

VISUALIZATION PLATFORM

**Source of data**: The data is taken from a Private Equity firm based on New York, US. Primarily the attributes of the dataset have 3 sources:

1. LinkedIn
2. Crunchbase and
3. Dealcloud

**Objective:** The client is a Private Equity investor. Iterative research and business analysis have been done on the data itself to find the prospect(companies) for the client. This data is a subset of the whole data and represents only some of the Client’s qualified companies. The Qin companies are eventually the prospects which we are looking for. Here are some conditions to make a company qualified out, -

1. If the company already exists inside Dealcloud, then the company may be marked as QO because it means that the client is already interacted with the particular company and came up with a decision.
2. FTE has to be within the range of 0 and 125.
3. Capital raised or last funding amount has to be lesser than $10m.
4. Whether the company is Solutions Co.
5. Whether the company is Services Co.
6. Software product not found.
7. Hardware product found.
8. Not in target geography (explained later)
9. Whether the company is acquired by other bigger companies.
10. Whether the company follows B2C model.
11. Custom Software development.
12. Reseller.
13. Partnership with bigger companies.
14. Third party logistics.
15. Website not working / not found.
16. Cyber security.
17. Infrastructure.
18. IT Support.
19. Consultancy.
20. Accountancy.
21. Public firm.
22. Market place.
23. Website is in different language other than English.

Although we have considered only the Qin companies as the data quality is much better for the Qin rows in comparison with the QO rows.

**Condition for being qualified in:** The company must have its own B2B software product/platform/ Cloud software, SaaS.

**Target geography:** The companies coming from the following states can only be considered as Qin, otherwise it will be QO.

**QIN areas:** Alabama, Arizona, Arkansas, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Pittsburgh, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Northern Virginia, Rest of Virginia, West Virginia, Wisconsin, Wyoming, Quebec, Ontario, Virginia.

