

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

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.

**Relevant variables:**

Under the numerical columns-

1. FTE (full time employee)
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 (Last funding 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 Subsector)
9. Subsector
10. State code
11. Status (depends on the FTE, 2 labels are there)

**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:**

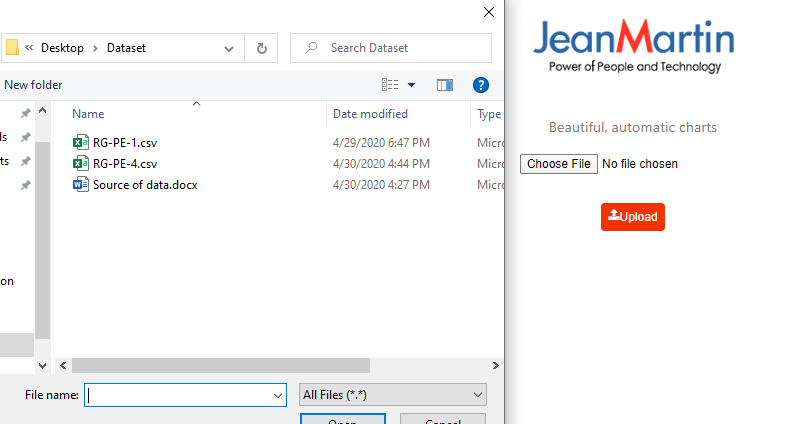
1. **Backend:**
2. Python (Version 3.7.4)
3. Django (Version 3.0.3)
4. **Frontend:**
5. Node (Version 10.19.0)
6. Yarn (Version 1.21.1)
7. IDE - Visual Studio(Version 1.42.1)
8. **Remote server:**
9. Ubuntu Cloud server (Version 18.04)



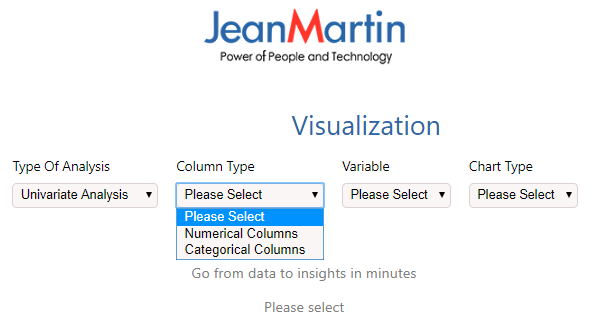
**API used:** Total 7 APIs have been used here in the backend, -

