INVESTMENT ASSIGNMENT

SUBMISSION

Name: Avishek Sen Gupta

Spark Investments

Introduction

This is the Assignment for the Investment case study for Spark Funds. **Spark Funds**, an asset management company. 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.

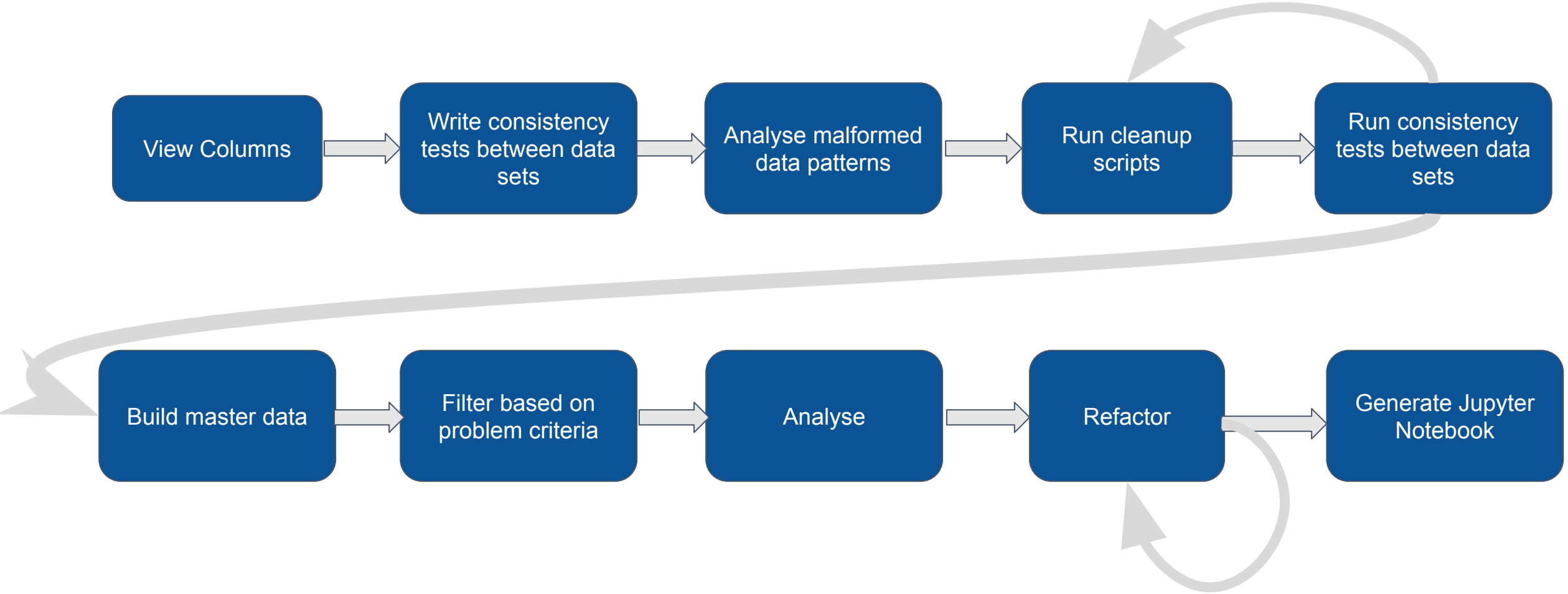
Business Understanding and Domain

- Spark Funds wants to invest **between 5 to 15 million USD** per round of investment
- Spark Funds wants to invest **only in English-speaking countries** because of the ease of communication with the companies it would invest in
- Spark Funds wants to invest **where most other investors are investing**. This pattern is often observed among early stage startup investors.

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'.

Problem Solving Workflow



Data Understanding: CRISP-DM #2

To understand the shape of the data and its cleanliness, we perform a couple of preliminary checks on the *companies* and *rounds2* dataset.

