Powered by Plain Concepts

#STechDay2021

SINGULARITY TECH DAY_2021

The era of AI and Cognitive Services

The hitchhiker's guide to your users

plain concepts

databricks

ORGANIZATION

SPONSORS

THANK YOU!

What makes a great product?

Someone may think that a product …

- fulfills a real, deeply-felt need
- delivers a singular value proposition
- is simple and intuitive
- has a great design → Craftspersonship
- uses the correct technology

Where is the user?

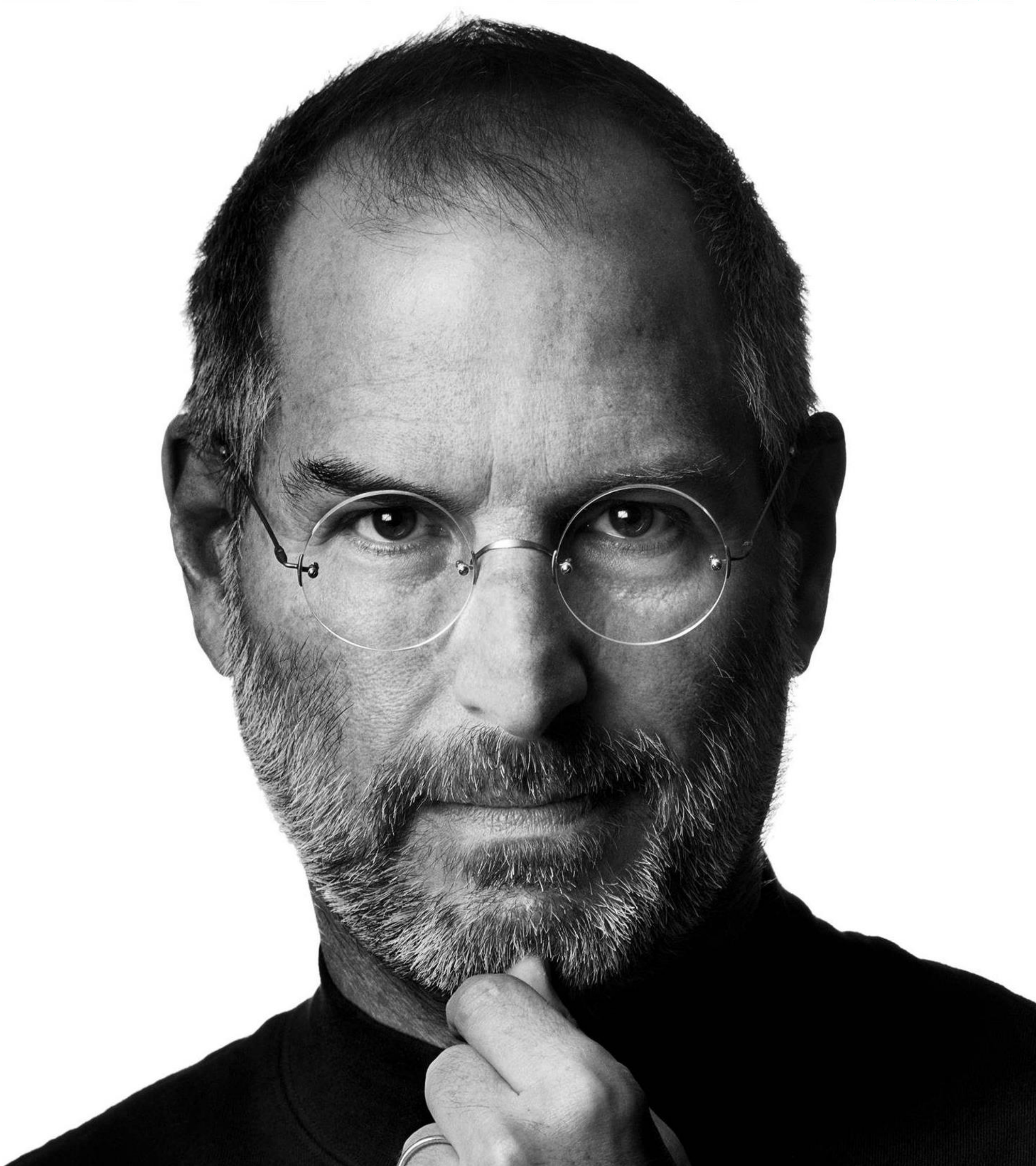
“

Customers don’t know what they want until we’ve shown them.

“

You've got to start with the customer experience and work backwards to the technology. You can't start with the technology and try to figure out where can I sell it.

Steve Jobs




WHAT
NOW?

How can I become customer centric?

- Being in customers shoes
  - DILO → **D**ay **I**n **L**ife **O**f
- Try, test, innovate
- Create new necessities
- Personalization over generalization
- Understanding customer psychology


Causality

Eduardo Matallanas

Head of AI @ Plain Concepts

Knowmad interested in:

- Changing live through AI
- Passionate in robotics
- Data lover
- Films & series
- Bike enthusiast & martial artist practitioner

@ematde

eduardo.matallanas@plainconcepts.com

Things we will talk today

- Customers, customers and more customers
- Causality
- How to combine AI+causality
- Some Statistics

Attention: Lots of questions with no answer 😜

How can I become user-centric?

- Data is needed
  - Monitor your users
  - Measure user behavior
- Use of AI/ML to
  - Find data correlation
  - Extract features form behavior

