

CapstoneProject_CarAccidentSeverity Final Submission

By

Sri Soumya Gudla

Business and Integration Arch. Analyst

Accenture.

1. Business Problem

Considering the traffic conditions in the Seattle city let's consider a business model like installing accident prone areas signboard at the areas where more severe accidents occur. To predict the top 10 vulnerable places it is required to collect the accident data. The dataset is obtained from SDOT traffic management system available from 2004 to present. In order to predict the most vulnerable places to accident the following factors are considered i.e. time of occurrence, weather, fatalities, traffic delay, property damage etc. So, considering these factors the data has to be segregated and cleaned for predicting the accident severity at those 10 vulnerable places.

2. Dataset

This dataset is about past accident data. The Data-Collisions.csv data set includes details of 194673 accidents from 2004 to 2020. It has a total of 38 fields. The following fields are as below

- OBJECTID:- ESRI unique identifier
- SHAPE:- ESRI geometry field
- INCKEY:- unique key for the incident
- COLDETKEY:-Secondary key for the incident
- ADDRTYPE:- Collision address type: Alley\Block\Intersection
- INTKEY :- Key that corresponds to the intersection associated with a collision
- LOCATION:-Description of the general location of the collision
- EXCEPTRSNCODE
- EXCEPTRSNDESC
- SEVERITYCODE:- A code that corresponds to the severity of the collision(3—fatality,2b—serious injury,2—injury,1—prop damage,0—unknown)
- SEVERITYDESC:-A detailed description of the severity of the collision
- COLLISIONTYPE:- Collision type
- PERSONCOUNT :- The total number of people involved in the collision
- PEDCOUNT: - The number of pedestrians involved in the collision. This is entered by the state.
- PEDCYLCOUNT: - The number of bicycles involved in the collision. This is entered by the state.
- VEHCOUNT: - The number of vehicles involved in the collision. This is entered by the state.
- INJURIES: - The number of total injuries in the collision. This is entered by the state.
- SERIOUSINJURIES: - The number of serious injuries in the collision. This is entered by the state.
- FATALITIES: - The number of fatalities in the collision. This is entered by the state.
- INCDATE: - The date of the incident.
- INCDTTM: - The date and time of the incident.
- JUNCTIONTYPE:- Category of junction at which collision took place
- SDOT_COLCODE:- A code given to the collision by SDOT.

- SDOT_COLDESC:- A description of the collision corresponding to the collision code.
- INATTENTIONIND:-Whether or not collision was due to inattention. (Y/N)
- UNDERINFL: - Whether or not a driver involved was under the influence of drugs or alcohol.
- WEATHER: - A description of the weather conditions during the time of the collision.
- ROADCOND: - The condition of the road during the collision.
- LIGHTCOND: - The light conditions during the collision.
- PEDROWNOTGRNT: - Whether or not the pedestrian right of way was not granted. (Y/N)
- SDOTCOLNUM: - A number given to the collision by SDOT.
- SPEEDING: - Whether or not speeding was a factor in the collision. (Y/N)
- ST_COLCODE:- A code provided by the state that describes the collision. For more information about these codes, please see the State Collision Code Dictionary.
- ST_COLDESC:- A description that corresponds to the state's coding designation.
- SEGLANEKEY: - A key for the lane segment in which the collision occurred.
- CROSSWALKKEY: - A key for the crosswalk at which the collision occurred.
- HITPARKEDCAR: - Whether or not the collision involved hitting a parked car. (Y/N)

So the data has to be prepared before building a model. Data Wrangling is the process of converting data from the initial format to a format that may be better for analysis. The steps followed are as shown below.

* Identifying and handling missing values

* Data Evaluation

* Data Standardization

Once data is cleaned it can be used to visualize and built the model for the business problem stated above.

3. Reference

https://github.com/srisoumya02/Coursera_Capstone/blob/main/CapstoneProject_CarAccidentAssgnment.ipynb?short_path=5871d64