The Influence Factors of Supermarket Profit: A Case Study of The United States

Lejing Xue[1, *]

[1]Eller College of Management, University of Arizona, AZ, 85705, USA

*Corresponding author: lejingx@arizona.edu

Abstract. A Supermarket's profit is the money it earns when its total revenue exceeds its whole expenses. Any profit made by a corporation goes to its owners, who can choose to distribute it to shareholders as income or reinvest it in the business to fund future growth. The goal of this study is to analyze how the profitability of a particular superstore in the United States is affected by several economic factors. The multiple linear regression method, which is a statistical tool for prediction, will be utilized extensively throughout the paper. This project will go through data wrangling techniques, where this paper checked for inconsistency, missing values, and duplicate records, and filtered noisy data by identifying outliers and removing them, sampling the data due to its size to a more convenient size for ease of analysis, to proceeding to visualization, in order to determine what the results of the analysis are. Our primary method of multiple linear regression yielded an R-squared value of 0.563 for our model. This value represented the coefficient that indicated how well the values fit compared to the originally used values. Each one of the analyses that was carried out will be based on a confidence interval of 95%. In conclusion of the study, this paper found out that the more the sales and higher the quality the more chances of getting profit on the sales made with discount accounting to a negative variation to the profit sales. This model can be used to demonstrate future trends in profit margins and be applicable apply to other supermarkets to avoid losses.

Keywords: Profit; sales; quantity; discount.

1. Introduction

This chapter will compose the background information of our outcome variable Profit, the benefits of profit to any organization, firm, or company, the problem that led to the proposed project the research questions, the aims of the project, and finally the hypothesis for the project.

A financial gain or the amount of money that remains after an organization has deducted all of its expenses is the definition of profit [1]. It is necessary for your supermarket to have more earnings than losses, just like it is essential for any other businessperson. The most important factors that influence a supermarket's capacity to turn a profit are competition from other supermarkets, the demand for dominance, the state of the economy, advertising, substitutes, relative costs, economies of scale, dynamically efficient, price discrimination, management, objectives of firms and exchange rate. Here will discuss all 13 key factors respectively [2].

In terms of competition from other supermarkets, when a supermarket has monopolistic power, it has very little competition, which lowers the level of demand inelasticity. Demand inelasticity occurs when a buyer's desires for a product do not fluctuate as much as the product's change in price [3]. Because of this, the supermarket is in a position to boost its earnings by raising the price. If there is a lot of competition in the market, it will be difficult to turn a profit. This is due to the fact that customers would only buy from businesses that offered the lowest prices. The concept of contestability is also an essential component. The term "market contestability" refers to the ease with which new businesses can break into an existing market. If it is simple to enter a market, then businesses will constantly have to deal with the possibility of competition, even if it is "hit and run" competition, which will result in lower profitability [4].

The strength of demand is the second factor to consider. For example, there would be a high level of demand for a product if it is fashionable. For instance, mobile phone firms were profitable during the period of rising demand and expansion in the market. Products like Spam (canned meat) that are seeing a decline in demand will lead to a reduction in the company's earnings. The state of the economy is another factor taken into consideration. Demand for most products will rise in a growing

economy, particularly for luxury goods with high-income elasticity of demand. This is due to the high-income elasticity of demand for luxury goods. For instance, luxury sports car producers will prosper when the economy is expanding, but they will struggle when the economy is contracting.

Taking into Account Advertising Demand for a product might become more inelastic as a result of a good advertising campaign that increases demand for that product. However, in order to cover the costs of advertising, the increased revenue will need to be sufficient. Word of mouth might be one of the most effective tactics at times. For instance, it was unnecessary for YouTube to undertake a lot of advertising because of its popularity [5]. If there are a lot of alternatives available or if they are expensive, then the demand for the product will be higher. Unlike the case of substitutes, the more substitutes there are, the less demand there should be because there are more choices. In a similar vein, complementary goods will be essential to the success of a company's bottom line.

