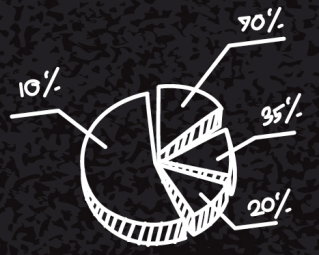
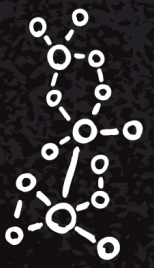




Bigger is Better, but at What Cost?

Towards Understanding the Economic Value of Data

$a^2 + b^2 = c^2$

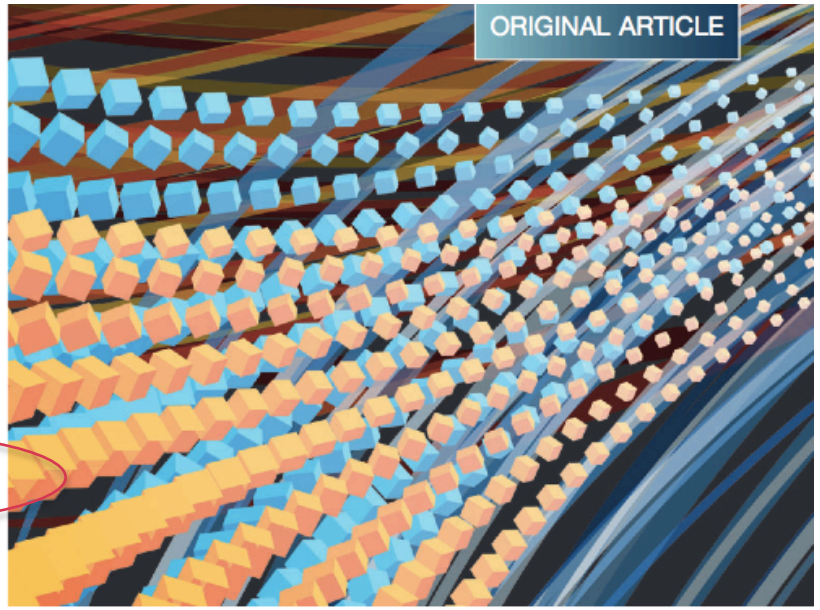
@delbrians

We are often told that bigger is better.

PREDICTIVE MODELING WITH BIG DATA:

Is Bigger Really Better?

Enric Junqué de Fortuny,[†] David Martens,[†] and Foster Provost[‡]



The answer is usually 'Yes'

but….

“With Great Power Comes Great Responsibility”

- Peter Parker

Big Data =
More Statistical Power

$$\pi(\tau) \approx 1 - \Phi(1.64 - \tau\sqrt{n}/\hat{\sigma}_D).$$

And yes
More responsibility

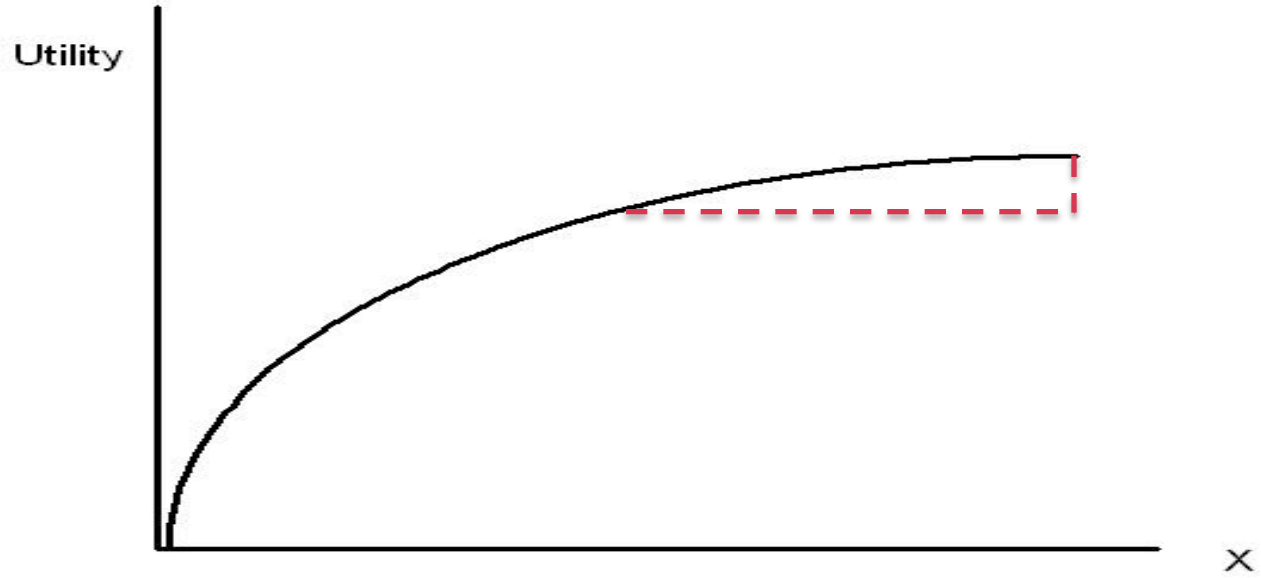
#1 responsibility as a manager/executive

