1. **A description of the problem and a discussion of the background. (15 marks)**

*(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.)*

Road accidents are one of the major causes of premature deaths. They are many factors that influence the severity of a road accident whenever it occurs, including weather conditions, speed of the vehicle, condition of the road, etc. In order to reduce the severity of a road accident or prevent it altogether, there is a need to understand how the different factors mentioned above affect the severity of a road accident. By understanding the relations between these factors, and developing a model that can predict the severity of a road accident with high accuracy, road users can use these predictions to adjust their driving or change their travels if possible, so that they can reduce the severity of the accidents or avoid it at all. This would greatly reduce the severity of accidents and some accidents will be avoided.

The audience of this project are the policy makers of Seattle and road users in general.

1. **A description of the data and how it will be used to solve the problem. (15 marks)**

*(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 that is used in this project is provided by SPD and recorded by Traffic Records. The data is from the year 2004. The target of the data is the severity of the accident. The data has different features that relate to the severity of the data like speed, number of vehicles involved, number of passengers, etc.

A data exploratory will be done on the data to understand the data and identify relationships. These relationships will be used in the creation of a machine learning model that will predict the severity of data using supervised classification algorithms.

This data will be split into training and testing data to be used to train the model and test for out of case accuracy.