



Machine Learning for Digital Advertising

Marc Garcia

Outreach digital - June 14th, 2016

Digital advertising: a data scientist point of view

What is the Internet? A lot of bored people...



...who use a gadget...



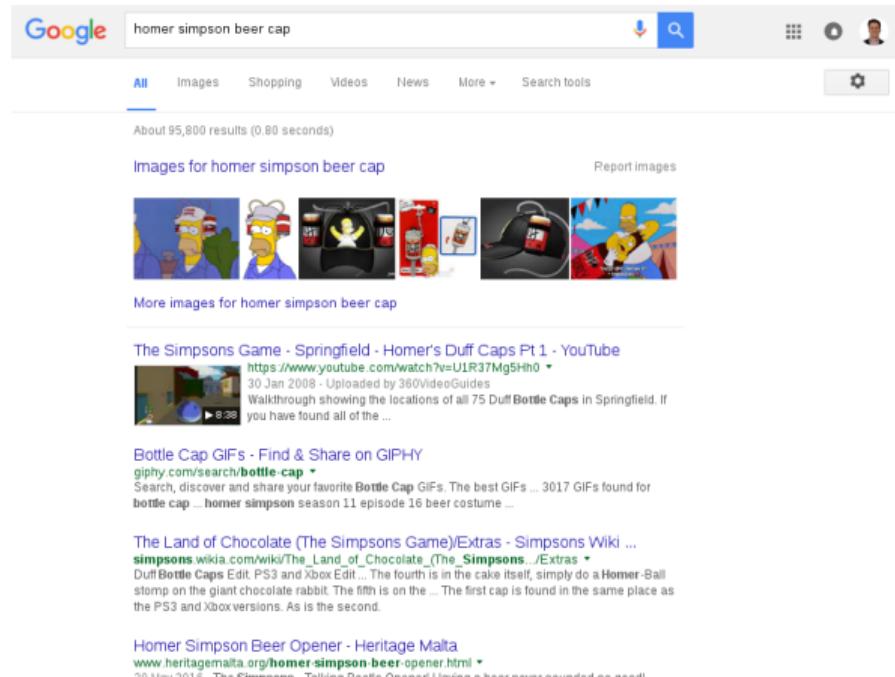
...to access useless sites.



The screenshot shows a Facebook page with the following details:

- Title:** When I Was Younger, I Put My Face Close To The Fan To Hear My Robot Voice
- Description:** Become a Fan
- Wall:** Wall, Info, Photos, Discussions, Reviews, Boxes, >
- Post 1:** When I Was Younger, I Put My Face Close To The Fan... + Fans
When I Was Younger, I Put My Face Close To The Fan... Just Fans
- Post 2:** When I Was Younger, I Put My Face Close To The Fan To Hear My Robot Voice Good morning...I see the assassins have failed.
Yesterday at 11:23am
- Engagement:** 885 people like this, View all 164 comments
- Post 3:** When I Was Younger, I Put My Face Close To The Fan To Hear My Robot Voice
Stop using BOB and start using THIS!!
LOOK IT'S YOUR MUM ☺ This is your mum, she is / secretly seeing Bob. But soon / you may call him Daddy :)
- Engagement:** 223 people like this, View all 115 comments
- Side Column:** Create an Ad, Watch the Winning Videos, ZOO.WORLD
- Bottom:** When I Was Younger, I Put My Face Close To The Fan To Hear My Robot Voice

Ok, some of them are looking for something.



Google search results for "homer simpson beer cap". The search bar shows the query. Below it, the "All" tab is selected, followed by other categories like Images, Shopping, Videos, News, More, and Search tools. It displays about 95,800 results in 0.80 seconds. The first result is an image search titled "Images for homer simpson beer cap" with a link to "Report images". Below the images, there's a link to "More images for homer simpson beer cap". The second result is a YouTube video titled "The Simpsons Game - Springfield - Homer's Duff Caps Pt 1 - YouTube" with a thumbnail showing Homer Simpson. The third result is a GIPHY link titled "Bottle Cap GIFs - Find & Share on GIPHY" with a thumbnail showing a bottle cap. The fourth result is a Wikipedia link titled "The Land of Chocolate (The Simpsons Game)Extras - Simpsons Wiki ..." with a thumbnail showing a chocolate rabbit. The fifth result is a website link titled "Homer Simpson Beer Opener - Heritage Malta" with a thumbnail showing a bottle cap.

homer simpson beer cap

All Images Shopping Videos News More Search tools

About 95,800 results (0.80 seconds)

Images for homer simpson beer cap Report images

More images for homer simpson beer cap

The Simpsons Game - Springfield - Homer's Duff Caps Pt 1 - YouTube
<https://www.youtube.com/watch?v=U1R37Mg5Hh0> ▾
30 Jan 2008 - Uploaded by 360VideoGuides
Walkthrough showing the locations of all 75 Duff Bottle Caps in Springfield. If you have found all of the ...

