# APPLYING MACHINE LEARNING IN MOBILE DEVICE AD TARGETING

Leonard Newnham
Chief Data Scientist

MOBILE VIDEO.
ARTIFICIAL INTELLIGENCE.
DATA.

# Introduction

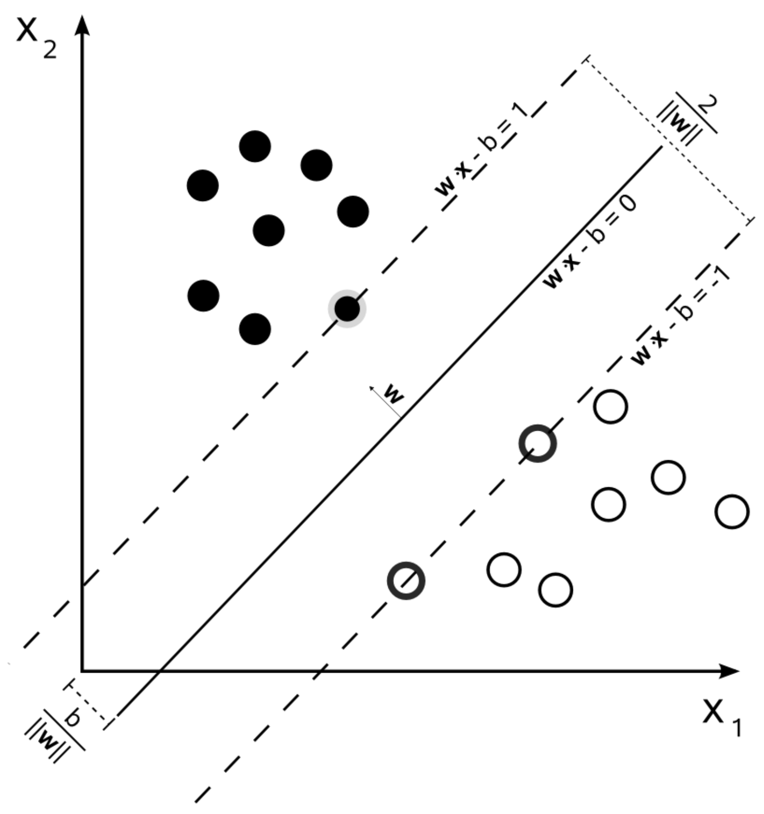
- Who is LoopMe?
- What we do
- The problem we solve
- Data
- Predictive models
- Bidders
- Future Research
- Lessons Learned

# Who is LoopMe?

- LoopMe is the world's largest mobile video platform, reaching over 1.25 billion consumers worldwide.

- London and Ukraine based start-up

- Machine Learning is at the core of everything we do



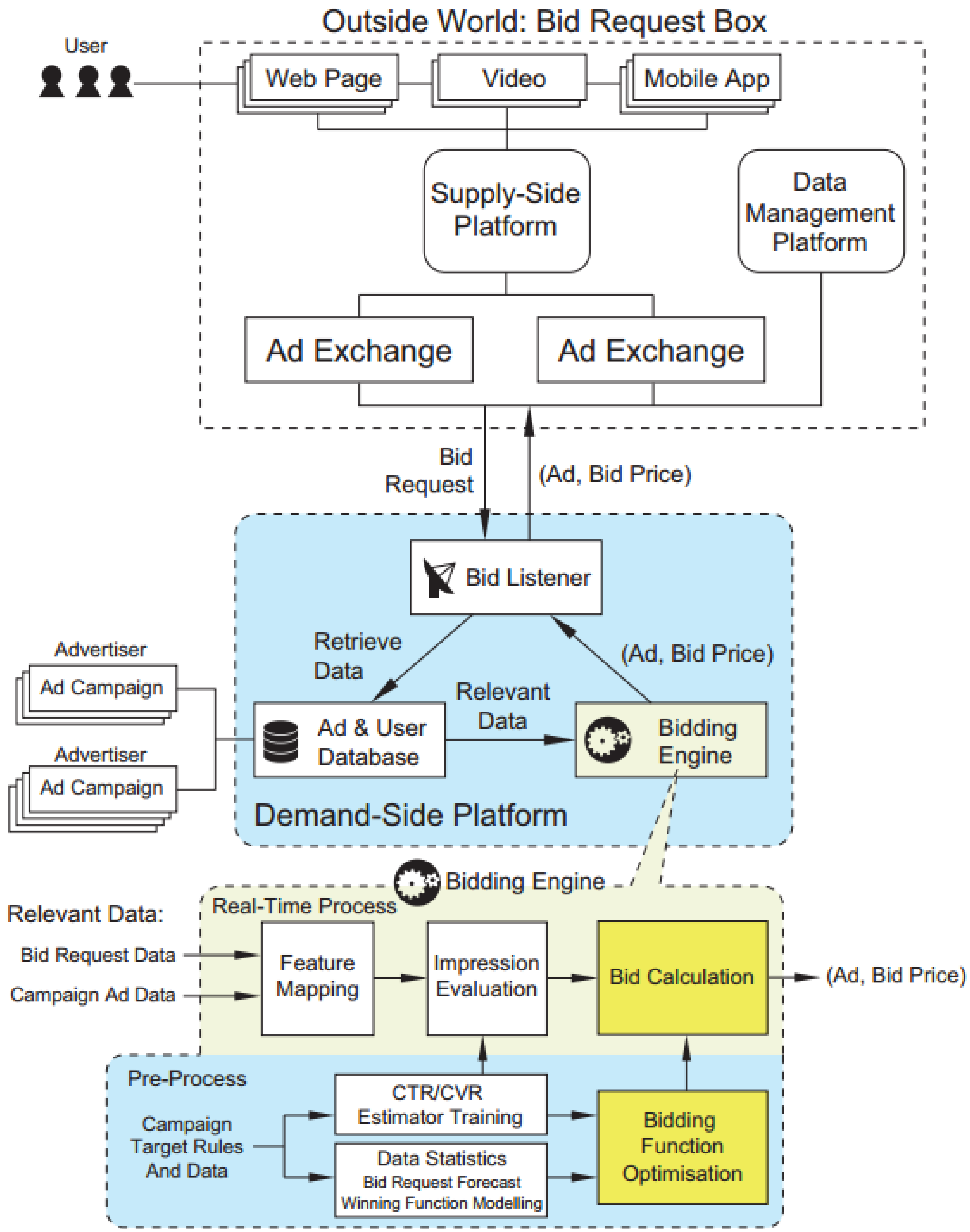
# What We Do

**Supply Side Platform**

- Real-Time Bidding Exchange
- Android/iOS SDK
- Web / VAST Tags

**Demand Side Platform**

Profile Assembly → Predictive Modelling → Real-Time Bidding




**Supply Side Platform**

**Demand Side Platform**
**- LoopMe**

# Ad Campaign

Contract with advertiser

- Number of impressions,
- Time period,
- Creative set,
- Country,
- White list
- Creative format,
- Optimisation goal: CPM, CPC, CPI

