



Concepts before Machinery

Harnessing the power of domain expertise for machine
learning based solutions

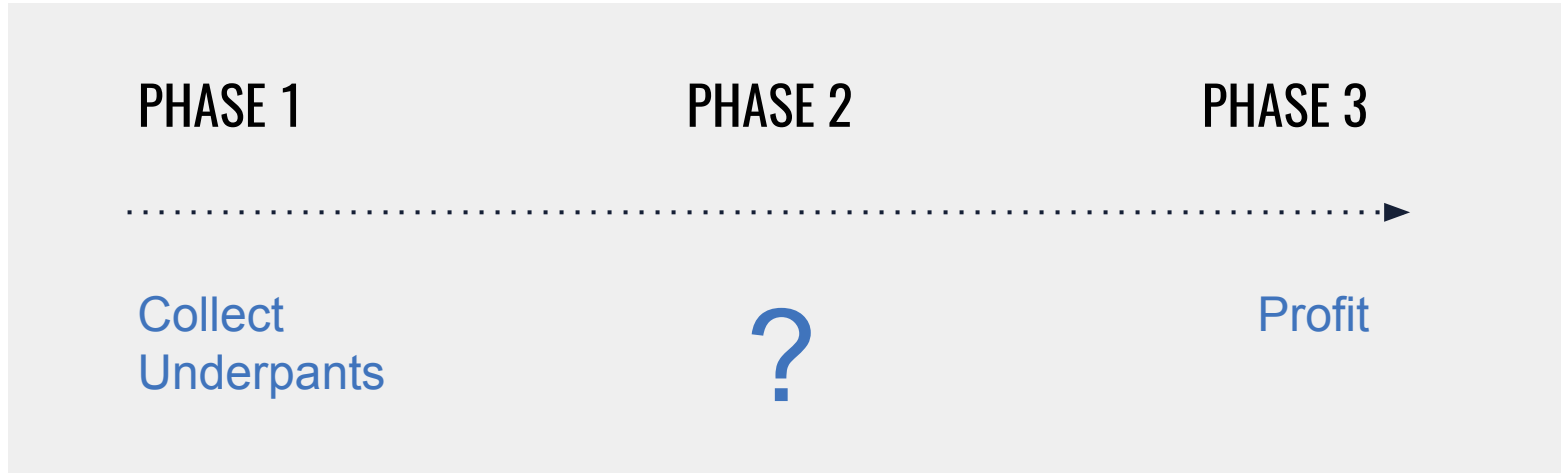
Ofer Ron

October 2017

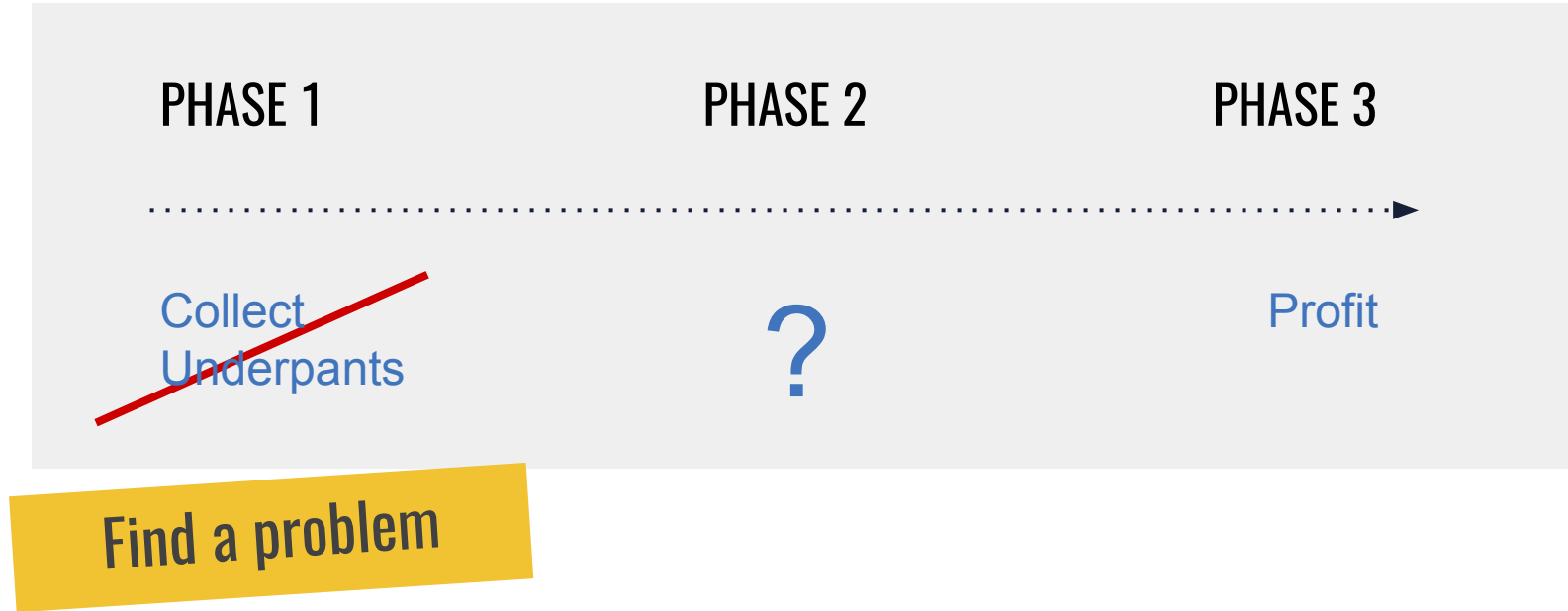
UNDERPANTS Gnomes



The Underpants Gnomes business plan



The machine learning product cycle



The machine learning product cycle



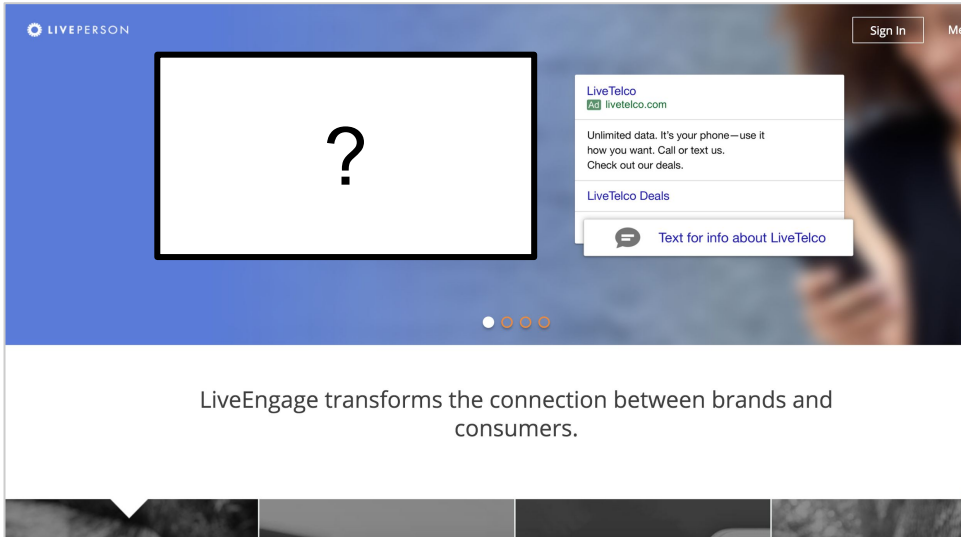
The reality

- Find a problem amenable to machine learning.
- Understand the KPIs and what drives them.
- Deploy a solution and (hopefully) profit!

Three tales of exploration

(Or how we learned to think before we modeled)

