**Homework 01 Questions Only**

**Due when the dropbox closes (see MyCourses)**

**Prof. Thomas Kinsman**

1. Considering the namespace colors, are all 12 namespace colors represented in the responses for favorite color? (The colors are: black, gray, white, red, green, blue, yellow, magenta, cyan (i.e. light blue), orange, brown, and pink.)   
   Are there any missing?
2. Build a 0R Rule for guessing someone’s favorite namespace color. What is the 0R Rule?
3. Using your 0R Rule, guess the Professor’s favorite color.
4. Considering the question of left-handed – search online for the correct proportion of the population that is left vs. right-handed. Do you think that the results of our quiz reflect valid data?
5. Considering the Star Wars Question, what kind of data cleaning should be done before testing to see if the user gave the correct answer? Should the machine test for case sensitive responses? Is there a pattern in the lower case responses?
6. Considering the Winnie-the-Pooh question, there were bad responses. What processing would automatically help the analysis? Use your Computer Science skills here. How do you match a pattern? What pattern might you use? (CS students, please help students from other disciplines that ask for help.)
7. Prof. Kinsman invented a new Boolean attribute. This attribute is set to true (or 1) if the person was involved in any team sports OR group activities (such as orchestra). What is the word we use for an attribute that is derived from other attributes?
8. Considering the first question about the hours of sleep one gets, could this be taken the wrong way?   
   The quiz was only taken once, and the person taking the quiz did not know what was coming,   
   do you think the answers given correctly represented the ultimate intent of the question?
9. Considering all of the sleep questions, do you think the order that the questions are asked matters?
10. Considering all of the questions about sleep, do you see any evidence of a mixture model?
11. I suspect that some students literally raced through the quiz. Such students are more likely to select the first choice of each option, to make spelling or typing mistakes, or to create incorrect data.   
    Can you find any evidence of racing? What evidence? Who are the suspects?
12. Do any attributes require deduplication? How would you propose this be performed?
13. Does the answer to the question about the person who invented the first compiler help you differentiate between students at all?
14. Some surveys *require* the subject to provide a phone number. Sometimes 867-5309 comes up a lot.   
    What message are these subjects trying to convey? How should such a survey question be changed?
15. Considering the processing path that the data went through, why would someone report that they get “July 8th” worth of sleep a night? How might this have happened? (Honest, this happens.)
16. Of students that expect to get an A, what is the most common day that they want homework due?
17. Of students that expect to get a B, what is the most common day that they want homework due?

Sort the data by favorite kind of cookie:

1. For students whose favorite cookie is chocolate chip, what grade are they most likely to expect?
2. For students whose favorite cookie is peanut butter, what grade are they most likely to expect?
3. For students whose favorite cookie is macadamia nut, what grade are they most likely to expect?
4. The survey lets the user refuse to answer the question about their expected grade.   
   This generates missing data. Is this better or worse than the alternatives? Why do you think that?
5. For students that refused to answer the question about their expected grade, could we guess the grade they expected to get using statistical analysis? What would you do to guess their expected grades?
6. Sort the data by the amount of time taken on the quiz. Do you notice any patterns for respondents that took less than 0 to 14 minutes, compared to those that took 15 to 30 minutes?
7. Do you notice any other patterns in the data, as presented?