



# Using Machine Learning in the delivery of ads

Ruth Garcia

NDR Conference – June 07, 2018

Iasi, Romania



## What is Skyscanner?



Skyscanner is a leading **global travel search site** offering a comprehensive and **free** flight search service as well as online comparisons for hotels, car hire and **now trains**.



**1. Decision Science:** Ensuring we have the right data; insights and information to make the most impactful, scientific decisions in every aspect of our operations.



**2. Building Data Products:** Leveraging our vast wealth of data to build more contextual, relevant products for Travellers and Travel Suppliers (our Partners)



# Machine Learning at Skyscanner

- Destination recommendations
- Itinerary mashups
- Advertising

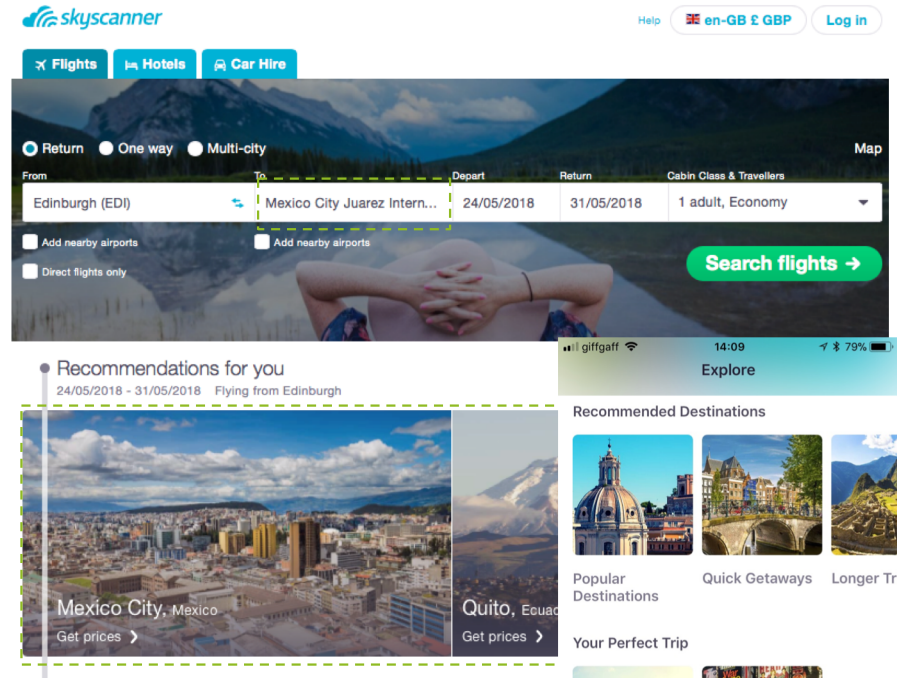
## Mashups

	06:03 MIA	2h 58 Direct	→	09:01 EWR	£192 <a href="#">Select →</a> 2 bookings required
	11:30 JFK	3h 14 Direct	→	14:44 MIA	



Mashups. A better blend of flights for your journey, offering more choice and savings.

## Recommendations



skyscanner en-GB £ GBP Log in

Flights Hotels Car Hire

Return One way Multi-city

From: Edinburgh (EDI) To: Mexico City Juarez Intern... Depart: 24/05/2018 Return: 31/05/2018 Cabin Class & Travellers: 1 adult, Economy

Search flights →

Recommendations for you  
24/05/2018 - 31/05/2018 Flying from Edinburgh

Recommended Destinations

- Mexico City, Mexico
- Quito, Ecuador

Popular Destinations Quick Getaways Longer Trips

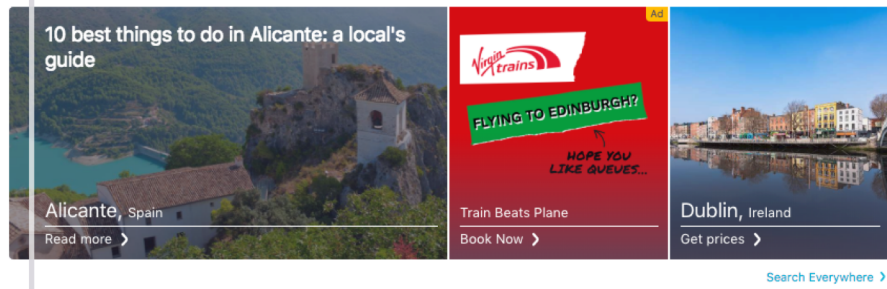
Your Perfect Trip

- Solo Travel
- Kid Free

Be Inspired

Search Explore My Travel Profile

## Advertising



10 best things to do in Alicante: a local's guide

Alicante, Spain

Read more >

Train Beats Plane

Book Now >

Dublin, Ireland

Get prices >

Search Everywhere >

# Ads at Skyscanner

skyscanner

en-GB £ GBP Log in

Flights Hotels Car Hire

Return One way Multi-city

From: London Heathrow (LHR) To: New York (Any) Depart: 12/06/2018 Return: 29/06/2018 Cabin Class & Travellers: 1 adult, Economy

Add nearby airports Direct flights only

**Search flights →**

## Popular destinations

June 2018 Flying from London Heathrow

10 best things to do in Alicante: a local's guide

Alicante, Spain

Read more >

**Virgin Trains**

**FLYING TO EDINBURGH?**

HOPE YOU LIKE QUEUES...

Train Beats Plane

Book Now >

Dublin, Ireland

Get prices >

Search Everywhere >

skyscanner

en-GB £ GBP Log in

Flights Hotels Car Hire

London (LHR) - New York, NY (Any) 1 adult Economy

Tue 12 Jun Fri 29 Jun

Show whole month Additional bag fees may apply

Get Price Alerts 1907 results sorted by Best Sort by Best

Stops: Direct £535, 1 stop £543, 2+ stops None

Departure times: 8 hours - 14.5 hours

Airlines: oneworld, SkyTeam, Star Alliance, Aer Lingus, Air Canada, Air France, American Airlines

Best £535 (7h 13 average), Cheapest £535 (7h 23 average), Fastest £535 (7h 13 average)

virgin atlantic: Discover 3 new ways to fly with Virgin Atlantic. Still get everything you'd expect on a great flight - but at a price you wouldn't.

