Attribution Models

Applied AI Course

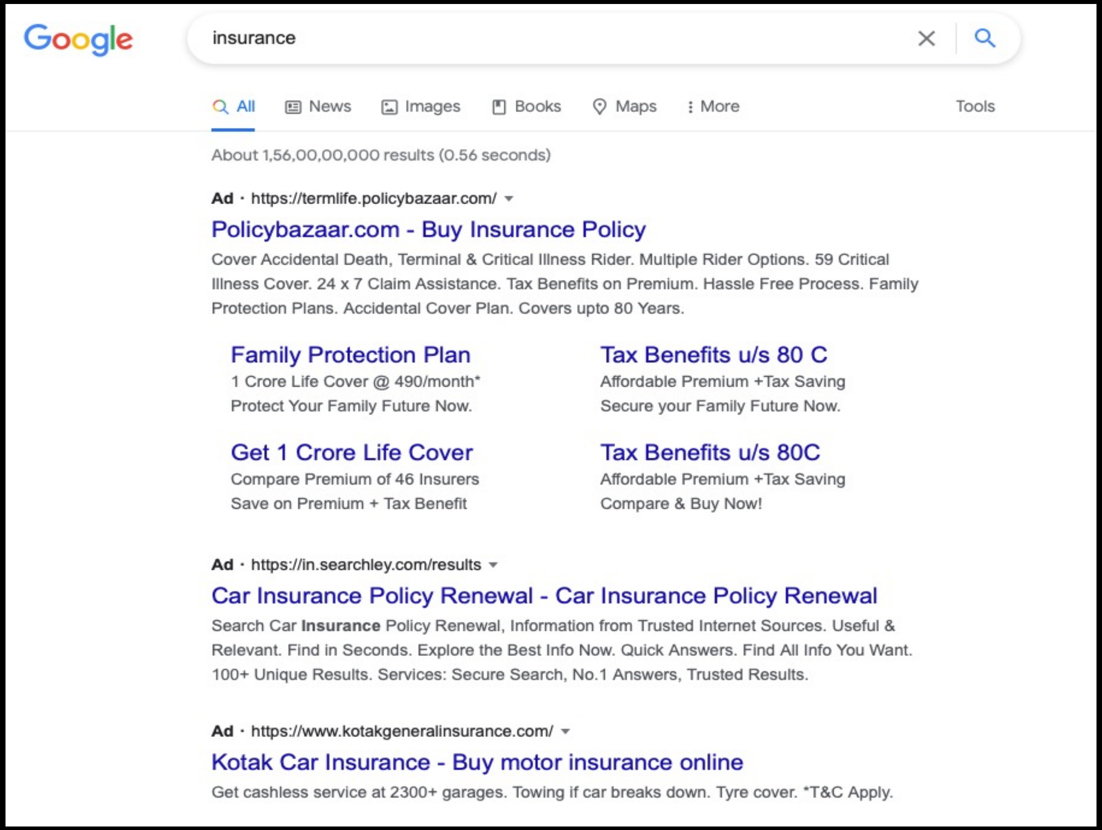
APPLIED ROOTS

"Half the money I spend on advertising is wasted; the trouble is I don't know which half" - John Wanamaker