- It seems that the common key between *companies* and *rounds2*, is the **permalink**. Thus, this is chosen as the primary key for identification.
- We'd see that the number of unique companies in the *companies* dataset and the *rounds* dataset are not the same number, implying that the reference company data is malformed or missing.
- Listing out the companies which were in *companies* but not in *rounds2* (and vice versa) was done, and then the patterns of mismatch were observed by looking at the mismatches manually.
- The *mappings* dataset was also not clean. A lot of sector names had their names mangled, with the two letters **na** replaced with **0**.

These are the patterns observed in the malformed data:

- **Pattern 1**: This is the most common type of malformed data. In this the permalink in the *companies* dataset is malformed, mostly because it uses Mandarin characters. The corresponding company name is well-formed. The corresponding permalink in the *rounds2* data set is also well-formed.
- **Pattern 2**: This is the second most common type of malformed data. In this case, the permalinks for a company are malformed in different ways in both the *companies* and *rounds2* dataset, and thus do not match.
- **Pattern 3**: The third common pattern of malformed data is where the *companies* dataset has the correctly generated permalink, but the corresponding permalink in the *rounds2* dataset is malformed.

Data Preparation: CRISP-DM #3

- For **Pattern 1**, the permalinks in the *rounds2* dataset were copied over to the *companies* dataset, after using suitable search patterns.
- For **Pattern 2**, the permalinks were regenerated by sanitising the (well-formed) company name and copied over to the **companies** and *rounds2* dataset.
- For **Pattern 3**, the permalinks in the *companies* dataset were copied over to the *rounds2* dataset, after using suitable search patterns.

- In all these scenarios, the corrected permalinks were stored in new columns (*company_permalink_lowercase* for *rounds2* and *permalink_lowercase* for *companies*), without touching the data in the original columns.
- As part of the data preparation, all permalinks were made lowercase, because the case was inconsistent within and across the *companies* and *rounds2* datasets.

- For the missing sector mappings, we added the missing sectors and their main sector classification manually, after the dataset was loaded.
- The tests for checking whether there were any unique company mismatches were run again after the data cleanup.

Data Analysis

Build Master

The *companies* and *rounds2* datasets were then joined on the permalink column. Spark Funds' constraints on linguistic preferences (English-speaking countries) were then applied.

Median Investment per Investment Type

The median investments in the four investment types (Seed, Angel, Venture, Private Equity) were then compared, and the one which fits Spark Funds' constraints is chosen. As it turns out, this appears to be Venture funds.