Example:


10,000 impressions per day for 28 days

LoopMe

# The Problem we Solve

Within milliseconds:

- Determine relevant creatives
- Score these creatives against KPI to optimise
- Determine whether to respond
- Determine how much to bid
- Respond

# How we do this

How do we do this:

- Collect data and build profiles
- Predictive models
- Bidding algorithms


data

# Base Data

User id
City
Country
Device width
Device height
Device OS version
Device OS
ISP
IP

Session depth
Orientation
Platform
Publisher
SDK
Time
App name
App company

Note:
All user IDs are anonymous
We don't store data if user opts out
PII data not stored

# Augmented Data - Segments

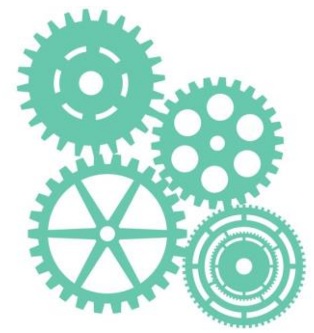
**Segments:**

Male/ Female
Age Range
Various game categories enthusiasts
Health and fitness fans
Messaging enthusiasts
Productivity apps users
Early adopters
etc…

**Three Sources**

- Simple rules:
  If owns App A, B or C
  then Computer Gamer

- Predictive models:

  profile data ->  -> p(Female) = 0.82

- Third party data:
  various commercial providers
  many small, free sources, eg weather.

# Augmented Data - Location

**Location with varying degrees of accuracy:**

- IP Address
  city, country, region

- Wifi name
  university or business name

- GPS coordinates
  nearby businesses and POIs
  Types of businesses visited

**From all three:**

Home/office location
Frequent Traveller
etc.

# Augmented Data – App and Creative Attributes

**App Data:**

- App name and App category
  Categorization of users based on apps used

**Creative attributes:**

- Age rating
  child, teen, adult, everyone, etc.

- Interactive
  whether interactive or not

- Audio
  has audio

- Type of content
  video, banner, pre-roll, click-to-play, etc.

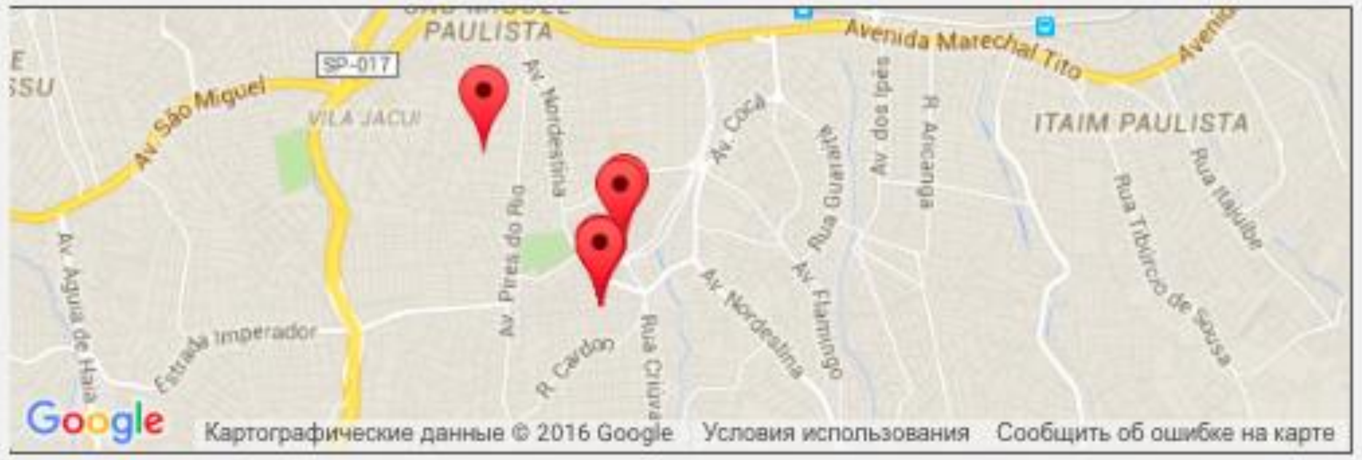
# Data: Example Profile

LoopMe — METRICS — PROFILES

Start typing device identifier

- 01cd410c-0611-4a2c-9171-92a8f0ba4cbd
- 03761a1e-4906-420d-bace-3365669786aa
- 03ea9e88-0a37-4312-9321-a011460d0ee1
- 0330c5df-cac4-47fa-b7b2-7682a2d6c575
- 01dd85ec-8912-4fb4-9a42-85091c406a75
- **03bcc1a6-846f-4182-aa79-e26b62cd3c2f**
- 02fb6bff-d3be-4c78-827c-f4b8b663fc93
- 0398f543-1332-40c8-b8a2-04edced9ef82
- 01cd4108-c67e-461c-9be8-5209c25b484a
- 00d740ee-59e4-4238-afd4-e165e59ab67b
- 04d9e4a5-01fd-42f9-ba0f-24e1ea705c4d
- 04ca6ee7-8431-4ac8-aff7-2ee6202200e3
- 00d7412a-3596-40c5-9d0e-af6c9fd15c36
- 03CEB314E78E4E548591E27D73C87A19
- 017dc890-2d45-4a42-be09-a03554473c1b
- 04d9e49c-b73d-4980-9539-af9871395464
- 01cd40c6-57b0-4e5d-b191-4b8df1a41566
- 0446FD0E-9D6F-4CFA-A457-ED6EDC44D4C5
- 000b96e2-200e-400f-a3d6-57bae2e63d09
- 00677fa2-cd0e-4384-95fc-fb2c22ea9bc8
- 030d4770-ae48-4e01-9425-f00ae6c89fa2
- 02cd1e3a-dfef-45b4-bd3f-7e45c4fbafdc
- 03d3776f-0376-4be7-b73d-7103ca5bcf2a
- --CC0fCwWBp1p5jHcuHmqSTpcmQ
- 0114d85d-8d43-49ae-9062-1e2254201cdc
- 023c3c8d-b3a4-408d-9bf8-d27f13b9fc73
- 02b5ece08bd4cb4bb3dd366d40626de0d0fedfb9

| | |
|---|---|
| App categories | Tools: 251 |
| App genres | — |
| Installed Apps | com.cleanmaster.security: 109<br>com.cleanmaster.mguard: 102 |
| Content Age Rating | — |
| Keywords | m_age:21,m_interest:11: 54<br>m_age:27: 29<br>m_age:35: 9<br>m_age:21,m_interest:43,m_interest:11,m_interest:21,m_interest:1,m_interest:8,m_interest:19,m_in<br>31 |
| Visited Sites | — |
| **Channels** | |
| API Names | rtb: 251 |
| Exchanges | mopub: 251 |
| Cellular home ip | — |
| Cellular home isp | — |
| Cable home ip | — |
| Cable home isp | — |
| ISP Names | Vivo: 251 |
| Wifi Names | — |
| **Geo/Location** | |
| Home country | BR |
| Home region | 27 |
| Home city | São Paulo |
| Home zip | — |
| Hard traveler | No |
| Visited countries | BR |
| Visited locations | |

