

INVESTMENT ASSIGNMENT SUBMISSION

Name:

Business Objective: Spark Fund – Finding the best investment

In this presentation we are presenting our analysis on investment data across the globe by various investment firms.

Given the constraints that Spark Funds wants to: (**business problem**)

- Invest in only countries where English is one of the official language
- Invest in the country where most other investment first are investing
- Invest in sectors that are most heavily invested
- Invest in range of 5-15 million (Find companies , sectors, type of funding that is in this range) most suitable.

Data Understanding – What is available

We have 3 datasets available

- Companies – list of companies that received funding
- Mappings – mapping of category list to sectors to which company belongs
- Rounds2 – where actual transactions of fundings are stored

Data Understanding – transformation

Using these 3 datasets a master data frame is constructed

- By joining rounds2 with companies – here lowercase is done to ensure all companies are joint
- Further based on business rules first category from category list is obtained
- Corresponding sector is found from mappings table
- Further master frame is joined to get one dataframe that contains companies, country_code, funding type and main sectors. These are used for analysis to derive results.

Checkpoint 2: Funding Type Analysis

Here we looked at unique funding types available. Total funding amount , means for each of these done.

Since from this assignment perspective we are interested in seed, venture, private equity and angel these are examined

It is found that venture has largest funding. And most appropriate from range being within 5-15 million

Checkpoint 3: Country Analysis

In country analysis data is sorted in descending order by each country.

Looked at top 4-5 countries and found that USA, GBR and IND are the countries which have English as official language and have largest funding.

So it is derived that spark fund should concentrate chiefly on these 3 countries.

Checkpoint 4 and 5 : Sector Analysis 1 and 2

From there we proceed to sector analysis

- For this first step is to prepare dataset by merging mapping dataset into it.
- This is done by taking first category from category_list and mapping it to sector
- This field is added to master_frame as main_sector
- Analysis 2
 - Since we have dataset D1, D2 and D3 for each country ,
 - Data is examined by grouping by main_sector for each of these countries and top 3 sectors are noted
 - Now under each top sector data is sorted by each company and top company names are derived

Checkpoint 6: Plots

Plot 1

- I chose pie chart to show % of investment by each funding type. Clearly 59% of funding comes from venture funds.
- Also plotted avg funding amount in separate bar chart. And It shows venture average is 11.75. this makes it most suited funding type for spark funds.

Plot 2

- Got data grouped by Country
- Sorted in descending order by funding amount
- Taken top 9 records and plotted on bar chart

Plot 3

- Created 3 DFs one for each country
- Sorted data by funding amount group by main_sector. Took top 3 records. So now these DFs have top 3 sectors for each country
- Created a pivot table with sector in columns and country in rows. This clearly indicates sector values and names.
- Also created a scatter plot for the same.

Problem solving methodology

Got data for companies and rounds2 into two separate dataframes

In order to join these 2 datasets I had to add a field with permalink in lowercase

Join two datasets into one

funding type analysis :

- by selected data by each funding type (venture, angel, seed and private_equity)
- Venture is maximum and mean for this type is 11 million

