CSE 519 -- Data Science
Quant Shop Video Review Form

Name Charuta Pethe **Email** cpethe@cs.stonybrook.edu
ID Number 111424850

Video episode title: White Christmas **Episode number (1-8)** 5

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

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. It is important to clearly define the problem statement before trying to find ways to solve it - otherwise, the problem becomes overwhelming.
    a. If the meaning of “White Christmas” is uncertain, gathering a large amount of data is of no use, as it will merely add to the confusion.
    b. Specifying the target cities early on gave direction to the team’s work.
2. One outcome can have an effect on another outcome, i.e. the results of different instances of the experiment can be correlated.
    a. I was familiar with the concept of correlation between independent variables and the outcome.
    b. However, I learned that correlation between the outcomes in 2 different cities also needs to be taken into account while building a model, as the probability of a White Christmas in one city can affect the probability in another city.
3. Proper visualization of data yields better insight.
    a. When the available data is plotted in a way most suitable to the context, it becomes more meaningful, and it becomes easier to understand its impact.

What did you learn from looking at their project report?

1. It is important to have an idea of exactly which/how much data is available, so that we can further decide which parts of the data can be used to build the model.
2. By understanding the correlation between different variables and the outcome, it becomes clear which variable affects the outcome the most.
3. Starting with a simple model (in this case, initially, the team simply took an average of the values over past years) and then considering more factors allows us to make the model increasingly more sophisticated and accurate.

Rate program quality from 1 (terrible) to 10 (amazing): 7 **Was it worth watching (Y/N)?** Y

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

This video comes across as a black box, where data is going in and predictions are coming out, but how it's being done is not clear. The actual process can be understood only after reading the report in detail.
It would help if more details about models and their actual implementation were included in the video.