INVESTMENT ASSIGNMENT

SUBMISSION

Name: Shubham Mishra

<Abstract>

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’.

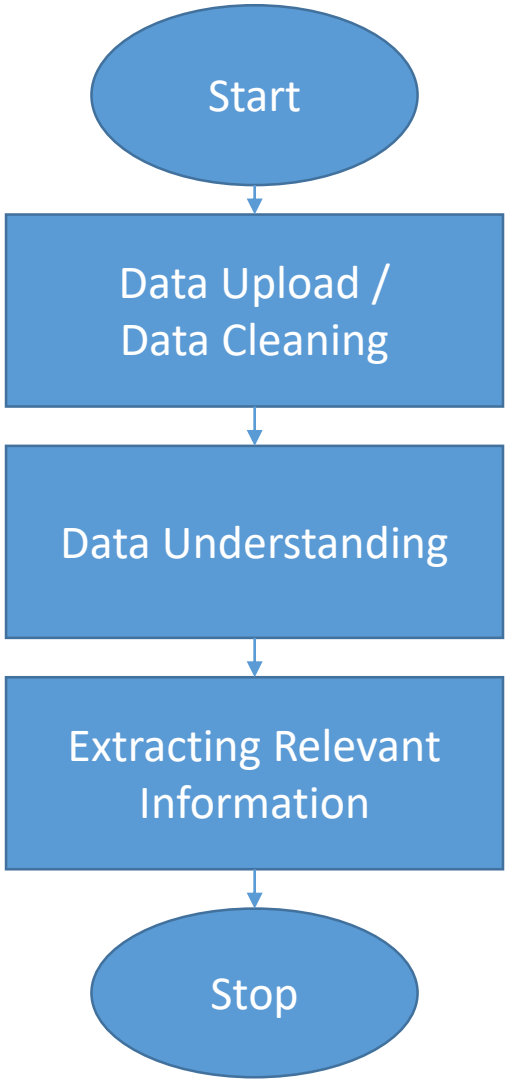
Goals :-

Goal 1 > 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.

Goal 2 > Identifying the countries which have been the most heavily invested in the past. These will be Spark Funds’ favourites as well.

Goal 3 > Understanding the distribution of investments across the eight main sectors. (Note that we are interested in the eight 'main sectors' provided in the **mapping file**.

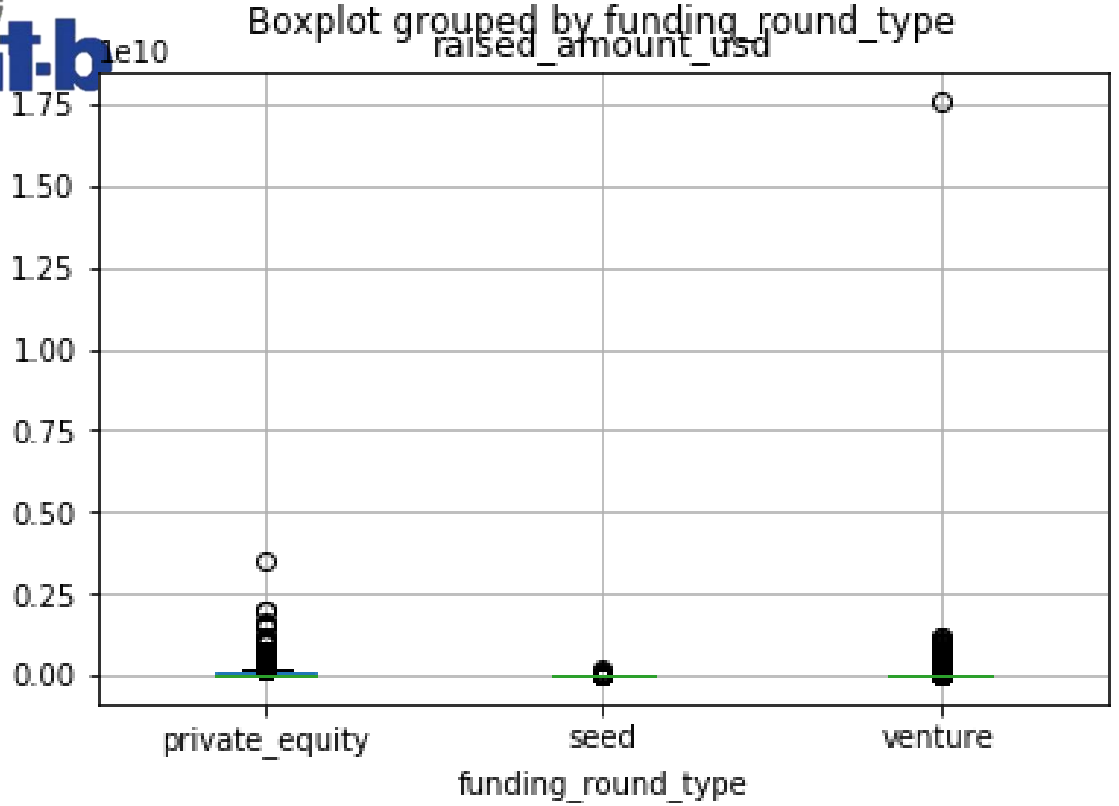
Work-Flow:


Start
Data Upload / Data Cleaning
Data Understanding
Extracting Relevant Information
Stop

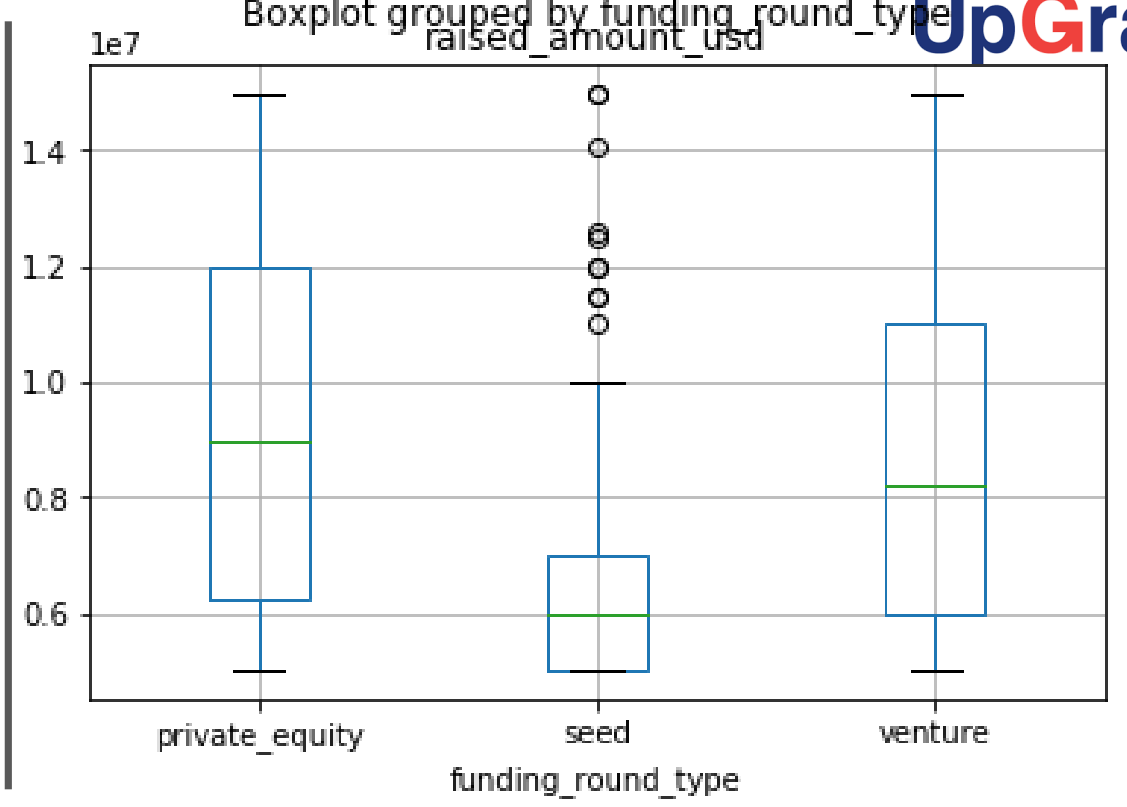
<Analysis>

Phase 1 :

- While Importing the datasets, I got to know that we needed to separate data using '\t' and the encoding type needed was –' ISO-8859-1'.
- Since Columns like "permalink" and "name" contained special characters, they needed to be removed.
- Even Case Sensitivity had to be resolved in order to merge the Datasets.
- While Checking for the Number of Nulls, I made a conclusive decision of removing all the rows having nulls except –"funding_round_code" as the amount of nulls contained by this field was huge(66789 records) and removing all those records would highly impact my dataset.
- Then I plotted a graph with funding Type – "seed", "venture" and "private_equity" vs "raised_amount_usd". Concluded that since the datasets has outliers boxplot was unable to give proper results
- Then I took the limit between 5-15 million usd and plotted the Graph.