Bottle Cap GIFs - Find & Share on GIPHY
giphy.com/search/bottle%20cap ▾
Search, discover and share your favorite **Bottle Cap** GIFs. The best GIFs ... 3017 GIFs found for bottle cap ... homer simpson season 11 episode 16 beer costume ...

The Land of Chocolate (The Simpsons Game)Extras - Simpsons Wiki ...
[simpsons.wikia.com/wiki/The_Land_of_Chocolate_\(The_Simpsons\)_Extras](http://simpsons.wikia.com/wiki/The_Land_of_Chocolate_(The_Simpsons)_Extras) ▾
Duff Bottle Caps Edit, PS3 and Xbox Edit ... The fourth is in the cake itself, simply do a Homer-Ball stomp on the giant chocolate rabbit. The fifth is on the ... The first cap is found in the same place as the PS3 and Xbox versions. As is the second.

Homer Simpson Beer Opener - Heritage Malta
www.heritagemalta.org/homer-simpson-beer-opener.html ▾
Malta 1912 The Commonwealth Timeline Project. Accessed 16 November 2012

What is our goal?

What is our goal?

Get their money!



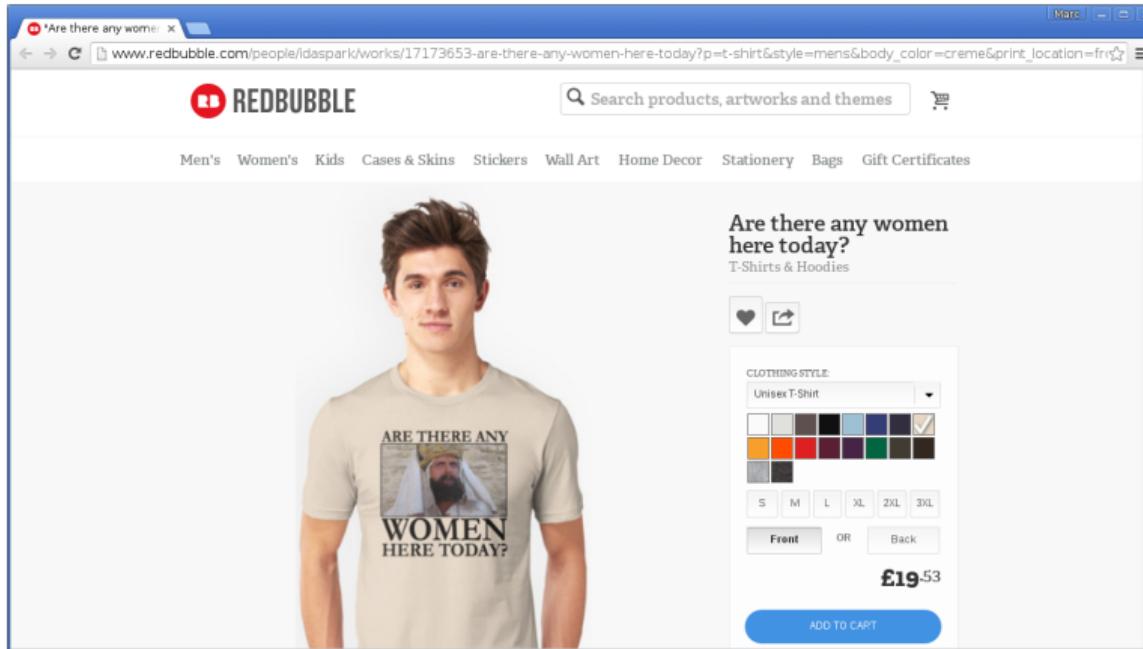
What is our goal?

That's the truth,

but let's say it in a more subtle way...

