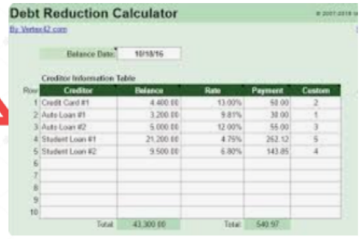
Brandenburgische
Technische Universität
Cottbus - Senftenberg

Explainable Machine Learning

Causal Inference Methods



Recapitulation Mitigation Bias

Brandenburgische Technische Universität Cottbus - Senftenberg

Mathematical Bias Mitigation Methods		
Demographic Parity	• same PR in groups • support the unprivileged group	
Equal Opportunity	• same TPR in groups • False Positives are not costly to the user nor the company	
Equalised Odds	• same TPR and FPR in groups • there remains a profit for the company	

Recapitulation Mitigation Bias

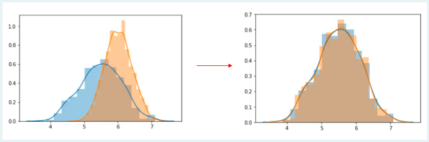
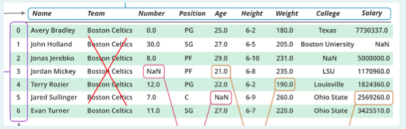
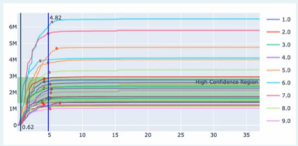
Brandenburgische Technische Universität Cottbus - Senftenberg

Bias Detection	
Identification of groups	
Metrics for each group separately	• confusion matrices, etc.

Data transformation with Imbalance Handling	• BinaryLabelDataset() for binary classification
	• RegressionDataset() for regressioparity
	  • protected_attribute_names=['AGE_GROUP'],
	  • favorable_label=0, unfavorable_label=1

Metrics for quantifying bias	• Statistical parity difference (SPD) - (0) $Pr(Y = f \mid D = \text{unprivileged}) - Pr(Y = f \mid D = \text{privileged})$ • Disparate impact (DI) - 1 $\frac{Pr(Y = f \mid D = \text{unprivileged})}{Pr(Y = f \mid D = \text{privileged})}$ • Smoothed empirical differential (SED) - 0

Mitigating bias phases	ML Bias Mitigation Methods

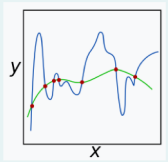
Preprocessing	• **Reweighting**
	  • fitted_cb_mdl = cb_mdl.fit(X_train, y_train, verbose=False, sample_weight=train_rw_ds.instance_weights)
	• **Disparate impact remover**
	 
	• Unawareness/Suppression
	  • delete features
	 
	• Feature Engineering – Guardrails
	 
	• Balancing/Resampling
	  • Upsampling/Downsampling
	 
	• Relabelling/Massaging
	• Learning fair representations
	  • Demographic Parity
	• Optimized preprocessing for discrimination
	  • changes labels and features

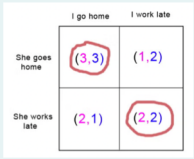
Recapitulation Mitigation Bias

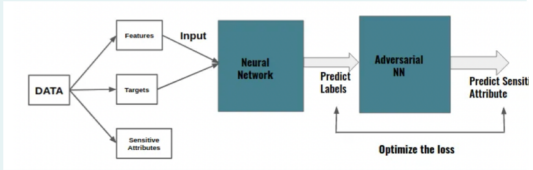
Brandenburgische
Technische Universität
Cottbus - Senftenberg

Mitigating bias phases	ML Bias Mitigation Methods

In-processing	• **Prejudice remover**
	  • **Ridge Regression and prejudice index (PI)**
	• **Gerry fair classifier**
	  • Nash equilibrium
	• Cost-sensitive training
	  • changes `class_weights` or `sample_weight`
	• Constraints
	  • Ordinalization + Guardrails
	• Adversarial debiasing
	• Exponentiated gradient reduction
	  • demographic parity or equalized odds
	  • changes weight of biased features







Recapitulation Mitigation Bias

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Mitigating bias phases	ML Bias Mitigation Methods

Post-processing	• **Equalized odds postprocessing**
	  • `EqOddsPostprocessing` changes labels using BinaryLabelDatasets
	• **Calibrated equalized odds postprocessing**
	  • like Equalized odds postprocessing + contraints for FNR or FPR
	• Prediction abstention
	  If confidence for a prediction <= threshold
	  -> not to make a prediction
	• Reject option classification
	• different rejection thresholds for different groups

Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Fair and win-win bank policy decisions

- Bank policy decisions have the potential to significantly affect cardholders' live
- Can impact to the level of life and death

Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

-> moral imperative to evaluate policy decisions with extreme care

The bank wants to understand the **causal relationships** between the policies and their effect

We want to use Causal Models

Causal Models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What are Causal Models good for?

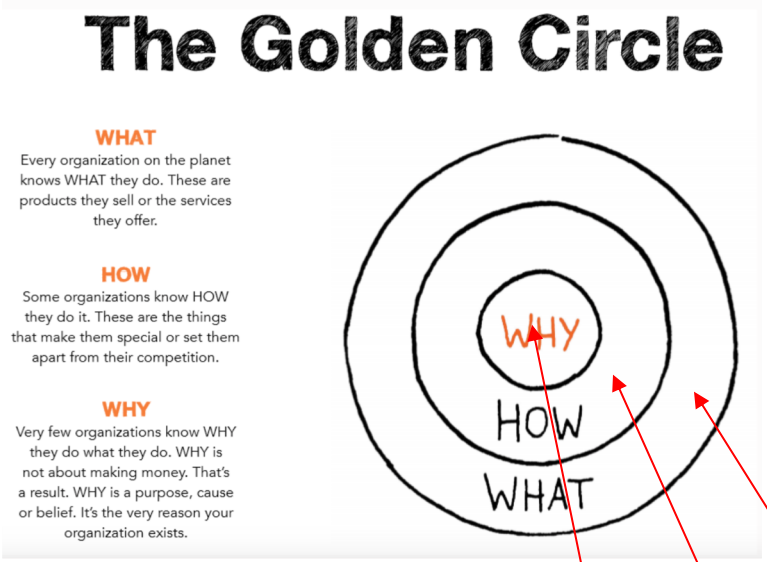
Causal Models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Understanding Cause and Effect Relationships
  - understand how changes in one variable may influence another
- Predictive Modeling
  - predictions about outcomes of certain interventions or changes in the system
- Policy and Decision Making
  - assess the consequences of various actions
- Healthcare and Medicine
  - evaluate the effectiveness of interventions
- Environmental Studies
  - assess the impact of human activities on ecosystems

...

Causal Interference



Brandenburgische Technische Universität Cottbus - Senftenberg

Causality vs. Prediction:

- Prediction
  - many machine learning techniques focus on prediction (what is the outcome)
- Causal inference
  - aims to uncover the mechanisms underlying the data generation process
  - **forecasting** future outcomes **+ understanding why and how** things happen

Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Causal inference in machine learning

- observes correlations
- to uncover the underlying cause-and-effect relationships between variables.
- powerful approach for applications in various domains
- -> to more informed and impactful decisions

Cause and Effect in Relationships

Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Causal inference graphs and models:
  - Combine all variables together and estimate effects to make more principled decisions

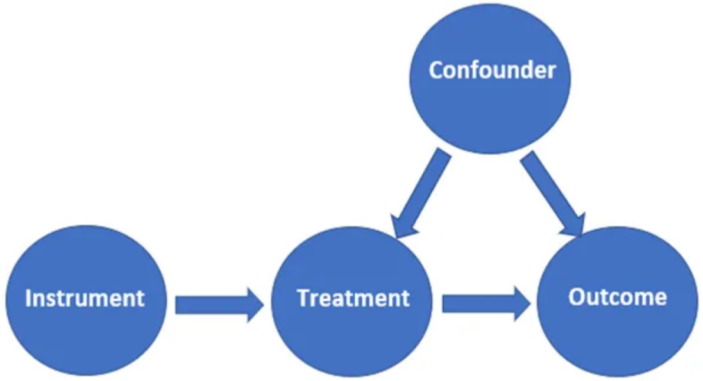
To properly assess the impact of a cause, whether by design or accident

-> **separate its effect from confounding variables**

Causal Inference Methods

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Creating a causal model
- Understanding heterogeneous treatment effects
- Testing estimate robustness



Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Decision-making will often involve understanding **cause** and **effect**

- Desirable effects:
  - you can decide to replicate its cause
- Unwanted effects:
  - avoid them
- change something on purpose to observe how it changes outcomes
- trace back an accidental effect to its cause
- simulate which change will produce the most beneficial impact

Causal Interference

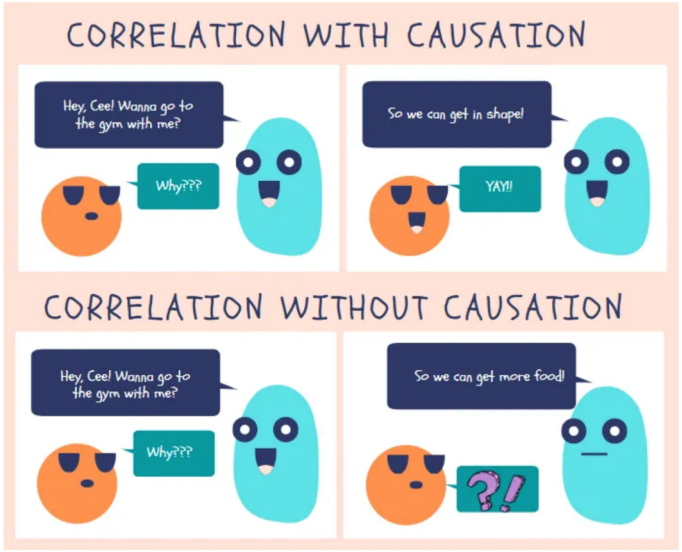
Brandenburgische
Technische Universität
Cottbus - Senftenberg

What does Causal Interference mean?

"What effect does changing variable X have on variable Y?" or

"Does variable X cause changes in variable Y?"

Causal Interference



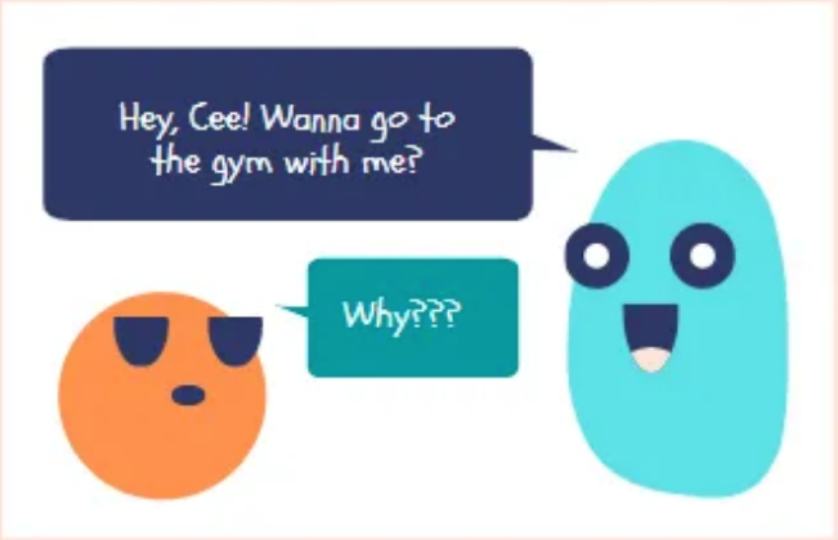
Brandenburgische Technische Universität Cottbus - Senftenberg

Correlation vs. Causation:

- Correlation
  - two variables tend to change together
  - but it doesn't necessarily mean that changes in one variable cause changes in the other
- Causation
  - If one variable's changes directly lead to changes in another

CORRELATION WITH CAUSATION

CORRELATION WITHOUT CAUSATION



Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

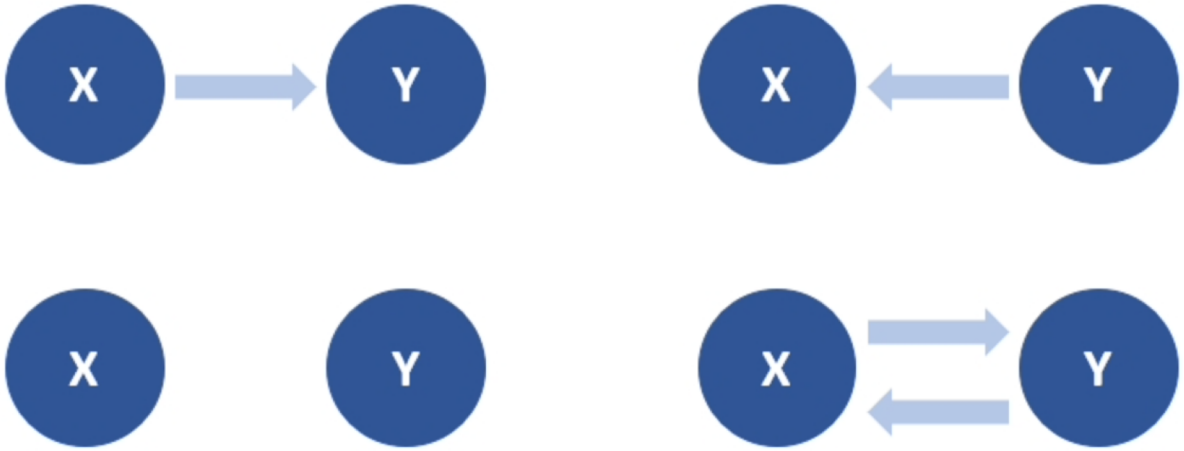
How can we identify a causal relationship?

Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Causal Relationship

-> either X -> Yor Y -> X applies (X,Y random variables)



What is the **Causal Direction** X -> Y or Y -> X?

Brandenburgische
Technische Universität
Cottbus - Senftenberg

So we are interested in the conditional probabilities for the two models X -> Yand Y -> X,

given our data x and y,

i.e. P(X -> Y|x,y): Probability that Y follows from X, if x and y occur or have occurred simultaneously

and P(Y -> X|x,y)

Brandenburgische
Technische Universität
Cottbus - Senftenberg

The Bayes factor relates these probabilities to each other:

$$O_{X \to Y} = \frac{P(X \to Y | x,y)}{P(Y \to X | x,y)}$$

We can transform it to

$$O_{X \to Y} = \frac{P(x,y | X \to Y)}{P(x,y | Y \to X)}$$

P(x,y|X -> Y): Probability that x and y occur simultaneously, if Y follows X

$$O_{X \to Y} = \frac{P(x,y | X \to Y)}{P(x,y | Y \to X)}$$

>1: *X* —> *Y*
<1: *Y* —> *X*

Probability that someone uses an umbrella x because it is raining y (X->Y) is greater than the probability that it is raining y because someone uses an umbrella x (e.g. against the sun) (Y->X)
>1: X->Y

Decision-making

Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

$$O_{X \to Y} = \frac{P(x,y|X \to Y)}{P(x,y|Y \to X)}$$

>1: $X \longrightarrow Y$
<1: $Y \longrightarrow X$

Probability that someone uses an umbrella x because it is raining y (X->Y) is greater than the probability that it is raining y because someone uses an umbrella x (e.g. against the sun) (Y->X)
>1: X->Y

What does it mean for the bank policies?

Decision-making

Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

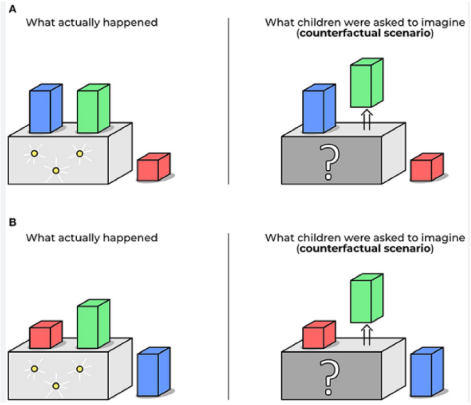
$$O_{X \to Y} = \frac{P(x,y|X \to Y)}{P(x,y|Y \to X)}$$

>1: *X* —> *Y*
<1: *Y* —> *X*

Probability that someone uses an umbrella x because it is raining y (X->Y) is greater than the probability that it is raining y because someone uses an umbrella x (e.g. against the sun) (Y->X)
>1: X->Y

Did the treatments decrease the default rate compared to the control group?

Causal Interference



Brandenburgische
Technische Universität
Cottbus - Senftenberg

Counterfactuals

- Causal inference **involves considering counterfactual** scenarios
  - **one variable's** value is **changed**
  - **other variables constant** -> impact to outcome
  - "what would have happened if"

- It aims to understand
  - how changing one variable affects the outcome
  - compared to what would have happened if the variable was not changed

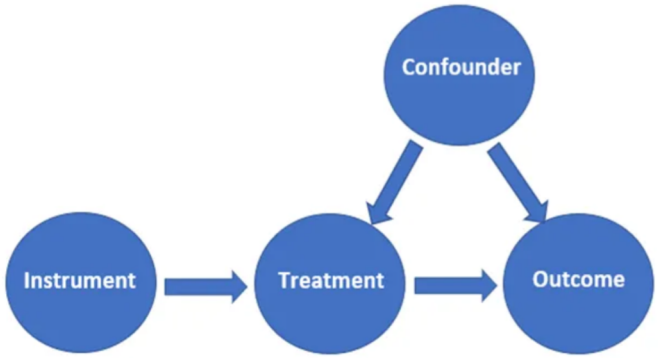
Causal Interference

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Confounding factors:

variables that are related to both the treatment and the outcome

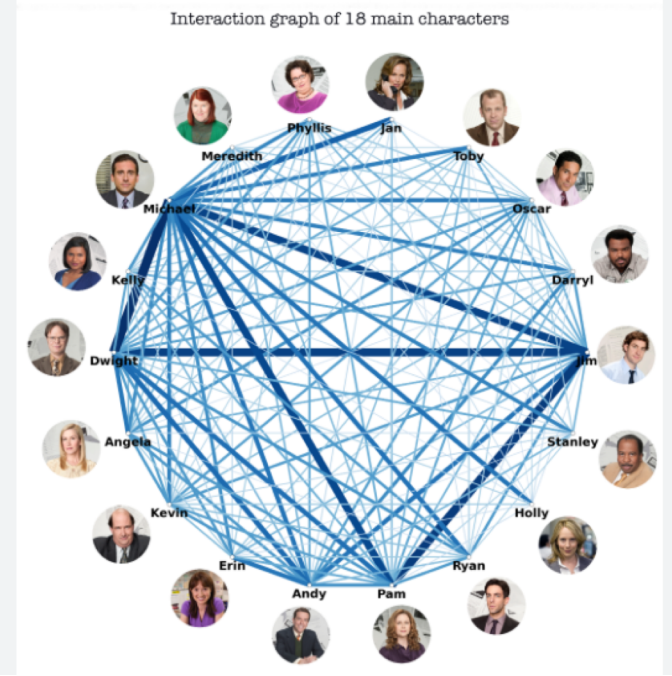
-> potentially leading to false causal conclusions



Brandenburgische
Technische Universität
Cottbus - Senftenberg

Causal Models:

- how variables interact and influence each other
- can help **simulate how changes** in one variable **affect other variables**



Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

The bank would like to develop a fair banking policy to achieve a win-win situation for customers and the bank

How would you get started?

Data

Brandenburgische
Technische Universität
Cottbus - Senftenberg

We use the same data as in Bias Mitigation

```
ccdefault_all_df = pd.read_csv('./data/credit/creditworthiness.csv',sep=',',encoding='latin-1')
```

- CC_LIMIT_CAT: ordinal; the credit card limit (_CC_LIMIT) separated into eight more or less equally distributed quartiles
- EDUCATION: nominal; the customer's educational attainment level (0: Other, 1:High School, 2: Undergraduate, 3: Graduate)
- MARITAL_STATUS: nominal; the customer's marital status (0: Other, 1: Single,2: Married)
- GENDER: nominal; the gender of the customer (1: Male, 2: Female)
- AGE GROUP: binary; denoting if the customer belongs to a privileged age group (1:privileged (26-47 years old), 0: underprivileged (every other age))
- pay_status_1… pay_status_6: ordinal; the repayment status for the previous six periods from April, pay_status_6, to August 2005, pay_status_1 (-1: pay duly, 1: payment is 1 month delayed, 2: payment is 2 months delayed… 8: 8 months delayed, 9: 9 months and above)
- paid_pct_1… paid_pct_6: continuous; what percentage of the bill due each month from April, paid_pct_6, to August 2005, paid_pct_1, was paid
- bill1_over_limit: continuous; the last bill's ratio in August 2005 over the corresponding credit limit
- IS_DEFAULT: binary; target; whether the customer has delayed repayment

- _...

Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What do we want?

- Develop a treatment
- -> Less important *IS_DEFAULT*
- Selling to your existing customers is far easier than acquiring new customers.

-> more bind a customer to you

Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Taiwanese bank started a lending policy experiment of 6 month

- experiment's focus:
- only persons considered salvageable (low-to-mid risk-of-default) customers

At the end:

- they want to understand
- **how** the **policies (treatments)** have **impacted customer behavior**

Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What do we want?

- Develop a treatment with a
- win-win result

What does a win-win result mean?

- High Lifetime value (LTV)

Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

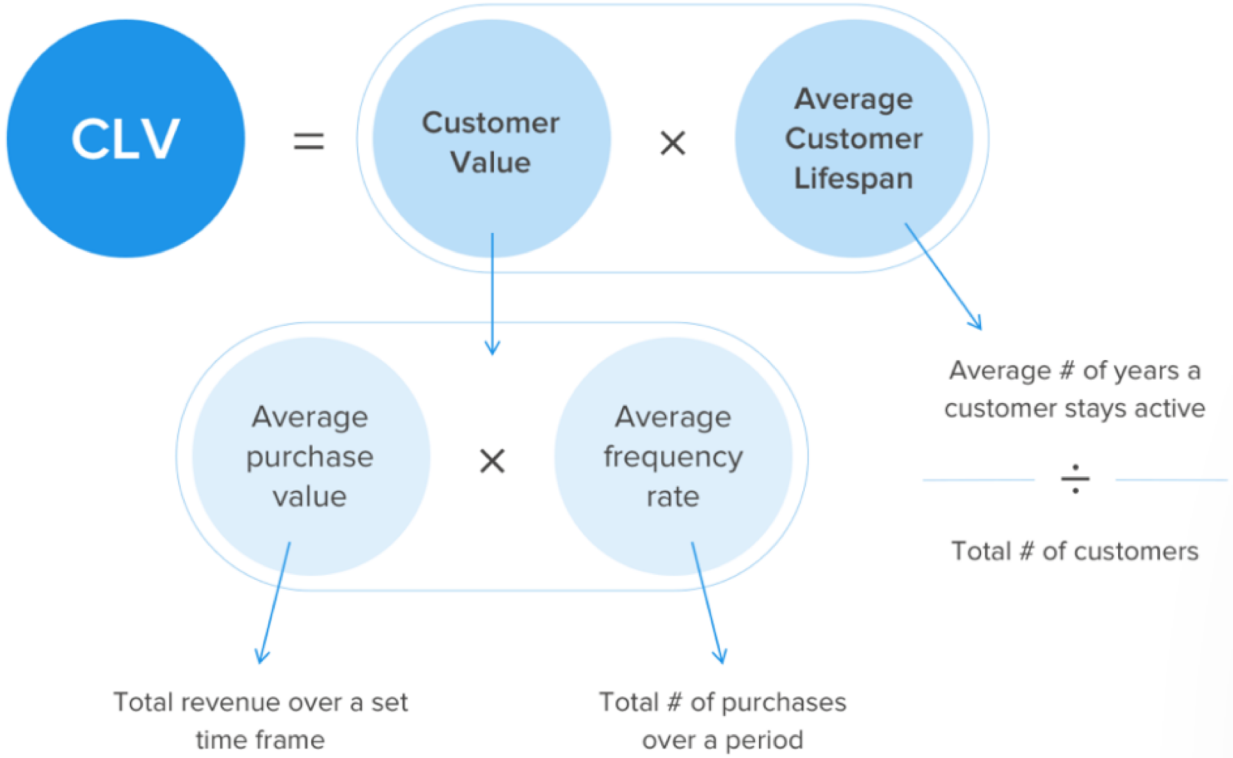
Lifetime value: what does it mean?