ON TOTAL DATASET



5-15 MILLION DATASET

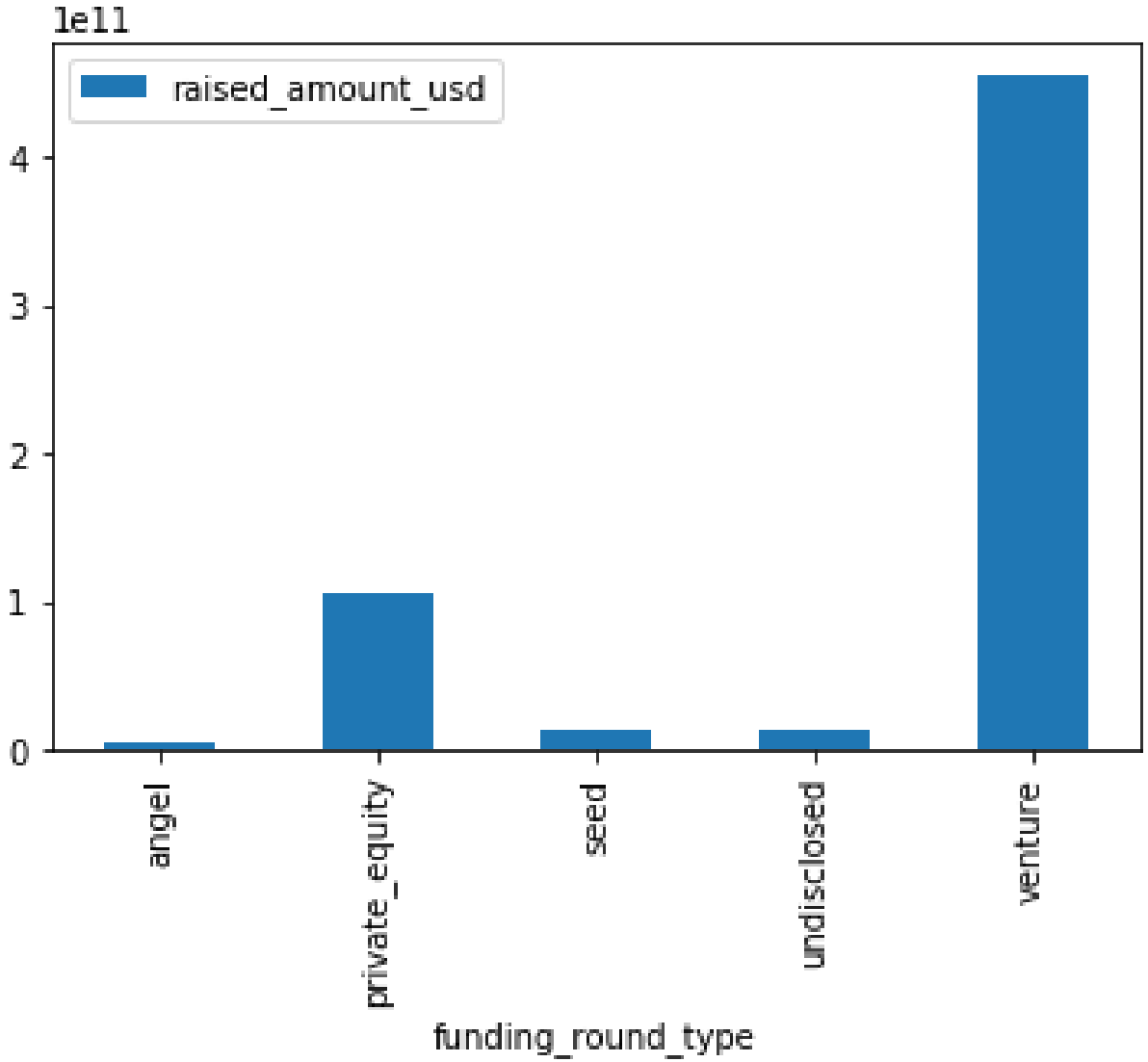
<Analysis>

Conclusion : venture & private_equity somewhat follows similar distribution.

