

Volatility of Industrial Growth towards various Economic Factors – A technical Analysis

Analytical CRM
Atif Feroz Jeelani
M.Sc. Data Analytics
x17169992
National College of Ireland

Abstract – Analysing the behaviour of potential growth of an Industry plays a vital role for the stakeholders including entrepreneurs, policymakers and researchers to make a suitable marketing strategy in terms of leveraging more profit. Analytical CRM helps in finding out the hidden pattern and helps in revealing the existing relationship between different industry variables. Creating an opportunity for investors and entrepreneurs to make the right business decisions. This literature provides a technical approach and focuses to find out the dynamic relationship between the growth of an industry to other factors affecting the same like employs, revenue, and expense which will give us a better understanding of the industry growth and thus, improving decision making. In this literature, we have implemented multiple linear regression to an Industry dataset including 500 different companies and presented results of the implementation

Keywords – Industry sector , Multiple linear regression , customer relationship management , Visual analytics , growth .

I. INTRODUCTION

Nearly every business plan desires growth in the future. The expansion of one's business is something every person whether an investor or good entrepreneurs aspire to achieve. Especially the early success of profit-making often fuels the dream of growth. When a level of success is achieved as a company owner, decisions of expanding the business into new industry sectors starts developing. As one company moves ahead of the start-up phase, identifying new opportunities to grow the business becomes a priority for long term success. So, it becomes very important for investors/entrepreneurs to identify the underlying factors and trends responsible for the growth of an industry. Key indicators such as profits, sales,

staff members, industry sectors can be looked upon to measure the business growth. [1]

A better understanding of Industry sectors and what services they are providing, their strategic planning and to observe their trends with Growth of the business helps in business intelligence of the companies for which analytical customer relationship management is used. The analysis and study of data about a company to analyse the likelihood of a business growth pattern or a customer sale patterns and understanding such trends and relations is known as Analytical Customer Relationship Management (CRM). To discover the underlying facts of a company data and observe the relations and trends between factors like revenue, profit, expense, and growth using several statistical approaches like regression techniques are extensively used in the business and investment domain. [2]

In this literature, we will be using a data of 500 different companies of U.S and analyse their hidden patterns and trying to reveal the existing relationships. The focus is to find a result to what affects the growth of a company and which industry sector outperforms in terms of profit and growth. This can benefit the investors, entrepreneurs and company owners in making the right business decisions. In this literature, we will try to portray a deep knowledge about the visual data analysis. We would be using data representations and regression technique to get a clear insight.

The report will be organized as followings. The literature is organized as the following: Section i) Introduction, section ii) Literature review of related work iii) Methodology section iv) Dataset description section, Business question, and Hypothesis v) Implementation of report, section vi) Results of the implemented work vii) Conclusion of the report viii) Discussion of future work research

II. Literature Work

A lot of research has been done in exploring and analysing the relationship of growth of a company with various other variables. Bruce Kirchhoff presented an empirical approach in identifying the growth factor. The research was conducted on German, U.S, Australian and Scottish economies. The result shows that age, location, size and legal form are related to business growth. The research uses Swedish data and shows a comparison of domestic vs international business and results depict that the industry sector plays a most important factor relating to growth. [3]

In another literature, Eric A. Nerlinger proposes a research explaining the variation of growth between New technology-based firms (NTB's) and non-innovative firm foundation. In his research, he uses a multivariate regression model to explain the growth of the firm. The regression results depict a strong relationship between growth rate and type of firm, founder-specific also external factors. These factors influence growth rates for both the form types. A further growth rate of NTBFs achieves a higher side as compared to non-innovative young firm. [5]

Stephen Hymer in his research has collected data of the 1000 largest manufacturing firms of the U.S from 1946 to 1955 on the growth rates. He has compared and computed the distribution of growth rates for different size classes of firms. Using statistical techniques he finds that the entire distribution of the growth rates does not vary for firms of different sizes. The law of proportionate effect depicts that growth rates are the same for large and small firms. In fact, growth tendency is found larger for smaller firms comparatively. [4]

III. METHODOLOGY

Methodology plays an important role in creating a pipeline for the researcher to proceed through the implementation of the research. In this research my implementation will be following a CRISP – Data mining model. The model provides a structured approach to plan a project. The relationship that exists between the task of the project is usually present in the life cycle of CRISP. Based on the result of the project the relation between the task is decided which can be dependent on the data and user.

