Road to Safety: Traffic Accident Analysis

▼ Source

Desights AI | Road to Safety: Traffic Accident Analysis

This challenge leverages the power of artificial intelligence to enhance road safety and understanding of traffic accidents. The competition invites participants from diverse backgrounds,

ᴐ https://desights-ai.translate.goog/challenge/18?_x_tr_sl=en& _x_tr_tl=fr&_x_tr_hl=fr&_x_tr_pto=sc&_x_tr_hist=true

1. General Trends

What are the overall trends in traffic accidents, fatalities, and serious injuries in Catalonia from 2010-2021?

The number of victims involved in car accident reduced from 3200 victim in 2010 to 2200 victims in 2021.

The number of Victims significantly decreased in 2019, and rapidly increased back in 2021 but didn't reached pre COVID levels.

Number of Victims involved in accidents per Total Victims, Fatalties, Serious Injuries and Light Injuries (Line)
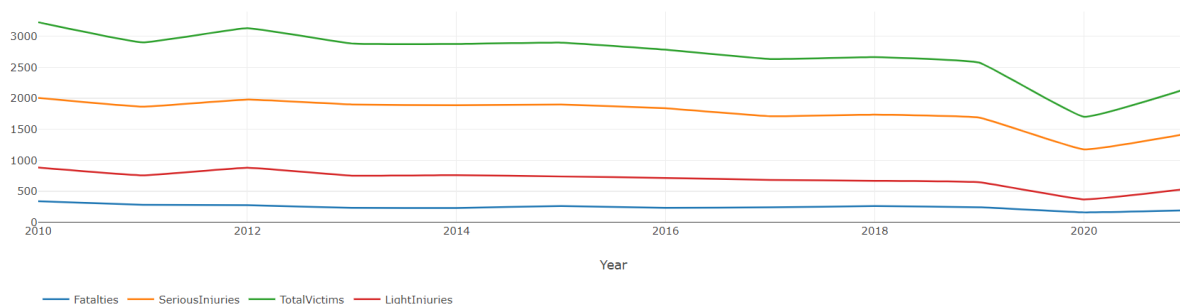
Number of Victims involved in accidents per Total Victims, Fatalties, Serious Injuries and Light Injuries (Scatterplot)
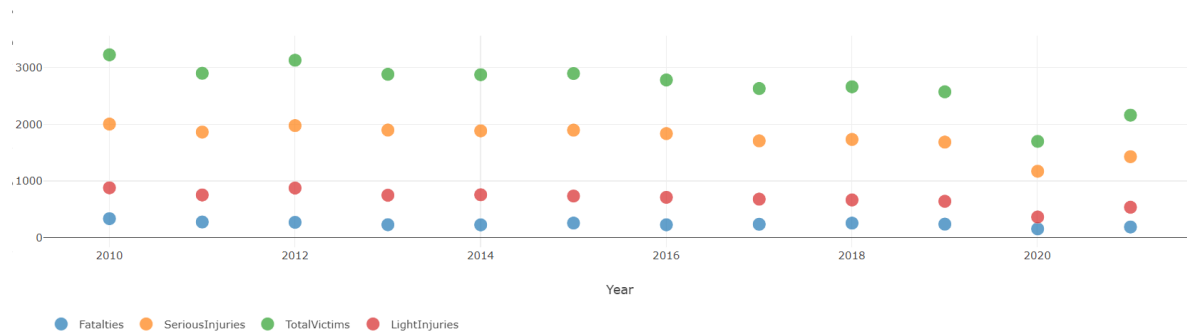

2. Accident Characteristics

What common characteristics (time of day, type of road, etc.) are observed in the most severe accidents?
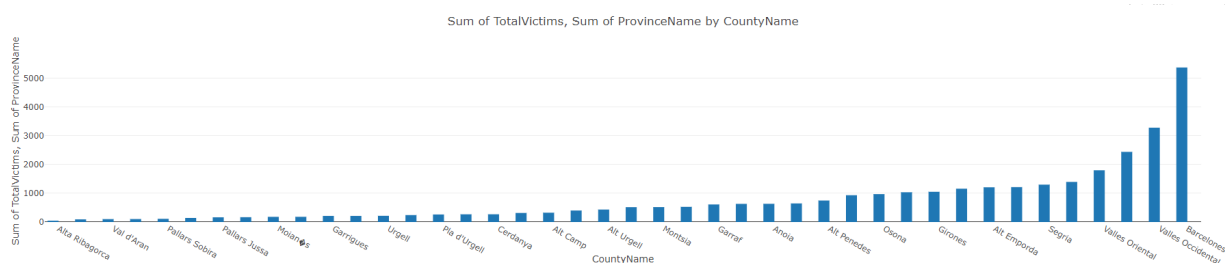
3. Geographical Insights

Which municipalities or counties in Catalonia have the highest incidence of traffic accidents? How does this correlate with population density or road network characteristics?