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Life Time Value (LTV) - Term from the management

- rule: it costs 5x as much to generate a new customer than to sell to your existing customers

complete customer lifetime value (CLV)



Mission

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Experiment: 3 Treatments

- Lower credit limit:
  - Some customers had their credit limit reduced by 25%.
- Payment plan:
  - They were given 6 months to pay back their current credit card debt.
  - The debt was split up into six parts, and every month they would have to pay one part.
- Both measures:
  - A reduction in credit limit and the payment plan.

Data

Brandenburgische
Technische Universität
Cottbus - Senftenberg

...

• _TREATMENT: nominal; the intervention or policy prescribed to each customer**(-1: not part of the experiment, 0: Control group, 1: Lower Credit Limit, 2:Payment Plan, 3: Payment Plan and Credit Limit)**

• _LTV: continuous; the outcome of the intervention, which is the **lifetime value** estimated in NT$ given the credit payment behavior over the previous 6 months

• _CC_LIMIT: continuous; the **original credit card limit** in NT$ that the customer had before the treatment. Bankers expect the outcome of the treatment to be greatly impacted by this feature.

• _risk_score: continuous; the risk score that the bank computed 6 months prior for each customer based on credit card bills' ratio over their credit card limit.
It's like bill1_over_limit except it's a weighted average of 6 months of payment history, and it was produced 5 months before to choose the treatment.

...

new target

Data

Brandenburgische
Technische Universität
Cottbus - Senftenberg

This data do we need later as **Confounders**

....

_spend: continuous; how much was spent by each customer in New Taiwan Dollar (NT$)

_tpm: continuous; **median transactions per month** made by the customer with the credit card over the previous 6 months

_ppm: continuous; **median purchases per month** made by the customer with the credit card over the previous 6 months

_RETAIL: binary; if the customer is retail, instead of a customer obtained through their employer

_URBAN: binary; if it's an urban customer

_RURAL: binary; if it's a rural customer

_PREMIUM: binary; if the customer is "premium". Premium customers get cashback offers and other spending incentives

Methods

Brandenburgische
Technische Universität
Cottbus - Senftenberg

1. Understanding experiment results
2. Create a Causal Model
3. Understanding Treatment Effects
4. Choosing Policies
5. Testing estimate robustness

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Plot the experiment results

1. Percentage of each experiment group the defaulted
2. sum of LTV of each experiment group
3. plot

   x-axis (experiment group)

   two y axes (% default, sum LTV)

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Plot the experiment results

1. Percentage for each group that defaulted (pct_s)

```
pct_s = ccdefault_causal_df[ccdefault_causal_df.IS_DEFAULT==1].\
                groupby(['_TREATMENT']).size() /\
        ccdefault_causal_df.groupby(['_TREATMENT']).size()
```

Number of default/ all

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Plot the Life Time Values (LTV)

2. the sum of lifetime values for each group (ltv_s) in thousands of NTD (New Taiwan Dollar - K$).

Number of default/ all

```
ltv_s = ccdefault_causal_df.groupby(['_TREATMENT'])\
                                        ['_LTV'].sum()/1000
```

Data - Treatment

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Define treatments

```
treatment_names = ['Lower Credit Limit', 'Payment Plan', 'Payment Plan &\nLower Credit Limit']
all_treatment_names = np.array(["None"] + treatment_names)
```

Understanding the results of the experiment

We want to know

- Did the treatments **decrease the default** rate compared to the control group?
- Has spending behavior **increased the estimated lifetime value**

-> visualize both in a single plot

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

3. Plot

Two y-axis
1. percentage for each group that defaulted (pct_s): 2. Lifetime value
4 groups-> 4 values

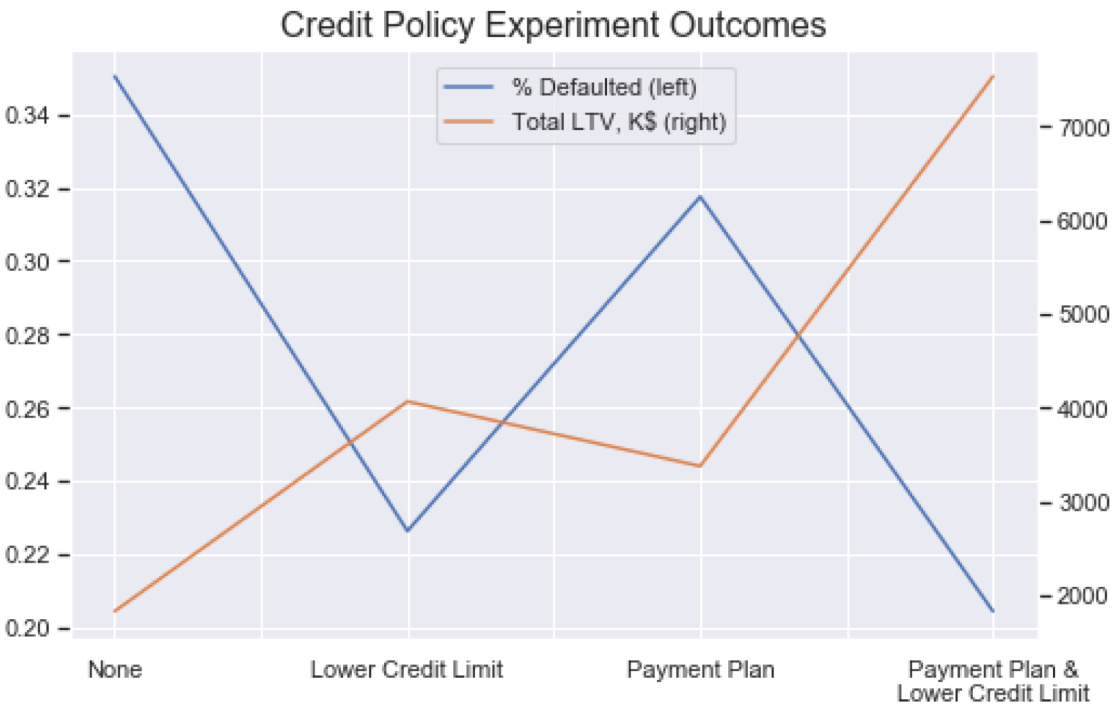
```
plot_df = pd.DataFrame({'% Defaulted (left)':pct_s, 'Total LTV, K$':ltv_s})
plot_df.index = all_treatment_names

ax = plot_df.plot(secondary_y=['Total LTV, K$'], figsize=(8,5))
ax.get_legend().set_bbox_to_anchor((0.7, 0.99))
plt.grid(False)
plt.title("Credit Policy Experiment Outcomes", fontsize=16)
plt.show()
```

x-axis

Legend: x position, y position

Data - Treatment

Brandenburgische
Technische Universität
Cottbus - Senftenberg



_TREATMENT:
0: Control group,
1: Lower Credit Limit,
2:Payment Plan,
3: Payment Plan and Credit Limit

Result:
All treatments are better than the control group

- Less default
- Higher lifetime value

The lowering of the credit limit

- decreases the default rate over 12%
- and more than doubles the estimated LTV

Payment plan

- only decreases the defaults 3%
- and increases the LTV by about 85%

both policies combined

- reduced the default rate nearly 15%
- And quadrupled the control group's LTV

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Now we have a good treatment

Important
how they distributed it among the credit cardholders

Bank **chose treatment** according to their **risk factor**
-> subjective & expensive

but
Lifetime value (_LTV) is largely affected by the credit limit available (_CC_LIMIT)

-> must take that into account

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Do _risk_score and _CC_LIMIT have a correlation?

-> plot

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

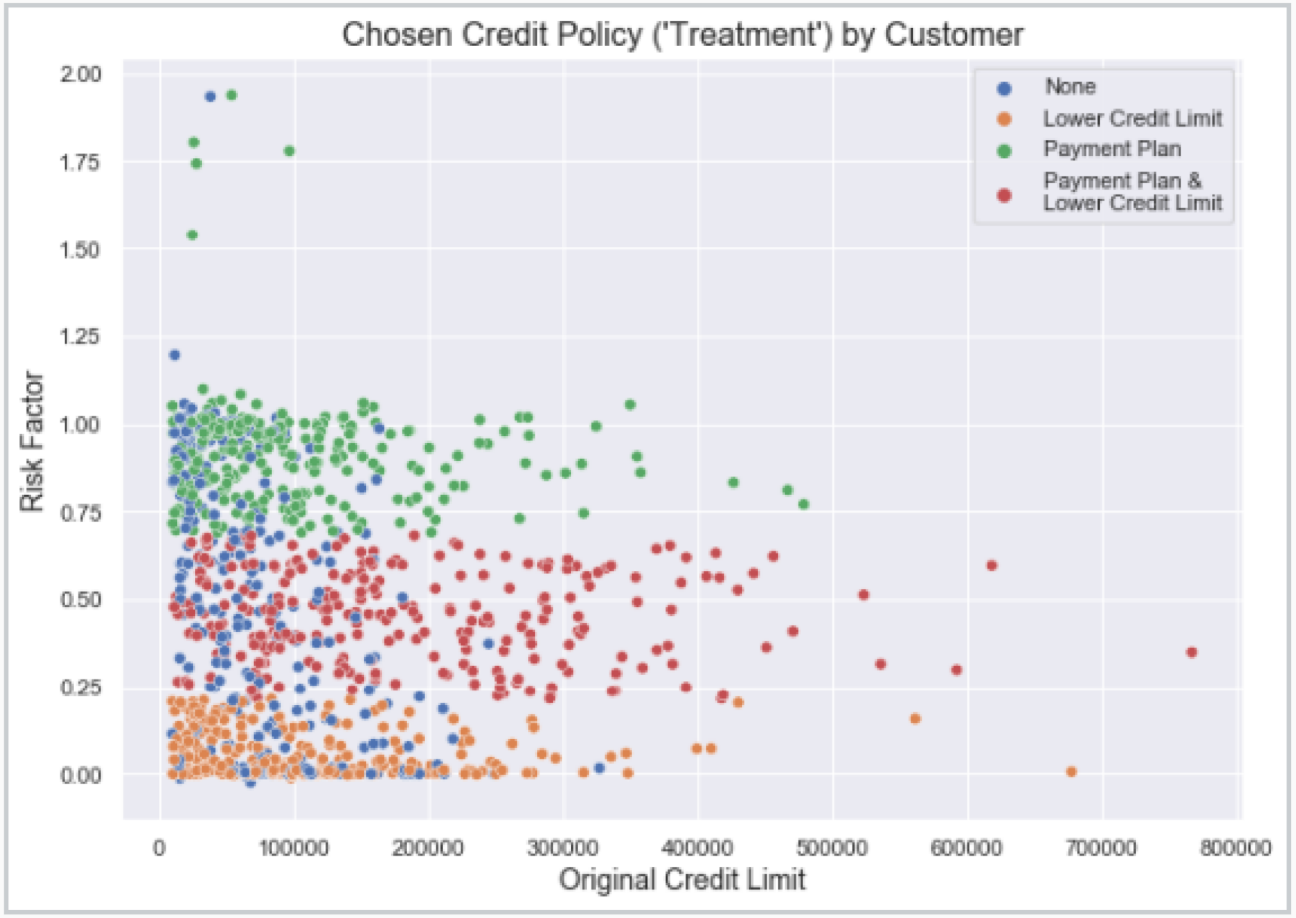
Understand the distribution *_CC_LIMIT* and *risk_score* is by plotting both variables against each other

```
plt.figure(figsize=(10, 7))
sns.scatterplot(
    x=ccdefault_causal_df['_CC_LIMIT'].values,
    y=ccdefault_causal_df['_risk_score'].values,
    hue=all_treatment_names[ccdefault_causal_df['_TREATMENT'].values],
    hue_order=all_treatment_names
)
plt.title("Chosen Credit Policy ('Treatment') by Customer", fontsize=16)
plt.xlabel("Original Credit Limit", fontsize=14)
plt.ylabel("Risk Factor", fontsize=14)
plt.show()
```

Hue *vector or key in data g*rouping variable that will produce points with different colors

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg



- control group (None) is spread out more vertically
- treatments based on risk level also meant that they unevenly distributed the treatments based on _CC_LIMIT

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What is the distribution of _CC_LIMIT and _LTV according to the Treatment?

How can we visualize it?