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:**

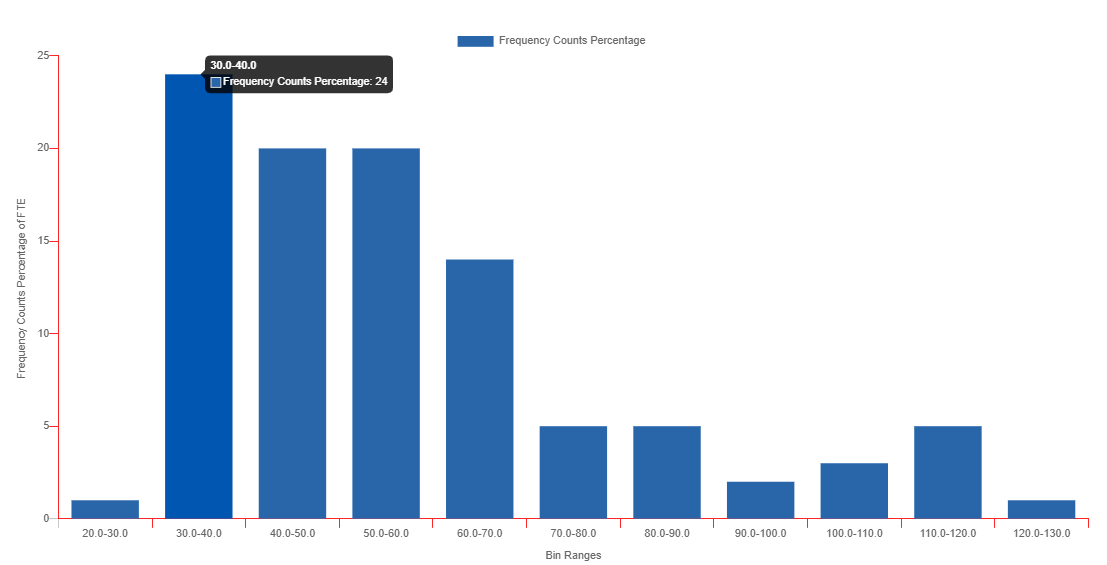
****

**Column type:**

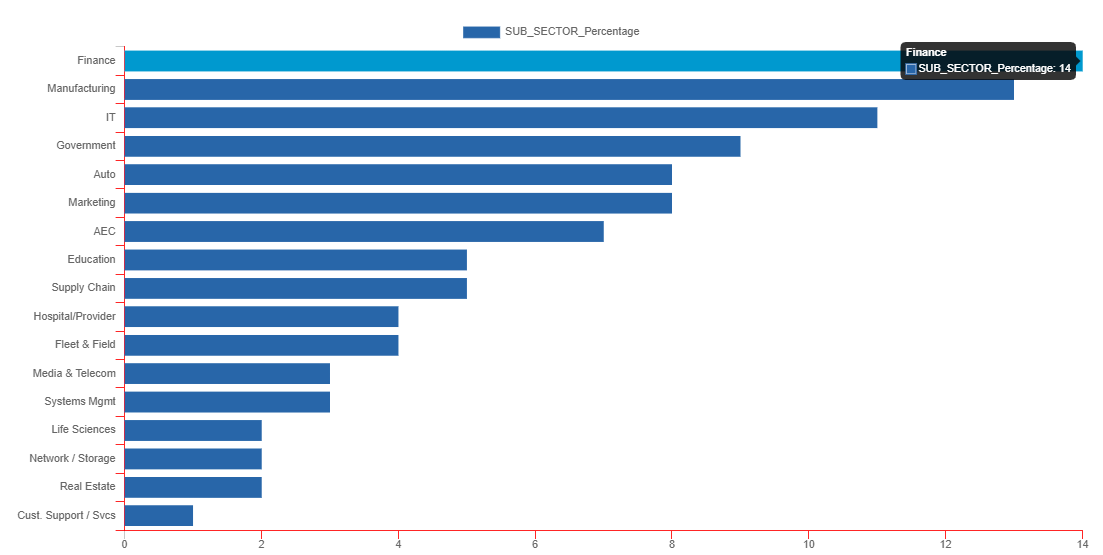
****



**Univariate analysis:**

****

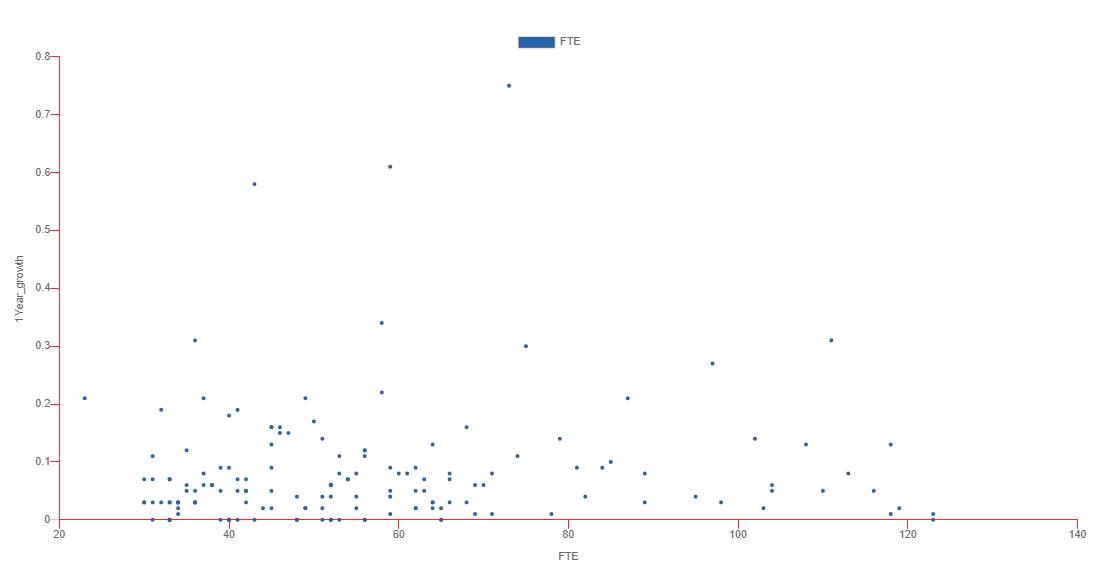
**Interpretation**: 24% of the companies which are qualified in, are within the FTE ranges between 30 to 40.

