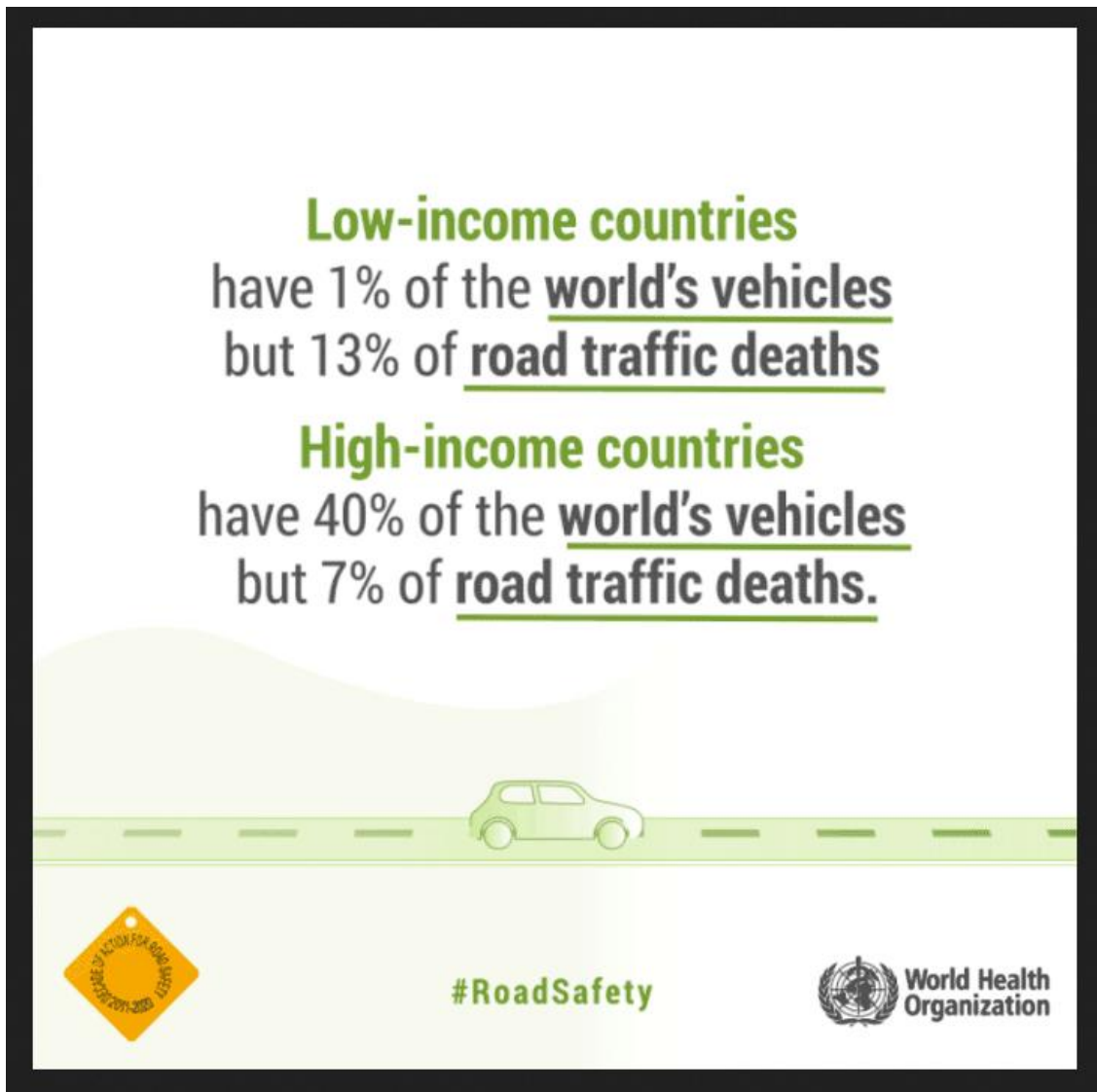


## The Capstone – Predicting Car Accident Severity (CRIPS-DM)

### Business Understanding



**Road traffic crashes** are globally among the leading causes of death and even exceed serious diseases such as HIV/AIDS [1]. The number of killed persons due to traffic crashes is estimated to be around **1.35 million** while less protected traffic participants such as pedestrians or cyclist are most affected. High-income countries count for high usage of vehicles; however, fatal rate is 3 times lower than in low-income countries but still – morbidity and mortality due to traffic crashes remain high and globally represent a high **social, health and economic burden**. In the global status report on road safety the WHO states an unacceptably high number of deaths on the world's roads [2]. Governments all over the world have taken several (legislative) measures to minimize behavioural risk factors such as speed and drink-driving, or to improve road infrastructure and vehicles safety, thus decreasing the occurrence of road traffic collisions. At the same time, methods to prevent fatal outcome in road traffic collision are taken (e.g. helmets or seatbelts). Post-crash care is another crucial factor in minimizing mortality and morbidity in traffic crashes, but highly dependent on available resources which explains the discrepancy between low- and high-income countries in providing prehospital care. The evidence-

based guideline *Save LIVES: a road safety technical package* has been provided by the WHO for decision-makers to reduce road traffic deaths [3]. The acronym **LIVES** sets the focus on the interventions speed management, **L**eadership, **I**nfrastructure design, **V**ehicle safety, **E**nforcement of traffic laws and post-crash **S**urvival.

In the context of globally rising motorization of transport and the **SDG target** to decrease road traffic deaths to 50 % by 2020, further research should be performed to find factors with enormous potential impact. This project aims to re-evaluate and identify factors leading to a high severity in collisions with serious and/or fatal injuries, and to describe the prediction strength. A prediction model on collision severity is also needed to efficiently distribute the limited resources of emergency health care to the places where needed the most.

Such a model enables decision-makers

- to focus on most efficient preventive strategies
- to allocate limited health care resources
- to provide a risk assessment for individual considerations as a traffic participant

By doing so a range of stakeholders such as **politicians, emergency health care managers, physicians and individual representatives of traffic participants** should be addressed in the evaluation process.

### References

[1] Peden M, Scurfield R, Sleet D, Mohan D, Hyder AA, Jarawan E, et al.: *World report on road traffic injury prevention*. Geneva 2004.

[2] World Health Organization (WHO): *Global status report on road safety 2018*. Geneva 2018.

[3] World Health Organization (WHO): *Save LIVES: A road safety technical package*. Geneva: World Health Organization Copyright 2017.

### Data Understanding

The data used to build this model is provided by the **Seattle Department of Transportation** which is the official governmental department focusing on transportation systems and providing safe access to places in Seattle. The dataset consists of all types of collisions from 2004 on combined with several attributes such as weather, road or light condition, involved traffic participants, infrastructural aspects and geographical data. Some of the attributes have already been linked to the occurrence of traffic collisions (e.g. speeding, weather, road or light condition) and will now be used to evaluate the collision severity. The latter is labelled by a severity code ranging from **1 (property damage only) to 3 (fatality)** in the provided dataset. Attributes of interest to predict severity in the first place without further in-depth assessment include the following points while individual factors which need further investigation such as inattention or drink-driving and the specific location were excluded.

- involvement of **vulnerable traffic participants**: pedestrians, bicycles
- **infrastructural factors**: junction type, road condition, pedestrian right of way, speeding
- **environmental factors**: weather condition, light condition

Being able to include the contextual data of a traffic collision into the model will ensure a comprehensive assessment of potential influencing factors. The choice of factors may enable conclusions for a more general setting without time or location specific data.