# MIS 344 Term Project

The final project selection starts in week four with lab 4 and will require independent work throughout the rest of the term (See BB for due date). The deliverables for the project include a series of reports as well as forecasting and predictive analysis models to support future trends. To showcase your work you will need to present your analysis in a Power Point presentation. This presentation should include visuals of graphs, charts, reports, analysis on your findings and references on where you got the data and supporting documentation.

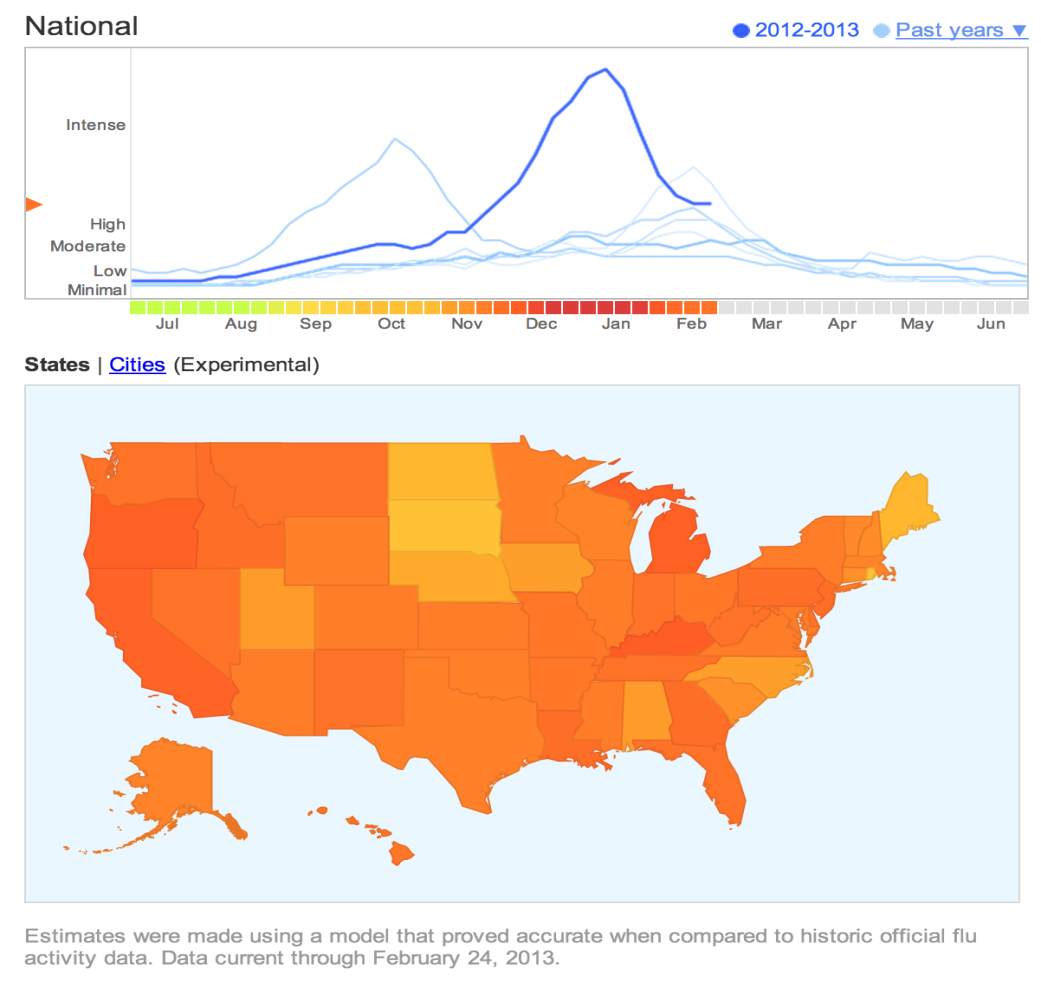
Below are a couple of scenarios that you can use as an example to see the types of projects that would be acceptable. You are welcome to use one of these scenarios as your term project.

Project Components:

1. Student captured a data set and moved the data set into SQL Server. This will require some research on the web for datasets that will help you with your analysis.
2. Student used Crystal to develop reporting solutions
3. Student demonstrated understanding of reported information and can articulate in a professional manner.
4. Student successfully created predictive modeling for their given scenario using regression forecasting or another algorithm
5. Student ran decision trees, neural nets, neighboring algorithms or other data mining algorithms to support their findings.
6. Student presented findings in a 15 minute presentation to the class