****

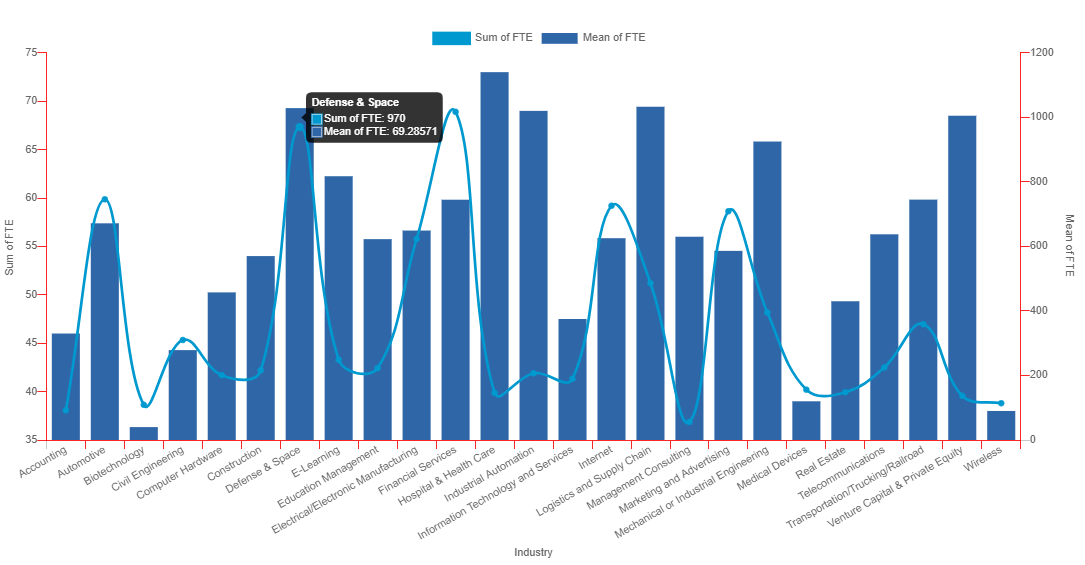
**Interpretation:** 14% of the companies are coming from the Finance sector.



**Bivariate Analysis:**

****

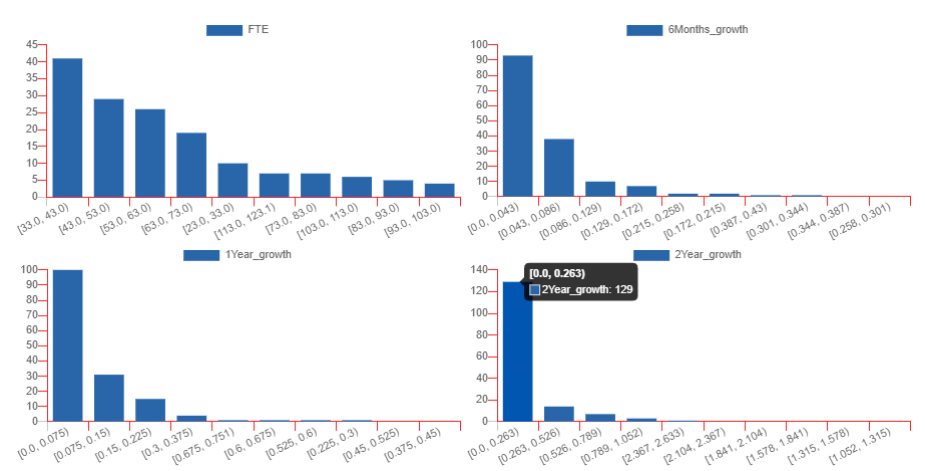
**Interpretation:** If we plot the 1-year growth percentage with the FTE, we will see that most of the growth percentages are coming within the FTE range of 30 to 70.

