**Data Description**

iPinYou prepares a set of processed iPinYou DSP bidding, impression, click, and conversion logs from some advertising campaigns during seven days for the training an purpose, and a set of iPinYou DSP bidding logs from the following three days from the same set of advertising campaigns for offline testing purpose. The training and validation set is released to participating teams. The testing set is withheld to the judges. Here is some description of the data in the log and the processing that iPinYou has done for a variety of purposes.

For each line of the bidding log, the first column is bid ID, which uniquely identifies one ad impression opportunity; in each line of the impression or click log, the bid ID is same as that in the corresponding bidding log. IPinYou applies a hash function to bid ID before the log is released.

**Log Files**

For each line of the bidding log, it contains information about the ad slot, the user, and bidding information.

Timestamp column is when the bid request arrives at the DSP server.

Ipinyou ID is the user cookie set by iPinYou. iPinYou applies a hash function to the cookie ID before the log is released for user privacy protection. User Agent is the user agent of browser through which the user visit the publisher’s website.

IP address is the first three bytes of user IP address. iPinYou removes the last byte for user privacy protection before the log is released. Region ID is the region (a.k.a., province or state) ID where the user visits. CityID is the city where the user visits.

Ad exchange column represents which ad exchange this ad impression comes from. Currently, iPinYou bid on AdX (Google DoubleClick AdX), TANX (Alibaba), Baidu(Baidu),Youku(Youku),Tencent (Tencent),Amx(Google Mobile). iPinYou uses 1, 2, 3 ,4,5 to represent these exchanges .

Domain is the name of domain where the ad impression will show up. iPinYou applies a hash function to domain before the log is released. URL is where the ad impression will show up. The publisher may ask the ad exchange not to pass this information to DSPs. iPinYou applies a hash function to URL before the log is released. Anonymous URL ID is the domain page ID set by the ad exchange when URL is not available to DSPs, mentioned previously. Only one of URL and Anonymous URL ID is meaningful.

Ad Slot ID is the unique ID to show the location where ad impression will show up on the web page. One web page can have multiple ad slots.

Ad slot width and Ad slot height show the width and height of the ad slot respectively. Both columns can be an array, which means the publisher allows multiple widths and heights of the ad creative.

Ad slot visibility column represent whether the ad slot is above the folder.It is an enumerated type with values of Na, FirstView, SecondView,ThirdView,FourthView,FifthView,SixthView,SeventhView,EighthView,NinthView,TenthView,OtherView

Ad slot format may be fixed ad slot (Fixed) , the popup window (Pop),Background,Float.Na presents when not specified.

Paying price is what the winning DSP actually pays on the exchange. iPinYou does a linear scale of bidding price before the log is released. The price unit is RMB Yuan per CPM.

Creative ID is the ad creative of the advertiser for which the DSP bid on the exchange.

Bidding price is what the DSP bids on the exchange. iPinYou does a linear scale of bidding price before the log is released. The price unit is RMB Yuan per CPM.

Advertiser ID is the advertisement owner's ID.

User Tags are the user's tags IDs in iPinYou's user profile database. Only a part of the user's tags are released for participants to model user interests.

Stream ID is the stream id.

**Region and City File**

iPinYou also provides two separate text files. One is the mapping file from the region id to region name, and the other is the mapping file from the city id to the city name. All region and city values are currently geo locations from the mainland China. Each line of the region file is a pair of the form (region id, region name); each line of the city file is a pair of the form (city id, city name).

**Tag ID to Tag Name Mapping File**

iPinYou also provides a mapping file from the user tag id to the tag name. The tag name includes the hierarchical structure of tags.

**Anti-spam Filtering**

iPinYou filters publishers with suspicious activities from the bidding logs. iPinYou uses sophisticated algorithms to detect spams.

**Log Format**

The following are the log format and examples. Columns with (\*) means that data in the column is modified before the log is released.

**Log example**

|  |  |
| --- | --- |
| Column | Example |
| Bid ID(\*) | 71e9a371b25950af4e28688743260f3b |
| Timestamp | 20130606001907400 |
| iPinyou ID(\*) | VhkSLxSELTuOkGn |
| User-Agent | Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729) |
| IP(\*) | 27.36.3.\* |
| Region | 216 |
| City | 222 |
| Ad Exchange(\*) | 3 |
| Domain(\*) | tMxYQ19aM98 |
| URL(\*) | a0ab67bbe87f2e024338b093856580 |
| Anonymous URL ID | null |
| Ad slot ID | LV\_1001\_LDVi\_LD\_ADX\_2 |
| Ad slot width | 300 |
| Ad slot height | 250 |
| Ad slot visibility | 0 |
| Ad slot format | 0 |
| Paying price (\*) | 100 |
| Creative ID | 44966cc8da1ed40c95d59e863c8c75f0 |
| Bidding price (RMB/CPM) (\*) | 300 |
| Advertiser ID(\*) | 3386 |
| User Tags(\*) | 282825713030 |
| Stream ID | 0 |