Sponsored: Check prices and save. Select →

Outbound: 19:50 LHR Direct → 22:30 JFK (Operated by British Airways), 10:25 JFK Direct → 22:10 LHR (Operated by American Airlines)

Return: 13:10 LHR Direct → 16:00 JFK (Operated by British Airways), 10:25 JFK Direct → 22:10 LHR (Operated by American Airlines)

18:05 LHR Direct → 21:00 JFK (Operated by British Airways), 10:25 JFK Direct → 22:10 LHR (Operated by American Airlines)

21 deals from £535, 19 deals from £535

Need a hotel in New York? Search hotels

United Airlines: Connecting you to even more cities. Book Now

The best things in life are free. And so are we. skyscanner. No commission. No fees. No bias.

Need a hotel in New York? Search hotels

Need a hotel in New York? from £708. Book flight

Hire a car in New York from £17 per day. Find a car

# Not always annoying ads

Show whole month Additional bag fees may apply

[Get Price Alerts](#) 1221 results sorted by Best Sort by Best

**Stops**

- Direct £258
- 1 stop £188
- 2+ stops £295

**Departure times**

Outbound 00:00 - 23:59

Return 00:00 - 23:59

**Journey duration**

2.5 hours - 24 hours

**Airlines**

Select all | Clear all

- oneworld
- SkyTeam
- Star Alliance
- Air Europa

Best	Cheapest	Fastest
£258 2h 25 (average)	£188 5h 35 (average)	£279 2h 20 (average)

**BRITISH AIRWAYS** Book flights to Valencia on ba.com  
No hidden fees. Our lowest prices guaranteed. Book direct now!

**Sponsored**  
**£230**  
[Select →](#)

<b>easyJet</b> 06:20 LGW <b>Direct</b> 2h 20 → 09:40 VLC	9 deals from <b>£258</b> <a href="#">Select →</a>
<b>easyJet</b> 11:05 VLC <b>Direct</b> 2h 30 → 12:35 LTN	<a href="#">Select →</a>
<b>easyJet</b> 09:10 LGW <b>Direct</b> 2h 20 → 12:30 VLC	7 deals from <b>£266</b> <a href="#">Select →</a>
<b>easyJet</b> 11:05 VLC <b>Direct</b> 2h 30 → 12:35 LTN	<a href="#">Select →</a>
<b>easyJet</b> 06:20 LGW <b>Direct</b> 2h 20 → 09:40 VLC	8 deals from <b>£279</b> <a href="#">Select →</a>
<b>easyJet</b> 21:10 VLC <b>Direct</b> 2h 20 → 22:30 LGW	<a href="#">Select →</a>

793 Hotels ready for you in Valencia

[Book now](#)

Need a hotel in Valencia?

[Search hotels](#)

**TUI**  
Discover your smile

BOOK YOUR JAMAICA HOLIDAY TODAY.

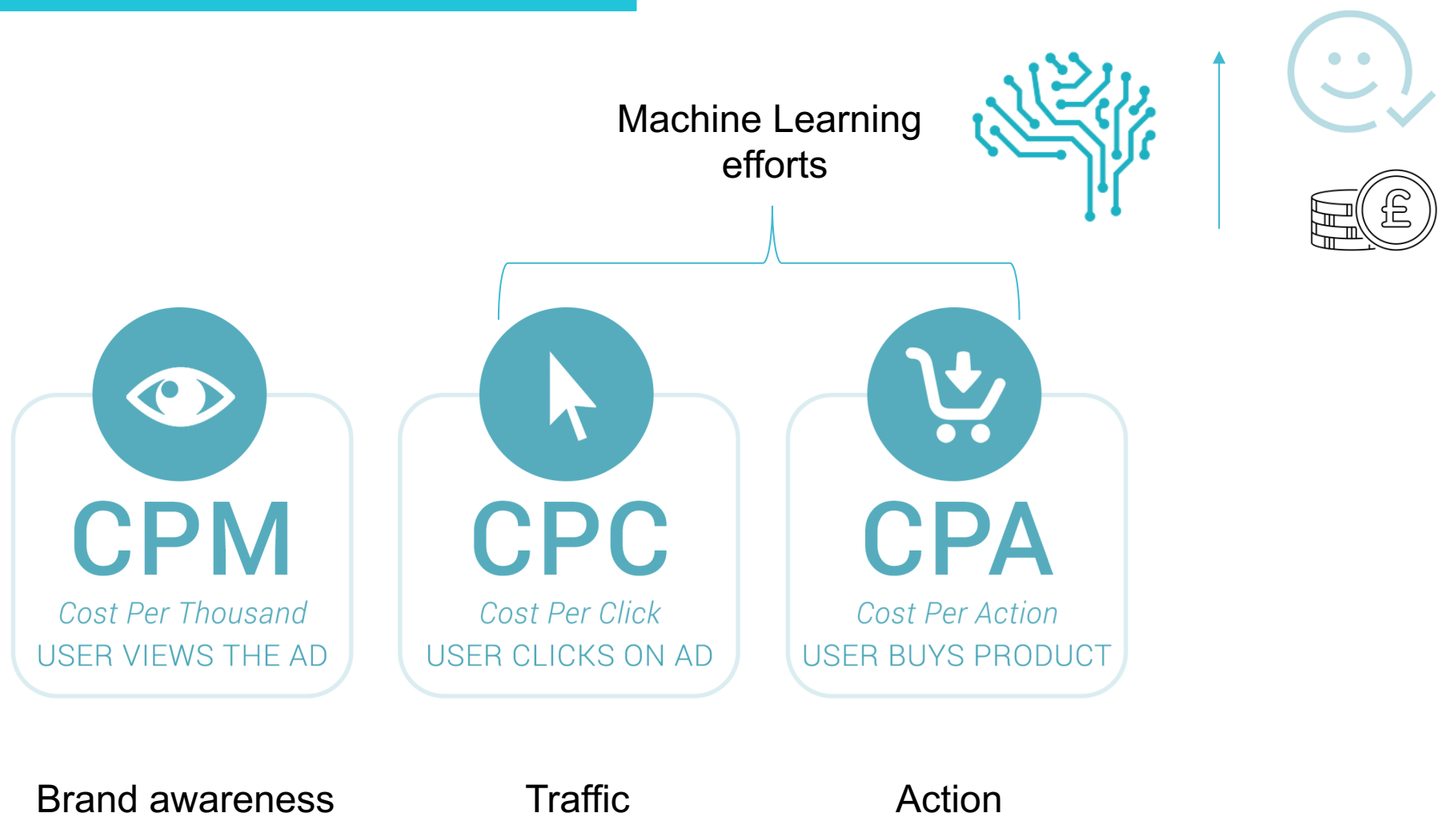
[BOOK NOW](#)