**Industry – Sector – Subsector mappings:**

|  |  |  |
| --- | --- | --- |
| **INDUSTRY** | **SECTOR** | **SUB-SECTOR** |
| Accounting | 2. Horizontal | Finance |
| Airlines/Aviation | 1. Vertical | Travel & Hospitality |
| Alternative Dispute Resolution | 2. Horizontal | GRC |
| Alternative Medicine | 4. Healthcare | Life Sciences |
| Animation | 1. Vertical | Media & Telecom |
| Apparel & Fashion | 1. Vertical | Retail |
| Architecture & Planning | 1. Vertical | AEC |
| Automotive | 1. Vertical | Auto |
| Aviation & Aerospace | 1. Vertical | Travel & Hospitality |
| Banking | 2. Horizontal | Finance |
| Biotechnology | 4. Healthcare | Life Sciences |
| Broadcast Media | 1. Vertical | Media & Telecom |
| Building Materials | 1. Vertical | AEC |
| Business Supplies and Equipment | 2. Horizontal | EAM / FM |
| Capital Markets | 2. Horizontal | Finance |
| Chemicals | 1. Vertical | Manufacturing |
| Civic & Social Organization | 1. Vertical | Non-profit |
| Civil Engineering | 1. Vertical | AEC |
| Commercial Real Estate | 1. Vertical | Real Estate |
| Computer Games | 1. Vertical | Media & Telecom |
| Computer Hardware | 3. Infrastructure | Systems Mgmt |
| Computer Networking | 3. Infrastructure | Network / Storage |
| Computer & Network Security | 3. Infrastructure | Security |
| Computer Software | 2. Horizontal | IT |
| Construction | 1. Vertical | AEC |
| Consumer Electronics | 1. Vertical | Retail |
| Consumer Goods | 1. Vertical | Retail |
| Consumer Services | 2. Horizontal | Cust. Support / Svcs |
| Cosmetics | 1. Vertical | Retail |
| Dairy | 1. Vertical | Retail |
| Defense & Space | 1. Vertical | Government |
| Design | 1. Vertical | Media & Telecom |
| Education Management | 1. Vertical | Education |
| E-Learning | 1. Vertical | Education |
| Electrical/Electronic Manufacturing | 1. Vertical | Manufacturing |
| Entertainment | 1. Vertical | Media & Telecom |
| Environmental Services | 1. Vertical | Energy & Natrl. Resources |
| Events Services | 1. Vertical | Travel & Hospitality |
| Facilities Services | 2. Horizontal | EAM / FM |
| Farming | 1. Vertical | Energy & Natrl. Resources |
| Financial Services | 2. Horizontal | Finance |
| Fine Art | 1. Vertical | Media & Telecom |
| Fishery | 1. Vertical | Energy & Natrl. Resources |
| Food & Beverages | 1. Vertical | Travel & Hospitality |
| Food Production | 1. Vertical | Travel & Hospitality |
| Fund-Raising | 1. Vertical | Non-profit |
| Furniture | 1. Vertical | Manufacturing |
| Glass, Ceramics & Concrete | 1. Vertical | AEC |
| Government Administration | 1. Vertical | Government |
| Government Relations | 2. Horizontal | GRC |
| Graphic Design | 1. Vertical | Media & Telecom |
| Health, Wellness and Fitness | 1. Vertical | Travel & Hospitality |
| Higher Education | 1. Vertical | Education |
| Hospital & Health Care | 4. Healthcare | Hospital/Provider |
| Hospitality | 1. Vertical | Travel & Hospitality |
| Human Resources | 2. Horizontal | HR |
| Import and Export | 2. Horizontal | Sales |
| Individual & Family Services | 2. Horizontal | Cust. Support / Svcs |
| Industrial Automation | 1. Vertical | Manufacturing |
| Information Services | 2. Horizontal | IT |
| Information Technology and Services | 2. Horizontal | IT |
| Insurance | 1. Vertical | Insurance |
| International Affairs | 1. Vertical | Government |
| International Trade and Development | 1. Vertical | Government |
| Internet | 2. Horizontal | IT |
| Investment Banking | 1. Vertical | Wealth/Asset Mgmt |
| Investment Management | 1. Vertical | Wealth/Asset Mgmt |
| Judiciary | 1. Vertical | Legal |
| Law Enforcement | 1. Vertical | Government |