Taking into account the relative expenses [6]. Profits will fall in direct proportion to an increase in costs, such as labor, raw materials, or rent, in direct proportion to the size of the increase in those costs. Alternatively, if the company is successful in increasing productivity by improving technological capabilities, profits should rise. This is due to the fact that productivity and income are inversely related. If a company relies on importing raw materials to operate, the current exchange rate is critical. a rise in the price of imported goods as the value of the currency falls A decline in the value of the currency, on the other hand, benefits companies that deal with exports because it increases their ability to compete.

Concerning issues associated with economies of scale [7]. If a company's fixed costs are high, it will need to produce in large quantities in order to produce at the smallest efficient scale and reap the benefits of economies of scale; otherwise, its average costs will be too high. The company will need to produce a large quantity in order to produce at the most efficient scale.

The acronym stands for the phrase "dynamically efficient." If a company is not dynamically efficient, its expenses will continue to rise inexorably over time [8]. State monopolies, for example, were frequently unwilling to cut costs by eliminating unnecessary labor or other overhead expenses. As a result, prior to privatization, they made very little profit; however, as a result of the market's workings and the incentives it provided, they became more effective. This was due to the fact that they made very little profit prior to privatization.

Because of differences in price [9]. If the company is able to differentiate its prices, then it will operate more efficiently. This strategy involves charging multiple prices for the same commodity in order for the company to be able to charge greater prices to customers whose demand is less elastic. This is significant for companies that operate airlines.

For Executives. Companies must have effective management in place in order to sustain long-term growth and profitability. Ineffective management, for example, can lead to a drop in employee morale, which can have a negative impact on both customer service and employee turnover. Furthermore, poor expansion strategies put businesses at risk of failure. Several financial institutions, for example, made the mistake of investing in high-risk subprime mortgages, resulting in significant losses. Tesco ran into financial problems as a result of its diversification into unrelated businesses such as garden centers [6]. As a result, the company spread itself too thin and lost sight of what made them successful in the first place.

Considering the supermarket's objectives. Supermarkets aren't always looking to make the most money [10]. Some businesses may wish to increase their market share, in which case they will have to forego profits in order to do so. As an example, Walmart and, to a lesser extent, Amazon take this approach.

Finally, consider the currency exchange rate. If a company's profitability is determined by exports [4] a decline in the value of the currency exchange rate will result in increased profitability. A lower currency exchange rate increases exporters' competitiveness in international markets. As a result, the company has the option of increasing sales or increasing its profit margin. If the company purchases raw materials from outside the country, depreciation will raise its production costs. A depreciation of the exchange rate, for example, would increase the cost of imports; as a result, businesses that

imported raw materials would face cost increases [11]. In our case, will going to be considering the superstore profits as essential in running the store in matters like maintaining the supplies, the best source of capital to expand the business, and alluring investors to invest in the business seeking more returns.

The analysis will be carried out using Jupyter notebook programming software where the results will be presented in table 1 format and graphs and all findings will be well explained and interpreted. Thereafter will draw conclusions that will be based on the following research questions and aim

RQ1: Is there a relationship between the outcome variable profit with the experiment variables sales, quantity, and discount and RQ2: Is there a linear regression model representing the outcome variable with all the predictor variables?

The aim is to build a multiple linear regression with profit as the outcome variable and sales, quantity, and discount as the experiment variables

1.1 Policy recommendation

The Policy Regarding Supermarket Competition One of the key elements of the new framework is the need to address a general shortage of larger full-line supermarket capacity. Other important aspects of the new framework include: encouraging the development of additional full-service independent supermarket chains in the ACT, encouragement of appropriate independent full-line chain stores to open in newly constructed and redeveloped Group Centers Support for a different wholesale grocery supply source, which would be encouraged by a site-specific approach, recognizing that market and competition signals are requiring larger format supermarkets in comparison to the overall size of shopping centers, including in some local shopping centers. The removal of artificial restrictions placed on grocery stores in the most appropriate local areas.

The growth of centers in ways that do not jeopardize the public's quality of life and enables those stores to provide a more competitive offer against full product lines stores in more populated areas New entrants should be allowed into both newly developed and established areas.

