ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: "Big" Data Analytics
Proposal – Natural Oils

By: Nathan Stefanick

Step 1:
"Identify a problem," well I would have to say that the biggest problem in my life right now is that I am a poor college student, and I need a solid source of revenue to fund my addiction to Jimmy Johns, my local sandwich delivery service. So I decided the best way to increase my funds is to invest, but what do I invest in? I want to make a lot of money, so I decided that my best investment would be in some sort of global commodity. At first I wanted to invest the beverage trade: tea, coffee, hot chocolate, etc. However doing some research into I found a much more interesting commodity, oils. There is a huge demand for oil that I never really thought about, of course there is oil in the food production industry, almost everything we eat is cooked in oil or uses some sort of oil as a non-stick agent. However natural oils also go into shampoos, conditioners and many more beauty products. Oils are also used in the health industry to manufacture sunscreen, anti-acne cream, deodorant, and toothpaste. So how do I make money off all these uses for oil? The basis of my project is to use the most basic principal of investment: "buy low, sell high." Using the data that I find, I will first create confidence intervals of the prices of the different types of oil in order to determine which oils have the most variability in prices. I will determine which oils are overall the most variable and high risk for investment. After that I will make multiple linear regression models. By that I mean I will make n-by-n graphs for the different kinds of oil that I collect and see if I can spot any trends in the data. After that I will use a random generator to pick out a random years to see if there is seasonal correlations in prices of oil.

Step 2:
In my search for data I came across a spreadsheet for the United Nations Conference on Trade and Development. This spreadsheet contains a very large amount of historical information. It has the prices of eight different kinds of oil produced in various places around the world. These are the free market commodity prices as determined by the UN. This data goes back to January 1960 and goes until August 2015; the data is recorded on a monthly basis. It luckily can be downloaded in either Excel or comma delimited form, which makes it easy to translate into different programs. I plan on using SPSS, a statistical analysis program that I used in a stat class previously. Using this program I have many statistical tools at my disposal. Much of what I can and will do is with linear regression and analysis of covariance.

Step 3:
What I hope to find by the end of this project is correlations between different types of oils. I would like to determine if the price of oil A increases, then the price of oil B will decrease. This herein would translate to real time buying and selling of stock in natural oil companies. If I were to see a spike in oil A's prices then I would know to sell stock in oil B because that would indicate oil B's prices were likely to decrease, so I want to cash out now. I will also make a guess, that is a very reasonable scenario. It is likely that industry is based off the crop supply for different oils. If a soybean crop was wiped out one year by a hurricane or other natural disaster, the supply will decrease exponentially, causing the price to skyrocket. This may cause customers to search for alternatives, thus increasing demand and changing the prices of other oils. I hope my result will help me determine when to buy and sell stock in a given company and ultimately make me not poor!