Startup Investment Project

Safa Eladib - Mohamad Doughan - Jean-Denis Drane

INTRODUCTION & CONTEXT

- In light of the global recession coupled with the sudden spread of the Covid-19 pandemic, a good friend of ours has unfortunately been forced to quit his job due to the economic and financial downturn that his company was facing. As a result, he thought that it would only seem reasonable in such dire times to venture in a personal business startup. Unluckily, our friend has no information on the global startup scene and has asked us to brief him on it so that he can go ahead and plan his next steps

- Key features to include in our debrief:
  - Market analysis on the leading 10 countries in terms of total startups launched
  - Market analysis on the leading 10 countries in terms of aggregate fundings for startups
  - Market analysis on the distribution of stage funding for leading 10 countries
  - Trending startup categories
  - A close-up on one of the categories expected to skyrocket (E-Learning)

IRON HACK

01
DATA ACQUISITION

- Dataset available in kaggle: https://www.kaggle.com/arindam235/startup-investments-crunchbase#investments_VC.csv

- DataSet reporting on startup businesses globally till 2014:
  - Shape: 54294 rows, 39 columns
  - Company information
  - Investment types
  - Investments rounds

- Tools used:
  - Jupyter Notebook
  - Spyder

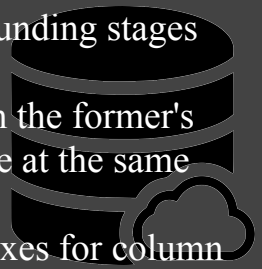
02
WEB SCRAPING

- Web Scraping to update our data with startups after the year 2014
  https://www.startupranking.com/
  - Limited our study on the top 200 startups
  - Cleaning and aggregating data
  - Complementing our analysis from initial dataset

- Tools used:
  - Selenium
  - Jupyter Notebook
  - MySQL
  - Spyder

03
DATA CLEANING

- Dropping all rows where the entire instances have a NAN value (> 5000 rows)
- Stripping all leading and trailing white spaces
- Creating a new DF with all startups that have duplicated name (~ 200 startups)
    - Dropped all duplicated names where they have the same Name, Homepage URL, and Country Code
- Drop all rows where the instances inside Category List, Market, Funding Total usd, Country Code, Region, & Market
- Amending the category list by adjusting the text string and sorting words alphabetically
- Amending the total funding column and converting the resulting data into a Float type
- Since we have total investment broken down into funding stages, it was a good practice to sum all funding stages for each startup and comparing the value with the Funding Total usd column
    - For every instance where the later is missing and the former has a value, replace the later with the former's
- Checking that the null values in the 3 columns (founded_quarter, founded_year, founded_month) are at the same instances
- Checked that all indexes in column founded_at where the value in NaN is a subset in the list of indexes for column founded_month where the value is NaN