APPLIED ROOTS

Digital marketing

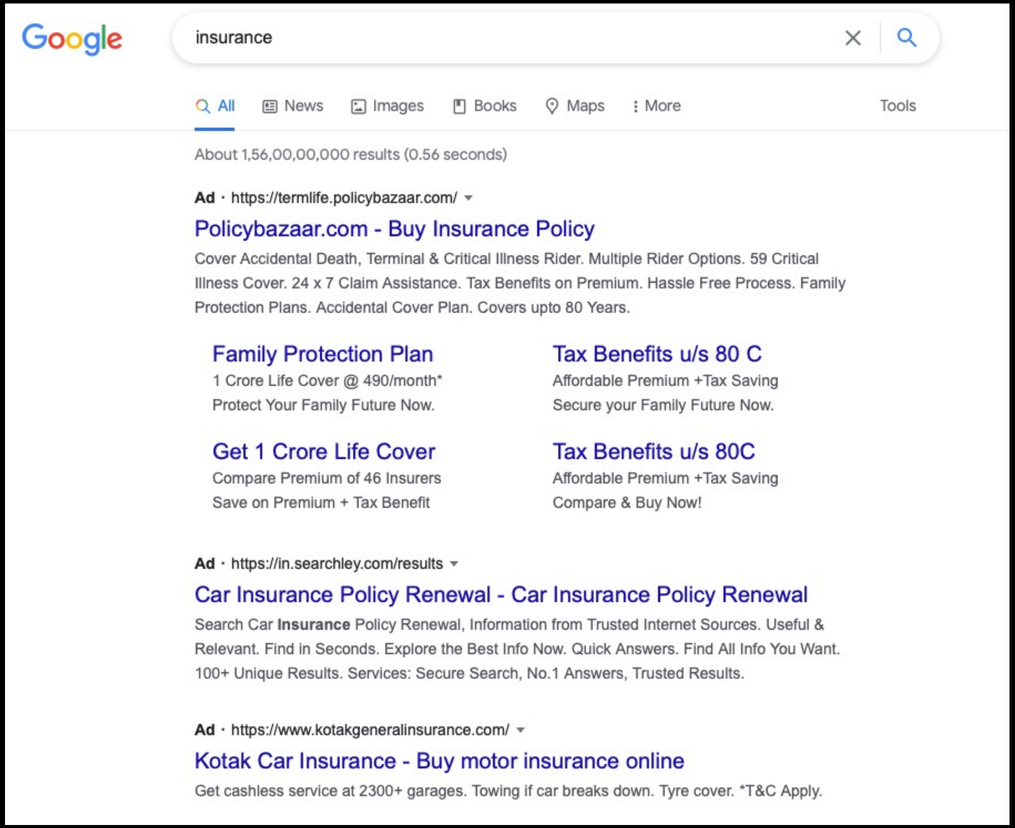
ad-click vs ad-impr



APPLIED ROOTS

Digital marketing

search vs display ads



APPLIED ROOTS

Traditional marketing :

TV → prime-time
→ TV-series
⋮

Print → Front-Page
→ Page3
→ Sports page

Outdoor → hoardings
→ Buses
⋮

APPLIED ROOTS

(Digital) Channels:

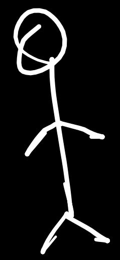
- email
- G-Search
- Youtube Video
- Youtube Search
- FB feed
- FB display ads
- Instagram Feed
- Twitter (promoted) Feed
- Amazon Search
- Amazon Display
- Google Display

⋮

APPLIED ROOTS

Marketing & Advertising:

Amazon



Facebook Ad Impression

Google Search Ad Impr + Click

Instagram Ad Impr

Youtube Search Ad Impr

Channels

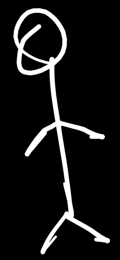
Time

Buys on Amazon for $100

APPLIED ROOTS

Marketing & Advertising:

Amazon



Facebook Ad Impression

Google Search Ad Impr + Click

Instagram Ad Impr

Youtube Search Ad Impr

Buys on Amazon $100

Time

How much contribution ?

APPLIED ROOTS

Big Ques:

How much did each ad impr | click contribute to the final purchase?

APPLIED ROOTS

Big players:

- Google Ads & Analytics
- FB
- Amazon

⋮

APPLIED ROOTS

Rule - based :

- gut - feeling driven
- Clicks more important than impressions
- Search ads are more influential than display ads.

APPLIED ROOTS

In the **Last Interaction** attribution model, the last touchpoint—in this case, the *Direct* channel—would receive 100% of the credit for the sale.

In the **Last Non-Direct Click** attribution model, all direct traffic is ignored, and 100% of the credit for the sale goes to the last channel that the customer clicked through from before converting—in this case, the *Email* channel.

In the **Last Google Ads Click** attribution model, the last Google Ads click—in this case, the first and only click to the *Paid Search* channel —would receive 100% of the credit for the sale.

In the **First Interaction** attribution model, the first touchpoint—in this case, the *Paid Search* channel—would receive 100% of the credit for the sale.

In the **Linear** attribution model, each touchpoint in the conversion path—in this case the *Paid Search*, *Social Network*, *Email*, and *Direct* channels—would share equal credit (25% each) for the sale.

In the **Time Decay** attribution model, the touchpoints closest in time to the sale or conversion get most of the credit. In this particular sale, the *Direct* and *Email* channels would receive the most credit because the customer interacted with them within a few hours of conversion. The *Social Network* channel would receive less credit than either the *Direct* or *Email* channels. Since the *Paid Search* interaction occurred one week earlier, this channel would receive significantly less credit.

In the **Position Based** attribution model, 40% credit is assigned to each the first and last interaction, and the remaining 20% credit is distributed evenly to the middle interactions. In this example, the *Paid Search* and *Direct* channels would each receive 40% credit, while the *Social Network* and *Email* channels would each receive 10% credit.