Travel smart

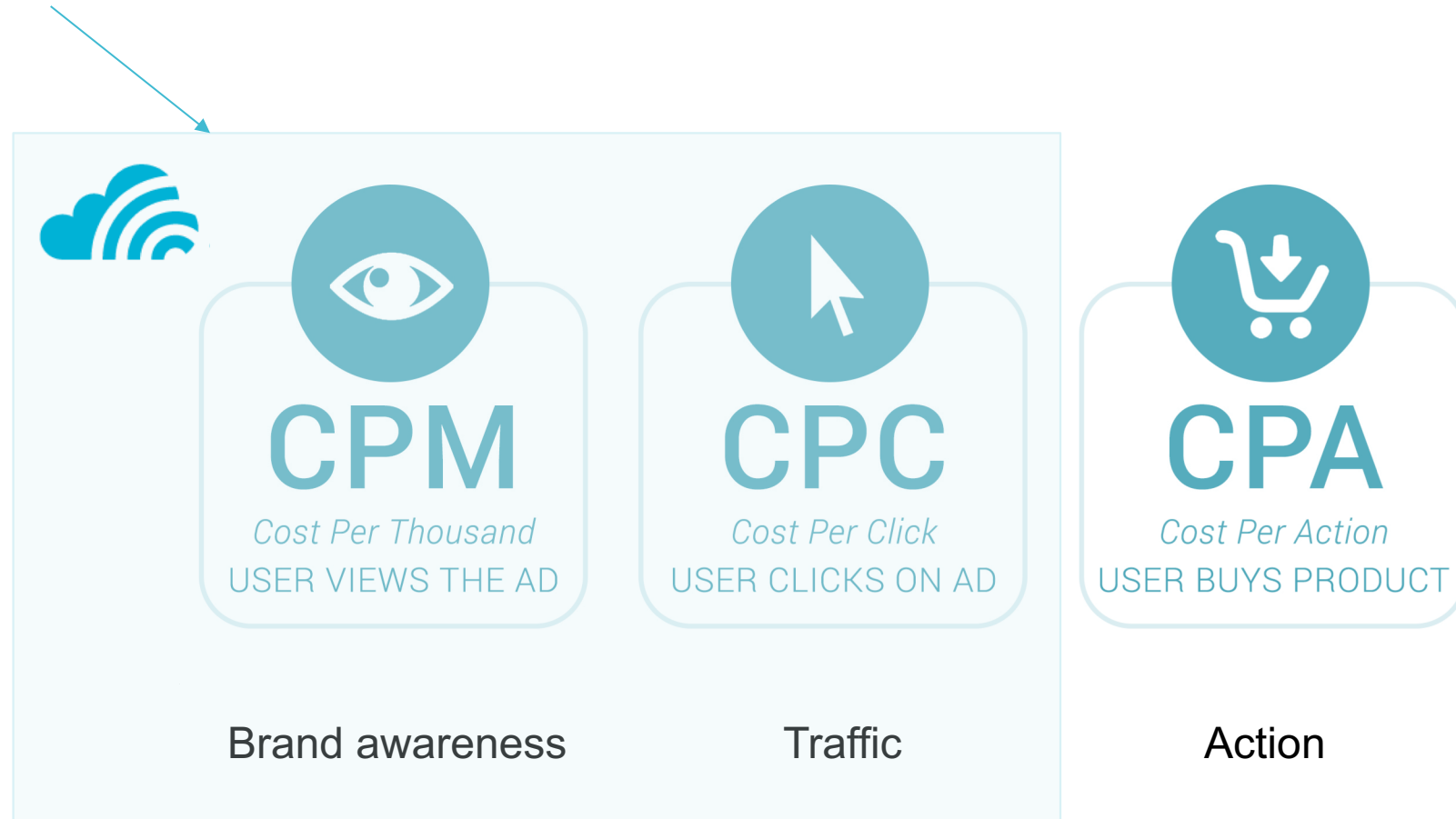
Find deals and inspiration on flights, and car

**skyscanner**  
The world's search en

# Advertising schemes



# Advertising schemes






# Advertising Schemes



**CPM**  
*Cost Per Thousand*  
USER VIEWS THE AD

Brand awareness



**CPC**  
*Cost Per Click*  
USER CLICKS ON AD

Traffic



**CPA**  
*Cost Per Action*  
USER BUYS PRODUCT

Action

# Ads at Skyscanner



## DoubleClick for Publishers

skyscanner

Flights Hotels Car Hire

Return One way Multi-city

From: London Heathrow (LHR) To: New York (Any) Depart: 12/06/2018 Return: 29/06/2018 Cabin Class & Travellers: 1 adult, Economy

Search flights →

### Popular destinations

June 2018 Flying from London Heathrow

#### 10 best things to do in Alicante: a local's guide

Alicante, Spain

Read more >

FLYING TO EDINBURGH?

HOPE YOU LIKE QUEVES...

Train Beats Plane

Book Now >

Dublin, Ireland

Get prices >

Search Everywhere >

skyscanner

London (LHR) - New York, NY (Any) 1 adult Economy

1907 results sorted by Best

Stops	Best £535 (7h 13 average)	Cheapest £535 (7h 23 average)	Fastest £535 (7h 13 average)
Direct £535	Sponsored: Discover 3 new ways to fly with Virgin Atlantic. Still get everything you'd expect on a great flight - but at a price you wouldn't.		
1 stop £543	Sponsored: Need a hotel in New York? Search hotels		
2+ stops None	Sponsored: Need a hotel in New York? Search hotels		

Departure times	Outbound 00:00 - 23:59	Return 00:00 - 23:59	Journey duration	Airlines
19:50 LHR Direct → 22:30 JFK	10:25 JFK Direct → 22:10 LHR	8 hours - 14.5 hours	13:10 LHR Direct → 16:00 JFK	10:25 JFK Direct → 22:10 LHR
18:05 LHR Direct → 21:00 JFK	10:25 JFK Direct → 22:10 LHR	8 hours - 14.5 hours	10:25 JFK Direct → 22:10 LHR	10:25 JFK Direct → 22:10 LHR