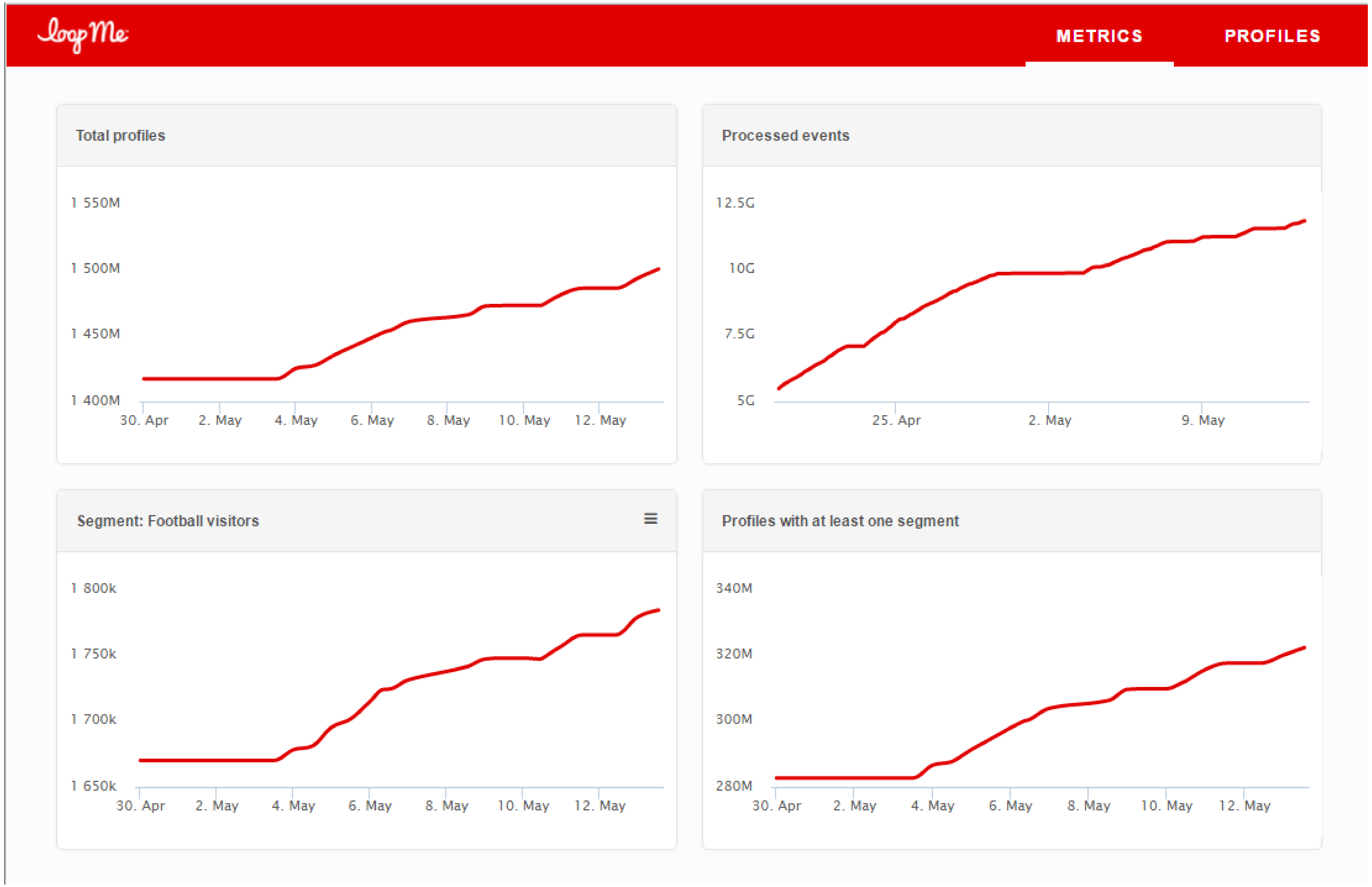



**-23.513613,-46.436603**
Time: 2016-05-05T09:26:15.000Z; App: 15876

**-23.513576,-46.436611**
Time: 2016-05-03T23:45:16.000Z; App: 15876

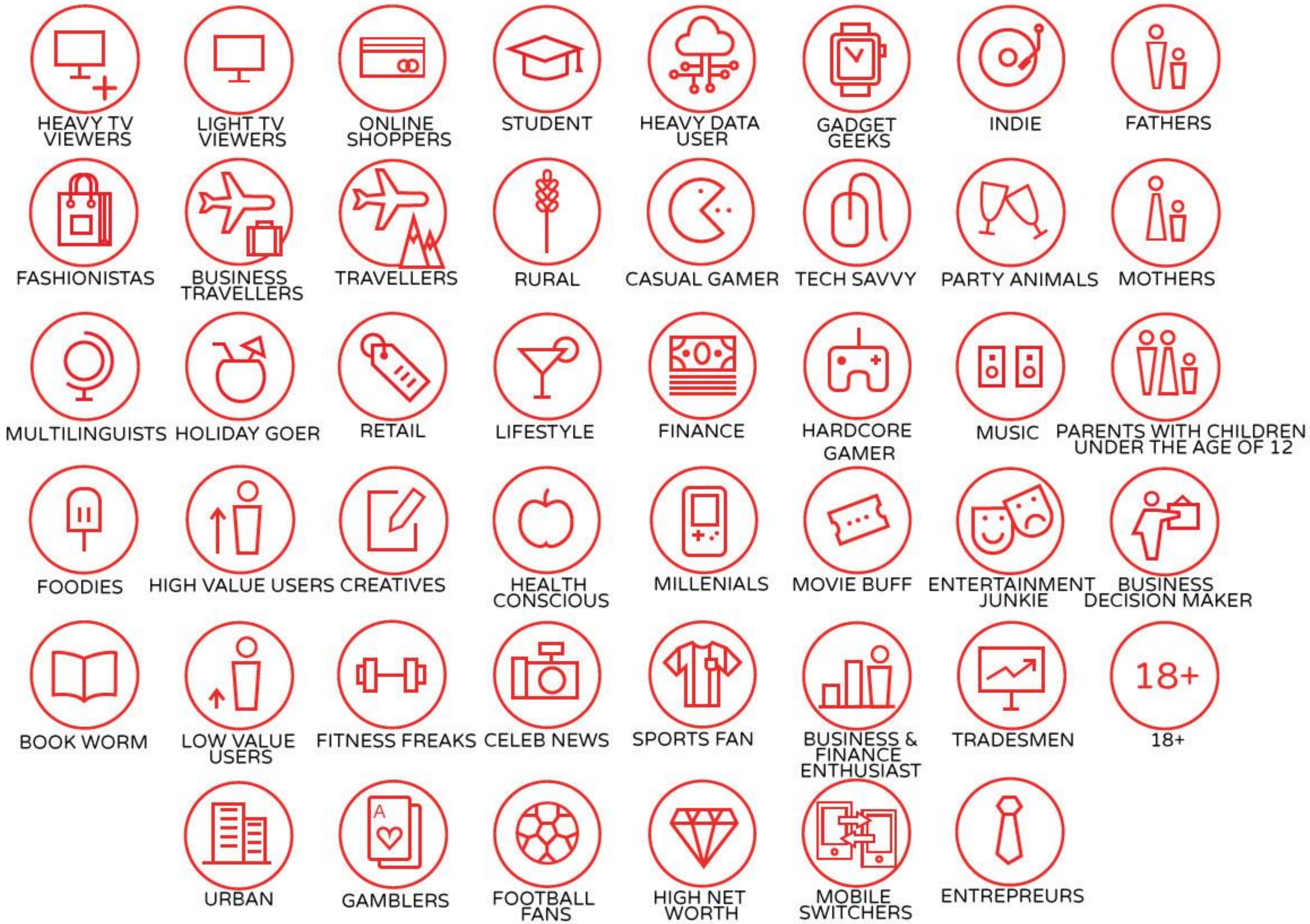
**-23.509518,-46.435055**
Time: 2016-05-03T20:41:23.000Z; App: 15876
Time: 2016-05-03T19:14:38.000Z; App: 15876
Time: 2016-05-03T16:01:08.000Z; App: 15876

# Data Growth


METRICS
PROFILES
Total profiles
1 550M
1 500M
1 450M
1 400M
30. Apr
2. May
4. May
6. May
8. May
10. May
12. May
Processed events
12.5G
10G
7.5G
5G
25. Apr
2. May
9. May
Segment: Football visitors
1 800k
1 750k
1 700k
1 650k
30. Apr
2. May
4. May
6. May
8. May
10. May
12. May
Profiles with at least one segment
340M
320M
300M
280M
30. Apr
2. May
4. May
6. May
8. May
10. May
12. May

# Data: Segments

Turning raw data into useful data



# Build Models

# Data Sets and Predictive Model Building

We Match Campaigns to People


# Data Sets and Predictive Model Building

**Of course, it is not always this easy…**

many dozens of factors involved

**However, we do need to model interactions**

Campaign A appeals to Person Type X
Campaign B appeals to Person Type Y

**And do it in real-time**

**There are several ways to do this…**

# Data Sets and Predictive Model Building

**Single simple model, eg logistic regression**

- Learns weight for each binary feature

- Problem
  Cannot easily learn
    campaign A appeals to men
    campaign B appeals to women

# Data Sets and Predictive Model Building

**One model for each campaign**

- Now we learn exactly what type of people campaigns A and B appeal to.
- Problems:
  1. takes a long time to get sufficient data for a new campaign
  2. No learning is transferred to new campaign
  3. Learning common to all campaigns is learned multiple times

# Data Sets and Predictive Model Building

**Single Model with interaction features between campaign and other variables**

New features:

campaign=A *AND* gender=M,   campaign=A *AND* gender=F,…

- Now we learn common learning just once and efficiently

- We can learn campaign A appeals to men and campaign B appeals to women

- Problems

  Much learning lost when a campaign is terminated and replaced by similar one

# Data Sets and Predictive Model Building

**Single Model with interaction features between campaign features and other variables**

say:

campaign A is for tennis rackets
campaign B is for tennis balls

both sporting products

Add attribute to a campaign of product type

If campaign B is replaced by campaign C, also sporting product
-> much learning common to all sporting products transferred

- Problems

  We get a lot of features. Easily >1 million for real application
  -> problem with noise.