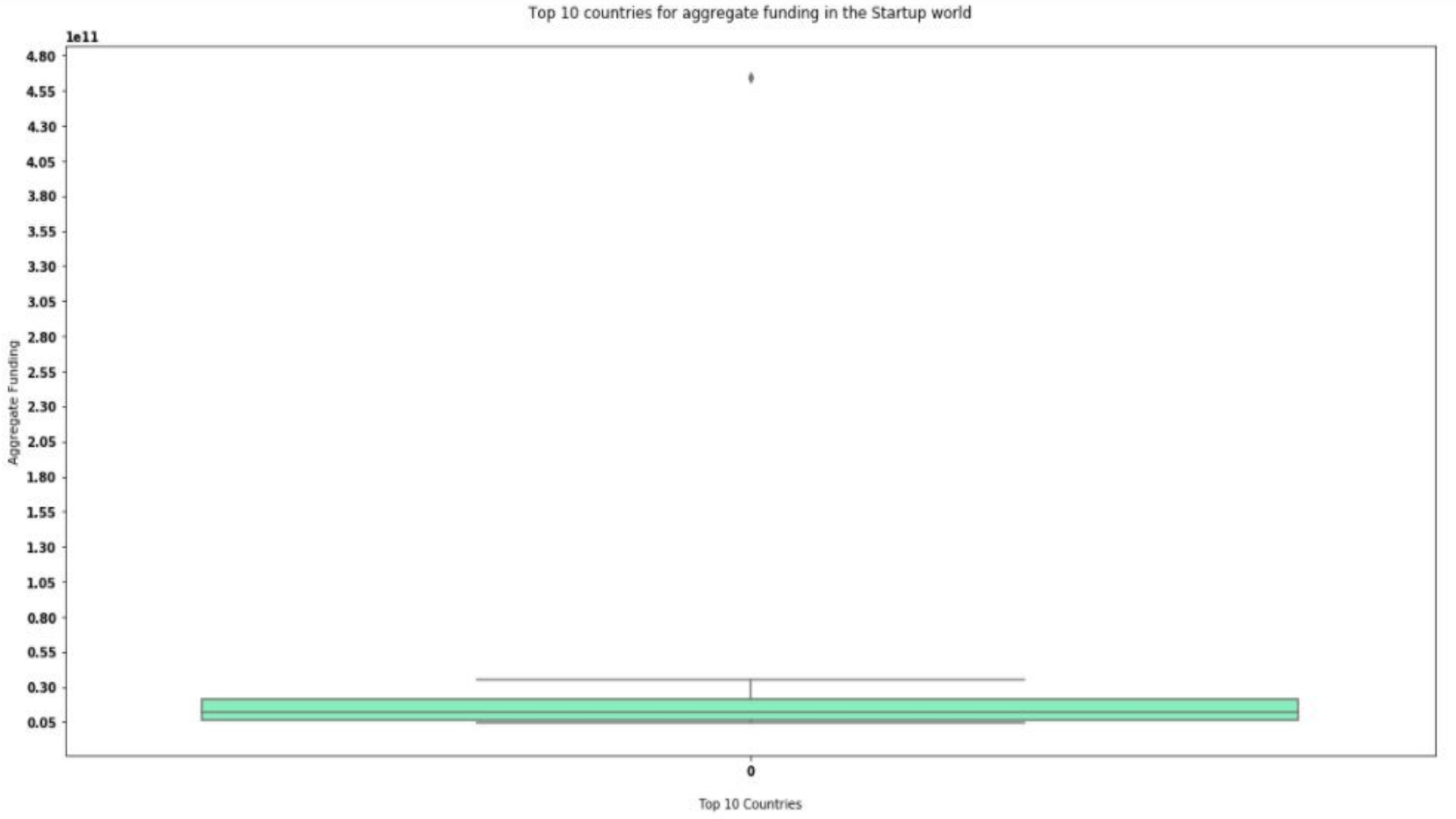


03
DATA CLEANING

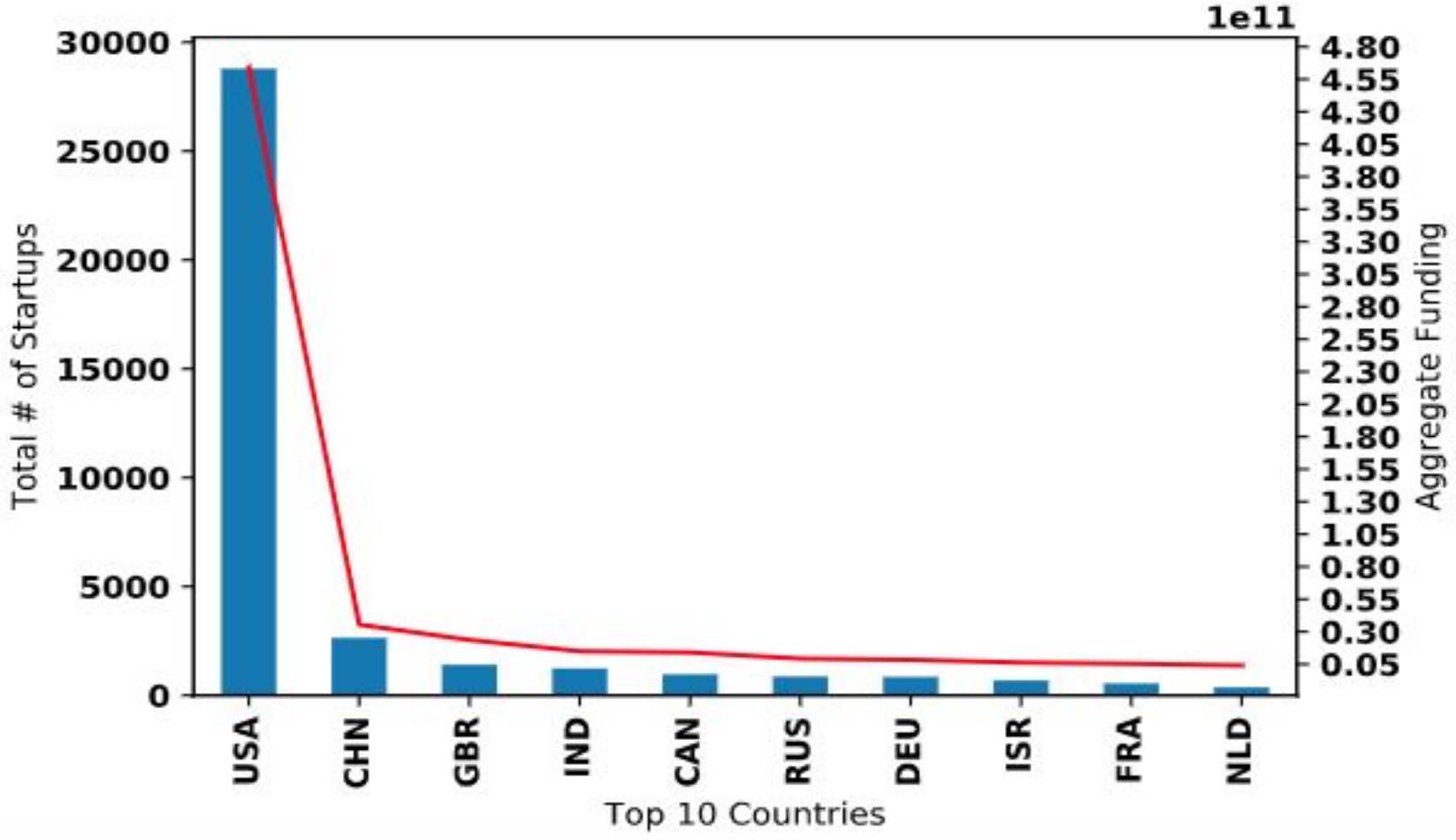
- Function to study the correlation between the 'founded_at' and 'first_funding_at' to check whether we can replace the missing values in column (founded_at) with the values of (first_funding_at)
- Check if all NaN instances in country and region are a subset of all NaN instances in City
- For any missing country code, if we have another instance that has same region and a valid Area -> Replace missing country code with the valid one
- Replacing missing values in the category list with values in market list
- Resetting index