Is that extra statistical power worth it?

Typical big data trade-off

Doubling the data often doubles the investment, but doesn't double the performance.

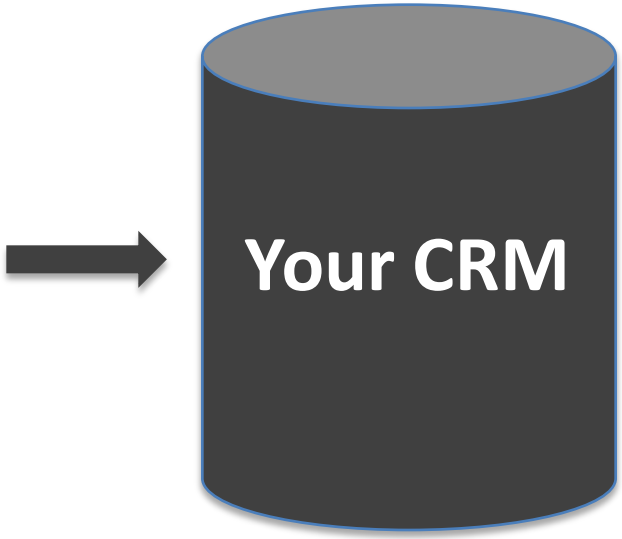
Understanding this tradeoff is key to understanding the economic value of data.



Common Data Sources

Data Vendors ($$$)

LiveRamp
ACXIOM
AddThis
Experian
exelate
bluekai
bizo



Customers(CDB*)

*Cost of Doing Business

2 Follow Up Questions

1. How much should you pay for this data?

2. How much is a customer's data worth?

In Important Inequality

The value of a customer ≠ **The value of a customer's data**

i.e., value(customer data) ≠ $\frac{\text{Total Revenue}}{\text{Total Customers}}$

An Import Equality

E[Value of Application | with Data]

– E[Value of Application | w/o Data]

Value of Data

An Import Equality

$$\text{E[Value of Application } \mid \text{ with Data]}$$

$$- \ \underline{\text{E[Value of Application } \mid \text{ w/o Data]}}$$

$$\text{Value of Data}$$

This equation offers important lessons about the value of data!

Lesson 1: *Data Has No Intrinsic Value*

The VOD is tied to applications/actions derived from data.

Show me an ad,
Recommend a product,
Give me a loan etc.

Caveat 1: *What is an outcome worth?*

With an application defined, need to know the value of various outcomes.

Lets assume a click is worth:

Caveat 2: *Application Scale Matters.*

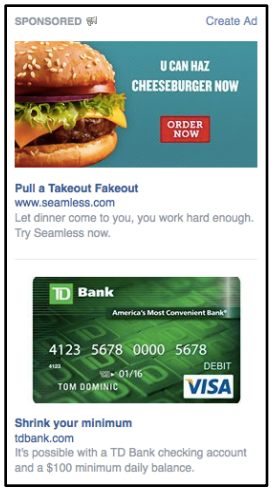
If data is free to duplicate, VOD depends on scale too.

Lesson 2: *The VOD is Counterfactual*

Data has $ value if using it generates more $ than a baseline strategy.

Without my data, you show this ad.

P(Click)=1%



VS.

With my data, you show this ad.