Rule-based

why?

https://support.google.com/analytics/answer/1662518?hl=en

APPLIED ROOTS

Data-driven models

- can we decipher from raw-data?
- many methods

APPLIED ROOTS

Other-names for Attribution - modeling:

Multi-channel attribution

Marketing mix modeling

APPLIED ROOTS

Granularity of attribution:

- For each sale/ conversion
- Across many sales/conversions in a day/week

APPLIED ROOTS

Regression-based MMM:

e.g: Spend on TV, radio, newspaper] every week

revenue/sales

APPLIED ROOTS

$x_i = \langle TV_i, Radio_i, newspaper_i \rangle$

$y_i = sales_i$

Linear regression model:

$sales_i = w_0 + w_1 TV_i + w_2 Radio_i + w_3 newspaper_i$

APPLIED ROOTS

weights for attribution:

$$Sales_i = 100K + 2 \cdot TV_i + 1 \cdot Radio_i + 0.5\, newspaper_i$$

$$\Downarrow$$

a \$1 Spent on TV gives 4 times more sales than \$1 on newspaper

Other regression models also work

Random forest

GBDT

Neural Nets

Feature Importance

model-specific

model-agnostic

[LIME, SHAP]

APPLIED ROOTS

Problems:

Feature importances are not perfect

but are very useful

Be careful:

- time-varying sales (due to festivals)

  aka Seasonality

  ↓

  add seasonality features

  day of week, Week of Year, isHolidays ....

APPLIED ROOTS

- Carry-over effects

⇓

time-gap between ads and sales

e.g: Home, Car, smartphone, Chocolate-bar

APPLIED ROOTS

Code - example:

https://towardsdatascience.com/building-a-simple-marketing-mix-model-with-ols-571ac3d5b64f

Game-Theoritic

- Shapley values [2012 Nobel prize in Economics]

(***) - Mathematic gaurantees & very rigorous foundations

- "The Beautiful mind" movie

APPLIED ROOTS

- Co-operative Games
  - economics: Team of 10 (players) build & sell a truck. How much of the value is to be attributed to each?

APPLIED ROOTS

each channel is a player

n - players (let)

$N = \{1, 2, 3, \dots n\}$

[Terminology]

A = coalition $\subseteq N$

APPLIED ROOTS

Characteristic Function:

$$v : 2^N \rightarrow R$$

e.g: $A = \{1, 3, 6\}$, $v(A) = \$100$

Payoff / Value generated by coalition A.

APPLIED ROOTS

$V(A)$ in Attribution modeling

- Total number of conversions/sales (simpler) generated
- $P(\text{conversion} \mid A)$ using an ML-model

APPLIED ROOTS

Shapley value of i^{th} channel/player = How much credit to give to channel i

given $v(A)$ & $N = \{1, 2, \ldots n\}$

$\phi_i(v)$

APPLIED ROOTS

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n-|S|-1)!}{n!} \left[v(S \cup \{i\}) - v(S) \right]$$

sum across all subsets

weighted

incremental value due to player i

APPLIED ROOTS

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n-|S|-1)!}{n!} \left[v(S \cup \{i\}) - v(S) \right]$$

$$\left[\frac{1}{{}^{n}C_{|S|}} \right]$$

APPLIED ROOTS

Problem: - Exponential time - complexity

- lots of data needed

Code Sample: https://medium.com/data-from-the-trenches/marketing-attribution-e7fa7ae9e919