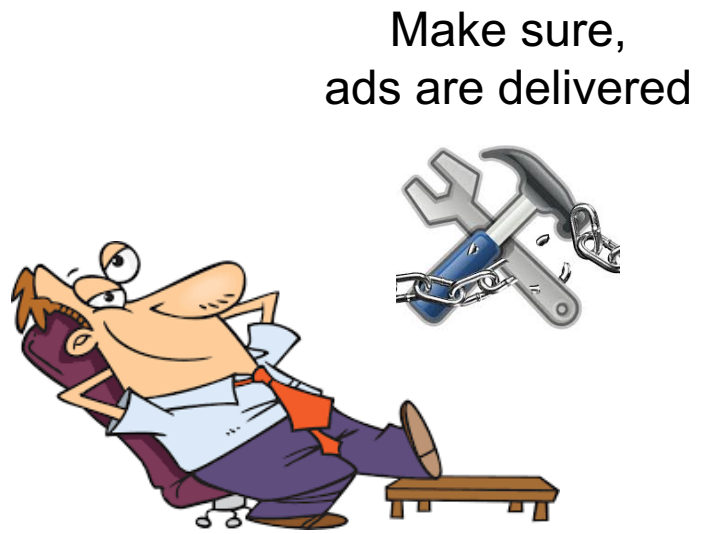
Need a hotel in New York? Search hotels

# Ads at Skyscanner



DoubleClick  
for Publishers

=



Make sure,  
ads are delivered

**No need to worry about  
how ads are delivered**

# Delivery of Ads



DoubleClick  
for Publishers



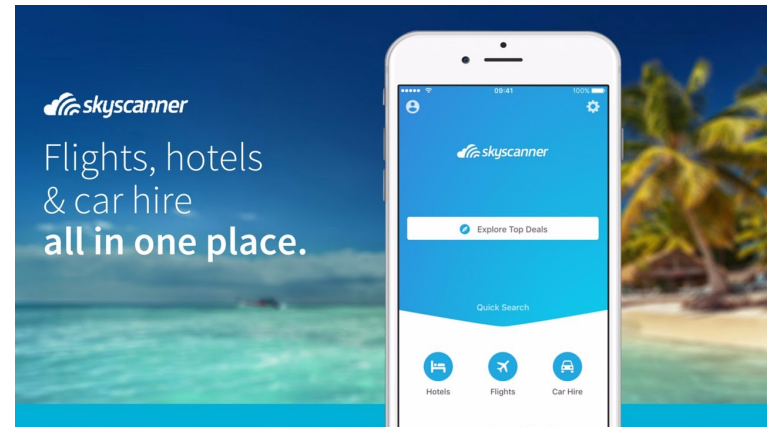
Black box  
algorithms



Technical  
difficulties







Data  
ownership

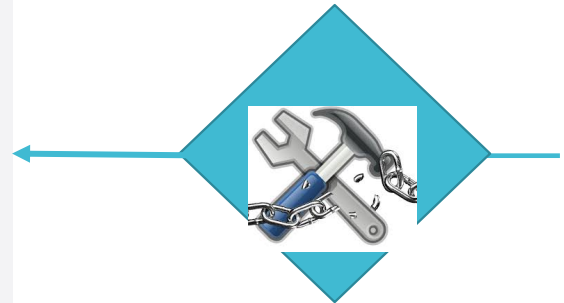


**Skyscanner Ads  
Manager**

# Click prediction algorithm

## Click prediction algorithm

Best £45 1h 33 (average)	Cheapest £45 1h 33 (average)	Fastest £90 1h 30 (average)
 <p>Get to Edinburgh nae bother Good brew, good food and BEAM on-board entertainment. Sorry planes, you're just not trains.</p>	<p>Sponsored</p> <p>Check prices and save</p> <p>Select →</p>	
 <p>08:10 LGW → 1h 35 Direct → 09:45 EDI</p>	<p>8 deals from £45</p> <p>Select →</p>	
 <p>18:55 EDI → 1h 30 Direct → 20:25 LGW</p>		
 <p>08:10 LGW → 1h 35 Direct → 09:45 EDI</p>		<p>8 deals from £45</p>



