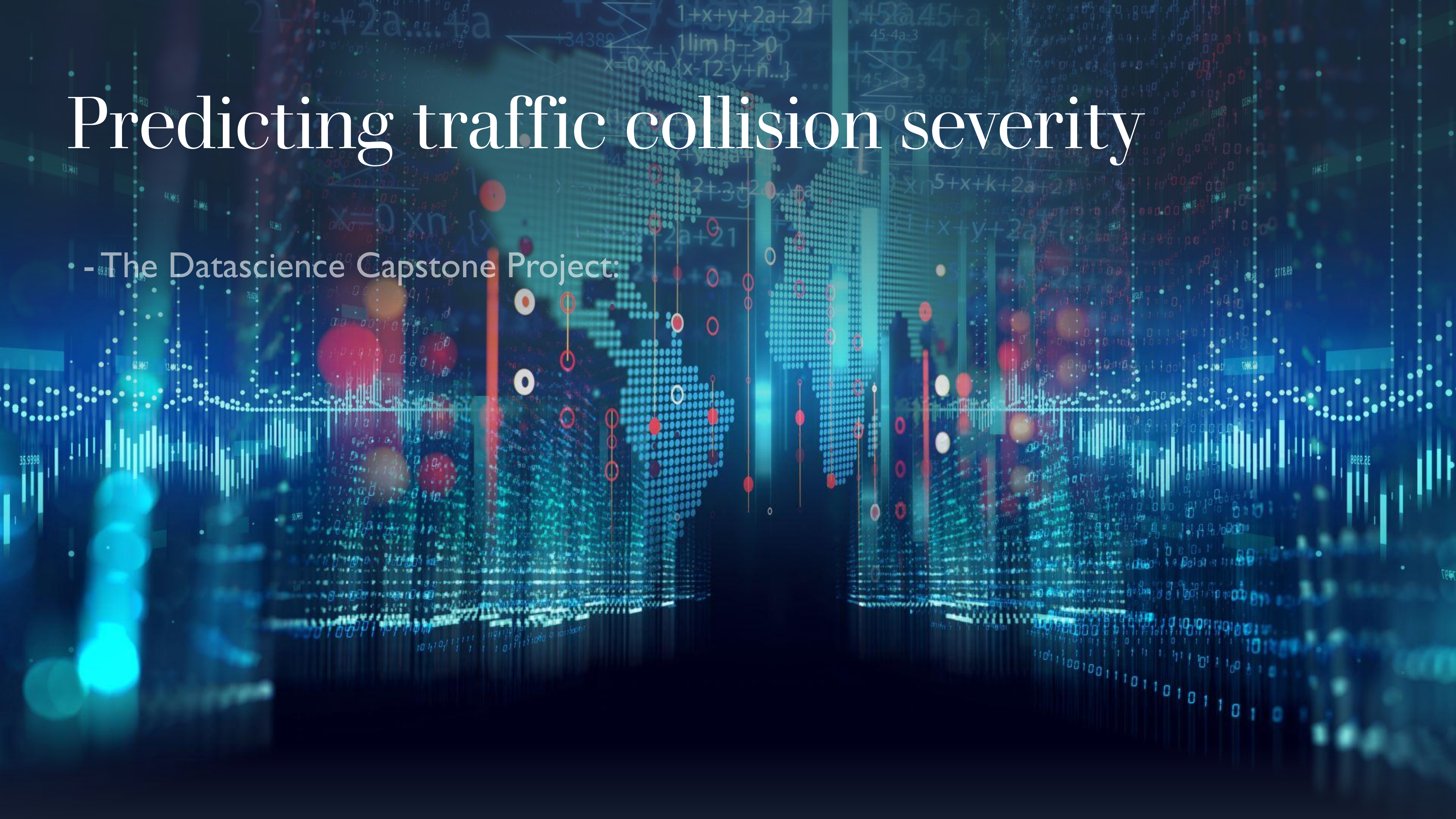
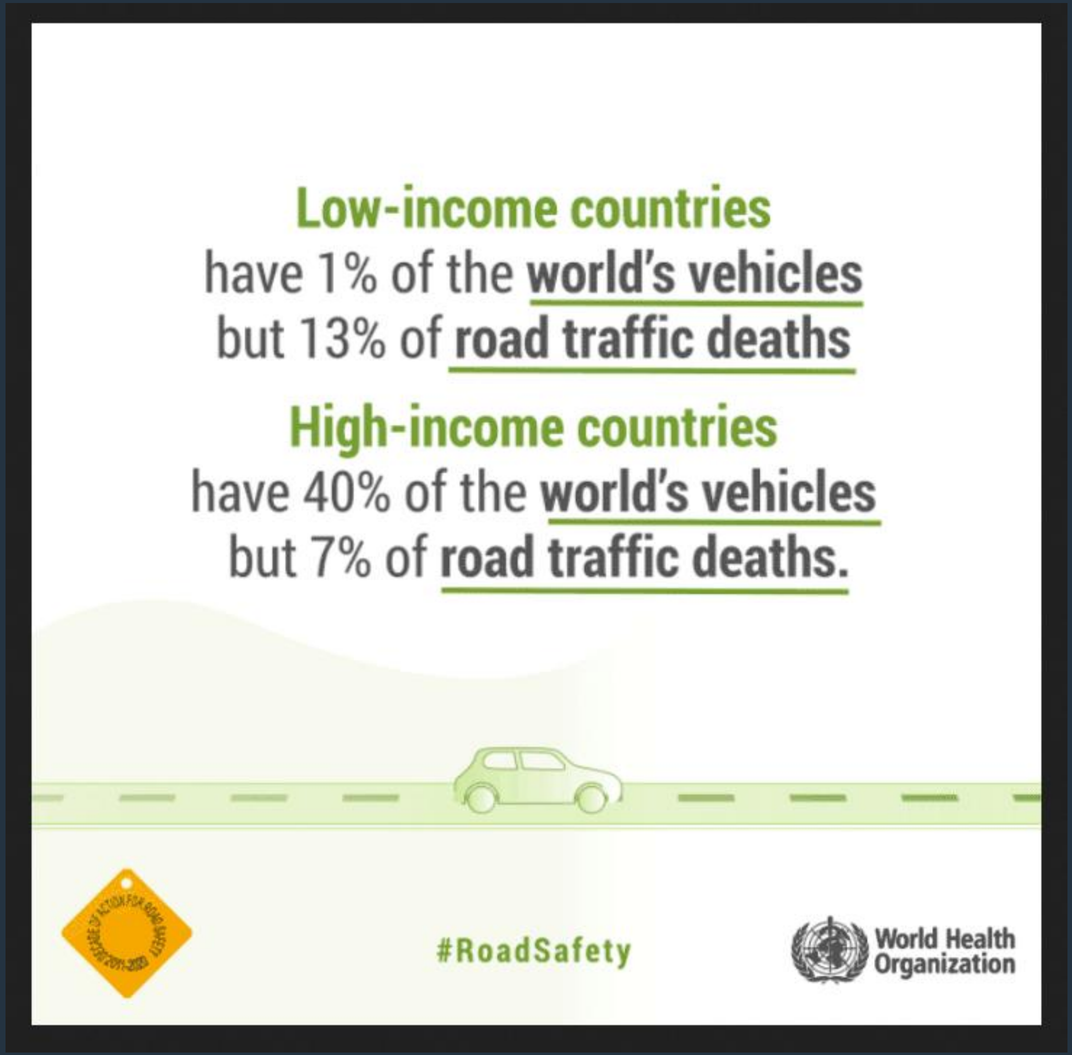
# Predicting traffic collision severity

