Star Digital ads marketing analysis

In this project, we want to solve the following questions:

1. Is online advertising effective for Star Digital?
2. Whether increasing the frequency of advertising increases the probability of purchase?
3. Which sites should Star Digital advertise on?

1. Is online advertising effective for Star Digital?

From our logistic regression model, the p-value of "test" is 0.0614, which is greater than 0.05. It means there is no strong correlation between seeing the advertising and purchase. However, when we take a look of the detail, online advertising has a positive effect on Star Digital overall.

```
Call:
glm(formula = purchase ~ test, family = binomial(), data = Star_Digital)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.186  -1.186   1.169   1.169   1.202

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.05724    0.03882  -1.474   0.1404
test         0.07676    0.04104   1.871   0.0614 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 35073  on 25301  degrees of freedom
AIC: 35077

Number of Fisher Scoring iterations: 3
```

2. Whether increasing the frequency of advertising increases the probability of purchase?

```
Call:
glm(formula = purchase ~ rowsums6, family = binomial(), data = Star_Digital)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-4.856  -1.120   0.139   1.222   1.249

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.166175   0.014385  -11.55   <2e-16 ***
rowsums6     0.030583   0.001381   22.15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 35077  on 25302  degrees of freedom
Residual deviance: 34247  on 25301  degrees of freedom
AIC: 34251

Number of Fisher Scoring iterations: 5
```

```
> exp(coef(fit))
(Intercept)    rowsums6
  0.8468978   1.0310552
```

From the outcome above, the exp value of Sum of imp1-6 is 1.0311, which means "For Per thousand impressions increases in Sum of imp1-6, the odds of purchasing Star Digital products increases by 3.11%.

```
Call:
glm(formula = purchase ~ imp_1 + imp_2 + imp_3 + imp_4 + imp_5 +
    imp_6, family = binomial(), data = Star_Digital)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.3241  -1.1070   0.0135   1.2385   4.0054

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.186300   0.014502 -12.846  < 2e-16 ***
imp_1       -0.011802   0.003987  -2.960  0.00307 **
imp_2        0.014579   0.001891   7.708 1.28e-14 ***
imp_3        0.215225   0.151530   1.420  0.15551
imp_4        0.189067   0.007632  24.773  < 2e-16 ***
imp_5       -0.489558   0.052361  -9.350  < 2e-16 ***
imp_6        0.018527   0.003253   5.694 1.24e-08 ***
---
```

```
> exp(coef(fit.full))
(Intercept)       imp_1       imp_2       imp_3       imp_4       imp_5       imp_6
  0.8300245   0.9882678   1.0146859   1.2401407   1.2081213   0.6128970   1.0186992
```

From our logistic regression model, imp_2(0.0146), imp_3(0.215), imp_4(0.189), and imp_6(0.0185) have **positive impacts** on sales.

From our Exp function, we could make the following conclusions:

For Per thousand impressions increases in imp_2, the odds of purchasing Star Digital products increases by 1.47%.

For Per thousand impressions increases in imp_3, the odds of purchasing Star Digital products increase by 24.01%.

For Per thousand impressions increases in imp_4, the odds of purchasing Star Digital products increases by 20.81%.

For Per thousand impressions increases in imp_6, the odds of purchasing Star Digital products increases by 1.87%.

Imp_1(-0.0118), and imp_5(-0.4896) have **negative impacts** on the sales.

For Per thousand impressions increases in imp_1, the odds of purchasing Star Digital products decreases by 1.12%.

For Per thousand impressions increases in imp_5, the odds of purchasing Star Digital products decreases by 38.71%.

So there is a frequency effect of advertising on purchase.

3. Which sites should Star Digital advertise on?

```
Call:
glm(formula = purchase ~ rowsums5, family = binomial(), data = Star_Digital)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.1213  -1.1351   0.1338   1.2203   1.2349

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.134265   0.013811  -9.722   <2e-16 ***
rowsums5     0.033883   0.001564  21.668   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Call:
glm(formula = purchase ~ imp_6, family = binomial(), data = Star_Digital)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6634  -1.1773   0.8747   1.1874   1.1874

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.023406   0.013400  -1.747   0.0807 .
imp_6        0.023144   0.003228   7.169 7.56e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


> exp(coef(fit.rowsums5))
(Intercept)    rowsums5
  0.8743583   1.0344638
> exp(coef(fit.6))
(Intercept)       imp_6
  0.9768658   1.0234143
```

From the outcome above, the exp value of SumofIMP1_5 is 1.0345, which means “For Per thousand impressions increases in SumofIMP1_5, the odds of purchasing Star Digital products increases by 3.45%. The exp value of imp_6 is 1.0234, which means “For Per thousand impressions increases in imp_6, the odds of purchasing Star Digital products increases by 2.34% < 3.45%. Therefore, Star Digital should **advertise on sites 1 to 5 as a whole.**