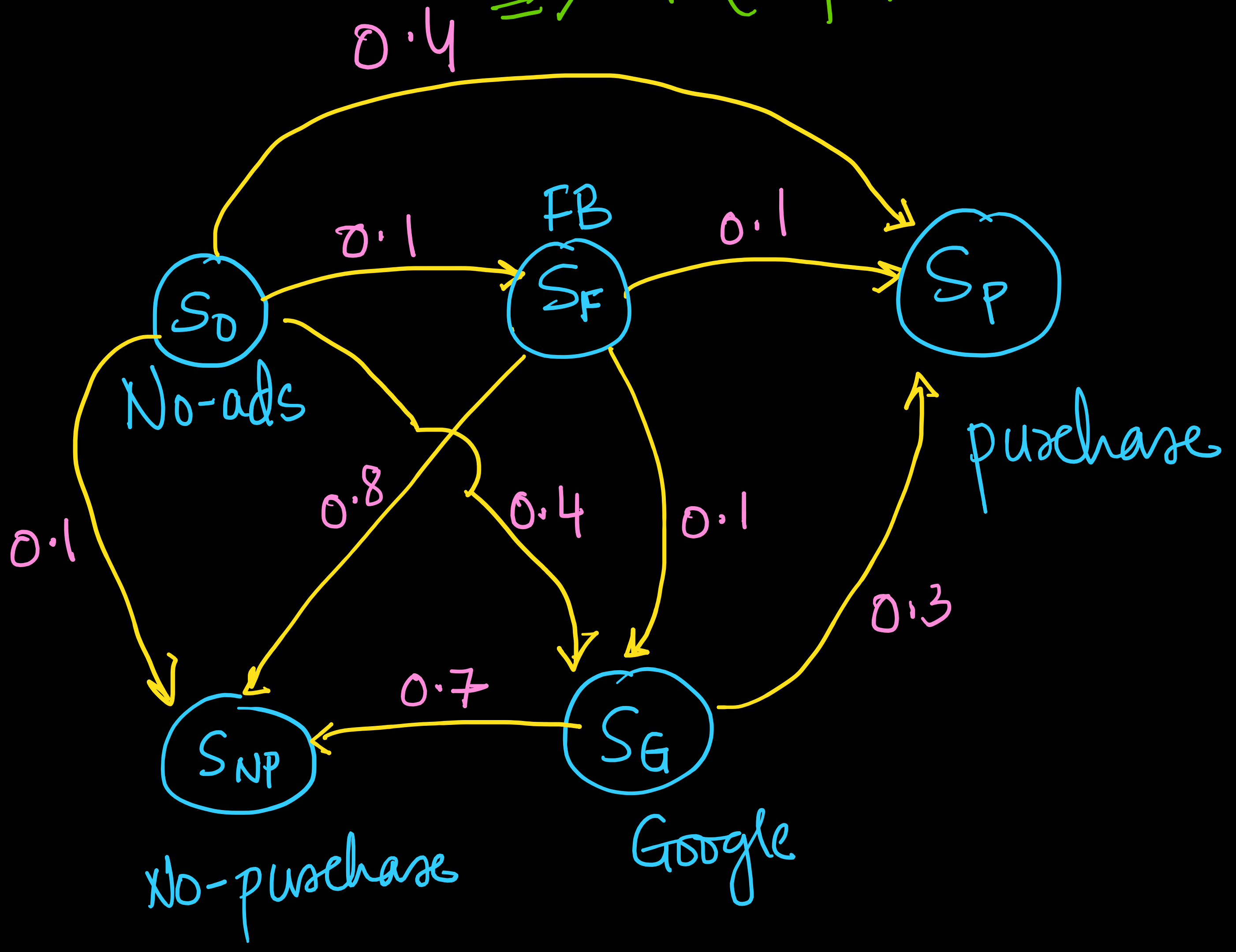
Attribution Models

- Game theoritic (SOTA)
- ML-based (popular)
- Markov-chains (Classical)
- Survival Analysis (Classical)
- ⋮

APPLIED ROOTS

Markov Chains:

