INVESTMENT ASSIGNMENT

SUBMISSION

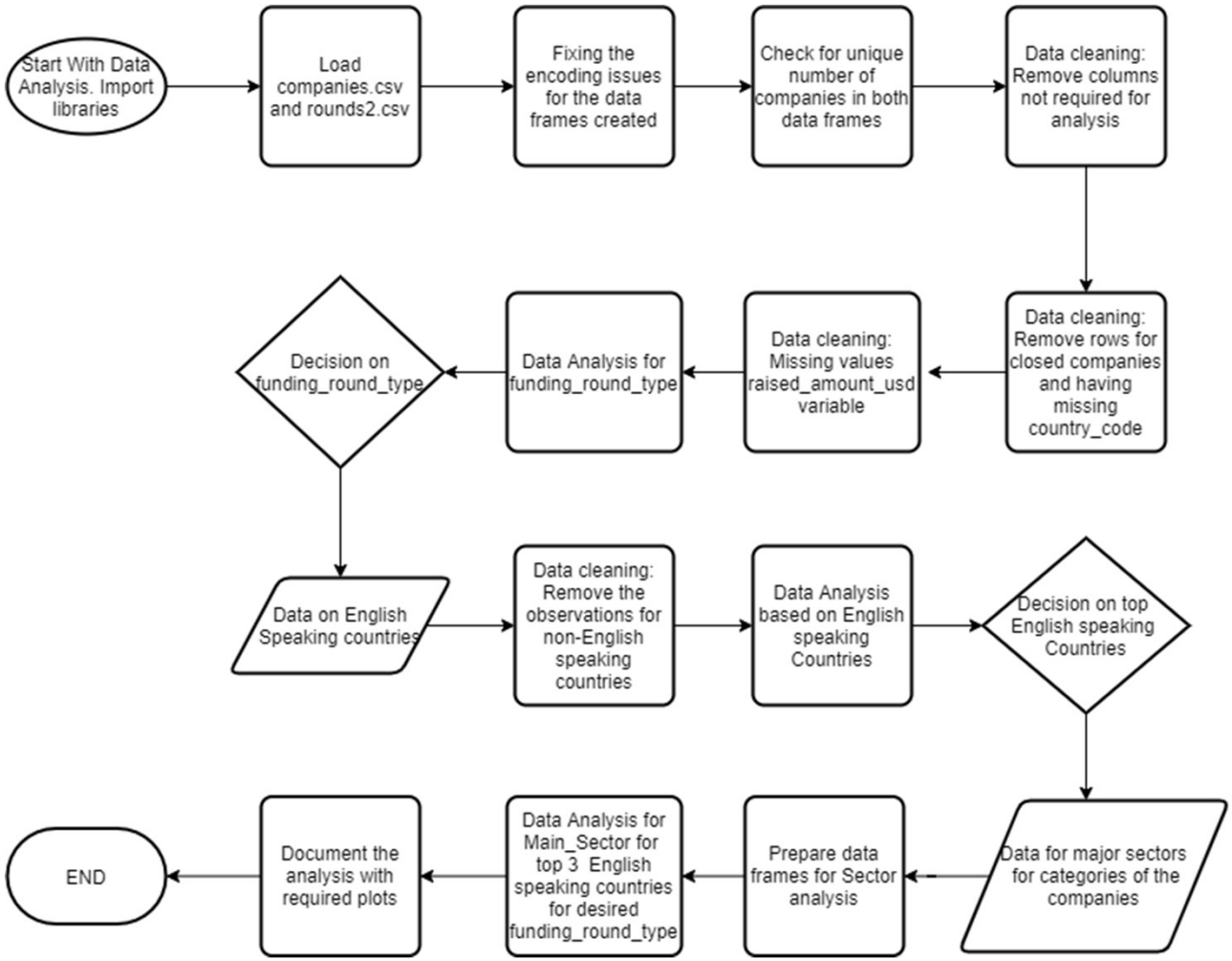
Name: Parthesh C Pujari

Abstract

This data analysis is for an asset management company – Spark Funds. Spark Funds wants to invest amount between 5 million - 15 million USDs in some companies of English-speaking countries.

For this data analysis the real investment data has been sourced from crunchbase.com. The business 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'.

