**Project Design Writeup**

**Project Problem and Hypothesis**

* What's the project about? Predicting city-cycle fuel consumption in miles per gallon.
* Where does this seem to reside as a machine learning problem? Linear Regression.
* What kind of impact do you think it could have? By predicting miles per gallon, we can better inform users on cars they are interested in.
* What do you think will have the most impact in predicting the value you are interested in solving for? It looks like weight, horsepower and acceleration might have a direct impact on mpg.

**Datasets**

* Description of data set available, at the field level (see table)
* 1. mpg: continuous   
  2. cylinders: multi-valued discrete   
  3. displacement: continuous   
  4. horsepower: continuous   
  5. weight: continuous   
  6. acceleration: continuous   
  7. model year: multi-valued discrete   
  8. origin: multi-valued discrete   
  9. car name: string (unique for each instance)

**Domain knowledge**

* What experience do you already have around this area? Work for Autotrader.com
* Does it relate or help inform the project in any way? It does not help but the outcome can be used for researching cars.
* What other research efforts exist?
  + Use a quick Google search to see what approaches others have made, or talk with your colleagues if it is work related about previous attempts at similar problems.

**Project Concerns**

* What are the assumptions and caveats to the problem? Weight, horsepower, acceleration and perhaps other variables have a direct impact on miles per gallon.
  + What is already implied about the observations in your data set? That some of the variables will have an impact on miles per gallon.
* What are the risks to the project? Handling incomplete date and outliers.
  + What's the cost of your model being wrong? The prediction could be off and users research can be impacted.
  + Is any of the data incorrect? Potential outliers can be present and will be identified through exploratory data analysis. Missing data is present and potential solutions are being researched.

**Outcomes**

* What do you expect the output to look like? Prediction of miles per gallon using a linear regression model based on weight, horsepower and maybe acceleration.
* What does your target audience expect the output to look like? Linear regression that predicts miles per gallon to a certain degree of confidence based on a couple of features.
* What gain do you expect from your most important feature on its own? To have a direct impact on miles per gallon.
* How complicated does your model have to be? Not complicated at all. It should be a simple linear regression model.
* How successful does your project have to be in order to be considered a "success"? As long as the prediction is higher than ~70%, the project would be considered a “success”.
* What will you do if the project is a bust? Will try and form the project into a classification model.