P(Click)=5%

The VOD Calculus

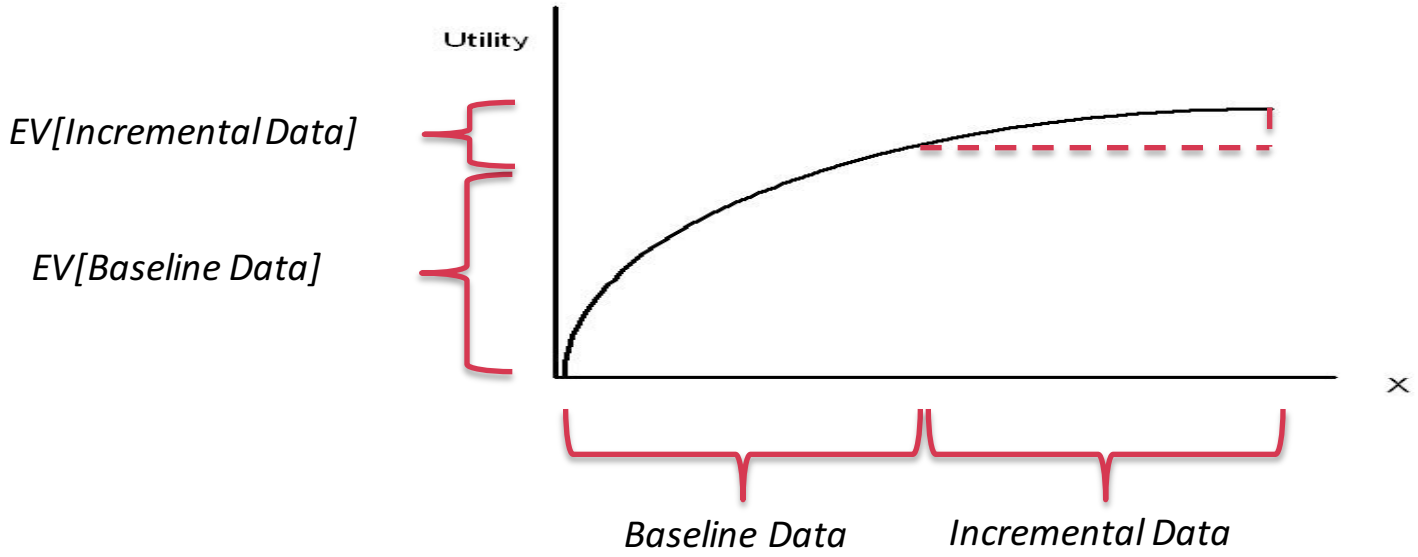
Let's put our equation to work.

$$V(\text{Click}) * [\, E(\text{Click} \mid \text{Data, Targeted Ad}) - E(\text{Click} \mid \text{noData, Default Ad}) \,]$$

