Case Study #3: Customer Segmentation at TripAdvisor with Recommendation AB Tests

ALICE Team[1] and Victor Lei[2]

Miruna Oprescu[1]

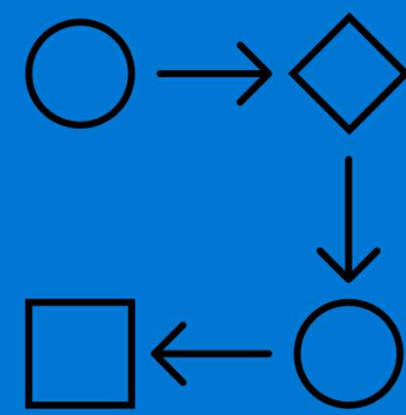
[1]MSR New England
[2]TripAdvisor

Microsoft

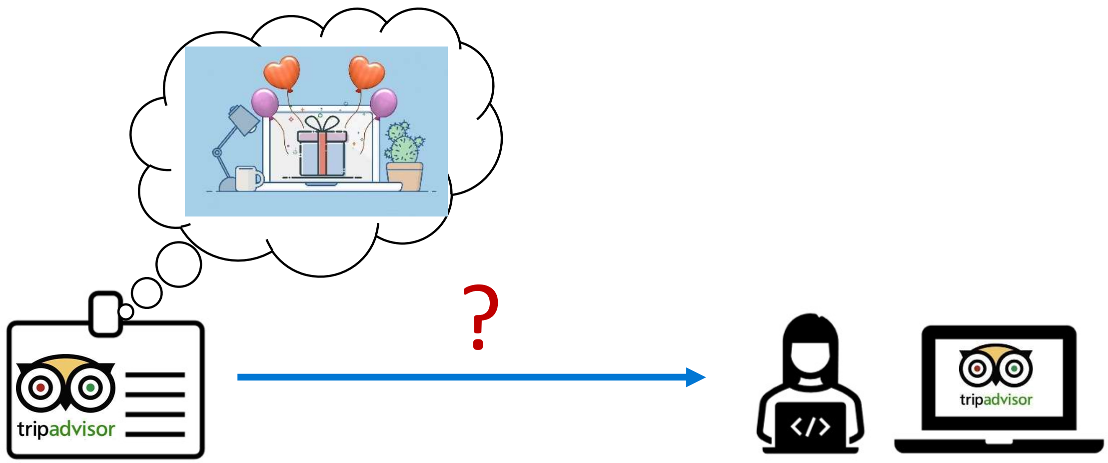
EconML

Microsoft

1. Case Study
2. Methodology
3. EconML Solution
4. Results & Takeaways



The Trip Advisor Membership Program



Trip Advisor Membership

User engagement

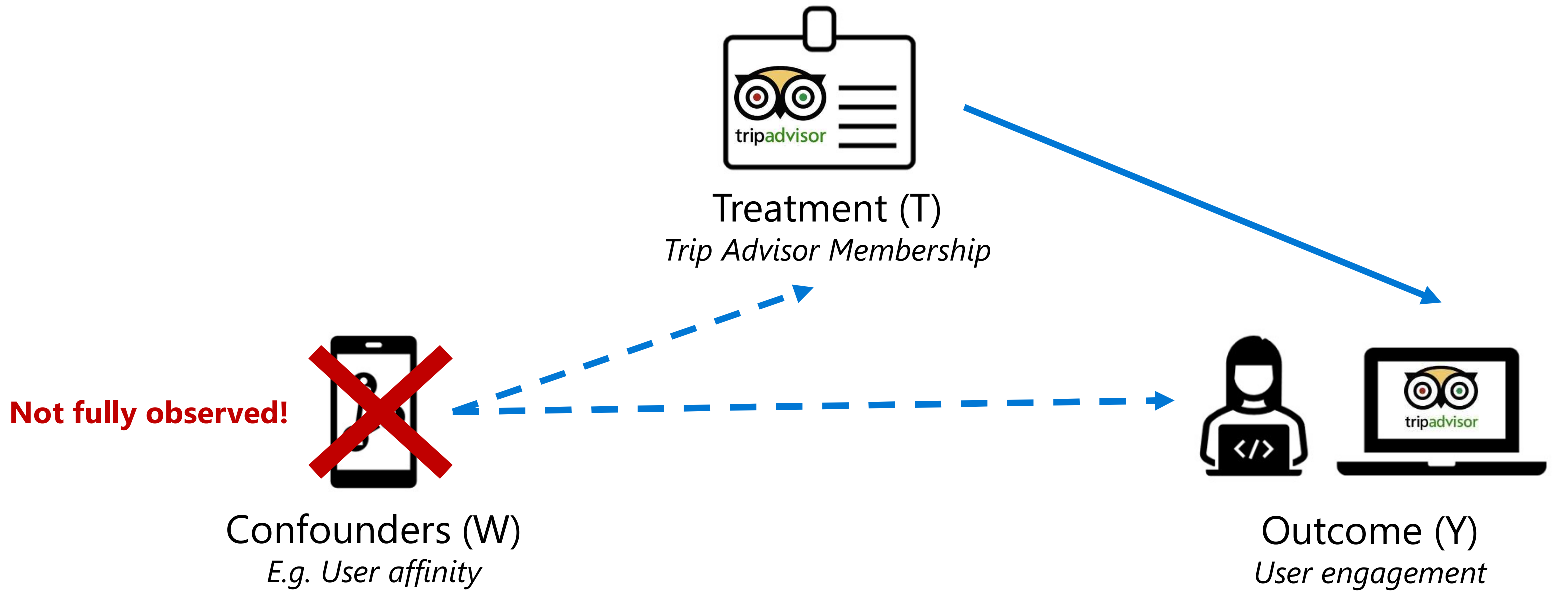