- The Datascience Capstone Project:



**This is the final project for the data science course: It is only a testing project, no real results and outputs are generated thus the developed models should not be used for any interpretation!**

# Need for predicting traffic collision severity



**Low-income countries**
have 1% of the **world's vehicles**
but 13% of **road traffic deaths**

**High-income countries**
have 40% of the **world's vehicles**
but 7% of **road traffic deaths.**

#RoadSafety

World Health
Organization

*Image source: https://www.who.int/violence_injury_prevention/road_safety_status/2018/CAR-2.gif?ua=1*

# Need for predicting traffic collision severity - Distribution of resources is cruial for post-crash survival

- Road traffic crashes: high morbidity and mortality rates
  - Especially high burden for low income countries
- Prevention strategies to reduce mortality
  - Interventions: speed management, infrustructure, vehicle safety, traffic legislation
- Post-crash survival
  - Health care resources as limited factor: districtuion of resources is crucial

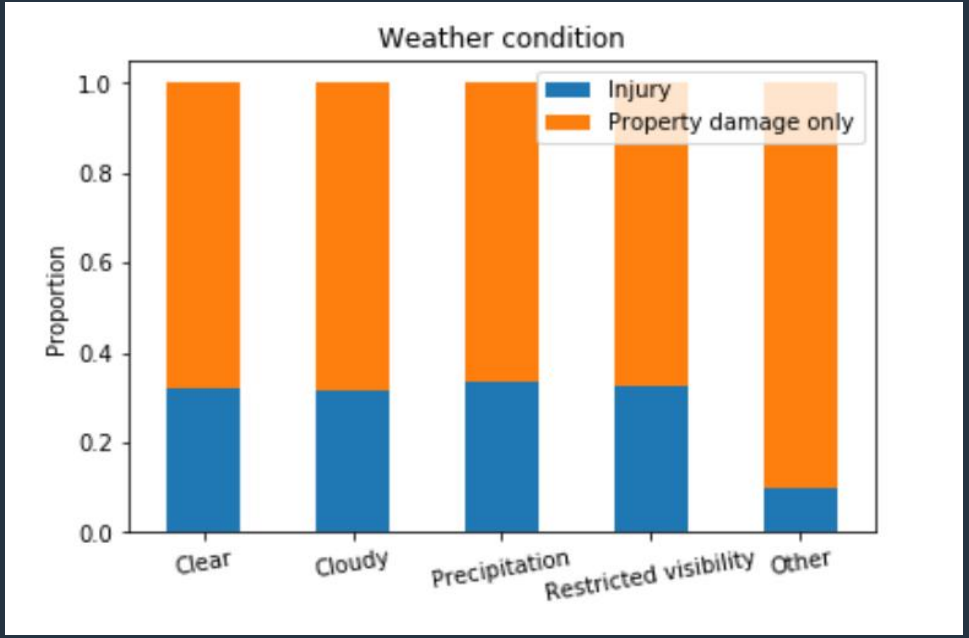
# Methodology - Data

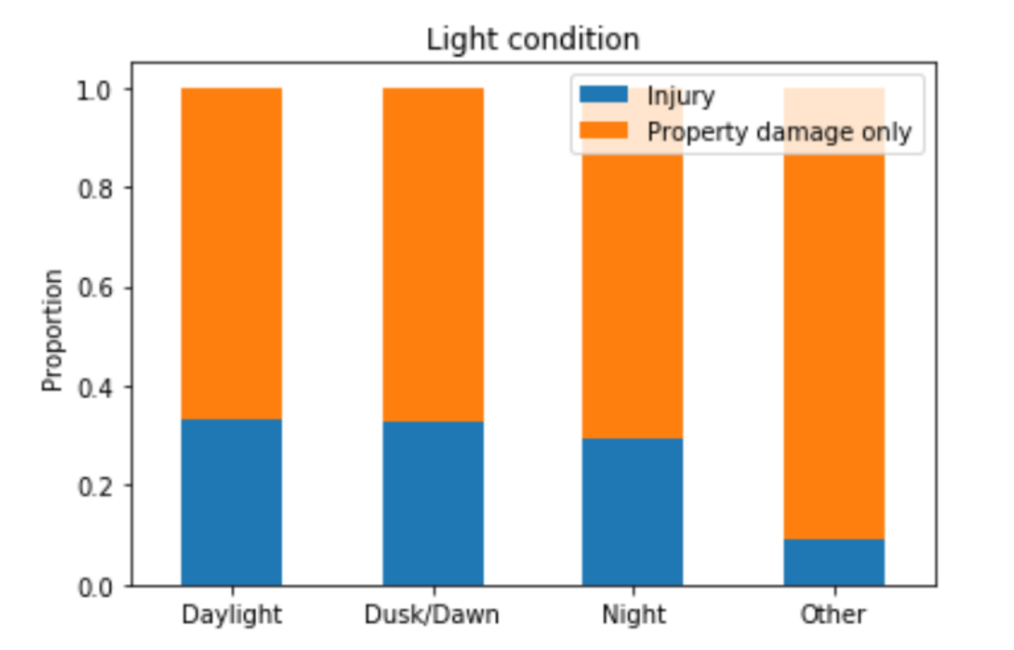
- Data source: Seattle Department of Transportation
  - https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv
- Target variable: Traffic collision severity
  - Property damage only
  - Injury
- Attributes: First available information
  - Involvement of objects/persons: Number of vehicles, number of persons, bicycles, pedestrians
  - Environmental factors: Weather and light condition
  - Infrastructural factors: Road condition, relation to junction

