Road Traffic Accidents Analysis

One of the most critical, global safety problems are road accidents. Based on "World Health Organization", there were 1.35 million road traffic deaths globally in 2016 with millions more facing injuries and consequences. Globally, road traffic crashes are a <u>leading cause of death</u> among young people and specifically, the main cause of death among those aged [15, 29] years. Road traffic injuries are currently estimated to be the 8th leading cause of death across all age groups globally and are predicted to become the seventh leading cause of death by 2030 and that's why I chose this topic.

The data I will be working with for this topic is named 'Road Traffic Accidents in Addis Ababa' and it is sourced from the Kaggle website (https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents/data) and it uses police reported accident for years [2017, 2020]. Addis Ababa is a good case study since it is a rapidly expanding urban center in a developing nation and it faces challenges like aging old infrastructure and socio-economic disparities which influence road safety outcomes. Our dataset has beneficial attributes like: Time and Day Details (Day of the week, time of accidents), Driver Information (Age, sex, educational level, driving experience), Vehicle Information (Type of vehicle, ownership, service year of the vehicle) and Accident Attributes (Causes of accidents, severity).

Findings from Addis Ababa could inform strategies in other urban regions including U.S. cities like Denver, Colorado, where comparable issues persist despite differing contexts. Colorado, particularly Denver and its surrounding areas also experiences traffic issues due to its growing population, mountainous terrain and winter weather conditions. Comparing findings from Addis Ababa to Colorado may lead to show

the universal traffic safety challenges, most impacting factors and probable solutions.

Some questions which we will try to answer in this project are:

"What are the most common causes of accidents?"

"How do factors like age, sex, and educational level of drivers correlate with accident frequency?"

"Are there specific causes that are more frequent for certain types of vehicles or driving experiences?"

In order to complete this project, we divide our road map into 4 major sub-sections:

- Data Exploration: EDA using pandas to get key statistics + Data Visualization (2 days)
- Feature Engineering: Create and Modify attributes of our dataset to achieve maximum results (2 days)
- High level analysis of data: Get relationships between variables + apply simple ML algorithms for accident prediction (4 days)
- Finalize the results: Connect all results and dots and get a comprehensive conclusion + Report (3 days)

References:

1. https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/road-traffic-mortality
2. https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents/data