**Introduction**

Across the US, crashes each year result in thousands of lives lost, injured victims, and billions of dollars in property damage. The National Highway Traffic Safety Administration (NHTSA) uses data from many sources, and one of their goals is to reduce both human and property damage.

Many different factors are provided to the NHTSA, and not all factors are important in determining the severity of accidents. The goal is to identify these important factors and take a proactive approach in minimizing both the cost and the severity of accidents.

**The Data**

* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “[A Countrywide Traffic Accident Dataset](https://arxiv.org/abs/1906.05409).”, 2019.
* Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. ["Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights."](https://arxiv.org/abs/1909.09638) In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.

The data is for the contiguous 49 US States and is continuously being collected starting from February 2016 to June 2020, using several data providers, including two APIs which provide streaming traffic event data. These APIs broadcast traffic events captured by a variety of entities, such as the US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks.

**Business**

Not all information provided to dispatch personnel is relevant to the severity of the accident. The goal is to identify and predict factors that contribute to injuries to people and the severity of the injuries. By identifying these factors, we aim to predict and minimize further serious accidents as well as giving emergency personnel the ability to prepare for the severity of the accident based on the information provided by dispatch personnel using our model.