INVESTMENT ASSIGNMENT

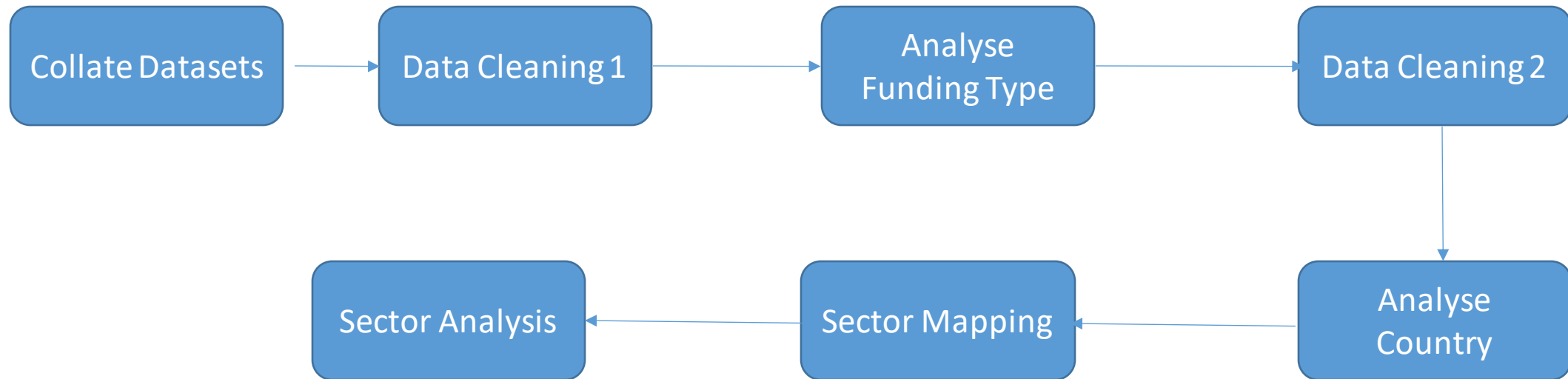# SUBMISSION

Name: Sujay Jayashankar

Problem Statement

Objective:

1. Analyse the data to understand the most suitable investment type for Sparks Fund, which has a budget between 5,000,000 and 15,000,000 USD.

2. For the chosen investment type, find the top three English speaking countries with maximum amount raised.

3. For these three countries, analyse the investments made in each sector.

Assumptions:

- Entries with missing data related to the above mentioned fields will be removed as these values can't be assumed.

- During matching of dataframes, data that do not match due to spelling errors will be corrected and matched, data that do not share any similarity will be removed.

Problem Solving Methodology

Funding Type Analysis

- Convert company unique ID attribute in 'companies' and 'ronuds2' dataframes into common string format

- Merge the two dataframes to include data of companies that are common to both dataframes

- Narrow down the analysis to only 'venture', 'seed', 'angel' and 'private equity' funding types

- Clean the dataframe in focus by removing entries with missing raised amount values