| Law Practice | 1. Vertical | Legal |
| Legal Services | 1. Vertical | Legal |
| Legislative Office | 1. Vertical | Government |
| Leisure, Travel & Tourism | 1. Vertical | Travel & Hospitality |
| Logistics and Supply Chain | 2. Horizontal | Supply Chain |
| Luxury Goods & Jewelry | 1. Vertical | Retail |
| Machinery | 1. Vertical | Manufacturing |
| Management Consulting | 2. Horizontal | Cust. Support / Svcs |
| Maritime | 1. Vertical | GIS |
| Marketing and Advertising | 2. Horizontal | Marketing |
| Market Research | 2. Horizontal | Sales |
| Mechanical or Industrial Engineering | 1. Vertical | Manufacturing |
| Media Production | 1. Vertical | Media & Telecom |
| Medical Devices | 4. Healthcare | Hospital/Provider |
| Medical Practice | 4. Healthcare | Hospital/Provider |
| Mental Health Care | 4. Healthcare | Hospital/Provider |
| Military | 1. Vertical | Government |
| Mining & Metals | 1. Vertical | Manufacturing |
| Motion Pictures and Film | 1. Vertical | Media & Telecom |
| Museums and Institutions | 1. Vertical | Travel & Hospitality |
| Music | 1. Vertical | Travel & Hospitality |
| Nanotechnology | 4. Healthcare | Life Sciences |
| Newspapers | 1. Vertical | Media & Telecom |
| Nonprofit Organization Management | 1. Vertical | Non-profit |
| Oil & Energy | 1. Vertical | Energy & Natrl. Resources |
| Online Media | 1. Vertical | Media & Telecom |
| Outsourcing/Offshoring | 2. Horizontal | Cust. Support / Svcs |
| Package/Freight Delivery | 2. Horizontal | Fleet & Field |
| Packaging and Containers | 2. Horizontal | Fleet & Field |
| Paper & Forest Products | 1. Vertical | Energy & Natrl. Resources |
| Performing Arts | 1. Vertical | Media & Telecom |
| Pharmaceuticals | 4. Healthcare | Pharmacy |
| Philanthropy | 1. Vertical | Non-profit |
| Photography | 1. Vertical | Media & Telecom |
| Plastics | 1. Vertical | Manufacturing |
| Political Organization | 1. Vertical | Government |
| Primary/Secondary Education | 1. Vertical | Education |
| Printing | 2. Horizontal | EAM / FM |
| Professional Training & Coaching | 1. Vertical | Education |
| Program Development | 2. Horizontal | Cust. Support / Svcs |
| Public Policy | 1. Vertical | Government |
| Public Relations and Communications | 2. Horizontal | Cust. Support / Svcs |
| Public Safety | 1. Vertical | Government |
| Publishing | 1. Vertical | Media & Telecom |
| Railroad Manufacture | 1. Vertical | Manufacturing |
| Ranching | 1. Vertical | Energy & Natrl. Resources |
| Real Estate | 1. Vertical | Real Estate |
| Recreational Facilities and Services | 2. Horizontal | Cust. Support / Svcs |
| Religious Institutions | 1. Vertical | Non-profit |
| Renewables & Environment | 1. Vertical | Energy & Natrl. Resources |
| Restaurants | 1. Vertical | Travel & Hospitality |
| Retail | 1. Vertical | Retail |
| Security and Investigations | 2. Horizontal | Cust. Support / Svcs |
| Semiconductors | 1. Vertical | Manufacturing |
| Shipbuilding | 1. Vertical | Manufacturing |
| Staffing and Recruiting | 2. Horizontal | HR |
| Supermarkets | 1. Vertical | Retail |
| Telecommunications | 1. Vertical | Media & Telecom |
| Think Tanks | 2. Horizontal | Cust. Support / Svcs |
| Translation and Localization | 2. Horizontal | Cust. Support / Svcs |
| Transportation/Trucking/Railroad | 2. Horizontal | Fleet & Field |
| Utilities | 1. Vertical | Utilities |
| Venture Capital & Private Equity | 2. Horizontal | Finance |
| Veterinary | 4. Healthcare | Hospital/Provider |
| Warehousing | 2. Horizontal | Fleet & Field |
| Wholesale | 2. Horizontal | Sales |
| Wireless | 3. Infrastructure | Network / Storage |
| Writing and Editing | 1. Vertical | Media & Telecom |

