**INTERNET MARKETING PROJECT (110 POINTS)**

1. (50 Points) Please add four additional metrics into your data using the formula provided

a. Net Revenue (Amount (total revenue) – Total Cost))

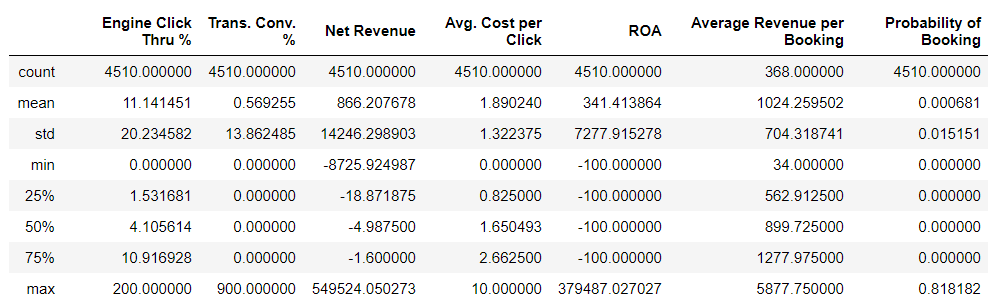
b. Return on Ad $ Spent (ROA) (Net Revenue / Total Cost) (Note: Set this variable as percentage; if Total Cost is 0, then set ROA as 0 for that observation.)

c. Average Revenue per Booking (Amount /Total Volume of Bookings) (Note: if Total Volume of Bookings is 0, then set null value for that observation)

**d. Probability of Booking (Engine Click Thru % (CTR) \* Trans. Conv. % (TCR) / 10000)**

**Please provide descriptive statistics (Count, Max, Min, Mean, and Std.) for variables (CTR, TCR, Net Revenue, Avg. Cost per Click, ROA, Average Revenue per Booking, Probability of Booking). Please report a summary statistics table and provide short descriptions of your observations and thoughts.**

**Descriptive Statistics**



**Observations:**

We observed that of the seven variables in the above descriptive analysis, six of them have count of 4510 rows while one of the variables, Average Revenue per Booking had only 368 rows which is attributed to the previous setting of 0 in variable to null.

**Engine Click Thru %:** This is the click through rate per campaign. With the mean of 11.14% and a max of 200%; this indicates the existence of a large range within the data and corresponding outliers. This result can be used to evaluate the performance of each campaign.

**Trans. Conv. %:** This is the rate of conversion of customers which explains the number of people who have seen the company’s advert and end up clicking on it. With the mean of 0.57% and a max of 900%; this indicates the existence of a large range within the data and corresponding outliers. This is a very important KPI for organizations.

**Net Revenue:** This is estimated from (Total Revenue – Total Cost). The maximum net revenue obtained in this dataset is $549,524.05 while the minimum is -$8,2725.92 which means there were cases where the cost was higher than the total revenue. Measures or strategies should be put in place to reduce or eliminate losses from high cost in search engine marketing campaigns.

**Avg. Cost per Click:** This is the average cost of clicks charged to companies for adverts placed on the internet. We observed a maximum of $10, a minimum of $0 and a mean of $1.89. Though the cost-per-click of search engine keywords have seen a continuous growth in recent time especially with the growth of new players in the market, nevertheless, there should be a strong focus on using analytics expertise in controlling cost.

**ROA:** This is the profit made on advertising campaigns which measures the effectiveness of the campaign. We observed that the maximum ROA is $379,487.2, minimum of -$100 and mean $341. This shows that there were case where no profit was recorded on a lot of campaigns, but high profit was made from a few campaigns. Resources should be made available to continuously optimize performance and ROA.

**Average Revenue per Booking:** We observed that a huge number of the Total Revenue and Total Volume of Bookings were 0 values in the dataset thereby having only a record of 368 counts. Nonetheless, the mean value is $1,024.26, a maximum of $5,877.75 and minimum of $34 which shows there were a few cases where high revenue was made from the campaign. There should be focus on ensure high revenue per campaign.

**Probability of Booking:** This is the probability of a customer/user seeing the advert online and proceed to making a booking or purchase. The maximum probability of 0.81, minimum of 0.00 and a mean of 0.000681 is not an encouraging result. Strategies need to be drawn up to increase the booking rate such as creating appealing adverts with corresponding impression which will lead high conversion rate.

