INVESTMENT ASSIGNMENT

SUBMISSION

Name: Phaneendra G

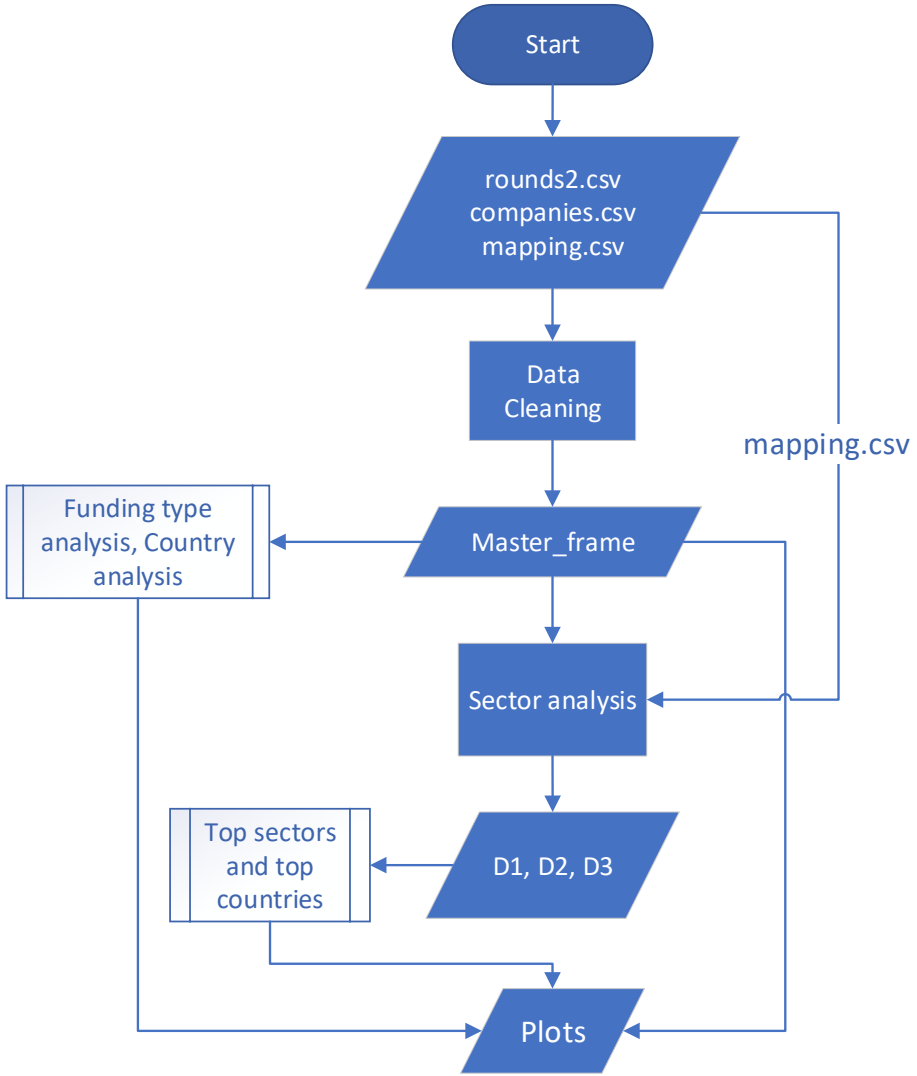
Abstract

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. For that we need to find the pattern where other investors are investing. This pattern is often observed among early stage startup investors. We have taken real investment data from '**crunchbase.com**', that is incredibly useful.

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 of data analysis are divided into investment type analysis, country analysis and sector analysis. Performing these analysis we will create a dataset that is used to take the investment decisions effectively.

Problem solving methodology



Data Cleaning

Initially during data cleaning we worked on “rounds2.csv” and “companies.csv” datasets. We found the unique companies in “rounds2” and “companies” datasets using the “permalink” attribute . There are more companies in “rounds2” than in “companies” so we merge both the datasets into one and the resultant data frame is considered as the master frame that contains entire data and that is used in the later stages.

Now considering the new and complete data frame named master frame, we need to handle the missing values. There are lot of missing values and some data that is not actually useful for our analysis. There are some missing values in the “raised_amount_usd”, “country_code” and “category_list” which are not useful for the further process, so we clean it. And also there are some columns that doesn’t add any importance like “funding_round_code”, “funded_at” , “homepage_url”, “state_code” , “region” and “city”.

These columns may be useful in more deeper communications with the companies, Spark wants to invest on. But at present we don’t need them, so we drop them all from the master frame.