**Here are some business objectives / problem statements:**

**How the companies’ growth rates are effected by the factors such as geography, Industry, Sector, Subsector and age of the company. Is there any correlation found between the Capital raised and the employee count? Should the investor target any particular industry? Any inference that can be portrayed while considering the geography?**

1. Top 2 states where QIN rate is highest. (Univariate - Category)
2. At which bucket, the FTE is coming highest and lowest? (Univariate - Numeric)
3. What is the Maximum and Average FTE count though out the data? (Univariate - Numeric)
4. At which range of 1-year growth, maximum no of companies is coming in? (Univariate - Numeric)
5. At which FTE range, 1-year growth is coming highest? (Bivariate - Numeric)
6. On which States, average Capital Raised is coming highest and lowest? (Bivariate - Both)
7. On which Subsectors, average 1-year growth is coming highest and lowest? (Bivariate - Both)
8. How many such companies are there in the dataset, where Sector = Horizontal and Subsector = Finance? (Multivariate - Category)
9. How many unique Sector and Subsector is there in the data? (Multivariate - Category)
10. Top 2 combination of Industry and Subsector, where the no of companies is highest and lowest. (Multivariate - Category)
11. Which industry is giving lowest average 1-year growth when the FTE is greater than 60? (Multivariate – Both)
12. For which Subsector and Industry, the average 1-year growth is lesser than its 6-month growth? (Multivariate – Both)
13. What are attributes having highest and lowest correlation between themselves? Interpret. (Correlation analysis)
14. For which designation of a company, client went for the maximum & minimum no of time? (Univariate - Category)
15. How much percentage of company lies in Canada? (Univariate - Category)
16. In which age group of the companies, the 6-month, 1-year and 2-year growth have grown exponentially and became the highest?

**Relevant variables:**

Under the numerical columns-

1. FTE (No of full-time employees of the company)
2. 6 months growth (FTE growth in % for 6 months)
3. 1-year growth (FTE growth in % for 1 year)
4. 2-year growth (FTE growth in % for 2 year)
5. Capital raised (Company’s last funding/invested amount)

Under the categorical columns-

1. State (US states)
2. City
3. Industry
4. Founded Year (of the company)
5. Title (Designation of higher official to contact)
6. Name of the company
7. Country (US/Canada)
8. Sector (Consists of multiple sectors – 4 labels are there)
9. Subsector (Consists of multiple Subsectors – Multiple labels are there)
10. State code
11. Status (depends on the FTE, - [15<FTE<30] and [30<FTE<125], - 2 labels are there)
12. Age of the company (Derived from Founded Year)

**Other variables:**

1. Qualified In: Whether the company is Qin or QO, - contains 2 labels, - Yes and No.
2. Reason for QO: Why the company is marked as QO.
3. Contact Name: Mainly CEO, if not found then COO/MD/President/Chairman/VP/CFO/CTO.
4. Title: Designation of the contact person.
5. Contact Email: Email of the contact person.
6. No of investors: How many companies has already invested on the company.
7. Investor name(s): Name of the investors.
8. Company LinkedIn link
9. Company website link
10. Description: About section from LinkedIn.
11. Email drafted: If a company is Qin, whether the email is drafted for the CEO of the company.
12. Date of drafting email.
13. Reason for not drafting email.
14. Type: All the companies considered here, are prospect.
15. Capital Raised (Currency): ‘USD’ for all companies.
16. Country: Either US or Canada.
17. US postal code.
18. Global postal code: Postal code from Canada.
19. Status: It is a category consists of 3 levels and it depends on the values of FTE, -
20. 3A. Prospect (30+ FTEs) - When FTE is greater than 30.
21. 3B. Prospect (15 - 30 FTEs) – When FTE is between 15 and 30.
22. 3C. Prospect (< 15 FTEs) - When FTE is lesser than 15.

