

CSE 519 -- Data Science
Quant Shop Video Review Form

Name Charuta Pethe
ID Number 111424850

Email cpethe@cs.stonybrook.edu

Video episode title: Winning the Baby Pool **Episode number (1-8)** 3

Of the eight episodes, how many have you previously watched? (0-7) 2

Give the month and year you started your graduate studies at Stony Brook August 2017

Did you learn anything by watching it? (Y/N) Y

Describe stuff you did learn below (if Y):

1. Attributes which are not obvious parameters that easily come to mind can also turn out to be important factors affecting the outcome of the prediction model. All possible data should be considered carefully.
2. Rounding off data values unnecessarily can lead to loss in resolution and reduce the usability of the available data. Even if the resolution of the original dataset is too high, it should be reduced only to a reasonable extent in order to obtain good predictions.

What did you learn from looking at their project report?

All available data should not be used as it is. When it is evident that some parameters in the available dataset will have no impact on the outcome, they should be discarded (in this case, variables such as hospital code and attendant).

Rate program quality from 1 (terrible) to 10 (amazing): 9 **Was it worth watching (Y/N)?** Y

(This is the first report I have seen so far that describes data preprocessing and the split between test data and training data.)

What could we do to make the videos better, and more educational?

Including the answer to this question in the video would be more helpful: What are the different ways of measuring error distribution and how did the team select root mean square?