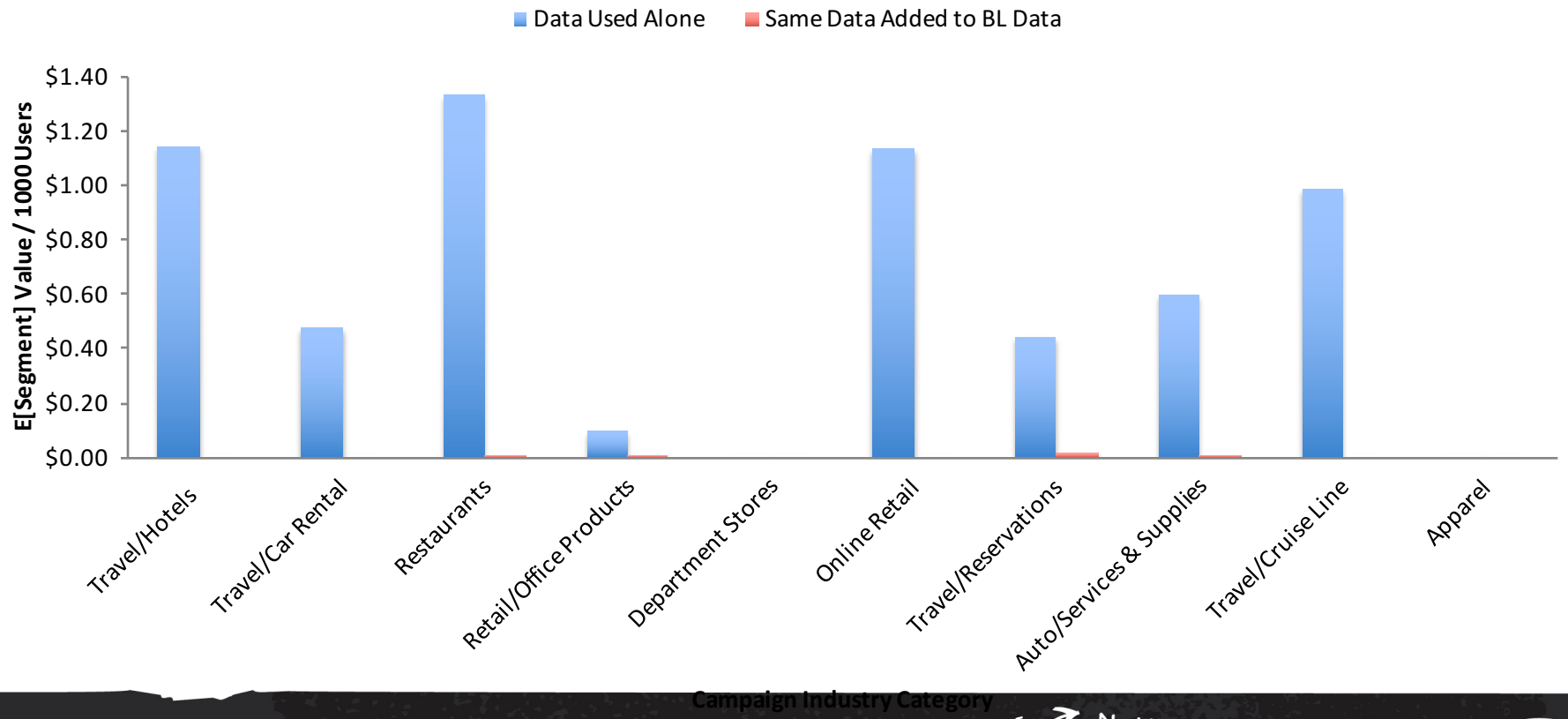
$$EV(\text{my Data in Targeting Ad}) = \$1 * [\, 5\% - 1\% \,] = \boxed{\$0.04}$$

Lesson 3: *The VOD is Relative, not Absolute.*

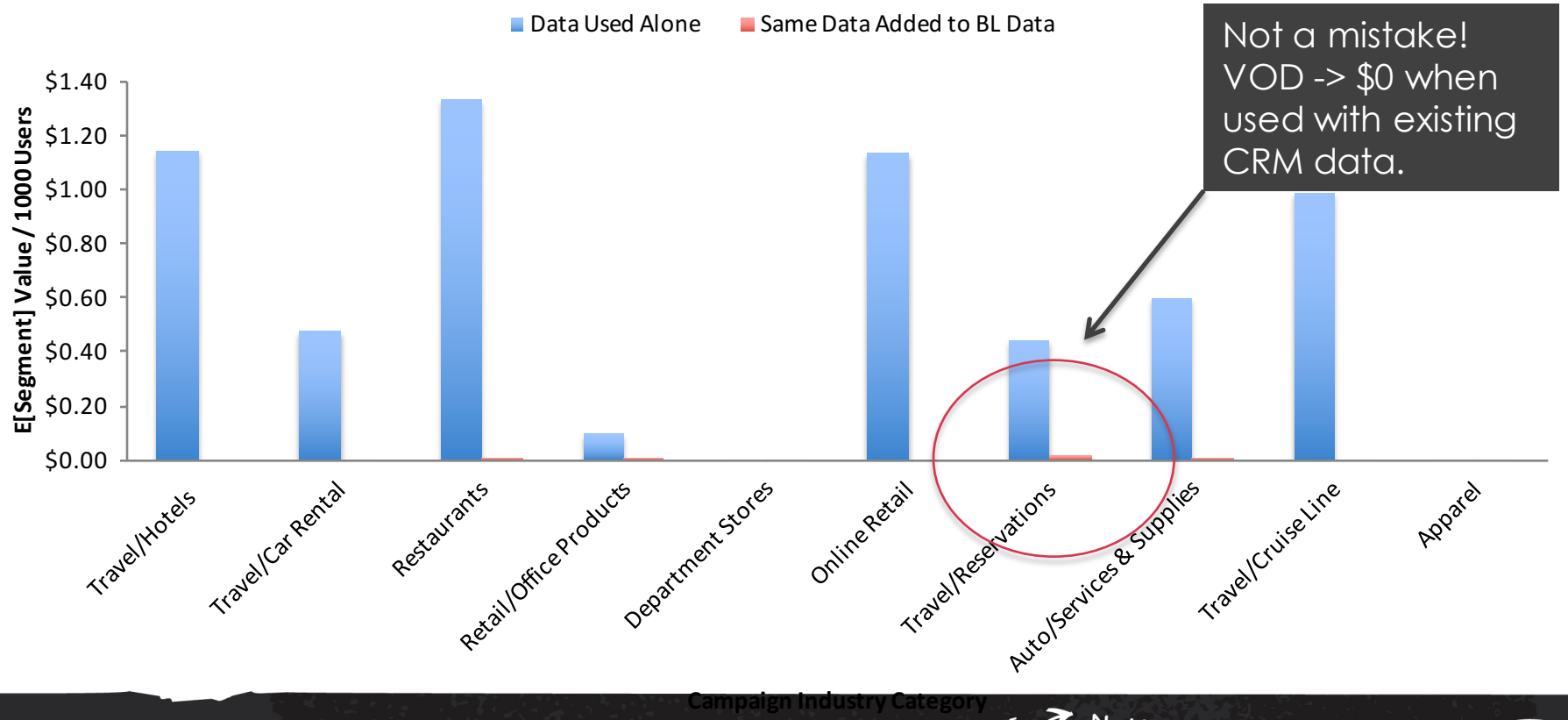
The value of a data point/source depends on what other data is being used along with it.



