Term Project - Predicting Risky Intersection Behaviour Driven by Misinterpretation of Traffic Signals

By – Shreepati Jha(B01756304)

Can drivers’ misinterpretation of yellow signals, combined with driving speed patterns and cognitive factors, predict the likelihood of running red lights during the yellow-to-red transition?

Dataset -

The Dataset used for the analysis was a primary survey dataset collected from auto/taxi drivers in the City of Hyderabad, India. The dataset consists of 276 observations and 44 variables, primarily focusing on the transportation behaviours, concerns, and patterns of drivers in Hyderabad. It includes both categorical and numerical data, capturing aspects such as the **age** and **vehicle type** of drivers, **work hours**, **job concerns**, and **vehicle ownership** status. The dataset also delves into driving behaviours, with variables like **driving frequency**, **driving distance**, and **response to red lights** at intersections. It includes details on various factors influencing driving decisions, such as **mobile payment usage**, **income range**, and **road sign awareness**. The variables also cover specific driving behaviours like speed during straight drives, turns, and interaction with signals (e.g., **Yellow Signal Meaning**, **Green-to-Yellow Confusion**). Other attributes focus on external factors, including **traffic impact**, **vehicle insurance**, and **road sign notices**, reflecting a comprehensive view of the drivers' experiences, challenges, and decision-making processes on the road. Additionally, the dataset also highlights some factors contributing to red-light violations and traffic-related behaviours. Missing data in certain variables, such as **Sign_X** and **Sign_GiveWay**, are also present in the dataset.

Based on the Exploratory data analysis of the variables, a novel research question evolved to study the **cognitive factors** (e.g., understanding traffic signals, sign recognition) with **behavioral metrics** (e.g., speed at intersections, lane choices) to model risky behavior. The hypothesis is to analyse the drivers’ *perception* of traffic rules and their *real-world actions* and the results could potentially lead to insights which help in education of drivers and adjustment of traffic policies.

Final reacrch quesiton formulated -

Can drivers' misinterpretation of yellow signals, combined with driving speed patterns and cognitive factors, predict the likelihood of running red lights during the yellow-to-red transition?

Variables selected for the Analysis -

Among the 44 variables we had, a few variables related to cognitive understanding, behaviour factors, demographic context and the traffic sign recognition ability were selected. a detailed insight into these variables are as follows:-

Binary Dependent Variable:	
Running_RedLight_Yellow	running a red light at while still yellow
Binary Independent Variables	
Behavioral Factors	
Speed_Straight	Speed while crossing straight through the intersection
Speed_LeftTurn	Speed while turning left
Speed_RightTurn	Speed while turning right
Change_Speed_Intersection	Change Speed at Intersections
Cognitive Misinterpretation	
Yellow_Signal_Meaning	Meaning of Yellow Traffic Signal
GreenToYellow_Confusion	Decision Making on Green-to-Yellow Transition
Amber_SufficientTime	Sufficient Amber Time
Demographic/Job Context:	
Age	Age
Work_Hours	Average Working Hours Per Week
Traffic Sign Recognition	
Sign_SpeedLimit	Traffic Sign - SPEED LIMIT
Sign_NoEntry	Traffic Sign - NO ENTRY

The categories of the variables have been discussed in detail in the R Markdown file provided.

Analysis Components

Multiple cross-tabulations and chi-square tests were computed among the variables to find insights. Other than that, a binary logit model was also estimated for the variables mentioned above.

Summary of cross tabulation variables and the result of chi-square tests:-

- Speed vs Running Red Light at Yellow (Not significant)
- Confusion at Yellow vs Speed Straight (Not significant)
- Speed Limit Sign vs Red Light Binary (Not significant)
- Amber Time Sufficiency vs Running Red Light (Not significant)
- Age vs Speed Behaviour (Right Turn only, significant)
- Work Hours vs Running Red Light at Yellow (Significant)
- Green-to-Yellow Confusion vs Red Light Binary (Significant)

Equation for the Full Binary logit model:-

P(RedLight_Binary = 1) = 1 / (1 + exp(-(β0 + β1 * Age + β2 * Work_Hours + β3 * Change_Speed_Intersection + β4 * Speed_Straight + β5 * Speed_LeftTurn + β6 * Speed_RightTurn + β7 * Yellow_Signal_Meaning + β8 * GreenToYellow_Confusion + β9 * Amber_SufficientTime + β10 * Sign_SpeedLimit + β11 * Sign_NoEntry)))

Final reduced logit model:-

P(RedLight_Binary = 1) = 1 / (1 + exp(-(β0 + β1 * Yellow_Signal_Meaning + β2 * GreenToYellow_Confusion)))

Conclusion

Though the binary logit model's intercept was significant and statistically significant for some predictors, it does not perform well overall, particularly in identifying risky behaviour. This suggests that further model refinement, perhaps with additional variables or different modelling techniques, is necessary for improving its predictive accuracy and usefulness.
Here the model's performance has been reported as bad accounting to the Low Pseudo R^2,Weak Sensitivity, High Specificity and Low AUC.