The Funding type analysis is the next and important task to do. We consider the best funding type for investing, by calculating the overall amount raised in a particular funding type. Then we can finalise that “Venture” is the best funding type for investing for the Spark funds. “Private equity” is the top funding type actually, but its average investment is not in the range of Spark Funds. So, we go for the next top funding type.

Next comes the Country analysis in which we take all the countries in which “Venture” plays the major role. The top 9 countries from the list are segmented into a new data frame as “top9” data frame, from which we select the best countries which speak English so that there is no communication gap for the investment process.

Sector Analysis

In the sector analysis we take the new dataset “mapping.csv” that has all the sectors in the master frame, that comes under some specific eight main sectors. This dataset is in the form of table that contains all the sectors as rows and the main sectors as the columns, that later got minimised or melted down into a data frame, that is compatible to merge with the master frame.

The new master frame contains the sectors as well as the main sectors of every individual sub sector. In the next stage we divided the master frame into three data frames D1, D2, D3 based on the countries we have selected from the country analysis. These data frames contains all the attributes of the master frame in addition with two new columns “Count” and “Sum” of the raised amount in each main sector.

Data Visualisation

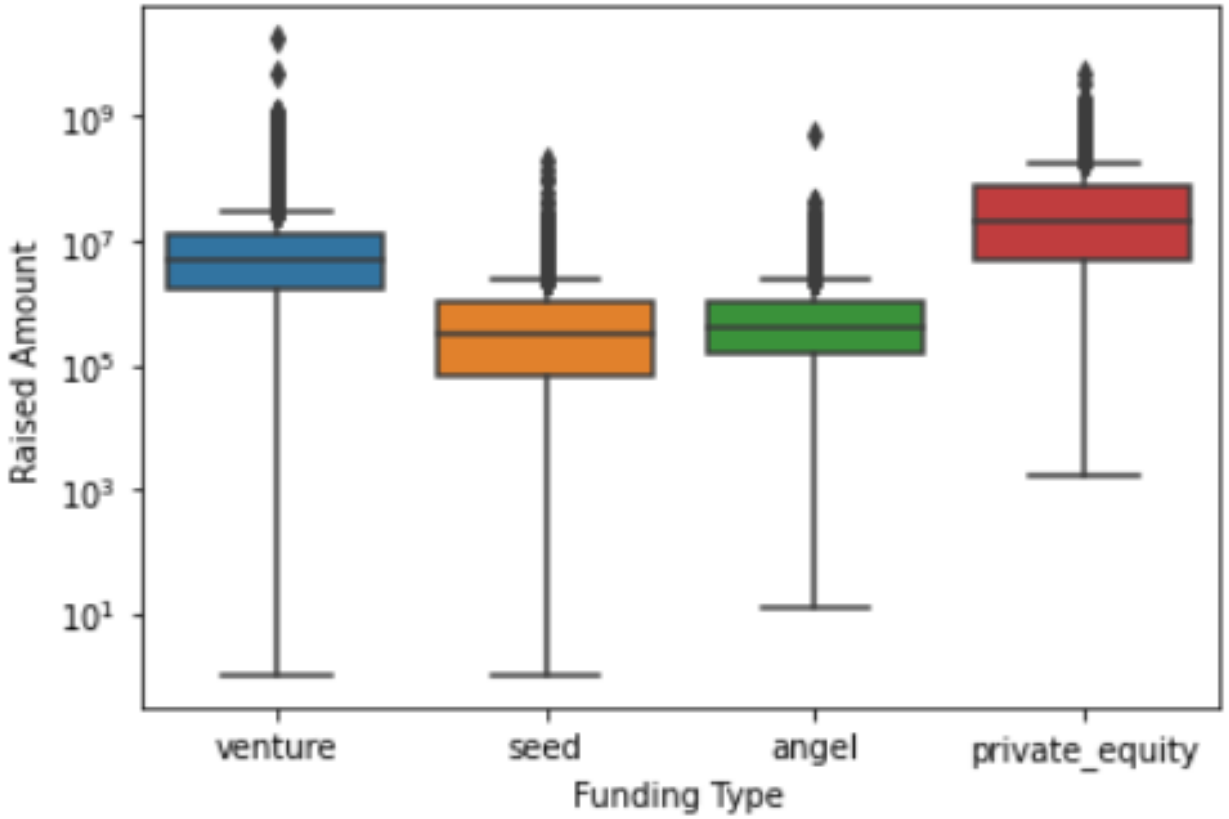
Before data visualisation we need to sort the entire data in a particular order, for that we need to extract more information in particular, from the data frames D1, D2, D3. We found out the top most main sectors in a particular country according to the count and the total investments in that sector. Then we need to find the best companies in those top sectors to invest in.

