Ad industry introduction and DS example

Ella

Advertising industry

Overview

- Definition
  - Uses various media outlets to motivate people to buy products and services or change their attitudes.
- Usage
  - convinces people to buy a particular goods or services
    - creates and maintain a brand identity or brand image;
    - communicates a change in their existing product line;
    - introduces of a new product or service;
    - increase the buzz-value of their brand .
  - aims at raising money for charity, for gaining support for political parties, or encouraging some actions.
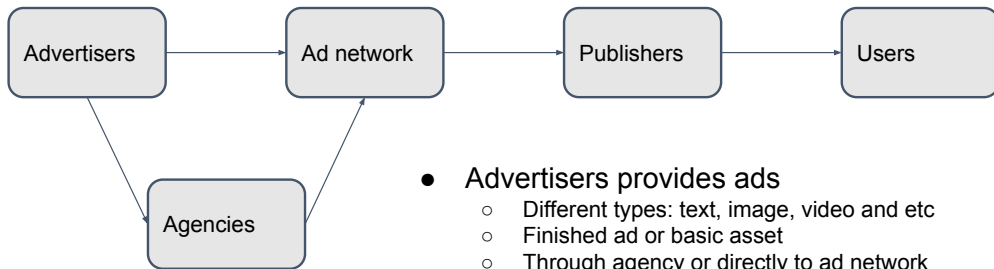
Advertising industry background

- Channels
  - Traditional: TV, Radio, magazine, newspaper, billboard
  - Digital: online/internet
- Spend

Global digital ad spending beat TV for the first time in 2017
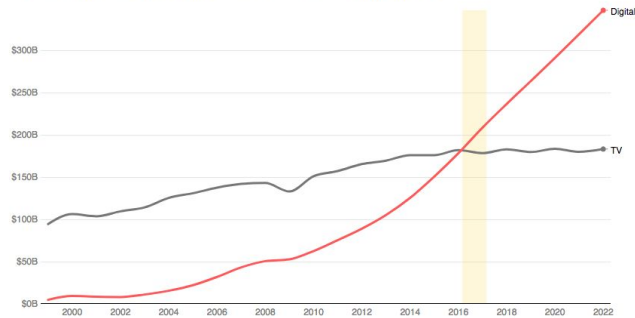
Stakeholders in advertising

```
Advertisers  →  Ad network  →  Publishers  →  Users
     ↓              ↑
         Agencies
```

- Advertisers provides ads
  - Different types: text, image, video and etc
  - Finished ad or basic asset
  - Through agency or directly to ad network
- Ad network
  - Connects multiple advertisers and publishers
  - Traditional or online (mainly)
- Publishers provides ad slots (inventory) to show ads
- Users are the actual ad viewers.
  - Related behavior

Give some examples

Advertisers

- Large advertisers
  - Have their own marketing/DS in house, sometimes work with ad agencies.
  - Usually have deals with ad network / publishers directly.
  - Generates their own creatives/advertisements
- Medium to small advertisers
  - Have limited resource.
  - Sometimes use ad agencies, who takes care of all ad creation and connects with ad networks.
- Performance driven v.s. Brand advertisers
  - Performance driven: focus on short term metrics like impressions, clicks, conversions, purchase/sales
  - Brand advertiser: aims to build long term connection with users.

http://www.businessinsider.com/companies-are-cutting-out-ad-agencies-and-going-in-house-2017-6

Important metrics (direct sale)

- Funnel
  - Impression -> Click -> Conversion -> Sale
- Cost type
  - By click: Favors advertisers because publisher only get paid once a click is made on an ad.
  - By impression: Favors publishers since no risk associated with a bad quality creative delivered by the advertiser
  - By conversion: Not widely used.
- Terminology
  - Cost per click: CPC
  - Cost per thousand impressions: CPM
  - Click through rate: CTR
  - Conversion rate: CVR
  - ROAS: return on ad spend
- Second auction price bidding

Why second price bidding: second price auctions induce truthful bidding.

https://economics.stackexchange.com/questions/12389/what-are-the-advantages-and-disadvantages-of-second-price-auctions

http://adclickmedia.com/blog/the-ultimate-guide-to-making-money-from-ad-networks/

Important metrics (brand)

- Brand advertisers
  - Creates sustainable advantage over competition by bonding meaning to the brand name.
  - Plants and grows associations and ideas about particular brand in the minds of people.