<Analysis>

On Further Analysis:

- Checking in which funding type max amount of investment took place.
- It came out to be venture, Hence I choose venture to be the most profitable Funding Type.

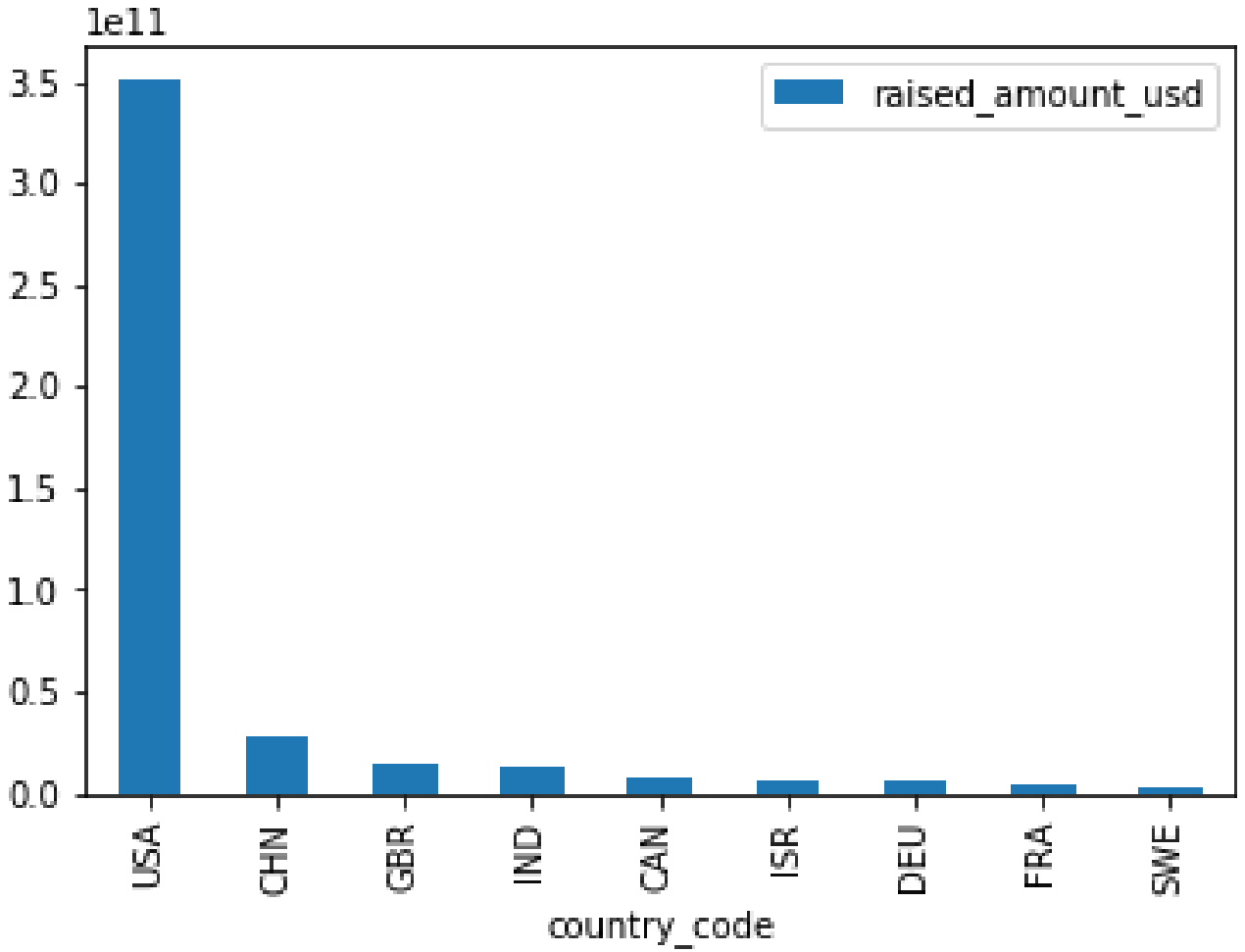


<Analysis>

Phase 2 :

