INVESTMENT ASSIGNMENT

SUBMISSION

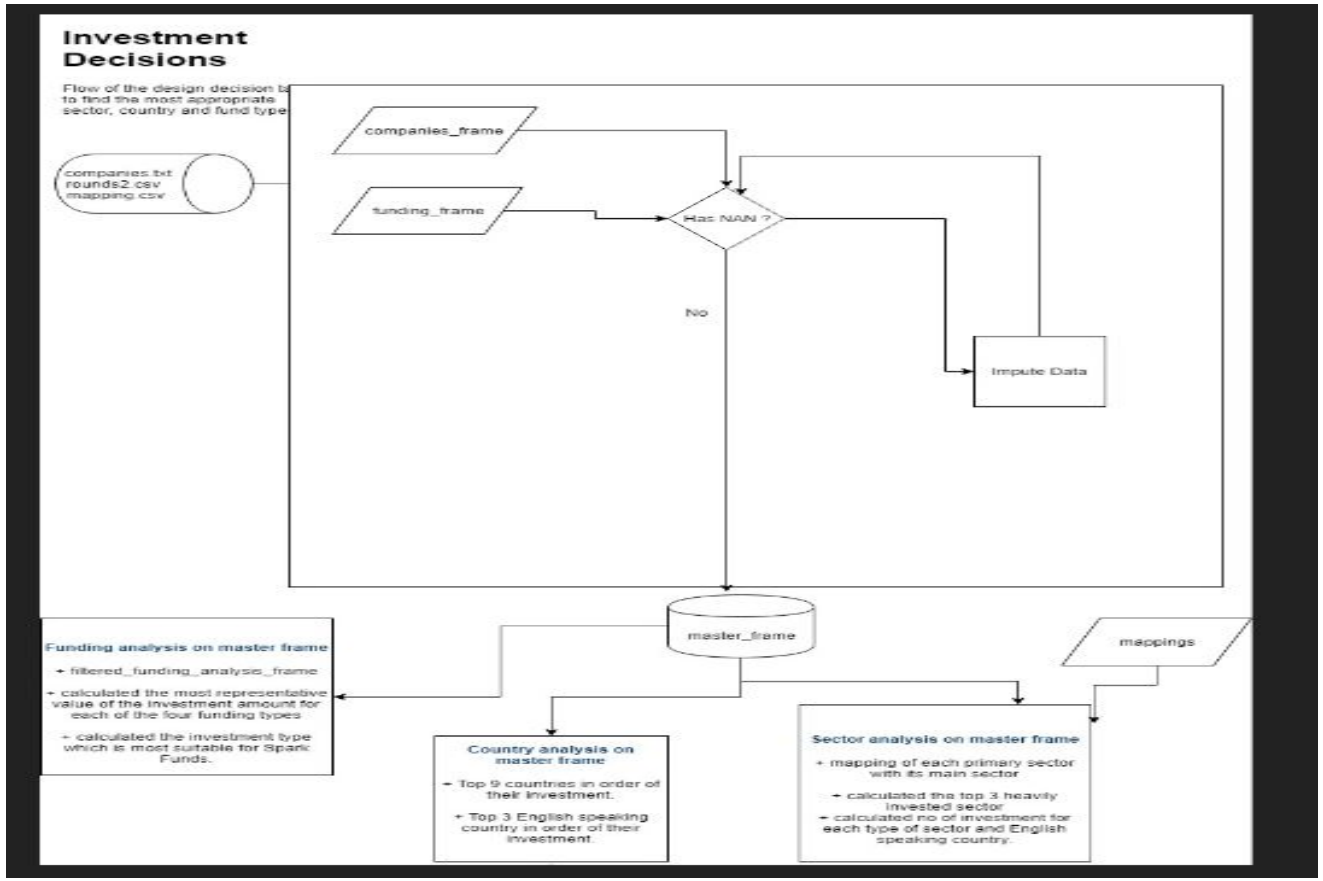
Name: Sourabh Mishra

<Abstract>

This assignment examines the strategy to find the best sector, country and suitable investment type. The objective of the case-study is to provide a report from which management can take decisions to invest where majority of the market is investing.