$$P(S_P | S_O) = 0.4$$

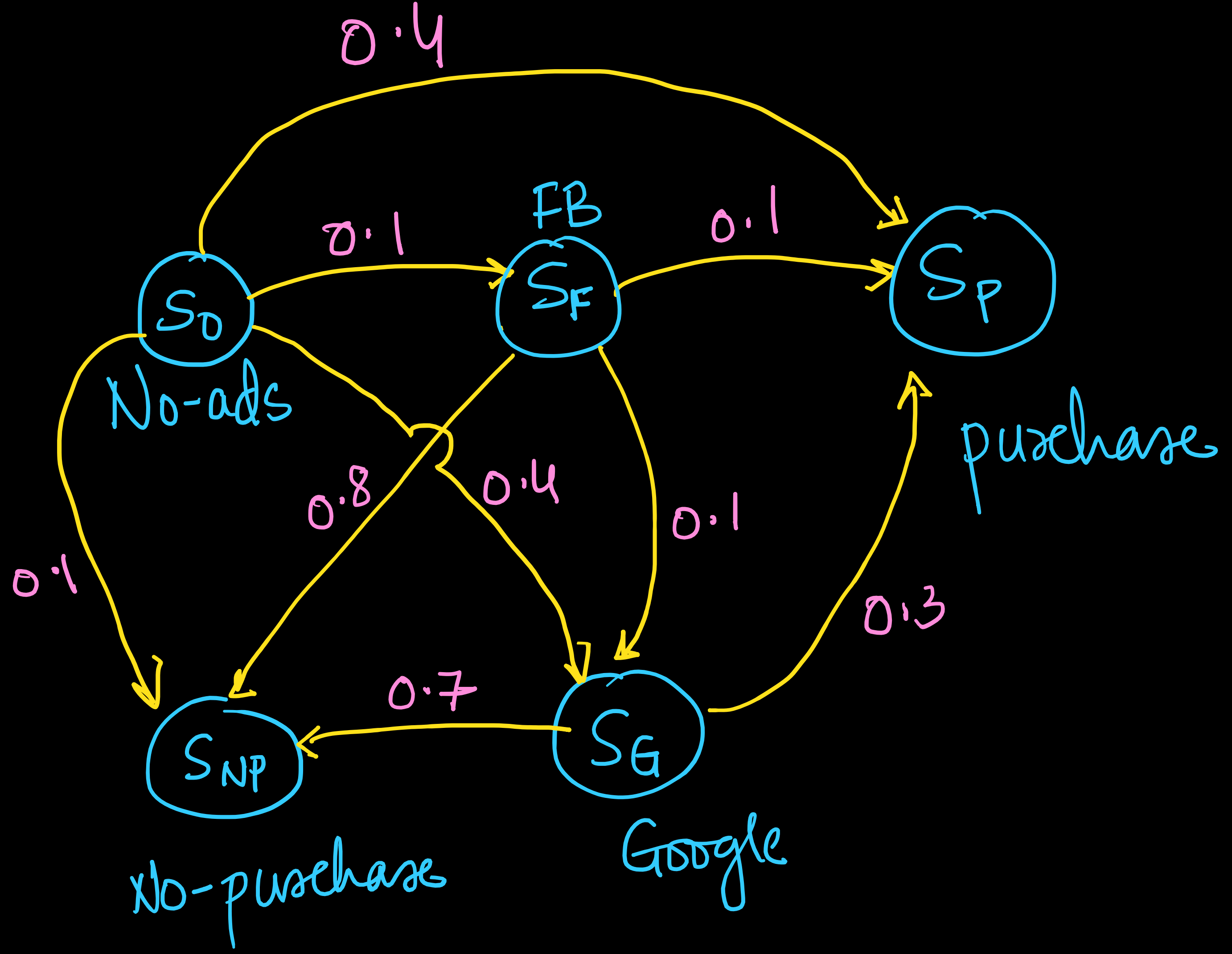


APPLIED ROOTS

Markov Property: 'memorless'

$$P(S_j \mid \underbrace{S_{j-1}, S_{j-2}, \ldots S_0}_{\text{history doesnot matter}}) = P(S_j \mid S_{j-1})$$

APPLIED ROOTS



$$P(\text{conversion or Purchase})$$

$$= 0{\cdot}4 +$$

$$0{\cdot}1 \times 0{\cdot}1 +$$

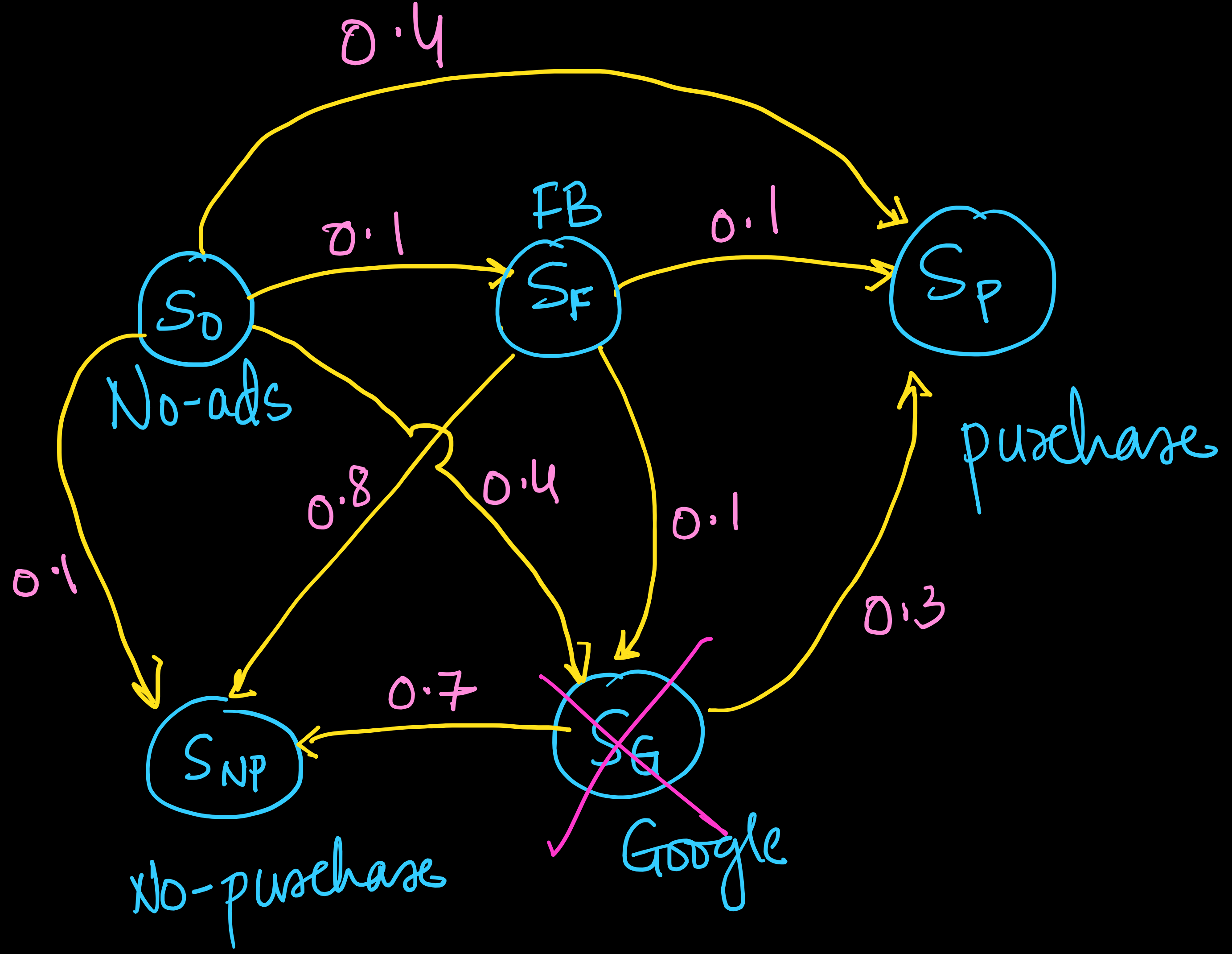
$$0{\cdot}1 \times 0{\cdot}1 \times 0{\cdot}3 +$$