- . Found out top 3 English Speaking Countries.
- 1)USA
- 2)Great Britain(GBR)
- 3) India(IND)

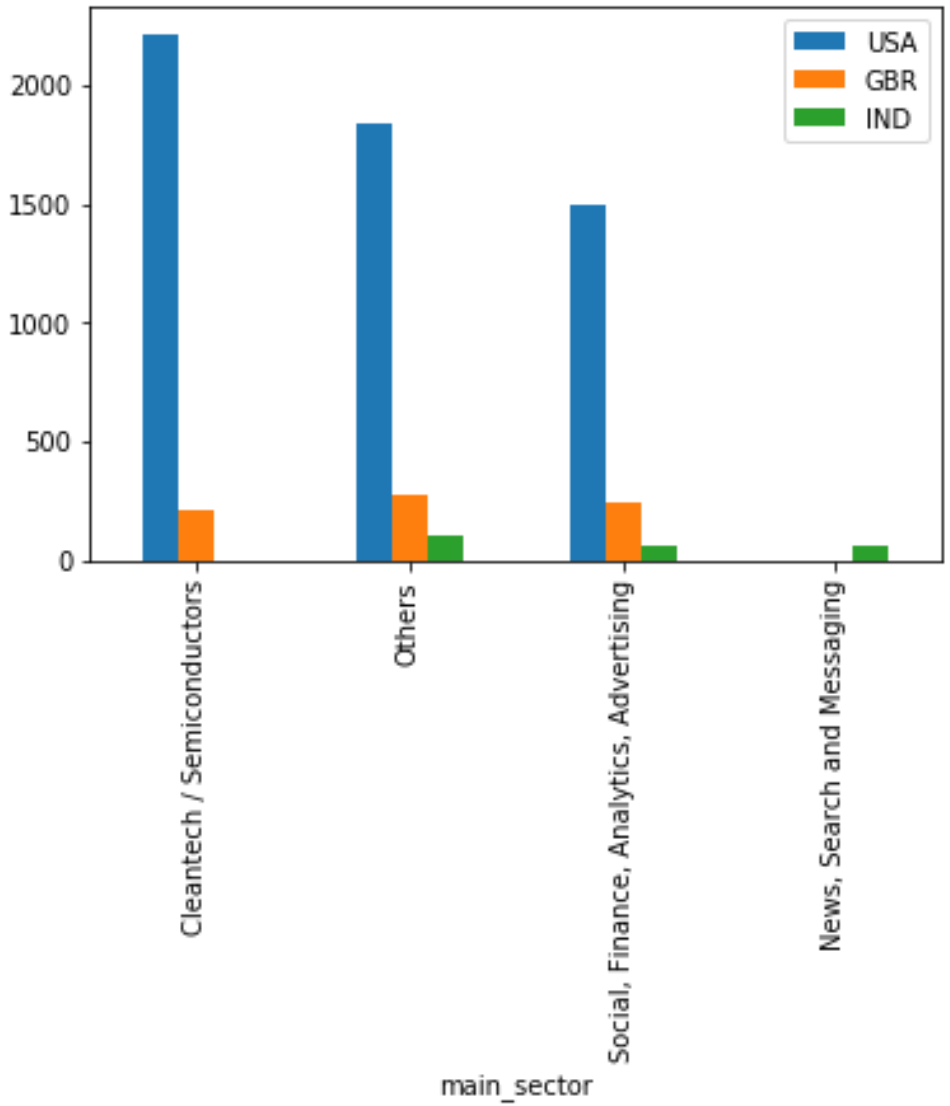
Ref- “https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language”