Number of accidents per Municipality
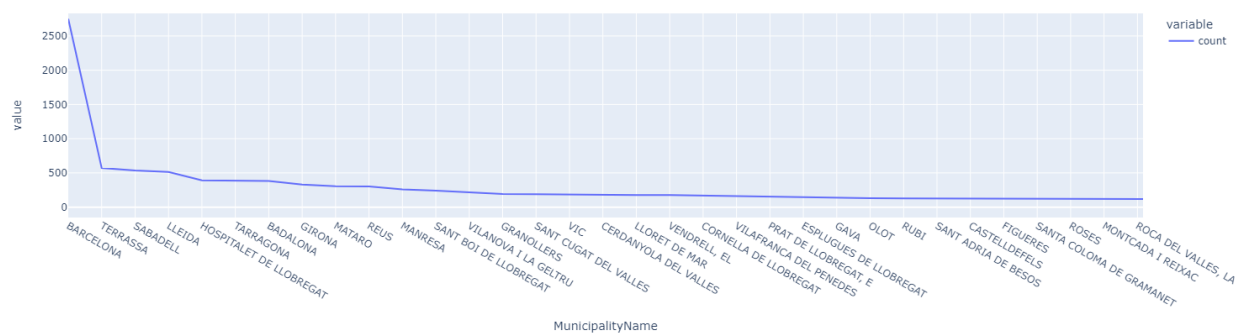
Number of accidents per County Name

Sum of TotalVictims, Sum of ProvinceName by CountyName

Without surprise the province that suffers from roads accident is Barcelona with a number of victims of accidents larger than 1500 victims nearly every years since 2010.

The second and third province most impacted by road accident are Girona and Tarragona with almost 500 hundred victims involved in road accidents every years, nearly 4 times less victims than Barcelona.

Number of accident and per Province
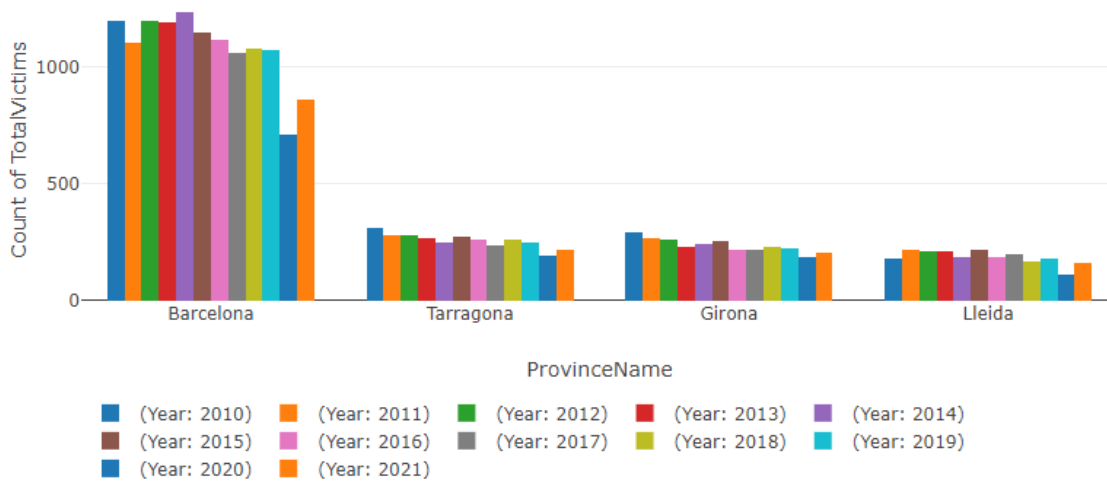
Number of victims of accident per Province

Sum of TotalVictims by ProvinceName
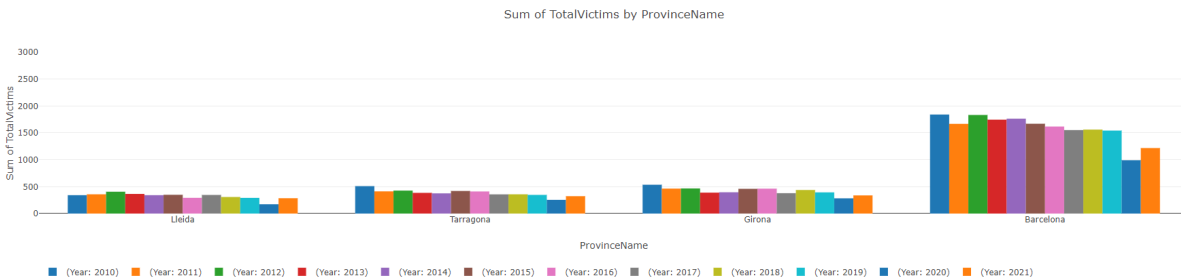
The cities that suffer the most from accident, both in terms of number of accident and number of victims of those are Barcelona, Tarragona, Lleida, Badalona, and Terrassa

Number of accident and per Municipality

Count of TotalVictims by MunicipalityName
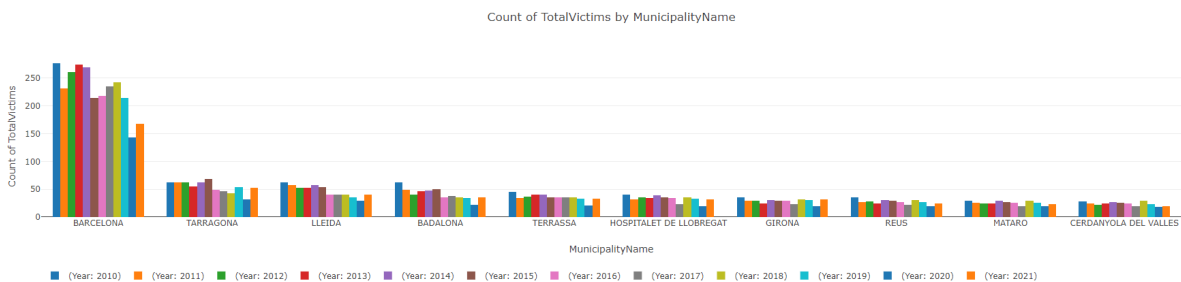
Number of victims involved in accident per Municipality

