adform

RTB Bid Landscape survival models powered by bitmaps

Łukasz Mączewski Senior Data Scientist
Przemysław Piotrowski Senior Data Engineer

AI&Analytics, Adform
14.05.2019

Agenda

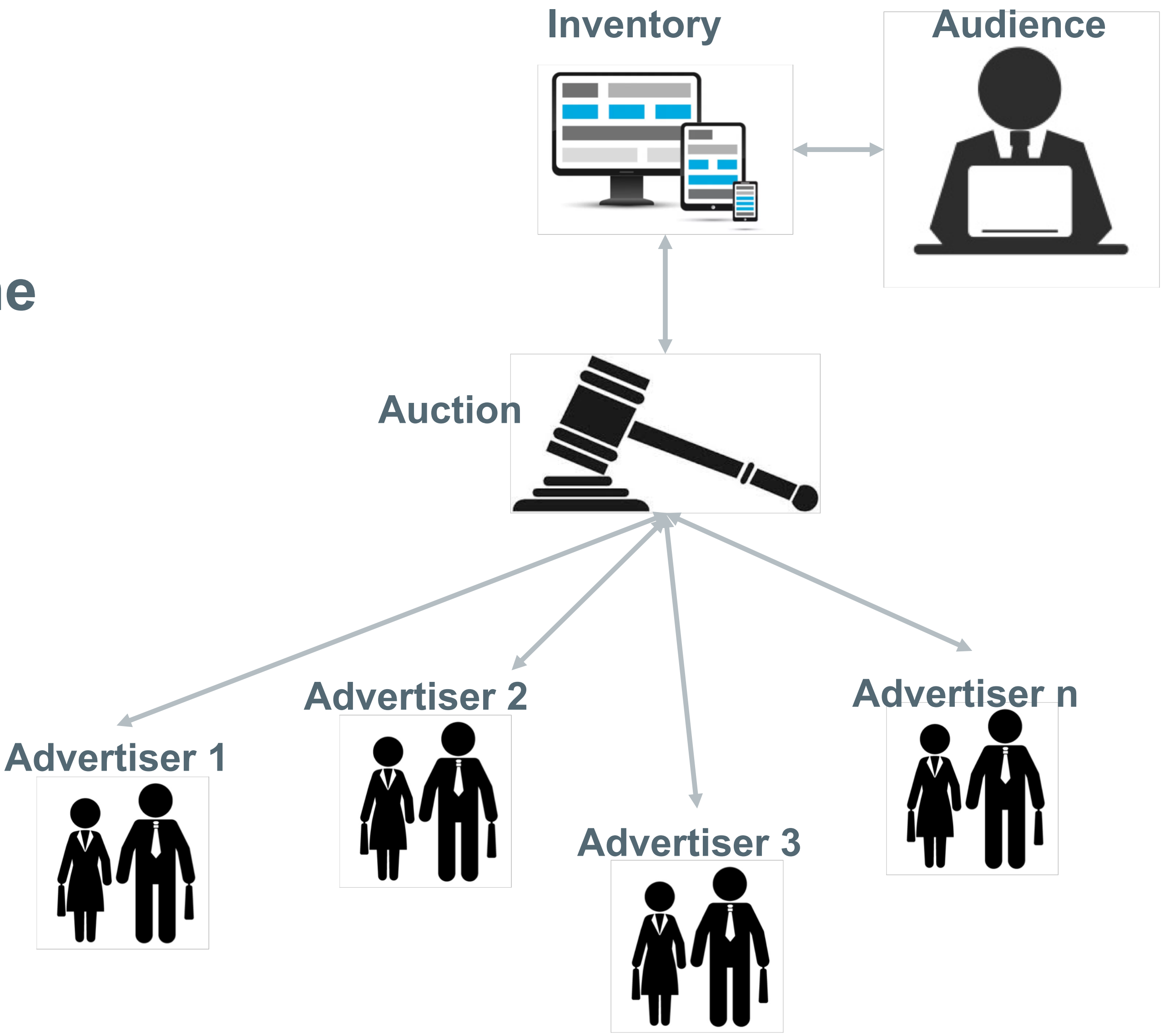
- Real Time Bidding – RTB
- Bidding Strategy
- Bid Landscape
- Bitmap population selection
- Survival analysis
- Validation
- Product demo

Real-Time Bidding (RTB)

RTB allows for fully automated ad inventory selling / buying in a real-time

Demand Side Platform is to estimate optimal
bid price according to:

- Targeting (audience)
- Business requirements (KPIs)
- Ad inventory (context, ad slot price)
- Campaign budget and it's spend plan



First price vs. second price auction

First price auction:

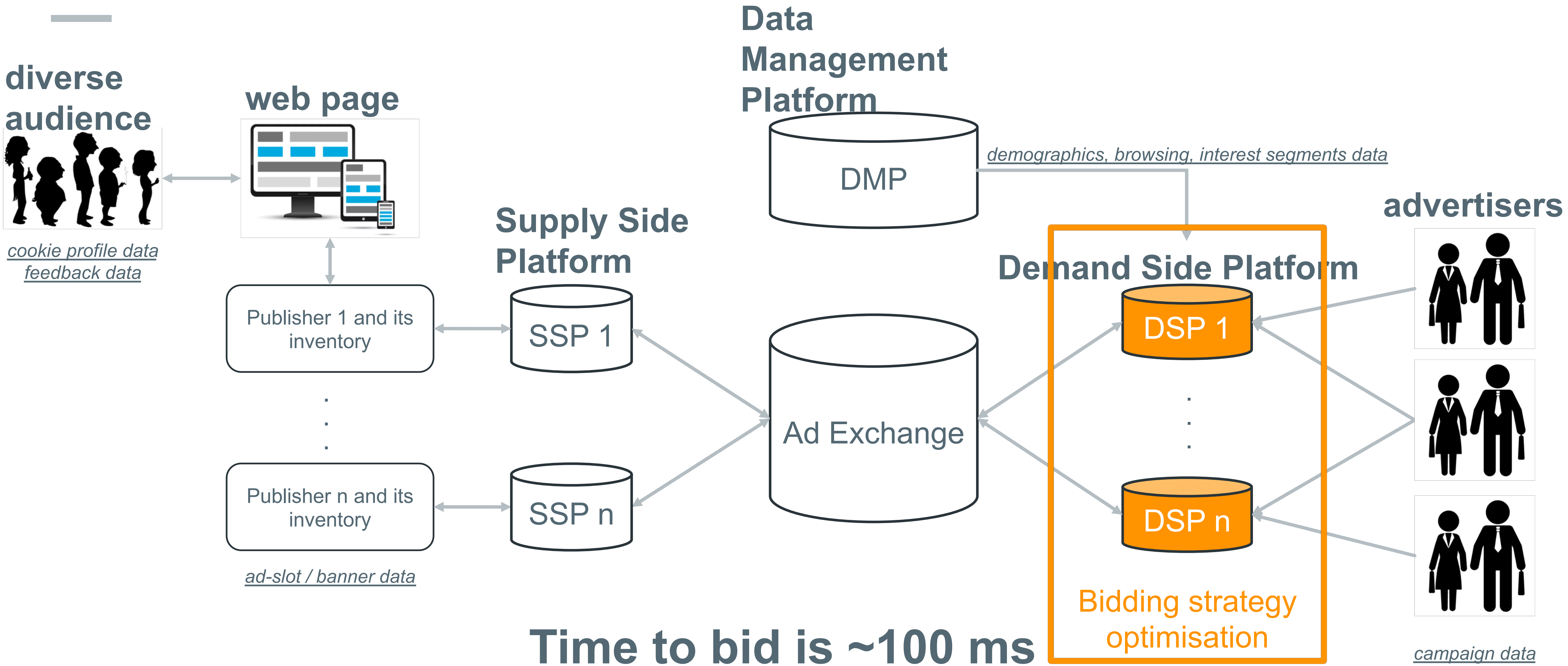
- Pay exact amount that was bidded
- Bid price is close to a real impression value
- Seller revenue maximization

Second price auction:

- Pay 0.00001$ more than second-highest bid price
- Bid as high as possible to maximize winning probability

In the past the second price auction was preferred, but currently the first price auction becomes more and more popular

RTB ecosystem



Time to bid is ~100 ms

Bidding strategy optimisation

DSP algorithms

Evaluation of bid request business potential

Bid Landscape

Estimation of winning probability at bid price

Budget throttle

Control budget spend

Bidding Strategy Optimization

Candidate impression valuation with respect to KPIs and budget constraints

Optimal bid price

Numbers

Adform provides an integrated Software as a Service platform for the buying, managing and serving of digital advertising.

Presence in EU, US, APAC

3 millions bid request per second

billions of engagements daily

5 millions scoring of algorithms per second

200 thousands of cookie profiles in DMP

Algorithms

Discovering knowledge from massive RTB data

Cookie profile/ Targeting:
- device type
- os type
- browser type
- screen size
- country
- regions
- city
- clicks made
- visited log points
- visited domains
- urls
- IAB categories
- ...

Data Management Platform:
- demography
- interests and preferences
- profession
- ...

Adslot/ banners/ creatives:
- inventory name
- domain
- URL
- banner position
- banner size
- banner type
- floor price
- video player properties
- ...

Feedback:
- acquisition/conversion
- click
- order status
- revenue
- viewability
- ...

Campaign:
- budget
- pricing type
- deals
- targeting
- brand safety
- ...