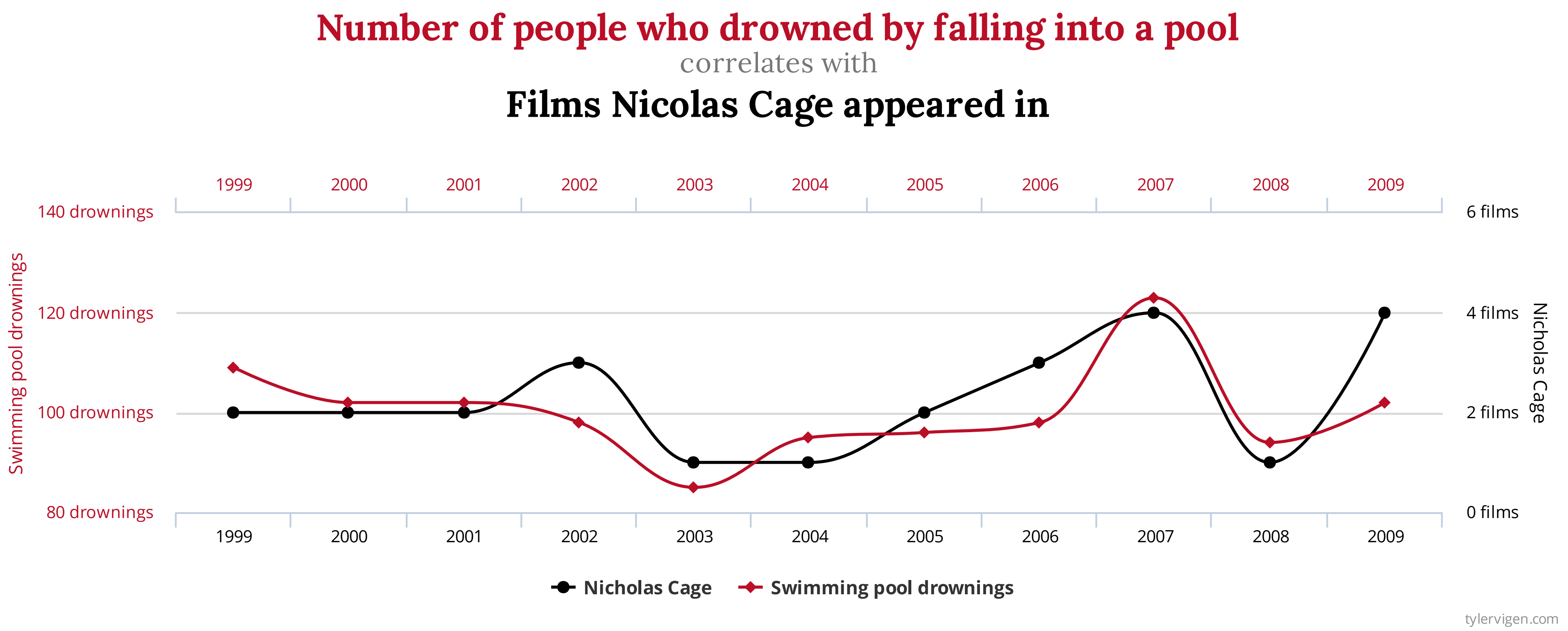
Why really users select our product?

Causality is the answer

Why do we need causality when we already have AI?

Why do we need causality when we already have AI?

- Not everything is correlation

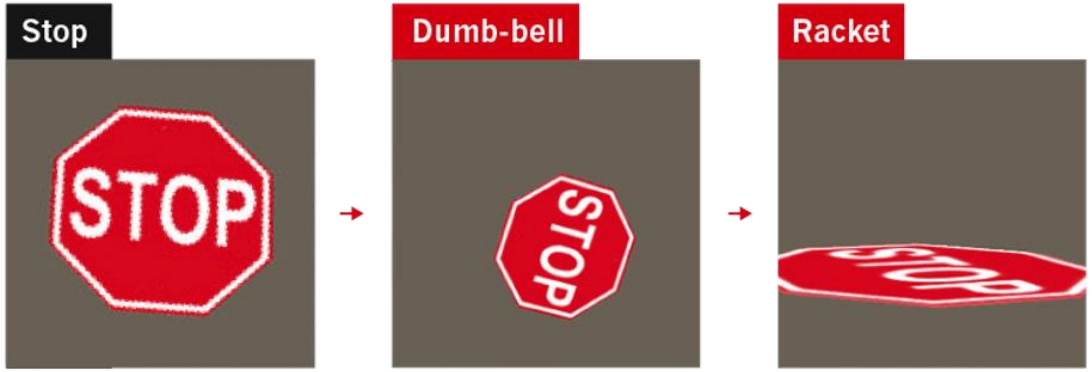


Why do we need causality when we already have AI?

- Not everything is correlation
    - AI often finds spurious ones

Why do we need causality when we already have AI?

- Not everything is correlation
  - AI often finds spurious ones
- Implies generalization problems



Incorrect predictions under changes in data

[Alcorn et al. CVPR 2019]

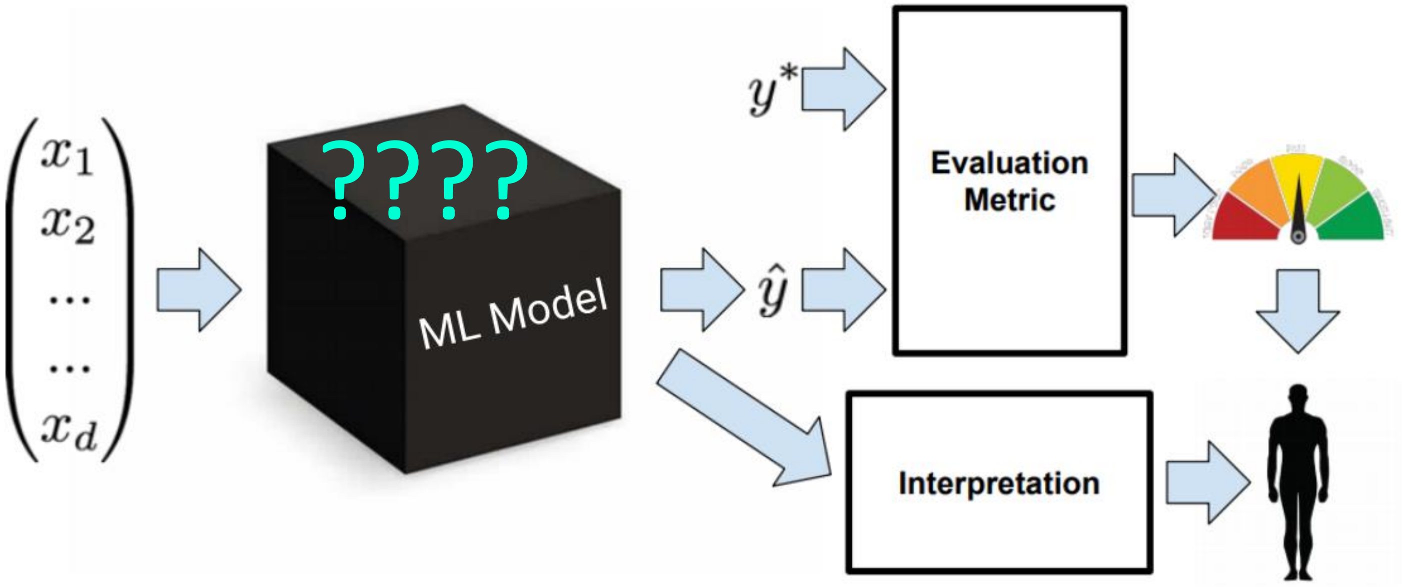
What color is the tray?	Pink
What colour is the tray?	Green
Which color is the tray?	Green
What color is it?	Green
How color is tray?	Green

Fooled by semantically equivalent perturbations

[Ribeiro et al. ACL 2018]

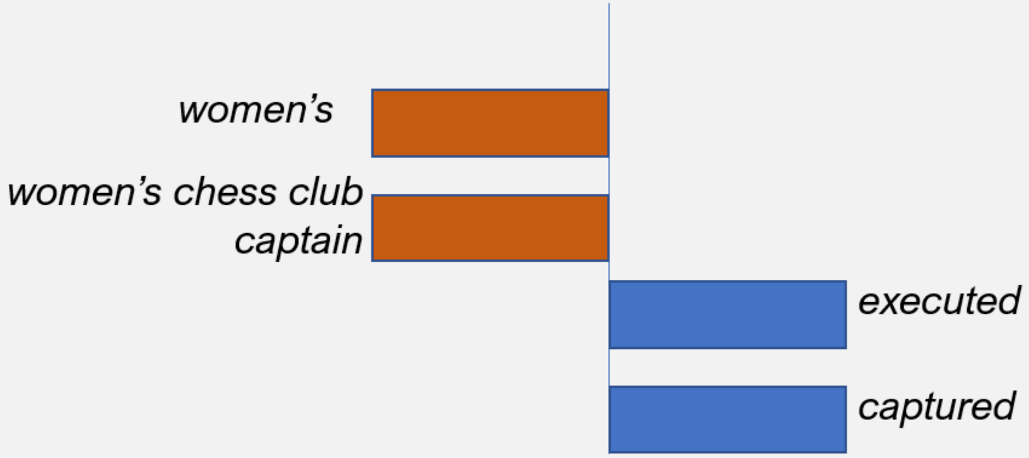
Why do we need causality when we already have AI?