Sum of TotalVictims by MunicipalityName
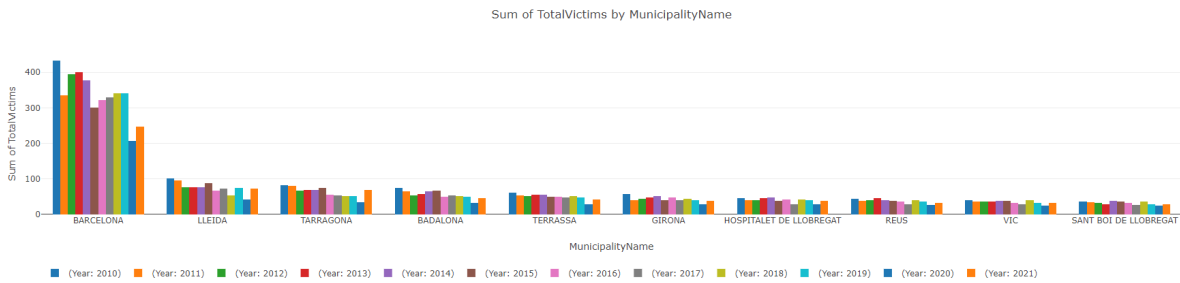
The most impacted Counties are Barcelones, Valles Occidental, Baix Llobregat, Maresme and Tarragones, both in number of accidents and victims.

Number of accident and per County
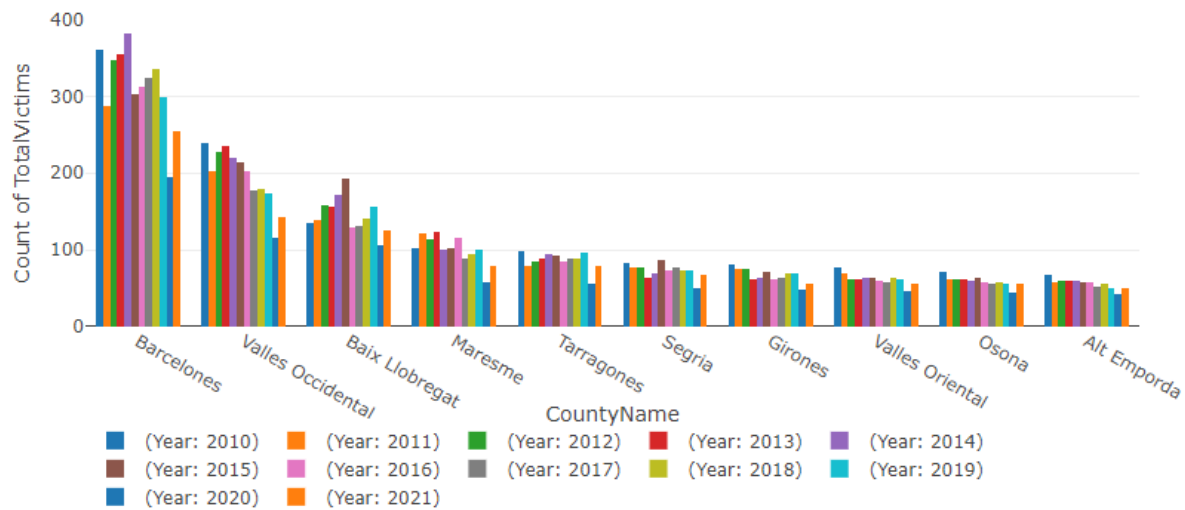
Number of victims involved in accident per County

Sum of TotalVictims by CountyName
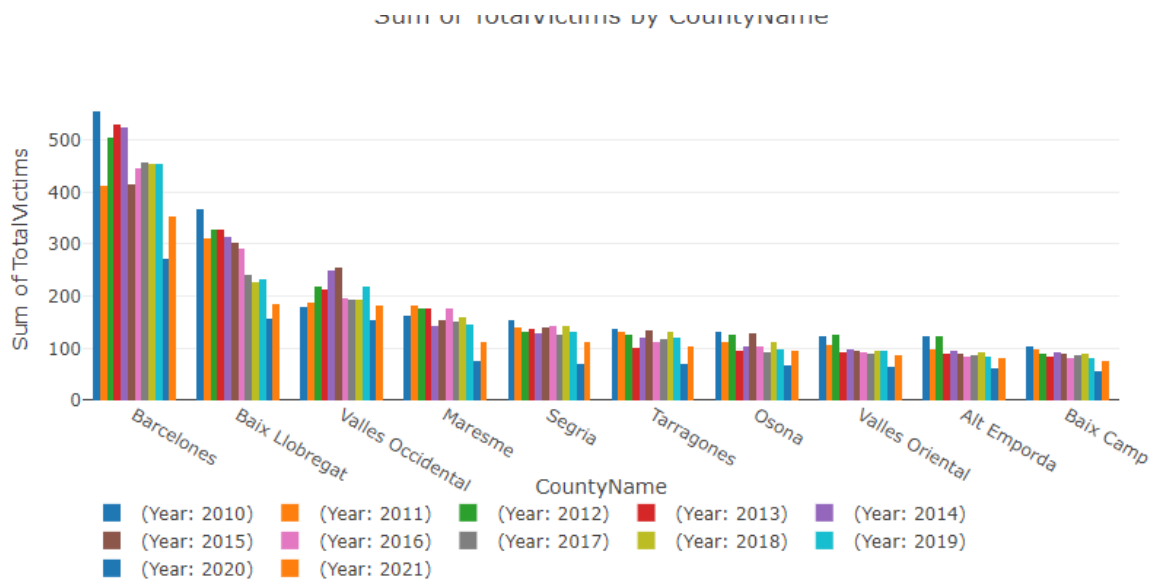
4. Yearly Trends

How have traffic accident patterns (frequency, severity) changed yearly from 2010 to 2021?