Tale No. 1: Choosing content



The problem

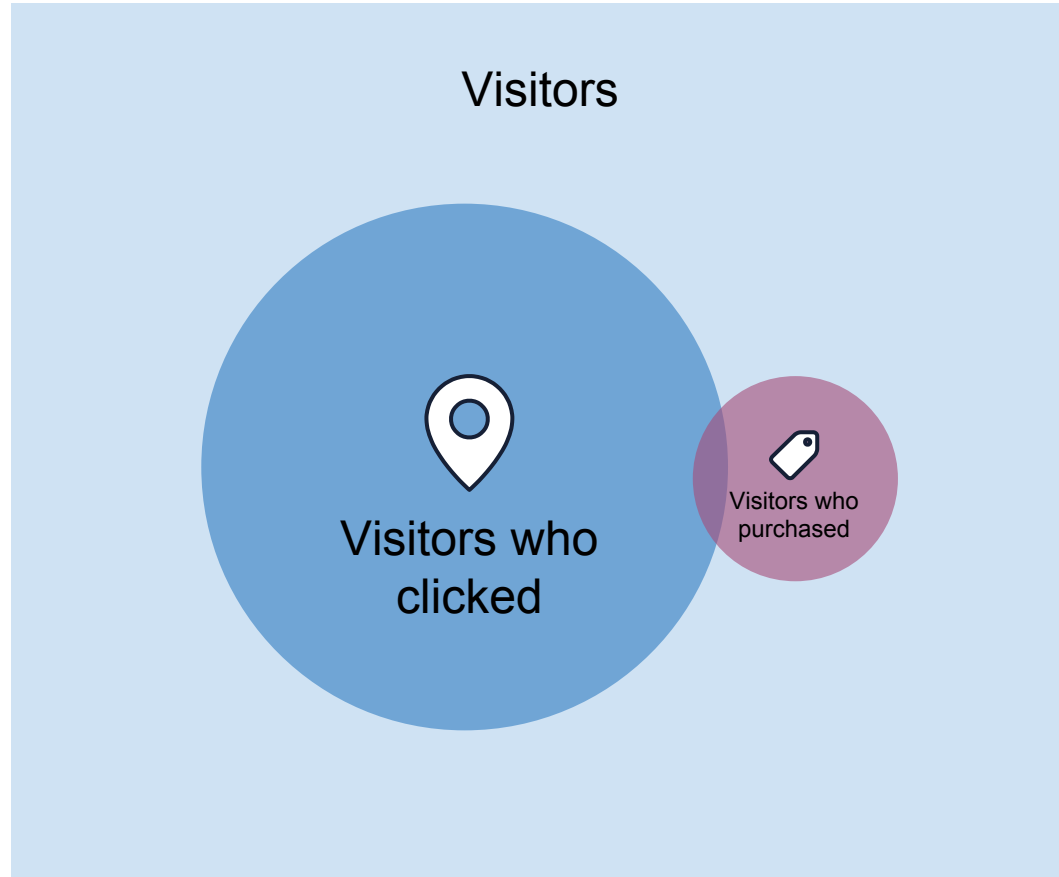
Given a model for

$$P(\text{Click} \mid \text{Content displayed}, \text{Visitor})$$

Build

$$P(\text{Revenue} \mid \text{Content displayed}, \text{Visitor})$$

The data



The Machinery