- Not everything is correlation
  - AI often finds spurious ones
- Implies generalization problems
- Not too explainable



Why do we need causality when we already have AI?

- Not everything is correlation
  - AI often finds spurious ones
- Implies generalization problems
- Not too explainable
- Fairness/Non-discrimination
  - Some correlations shift data distributions

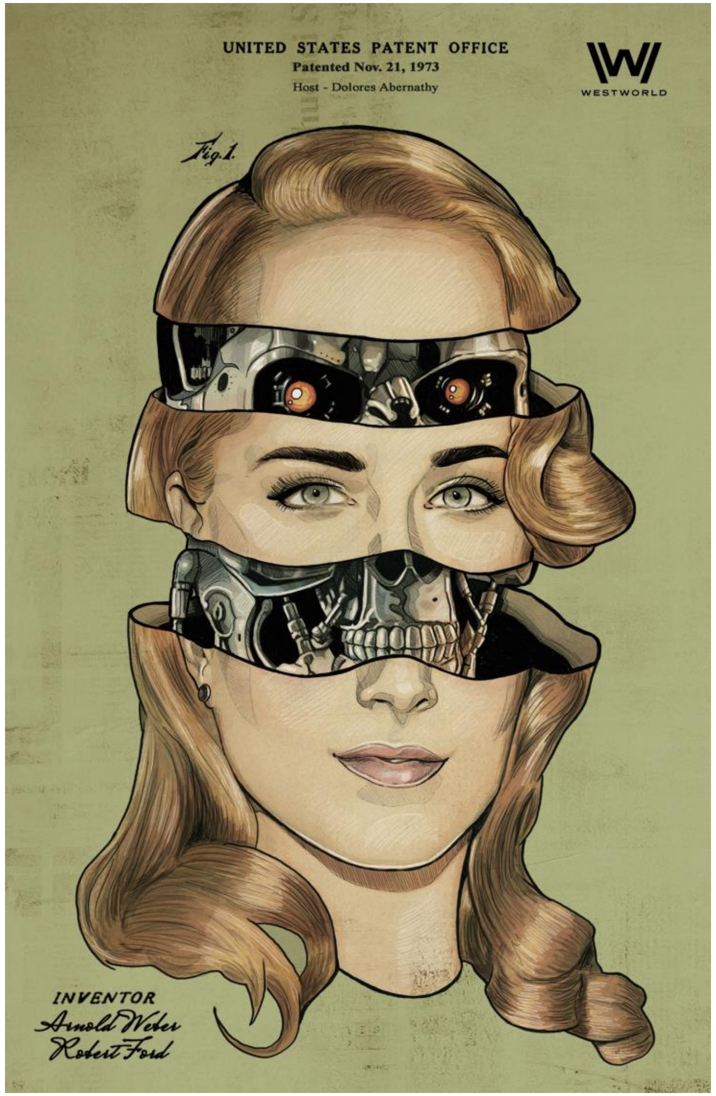


Bias in ML model for hiring decisions

[Reuters 2018, Weblink]

Why do we need causality when we already have AI?

- Not everything is correlation
  - AI often finds spurious ones
- Implies generalization problems
- Not too explainable
- Fairness/Non-discrimination
  - Some correlations shift data distributions
- Decision-making → "Human-like" AI



What is causality?

