US-Accident: A Countrywide Traffic Accident Dataset

About the Dataset:

The dataset it taken from the Kaggle Website and was uploaded by Sobhan Moosavi. It is a United States country wide collection of data from 49 states that is continuously collected starting from February 2016. Gathering data is done by several data providers and multiple API (Application Programming Interface) that provide streaming of data. Those API broadcast traffic incidents that is then recorded by State Department of transportation, law enforcement agencies, traffic cameras and sensors within its road network.

As of this data it has collected data starting from February 2016 to March 2023 and has a record for about 7.7 million accidents. However as stated by the contributor this data set may have missing information and data for certain days possibly by network connectivity issues at the time of data collection. Unfortunately, has stated that this information shall no longer be updated and the data set shall be the current one.

Overview of the Problem and How Data Science could help:

On the research at the National Highway Traffic Safety Administration that in 2019 that $340 billion was the cost of motor vehicle crashes. Equivalent to $1,035 for each of the 328 million people in the United States and 1.6% of the $21.4 trillion U.S gross domestic product for 2019 itself.

Losses include medical cost, lost of productivity, legal and court cost, emergency service costs, insurance administration cost, congestion costs, property damage and workplace losses. Thus, when quality of life valuation is considered, the total value of societal harm was nearly $1.4 trillion.

The target feature looking at is the Severity that has a number of 1 being the shortest traffic delay and 4 that has a significant impact on traffic. And the independent variables are what presence in that road has a Bump, Crossing, Junction, No Exit, Railway, Roundabout, Station and Stop.

I tend to use Data Science to predict what such presence contributes to the Severity. Finding areas, city of County or States with possible increase occurrence. Identifying the presence factors may help to provide appropriate further studies and solutions to minimize adverse effect of severity.

Attributes Description

#	Attribute	Description	Nullable
1	ID	This is a unique identifier of the accident record.	No

#	Attribute	Description	Nullable
2	Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).	No
3	Start_Time	Shows start time of the accident in local time zone.	No
4	End_Time	Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.	No
5	Start_Lat	Shows latitude in GPS coordinate of the start point.	No
6	Start_Lng	Shows longitude in GPS coordinate of the start point.	No
7	End_Lat	Shows latitude in GPS coordinate of the end point.	Yes
8	End_Lng	Shows longitude in GPS coordinate of the end point.	Yes
9	Distance(mi)	The length of the road extent affected by the accident.	No
10	Description	Shows natural language description of the accident.	No
11	Number	Shows the street number in address field.	Yes
12	Street	Shows the street name in address field.	Yes
13	Side	Shows the relative side of the street (Right/Left) in address field.	Yes
14	City	Shows the city in address field.	Yes
15	County	Shows the county in address field.	Yes
16	State	Shows the state in address field.	Yes
17	Zipcode	Shows the zipcode in address field.	Yes

#	Attribute	Description	Nullable
18	Country	Shows the country in address field.	Yes
19	Timezone	Shows timezone based on the location of the accident (eastern, central, etc.).	Yes
20	Airport_Code	Denotes an airport-based weather station which is the closest one to location of the accident.	Yes
21	Weather_Timestamp	Shows the time-stamp of weather observation record (in local time).	Yes
22	Temperature(F)	Shows the temperature (in Fahrenheit).	Yes
23	Wind_Chill(F)	Shows the wind chill (in Fahrenheit).	Yes
24	Humidity(%)	Shows the humidity (in percentage).	Yes
25	Pressure(in)	Shows the air pressure (in inches).	Yes
26	Visibility(mi)	Shows visibility (in miles).	Yes
27	Wind_Direction	Shows wind direction.	Yes
28	Wind_Speed(mph)	Shows wind speed (in miles per hour).	Yes
29	Precipitation(in)	Shows precipitation amount in inches, if there is any.	Yes
30	Weather_Condition	Shows the weather condition (rain, snow, thunderstorm, fog, etc.)	Yes
31	Amenity	A POI annotation which indicates presence of amenity in a nearby location.	No
32	Bump	A POI annotation which indicates presence of speed bump or hump in a nearby location.	No
33	Crossing	A POI annotation which indicates presence of crossing in a nearby location.	No

#	Attribute	Description	Nullable
34	Give_Way	A POI annotation which indicates presence of give_way in a nearby location.	No
35	Junction	A POI annotation which indicates presence of junction in a nearby location.	No
36	No_Exit	A POI annotation which indicates presence of no_exit in a nearby location.	No
37	Railway	A POI annotation which indicates presence of railway in a nearby location.	No
38	Roundabout	A POI annotation which indicates presence of roundabout in a nearby location.	No
39	Station	A POI annotation which indicates presence of station in a nearby location.	No
40	Stop	A POI annotation which indicates presence of stop in a nearby location.	No
41	Traffic_Calming	A POI annotation which indicates presence of traffic_calming in a nearby location.	No
42	Traffic_Signal	A POI annotation which indicates presence of traffic_signal in a nearby loction.	No
43	Turning_Loop	A POI annotation which indicates presence of turning_loop in a nearby location.	No
44	Sunrise_Sunset	Shows the period of day (i.e. day or night) based on sunrise/sunset.	Yes
45	Civil_Twilight	Shows the period of day (i.e. day or night) based on civil twilight.	Yes
46	Nautical_Twilight	Shows the period of day (i.e. day or night) based on nautical twilight.	Yes
47	Astronomical_Twilight	Shows the period of day (i.e. day or night) based on astronomical twilight.	Yes