CS 523 – Database Topics
Mini Project 3
Due Date: December 5, 2015

Background:

   For the third mini project this semester you are to perform some data mining activities on your data from the previous mini projects. You can choose between a) researching and implementing one data mining run on your data, or b) finding a data mining package and performing several data mining runs on your data. As always, there are two parts: 1) a written paper describing your work, and 2) a presentation to the class on your findings.

Paper Requirements

1. **Introduction**. *This is repeated from Mini Project 1*
2. **In Depth Description** *This is repeated from Mini Project 1*
3. **Preprocessing.** Describe any cleanup activities that were done (*some of this might have been done in Mini Project* 1).
4. **STAR Schema Design**. *This is repeated from Mini Project 2.* Describe what your fact table is and what dimensions exist. Also, list any rollups you may have. Describe how you converted your raw project data into this.
5. **OLAP Queries.** *This is repeated from Mini Project 2.* Develop at least 5 different and distinct OLAP queries to perform on your data. For each one, describe what it is to do and why its results may be important.
6. **OLAP Implementation**. *This is repeated from Mini Project 2.* Show the commands used to implement your STAR schema as well as your data cleansing activities (show screenshots from a running database). Also, supply the SQL code used to run the OLAP queries and at least a sample of each output.
7. **OLAP Conclusions**. *This is repeated from Mini Project 2.* Describe what conclusions may be drawn from the actual results of running your SQL OLAP queries.
8. **Data Mining Introduction**. Describe what you have decided to do and why.
9. **Data Mining In Depth Description** (could be multiple sections).
   a. If you choose to implement your own algorithm, describe what you have found, how it works, what input parameters are required, and what the output should look like.
   b. If you choose to find an existing package, describe what it is, what it does, what algorithms are available, etc.
10. **Data Mining Implementation**.
    a. Describe your code in detail; describe your input dataset; show your results.
    b. Describe the interface to the program; present which algorithms you have chosen to use; describe the input datasets; show your results.
11. **Data Mining Conclusions**. Describe what conclusions may be drawn from the actual results of running your data mining algorithm(s).
12. **Bibliography**. Add to your bibliography all books, papers, and websites you have used to research the algorithm/program you have chosen to use.

Presentation Requirements

1. I would suggest outlining your presentation following the outline of your paper. Use a large enough font so that all results are readable.
2. The implementation may be presented via screenshots of the program(s) running.
3. I suggest using PowerPoint to put together the presentation.
4. Aim for a 10 minute presentation.

NOTE 1: I found a copy of the slides very useful, so please include these along with the written report when you turn in your documentation for this project.

NOTE 2: When taking screenshots, invert the colors so that the background is white and the letters are black (white letters on a black background is hard on printer toner cartridges).