Sum of SeriousInjuries, Sum of TotalVictims, Sum of Fatalties by Year
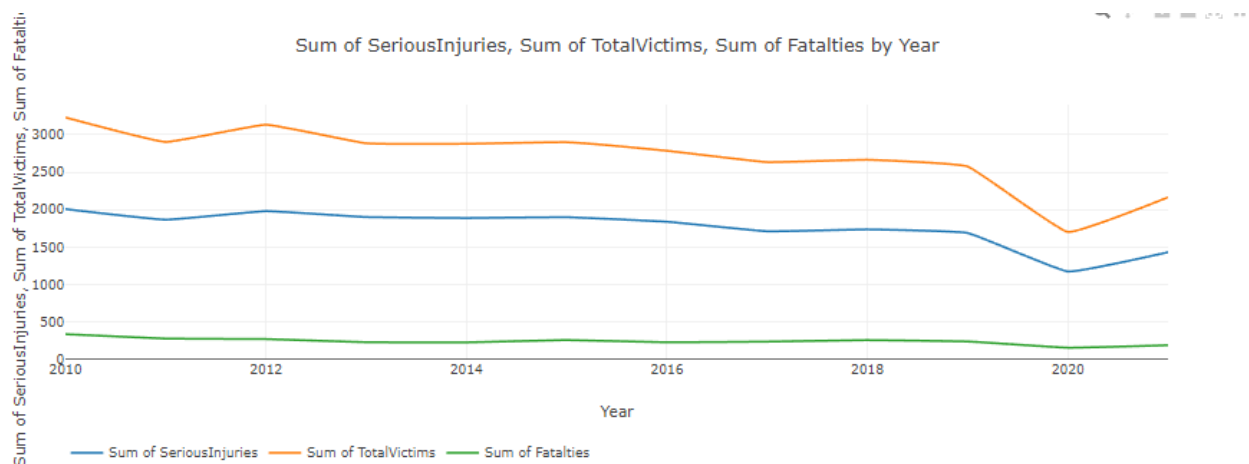
5. Day and Time Patterns

On what days of the week and times of day do most accidents occur? Are there notable differences between weekdays and weekends?

TotalVictims by day_name
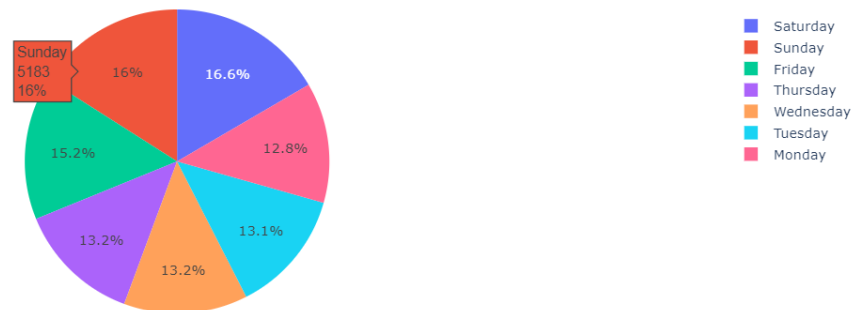
Most of the accident and most of the victims of accident happens on weekend day with respectively:

Friday 2918 big accident, 399 deadly, involving 4216 victims and 718 death

Saturday 2486 big accidents, 436 deadly, involving 4490 victims and 893 deaths

Sunday 2760 big accidents , 429 deadly, involving 4155 victims and 1028 death

Accident on Sunday ar more severe since they are more deadly both in number of deadly accidents and in number of death (1028) than any other day.

One can alsa notice that the number of victims during weekday Tuesday to Thursday are very similar, respectively 4251, 4269 and 4277 victims involved.

Number of victims of accident per weekday
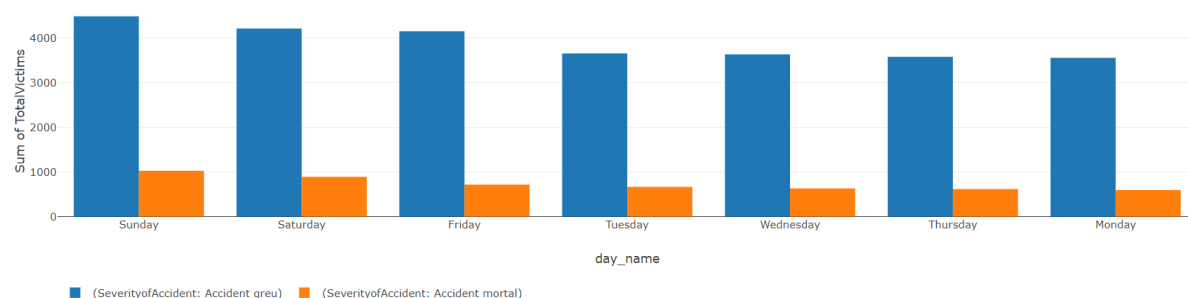
Number of accidents per weekday
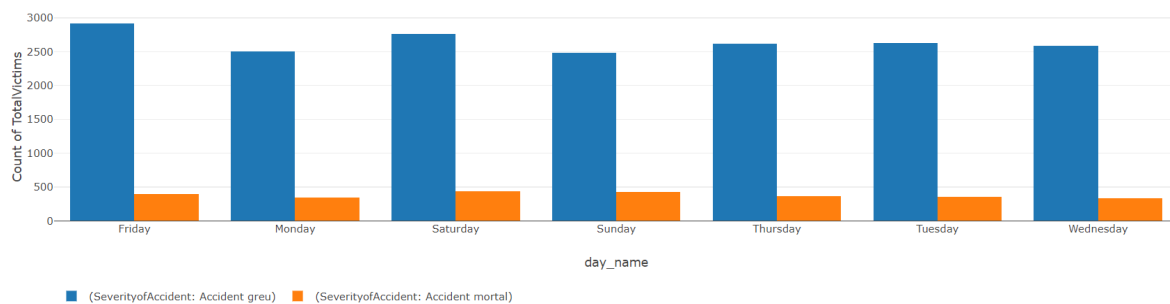
The importance of accidents during the weekend day could be explained by the difference of itineraries, where drivers tend to go to different locations the weekend, compared to weekdays where usual work-home are taken.

