INTERNET MARKETING PROJECT (110 POINTS)

DUE DATE LISTED ON CANVAS COURSE SITE

- This project will analyze the data that comes with the Air France case, interpret the results, and make recommendation from the Case: **Air France Internet Marketing: Optimizing Google, Yahoo!, MSN, and Kayak Sponsored Search**.

- You may need to review the case to get familiar with the context.

- **Please follow the instructions very carefully to do this project!** Please do the following analyses using the data in the "DoubleClick" sheet of the excel file and answer the corresponding questions.

Required submissions are **Different** from prior project's

You will submit **two** files: one WORD document as a detailed report, and a Jupyter notebook as your source codes. No PPT or oral presentation will be required.

Report in WORD doc: Please copy/summarize your _key_ results for each question to a word file along with your answers to produce the final report. Don't copy the codes. You can also reformat the outputs from Python in this report. Please pay attention to your writing. Proof read your work before submission.

Jupyter notebook: all the codes used to get the analysis results should be included with code outputs displayed.

1. 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.**

    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). You should also check the multicollinearity issue for a legitimate behavioral analytics model. 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. but keep in mind our priority #1: does the IV possibly influence DV?)

Bonus questions

1. (**5 points**) 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*.**

Publisher Name	Sum of Net Revenue	Average Cost Per Click	Total Volume of Bookings	Average Revenue Per Booking	ROA	Probability of Booking	Sum of Click Charges	Cost/Booking
Google - Global	$808,603	$1.66	797	$1,166	669%	0.0441%	$120,947	$151.75
Google - US								
MSN - Global								
MSN - US								
Overture - Global								
Overture - US								
Yahoo - US								
Grand Total								

Hint: you may need to take more than one step to get some of those measures.

2. (**15 points**) 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?