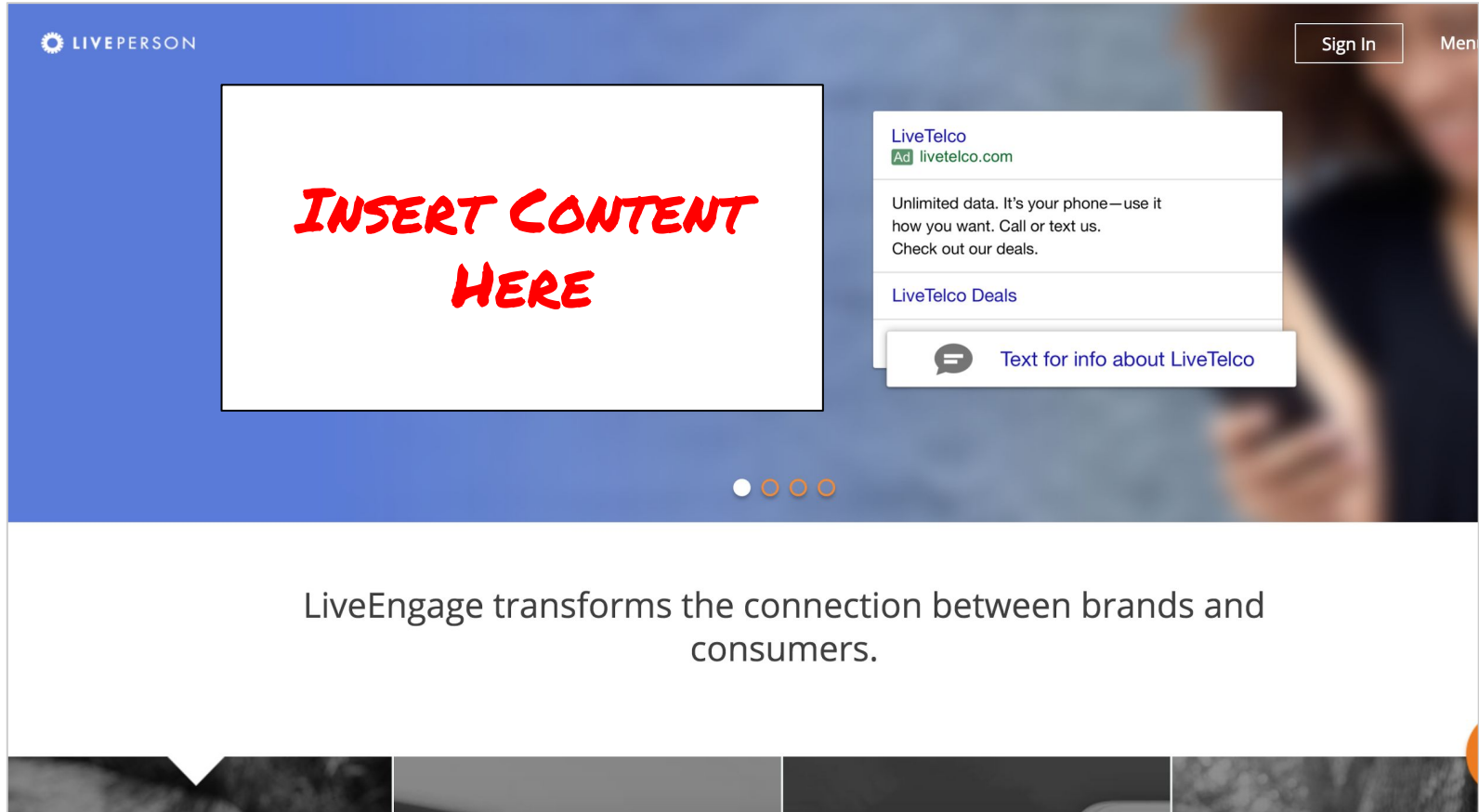
Revenue data is scarce.

We built a flow which evolved the model from clicks to revenue as more data was collected.

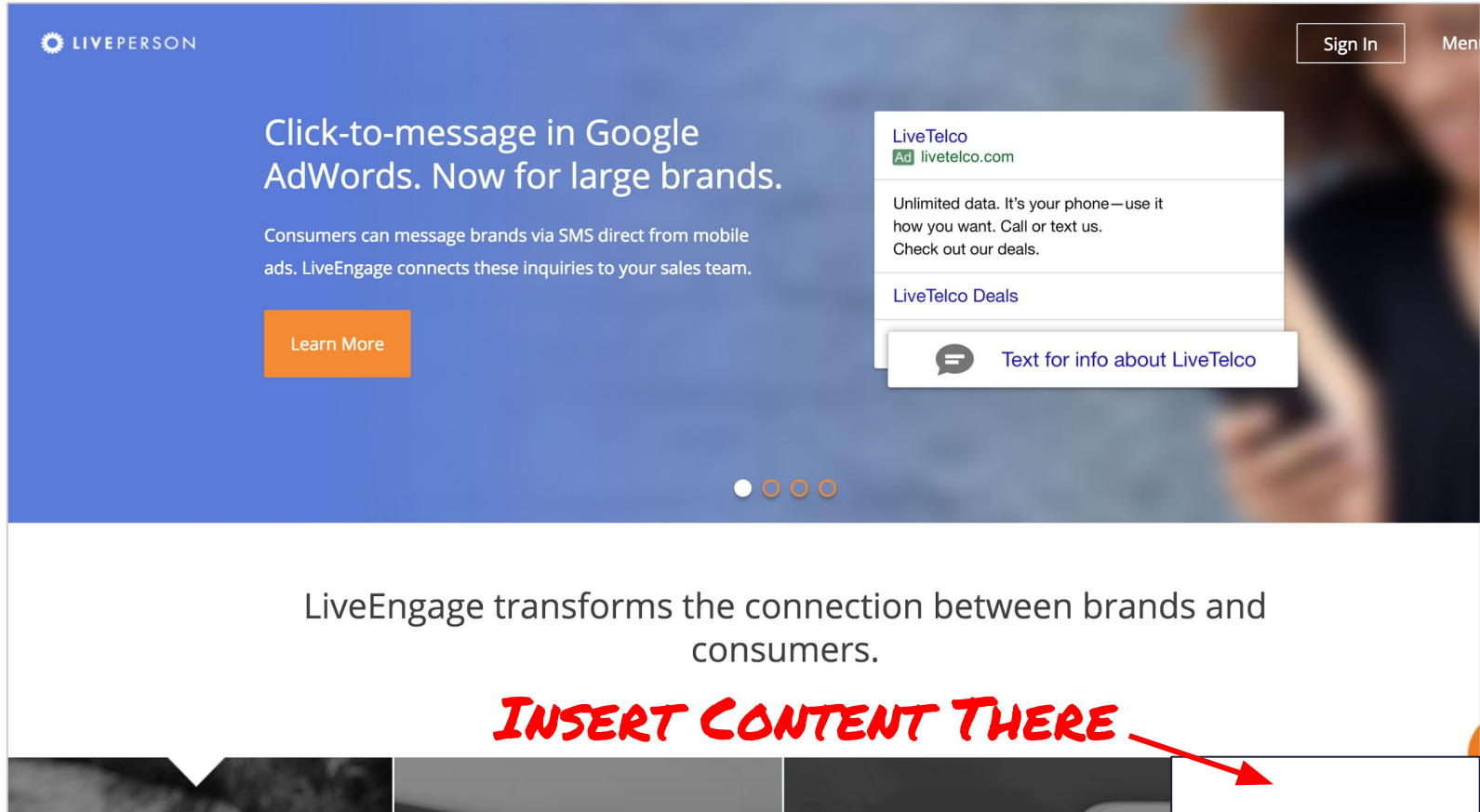
Lab: revenue predicted correctly.

Field: in some sites, revenue dropped after deployment.