Dangerous driving behavior related to substance abuse or speed risk taking, might also occur more frequently than weekday.

6. Environmental Impact

How do different weather conditions affect the likelihood of accidents? Is there a correlation between visibility, road conditions, and accident severity?

7. Road and Traffic Features

What impact do road features (such as speed limits and road types) and traffic density have on the occurrence of accidents?

We can observe that most of the accident occurs in "carretera convencional with speed limit of 100km/h where occurs 6 times more accident, and involving 8 times more victims than road where speed limit is below or above 80 km/h

Number of accidents per Road Speed Limit and Type of Road

	RoadSpeedLimit	TypeofRoad	TotalVictims
61	100.0	Via urbana(inclou carrer i carrer residencial)	6344
60	100.0	Carretera convencional	4782
71	999.0	Carretera convencional	1017
16	30.0	Via urbana(inclou carrer i carrer residencial)	746
29	50.0	Carretera convencional	668
47	80.0	Carretera convencional	658
35	60.0	Carretera convencional	503
21	40.0	Carretera convencional	483
22	40.0	Via urbana(inclou carrer i carrer residencial)	470
56	100.0	Altres	386

Number of victim per Road Speed Limit and Type of ROAD

	RoadSpeedLimit	TypeofRoad	TotalVictims
60	100.0	Carretera convencional	9192
61	100.0	Via urbana(inclou carrer i carrer residencial)	8421
71	999.0	Carretera convencional	1362
47	80.0	Carretera convencional	1264
29	50.0	Carretera convencional	1028
16	30.0	Via urbana(inclou carrer i carrer residencial)	954
35	60.0	Carretera convencional	952
21	40.0	Carretera convencional	751
22	40.0	Via urbana(inclou carrer i carrer residencial)	635
56	100.0	Altres	558

8. Vehicle Types and Accident Severity

Does the involvement of specific types of vehicles (like heavy trucks and motorcycles) correlate with more severe accidents?

The "Motorcycle" factor does not seem to have a significant impact on number of accident, since there are only few accident with the biggest number of Motorcycles, for examples :

- there are 2 accident with 7 motorcycles involved

- There 2 (1+1) accidents with 6 motorcycles involved

However, the number of accident rise significantly when it involves "only" 2 motorcycles, with hundred of accidents.

N**umber of accident per number of motorcycles involved**

	Motorcycles	LightVehiclesInvolved	HeavyVehiclesInvolved	OtherUnitsInvolved	TotalVictims
96	7	1	0	0	2
94	6	0	0	0	1
95	6	1	0	0	1
91	4	0	0	0	3
92	4	0	1	0	1
93	4	1	0	0	1
86	3	0	0	0	16
87	3	1	0	0	8
88	3	1	1	0	1
89	3	2	0	0	4
90	3	3	0	0	1
75	2	0	0	0	235
76	2	0	0	1	1
77	2	0	1	0	8
78	2	1	0	0	77
79	2	1	0	1	1
80	2	1	1	0	3
81	2	2	0	0	11
82	2	2	1	0	1
83	2	3	0	0	5

The same observations can be made for accidents with Light Vehicles Involved since there are very few accidents with those Vehicles involved.

However, the number of accidents rises significantly when there are 4 Light Vehicule Involved

Number of accident per number of Light Vehicule Involved

	Motorcycles	LightVehiclesInvolved	HeavyVehiclesInvolved	OtherUnitsInvolved	TotalVictims
54	0	17	4	0	1
53	0	14	1	0	1
51	0	9	1	0	1
52	0	9	2	0	1
49	0	8	0	0	1
50	0	8	1	0	1
74	1	8	0	0	1
85	2	8	2	0	1
46	0	7	0	0	6
47	0	7	0	1	1
48	0	7	1	0	2
73	1	7	0	0	2
44	0	6	0	0	11
45	0	6	1	0	5
41	0	5	0	0	31
42	0	5	1	0	5
43	0	5	2	0	4
72	1	5	0	0	3
38	0	4	0	0	110
39	0	4	1	0	15

The same observations can be made for accidents with Heavy Vehicles Involved since there are very few accidents with those Vehicles involved.