What is our goal?

We want to attract them to our site...



What is our goal?

...and sell our product.

Payment details



VISA VISA Electron MasterCard Maestro Discover

Card number

Expiry date MM/YY

 /

CVV

Name on card

How does it work?

How does it work?

The user journey



How does it work?

The user journey



How does it work?

The user journey



How does it work?

The user journey



How does it work?

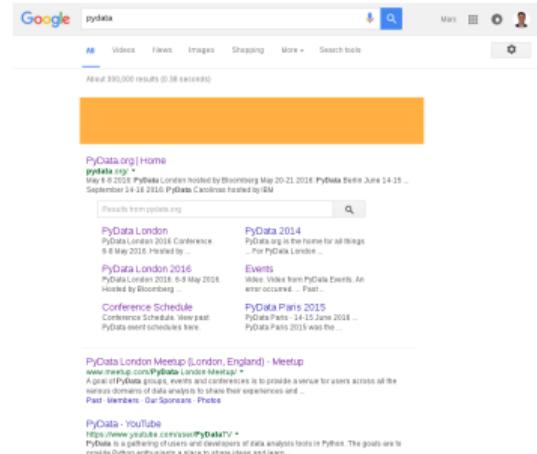
The user journey



Ad exchange point of view

How the ad exchange works?

A publisher has an available ad space.



How the ad exchange works?

Advertisers bid for their ads on the cost for a click.

Machine Learning Python - Develop New Skills. Be In Demand
 udacity.com/Machine-Learning-Python •
Launch A Career In Technology Today
Expert Mentors - Get Hands-On Experience - Built By Silicon Valley
Best-in-Class Curriculum
Corporate Training
Nanodegree Programs
Career Support

bid: £0.35

Free Russian Wife Search - 30,000 Beautiful Brides are Waiting
 www.dream-marriage.com/Russian Women •
10 Years in Business. Join for Free
Trusted Since 2003 - Live Video Chat - Confirmed Profiles - New Women Daily
Search Gorgeous Women - Photo Albums - Girl of The Week

bid: £0.20

How the ad exchange make money?

$$\text{income} = \text{bid} \cdot \frac{\text{clicks}}{\text{impressions}} \quad (1)$$

- Impressions are constant regardless of the ad
- Bids are known for each ad
- **Clicks** are unknown, but can be **predicted**

Ad exchange point of view

What is the best option?

Machine Learning Python - Develop New Skills. Be In Demand
[udacity.com/Machine-Learning-Python](https://www.udacity.com/Machine-Learning-Python) •
Launch A Career In Technology Today
Expert Mentors - Get Hands-On Experience · Built By Silicon Valley
Best-in-Class Curriculum
Corporate Training
Nanodegree Programs
Career Support

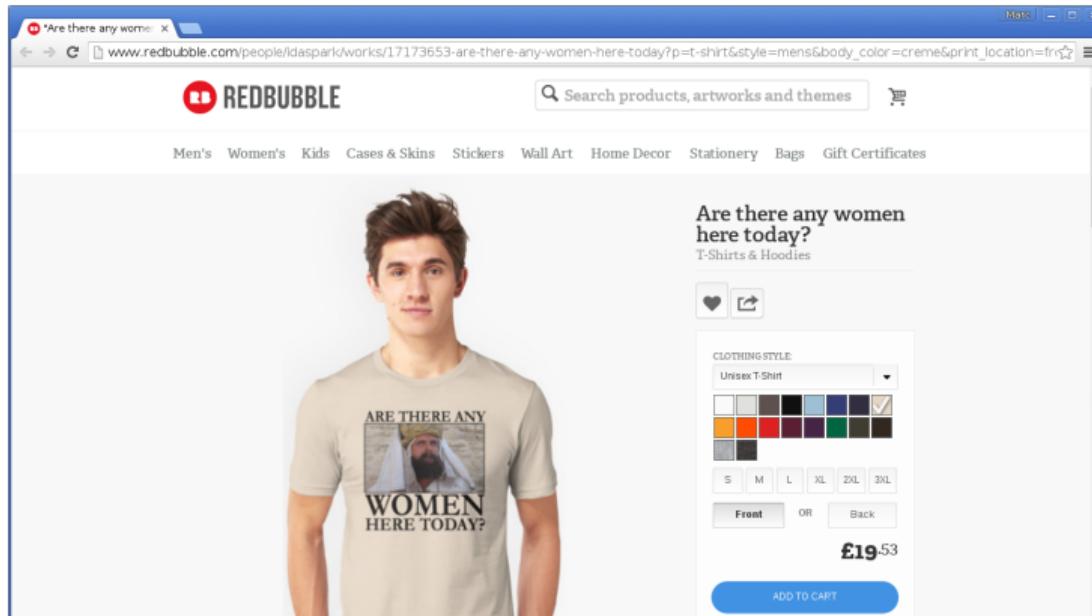
bid: £0.35
expected clicks (for 1000 impressions): 25
expected profit (for 1000 impressions): £8.75