****

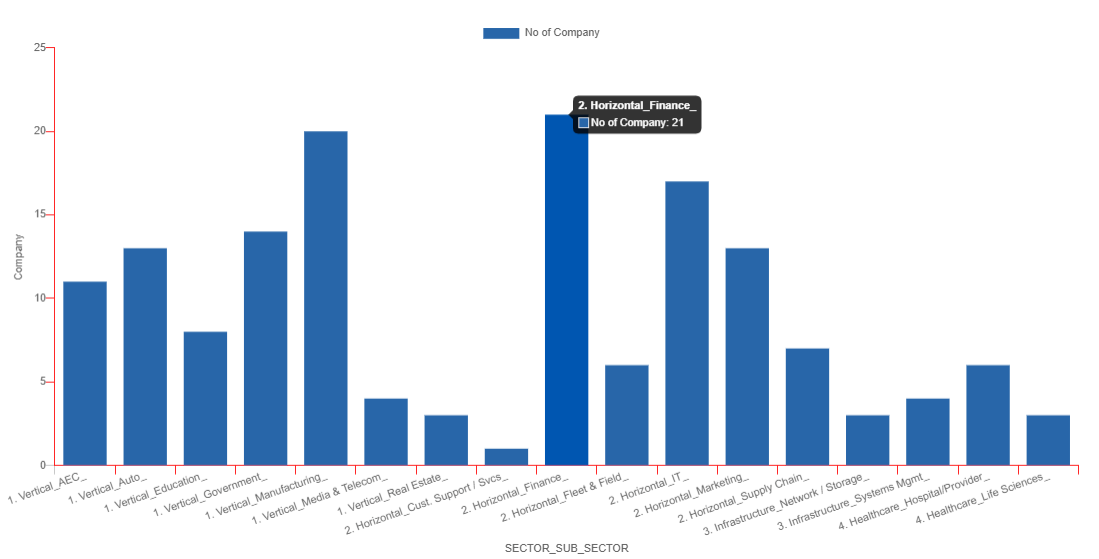
**Interpretation:** The sum and average no of FTE, both are coming higher in the Defense industry in comparison with other.



