**Cruise Recommendation and Favorability Prediction**

Problem Statement and Hypothesis

The client would like to understand how does the marketing spend in different marketing funnels impacted their brand image. My original question is: How does the customer attribution impacted the brand?

By utilize machine learning, I would like to use these attribution information to predict the probability of the customers’ probability to recommend the brand and increase favorability.

The hypothesis is that the probability of favorability will reach 80% when the subjects selected and are in favor of 50% of the attribution questions (Yes/No).

Data: What, When, Where, Why, How

It is a live survey data that feeds to the online banner ads. Once the person is exposed to the ad, we will send a link for them to fill the survey up. We are going to answer the data question above.

Data Structure:

Demographic Information (Age, Income, Career, etc)

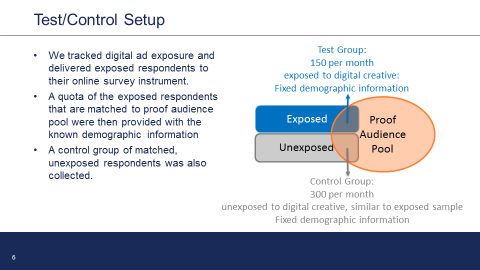
The data is mostly on a scale either 0-5 or 0-10

* Demand Generation Question (1 question)
* Attribution Questions (10 Questions)
* Overall Consideration and Recommendation Questions (5 questions)
* Competitor Benchmarking Questions
* Current Data Structure: (2454,93)

Data Range:

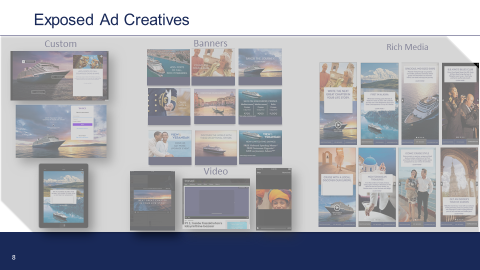
* 1/20 to Present

Below is the test and control setup, media strategy by layers, partner sites, and tech stack.





Below are the creatives involved in the testing.



* Description of any pre-processing steps you took.
  + Features have a mixture of text and number, and have already processed most of the data into the right format.
* What you learned from exploring the data, including visualizations
  + Drive the data visualization from tableau
* How you chose which features to use in your analysis
  + I will ask the question. Does the data have some strong correlation between two features. If it does, then I will investigate further whether I would use it by running a regression. Does that improve how my model perform? If it does, I will look at how big of the impact it is? Not all will be added into the model.
* Details of your modeling process, including how you selected your models and validated them
  + Initial regression analysis to select the features, and the push the features into logistic regression. Split test and train them.
* Your challenges and successes
  + Couple challenges:
    - The data needs to be matched up with the reference table, and make it meaningful.
    - Not all the data is statistically significant yet @ 90%
    - Can not change survey questions anymore.
* Possible extensions or business applications of your project
  + Clustering and classifying the cookie data.
* Conclusions and key learnings
  + Still in work.