The company’s decision makers have some questions:

1. Does this program work, i.e. does it compel users to spend more time on the website?
2. For what kind of users does it work best?

The Trip Advisor Membership Program

Proposal #1: Use existing data on members and non-members

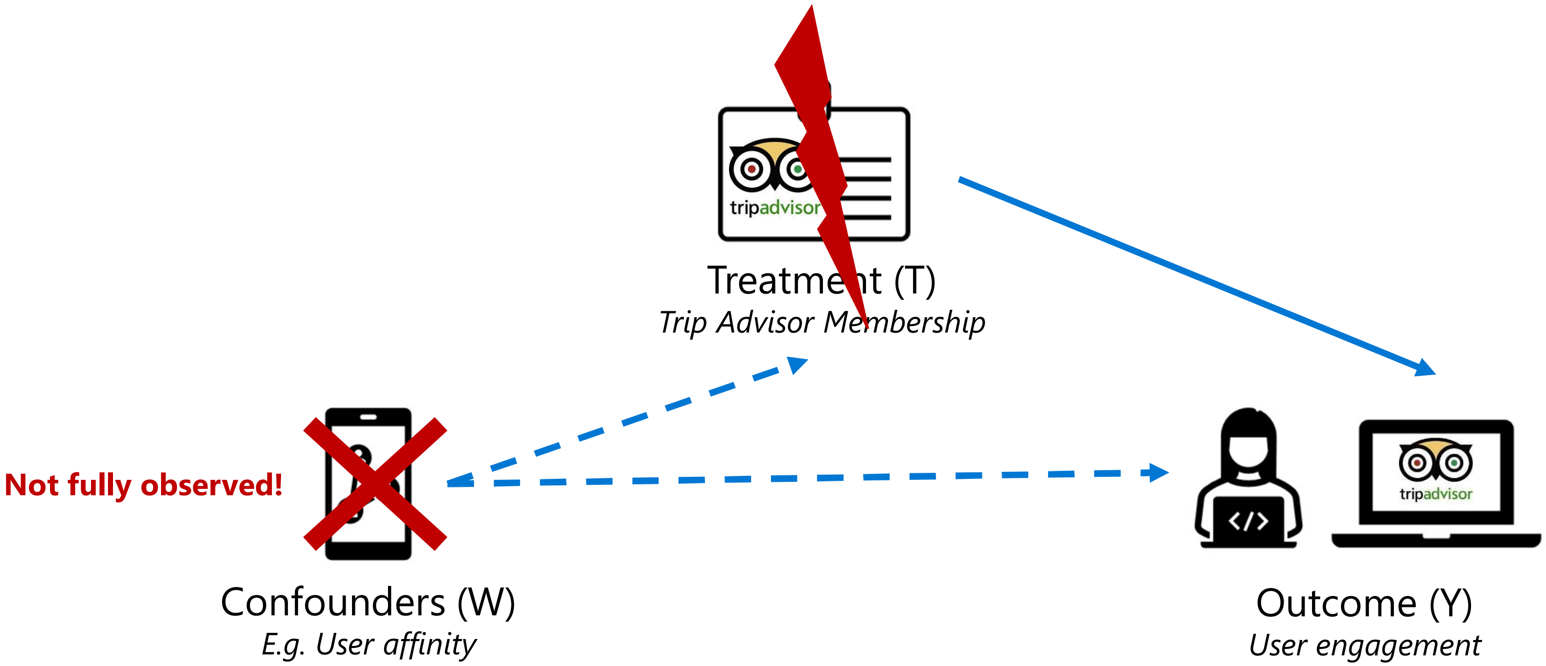
Problem: Confoundedness. *We cannot simply compare members with non-members because the customers who chose to become members are likely already more engaged than other users.*