$$0{\cdot}4 \times 0{\cdot}3$$

$$= 0{\cdot}533$$

APPLIED ROOTS

Removal effect of S_G



$P(\text{purchase})$

$= 0.4 +$

0.1×0.1

$= 0.41$

APPLIED ROOTS

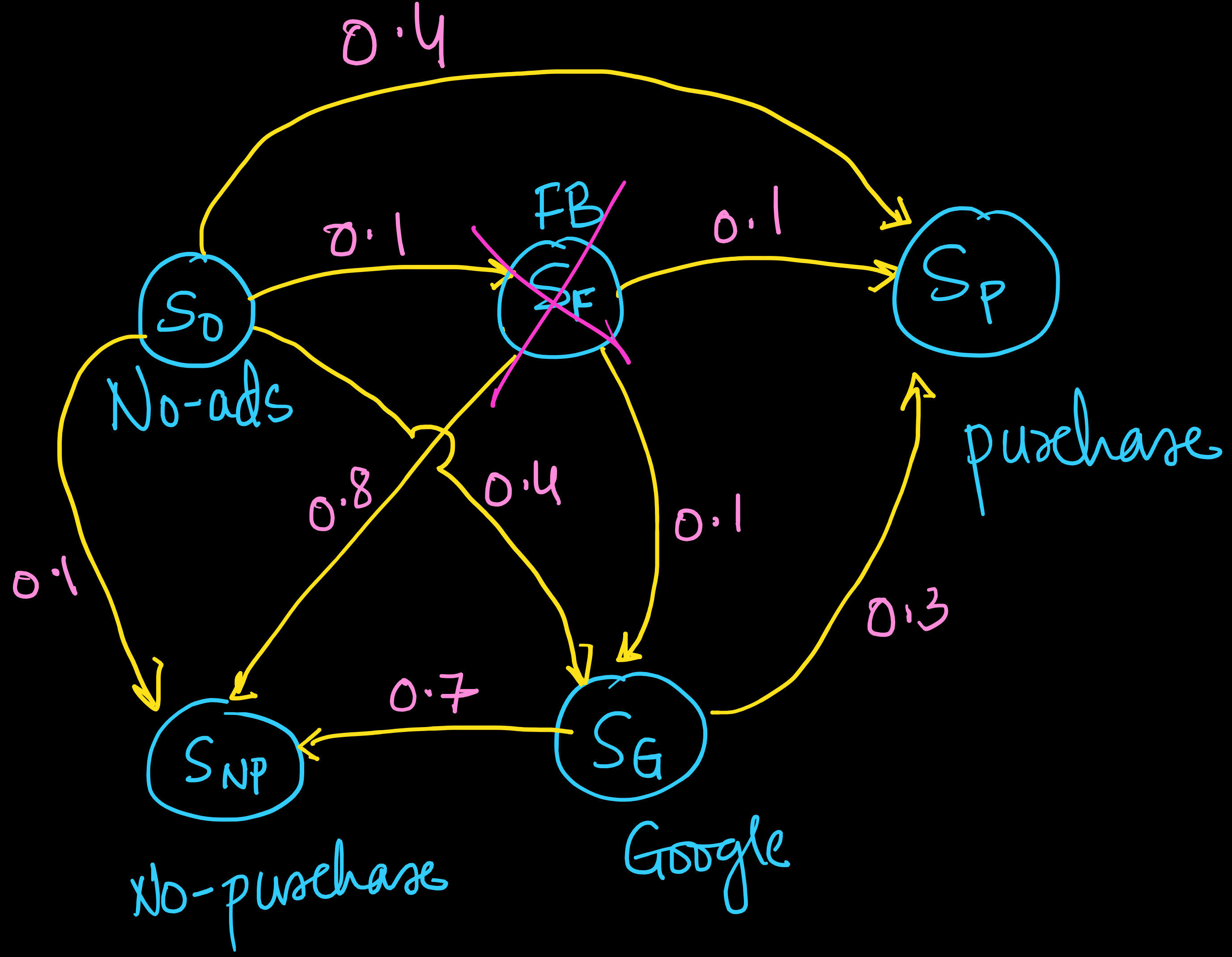
Effect of removing $S_G = 1 - \left(\frac{0.41}{0.533}\right)$