The optimal site



The average site



The data for the **optimal** site



The data for the **average** site



The solution

$P(\text{Revenue} \mid \text{Visitor, Content clicked}) \times$

$P(\text{Content Clicked} \mid \text{Visitor, Content displayed})$

The solution

$P(\text{Revenue} \mid \text{Visitor, Content clicked}) \times$

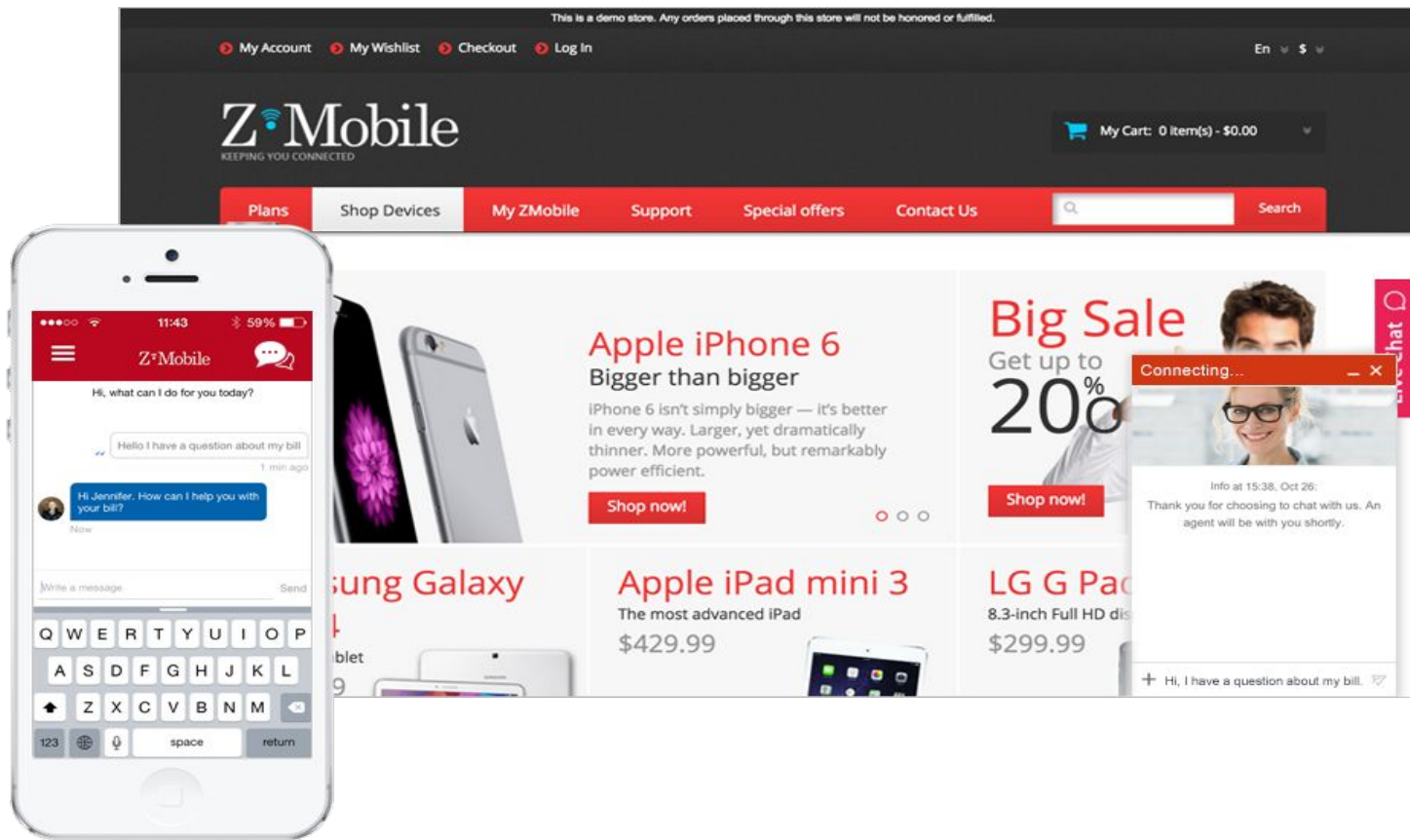
$P(\text{Content Clicked} \mid \text{Visitor, Content displayed})$

The moral

Control for the noise in your system explicitly.

Tale No. 2: Who should brands chat with?