The Trip Advisor Membership Program

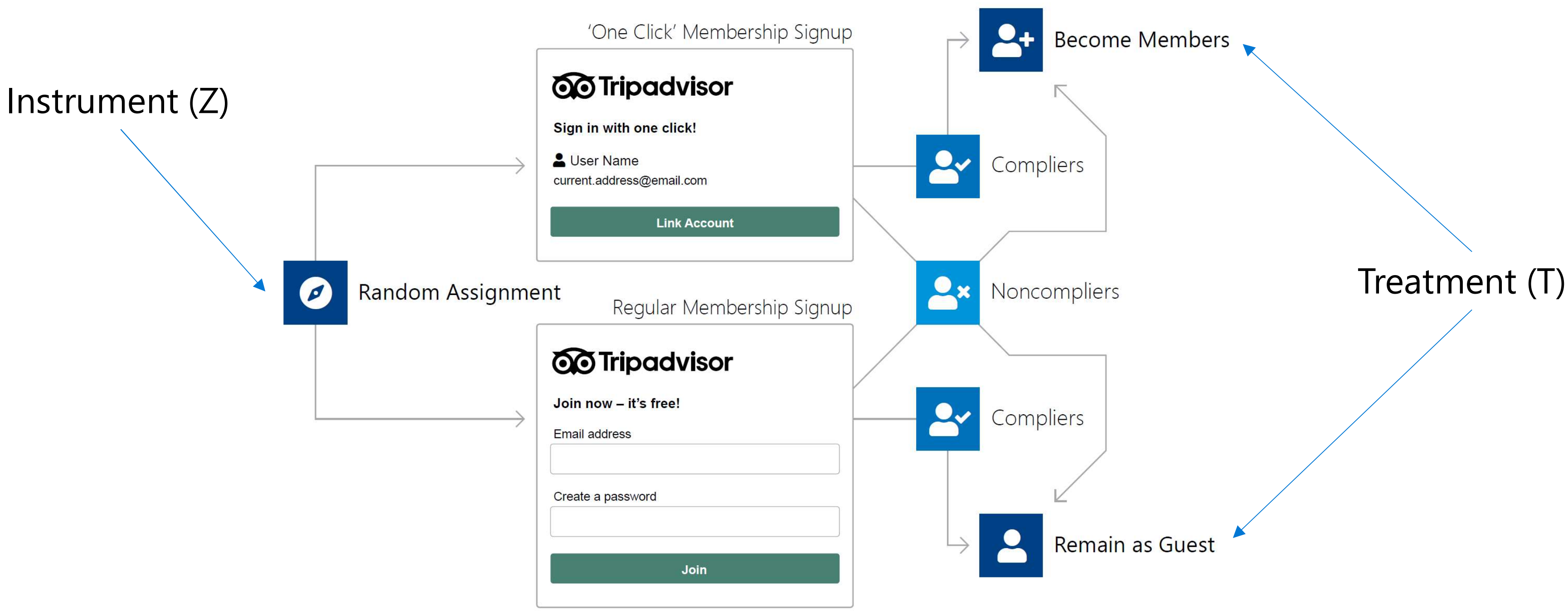
Proposal #2: Run an A/B test

Problem: Imperfect Compliance. *A direct A/B test is infeasible because the website cannot force a random subset of users to become members.*



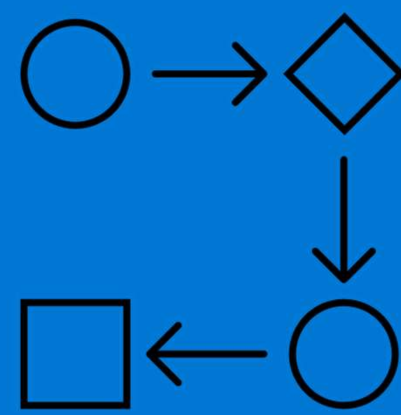
The Trip Advisor Membership Program

Proposal #3: Run a recommendation A/B test that offers signup incentives to a randomized group of users.



Trip Advisor Solution: Use existing A/B test that offered an easier sign-up flow to a subset of users.