Google brand lift intro: https://www.youtube.com/watch?v=ZnXtoSF-kwU
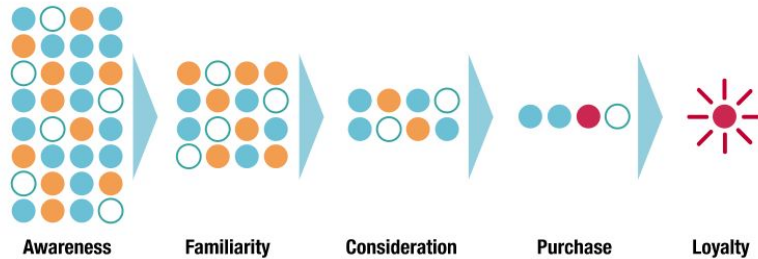
Publishers

- Purpose: to monetize from their own website's traffic
- Apply to ad network to be eligible publisher
  - Decent traffic
  - Update content
  - Follow ad network guideline
- Place ad tag on ad zone.
  - Ad tag: small piece of code
  - Function: to identify that zone and to track everything associated with ads shown and potential clicks.
- Get payment periodically from ad network

http://adclickmedia.com/blog/the-ultimate-guide-to-making-money-from-ad-networks/

Ad networks

- Connects advertisers to publishers that want to host advertisements as middleman.
  - Aggregate ad space (inventory) from publishers
  - Match advertiser demand with supply
- Why ad networks are necessary
  - Save time for business/advertisers from searching for optimal placement spots to show their ads.
  - Save time for publishers from searching for partner to fill all of their inventory.
- Digital ad networks examples
  - Google Adwords (other roles of Google?)
  - Facebook audience network
  - Criteo
  - …

Adword wiki: https://en.wikipedia.org/wiki/AdWords
Basics of ad network:
http://adclickmedia.com/blog/the-ultimate-guide-to-making-money-from-ad-networks/

Ad networks main functions

- ● Automatic ad serving
  - ○ Why is it necessary?
  - ○ To perform complex rules and algorithms to select the most suitable **ad** for each **ad slot** to display to each **viewer**.
  - ○ Rules are defined by publisher, advertiser and ad server. Including
    - ■ targeting criteria
    - ■ viewing frequency
    - ■ location of ad placement
    - ■ ad format
    - ■ ad priority
    - ■ earning potential and etc
  - ○ Given above constraints, results needs to be returned within milliseconds.
    - ■ Long page load time will turn viewers away.

https://www.adspeed.com/Blog/ad-serving-1962.html

Major steps during ad serving

1 User visits a webpage

2 Ad request is sent to an ad server

3 Orders advertisers for bids

4 Winning bid is selected

Winner pay second price

5 Ad created is served (one impression)

sovrn

Ad networks main functions (cont'd)

- Ad tracking
  - Records all interactions between the viewers and the ad for further processing and analysis: impressions, clicks, conversions.
  - Measure and evaluate each ad for its performance.
- Ad management
  - Different advertisements (creatives, ad groups, campaigns) are generated for different purposes, i.e., targeting groups or products and etc.
- Reporting
  - Via ad tracking, reporting system presents numbers and metrics in multiple ways using tables, grids, charts and other visual elements.
  - Typical reports include impressions, clicks, click-through rates (CTRs), revenue/expense, conversions, and events.

https://www.adspeed.com/Blog/ad-serving-1962.html

Ad networks main functions (cont'd)

- Ad billing
  - Pays the publisher
  - Bills the advertiser
- Application Program Interface (API)
  - Provides more advanced setups and integrations.
  - Helps the publisher, advertiser, ad network or agency build customized processes, automate tasks and integrate the ad server with other business systems such as CRM, ERP or other external advertising systems.

Users

- Purpose of cookies
  - Small files a web server automatically sends to user computer's web browser when browsing certain websites.
  - When visit back, user's browser automatically sends the original cookie allowing that website to recognize user and tailor online experience accordingly. So ads are better targeted to their interests.
- Types
  - First party cookies: website visited by user generates.
  - Third party cookies: Other entities (e.g., ad network company) use first party cookies.
  - Example: blogger can have a "Like" button on their website, which will generate a cookie on visitor's computer, that cookie can later be accessed by Facebook to identify visitor and see which websites he visited. Such cookie is considered to be a third-party cookie.

https://blog.adroll.com/product/how-do-advertising-cookies-work
http://www.ravelrumba.com/blog/third-party-cookies/

Problem example

- Ads platform wants to predict click through rate of display ads of recommended content