- Candidates
- 
  - 
  - 
  - 

Ranking algorithm

# Cloud services and tools at Skyscanner

## Languages



## AWS services



S3

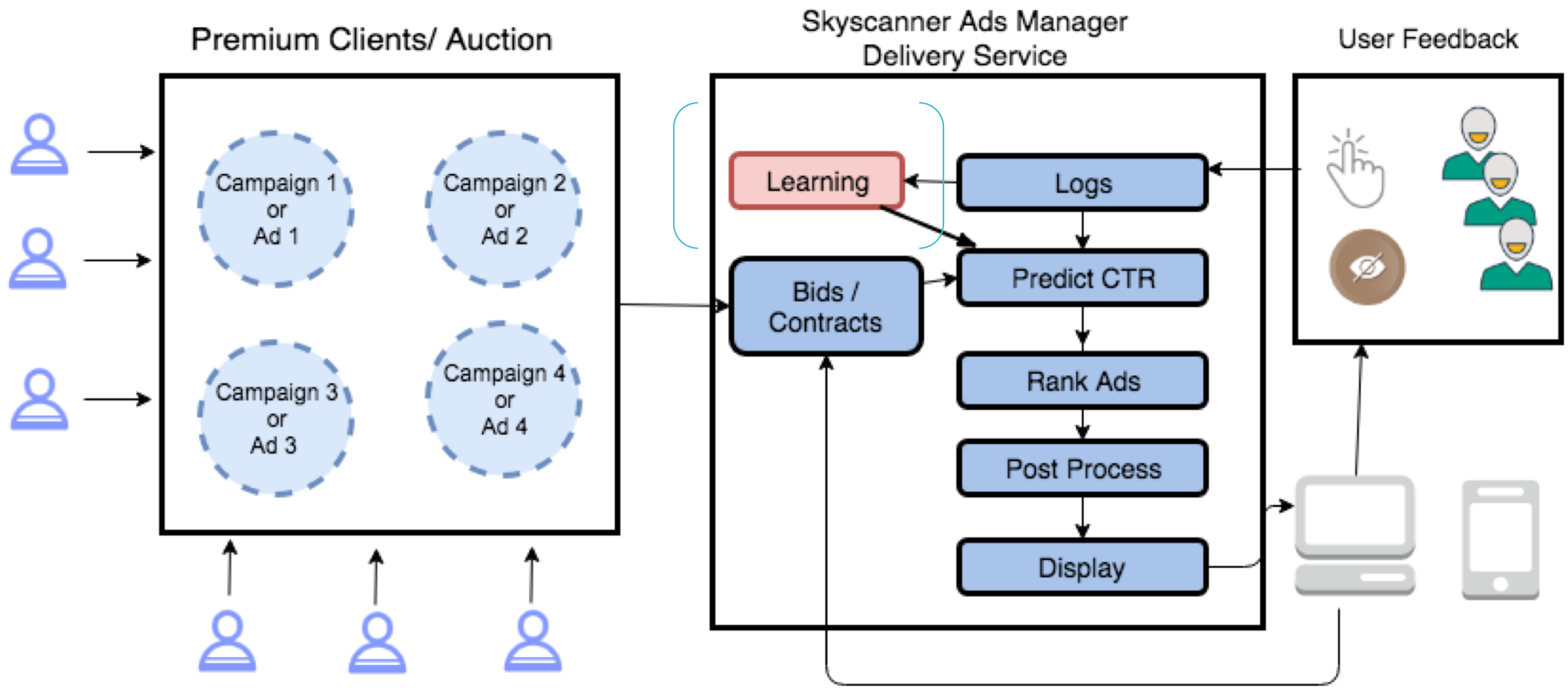


Amazon  
Athena



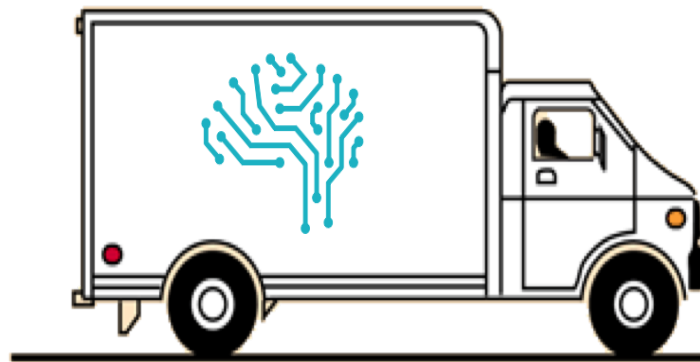
Batch

# Overview of an Advertising System



# Expectation vs. reality

## Expectation



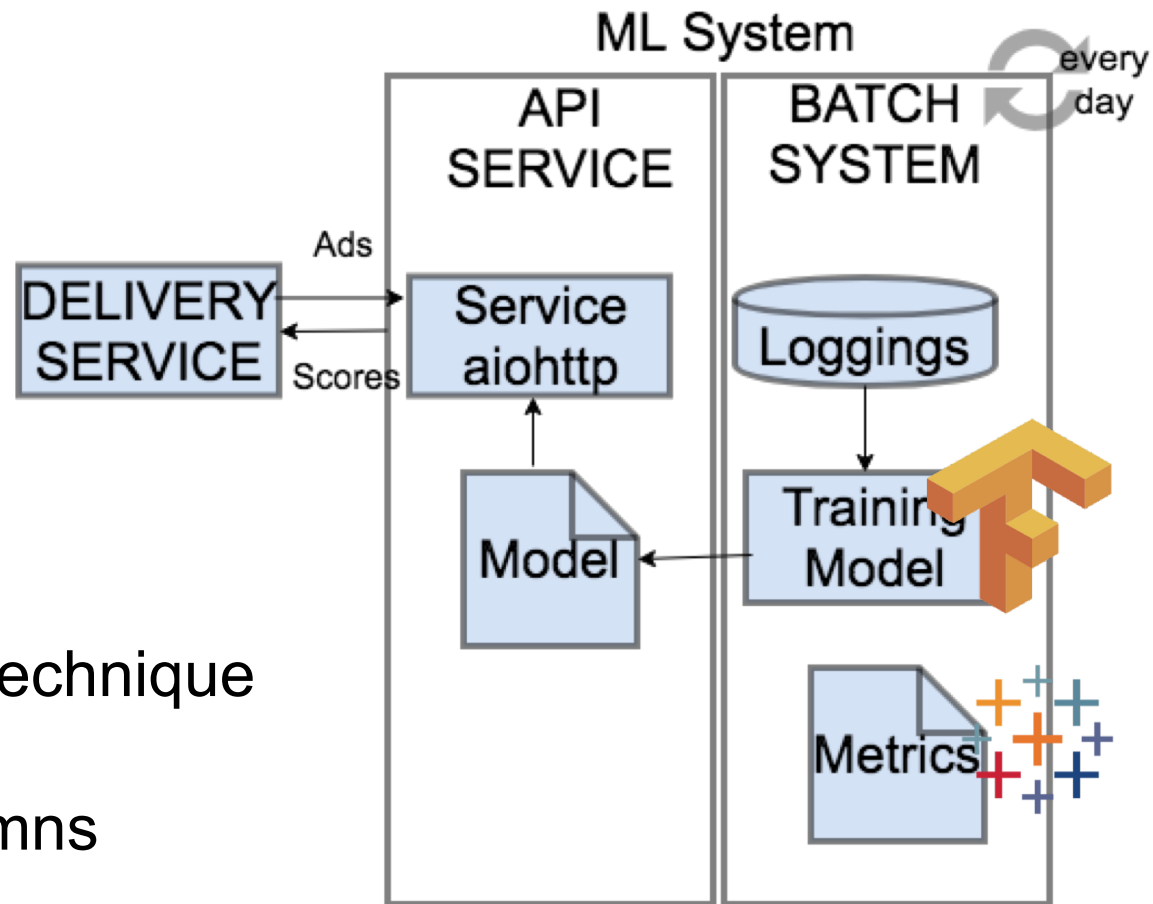
### Features:

- User history,
- User features,
- Route features ,
- Ad features with, colors, text

### Tensorflow

- Optimization technique
- Embeddings
- Crossed columns
- Hashing

Flexible but with the risk of not being fast enough.





