

MARKET ANALYSIS OF SALES

Impact of Advertisement Platforms on Sales

Benard Kipngeno

11 May 2021

1. Aim

The objective of this study is to investigate the influence of the following advertisement platforms; youtube,facebook and newspaper on the product sales promotion and marketing in the United states. To achieve the aim of the research,marketing data was extracted from datarium R package and saved as Excel CSV format. It is through this study that the marketing performance of each platform will be measured as well as identifying the best marketing platform among the three. The underlying customers patterns in the preference of advertisement platforms will be revealed in this research. This study also aims at improving on the previous research conducted on the factors affecting market of United States products. The following are peer reviewed references;<https://www.sciencedirect.com/science/article/abs/pii/S0747563216300450>,<https://www.ingentaconnect.com/content/hsp/jdsmm/2016/00000003/00000004/art0009>,<https://www.wakeenpages.com/blog/the-impact-of-newspaper-advertising/>.

2. Question and problem

The main question of this research is how the advertisements channels i.e. youtube,facebook,newspapers affect product sales? What is the magnitude of each platform on the product sales promotion? Does the advertisement platforms explain the variation in the product sales?

3. Scope of the research

This study focuses on the impact of the three advertisements channels; youtube,facebook and newspapers on the product sales promotion in United states but the findings may be generalized to other countries. These platforms are also gaining popularity exponentially in the rest of the world and with time the generalizations of this findings may also extend to other parts of the world.

4. Methods

The data used in this research was obtained from the *datarium R package*, it's called "*marketing*". Exploratory data analysis was done to understand the underlying patterns in the data. Descriptive statistics were generated and Graphs such as histogram and boxplot was drawn to visualize distribution of the variable sales and check for the outliers in the data. To answer research questions statistical test such as hypothesis testing of correlation and plotting scatter plot were carried out. Multiple linear regression model was used understand the impact of each independent variable on the dependent variable.

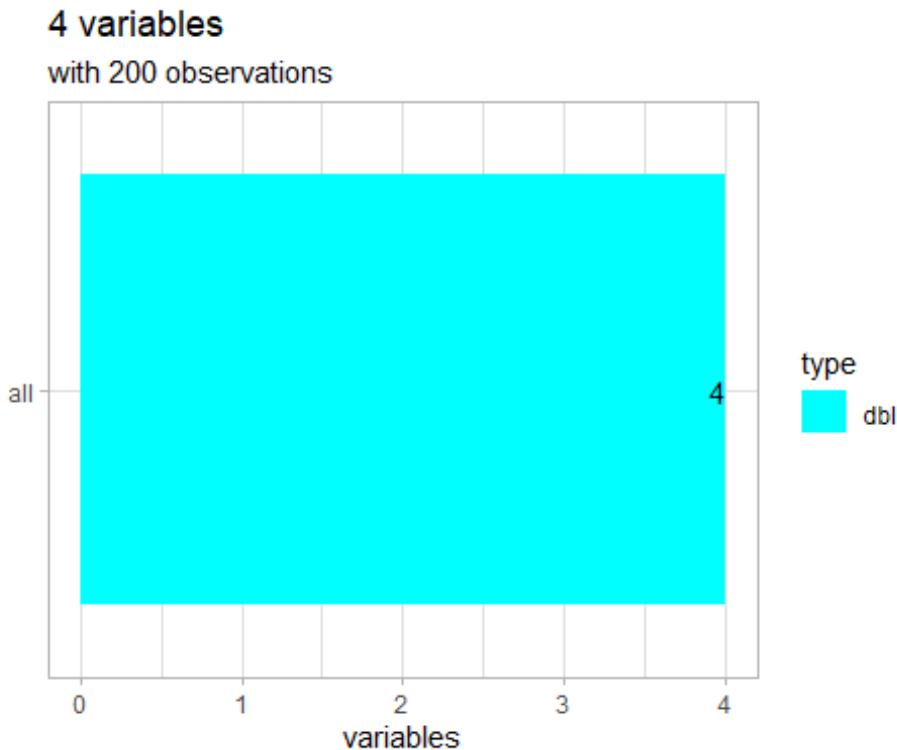
5. Results

5.1 Exploratory Data Analysis of Marketing Data

5.1.1 Descriptive Statistics

5.1.1.1 Examining the number of variables and observation in the data

The figure below describe the data visually;



The data consist of four continuous variables that is youtube,facebook,newspapers and sales and 200 observations which are recorded in thousands of dollars except sales which is recorded in billions of dollars. The recorded observations of the advertisement platforms are the cost incurred in marketing a product through it while sales are the revenue from the sales of the product.