1. Case Study
2. Methodology
3. EconML Solution
4. Results & Takeaways

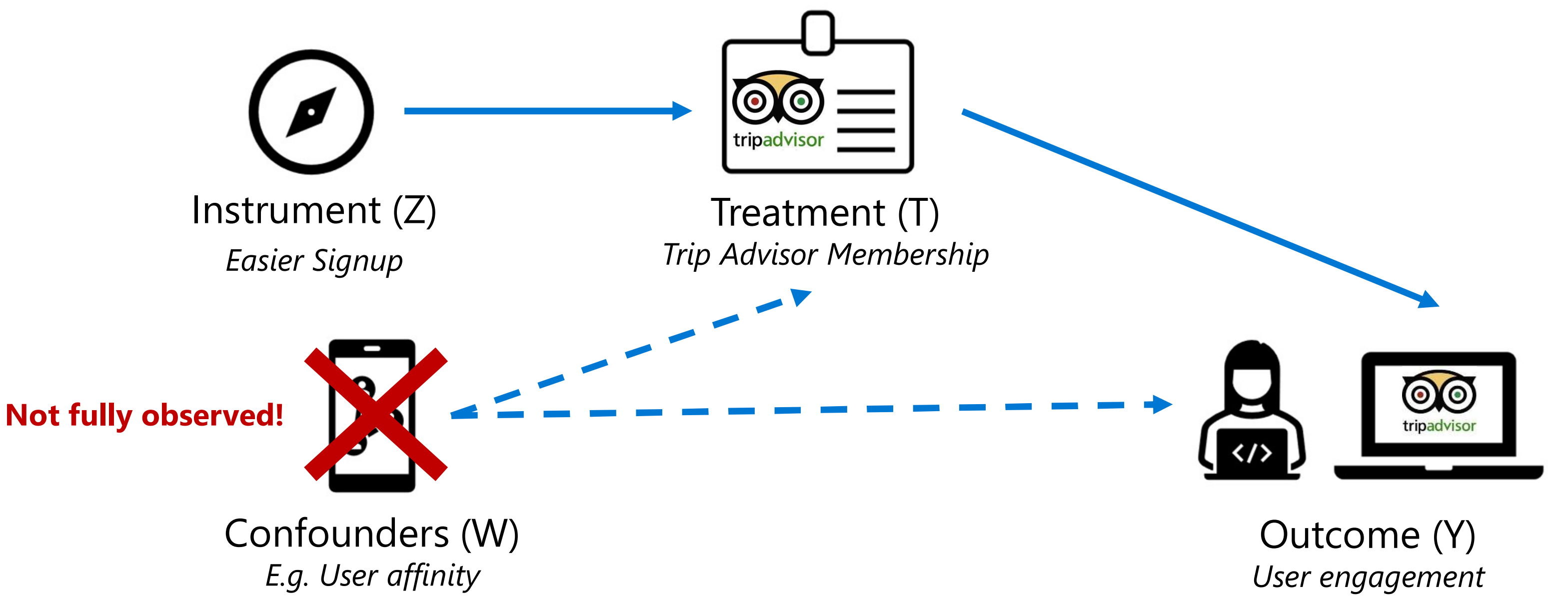


Instrumental Variables (IVs)

Instrumental Variable