[6] Different stages of CRISP-DM are as above.

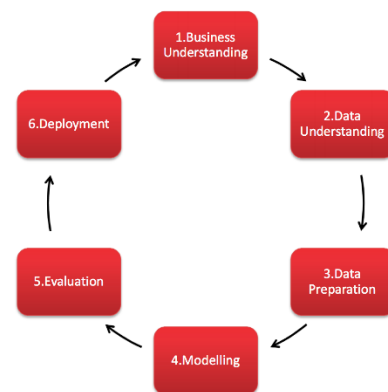


Fig : CRISP -DM [6]

In order to study and analyse the data and find the underlying existing relationship relating to the growth of companies and leverage more profit that will help different people making a good business decision. The methods that were followed are as below :

1. Visual Analytics :

The reason behind visual analytics is to get a clear picture of the Industry sectors that are making more profit in terms of revenue and expenses. Also to study which industry sectors have progressed best in terms of growth. Sometimes statistical methods make it difficult for a business user to understand and relate with a business problem and end up making the wrong decision. Visual analytics helps in perceiving new aspects of data. Visualization can help the user achieve new knowledge by incorporating interactivity. Hence, a visualization makes the study of data simple and interactive for the business user. Data is represented in the form of graphs, maps, bars, etc.

There are various tools that can be used for drawing a powerful and intuitive visual analytics. For this project, I am using R and Tableau software for the study and exploration of data visually. I will be considering boxplot and scatter plot using R also using simple bar chart in tableau. The data in R and tableau both needs data pre-processing like removing of Null vales and imputations need to be done. After that the connection between excel spread sheet and tableau worksheet is made. The data is pulled column by column to the sheet of

representation where graph is constructed with colours and labels making it simple for user understandability.

2. Regression Model :

To analyse the relation and correlation among the dependent and independent variable is a process of statistics which can be described under regression analysis. The technique of regression is basically used to explore whether there exists a relationship between the independent and dependent attributes. The model of regression is to be used generally on three categories of variables that is nominal, categorical and ratio/interval variables. Different models of regression are used based on the variable types to be considered like Multiple linear regression, linear regression, logistic regression, etc.

In this paper we are going to use Multiple linear regression to study and find out if there exists any relation between Industry sector, expenses, employ number, revenue and profit (taken as an independent variable) to the growth of the company (dependent variable). Using this regression model we will be able to check how much variability is shown by the independent variables on the dependent variable and which attributes shows maximum variability which affects growth. The results obtained can be used to attain an improved business decision making.

IV. DATA SET DESCRIPTION

The dataset used to find out the factors affecting growth and analyse the relationship is taken of 500 different U.S companies. It consists of various variables that be used to obtain knowledge. The dataset consists of 11 attributes having 500 rows. The data includes categorical variable as well as continuous variables. Some of the attributes are :

- Industry sector
- Employ Number
- Expense
- Revenue
- Profit
- Growth
- Company name
- Year

- City
- State

The dataset is a publicly available and can be used by anyone for research purposes. The data was downloaded from :

<https://www.superdatascience.com/pages/rcourse-advanced>.

Data Pre-processing

The downloaded data had to be pre-processed before making it to obtain results using Tableau, R and IBM SPSS . For these software to make statistical calculations and data visualizations data needed to be cleaned in R as it had many missing values which had to be properly understood and imputed. Imputation measures like replacing null with median values was done. Changing of string type from factor to numeric was done for statistic calculation. Growth variable was chosen as dependent variable.

Business Question :

- How much variability does the dependent variable (Growth) show on Independent variables (industry sector, employ, profit, revenue and expense)?
- Which factor is highly correlated in depicting the factor of Growth of a company?
- Which company sector should be preferred by an investor or entrepreneur for extension of business

Hypothesis :

H0 : There exists no correlation between Growth of company and Industry Sector.

H1 : There exists correlation between Growth and Industry Sector.

V. IMPLEMENTATION

The implementation part of this project constitutes the study, representation and analysis of the taken dataset. It compares various Industry attributes like revenue, expense, profit, industry sector with growth of the industry which will prove beneficial in business decision making. There are three steps of implementation part . First two gives us the

information of data using visualization and final part is using regression model giving us more statistical significant results.

1. Visual Data Analytics :

In this step, we first uploaded the data inside tableau. We pulled down only the desired columns. The columns included are Growth and the city/state .

