General information about the Midterm:
- it's a mixture of interpretation / conceptual questions and computations

Things you should bring:
- the textbook (the Midterm is open book)
- 1-2 pages of your own notes
- Your journal
- a calculator (there will only be a few calculations on the Midterm)

Things you should review:
- homework problems and answers from Chapters 2 through 6
- things we covered in the in-class slides
- textbook Chapters 2 through 6 (to refresh your memory)

Areas you should know about or things you should know how to do:
- difference between data mining and statistics
- differentiate among different machine learning approaches
- how to compute various distance measures, & the best applications for them
- how to compute mean and variance, and various normalization measures
- identify the type of a variable (nominal, ordinal, interval, ratio)
- equal frequency and equal width bins
- draw a histogram
- how to compute the number of cuboids in a data cube
- appropriate uses of different cubing computation methods
- how to compute the number of aggregate and closed cells in a cube
  - (review Chapter 5, Exercise 1)
  - Also, with levels in a concept hierarchy
- how to compute measures like chi-squared, lift, and all-confidence
  - compute the expected frequency for chi-square
- how to compute support and confidence
- when to use different sampling methods
- when to use Multiway, BUC, StarCubing, and shell-fragment approaches
- how to draw a simple star or snowflake schema
- perform basic OLAP operations for a query
- what is null invariance and its importance