**Multivariate analysis:**

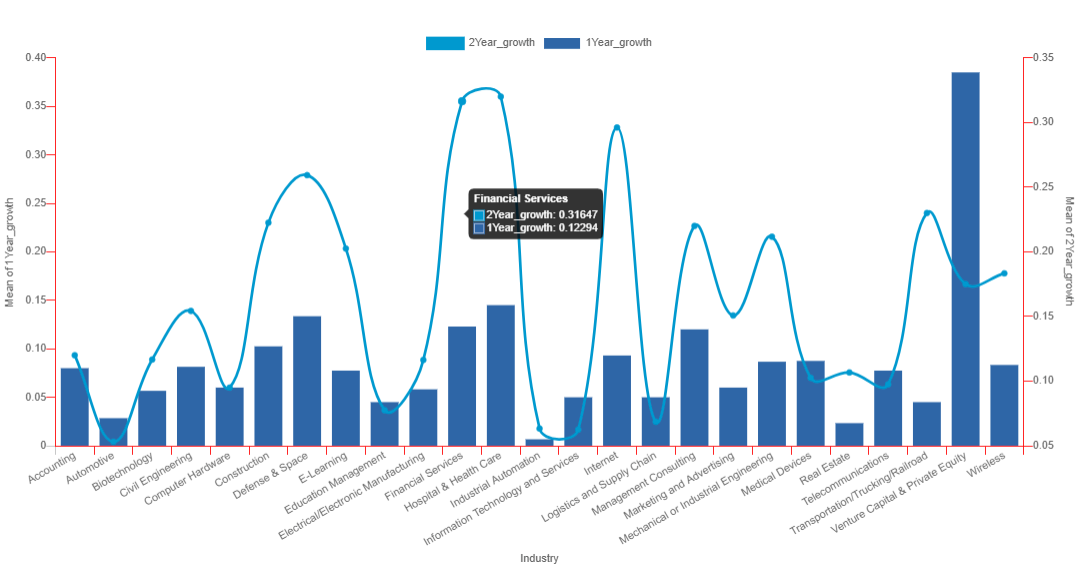
****

**Interpretation:** This is a histogram.

****

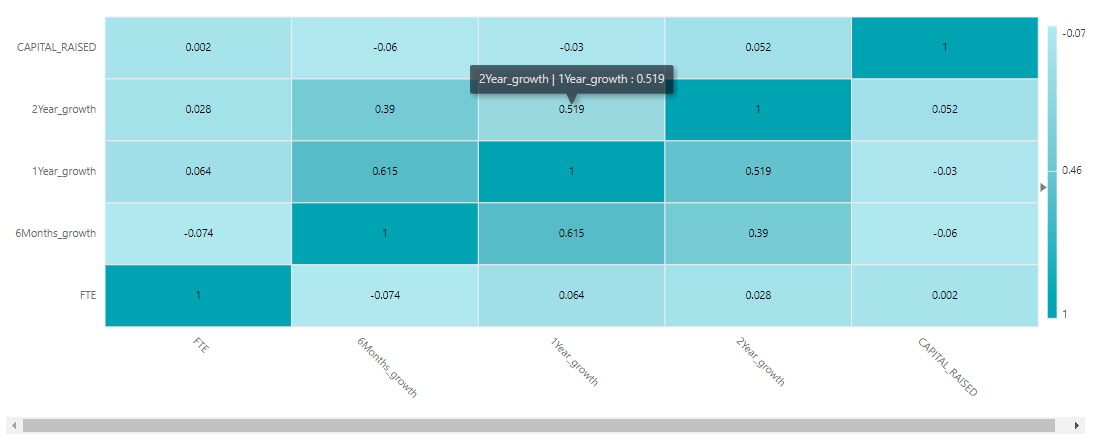
**Interpretation:**  Number of companies when Sector is Horizontal and Sub sector is Finance is 21 which is the highest in the lot.



****

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

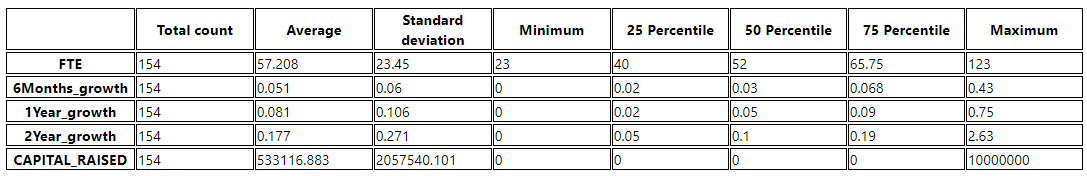
**Correlation analysis:**

****

**Interpretation:** We can see a very high correlation between 6-months and 1-year growth, and similarly 1-year and 2-year growth. 6-months and 1-year growth is negatively correlated with capital raised.



**Descriptive statistics:**

****

**Some Problem statements:**

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 combinations 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 it’s 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. At which range of age of the company, the 6-month, 1-year and 2-year growth have grown exponentially and became the highest? (Multivariate both)