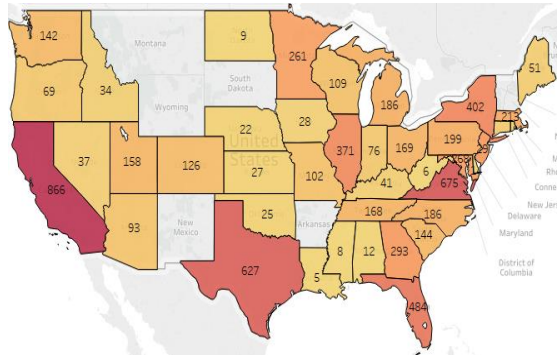


Fig . 2 Map showing growth rate in U.S

The above map drawn in tableau depicts the growth parameter of different companies in the country of U.S. In the above analytics it was noticed that companies established in California have drawn highest growth followed by Virginia and Texas. The red colour depicts the highest growth rate and yellow shows the lowest. Looking at this visualization it becomes quite clear for a business person to know the states in the U.S where business has flourished with highest growth rate making it easy to decide where the expansion of the business can be established considering the growth of a company.

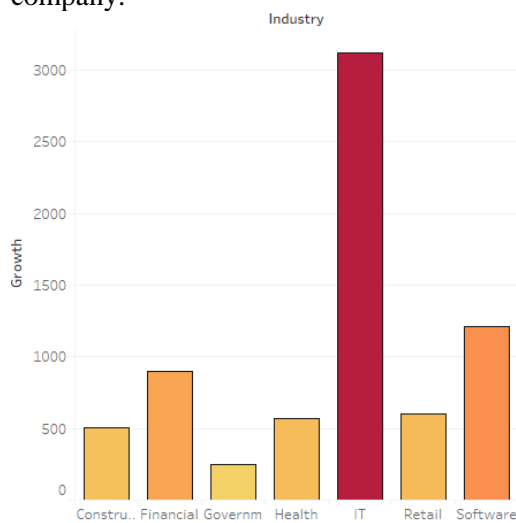


Fig. (3) Bar chart showing Growth by Industry

Similarly, another visualisation was also drawn using tableau. The parameters used are the different industry sectors and their growth. These columns were pulled down in a simple bar graph. The graph illustrates clearly that among the different industry sectors I.T industry top's the league with respect to growth followed by Software companies. I.T industry sector gives a clear swipe to others making it easy for the investors and entrepreneurs to decide in decision making.

In our second phase of implementation we have used R for visual analytics . The .csv file was imported in R and cleaned. Steps for proper imputations were followed. In R we conducted a similar test using ggplot2 to analyse growth by industry sector using a boxplot . We used jitters to give us details about outliers and removed them. The boxplot gives us another representation of Industry sector and growth relation. Here we can see I.T industry is proved to be best in terms of growth with small differences between software and financial companies.

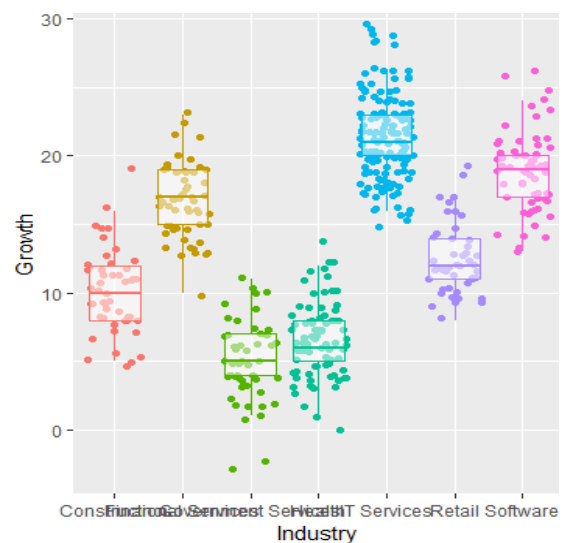


Fig . (4) Box-Plot showing Growth by Industry

In another visual analytics, we used ggplot2 to represent various industry trends in terms of expense and revenue relationship . We are using ggplot and using geom_point to represent x axis as revenue and y as expenses and colour for industry. This generated visualization gives us trends for revenue expense relationship. Here we can notice that Financial companies keep on generating revenue at constant pace with less

expense while as I.T companies can be expensive in the beginning and generates revenue slowly. Government sectors depicts increasing expense and low revenue.

