INVESTMENT ASSIGNMENT

SUBMISSION

Name:

<Abstract>

The motive of this analysis:

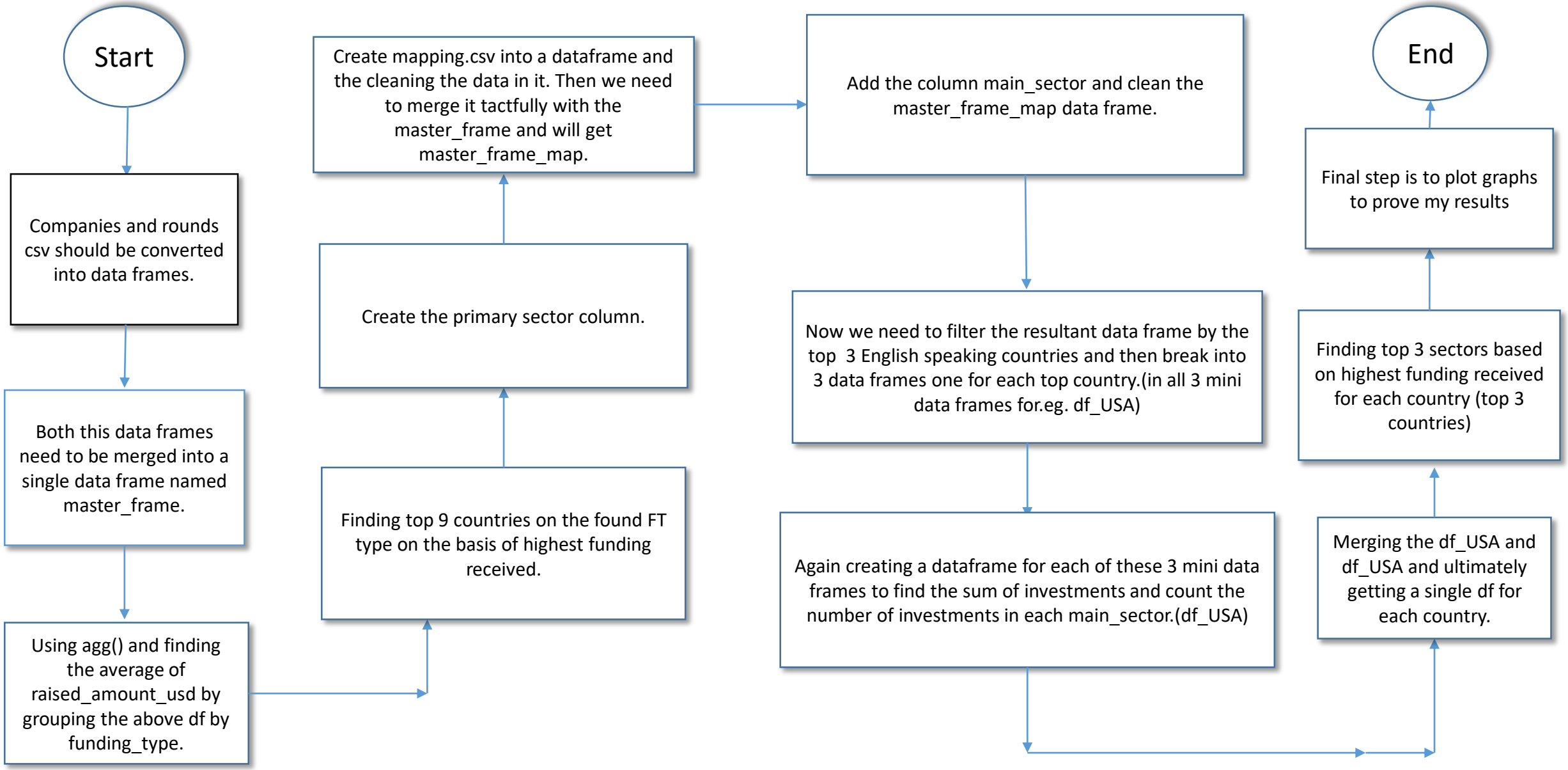
- Finding the FT type in which we should be investing given that the average of raised_amount_usd using group by funding_round_type lies between 5-15 million USD.

- Finding the Top 3 countries on the basis of highest funding received so that we can invest and also it should be an English speaking(official language) country of course.

- Finding the top three countries for each of the three top English speaking countries (based on the highest number of investment received) where we would like to invest.

- Proving the above details by plotting graphs.

