Traffic Accident Severity Prediction

Team Name: The Collective

Team Members and Email id's:

1. Susmitha Srirangam - 00764501 - ssrir5@unh.newhaven.edu
2. Kaivalya Reddy Maddireddy - 00760616 - kmadd3@unh.newhaven.edu
3. Thatikonda Akanksha - 00764338 - athat2@unh.newhaven.edu

Research Question: The aim is to predict the severity of accident based on the data collected by applying various data mining techniques.

Dataset information:

This dataset contains eleven attributes with values recorded based on time while considering hours of travel like rush hour, workdays, weekends, surface conditions, weather conditions, speed limits and predicting the severity of accidents weather they are fatal or non-fatal.

Attribute Information:

- **Rush Hour:** Shows weather the hour of travel is comes under rush hour or non-rush hour so, we can get to know the traffic conditions of the area.
- **Work Zone:** Shows if the area of travel is coming under work zone or not from this, we can decide how the timings impact traffic in this zone
- **Workday:** Shows if the day is workday or not because of which we can determine the which zones can see traffic during the working days.
- **INT_HWY**
- **LGTCON_day**
- **Level:** The level of road from the surface which can also tend to accidents when having the impact of weather
- **Speed Limit:** Travel speed of vehicles during the traffic and while having the impact of other conditions mostly lead to accidents
- **Surface Conditions:** Shows weather the surface condition of the road is good or wearied off including the climate conditions like snow, rainy or greasy roads

- **Traffic Two Way:** This shows if the traffic is one way or two ways to rule out and analyze the conditions opposite side
- **Maximum Severity:** With analysis of all these conditions the severity of the accident is determined weather it is fatal or non-fatal.

Data Exploration procedures using R programming:

Data exploration is the process of analyzing data to comprehend and list its key features using statistical and graphical techniques. During this process, we dig into data to see what story the data have, what we can do to enrich the data, and how we can link everything together to find a solution to a research question.

The procedure for data exploration and visualization using R Data loading is simple and can be done with a range of file formats thanks to the availability of predefined libraries and simple syntax. XLS, TXT, CSV, and JSON are simple.

Changing variables: In R, changing a variable's data type requires adding a character string to a numeric vector and converting the vector's elements to the character string.

Dataset transposition: R offers code to convert datasets from wide to narrower structures. Data frame sorting is conducted by using order as an index.

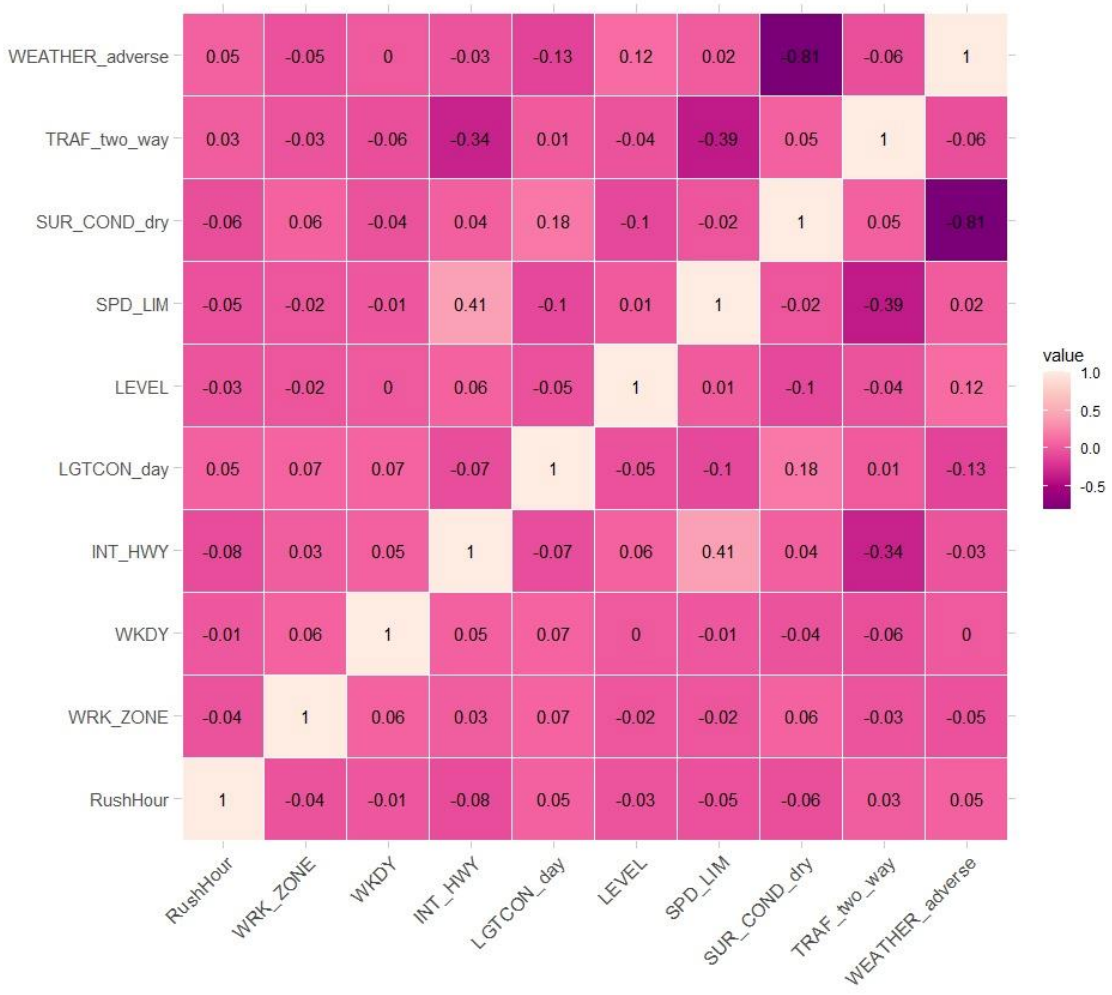
Plot or histogram data: To better understand the distribution between categories, create frequency tables.