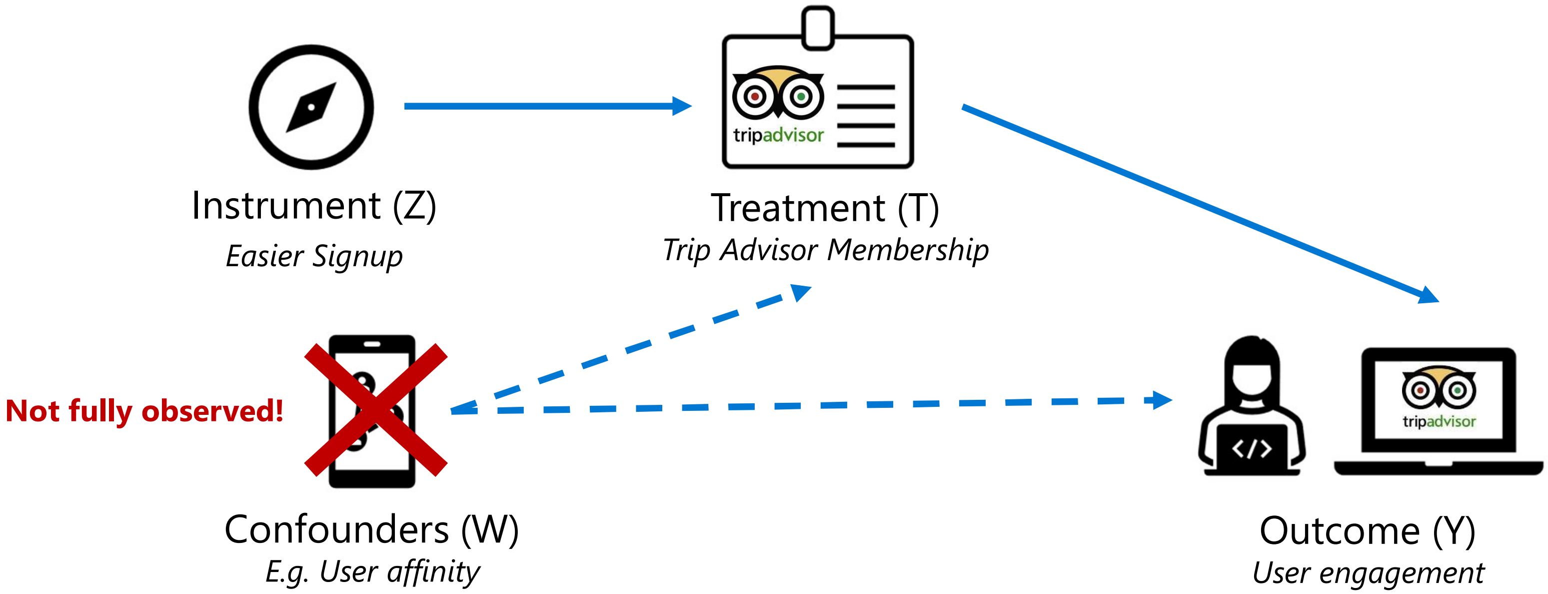
A random variable Z that affects the treatment assignment T but does not affect the outcome Y other than through the treatment

