

# INVESTMENT ASSIGNMENT SUBMISSION

Name: KUMARAGURU MUTHURAJ

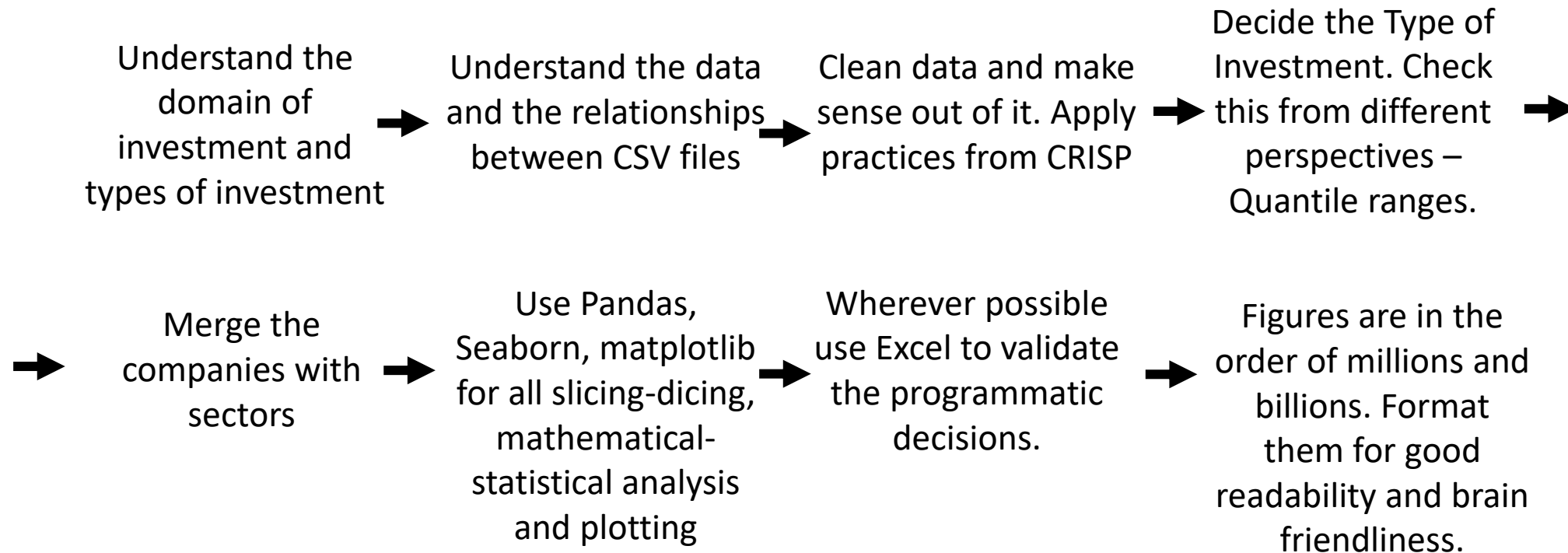
## **Abstract of the analysis**

**Spark Funds CEO will be enabled with the right set of data, based on data analytics, done out of investment data across the globe to help her arrive at a conclusion on list of prospective companies to invest, based on the following factors:**

- Investment type – Venture, Angel, Seed or Private Equity, for investment range of USD5M to 15M.**
- Sectors that are most sought by global investors.**
- Top 3 English speaking countries for ease of communication.**

**The analysis is done with 95000 investments done across the globe in different countries.**

Data from **crunchbase.com** – Companies, Investments done in companies and Company category mapping data.



## Critical Analysis – Cleaning data

- The final decisions are made out of number of investments made. Eliminating perceived bad data has to be done with care.
- Only data that is bad and impacts the decisions are removed like NaN values in investment amounts.
- Blank values in categories are to be retained as such and mapped to BLANKs main\_sector.
- Fixing the category name spelling errors was mandatory for a smooth merging of data.
- Getting a list of English speaking countries and cleaning them was important as we might miss a few that are listed on top.