5.1.1.2 Checking for missing values and unique values

The table below shows features of the variables in the data;

variable	type	na	na_pct	unique	min	mean	max
youtube	dbl	0	0	190	0.84	176.45	355.68
facebook	dbl	0	0	167	0.00	27.92	59.52
newspaper	dbl	0	0	172	0.36	36.66	136.80
sales	dbl	0	0	121	1.92	16.83	32.40

The data had no missing values.

5.1.1.3 Exploring data by measures of central tendencies

The table below shows summary statistics of each variable;

youtube	facebook	newspaper	sales
Min. : 0.84	Min. : 0.00	Min. : 0.36	Min. : 1.92
1st Qu.: 89.25	1st Qu.: 11.97	1st Qu.: 15.30	1st Qu.: 12.45
Median : 179.70	Median : 27.48	Median : 30.90	Median : 15.48
Mean : 176.45	Mean : 27.92	Mean : 36.66	Mean : 16.83
3rd Qu.: 262.59	3rd Qu.: 43.83	3rd Qu.: 54.12	3rd Qu.: 20.88
Max. : 355.68	Max. : 59.52	Max. : 136.80	Max. : 32.40

The revenue accumulated on average is 20.90 billion dollars. The cost incurred in youtube advertisements is almost 5 times the cost incurred in the facebook advertisement and 6 times the cost incurred in newspaper advertisements.

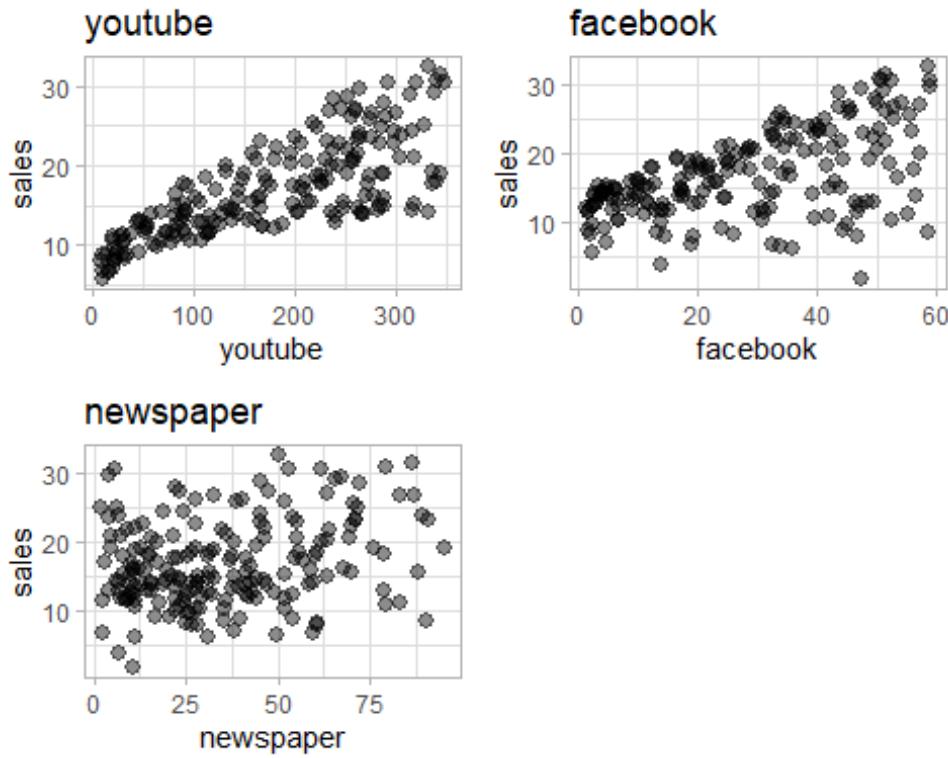
5.1.1.4 Exploring data to understand distribution of the variables



The above normal distribution curves depicts that the variables are approximately distributed, therefore the assumption of normal distribution is not violated.

