INVESTMENT ASSIGNMENT

SUBMISSION

Name: NAMITA RANI

Abstract

Problem Statement ->

An asset management company, Spark Funds wants to make investments in few companies and has below 2 constraints for investment:

- It wants to invest between **5 to 15 million USD** per round of investment.
- It wants to invest only in **English-speaking countries** due to ease of communication.

To make an investment Spark Funds wants to understand the global trends in investments by identifying the best sectors, countries and a suitable investment type for making investments using data analysis. The overall strategy is to invest where other companies are investing, implying that as the 'best' sectors and countries so that she can take the investment decisions effectively.

To achieve this objective, 3 types of data analysis are conducted:

1. **Investment Type Analysis**: Comparing the typical investment amounts in the different investment types so that Spark Funds can choose the type that is best suited for their strategy.
2. **Country analysis**: Identifying the countries which have been the most heavily invested in the past. These will be Spark Funds' favorites as well.
3. **Sector analysis**: Understanding the distribution of investments across the different sectors.

Data given for analysis is 3 files , named **Companies,Rounds2 and mapping** file having companies details, Funding rounds details and Sector classification respectively.

<Problem Solving Methodology and Analysis>

While looking/analysing the data , below observations are made. This also shows how the case study is proceeded.

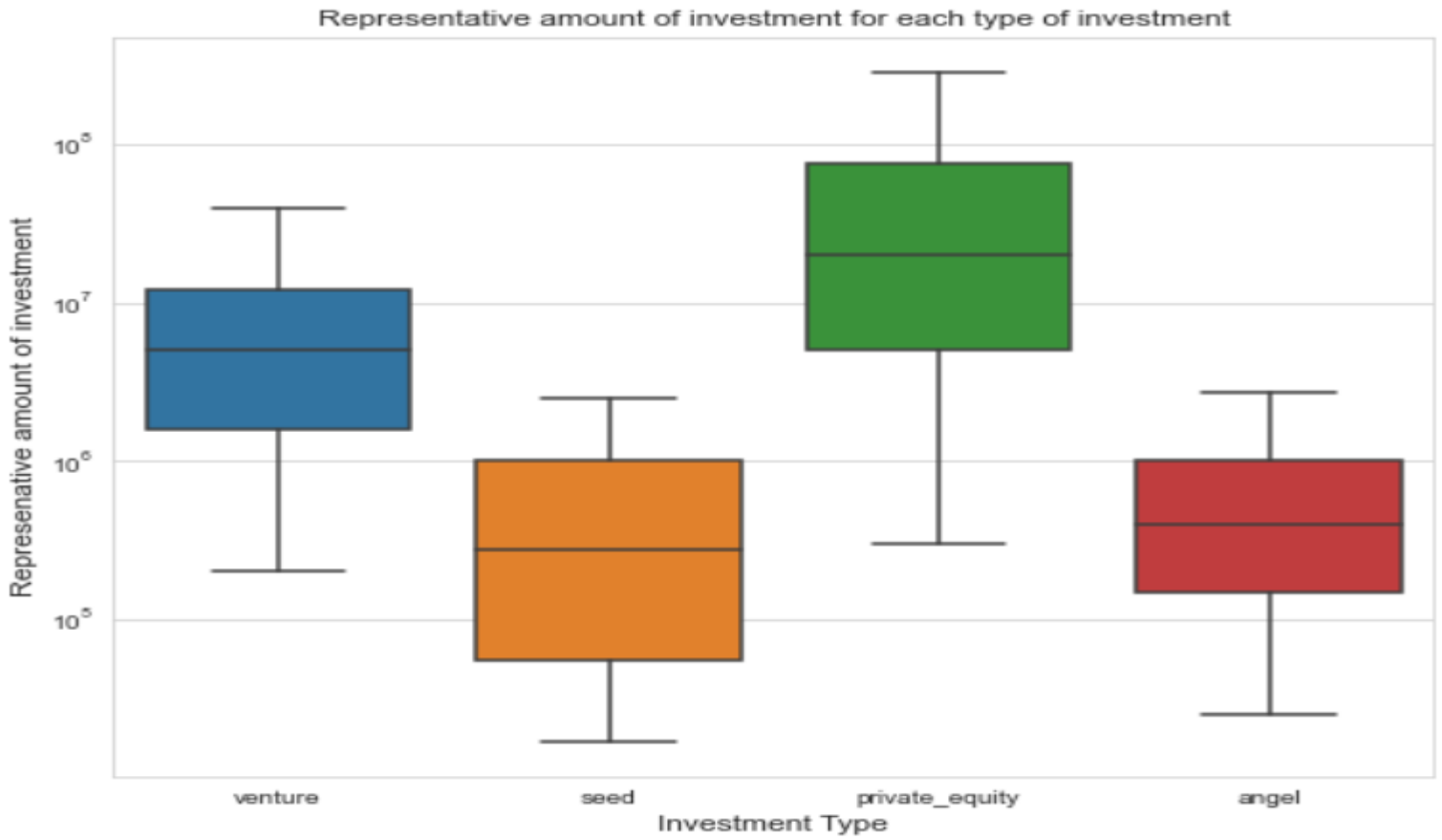
1. Data in given files(companies and rounds2) is not case consistent. So, it is rectified.
2. Master data frame is created having information related to money raised in funding across different investment types in different countries
3. It's found that there were rows with no information on amount of investment. These rows were removed (imputing with mean/median is not suitable here as standard deviation of the amount variable is very large, it'd create ambiguities in the outcome if imputed).
4. As Spark Funds wants to invest in one of the 4 investment types seed, venture, angel, private equity; master frame is further filtered based upon these types and the rows with no country and category information are removed as they don't add much value to the analysis. Because anyhow, after targeting the suitable investment type , we are expected to perform country and category analysis on that data. Hence, Imputing these columns with mode may add variations in the analysis.
5. Investment type data for 4 categories is analyzed and outliers are acknowledged for investment amount.
6. We have not removed the outliers and hence used median of investment amount as the measure of central tendency for each investment type. So, now we have the representative value of investment amount for each investment type. This helps in deducing that Spark funds would invest in Venture type .
7. Top 9 countries are found based upon the total amount of investments. Out of them top 3 English speaking countries are chosen.
8. The data for Venture type is further clubbed with information from Mapping file. Mapping file has main sector mapped against each category. This helps in pointing which company category is linked to a particular main sector.
9. For each sector we have the company which received the total highest investment in top 3 countries.
10. Based on this, We can suggest Spark Funds where it can make the investment.