## Critical Analysis – Identifying investment type

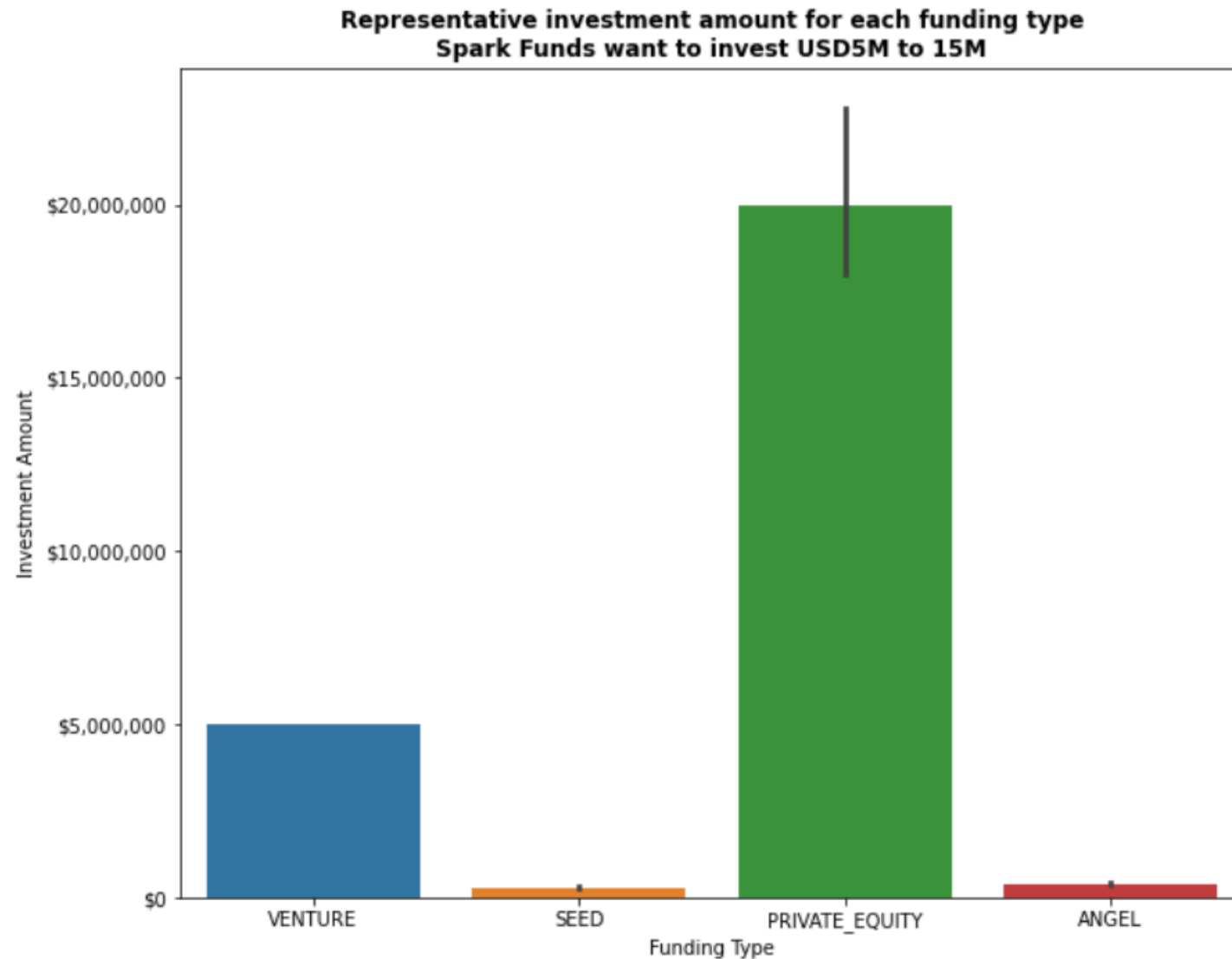
- **Representative amount for each of the 4 funding types ['ANGEL', 'SEED', 'VENTURE', 'PRIVATE\_EQUITY']** - Records with NaN values investment amounts are removed. Median is the most representative amount for a population of numeric values. The median along with the boxplots are calculated and displayed for each of the following range of Quantiles.
  - 100% data is considered without removing outliers. [400K, 275K, 5M, 20M]
  - Data between 0 to 95th percentile. [368K, 250K, 4.75M, 18M]
  - Data between 5 to 95th percentile. [400K, 280K, 5M, 20M]
  - Data between 25 to 75th percentile. [439K, 300K, 5M, 20M]

The result is that the Median for the 4 cases don't change drastically. The data clearly indicates that the only investment type that falls close/under 5M to 15M is **VENTURE**. We choose to move with this type.
- **Because, the decisions are based out of investment counts, removing outliers is NOT a good idea, for the reason that they will adversely impact the sector that we arrive at. Outliers for the type VENTURE is not removed for processing further.**

