NYU

Airbnb: Inferring Causal Effects of Marketing Ads on User Bookings

Diana Liang, Andrew Pagtakhan
Dec 4th, 2020

Agenda

- Motivation
- Data
- Estimands
- Methods
- Assumptions
- Diagnostics
- Results
- Discussion

Motivation

What is Airbnb?

Airbnb is an American vacation rental online marketplace company based in San Francisco, California, United States. Airbnb offers arrangement for lodging, primarily homestays, or tourism experiences

Logo: https://www.airbnb.com/
Source: Airbnb - Wikipedia

Research question

What is the effect of direct marketing ads vs. indirect marketing ads on user bookings?

Terminology

•*Direct marketing*: direct attempts requesting a customer to purchase a product**.** E.g., promoted or paid advertising on social media, video ads on YouTube

•*Indirect marketing*: takes an awareness approach to building customer loyalty**.** E.g., search engine optimization, sponsorships, online reviews

Motivation

Why does this matter?

- This project motivates other general questions – how much can marketing influence a potential customer's purchasing habits, and in turn general consumption behaviors and trends?
- Are there potential ethical concerns that arise?

Related research:

Airbnb: Online targeted advertising, sense of power, and consumer decisions - ScienceDirect

Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study - Daniel Guttentag, Stephen Smith, Luke Potwarka, Mark Havitz, 2018 (sagepub.com)

Data

Dataset

- User bookings data from 2010 – 2015 data with 4 main categories (see below table)
- ~106K users, 17 variables

Category	Example(s)
User Demographics	• Age • Gender • Language
Platform Identifiers	• Browser (Chrome, Firefox, etc.) • Device Type (Windows Desktop, iPhone, etc.)
Marketing Data	• Channel (direct, seo, etc.) • Provider (baidu, bing, google, etc.)
Bookings Data	• Destination country (US, CA, etc.)

Source: Airbnb New User Bookings | Kaggle

Dataset used for analysis

Due to computational limits, data was sampled to 10K users

	Category	Example(s)	Data used for analysis
Pre-treatment	User Demographics	• Age • Gender • Language	• Filtered to ages 18 – 80 • Bucketed language into English and Other
	Platform Identifiers	• Browser (Chrome, Firefox, etc.) • Device Type (Windows Desktop, iPhone, etc.)	• Grouped top 5 browsers by frequency (Chrome, IE, Firefox, Safari, Mobile Safari), bucketed rest into "Other"
Treatment	Marketing Data	• Channel (direct, seo, etc.) • Provider (baidu, bing, google, etc.)	• Binarized channel into Direct vs. Indirect marketing
Outcome	Bookings	• Destination country (US, CA, etc.)	• Binarized destination country into booked / not-booked

Data Exploration

[NYU_Causal_Airbnb - | Tableau Public]

	Overall (N=10000)
Age	
Mean (SD)	36.1 (11.0)
Median [Min, Max]	33.0 [18.0, 80.0]
Browser Group	
Chrome	3477 (34.8%)
Firefox	1576 (15.8%)
IE	773 (7.7%)
Mobile Safari	847 (8.5%)
Other	1129 (11.3%)
Safari	2198 (22.0%)
Gender	
FEMALE	5234 (52.3%)
MALE	4750 (47.5%)
OTHER	16 (0.2%)
Language (1 = English, 0 = Other)	
Mean (SD)	0.961 (0.194)
Median [Min, Max]	1.00 [0, 1.00]
Device	
Android Phone	76 (0.8%)
Android Tablet	45 (0.4%)
Desktop (Other)	67 (0.7%)
iPad	617 (6.2%)
iPhone	901 (9.0%)
Mac Desktop	4604 (46.0%)
Other/Unknown	326 (3.3%)
SmartPhone (Other)	4 (0.0%)
Windows Desktop	3360 (33.6%)

Estimands

Estimand of Interest

- We are interested in making inferences for the set of users who received the direct marketing ads
- Average Treatment Effect for the Treated (ATT)
- What is the treatment effect of this set of users as a result of direct marketing?

