* 16, 259 noise 0
* 1, 639 Real Pulsars 1
* Each row lists the variables first, and the label is the final entry

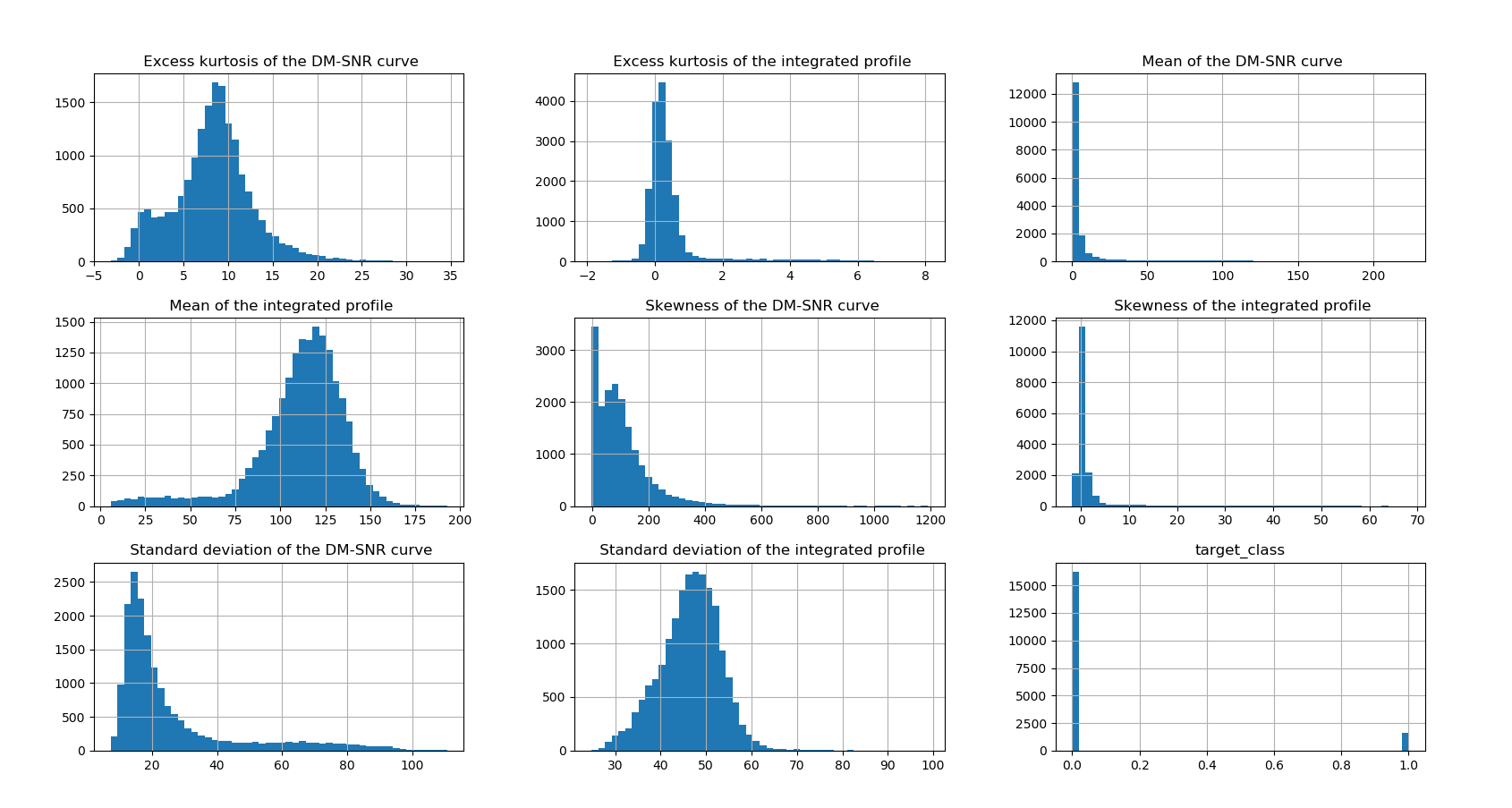
8 continuous variables and 1 class variable

1. Mean of the integrated profile.
2. Standard deviation of the integrated profile.
3. Excess kurtosis of the integrated profile.
4. Skewness of the integrated profile.
5. Mean of the DM-SNR curve.
6. Standard deviation of the DM-SNR curve.
7. Excess kurtosis of the DM-SNR curve.
8. Skewness of the DM-SNR curve.
9. Class

HTRU 2 Summary 17,898 total examples. 1,639 positive examples. 16,259 negative examples.

ML Checklist

* Framing the Problem
  + We want to be able to identify true pulsars, which are hidden among many noisy cases.
  + Performance: how accurate our model will identify pulsars.
  + Current solutions using LR or other forms on the Kaggle kernels.
  + Form of supervised learning due to labeled data.
    - This is univariate (one output) supervised learning.
  + Local project so no need for batch learning and maintenance.
* Measure of Performance
  + Root Mean Square Error to measure performance
* Algorithm Choices: Random Forest, Support Vector Machines, ?

Original Data as Histogram