CAR ACCIDENTS IN SWITZERLAND: A REVIEW OF LEADING FACTORS

IBM Data Science Professional Certificate – Capstone Project

August 2020

Cédric M. G.

INTRODUCTION

- Car accidents are still counted in thousands in Switzerland in 2020
- The direct and indirect consequences of such events (injuries, death, psychological damages, material damages, etc.) are sizeable
- Since 1992, the Swiss Federal Statistics Office (OFS) is collecting data on car accidents country-wide and making such information available to the public. This analysis will leverage this data

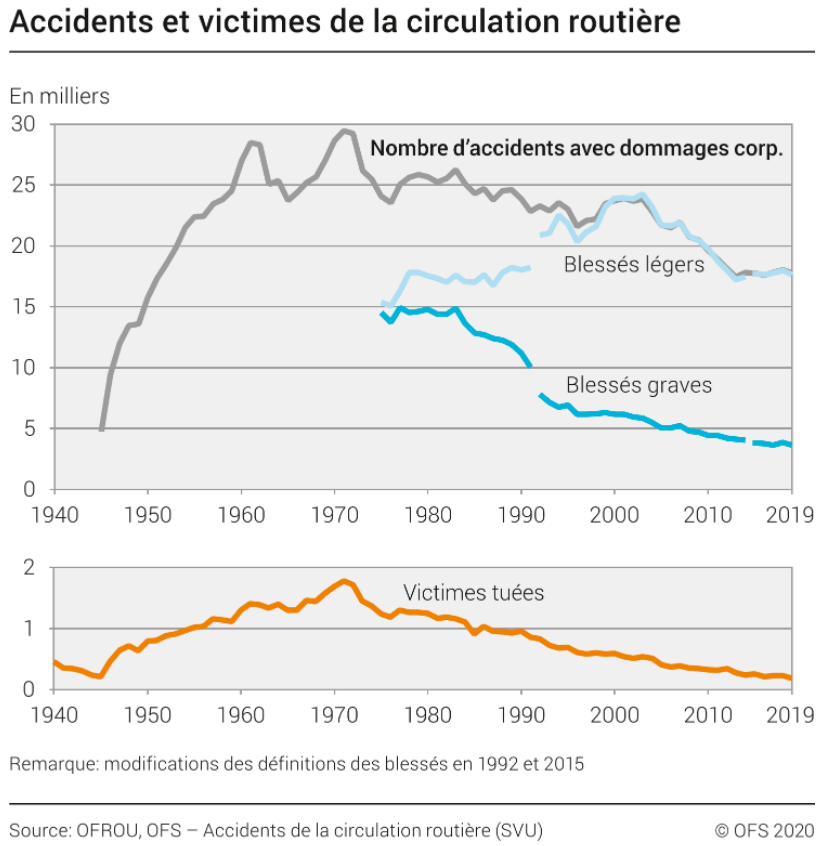
OBJECTIVE AND EXPECTED OUTCOMES

- The objective is to explore a year 2019 dataset from the Swiss Federal Statistics Office (OFS) and determine what are the key factors that drive the outcome of an accident for the involved car(s)' passengers: light injuries, severe injuries, fatal outcome
- The outcomes of this analysis can be used as a prescriptive tool to :
  - Have the appropriate medical emergency resources allocated for the times, locations and circumstances when accidents are most likely to occur, with a particular emphasis on the severe and life-threatening cases.
  - Design prevention measures based on those accident factors identified as having the largest influence on accident outcomes.
- The interests of this study are as follows
  - Being able to reduce injuries and deaths through prevention campaigns
  - Being able to allocate medical emergency resources more efficiently
  - Reducing the suffering and deaths of the thousands of people affected by road accidents.
  - Reducing the economic cost of accidents for society as a whole, whether materially or in terms of human resources

DATA SOURCE

- The data source is the Swiss Federal Statistical Office's database on car accidents.
- The data spans over 1992-2019, but we will focus here on 2019 for simplicity purposes.
- Over 17'000 accidents in Switzerland in 2019, a large enough dataset for our purpose.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



