## 

## 

## Problem Statement :

* Predict CVU value for future correctly on basis of present data.
* Scope of the current problem is to successfully incorporate app ownership data in the CVU prediction model.

## Features:

* Site app id
* App id
* Display marketing name
* Os version id
* Appownership feature
* CVU(target feature)

## Constraints:

* Country\_id = 94
* Device\_os\_id = 3 or device\_os\_id = 5
* Device\_type\_id = 1 or device\_type\_id = 3

# Data preparation

1.Request data for 8 days from 1-8 May.

* Request\_date
* Request\_uid
* Display\_marketing\_name
* Os\_version\_id
* Market\_id
* Device\_os\_id
* Country\_id
* Device\_type\_id

2 Download data for 8 days from 1-8 May.

* Request\_date
* Request\_uid
* App\_id

3 Appsbag(app downloaded by the user) at user level from last 30 days of purchase download and custom data.

4 Taking join of download and request data on request\_uid and date level to get the other features(lets call it df).

5.Adding appsbag to both df and request data on request uid level.

6 Creating Appownership feature for download data and groupby on all cuts(except request\_uid) to find the download counts.

7 Apps list it is the list of unique apps in 8 days of data.

8 Replicating the request data for the apps list and than generating feature for each corresponding cuts and than grouping on all cuts(except request\_uid) to get the user count.

9 Finally joining newly generated download data and request data on all cuts to get the final dataframe(CVU = downloa\_counts/user\_count).

## Results:

## Without Appownership feature:

Model used : Logistic Regression

Evaluation : used weighted RMSE

* Train error : 0.38621907 × 10-2
* Test error : 0.385392154 × 10-2

Model used : Bining

Evaluation : used weighted RMSE

* Error : 0.02996773 × 10-2

## With Appownership feature:

Model used : Logistic Regression

Evaluation : used weighted RMSE

* Train error : 0.3862679 × 10-2
* Test error : 0.385459621 × 10-2

Model used : Bining

Evaluation : used weighted RMSE

* Error : 0.03334398× 10-2

## Data preparation for NFFM

1. Take purchase custom and download data for 56 days (18April-12June) and groupby on request uid level to get the appsbag for each user (let’s call it df1)..
2. Rrnob data for 8days (13June-20June) and join it with df1 to get the appsbag for each user.
3. Download data for 8days (13June-20June) and join it with df1 to get the appsbag for each user.
4. Take a join of download and rrnob data on date,request\_uid and appsbag level (let’s call it df2).
5. Groupby df2 on all cuts to get the download count (let’s call it df3).
6. Groupby rrnob data on all cuts to get the user count (let’s call it df4).
7. Join df3 with df4 on all cuts to get the CVU value.