# Data Sets and Predictive Model Building

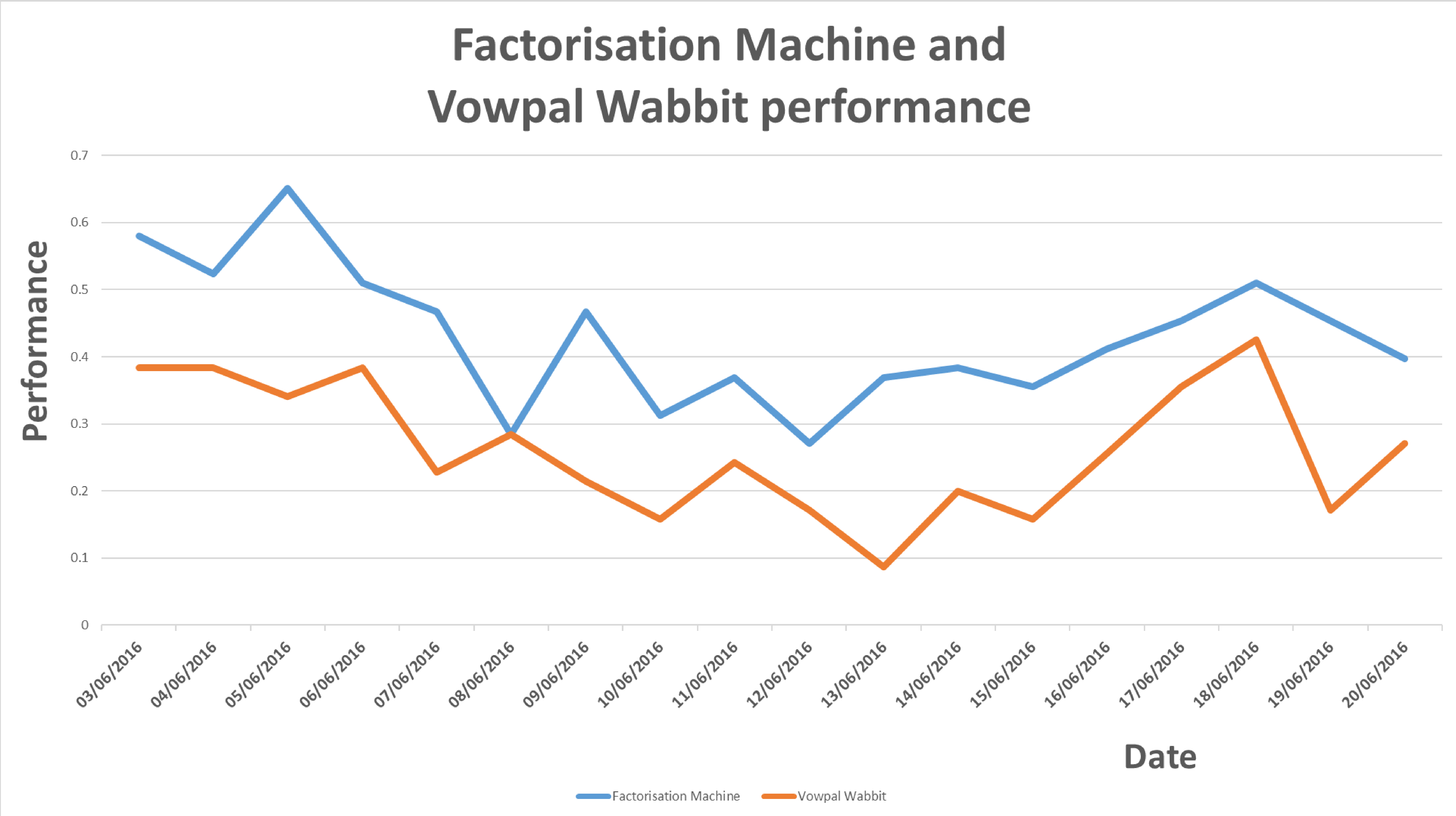
**Factorisation Machine**

- Learn
  - bias
  - 1-way interactions
  - 2-way interactions and factorise these

Have all interactions between campaign features and other variables

Result: manageable model size

# LibFM Performance



# Bidders

# Bidding algorithms

- Ad Exchanges run second price auctions
- For CPC – cost per click, a simple strategy

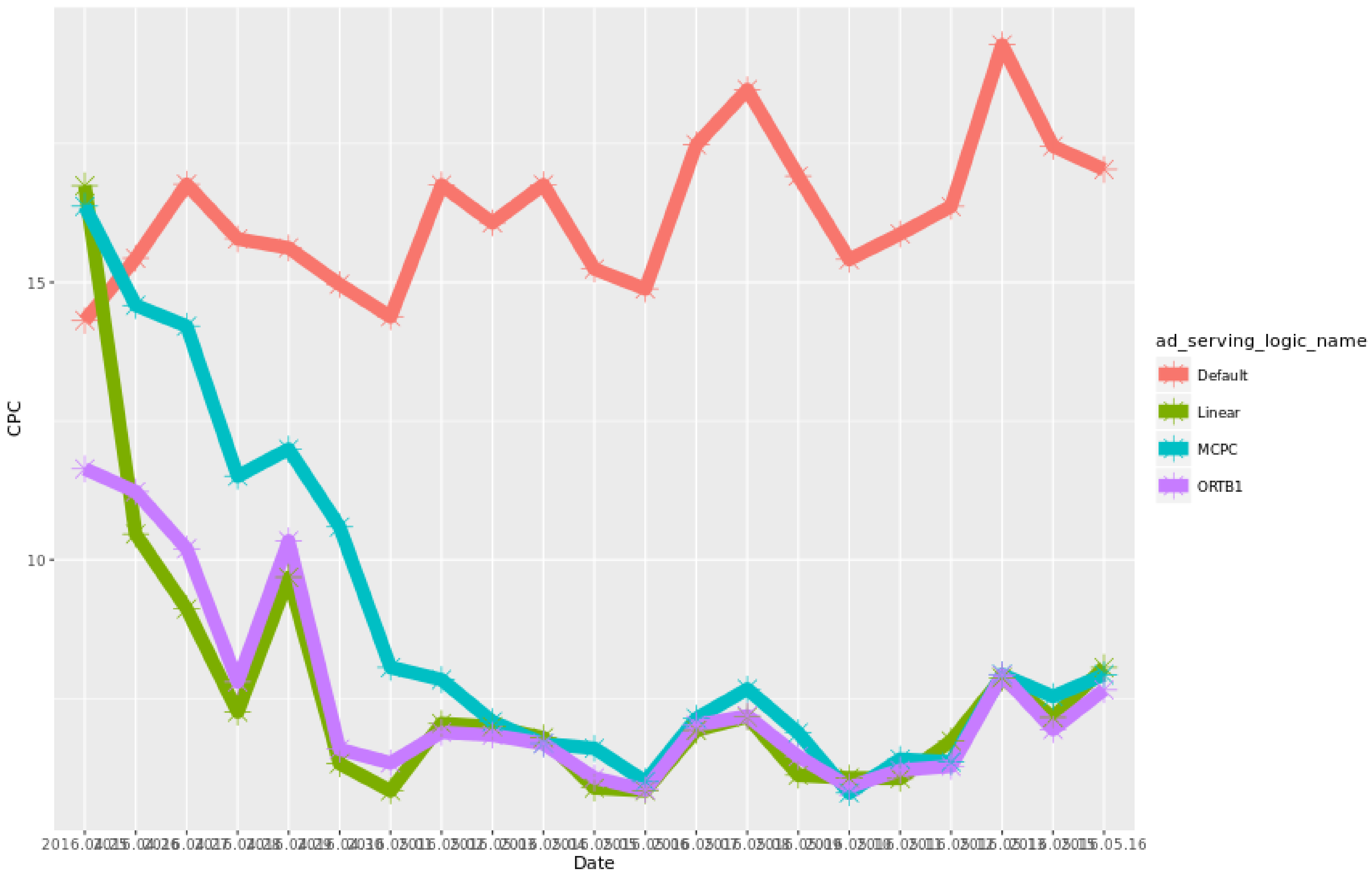
  Expected return = p(click) * CPC
  Bid this amount