2. Data and Methodology

This chapter will go through what the data is all about from the data attributes to data wrangling techniques, checking for missing variables, inconsistency, duplicate data, and outliers in each well-explained procedure will be conducted and explained also correlation analysis will be done to determine the relationship, principal analysis to determine the amount of variation on the data and visualization for normality test to scatter plots.

2.1 Data

The initial step was importing the data into python the super store dataset downloaded from the Kaggle dataset using this link below: https://www.kaggle.com/datasets/ibrahimelsayed182/sample-super-store is a record of 9994 data entries with the following attributes Ship mode, Segment - Segment Category, Country, City, State, Postal code, Region, Category -Categories of product, Sub-Category, Sales -Number of sales, quantity, discount, and Profit

Due to the nature of the data having 9994 data entries, it was too big to enable effective analysis as the majority of the time a memory error was displayed in each analysis. These purposed to use a sample of the data that’s 50% for the analysis which was randomly selected with a record of 4997 data entries as it didn’t generate any error and was more ideal. Afterward, this paper proceeded with data wrangling which is the process of cleaning, organizing, and transforming raw data into a meaningful desired format for analysis and decision-making.

Table 1. Table of summary statistics

	Postal Code	Sales	Quantity	Discount	Profit
count	4997	4997	4997	4997	4997
mean	55044.26	230.07	3.79	0.15	27.08
std	31971.27	606.30	2.19	0.207	255.85
min	1040.00	0.5560	1.00	0.00	-6599.978
25%	23223.00	17.90	2.00	0.00	1.814
50%	55901.00	56.56	3.00	0.10	8.922
75%	90004.00	212.94	5.00	0.20	30.09
max	99301.00	17499.95	14.00	0.80	8399.97

From the above results, this paper can see a summary statistic of the numeric variables taking into account the count of the variables, the mean, standard deviation, maximum and minimum values, and the quartiles

The sum of the count of missing values was taken into consideration to avoid working with missing data. The record showed that no missing data and inconsistency was evident hence this paper proceeded to check and remove duplicates and outliers. To identify whether or not the data had been duplicated the data. The duplicate command was used where the duplicates were identified and later on removed and proceeded to check for outliers in the numeric variables except for the postal code variable as it was the address of each state.

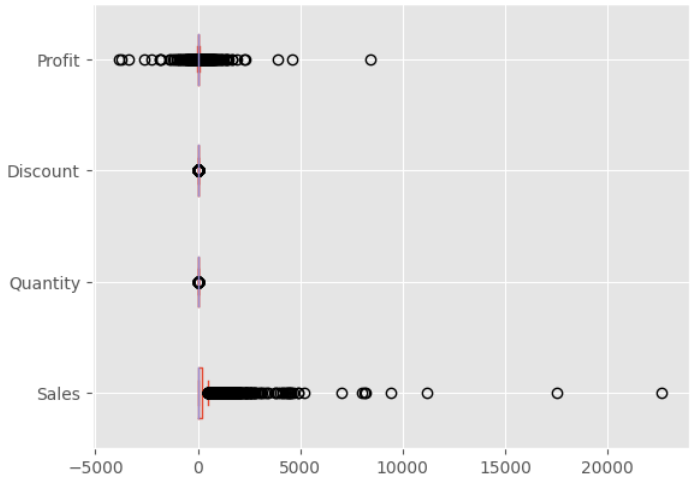

Fig. 1 Outliers plot

In figure 1, Outliers are considered as data values that differ considerably from the bulk of any given data set and were present in order not to affect the results of the analysis, they were purposed to be removed by converting them into missing values that were easier to drop or remove.

2.2 METHODOLOGY

In this chapter will go through the essential statistical approach for the project that's building a multiple linear regression from linear regression to estimation of the model parameters and finally defining the model.