Estimand of Interest (Formula)

$$E[Y(1)|Z=1] = \alpha + \beta_1 Age + \beta_2 Browser + \beta_3 Gender + \beta_4 Language + \beta_5 Device + \tau$$

$$E[Y(0)|Z=1] = \alpha + \beta_1 Age + \beta_2 Browser + \beta_3 Gender + \beta_4 Language + \beta_5 Device$$

$\beta_i =$ coefficient each covariate, $\forall i$

$z =$ treatment group (1 = served direct ad, 0 = served indirect ad)

$$\boxed{\tau = E[Y(1)|Z=1] - E[Y(0)|Z=1]}$$

Average Treatment Effect on the Treated (ATT)

NYU

Methods

Propensity Score Matching

In order to get a proxy for the counterfactual set of users to compare to the treatment group, the following two steps are used:

1. **Propensity Score Modeling**: a one-number summary of the covariates
2. **Matching**: assign different weights to the control units to make them more similar to treatment units

Propensity Score Modeling

Various propensity score methods were analyzed:

1. Logistic Regression
2. Probit Regression
3. CART
4. Random Forest
5. GBM
6. BART

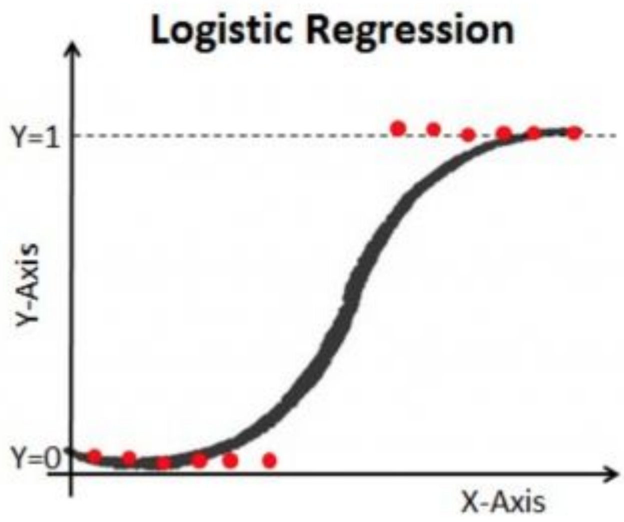
Propensity Score Modeling

Logistic Regression:

- Generalized linear model (GLM) with a logit link function

Probit Regression:

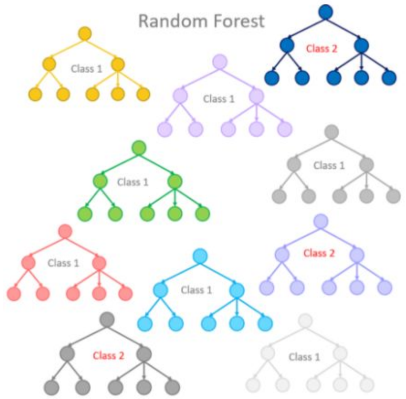
- GLM with a probit link function



Source: https://medium.com/@ODSC/logistic-regression-with-python-ede39f8573c7

CART (**C**lassification **A**nd **R**egression **T**rees):

- A single decision tree
- Easy to overfit



Source: https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147

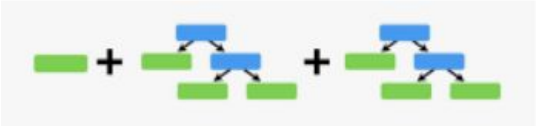
Random Forest:

- Fits a number of trees on various sub-samples of the dataset
- Uses averaging to improve the predictive accuracy and control over-fitting
- A large number of relatively uncorrelated models trees operating as a committee will outperform any of the individual constituent trees

Propensity Score Modeling

GBM (**G**radient **B**oosting **M**achine)**:**

- Builds an ensemble of shallow trees in sequence with each tree learning and improving on the previous one