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg

We use a histogram

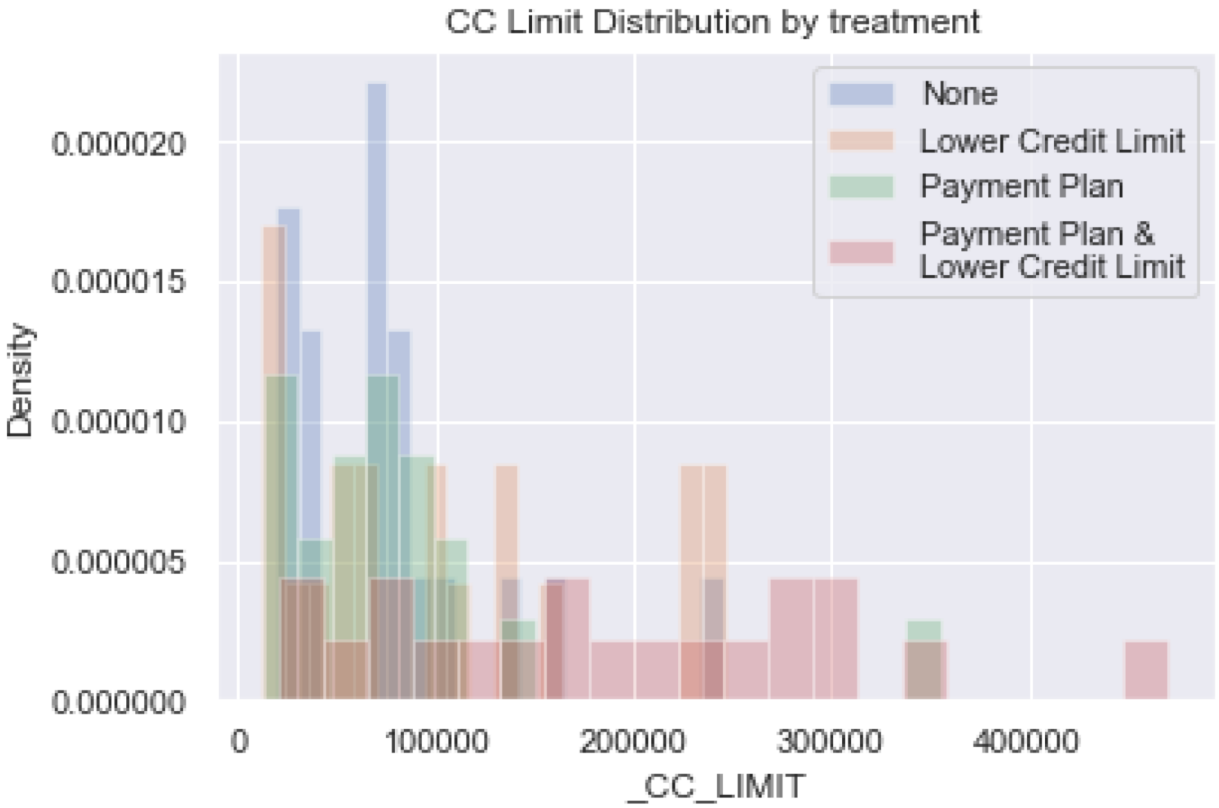
Color saturation

density=**True**
return a probability density :
each bin will display the bin's raw count
divided by the total number of counts

```
kwargs = dict(alpha=0.5, bins=n, density=True)

i = 0
for name in all_treatment_names:
    x = ccdefault_all_df[ccdefault_all_df._TREATMENT == i]['_CC_LIMIT'][:n].values
    # Plot
    plt.hist(x, **kwargs, label=name)
    i = i+1
plt.xlabel('_CC_LIMIT')
plt.ylabel('Density')
plt.title('CC Limit Distribution by treatment')
plt.legend()
plt.show()
```

Data - Treatment
Understanding experiment results

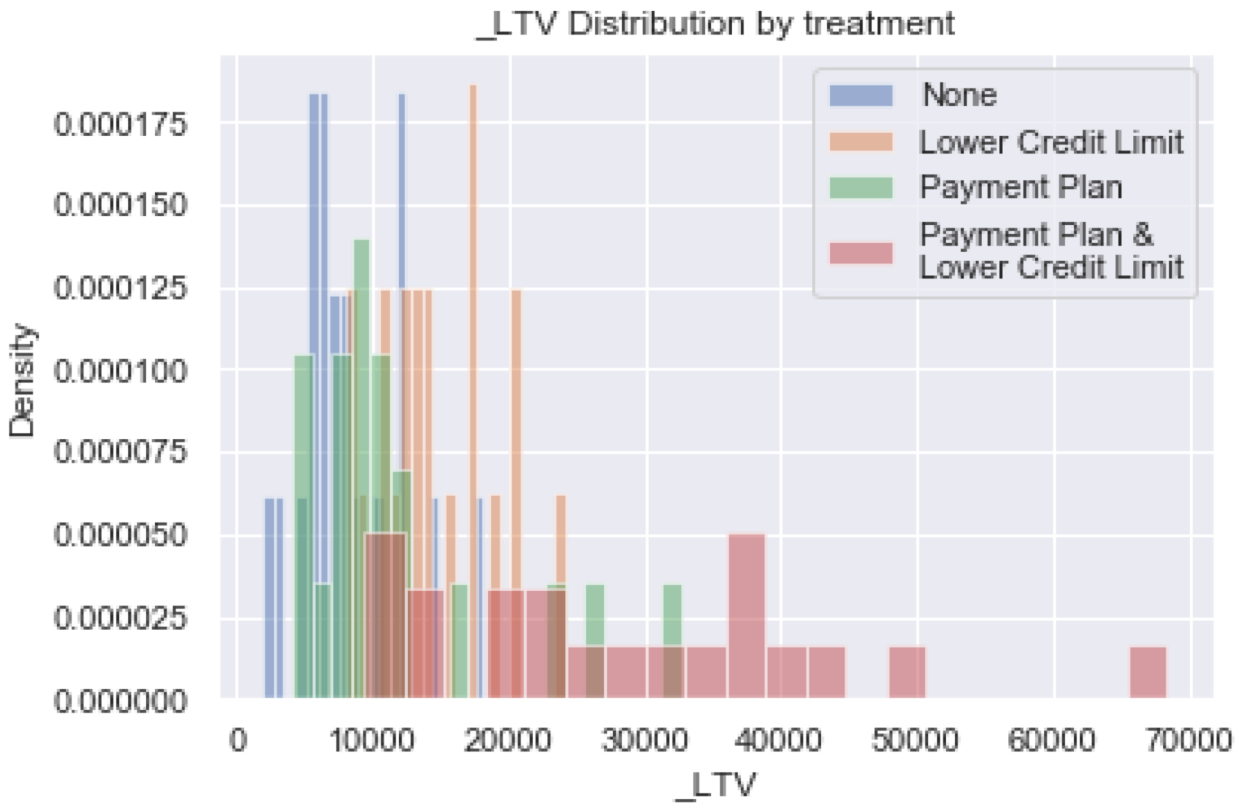
Brandenburgische
Technische Universität
Cottbus - Senftenberg



The treatment depends on the credit card limit _CC_LIMIT

Data - Treatment
Understanding experiment results

Brandenburgische
Technische Universität
Cottbus - Senftenberg



The outcome Lifetime value _LTV depends on the treatment

Causal models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What are Causal Models and what they are good for?

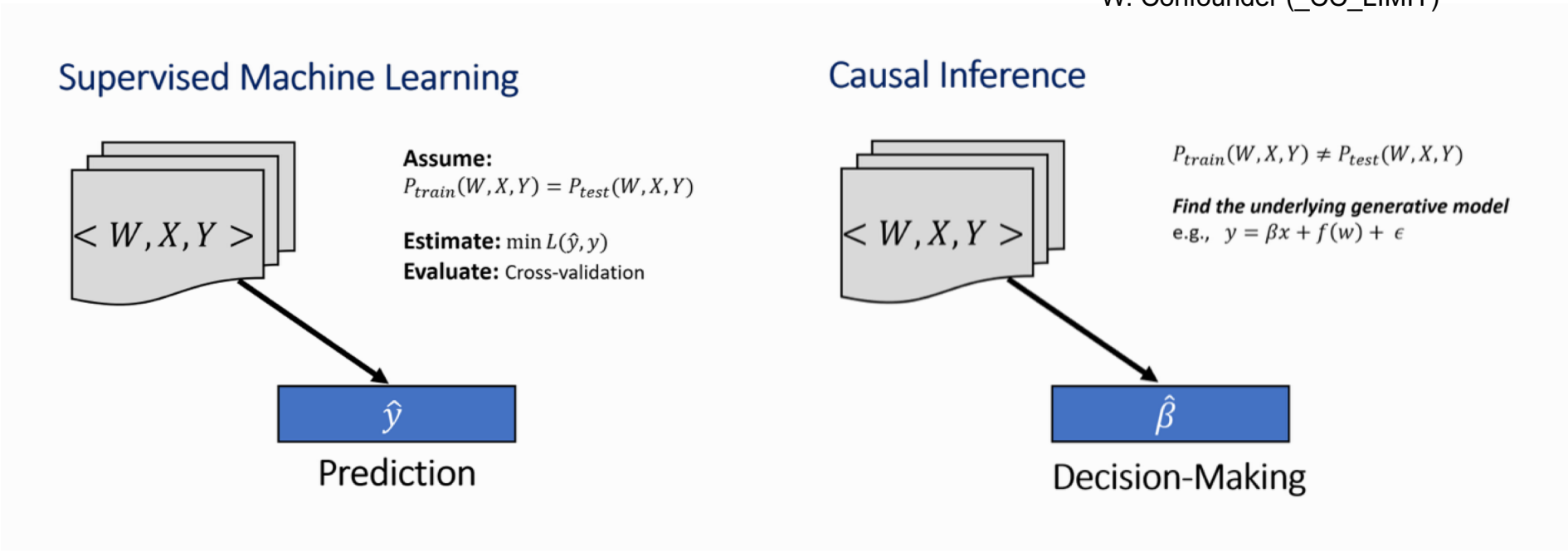
Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Difference between prediction and causal inference

- Predict outcome when changing the Effect Modifiers

W: Confounder (_CC_LIMIT)

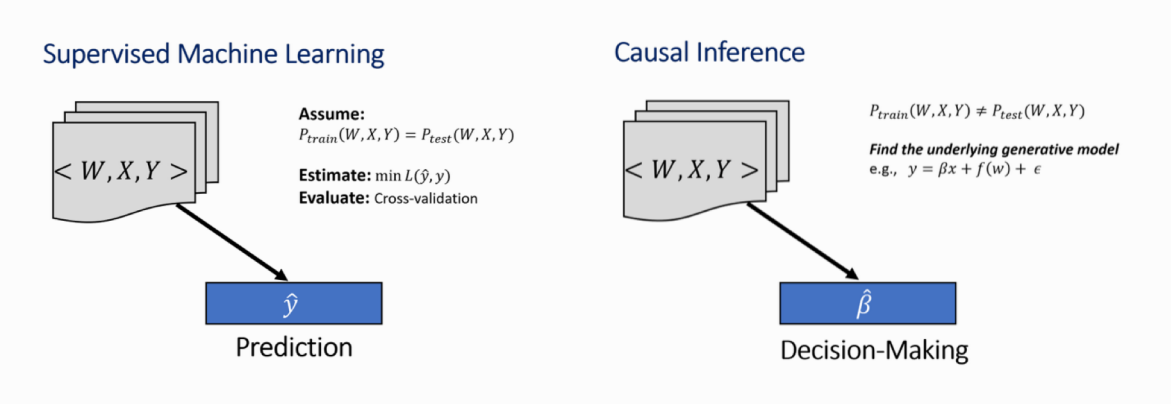


https://www.pywhy.org/dowhy/v0.8/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Creating a causal model

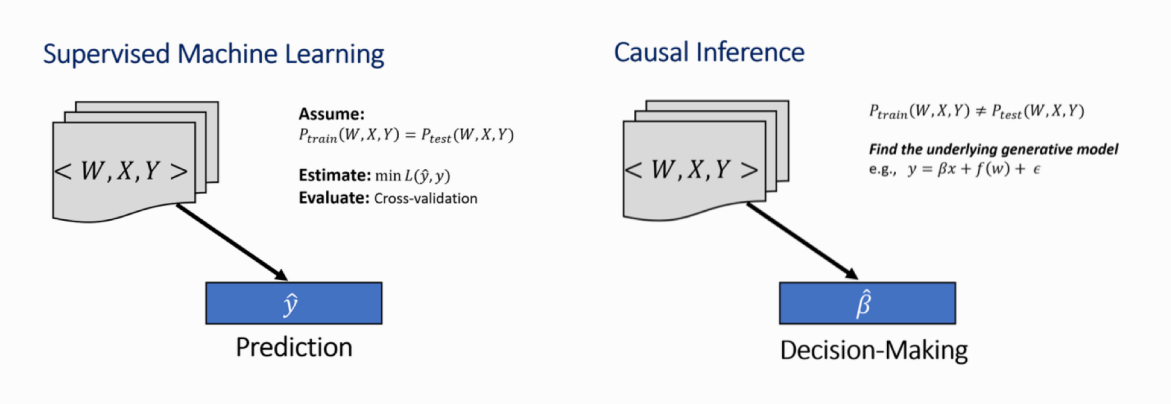
Brandenburgische
Technische Universität
Cottbus - Senftenberg

What result do we want to get using Causal Interference?



Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg



3 Equations (for 3 treatment)

_LTV = intercept + coef * _CC_LIMIT

https://www.pywhy.org/dowhy/v0.8/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Understanding causal models

The causal model we will build can be separated into four components:

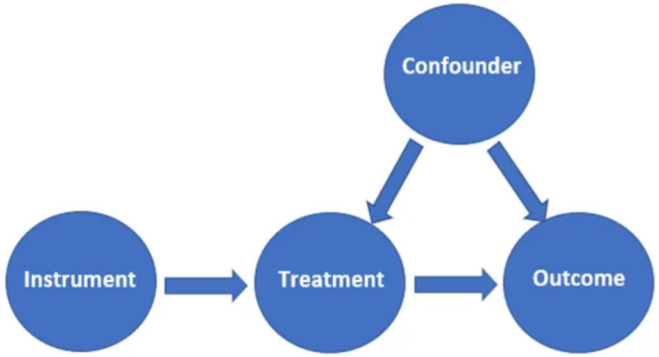
- Outcome (Y):
  - The outcome variable(s) of the causal model.
- Treatments (T):
  - The treatment variable(s) that influences the outcome.
- Effect modifiers (X):
  - The variable(s) that influences the effect's heterogeneity conditioning it. Effect modification occurs when the effect of a factor is different for different groups, It can sit between the treatment and the outcome.
- Confounders (W):
  - Also known as common causes or controls. They are the features that influence both the outcome and the treatment.

Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Heterogeneous treatment effect modifier

- The varying feature that **directly impacts the treatment effect**
- → *_CC_LIMIT*



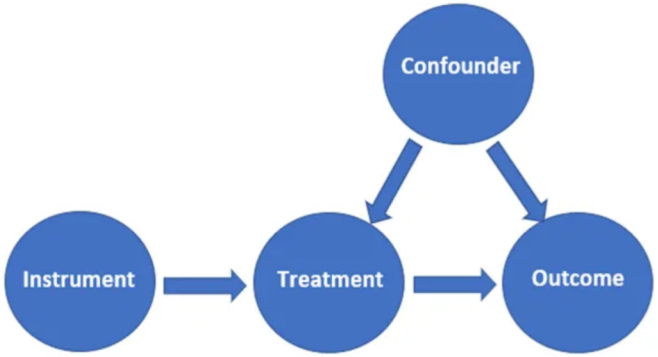
- -> We can create a causal model that includes both the *_TREATMENT* feature and the effect modifier *_CC_LIMIT*.

Understanding causal models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What are our

- Effect modifier/Instrument (X)
- Outcome (Y)
- Treatment (T)
- Controls/Confounders (W)?

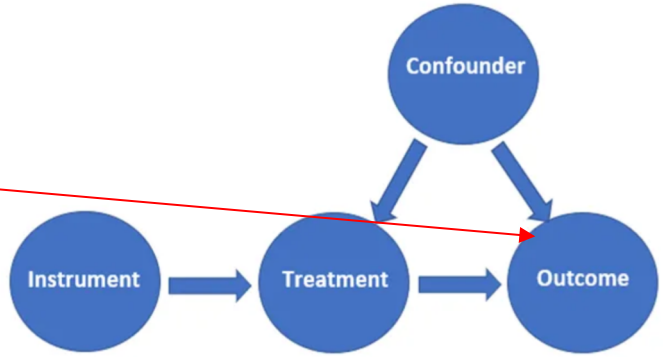


Understanding causal models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What are our

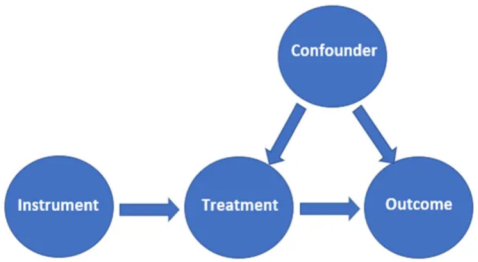
- Effect modifier/Instrument (X)
  - _CC_LIMIT
- Outcome (Y)
  - Lifetime value _LTV
- Treatment (T)
  - _TREATMENT
- Controls/Confounders (W)?



Understanding causal models

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Outcome (Y)
  - Lifetime value _LTV
- Treatment (T)
  - _TREATMENT
- Effect modifier (X)
  - _CC_LIMIT
- Controls (W)?



_spend: continuous; **how much was spent** by each customer in New Taiwan Dollar (NT$)

_tpm: continuous; **median transactions per month** made by the customer with the credit card over the previous 6 months

_ppm: continuous; **median purchases per month** made by the customer with the credit card over the previous 6 months

_RETAIL: binary; if the customer is retail, instead of a customer obtained through their employer

_URBAN: binary; if it's an urban customer

_RURAL: binary; if it's a rural customer

_PREMIUM: binary; if the customer is "premium". Premium customers get cashback offers and other spending incentives

Creating a causal model

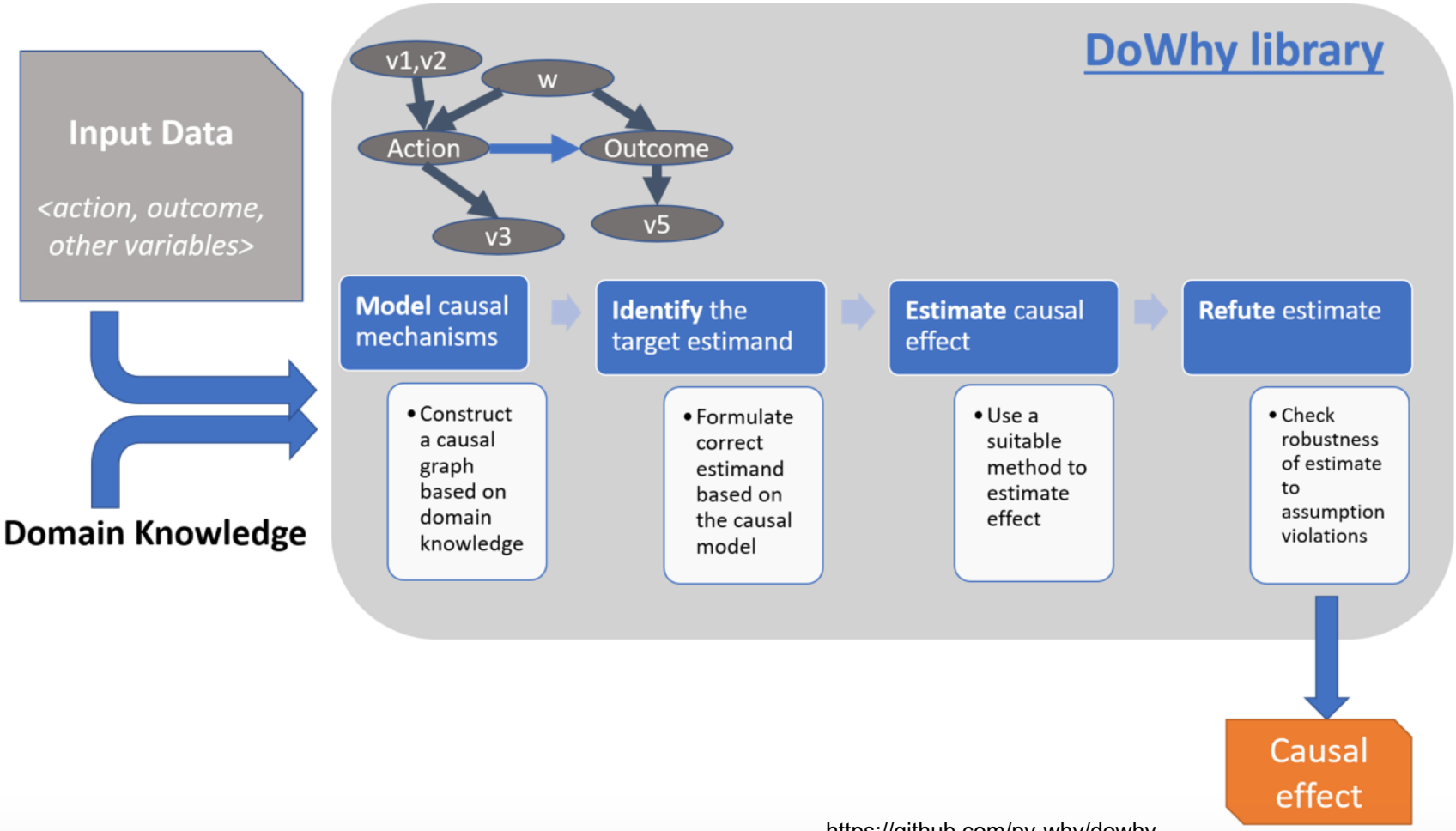
Brandenburgische
Technische Universität
Cottbus - Senftenberg

- Outcome (Y)
  - Lifetime value _LTV
- Treatment (T)
  - _TREATMENT
- Effect modifier (X)
  - _CC_LIMIT
- Controls/Confounders (W)

```
Y = ccdefault_causal_df[['_LTV']]
T = ccdefault_causal_df[['_TREATMENT']]
X = ccdefault_causal_df[['_CC_LIMIT']]
W = ccdefault_causal_df[['_spend','_tpm','_ppm','_RETAIL','_URBAN','_RURAL','_PREMIUM']]
```

Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg



https://github.com/py-why/dowhy

Creating a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

DoWhy library

The goal is to derive the Conditional Average Treatment Effect (CATE)

Estimating the causal effect requires four steps:

1. Create a **causal model** from the data and given graph
2. **Identify causal effect** and return target estimands
3. **Estimate the target estimand** using a statistical method
4. Refute the obtained estimate using **multiple robustness checks**

https://www.pywhy.org/dowhy/v0.11.1/user_guide/causal_tasks/estimating_causal_effects/index.html

Create a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Use Deep Reinforcement Learning method **LinearDRLearner** with two models

We have multiple treatments:
-> `multi:softmax`

```
from econml.dr import LinearDRLearner

drlearner = LinearDRLearner(
                model_regression=xgb.XGBRegressor(learning_rate=0.1),
                model_propensity=xgb.XGBClassifier(learning_rate=0.1, max_depth=2,
                                    objective="multi:softmax"),random_state=rand,
)
```

model_regression (*scikit-learn regressor or 'auto', default 'auto'*) – Environment
Estimator for E[Y | X, W, T].
Trained by regressing Y on (features, controls, one-hot-encoded treatments) concatenated.
The one-hot-encoding excludes the baseline treatment. Must implement *fit* and *predict* methods.

model_propensity (*scikit-learn classifier or 'auto', default 'auto'*) – Agent
Estimator for Pr[T=t | X, W].
Trained by regressing treatments on (features, controls) concatenated.
Must implement *fit* and *predict_proba* methods.