<High Level Steps Performed>

1. Gathering Requirement from each Checkpoint and proceeding in order.
2. Preparing the data in the format required for further analysis. E.g. In this case company unique id in companies file is made case consistent to match it with the company ids present in rounds2 file.
3. Once the data is in the required format, it's cleaned to filter out the irrelevant information(source of truth : Business Requirement and Constraints)
4. Manipulating the data E.g. Merging 2 different frames to get linked info.
5. Once the data is ready, it is analyzed w.r.t requirements and results are deduced for further analysis.
6. Steps 2-5 are repeated as we proceed further in each checkpoint depending upon the requirements and data is handled to get the outcomes.
7. Presentation of deduced outcomes are plotted to show the results pictorially as visual representation makes more sense to Business.

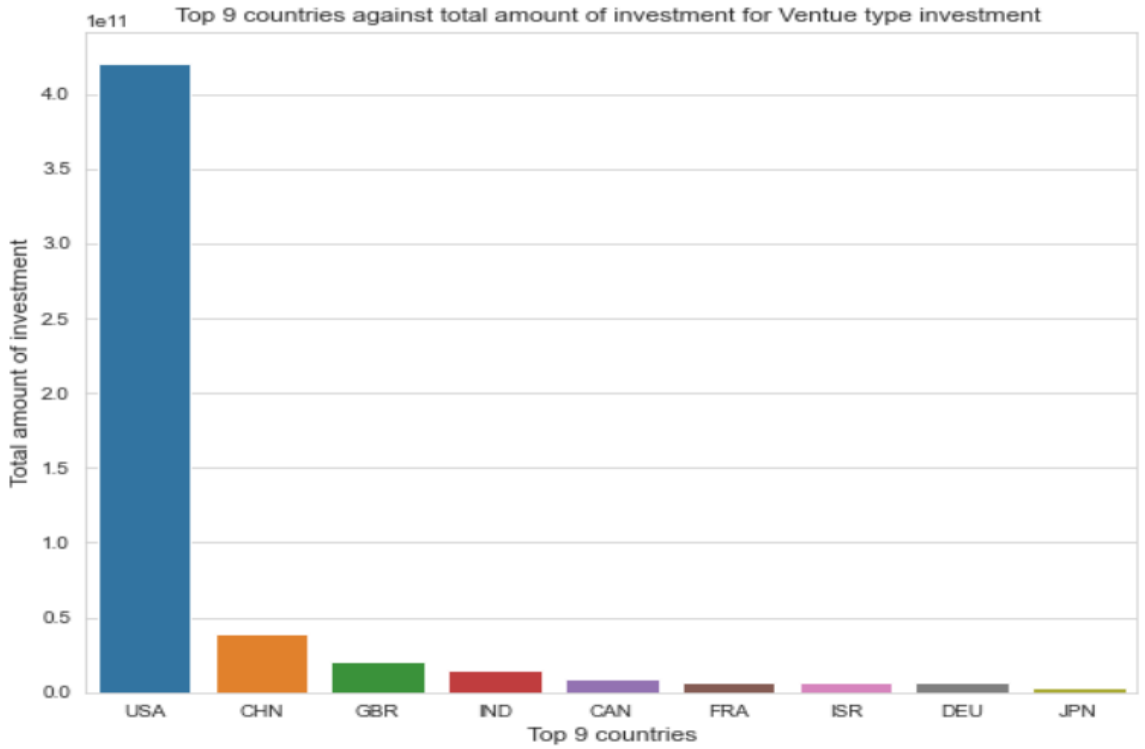
<Results>

Plot 1 : This shows **the representative amount of investment(median)** in each funding type. This shows that venture type investment is most suitable as the representative value for this is 5 Million USD. This fits as per the constraint from Spark Funds