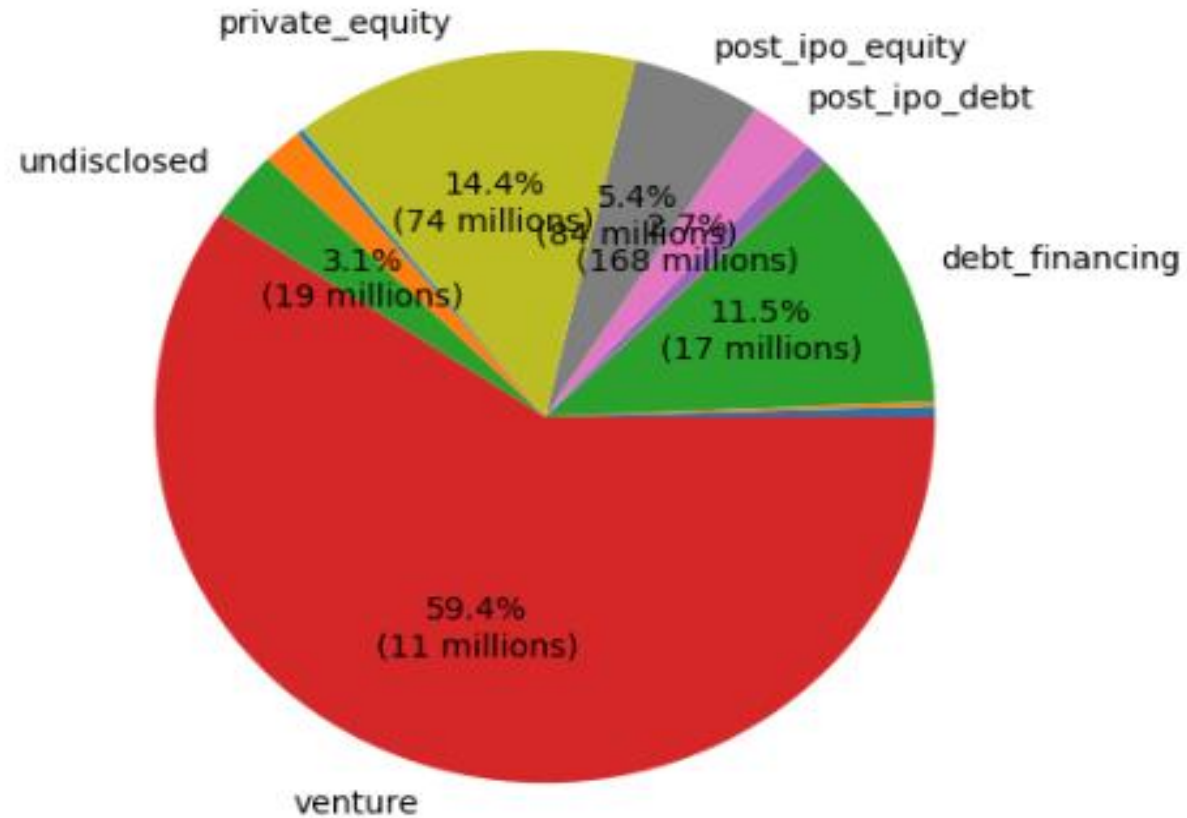
Country Analysis

- Used groupby countrycode for master dataframe and examined the data by sorting it descending by funding amount
- Looked up manually top countries that are present in wiki with English being official language.
- It can be seen that USA, GBR and IND are the countries where English is spoken and these countries received maximum funding.

Problem solving methodology

- Problem is solved by getting 3 datasets in python in memory
- Join these datasets by finding common link like company name and category list
- Group by attributes like country code and funding_round_type to examine the data
- Looking up wiki identify 3 countries that are to be focused upon
- Filter and create datasets one for each country with funding round type that we are interested in.
- Use group bys, sorting for examining the data
- Finally use matplotlib to plot the data

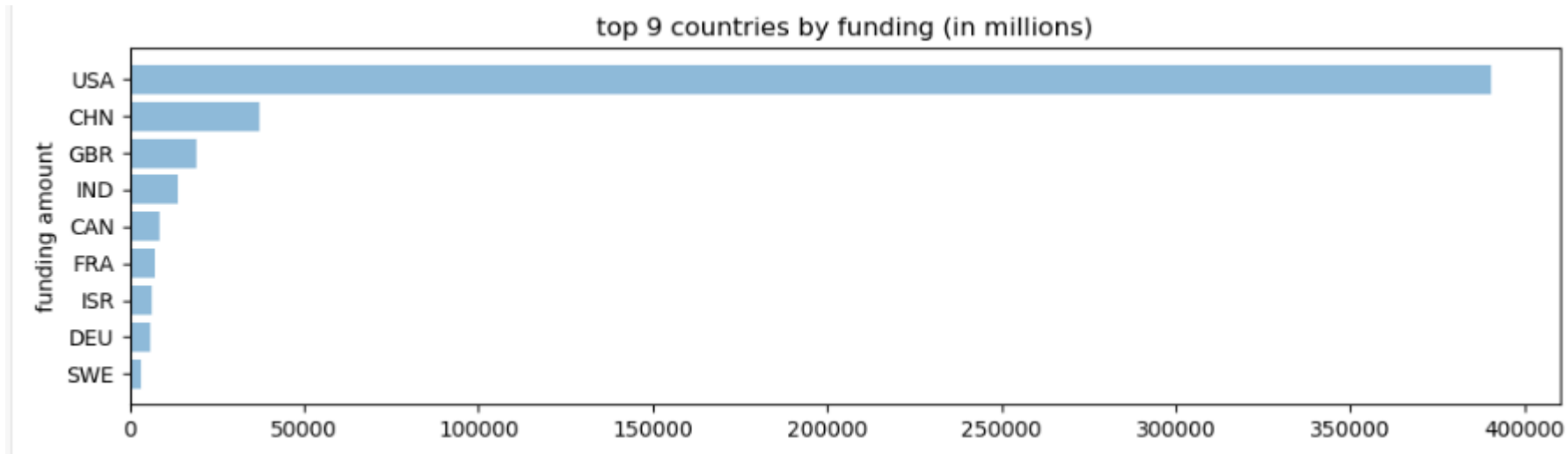
Results



Graph one shows more than half of funding has come as Venture rounds.

