Making Catalonia Safer:
Traffic Accident Analysis

Introduction

This report aims to inform policymakers and stakeholders about the state of traffic accidents and whether the situation of traffic accidents is getting better or worse from 2010-2021in Catalonia. By analyzing the traffic accident data in Catalonia, we hope to identify insights and provide future improvements to reduce traffic accidents and enhance road safety.

Moreover, we look forward to producing innovative visualizations that effectively communicate road accident spatial and temporal patterns. These visual tools will be crucial in understanding and addressing the complexities of road safety. Besides this, the report will provide a forecasting model that aims to predict the likelihood and severity of road accidents in the next years.

We hope that our findings will help drive improvements in road safety measures such as improved traffic regulations as well as infrastructure development.

Key Findings

There are some interesting findings from the analysis of the traffic accident dataset in Catalonia, the most shocking one being that good driving conditions, such as weather conditions, lighting conditions and road conditions, did not work to reduce the occurrence of a traffic accident.

Besides this, it is well known that driving at night might be more dangerous than driving during the day or evening, however, the dataset shows that more traffic accidents actually happen during the day and evening as opposed to late nights and early mornings.

After a thorough study of the trends in number of traffic accidents, fatalities and serious injuries, we can confidently say that the state of traffic accidents is improving and will continue to improve in the future. This improvement can be seen in the number of traffic accidents as well as the number of fatal traffic accidents.

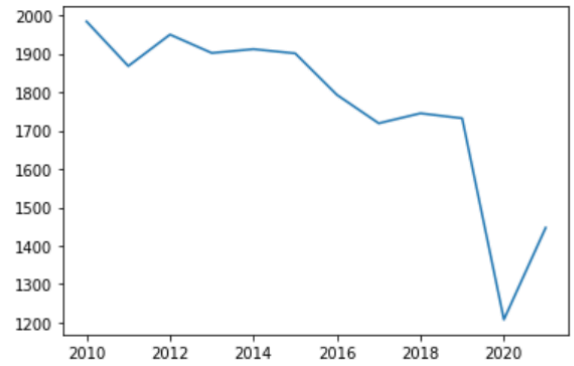
Not surprisingly, we saw that COVID had a massive impact on the number of traffic accidents as there was less traffic congestion in general. In addition, we saw a decrease in overall traffic accidents after peak COVID times as the effect of the pandemic on traffic activity was still present.

To provide some context, the data visualizations and graphs and labelled using Figure x.x and the Figure is labelled by order of which evaluation criteria it is answering. We did this to avoid using titles for every distinct evaluation criteria, so it would read more like a typical report and not a question/answer format. For example, Figure 2.1 would address question 2, Accident Characteistics.

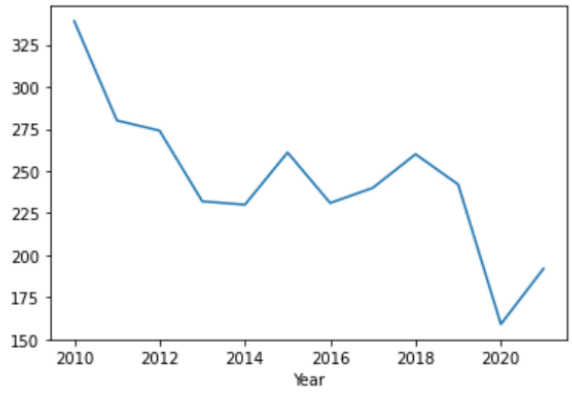
In-depth Analysis

Total Number of Traffic Accidents, Fatalities and Serious Injuries

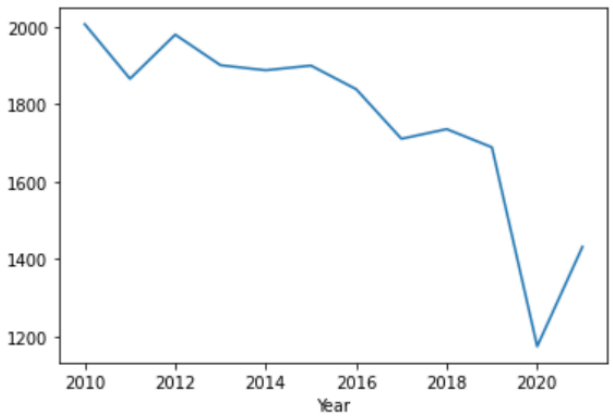
Number of Traffic Accidents (Figure 1.1)



Number of Fatalities by Year (Figure 1.2)



Number of Serious Injuries by Year (Figure 1.3)