Discovering knowledge from massive RTB data

Data summary

Huge class imbalance + Mostly categorical data + High cardinality + Sparsity = High risk of overfit

Data sampling

Ensembling

Preselection of features

Reduction of dimensionality

Regularisation

Cross validation

RTB models

CPC
Cost per Click

CPL
Cost per Lead

netCPL
Confirmed lead

CPV
Cost per Viewability

ROAS
Return on Ad Spend

VCR
Video Completion Rate

Adform machine learning pipeline

Train/Test data sets preparation (dedicated ETLs)

Data preprocessing and feature engineering

Feature selection

Building model and hyperparameters optimization

Offline backtests

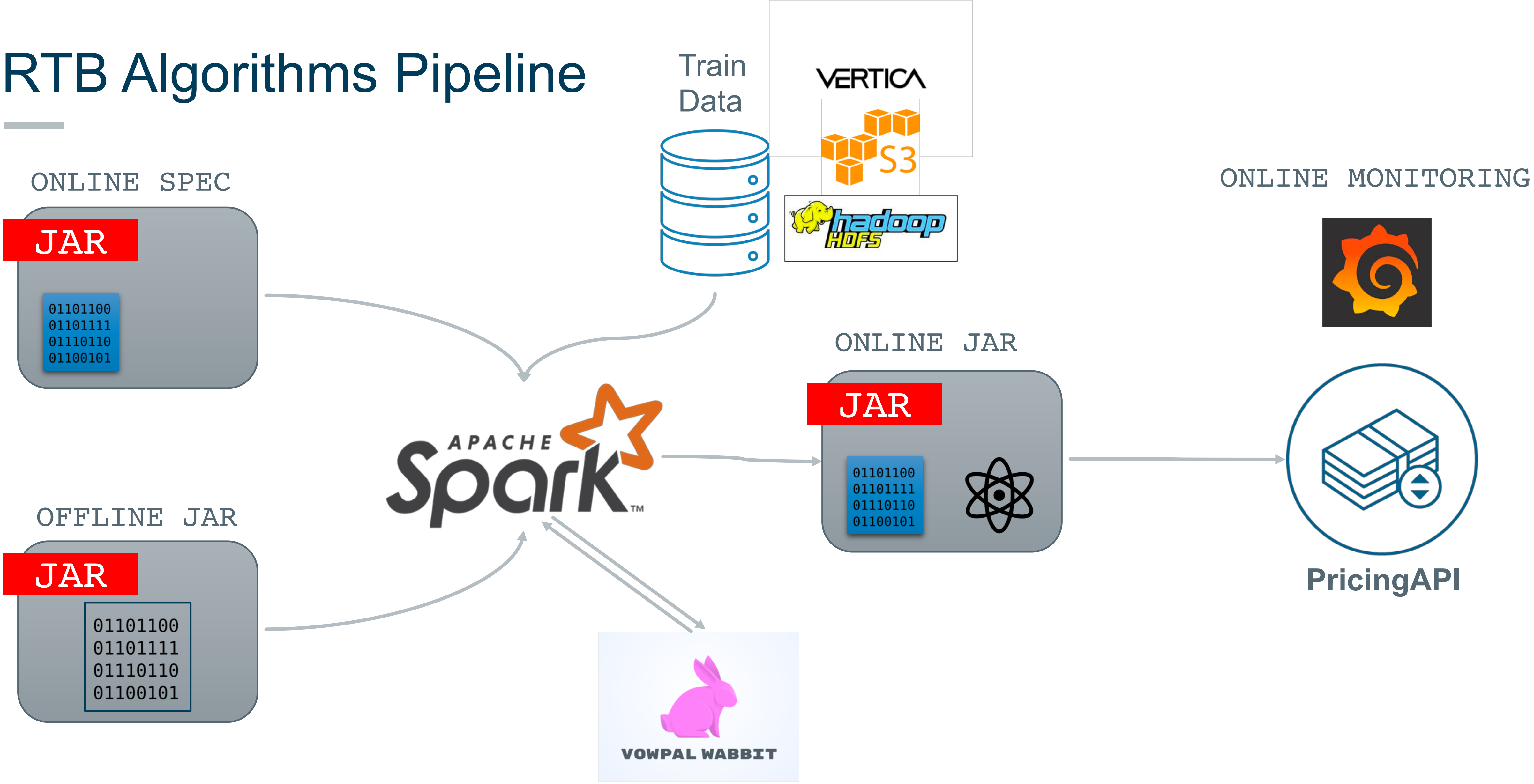
Online A/B tests

Deployment

Live monitoring

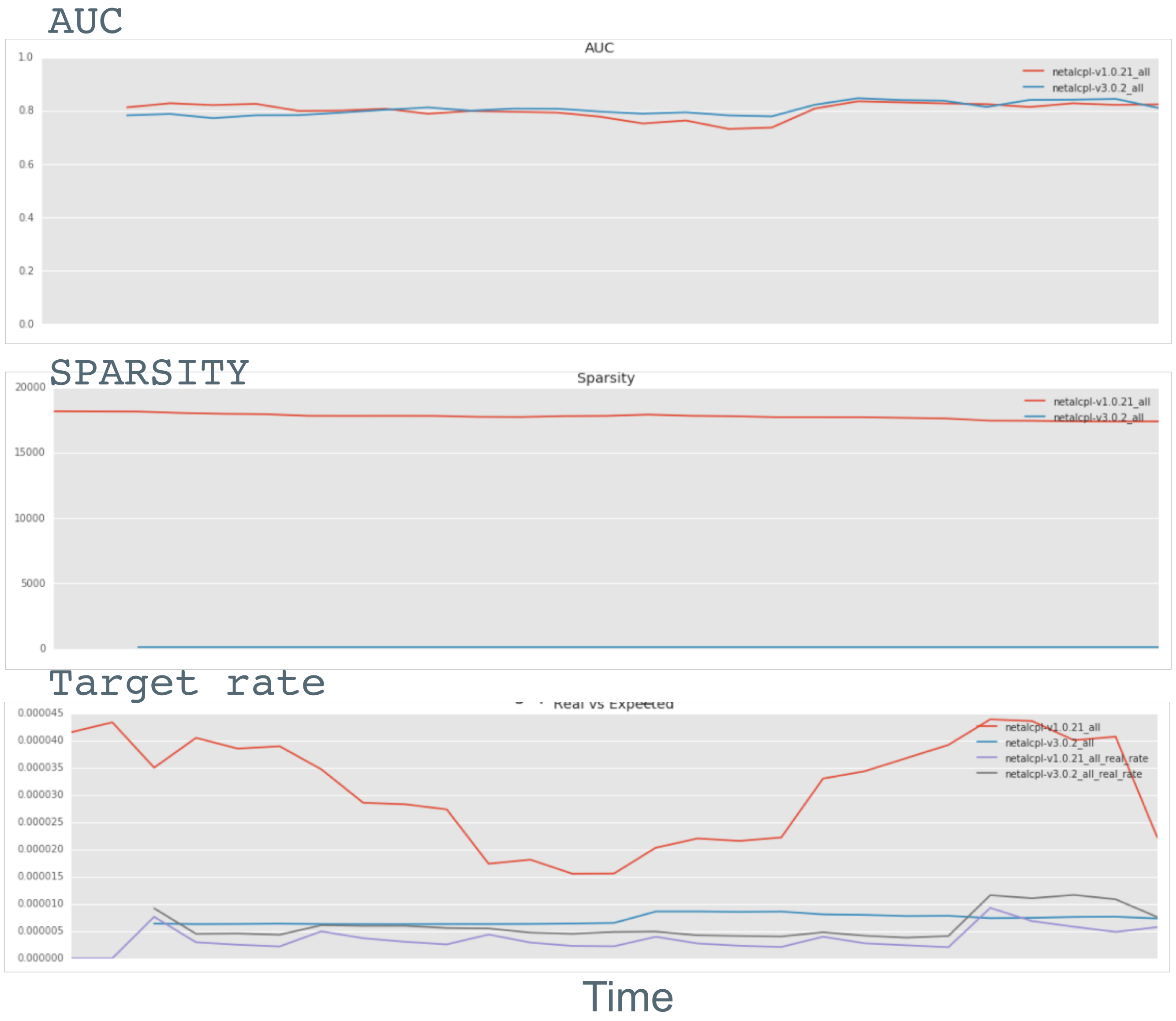
After successful model deployment it is periodically retrained

RTB Algorithms Pipeline



Backtests

- Offline procedure to check performance of two competing models in time
- Sequential analysis of historical data:
  - Data divided into training and test sets
- Time series analysis of critical performance KPIs:
  - AUC
  - Sparsity
  - Target rates
- If the backtests succeed the algorithm is submitted to online A/B testing procedure

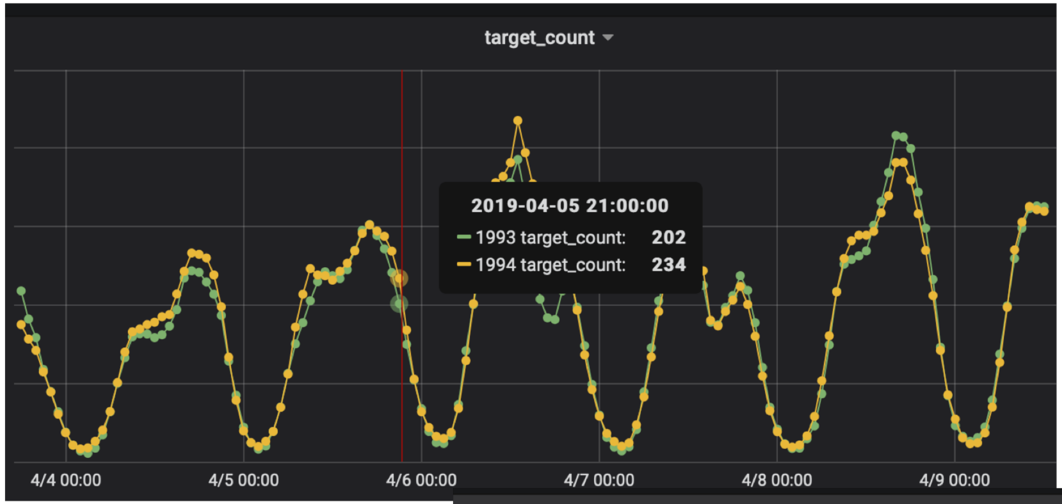


A/B test online monitoring

Split traffic to `hash(cookie_id)%1024` buckets

For each experiment monitor

- Cost per target
- Target rate
- Target count
- Number of impressions
- ... many more



adform

Questions?

Bid Landscape

User Stories

As a Trafficker I want to know **volume of impressions** my campaign would win in **RTB** given the selected inventory, targeting and **budget settings** (CPM).

Bid Shading

Demand Side Platform **protects** Advertiser from overpaying while
AdTech ecosystem **migrating** from second-price to **first-price** RTB auctions.

As a DSP I want to enable Advertisers winning impressions for lowest possible price.