In our case will use a multiple linear regression model which is of the form.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \quad (1)$$

where y denotes sales profit, x1 denotes the sales, x2 denotes the quantity and x4 denotes the discount its a multiple linear regression model with 3 regressors with β_o being the intercept and ϵ

is the random error component for the multiple linear models that assumes to have a mean of zero and unknown variance.

$$\epsilon \sim N(0, \sigma^2) \tag{2}$$

The estimation of $\beta_0, \dots, \beta_4$ is as follows Let:

$$\mathrm{y_i} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_4 x_{i4} + \epsilon_i, i = 1, \dots, n \tag{3}$$

$$= \beta_0 + \sum_{j=1}^{k} \beta_j \, x_{ij} + \epsilon_i, i = 1,2, \dots, n \dots\dots\dots\dots\dots \tag{4}$$

The least-squares function is:

$$\mathrm{S}(\beta_0, \beta_1, \dots, \beta_4) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(yi - \beta_o - \sum_{j=1}^{k} \beta_i \, x_{ij}\right)^2 \dots\dots \tag{5}$$

$$\frac{\partial S}{\partial \beta_0} |_{\beta_o \beta_1, \dots, \beta_4} = -2 \sum_{i=1}^{n} \left(y_i - \bar{\beta_0} - \sum_{j=1}^{k} \bar{\beta_1} \, x_{ij}\right) = 0 \dots\dots \tag{6}$$

And

$$\frac{\partial S}{\partial \beta_1} |_{\beta_o \beta_1, \dots, \widehat{\beta_4}} = -2 \sum_{i=1}^{n} \left(y_i - \widehat{\beta_0} - \sum_{j=1}^{k} \widehat{\beta_1} \, x_{i,j}\right) x_{i,j} = 0, j = 1,2, \dots, k \dots \tag{7}$$

Hypothesis:

Here will formulate a null hypothesis that there is no relationship between the outcome variable and the experiment variables Vs the alternative hypothesis that there is a relationship between the four variables

Our multiple linear model being

$$\text{profit} = \beta_0 + \beta_1 \text{Sales} + \beta_2 \text{quantity} + \beta_3 \text{discount} + \epsilon \tag{8}$$

the null hypothesis states

$$\beta_1 = 0, \beta_2 = 0, \beta_3 = 0 \dots\dots \tag{9}$$

vs the alternative hypothesis

$$\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0 \tag{10}$$

The second hypothesis would be, there is no multivariable linear regression defining the variables vs the alternative hypothesis there is a multivariable linear regression model

Our multivariable linear regression model being

$$\text{profit} = \beta_0 + \beta_1 \text{Sales} + \beta_2 \text{quantity} + \beta_3 \text{discount} + \epsilon \tag{11}$$

the null hypothesis states

$$\beta_1 \text{to} \, \beta_3 = 0 \tag{12}$$

vs the alternative hypothesis

$$\beta_1 \neq 0 \text{ to } \beta_3 \neq 0 \tag{13}$$

3. Results

3.1 Data Analysis Results

A scatter plot of profits against sales was plotted to show the relationship

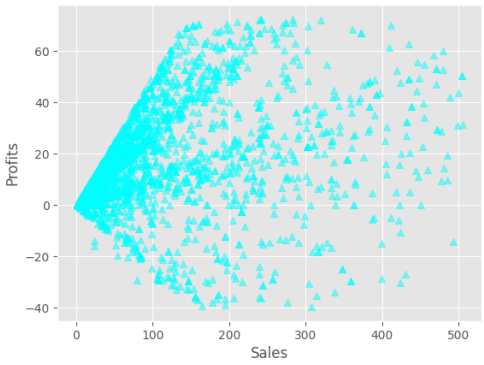

Fig. 2 Scatter plot of Profits

To check for normality in the dataset both the histogram and density plot were plotted as essential when estimating the regression coefficients and also during the prediction. The plots were as follows:

In each plot, this paper expect to see a bell-shaped figure 2 to prove that normality is evident in each variable

This paper begin with the outcome variable which is quite evident that normality is present as the shape of the data is bell-shaped.