**e. Please make a Histogram for any of the variables of your own interests in the data. Then report any insights you may be able draw from the charts.**

**2. (60 points) Please conduct regression analysis to study what factors influence the Total Cost. Basically, Total Cost is your dependent variable (Y) and your task is to determine what the important independent (explanatory) variables are. You should use the domain knowledge you have learnt from the case, personal experiences, and external research to guide your variable selections. You may try different set of independent variables in the data set to see which one(s) has significant results and thus support your belief (you may need to create dummy variables for some of the non-numerical variables). Please report 1) the final set of independent variables you have chosen and why you have chosen them; and 2) the estimated regression equation with simple explanations for each estimated coefficient (β) and its associated relationship (include significance, direction of the impact, magnitude of the impact, and justification of the identified relationship). (Hint: feel free to explore the data in any way you want, e.g. correlation matrix, scatter plots, etc.).**

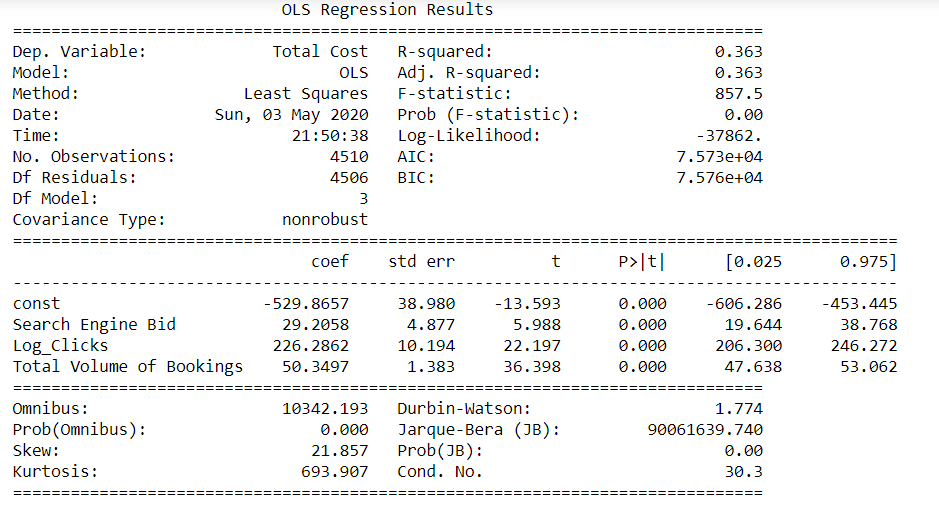
**1) The final set of independent variables you have chosen and why you have chosen them.**

The final set of independent variables we selected are:

* **Search Engine Bid:** Advertising companies bid on keywords which users or customers can enter when searching for such services on the internet, this invariably gives the advertising company a greater chance of their products/services to appear in the reach result page. This is at a cost to the company offering the products or services. In an ideal situation, the higher the cost of bid, the higher the cost to the company.
* **Log\_Clicks:** This the log transformation of the variable Clicks. Log transformation was carried out on Clicks to limit the wide range in its values. Clicks are the number of clicks made on an advert on the internet. Clicks also helps companies to understand how well their advert is appealing to their target market however for every click on an ad, the companies are charged thereby increasing the total cost of expenses of the company
* **Total Volume of Bookings:** This is number of booking made per campaign. In some cases, when the volume of transaction is high it could also cause the number of clicks on the advert to increase there by having impact on cost.

These variables were selected using our domain knowledge and the use of the correlation matrix to check for correlation between the dependent and independent variables and to avoid multicollinearity.

**2) The estimated regression equation with simple explanations for each estimated coefficient (β) and its associated relationship (include significance, direction of the impact, magnitude of the impact, and justification of the identified relationship).**



**Estimated Regression Equation:**

Total Cost = -529.87 + 29.21(Search Engine Bid) + 226.29(Log\_Clicks) + 50.35(Total Volume of Bookings)

**INTERPRETATION:**

* All the variables used for the regression analysis is statistically significant
* If the values of all the independent variables are 0, then the model predicts that Total Cost will be -$529.87
* For every $1 increase in the Search Engine Bid, the Total Cost will increase by $29.21
* For every increase in the number of Clicks, the Total Cost will increase by $226.29
* For every single increase in the number of bookings (Total Volume of Bookings), the Total Cost will increase by $50.35
* R2= 0.363, this captures how variation in Total Cost is explained by the selected independent variables
* Adj R-Square = 0.363, this also captures the fitness of the model
* F-stats = 0.00, this captures the whole model fit and suggests that a valid linear relationship exists between the dependent and independent variables.