**Types and explanation of the analysis implemented inside the visualization platform:**

Every analysis starts from EDA or in other words, - Exploratory data analysis. All of the followings are the part of it. The sequence is also maintained accordingly.

1. **Univariate analysis**: This is the simplest form of analyzing data. “Uni” means “one”, so in other words your data has only one variable. It doesn't deal with causes or relationships and its major

purpose is to describe; it takes data, summarizes that data and finds patterns in the data. There are two types of univariate analysis, -

1. Numerical
2. Categorical
3. **Bivariate analysis**: Bivariate analysis is the simultaneous analysis of two variables (attributes). It explores the concept of relationship between two variables, whether there exists an association and the strength of this association, or whether there are differences between two variables and the significance of these differences. There are three types of bivariate analysis, -

A. Numerical & Numerical

B. Categorical & Categorical

C. Numerical & Categorical

1. **Multivariate analysis:** This is similar to bivariate but considers more than 2 variables. There are three types of bivariate analysis, -

A. Numerical & Numerical

B. Categorical & Categorical

C. Numerical & Categorical

1. **Correlation analysis:** Correlation is the analysis that measures the strength of association between two variables and the direction of the relationship. In terms of the strength of relationship, the value of the correlation coefficient varies between +1 and -1. A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. The direction of the relationship is indicated by the sign of the coefficient; a + sign indicates a positive relationship and a – sign indicates a negative relationship.

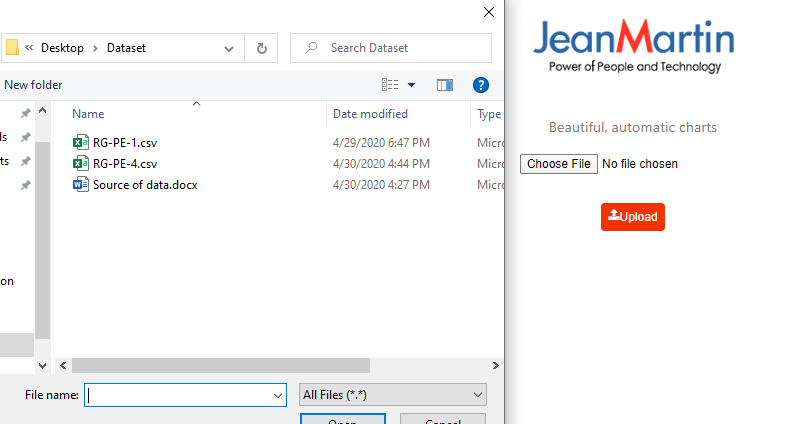
**Technical requirements:**

1. **Backend:**
2. Python (Version 3.7.4)
3. Django (Version 3.0.3)
4. **Frontend:**
5. Node (Version 10.19.0)
6. Yarn (Version 1.21.1)
7. IDE - Visual Studio(Version 1.42.1)
8. **Remote server:**
9. Ubuntu Cloud server (Version 18.04)

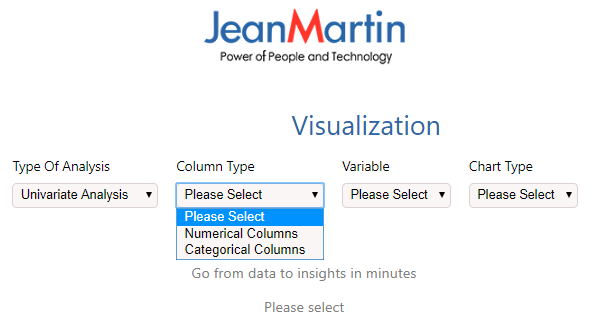
**API used:** Total 7 APIs have been used here in the backend, -

1. File upload: This service will upload a csv file from the local system to the remote Ubuntu server.
2. Column type: This response will provide the numerical and categorical columns separately.
3. Univariate: For single variable analysis.
4. Bivariate: For 2 variable analysis.
5. Multivariate: For more than 2 variable analysis.
6. Correlation: To find the correlation between the numeric variables.
7. Descriptive statistics: This will give some statistical metrics such as Max, Min, Mean, Std etc.

