DATA science project –

Prediction of Accident Severity

By Narendra Jayaram
Dated 9/13/2020

Attributes

- In total, there are 37 attributes (columns)
- Not all attributes are useful, so you need to decide what to keep and what to drop
- Some attributes have missing data
- You have numerical and categorical types of data

Location	Road condition
Weather condition	Junction junction
Car Speeding	number of people involved
Light conditions	number of vehicles involved in

Resources

- Kaggle.com
- Public open data such open.canada .ca, data.gov.uk
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.
- Antoine H´ebert, Timoth´ee Gu´edon, Tristan Glatard, and Brigitte Jaumard. "High-Resolution Road Vehicle Collision Prediction for the City of Montreal". https://arxiv.org/pdf/1905.08770.pdf, Nov 2019

INTRODUCTION:

What's new?

- In February 2020, we released phase 2 of our **Bike and Pedestrian Safety Analysis**, to look at bicycle and pedestrian incident trends. This tool helps us proactively make safety enhancements across the city. This groundbreaking approach helps us prioritize locations, anticipate issues, and make decisions informed by data.

- On December 10, Mayor Jenny A. Durkan **announced a series of steps to improve safety on City streets** and reaffirm the City's commitment to achieving the Vision Zero goal of ending traffic deaths and serious injuries by 2030. Mayor Durkan announced the City will **reduce speed limits**

to 25 miles per hour (mph) throughout the city, **double the number of safety-enhanced traffic signals**, invest in engineering changes to create safer streets, create a new crash review task force, and launch additional traffic safety education and enforcement tactics. Read our **2019 update** to learn more.

While we're excited to implement these steps, we also want to take a moment to remind everyone that we all have a role to play in improving safety. As you're traveling Seattle's streets, look out for yourself and for each other. Recognize that every intersection is unique, so stay alert. If you're driving, pay attention, slow down, and expect people are walking and biking in every part of the city at all times of the day.

Tasks Included in current notebook:

- Loading of data.
- Preprocessing.
- Visualizations.

++

1. **Data understanding:**

Data 2018 was one of the lowest years on record for transportation-related fatalities. Still, 14 people lost their lives last year and more than 170 people sustained serious injuries. As of December 2019, twenty-five people have been killed on our streets and another 153 people have been seriously injured. Impairment, distracted driving, speeding, and failure to stop for pedestrians continue to be the top contributing causes to collisions in Seattle. Pedestrians account for the majority of victims in fatal collisions in Seattle. Pedestrians make up just 4% of total collisions annually but more than 50% of fatalities. And we know that older pedestrians are more at risk. In 2019, the median age of pedestrians killed in collisions is 62 years old. More than 90% of severe collisions occur on our arterial streets. Arterials carry the highest volumes of people walking, biking, driving, rolling, or riding transit and more than 80% of our arterial streets have speed limits higher than 30 miles per hour.

This project will look at predicting the probability and severity of vehicular accidents based on weather and other characteristics, using historic collision data.

Data Set Summary Data Set Basics Title Collisions—All Years Abstract All collisions provided by SPD and recorded by Traffic Records. Description This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present. Supplemental Information Update Frequency Weekly Keyword(s) SDOT, Seattle, Transportation, Accidents, Bicycle, Car, Collisions, Pedestrian, Traffic, Vehicle Contact Information Contact Organization SDOT Traffic Management Division, Traffic Records Group Contact Person SDOT GIS Analyst Contact Email DOT_IT_GIS@seattle.gov

Stakeholders:

- Public Development Authority of Seattle
- Transport authority

The dataset used for this project is based on car accidents which have taken place within the city of *Seattle, Washington* from the year *2004* to *2020*. This data is regarding the *severity of each car accidents* along with the time and conditions under which each accident occurred. Records are assigned to two groups – 'property damage only' and 'injury' collisions – which makes it ideal for a supervised learning classification model. There are approximately 195,000 samples in the dataset

The model aims to predict the severity of an accident, considering that, the variable of Severity Code was in the form of *1* (Property Damage Only) and *2* (Physical Injury)

The dataset describes each collision using 36 different features, which provides a robust set of independent variables as possible parameters for training the model.

