# Normalization/Feature Engineering

Perhaps normalization of *timezone* is a bit tricky.

The range is -720<-->+720 (you can divide by 60 to get the hours -12<-->+12)

*sh\_version* and *client\_version* are multi-dotted strings (1.2.4). Over time these versions will change, so its worth tracking these. To normalize is tricky because the next version after 1.2.4 might be 1.3 or 1.3.0.

So:

* 1.2.4 could be 1\*10^2 + 2\*10^1 + 4\*10^0 = 123
* 1.3.0 = 130

*model* is a finite number of strings. In this case models of iPhones (4/4s/5/6/6a etc)

*os\_version* is similar small set of strings - not too much diversity

# Field Explanation

1 # unique ID for our reference only - can be ignored

row = str(install.installid)

2 # Does the user allow push notifications? TRUE/FALSE **ok**

row = row + "," + str(len(install.access\_data)>0)

3 # Has the user opted in to Ad Targeting TRUE/FALSE **ok**

row = row + "," + str(len(install.advertising\_identifier)>0)

4# How many campaigns did this user match? INTEGER **ok**

row = row + "," + str(len(install.campaignmatch\_set.all()))

5# Name of Telco carrier for this user’s phone STRING **ok**

row = row + "," + str(install.carrier\_name)

6# What version of the Customer App are they on? DOTTED NUMBER **ok**

row = row + "," + str(install.client\_version)

7# When did they start using the App DATETIME **ok**

row = row + "," + str(install.created)

8# Do they allow iBeacons TRUE/FALSE **ok**

row = row + "," + str(install.feature\_ibeacons)

9# Do they allow Geolocation TRUE/FALSE **ok**

row = row + "," + str(install.feature\_locations)

10# Do they allow push (different to access\_data) TRUE/FALSE **ok**

row = row + "," + str(install.feature\_push)

11# Screen Height INTEGER **ok**

row = row + "," + str(install.height)

12# Do they allow Bluetooth? TRUE/FALSE **ok**

row = row + "," + str(install.ibeacons)

13# Do they have a unique GLOBAL identifier? TRUE/FALSE **ok**

row = row + "," + str(len(install.identifier\_for\_vendor)>0)

14# How many crashes did the user have? INTEGER **ok**

row = row + "," + str(len(install.installcrash\_set.all()))

15# Not sure this is useful. INTEGER **ok (0 or 1)**

row = row + "," + str(len(install.installdailyusage\_set.all()))

16# IP Address. Probably useless. DOTTED NUMBER **ok**

row = row + "," + str(install.ipaddress)

17# Date of Last Update by SDK (still operating) DATETIME **ok**

row = row + "," + str(install.last\_seen)

18,19# not included

# row = row + "," + str(install.latitude)

# row = row + "," + str(install.longitude)

18# Probably useless. We should be ignoring rows that <>’prod’ as they are developer tests TRUE/FALSE **ok**

row = row + "," + str(install.mode=='prod')

19# User’s Device Model STRING eg iphone **ok**

row = row + "," + str(install.model)

**(skip att#20 for rows that have more than 36 attributes)**

20**# User has blocked push after initially accepting TRUE/FALSE (not T/F in data, it's date time)**

row = row + "," + str(install.negative\_feedback)

**21# Device Operating system STRING (ios only?) ok (ios only)**

row = row + "," + str(install.operating\_system)

22# Device Operating system version DOTTED NUMBER **ok**

row = row + "," + str(install.os\_version)

23# Push Token revoked TRUE/FALSE **(Datetime)**

row = row + "," + str(install.revoked)

24# User has signed in TRUE/FALSE **ok**

row = row + "," + str(len(install.sh\_cuid)>0)

25# Ignore TRUE/FALSE **ok**

row = row + "," + str(install.smart)

26# Did the user remove this App TRUE/FALSE **ok (datetime)**

row = row + "," + str(install.uninstalled)

27 Screen Width INTEGER **ok**

row = row + "," + str(install.width)

email\_exists = str(len(install.installtag\_set.filter(key='sh\_email')[0].string)>0)

28 did the user give us their email TRUE/FALSE **ok**

row = row + "," + email\_exists

if email\_exists:

freemails = ['gmail.com', 'hotmail.com', 'live.com', 'yahoo.com','yahoo.co.uk']

29 # is the email address a “free” email address TRUE/FALSE **ok**

row = row + "," + str(install.tags.get('sh\_email').split('@')[1] in freemails)

else:

row = row + ",FALSE"

30 age in days of App launch from App installation INTEGER **ok**

row = row + "," + str((install.installtag\_set.filter(key='sh\_app\_launch')[0].datetime - install.installtag\_set.filter(key='sh\_first\_run')[0].datetime).days)

31# How many push has this user received INTEGER **ok**

row = row + "," + str(install.tags.get('sh\_push\_count'))

32# what timezone is the user in INTEGER **ok**

row = row + "," + str(install.tags.get('sh\_utc\_offset'))

33# what SDK version - probably irrelevant DOTTED NUMBER **ok**

row = row + "," + install.tags.get('sh\_version')

34# is the user a “free” or “paid” user STRING **ok**

row = row + "," + install.tags.get('user\_type',"Free")

35# How many question did the user answer? INTEGER **ok**

row = row + "," + str(int(install.tags.get('questions',0)))

36# How many question did the user get correct? INTEGER **ok**

row = row + "," + str(int(install.tags.get('questions\_correct',0)))

36 max