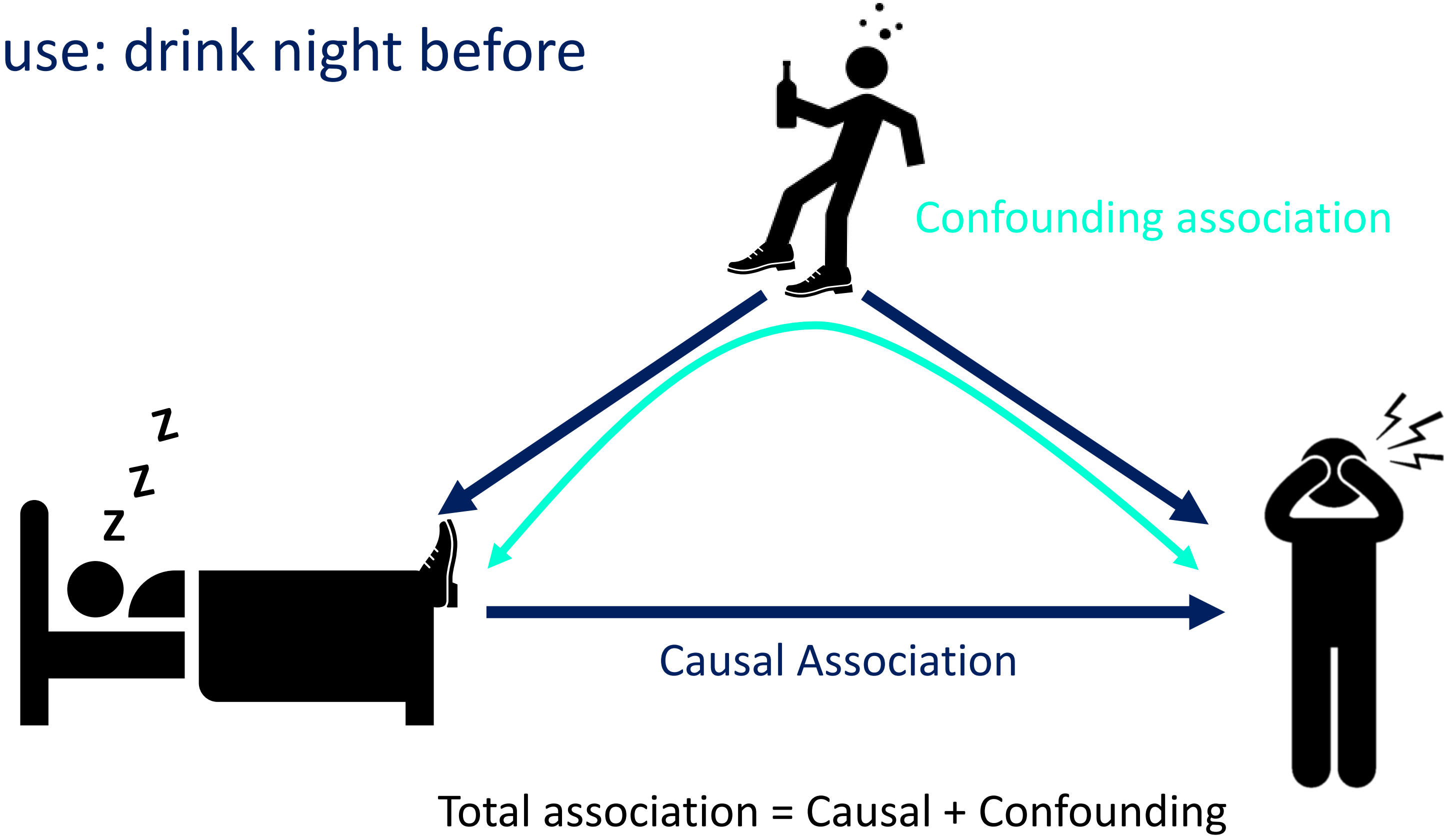
“

There is only one constant. One universal. It is the only real truth. Causality. Action, reaction. Cause and effect.

The Merovingian

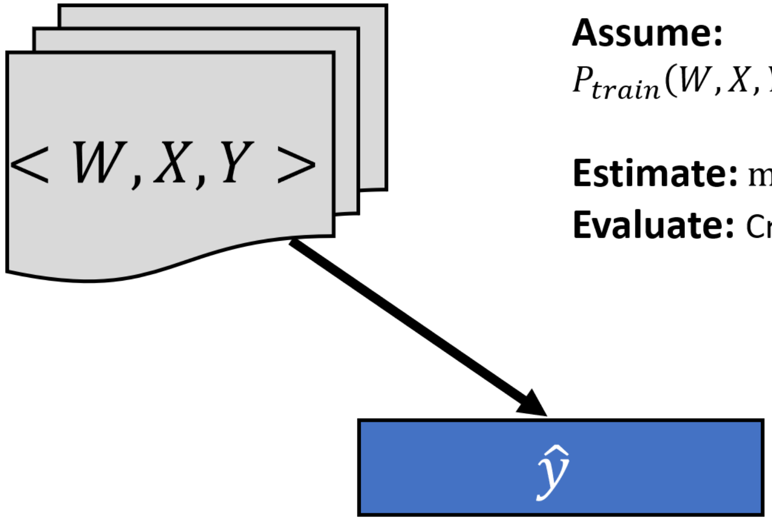
What is causation?

- Inferring the effect of a treatment/policy T in some outcome Y
- Example: Sleeping with shoes causes waking up with headache.
  - Common cause: drink night before



Difference between AI and Causal inference

Supervised Machine Learning

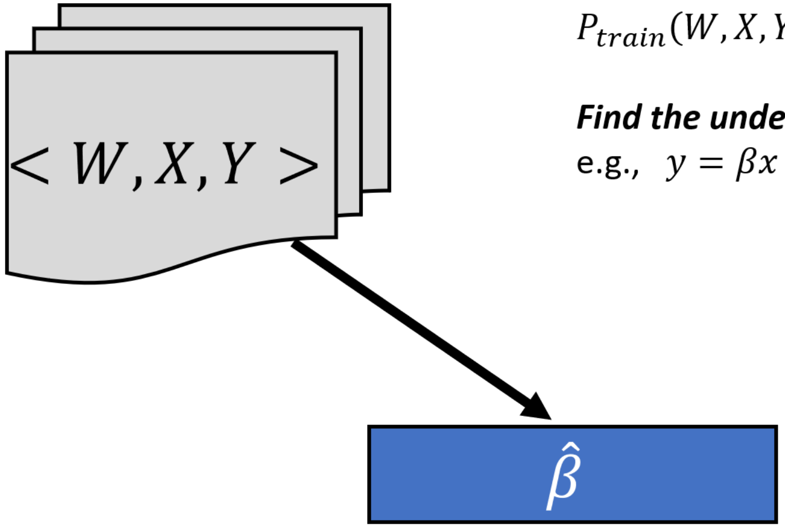


Prediction

Assume:
$P_{train}(W, X, Y) = P_{test}(W, X, Y)$

Estimate: $\min L(\hat{y}, y)$
Evaluate: Cross-validation

Causal Inference



Decision-Making

$P_{train}(W, X, Y) \neq P_{test}(W, X, Y)$

Find the underlying generative model
e.g., $y = \beta x + f(w) + \epsilon$

What is causation?

- Formally, T causes Y iff changing T leads to a change in Y, keeping everything else constant

Real World: **do(T=1)**

Counterfactual World: **do(T=0)**

Causal effect is the magnitude by which Y is changed by a unit change in T:

$$E[Y|do(T=1)] - E[Y|do(T=0)]$$

Ladder of Causation

3 IMAGINING

COUNTERFACTUALS

Activity: Imagining, Retrospection, Understanding

Questions: What if I had done...? Why? (Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

Examples: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2 DOING

INTERVENTION

Activity: Doing, Intervening

Questions: What if I do...? How? (What would Y be if I do X? How can I make Y happen?)

Examples: If I take aspirin, will my headache be cured? What if we ban cigarettes?

1 SEEING

ASSOCIATION

Activity: Seeing, Observing

Questions: What if I see...? (How are the variables related? How would seeing X change my belief in Y?)

Examples: What does a symptom tell me about a disease? What does a survey tell us about the election results?

Adapted from Judea Pearl, "The Book of Why: The New Science of Cause and Effect," chapter 1, page

Don’t panic

Let’s continue with an example

What makes great a reservation?