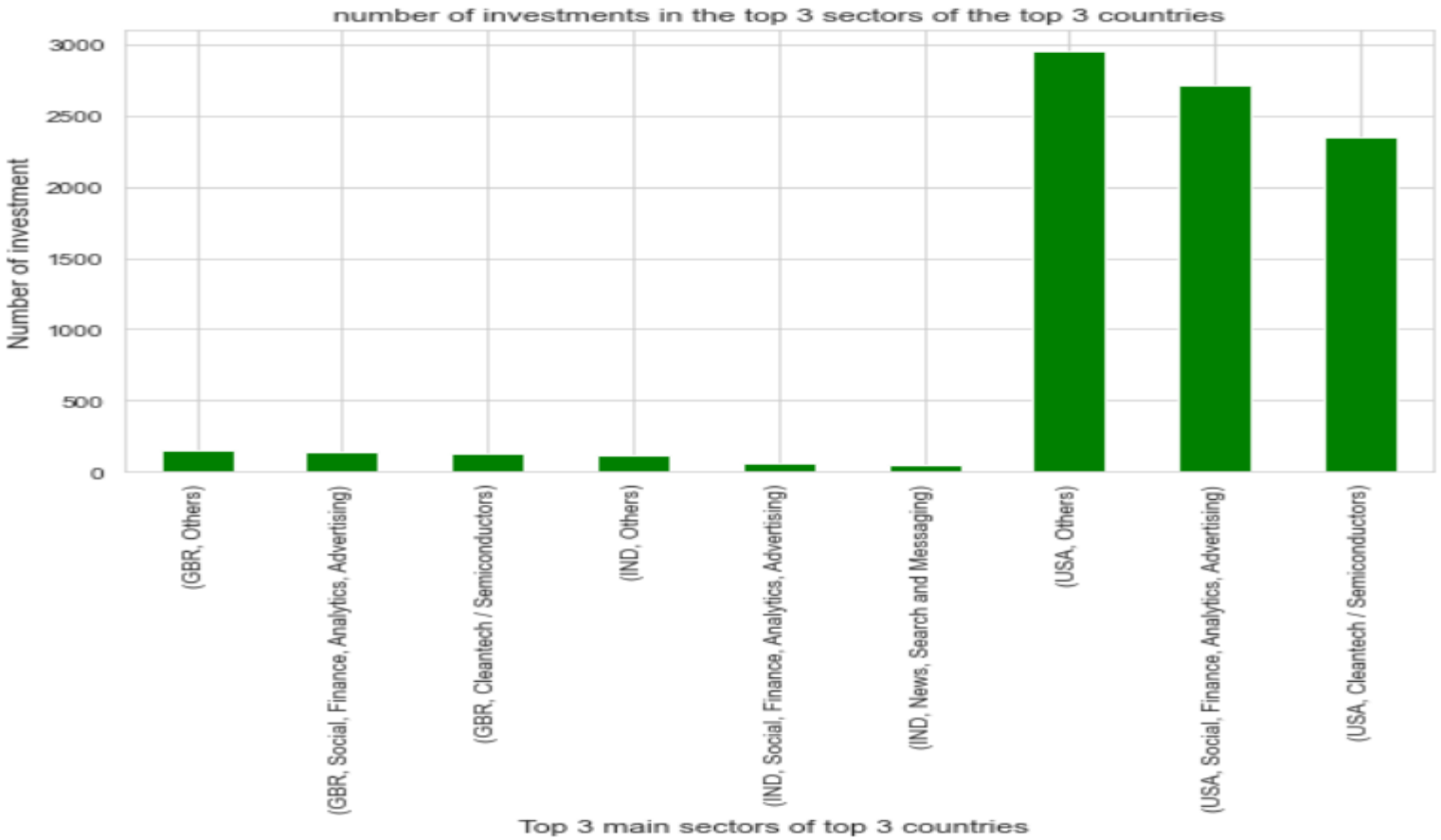
Results>

Plot 2 : This shows top 9 countries against the total amount of investments of funding type Venture. This shows USA,CHN,GBR are top 3 countries in terms of total amount of investment. But since Spark Funds is only considering English Speaking, it's would be USA,GBR,IND as top 3 countries for investment.



Results>

Plot 3 : This shows number of investments in the top 3 sectors of the top 3 countries for venture type investment. This represents top 3 sectors for USA,GBR and IND and the number of investments for each sector.



<Conclusions>

So, this is concluded that Spark Funds can invest in venture Type investment across USA,GBR and IND.

Top 3 sectors based upon number of investments across each country are below:

Top Sector Name : Others in all top 3 countries

Second Sector Name : Social, Finance, Analytics, Advertising in all top 3 countries

Third Sector Name: Cleantech / Semiconductors for USA and GBR and News, Search and Messaging for IND.