<Problem solving methodology>



<Analysis>

- Firstly the given data files companies.csv and rounds.csv was converted into respective data frames.
- Then found the primary key of companies.csv(permalink) and foreign key of rounds.csv(company_permalink) and then both the data frame was merged into a single dataframe named master_frame.
- Found the average of the column raised_amount_usd column using group by funding_round_type to find for which FT my condition i.e. investment amount should be between 5-15 million USD satisfies(as mentioned by Sparks Fund).
- Then proceeded with finding top 9 countries with highest funding received.
- Then chose all the first values of the category_list column by using list comprehension in the master_frame and created a column named primary sector in master_frame.
- Converted the mapping.csv file into a data frame and removed the unwanted data from it i.e. the blank row(1st row and the Blank Column) so that it now had 8 main sectors as mentioned and then merged it with the master_frame which was named as master_frame_map.

.

<Analysis>

- Then it was required to add a column named main_sector for that dropped all the binary columns and for each row, the column which had 1 in it, it's column name was put in the main_sector column for that particular row.
- Filtered the above data frame only for the top 3 English speaking countries whose names we already knew and then broke them into 3 different data frames each for a top English speaking country. The data frames were named as df_"country name".
- Found the sum and count of respective main_sector for each of the 3 countries and named them as df_"country name"_final.
- Merged df_"country name" and df_"country name"_final data frames with the common column as main_sector.
- Then found the top 3 sectors for each of the top 3 countries based on the number of investment received and also found the top 2 companies for these corresponding sectors which received highest funding.
- At last plotted 3 graphs.

<Analysis>

- First graph is the plot of "The fraction of total investments (globally) in angel, venture, seed, and private equity," vs "The average amount of investment in each funding type" which gave us the funding type in which Sparks Funds can invest.

- Second graph is the plot of "The top 9 countries" vs "The total amount of investments of funding type FT" which gave us the top 3 countries in which Sparks Funds should invest.

- Third graph is the plot of "the number of investments in the top 3 sectors" vs "Top 3 countries" for the suitable funding type FT which displays the top 3 sectors for each country in which Sparks Funds should invest.