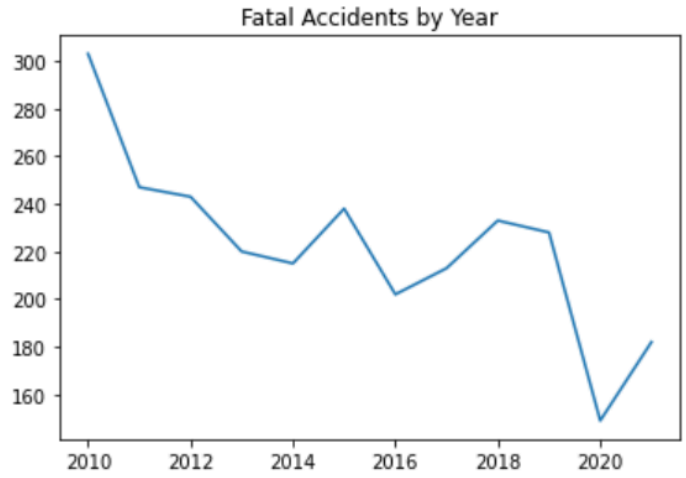


Overall, there is a downward trend of traffic accidents (Figure 1.1), fatalities (Figure 1.2) and serious injuries (Figure 1.3). The exception here is in the last year where the 3 numbers increased from the last year. Thinking about what was happening in the world at the time, you could deduce that the exceptionally low numbers in 2020 were caused by COVID-19 preventing people from even leaving their homes, which reduced the likelihood of accidents happening. This finding shows that congestion plays a big part in traffic accidents.

The 3 statistics would be very correlated as every traffic accident has a certain probability of a fatality or serious injury happening, so in general, lower traffic accidents would result in lower fatalities and serious injuries.

Severity of Accident

Figure 2.1



For the most severe accidents, we filtered the dataset to rows where the “Severity of Accident” would be “Fatal Accident” or “Accident mortal”. Looking at each of the columns, we started to discover some common characteristics among this subset of 2673 fatal accidents. The numbers are shown in Figure 2.1. The data shows that there is a steady decline of the number of severe accidents, excluding the year of COVID in 2020.

Speed Limit for Fatal Accidents

Figure 2.2

Road Speed Limit	
100.0	1791
80.0	182
50.0	127
60.0	107
40.0	94
30.0	74
70.0	68
90.0	38
999.0	24
120.0	22
20.0	8
110.0	3

Looking at the type of roads on which these accidents occurred (Figure 2.2), we can see that a majority of them happened on highways compared to urban areas (1830 vs 843). This makes sense in the context of fatal accidents as the higher speeds on the highway would result in more damage

compared to slower speeds in the urban areas. However, taking a look at the speed limit on the roads(Figure 2.2), we can see that more fatal accidents happen in the medium-level speed limits. Taking these 2 insights, we can say that the worst accidents happen on highways with mid-level speed limits.

Time of Day for Fatal Accidents

Besides this, looking at the time of day, we can see that an overwhelming majority of them happened in earlier in the day, which be mornings and afternoons. This finding is contrary to what we would have thought would be the results, as you would think that lower visibility as well as drowsiness in the night time would cause more fatal accidents to happen. We will elaborate on what the possible solution for this would be at the end of this study.

Population

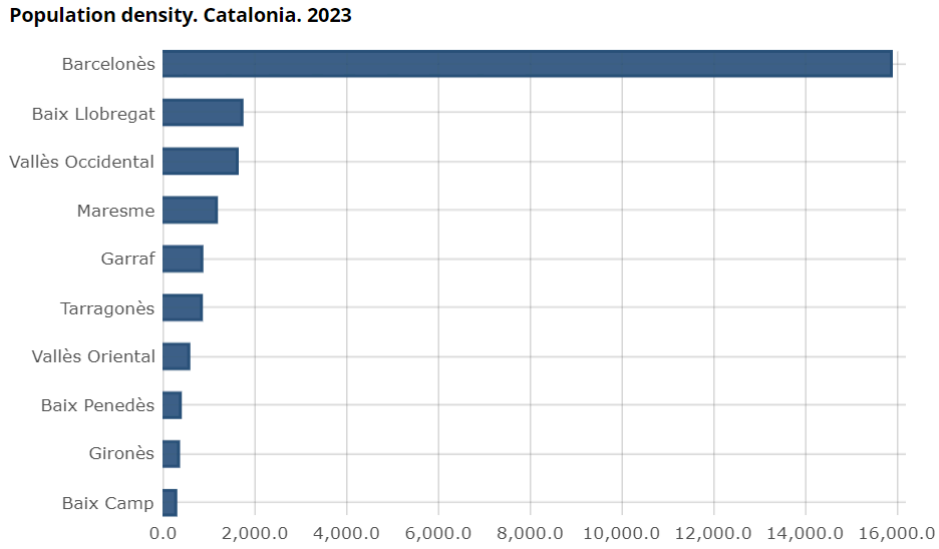
Population of Count (Figure 3.1)

	Population	Surface area km²	Density inh./km²	Ranking by number of inh.
Barcelona	1,655,956	101.4	16,339.0	1
Hospitalet de Llobregat, l'	276,617	12.4	22,307.8	2
Terrassa	225,274	70.2	3,210.9	3
Badalona	224,301	21.2	10,590.2	4
Sabadell	217,968	37.8	5,767.9	5
Lleida	142,990	212.3	673.5	6
Tarragona	138,326	57.9	2,389.9	7

Breaking down the number of accidents by county (Figure 3.1), we can see that Barcelones has the highest amount of accidents, with Valles Occidental, Baix Llobregat, Valles Oriental and Maresme being next. However, taking the population of each county into account, we can see that the number of traffic accidents is correlated to the population of each of the counties.

