Anneil Basnandan

Data Science – NYC DAT 38

Part 2: Project Design Write-up

**Project Title**

Predict Reservations Resulting from Amount Spent on Paid Search

**Project Problem and Hypothesis**

Null Hypothesis: There is not a relationship between the amount spent on Paid Search and the number of web reservations.

Follow up Question: If there is a relationship, how much does each amount of money spent on Paid Search convert to web reservations?

I think that there is a relationship between amount spent on Paid Search and the Number of Reservations. I want to use past spending and conversions to predict future conversions (a continuous variable). I will perform a regression.

**Dataset**

Fields:

* Month
* Number of Sessions
* Potential Lifetime Revenue
* Number of Reservations
* Conversion Rate
* Per Visit Value
* Cost of Paid Search
* Number of Clicks
* Click through Rate
* Cost per Click
* Revenue per Click
* Return on Advertising Spending
* End of Month Occupancy

**Domain Knowledge**

I use Google Analytics periodically to determine the number of reservations generated by Paid Search. My boss has had past analysts perform a similar analysis. They found a logarithm regression model, but the r-squared value is 0.55. He would like to know if there is a better model since we have more data. According to that model, $80,000 per month is the ideal amount to spend on paid search. In the beginning of 2016, we have spent more than $100,000 each month on paid search; my boss thinks that the old model is not working as it should.

**Project Concerns**

* How will I manipulate this data set?
* Does it need to be de-seasonalized so that I can look at all months at one time (i.e., maybe use time series seasonal decompose in statsmodel)?
* What are the best predictors of conversions given this data set?
* I have no data on if/how paid search has changed over time and when that occurred.
* How can we keep occupancy high (more conversions) without overspending?
* It is possible that there is data corruption in Google Analytics.
* Is it possible to build a better model than what was created before? This would be considered a “success”.
* Are there other elements that should be considered in the model given the data that I have access to?