X: Modifier _CC_LIMIT
W: Confounder all other features
T: Treatment

Create a causal model

Brandenburgische
Technische Universität
Cottbus - Senftenberg

```
Y = ccdefault_causal_df[['_LTV']]
T = ccdefault_causal_df[['_TREATMENT']]
X = ccdefault_causal_df[['_CC_LIMIT']]
W = ccdefault_causal_df[['_spend','_tpm','_ppm','_RETAIL','_URBAN','_RURAL','_PREMIUM']]
```

Fit the models

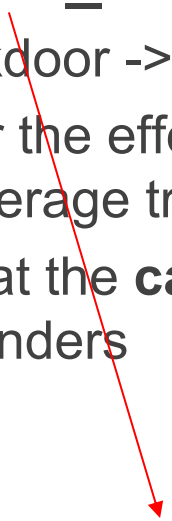
```
causal_mdl = drlearner.dowhy.fit(Y, T, X=X, W=W,
                                 outcome_names=Y.columns.to_list(),
                                 treatment_names=T.columns.to_list(),
                                 feature_names=X.columns.to_list(),
                                 confounder_names=W.columns.to_list(),
                                 target_units=X.iloc[:550].values)
```

Identify causal effect and return target estimands

Brandenburgische Technische Universität Cottbus - Senftenberg

Now we want to understand heterogeneous treatment effects

- DoWhy **identified_estimand_**
  - Identify the **estimand** – backdoor -> (*_TREATMENT*)
  - the probability expression for the effect *(_LTV*) to be estimated (the identified estimand) - average treatment effect (ATE)
  - backdoor criterion implies that the **causal effect can be identified** by conditioning on all confounders



```
identified_ate = causal_mdl.identified_estimand_
print(identified_ate)
```

Identify causal effect and return target estimands

Brandenburgische Technische Universität Cottbus - Senftenberg

Identify causal effect and return target estimands

```
Estimand type: EstimandType.NONPARAMETRIC_ATE

### Estimand : 1
Estimand name: backdoor
Estimand expression:
      d
————————————(E [_LTV|_RETAIL,_RURAL,_PREMIUM,_spend,_ppm,_CC_LIMIT,_URBAN,_tpm
d[_TREATMENT]
])

Estimand assumption 1,
Unconfoundedness: If U→{_TREATMENT} and U→_LTV
then P(_LTV|_TREATMENT,_RETAIL,_RURAL,_PREMIUM,_spend,_ppm,_CC_LIMIT,_URBAN,_tpm,U)
= P(_LTV|_TREATMENT,_RETAIL,_RURAL,_PREMIUM,_spend,_ppm,_CC_LIMIT,_URBAN,_tpm)

### Estimand : 2
Estimand name: iv
No such variable(s) found!

### Estimand : 3
Estimand name: frontdoor
No such variable(s) found!
```

Assumption: No confounder

probability expression for the effect *(_LTV)*

Estimate the target estimand using a statistical method

Brandenburgische Technische Universität Cottbus - Senftenberg

Predict the target using the estimand with different values

```
drlearner_estimate = causal_mdl.estimate_
print(drlearner_estimate)
```

```
## Realized estimand
b: _LTV~_TREATMENT+_RETAIL+_RURAL+_PREMIUM+_spend+_ppm+_CC_LIMIT+_URBAN+_tpm | _CC_LIMIT
Target units:

## Estimate
Mean value: 7227.831176285025
Effect estimates: [[6760.04077377]
 [7338.73597439]
 [7365.03622855]
 ...
 [7224.08760433]
 [7504.29508224]
 [7221.24571753]]
```

Output Treatment Modifier

Live Time Values

https://www.pywhy.org/dowhy/v0.8/example_notebooks/tutorial-causalinference-machinelearning-using-dowhy-econml.html

Identify causal effect and return target estimands

Brandenburgische Technische Universität Cottbus - Senftenberg

Print coefficients and intercepts

```
idxs = np.arange(0, causal_mdl._d_t[0]) # 3 TREATMENTS
print("idxs\n",idxs)
coefs = np.hstack([causal_mdl.coef_(T=i+1) for i in idxs])
print("coefs\n",coefs)
intercepts = np.hstack([causal_mdl.intercept_(T=i+1) for i in idxs])
print("intercepts\n",intercepts)
```

```
coefs
 [0.00640065 0.04209277 0.06804091]
intercepts
 [6505.34627058 1245.33441271 5576.02328924]
```

We can calculate the proposed _LTV

$$\_LTV = intercept + coef * \_CC\_LIMIT$$

Understanding heterogeneous treatment effects

Brandenburgische Technische Universität Cottbus - Senftenberg

Summery of coefficients, intercepts, their std, pvalue

```
for i in range(causal_mdl._d_t[0]):
    print("Treatment: %s" % treatment_names[i])
    display(causal_mdl.summary(T=i+1))
    print("\r\n")
```

```
Causal Estimate is 3604.8891161642023
Treatment: Lower Credit Limit
<class 'econml.utilities.Summary'>
"""
                          Coefficient Results
=============================================================
          point_estimate stderr zstat pvalue ci_lower ci_upper
-------------------------------------------------------------
_CC_LIMIT          0.006   0.02 0.326  0.744   -0.032    0.045
                        CATE Intercept Results
=============================================================
                point_estimate  stderr  zstat pvalue ci_lower ci_upper
-------------------------------------------------------------
cate_intercept       6505.346 1312.803 4.955    0.0 3932.299 9078.393
-------------------------------------------------------------
```

```
Treatment: Payment Plan
<class 'econml.utilities.Summary'>
"""
                          Coefficient Results
=============================================================
          point_estimate stderr zstat pvalue ci_lower ci_upper
-------------------------------------------------------------
_CC_LIMIT          0.042  0.019  2.23  0.026    0.005    0.079
                        CATE Intercept Results
=============================================================
                point_estimate  stderr  zstat pvalue  ci_lower ci_upper
-------------------------------------------------------------
cate_intercept       1245.334 1225.572 1.016   0.31 -1156.742  3647.41
-------------------------------------------------------------
```

```
Treatment: Payment Plan &
Lower Credit Limit
<class 'econml.utilities.Summary'>
"""
                          Coefficient Results
=============================================================
          point_estimate stderr zstat pvalue ci_lower ci_upper
-------------------------------------------------------------
_CC_LIMIT          0.068  0.019 3.629    0.0    0.031    0.105
                        CATE Intercept Results
=============================================================
                point_estimate  stderr  zstat pvalue ci_lower ci_upper
-------------------------------------------------------------
cate_intercept       5576.023 1238.128 4.504    0.0 3149.337  8002.71
-------------------------------------------------------------
```

Understanding heterogeneous treatment effects

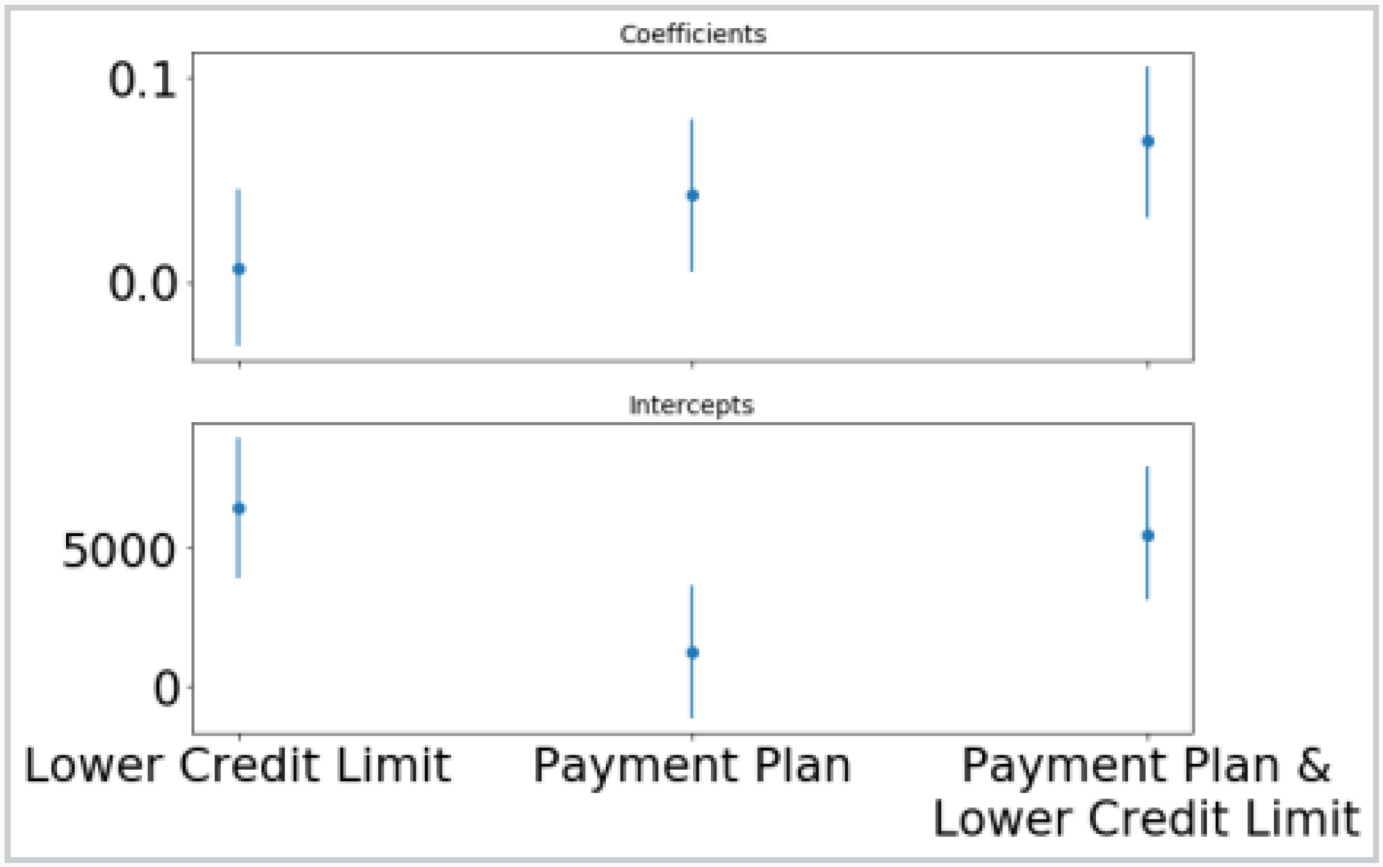
Brandenburgische
Technische Universität
Cottbus - Senftenberg

We can visualize the coefficients and intercepts for the 3 treatments

```
idxs = np.arange(0, causal_mdl._d_t[0])
coefs = np.hstack([causal_mdl.coef_(T=i+1) for i in idxs])
intercepts = np.hstack([causal_mdl.intercept_(T=i+1) for i in idxs])

coefs_err = np.hstack([causal_mdl.coef__interval(T=i+1) for i in idxs])
coefs_err[0, :] = coefs - coefs_err[0, :]
coefs_err[1, :] = coefs_err[1, :] - coefs

intercepts_err = np.vstack(
    [causal_mdl.intercept__interval(T=i+1) for i in idxs]
).T
intercepts_err[0, :] = intercepts - intercepts_err[0, :]
intercepts_err[1, :] = intercepts_err[1, :] - intercepts

plt.figure(figsize=(10, 7))
ax1 = plt.subplot(2, 1, 1)
plt.errorbar(idxs, coefs, coefs_err, fmt="o")
plt.xticks(idxs, treatment_names)
plt.setp(ax1.get_xticklabels(), visible=False)
plt.title("Coefficients", fontsize=14)

plt.subplot(2, 1, 2)
plt.errorbar(idxs, intercepts, intercepts_err, fmt="o")
plt.xticks(idxs, treatment_names)
plt.title("Intercepts", fontsize=14)
plt.show()
```

Understanding heterogeneous treatment effects

Brandenburgische
Technische Universität
Cottbus - Senftenberg



Coefficients -> best LTV: Payment Plan & Lower Credit Limit
Impact of Intercepts?

Understanding heterogeneous treatment effects

Brandenburgische
Technische Universität
Cottbus - Senftenberg

We could compute the best LTV for each customer

But

Additional costs to the 3 treatments:

- Setting up a payment plan requires administrative and legal costs of about $1,000 per contract,
- lowering the credit limit -> estimated at $72 per average payments per month (_ppm) over the lifetime of the customer.