# Expectation vs. reality

## Reality



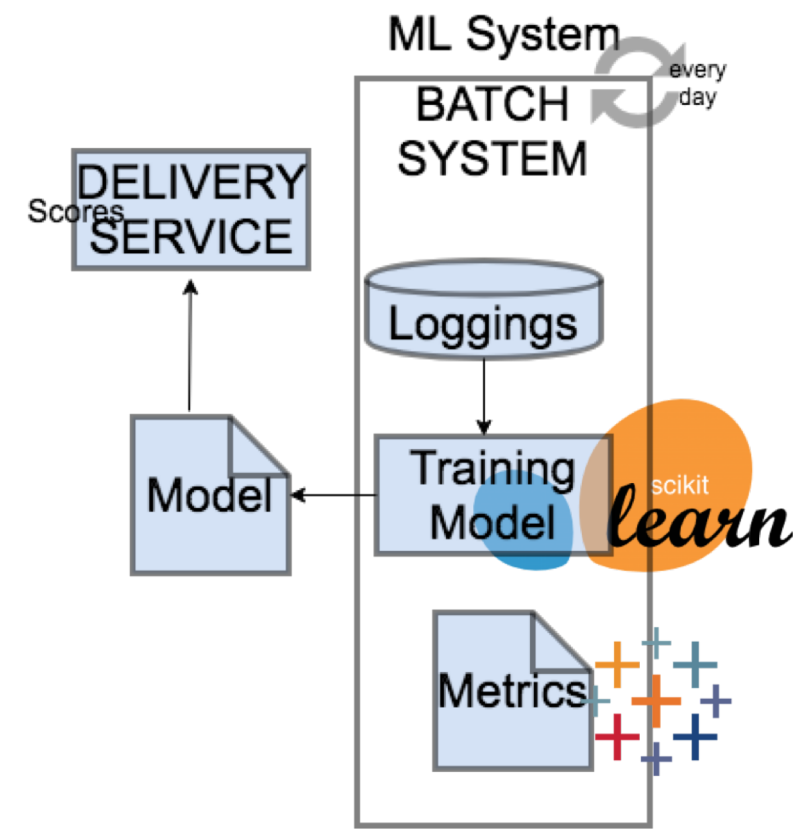
### Features:

- ~~User history,~~
- User features,
- Route features ,
- Ad features with, colors, text

### Tensorflow

- ~~Optimization technique~~
- ~~Embeddings~~
- ~~Crossed columns~~
- ~~Hashing~~

Not flexible but fast and easier to implement.



# Challenges in building the model

**Goal:** model, that estimates the likelihood of whether an impression will result in a click or not.

The screenshot displays a flight search interface with three filter tabs at the top: 'Best' (£45, 1h 33 average), 'Cheapest' (£45, 1h 33 average), and 'Fastest' (£90, 1h 30 average). Below the filters, there is a sponsored advertisement for Virgin Trains with the text 'Get to Edinburgh nae bother' and 'Good brew, good food and BEAM on-board entertainment. Sorry planes, you're just not trains.' To the right of the ad is a 'Sponsored' label and a 'Check prices and save' button. Below the ad, there are two flight options from easyJet. The first option is from LGW to EDI, departing at 08:10 and arriving at 09:45, with a 1h 35m duration and 'Direct' service. The second option is from EDI to LGW, departing at 18:55 and arriving at 20:25, with a 1h 30m duration and 'Direct' service. To the right of the flight options, there is a 'Select →' button with a hand icon pointing to it, and a price of £45. Below the flight options, there is another flight option from easyJet from LGW to EDI, departing at 08:10 and arriving at 09:45, with a 1h 35m duration and 'Direct' service. To the right of this option, there is a 'Select →' button and a price of £45.

Filter	Price	Average Duration
Best	£45	1h 33 (average)
Cheapest	£45	1h 33 (average)
Fastest	£90	1h 30 (average)

**Virgin Trains** Get to Edinburgh nae bother  
Good brew, good food and BEAM on-board entertainment. Sorry planes, you're just not trains.

**Sponsored** Check prices and save  
Select →

**easyJet** 08:10 LGW → 09:45 EDI 1h 35 Direct  
**easyJet** 18:55 EDI → 20:25 LGW 1h 30 Direct

8 deals from £45  
Select →

**easyJet** 08:10 LGW → 09:45 EDI 1h 35 Direct  
8 deals from £45  
Select →

# Challenges: Which model to use?

## Model Possibilities (easy to read in node.js):

- **Logistic regression**
  - Random Forest : gets lost
  - Neural networks: too slow hard to put it in json
- } Gridsearch for hyperparameteres

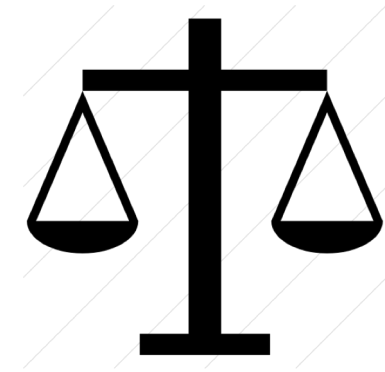
## Solvers:

- Logistic regression: Liblinear, sag
  - **SGDClassifier**
- } Train all data at once
- } **Train data in batches**  
Saves memory

# Challenges: Categorical values

## Categorical values:

- One hot encoding
- **One hot encoding grouping less frequent features**
- Hashing trick



- One hot encoding

- One hot encoding grouping
- Less frequent features

Creatives	C1	C2	C3
C1	1	0	0
C2	0	1	0
C3	0	0	1

Creatives	C1	C2	C3	C4
C1	1	0	0	0
C2	0	1	0	0
C3	0	0	1	0
C4	0	0	0	1

# Challenges: Categorical values

- **Hashing Trick: Same hashing function in training, testing and production**

id	features
123	creative1, advertiser2,mobile, etc.
321	creative2, advertiser4,mobile, etc.