- Total budget and campaign lifetime mean this is not optimal…

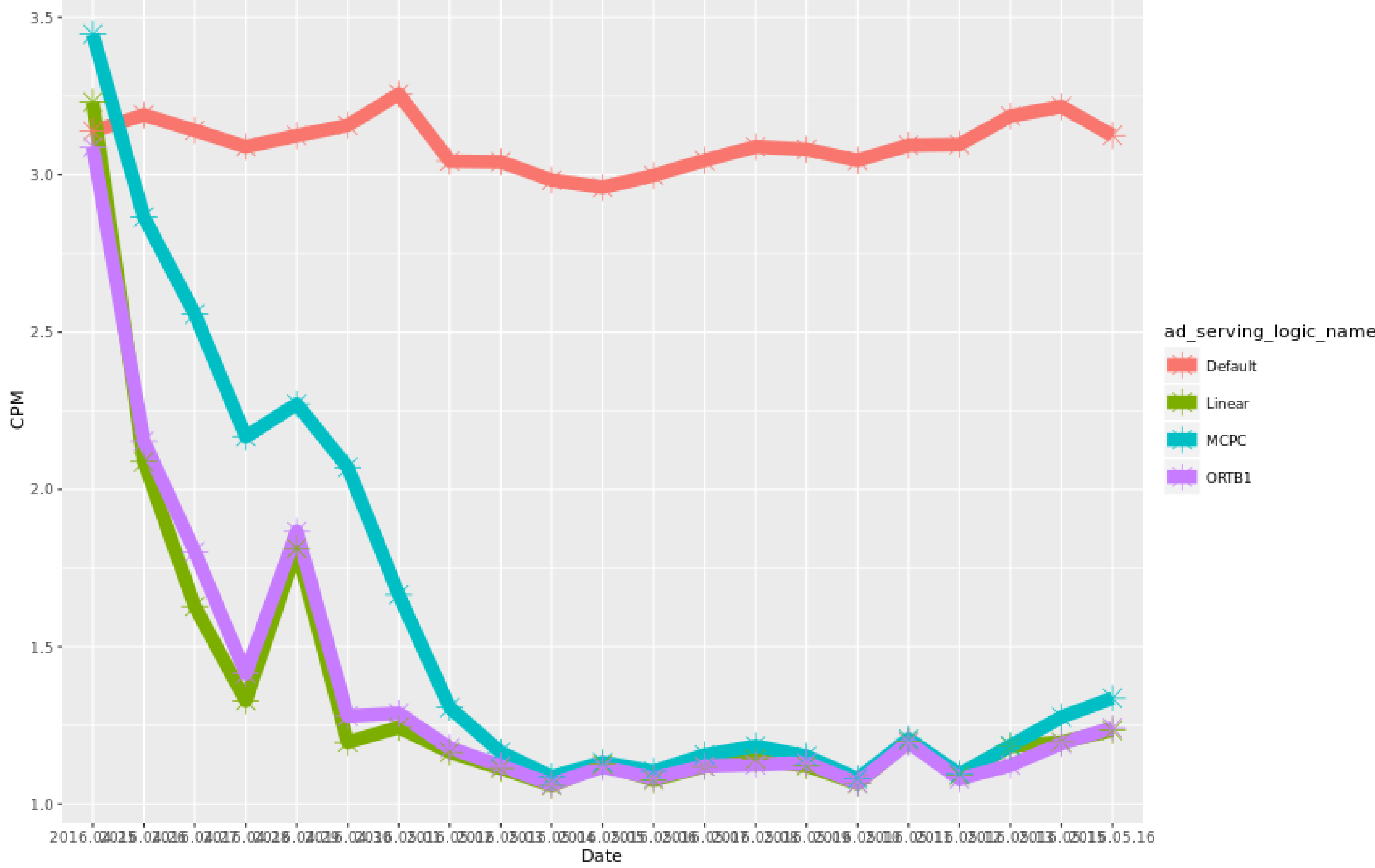
# Bidding algorithms

Currently experimenting with different bidder types.
Open area for research

## CPM chart




## CPC chart



# Future Research

# Future Research Areas

## Digital fingerprinting

- where there is no persistent device ID
- mobile web

## Changing behaviour

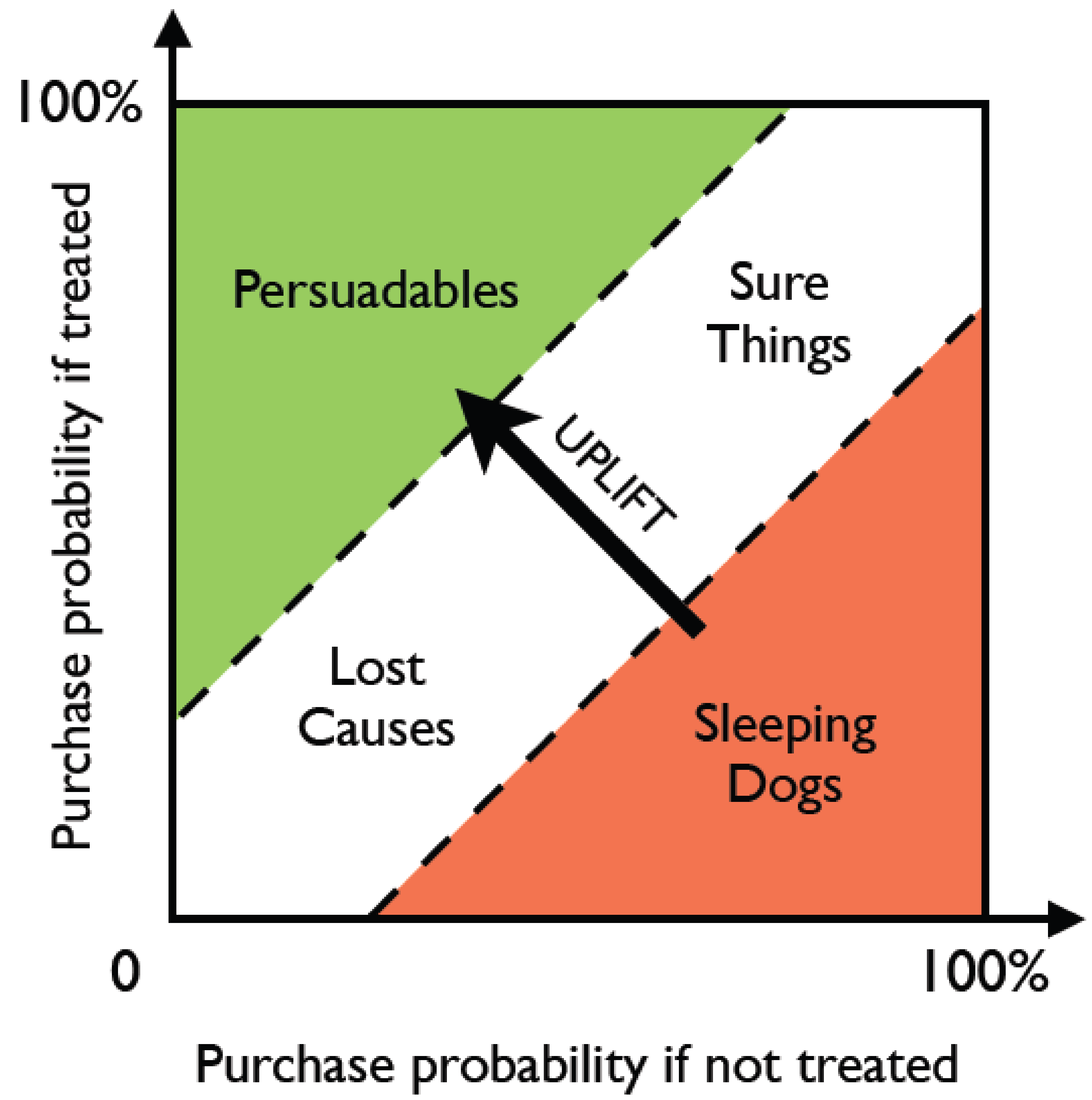
- How do we target the individuals where we can change behaviour?
- Not just those who click the most