However, the number of accidents rises significantly when there are 2 Heavy Vehicles Involved

Number of accident per number of Heavy Vehicule Involved

|	Motorcycles	LightVehiclesInvolved	HeavyVehiclesInvolved	OtherUnitsInvolved	TotalVictims
37	0	3	16	0	1
30	0	2	6	0	1
22	0	1	5	0	1
11	0	0	4	0	5
21	0	1	4	0	1
29	0	2	4	0	1
54	0	17	4	0	1
10	0	0	3	0	14
20	0	1	3	0	8
28	0	2	3	0	3
36	0	3	3	1	1
8	0	0	2	0	99
9	0	0	2	1	1
18	0	1	2	0	52
19	0	1	2	1	1
27	0	2	2	0	22
34	0	3	2	0	3
35	0	3	2	1	1
40	0	4	2	0	3
43	0	5	2	0	4

9. Temporal Clustering

Are there specific periods (months, years) where accident patterns cluster significantly? What might be the causes for these clusters?
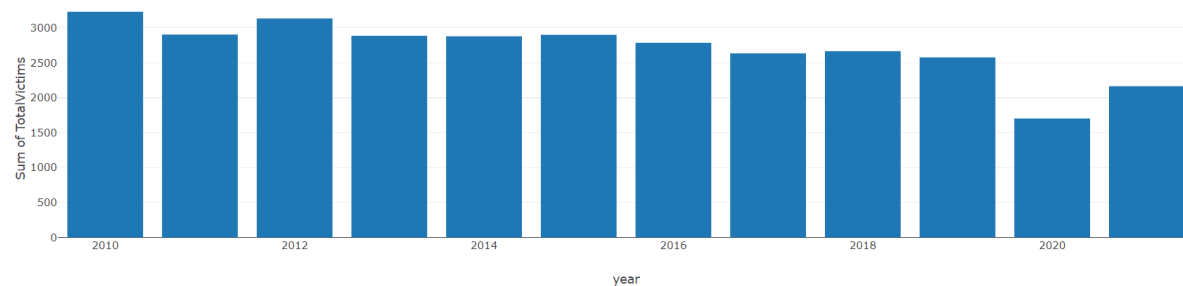
Most important number of victims of car accident occur during summer month (June, July , August).

Since Catalunya is attracting many tourist during summer, it could explain importance of affluence during this period and the importance a car accidents.

Number of victims per month
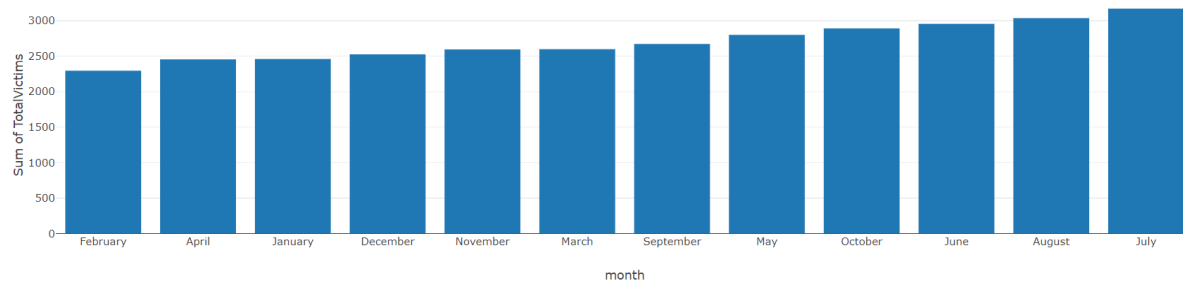
Number of victims per Year

10. Time-Series Forecasting

Based on past trends, create a model to forecast the number of accidents, fatalities, or serious injuries for the upcoming year.

Clearly describe the forecasting model you have developed. This should include the type of model, its structure, and any specific features or techniques it utilizes. Discuss the factors that influenced your decision, such as the model's accuracy, efficiency, suitability to the data characteristics, or its ability to handle the complexities of the dataset.

For time serie forecatsing we choose **Nbeats** model that is implemented in the DARTS framework.

This choice of model is justified by its support for multivariate input as this model accept many feature as input to forecast one output, and its modularity when it come to adjust parameters.

It is a powerful and flexible forecasting model with advantages in terms of flexibility, interpretability, and parallelization.

Last but not least this model architecture was the winning solutions for the M4 forecasting competition

The goal of the challenge is to predict the number of accidents for the upcoming year.

To prepare the date for the forecast, we aggregated the feature by month , not by year. We do not aggregate feature on a yearly basis since it will shorten the dataset for training the model, and obliterate its forecasting abilities.

We also choose to use only numerical features. This choice is a pragmatic one since aggregating categorical data, such as Weather condition or Road conditions takes a considerable amount of time in preprocessing, and might in the end add noise to our model and impact the accuracy of predictions.

We kept the following features since it was aggregatable as numerical (non categorical values):

```
Features used to forecats
Fatalties, SeriousInjuries, LightInjuries, TotalVictims,
UnitsInvolved, PedestrainsInvolved, BicyclesInvolved,
MopedsInvolved, Motorcycles, LightVehiclesInvolved,
HeavyVehiclesInvolved, OtherUnitsInvolved,
UnspecifiedUnitsInvolved,
Target
NbrAccidents
```

The chosen model was a N-Beats model implementation that support multivariate dataset. This model has the advantage of having lookback period of N time and the abilty to perform prediction X time in the future.