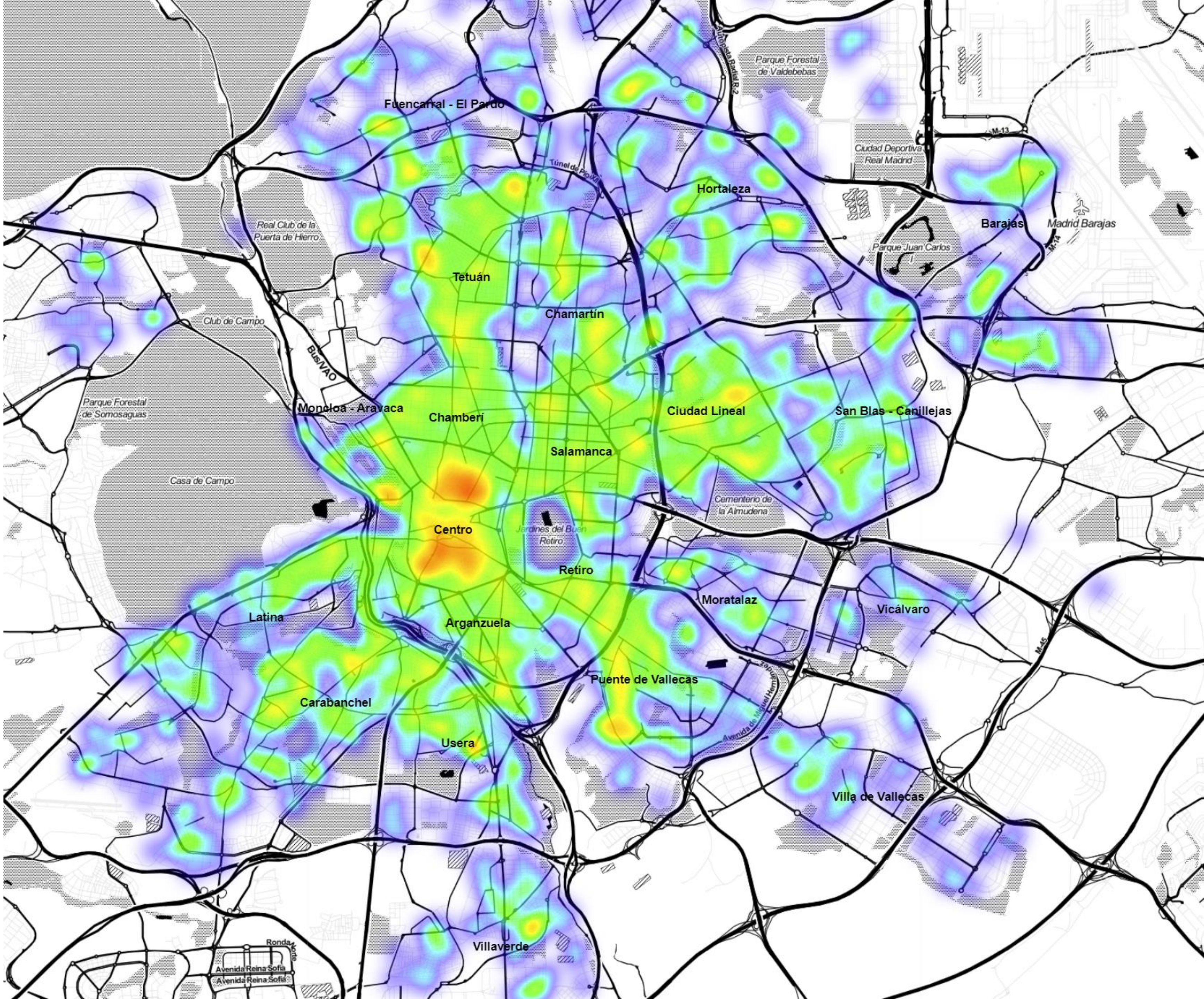
Example data

- Data characteristics:
  - Location → Madrid
  - Historic listings until 7^{th} Nov 2021
  - User reviews
- Scrapped data from Airbnb Website
  - Inside Airbnb