Free Russian Wife Search - 30,000 Beautiful Brides are Waiting
www.dream-marriage.com/Russian-Women •
10 Years in Business. Join for Free
Trusted Since 2003 · Live Video Chat · Confirmed Profiles · New Women Daily
Search Gorgeous Women · Photo Albums · Girl of The Week

bid: £0.20
expected clicks (for 1000 impressions): 47
expected profit (for 1000 impressions): £9.40

E-commerce point of view

T-shirt e-commerce example

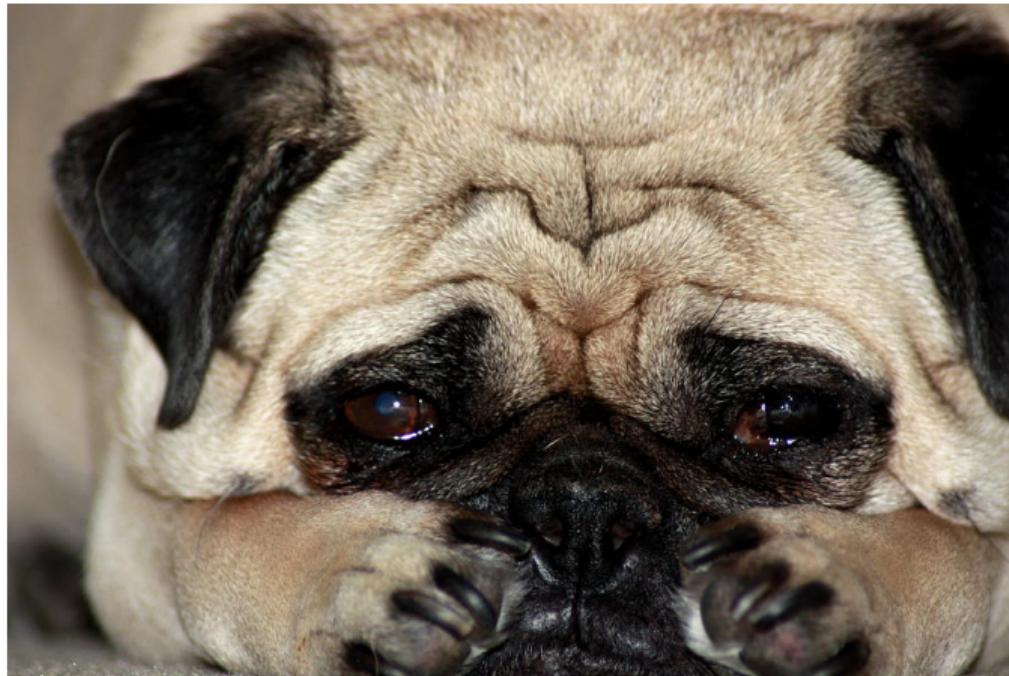


Income statement¹

Sales	£23.89	100%
Cost of goods sold	£11.72	60%
Delivery	£4.36	18.25%
Payment gateway	£0.84	3.5%
Total direct costs	£19.53	81.75%
Available for marketing	£4.36	18.25%
Average cost-per-click	£0.20	
Conversion rate (break-even)	22	

¹Excludes many costs (product returns, freight-in, development, support, hardware, salaries, etc.)

Not so easy to make money...



An optimization problem

What can be optimized?

- Higher prices (added value, branding...)
- Lower production costs
- Leverage marketing costs
 - Virality (share on social networks...)
 - Recurrence (newsletter to customers...)
- **Conversion rate optimization**
 - Targeting / market segmentation
 - Ad optimization
 - Website optimization (A/B testing...)

