**Data Analysis**

The increased competition in the digital market in UK is the biggest SMEs are experiencing, thus it is significant to understand how SMEs can maintain a stronger competitive advantage (Scuotto et al., 2021). To achieve this, the following statistical variables are to be analyzed: total online sales, digital marketing investment/expenditure and ROI.

The following hypothesis were formed and tested on our SME dataset;

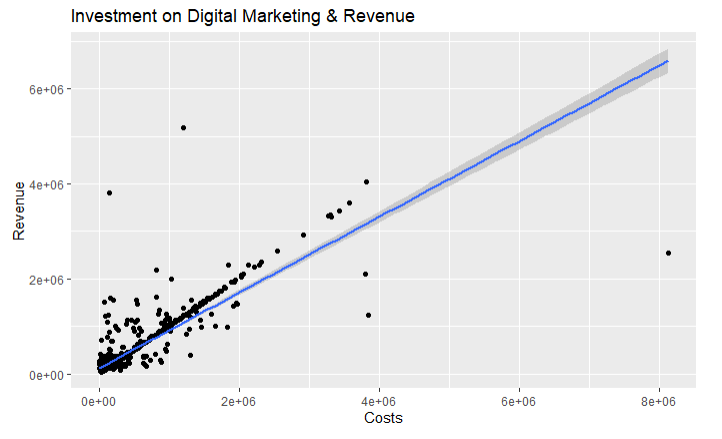
* Null Hypothesis 1 - Higher investment on digital marketing by SMEs leads to higher ROI/revenue.
* Null Hypothesis 2 - Higher investment on digital marketing leads to higher (online) sales by SMEs.

**Data cleaning**

In R, cleaning data can be carried out using various packages and functions (Gendron, 2016). For instance, dpyr library provides various functions such as filtering, sorting. In our case, we filtered out dataset by selecting specific variables of interest on which we want to perform analysis on. Sorting was done to arrange values by the online sales column.

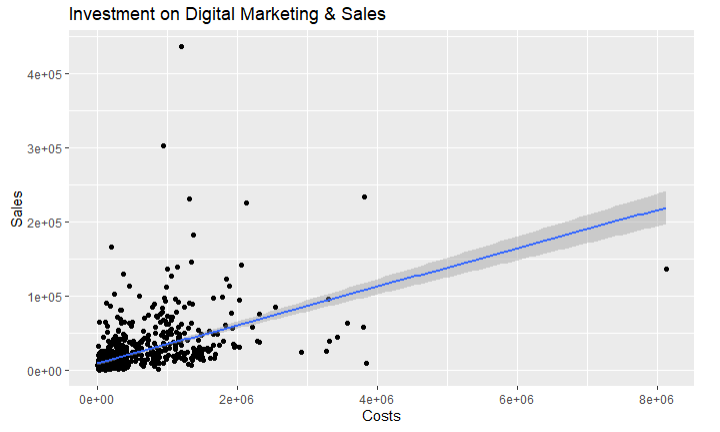
**Hypothesis Testing**

**Null Hypothesis 1** - Higher investment on digital marketing by SMEs leads to higher ROI/revenue



From the above plot, there is a positive correlation between costs and revenue. This observation implies that higher investment on digital marketing does lead to higher ROI for SMEs. Thus, we accept the null hypothesis.

**Null Hypothesis 2** - Higher investment on digital marketing leads to higher (online) sales by SMEs.



From the above plot, there is a positive correlation between costs and sales. This observation implies that higher investment on digital marketing does lead to higher (online) sales for SMEs. Thus, we accept the null hypothesis.

**Linear Regression**

Two linear regression models were created.

**Model 1**

Linear regression was done using costs as the capital investment (predictor) and sales as the depended variable. The following results were found;

Call:

lm(formula = sales ~ costs, data = select\_data)

Residuals:

Min 1Q Median 3Q Max

-99281 -9267 -5250 1405 395550

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 9.885e+03 1.071e+03 9.233 <2e-16 \*\*\*

costs 2.585e-02 1.469e-03 17.601 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 26620 on 972 degrees of freedom

(26 observations deleted due to missingness)

Multiple R-squared: 0.2417, Adjusted R-squared: 0.2409

F-statistic: 309.8 on 1 and 972 DF, p-value: < 2.2e-16

**Results interpretation of model 1**

The results implies that there is a positive correlation between costs and sales. The cost coefficient (2.585e-02) is statistically significant, as indicated by the p-value (<2e-16). This indicates that there is a statistically significant relationship between costs and sales. The adjusted R-squared value of 0.2409 indicates that 24.09% of the variation in sales can be explained by costs. The F-statistic of 309.8 and p-value of <2.2e-16 indicate that the model is statistically significant.

**Model 2**

Linear regression was done using costs (predictor) and revenue (depended) variable. The following results were found;

Call:

lm(formula = revenue ~ costs, data = select\_data)

Residuals:

Min 1Q Median 3Q Max

-4046366 -84575 -57551 31530 4093500

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.246e+05 1.205e+04 10.34 <2e-16 \*\*\*

costs 7.961e-01 1.653e-02 48.15 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 299600 on 972 degrees of freedom

(26 observations deleted due to missingness)

Multiple R-squared: 0.7046, Adjusted R-squared: 0.7043

F-statistic: 2319 on 1 and 972 DF, p-value: < 2.2e-16

**Results interpretation of model 2**

The linear model used to predict revenue from costs has a relatively high R-squared value (0.7046). This indicates that the model fits the data well. The intercept coefficient estimate is 124600, which means that the average revenue are 124600 when the costs are 0. The slope coefficient, that is the estimated effect of costs on revenue, is 0.796. This means that for every additional unit of cost, the revenue increase by an average of 0.796 units. The p-value for the slope coefficient is less than 0.001, which means that the coefficient is statistically significant. This indicates that increasing investments in digital marketing does lead to higher revenue for SMEs.

**Conclusion**

In conclusion, it is evident from the hypothesis conducted that higher digital marketing investment results in higher ROI and online sales for SMEs. Further, the results of linear regression models indicates that costs coefficient is statistically significant.

**Implications**

The findings shows that investing in digital marketing can be an effective way for SMEs to increase their ROI and online sales. Thus, it is recommended that SMEs should invest heavily in digital marketing for increased profitability and so as to gain competitive edge.

**Reference**

Gendron, J. (2016). *Introduction to R for Business Intelligence*. [online] *Google Books*. Packt Publishing Ltd. Available at: https://books.google.co.ke/books?hl=en&lr=&id=XkvWDQAAQBAJ&oi=fnd&pg=PP1&dq=In+R+statistical+language [Accessed 13 Mar. 2023].

Scuotto, V., Nicotra, M., Del Giudice, M., Krueger, N. and Gregori, G.L. (2021). A microfoundational perspective on SMEs’ growth in the digital transformation era. *Journal of Business Research*, 129, pp.382–392. doi:https://doi.org/10.1016/j.jbusres.2021.01.045.