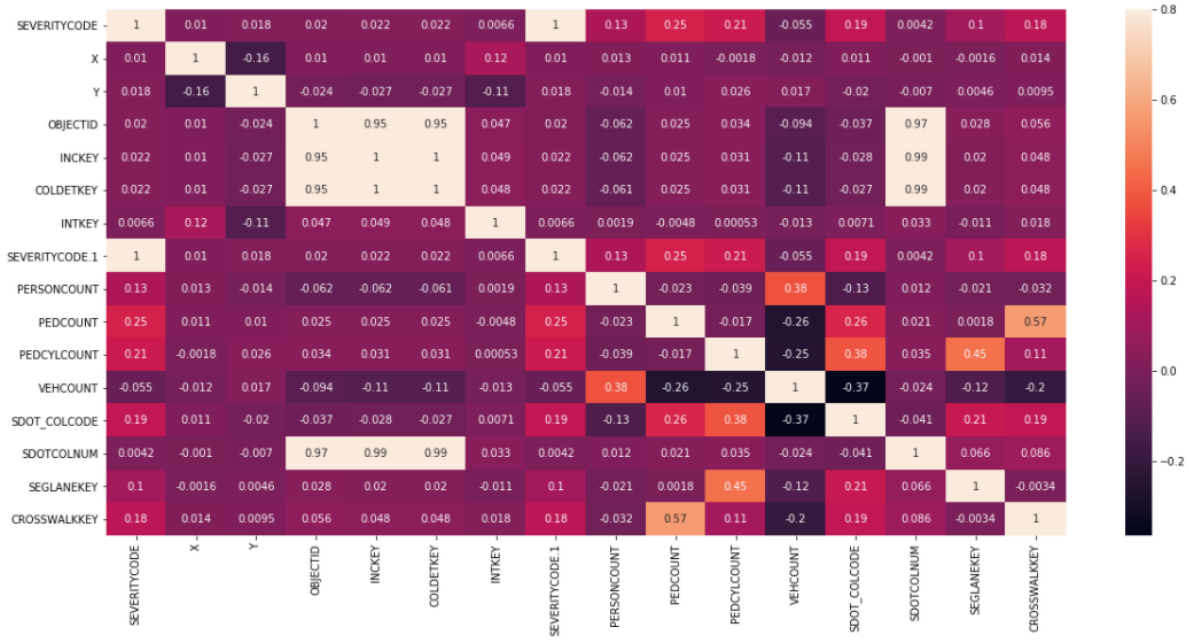
These features identify: •

- the date, time, position and location details of the collision; •
- codes used by the state and DOT to categorize the collision; •
- characteristics of the incident, such as type of collision and the number of people, pedestrians, bicycles and vehicles involved; •
- environmental factors like weather and road condition; and •
- the existence of driver behaviors including speeding, inattentiveness and the presence of drugs or alcohol

A model for predicting severity will be built using historic collision records collected by the Seattle Police Department and maintained by the SDOT Traffic Management Division

```
Type 1 collisions (property damage) =  136485
Type 2 collisions (with injury) =  58188
Ratio of Collisions with property damage only to those with injuries: 2.35
Ratio of Collisions with injuries to those with property damage: 0.43
```

Correlation of data




```
PEDROWNOTGRNT     190006
EXCEPTRSNDESC     189035
SPEEDING          185340
INATTENTIONIND    164868
INTKEY            129603
EXCEPTRSNCODE     109862
SDOTCOLNUM         79737
JUNCTIONTYPE        6329
Y                   5334
X                   5334
LIGHTCOND           5170
WEATHER             5081
ROADCOND            5012
COLLISIONTYPE       4904
ST_COLDESC          4904
UNDERINFL           4884
LOCATION            2677
ADDRTYPE            1926
ST_COLCODE            18
dtype: int64
```

Categorical features

```
Index(['REPORTNO', 'STATUS', 'ADDRTYPE', 'LOCATION', 'EXCEPTRSNCODE',
       'EXCEPTRSNDESC', 'SEVERITYDESC', 'COLLISIONTYPE', 'INCDATE', 'INCDTTM',
       'JUNCTIONTYPE', 'SDOT_COLDESC', 'INATTENTIONIND', 'UNDERINFL',
       'WEATHER', 'ROADCOND', 'LIGHTCOND', 'PEDROWNOTGRNT', 'SPEEDING',
       'ST_COLCODE', 'ST_COLDESC', 'HITPARKEDCAR'],
      dtype='object')
```

Numerical Features

```
Index(['SEVERITYCODE', 'X', 'Y', 'OBJECTID', 'INCKEY', 'COLDETKEY', 'INTKEY',
       'SEVERITYCODE.1', 'PERSONCOUNT', 'PEDCOUNT', 'PEDCYLCOUNT', 'VEHCOUNT',
       'SDOT_COLCODE', 'SDOTCOLNUM', 'SEGLANEKEY', 'CROSSWALKKEY', 'DATETIME'],
      dtype='object')
```

```
Numerical features : 16
Categorical features : 22
```