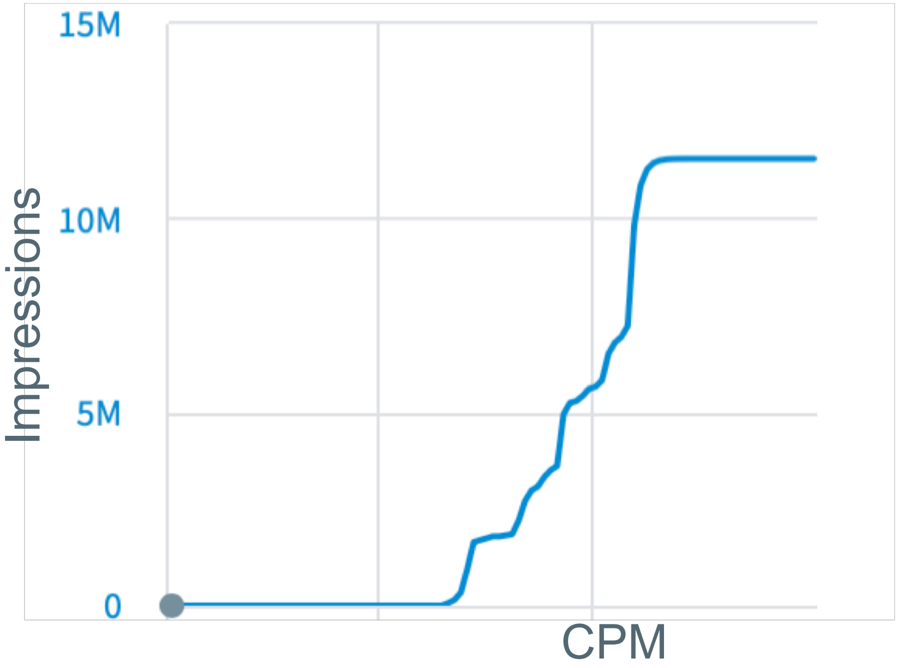
Forecasting & Bid Landscape

Forecasting

- Total number of matching Bid Requests for given Campaign Targeting
- Bid Opportunities

Bid Landscape

- Predicting **probability** of winning auction with a given bid price and Campaign Settings
- Win rate evolution with a bid price



Bid Landscape

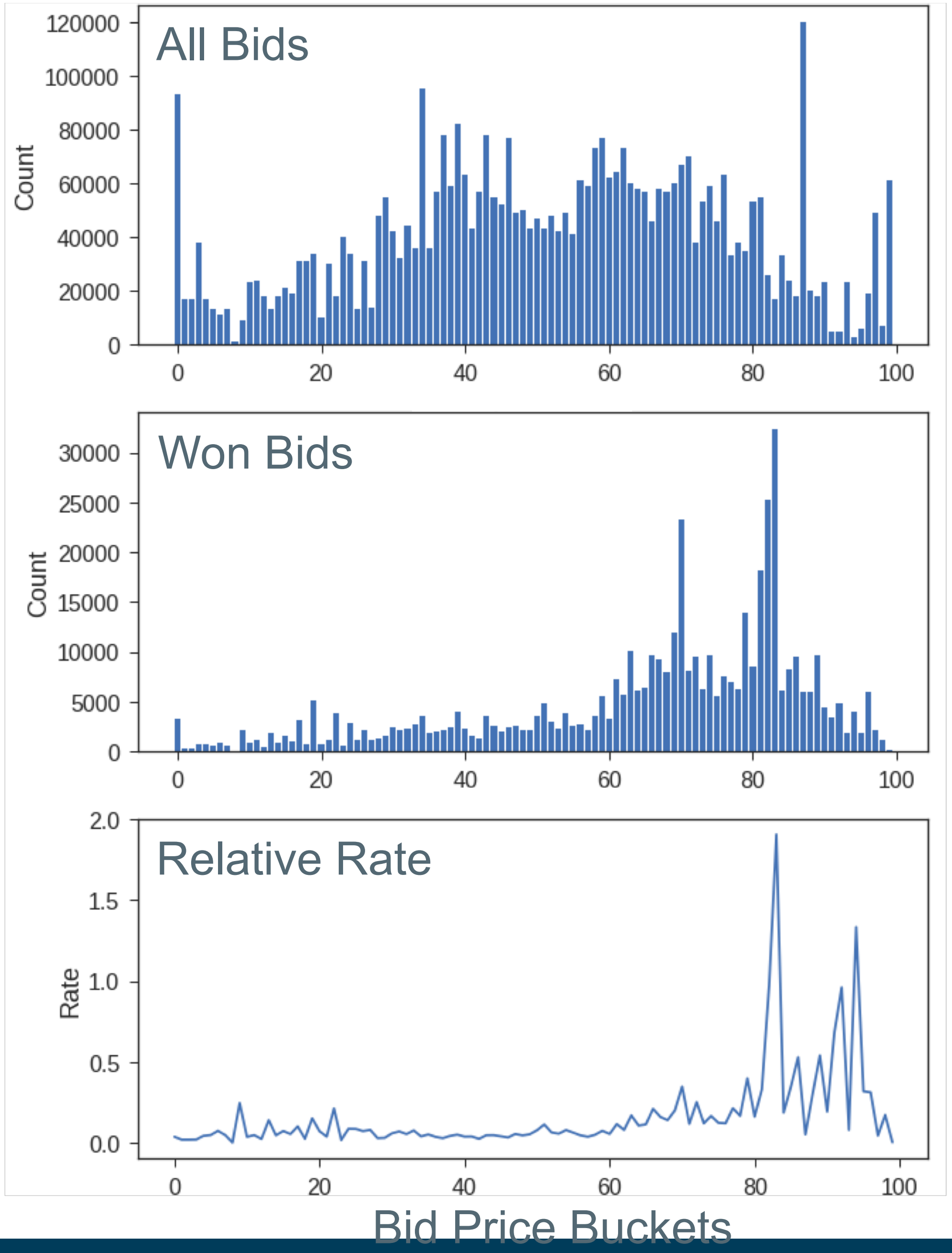
- Building Bid Landscape model using historical data regarding won and failed auctions:

  For lost auctions win price is unknown

- Right Censored Data

  Survival Analysis

  The exact moment of failure for majority of objects is unknown



Population selection for Bid Landscape modelling

The Data

Input

Bids and Impressions with features

Impression is a Bid that won the auction

Output

cardinality of selected (Campaign Settings)

Bids bucketed by **bid price**

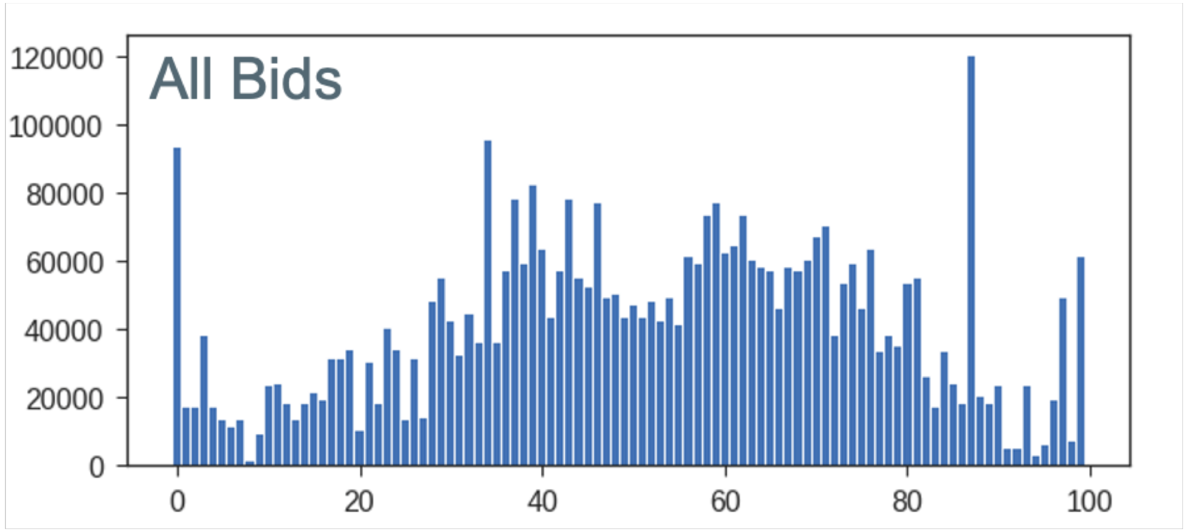
Impressions bucketed by **bid price**

Impressions bucketed by **win price**

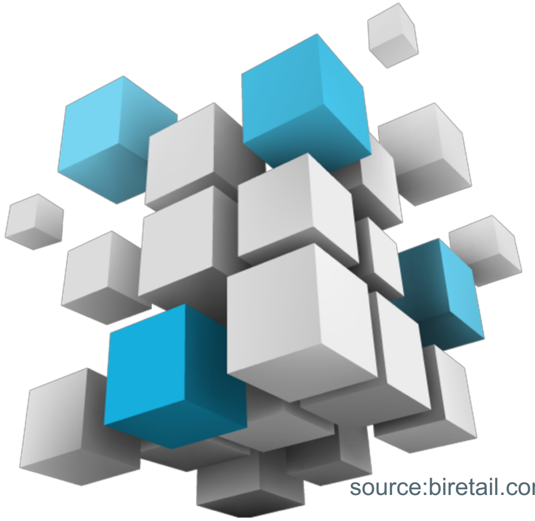
100 buckets in our case

How?

Population Selection – select relevant population from high-dimensional space for survival analysis.



100000	2000	5000	100	4000	3000	...	4500



source:biretail.com

Bitmaps (bitset, bitvector)

Bid	iOS	waw	11:00	Matching bids
0	0	1	1	0
1	0	0	1	0
2	1	0	0	0
3	0	0	0	0
4	1	1	1	**1**
5	1	0	0	0
6	0	0	0	0
7	1	1	1	**1**
…	…	…	…	…
100mln	0	0	0	0

Bid Price Buckets

Counts

9900

790

2000

….