- Represent each funding type by the median of raised amount

- Decide on appropriate funding type to pursue and remove data for all other funding types
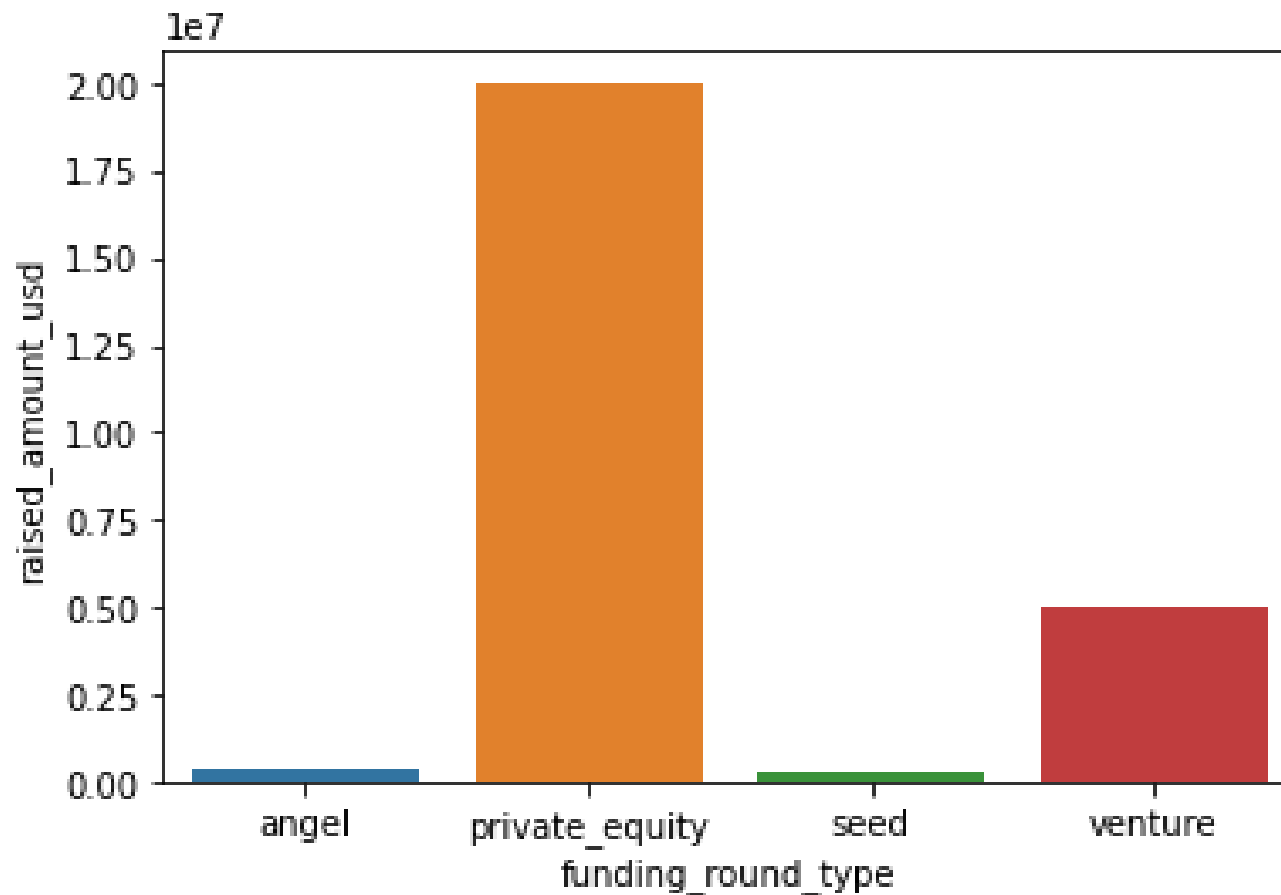
Country Analysis

