**Team Assignment, MKTG 562 (Pricing)**

Group: Aditya, Manasi, Mervin and Soundari

*Question 1: Using a linear regressions framework with unit sales (and/or log unit sales) as the dependent variable, investigate the effect of regular price, feature and rating on sales for each of the three apps. How good are these models?*

**Answer:**

In our analysis, we first build the three types of regression models (linear, semi-log and log) for each competitive app. We then compare these models for every app based on 1) How many of the features are statistically significant 2) Which model has a higher R-squared/ Adjusted R squared value. To interpret and compare the coefficients of every feature in a model, we apply the formulas for interpreting coefficients explained in class i.e.  
1. Linear Regression Model:

Coeff.feature1 = (beta1.feature1)\*mean(feature1)/mean(UnitSales)

2. Semi Log Model:

Coeff.feature1 = (beta1.feature1)\*mean(feature1)

3. Log Model:

Interpreted coefficients as is.

Note: Find the analysis for every app in R markdown file ‘Question 1: EX3 Pricing’

App 1

The following table explains the regression models run for the first App.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.886 | 0.83 | 0.837 |
| Adjusted R Sq | 0.883 | 0.83 | 0.833 |
| Statistical Significance | Intercept **not sig**  Price, Rating, Featuring **sig** | Intercept,  Price, Rating, Featuring **all sig** | Intercept,  Price, Rating, Featuring **all sig** |
| Interpretation | After considering the mean of features and mean of sales in the coefficient, we find that all the three models tell us that for unit increase in price, the sales go down by about 0.7, for featuring the app the sales increase by a negligible 0.07 amount and finally for unit increase in rating the sales increase by around 1.2-1.3 (find detailed analysis in R Markdown) | | |

Conclusion: Since the adjusted R squared and multiple R squared are more for the **Linear Regression** as compared to **log and semi log,** and all the features are equally significant for linear, semi log and log regressions, we go for the regression with higher R sq value and which is also the simpler option, that is the Linear Regression. Also, the intercept is not significant in the Linear Regression, which may be an issue while predicting the unit sales from the three features. However, in our analysis in Question 1, we are merely investigating the effect of featuring on homepage, price of app and rating of app on the unit sales, Linear Regression works well.

App 2

The following table explains the regression models run for the second App.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.67 | 0.39 | 0.39 |
| Adjusted R Sq | 0.66 | 0.37 | 0.38 |
| Statistical Significance | Intercept,  Price, Rating, Featuring **all sig** | Intercept,  Price, Rating, Featuring **all sig** | Intercept,  Price, Rating, Featuring **all sig** |
| Interpretation | After considering the mean of features and mean of sales in the coefficient, we find that all the three models tell us that for unit increase in price, the sales go down by about 2.6, for featuring the app the sales increase by a negligible 0.1 amount and finally for unit increase in rating the sales decrease by around 5.1 (find detailed analysis in R Markdown). This is interesting because the analysis tells us that the ratings are reason for the app sales going down. It will be worth trying to see the effect of other ratings of competitive apps for the sales of this app | | |

Conclusion: Since the adjusted R squared and multiple R squared are largely higher for the **Linear Regression** as compared to **log and semi log,** and all the features are equally significant for linear, semi log and log regressions, we go for the regression with higher R sq value and which is also the simpler option, that is the Linear Regression.

App 3

The following table explains the regression models run for the second App.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.47 | 0.44 | 0.44 |
| Adjusted R Sq | 0.46 | 0.43 | 0.42 |
| Statistical Significance | Intercept,  Price, Rating **not sig**  Featuring **sig** | Intercept **not sig**  Price, Rating, Featuring **all sig** | Intercept,  Price, Rating, Featuring **all sig** |
| Interpretation | After considering the mean of features and mean of sales in the coefficient, we find that all the three models tell us that for unit increase in price, the sales go down by about 2.6, for featuring the app the sales increase by a 0.25 amount and finally for unit increase in rating the sales increase by around 8.9 (find detailed analysis in R Markdown). This is interesting because the analysis tells us that the ratings are a big reason for the app sales going up. It will be worth trying to see the effect of other ratings of competitive apps for the sales of this app | | |

Conclusion: Although the adjusted R squared and multiple R squared are largely higher for the **Linear Regression** as compared to **log and semi log,** and the features are not statistically significant for linear whereas, they are significant for semi log and log regressions. Hence we go for the Semi Log regression with higher R sq value and which is also the simpler option between semilog/log.