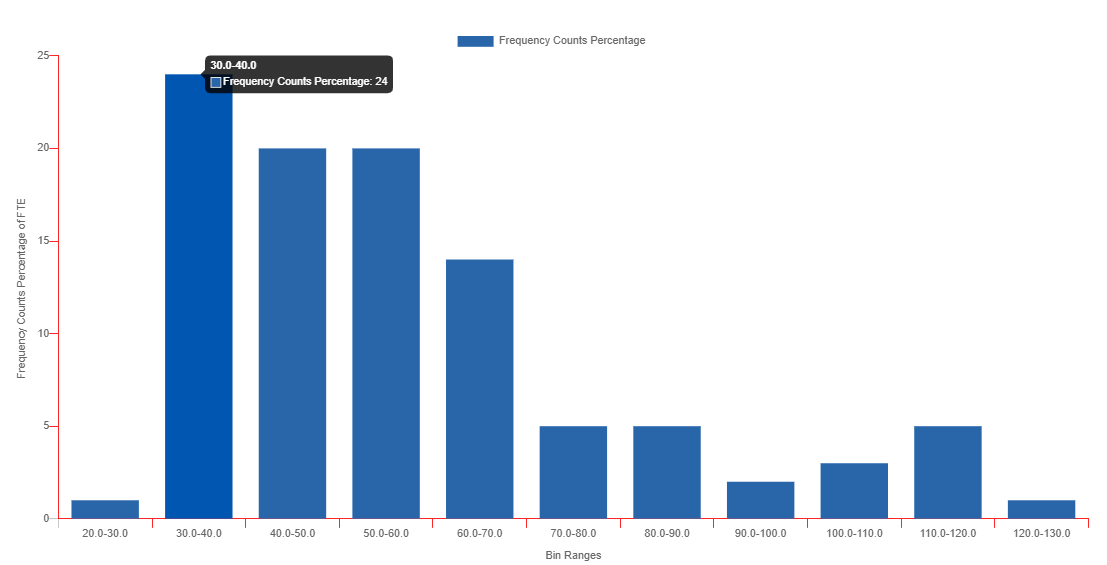
**File upload:**

****

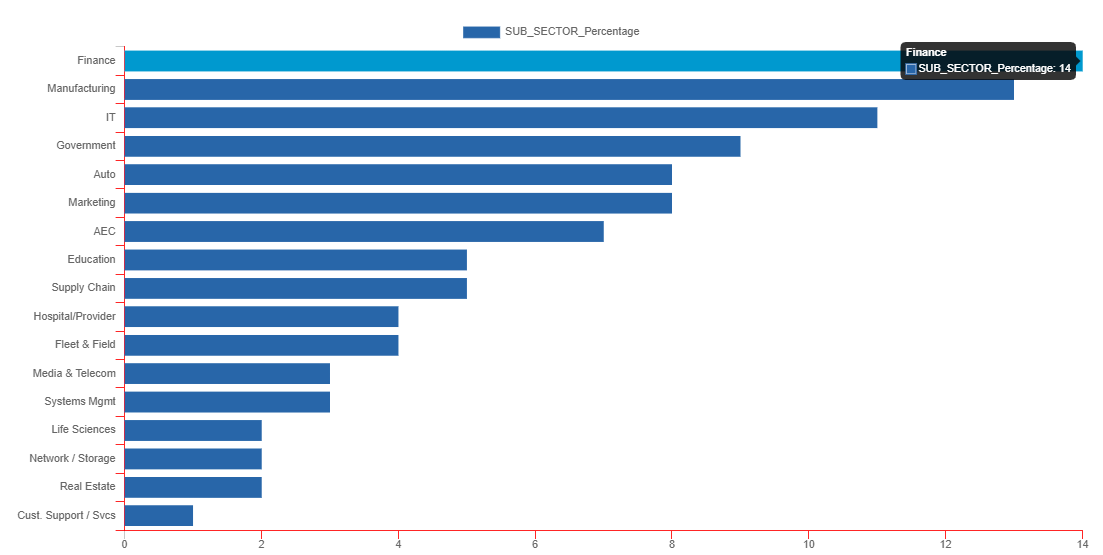
**Column type:**

****

**Univariate analysis:**

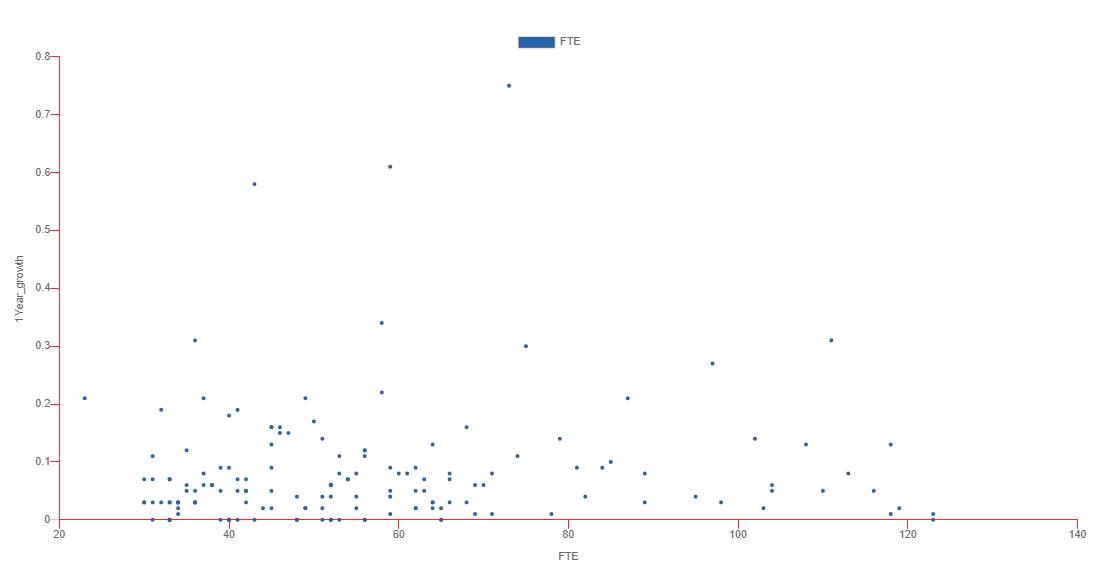
****

**Interpretation**: 24% of the companies which are qualified in, are within the FTE ranges between 30 to 40.

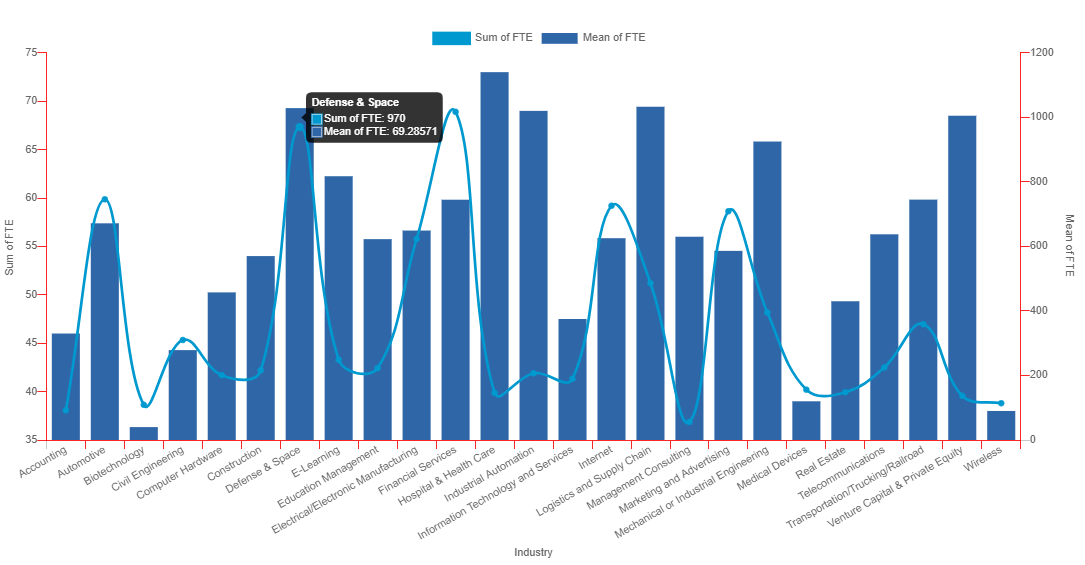
****

**Interpretation:** 14% of the companies are coming from the Finance sector. Finance, Manufacturing and IT can be the target Industry.