- Remove entries with missing country codes

- Analyse the top 9 countries with highest raised amounts

- Filter out the Countries where Egnlish is not an official language

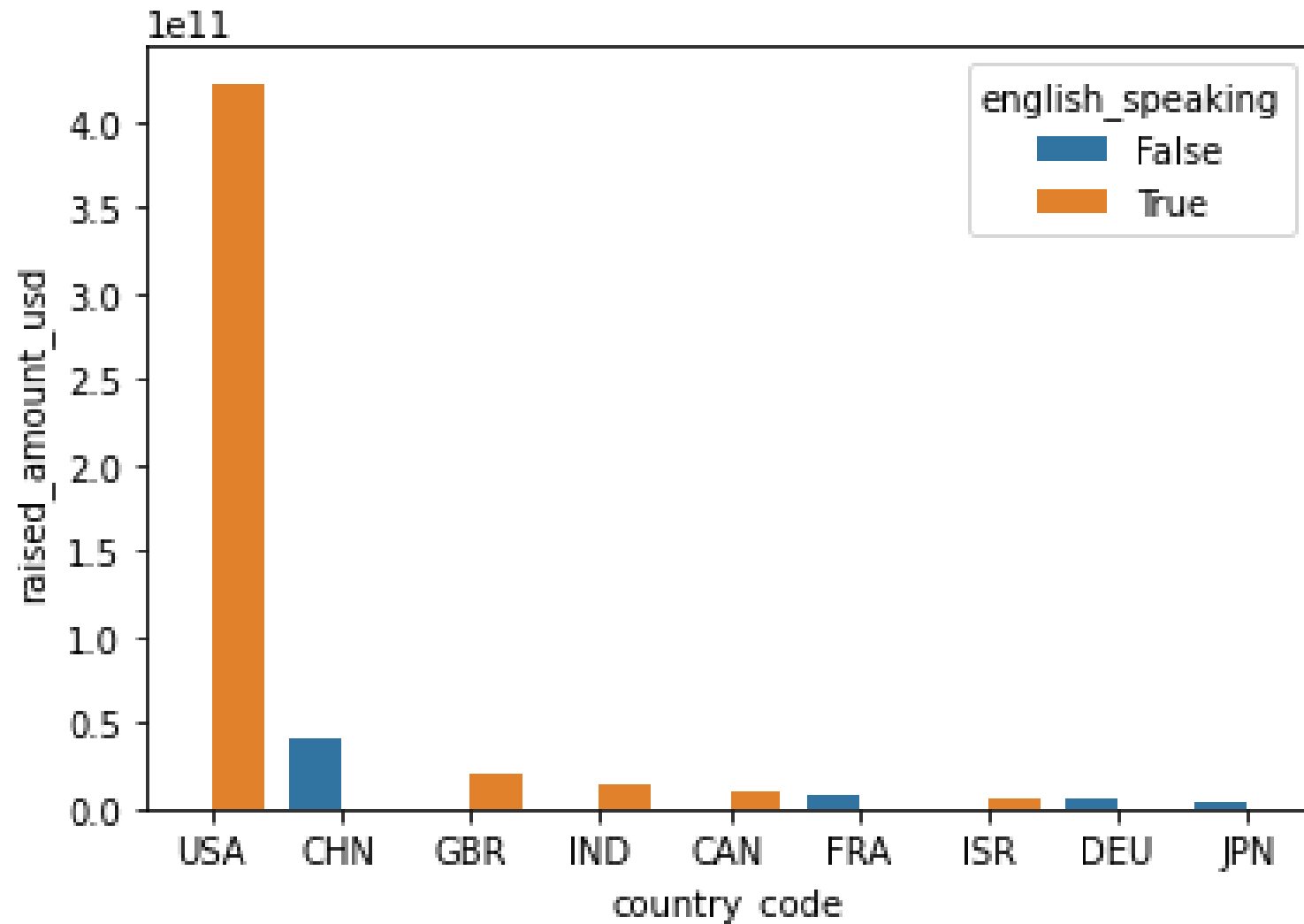- Select the top 3 English speaking countries and remove data for all other countries

Sector Analysis

- Understand the mapping file and add atribute to simplify the merge with dataframe in focus

- Clean any typos and errors in 'category_list' attribute in mapping file

- In the dataframe in focus, add a new attribute to identify a primary sector, this will be the first category listed in each row of the dataframe

- Map the mapping file with the dataframe in focus to add a new attribute for main sector

- Delete all entries which have a missing main sector

- Filter the dataframe to contain entries whose raised amounts are within 5,000,000 and 15,000,000

- Divide the dataframe into 3 dataframes each representing a top english speaking country

