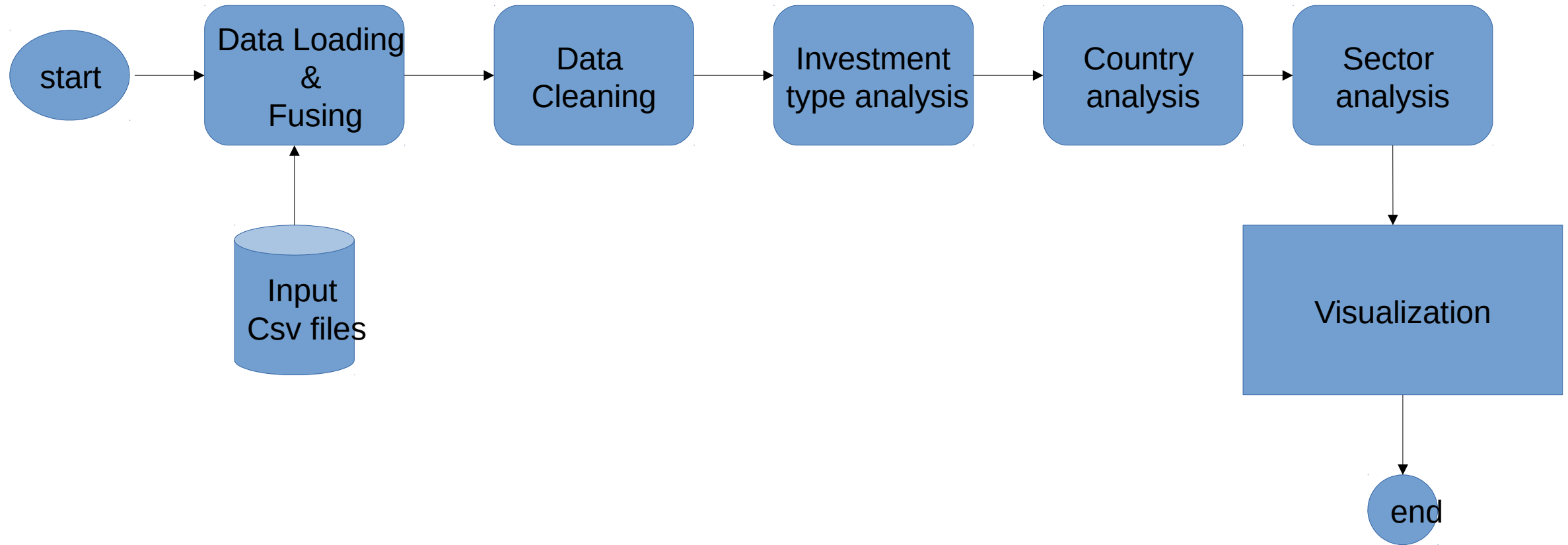
# INVESTMENT ASSIGNMENT SUBMISSION

Name: Krishna Kumar

## Objective

- The CEO of Spark Funds wants to understand the global trends in investments so that she can take the investment decisions effectively.
- Spark Funds has two minor constraints for investments:
  - It wants to invest between 5 to 15 million USD per round of investment.
  - It wants to invest only in English-speaking countries because of the ease of communication with the companies it would invest in.
- Spark Funds has a strategy for investments:
  - Spark Funds wants to invest where most other investors are investing. This pattern is often observed among early stage startup investors.
- 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'.

## Process Flow



# Data Pre-Processing

- This stage includes data loading, fusing and cleaning.
- The loaded input files (companies.csv, rounds2.csv and mapping.csv) shall be fused into single data frame during processing.
- Data cleaning is performed by dropping unwanted columns like
  - **funding\_round\_code**
  - **founded\_at**
  - **homepage\_url**
  - **state\_code**
  - **City**
  - **region**
- Data rows with less than 20% NULL valued cells in the essential columns can be removed.
  - **raised\_amount\_usd with 17% NULLs**
  - **category\_list with 1% NULLs**
- Impute values to cells with NULL in essential columns like
  - **country\_code with 'UNK' for unknown**