Population Density

Figure 3.2



Comparing the number of accidents to the population density of each county in Catalonia (Figure 3.2), we see that the top 5 counties with the most accidents, which are Barcelones, Valles Occidental, Baix Llobregat, Valles Oriental and Maresme, feature in the top 7 most densely populated counties.

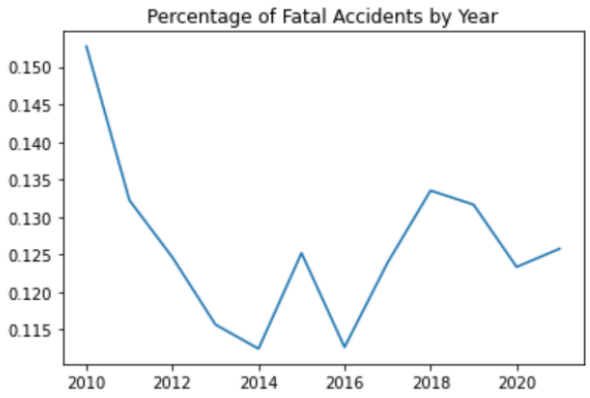
Based on the findings, it is clear that congestion, which correlates to population density, plays a big role in the number of traffic accidents in an area.

Looking into the types of roads on which accidents occurred, they happened as much on urban roads (10676) as they did on conventional roads (9019). However, looking deeper into the traffic accidents in specific counties, such as Barcelones, we can see that a majority of the accidents happened on urban roads compared to other types of roads.

Percentage of Fatal Accidents

The graph below (Figure 4.1) shows the percentage of the total number of traffic accidents that were fatal accidents every year in Catalonia.

Figure 4.1

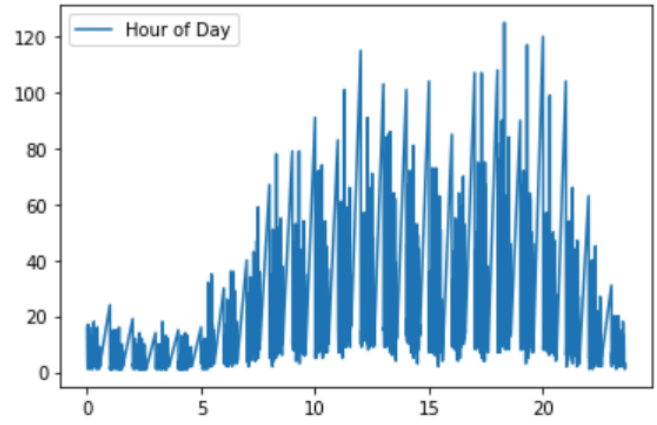


Taking the information about the total number of traffic accidents into account, we can say that 2010 was the most dangerous year on the road as there were high numbers of traffic accidents as well as a high proportion of fatal accidents. Overall, it seems like the roads becoming safer with time as fewer traffic accidents as well as fewer fatal accidents are happening.

Time of Day

Putting the time of day of each accident into a graph (Figure 5.1), you can clearly see which times of the day traffic accidents tend to happen. There is a spike in traffic accidents at around lunchtime (12:00-2:00 p.m.) as well as dinner time (6:00 - 8:00 p.m.) with a decent amount of traffic from 8 a.m. to 10 p.m. The early morning and the late nights seem to be relatively quiet, which would make sense as fewer people are on the road during these hours.

Figure 5.1



Day of the Week

Looking at the day of the week at which traffic accidents have occurred, the number of accidents peaked on Friday and Saturday with 3317 and 3196 accidents respectively. All the other days in the week are pretty even, averaging at around 2929 accidents.