# Methodology – Pre-processing

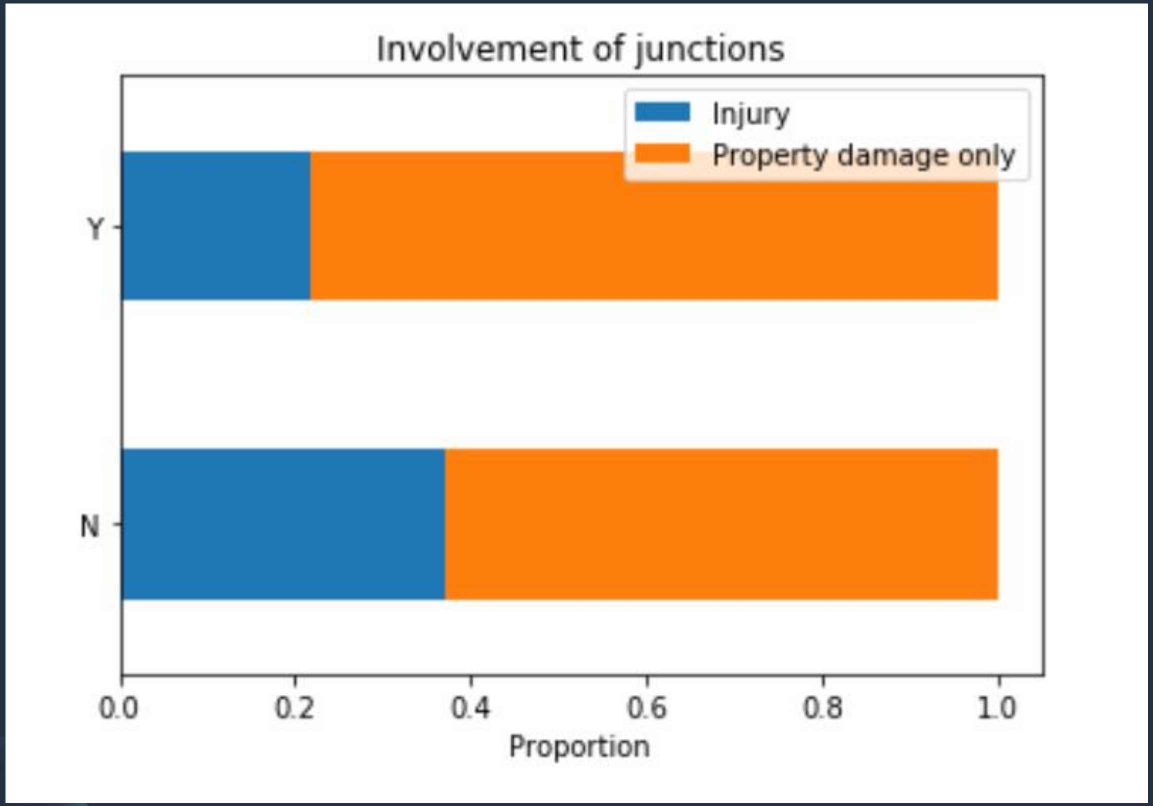
- Missing values: Missing mechanism for variables with rates > 5 % is MAR
- Feature selection
  - Reducing redundancy
  - Excluding identifier
  - Excluding geographical and time data
- Balancing
  - Inclusion in the final model algorithm