The analysis of the whole case study has three quite different sequential dimensions. Firstly, we have a phase of gathering the data related to case study. We have to make sure the data formatting is proper and all values are valid values. The robustness of the data employed is a crucial step in the data cleaning process. Secondly we have the data cleaning process, in which i made sure any empty or NAN values are handled properly. I also evaluated the contribution of different data attributes to make the final decision. Finally, we have analysis phase, where i have analysed the data based on the objective of the case-study.

<Problem solving methodology>



https://drive.google.com/file/d/1guLqQrq53w7dBI82thJ8fI4Tjq-JxeXQ/view?usp=sharing

<Analysis>

1. Funding Type Analysis:

After merging company and fund rounds data and keeping it as master_frame. I removed unncessary columns like ‘state_code’, ‘funded at’, ‘country_code’, ‘city’ etc as these fields are not required for funding type analysis.
I created a list of all the funding type in which ‘Spark Funds’ is interested such as ‘venture’, ‘angel’, ‘seed’ and ‘private equity’ and filtered the funding_analysis_frame based on it.

Using the aggregate functions i computed the most representative amout for each type of investment.
Based on the funding capacity of company i computed the best fund type as ‘Venture’ funds.

From Analysis point of view, these were the outcomes:

1) Private equity was the most funded type.
2) seed fund type was the least funded type.
3) The standard deviation of funding amount was quite high. Which reflected in the difference of its mean and median.

<Analysis>

2. Country Analysis:

Here based on the funding type selection from previous analysis as ‘venture’, i filtered the master frame and selected only those data which has fund type as Venture and selected top9 heavily invested countries.

I also did some webscapping to find the english speaking countries from wikipedia
https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language.

Based on the output table from scraped data i merged it to the top 9 countries and selected top3 most invested english speaking country.

Analysis points:

1) USA investment is 20 times better than England.
2) India was 4th in the list of most invested country.

<Analysis>

3) Sector Analysis

First of all i mapped the primary sector with the main sector using the mappings file. Then as we only need to do analysis for fund type as ‘Venture’ and top3 english speaking countries. I filtered the data based on these constraints.
Also i removed those rows which contains less than 5M or more than 15 M investment.

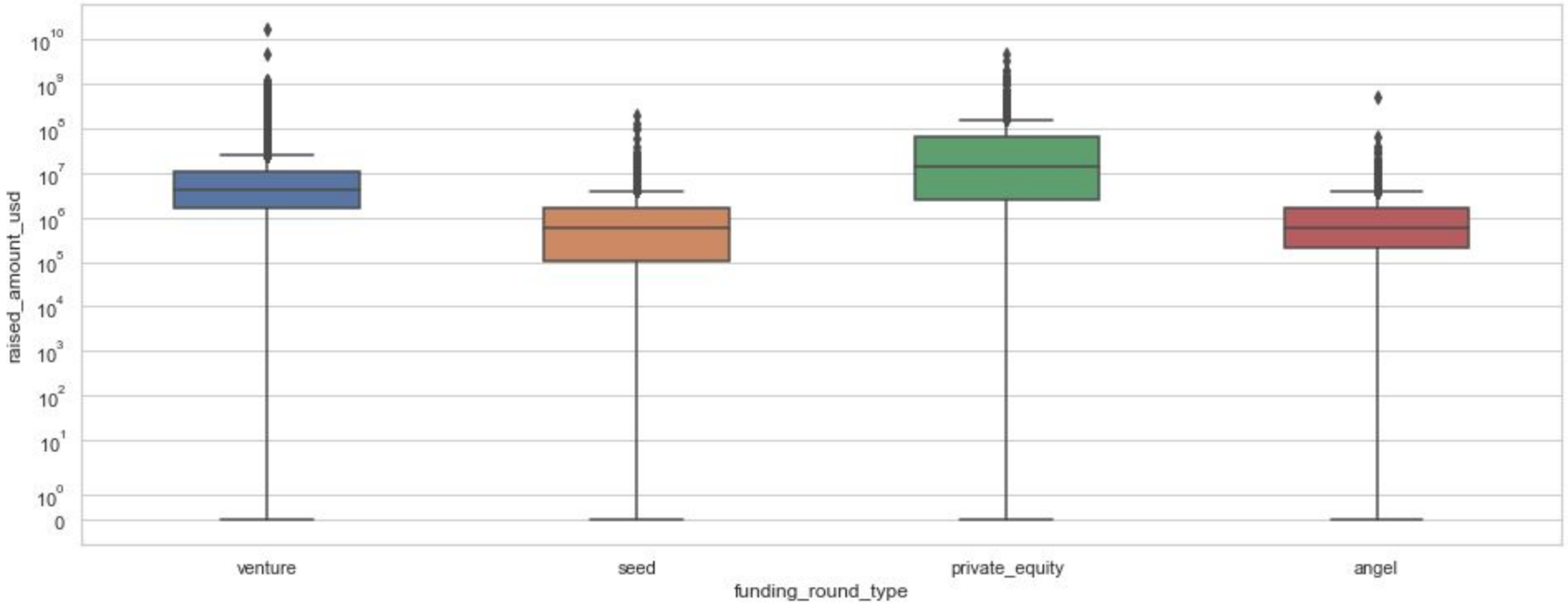
As per the problem statement i created three data frame namely D1, D2 and D3 for each top 3 country.