DATASET VARIABLES

- **Variables in the dataset**
- Types of accidents : TYPE_ACCIDENT
- Type of road: TYPE_ROAD
- Severity of the accident: SEVERITY (the dependent variable)
- Month of the accident: MONTH
- Day of the week: DAY
- Time of the accident: TIME_ACCIDENT

DUAL METHODOLOGY

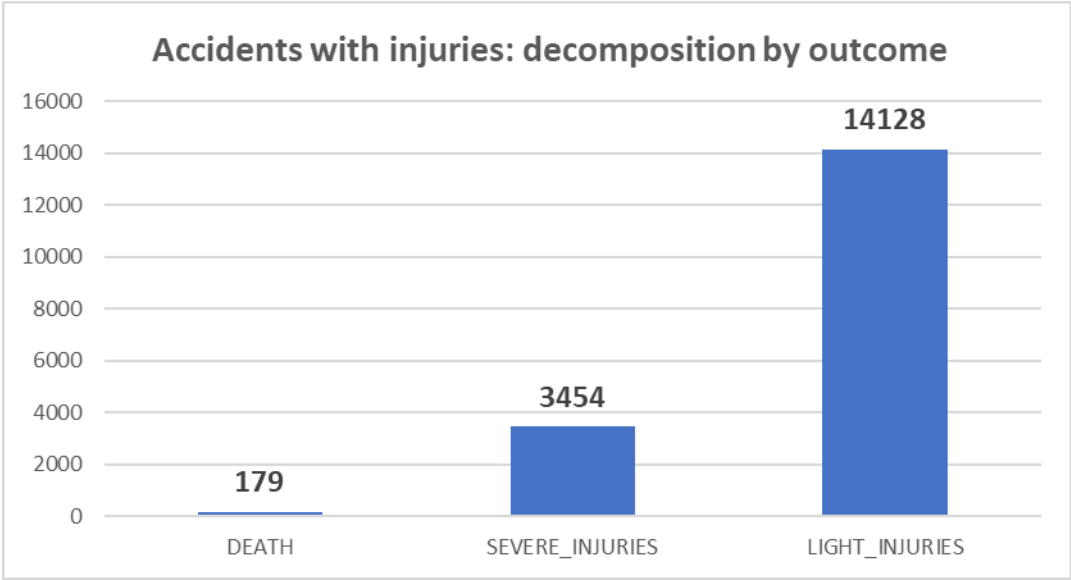
- Traditional analysis methods: explaining the accidents via the independent variables through simpler statistics
- Machine Learning: typical classification problem. Here, use of Decision Tree, suitable for non-numerical classifiers.

KEY RESULTS FROM THE REPORT

- The summer is a critical period.
- Accidents seem to overly take place in the late afternoon and evenings, possibly when people come back from work.
- Skidding is still the number one factor of accidents, suggesting that people get distracted and/or that their judgement of road conditions and vehicle physics should be improved.
- In the weekends, there is an over-proportionate number of skidding-generated accidents, suggesting that measures need to be taken in prevention or those, f.ex. in the areas of alcohol consumption and judgement alteration.

KEY RESULTS - FIGURES

- Large majority of injuries vs. deaths but still too many of them.



KEY RESULTS – FIGURES (CONT'D)

- Skids are a real concern as they are the leading cause and by quite a margin.
- Back collisions are very prevalent and possibly due to inattention and fatigue during commutes.
- Pedestrians are still victims of accidents, especially of course in urban areas and led by an ever-increasing pace of driving in those conditions.

Accident Causes	Death	Severe injuries	Light injuries	Grand Total
ANIMAL	1	27	41	**69**
BACK	7	284	3'281	**3'572**
FRONTAL	14	159	471	**644**
INTERSECTION	10	139	691	**840**
OTHERACC	1	39	115	**155**
OVERTAKE	13	474	2'516	**3'003**
PARKING	3	38	213	**254**
PEDESTRIANS	35	551	1'546	**2'132**
SKID	90	1'499	4'208	**5'797**
TURNING	5	244	1'046	**1'295**
Grand Total	**179**	**3'454**	**14'128**	**17'761**

KEY RESULTS – FIGURES (CONT'D)