Country Analysis

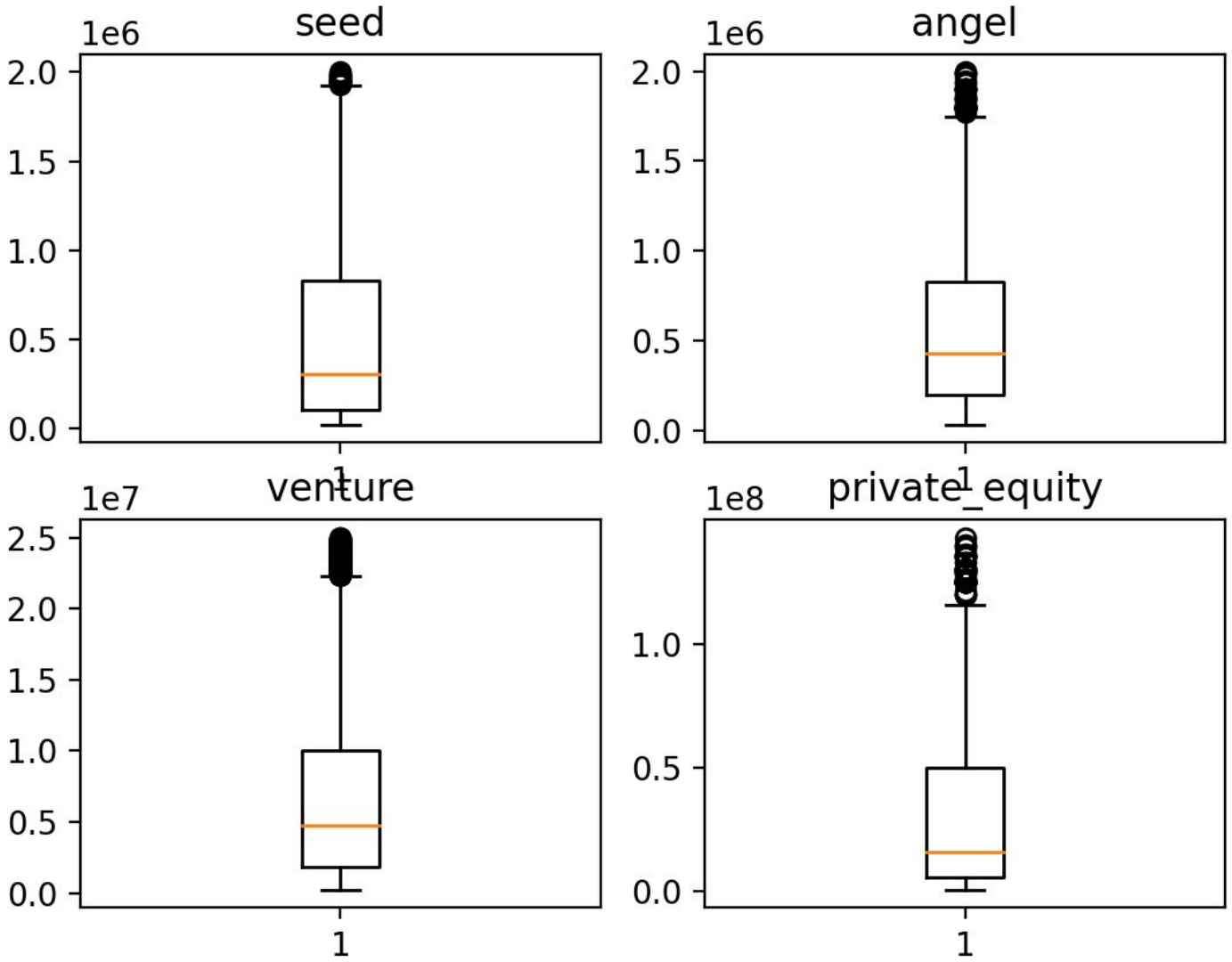
- The top 9 most heavily invested countries are selected, based on the amount raised, using the constraint of investment amount (USD 5 million - USD 15 million).
- The aggregate stats for the top 3 countries for investment amount and number of investments are calculated.