- -> plot _CC_LIMIT, Treatment, payment per month (_ppm)

Understanding heterogeneous treatment effects

Brandenburgische Technische Universität Cottbus - Senftenberg

Additional costs to the 3 treatments:

- payment plan. $1,000
- lowering the credit limit: $72 per average payments per month (_ppm) over the lifetime of the customer.

```
cost_fn = lambda X: np.repeat(np.array([[0, 1000, 1000]]), X.shape[0], axis=0) +\
                    (np.repeat(np.array([[72, 0, 72]]), X.shape[0], axis=0) *\
                     X._ppm.values.reshape(-1,1))
```

Lower Credit Limit. Payment Plan. Lower Credit Limit& Payment Plan

Understanding heterogeneous treatment effects

Brandenburgische Technische Universität Cottbus - Senftenberg

Opt_LTV = new _LTV – additional costs

Recommended Treatment = Treatment-index(max(Opt_LTV))

```
treatment_effect_minus_costs = causal_mdl.const_marginal_effect(X=X.values) -\
                                    cost_fn(ccdefault_causal_df)
treatment_effect_minus_costs = np.hstack([np.zeros(X.shape), treatment_effect_minus_costs])


recommended_T = np.argmax(treatment_effect_minus_costs, axis=1)
```

index of highest of 4 LTV

Understanding heterogeneous treatment effects

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Plot
_CC_LIMIT, Treatment, payment per month (_ppm)
with additional costs

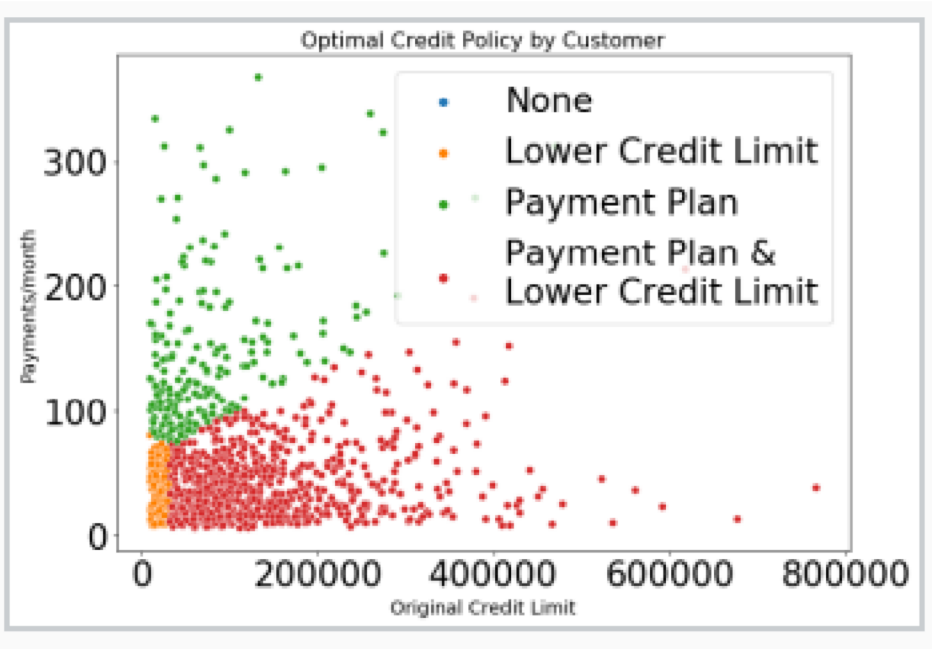
Understanding heterogeneous treatment effects

Brandenburgische Technische Universität Cottbus - Senftenberg

Plot best Treatment and original _CC_LIMIT per client

```
plt.figure(figsize=(10, 7))
sns.scatterplot(
    x=ccdefault_causal_df['_CC_LIMIT'].values,
    y=ccdefault_causal_df["_ppm"].values,
    hue=all_treatment_names[recommended_T],
    hue_order=all_treatment_names
)
plt.title("Optimal Credit Policy by Customer", fontsize=16)
plt.xlabel("Original Credit Limit", fontsize=14)
plt.ylabel("Payments/month", fontsize=14)
plt.show(
```

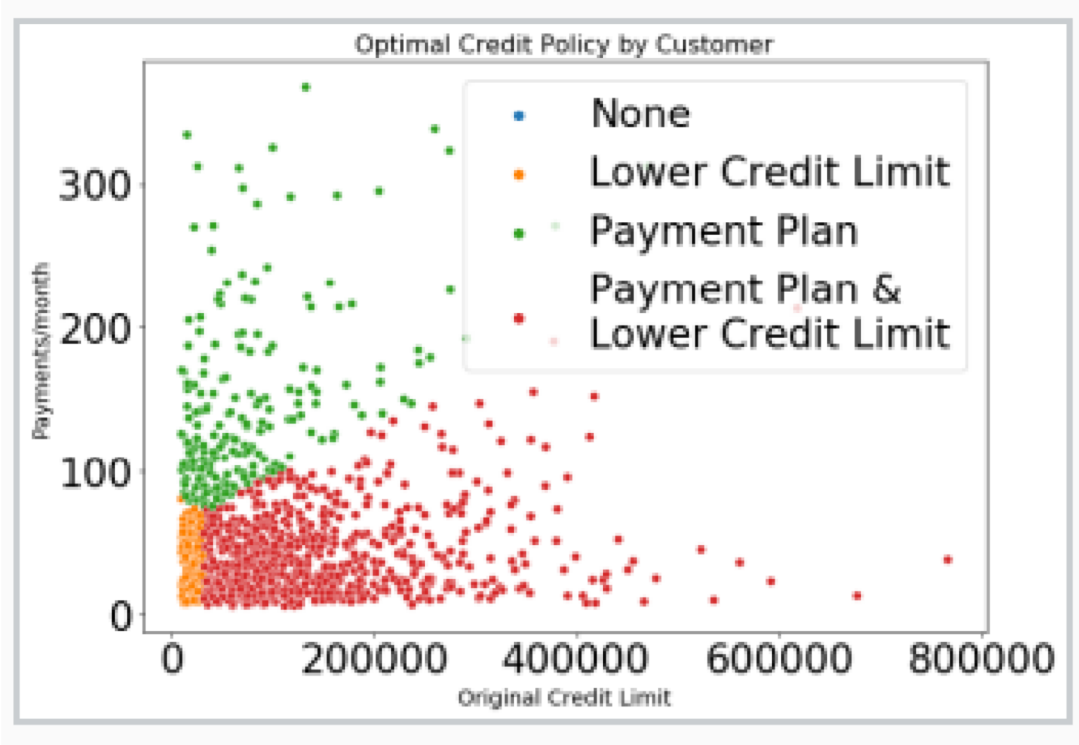
Colors: Recommended treatments per client



Understanding heterogeneous treatment effects

Brandenburgische Technische Universität Cottbus - Senftenberg

- ‘None’ (no treatment) is never recommended
- can deduce that all treatments are beneficial to customers



Fairness

Brandenburgische
Technische Universität
Cottbus - Senftenberg

What about fairness?

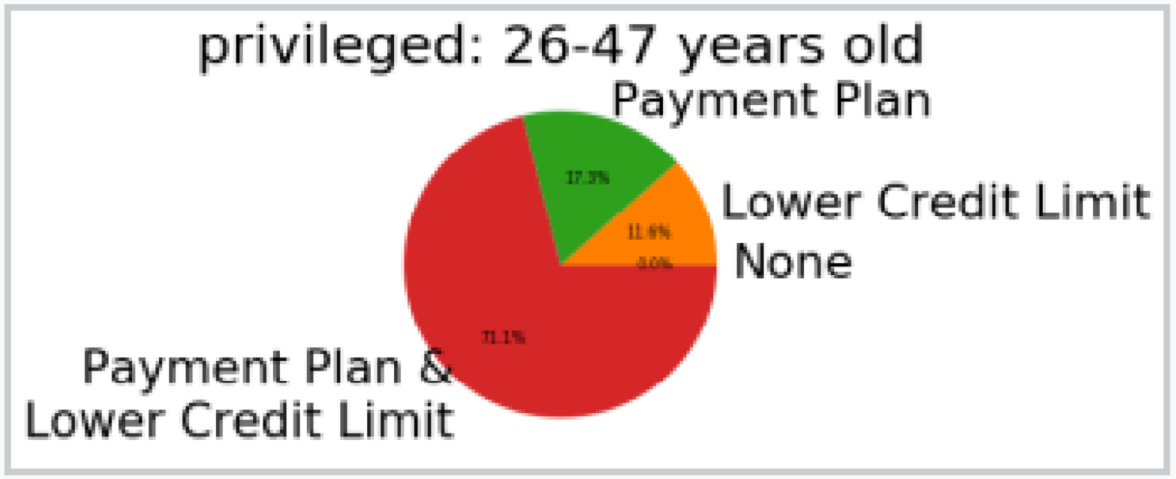
Fairness

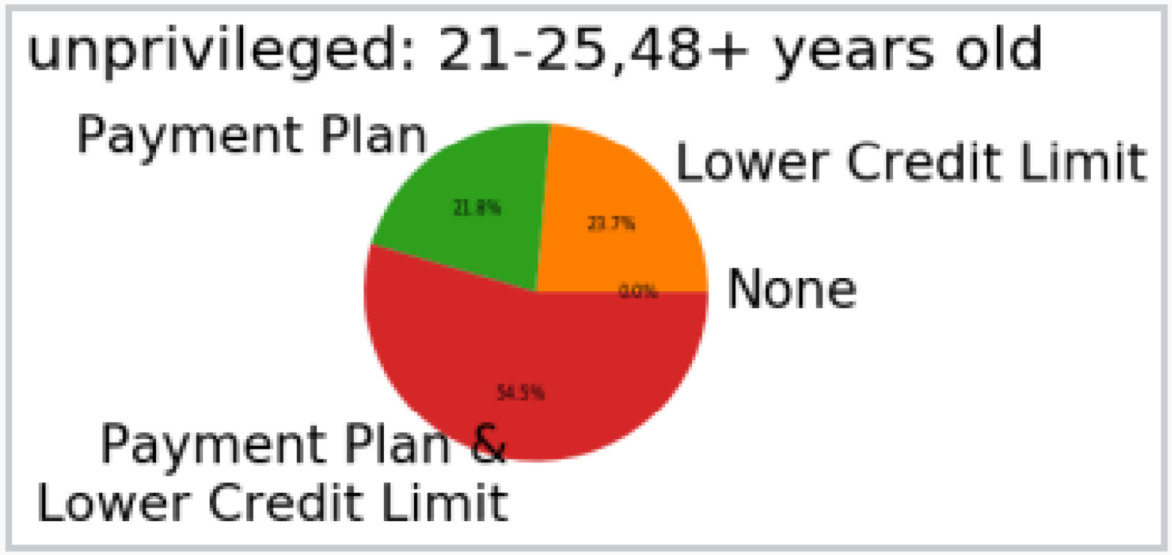
Brandenburgische
Technische Universität
Cottbus - Senftenberg

Are the treatments that are expensive and inexpensive for the bank **fair distributed** among **unprivileged** and **privileged** customers?

How can we determine that?

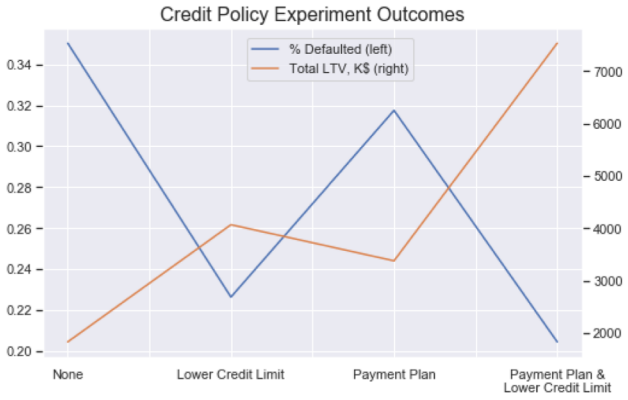
Fairness





Brandenburgische
Technische Universität
Cottbus - Senftenberg

The privileged clients get more better, expensive treatments (payment plan & lower limit)
-> higher probability for repay



Fairness

Brandenburgische
Technische Universität
Cottbus - Senftenberg

- This **disparity** is primarily due to the **bank's costs** being a factor, so if the bank were to absorb some of these costs, it could make it fairer.

- But **what** would be a **fair solution?**
- Choosing credit policies is an example of procedural fairness, and there are many possible definitions.
- Does equal treatment mean equal treatment or proportional treatment?
- What if a customer prefers one policy over another? Should they be allowed to switch?
- -> There are so many ways to go about it

Refute the obtained estimate using multiple robustness checks

Brandenburgische
Technische Universität
Cottbus - Senftenberg

How robust is our estimated causal effect?