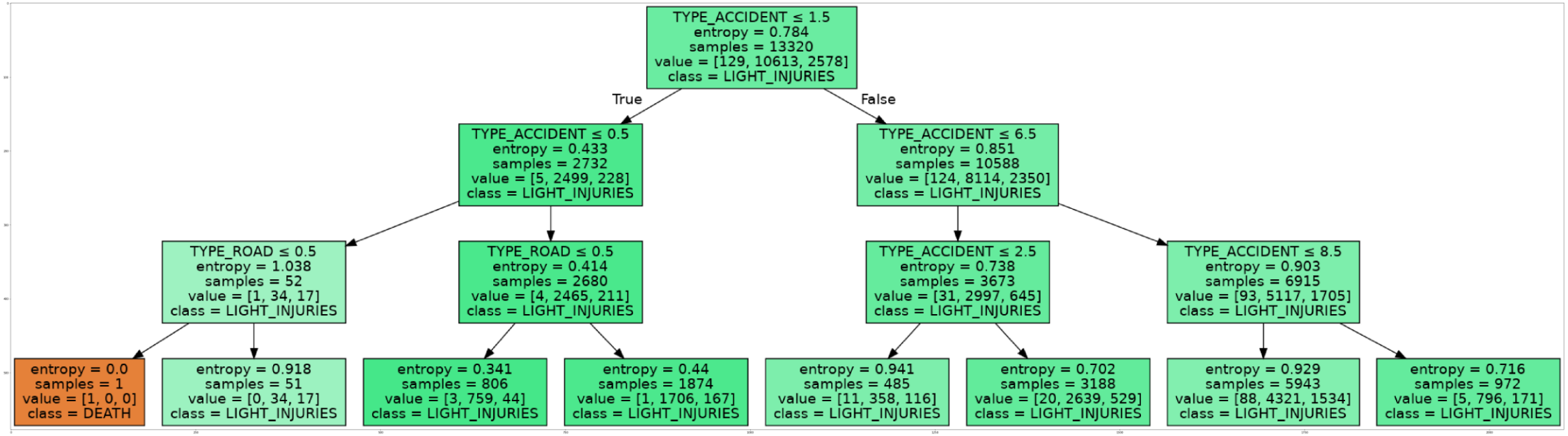
- Most accidents take place in the afternoon (noon-6pm), and very few during the night.
- Impact of work-home commuting.
- Fatigue, distractions.
- Quite a few accidents in the morning, possibly with the same reasons (commuting).

Accident time	Death	Severe injuries	Light injuries	Grand Total
AFTNOON	77	1'528	6'352	7'957
EVENING	38	748	2'847	3'633
MORNING	45	935	4'120	5'100
NIGHT	19	243	809	1'071
TOTAL	**179**	**3'454**	**14'128**	**17'761**

KEY RESULTS – MACHINE LEARNING

- Good accuracy of 79% on the tree
- HOWEVER, it appeared quickly that the classifiers are not able to perform their work as initially thought. 79% is barely the proportion of light injuries, meaning the model is only able to re-bucket these in light injuries. It is not able to allocate severe injuries and deaths to specific buckets.
- Model results are not conclusive and driven by the dummy-zation of the categorical variables. Example: the 10 types of accidents are coded from 1 to 10. This makes the algorithm useless to a large extent.
- Recommendation to drop these results and stick to the traditional analysis methods, in that case.

DECISION TREE - ILLUSTRATION



CONCLUSION

In conclusion, the reader will have taken note that key factors. of accidents in Switzerland over the year 2019 have been as follows :

- The summer is a critical period.
- Accidents seem to overly take place in the late afternoon and evenings, possibly when people come back from work.
- Skidding is still the number one factor of accidents, suggesting that people get distracted and/or that their judgement of road conditions and vehicle physics should be improved.
- In the weekends, there is an over-proportionate number of skidding-generated accidents, suggesting that **measures need to be taken in prevention or those, f.ex. in the areas of alcohol consumption and judgement alteration**.

Measures should therefore be implemented in priority against these factors. Additionnally, it would be commendable to organize the allocation of emergency medical resources and personal for those times identified as critical, such as for instance the summer, the weekdays and the afternoons.

THANK YOU