Conversion rate optimization: the numbers

$$\text{income} = \text{average_price} \cdot \text{margin} \cdot \text{num_sales} - \text{average_cpc} \cdot \text{num_clicks} \quad (2)$$

Conversion rate optimization: Features



Shop for monty python t-shirt on Google

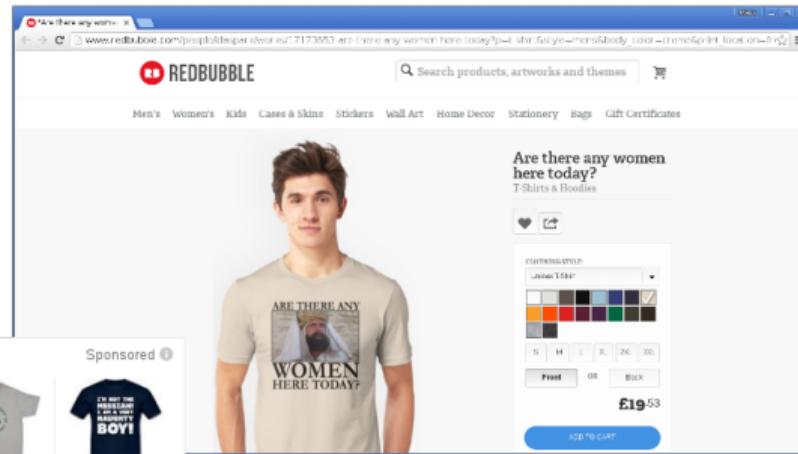
				
T-Shirt: Monty Python- ... £20.99 AllPosters.co... 	Monty Python 'Dead Parrot ... £20.00 Freemans 	Men's White Monty Python T- £4.99 TruffleShuttle.... 	Monty Python Life Of Brian T- £26.00 notonthehigh... 	Funny Monty Python Parody £17.99 Spreadshirt

 Special offer

T Shirts Monty Python - Latest Must-Have Fashion

 www.amazon.co.uk/clothing ▾
Free UK Delivery on Eligible Orders

Are there any women here today? T-Shirts & Hoodies



The screenshot shows a RedBubble product page for a t-shirt. The shirt features a portrait of a man in a white robe with the text "ARE THERE ANY WOMEN HERE TODAY?" printed on it. The page includes a color palette for customization, size selection (S, M, L, XL, XXL), and a "ADD TO CART" button.



Features: Users

What Facebook, Google... know about our users?

- Age
- Gender
- Location
- Interests / Pages visited
- Friends / Interactions
- Sociocultural status
- ... **mostly everything**

What do we know about them?

- Not much individually
- **We can access aggregated data**



An optimization problem

Features: Ads

ASDA
Save money. Live better.

Tempting prices on BBQ favourites.

£2 each

Extra Special chocolate pork sausages 12 pack

Extra Special caramelised onion pork sausages 6 pack

Shop now >

What Science Knows About Attraction And Love
About ...
Glimpse into the ...

21 Women So Beautiful They Put Nature To Shame
They Put ...
Feminine energy holds ...

24 Reasons Why Cuba Is The Best Place To Live
We are going to introduce ...

Datadog, Inc. @datadoghq · Apr 26
This is how you should monitor Apache Kafka metrics:

Read our new guide about key Apache Kafka performance metrics.
datadoghq.com

4 5 ...
Promoted



SWEDISH GIRLS
READ MORE

21 Women So Beautiful They Put Nature To Shame
READ MORE

24 Reasons Why Cuba Is The Best Place To Live
READ MORE

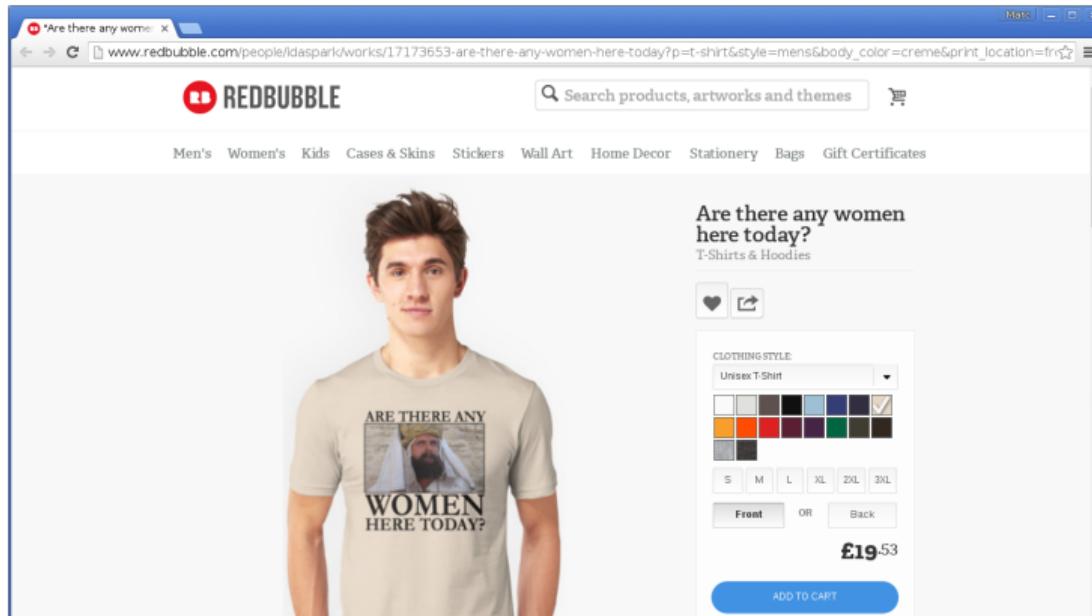
Jaw-Dropping Cheerleading Slip-Ups That Will Make You Do See the side of ...
READ MORE

Hollywood Icons Who Lost Their Virginity To Fellow Celebs Hear about celebs getting ...
READ MORE

A Side Of Cuba The World Hasn't Seen Until Now
Now that it's easier to ...

An optimization problem

Features: Website



Features: Summary

Users

- Age
- Gender
- Location
- Interests
- Friends
- Sociocultural status
- ...

Ads

- Source
- Message
- Type (image, text,...)
- Placement
- Site
- ...

Website

- Layout
- Colours, font size...
- Content density
- Offers, reviews...
- Conditions (delivery, returns)
- Steps to purchase
- ...

Other

- User behaviour
- Day of week
- Time
- Weather
- Social environment
- ...

An amazing dataset

Age	28	Web layout	basic
Gender	female	Web main colour	green
Location	Lewisham	Conditions	free delivery
Occupation	receptionist	Returns	free
Interests	hiking, salsa	Product type	shoes
Education	university	Quality	high
Device	smartphone	Price	£79
Source	Google search	Day	business day
Placement	top	Time	evening
Ad type	image	Purchase	true
Message type	sales offer		

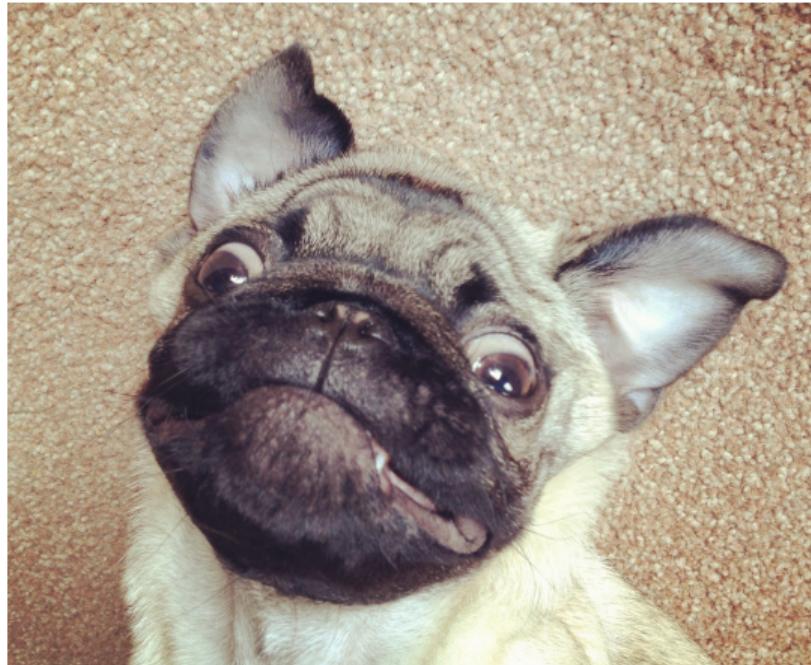
What could we get from a dataset like this?