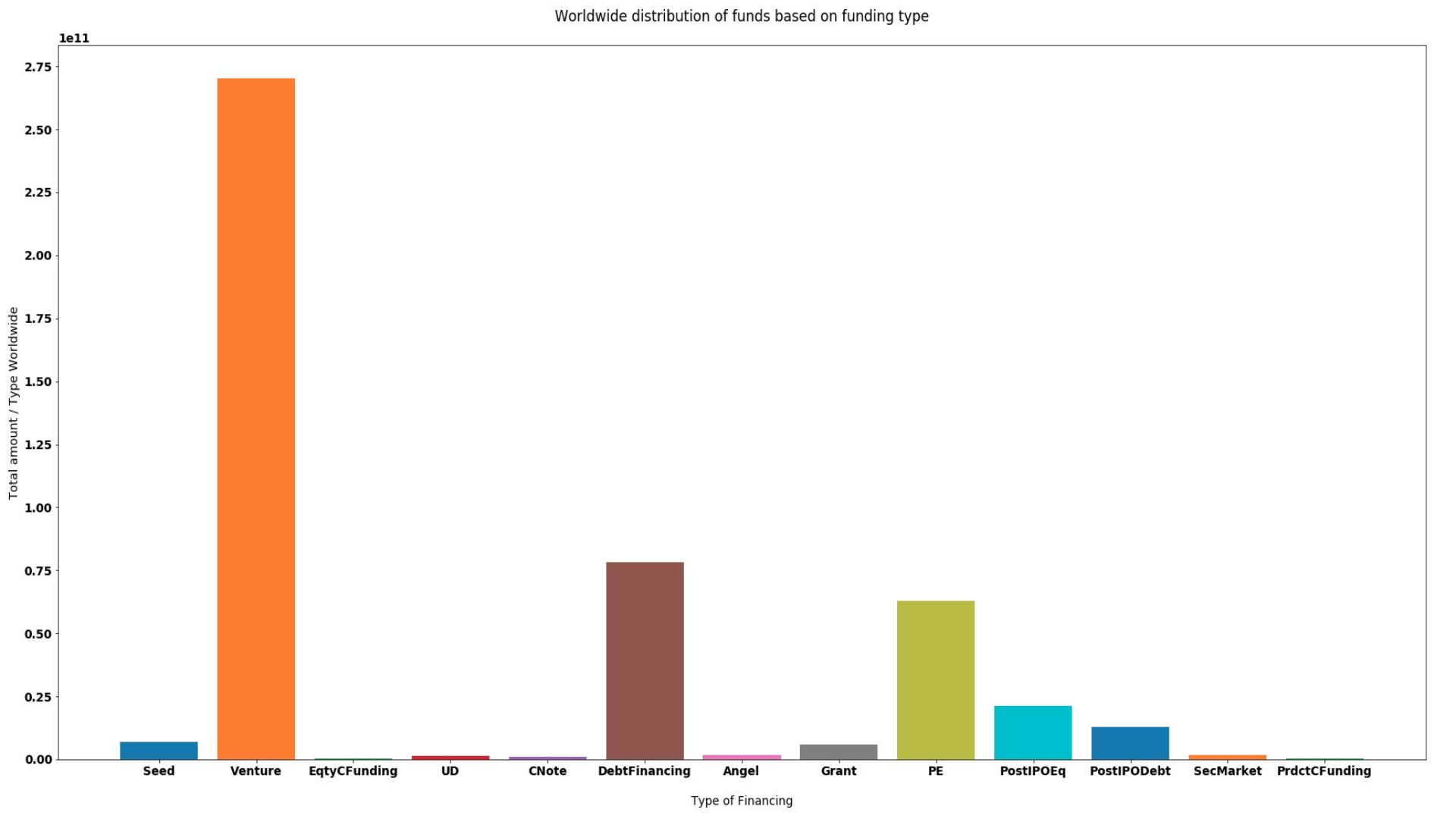
04

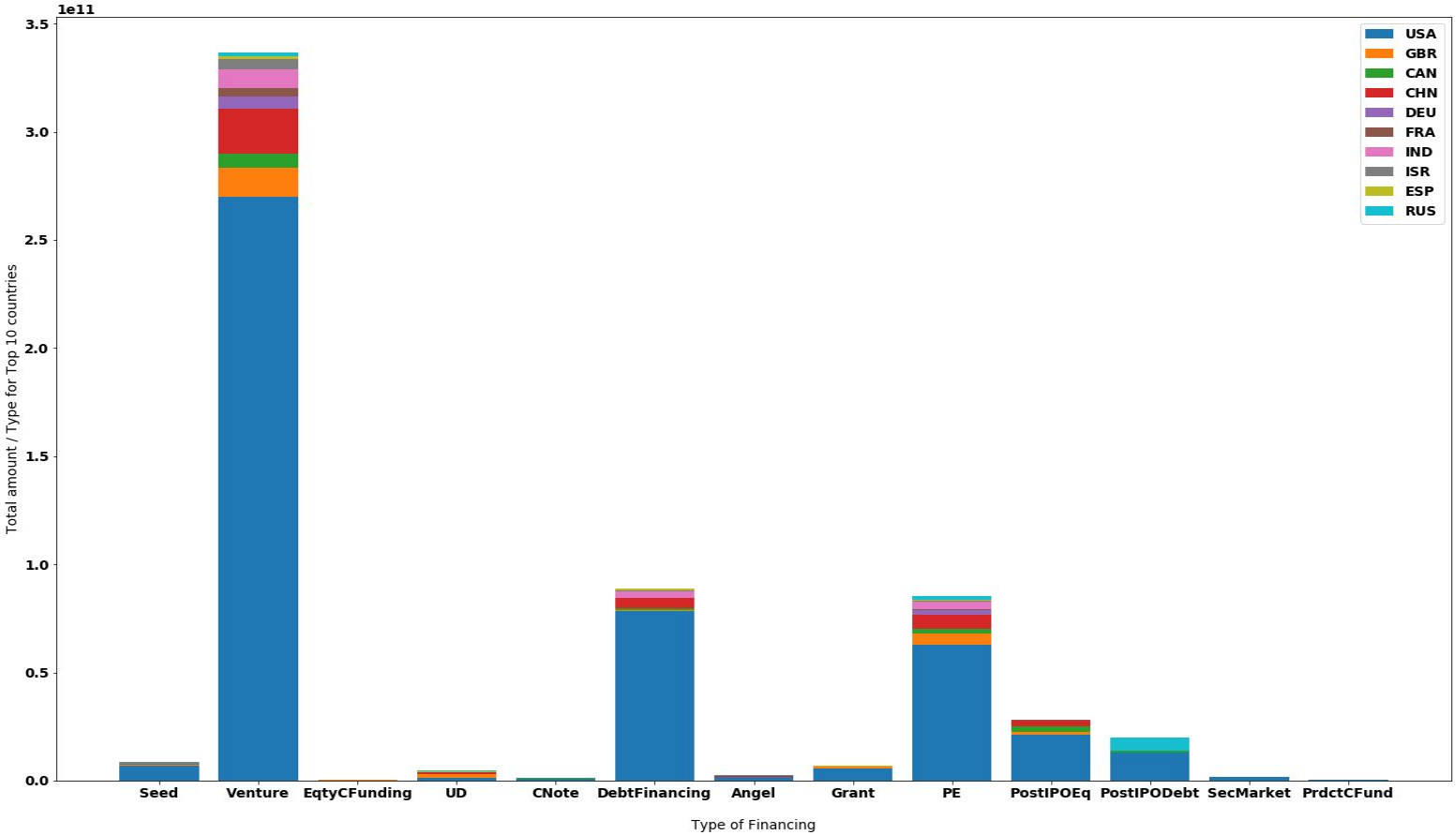
ANALYSIS & VISUALIZATION

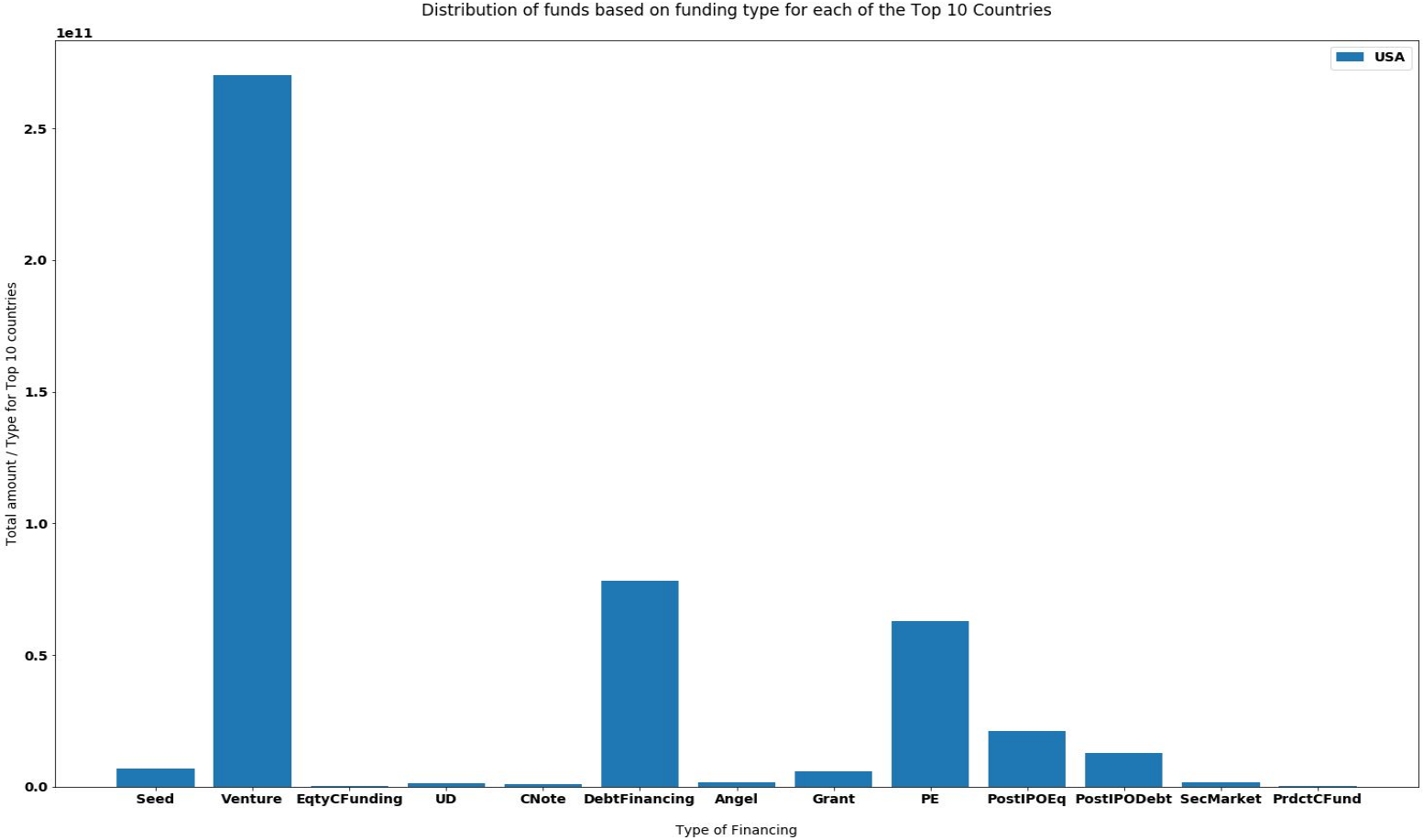

Top 10 countries for aggregate funding in the Startup world
1e11
4.80
4.55
4.30
4.05
3.80
3.55
3.30
3.05
2.80
2.55
2.30
2.05
1.80
1.55
1.30
1.05
0.80
0.55
0.30
0.05
Aggregate Funding
0
Top 10 Countries

