Cruise Recommendation Prediction and Segmentation Analysis

GA – DAT2
Owen Ho
Initial Presentation

CONTENT

- Key Question and Hypothesis
- Methodology
- Data Structure
- Current & Potential Analyses

The question we set out to answer:
How does customer attribution from digital media impact the brand recommendation and favorability?

@nd who are they?

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 (binary).
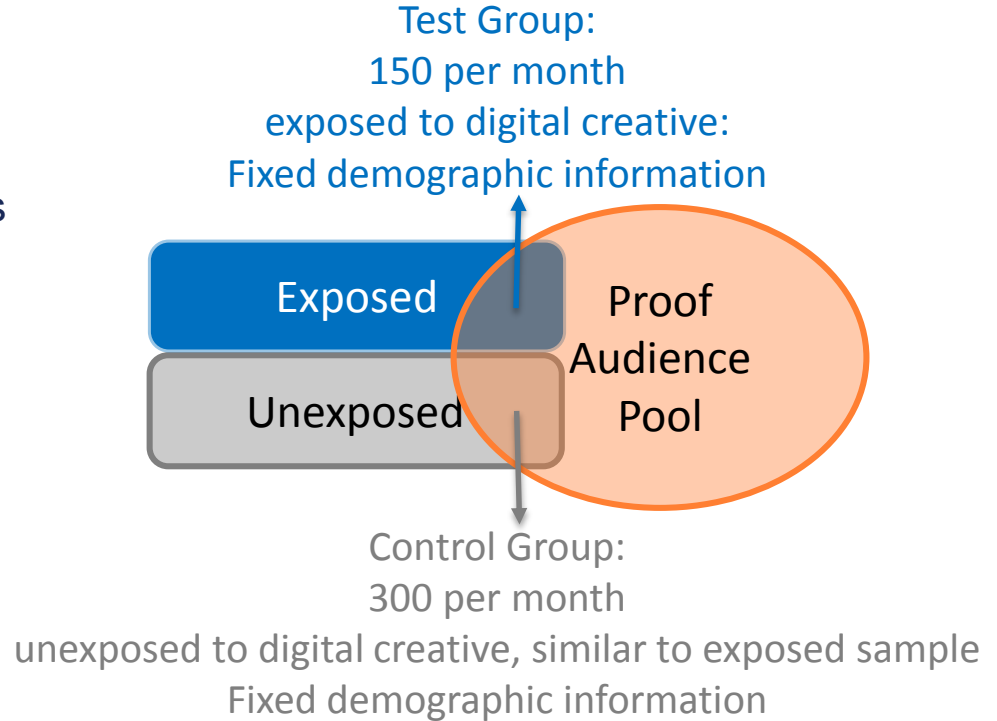
Methodology

Data Collection Journey

- How did I set it up?

- Who did I partner with?

- What did I provide to our partner and our prospects?

- What is the current data?

Test/Control Setup

- We tracked digital ad exposure and delivered exposed respondents to their online survey instrument.

- A quota of the exposed respondents that are matched to proof audience pool were then provided with the known demographic information

- A control group of matched, unexposed respondents was also collected.

Test Group:
150 per month
exposed to digital creative:
Fixed demographic information

Exposed

Unexposed

Proof Audience Pool

Control Group:
300 per month
unexposed to digital creative, similar to exposed sample
Fixed demographic information

Strategy, Sites, & Tech Stack

Awareness
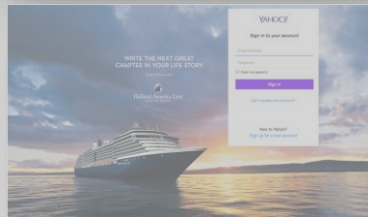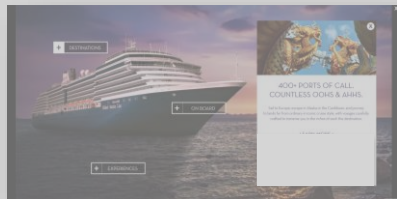- Brand Building
- Destination Awareness
- New York Times
- Food & Wine

Direct Response
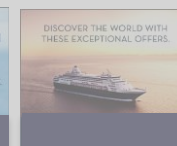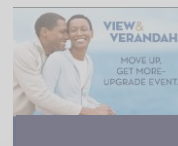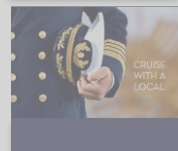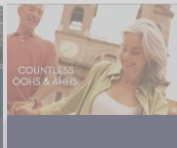- Destination Direct Response
- Promotions

