Startup Profit Prediction

Term Paper

Rainer Naftal

Andrew Nzioki

Josphine Alphonce

Paul Mwasame

Ken Okoth

**Background**

Nine out of every ten startups fail for several reasons, such as making products rejected by the consumers, implying pore market research. Additionally, 42% of failed startups identify lack of market as the reason behind. Other factors that play a part in the failure include entrepreneurs failing to follow the due business process, unattainable growth, inability to recover. Additionally, entrepreneurs overestimate the value of the intellectual property before product-market fit by approximately 255%. Further finance problems also are among critical factors that contribute to startup failures. More than 50% of the interviewed founders didn't have a budget for their project, and 75% were self-funded, yet only 16% point at financial problems as the reason for failure, and to validate concept finance is needed startups (Patel, 2015).

Notably, tech problems are also among the major factors plaguing startup failures. As such, interviewed startups have some technology in their core. However, the biggest mistake is over-investing inexpensive technology (developer time) before the marketing assumptions have been validated(Da Silva, 2018).

‍Additionally, operational problems are understandably rare. This might not be the case for startups that work with physical products. Some heavily-regulated food and finance still present legal obstacles making it difficult for startups. The above reasons contributing to the startup's failure form a basis for researching and analyzing factors or variables contributing to startup success. Hence its profitability.

**Variable definition/ declaration/validation**

The variables used in this dataset areR&D spending, Administration Spending, Marketing Spending, and location features, while the target variable is: Profit. R&D spending, Administration Spending, Marketing Spending, and location features are the independent variables, while profit is the dependent variable.

**Research & Development spending**

When a company is considering launching a product, a certain effort and amount are spent on research to understand the market better. Therefore, the amount is spent on research is a critical success factor. R&D is important for businesses because it provides powerful knowledge and insights that improve existing processes where efficiency can be increased and costs reduced. It also allows businesses to develop new products and services to survive and thrive in competitive markets (Fuertes-Cullen, 2020). Further, R&D spending helps the company to have; Increased market participation, cost management benefits, advancements in marketing abilities, and trend-matching.

**Administration spending:**

This is the amount that startups are spending on the Admin panel. Forecasting administration spending would enable the business people to budget on their amount, avoiding financial problems.

**Competition**

The type of industries the company ventures into might have other established and monopolized players. Due to the size of the players and the size of capital needed to compete fairly in that industry, some startups may fail.

**Marketing spending**

This is the amount that startups spend on marketing strategies that might determine how the market receives their product. Analyzing the market and forecasting marketing spending as a factor contributing to profitability is important because marketing provides a clear sense of direction for marketing efforts and gives direction on the effective allocation of scarce financial resources(Yohn,2020). Further, forecasting the marketing spending can aid in planning and coordination with other essential departments such as HR. Moreover, knowing how much you have in the budget to spend better helps determine which marketing avenues to go after to maximize your marketing ROI.

**Location of the startup**

Based on the business location, the startup could experience success cause ideally good location should be convenient for your customers, employees, and suppliers - without being too expensive for your business. It would be best to weigh up the advantages and disadvantages of various locations when deciding on a suitable place for your business property.

All the above variables determine the profitability of a startup, and profit equals a company's revenues minus expenses. Earning a profit is important to a small business because profitability impacts whether a company can secure financing from a bank, attract investors to fund its operations, and grow its business. Companies cannot remain in business without turning a profit.

**Model Specification**

We will use multiple linear regression in modeling the variables. Multiple linear regression refers to a statistical method that uses at least two independent variables to predict the outcome of a dependent variable.

The model for the startup prediction of a business will be as follows

*Where y is the dependent variable*

are the model parameters

are independent variables

is the error term.

The variables used for the startup prediction were as follows

* R&D spend – x1
* Administrative Spending – x2
* Marketing spending – x3
* Profits – y

Therefore, our model becomes:

We chose multiple linear regression because it will enable us to determine the model's variation and the relative contribution of each independent variable in the total variance. This enables researchers to analyze the strength of the relationship between a dependent variable and the predictor variables and the significance of every predictor to the relationship. Additionally, with this model, a researcher can identify the outliers and anomalies easily

The purpose of this analysis is to establish how different quantities/ factors influence the profits of a startup

**DATA COLLECTION|EXTRACTION|ESTIMAION, AND ANALYSIS**

**Descriptive Statistics.**

**Statistics**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **N\*** | **Mean** | | **SE Mean** | | **StDev** | | **Variance** | **CoefVar** | **Minimum** | **Q1** |
| R&D Spend | 50 | 0 | 73722 | | 6492 | | 45902 | | 2107017150 | 62.26 | 0 | 36107 |
| Administration | 50 | 0 | 121345 | | 3962 | | 28018 | | 784997271 | 23.09 | 51283 | 102580 |
| Marketing Spend | 50 | 0 | 211025 | | 17294 | | 122290 | | 14954920097 | 57.95 | 0 | 125325 |
| Profit | 50 | 0 | 112013 | | 5700 | | 40306 | | 1624588173 | 35.98 | 14681 | 87769 |
| **Variable** | **Median** | | | **Q3** | | **Maximum** | |
| R&D Spend | 73051 | | | 105066 | | 165349 | |
| Administration | 122700 | | | 145191 | | 182646 | |
| Marketing Spend | 212716 | | | 300633 | | 471784 | |
| Profit | 107978 | | | 142254 | | 192262 | |