what %age of conversions would be lost if we donot use Google-ads

$= 0.2307$

APPLIED ROOTS

Removal effect of S_F



$$P(\text{purchase})$$

$$= 0.4 + 0.4 \times 0.3$$

$$= 0.52$$

APPLIED ROOTS

Effect of removing $S_F = 1 - \left(\frac{0.52}{0.533}\right)$

$= 0.024$

what %age of conversions would be lost if we donot use FB-ads

APPLIED ROOTS

$\Rightarrow$ Direct $\longrightarrow$ $100 - (23.07 + 2.4) = \boxed{74.53\%}$

Google Ads $\longrightarrow$ $\boxed{23.07\%}$

FB ads $\longrightarrow$ $\boxed{2.4\%}$

APPLIED ROOTS

What's the drawback of Markov Chains?

APPLIED ROOTS

$$P(s_j \mid \underbrace{s_{j-1}, s_{j-2}, \ldots s_0}_{\text{ignoring all these}}) = P(s_j \mid s_{j-1})$$

APPLIED ROOTS

Advantage of Markov chains

Computationally easy → Paths in a Graphs

→ remove & recompute probabilities

APPLIED ROOTS

Trade off between $\underbrace{\text{computation \& rigour}}$

Markov chains

Shapley values

APPLIED ROOTS

Dataset + Code

https://towardsdatascience.com/marketing-channel-attribution-with-markov-chains-in-python-part-2-the-complete-walkthrough-733c65b23323

APPLIED ROOTS

Survival Analysis for Attribution.

- healthcare analytics
- → survival of a patient
- → In marketing : Conversion

'Death' in Health care

↑

event of Interest

APPLIED ROOTS

Survival Function:

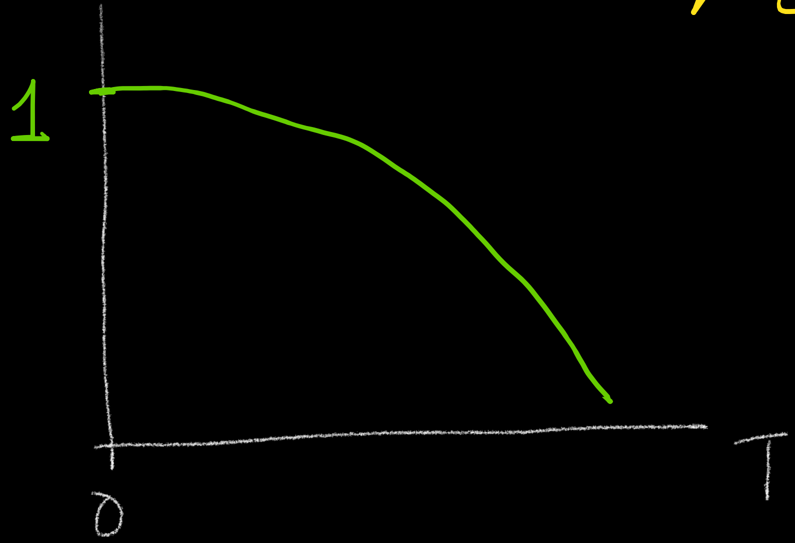
$$S(t) = P(T > t) \longrightarrow \text{Probability of Survival @ time } t$$

Note $T \geq 0$

APPLIED ROOTS

CDF

$$S(t) = P(T > t) = 1 - F(t)$$



$$\Rightarrow S(t) = 1 - P(T \le t)$$

$$S'(t) = -f(t)$$

PDF

Hazard Function:

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T \leq t+dt \mid T > t)}{dt}$$

APPLIED ROOTS

$$
\begin{aligned}
\lambda(t) &= \lim_{dt \to 0} \frac{Pr(t \leq T \leq t + dt | T > t)}{dt} \\
&= \lim_{dt \to 0} \frac{Pr(t \leq T \leq t + dt) / Pr(T > t)}{dt} \\
&= \lim_{dt \to 0} \frac{(F(t + dt) - F(t)) / S(t)}{dt}
\end{aligned}
$$

PDF

CDF

$$
\begin{aligned}
&= \frac{f(t)}{S(t)} \\
&= -\frac{S'(t)}{S(t)}.
\end{aligned}
$$