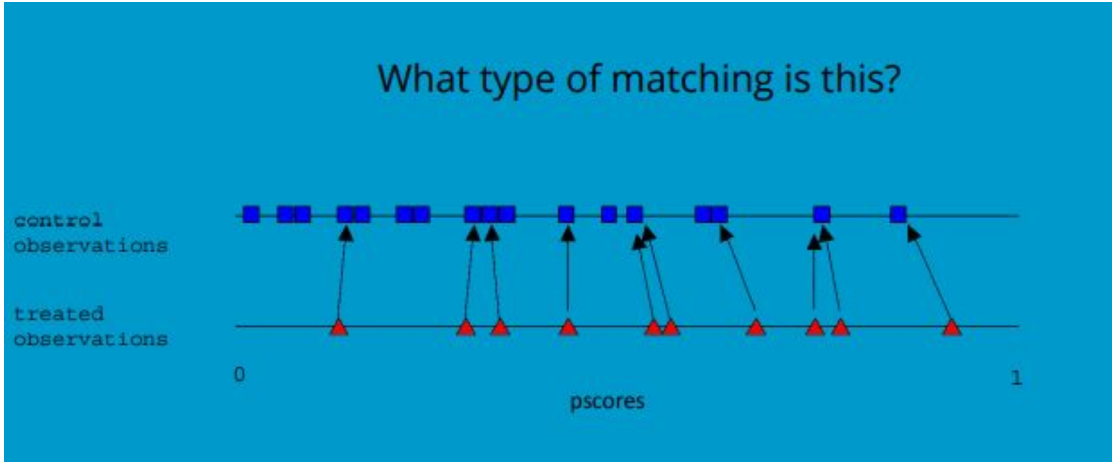
BART (**B**ayesian **A**dditive **R**egression **T**rees)**:**

- “Sum-of-trees” model where each tree is constrained by a regularization prior to be a weak learner
- Fitting and inference are accomplished via an iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior

Matching

For each of the 6 models, we applied 2 different matching schemes:

- 1-1 Nearest Neighbor Matching with Replacement
  - For each member of the inferential group we find exactly one match from among the control group
  - Control units can be reused as long as there is a good match



Source: Hill, Jennifer. NYU Causal Inference - Lecture 4.

Matching

- Inverse Probability of Treatment Weighting (IPTW)
  - Weight the sample to be representative of a pseudo-population that is unconfounded and represents the population of interest
  - The treatment group observations would be assigned weights of 1 since they are already representative of themselves
  - The control observations would be assigned (unnormed) weights equal to p/(1-p)

NYU

Assumptions

Assumption 1: Ignorability

- It requires adjustment for all confounding covariates
- Fairly plausible:
  - All user demographics are captured in this analysis
  - However there is no guarantee that all pre-treatment covariates are observed

Source: Hill, Jennifer. NYU RAOS - Ch. 20.

Assumption 2: Overlap

- Plausible
  - There is sufficient overlap between treated and non-treated users on all variables

Assumption 3: Appropriate specification of the propensity score is achieved

- Plausible
    - We use 6 different model specifications and choose the one with the best balance