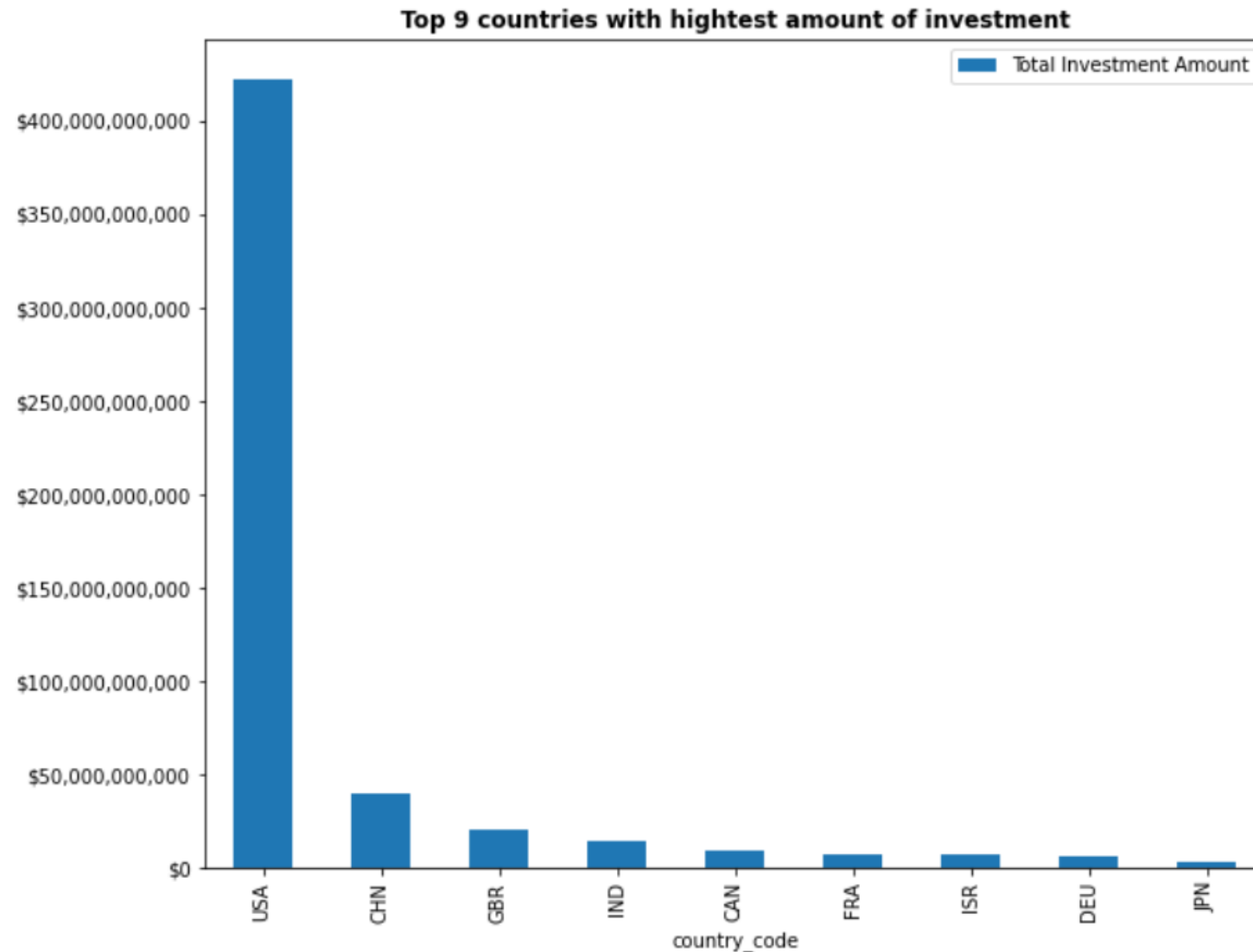
## Critical Analysis – Choice of programming constructs

- Pandas is extensively used to work through this analytics assignment.
- Dataframes are completely sliced-diced with Pandas, Pivot-tables extensively. Seaborn and Matplotlib are used for display.
- Usage of built-in functions are done in most of the places for performance.
- Formatting the dollar amounts are done to give a more realistic, brain-friendly display that is easy to understand.
- This entire analysis fits in to the case of Segmented Univariate Quantitative analysis, where investment amount was analysed from different sectors, countries and investment types.