Flowchart for solving problem



Data Analysis

- After loading the companies.csv and rounds2.csv, some encoding issue were seen. These were resolved by using encoding='iso-8859-1' for both the files. After this columns like permalink, name and company_permalink in both the data frames created showed invalid characters. This was fixed using following command for respective columns

  companies.permalink = companies.permalink.str.encode('utf-8').str.decode('ascii','ignore')

- This fixed both the data frames and were merged for further data analysis

- Data Cleaning:
  - Deleted the rows for companies whose status is closed
  - Deleted the rows for companies whose country_code is missing
  - Deleted the columns which are not required for data analysis like 'funding_round_code', 'founded_at', 'homepage_url', 'state_code', 'region', 'city' etc.

Analysis continued…

- Imputing Missing Values

  The 'raised_amount_usd' columns has 17% missing values. From the boxplots it is seen there are outliers for this variable and hence for imputing the missing values consider median of the funding type category. Hence for example missing value of raised_amount_usd for venture funding type will be imputed with median of raised_amount_usd for venture.

- Funding type Analysis

  Spark Funds has decided to check for venture, seed, angel and private_equity funding types for investment. The Representative Investment amount for these funding types was taken as "median" and used for this analysis. The master data frame was filtered with funding type = venture for further data analysis

- Country Analysis

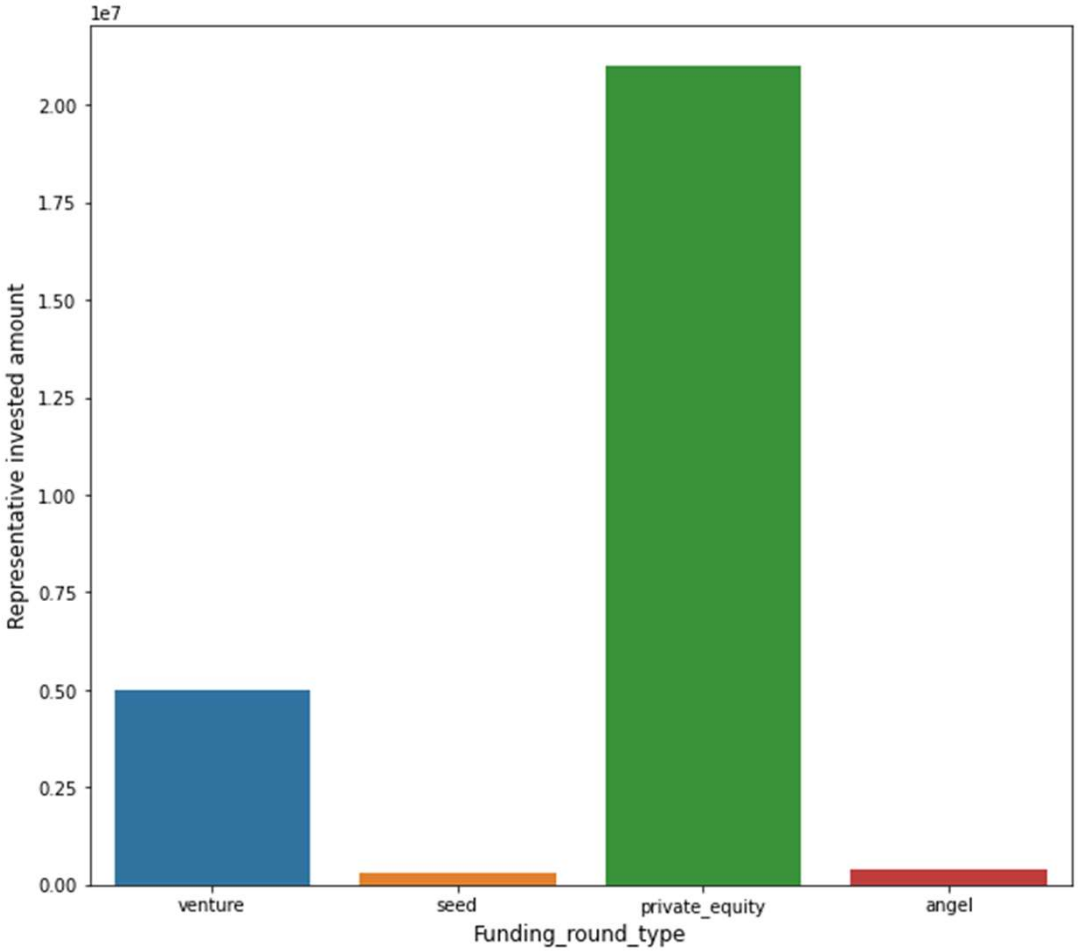
  From the input data of English speaking countries, the data frame was filtered with observations of English speaking countries. New column giving this English speaking information was added. Top 9 English-speaking countries were found based on the total investment attracted by that country and a data frame was created for this Top 9 countries. Top 3 countries were selected from this data. Data frame was created with venture as funding type and the top 3 English speaking countries.