Time of Day by Day of the Week

Figure 5.2

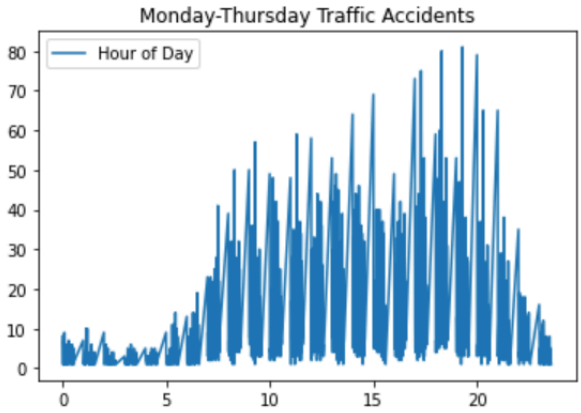


Figure 5.3

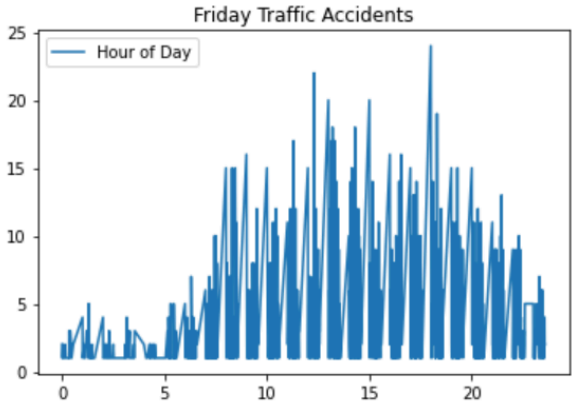


Figure 5.4

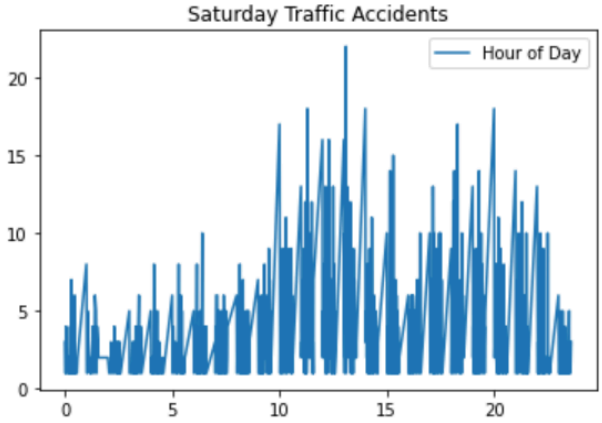
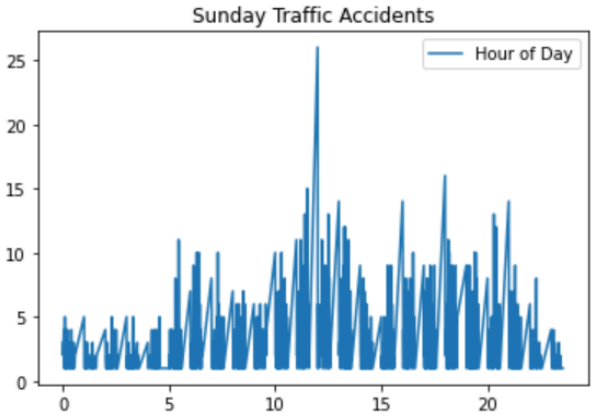


Figure 5.5



The graphs above show the difference in the number of traffic accidents based on the hour of the day for different days of the week. Note that that the numbers of the axis are not important here as we are trying to figure out how the trend changes throughout the day. Monday - Thursday(Figure 5.2) tends to have more constant traffic compared to the other days, this constant traffic is seen between general working hours. Similar to the general time of day graph, the weekends(Figure 5.3, 5.4 and 5.5) tend to have more spikes in traffic accidents during lunch and dinner times.

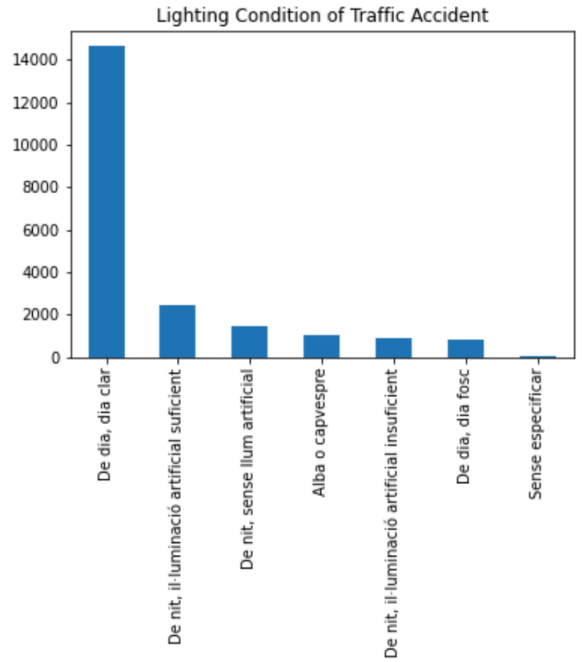
Overall, the amount of traffic accidents corresponds very well to the overall traffic and congestion on the road, which shows that a leading cause of traffic accidents is the presence of more cars. One

observation we thought we would notice is a high number of traffic accidents at night compared to the day, due to there being an absence of light which affects visibility, however, we did not see this trend in the dataset

Lighting Conditions

Looking at the lighting conditions of each traffic accident (Figure 6.1), you can see that “daytime and clear day” is the most common by a big factor, 14626 vs 6535 (all other weather conditions). This is quite shocking as you would think that these ideal lighting conditions would reduce the possibility of a traffic accident and other extreme weather conditions would increase the possibility of a traffic accident. One possible explanation for this could be that people who are driving in extreme measures acknowledge this and drive more carefully, which in turn drastically reduces the chance of a traffic accident happening.