id	Feat_1	Feat_2	Feat_3	....	Feat_k
123	1	0	1	....	0
321	1	0	0	....	1

We gain



**feature space reduction**  
(k is determined upfront)

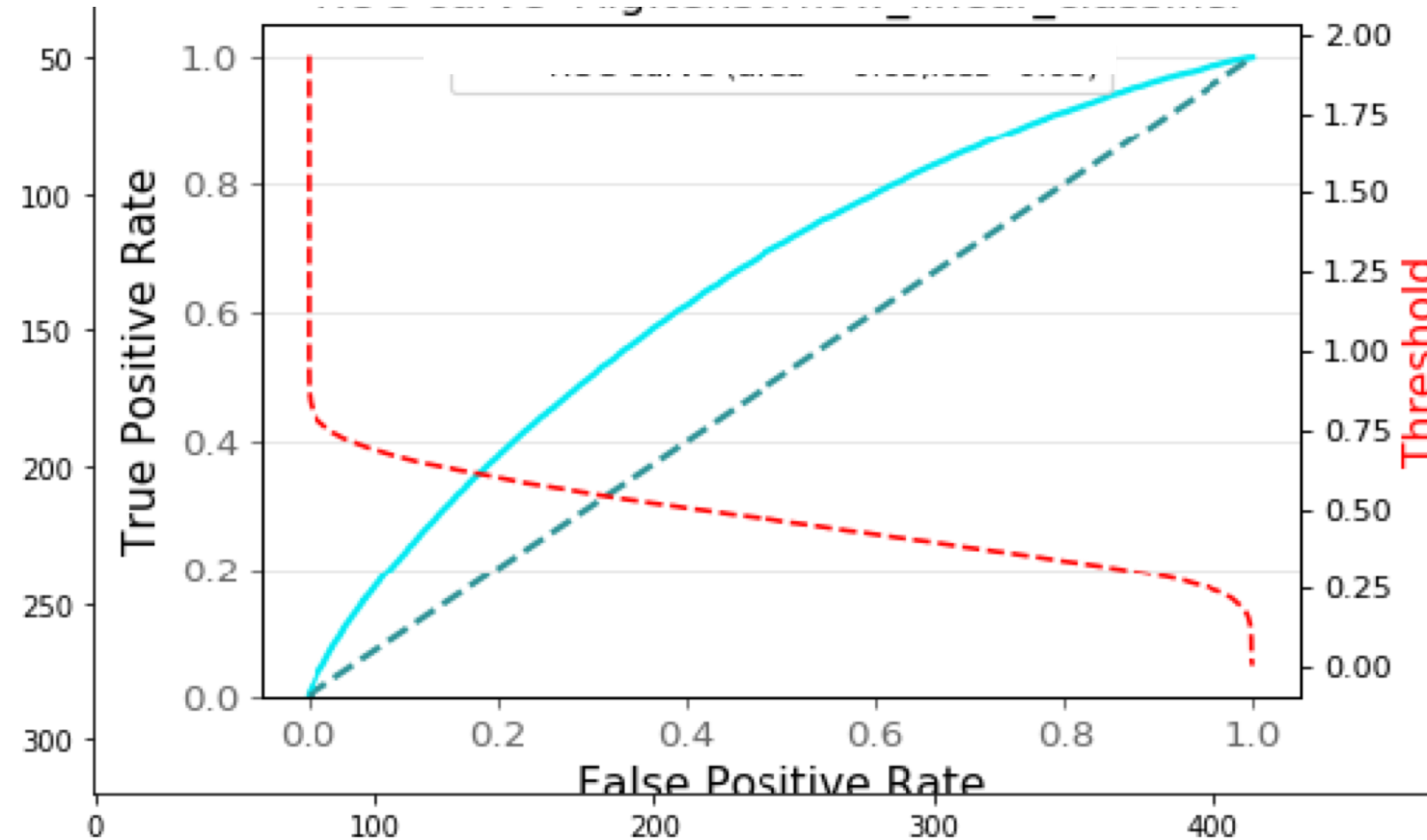


**new categories**  
n-grams are assigned  
to existing bins



**collisions**  
usually don't bother us

# Machine Learning Performance



**AUC:** if caring about ranking

**Log-Loss:** if caring about the value of CTR

Other metrics:

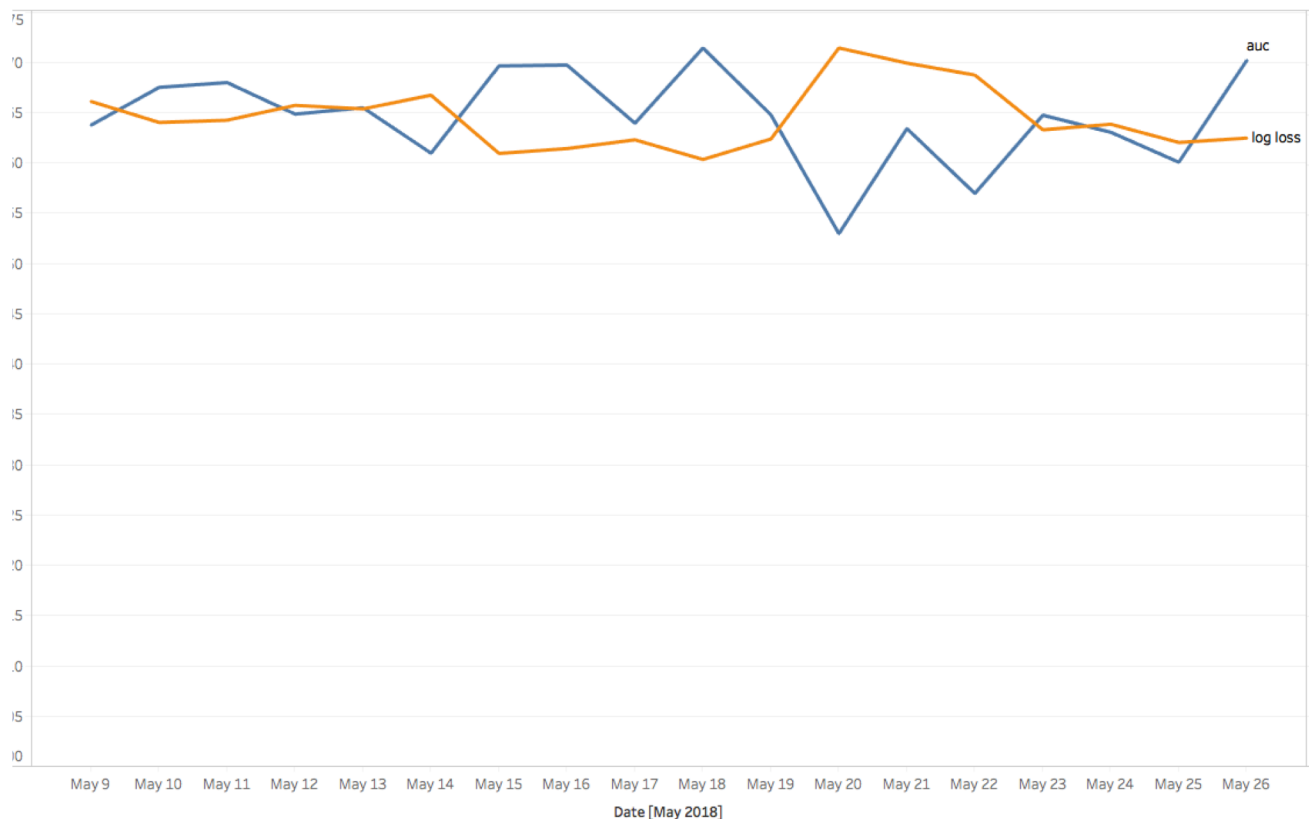
**Precision at 1:** based on target groups.