5.1.1.5 Examining the relationship that exists between dependent variable and independent variables

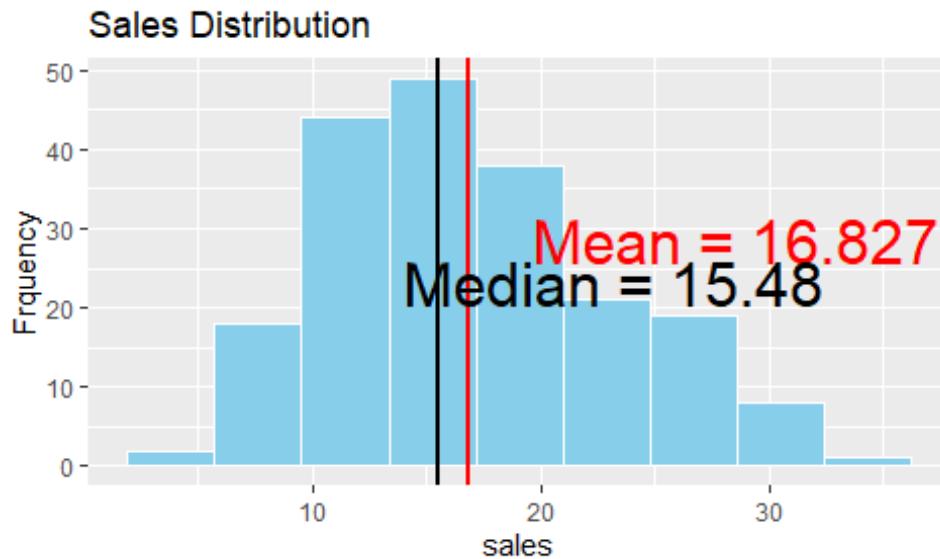
The graphs below show the relationship between the target variable and the independent



variables.

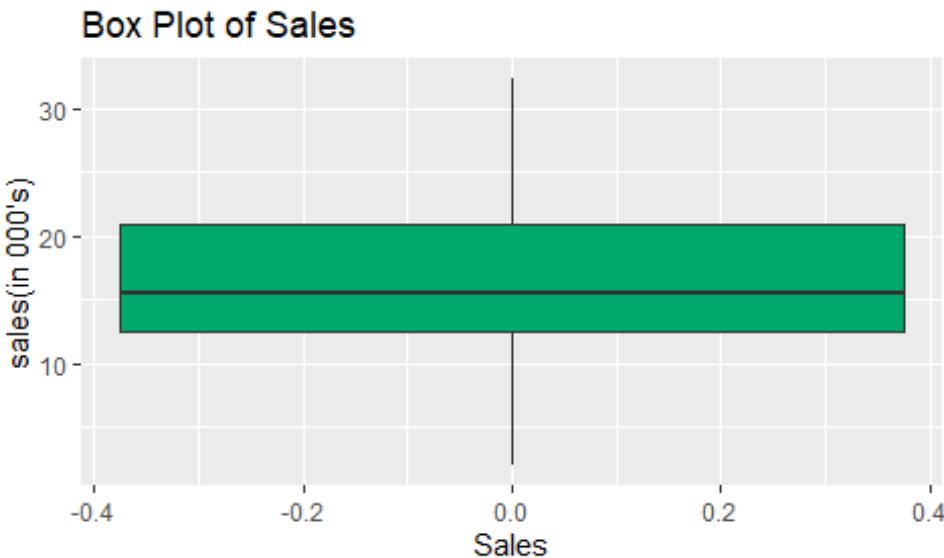
5.1.1.6 Checking for the sales distribution using histogram

The histogram below shows the distribution of sales, it clearly shows that the observation of the sales nearly follows normal distribution.



5.1.1.7 Box plot to check for outliers in the sales

The box plot below shows that there are no outliers in the variable sales since all observations are within the minimum and maximum range of the box plot.



5.1.1.8 Checking the Multi-collinearity of independent variables

The tables below show the correlation coefficient of the independent variables. According to the results obtained, there is no correlation coefficient between independent variables that exceeds 0.5, therefore multi-collinearity does not exist. This implies that linear regression assumption is not violated and all three variables can be used in the linear model predicting sales.

	youtube	facebook	newspaper
youtube	1.0000000	0.0548087	0.0566479
facebook	0.0548087	1.0000000	0.3541038
newspaper	0.0566479	0.3541038	1.0000000

5.2 Data Analysis

5.2.1 Correlation of dependent variable against the independent variables

Correlation is the most appropriate statistical analysis test used to check if there exists linear relationship between dependent variable sales and independent variables youtube, facebook and newspapers and it also applies only to continuous variables.