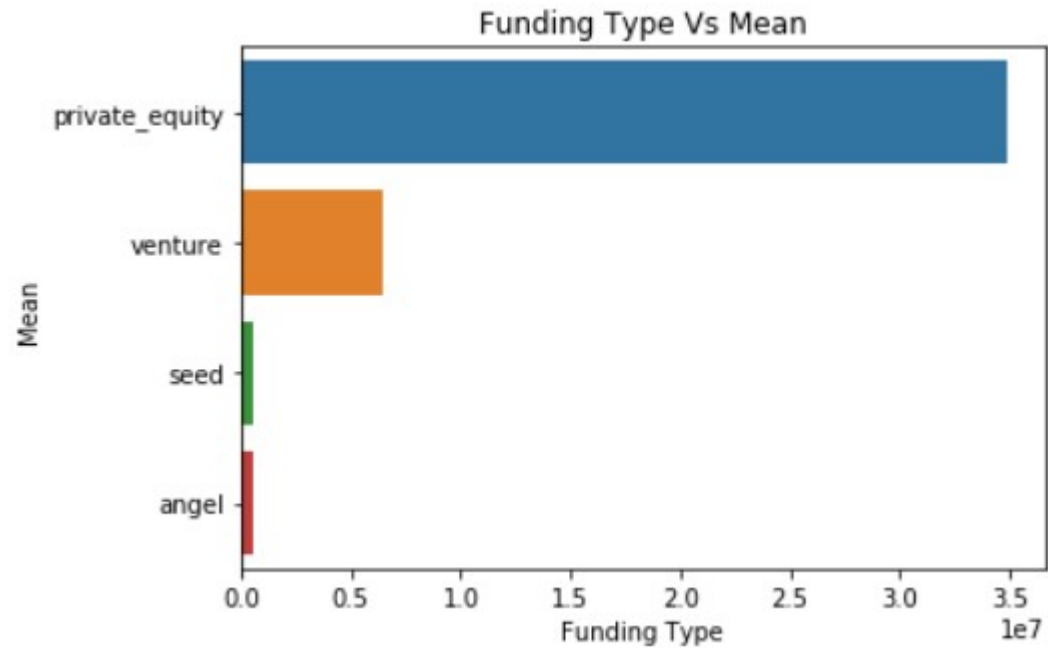
## Exploratory Data Analysis

- After Data pre-processing stage, we start EDA with 3 objective columns from master data frame and 1 column merged to it from sector data.
  - **funding\_round\_type**
  - **raised\_amount\_usd**
  - **category\_list**
  - **main\_sector**
- **‘raised\_amount\_usd’** is grouped for each funding type and outliers are removed to analyze the mean value of amount.
- Based on this a suitable ‘Funding type’ is chosen (**Venture**)
- Based on chosen FT grouping is performed for different countries and sorted to find the top investment countries with the constraint on range of investment amount.