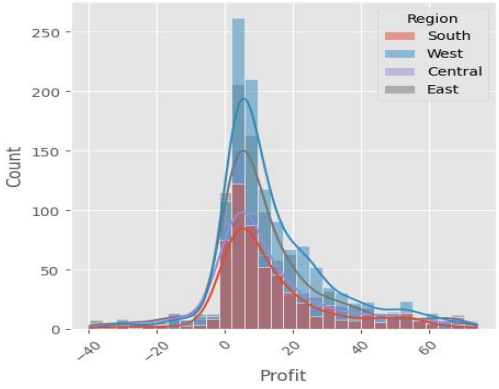

Fig. 3 Histogram of profits

Proceeding to the other variables this paper see that normality isn't present (figure 3).

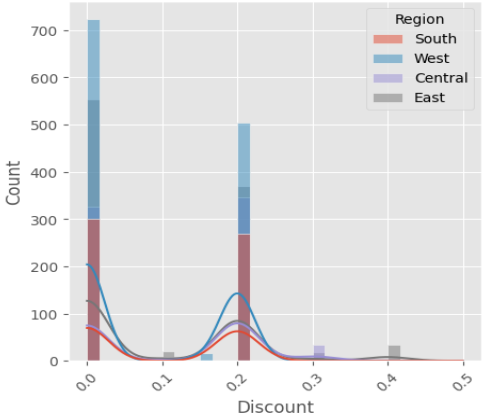

Fig. 4 Histogram of Discount

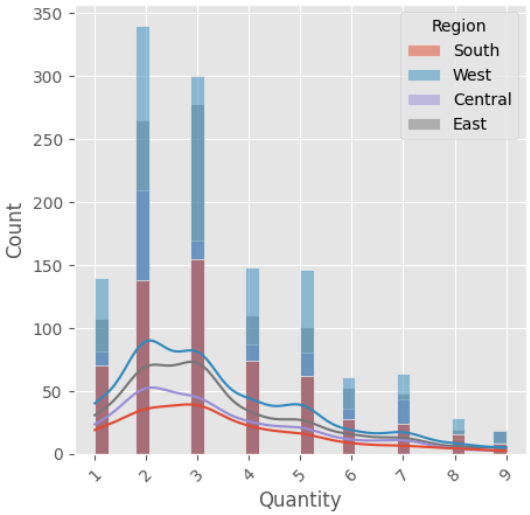

Fig. 5 Histogram of Quantity

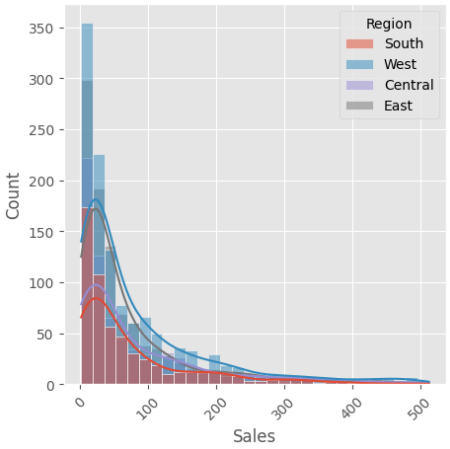

Fig. 6 Histogram of Sales

The figure 4-6 for the experiment variables lack show an absence of normality

The Karl Pearson coefficient of correlation was an effective way of measuring the relationship between the outcome variable that's profit and the experiment variables that's sales, quantity, and discount. The Pearson product-moment correlation coefficient is defined as a measure of the strength of a linear association between two variables and is denoted by r. Here a null hypothesis was formulated to test the relationship that's $\rho \neq 0$ vs an alternative hypothesis $\rho = 0$ with a 0-coefficient indicating that there is no relationship between the variables. The relationship was as follows:

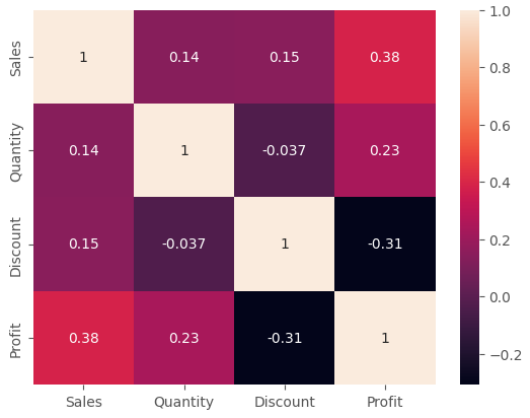

Fig. 7 Correlation matrix

Basing on the graph the coefficient values ranged from +1 to -1 with values greater than 0 indicating a positive relationship between the respective variables and values less than 0 indicating a negative relationship between the respective variables. Sales were highly correlated to profit as compared to other experiment variables with a positive coefficient value of 0.38 this indicated the more the sales the more the profits, the correlation between quantity and profit had a weak positive relationship of 0.23 meaning the better the quality of the commodity the higher the profit rates of the product and the correlation coefficient between discount and profit had a negative relationship of -0.31 this meant that the more the discount offered on each commodity the lesser the profit.

To determine how much variation was evident or captured the principal component analysis was used where a 97.22% variance was captured for the first principal and a 2.75% for the second principal.

3.2 Regression Analysis Results

Will begin by splitting the data into training and testing sets with a split ratio of 0.4 which is 40% of the testing set that will be, later on, used for analysis and prediction with the regression.

The multiple regression output was as follows table 2:

Table 2. Regression Outputs

Dep. Variable:	Profit	R-squared (uncentered): 0.563	
Model:	OLS	Adj.R-squared (uncentered): 0.563	
Method:	Least Squares	F-statistic: 916.1	
Date:	Mon, 26 Sep 2022	Prob (F-statistic): 0.00	
Time:	05:33:05	Log-Likelihood: -8742.5	
No. Observations:	2135	AIC: 1.749e+04	
Df Residuals:	2132	BIC: 1.751e+04	
Df Model:	3		
Covariance Type:	nonrobust		

	Coef	std err	t	P>|t|	[0.025	0.975]
Sales	0.0851	0.003	25.142	0.000	0.078	0.092
Quantity	2.9804	0.110	27.082	0.000	2.765	3.196
Discount	-49.4835	2.769	-17.872	0.000	-54.913	-44.054

Omnibus:	211.945	Durbin-Watson:	2.002
Prob(Omnibus):	0.000	Jarque-Bera (JB):	958.768
Skew:	-0.373	Prob(JB):	6.40e-209
Kurtosis:	6.197	Cond. No.	1.07e+03

Notes: R^2 is computed without centering (uncentered) since the model does not contain a constant. Standard Errors assume that the covariance matrix of the errors is correctly specified. The condition number is large, 1.07e+03. This might indicate that there are strong multicollinearity or other numerical problems.

The regression coefficient values for our model, were 0.0936, 2.8901, and -51.3530 which showed the variation of the dependent variable whether increasing for positive values and decreasing for

negative values. The t-value represented the model's coefficients divided by the standard error, these we essential for testing the hypothesis that the coefficient values are not equal to zero as all the coefficient values were statistically significant as their p-values were lesser than the assumed alpha 0.05. The model registered an R-square of 0.577 Our model, therefore, simplifies to Profit = $0.0936(sales) + 2.8901(Quantity) - 51.3530(Discount)$. It follows from the correlation test that profits were highly determined by sales as there was a high correlation relationship between the variables

4. Conclusion

This study was conducted on the basis of the following research question Is there a multiple linear regression model representing the outcome variable profit with the factors that influence the profitability of a supermarket in the United States that's sales, discount, and quality of the product? To answer the research question, the following hypotheses are proposed:

There is no relationship between the outcome variable profit, and the experiment variables sales, discount, and quality of the product and there is no multiple linear regression defining the variables. Having conducted the analysis which gave the results presented in the data analysis and results in the section this paper concludes that there is a multiple linear regression model representing the outcome variable with all the predictors as the f score 0.00 was less than the assumed alpha. Here can see that the aim of attaining profit the more the sales made by the supermarket and the higher the quality of the products distributed to different destinations the higher the profit brought to the supermarket with discounts posing a negative variation to the supermarket gains.

