Accident prediction based on certain parameters

# Introduction

At this moment, solutions based on technology are becoming overwhelmingly popular. Some examples are autonomous transportation, Neuroscientific solutions to multiple brain-related problems, etc.

In my following analysis, I wanted to focus on accidents based on vehicles who are driven by humans. While autonomous vehicles could help humans limit traffic accidents and reduce congestion, there is still a lot of fundamental psychological problems to be solved. (for instance, the trolley problem.)

In order to help manufacturers of autonomous vehicles, I wanted to do my part in this technological change and analyze how certain parameters have an impact on the severity of injury when having an accident. Take into account that this report does not predict whether an accident will happen, it only predicts on what the severity is of an accident when an accident would happen.

Why is this important?

When able to predict whether an accident in certain conditions will have serious consequences, autonomous vehicles could take this information into account and take measures. For instance by braking better or not enabling certain vehicle stances.

In Belgium, there are 52 deaths/million/year related to car accidents. When Autonomous vehicles will become widespread and consider the possibility of serious accidents when driving, I hope that the amount of people being impacted by vehicle accidents will be lower.

# Data

Data was found on the statbel.fgov.be which is the statistical service of the Belgian Government. Link to the data is the following: <https://statbel.fgov.be/nl/open-data/verkeersslachtoffers-2019>. The dataset is called TF\_ACCIDENTS\_2019 with the following characteristics:

DT\_DAY: Day of the year (date)

DT\_HOUR: Hour of the day (integer)

CD\_DAY\_OF\_WEEK: Day of the week (integer)

TX\_DAY\_OF\_WEEK\_DESCR\_FR: French description of the week (string)

TX\_DAY\_OF\_WEEK\_DESCR\_NL: Dutch description of the week (string)

CD\_BUILD\_UP\_AREA: Numerical description of place of accident (integer)

TX\_BUILD\_UP\_AREA\_DESCR\_NL : Dutch description of the place of the accident (string)

TX\_BUILD\_UP\_AREA\_DESCR\_FR: French description of the place of the accident (string)

CD\_COLL\_TYPE: Numerical description of the collision type. (integer)

TX\_COLL\_TYPE\_DESCR\_NL: Dutch description of the collision type (string)

TX\_COLL\_TYPE\_DESCR\_FR: French description of the collision type (string)

CD\_LIGHT\_COND: Numerical description of the light conditions (integer)

TX\_LIGHT\_COND\_DESCR\_NL: Dutch description of the light conditions (string)

TX\_LIGHT\_COND\_DESCR\_FR: French description of the light conditions (string)

CD\_ROAD\_TYPE : Numeric road type (integer)

TX\_ROAD\_TYPE\_DESCR\_NL: Dutch description of the road type (string)

TX\_ROAD\_TYPE\_DESCR\_FR: French description of the road type (string)

CD\_MUNTY\_REFNIS: Municipality code (integer)

TX\_MUNTY\_DESCR\_NL: Dutch description of the municipality (string)

TX\_MUNTY\_DESCR\_FR: French description of the municipality (string)

CD\_DSTR\_REFNIS: District code (integer)

TX\_ADM\_DSTR\_DESCR\_NL: Dutch description of the district (string)

TX\_ADM\_DSTR\_DESCR\_FR: French description of the district (string)

CD\_PROV\_REFNIS: provincial code (integer)

TX\_PROV\_DESCR\_NL: Dutch description of the provincial (string)

TX\_PROV\_DESCR\_FR: French description of the provincial (string)

CD\_RGN\_REFNIS: Regional code (integer)

TX\_RGN\_DESCR\_NL: Dutch description of the regional code (string)

TX\_RGN\_DESCR\_FR: French description of the regional code (string)

MS\_ACCT: Boolean stating whether an accident has happened (1 is true). (Boolean)

MS\_ACCT\_WITH\_DEAD: Boolean stating whether a deadly accident has happened (1 is true). (Boolean)

MS\_ACCT\_WITH\_DEAD\_30\_DAYS: Boolean stating whether an accident has led within 30 days to a death (1 is true). (Boolean)

MS\_ACCT\_WITH\_MORY\_INJ: Number of people who have minor injuries. (integer)

MS\_ACCT\_WITH\_SERLY\_INJ: Number of people who have major injuries. (integer)

MS\_ACCT\_WITH\_SLY\_INJ: Number of people who have severe injuries. (integer)

# Methodology

In short, the steps done to conclude my findings are the following:

1. Imported data from downloaded data found from https://statbel.fgov.be/nl/open-data/verkeersslachtoffers-2019
2. Analysis of data
   1. Checks for correlation. (Pearsons; spearman; kendall; phi)
   2. Missing values
   3. Cardinality
3. Cleaning of data
   1. Drop empty rows
   2. Delete columns with high correlation to other columns
   3. Every parameter should have the correct datatype
   4. Encode variables to numerical
4. Linear regression used because of the predictive power of Linear regression
   1. Train\_test\_split
   2. R^2 calculation
   3. T-test
   4. VIF\_calculation

# Result

The model:

[typeofaccidentinjury] = constant + [Day]+ [collisiontype]+[areadescription]+[Lightcondition]+[areadescription]+[roadtype]

Results from Linear regression only have explanatory power of 0.014.

Based on the t-test and cut-off point 0.05; the explanatory parameters are alle significant and help explain the [typeofaccidentinjury], yet they only make up a small percentage of explaining [typeofaccidentinjury]

# Discussion

Based on the explanatory power of the model; the model is not explaining a big part of [typeofaccidentinjury]. The model cannot be used as an entity itself (as for this; we only have explanatory power of 1.4 %.

Further research needs to indicate important values which indicate the [typeofaccidentinjury], I’m thinking then of speed at time of impact, security indicator for the car, etc.

The model discussed in this report only supplements another model and cannot be used by itself.

# Conclusion

This report introduces a additional model to add explanatory power in order to predict the type of accident injury whenever an accident happends.