**Introduction/Business Problem**

Most major metropolitan areas are subject to multiple and seemingly unpredictable accidents that cause damage to people, automobiles and infrastructure. The severity of these accident’s places strain on city services not to mention the toll on human beings.

A probabilistic machine learning program can help to mitigate the effects that these accidents have on the overall traffic flows of cities. Private businesses, transportation services, and local governments would all benefit from optimizations to traffic patterns based on the predictability of accident severity however city planners would reap the most immediate benefits from this model. Costs in money, man-hours, and secondary accidents would be reduced.

**Data**

The dataset is a collation of accident reports for the Seattle metropolitan area. It is a flat file (csv) that can read into various data science codes but Python will be the analytical language for this model. The dataset contains 194,673 rows with 37 distinct fields. There are more than enough data points to construct training and testing cohorts.

Not all of the fields are as relevant to building the model but several fields may be used: latitude and longitude coordinates for mapping. Addresses and categorical data that lists the accident severities as well as date-time inputs. There are weather conditions as well. Much of the categorical data will have to be transformed for numeric inputs and there are conditions that lend themselves to clustering.

There is a column that can be utilized for Unique ID. The data will have to be tidied prior to analysis and model construction, but the dataset does provide enough relevant information to conduct a supervised machine learning model that can predict accident severity. The exploratory analysis provided an initial evaluation of what fields will be used in constructing the model, further fields may discarded as the analysis evolves:

|  |  |  |
| --- | --- | --- |
| Feature | Data Type | Comment |
| SEVERITYCODE | Int | Target |
| (X)Longitude | Float | Keep |
| (Y)Latitude | Float | Keep |
| OBJECTID | Int | Unique ID |
| INCKEY | Int | Drop |
| COLDETKEY | Int | Drop |
| REPORTNO | Int | Drop |
| STATUS | Text | Drop |
| ADDRTYPE | Text | Keep |
| INTKEY | Int | Evaluate |
| LOCATION | Text | Keep, Evaluate |
| EXCEPTRSNCODE | Text | Drop |
| EXCEPTRSNDESC | Text | Drop |
| SEVERITYDESC | Text | Keep, evaluate |
| COLLISIONTYPE | Text | Keep, evaluate |
| PERSONCOUNT | Int | Keep, evaluate |
| PEDCOUNT | Int | Keep, evaluate |
| PEDCYLCOUNT | Int | Keep, Evaluate |
| VEHCOUNT | Int | Keep, evaluate |
| INCDATE | Text | Keep, transform to Day of the week for analysis |
| INCDTTM | Text | Keep, & Transform to ranges |
| JUNCTIONTYPE | text | Keep, evaluate |
| SDOT\_COLCODE | Int | Drop |
| SDOT\_COLDESC | text | Drop |
| INATTENTIONIND | Text | Keep, evaluate |
| UNDERINFL | Text | Keep, Evaluate |
| WEATHER | Text | Keep |
| ROADCOND | Text | Keep |
| LIGHTCOND | Text | Keep |
| PEDROWNOTGRNT | Text | Keep, Evaluate |
| SDOTCOLNUM | Int | Drop |
| SPEEDING | Text | Keep, evaluate |
| ST\_COLCODE | Int | Drop |
| ST\_COLDESC | Int | Drop |
| SEGLANEKEY | Int | Drop |
| CROSSWALKKEY | Int | Drop |
| HITPARKEDCAR | Text | Drop |