Highly recommendations should be made to supermarkets to maximize the sales and the quality of each commodity to increase its profit also even though giving discounts is a form of persuading the customers to buy more commodities it negatively affects the chances of profitability in the supermarket. Develop Positive Relationships with Our Customers and Distributors In order for the supermarket to fulfill its strategic goal, that must build a functional and long-term relationship with customers. This can be accomplished by improving master shopping cards, awarding points that can be redeemed for future purchases, providing free delivery services, developing an online shopping platform for our online customers, and creating a database for our clients so that this study can maintain regular communication with them. To maintain a favorable relationship with our Distributors, this paper must have a dependable inventory system that is up and running and that this paper can share with them.

Significant Distinctiveness and Skill the Supermarket will spend in order to supply our customers with one-of-a-kind products from our suppliers and to gain new rebranding knowledge. This will necessitate working with our numerous vendors to provide unique items that are suited to our consumers' preferences and preferences. The supermarket will conduct a survey to learn more about our customers' preferences in order to better serve them.

Expansions of the Market The supermarket will investigate new areas and markets where its target customers are located, and then, based on the findings of this investigation, it will open a new branch in close proximity to these suburbs. The Supermarket will also investigate regions where it can compete fairly with its competitors and will open outlets to tempt customers who already shop at those competitors' facilities.

Pricing The supermarket will look for the most dependable suppliers and wholesalers who offer high-quality products at cheap pricing. The flexible margins will be passed on to the customer, ensuring that the supermarket's products are priced competitively in comparison to those offered by competing businesses.

Marketing Techniques This is related to the grocery store's marketing. This paper are employing the necessary strategies and channels to successfully sell the supermarket. This can be performed by outlining all of our marketing tactics and recruiting them all to help fuel the growth of our store. 6. Techniques: This includes activities that must be conducted on a daily basis as part of the

supermarket's operations. This paper will use the framework that is currently in place to do these tasks. The strategy and long-term vision of the Supermarket will serve as the foundation for the creation of its techniques. Because cultivating customer relationships is our major priority, the techniques this paper adopt will prioritize establishing a favorable reputation and fostering deep ties with clients, giving customers the feeling that our solutions provide genuine value.

References

[1] Littleton, A.C. What Is Profit? Accounting Review, 1928, Pp.278-288.

[2] Burja, C. Factors Influencing The Companies 'profitability. Annales Universitatis Apulensis: Series Oeconomica, 2011, 13(2), P.215.

[3] Wood, A. A Theory Of Profits. Cup Archive 1975

[4] Martín-Oliver, A. And Salas-Fumás, V. The Output And Profit Contribution Of Information Technology And Advertising Investments In Banks. Journal Of Financial Intermediation 2008, 17(2), Pp.229-255.

[5] Suhaimi, R., Abdullah, F. And Saban, G., December. Factors Affecting Profit Efficiency Of Commercial Banks In Malaysia. In 2010 International Conference On Science And Social Research 2010 (Cssr 2010) (Pp. 1059-1064).

[6] Tey, Y.S. And Brindal, M.,. Factors Influencing Farm Profitability. Sustainable Agriculture Reviews, 2015 Pp.235-255.

[7] Chevalier, Michel. "Increase in sales due to in-store display." Journal of marketing research 12.4 (1975): 426-431.

[8] Thanassoulis, John. "Supermarket Profitability Investigation." (2009).

[9] Chintagunta, Pradeep K., Jean-Pierre Dubé, and Vishal Singh. "Balancing profitability and customer welfare in a supermarket chain." Quantitative Marketing and Economics 1.1 (2003): 111-147.

[10] Gutman, Graciela E. "Impact of the rapid rise of supermarkets on dairy products systems in Argentina." Development Policy Review 20.4 (2002): 409-427.

[11] Reynolds, Paul L., and Geoff Lancaster. "A scheme to increase profitability in entrepreneurial SMEs." Journal of Small Business and Enterprise Development (2006).