Data fields

* id: ad identifier
* click: 0/1 for non-click/click
* hour: format is YYMMDDHH, so 14091123 means 23:00 on Sept. 11, 2014 UTC.
* C1 -- anonymized categorical variable
* banner\_pos
* site\_id
* site\_domain
* site\_category
* app\_id
* app\_domain
* app\_category
* device\_id
* device\_ip
* device\_model
* device\_type
* device\_conn\_type
* C14-C21 -- anonymized categorical variables

1. Exploratory Data Analysis ~ Sampling the data, memory optimization, analysing the click through distribution with the help of data visualizations and various other Key Performance Indicators(KPIs)

2. Model development - USing Logistic Regression ~ choosing among L1/L2 regularization techniques to in order to reduce the dimensional space and to keep a check on overfitting of the model

3. Evaluating the results -! Considering various metrics such as Classficiation reports, accuracy scores, confusion matrices and ROC/AUC scores to evaluate the performance of the built model

### CTR analysis ~ Click v/s No click distribution

### Click through rate on a set of 1 million records of click stream data sampled at random from the population of 40 million records is 16.9 ~ 17%. CTR effectively = 17%.

# Banner Position

### Banner positions representing attractive and appealing designs that might highly affect a user's behavior and in turn trigger their decision to click. Or not. Hence making it an effective metric to predict clicks

# Part 2: Developing the Prediction model

### Using the key metrics discussed above as a part of the EDA to put together a predictive model in order to forecast the possibility of a click

### Feature Selection ~ To reduce the dimensional space occupied and to deal with overfitting, use GRID SEARCH cross validation and regularization to obtain trade off b/w number of features and F-1 score

### The model has an 83% accuracy score and 0.5 is the area under the receiver operating characteristic curve. ROCAUC implying the expected position of positives drawn before a uniformly drawn random negative

.