Analysis points:

1) There are only two sectors which is top across three countries.

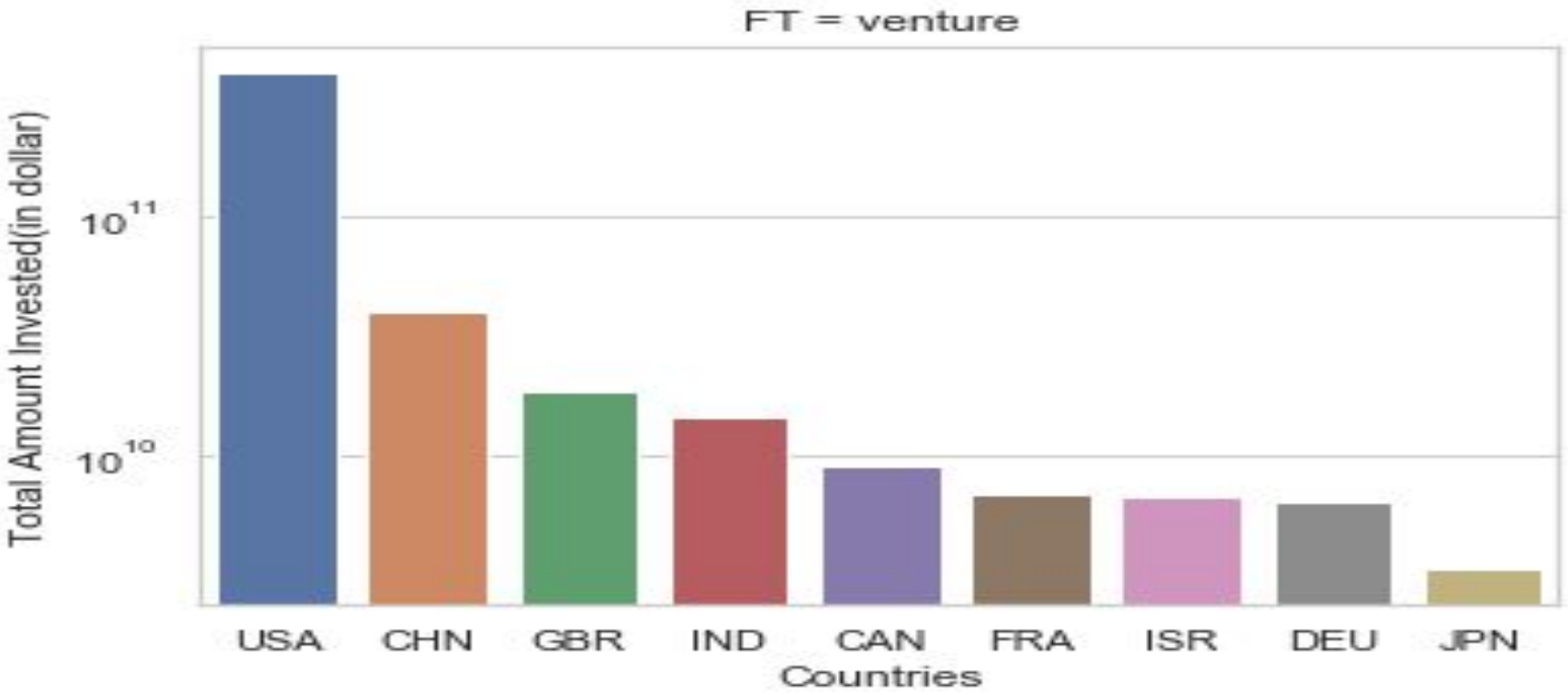
<Results>

Plot 1



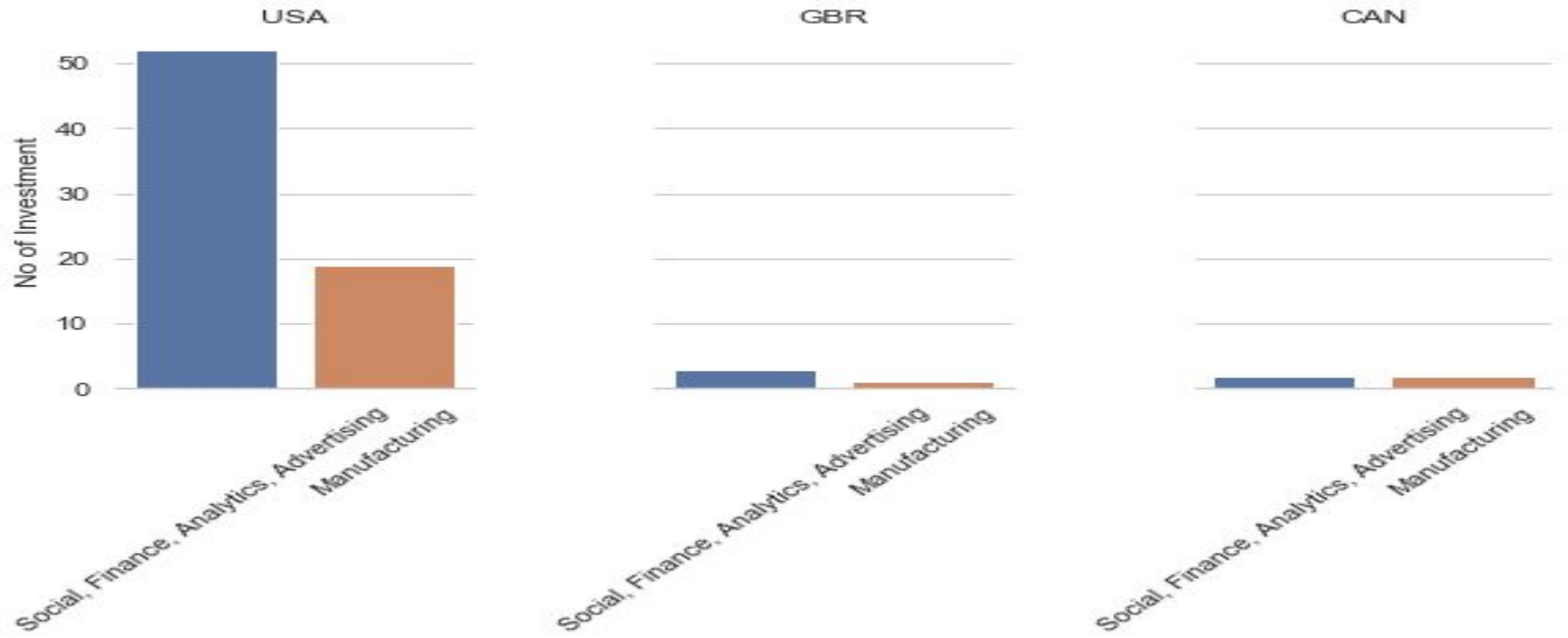
<Results>

Plot 2



<Results>

Plot 3



<Conclusions>

Assumption made for this case study:

1) Evaluating machine already has Anaconda installed and following packages are already installed: Pandas, numpy, requests, json, seaborn and matplotlib
2) Used ‘unicode_escape’ encoding to parse the input data.
3) Took %d-%m-%Y as the default format and parse all data accordingly.
4) I have removed closed status type company data.

Based on previous 3 analysis:

It appears the most of the investors are investing in venture funds. Also most of the investment is happening in USA and in Social, Finance, Analytics, Advertising sector.