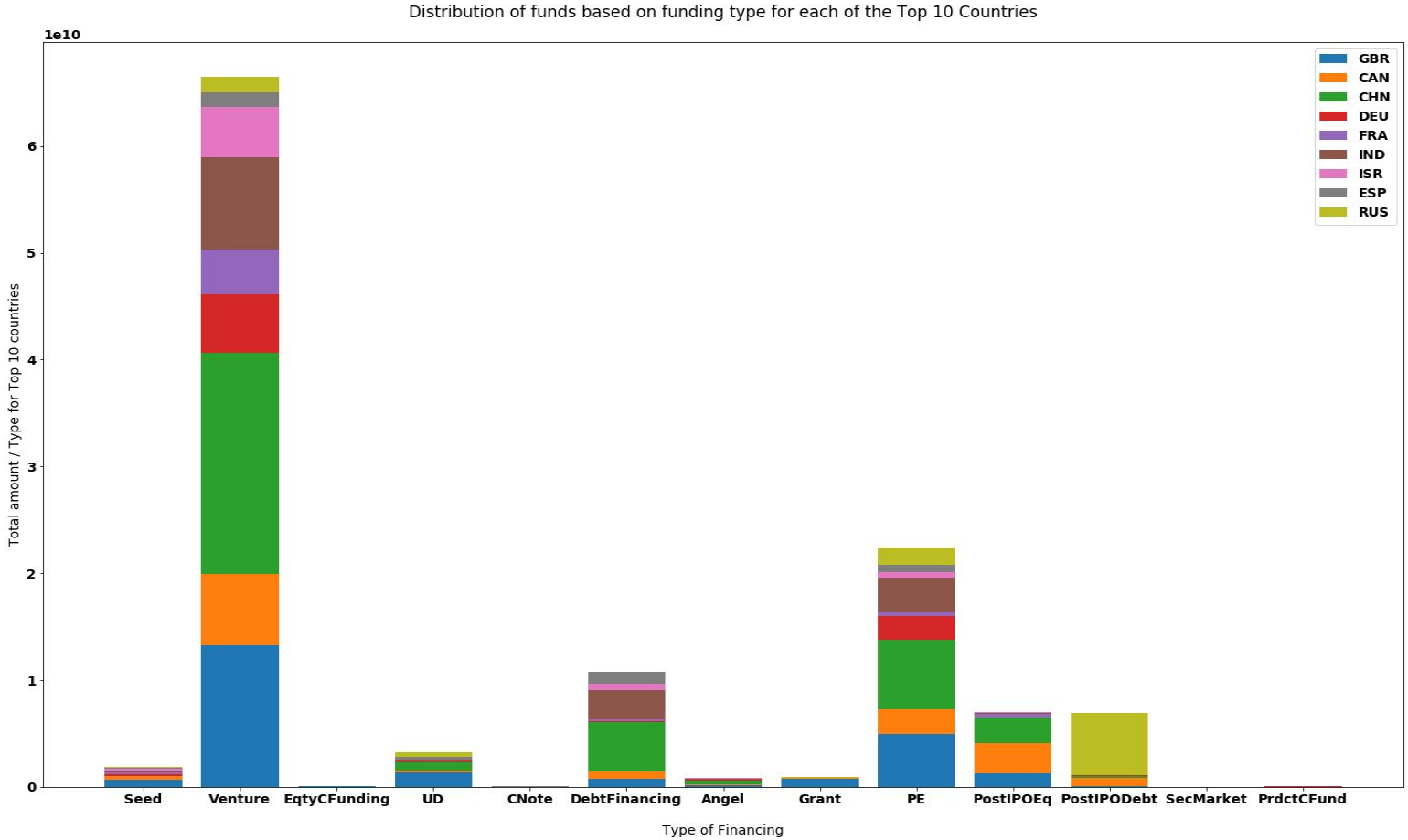
Relation between the Total # of startups and Average Funding for Top 10 countries