**Coefficients**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Term** | **Coef** | **SE Coef** | **T-Value** | **P-Value** | **VIF** |
| Constant | 50122 | 6572 | 7.63 | 0.000 |  |
| R&D Spend | 0.8057 | 0.0451 | 17.85 | 0.000 | 2.47 |
| Administration | -0.0268 | 0.0510 | -0.53 | 0.602 | 1.18 |
| Marketing Spend | 0.0272 | 0.0165 | 1.66 | 0.105 | 2.33 |

**Regression Equation**

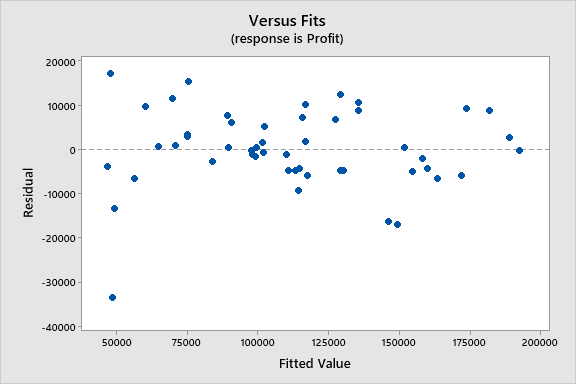
|  |  |  |
| --- | --- | --- |
| Profit | = | 50122 + 0.8057 R&D Spend - 0.0268 Administration + 0.0272 Marketing Spend |

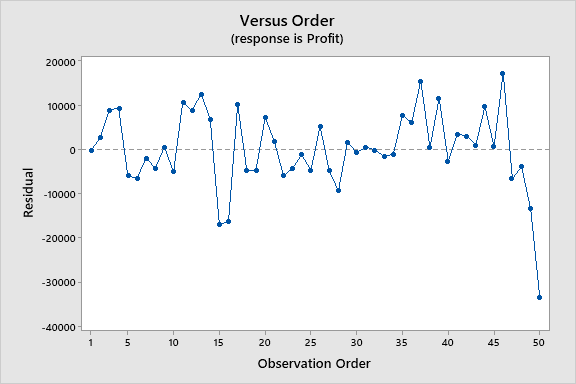
**Analysis of Variance**

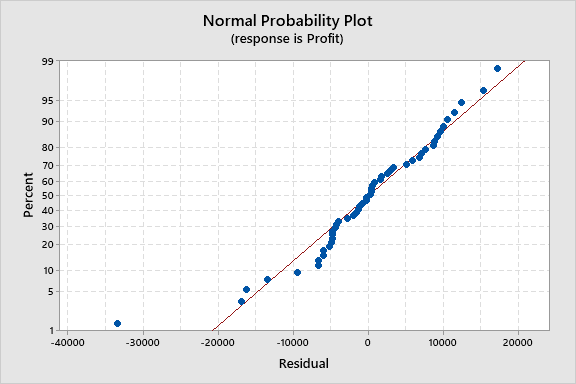
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| Regression | 3 | 75683964196 | 25227988065 | 295.98 | 0.000 |
| R&D Spend | 1 | 27147076244 | 27147076244 | 318.49 | 0.000 |
| Administration | 1 | 23538549 | 23538549 | 0.28 | 0.602 |
| Marketing Spend | 1 | 233485362 | 233485362 | 2.74 | 0.105 |
| Error | 46 | 3920856301 | 85236007 |  |  |
| Total | 49 | 79604820497 |  |  |  |

**Model Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **S** | **R-sq** | **R-sq(adj)** | **R-sq(pred)** |
| 9232.33 | 95.07% | 94.75% | 94.15% |







Result Interpretation.

From the coefficient analysis, the R&D spend, and the marketing spends have a p-value of 0.00 and 0.105, respectively, less than 0.05, indicating a significant association between the response variable (profit) and the terms.

The R2 value of 95.07% indicates that the model provides a good fit for the data.

The residuals versus fits plot verify the assumption that the residuals are randomly distributed and have constant variance.

The residuals versus order plot verifies the assumption that the residuals are independent of one another. The residuals fall randomly around the centerline.

In the normal probability plot, the points have a slight deviation from the straight line, and there is evidence of outliers.

Conclusion.

The coefficient for the predictor variables does not equal zero. There are factors from the above analysis; there is sufficient evidence that the model.

**Reference**

Fuertes-Callén, Y., Cuellar-Fernández, B., & Serrano-Cinca, C. (2020). Predicting startup survival using first year’s financial statements. *Journal of Small Business Management*, 1-37.

Da Silva, C. M. P. (2018). Startups: A study to test survivorship and bankruptcy, forecasting models.

Yohn, T. L. (2020). Research on the use of financial statement information for forecasting profitability. *Accounting & Finance*, *60*(3), 3163-3181.

Patel, N. (2015, September 2). *90% of startups fail: Here's what you need to know about the 10%*. Retrieved March 21, 2022, from https://www.forbes.com/sites/neilpatel/2015/01/16/90-of-startups-will-fail-heres-what-you-need-to-know-about-the-10/?sh=5c0078f16679