Correlation matrix: This used to summarize data. Here we analyze relation between two or more variables by calculating similarities between them.

Outliers Detection: Outliers in the data is the observations made to find out if there are any exception cases present in the input data. Here we will find out if the observations are large or small value compared to most observations. Boxplot and Cleveland dot plots are tools for outlier detection.

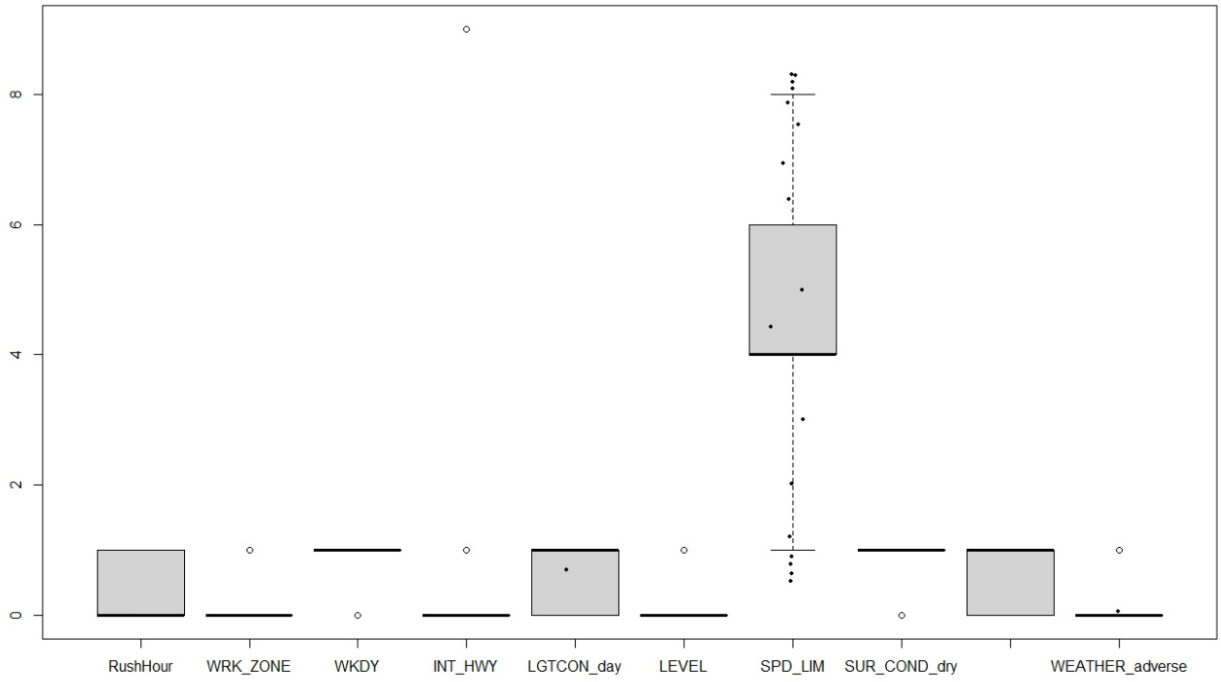
Description of data exploration techniques used:

Here we have applied correlation to find out similarity between the unique features so, we can eliminate exploring any similar attributes and in turn narrowing down the search.



According above graph the similarities between the attributes is exceptionally low so, we can use all the attributes in predicting the severity of the accidents occurred. By considering all these features our goal is to predict the future possibility of accidents that might occur.

Outliers Detection: Our next process is to find if there are any outlining data present in the dataset. Outlining data might occur due to some conditions like road works or any other blockings that might occur due to accidents happened on that day on the road.



GitHub Repository:

https://github.com/susmitha7599/TheCollective