Example of IV: Cohort assignment in recommendation A/B test



Challenges and Limitations of Typical IV Methods

- They might not scale like typical ML algorithms
  - Weak instruments require large datasets for effect estimation
- They do not account for both complex effect or compliance heterogeneity
  - Not accounting for heterogeneity in compliance and effect can lead to biased average effects



Trip Advisor Experiment

For random half of 4 million users, easier sign-up flow was enabled

- Easier sign-up incentivizes membership

For each user we observe:

- $T \rightarrow$ treatment, *e.g. user membership*
- $Y \rightarrow$ outcome, *e.g. number of visits in the 14 days post experiment*
- $X \rightarrow$ features that capture heterogeneity, *e.g. user features*
- $Z \rightarrow$ instrumental variable, *e.g. assignment in A/B test*

Structural equations:

$$Y = \theta_0(X) \cdot T + f_0(X) + e$$
$$T = h_0(X, Z) + \eta$$

Doubly Robust IV (DRIV)[1]

Simplifications: $T, Z \in \{0,1\}, P(Z=1|X) = 1/2$.

Classification

1. Consider the compliance score:

$$\Delta(X) = (2Z-1)\frac{\mathbb{P}(T=1|Z=1,X) - \mathbb{P}(T=1|Z=0,X)}{2}$$

   Define the following residuals: $\tilde{Y} = Y - \mathbb{E}[Y|X]$, $\tilde{T} = T - \mathbb{E}[T|X]$.

2. Estimate preliminary $\hat{\theta}(X)$ with DMLIV:

$$\hat{\theta} = \underset{\theta(\cdot)}{\operatorname{argmin}}\, \mathbb{E}\left[\left(\tilde{Y} - \theta(X)\cdot\Delta(X)\right)^2\right]$$

Regression

3. Estimate robust final $\theta(X)$:

$$\min_{\theta(\cdot)} \mathbb{E}\left[\left(\hat{\theta}(X) + \frac{\tilde{Y} - \hat{\theta}(X)\cdot\tilde{T}}{\Delta(X)} - \theta(X)\right)^2\right]$$

[1] V. Syrgkanis, V. Lei, et al. *Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments*. NeurIPS, 2019.

Doubly Robust IV (DRIV)[1]

Simplifications: $T, Z \in \{0,1\}, P(Z=1|X) = 1/2$.

Classification

1. Consider the compliance score:

$$\Delta(X) = (2Z-1)\,\frac{\boxed{\mathbb{P}(T=1|Z=1,X)} - \boxed{\mathbb{P}(T=1|Z=0,X)}}{2}$$

Define the following resic ... $- \boxed{\mathbb{E}[T|X]}$.

2. Estimate preliminary $\hat{\theta}(X)$ with DMLIV:

$$\hat{\theta} = \underset{\theta(\cdot)}{\operatorname{argmin}}\, \mathbb{E}\boxed{\left[\left(\tilde{Y} - \theta(X)\cdot\Delta(X)\right)^2\right]}$$

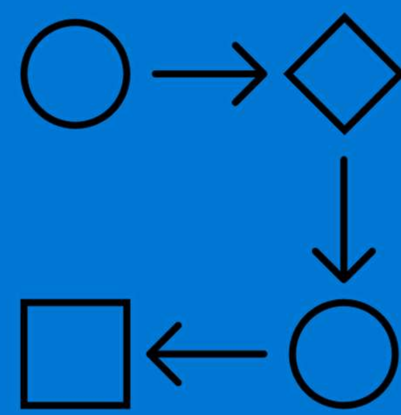
3. Estimate robust final $\theta(X)$:

Regression

$$\min_{\theta(\cdot)} \mathbb{E}\boxed{\left[\left(\hat{\theta}(X) + \frac{\tilde{Y} - \hat{\theta}(X)\cdot\tilde{T}}{\Delta(X)} - \theta(X)\right)^2\right]}$$

[1]V. Syrgkanis, V. Lei, et al. *Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments*. NeurIPS, 2019.