100

Nested queries on millions of features

AND, OR, NOT, XOR

Bid	iOS	waw	11:00	3 mln other features	Matching bids
0	0	1	1		0
1	0	0	1		0
2	1	0	0		0
3	0	0	0		0
4	1	1	1		**1**
5	1	0	0		0
6	0	0	0		0
7	1	1	1		**1**
…	…	…	…		…
100mln	0	0	0		0

Bid Price Buckets

Counts
9900
790
2000
….
100

Buckets' cardinality

Bids

	iOS	waw	16:00
0	0	1	1
1	0	0	1
2	1	0	0
3	0	0	0
4	1	1	1
5	1	0	0
6	0	0	0
7	1	1	1
…	…	…	…
100mln	0	0	0

10000	3000	...	4500

Impr. Win Price

	iOS	waw	16:00
0	0	1	1
1	0	0	1
2	1	0	0
3	0	0	0
4	1	1	1
5	1	0	0
6	0	0	0
7	1	1	1
…	…	…	…
100mln	0	0	0

2000	100	...	0

Impr. Bid Price

	iOS	waw	16:00
0	0	1	1
1	0	0	1
2	1	0	0
3	0	0	0
4	1	1	1
5	1	0	0
6	0	0	0
7	1	1	1
…	…	…	…
100mln	0	0	0

1000	800	...	400

Survival Analysis

Roaring Bitmap

Roaring Bitmap – state of the art compressed bitmap algorithm.

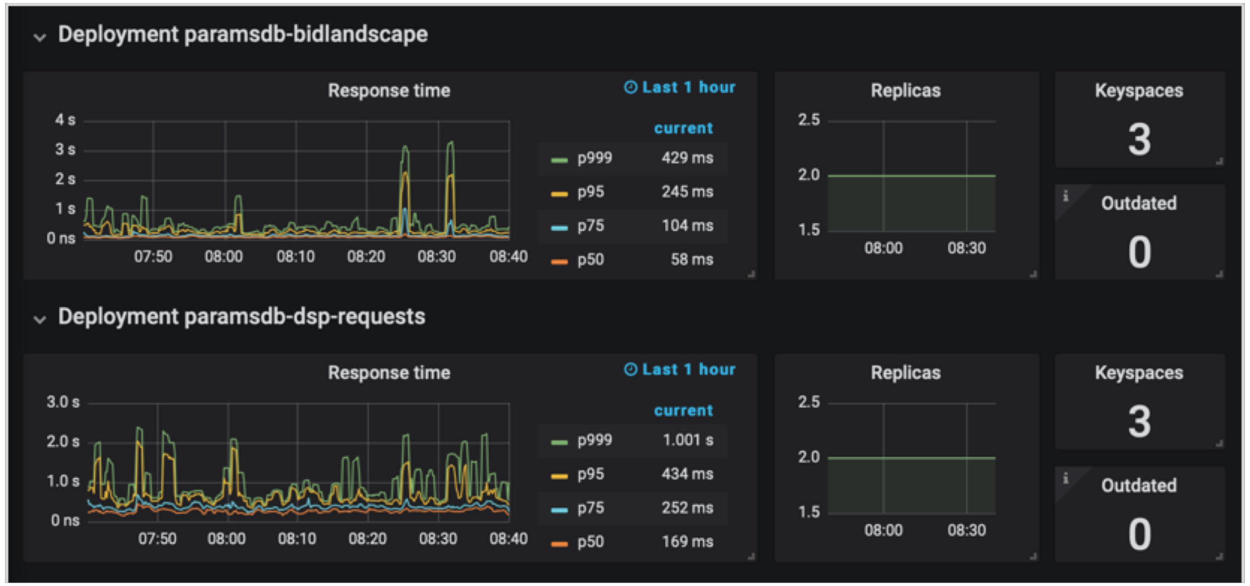
Hybrid approach between **sparse** and **dense** bitmaps representation.

roaringbitmap.org/about/

„Use Roaring for bitmap compression whenever possible.
Do not use other bitmap compression methods” [1]

[1] Wang, Jianguo, et al.
"An experimental study of bitmap compression vs. inverted list compression." 2017.

Bitmap Execution Database



- **Roaring Bitmap** – state of the art compressed bitmap algorithm
- Whole 7 days of Internet bid opportunities – 12GB (sample 1/1000) Naive 50TB
- Bids&Impressions – 5GB (sample 1/100)
- Majority of responses under 1 second

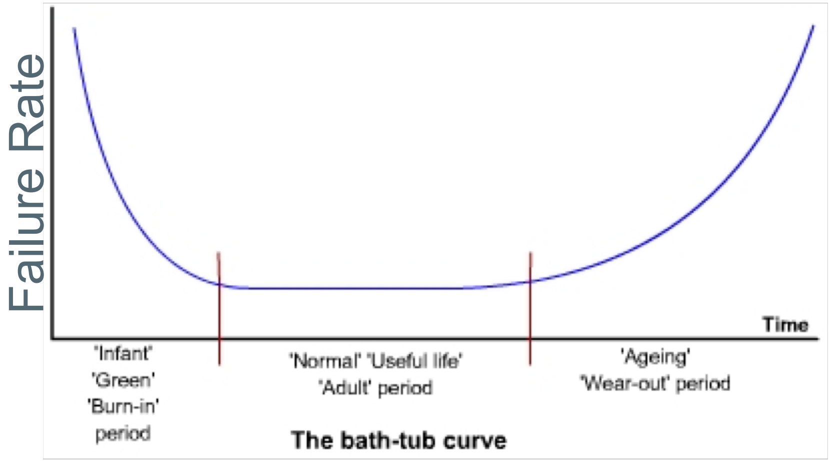
Nested set operations language

```
And(
  Or("inventory|1", "inventory|2", "inventory|3"),
  AndNot(
    Universe,
    Or("category|moto", "category|fishing")
  ),
  Or(
    "hour_of_week|41",
    "hour_of_week|42",
    "hour_of_week|43",
    "hour_of_week|44"
  ),
  Or("tld|com", "website|example.com"),
  And(
    "country|poland",
    AndNot(
        Universe,
        "city|warsaw"
    )
  ),
  ...
)
```

Survival analysis

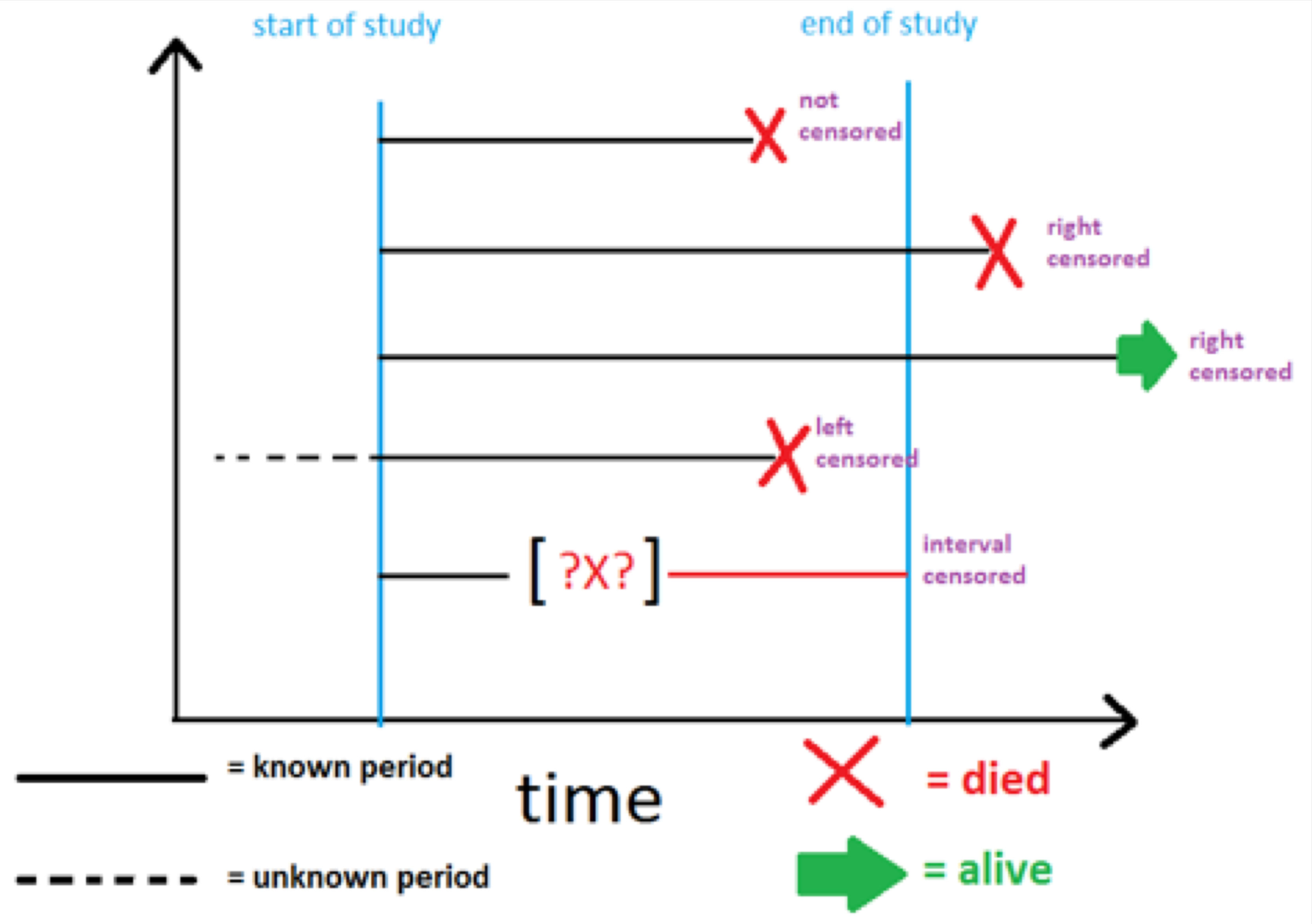
Principles of survival analysis

- The basic assumption in survival analysis is that the risk of failure changes with time



- Two major approaches for survival probability modelling can be distinguished:
  - Parametric – can be extrapolated for new experiences (Weibull)
  - Non-parametric – very flexible but can't be extended for new experience (Kaplan-Maier)

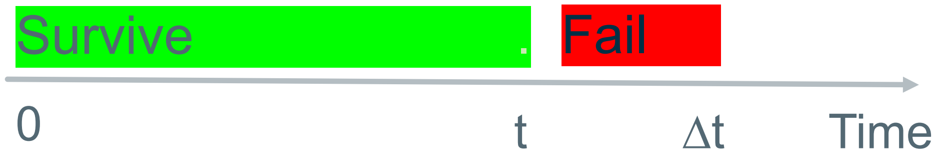
Data censoring in survival analysis



Weibull based survival probability estimation

- Weibull distribution is derived from hazard function:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}.$$



where:

- *S(t)* - survival function (cdf)
- *F(t) = 1 – S(t)* - failure function (cdf)
- *f(t) = dF(t)/dt - pdf*

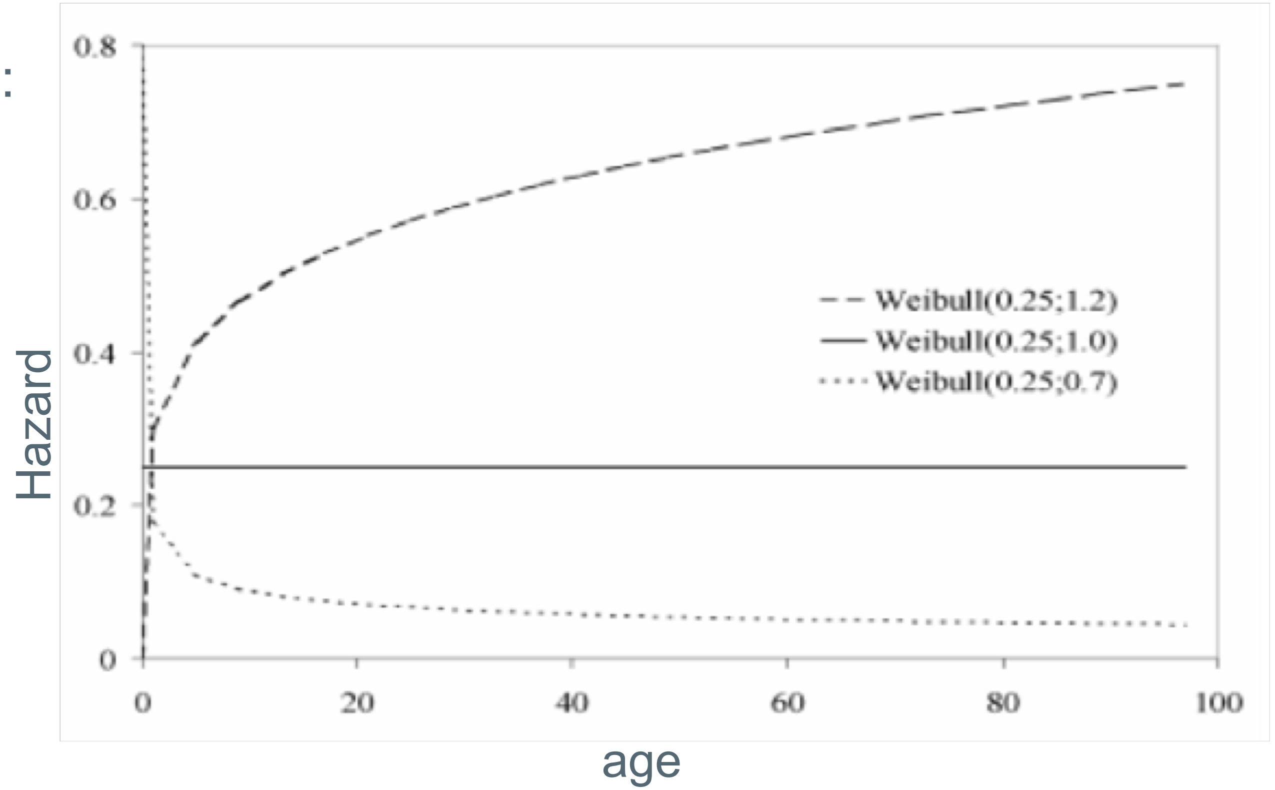
- Weibull distribution is defined by 2 parameters:

$$h(t) \;=\; p\lambda^p t^{p-1}$$

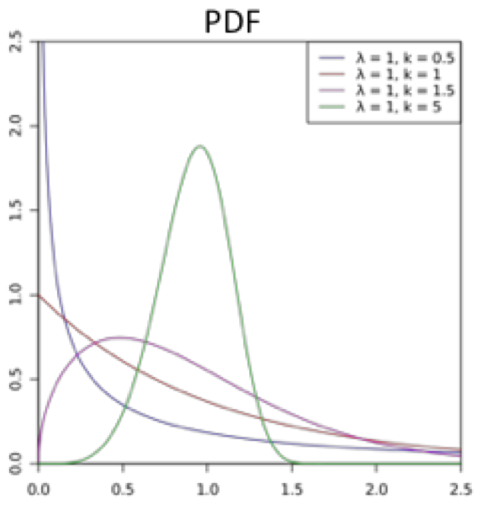
- Shape parameter *p* - failure mode type
- Scale parameter λ - position

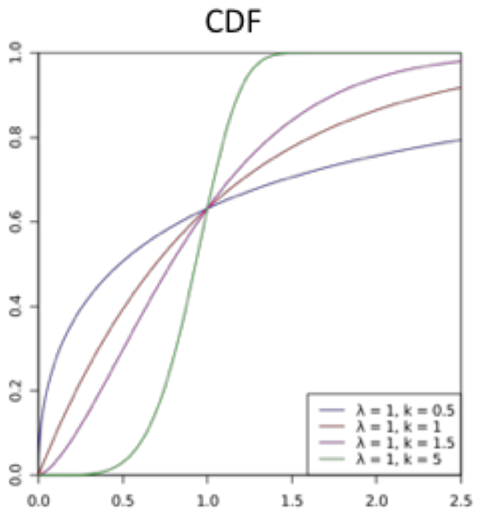
Weibull based survival probability estimation

- One can distinguish 3 classes of failure types:
  - Infant mortality $p < 1$
  - Random failures $p = 1$
  - Ware out $p > 1$



Weibull examples





Parametric approach easy to extrapolate

Kaplan-Maier based survival probability estimation

- In Kaplan-Meier method we are to predict chance of surviving time Δt after surviving time *t:*

Survive . Survive.

0 t Δt Time

$$S(t,\Delta t) = S(t) \cdot S(\Delta t|t)$$

Kaplan and Meier (1958) extended the estimate to *censored* data. Let

$$t_{(1)} < t_{(2)} < \ldots < t_{(m)}$$

denote the distinct *ordered* times of death (not counting censoring times). Let d_i be the number of deaths at $t_{(i)}$, and let n_i be the number alive *just before* $t_{(i)}$. This is the number exposed to risk at time $t_{(i)}$. Then the Kaplan-Meier or *product limit* estimate of the survivor function is

$$\hat{S}(t) = \prod_{i:t_{(i)}<t} \left(1 - \frac{d_i}{n_i}\right).$$

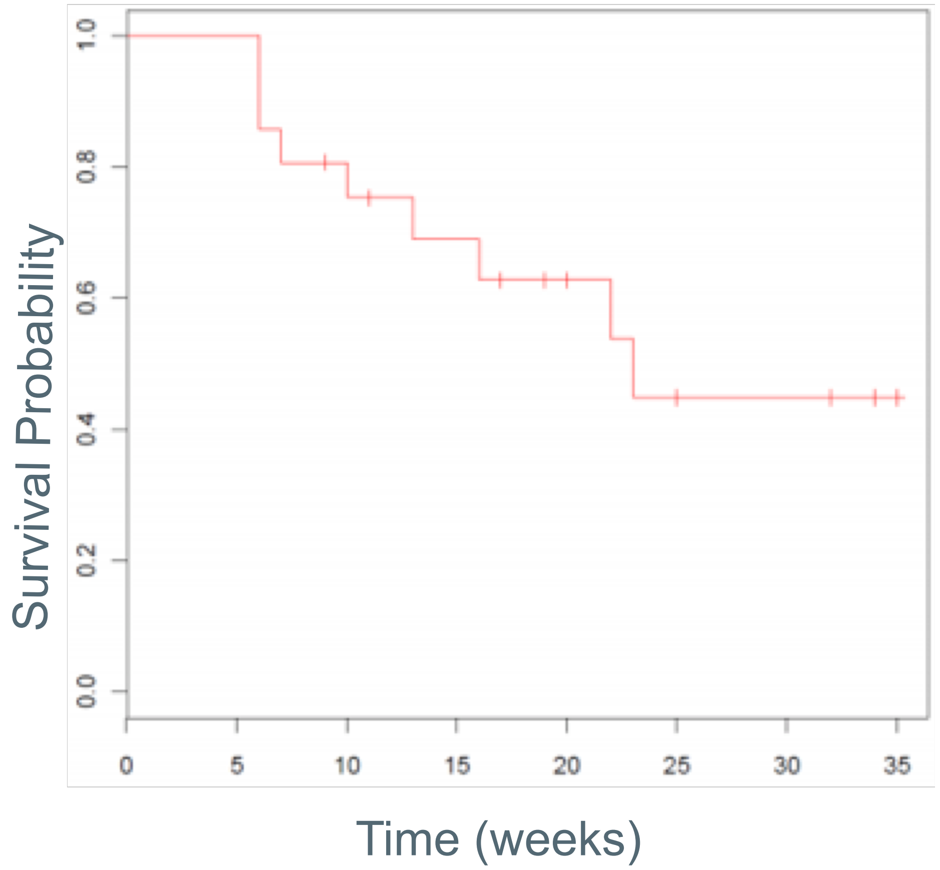
No parameters to be estimated, very flexible and easy in implementation

Kaplan-Meier curve calculation

$$S(t,\Delta t) = S(t) \cdot S(\Delta t|t)$$

$t_{(j)}$	n_j	m_j	q_j	$\hat{S}(t_{(j)})$
0	21	0	0	1
6	21	3	1	1×18/21=.8571
7	17	1	1	.8571×16/17=.8067
10	15	1	2	.8067×14/15=.7529
13	12	1	0	.7529×11/12=.6902
16	11	1	3	.6902×10/11=.6275
22	7	1	0	.6275×6/7=.5378
23	6	1	5	.5378×5/6=.4482

KM Curve for Group 1 (treatment)

