

Chike Okwudiafor
Obehi Irekponor

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)

c. Average Revenue per Booking (Amount /Total Volume of Bookings) (

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

The Four Additional Metrics

	Probability of Booking	Average Revenue per Booking	ROA	Net Revenue
0	0.818182	975.327778	379487.027027	8775.6375
1	0.166667	1574.200000	251772.000000	1573.5750
2	0.111111	390.150000	100583.870968	389.7625
3	0.004988	832.575000	71906.491546	1662.8375
4	0.003145	935.000000	42400.000000	932.8000

Descriptive Statistics

	Engine Click Thru %	Trans. Conv. %	Net Revenue	Avg. Cost per Click	ROA	Average Revenue per Booking	Probability of Booking
count	4510.000000	4510.000000	4510.000000	4510.000000	4510.000000	368.000000	4510.000000
mean	11.141451	0.569255	866.207678	1.890240	341.413864	1024.259502	0.000681
std	20.234582	13.862485	14246.298903	1.322375	7277.915278	704.318741	0.015151
min	0.000000	0.000000	-8725.924987	0.000000	-100.000000	34.000000	0.000000
25%	1.531681	0.000000	-18.871875	0.825000	-100.000000	562.912500	0.000000
50%	4.105614	0.000000	-4.987500	1.650493	-100.000000	899.725000	0.000000
75%	10.916928	0.000000	-1.600000	2.662500	-100.000000	1277.975000	0.000000
max	200.000000	900.000000	549524.050273	10.000000	379487.027027	5877.750000	0.818182

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.

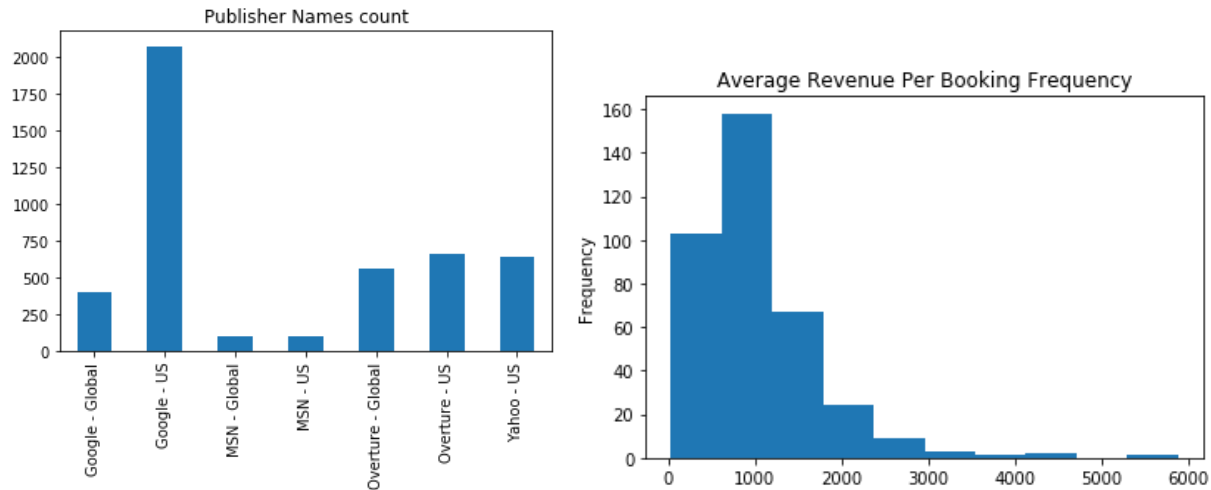
Avg. Cost per Click: This is the average cost of clicks charged to companies for adverts placed on the internet. We observed a maximum of \$10, a minimum of \$0 and a mean of \$1.89. Though the cost-per-click of search engine keywords have seen a continuous growth in recent time especially with the growth 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.



Average Revenue Per Bookings:

As seen from the charts, most of the revenue for the bookings came within these points, signaling that \$1,000 was the most frequent with about 160 booking frequency.

Publisher Names Frequency: This analysis was done to see what Publishers stood out in relation to their frequency, it showed an interesting insight and confirms that Google Us was the biggest publisher in this regard signaling that there is a chance that more ads were ran through that. It is also interesting to point that most of Air France ads were in the USA rather than globally.

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 is 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 thereby 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).