ADs by outbrain
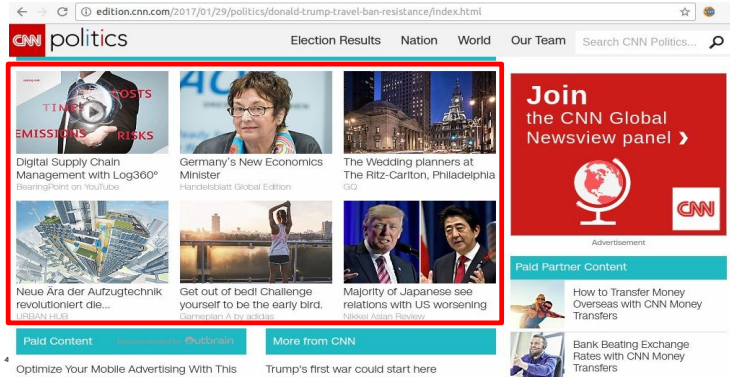
Outbrain ad click prediction project

- ● Who is Outbrain?
  - ○ Uses behavioral targeting to recommend contents to users.
  - ○ Works with 35,000 publishers and serves over 250 billion recommendations, 15 billion page views per month. Reach over 87% of U.S. Internet users.
- ● Project Goal
  - ○ Use multiple data source to predict whether user will click certain ad or not and rank ad candidates.
  - ○ [Data](#)

Provided data

Page Views
- uuid
- document_id
- timestamp
- platform
- geo
- traffic_source

Events
- display_id
- uuid
- document_id
- timestamp
- platform
- geo

Clicks
- display_id
- ad_id
- clicked

Promoted
- ad_id
- document_id
- campaign_id
- adv_id

Doc Meta
- document_id
- source_id
- published_id
- publish_time

Doc Topics
- document_id
- topic
- confidence

Doc Entites
- document_id
- entity
- confidence

Doc Categories
- document_id
- category
- confidence

Uuid: user id
Display id: ad slot
Document id:
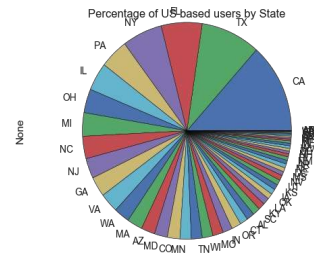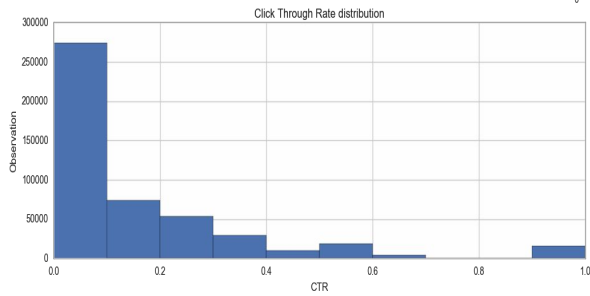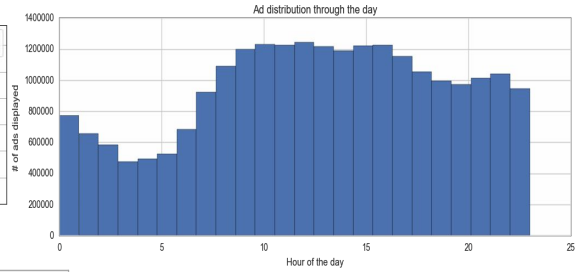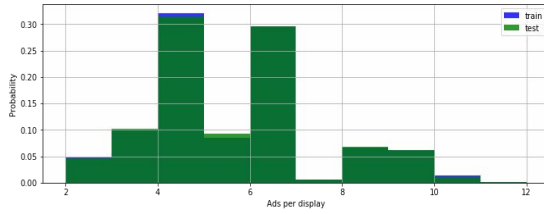website of ad slot

Provided data (cont'd)

Data exploration

- Ads shown
  - Distribution of number of ads shown per ad slot?
  - Ad repetition across ad slots
- Clicks
  - CTR distribution by date or by hour
- Traffic (user location)
  - Breakdown by geo
  - Breakdown by platform
- Topics, Categories and Entities

  ...

Data exploration

Feature selection

- **Feature processing**
  - Collapse if too many levels for categorical features, or consider top N levels.
  - Numerica features: if spread too wide, use log transformation
  - Missing value imputation
- **Feature selection**
  - Model with regularization to include all features
  - Tree based model and reply on feature importance plot

Model

- Logistic regression
- Boosting tree
  - Hyper parameters tuning
- Ranking SVM

Appendix