Figure 6.1



Influence of Environment

Out of all the traffic accidents influenced by external environment, It is clear that the influence of visibility has the biggest impact on traffic accidents. Looking at the presence of fog column, we can see that fog is not a main cause of the influence of visibility, as only about 5% of the influence of visibility accidents reported the presence of fog. This tells us that many of the visibility issues are caused by either physical barriers to visibility or eyesight problems.

Road Surface Conditions

Looking at the road surface conditions of the traffic accidents, an overwhelming majority of it is on dry and clean roads, about 95% of traffic accidents happened on dry and clean roads. The other categories included wet, slippery and flooded roads. Much like the lighting conditions, we can see that much more accidents happen in ideal driving conditions compared to worse driving conditions. Again, we can see this pattern of people being more careful in these types of driving conditions and paying extra attention to the roads, which drastically reduces the chance of a traffic accident happening.

Special Roads

Of the special lanes in which accidents occurred (Figure 7.1), the bus and bike lanes are the most common lanes where traffic accidents occur in general. You would think that there would be a higher chance of traffic accidents in the acceleration lane as this is usually where people are driving more recklessly, but it seems like being in the wrong lane is a bigger cause of traffic accidents

Figure 7.1

Special Lane Presence	
No n'hi ha	18567
Carril bus	300
Carril bici	184
Carril acceleració	158
Altres	152
Carril central	148
Carril d'alentiment	89
Carril lent	79
Carril avançament	42
Habilitació voral/carril addicional	37
Carril reversible	22
Sense Especificar	21
Carril habilitat en sentit contrari habitual	13

Direction of Road

Looking at the direction of the road on which traffic accidents occurred, we can see that they occur much more on 2-way roads (12546) compared to 1-way roads (4573). It is well-known that a two-way highway is one of the most dangerous roads to drive on, so this finding is in line with what we thought we would have discovered.

Special Traffic Measures

Roadwork/Construction is the leading cause of accidents in the special traffic measures column with 277 accidents. This shows us that there might not be sufficient warnings about upcoming construction to prepare people for the sudden obstacles ahead of them.

Heavy Vehicle Involvement

Doing a correlation analysis between the columns “Heavy Vehicle Involvement” and “Fatalities”, we can see that there is a 0.1579 correlation. Relative to the other correlations between Fatalities and other variables, “Heavy Vehicle Involvement” is quite high. This means that if there is a heavy vehicle involved in a given traffic accident, there is a good chance that there will be fatalities as a result of the traffic accident.

Motorcycle Involvement

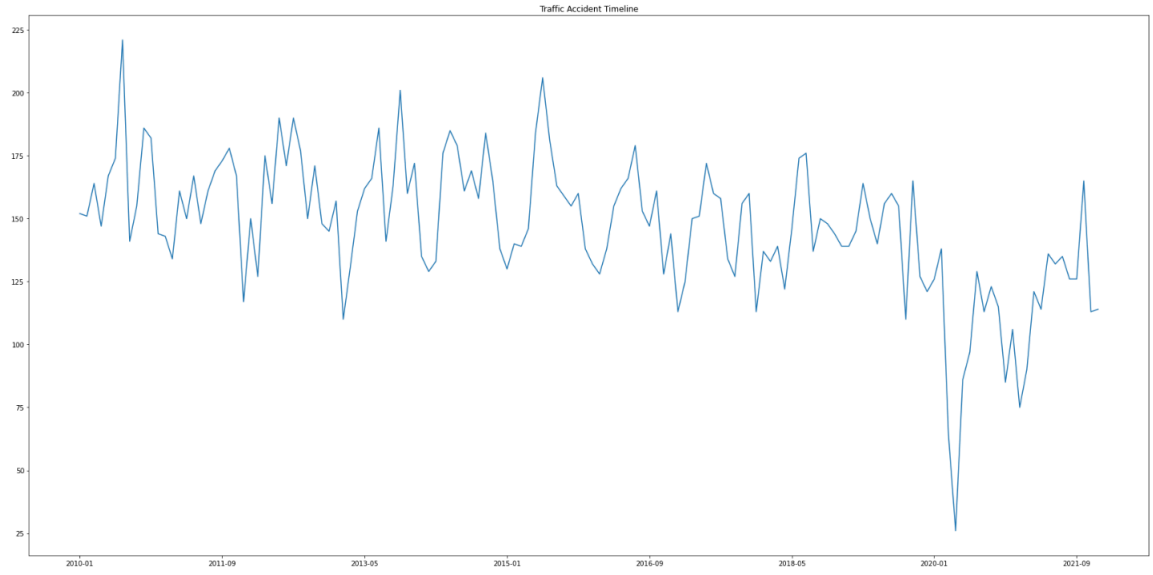
Looking at how motorcycles are involved with fatalities, there was no strong correlation between the two variables. Digging deeper into the numbers, we can compare the percentage of traffic accidents that involved motorcycle between fatal accidents and all the traffic accidents in general.

