1. Clearly define a problem or an idea of your choice. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem:

The problem I would like to tackle is police and other first responders not having the resources to respond quickly to accidents. Studies have shown that a major factor in determining whether a victim of a car accident will live or die is how quickly they receive medical attention. If first responders were able to predict, based on weather and road conditions, when serious accidents are more likely to occur, they could increase the number of people working that day in order to respond to accidents closer to when they happen. My audience is first responders, specifically police officers and EMT’s in charge of scheduling shifts and determining how many people are working on a given day. They should care about this problem because there is a correlation between speed of responding to an accident and the mortality rate of a victim of the accident.

1. Describe the data that you will be using to solve the problem or execute your idea. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using.

The data I am using is a set of accidents that have occurred in Seattle city. The label for the data is “severity,” which in this case is used to describe the fatality of an accident. This data consists of both numerical and categorical data. With the categorical data, such as the weather condition at the time of the crash, I will have to utilize one hot encoding in order to analyze it. Because I am looking at this data in terms of predicting what weather conditions would require the largest number of first responders to be working in order to respond to accidents fastest, the data in this database that interests me most is weather and road condition. I can compare the number and severity of accidents in good weather conditions vs. bad ones. I am also interested in seeing how weather effects the tendency of drivers to engage in poor behavior. For example, are drivers who are speeding more likely to be involved in an accident in good weather or bad weather. I would also want to look at what type of accidents are the most common in different weather conditions, and how those types of accidents correlate with accident severity.