**Homework 1: Linear Regression**

**Stock Market Prediction from Economic Indicators**

**Goal:**

The goal of homework 1 was to use linear regression to predict stock market values using several economic indicators.

**Summary:**

*Data:*

There are five (5) predictor variables:

1. Producer Price Index -Rail Transportation
2. Price of Gold
3. Price of Oil
4. Consumer Price Index
5. Consumer Confidence Index

The data came from several sources and needed preprocessing to facilitate use in the linear regression model. Preprocessing included dropping unneeded data, formatting titles, merging, and validation.

*Model***:**

Fitting the model was a process of dividing the data into a train and a test set and running the ordinary linear regression model. During the training process, using backwards selection, I found that two of the variables (Price of Gold and Producer Price Index) were not significant. Running the residuals plot on the remaining three variables revealed a non-random pattern to the error terms, which suggests that I need to transform the data. I also investigated several potential sources of error and found the following.

Data Linearity: The cosine-like pattern in the residual plot suggests the data is not linear.

Error Constancy: There was no obvious evidence of non-constant error. For example the data did not contain evidence of heteroskedasticity.

Outliers: There were some outliers in the data that need to be dropped to improve the models accuracy.

High Leverage Points: There were no high leverage points in the data

Non-Correlation of Error Terms: There was no evidence of non-correlation among the error terms

Multicollinearity: There was no evidence of multi collinearity. The Variance Inflation Factors for the predictors was near

**Lessons Learned:**

1. I spent too much time on the data: 2-3 hours searching for data and then 4-5 hours processing. I ran out of time. It appears I can use someone elses data from their project and just not look at their work. I will do that next time.
2. I forgot about the rubric and did not read it until the end. That will happen first next time.
3. I am excited for the next project. My focus will be better placed and my results will reflect that.