Of the 21161 traffic accidents in the dataset, about 8058 of them involved one or multiple motorcycles, which is about 38% of the accidents. On the other hand, of the 2673 fatal accidents in the dataset, about 793 of them involed one or multiple motorcycles, which is about 29.6% of the accidents. Based on the evidence, fatal accidents are not more likely to involve motorcycles.

Timeline of Traffic Accidents

Graphing the number of traffic accidents every month from 2010 - 2021(Figure 9.1), we can see that there is a cyclical cycle of accidents each year. Besides this, one insight that we had already talked about previously is that the overall number of traffic accidents are decreasing as time goes on. We can now see an anomaly around March and April of 2020, which coincides with COVID-19 and the lockdown.

Figure 9.1



Breaking it down by year (Figure 9.2) and month (Figure 9.3), there are some interesting insights for the month of the year. From the diagram, we can see that there are 2 distinct spikes in traffic accidents in May - July as well as in October.

Figure 9.2

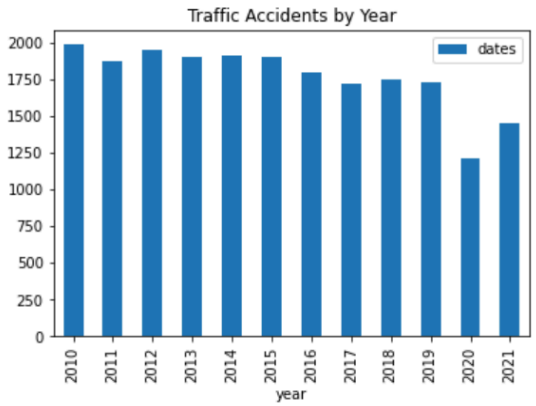
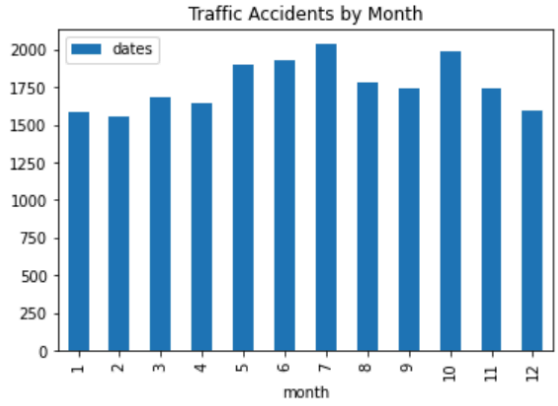
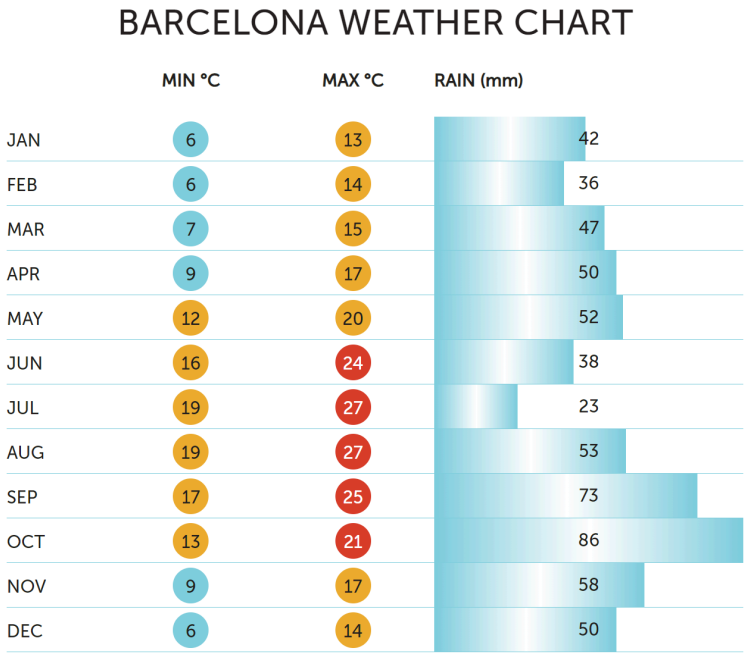


Figure 9.3



One possible explanation for these spikes in traffic accidents could be the increase in the amount of tourists that visit Catalonia. The first thought was to look up the number of tourists coming into Catalonia by month, however, this information was not readily available online. Looking at popular tourist sites and tourist recommendations, most websites recommend visiting Catalonia during late Spring, which is around May - June. This is because the weather is exceptionally good during this period, with warmer temperatures and less rain, as seen in Figure 9.4.