**Mean Reciprocal Rank**

# Challenges: How to monitor performance?

## Updating model based on different Sampling methods.

prediction-graph



click-prediction

to	algorithm	sampling	days
2018-05-10	logistic_regression_balanced	random	7
2018-05-11	logistic_regression	undersampling	6
2018-05-12	logistic_regression_balanced	not-sampled	5
2018-05-13	logistic_regression	undersampling	7
2018-05-14	logistic_regression_balanced	not-sampled	7
2018-05-15	logistic_regression	undersampling	3
2018-05-16	logistic_regression_balanced	not-sampled	6
2018-05-17	logistic_regression_balanced	not-sampled	7
2018-05-18	logistic_regression_balanced	not-sampled	7
2018-05-19	logistic_regression_balanced	not-sampled	7
2018-05-20	logistic_regression_balanced	not-sampled	7
2018-05-21	logistic_regression_balanced	undersampling	7
2018-05-22	logistic_regression_balanced	undersampling	5
2018-05-23	logistic_regression_balanced	not-sampled	3
2018-05-24	logistic_regression_balanced	not-sampled	7
2018-05-25	logistic_regression	random	7
2018-05-26	logistic_regression	random	7
2018-05-27	logistic_regression_balanced	random	5

# The road ahead: Balancing exploitation and exploration



## EXPLOITATION

Playing the machine that (currently) pays out the most.



**Choose ad based on ONLY CTR**



## EXPLORATION

Playing the other machines to see if any pay out more.



**Choose ad based on OTHER criteria**



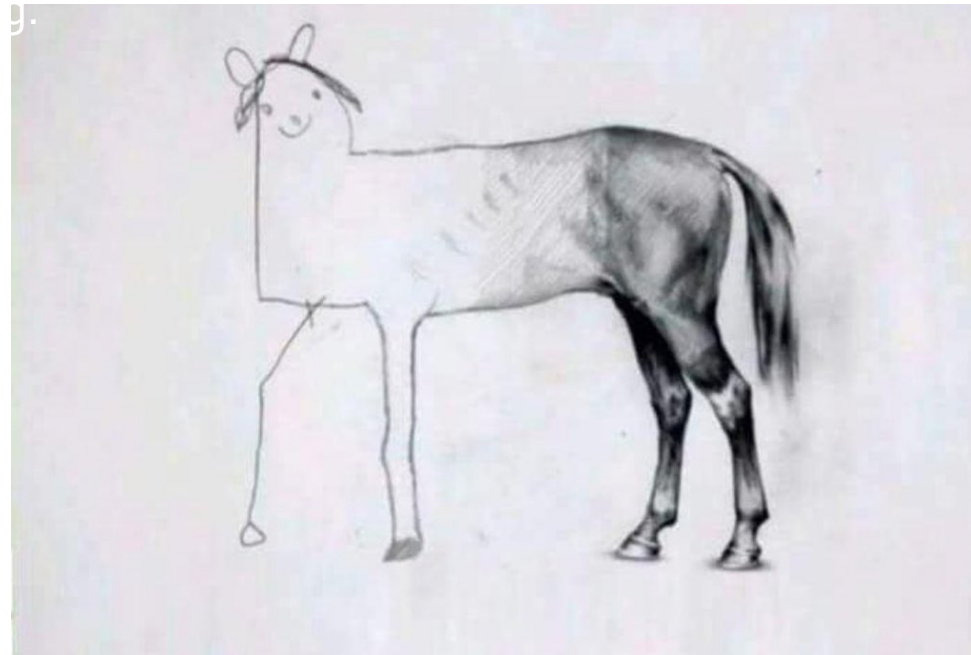
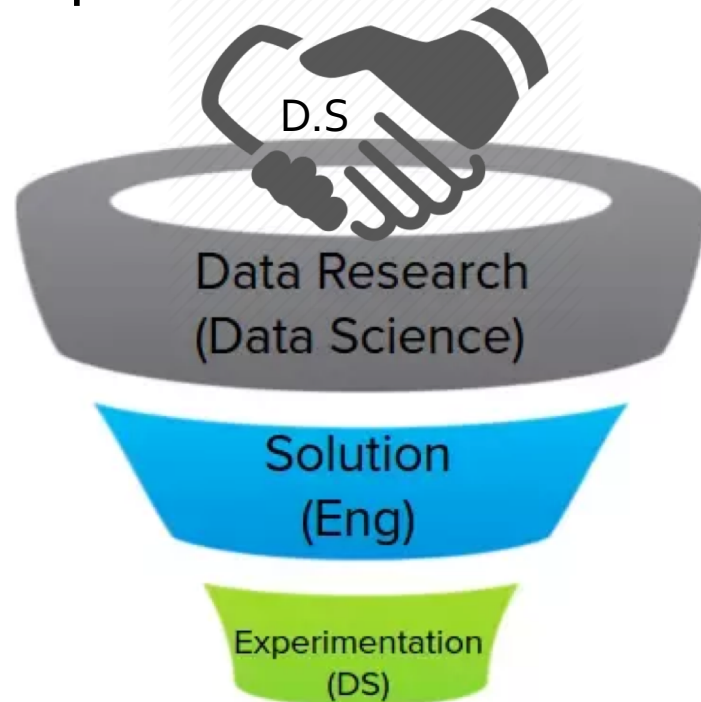
# The road even further ahead: quality

Post-click, colors, images, text,  
landing page



# Conclusions

1. Machine Learning can bring many benefits to the product the challenge is to prove it
2. Bringing ML into production could hard at the beginning
3. Cooperation between data scientists and engineers is crucial



Thank you

@ruthygarcia

Questions?