Refute the obtained estimate using multiple robustness checks

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Robustness methods

(DoWhy libraries)

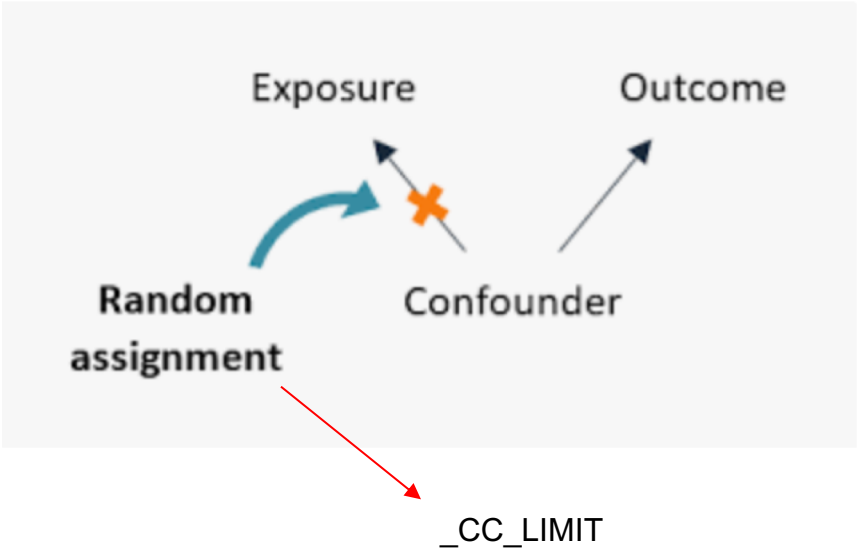
- Random confounder
- Placebo treatment refuter
- Data subset refuter

Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Random confounder (common cause)

- Adding a randomly generated confounder
- If the estimate is robust
- -> the **average treatment effect (ATE)** should not change too much.



Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Random confounder (common cause)

- If the estimate is robust
- -> the **average treatment effect (ATE)** should not change too much.

```
ref_random = causal_mdl.refute_estimate(method_name="random_common_cause")
print("ref_random\n",ref_random)
```

```
ref_random
 Refute: Add a random common cause
Estimated effect:7227.831176285025
New effect:7206.629398196633
p value:0.84
```

What does the p-value mean?

Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

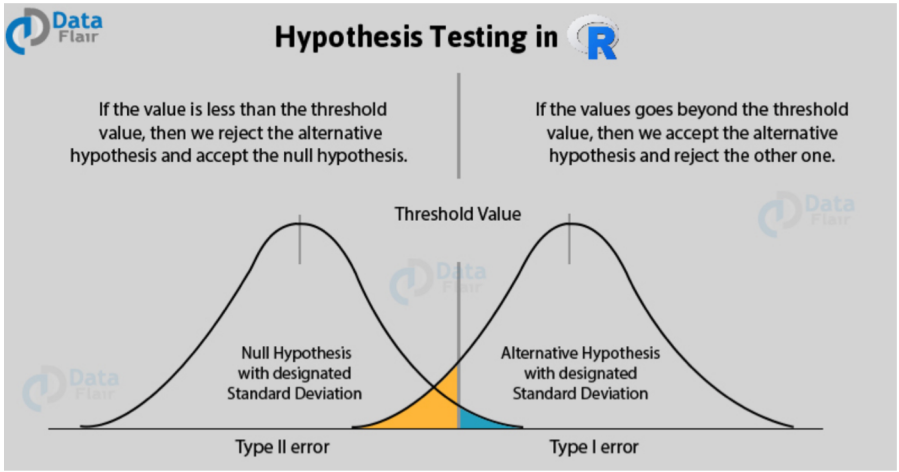
P-Value

What does the p-value mean?

- Compares the means of two groups
- most commonly used p-value is 0.05 (threshold)

Null hypothesis == TRUE:
Groups have no significant difference
p-value > 0.05

Null hypothesis is rejected (FALSE)
Groups have a significant difference
p-value <= 0.05



What does our p-value mean?

- Compares the average treatment effects
- -> robust

```
ref_random
 Refute: Add a random common cause
Estimated effect:7227.831176285025
New effect:7206.629398196633
p value:0.84
```

Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Placebo treatment refuter

- **Adding a placebo treatment**

- If the estimate is robust
- -> the ATE difference should be close to zero.

_TREATMENT[i] = random()

Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Placebo treatment refuter

- If the estimate is robust
- -> the ATE difference should be close to zero.

```
ref_placebo = causal_mdl.refute_estimate(\
                    method_name="placebo_treatment_refuter",\
                    placebo_type="permute", num_simulations=20)
print("ref_placebo\n",ref_placebo)
```

```
ref_placebo
 Refute: Use a Placebo Treatment
Estimated effect:7227.831176285025
New effect:-42.2256923470972
p value:0.4721052576185495
```

p-value > 0.05, -> not reject the null hypothesis
-> the estimated causal effect **is not very robust**
Possible Solution:
adding relevant confounders or by using a different causal model

Refute the obtained estimate using multiple robustness checks

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Data subset refuter

- Removing a random subset of the data

- If the estimator
- generalizes well
- -> the ATE should not change too much

Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Data subset refuter

- If the estimator
- generalizes well
- -> the ATE should not change too much

```
res_subset=causal_mdl.refute_estimate(method_name="data_subset_refuter", subset_fraction=0.9)
print("res_subset",res_subset)
```

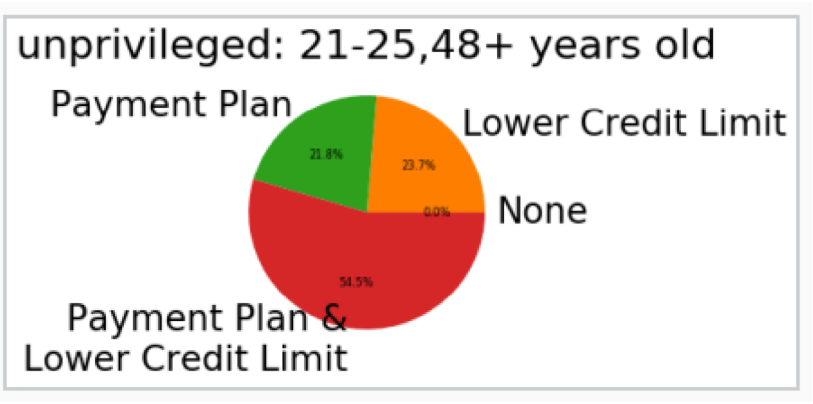
res_subset Refute: Use a subset of data
Estimated effect:7227.831176285025
New effect:7395.237570192716
p value:0.94

Estimator OK

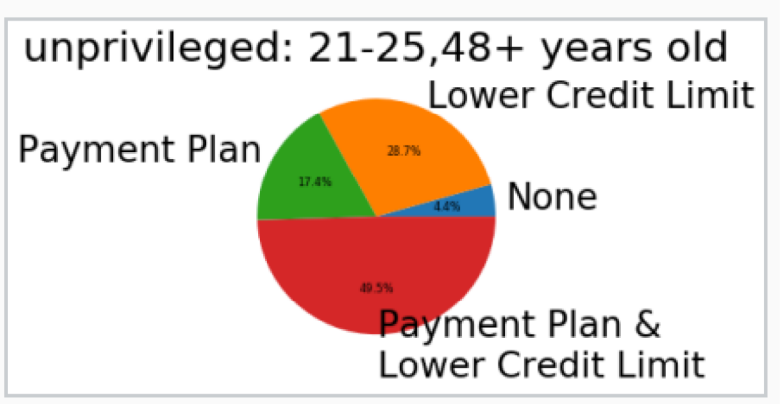
Refute the obtained estimate using multiple robustness checks

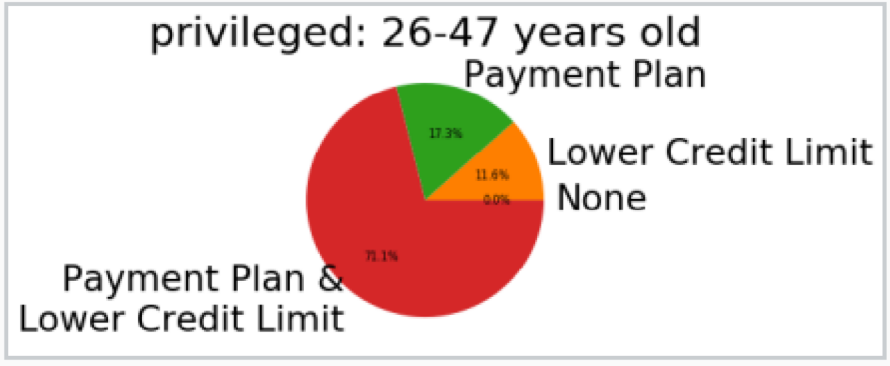
Brandenburgische Technische Universität Cottbus - Senftenberg

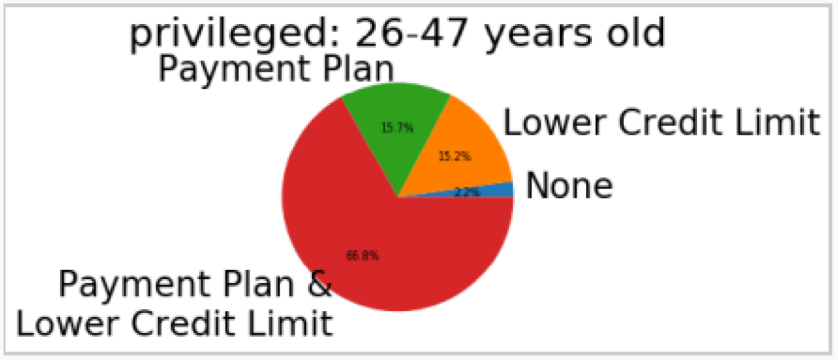
Does the model change, if we use AGE_GROUP as confounder?



Confounder: AGE_GROUP







Refute the obtained estimate using multiple robustness checks

Brandenburgische Technische Universität Cottbus - Senftenberg

Does the model change, if we use AGE_GROUP as confounder?

ref_random
Refute: Add a random common cause
Estimated effect:7227.831176285025
New effect:7206.629398196633
p value:0.84

ref_placebo
Refute: Use a Placebo Treatment
Estimated effect:7227.831176285025
New effect:-42.2256923470972
p value:0.4721052576185495

res_subset Refute: Use a subset of data
Estimated effect:7227.831176285025
New effect:7395.237570192716
p value:0.94

less robust with AGE_GROUP

Confounder: AGE_GROUP

ref_random
Refute: Add a random common cause
Estimated effect:8069.357551000912
New effect:7565.416905380992
p value:0.44

ref_placebo
Refute: Use a Placebo Treatment
Estimated effect:8069.357551000912
New effect:-42.01835579208754
p value:0.47588040115237695

res_subset Refute: Use a subset of data
Estimated effect:8069.357551000912
New effect:7942.358255012813
p value:0.78

Summery

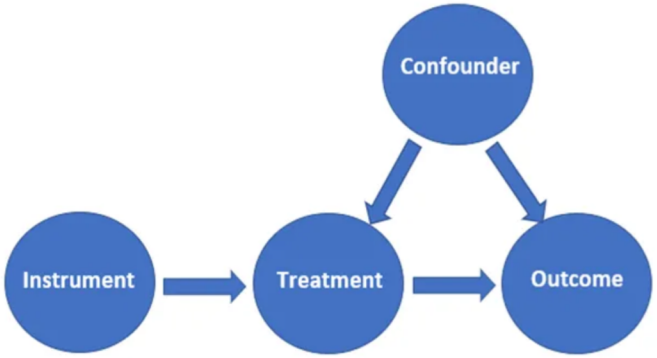
Brandenburgische Technische Universität

Method	Description
Prediction	• focus on prediction (what is the outcome)
Correlation	• two variables tend to change together • but it doesn't necessarily mean that changes in one variable cause changes in the other
Causation/ Causality/	• If one variable's changes • -> directly lead to changes in another
Causal inference	• **involves considering counterfactual** scenarios • hypothetical situations where one variable's value is changed • while keeping other variables constant aims to uncover the mechanisms underlying the data generation process **forecasting** future outcomes **+ understanding why and how** things happen

Summery

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Method	Description
Causal inference	1. Create a **causal model** from the data and given graph **2. Identify causal effect** and return target estimands **3. Estimate the target estimand** using a statistical method 4. Refute the obtained estimate using **multiple robustness checks**



Summery

Robustness methods

Brandenburgische
Technische Universität
Cottbus - Senftenberg

Method	Action	impact
Data subset refuter	Removes subsets	Less strong
Random confounder	Changes confounder randomly	Less strong
Placebo treatment refuter	New treatment values	strong