Figure 9.4



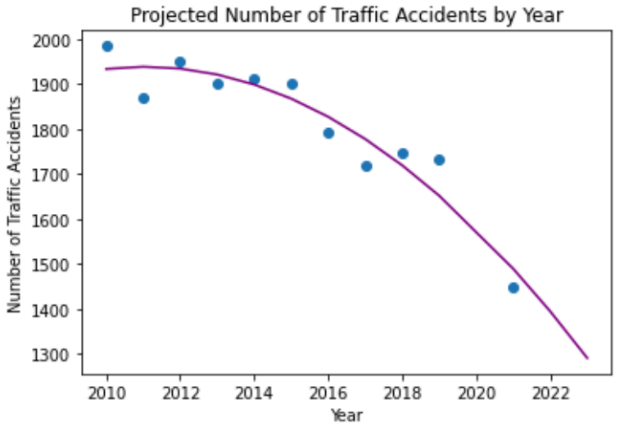
Doing more research into what happens in Catalonia during October, we came across multiple wine events. As this is the end of summer in Catalonia, there are many wine festivals and events that are held during this time of the year. This could lead to more traffic than usual as well as a higher possibility of light drinking when driving.

Projecting Traffic Accidents, Falatilies and Serious Injuries

Looking at the trend of the number of traffic accidents over the past years, it seems like the correlation would follow a quadratic regression. Knowing that COVID happened in 2020, we excluded the data points of this year as it would not reflect a normal year of traffic activity.

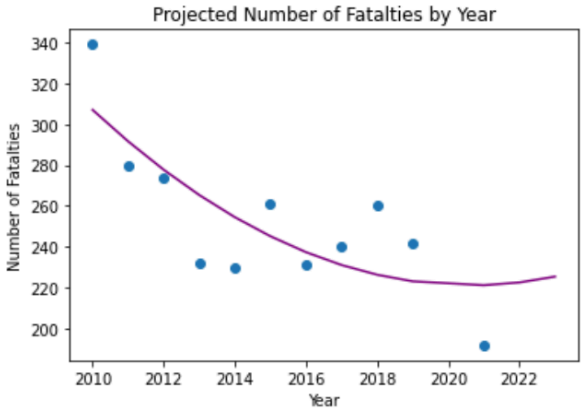
Using sklearn's Polynomial Regression on the past data, we predicted the number of traffic accidents for the next 2 years (Figure 10.1). For 2022 and 2023, the projected numbers of traffic accidents are 1394 and 1290 respectively.

Figure 10.1



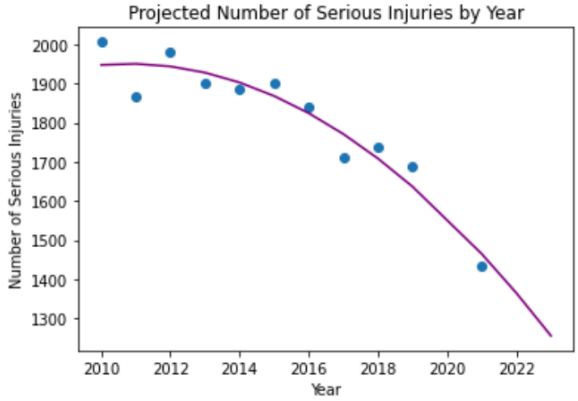
Using sklearn's Polynomial Regression on the past data, we predicted the number of fatalities for the next 2 years (Figure 10.2). For 2022 and 2023, the projected numbers of fatalities are 223 and 225 respectively.

Figure 10.2



Using sklearn's Polynomial Regression on the past data, we predicted the number of serious injuries for the next 2 years (Figure 10.3). For 2022 and 2023, the projected numbers of serious injuries are 1364 and 1255 respectively.

Figure 10.3



Correlation Analysis on Fatalities and Serious Injuries

Next, we looked at which variables would contribute to how severe the traffic accidents were. We did this by running a 2 correlation analysis. The first would compare the number of fatalities and the other variables and the second would compare the number of serious injuries and other variables. The goal of this is to figure out which variables had the highest correlation to fatalities and serious injuries respectively.

For the categorical variables in the dataset, we used the get_dummies function that pandas provides that converts them into numerical representations of the categories. After doing this, we took the top 20 highest correlated factors for both Fatalities and Serious Injuries.

For context, after running get_dummies(), which divides the values of each categorical column into individual columns, there are 156 columns. Besides this, we excluded categories that had too many types of values, such as Counties and Municipals, as it would be hard to comb through the data with so many different columns.

Figure 10.4

```
Severity of Accident _Accident greu                              0.915672
Severity of Accident _Accident mortal                            0.915672
Serious Injuries                                                 0.355594
Total Victims                                                    0.231872
Subzone within Area_Carretera                                    0.174473
Kilometer Point                                                  0.169738
Type of Road_Via urbana( inclou carrer i carrer residencial)     0.169420
Subzone within Area_Zona urbana                                  0.169149
Heavy Vehicles Involved                                          0.157959
Type of Road_Carretera convencional                              0.156603
Road's Altimetric Layout_0.5                                     0.125488
Subtype of Accident_Col·lisió frontal                            0.121421
Road Ownership _Estatal                                          0.115501
Lighting Conditions _De nit, sense llum artificial               0.114221
Road Ownership _Autonòmica                                       0.097234
Road's Altimetric Layout_Pla                                     0.096016
Intersection Characteristics _En secció                          0.093675
Intersection Characteristics _Dintre intersecció                 0.089334
Road's Altimetric Layout_Rampa o pendent                         0.088781
Direction of Road_Doble sentit                                   0.087241
Name: Fatalties, dtype: float64
```

