


A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light greenish-blue. They are positioned diagonally, with the blue one partially covering the green one.

# Predicting Accident Severity

Using Linear Regression to Predict Accident  
Severity



# Predicting Severity is good for drivers and EMTs Alike

- Vehicle related accidents are one of the leading causes of death in the United States and Worldwide
- By delivering insights about how bad an accident would be if it were to take place, it can help drivers be safer and can help EMT divisions be more prepared to respond to a large accident
- Not only could a model like this be used to save lives and save EMTs valuable response time



# Data Acquisition, Description, and Cleaning

- The data used to train this model was acquired from the Seattle Police Department and was recorded by Traffic Records
- The dataset provided has 194,674 rows, 38 attributes and contains accident data from the Seattle area from 2004-present. The data is updated every week
- In this dataset, many of the values are string values. This means that in order to perform Linear Regression, these values must be changed to Integers. Additionally, there are over 15,000 NaN values in the dataset. These values had to be substituted with zeros.



# Results

- Ultimately, the results received through this model were underwhelming
  - Since the dataset was originally in string form, I had to convert it into integer values
  - This conversion from string to integer probably was the reason for the disappointing results received from this model
- The dataset would have likely paired a lot better with a Logistic Regression Model rather than a Linear Regression model



# Conclusion and Future Directions

- The model produced did not return optimal results and came up short of expectations
  - The model's future performance could easily have been improved by utilizing Logistic Regression rather than Linear Regression would have suited the dataset a lot better
- Rather than using single attribute regression, multiple regression would be better
  - This would allow for the model to capture the complexity of a vehicle accident and would be a good way to take this model in the future