5.2.2 Correlation between dependent variable against all independent variables

5.2.2.1 Correlation Table

Variable	sales
sales	1.00

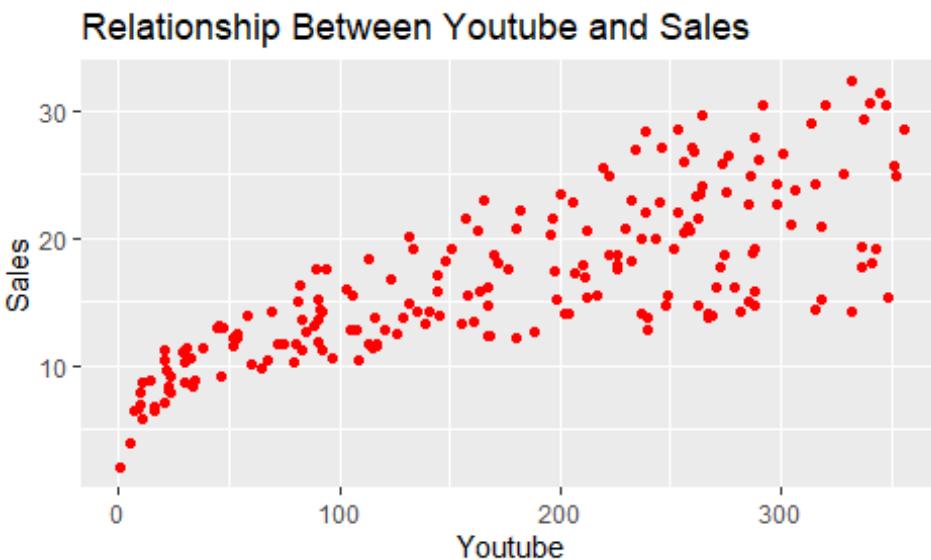
youtube	0.78
facebook	0.58
newspaper	0.23

5.2.3 Hypothesis testing of correlations

5.2.3.1 Testing the significance of the relationship between Sales and Youtube

To achieve this, null and alternative hypothesis were formulated as follows; Null hypothesis states that there is no linear significant relationship between the sales revenue and the cost incurred in youtube advertisements.

Alternative hypothesis states there is linear significant relationship between the sales revenue and the cost incurred in youtube advertisements. Below are the scatter plot and the correlation hypothesis testing results;



```
## 
## Pearson's product-moment correlation
## 
## data: marketing$sales and marketing$youtube
## t = 17.668, df = 198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7218201 0.8308014
## sample estimates:
##        cor
## 0.7822244
```

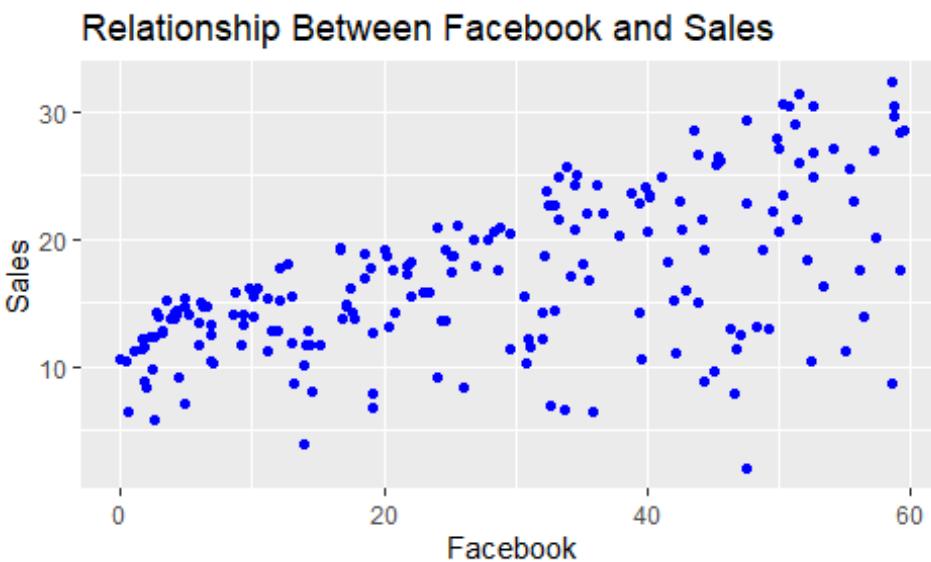
According to the results above, at 95% confidence level p-value(2.2e-16) is less than alpha-value(0.05), therefore we reject null hypothesis. This implies that there is linear significant relationship between the sales revenue and the cost incurred in youtube