- Women aged 25 to 35 buy shoes on rainy Sundays
- Users leave pages with more than 800 characters on the layout quite fast
- People from city A buy more product X, while people from city B, prefer product Y
- Ads with red colours have more clicks, but we get less conversions from them

And all this means higher profits.

An optimization problem

Amazed?



Introduction to machine learning

So, what is machine learning?

Different definitions

- A fancy word for statistics
- The science of programming computers, so we don't tell them what to do, but we tell them how to learn what to do
- A set of techniques to implement statistical models that are able to predict, and find patterns in the data

Similar concepts: Big data, data science, data mining, deep learning, artificial intelligence...

What can machine learning do?

It can accurately compute the probability of a certain user, to click on a certain ad.

Machine Learning Python - Develop New Skills. Be In Demand
www.udacity.com/Machine-Learning-Python •
Launch A Career In Technology Today
Expert Mentors - Get Hands-On Experience - Built By Silicon Valley
Best-In-Class Curriculum
Corporate Training
Nanodegree Programs
Career Support

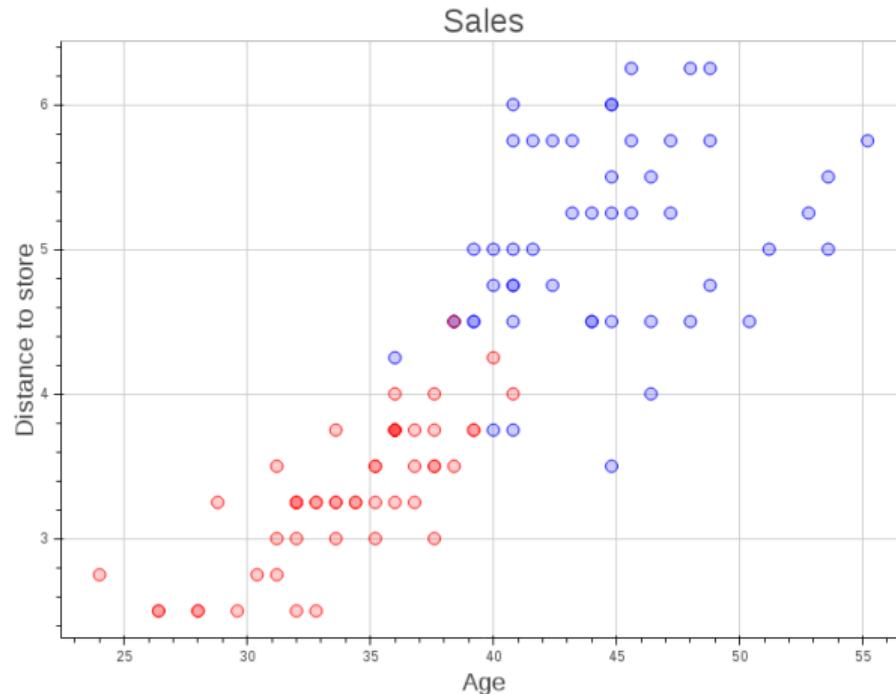
bid: £0.35
Click probability: 0.025
Expected profit: £0.00875

Free Russian Wife Search - 30,000 Beautiful Brides are Waiting
www.dream-marriage.com/Russian-Women •
10 Years in Business. Join for Free
Trusted Since 2003 - Live Video Chat - Confirmed Profiles - New Women Daily
Search Gorgeous Women - Photo Albums - Girl of The Week

