Oscar Solorzano
Kyle Matyac
CSc 177

# Progress Report

## Learning Experience so Far

We have learned how to write a CSV file reader that inserts the values obtained into a MySQL relational database. We have also learned some of the process of outputting a MySQL resultset to a web frontend and how we will be using that data to provide something meaningful to whoever reads it.

## Data Mart for Electricity Rates Based on Location, Company, and Zone

The purpose of the data mart project is to take the data we've obtained as a csv and insert it into a MySQL database so that we can query it based on location, company, or zone and then output meaningful results. The current plan is to output this data onto tables or bar graphs for easy visualization. For location data, the current plan is to color in a google map that depicts location vs. electricity price rates. This would then be done again for each zone type.

The data mart has been completed and filled according to a star schema that separates utility data from location data. A web based user query tool is being developed that will be located on our team website. This functionality is no different from our initial proposal.

## Analysis of Electricity Rates Based on Location, Company, and Zone

Most of the data mining portion of our project will try to answer questions regarding how location and zones impact electricity price rates. We anticipate higher rates in places with higher populations and lower rates in less developed areas. Hopefully with our visualization, we will be able to pinpoint these locations and any severe outliers. We will most likely be using a clustering algorithm to look for outliers in the spatial data. Association rules will be created for outliers when looking at companies. This is all still within the scope of the original proposal.

For the data mining portion of the project, we haven't decided on what algorithm to use specifically. From class discussion, a version of K-Means seems to make the most sense to differentiate clusters and find outliers. We still need to run our data through a mining tool such as Weka to have a more meaningful explanation of the algorithms we will use.

## Schedule

April 13th: Complete query tool
April 20th: Finish google maps "heat map" for electricity rates.
April 27th: Complete Data mining analysis
May 11th: Finalize project report documents and PPT presentation