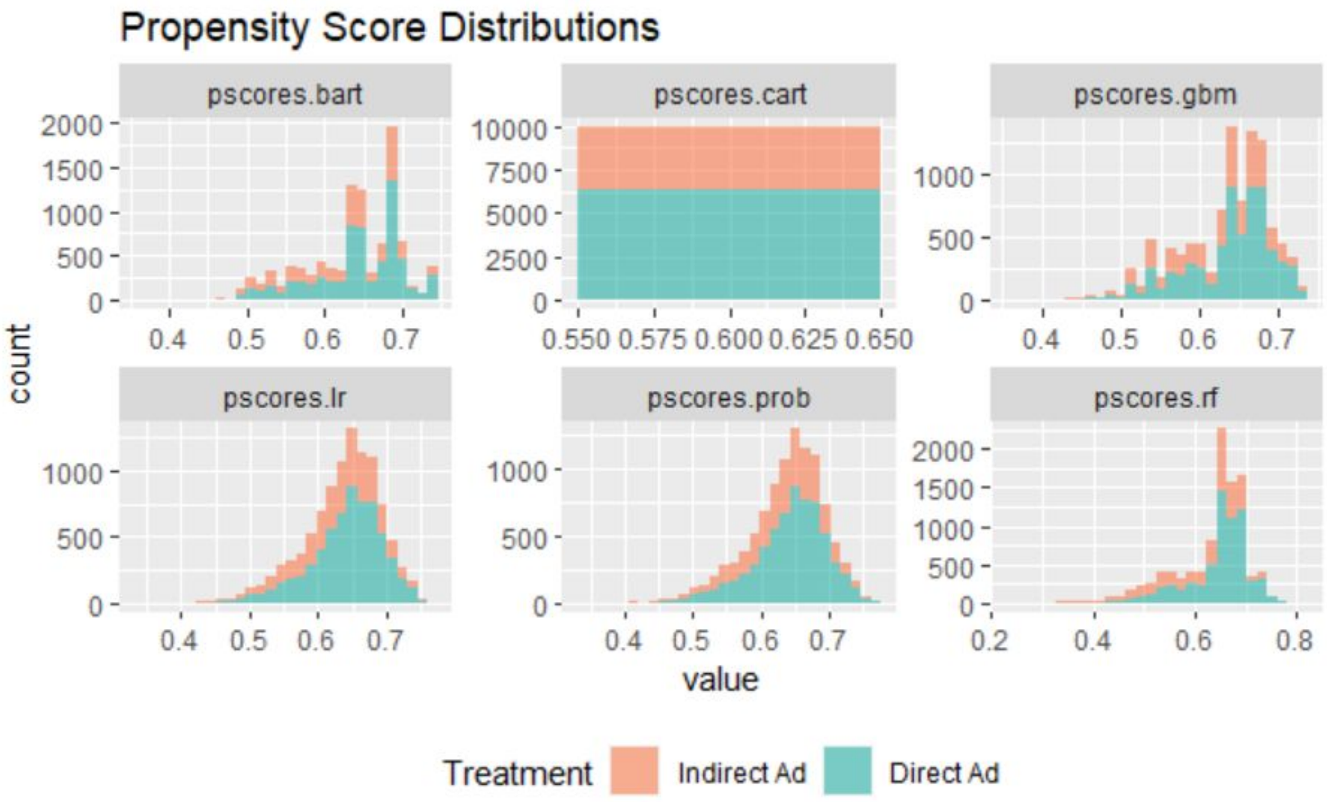
Assumption 4: SUTVA

- Fairly plausible
  - Reasonable to assume that there is no interference among users and no hidden versions of marketing ad treatment