Analysis continued…

- Sector Analysis

  For this analysis first we load the given mapping.csv file which help us to map the company's Primary category to its Main Sector. Now the data frame from the previous analysis is filtered for venture funding type and top 3 countries for attracting most of the investment for this funding type. This data frame is further filtered to have only top 3 sectors based on the number of investments that sector attracts.

- Plots were included to back each of the data analysis




```
funding_round_type
angel 415000.0
private_equity 21000000.0
seed 310000.0
venture 5000000.0
Name: raised_amount_usd, dtype: float64
```

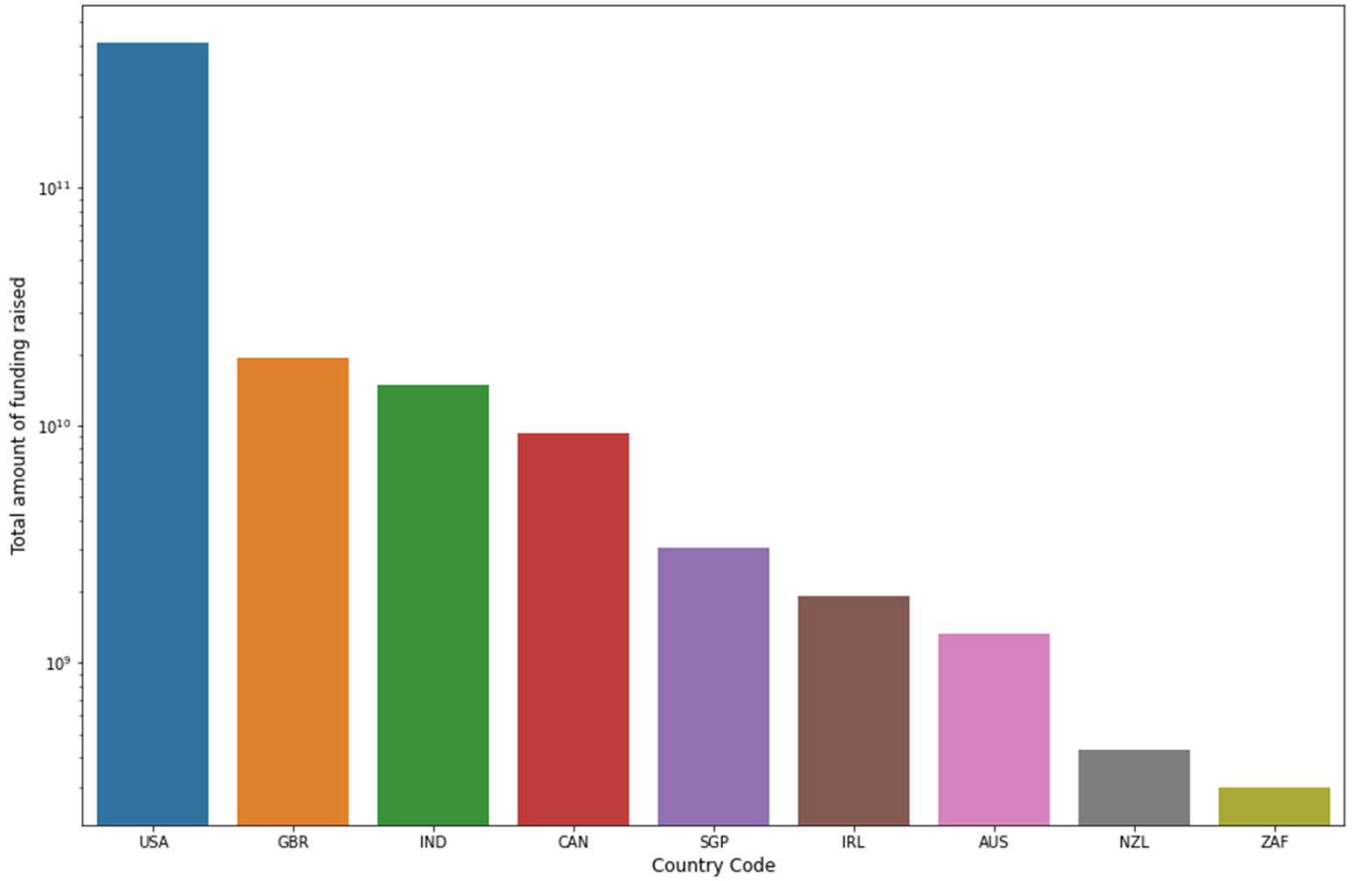
- **Funding type Analysis**

From the plot of Representative Invested amount to funding_round_type (Selected by Spark funds), it is clear that Spark funds can invest only in Venture funding type as arounf 5 million USDs have been invested in this funding type.

