Capstone Project – Car accident severity

Data Description:

The data set provided includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present. The Data set contained 38 features, only 15 of them will be used in our model.

The features that will be used in our classification model are:

1. ADDRTYPE: Collision address type: Alley, Block or Intersection.
2. COLLISIONTYPE: Collision type.
3. PERSONCOUNT: The total number of people involved in the collision
4. PEDCOUNT: The number of pedestrians involved in the collision.
5. PEDCYLCOUNT: The number of bicycles involved in the collision.
6. VEHCOUNT: The number of vehicles involved in the collision.
7. INATTENTIONIND: Whether or not collision was due to inattention. (Y/N)
8. UNDERINFL: Whether or not a driver involved was under the influence of drugs or alcohol.
9. WEATHER: A description of the weather conditions during the time of the collision.
10. ROADCOND: The condition of the road during the collision.
11. LIGHTCOND: The light conditions during the collision.
12. PEDROWNOTGRNT: Whether or not the pedestrian right of way was not granted. (Y/N)
13. SPEEDING: Whether or not speeding was a factor in the collision. (Y/N)
14. HITPARKEDCAR: Whether or not the collision involved hitting a parked car. (Y/N)
15. SEVERITYCODE: A code that corresponds to the severity of the collision: (3—fatality, 2b—serious injury, 2—injury, 1—prop damage, 0—unknown)

Some columns will need some pre-processing like replacing missing values and grouping relevant values together in one group, which will be discussed in the next section.