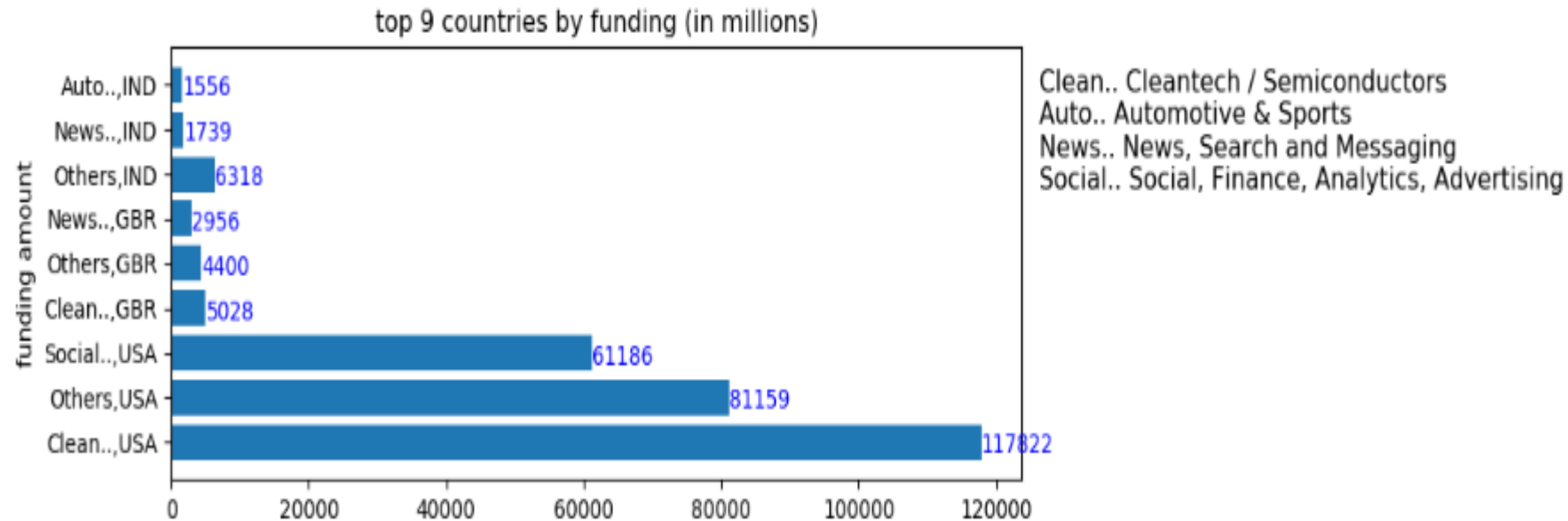
Graph two shows that venture average funding is 11 million which is best suited for Spark Fund

Analysis – Top 9 countries by funding amount



This shows USA has maximum funding being done following by GBR and India which are English speaking Countries. If Investors are from US then it may make most sense to stick to US market.

Results – Plot 3



This show cases Top 3 sectors for each country. For USA its CleanTech, Others and Social. CleanTech being highest. However if you look at India all the 3 sectors investments are not that high. It would be recommended that they stick to Any of the 3 sectors in USA since difference is not that high. CleanTech is most preferred.

Conclusions

- USA, GBR (UK) and IND (India) are the top countries where Spark should invest
- While in USA and UK (GBR) Cleantech / Semiconductors is the sector where most investments are going, In India it is Others..
- Top Companies where investments have gone in are like BioDesix and Tigo-Energy in USA, in UK it is Eusa-Pharma and Horizon-Discovery. In India its Gupshup, Imimobile.
- Based on top companies across these 3 top sectors, Spark fund should decide as to where the investments opportunities exist.
- Analysing funding history and returns of these companies will also help Spark further zero down the companies where it can get best return on its investments.