Exposed Ad Creatives

Exposed Ad Creatives

Custom

Banners
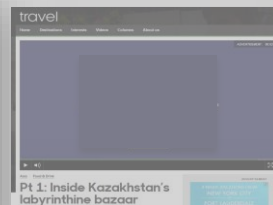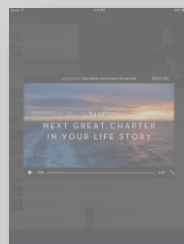
Rich Media
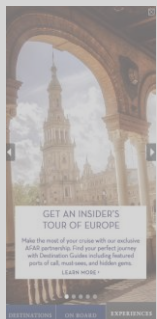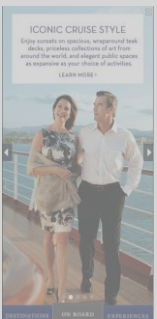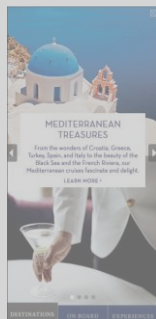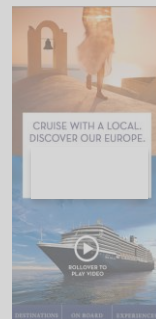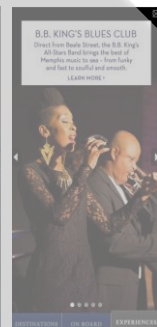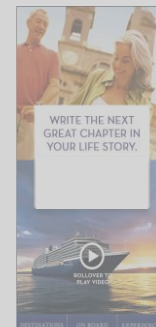
Overview of Respondents

Test Group

- Exposed to the ads
- Sample Size = 1468
- Median Age: 54
- Median HHI: $125,000 to $149,999
- 65% Female
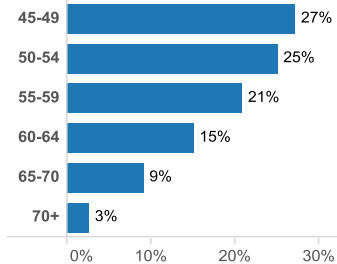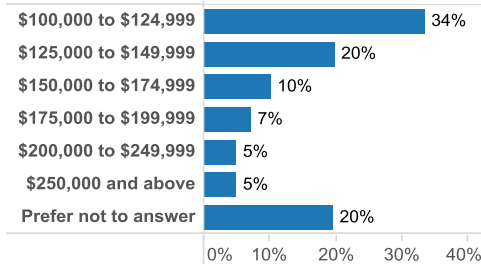
Control Group

- Not exposed to the ads
- Sample Size = 990
- Median Age: 57
- Median HHI: $125,000 to $149,999
- 59% Female

Test Group — Age

Age	%
45-49	27%
50-54	25%
55-59	21%
60-64	15%
65-70	9%
70+	3%

Test Group — HH Income

HH Income	%
$100,000 to $124,999	34%
$125,000 to $149,999	20%
$150,000 to $174,999	10%
$175,000 to $199,999	7%
$200,000 to $249,999	5%
$250,000 and above	5%
Prefer not to answer	20%

Control Group — Age

Age	%
45-49	18%
50-54	22%
55-59	21%
60-64	20%
65-70	12%
70+	6%

Control Group — HH Income

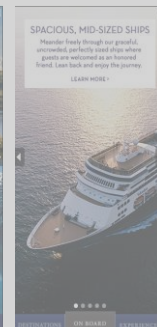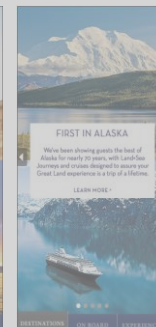| HH Income	%
$100,000 to $124,999	32%
$125,000 to $149,999	16%
$150,000 to $174,999	7%
$175,000 to $199,999	4%
$200,000 to $249,999	4%
$250,000 and above	5%
Prefer not to answer	31%

Data Structure

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)

Current State of Data

Pros:

- The data does not have a whole lot of missing data due to the force function.

- API has been setup to fetch data.

Cons:

- The data needs transformation from number scale to binary at the adequate level.

    - The trick is where is the cutoff from a scale from 0 to 10.

- 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.

Potential Analysis & Next Step

Current Analysis

- The initial analysis would be logistic regression on the possibility of recommending the brand.

  - Transform the data into binary form.

  - Run logistic regression and fit the model.

  - Perform prediction based on the new survey data.

  - Train the model again.

- The subsequent analysis can be a supervised classification among demographic and psychographic attribution.

  Balance the bias and variances, and run through the # of K.

  New survey data will be coming in on a weekly basis. I can use the new data to "classify" the potential customers.

Potential Analysis and Moving Forward

- Segmentation by Clustering
    - Normalize the data by utilizing standard scaler among demographic data and attribution data (not binary)
    - Perform K mean clustering to find out the hidden segmentation.
    - Conclude the audience in a limited group by utilizing

% of Respondents