**Bivariate Analysis:**

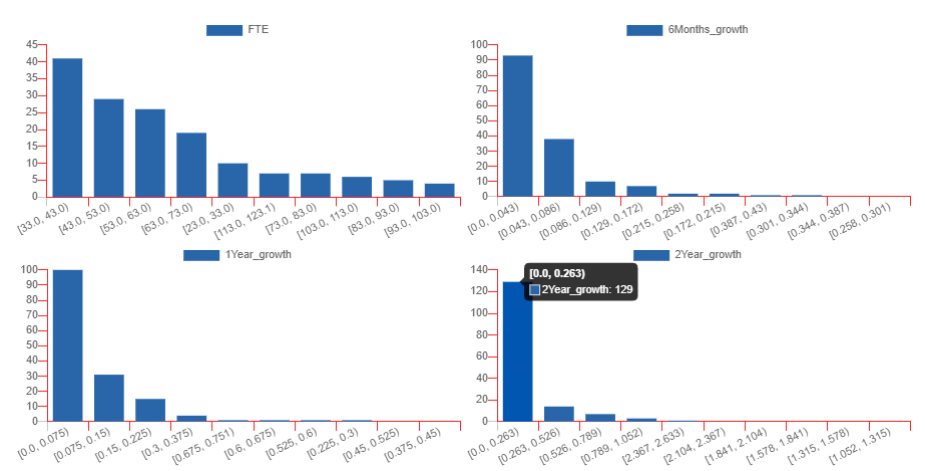
****

**Interpretation:** If we plot the 1-year growth percentage with the FTE, we will see that most of the growth percentages are coming within the FTE range of 30 to 70 which can be considered as the potential factor of a company to invest.

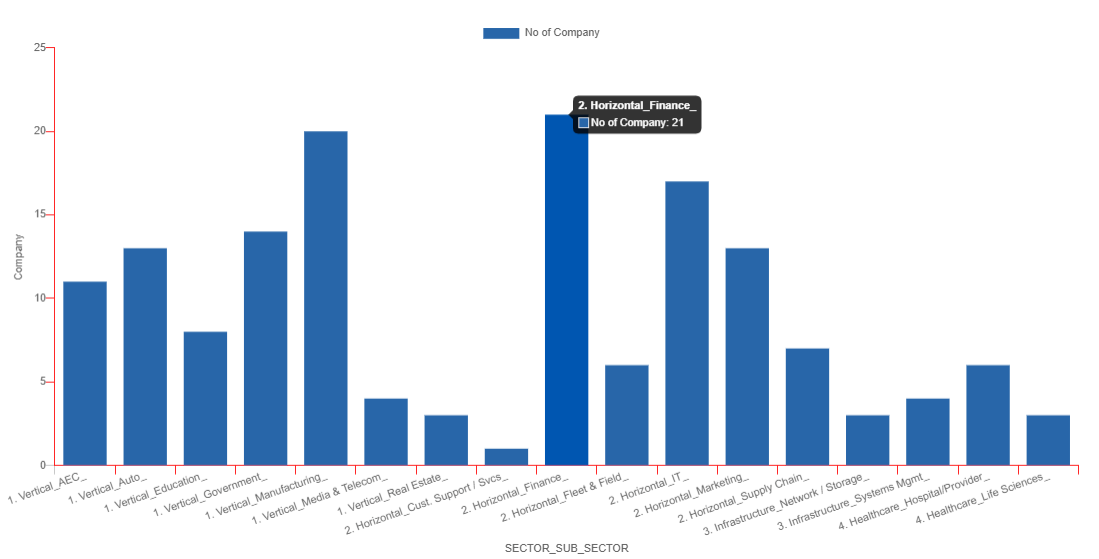
****

**Interpretation:** The sum and average no of FTE, both are coming higher in the Defense industry in comparison with other which implies Defense industry hired lot of people in comparison with the other. Potential growth may be shown there.