advertisements. The correlation coefficient is 0.78, therefore there is a strong positive linear significant relationship between the sales revenue and the cost incurred in youtube advertisements. This is also depicted clearly in the scatter plot by an upward increasing pattern.

5.2.3.2 Testing the significance of the relationship between Sales and Facebook

To achieve this, null and alternative hypothesis were formulated as follows; Null hypothesis : states that there is no linear significant relationship between the sales revenue and the cost incurred in facebook advertisements.

Alternative hypothesis states there is linear significant relationship between the sales revenue and the cost incurred in facebook advertisements. Below are the scatter plot and the correlation hypothesis testing results;



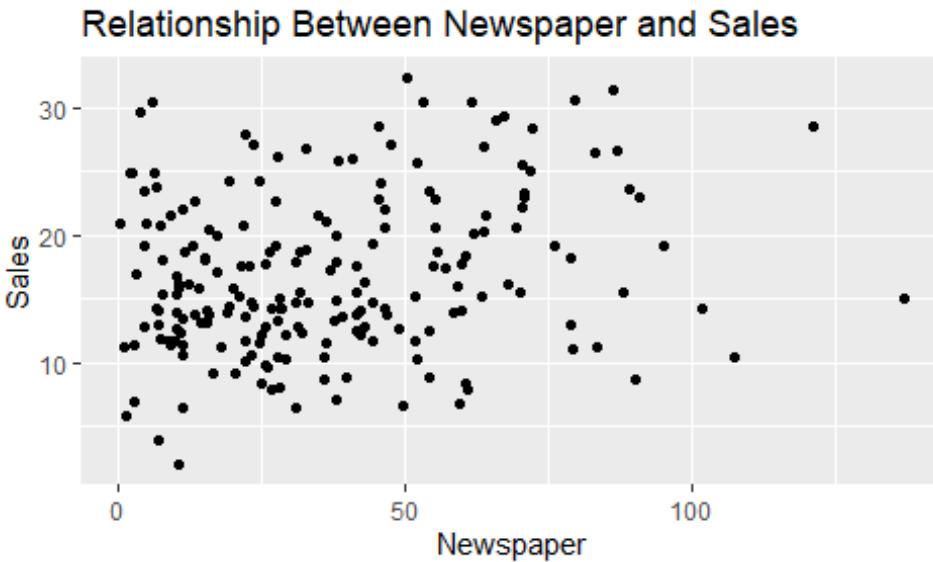
```
## 
## Pearson's product-moment correlation
## 
## data: marketing$sales and marketing$facebook
## t = 9.9208, df = 198, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4754954 0.6620366
## sample estimates:
##      cor
## 0.5762226
```

According to the results above, at 95% confidence level p-value(2.2e-16) is less than alpha-value(0.05), therefore we reject null hypothesis. This implies that there is linear significant relationship between the sales revenue and the cost incurred in facebook advertisements. The correlation coefficient is 0.58, therefore there is a moderate positive linear significant relationship between the sales revenue and the cost incurred in facebook advertisements. This is also clearly in the scatter plot as it is an direct relationship.

5.2.3.3 Testing the significance of the relationship between Sales and Newspaper

To achieve this, null and alternative hypothesis were formulated as follows; Null hypothesis : states that there is no linear significant relationship between the sales revenue and the cost incurred in newspaper advertisements.