Private_equity is out of reach as it has collected more than 20 million funds which is beyond higher limit of Spark funds

Angel and Seed funding types together too has attracted less funds than 5 million.

Hence Spark funds can invest in **“venture”** funding type

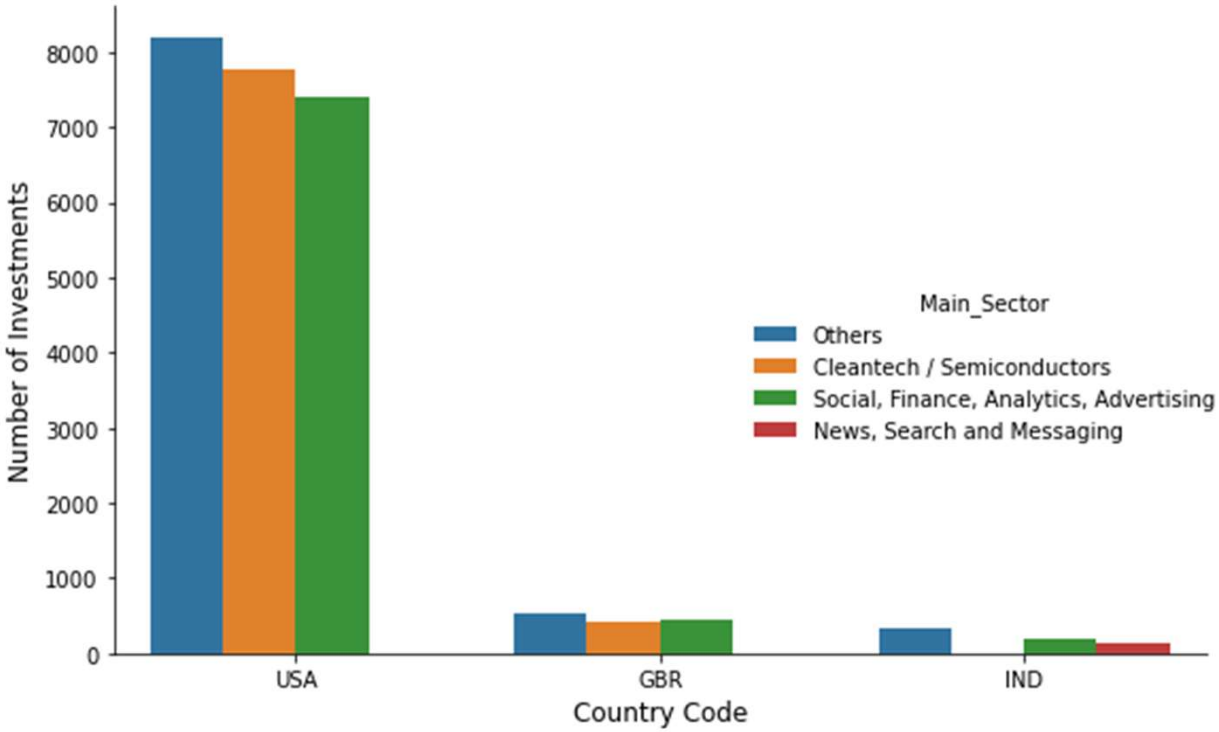


- **Country Analysis**

From the plot of Total amount of funding raised to country_code, it is clear that the top three English-speaking countries to attract most of the investments are

1) USA
2) GBR
3) IND

Y-axis is on log scale for clarity



- **Sector Analysis**

Sector Analysis for top 3 English-speaking countries for venture funding type.

According to the plot it is clearly seen that USA has attracted maximum number of investments

Others sector has raised maximum number of investments

Social, Finance, Analytics, Advertisement sector has raised number of investments after Others sector

Conclusions

- Considering the funding type analysis Spark funds has the option to invest in venture funding type considering it can invest between 5-15 million USDs
- From the country analysis USA seems to be most favorable country to invest.
- With Sector analysis, Others sector is invested the greatest number of times in all the countries for venture funding type.
- The second ranking sector - Social, Finance, Analytics, Advertisement is another good option as it has raised almost comparable number of investments as Others and similar trend is seen in other countries too.
- It will be useful to perform such analysis for private_equity funding type too as for the investment range of Spark funds it can invest in private_equity companies and its funding can remain in Inter-quartile range between 25 percentile and median of the raised amount for private_equity. In that case accordingly we must apply filters to the data and restricting the number of investments only in that range.