*Question 2: An important source of variation in sales often comes from competitive marketing activity. Investigate the impact on sales of each app from the changes in the marketing activity of competing apps. What, if any, competitive terms would you want to include in your final models?*

In order to find if the other app’s marketing strategy affects the sales of the app that we are analyzing, we have used three models linear regression, semi-log model and log-log model.

As for the previous question, we have compared these models for every app based on 1) the number of features that are statistically significant 2) higher R-squared/ Adjusted R squared value. Then we have compared the coefficients of every feature in a model using the formulas for interpreting coefficients as explained in class.

**App 1 effects based on features of App 2 and App 3**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.0569 | 0.0720 | 0.071 |
| Adjusted R Sq | 0.0035 | 0.0194 | 0.019 |
| Statistical Significance | Only Rating of App 2 **sig** | Only Rating of App 2 | Only Rating of App 2 |
| Interpretation | Only the rating of the 2nd App is affecting the sales of the first app. But the R sq values are not reliable | | |

**App 2 effects based on features of App 1 and App 3**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.201 | 0.259 | 0.256 |
| Adjusted R Sq | 0.155 | 0.217 | 0.214 |
| Statistical Significance | Featuring of App 1 | Featuring of App 1 and Price of App 3 | Featuring of App 1 and Price of App 3 |
| Interpretation | Not reliable R sq values | | |

**App 3 effects based on features of App 1 and App 2**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.009 | 0.129 | 0.131 |
| Adjusted R Sq | -0.046 | 0.079 | 0.081 |
| Statistical Significance | Nothing **sig** | Nothing **sig** | Nothing **sig** |
| Interpretation | Nothing is reliable or statistically significant | | |

From all the above analysis, we can see that none of the values are significant. The R squared value falls under 0.2 for all the apps. Thus the marketing strategies of competing apps alone does not affect the sales of any app. Hence we would prefer to analyze in depth including its own price and features.

*Question 3: Propose the “best” regression model for each app, taking into account own-effects and competitive-effects. Comment on the quality of these final models.*  
**App 1 effects based on features of its own and competitive apps(App 1 and App 2)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.891 | 0.847 | 0.848 |
| Adjusted R Sq | 0.881 | 0.834 | 0.834 |
| Statistical Significance | The price of the 2nd App affects the sales of first App | The price of the 2nd App affects the sales of first App | The price of the 2nd App affects the sales of first App |
| Interpretation | Only the rating of the 2nd App is affecting the sales of the first app. | | |

**App 2 effects based on features of its own and competitive apps(App 1 and App 3)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.778 | 0.598 | 0.599 |
| Adjusted R Sq | 0.758 | 0.562 | 0.564 |
| Statistical Significance | Price and Featuring of App 1 | Price and Featuring of App 1 | Price and Featuring of App 1 |
| Interpretation | Reliable R sq values for Linear Regression | | |

**App 3 effects based on features of its own and competitive apps(App 1 and App 2)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Linear Regression | Semi Log Regression | Log Regression |
| Multiple R Sq | 0.489 | 0.569 | 0.570 |
| Adjusted R Sq | 0.445 | 0.532 | 0.532 |
| Statistical Significance | No **sig**  of other features | Price of App 1 | Price of App 1 |
| Interpretation | Reliable R sq values for Linear Regression | | |

**Answer 3:**

For the first App, the best regression model will be a **SEMILOG** model. This is because the R sq and adjusted R sq values for all the three models are reliable and comparable, but SemiLog and Log explain more statistically significant values affecting the sales of App 1 that is the **price of App 2**. Hence we go for the simpler choice, the semi log model.

For the second App, the best regression model will be a **Linear Regression** model. This is because the R sq and adjusted R sq values for linear regression are much more than semilog or log regression. All models explain the same statistically significant values affecting the sales of App 2 that is the price of App 1 and Featuring of App 1 on homepage. Hence, we use the simpler more reliable model, Linear Regression.

For the Third App, the best regression model will be the **SemiLog model.** This is because the R sq and Adjusted R sq values of SemiLog model are better and they explain the same amount of statistical significance, that is price of App 1 affecting the sales of App 3.

Interpretation of Coeff:

For App 1, in the semi log regression model, one unit increase in the price of App 2 results in 0.286 increase in the Sales Units of App 1.

For App 2, for the linear regression model, for unit increase in price of app 1 results in 2.23 increase in sales of App 1 and for featuring of app 1, there is -0.028 decrease in sales of App 2.

For App 3, for semi log regression, for unit increase in price of App 1 there is a large 6.923 increase in sales of App 3.