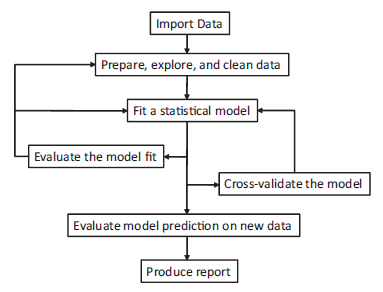
**Predicting Usage for a TV Remote Mobile App**

* Given usage data from a mobile app, predict which users will use the app for the most number of days.
* Identify which factors predict usage level of a user.
* Given first two weeks of usage, predict the total number of sessions for the user.



**Step 1 – Import Data:**

*1412725 obs. of 7 variables:*

*$ ID : int 22129605 22129605 22129605 22129605 22129605 22129605 22129605 22129605*

*$ Age : int 24 24 24 24 24 24 24 24 24 24 ...*

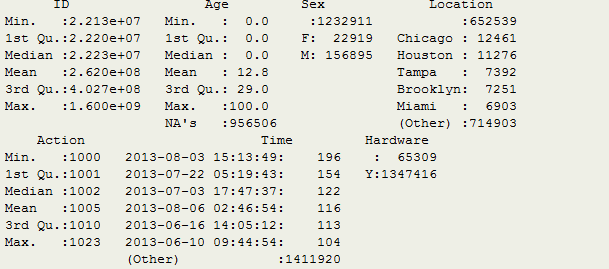
*$ Sex : Factor w/ 3 levels "","F","M": 2 2 2 2 2 2 2 2 2 2 ...*

*$ Location: Factor w/ 1950 levels "","Abbeville",..: 1 1 1 1 1 1 1 1 1 1 ...*

*$ Action : int 1000 1000 1000 1001 1001 1002 1002 1010 1010 1010 ...*

*$ Time : Factor w/ 427237 levels "2013-06-01 03:01:20",..: 1945 2920 3238 2104 3099 1960 2992…*

*$ Hardware: Factor w/ 2 levels "","Y": 2 2 2 2 2 2 2 2 2 2 ...*

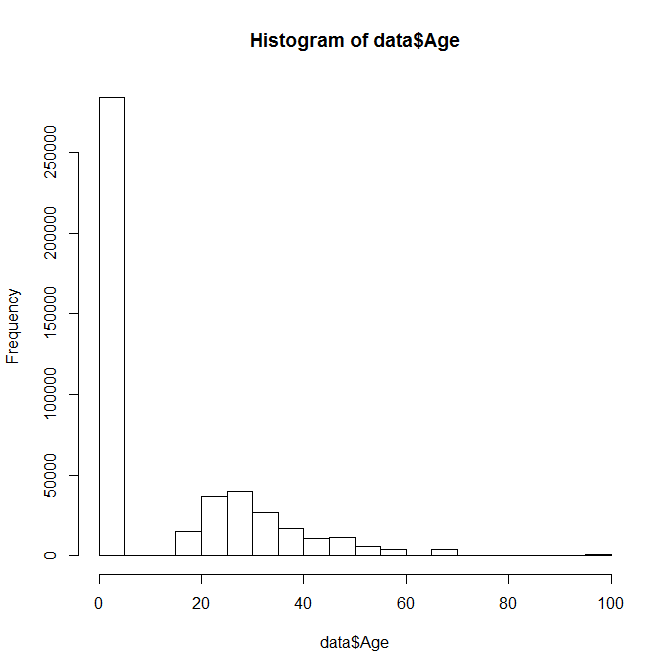


**Step 2 – Data Preparation and Exploration:**

*How to deal with sparse information?*

Age data, 2/3 of age data is blank. ½ of location data is blank. Almost all gender data is missing.

For example, here is the distribution of the age data, zero indicating missing value:



**Visualization and Results:**

Distributions of usage characteristics and user demographics

Trend charts

Identified key metrics from the data

Prediction of usage by personal characteristic

Forecasting of adoption using a network model

7. Discuss briefly why the project is important in its business setting.

First, the demographic characteristics (age, sex, and location) of users can help identify target personas for this app. Second, adoption and usage modeling of a new offering is of interest to many stakeholders from development to capital. Additionally, heavy users can be identified and network modeling can not only augment the adoption forecast but also identify additional marketing targets.