Alternative hypothesis states there is linear significant relationship between the sales revenue and the cost incurred in newspaper advertisements. Below are the scatter plot and the correlation hypothesis testing results;



```
##  
## Pearson's product-moment correlation  
##  
## data: marketing$sales and marketing$newspaper  
## t = 3.2996, df = 198, p-value = 0.001148  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.0924875 0.3557712  
## sample estimates:  
## cor  
## 0.228299
```

According to the results above, at 95% confidence level p-value(0.001148) is less than alpha-value(0.05), therefore we reject null hypothesis. This implies that there is linear significant relationship between the sales revenue and the cost incurred in newspaper advertisements. The correlation coefficient is 0.23, therefore there is a weak positive linear significant relationship between the sales revenue and the cost incurred in newspaper advertisements. This weak relationship is also clearly observed in the scatter plot as there is no clear pattern taken by the relationship.

5.2.4 Multiple Linear Regression

To answer the objective of this study, further investigation were performed that is more certain that the sales revenue is influenced by the medium of the advertisement. Multiple linear regression is the most appropriate to investigate the magnitude of each platform on sales revenue relative to others. The dependent variable is sales revenue predicted by independent variables youtube,facebook,newspapers. This results of the linear regression are shown below.

5.2.4.1 Regression Analysis

```
##  
## Call:  
## lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max  
## -10.5932  -1.0690   0.2902   1.4272   3.3951  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.526667  0.374290  9.422 <2e-16 ***  
## youtube     0.045765  0.001395 32.809 <2e-16 ***  
## facebook    0.188530  0.008611 21.893 <2e-16 ***  
## newspaper   -0.001037  0.005871 -0.177    0.86  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 2.023 on 196 degrees of freedom  
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956  
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

5.2.4.2 Accessing significance of the model

The multiple linear regression model is significant and fit for predicting sales revenue since at 95% confidence level,p-value(2.2e-6) is less than alpha-value(0.05).

5.2.4.3 Interpretation of coefficient of determination(R-squared)

89.72% of the variation in the sales revenue is explained by the independent variables youtube, facebook and newspapers while the remaining 11.28% cannot be accounted for by the model.

5.2.4.4 Accessing the magnitude of each independent variables on the dependent variable

The independent variable YouTube and facebook are the only significant independent variables because their p-value is less than the significance level 0.05. Newspaper is insignificant in the model because its p-value is greater than significance level 0.05. This implies that YouTube and facebook are the only variables to be considered when choosing advertisement medium for product sales promotion and marketing. Holding all other factors constant,one unit change in the cost incurred in YouTube advertisements results in

0.045765 in the sales revenue. Holding all other factors constant, one unit change in the cost incurred in facebook advertisements results in 0.188530 in the sales revenue.

5.3 Variable Selection using k-fold Cross-validation Method

A more rigorous way of selecting models based on the prediction error computed on a new test data using k-fold cross-validation techniques. id: model id, object: regsubsets object data: data used to fit regsubsets, outcome: outcome variable.

5.3.1 CV Errors

```
## [1] 3.898148 2.037649 2.054538
```

5.3.2 Select the model that minimize the CV error

```
## [1] 2

## (Intercept) youtube facebook
## 3.50531989 0.04575482 0.18799423

##
## Call:
## lm(formula = sales ~ youtube + facebook, data = marketing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -10.5572 -1.0502  0.2906  1.4049  3.3994 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 3.50532   0.35339  9.919 <2e-16 ***
## youtube     0.04575   0.00139 32.909 <2e-16 ***
## facebook    0.18799   0.00804 23.382 <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.018 on 197 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8962 
## F-statistic: 859.6 on 2 and 197 DF,  p-value: < 2.2e-16
```

K-fold cross validation Method affirms that the best explanatory variables of sales are Facebook and YouTube only. YouTube turned out to be the most significant explanatory variable of sales.

6. Main Conclusion

The investigation of this study establishes that YouTube and facebook are the advertisement media that should be considered by any company or organization when making decision on which platform to invest in advertising their product in order to promote their sales in the market. This also confirms that company that deals in product

sales or any marketing ought to conform to market dynamics since most of the customers spend their time online hence its the strategic place to channel adverts of products.

7. References

<https://search.proquest.com/openview/8e47927e86a14cc15df27ede8444480f/1?pq-origsite=gscholar&cbl=1818401>

https://books.google.co.ke/books?hl=en&lr=&id=Us4YE8lJVYMC&oi=fnd&pg=PP2&dq=regression+analysis&ots=WVGobxRjZe&sig=gOPy2Imq82AR1SKB96woOo9VNPM&redir_esc=y#v=onepage&q=regression%20analysis&f=false