ECON 481: Final Project Proposal

Group Members:

Donald Yin donaldy@uw.edu
Julien Johnson julienmj@uw.edu
Anas Fathul (muhdanas@uw.edu)

Background

We are interested in finding out which conditions cause more car crashes in Washington state. In particular, our research focuses on features consisting of geographical location, income levels, available funding (if possible to garner this data), weather conditions, and road quality. Among these, we will have subsets of features to create a better predictive model and draw correlations to our questions.

Primary Questions that we want to answer

1. Are there specific areas (county, city, or specific road) that are more prone to collision? Are these characterized by fewer or less severe collisions?
2. What are the key predictors contributing to road collisions in Washington state at various scales (County, City level, weather, etc.)?
3. Can we develop a model to predict high-risk scenarios or high-risk drivers based on historical data and define scenarios that are more likely to involve severe collisions?

Dataset

Our primary dataset, Washington collision is provided by the Washington State Patrol Collision Analysis Tool (CAT). It compromises data submitted by law enforcement officers regarding reported traffic collisions within Washington State. It has a lot of methods to filter out the data (from county level to specific type and severity of accidents) and has data date by date going back several years.

The other data set is provided by the Washington State Department of Transportation, Traffic Counts (AADT) for 2022. This dataset provides the Annual Average Daily Traffic for each state road.

We will import these data sets and clean and adjust them as necessary so that we have the proper inputs for the columns and each county where the crash has also has their median household values.

Methods

We will likely employ a lot of linear regression in our modeling, adding in more or less explanatory variables and seeing whether the additional variables have higher t-values in comparison to when they weren't there. We can regress over the county, time of day, weather conditions, funding of the area, etc.

We also plan on creating a model to predict whether a random scenario is a high risk scenario or more dangerous based on the severity of the car collision from whether there was no injury to potential fatalities.