1. Case Study
2. Methodology
3. EconML Solution
4. Results & Takeaways



Trip Advisor Data

- Data features:

Feature Name	Type	Details
days_visited_exp_pre	X	#days a user visits the attractions pages
days_visited_free_pre	X	#days a user visits the website through free channels (e.g. domain direct)
days_visited_{fs,hs,rs,vrs}_pre	X	#days a user visits the flights/hotels/restaurants/vacation rental pages
locale_en_US	X	whether the user access the website from the US
os_type	X	user's operating system (windows, osx, other)
revenue_pre	X	how much the user spent on the website in the pre-period
easier_signup	Z	whether the user was exposed to the easier signup process
became_member	T	whether the user became a member
days_visited_post	Y	#days a user visits the website in the 28 days after the experiment

- Data sample:

	days_visited_exp_pre	days_visited_fs_pre	revenue_pre	os_type_osx	easier_signup	became_member	days_visited_post
0	1	7	0.01	0	0	0	1
1	10	27	2.26	0	0	0	15
2	18	8	0.03	0	0	0	17
3	17	23	418.77	0	0	0	6
4	24	22	1.54	0	0	0	12

DRIV @EconML

EconML

```
> pip install econml
```

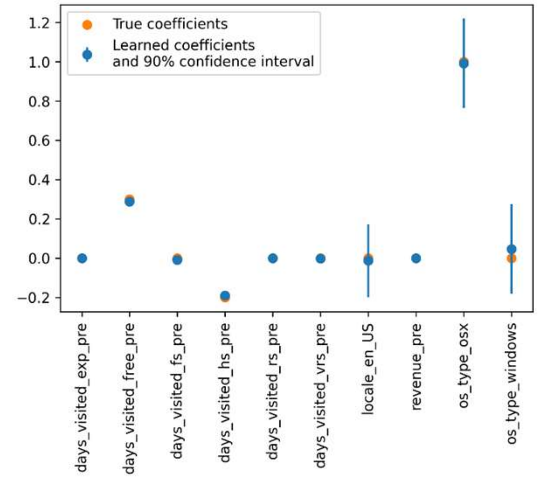
https://github.com/microsoft/EconML

DRIV In Action:

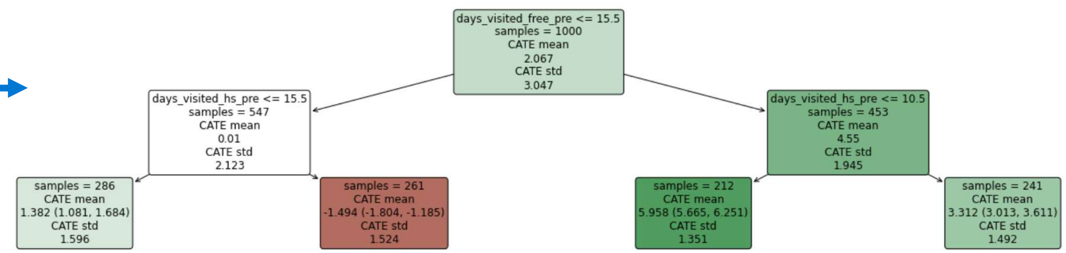
```
# Define the DRIV algorithm and nuisance functions
model = LinearIntentToTreatDRIV(
                model_Y_X = RandomForestRegressor(),
                model_T_XZ = RandomForestClassifier(),
               flexible_model_effect = RandomForestRegressor(),
              featurizer = PolynomialFeatures(degree=1, include_bias=False))
# Fit estimator and calculate treatment effects
model.fit(Y, T, Z=Z, X=X, inference="statsmodels")
te_pred = model.effect(X_test)
```

