GROUP (UP TO 2 PEOPLE) WEB ANALYTICS ASSIGNMENT #1 (50 POINTS)

DUE DATE: Wednesday, 2/17/2016

- This assignment will analyze the data (AirFrance.xls), interpret the results, and make recommendation from the Case: **Air France Internet Marketing: Optimizing Google, Yahoo!, MSN, and Kayak Sponsored Search** (KEL 319). **Please read the case first and follow the instructions in the following very carefully to do this assignment!**

- Please do the following analyses using the data in the "DoubleClick" sheet of the excel file and answer the corresponding questions. Please copy/summarize your ***key*** results for each question to a word file along with your answers to produce the final report.

1. (**8 Points**) Please add four additional metrics into your data using the formula provided
   - Net Revenue (Amount (total revenue) – Total Cost))
   - 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.*)
   - Average Revenue per Booking (Amount /Total Volume of Bookings) (*Note: if Total Volume of Bookings is 0, then set "NA" for that observation*)
   - 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, and. **Please report a summary statistics table and provide short descriptions of your observations and thoughts.**

   - Please make a Scatter Plot (with Trend line) and 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. a) (**10 points**) Use pivotTables to 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.*) **Please report your summary excel table and 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: use the "sum" function in pivotTable to get the Sum of Net Revenue, Sum of Total Volume of Bookings, Sum of Click Charges, Sum of Clicks, Sum of Amount, and Sum of Impressions. Then use those variables to further calculate Average Cost Per Click, Average Revenue Per booking, and Probability of Booking. **Or** you can use the ***calculated fields*** (more straightforward) as shown in the following tutorial http://www.excel-easy.com/examples/calculated-field-item.html (This is also included in the "Analytics Tutorial Part II").

b) (**10 points**) Based on your results in a), graph publishers on a bubble chart using the following dimensions: X=Probability of Booking, Y=Avg. Cost Per Click, Bubble Size=Total Costs or Total Funding (Sum of Click Charges).

Base on your bubble chart, can you categorize different publishers into four different types 1) for search engines with High probability of booking and Low CPC; (2) for search engines with Low probability of booking and Low CPC; (3) for search engines with High probability of booking and High CPC, and (4) for search engines with Low probability of booking and High CPC. **Please summarize and report your recommendations for each different type (keep your answers brief).**

3. (**5 points**) Use the following pivot tables to study tactics of campaigns with a high ROA for Google U.S., Based on this table, please answer what are the characteristics of best campaigns such as campaign category, keyword combination, match type, bid strategy, etc.? And make your own recommendations.

Bid Strategy Data to improve campaigns within high CPC publishers

Most common characteristics of these high ROA campaigns should be considered for improving the performance of other campaigns

TOP 10 Campaigns

Publisher Name	Google - US

Average of Return on Ad Dollar Spe					
Campaign	**Keyword**	Match Type	Avg. Pos.	Bid Strategy	Total
Geo Targeted San Francisco	paris cheap airline	Broad	1.00	Position 5-10 Bid St	32237%
Air France Branded	air france us	Broad	1.02		23081%
Geo Targeted New York	france airline ticket	Broad	1.54	Position 5-10 Bid St	18307%
Geo Targeted Miami	france airfare sale	Broad	1.00	Position 5-10 Bid St	16915%
Geo Targeted DC	france flights	Broad	1.20	Position 5-10 Bid St	15214%
Geo Targeted Detroit	international airfares	Broad	3.01	Position 5-10 Bid St	13899%
Geo Targeted Boston	paris cheap ticket	Broad	1.83	Position 5-10 Bid St	13070%
Geo Targeted Houston	paris cheap flights	Broad	1.32	Position 5-10 Bid St	12895%
Google_Yearlong 2006	rabat flights	Broad	1.14	Postiion 1-4 Bid Stra	9864%
Geo Targeted Philadelphia	paris flight	Broad	1.71	Position 5-10 Bid St	8787%
Geo Targeted Chicago	paris ticket	Broad	1.26	Position 5-10 Bid St	6290%
Geo Targeted Seattle	paris tickets	Broad	1.69	Position 2-5 Bid Stra	4124%
Geo Targeted Los Angeles	france air flight	Broad	1.14	Position 5-10 Bid St	3793%
French Destinations	air france to nice	Broad	1.08	Position 2-5 Bid Stra	2312%
Paris & France Terms	air france tickets paris	Broad	1.01	Position 2-5 Bid Stra	1467%

4. (**10 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 try different set of independent variables in the data set to see which one(s) has significant results (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). (***Hint:*** you may start with doing pair-wise correlations between Y and other variables to see what variables are significantly correlated with Y.)

5. (**7 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.
   - Kayak Trans. Conv. Rate
   - Average Publisher TCR
   - Kayak CPC
   - Average Publisher CPC

   Compare the calculations with what you have derived from Q2, what recommendation you would like to make about marketing in Kayak relative to other publishers?

6. (**3 points**) **Bonus question** ☺ Can you generate similar pivot tables as shown in Q3 for MSN US to study tactics of campaigns with a high ROA and also identify what are the characteristics of best campaigns for MSN US?