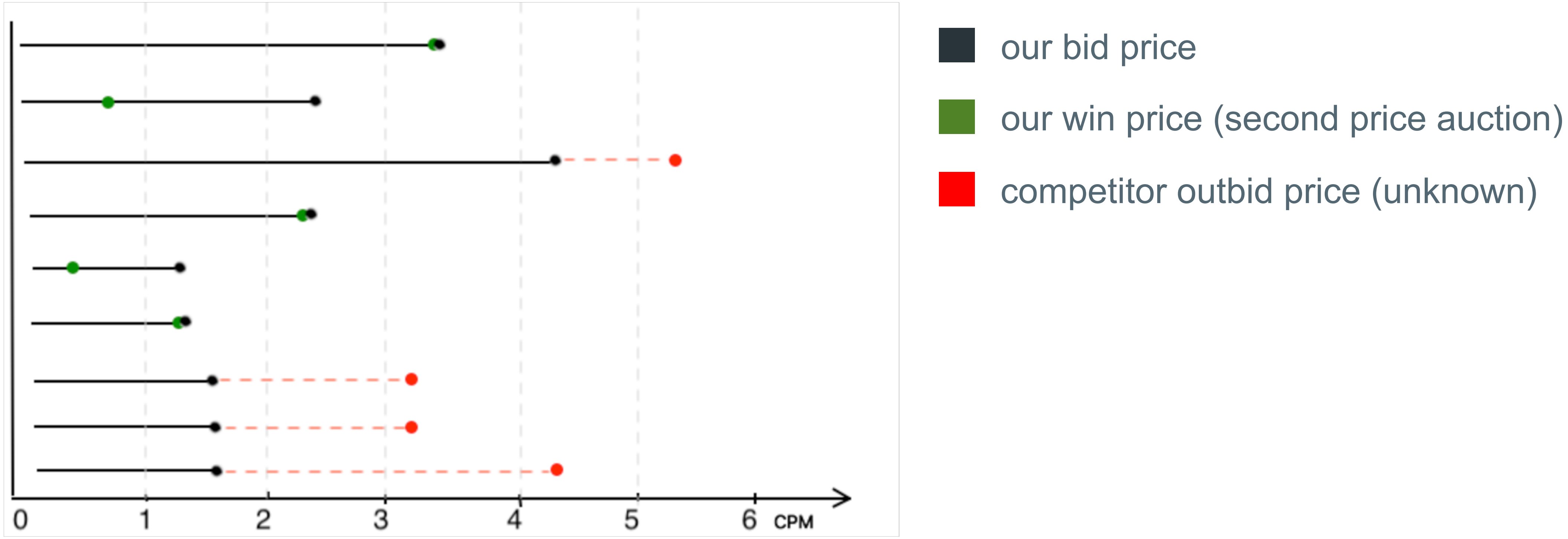


Elements of survival analysis in Bid Landscape modelling

RTB auctions modelling with survival methods

From survival models to Bid Landscape:

- time -> cost per 1k bought impressions (CPM)
- died -> won bid (left censored, known win price)
- alive -> lost bid (right censored, unknown exact win price)



Bid Landscape vs. Survival Analysis

Survival analysis

- Risk of failure evolves with time
- Incomplete knowledge regarding failure time, right censoring:
  - Majority of object survive
- Multiple competing failure modes
- Very high disproportion between positive (failures) and negative (survived) class representatives
- Usually very limited data
- Wrong predictions may have catastrophic consequences – quality monitoring required

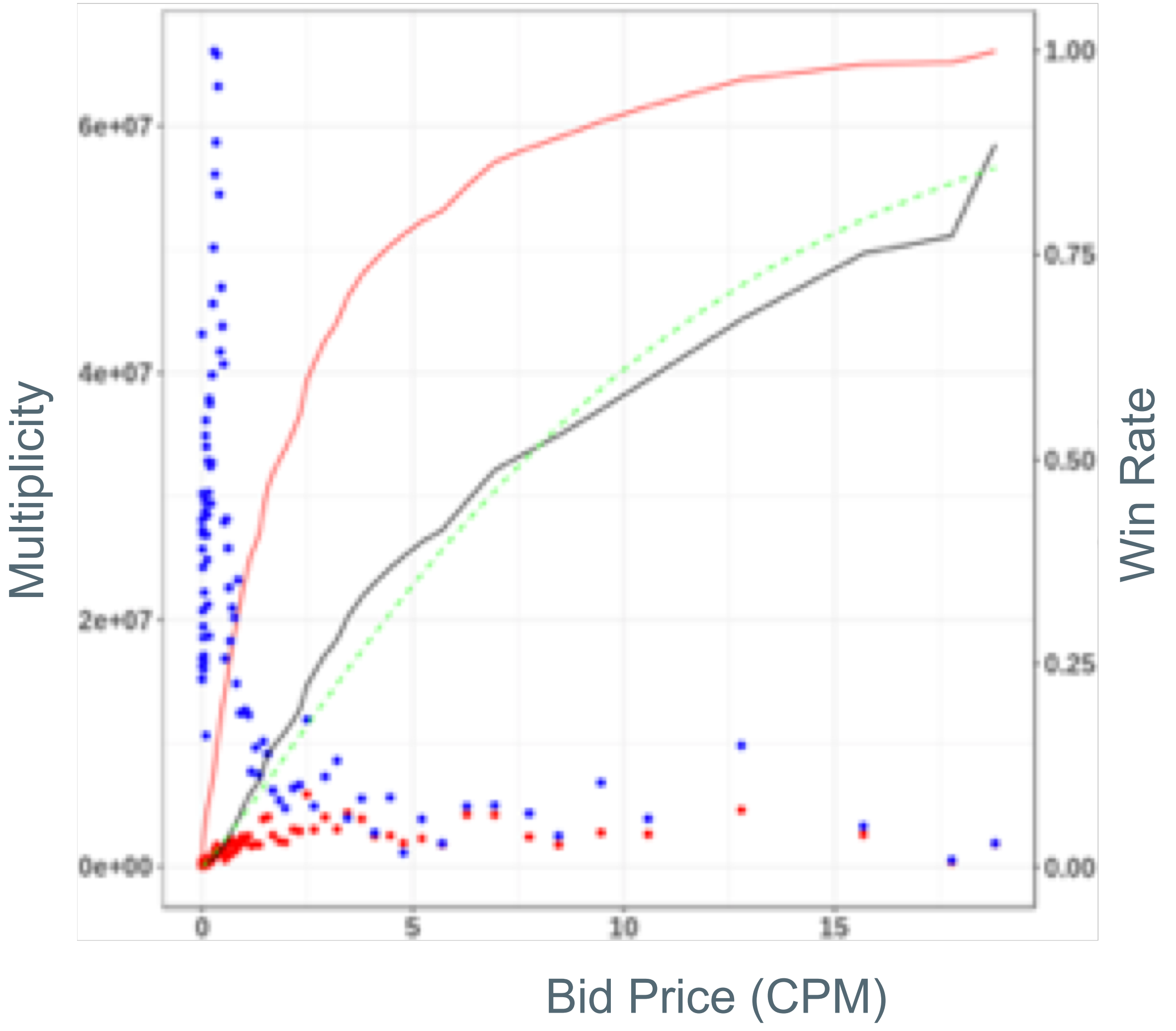
Bid Landscape analysis

- Winning probability evolves with price
- Incomplete knowledge regarding win price, right censoring:
  - Only small fraction of bids are won
- Different inventory sources
- Very high disproportion between positive (won) and negative (lost) class representatives
- Usually huge data sets – subsampling
- Wrong predictions can lead to not efficient campaign money spend – quality monitoring required

Survival model for Bid Landscape modelling

- Let's apply survival techniques to auction data:
  - Instead of time a bid price is to be used
- Legend:
  - **number of lost auctions**
  - **number of won impressions**
  - win price CDF
  - Kaplan-Maier estimate
  - Weibull estimate

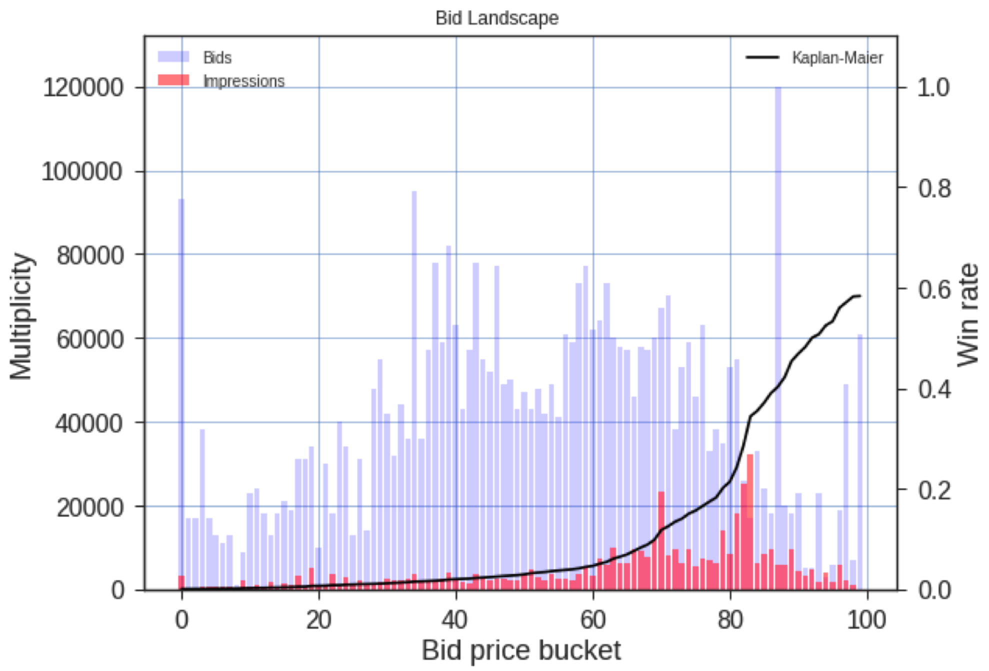
Adform Bid Landscape exploits Kaplan-Maier estimate exclusively