LivePerson: We connect people and brands



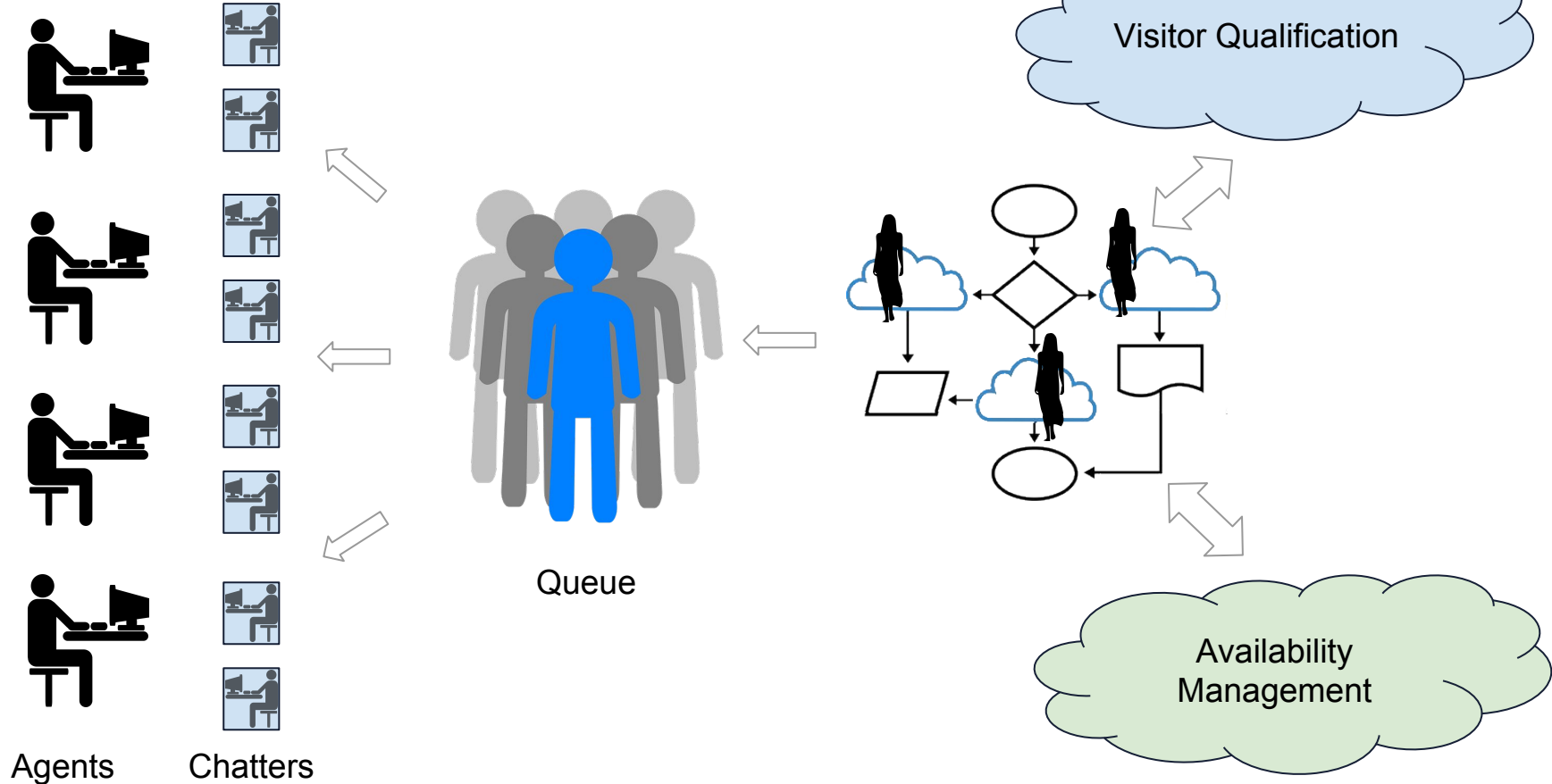
Resource optimization for ecommerce

Chat agents are expensive.

We have many more visitors than we have ability to chat with.

Solution: chat with visitors generating the most revenue per agent labor hour.

The system



First iteration

Model:

$$P(\text{Convert} \mid \text{Visited site})$$

Lab: Model performed well (measured by AUC).

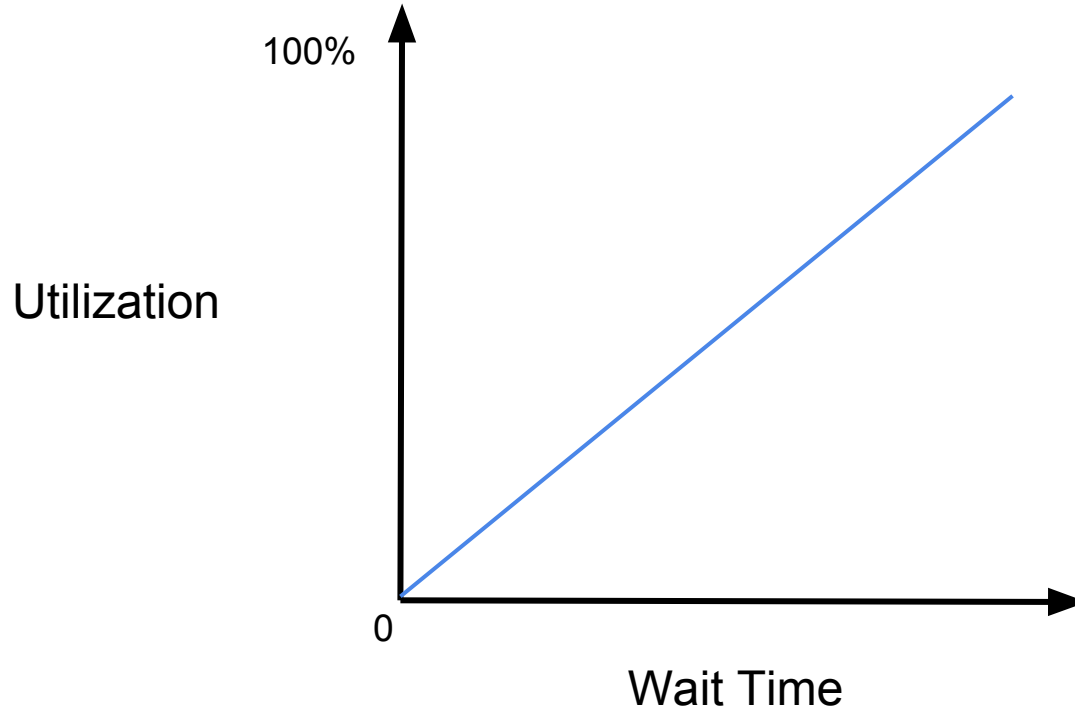
Field: in most sites, drops in conversion and chat volume.