SINGULARITY TECH DAY_2021

Assumptions

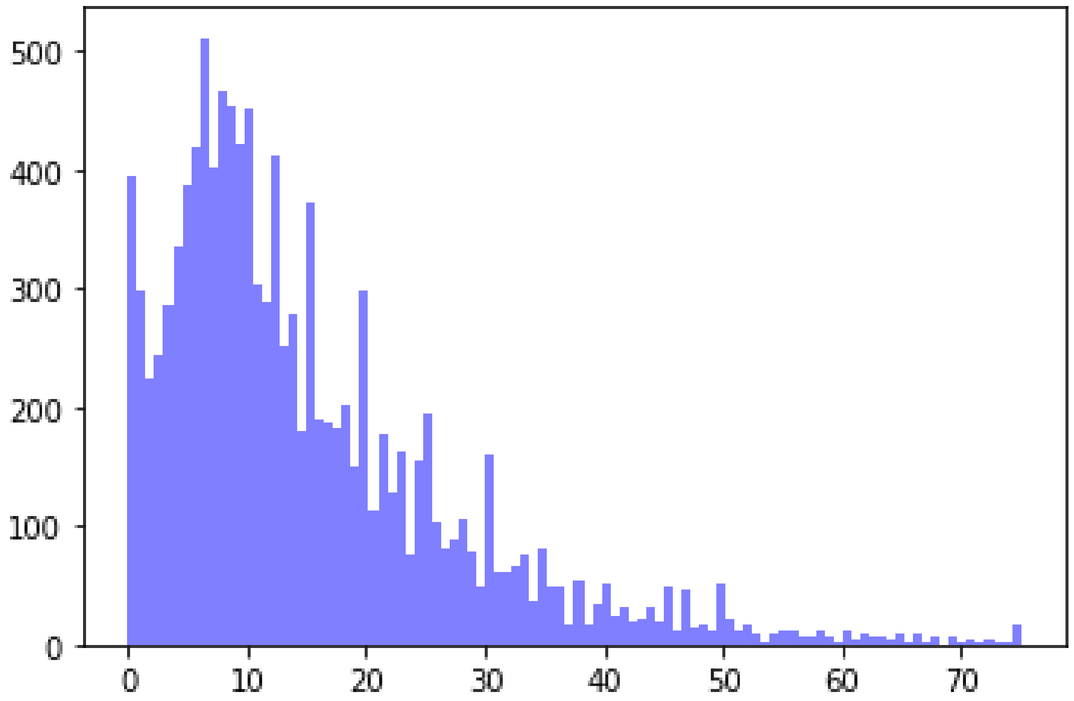
- Location



Assumptions

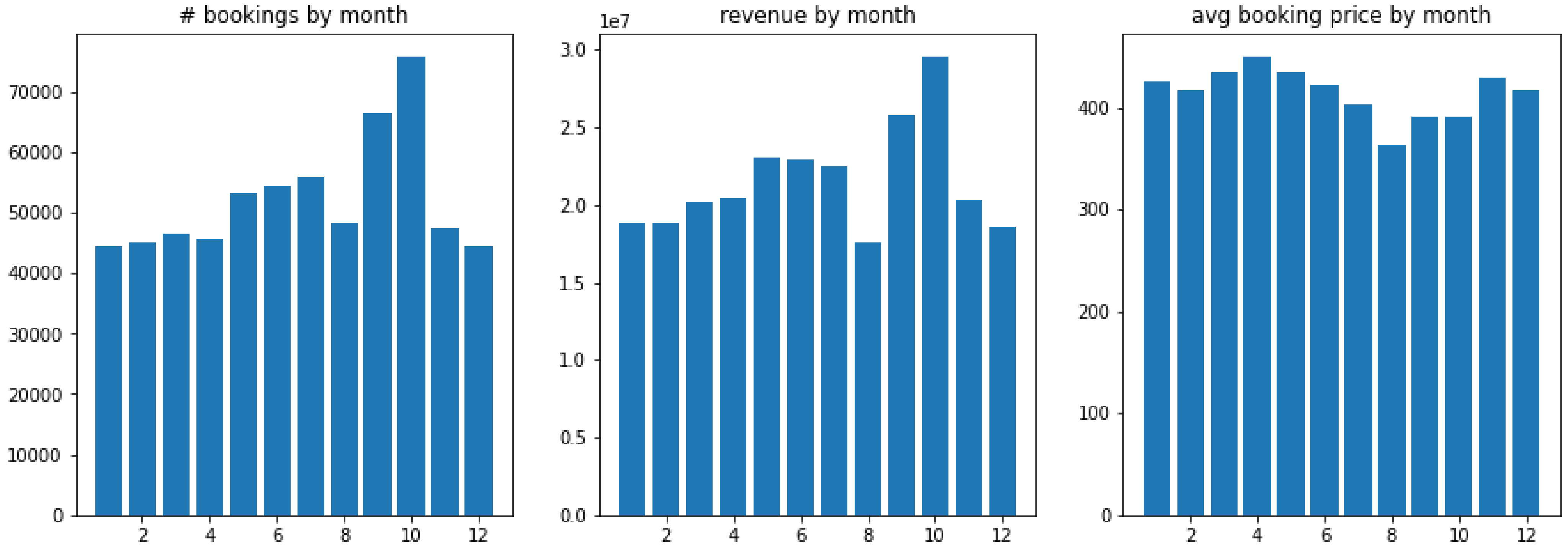
- Location
- Price

Madrid Price per person per night distribution



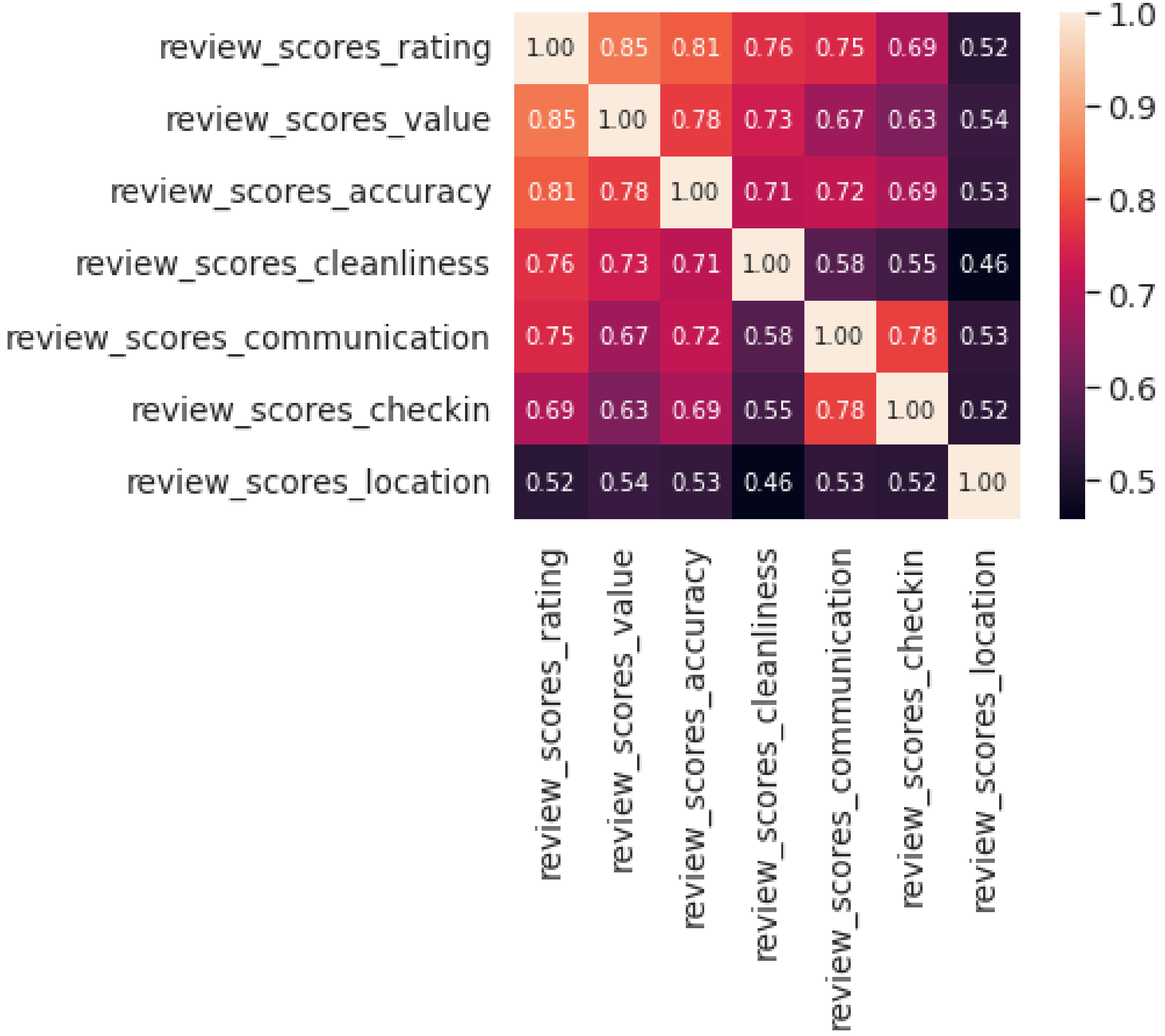
Assumptions

- Location
- Price
- Estimated Revenue



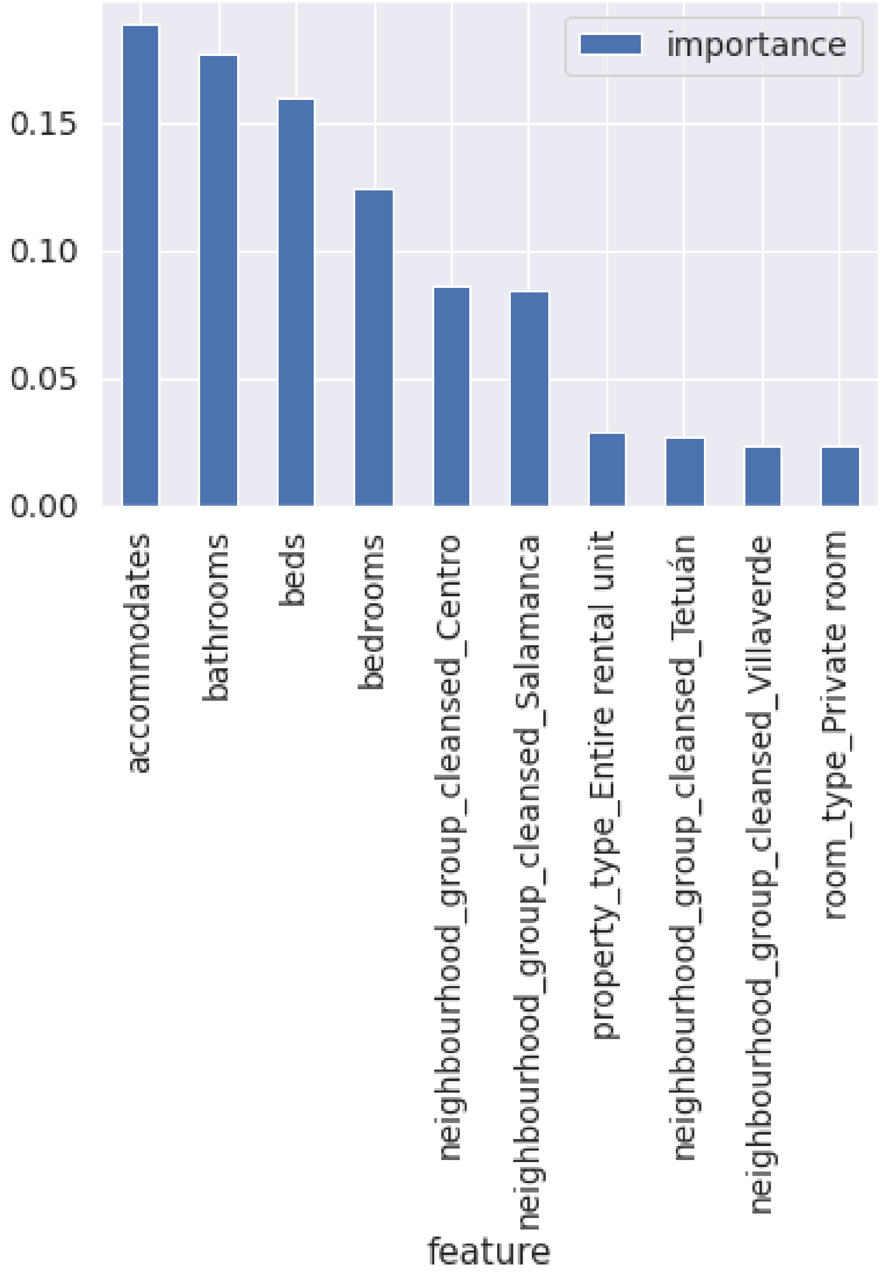
Assumptions

- Location
- Price
- Estimated Revenue
- Review scoring



Assumptions

- Location
- Price
- Estimated Revenue
- Review scoring
- Accommodation features
  - #beds, #bathrooms, #accommodates, etc.



Assumptions

- Location
- Price
- Estimated Revenue
- Review scoring
- Accommodation features
    - #beds, #bathrooms, #accommodates, etc.
- Host info
    - Verified, picture, >1 accommodation, etc.

Assumptions

- Location
- Price
- Estimated Revenue
- Review scoring
- Accommodation features
  - #beds, #bathrooms, #accommodates, etc.
- Host info
  - Verified, picture, >1 accommodation, etc.
- Other features
  - Review f, #reviews
  - Availability, instant bookable

What can I use to model causation?

Microsoft

Discover DoWhy

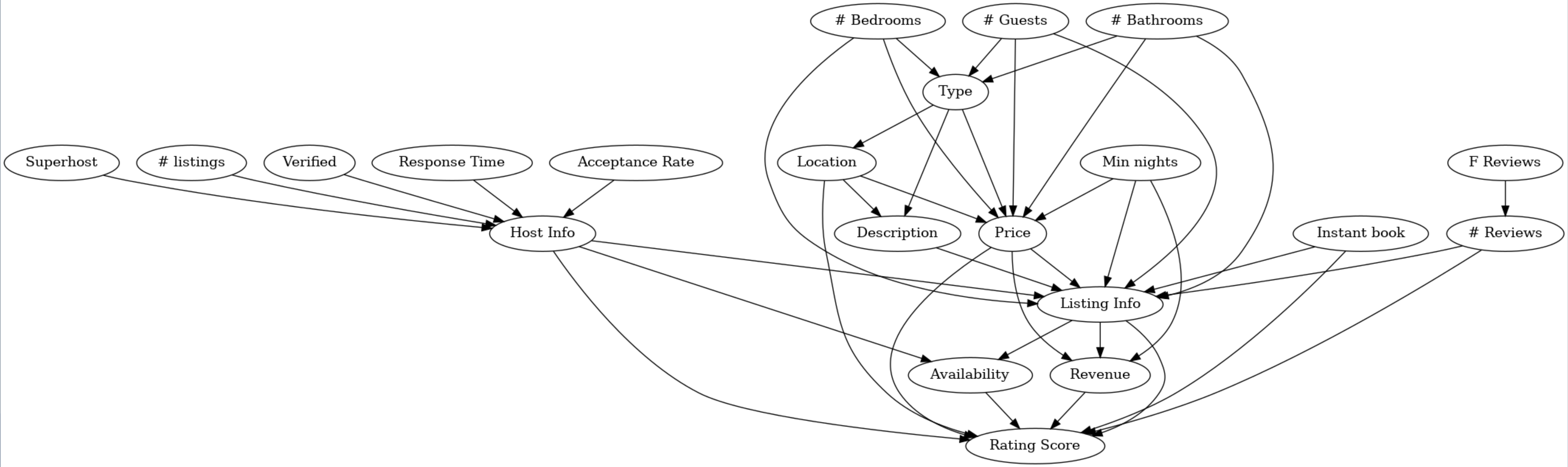
A software library for causal inference

DoWhy framework

1	2	3	4
Model	**Identify**	**Estimate**	**Refute**
Create a causal graph to encode assumptions	Formulate desired quantity from given relationships	Compute the causal effect on the data	Validate your assumptions

Step 1: Create the Causal Model