OLS Regression Results						
=====						
Dep. Variable:	Total Cost	R-squared:	0.363			
Model:	OLS	Adj. R-squared:	0.363			
Method:	Least Squares	F-statistic:	857.5			
Date:	Sun, 03 May 2020	Prob (F-statistic):	0.00			
Time:	21:50:38	Log-Likelihood:	-37862.			
No. Observations:	4510	AIC:	7.573e+04			
Df Residuals:	4506	BIC:	7.576e+04			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-529.8657	38.980	-13.593	0.000	-606.286	-453.445
Search Engine Bid	29.2058	4.877	5.988	0.000	19.644	38.768
Log_Clicks	226.2862	10.194	22.197	0.000	206.300	246.272
Total Volume of Bookings	50.3497	1.383	36.398	0.000	47.638	53.062
=====						
Omnibus:	10342.193	Durbin-Watson:	1.774			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	90061639.740			
Skew:	21.857	Prob(JB):	0.00			
Kurtosis:	693.907	Cond. No.	30.3			

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 log of Clicks, the Total Cost will increase by \$226.29. which means that we would divide the coefficient by 100. This tells us that a 1% increase in Clicks increases the dependent variable by (226/100)\$2.26. For example: For every 10% increase in number of Clicks, Total Cost increases by about \$9.35.
- For every single increase in the number of bookings (Total Volume of Bookings), the Total Cost will increase by \$50.35
- $R^2 = 0.363$, 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.

BONUS QUESTIONS

- 1) **Summarize metrics for each publisher. Please report the summary table including the variables as shown below (The answer for Google-global is already provided ☐, and please fill-in others.) In addition, discuss Key Observations and Takeaways.**

Note: You might notice that the probability figures are not the same as yours. However, it is possible to have those figures.

Publisher Name	Sum Of Net Reveue	Average Cost Per Click	Total Volume of Bookings	Average Revenue Per Booking	ROA(%)	Probabilit y of Booking	Sum of Click Charges	Cost/Booking
Google - Global	\$ 808,603.10	\$ 1.66	797	\$ 1,166.31	668.56	0.0389%	\$ 120,946.71	\$ 151.75
Google - US	\$ 1,391,841.00	\$ 1.84	1550	\$ 1,126.12	393.58	0.0626%	\$ 353,640.60	\$ 228.16
MSN - Global	\$ 133,363.90	\$ 1.08	129	\$ 1,128.09	1096.71	0.0831%	\$ 12,160.36	\$ 94.27
MSN - US	\$ 165,451.30	\$ 1.49	140	\$ 1,296.78	1027.74	0.0604%	\$ 16,098.49	\$ 114.99
Overture - Global	\$ 365,788.80	\$ 1.06	372	\$ 1,156.14	568.92	0.0084%	\$ 64,295.86	\$ 172.84
Overture - US	\$ 205,457.20	\$ 1.19	289	\$ 1,202.19	144.71	0.0025%	\$ 141,976.07	\$ 491.27
Yahoo - US	\$ 836,091.10	\$ 1.01	662	\$ 1,332.76	1809.81	0.2937%	\$ 46,197.82	\$ 69.79
Grand Total	3906596.4	9.332403	3939	8408.40306	815.717804	0.005496	755315.9219	1323.053525

Observations

- According to ROA section, the highest ROA is from Yahoo-US. Also, it seems to have an incredibly low Cost Per Click which indicates a good value for the marketing cost made.

In addition, Yahoo performs very well across all the other platform reach the third highest volume of bookings and the highest average revenue of booking. Therefore, Yahoo-US produces the best marketing campaign results and should be looked at for a better fit for Air France.

- It is also worth to point that Google-US has the highest Sum of Net revenue, however their CPC is also high, and their ROA is low. There needs to more investigation done to figure how Google -US can achieve such a high Net Revenue. In addition, Air France should try to find a way to reduce the Cost per Click while keeping the Net revenue high.

2) Based on the one-week summary data provided for Kayak in “kayak” sheet of the excel file, please calculate the following metrics and clearly show your calculation process.

a. Kayak Trans. Conv. Rate

b. Average Publisher TCR

c. Kayak CPC

d. Average Publisher CPC

Compare the calculations with what you have derived from the Bonus question #1, what recommendation you would like to make about marketing in Kayak relative to other publishers?

Average Publisher TCR: 0.5692550746242099

	Kayak Trans. Conv. Rate	Kayak CPC
0	7.326523	1.256475

Average Publisher CPC: 1.8902395796868083

Answer: The average Transaction conversion Rate indicates the average of how many clicks converted into transactions. A higher conversion rate would mean higher revenue and higher customer relevance. It seems the weekly Kayak Conv. Rate is way better than the average publisher TCR. Also, the Cost Per Click for Kayak is lower than the Average Publisher CPC signifying a lower spend on the ad click. We recommend Kayak should be considered very seriously as a publisher as every indication shows they perform way better than the other publishers.