Sector Analysis

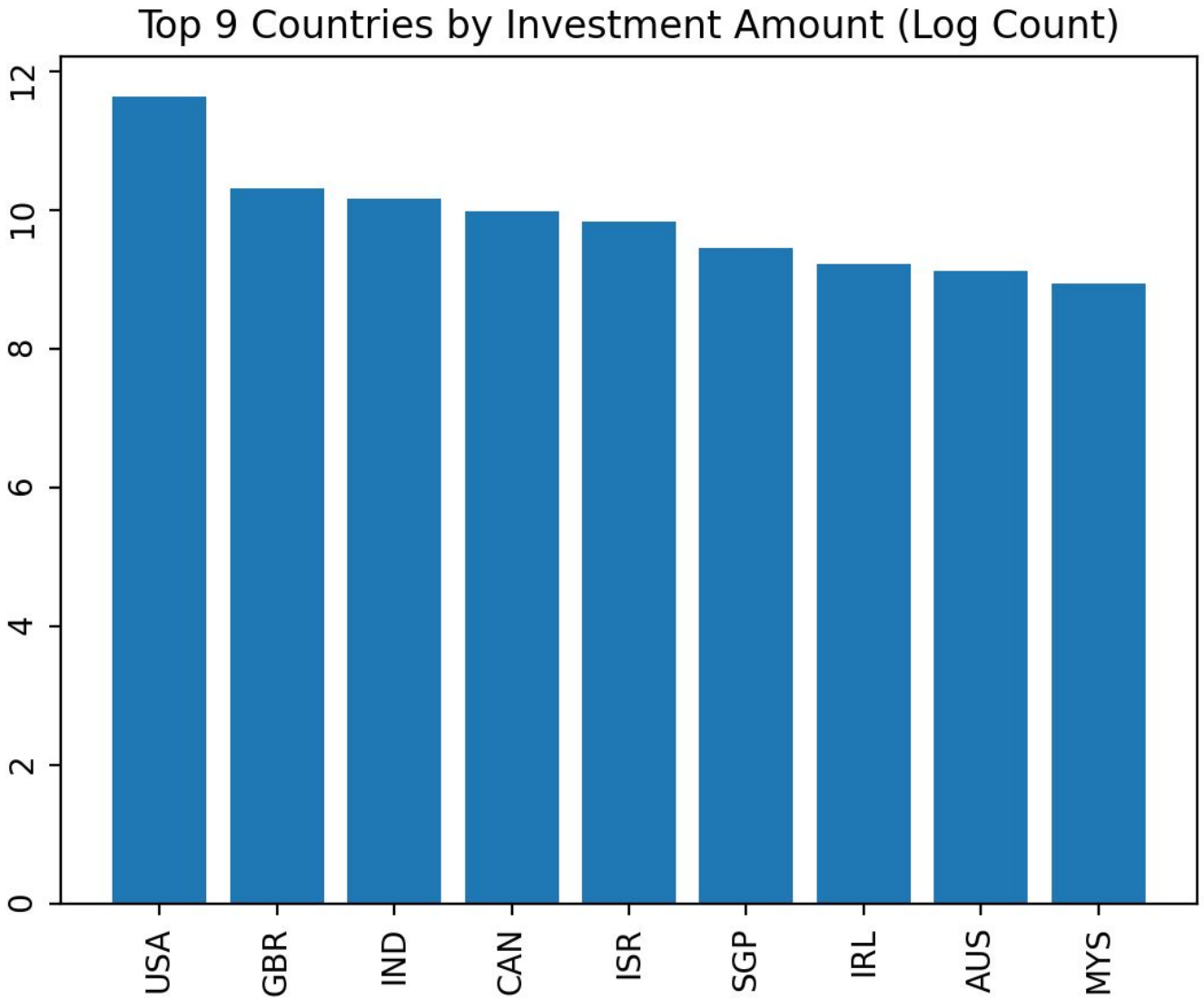
The breakdown of these stats by sector per country is also done separately.

Plots

Finally, the questions around highest-invested company in the top 2 sectors, with the relevant plots are created.



Investment Stats per Investment Type (after removing outliers greater than 90th Quartile


Top 9 Countries by Investment Amount (Log Count)
0
2
4
6
8
10
12
USA
GBR
IND
CAN
ISR
SGP
IRL
AUS
MYS

Investment for USAInvestment for GBR Investment for IND
0
500
1000
1500
2000
2500
3000
0
20
40
60
80
100
120
140
0
20
40
60
80
100
Others
Social, Finance, Analytics, Advertising
Cleantech / Semiconductors
Others
Social, Finance, Analytics, Advertising
Cleantech / Semiconductors
Others
Social, Finance, Analytics, Advertising
News, Search and Messaging

Conclusions and Advice

Based on Spark Funds' constraints, the investment type best suited to the constraint of USD 5 million - USD 15 million investment range, is Venture Investments, with a median of $4.75 million, and a 75th percentile of $10 million.

If they have some appetite for risk, private equity investments are also possible, because it has a median of just under $16 million, and a 25th percentile of $5.46 million. However, the number of venture investments (~35000) far exceeds the number of private equity investments (~1300), and thus venture investment wins out.

The top 3 English-speaking countries to invest in, in descending order, are:

- United States of America
- Great Britain
- India

The sectors which have received the most investment (count-wise) in these countries are:

- USA: Others, Social / Finance / Analytics / Advertising, Cleantech / Semiconductors
- GBR: Others, Social / Finance / Analytics / Advertising, Cleantech / Semiconductors
- IND: Others, Social / Finance / Analytics / Advertising, News / Search / Messaging