<Results>

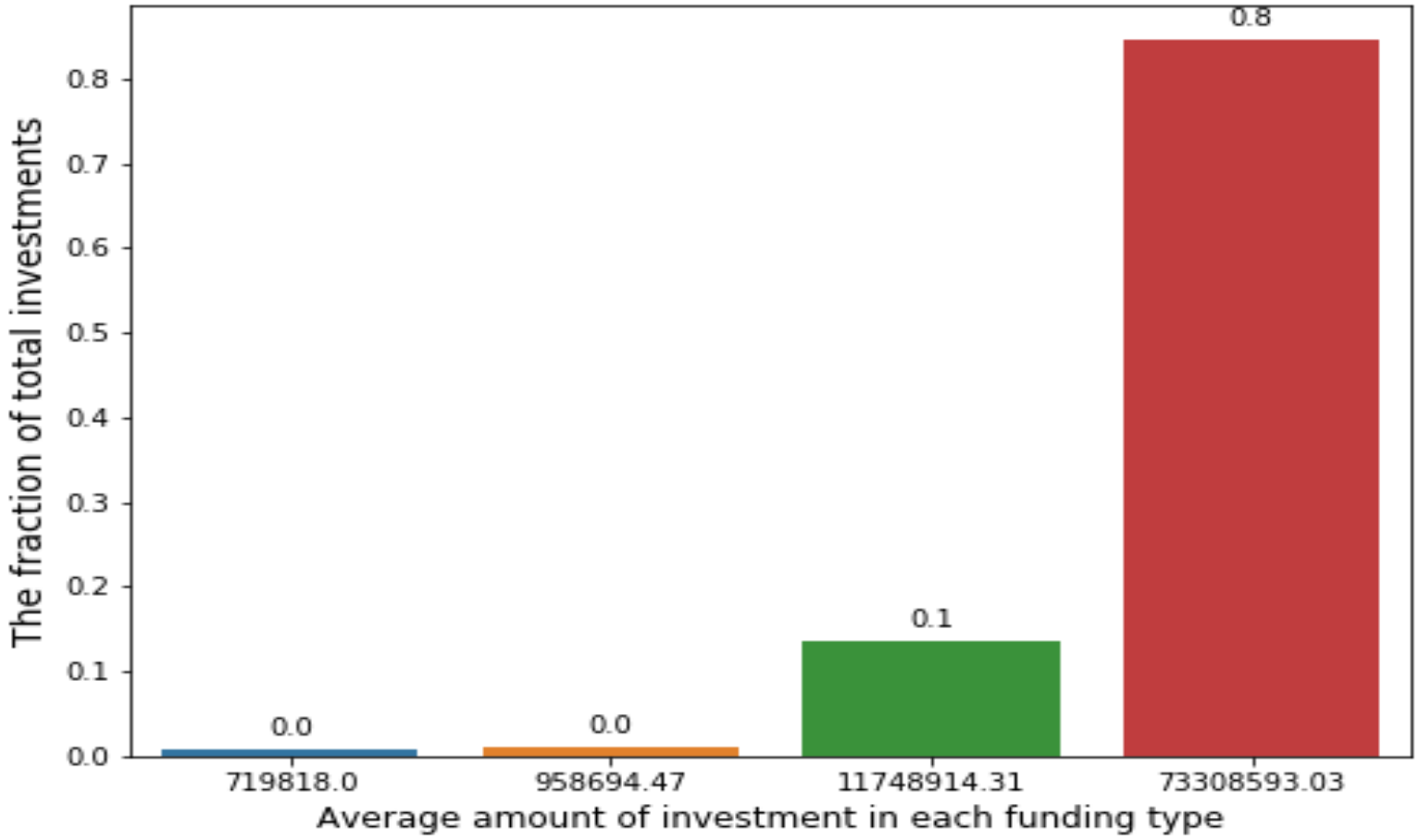
Plot 1

Blue bar: private_equity

Yellow bar: angel

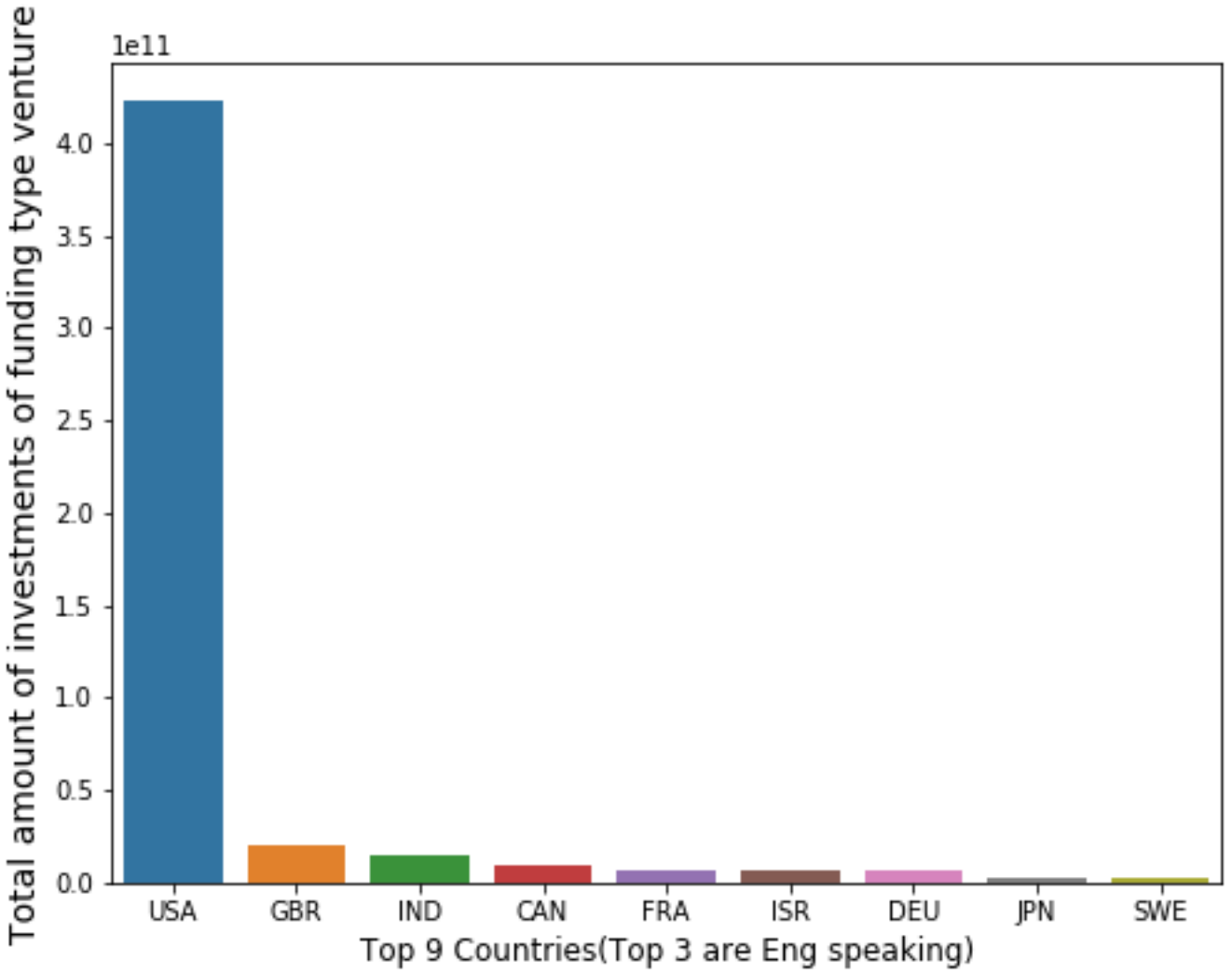
Green bar: seed

Red bar: venture



<Results>

Plot 2



<Results>

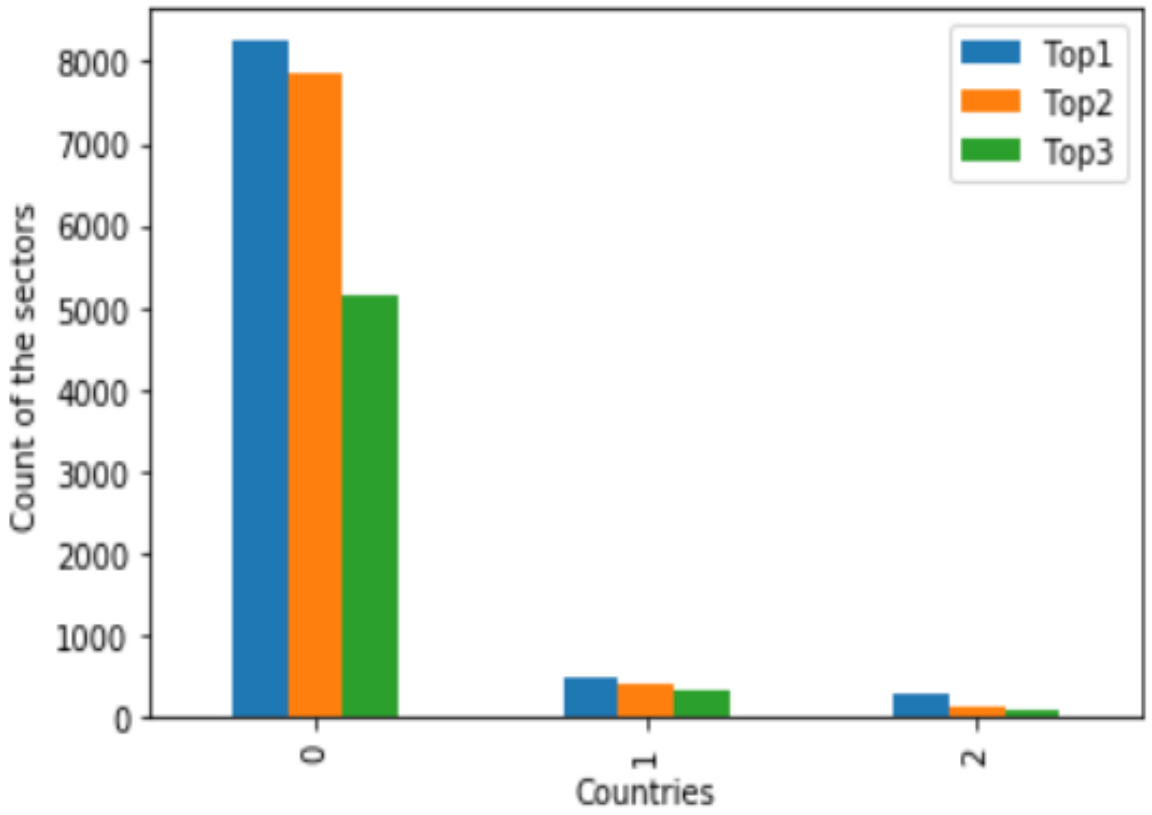
Plot 3

X axis labels:

0 is USA

1 is GBR

2 is IND



<Conclusions>

- ✓ The funding type in which we(Spark Funds) should invest is **Venture**.

- ✓ The top 3 English speaking countries where it would be best invest are **USA** (United States of America), **GBR** (Greater Britain), **IND** (India).
- ✓ The top 3 sectors for each of the above countries are:

  - ✓ USA:
    1. Top1= **Others**
    2. Top2= **Cleantech / Semiconductors**
    3. Top3= **Social, Finance, Analytics, Advertising**
  - ✓ GBR:
    1. Top1= **Others**
    2. Top2= **Cleantech / Semiconductors**
    3. Top3= **Social, Finance, Analytics, Advertising**
  - ✓ IND:
    1. Top1= **Others**
    2. Top2= **News, Search and Messaging**
    3. Top3= **Social, Finance, Analytics, Advertising**