Chat Volume

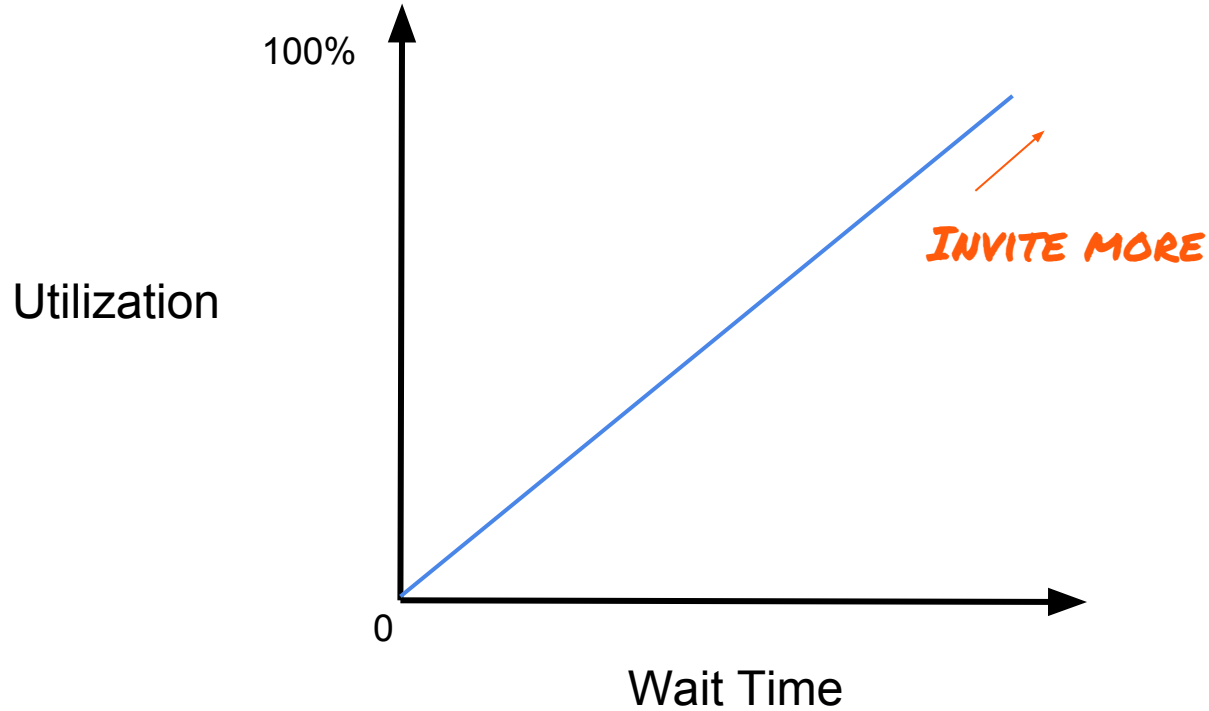
Call center queues are measured by:

- **Utilization:** are the agents working at full capacity?
- **Wait times:** how long do people wait for an agent? Do they abandon?

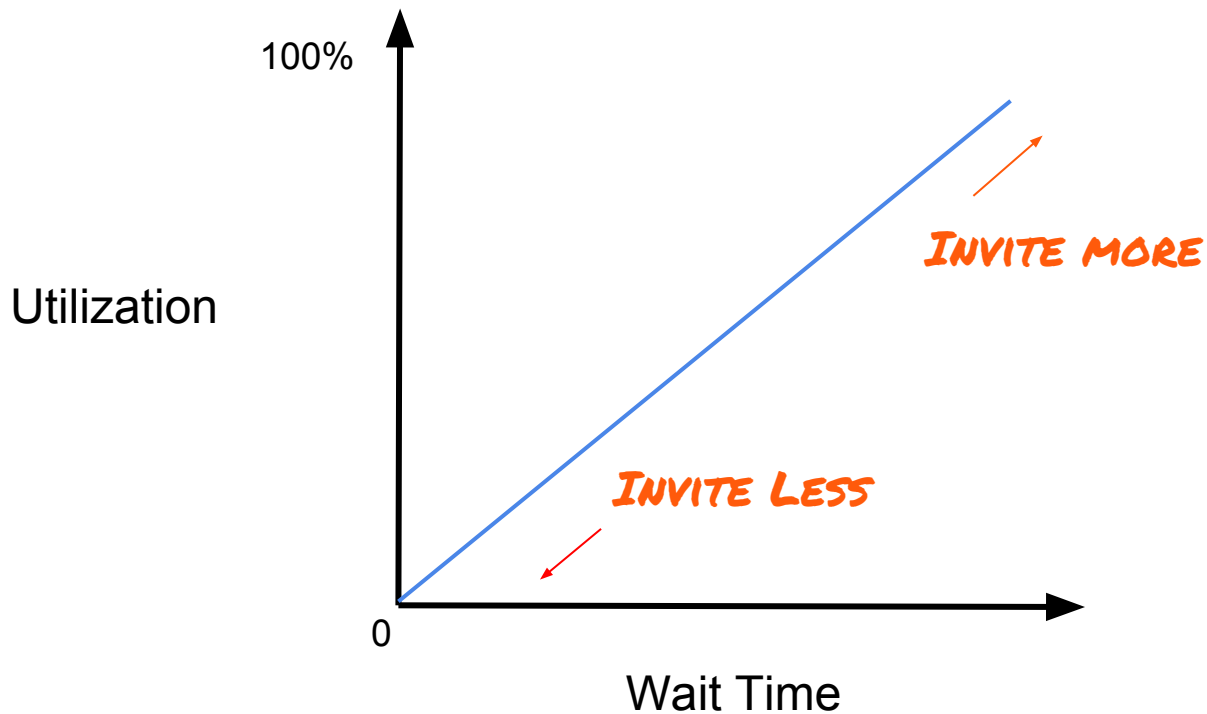
Utilization/Wait time Tradeoff



Utilization/Wait time Tradeoff



Utilization/Wait time Tradeoff



Invites aren't free

Model for invites explicitly

$$P(\textit{Accepted} \mid \textit{Invited}) \times P(\textit{Converted} \mid \textit{Visited})$$

