# 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

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

row = str(install.installid)

**AllowPush**

1# Does the user allow push notifications? TRUE/FALSE

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

**AdOptedIn**

2 # Has the user opted in to Ad Targeting TRUE/FALSE

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

**NumCampaignMatch**

3# How many campaigns did this user match? INTEGER

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

**Carrier**

4# Name of Telco carrier for this user’s phone STRING

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

**AppVersion**

5# What version of the Customer App are they on? DOTTED NUMBER

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

**StartDate**

6# When did they start using the App DATETIME

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

**AllowiBeacon**

7# Do they allow iBeacons TRUE/FALSE

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

**AllowGeo**

8# Do they allow Geolocation TRUE/FALSE

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

**AllowFeaturePush**

9# Do they allow push (different to access\_data) TRUE/FALSE

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

**ScreenHeight**

10# Screen Height INTEGER

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

**AllowBT**

11# Do they allow Bluetooth? TRUE/FALSE

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

**HaveUniqueGlobalID**

12# Do they have a unique GLOBAL identifier? TRUE/FALSE

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

**NumCrash**

13# How many crashes did the user have? INTEGER

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

**DailyUsage**

14# Not sure this is useful. INTEGER

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

**IP**

15# IP Address. Probably useless. DOTTED NUMBER

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

**LastUpdateDate**

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

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

# not included

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

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

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

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

**DeviceModel**

18# User’s Device Model STRING

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

**BlockPush**

19# User has blocked push after initially accepting TRUE/FALSE (DATETIME)

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

**OS**

20# Device Operating system STRING

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

**OSVersion**

21# Device Operating system version DOTTED NUMBER

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

**RevokePush**

22# Push Token revoked TRUE/FALSE (DATETIME)

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

**SignIn**

23# User has signed in TRUE/FALSE

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

24# Ignore TRUE/FALSE

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

**Uninstalled**

25# Did the user remove this App TRUE/FALSE

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

**ScreenWidth**

26# Screen Width INTEGER

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

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

**EmailExist**

27# did the user give us their email TRUE/FALSE

row = row + "," + email\_exists

if email\_exists:

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

**EmailAddress**

28# is the email address a “free” email address TRUE/FALSE

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

else:

row = row + ",FALSE"

**InstallDays**

29# age in days of App launch from App installation INTEGER

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

**PushCount**

30# How many push has this user received INTEGER

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

**Timezone**

31# what timezone is the user in INTEGER

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

**SDK**

32# what SDK version - probably irrelevant DOTTED NUMBER

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

**UserType**

33# is the user a “free” or “paid” user STRING

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

**Questions**

34# How many question did the user answer? INTEGER

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

**CorrectQuestion**

35# How many question did the user get correct? INTEGER

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