

Deconstructing Data Science Questions

A Case Study on Personalization

Leo Guelman

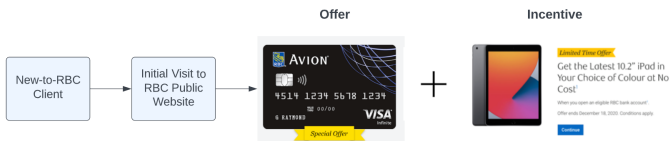
RBC Royal Bank

Talk Overview

- The practice of Data Science often places significant focus on adopting sound technical methods.
- However, much less effort is often placed on business problem identification.
- Using Personalized Marketing problem as a motivating example:
 - ▶ Show (fiction) story on how 4 different Data Scientists might approach this problem and get different conclusions from data.
 - ▶ Show that the different conclusions emerge as a result of slightly different interpretation of the business question being asked.

The Business Setting

- **Business Objective:** Cross-sell a credit card to new-to-RBC clients.
- **Offer/Incentive:**



- **Goal:** Maximize the expected profitability of the campaign on the new-to-RBC client segment.

Data Generating Process

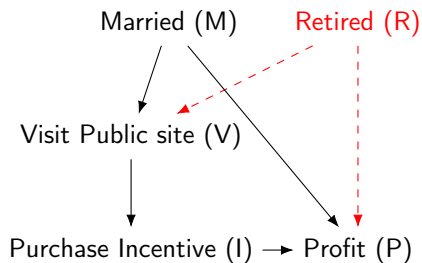


Figure: Causal DAG

Data Generating Process

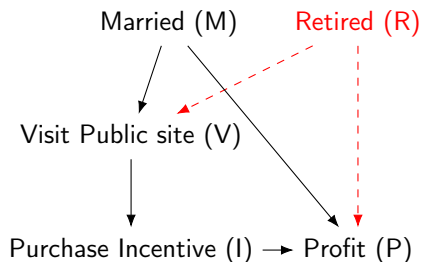


Figure: Causal DAG

$$V := M \oplus R$$

$$I := V$$

	R = 0		R = 1	
	M = 1	M = 0	M = 1	M = 0
I = 1	0.25	0.50	0.45	0.05
I = 0	0.50	0.10	0.05	0.30

Table: $E[P|M, R, I]$. Highlighted cells reflect (new-to-RBC) client's 'natural' choice to visit the Public site or not.

Data Scientist 1: Empirical Decision Criterion (EDC)

$$\text{EDC} \rightarrow \underset{I \in \{0,1\}}{\operatorname{argmax}} E[P|I, M]$$

$$E[P|I = 1, M = 1] = 0.25$$

$$E[P|I = 0, M = 1] = 0.05$$

$$E[P|I = 1, M = 0] = 0.05$$

$$E[P|I = 0, M = 0] = 0.10$$

	$R = 0$		$R = 1$	
	$M = 1$	$M = 0$	$M = 1$	$M = 0$
$I = 1$	0.25	0.50	0.45	0.05
$I = 0$	0.50	0.10	0.05	0.30

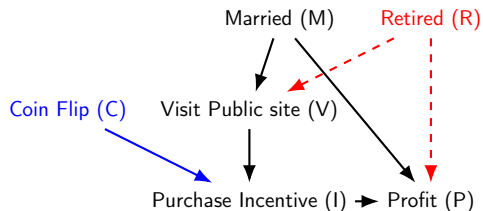
Table: $E[P|M, R, I]$.

Decision Rule:

- If Visit Site \wedge Married \rightarrow Purchase Incentive $\rightarrow E[P] = 0.25$
- If Visit Site \wedge Not Married \rightarrow No Purchase Incentive $\rightarrow E[P] = 0.05$

$$\text{Expected profit} = \boxed{0.15} = (0.25 + 0.05)/2.$$

Data Scientist 2: Post-Visit Randomization (PVR)



$$V := M \oplus R$$

$$I := V \wedge C$$

Figure: Causal DAG with post-visit randomization.