APPLIED ROOTS

$$\frac{d \log(g(t))}{dt} = \frac{g'(t)}{g(t)}$$

So, $\lambda(t) = -\frac{d}{dt} \log(S(t))$

$$\Rightarrow S(t) = \exp\left(-\int_0^t \lambda(x)\, dx\right)$$

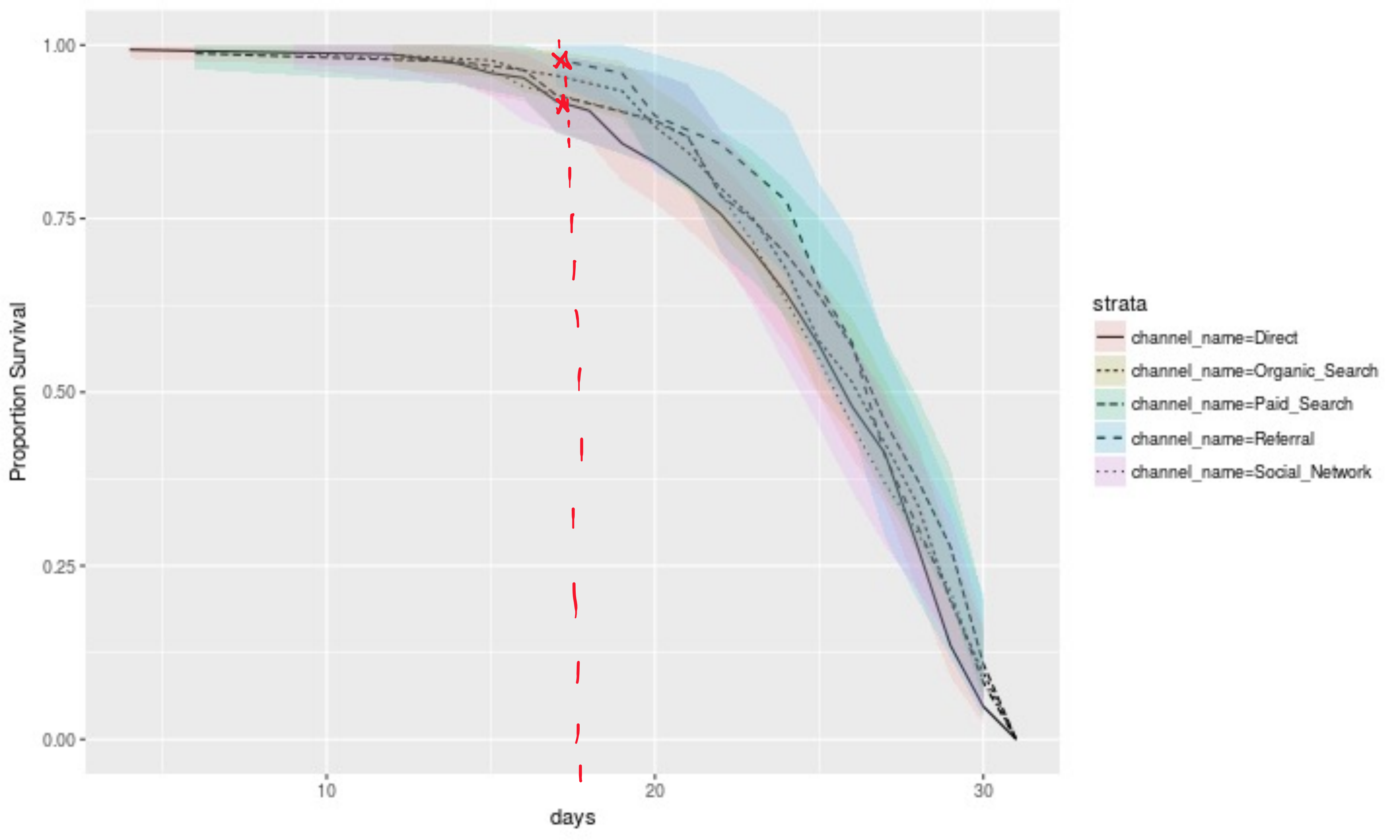
APPLIED ROOTS

if we know $\lambda(t)$, we can compute

$S(t)$

APPLIED ROOTS

Univariate



https://ahsanijaz.github.io/2016-10-21-ChannelAttribution/

APPLIED ROOTS

Cox-Proportional Hazards Regression Model:

$$\underbrace{\lambda(t|x)}_{\text{Co-variates or feature-vec}} = \underbrace{b_0(t)}_{\text{baseline Hazard with } x=\vec{0}} \cdot \underbrace{\exp\left\{\sum_{i=1}^{d} b_i x_i\right\}}_{\text{partial Hazards}}$$

APPLIED ROOTS

$$\lambda(t|x) = b_0(t) \cdot \exp\left\{\sum_{i=1}^{d} b_i x_i\right\}$$

$\downarrow$

multiplicative linear-model

b_i's : importance of each channel

Regression + feature Importance in the Survival analysis Framework.

APPLIED ROOTS

APPLIED
ROOTS

APPLIED ROOTS