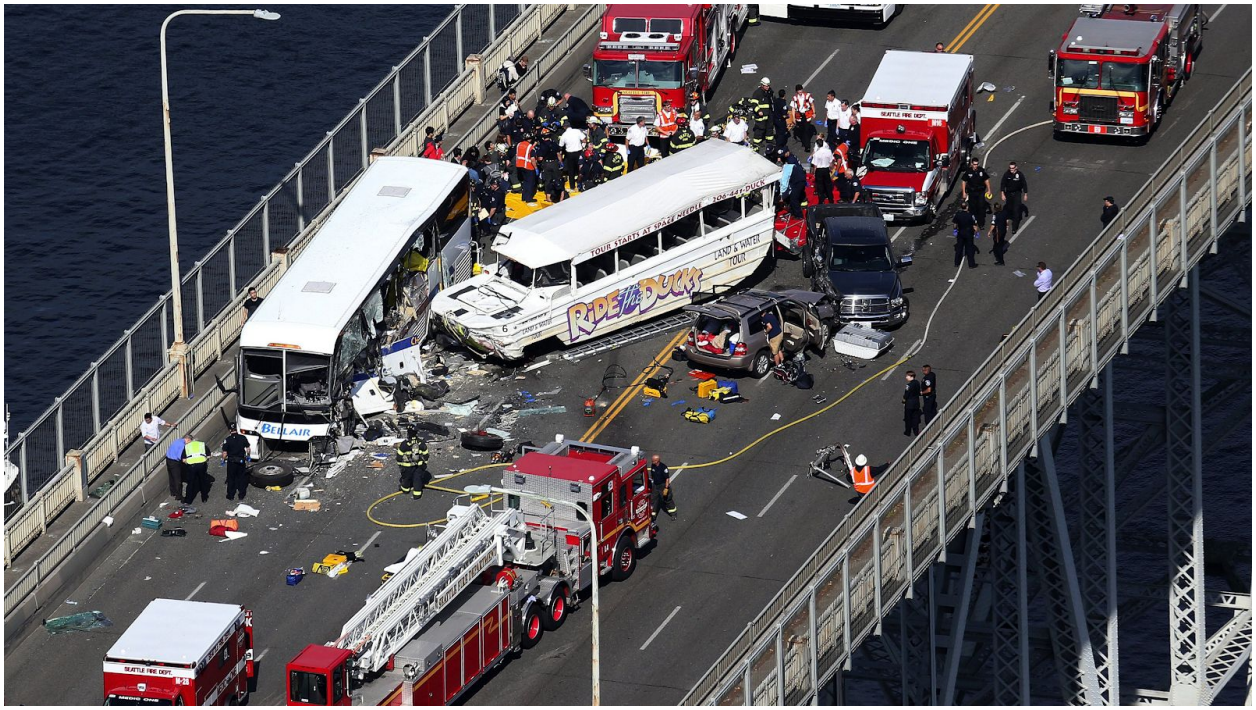
An Investigation on Road Accident And Severity in Seattle



HARSHIT JAIN

DATA SCIENCE AND PROFESSIONAL SPECIALISATION
IBM

Table Of Contents:

Business Problem

The Seattle government is going to prevent avoidable car accidents by employing methods that alert drivers, health system, and police to remind them to be more careful in critical situations.

In most cases, not paying enough attention during driving, abusing drugs and alcohol or driving at very high speed are the main causes of occurring accidents that can be prevented by enacting harsher regulations. Besides the aforementioned reasons, weather, visibility, or road conditions are the major uncontrollable factors that can be prevented by revealing hidden patterns in the data and announcing warning to the local government, police and drivers on the targeted roads.

The target audience of the project is local Seattle government, police, rescue groups, and last but not least, car insurance institutes. The model and its results are going to provide some advice for the target audience to make insightful decisions for reducing the number of accidents and injuries for the city.

Data

The data was collected by the Seattle Police Department and Accident Traffic Records Department from 2004 to present.

The data consists of 37 independent variables and 194,673 rows. The dependent variable, “SEVERITYCODE”, contains numbers that correspond to different levels of severity caused by an accident from 0 to 4.

Severity codes are as follows:

0: Little to no Probability	(Clear Conditions)
1:Very Low Probability	Chance or Property Damage
2: Low Probability	Chance of Injury
3: Mild Probability	Chance of Serious Injury
4: High Probability	Chance of Fatality

Attribute	Data type, length	Description
OBJECTID	ObjectID	ESRI unique identifier
SHAPE	Geometry	ESRI geometry field
INCKEY	Long	A unique key for the incident
COLDETKEY	Long	Secondary key for the incident
ADDRTYPE	Text, 12	Collision address type: • Alley • Block • Intersection
INTKEY	Double	Key that corresponds to the intersection associated with a collision
LOCATION	Text, 255	Description of the general location of the collision
EXCEPTRSNCODE	Text, 10	
EXCEPTRSNDESC	Text, 300	
SEVERITYCODE	Text	A code that corresponds to the severity of the collision: • 3—fatality • 2b —serious injury • 2—injury • 1—prop damage • 0—unknown
SEVERITYDESC	Text	A detailed description of the severity of the collision
COLLISIONTYPE	Text	COLLISION TYPE
PERSONCOUNT	Double	The total number of people involved in the collision
PEDCOUNT	Double	The number of pedestrians involved in the collision. This is entered by the state.
PEDCYLCOUNT	Double	The number of bicycles involved in the

		collision. This is entered by the state.
VEHCOUNT	Double	The number of vehicles involved in the collision. This is entered by the state
INJURIES	Double	The number of total injuries in the collision. This is entered by the state.
SERIOUSINJURIES	Double	The number of serious injuries in the collision. This is entered by the state.
FATALITI ES	Double	The number of fatalities in the collision. This is entered by the state

INCDATE	Date	The date of the incident
INCDTTM	Text	The date and time of the incident
JUNCTIONTYPE	Text	The date and time of the incident.
SDOT_COLCODE	Text	A code given to the collision by SDOT.
SDOT_COLDESC	Text	Text, 300 A description of the collision corresponding to the collision code.
INATTENTIONIND	Text	Whether or not collision was due to inattention. (Y/N)
UNDERINFL	Text	Whether or not a driver involved was under the influence of drugs or al Whether or not a driver involved was under the influence of drugs or alcohol.
WEATHER	Text	A description of the weather conditions during the time of the collision.
ROADCOND	Text	The condition of the road during the collision.
LIGHTCOND	Text	The light conditions during the collision.
PEDROWNOTGRNT	Text	Whether or not the pedestrian right ofway was not granted. (Y/N)

SDOTCOLNUM	Text	A number given to the collision by SDOT.
SPEEDING	Text	Whether or not speeding was a factor in the collision. (Y/N)
ST_COLCODE	Text	A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code Dictionary .
SEGLANEKEY	Text	A key for the lane segment in which the collision occurred.
CROSSWALKKEY	Long	A key for the crosswalk at which the collision occurred.
HITPARKEDCAR	Long	Whether or not the collision involved hitting a parked car. (Y/N)

Furthermore, because of the existence of null values in some records, the data needs to be preprocessed before any further processing.