Example of that last point…



Example of that last point…



Takeaways

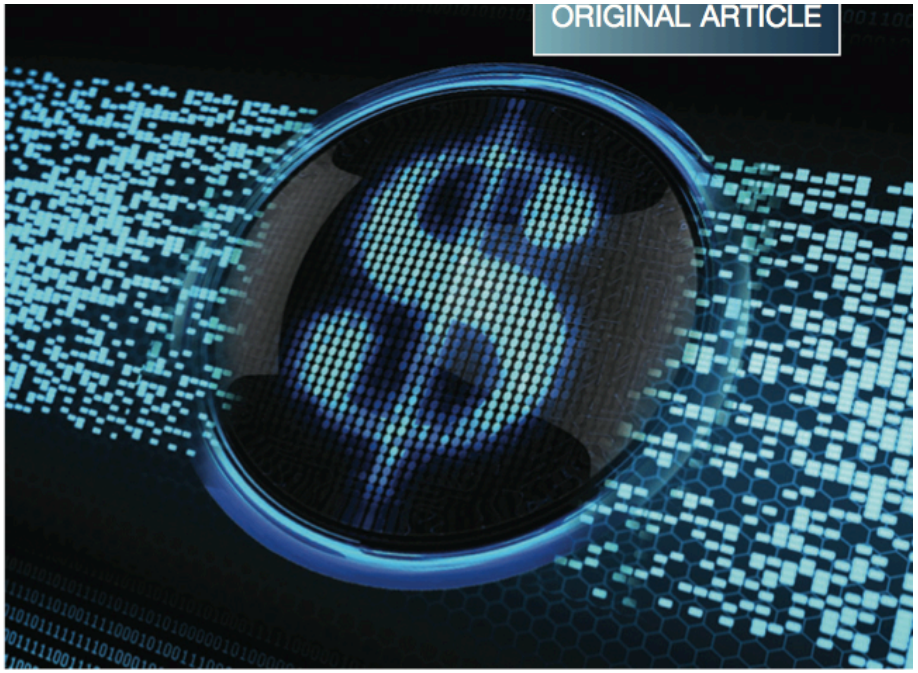
For data buyers/users:

- Use the expected value framework to determine optimal buying price or evaluate ROI
- Your own CRM data is likely worth the most
- Better models means data is worth more

For data sellers:

- The optimal selling price is a function of the buyers individual needs
- Auction mechanisms enable fair pricing

Want technical details? Big Data Journal, June 2014



BIGGER IS BETTER, BUT AT WHAT COST?

Estimating the Economic Value of Incremental Data Assets

Brian Dalessandro, Claudia Perlich, and Troy Raeder
Dstillery, New York, New York