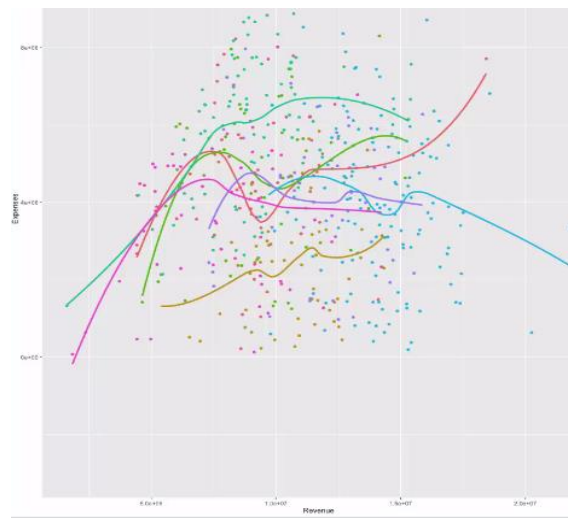


Fig. (5) Trend Analysis of Revenue and Expenses

2 Multiple linear regression :

For further analysis of the data, regression model was used to analyse a better and statistically significant study. For this purpose, the technique of Multiple linear regression was used in IBM SPSS tool. Using the multiple linear regression technique we tried to analyse and study if there existed relationship between growth (dependent variable) and revenue, expense, profit, number of employs and industry sector(independent variable) and if yes which factor contributes as the strongest unique contributor in explaining the variance of growth.

The design of the model includes various phases. First of all the industry variable consisted of 7 levels. In order to use it in the regression model 6 dummy variables had to be created and values 0 and 1 had to be allotted .With 0 being no industry 1 being respective industry. Removal of outliers had to take place and check for preliminary tests were taken. After these measures 2 test were carried out one with and one without industry sector as independent variable. The following results were generated in the SPSS statistics viewer :

Test without industry variable.

Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change
1	.499 ^a	.249	.244	5.982	.249	54.596

a. Predictors: (Constant), Employees, Expenses, Revenue

Fig. (6) Model Summary

In the model summary we check the R square value which tells us 24.9 percent variance in dependent variable is explained by the model. Which is average value

Correlations						
		Growth	Profit	Expenses	Revenue	Employees
Pearson Correlation	Growth	1.000	.495	-.249	.435	-.065
	Profit	.495	1.000	-.565	.838	-.031
	Expenses	-.249	-.565	1.000	-.023	.014
	Revenue	.435	.838	-.023	1.000	-.029
	Employees	-.065	-.031	.014	-.029	1.000
Sig. (1-tailed)	Growth	.	.000	.000	.000	.072
	Profit	.000	.	.000	.000	.245
	Expenses	.000	.000	.	.308	.382
	Revenue	.000	.000	.308	.	.262
	Employees	.072	.245	.382	.262	.

Fig. (7) Correlations

From the above correlation matrix we can check the correlation each variable is showing on the other. Here Profit and Revenue illustrates positive correlation while as expense and employees are negatively correlated with respect to growth.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
1	(Constant)	7.833	1.112	7.043	.000	5.648	10.018
	Expenses	.000	.000	-.239	.000	.000	.000
	Revenue	.000	.000	.428	.000	.000	.000
	Employees	-.001	.001	-.050	.201	-.002	.000

a. Dependent Variable: Growth

Fig. (8) Coefficients

The coefficients generated explains how uniquely independent variable has explained variance in growth. In the standard coefficient beta, revenue with value .428 is the strongest in explaining the variance in growth. Followed by revenue and employees making less significant contribution. Thus with unit change in revenue there will be increase in growth. Also with sig value less than 0.5 makes it statistically significant.

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5861.389	3	1953.796	54.596	.000 ^b
	Residual	17678.364	494	35.786		
	Total	23539.753	497			

a. Dependent Variable: Growth

b. Predictors: (Constant), Employees, Expenses, Revenue

Fig. (9) ANOVA

ANOVA tables describes the fitness of the model.

F-ratio at (3,494) = 54.596 . The model reaches statistical significance as Sig. = .000 ; this means $p < 0.005$

Test with industry variable.

This multiple regression test describes the variance shown by independent variable when Industry sector is included

Model Summary						
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change
1	.846 ^a	.716	.714	3.680	.716	311.323

a. Predictors: (Constant), Industry3, Employees, Expenses, Revenue

Fig. (10) Model Summary

Model summary explains R square of 71.6 percent .Which is an excellent result.