```
model= dowhy.CausalModel(
        data = pd_listing_causal,
        graph=causal_graph.replace("\n", " "),
        treatment='high_rating',
        outcome='review_scores_rating')
model.view_model()
from IPython.display import Image, display
display(Image(filename="causal_model.png"))
```



Step 2: Identify the cause

```
#Identify the causal effect
identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)
print(identified_estimand)


Estimand type: nonparametric-ate

### Estimand : 1
Estimand name: backdoor
Estimand expression:
     d
────────────── (Expectation(review_scores_rating|Assumptions))
d[high_rating]

Estimand assumption 1, Unconfoundedness: If U→{high_rating} and U→review_scores_rating then
P(review_scores_rating|Assumptions) = P(review_scores_rating|Assumptions)

### Estimand : 2
Estimand name: iv
No such variable found!

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

Step 2: Identify the cause

```
#Identify the causal effect
identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)
print(identified_estimand)
```

review_scores_rating

depends of

- Price
- Location
- Listing Info
    - Description
    - # bathrooms
    - # bedrooms
    - # accommodates
    - Type
- Host Info
    - Response Time
    - Acceptance rate
    - Verified Identity
    - Is superhost
- Number of Reviews
- Reviews per month
- Instant bookable
- Minimum nights

Step 3: Estimate identified cause