## Venture capital type funding is best suited

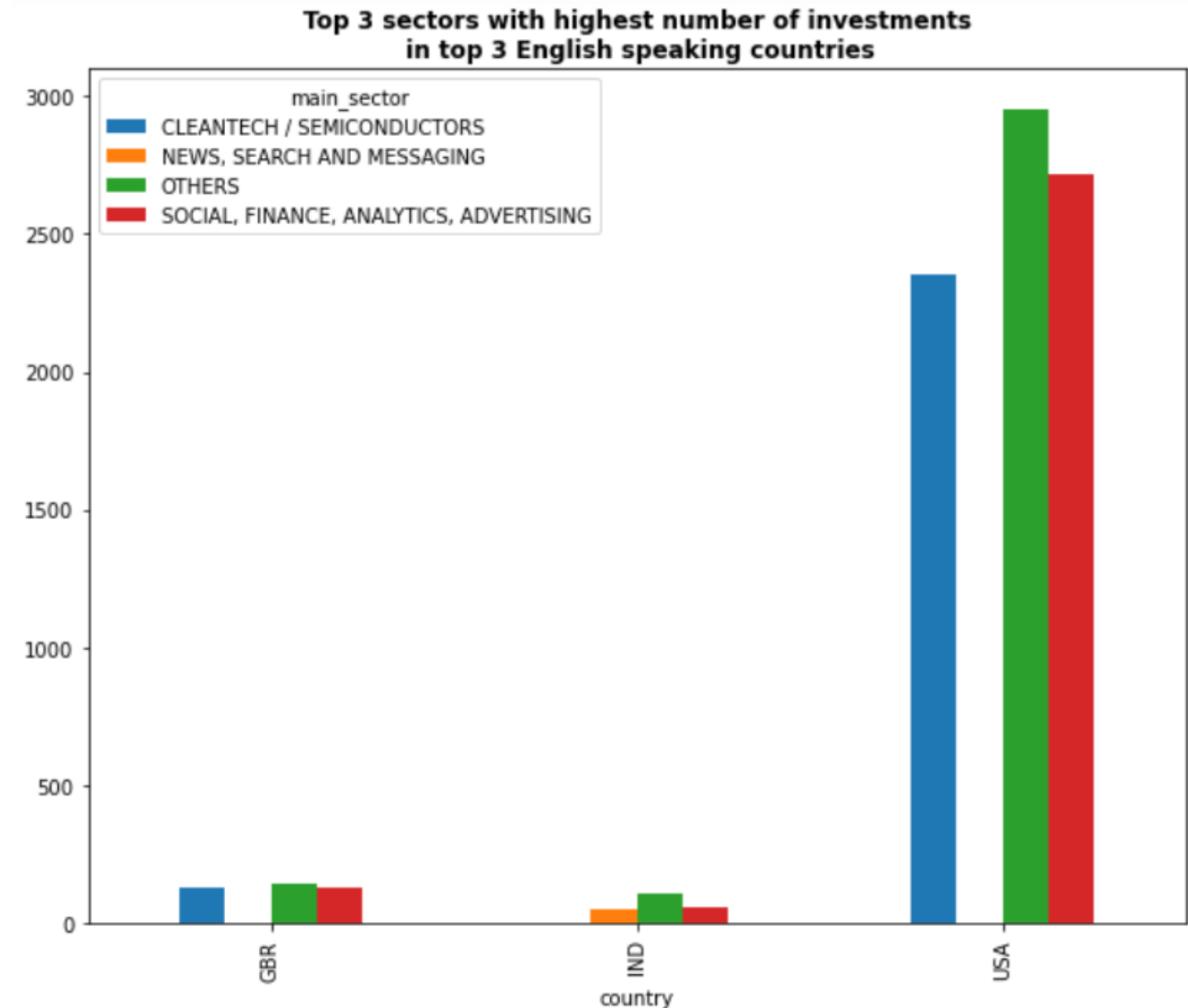


## Top 9 countries where investments are heavily made





## Top 3 sectors in top 3 English speaking countries. USA is most sought and Britain – India trail by multi-fold times



## Conclusions

- The top countries to choose from for Spark funds to invest are USA, Great Britain and India.
- The investment type is Venture, because this is the type that falls between 5M to 15M out of the pattern seen in the investment data.
- The top 9 countries where investments were made actively are USA, China, Great Britain, India, Canada, France, Israel, Germany and Japan. Of these, the top 3 English speaking countries to go after are USA, Great Britain and India.
- The top sectors in these 3 countries are cleantech - semiconductors, news search – messaging, social-finance-analytics-advertising and others.