According to this model independent variable show strong variance in growth of company.

Coefficients ^a							
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	20.547	.817	25.159	.000	18.942	22.151
	Expenses	.000	.000	-.035	1.418	.157	.000
	Revenue	.000	.000	.160	6.194	.000	.000
	Employees	.000	.000	-.010	-.420	.675	-.001
	Industry3	-2.503	.088	-.765	-28.504	.000	-2.675

a. Dependent Variable: Growth

Fig. (11) Coefficients

In the coefficients box, it is clearly seen the type of industry sector highly affects the growth of a company as standard coefficients obtained is .765 which is the strongest in explaining growth. Followed by revenue . Also industry variable with .sig .000 explains its statistically significant .

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16863.606	4	4215.901	311.323	.000 ^b
	Residual	6676.147	493	13.542		
	Total	23539.753	497			

a. Dependent Variable: Growth

b. Predictors: (Constant), Industry3, Employees, Expenses, Revenue

a. Dependent Variable: Growth

b. Predictors: (Constant), Industry3, Employees, Expenses, Revenue

Fig.(12) ANOVA

At the and the anova tell us about the goodness of fit about the model. F-with ratio at (3,493) = 311.323 . The model is statistical significant as Sig. = .000 ; this means $p < 0.005$

VII. RESULTS AND CONCLUSIONS

From the results we derive in the implementation part of the literature, we can easily conclude that our alternative hypothesis (H1) was proved to be statically significant i.e there exists a relationship between the growth of a company and different industry sectors and other various parameters like revenue. From fig. (10) and (11) we conclude there exists a strong relationship between these parameters and in fact industry sector was proven to be uniquely and strongly explaining variability in growth at .765 beta value. Also from the visual analytics, we get to know the trends analysis of Revenue and expense that explains how the performance of the different industry. I.T industry proves to be expensive at the beginning and generates revenue slowly whereas Financial companies could be seen performing at a constant phase. The boxplot visualization determines that I.T companies in terms of growth has overpowered the rest and determines the highest yielding growth followed by Software and Financial companies. Also, from the map generated it was depicted that California, Virginia, and Texas are best suited for establishment of a company considering the rate of growth is high in these states in the U.S

The final conclusion that be inferred from the results of this literature is that once after visual analytics and then statistically analysing the data using regression technique we can study the data for the purpose of business growth.

VIII. FUTURE WORK

In this literature Multiple linear regression was used to analyse the underlying relationships. In future work we can use other machine learning models and statistical techniques to give us more detail.

The results from other techniques like logistic regression or decision tree can be used and compared to find out which technique is better.

REFERENCE

[1]"The Advantages of Expanding Business", *Smallbusiness.chron.com*, 2019. [Online]. Available: <https://smallbusiness.chron.com/advantages-expanding-business-21144.html>. [Accessed: 23- Mar-2019].

[2]Wanghualin, "Data Mining and Its Applications in CRM", *2010 Second International Conference on Computer Research and Development*, 2010. Available: 10.1109/iccrd.2010.184 [Accessed 23 March 2019].

[3]P. Davidsson, B. Kirchhoff, A. Hatemi-J and H. Gustavsson, "Empirical Analysis of Business Growth Factors Using Swedish Data", *Journal of Small Business Management*, vol. 40, no. 4, pp. 332-349, 2002. Available: 10.1111/1540-627x.00061 [Accessed 23 March 2019].

[4]S. Hymer and P. Pashigian, "Firm Size and Rate of Growth", *Journal of Political Economy*, vol. 70, no. 6, pp. 556-569, 1962. Available: 10.1086/258716 [Accessed 23 March 2019].

[5]M. Almus and E. Nerlinger, *Small Business Economics*, vol. 13, no. 2, pp. 141-154, 1999. Available: 10.1023/a:1008138709724 [Accessed 23 March 2019].

[6]"What is the CRISP-DM methodology?", *Smart Vision - Europe*, 2019. [Online]. Available: <https://www.sv-europe.com/crisp-dm-methodology/>. [Accessed: 23- Mar- 2019].

APPENDIX

Worklog

TIME	WORK DONE
WEEK 1	Research and finalizing dataset
WEEK 2	Literature Review
WEEK 3	Feedback and review
WEEK 4	Selecting data model and data pre-processing
WEEK 5	Methodology
WEEK 6	Implementation
WEEK 7	Evaluation of results and Reporting