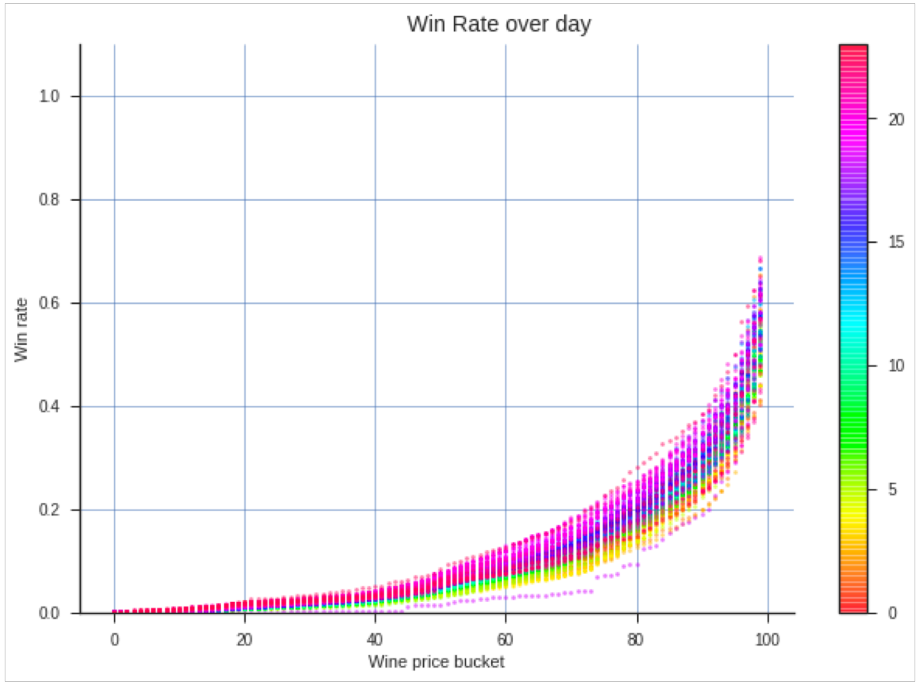
Bid Landscape validation in Adform

Bid Landscape validation – qualitative

Win rate evolution with bid price



Win rate evolution over day

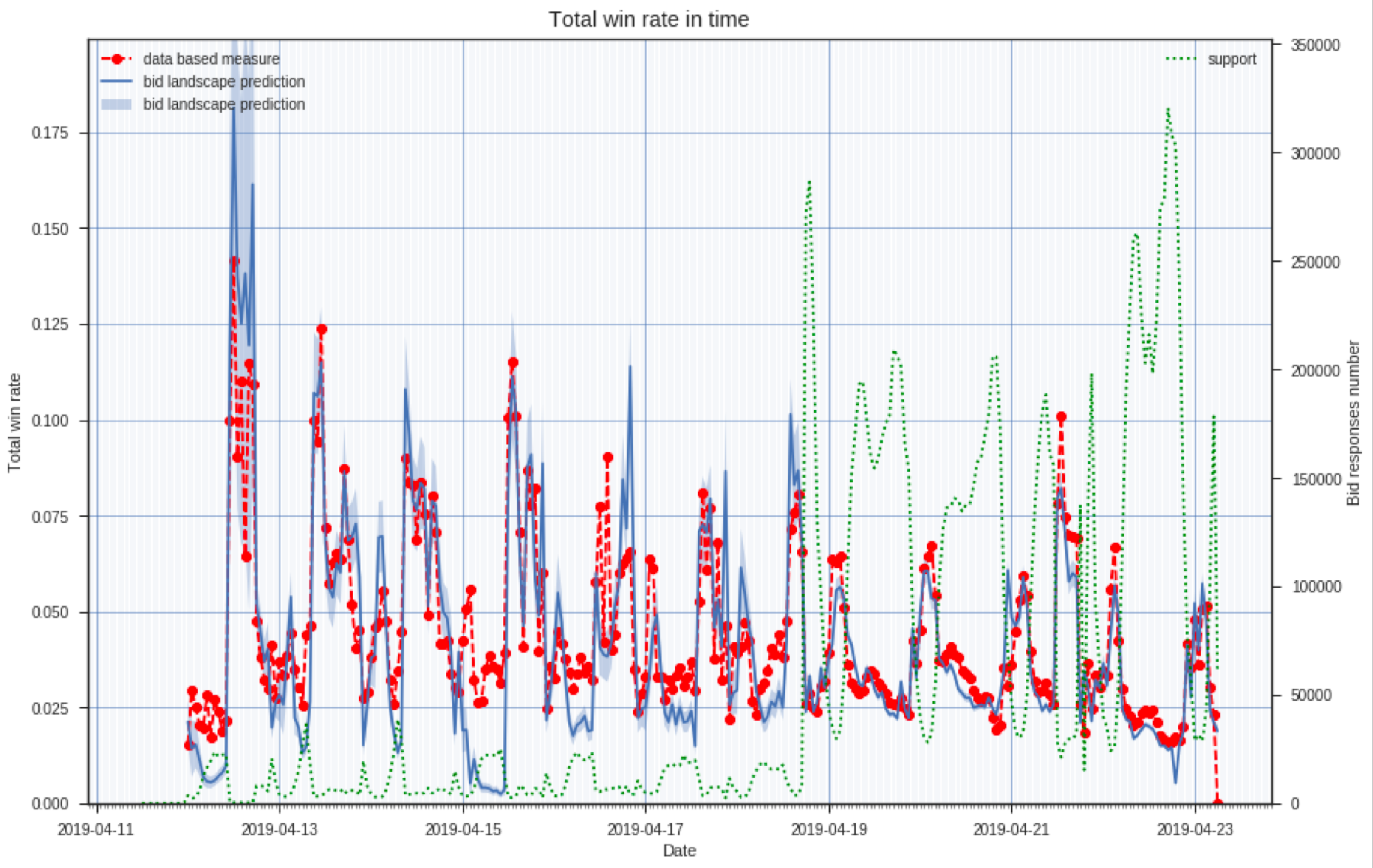


The bid landscape predictions are in line with the intuition:

- The win rate increases with bid price
- The win rate changes periodically over a day

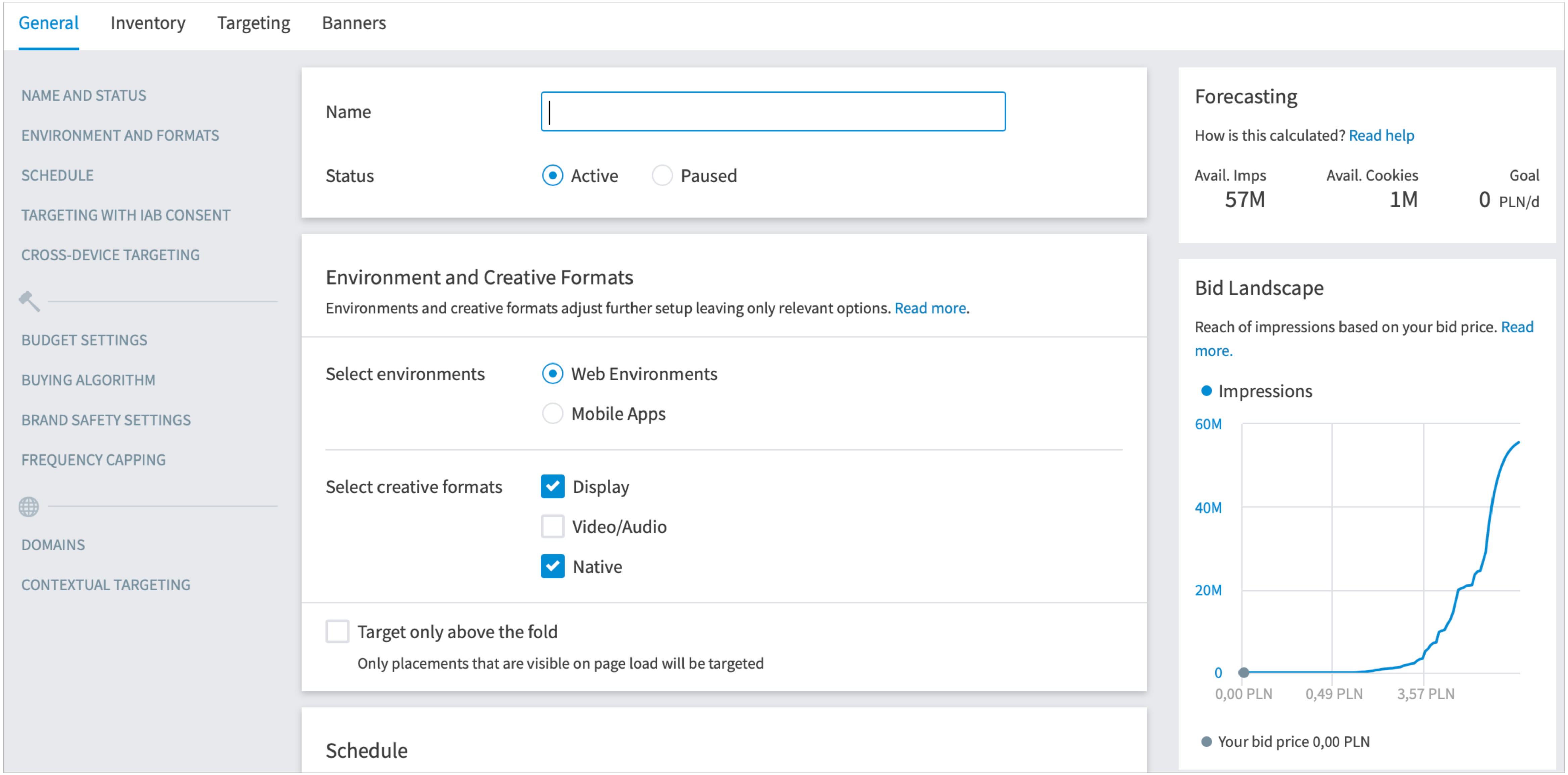
Bid Landscape validation – quantitative (real RTB data)