N BEATS models architecture
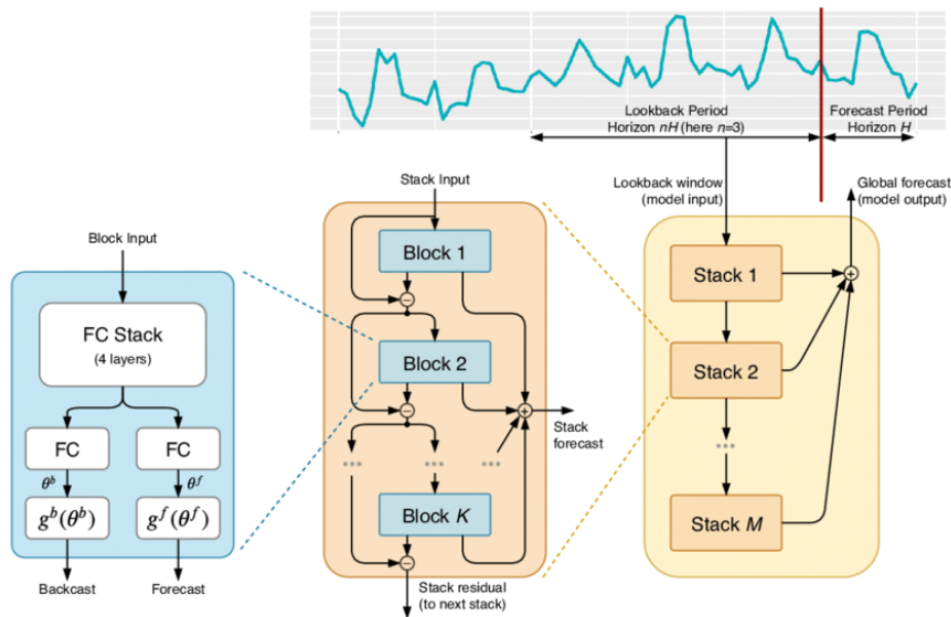
N-BEATS architecture, reproduced from Figure 1 of Oreshkin et al. (2020), with permission.

Chosen metric to evaluate our model are:

- Mean Absolute Error (MAE): which is a measure of how far off, on average, your predictions are from the actual values.For example: If your MAE is 2, it means your predictions are, on average, off by 2 units from the real values.

- Mean Absolute Percentage Error (MAPE) is similar to MAE but expressed as a percentage. It tells you the average percentage by which your predictions deviate from the actual values.If your MAPE is 10%, it means your predictions, on average, deviate by 10% from the true values.

MAE treats all errors equally, while MAPE gives more weight to larger errors relative to the actual values. Lower values for both metrics indicate better performance, with 0 being a perfect score.

We choose to train our model on 100 epoch (iteration) to compare our results, see below:

The predictions of number of accident for the next year is :

- **1183 accident** (model trained on 100 epochs)
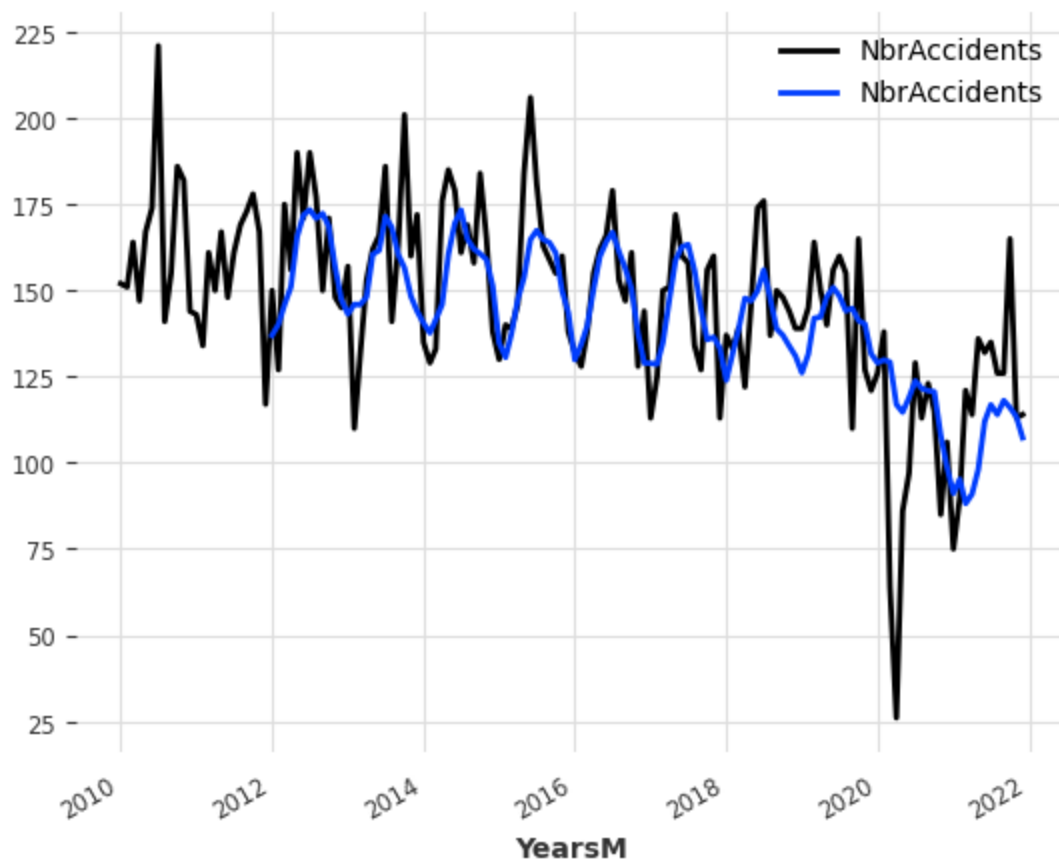
MAE 18.60706032878687
MAPE: 15.149142162048596

