Air France Internet Marketing: Optimizing Google, Yahoo!, MSN, and Kayak Sponsored Search

ISYS 622 – Project 3

Written by: Tanner Hefflefinger and Morgan Kaiser

05 May 2020

Project 3

Part 1.1 – Descriptive Stats

*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.*

**Engine Click Thru % (CTR)**

* On average, 11% of users who saw the ad clicked on the ad

**TCR**

* has a strong positive relationship with ROA

**Net Revenue**

* Has an extremely wide set of values, ranging in dollars from -8726 to 549,524.
* The average is around $866.

**Avg. Cost per Click**

* The average cost per click = 1.89 dollars
* Average spending of 0.83 dollars or more per click accounts for 75% of the data

**Return on Ad Dollar Spent (ROA)**

* For one publisher, the return on spending money for an advertisement amounted to $379,487

**Average Revenue per Booking**

* 25% of Average Revenue per Booking is $1,278 and up

**Probability of Booking**

* Does not exceed 81.2%
* strong, positive relationship with TCR and ROA

Part 1.2 – Histograms

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated*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.*

Part 2 – Regression Analysis

*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).*

Originally, we selected Clicks, Impressions, Amount, Net Revenue, and Total Volume of Bookings as our IVs due to high correlation with the DV, Total Cost. Here were the OLS results:

A close up of text on a white surface

Description automatically generated

We received errors regarding strong multicollinearity, thus, we wanted to see if we could decrease its magnitude by calculating the variance inflation factor (VIF) for the numerical variables. By combining research from statsmodels and Investopedia, we decided to drop the variables with VIF > 5 then re-run the OLS model with new IVs. This comes from Morgan’s previous exploration with multicollinearity. To avoid self-plagiarism, the code source was included in the attached Jupyter Notebook, via email to Dr. Zhou, and here (<https://github.com/mokaiser/breast-cancer-diagnosis-a1/blob/master/multicollinearity.ipynb>). The following is a dataframe of VIF calculations for A screenshot of a cell phone

Description automatically generatedselected variables:

It seems that **Clicks, Click Charges, Trans. Conv. %, Amount, Total Volume of Bookings, and Net Revenue are above the threshold (VIF > 5)**. This means we should **drop these columns** as they will cause a negative effect on our OLS regression model.

**Search Engine Bid, Avg. Cost per Click, Impressions, Engine Click Thru %, Avg. Pos., Total Cost/Trans, ROA, and Probability of Booking** were our new IVs. The following is our results **after re-running the OLS model** with these new IVs:

A close up of text on a white background

Description automatically generated

**Regression Results**

*Although the coefficients are still extremely large, we have significantly reduced the magnitude error associated with multicollinearity. Please also note we are only including variables with P>|t| = 0.05 or less.*

**Equation**

Total\_Cost = -123.90 + 40.92SearchEngineBid + 0.0030Impressions + 1.50TotalCostPerTrans

**Variable Breakdown**

**Search Engine Bid**

* For every 1 unit increase of Total\_Cost, Search Engine Bid would increase by ~40 units
* P < 0.05, statistically significant

**Avg. Cost per Click**

* For every 1 unit increase of Total\_Cost, Avg Cost per Click would decrease by ~4 units
* P > 0.05, NOT statistically significant

**Impressions**

* For every 1 unit increase of Total\_Cost, Impressions would increase by ~0.0030 units
* P < 0.05, statistically significant

**Engine Click Thru %**

* For every 1 unit increase of Total\_Cost, Engine Click Thru % would decrease by ~1 unit
* P > 0.05, NOT statistically significant

**Avg. Pos.**

* For every 1 unit increase of Total\_Cost, Avg. Pos would increase by ~11 units
* P > 0.05, NOT statistically significant

**Total Cost/Trans**

* For every 1 unit increase of Total\_Cost, Total Cost/Trans would increase by ~1 unit
* P < 0.05, statistically significant

**ROA**

* For every 1 unit increase of Total\_Cost, ROA would increase by ~0.001 unit
* P > 0.05, NOT statistically significant

**Probability of Booking**

* For every 1 unit increase of Total\_Cost, Probability of Booking would decrease by ~657 units
* P > 0.05, NOT statistically significant

Part 3.1 – Bonus #1

*Summarize metrics for each publisher. Please report the summary table including the variables as shown below. In addition, discuss Key Observations and Takeaways.*

A screenshot of a cell phone

Description automatically generated

* Overture has the lowest probability of booking for users based in the United States.
* Google has the highest click charges domestically and internationally.
* Overture had the highest overall click charges per booking in the US, over Google and MSN

Part 3.2 – Bonus #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: Kayak Trans. Conv. Rate, Average Publisher TCR, Kayak CPC, and Average Publisher CPC. Comparing 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?*

* The Transaction Conversion Rate is 7.33%
* Publisher TCR
  + MSN Global - 1.15%
  + MSN US - 1.30%
  + Overture US - 0.24%
  + Google Global - 1.1%
  + Yahoo US - 1.45%
  + Google US - 0.81%
* Kayak paid approximately $.80 per click
* Publisher CPC
  + MSN Global – $0.92 per click
  + MSN US - $0.67 per click
  + Overture US - $0.84 per click
  + Overture Global - $0.95 per click
  + Google Global - $0.60 per click
  + Yahoo US - $0.99 per click
  + Google US - $0.54 per click

**Marketing Recommendations**: Kayak is doing much better than other publishers with click conversion percentage, have a 7.8% conversion rate compared to the closest publisher, Yahoo, at 1.45%. For Cost Per Click Kayak is in the middle of the pack at $0.80 per click, the publishers beating them include MSN US, Google Global and Google US. MSN and Google likely have a lower CPC due to shear amount of clicks over Kayak. They each have substantial platforms that dwarf Kayak so a better strategy for Kayak to lower their own CPC would likely be to either expand the platform as a whole or reach out to site with greater traffic for advertising purposes.