DRIV @EconML

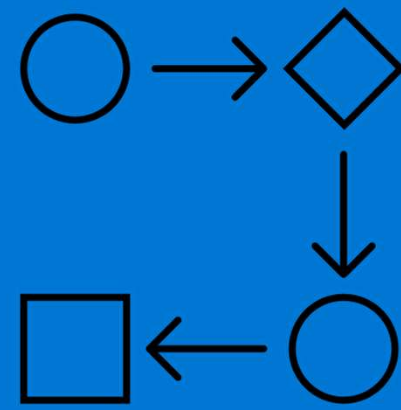
```
# Get model coefficients and confidence intervals
coefs = model.coef_
coef_error = model.coef__interval(alpha=0.05)
```



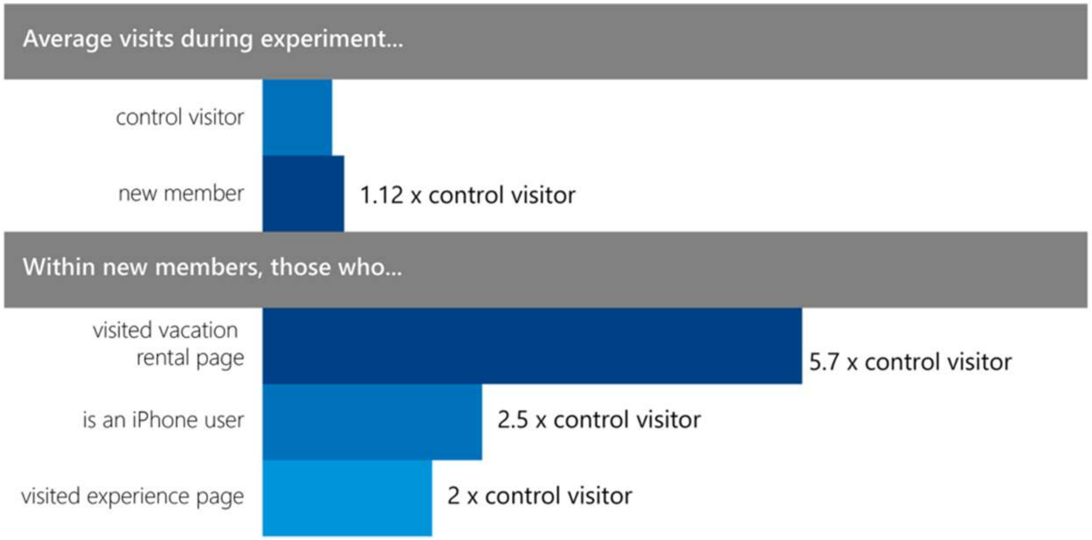

```
# Treatment effects interpreter
intrp = SingleTreeCateInterpreter(
            include_model_uncertainty=True,
            max_depth=2,
            min_samples_leaf=10)
intrp.interpret(model, test_customers)
intrp.plot(feature_names=column_names)
```



1. Case Study
2. Methodology
3. EconML Solution
4. Results & Takeaways



Results in a Nutshell



- Large heterogeneity based on which pages were recently visited
- Large heterogeneity based on platform of access (e.g. iPhone, Linux etc.)
- Results enable better targeting of the right user populations and improvements of membership offering for user segments with small effects

Key Takeaways

- **Recommendation A/B Tests** are useful when there are unobserved confounders and compliance cannot be enforced
- For Trip Advisor, these results enable **better targeting of the right user populations** and improvements of membership offering for user segments with small effects
- **EconML** provides an end-to-end solution to the intent-to-treat scenario
    - Jupyter Notebook: github.com/microsoft/EconML
    - TripAdvisor Case Study: aka.ms/econml

Thank You!

Microsoft

EconML