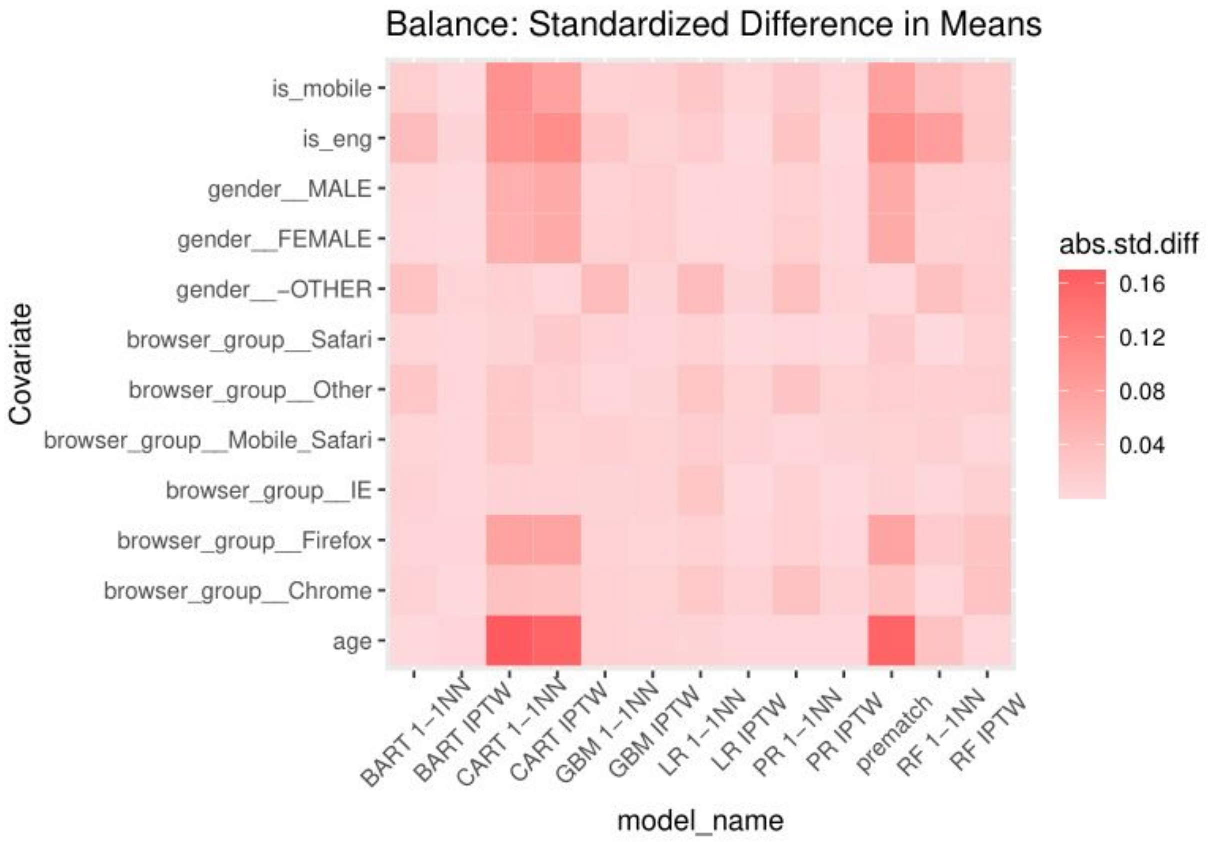
NYU

Diagnostics

Propensity Scores



Balance across all methods



Balance across all methods

Table 1: Absolute Standardized Difference in Means (across covariates

model_name	mean.bal
BART IPTW	0.0024449
PR IPTW	0.0034908
LR IPTW	0.0038071
GBM IPTW	0.0089942
BART 1-1NN	0.0123544
GBM 1-1NN	0.0129337
RF IPTW	0.0166580
PR 1-1NN	0.0167739
LR 1-1NN	0.0171875
RF 1-1NN	0.0223270
prematch	0.0525743
CART IPTW	0.0525743
CART 1-1NN	0.0548722

Table 2: SD Ratios (Absolute values, centered on 1)

model_name	sd.ratio
PR 1-1NN	0.0024309
GBM IPTW	0.0045362
GBM 1-1NN	0.0054187
BART 1-1NN	0.0076500
RF IPTW	0.0089785
LR 1-1NN	0.0096122
BART IPTW	0.0116685
LR IPTW	0.0347982
PR IPTW	0.0354177
RF 1-1NN	0.0679157
CART IPTW	0.1242490
CART 1-1NN	0.1451647
prematch	NA

Decide a PSM method

- We decide to adopt IPTW with the propensity scores estimated from BART, since it has:
    - Standardized Difference in Means closest to 0 across all covariates
    - Reasonable SD ratio of age, though not closest to 1
    - The SD ratios only include one covariate, so this metric is not as informative

Results

Estimate Causal Effects

- To estimate the causal effect, we fit a logistic regression using IPTW with the propensity scores estimated from BART
- Average probability of booking a stay when exposed to:
  - direct ad: 54%
  - indirect ad: 45%
- Interpretation:
  - The average estimated probability of booking a stay for users who received a direct ad is 19% higher than a similar group of users who received an indirect marketing ad (counterfactual)
  - Within-unit comparison: a user will be 19% more likely to book if receiving direct marketing than indirect marketing

NYU

Discussion

Limitations

- Representativeness:
  - No insight into representativeness of Airbnb’s population
  - Potential bias: most users booked in US or Canada
- Covariates:
  - Other useful characteristics that were not in the data (i.e. income, zip code, family size, purchasing habits…)
- SUTVA
  - Word-of-mouth referrals or interactions between users about Airbnb deals

Final Thoughts

- Challenging to infer true causal effects
- Quantifying effects under causal framework
- Social Impact: effects on individual decision-making?

NYU

Thank you!

For more details:

https://github.com/ajpag/Airbnb-Causal-Inference
https://public.tableau.com/profile/andrew.pagtakhan#!/vizhome/NYU_Causal_Airbnb/Dashboard?publish=yes