**Scenario 1** Flu Data

Data download from: <http://www.google.org/flutrends/>



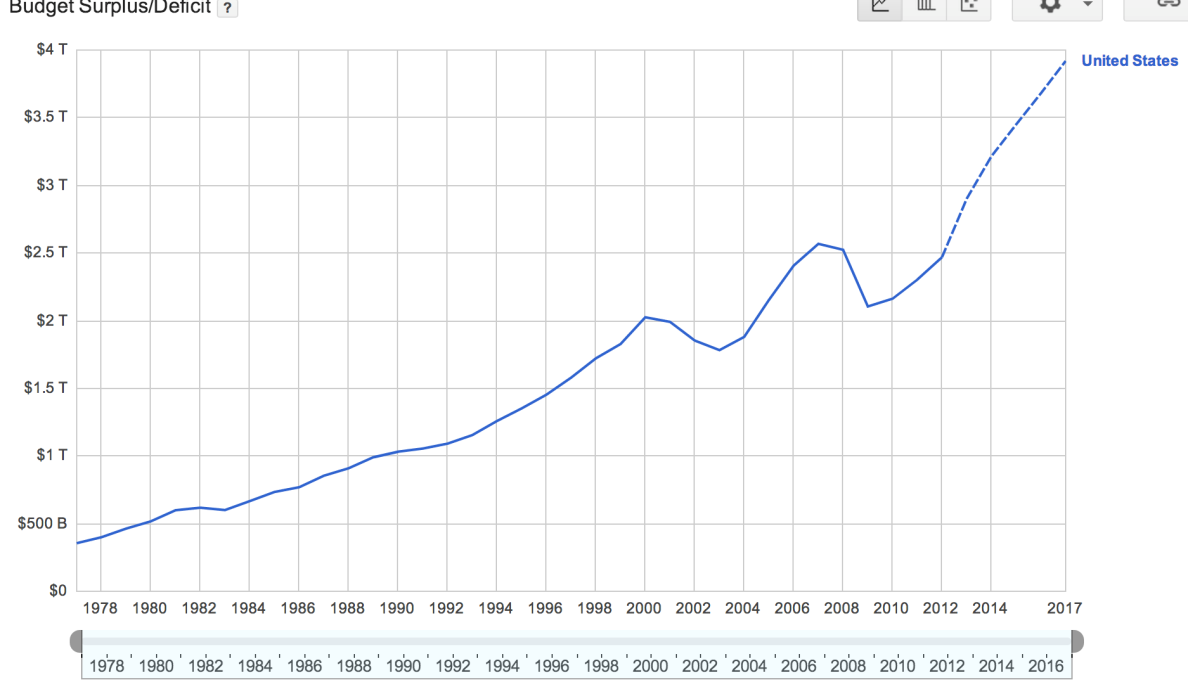
For this scenario the CDC is looking for trends annually by state, what are the typical entry points for the flu virus? Why is that? You will need to do outside research to support your claim, don’t just say because it is a big city and has an airport. References that are acceptable are peer reviewed articles which you can find on Google scholar and the ACM Digital library.  
  
Beyond historical analysis how does the CDC predict what will happen next year. How many flu shots should they produce? What time of year and which cities should receive shots sooner. Are other countries representing a greater risk then the United States? What should we do about them? Think of anything that may be relevant to the flu virus and be ready to defend your findings.

**Scenario 2** Financial Health of the U.S. Government

When will we get out of the recession? Well Google has pulled in data from the office of the president to analyze the current financial health on a number of different metrics. Your task is create a dataset that supports this scenario. You will download the data in excel form from the office of the president and transform that data into a relational model. From there you will actively report on the data and test data mining algorithms to support Google’s forecasting model. You will notice Google predicts the current spending and income of the U.S. economy through 2017. How can Google predict so far? What is it they are doing? Is this something you can replicate? If so, how? If not how would you present this information to management?

Management needs to know this information in order to determine their own financial health, investment strategies and overall strategic imperatives for the firm. The image below predicts the budget deficit to climb over the next four years quite substantially, what does this mean for everyone?

Data can be viewed: <http://www.google.com/publicdata/explore?ds=z6tggkh2adod2s_>



**Scenario 3** Speedtest.net

You work for Nero a very popular Internet provider. Your company has purchased data from speedtest.net a very popular speed testing web site. Your task is to analyze historical data. There are over 1.5billion records in this data set of worldwide connection speeds. This data will need to be handled with SQL server and will likely need to be normalized down to something easily manageable. There will be a little more work on the database architecture side with this project. Learning to apply a data mart to this project may be necessary.

Once data is in a manageable state you can then get to business. Your CEO needs to know where the weak spots are currently in the United States and Europe. Further you need to trend that data back and apply a linear regression model to predict the future drop outs or weak spots.

Nero is thinking of going global and expanding operations to Europe and maybe into the Middle East and Russia. Your task is to put a package of analytical facts together to direct Nero’s executive team in their decision making.

Data can be downloaded at <http://www.netindex.com/source-data/>

Google’s analysis of this data can be viewed at <https://www.google.com/publicdata/directory>