Real auction data versus Bid Landscape predictions

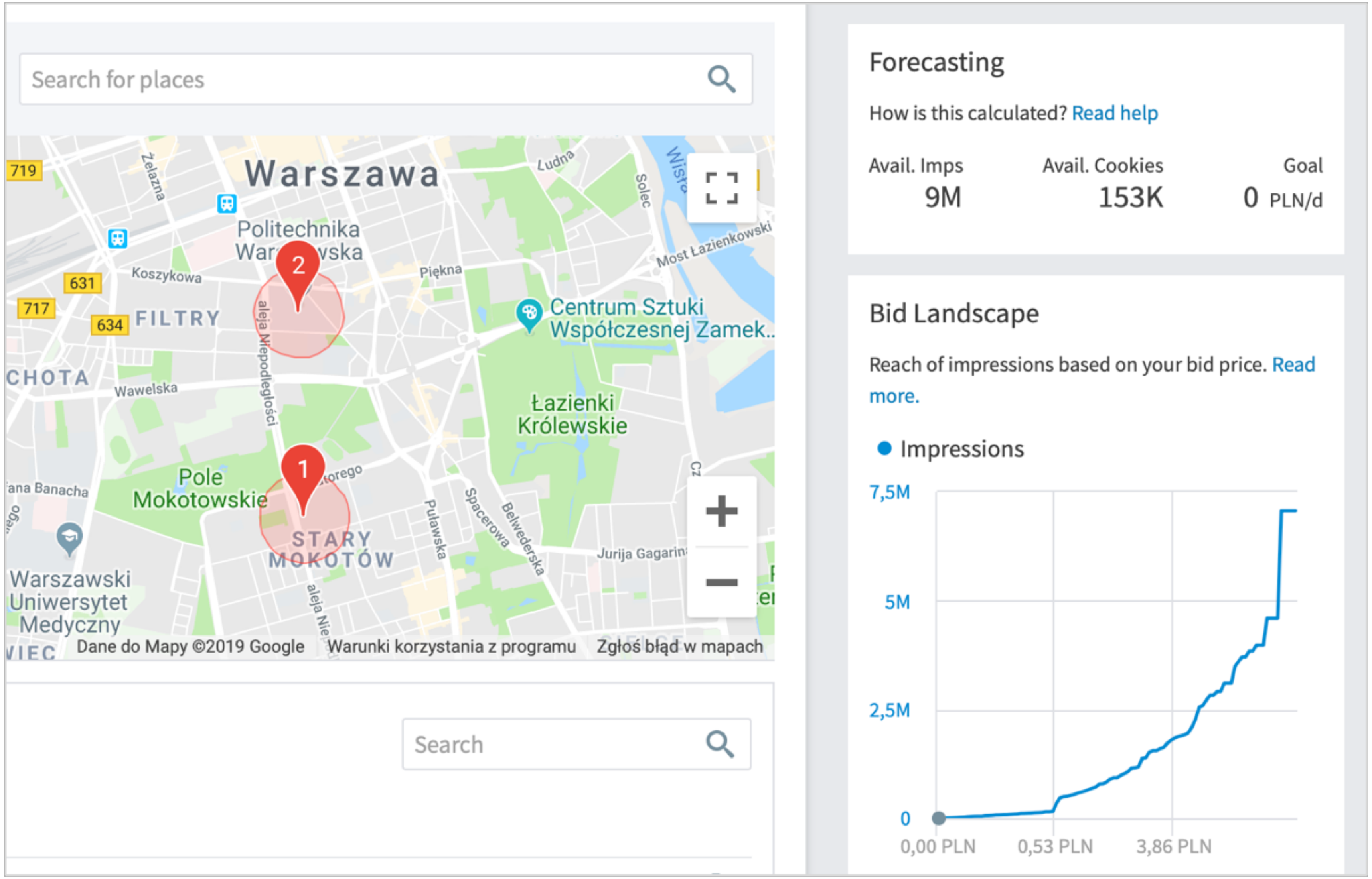


Bid Landscape as Adform product

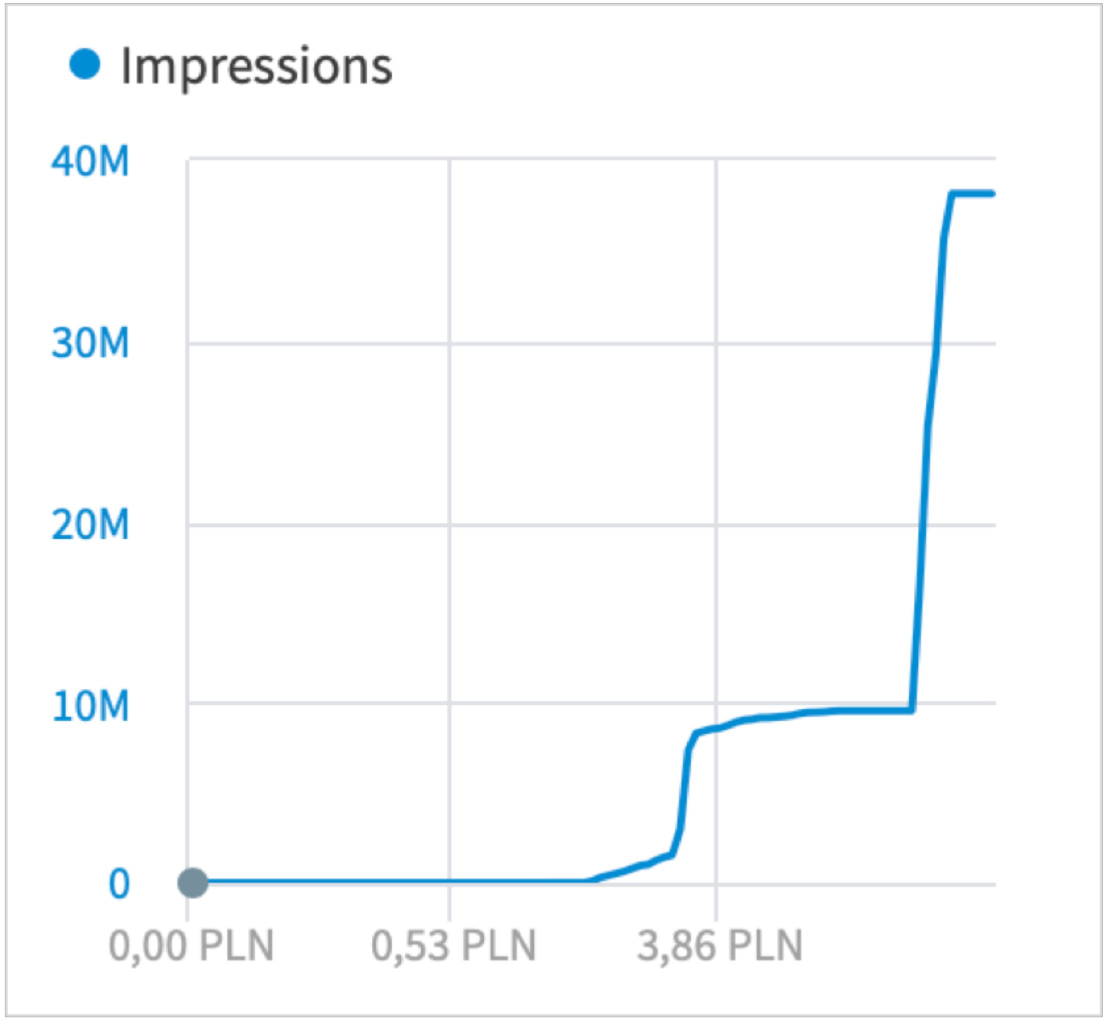
Interactive tool used in UI



Hyperlocal precision

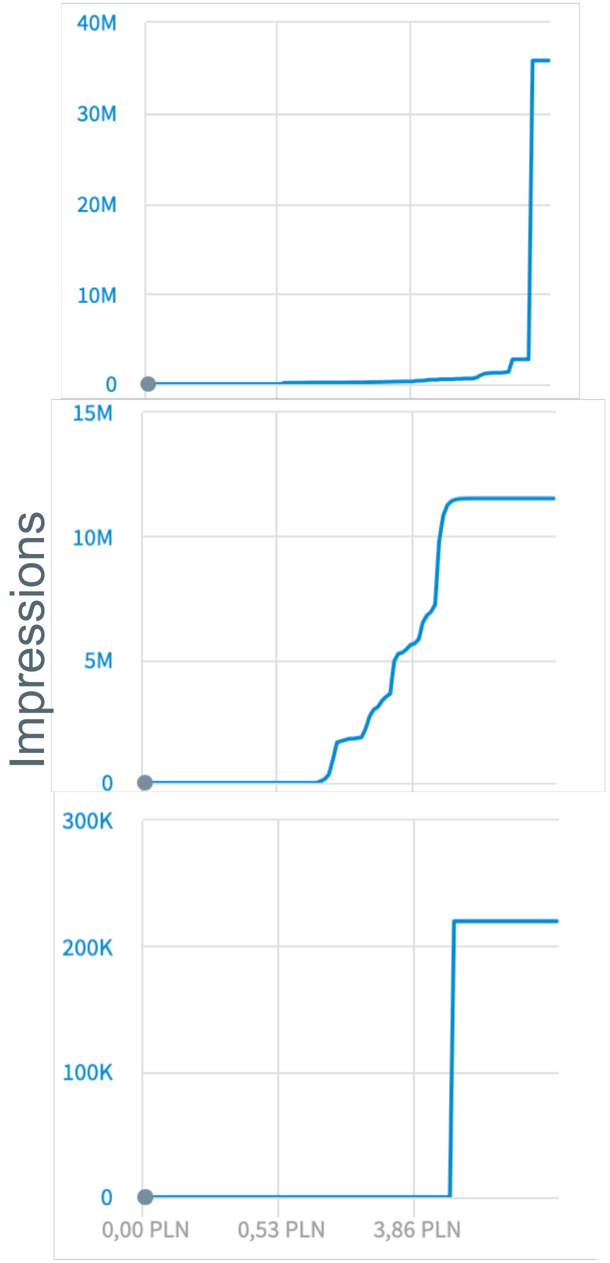


Floor Price detection



Bid Multipliers optimisation

- Campaign targeting Warsaw people to be run on three different domains
- Global CPM makes no sense, needs to be adjusted



Source of traffic:

Video oriented website

Polish accountants' portal

Jobs for professionals

Bid Landscape dynamic CPM algorithm

Given the Campaign Settings including hourly budget maximize bought relevant impressions by using lowest CPM.

First tests are showing multiplied Viewable Impressions, Clicks comparing to same Campaign using static price.

Summary and conclusions

- The Adform Bid Landscape is composed of two base elements ensuring speed and quality:
  - Bitmaps – fast and efficient population selection
  - Kaplan-Maier win rates – simple and precise estimate
- The Adform Bid Landscape is in production for more than one year
  - It is mostly used by Adform customers optimising the CPM campaigns
  - Currently we develop new RTB algorithm that is based on BL predictions
  - In the short time perspective we are to extend usage of BL in other RTB models

Thank You

adform

Questions?