# Environmental factors seem not to be crucial for more severe collisions

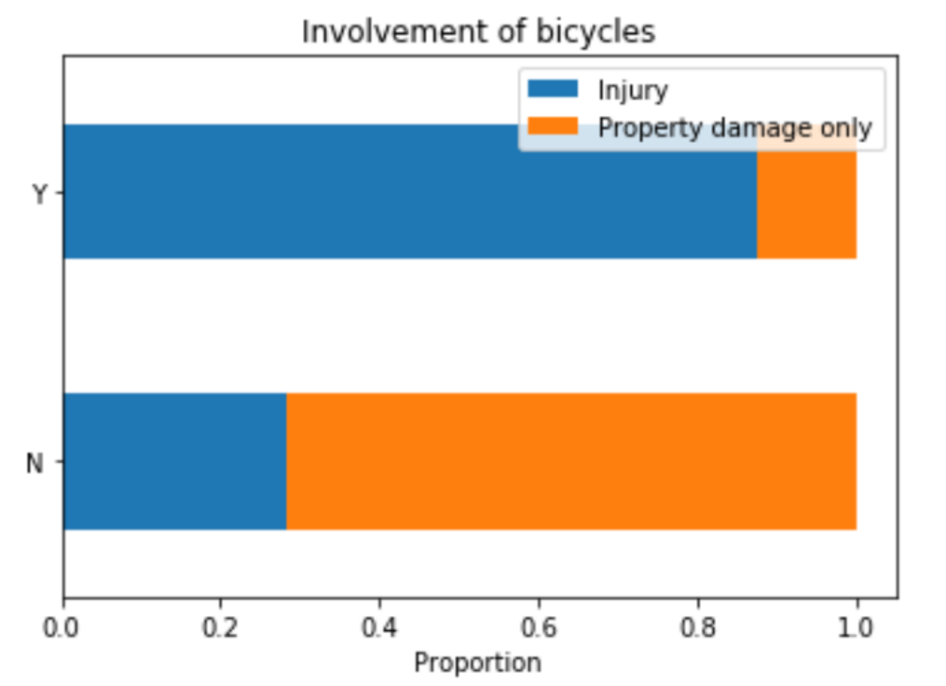


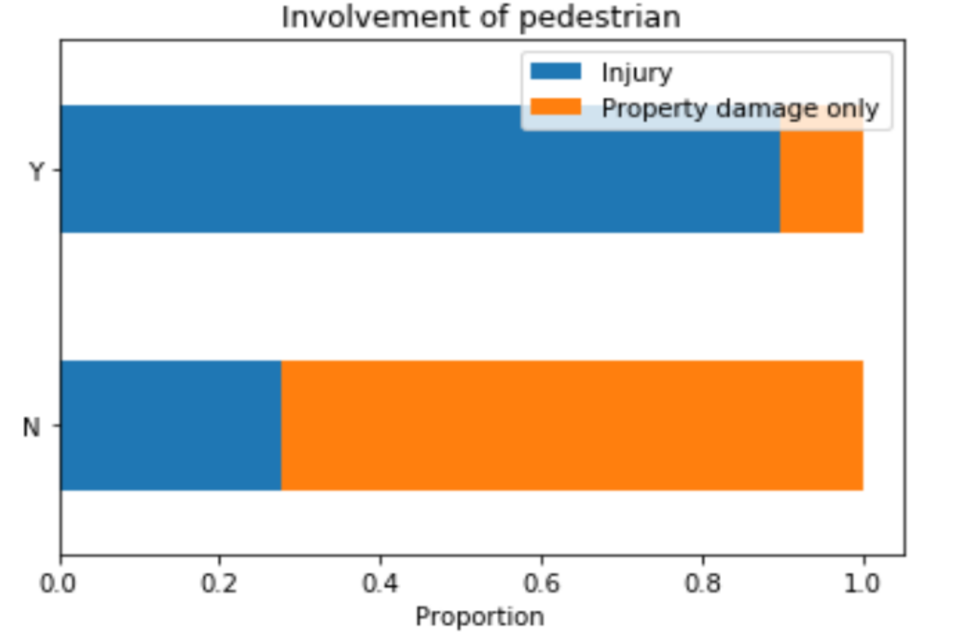


# Junctions are rather related to collisions with property damage only



# Involvement of pedestrians and/or bicycles are crucial for traffic collision severity


Involvement of bicycles


Involvement of pedestrian

# Modelling – Imbalanced dataset

- Balanced modelling shows a higher recall (true positive values) which is crucial for resource distribution
- Balanced modeling shows a lower precision which can be tolerated in the context of the resource distribution

# Modelling – Evaluation metrics of balanced logistic regression

- Log loss 0.60
- Jaccard similarity score 0.65
- Precision for minority category (injury): 0.43
- Recall for minority category (injury): 0.65
- F1-Score for minority category (injury): 0.52

# Need for predicting traffic collision severity - Conclusions

- Attributes can be used as targets for preventive strategies
- Resources can be allocated in the context of traffic collision severity
- Model needs to be evaluated and finalized as prediction performance is not as good as needed for the objective of re-distribute health care resources
- Only preliminary results!