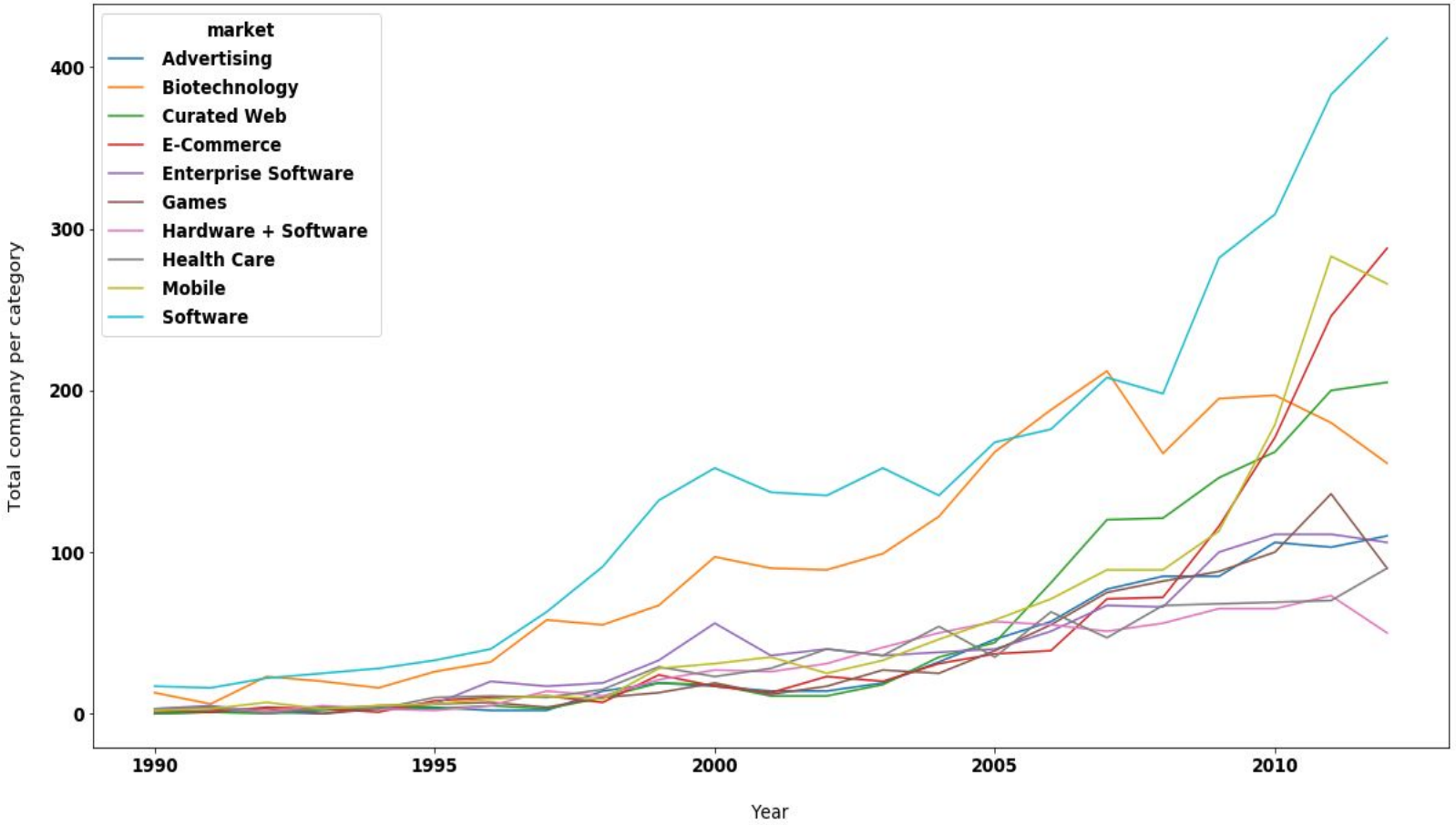


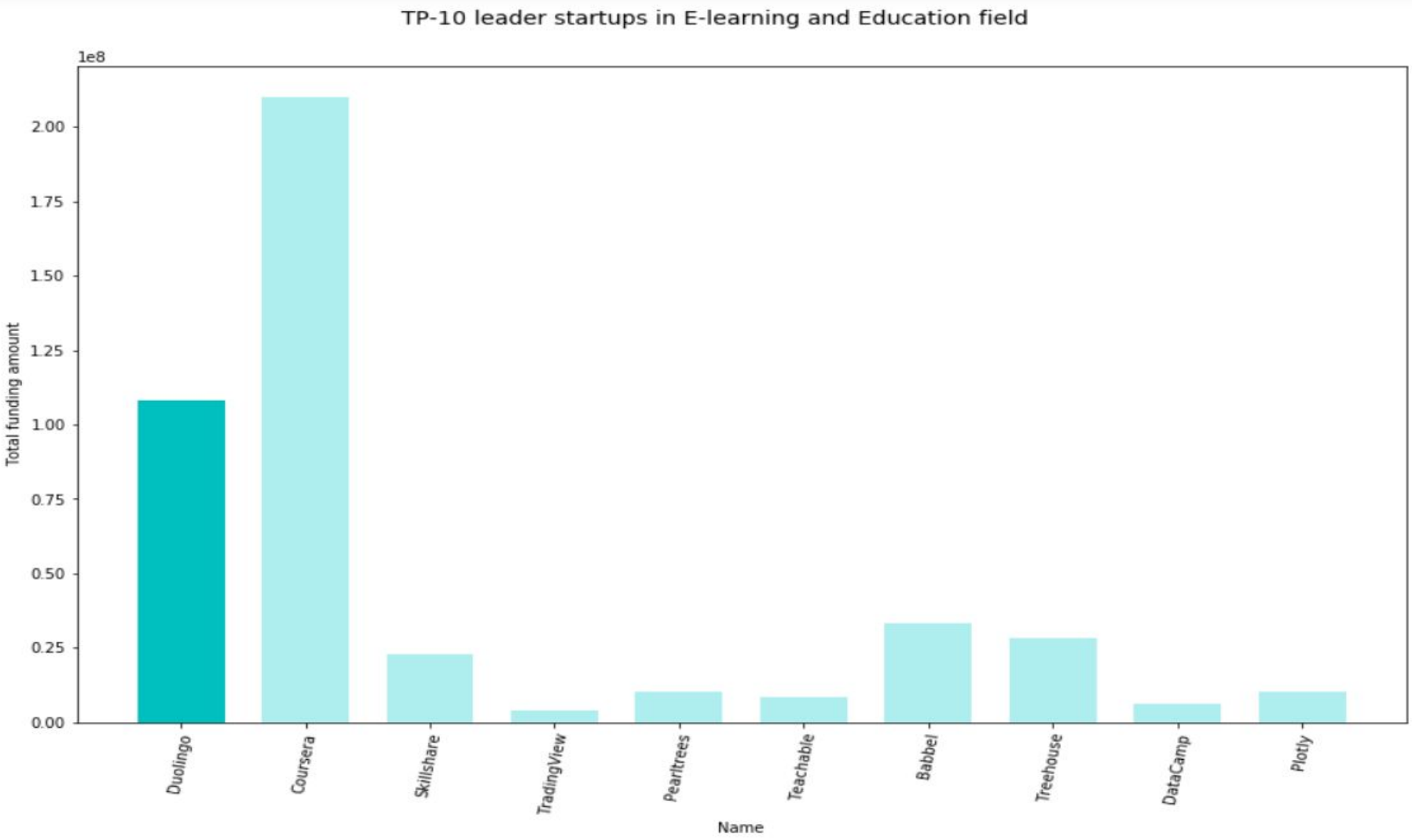

Worldwide distribution of funds based on funding type
1e11
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total amount / Type Worldwide
Seed
Venture
EqtyCFunding
UD
CNote
DebtFinancing
Angel
Grant
PE
PostIPOEq
PostIPODebt
SecMarket
PrdctCFunding
Type of Financing


Distribution of funds based on funding type for each of the Top 10 Countries
1e11
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Total amount / Type for Top 10 countries
USA
GBR
CAN
CHN
DEU
FRA
IND
ISR
ESP
RUS
Seed
Venture
EqtyCFunding
UD
CNote
DebtFinancing
Angel
Grant
PE
PostIPOEq
PostIPODebt
SecMarket
PrdctCFund
Type of Financing


Distribution of funds based on funding type for each of the Top 10 Countries
1e11
2.5
2.0
1.5
1.0
0.5
0.0
Total amount / Type for Top 10 countries
USA
Seed
Venture
EqtyCFunding
UD
CNote
DebtFinancing
Angel
Grant
PE
PostIPOEq
PostIPODebt
SecMarket
PrdctCFund
Type of Financing


Distribution of funds based on funding type for each of the Top 10 Countries
1e10
Total amount / Type for Top 10 countries
Type of Financing
GBR
CAN
CHN
DEU
FRA
IND
ISR
ESP
RUS
Seed
Venture
EqtyCFunding
UD
CNote
DebtFinancing
Angel
Grant
PE
PostIPOEq
PostIPODebt
SecMarket
PrdctCFund


market
Advertising
Biotechnology
Curated Web
E-Commerce
Enterprise Software
Games
Hardware + Software
Health Care
Mobile
Software
Total company per category
0
100
200
300
400
1990
1995
2000
2005
2010
Year


TP-10 leader startups in E-learning and Education field
1e8
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Total funding amount
Duolingo
Coursera
Skillshare
TradingView
Pearltrees
Teachable
Babbel
Treehouse
DataCamp
Plotly
Name

Challenges & Opportunities

All our knowledges on python and more...

Team work - Use of trello to organize tasks and plan the main milestones

Merging code

Time Constraints

IRON HACK

NOW LET SEE THE CODE