- Analyse the number of investments made in each sector for each country

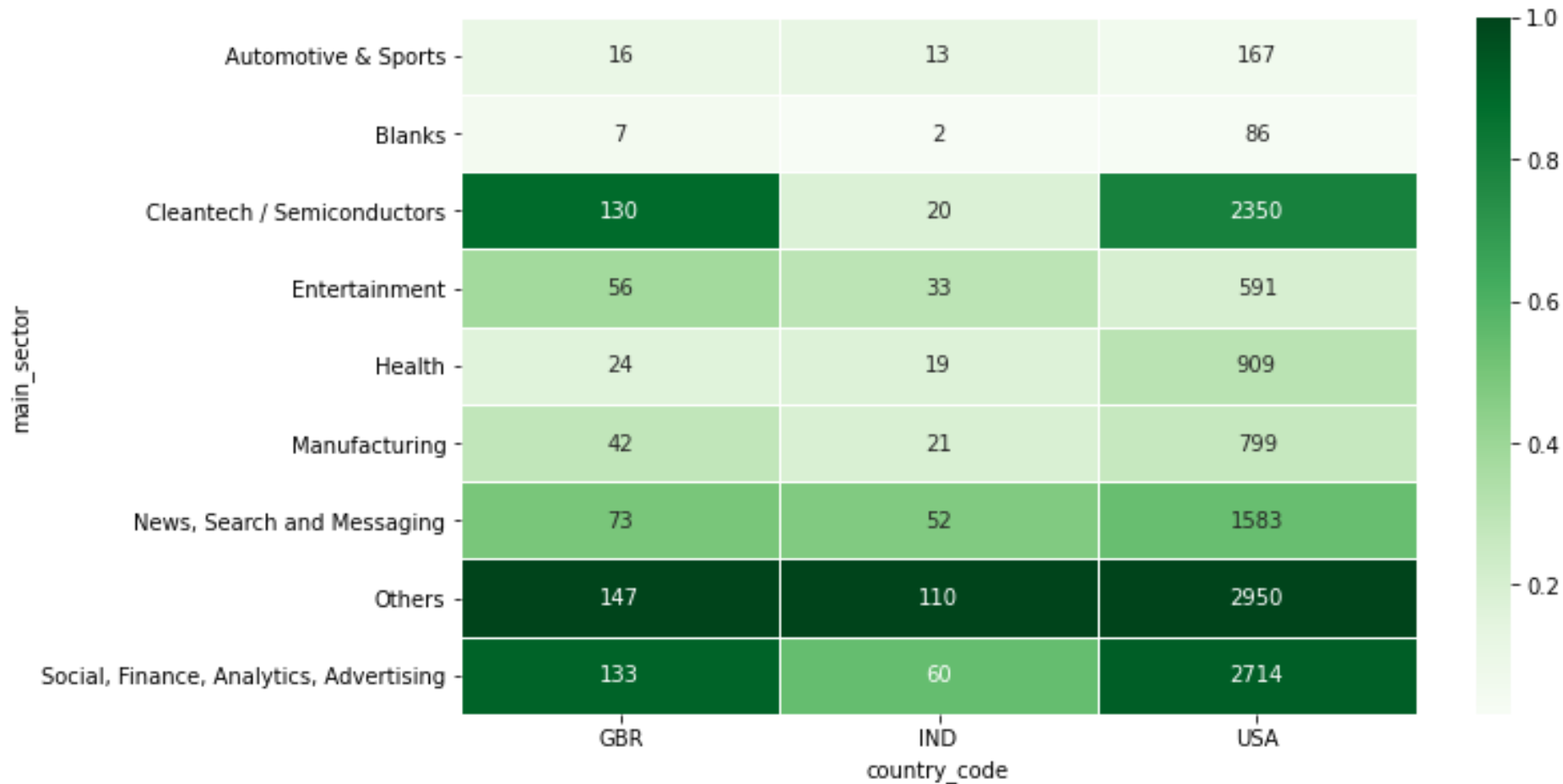- Analyse the total amount raised in each sector for each country

Result of Funding Type Analysis

Result of Country Analysis

Result of Sector Analysis

Conclusions

- The funding type 'venture' best suits the business budget range.

- The top three English speaking countries are USA, GBR and India.

- The sectors in which most number of investments are made in USA are (Desc. order) :
  - Companies in 'Others' sector
  - Companies in 'Social, Finance, Analytics, Advertising' sector
  - Companies in 'Cleantech / Semiconductors' sector

- The sectors in which most number of investments are made in GBR are (Desc. order) :
  - Companies in 'Others' sector
  - Companies in 'Social, Finance, Analytics, Advertising' sector
  - Companies in 'Cleantech / Semiconductors' sector

- The sectors in which most number of investments are made in India are (Desc. order) :
  - Companies in 'Others' sector
  - Companies in 'Social, Finance, Analytics, Advertising' sector
  - Companies in 'News, Search and Messaging' sector