<Analysis>

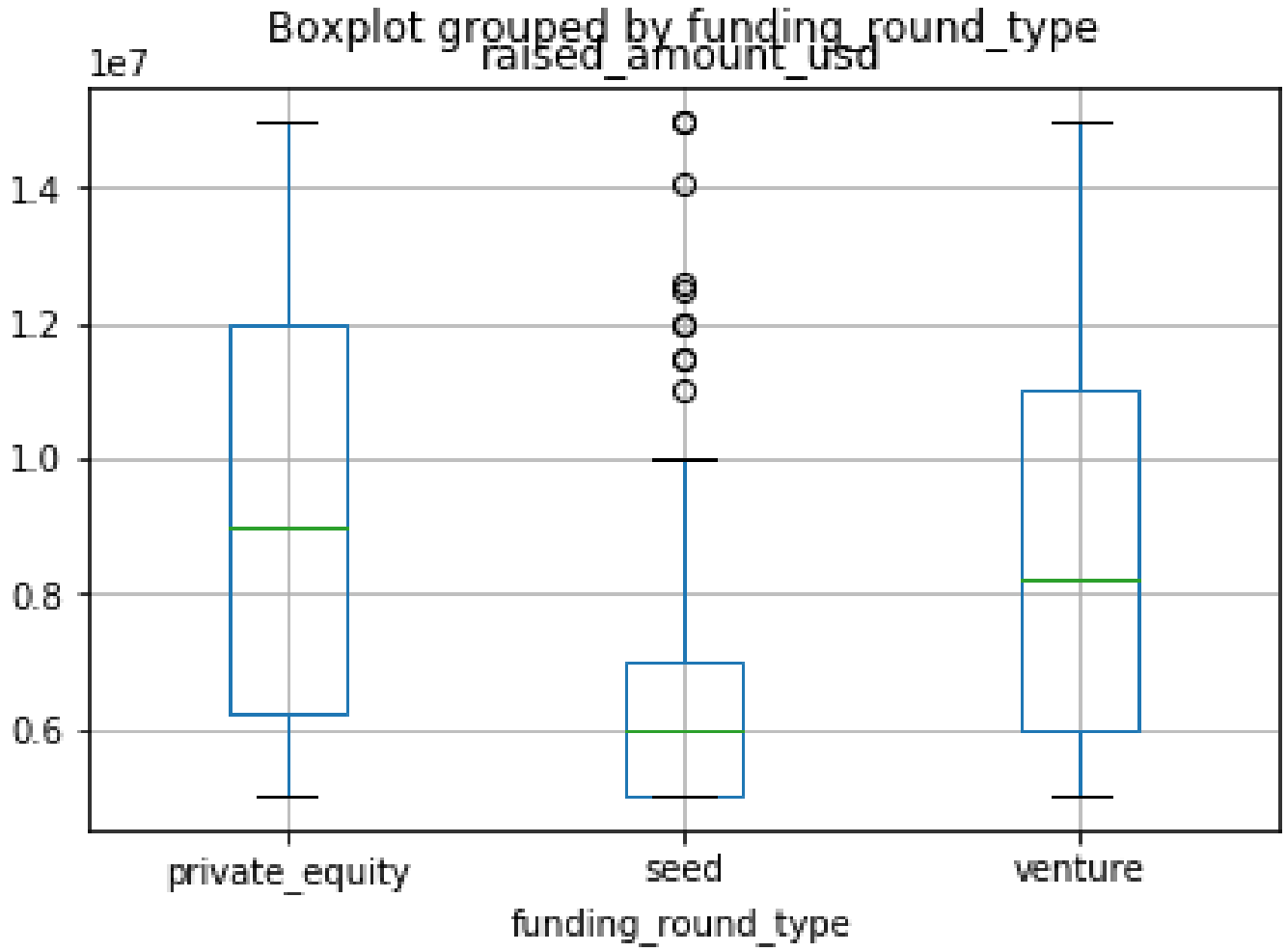
Phase 3 :

- Imported mappings.csv file to map the datasets with the primary_sector.
- Observation: ‘category_list’ data was corrupted in mappings.csv file. ‘na’ was replaced with 0. Hence, corrected it.
- Created 3 Data frames for USA, GBR, and IND.
- From each Data Frame extracted number of Investments for each Main Sector.
- Then by Merging the Data, Plotted a graph of top 3 sectors of the top 3 countries.