- **1142 accidents** (model trained on 50 epochs)
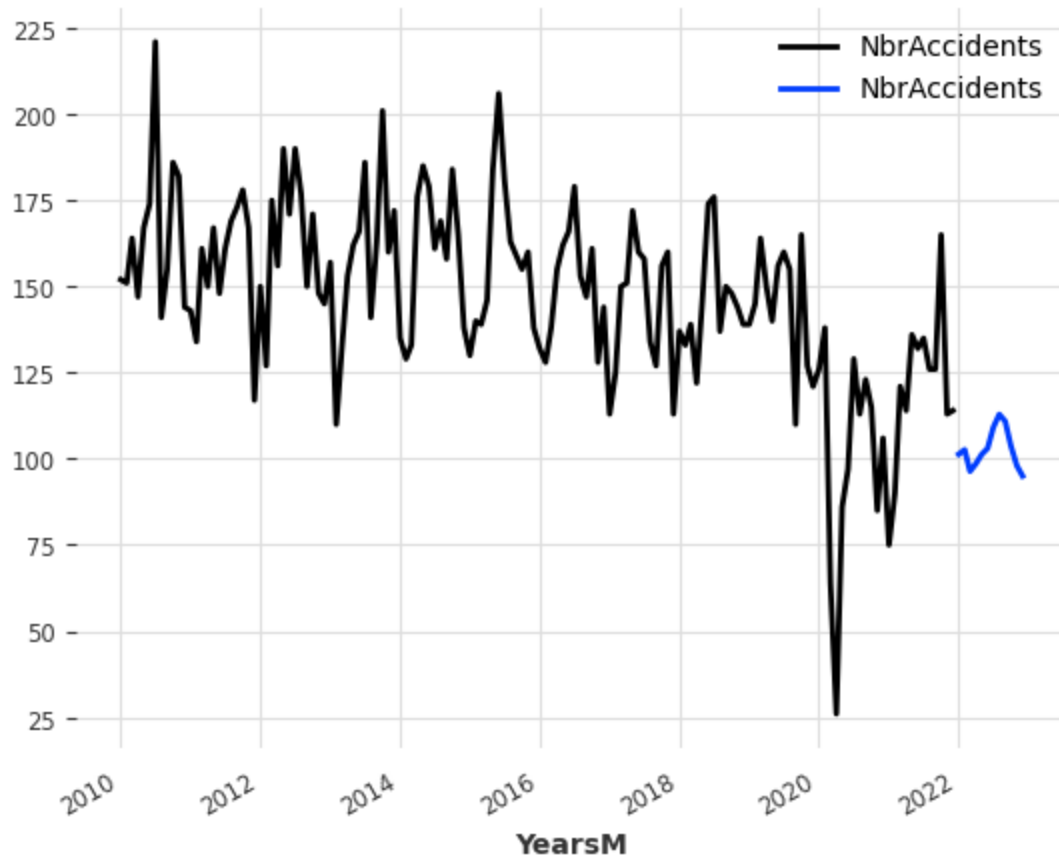
MAE 14.084361689181046
MAPE: 12.87823968264282

Looking at the forecast, it seems that the model trained on 50 epochs outputs more realistic result.
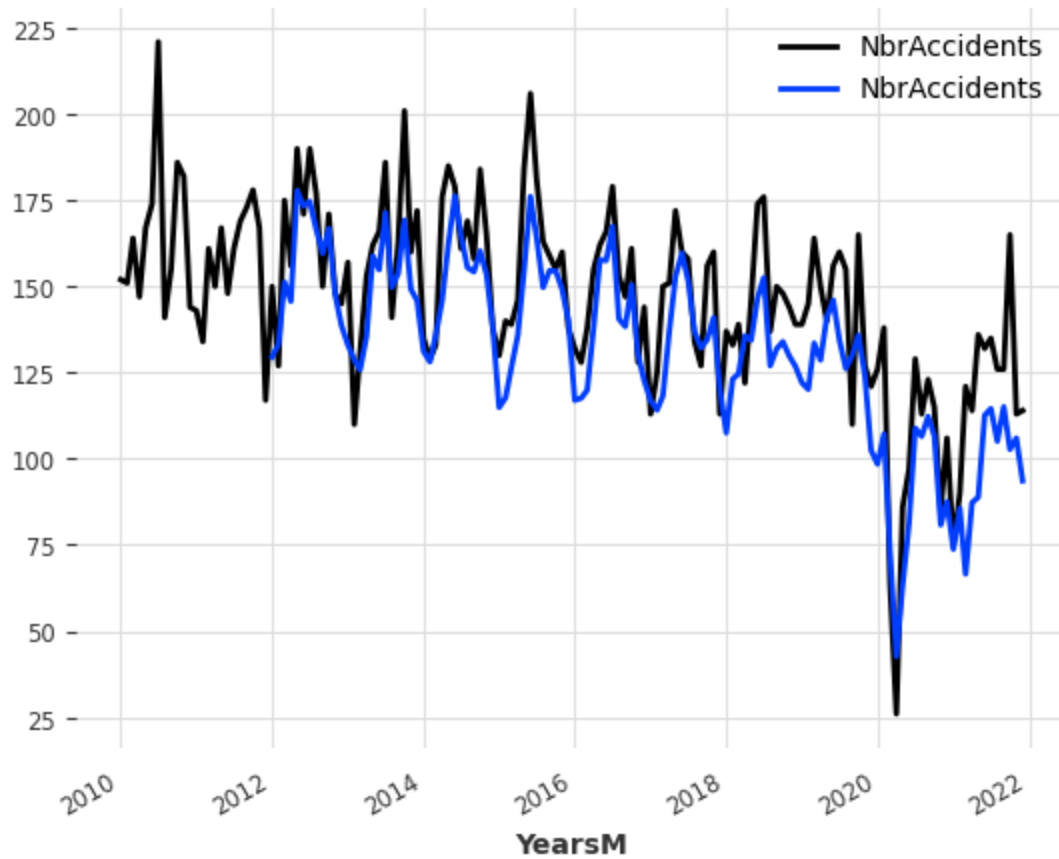
Historical forecast (50 epochs)

Prediction for the next 12 months (50 epochs)

<u>Historical forecast (100 epochs)</u>

Prediction for the next 12 months (100 epochs)