Data Scientist 2: Post-Visit Randomization (PVR) - cont'd

$$\text{PVR} \rightarrow \underset{I \in \{0,1\}}{\operatorname{argmax}} E[P | \operatorname{do}(I), M, V = 1]$$

	$R = 0$		$R = 1$	
	$M = 1$	$M = 0$	$M = 1$	$M = 0$
$I = 1$	0.25	0.50	0.45	0.05
$I = 0$	0.50	0.10	0.05	0.30

Table: $E[P | M, R, I]$.

$$E[P | \operatorname{do}(I = 1), M = 1, V = 1] = 0.25$$

$$E[P | \operatorname{do}(I = 0), M = 1, V = 1] = 0.50$$

$$E[P | \operatorname{do}(I = 1), M = 0, V = 1] = 0.05$$

$$E[P | \operatorname{do}(I = 0), M = 0, V = 1] = 0.30$$

Decision Rule:

- If Visit Site \wedge Married \rightarrow No Purchase Incentive $\rightarrow E[P] = 0.50$
- If Visit Site \wedge Not Married \rightarrow No Purchase Incentive $\rightarrow E[P] = 0.30$

$$\text{Expected profit} = \boxed{0.40} = (0.50 + 0.30) / 2.$$

Data Scientist 3: RCT on New-to-RBC Clients

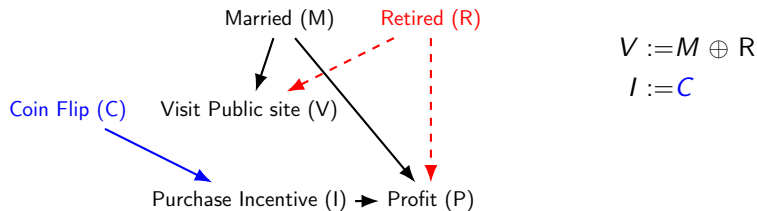


Figure: Causal DAG with RCT.

Data Scientist 3: RCT on New-to-RBC Clients - cont'd

$$\text{RCT} \rightarrow \operatorname{argmax}_{I \in \{0,1\}} E[P|do(I), M]$$

$$E[P|do(I=1), M=1] = 0.350 = (0.25 + 0.45)/2$$

$$E[P|do(I=0), M=1] = 0.275 = (0.50 + 0.05)/2$$

$$E[P|do(I=1), M=0] = 0.275 = (0.50 + 0.05)/2$$

$$E[P|do(I=0), M=0] = 0.200 = (0.10 + 0.30)/2$$

	$R = 0$		$R = 1$	
	$M = 1$	$M = 0$	$M = 1$	$M = 0$
$I = 1$	0.25	0.50	0.45	0.05
$I = 0$	0.50	0.10	0.05	0.30

Table: $E[P|M, R, I]$.

Decision Rule:

- If Married \rightarrow Purchase Incentive $\rightarrow E[P] = 0.35$
- If Not Married \rightarrow Purchase Incentive $\rightarrow E[P] = 0.275$

$$\text{Expected profit} = \boxed{0.315} = (0.35 + 0.275)/2.$$

Data Scientist 4: Counterfactual Decision Criterion

Profit = 0.4375

Summary

- Summarize results so far from the various criteria and emphasize no result is right or wrong, they answer different questions, but some questions are more preferable than others in specific context.
- Compare to oracle

Extensions

- Multiple Actions
- Online policy optimization (e.g., MABUC)

The Personalization Paradigm

- Illustration that shows that what is good for person 1 might be neutral or even harmful for person 2.
- Motivating illustration in healthcare, marketing, client-level decision making.

Notes

For each Data Science articulate what is the question being asked!!!!
The conclusion should be that although DS 4 achieves highest reward, whether that is the right solution, depends on the question being asked.