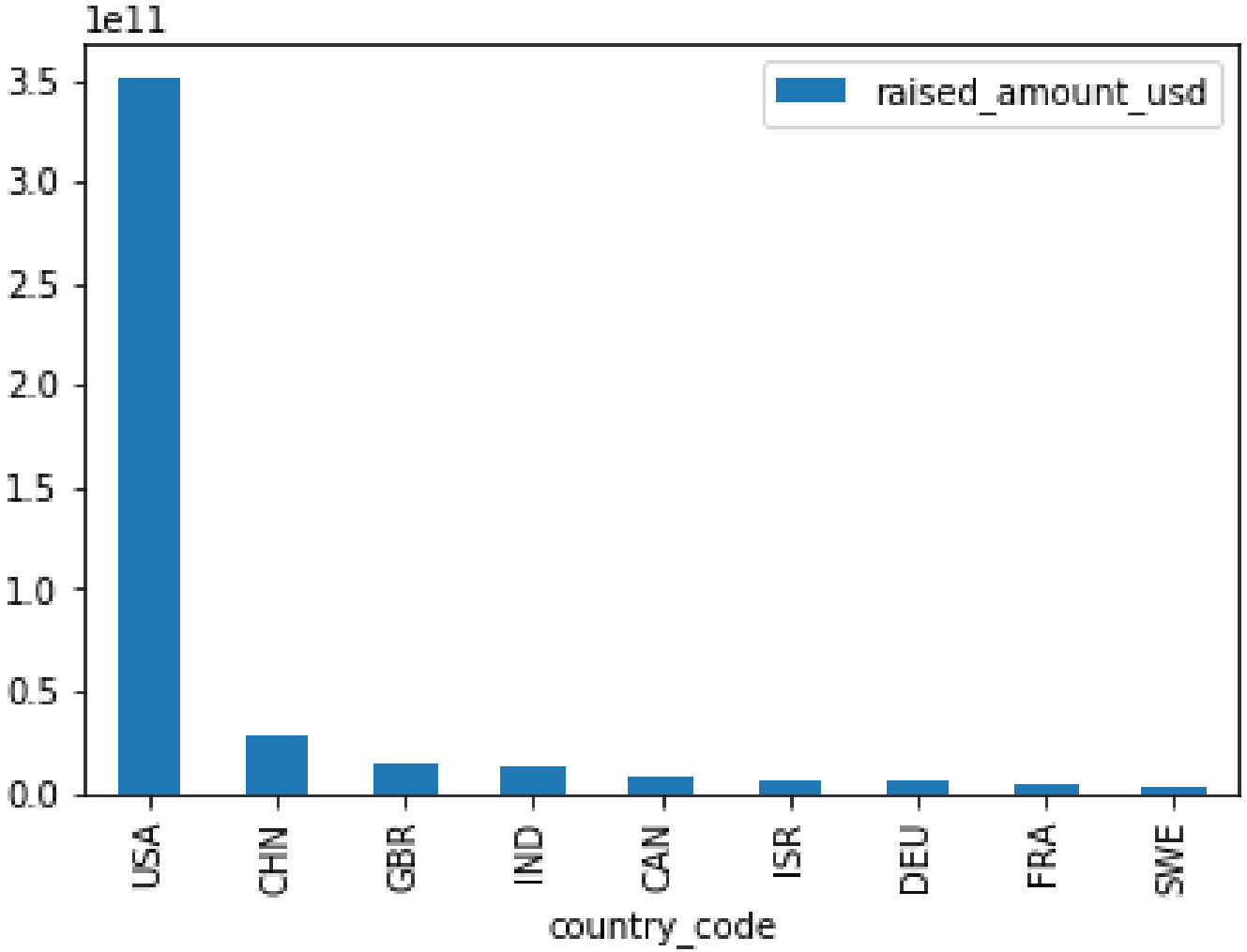
<Results>

Plot 1



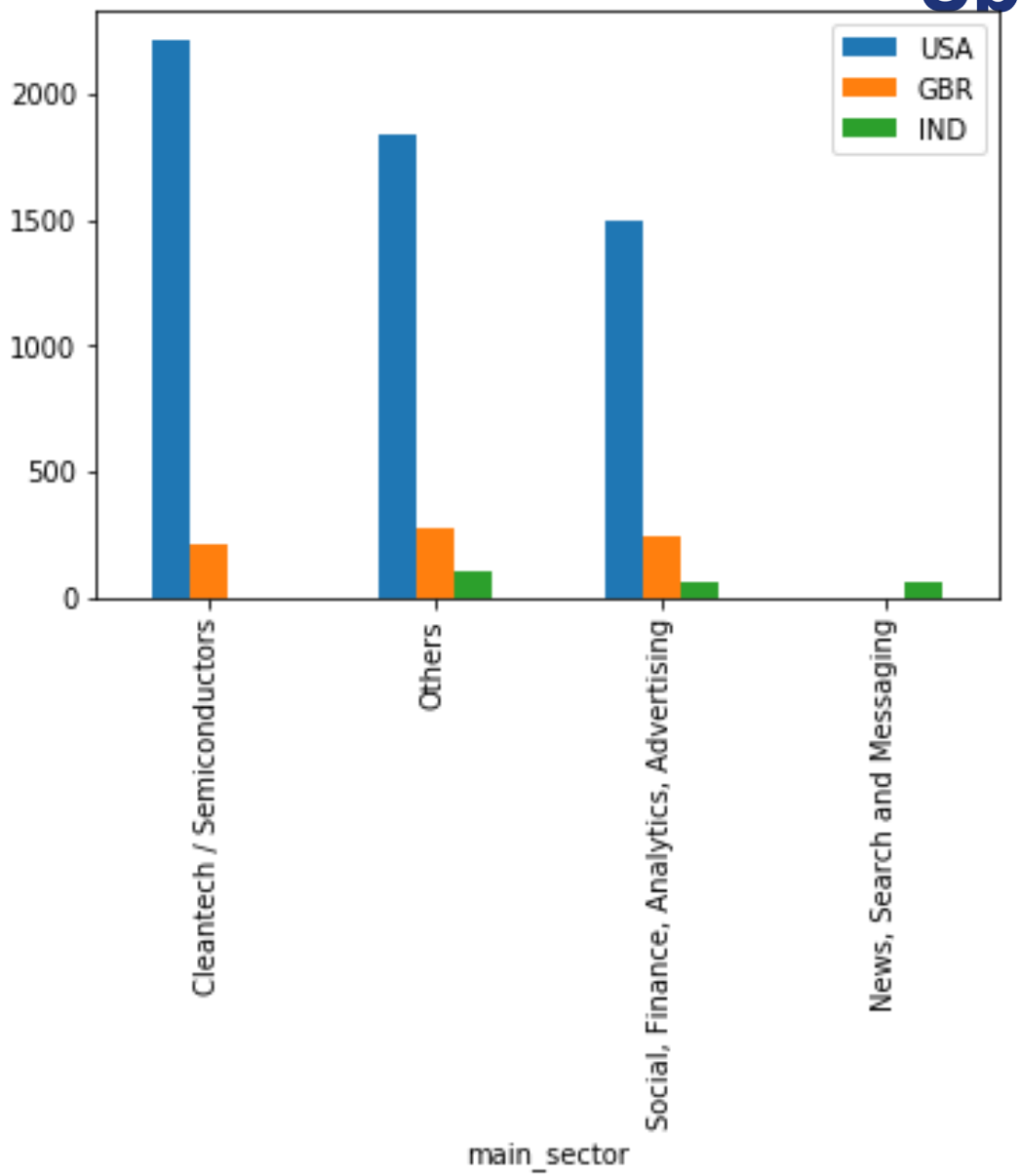
<Results>

Plot 2



<Results>

Plot 3



<Conclusions>

1. “permalink” was the unique key for merging the companies and rounds2 datasets.
2. 4 companies were present in rounds2 but not in companies.
3. “venture” is the most suitable type to invest in.
4. Top-3 English Speaking Countries are - “USA”, “Great Britain” and ”India”.
5. Top 3 sectors of USA are – “Cleantech / Semiconductors”, “Others” and “Social, Finance, Analytics, Advertising”.
6. Top 3 sectors of Great Britain are – “Others”, “Social, Finance, Analytics, Advertising” and “Cleantech / Semiconductors”.
7. Top 3 sectors of India are - “Others”, “Social, Finance, Analytics, Advertising” and “News, Search and Messaging”.
8. (USA)In “Cleantech / Semiconductors” > “Freescale Semiconductor” received highest investment.
9. (USA)In “Others” > “SoFi” received highest investment.
10. (GBR)In “Others” > “OneWeb” received highest investment.
11. (GBR) In “Social, Finance, Analytics, Advertising” > “Powa Technologies” received highest investment.
12. (IND) In “Others” > “Flipkart” received highest investment.
13. (IND) In “Social, Finance, Analytics, Advertising” > “ShopClues.com” received highest investment.