Invites aren't free

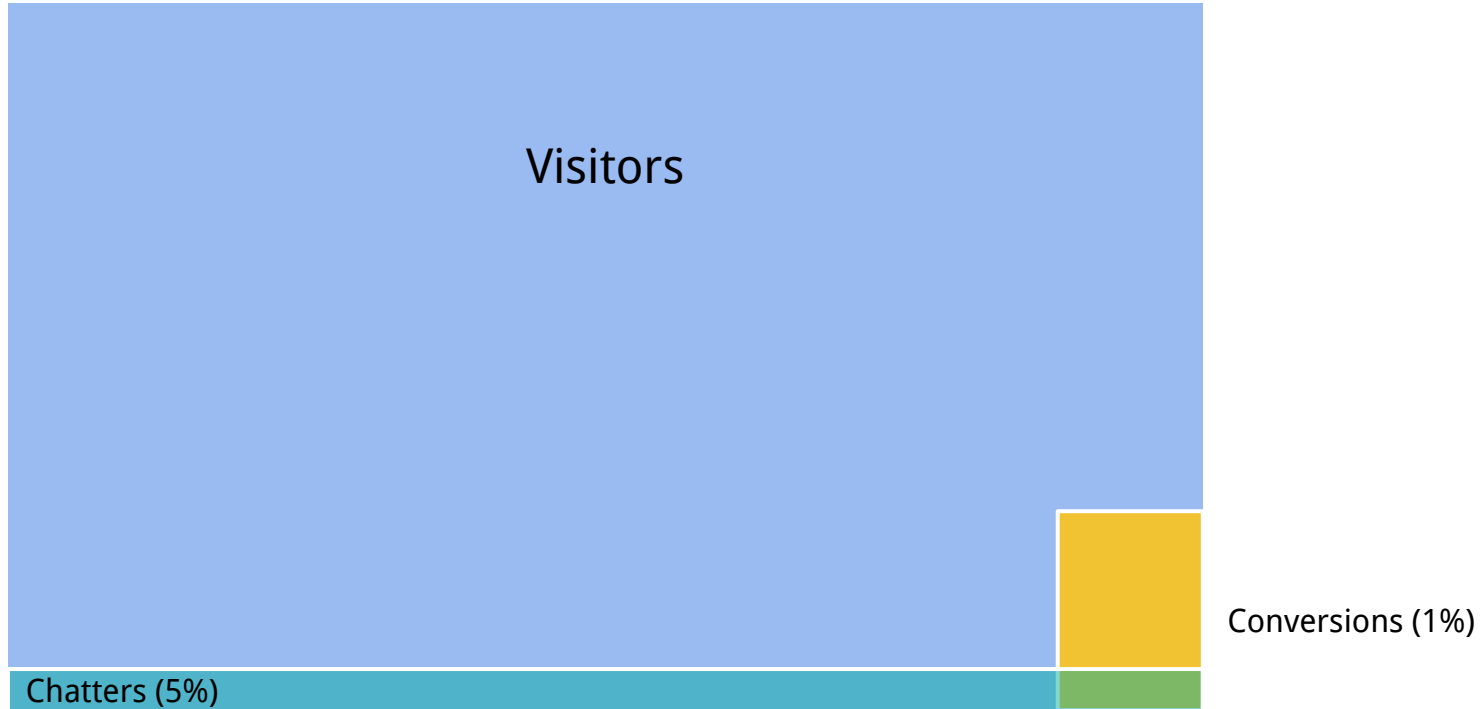
Model for invites explicitly

$$P(\textit{Accepted} \mid \textit{Invited}) \times P(\textit{Converted} \mid \textit{Visited})$$

