PRELIMINARY RESULTS

Before discussing our preliminary results, it is crucial to keep in mind the goal of our modeling. Overall, our goal is to evaluate the effectiveness, both in terms of cost and efficacy, of various safety strategies that can be implemented on the stretch of Highway I-40 between Memphis and Nashville. In order to accomplish this, we have transformed a dataset which only included crash severity and crash location into a dataset which accounts for many geospatial features on the stretch of I-40 that goes through Madison and Henderson counties by manually tagging feature categories from Google Maps. Our exploratory data analysis has given us some key insights into the the features which most influence crash severity and have determined some safety strategies we can implement by manipulating these features (for instance, if a crash happened near a guardrail or not). Now, our methodology to evaluate the effectiveness of our safety interventions is as follows. For each strategy, we will conduct a hypothesis test where we choose a segment of the highway to implement the strategy, hold out the crash data from that segment, and train and validate an assortment of models on the remaining data to predict if a crash is injurious or not. Then, we run our models on the crash data for the held out segment, first as-is and then with the features changed according to our safety intervention, and then seeing if the models predict a statistically significant decrease in number of injurious crashes. Once we know the outcomes of the hypothesis tests for each strategy, We will perform a cost-benefit analysis for each strategy and then rank them by overall effectiveness in terms of both cost and predicted decrease in injury rates.

Of course, the strength of this type of methodology depends quite a bit on the strength of our models and their accuracy. Unfortunately, it does look like our models are lacking somewhat in predictive power. However, keeping in mind how unpredictable crash outcomes can be, I think our modeling strength is surprisingly robust. We do not yet know the outcome of our hypothesis testing, but one of our models trained by CatBoost has an f1-score of .57 for predicting non-injurious crashes and an f1-score of .37 for predicting injurious crashes. While in terms of pure predictive power this may not be the best, it may well turn out that it is more than sufficient for our hypothesis testing.

Perhaps our biggest success so far is that we were able to create a fairly interesting dataset just by adding geospatial features to a dataset with only crash severity and the GPS location of the crash. It was quite tedious to label everything by hand, but we managed to find a strategy to accurately tag our data. Thankfully, our project has experienced any significant failures so far. Perhaps we could have a more robust dataset if we had specific safety strategies in mind from the get-go, but none of us are experts in highway safety, so it's reasonable to expect some ambiguity at the beginning of the project.