After the entire sorting process, the process of data visualisation starts, for that we got some conditions from the industry. The first plot shows the fraction of global data of given 4 funding types, which shows the funding type suitable for Spark Funds.

The second plot gives you a glance of the entire data of the top 9 countries investing in the preferred funding type. And also gives you the selected English speaking countries data from top9.

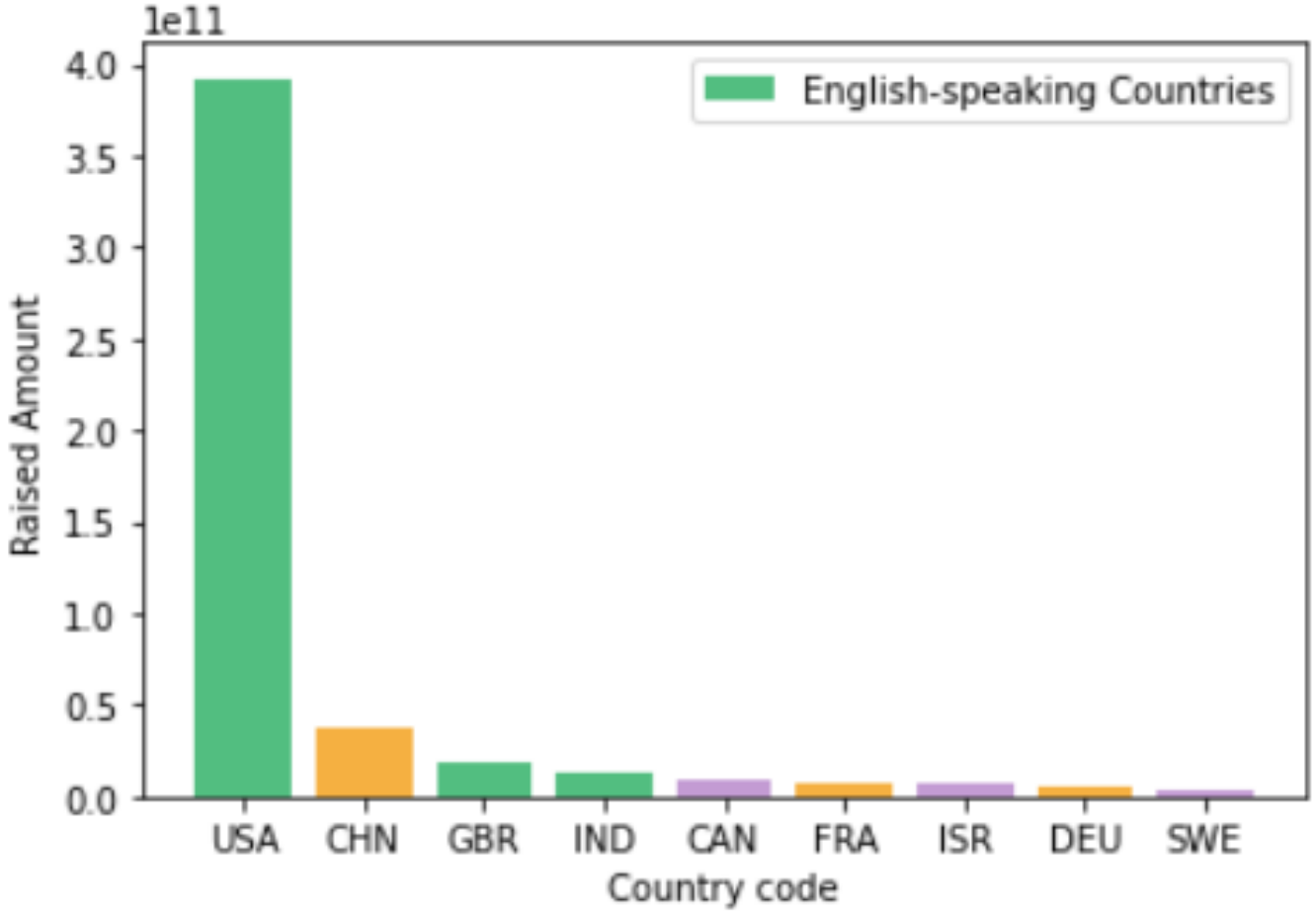
The third plot provide the data of all the top sectors in the selected countries which use the preferred funding type based on no.of times the sector is invested.