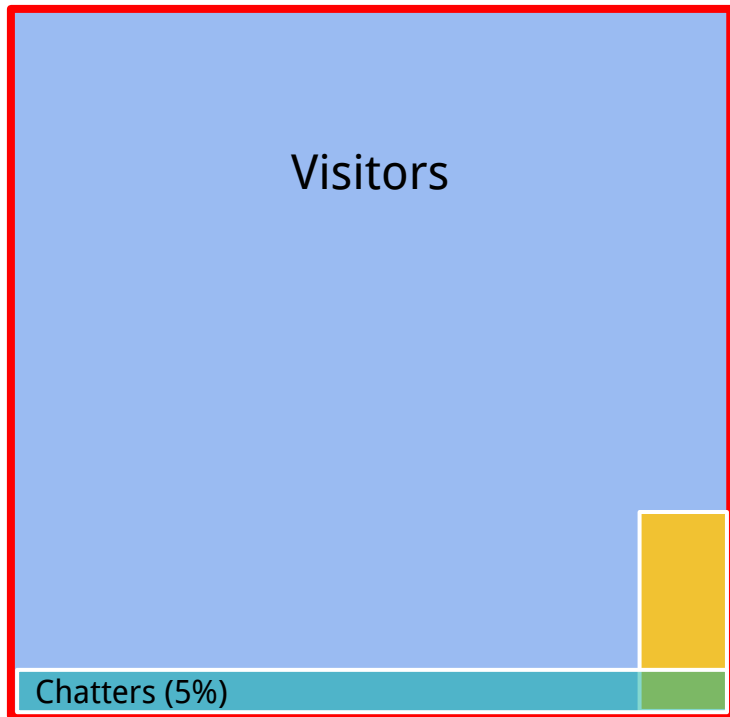
OR ANY OTHER MONOTONIC FUNCTION



Increasing conversions - the Data

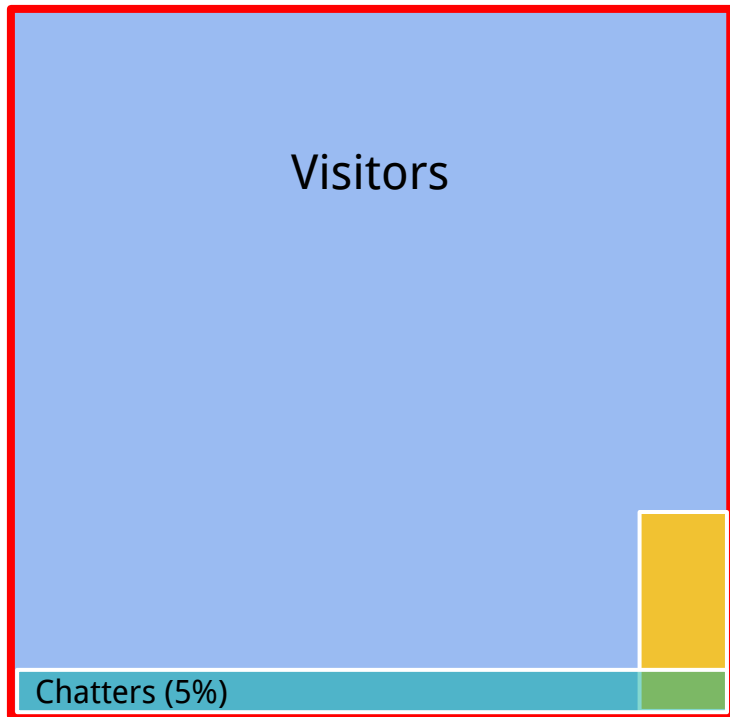


Increasing conversions - the Model



$$P(\text{Converted} \mid \text{Visited})$$

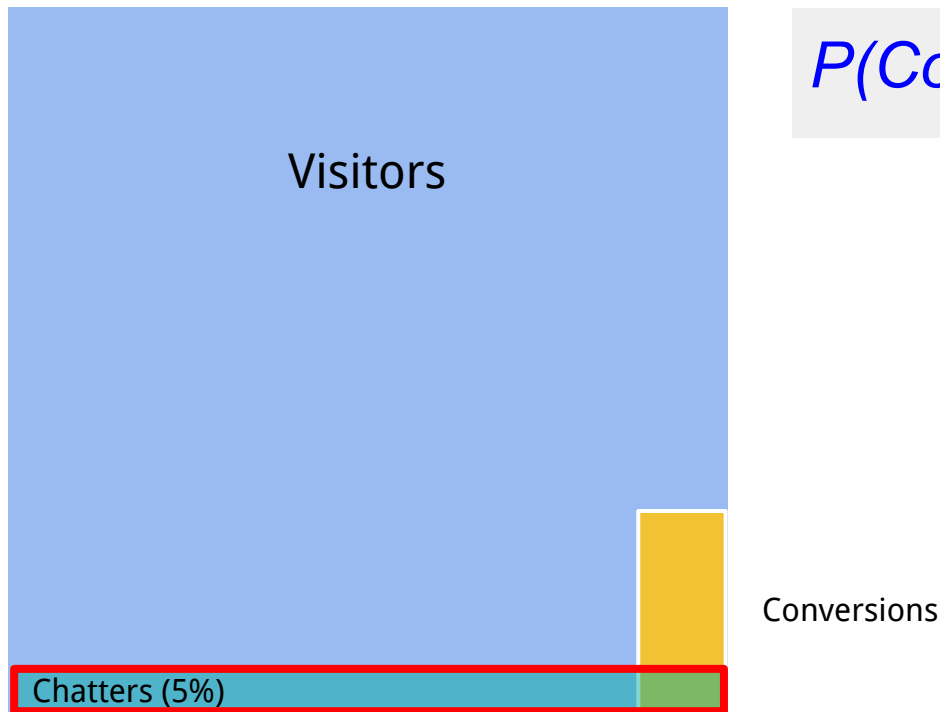
Increasing conversions - the Model



$$P(\text{Converted} \mid \text{Visited})$$

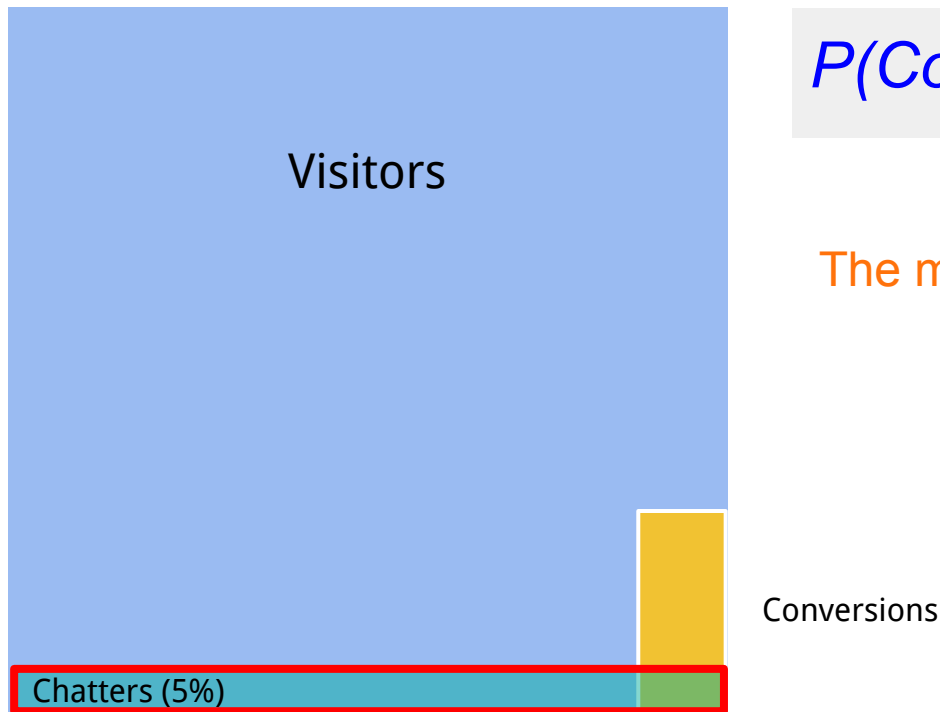
Chat => Convert signal is too small

Increasing conversions - the “large” model



$$P(\text{Converted} \mid \text{Chatted})$$

Increasing conversions - the small model



$$P(\text{Converted} \mid \text{Chatted})$$

The model won't generalize to all visitors

Why not both?

$F(P(\text{Converted} \mid \text{Chatted}), P(\text{Converted} \mid \text{Visited}))$

The moral

Understand the cost of a wrong prediction

Tale No. 3: How does the consumer feel?

Sentiment detection for conversations



The system - first iteration

The product: a line by line sentiment detection algorithm for conversations.

Machinery: Stanford CoreNLP library.

Lab: recall 30%, precision 30%.

Field: insufficient quality for product.

The system - second iteration

Retrain the model after tagging a large corpus at the phrase level.

am surprised I can chat with you

Very Negative

Negative

Neutral

Positive

Very Positive

[Save This Sentiment](#)

Lab: recall 30%, precision 40%.

Field: still insufficient quality for product.

The realization

- Stanford CoreNLP was conceived and trained for medium documents of high quality (movie reviews).
- Chat lines are very short and noisy documents.

The chat domain - an example

AUTOMATIC
RESPONSE

Hello, thank you for choosing ORG live chat. One of our agents will be with you shortly.

AUTOMATIC
RESPONSE

You are now through to Erika. How can I help you?

AGENT

Hi, Steve.

CUSTOMER

When will "MY ACCOUNT" be available ?

AGENT

I understand that you would like to access your My Account.

AGENT

We are working to bring back My Account and other services as soon as is possible. We know not having My Account is an inconvenience but until our engineers and external security teams are ready we will not return the site live. Until then you can chat with us or call our team to help.

CUSTOMER

That os not good enough

CUSTOMER

it has been 9 days

AGENT

I know how important this is for you. Let me go ahead and access your account so I can assist you with your query. Would that be Ok?

CUSTOMER

No - I don't trust you

CUSTOMER

Sd off goodbye

The solution

A custom built lexical model based and evaluated on actual conversation data.

Lab: recall 40%, precision 60%.

Field: Working, and well.

The moral

Always understand the assumptions and data that underlie the machinery you use.

Recap

- Understand your data and your funnel.
- Understand your impact on the whole system.
- Before using machinery review its assumptions.



Thanks

Ofer Ron

October 2017