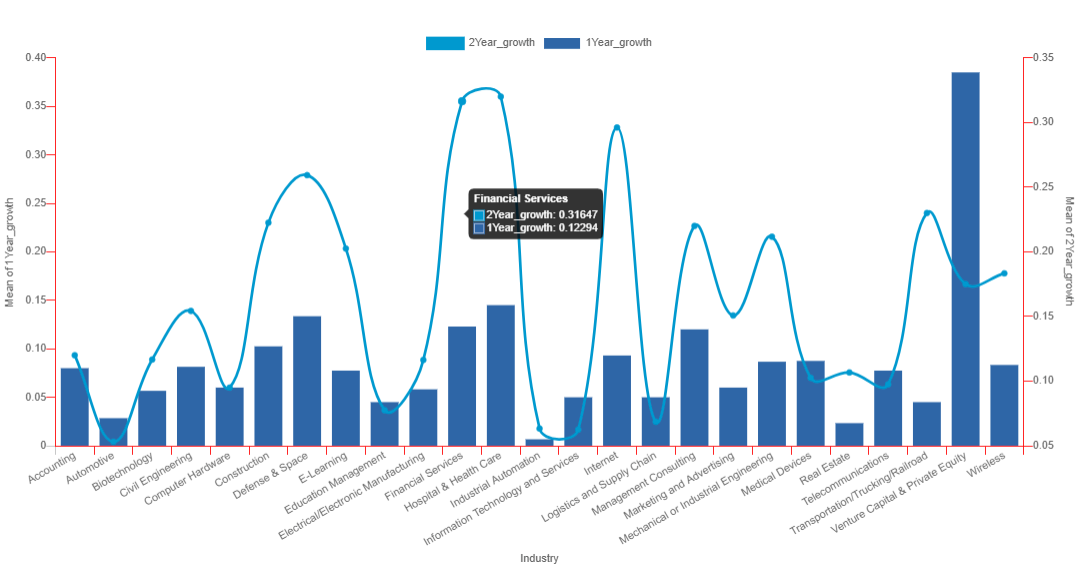
**Multivariate analysis:**

****

**Interpretation:** This is a histogram.

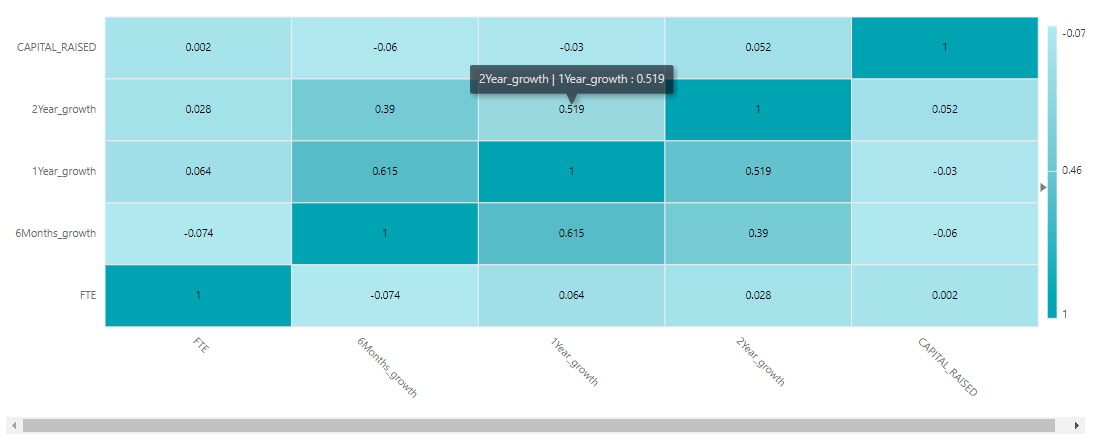
****

**Interpretation:**  Number of companies when Sector is Horizontal and Sub sector is Finance, - is 21 which is the highest in the lot. As it is already discussed Finance sector is clearly showing that it consists of maximum number of qualified in companies.

****

**Interpretation:** This is the comparison between 1-year and 2-year growth. We can see a huge growth in Financial Services.

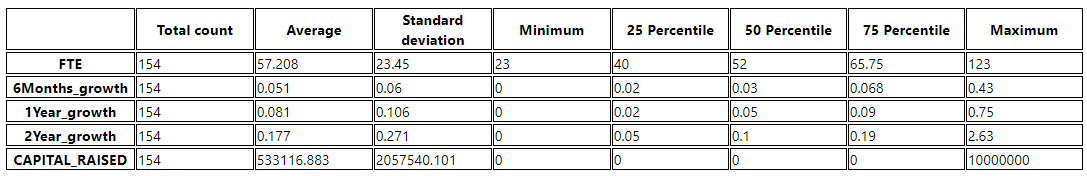
**Correlation analysis:**

****

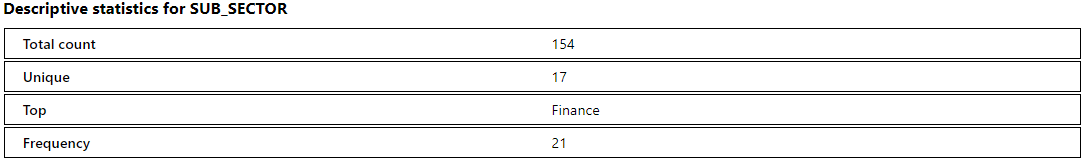
**Interpretation:** We can see a very high correlation between 6-months and 1-year growth which is obvious, and similarly 1-year and 2-year growth. 6-months and 1-year growth is negatively correlated with capital raised although with minimal percentage. Hence, it must not be inferred that decrease of 1 variable might not increase another one and vice versa.

**Descriptive statistics:**

**For numeric variables:**

****

**For categorical variables:**

****

**Developed by:**

1. *Sourav Biswas (Sr Data Analyst & Backend developer, Jean Martin Systems India Pvt Limited)*
2. *Anand Ramakrishnan (Frontend & UI developer, Jean Martin Systems India Pvt Limited)*
3. *Arivalagan Selvaraj (Frontend & UI developer, Jean Martin Systems India Pvt Limited)*