1. Introduction: Discuss
   1. The Business Problem: The focus of this project is the development of a model that can predict the severity of traffic accidents, given weather and road conditions. For this project, severity is one of two categories: either property damage or injury.
   2. Key Stakeholders: The immediate stakeholders of the project are public safety officials hoping to warn drivers of the potential severity of an accident to get them to drive more safely or change their travel plans.
2. Data: Describe
   1. The data that will be used to solve the problem:
      1. The dataset is rich, with more than 150k observations and many attributes. Beyond the weather and road conditions, there are other variables which seem likely to contribute to the severity of an accident such as light conditions, whether the driver was speeding, distracted, or under-the-influence of a substance.
      2. Key to the development of the model is the fact that the severity label indicates the severity of the observed accidents as either fatal or property damage.
      3. Data preparation will require the handling of missing data, as well as re-structuring certain attributes such as weather conditions. As an example, weather exists in the dataset as a single attribute with several categories. This will be re-structured using indicator variables to correlate each separate weather condition with the severity of an accident. The same will be done with other categorical variables. The dataset is also unbalanced with regards to the labeled observation. The dataset will be balanced to remove a source of bias in the predictive model.
   2. The source of the data:
      1. The data to be used for the testing and development of the predictive model will be the shared dataset compiled by the Seattle Department of Transportation, Traffic Management Division, Traffic Records Group.
      2. All collisions provided by SPD and recorded by Traffic Records. Timeframe: 2004 to 05/2020
      3. A more detailed description of the data may be found at: <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>
3. Methodology section: Discuss and Describe
   1. Exploratory data analysis
      1. Data shape
      2. Data Types
      3. Missing Values
      4. Unique Values
      5. Spatial plot of sample by severity
   2. Inferential statistical testing
      1. Pearson correlation (df.corr)
   3. Machine learning algorithms and why
      1. SVM
      2. Decision Tree
4. Results section: Discuss
   1. the results
5. Discussion section: Discuss
   1. Any observations you noted
   2. Any recommendations you can make based on the results.
6. Conclusion section:
   1. conclude the report.