

PROBLEM STATEMENTS

Build with Data

Apply your knowledge of data science to process and make sense of structured data, and derive meaningful insights and actions. These data will contain information about mobile internet access patterns and usage that is relevant to advertising.

Sample applications

Here are some example applications that can be built

- ¥ Geographic heat-maps for visualizing online activity over different periods of time
- ¥ Innovative visual representations for displaying high cardinality data.
- ¥ Searchable/Query-able indexes and trends for various attributes (Eg: What was the most popular app in Germany last month?)
- ¥ Analytics and real-time / streaming summarization across different dimensions
- ¥ Predictive modelling / machine-learning for click through rates and conversion rates.
- ¥ User behavior inference / collaborative filtering. Those who install apps X & Y, also install app Z

Note that these are just suggestions; you are free to come up with your own applications and showcase them!

Dataset

Get the sample data [here](#)

Full dataset will be provided to you in an USB drive at the venue on November 27th, 2015!

Each row in the data set represents an online access from a mobile phone for some user somewhere in the world, where Applift has the opportunity to show ads
Here is a complete list of data attributes:

BidId

Unique id that represents this row

TrafficType

"site" or "app" to represent if the access was through mobile website or mobile app

PublisherId

Unique id for that publisher

AppSiteId

Unique id for that app or site

AppSiteCategory

Category of the app or site. Eg: "Arts & Entertainment", "Hobbies and Interests" etc.

Position

Position of the slot where the ad is shown like "Footer", "Above the Fold" etc

BidFloor

Minimum CPM dollar amount that needs to be paid to show ads on this request.

Timestamp

Unix timestamp of the request

Age

Age of the person who made this access. "0" if unknown.

Gender

M or F. Maybe "0" if unknown

OS

Android, iOS etc.

OSVersion

Versions of the OS. Eg: For Android, they maybe 4.4, 5.0 etc

Model

Model of the phone or tablet used. Eg: Apple iPad or HTC PH39100

Manufacturer

Manufacturing brand of the phone. Eg: Apple or HTC

Carrier

Mobile carrier used. Eg: Airtel or AT&T Mobility

DeviceType

"Mobile/Tablet" if device is a mobile or a tablet. "Phone" if it is confirmed to be a mobile. "Tablet" if it is confirmed to be a tablet

DeviceId

Some unique id representing that device

DeviceIP

Some unique id representing the IP address used by that device (Hashed value)

Country

Three letter country code. Eg: USA

Latitude

Latitude of the location from which the request originated. Maybe 0 if unknown

Longitude

Longitude of the location from which the request originated. Maybe 0 if unknown

Zipcode

Pincode (also called zipcode) of the location. Format may differ by country. Eg: 62243 for USA. Maybe 0 if unknown

GeoType

Method used for obtaining the lat/long. Typically a represents the accuracy of the location: Eg: GPS, IP etc

CampaignId

Some unique-id representing the campaign that is showing the advertisement

CreativeId

Some unique-id representing the advertisement itself

CreativeType

0 for unknown. 1 for image-only. 2 for image-or-html. 3 for video. 4 for in-feed native advertisements

CreativeCategory

Category of the advertisement that is shown. For hierarchical category information, a "#" will be used to separate parent and child category. Eg: Air Travel#Business Travel

ExchangeBid

The CPM dollar amount that we are willing to pay to show the advertisement

Outcome

0 is ad was not shown. "w" if an ad was shown (representing an ad auction win). "c" if the ad was shown and it was clicked

You can also get the attributes and their explanation in a CSV file [here](#)

Please note that the data is artificially generated; it does not contain or represent the content of real data we process and nor does it contain any personally identifiable information. (PII)

Judging criteria

- ¥ **Product:** The utility of the product i.e. the value of the problems that you have solved for: interface, scalability and creativity
- ¥ **Analysis:** You will be judged on technical know-how, sophistication of the techniques used, and impact
- ¥ **Visualization:** Judged on the innovative ways used to analyze the data, intuitive feel, and aesthetics.
- ¥ **Quality** of the solution.
- ¥ **Degree of Complexity:** Technical achievement and Mastery

Resources

¥ [Real-time Advertising at 20 Billion requests per day](#) - Guna Kakulapati, CTO Applift India

¥ [Fast Interactive analytics on 100TB data using Apache Spark](#) - Dipa Dubashi, HEAD - Product Management SIGMOID

Support

In case you have any queries or require further information send a word to support@venturesity.com