Total Investments in Funding types

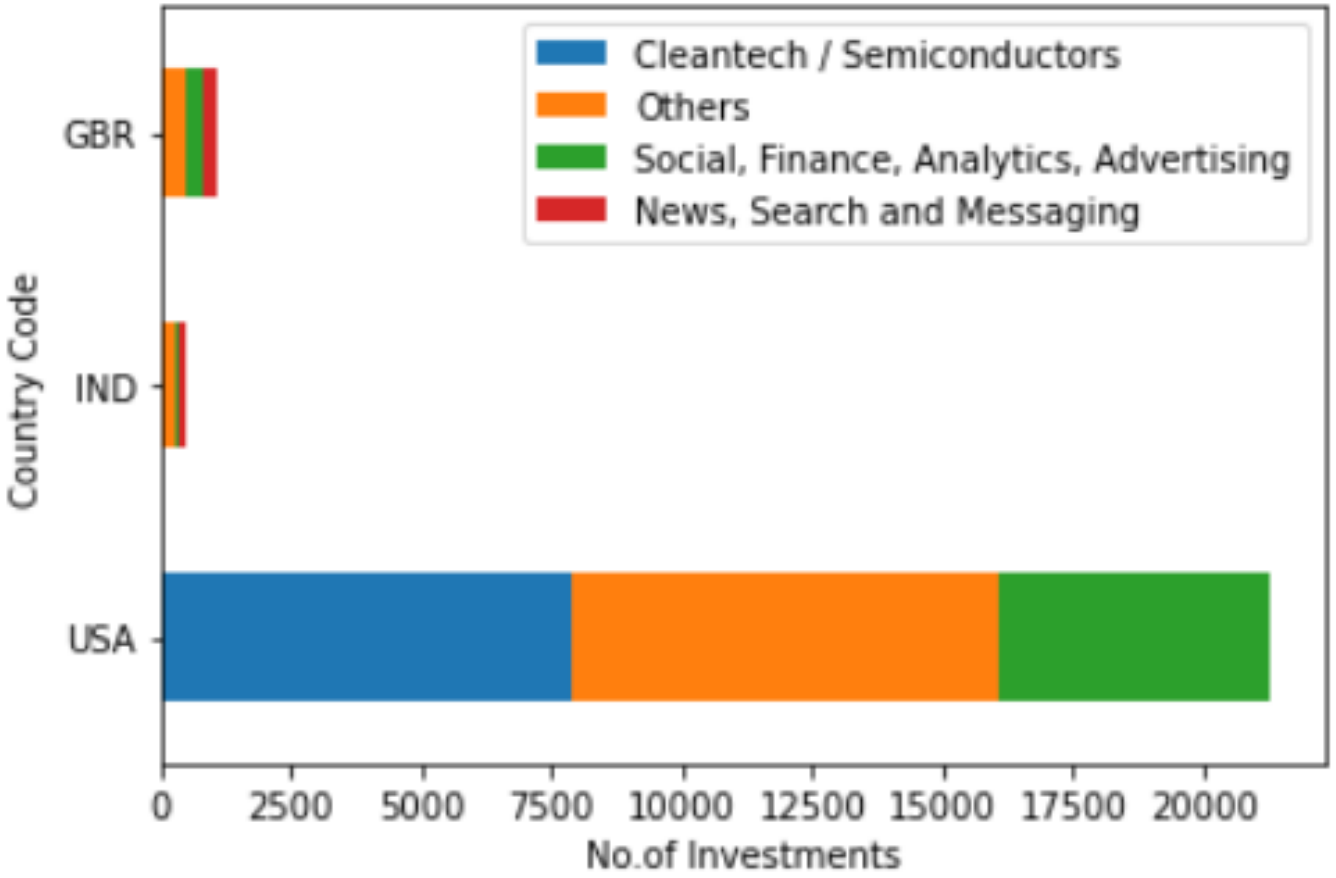


The “Private equity” may be the highest but its average investment amount is more than the amount Spark Funds could invest. So, we go for the next highest Funding type that is in the range of Spark Funds is the “Venture”.

Top 9 countries invested in Venture



Top sectors of top countries



Conclusion

- The Best preferred Funding type for Spark Funds is **Venture**. Which is concluded based on the total global investments.
- From the Top 9 countries, the selected English-Speaking Countries are **USA** (United States), **GBR** (United Kingdom) and **IND** (India).
- The Top sectors in these top 3 countries are plotted in the chart based on the number of investments.