Machine Learning Summary

1. **How do you frame your main question as a machine learning problem? Is it a supervised or unsupervised problem? If it is supervised, is it a regression or a classification?**

   My problem is a supervised classification problem.

   Given time and location of crime, I will see if I can predict:

   - *whether a crime is likely to be violent or non-violent.*

A model of this problem could be used to provide decision support to police departments. As a resident and homeowner in the Phoenix area, I personally would want to know this information for the safety of myself and my child. This type of prediction could be used in a program that assesses a community's crime rating to provide decision support to home buyers, and could be listed on websites used to aid in this process, such as Zillow.

2. **What are the main features (also called independent variables or predictors) that you'll use?**

   Predictors:

   - *Time of Day (Hour)*
   - *Part of Day (Morning, Afternoon, Evening, Night)*
   - *Season (Winter, Spring, Summer, Fall)*
   - *Month*
   - *Zip-code/ Median Property Value (correlated)*
   - *Extras: Weather data (temperature (F) & weather condition)*

3. **Which machine learning technique will you use?**

   I will be using classification trees.

4. **How will you evaluate the success of your machine learning technique? What metric will you use?**

   I will evaluate the success of my technique using a confusion matrix and K-fold cross validation.