Figure 10.5

```
Severity of Accident _Accident greu                 0.484911
Severity of Accident _Accident mortal               0.484911
Total Victims                                       0.429488
Fatalties                                           0.355594
Light Vehicles Involved                             0.177043
Light Injuries                                      0.128286
Subtype of Accident_Col·lisió frontal               0.121585
Type of Accident _Col.lisió de vehicles en marxa    0.102041
Units Involved                                      0.100722
Heavy Vehicles Involved                             0.065623
Subtype of Accident_Atropellament                   0.063671
Type of Accident _Atropellament                     0.063671
Day of the Week Grouping _CapDeSetmana              0.055424
Day of the Week Grouping _Feiners                   0.055424
Day Type_dill-dij                                   0.055290
Road Ownership _Autonòmica                          0.054613
Day Type_dg                                         0.047327
Motorcycles                                         0.042630
Type of Road_Carretera convencional                 0.041751
Subzone within Area_Carretera                       0.040221
Name: Serious Injuries , dtype: float64
```

As expected, some of the variables were highly correlated to Fatalities and Serious Injuries. One observation is that Fatalities have a much higher correlation (0.1744) to Subzone within Area_Carretera than Serious Injuries(0.04). This tells us that driving on the Carretera(highway) has a much higher risk of getting into traffic accidents that lead to fatal injuries. Again, we see that a two-way highway is the most dangerous road to drive on as it leads to the most fatalities.

Combing through most of the highest correlated variables, many of the high correlations make sense, in that what you would think causes fatalities and serious injuries in traffic accidents does show a high correlation to these variables. However, one correlation that we are curious about is between Fatalities and Kilometer Point, we have no possible explanation for this correlation but would be curious to look further into it in the future.

Conclusion

There were many findings from the in-depth analysis we did on multiple variables in the dataset, however, how can we use this information to help improve the state of traffic accidents in Catalonia?

The key takeaway we see from the data is that drivers are getting into a lot of traffic accidents where there are good driving conditions, which should not be happening as often compared to when there are bad driving conditions. The explanation that we came up with was that people take good driving conditions, such as good visibility, for granted and make mistakes because of this.

An argument against this could be that there is higher congestion during the mornings and afternoons compared to nights resulting in more fatal accidents, however, we do not have data to show this.

Regardless, we suggest that the Catalonia government constantly remind people to pay as much attention to the road in ideal driving conditions as they do in bad driving conditions.

Traffic Spikes

Digging deep into spikes of traffic accidents in the year, we found 2 distinct spikes in the number of traffic accidents in May - July as well as in October. Furthermore, some possible explanation for this was tourism during May - July, as the weather is the best during this time of the year, and the abundance of festivals centered around wine during early October.

There are two solutions that we propose to the Catalonia government to try to combat these spikes in traffic accidents.

The first is that the government try to provide incentives for tourists to travel to Catalonia during other times of the year by making off-peak season cheaper and subsidizing hotels and attractions. In addition to this, the government could also provide tourist attractions during other times of the year so that tourist congestion is not as impactful on the number of traffic accidents. At the end of the day, tourism is still going to be necessary for the Catalonia economy, so we think that the best solution is to tackle the distribution of tourism throughout the year instead of trying to reduce the amount of tourism

Secondly, the Catalonia government can potentially reduce the October traffic accident spike by enforcing no drunk driving, taking notice of any big alcohol-centered festivals and actively patrolling in these areas more often. You are not going to shut down these festivals as a government, so the next best thing to do is to keep it under control to reduce the likelihood of drunk driving or tipsy driving.

All in all, the traffic accident situation in Catalonia is getting better, however, there are definitely still some specific areas of improvement that would significantly decrease the occurrence of traffic accidents and make Catalonia safer.

Future Improvements

Given that there is such a wide range of categories in this dataset, we decided to only pursue the findings that we could provide possible explanations for. However, I find this dataset very interesting and will be making improvements to it in the future and building upon what I have submitted in my own time.

Besides this, since I am Chinese, I do not speak any Catalan or Spanish. Because of this, some of the translations I make in the study might be inaccurate or incorrect as I am using ChatGPT to translate the words from Catalan to English.

Bibliography

Idescat. (n.d.). Population. Idescat. https://www.idescat.cat/indicadors/?id=aec&n=15227&lang=en

Responsible Vacation. (n.d.). Best Time to Visit Catalonia. Responsible Vacation. https://www.responsiblevacation.com/vacations/catalonia/travel-guide/best-time-to-visit-catalonia

Wine Tourism. (n.d.). Wine Festivals in Spain. Wine Tourism. https://www.winetourism.com/wine-festivals-in-spain/