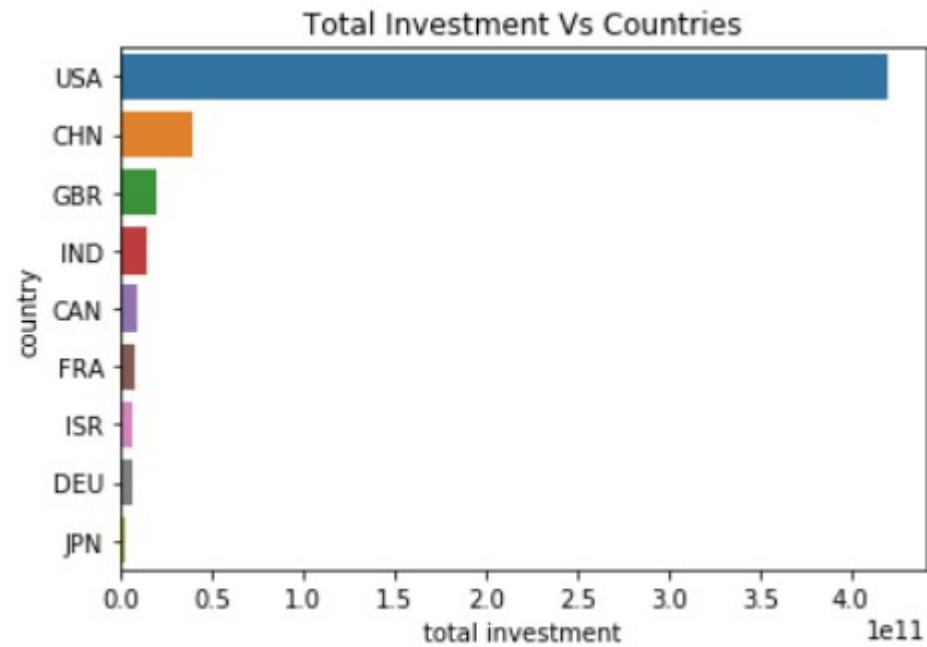
## Exploratory Data Analysis

- Based on chosen countries with language constraints, investment amount constraint and FT constraint the major sectors of investments are analyzed.
- Various grouping for top 3 countries based on highly invested sum and counts are performed for overall and major sectors.
- Finally visualization of results to present to CEO is completed.

# Funding Type Vs Mean Plot

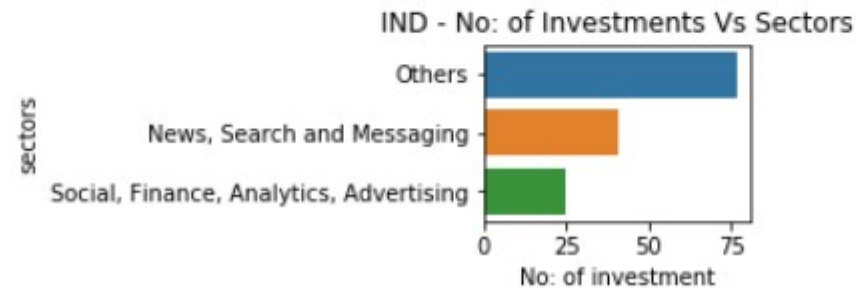
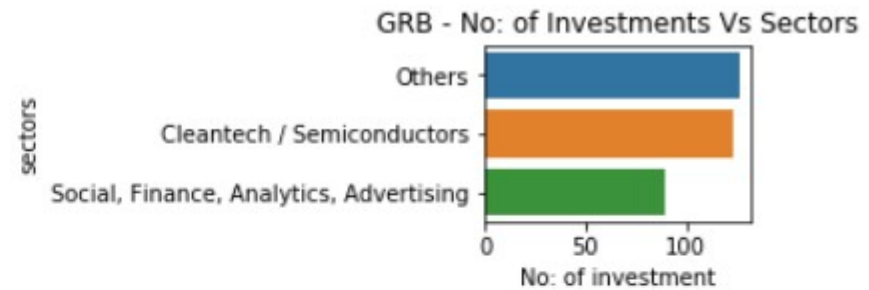
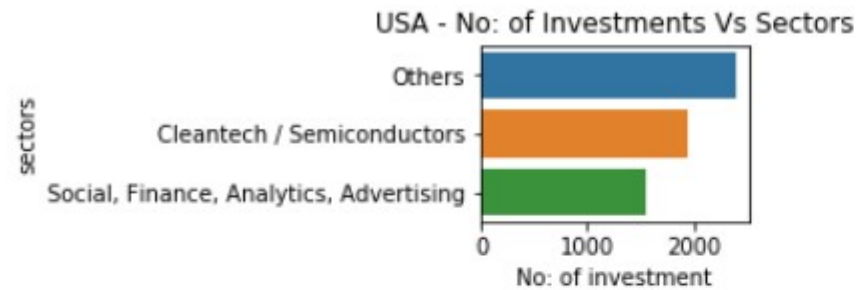


# Total Investment Vs Countries





# Number of Investments Vs Sectors for top 3 Countries



## Final Note

- Based on analysis, Spark Funds can invest in countries like
  - USA
  - Britain
  - India
- It can Invest into
  - Venture funding type
- It shall consider companies working on top 3 sectors of respective countries.