Web Application on Shiny using Data Analysis and Visualization

Data Mining Applications, Northeastern University

Professor: Justin Grosz

Submitted by: Jonathan Vinay Gabriel

Introduction

The City of Austin has a lot of issues and complaints from cyclists relating to the accidents and incidents taking place due to motor vehicles. The report presents the analytical application created for the cyclists in City of Austin to determine the risk of cycling based on the Day of the week, time, roadway part and the surface condition. The application was created using Shiny(R).

Steps Involved in Creating the Application

To create the analytical application, I began by loading the dataset and exploring the City of Austin's bike crash dataset. I checked for missing values and created bar plots to explore each variable in the dataset. This revealed that Non-Capacitating Injury had the highest frequency in the dataset, indicating that most of the incidents were not severe. Furthermore, the crash time was most frequent after 3 pm in the evening. Additionally, most crashes took place on dry surface conditions, with the highest frequency of the roadway part mentioned as Main/Proper Lane.

Further, I wanted to explore and visually demonstrate that not wearing helmets was the reason for most of the incidents. The bar plot of Crash Severity by Person Helmet showed that cyclists not wearing helmets were more prone to accidents (Figure 1).

To better understand the importance of variables, I reloaded the dataset under a different name and converted categorical variables to numerical ones. Then, I created a correlation matrix and used a heatmap to interpret the relationships. However, the results were not very significant. Hence, I proceeded to perform the variable importance test, which provided the results shown in

Figure 2. These results were helpful to understand which variables were less significant in calculating the risk score.

After loading the dataset for the third time as "trial_dataset", I dropped all the irrelevant variables and created dummy variables for the remaining ones. Using these dummy variables, I conducted the feature importance test again, this time to analyze which inputs in each of the variables "Day of Week", "Roadway Part", "Surface Condition", and "Timing" had the highest importance with respect to the target variable "Crash Severity". This analysis helped in assigning risk scores for creating the application.

Analytical Application

"Bike Crash Risk Score Calculator" assesses the risk of cycling in the City of Austin based on factors such as Time of the day, Day of the Week, Road Conditions (Dry, Wet, Standing Water, Ice, Sand, Mud, and Dirt), and Location (Main/Proper Lane, Entrance/On Ramp, Service/Frontage Road).

This application assists cyclists in analyzing the percentage of risk involved based on the provided factors. It enables them to take necessary precautions to ensure safety and prevent accidents.

The tool predicts the risk of cycling based on the user's input within the City of Austin. The risk score is determined using the results of a feature importance test conducted with the target variable "Crash Severity."

Bike Crash Risk Score Calculator

Risk Score: 20

Time of Day:

0   12   23

0 3 6 9 12 15 18 21 23

Day of Week:

Monday ▼

Road Conditions:

Dry ▼

Location:

Entrance/On Ramp ▼

Calculate Risk Score

Bike Crash Risk Score Calculator

Risk Score: 20

Time of Day:

0   12   23

0 3 6 9 12 15 18 21 23

Day of Week:

Monday ▼

Road Conditions:

Dry ▼

Location:

Entrance/On Ramp ▼

Calculate Risk Score

Based on the inputs provided, the risk score is generated.

References

Brownlee, J. (2019, August 22). *Feature Selection with the Caret R Package*. Retrieved from
    https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/
Shiny Gallery. (2024). Retrieved from https://shiny.posit.co/r/gallery/

Appendix

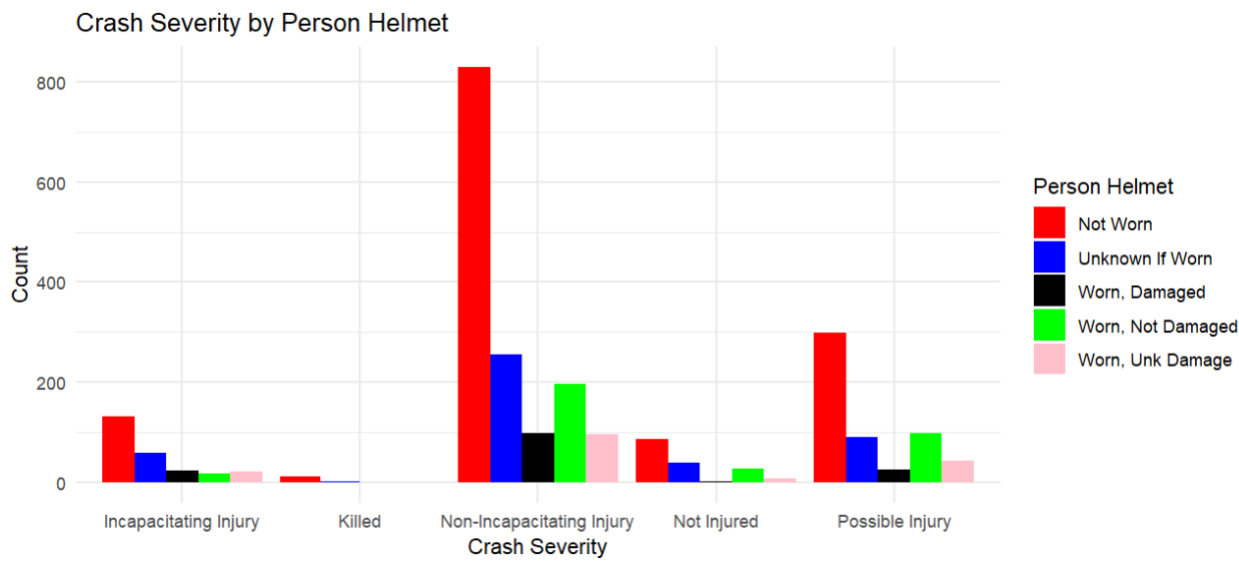Figure 1

```
lm variable importance

                                Overall
`\\`Roadway Part\\``            3.3599
`\\`Surface Condition\\``       2.5455
`\\`Person Helmet\\``           2.1729
`\\`Speed Limit\\``             1.1772
`\\`Active School Zone Flag\\`` 0.8793
`\\`At Intersection Flag\\``    0.5244
`\\`Day of Week\\``             0.4540
`\\`Crash Time\\``              0.4319
`\\`Construction Zone Flag\\``  0.0200
```

Figure 2