In my Capstone project, I am examining the relationship between “Fatality” (dependent variable) and how it is influenced by possible predictors (independent variables).

-----------------

**Dependent variable**

The categorical “**Fatality**” variable is defined by two levels non-fatal or fatal, dependent on whether a death occurred from each unique traffic incident.

The variable originates from the “Collision Severity” variable that is defined by 5 ordinal levels:

1. Property Damage
2. Injury (Minor)
3. Injury (Moderate)
4. Injury (Severe)
5. Fatal

With exception to the predefined “Fatal” level, I consolidated all other levels to the “Non-Fatal” level:

1. Non-Fatal
2. Fatal

From the observed data set of 39,021 unique cases:

* 38,810 cases were non-fatal
* 211 cases were fatal

--------------------

**Independent variables**

My independent variables will be responsible for predicting the fatal or non-fatal outcomes of the “Fatality” variable. Each has its own set of levels, listed under each variable.

**Primary Collision Factor or PCF Violation** - the violation cited by the CHP officer to the driver

* Driving or Bicycling Under the Influence
* Impeding Traffic
* Unsafe Speed
* Following Too Closely
* Wrong Side of Road
* Improper Passing
* Unsafe Lane Change
* Improper Turning
* Automobile Right of Way
* Pedestrian Right of Way
* Pedestrian Violation
* Traffic Signals and Signs
* Hazardous Parking
* Lights
* Brakes
* Other Equipment
* Other Hazardous Violation
* Other Than Driver (or Pedestrian)
* Unsafe Starting or Backing
* Other Improper Driving
* Pedestrian or Other Under the Influence
* Fell Asleep

PCF spans numerous categories have no ordinal scale. However, I can identify levels that can be perceived as more influential on a fatality. I predict that variables that involve faster speeds and involvement of pedestrians have a higher likelihood of resulting in a fatality.

**Lighting** - lighting conditions at the time of the incident

* Daylight
* Dusk - Dawn
* Dark - Street Lights
* Dark - No Street Lights
* Dark - Street Lights Not Functioning

I predict that incidents that occur in the night will have a greater role in fatalities as visibility is poorer in darker conditions.

**Road Surface** - road conditions at the time of the incident

* Dry
* Wet
* Snowy or Icy
* Slippery (Muddy, Oily, etc.)

San Francisco is not known for having weather extremes. Rain comes occasionally most of the year as California is still considered in “drought” conditions. My prediction is that most fatalities will be in dry conditions.

**Pedestrian Action** (if applicable) - actions of pedestrians at the time of the incident

* No Pedestrian Involved
* Crossing in Crosswalk at Intersection
* Crossing in Crosswalk Not at Intersection
* Crossing Not in Crosswalk
* In Road
* Including Shoulder
* Not in Road
* Approaching/Leaving School Bus

My prediction for pedestrian involvement on fatality will be levels where pedestrians and/or vehicles are in violation in areas where there is a disparity in speed. The greater the speed of a vehicle against a stationary pedestrian, the deadlier the outcome.

**Intersection** - whether the collision occurred in an intersection

* No
* Yes

Intersections are areas where people should be more aware of their surroundings as they are to follow traffic lights and need to non-verbally communicate with others to show intent. Speeds are usually slower as well. Non-intersections typically have a flow where every person understands their direction and intent in most cases. But because speeds can be typically much higher, miscommunication can lead to fatalities in various scenarios.

**Type of Collision** - the manner in which the the vehicle struck either another car, person, object, etc.

* Head-On
* Sideswipe
* Rear End
* Broadside
* Hit Object
* Overturned
* Vehicle/Pedestrian

To retierate speed disparity, collision types that involve differences in speed of two objects result in a more severe collision, and as a result, fatalities.