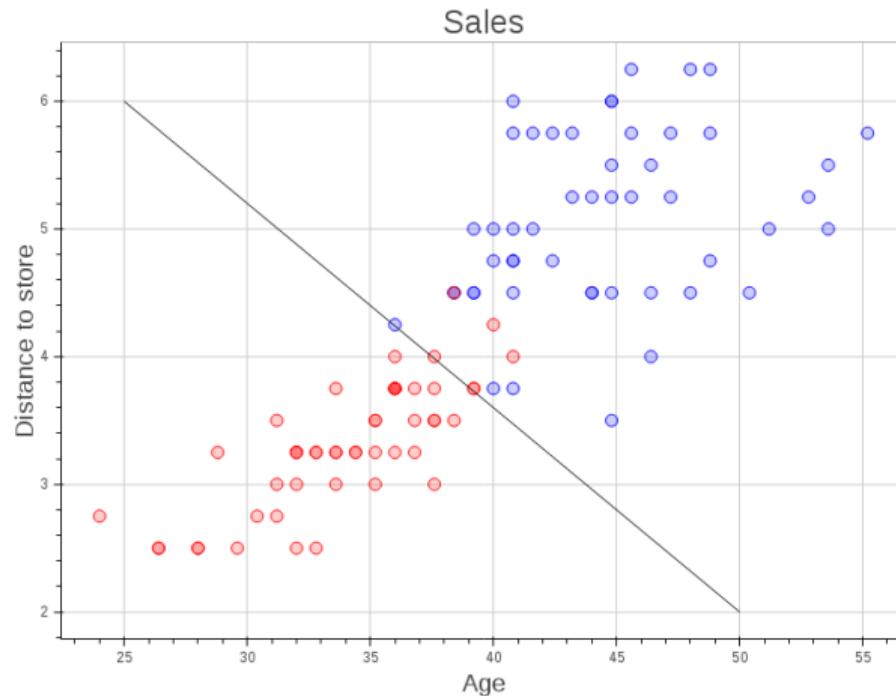
bid: £0.20
Click probability: 0.047
Expected profit: £0.00940

A small difference in profit but at scale: multiply by 1bn daily Fb users: **£650,000/day**

How does machine learning work?



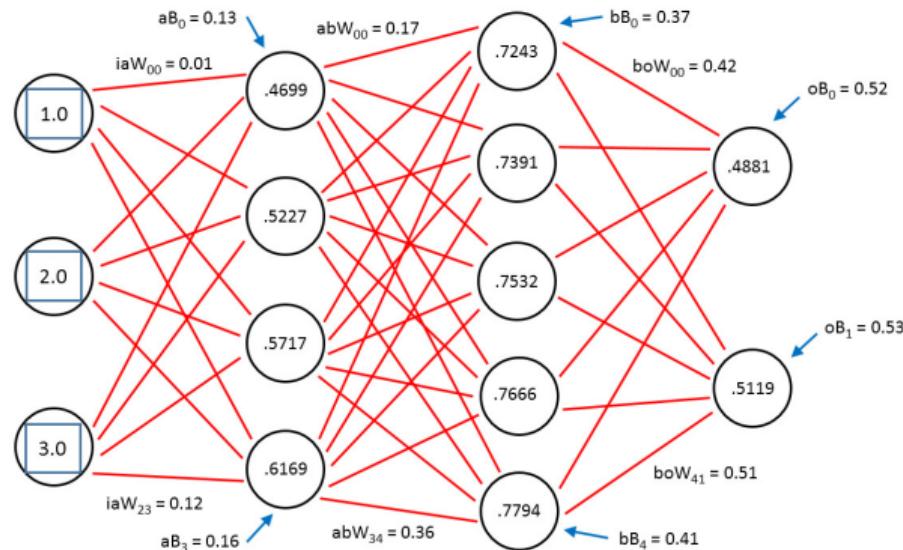
How does machine learning work?



Some maths

$$Pr_{click} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2)}} \quad (3)$$

More complex models: neural networks



Machine learning for digital advertising

Some problems

- Data is expensive (average of £0.20 per row) :(
- No public data
 - So, we need to obtain it all ourselves...
 - ...and it's expensive :(
- Unbalanced / sparse data (1 purchase for 100 clicks)
 - So, we need 100x the data...
 - ...and it's expensive :(
- High (or very high) dimensionality
 - So, we need a huge number of rows to avoid the curse of dimensionality
 - ...and it's expensive :(

More problems

- We bias the data when we know who is more likely to buy
 - Exploration an exploitation problem
 - We have to buy "bad" clicks to avoid it
- Many categorical variables
 - They don't play well with many models
- Data seasonality
 - Black Friday, sales, Christmas...
- Access to data
 - Google, Facebook do not give access to raw data
 - We have to use aggregated data
 - And they do not let you cross variables
 - We need to assume independence among them

He is very sad now



What models can be used?

Some techniques that gave good results:

- Gradient boosting trees
- Active learning (semisupervised learning)
- Online learning (for data freshness)

Questions?



@datapythonista

<http://datapythonista.github.io>

<https://www.linkedin.com/in/datapythonista>