```
estimate = model.estimate_effect(identified_estimand,
                                 method_name="backdoor.propensity_score_stratification",target_units="ate")
print(estimate)
```

*** Causal Estimate ***

Identified estimand

Estimand type: nonparametric-ate

Estimand : 1

Estimand name: backdoor
Estimand expression:

```
  d
──────────────(Expectation(review_scores_rating|Assumptions))
d[high_rating]
```

Estimate Mean value: 0.6019096196672494

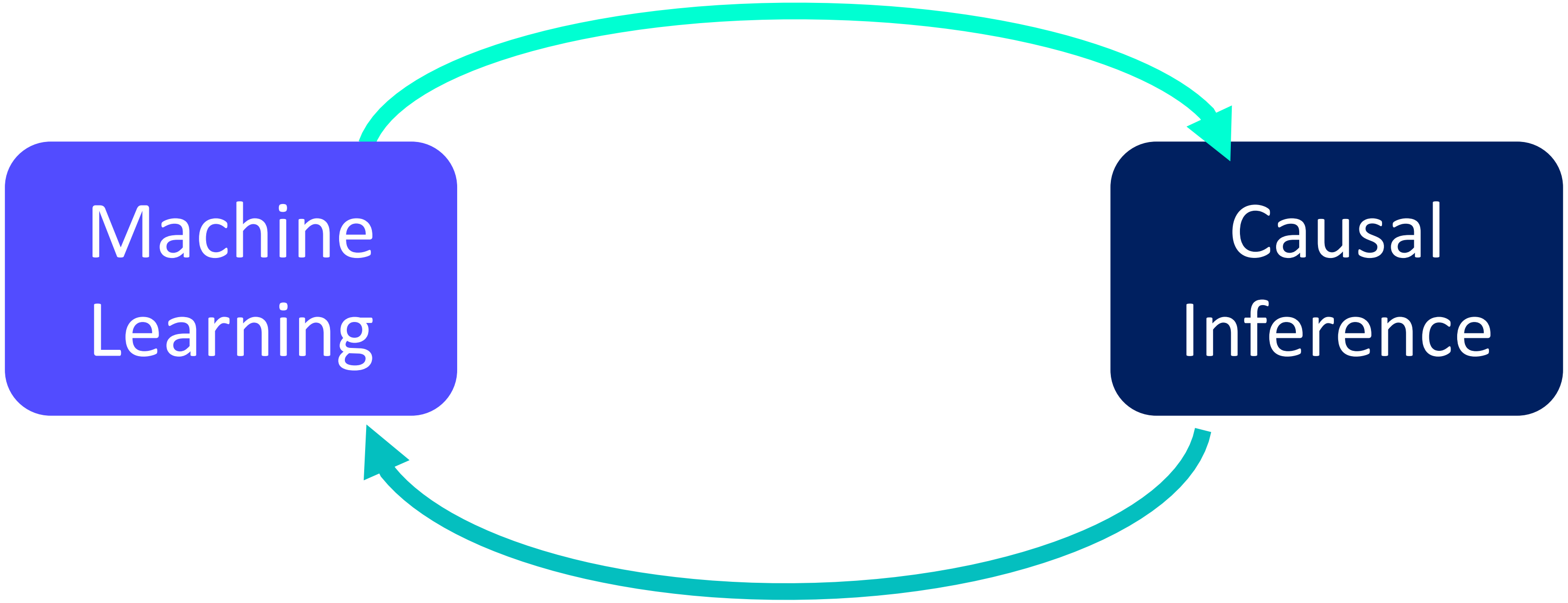
Step 4: Refute obtained results

```
refute_results=model.refute_estimate(identified_estimand, estimate,
        method_name="random_common_cause")
print(refute_results)
```

Refute: Add a Random Common Cause
Estimated effect: 0.6019096196672494
New effect: 0.6035238148349483

ML 🤝 Causal Inference

- Better estimations
- Refutation of causal effect



- ↑Generalization and robustness
- Beginning for fairness and explanation

Conclusion

- Causal inference is key for a strong AI approach
  - Tranferability, generalization, fairness, explanability
- Causal + ML is the beginning of a beautiful friendship
  - Opens heterogeneous Treatment effects → Uplifting
- Challenges:
  1. Counterfactual world cannot be observed → estimate effects and challenges in validation
  2. Data alone is not enough → multiple model representation, need assumptions and domain knowledge

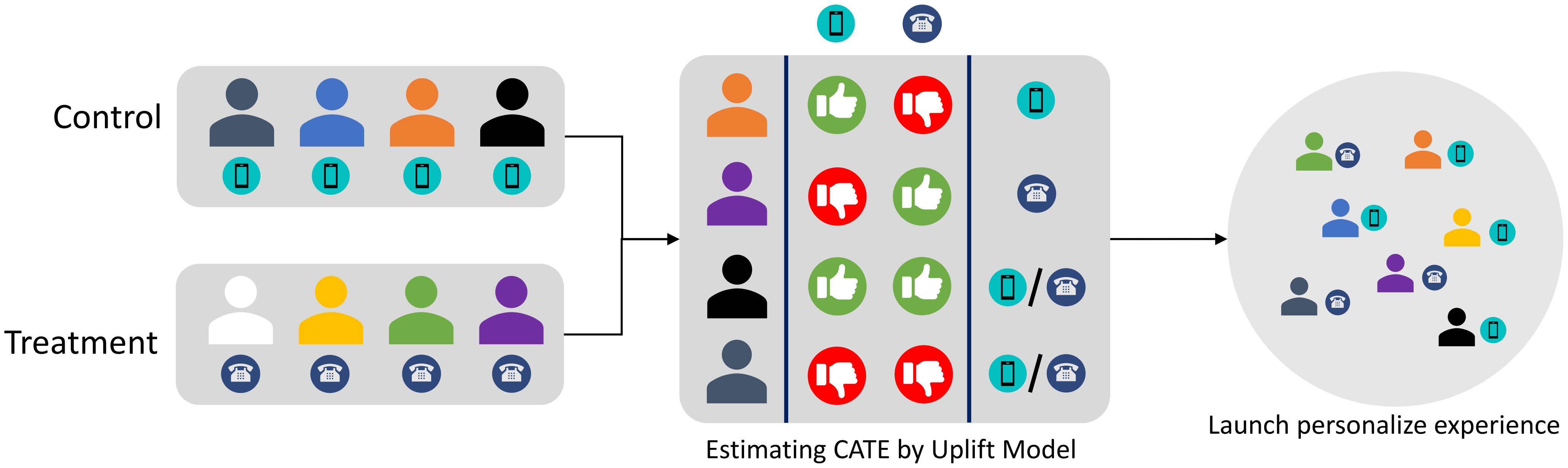
One more thing…

How to apply personalization?

- Uplifting modeling optimizes for incremental effect

$$\Delta(P(Behavior \mid Intervention), P(Behavior \mid No\ Intervention))$$

- Uplift modeling enables personalized treatment



SINGULARITY
TECH DAY_2021

Thank you!

databricks

plain concepts