## Low Frequency Events

- beyond clicks and installs
- advertisers are interested in people taking actions

# Changing Behaviour

- Traditional “response” models have a tendency to direct resources towards customers who would have bought anyway
- This often results in strong models but comparatively few incremental sales
- The customers who spend most after being subject to a marketing intervention are *not necessarily* the ones whose spending increases most as a result of that intervention

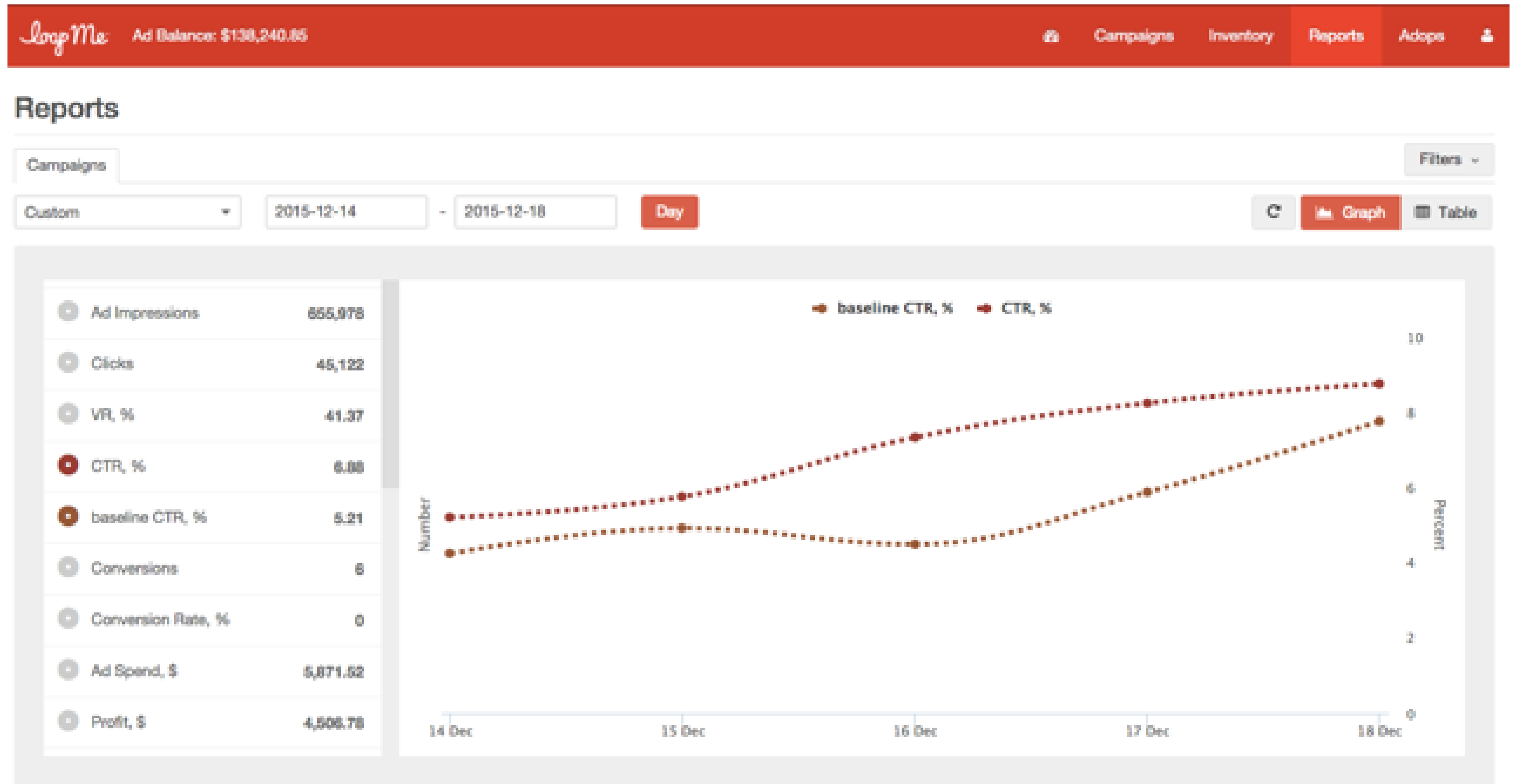



Lessons
Learned

# Concurrent and Persistent Control Groups

- First time visitor seen, randomly assign to customer group, typically:

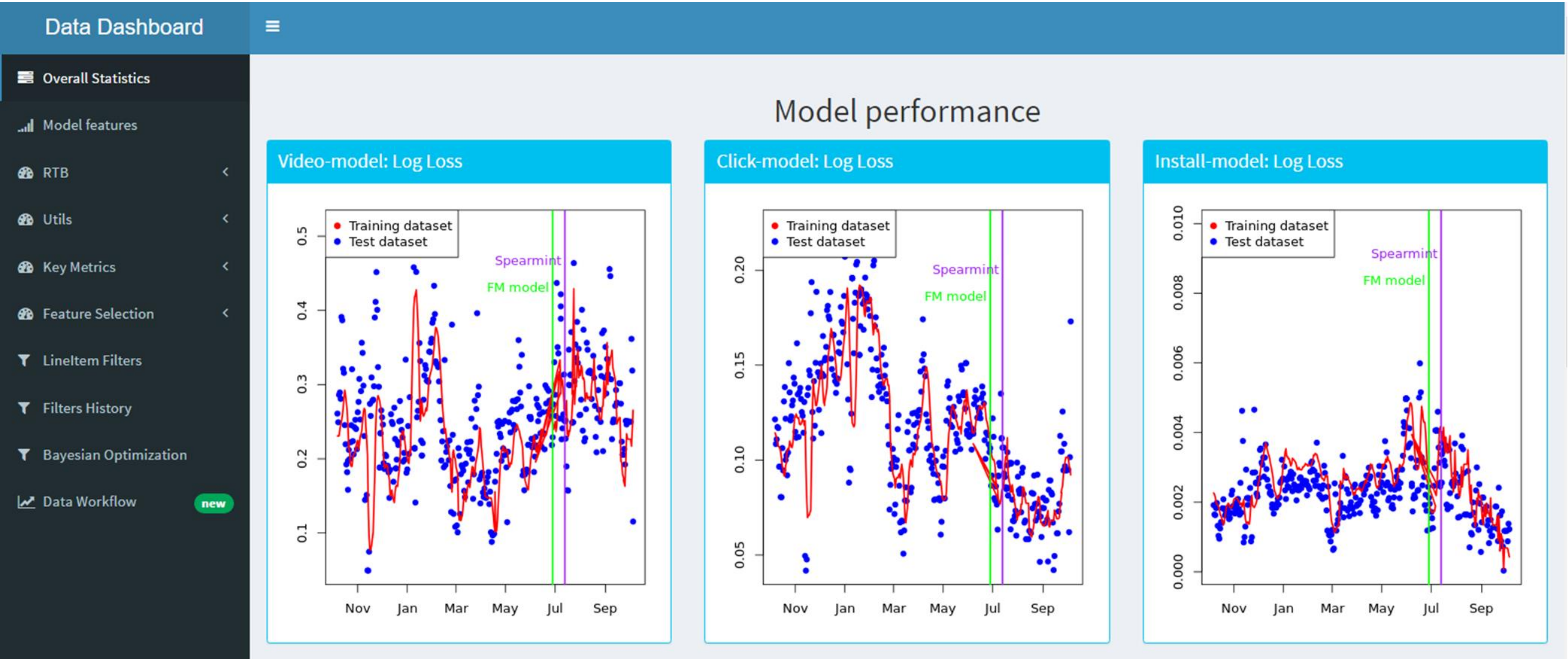
  90% AI group – always receive best prediction of AI
  10% baseline group – business-as-usual, served without using AI




- Concurrent control groups give the most accurate measure of uplift eliminates errors due to changes over time

- Uplift = AI performance / baseline performance

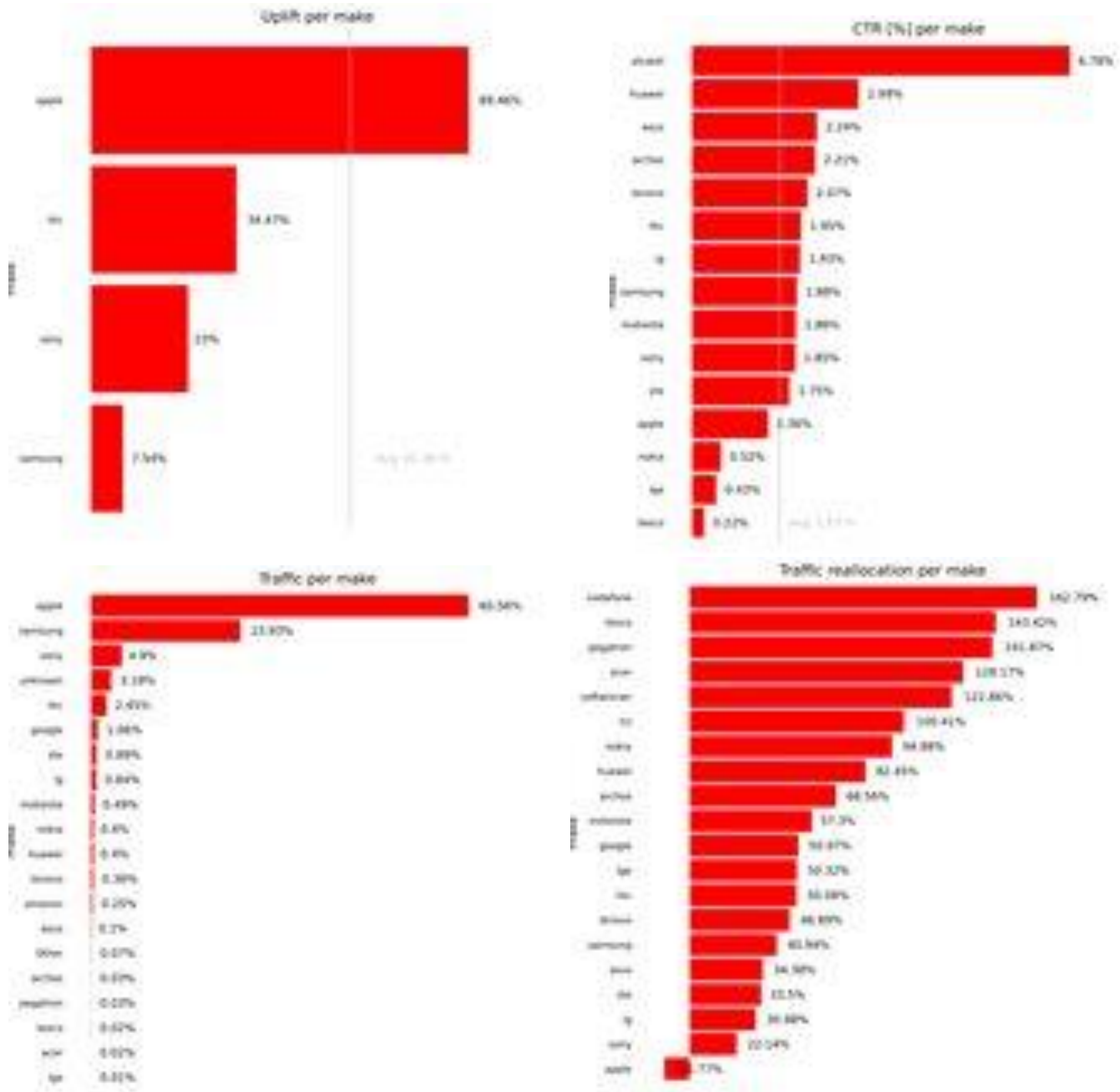
# Measure Everything



# Visualisation

**Improved Visualisation:**

- AI uplift
- AI audience insights
- What AI has learned



# Lessons Learned

- Use concurrent control groups
- Measure everything
  dashboard pages for various views of the system
- Visualisation of results
- Investigate every issue
- Don’t rely only on high level metrics like log-loss
  Look at the detail as well

# THANK YOU

leonard@loopme.com