Analysis of Traffic Accidents in Stamford, CT

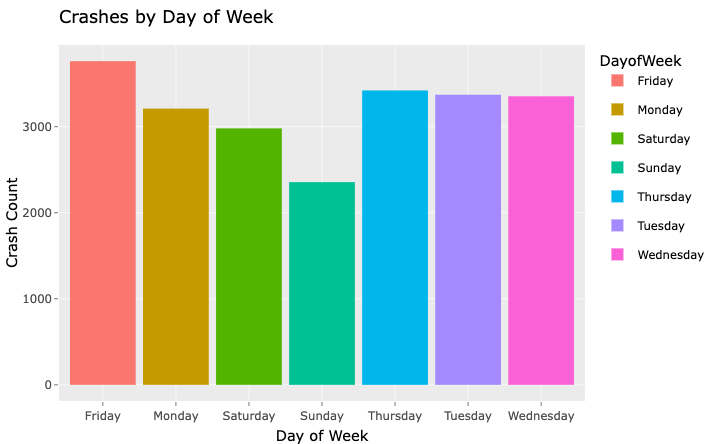
*Holli Tai & Joaquin Sanchez-Gomez*

Final Exam

This document analyzes the accidents that happened in Stamford, CT between 2015 and 2021, as well some information corresponding to the year 2022. This analysis has three different approaches: 1.- temporal, which refers to the year, day of the week and time the accidents occurred; 2.- the fatality of the damages and the weather conditions; 3.- using a model that allows us to figure out the causes behind an accident with human losses. For more clarification see our data on the dashboard provided.

In this regard, it was decided to analyze the incidents that occurred according to the year, day of the week and time of the event. In this first analysis there is a bias in the annual information, since the data reveals incomplete information for 2022, as well as a decrease in incidents in 2020 and 2021, which are associated with lockdown due to the covid-19 pandemic. On the other hand, the days with the highest number of incidents are during weekdays, while there is a lower incidence during weekends. In the first instance, this is related to the fact that there is much more traffic between Monday and Friday, because schools and workplaces operate normally. Additionally, most accidents happen on Friday; This is also linked to be the day when social gatherings increase, which are also associated with the consumption of alcoholic beverages.

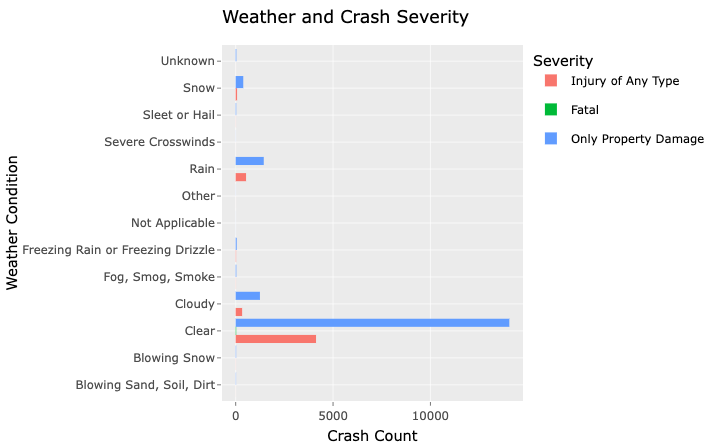
**Graph I. Crashes by day of week in Stamford, CT**



Source: own elaboration based on official data.

From a crash severity perspective, these are typified into three levels: death, injuries and property damages. Fortunately, the incidents that implied fatal outcomes represent a smaller proportion, since only 28 of them are registered. Subsequently, the accidents that produced some kind of injury were the second most frequent with 5,102 cases and, finally, the most frequent accidents produced some property damage, which accounts 17,321 cases, according to the data. In contrast to the weather conditions at the time of the incidents, most occurred on “good weather” conditions, as accidents in snow storms or rain were only a small share. This is partly due to a transit bias, since in adverse conditions for mobility, people refuse to move. The flows are much higher in favorable weather conditions, for this reason most of them occur in “clear conditions”.

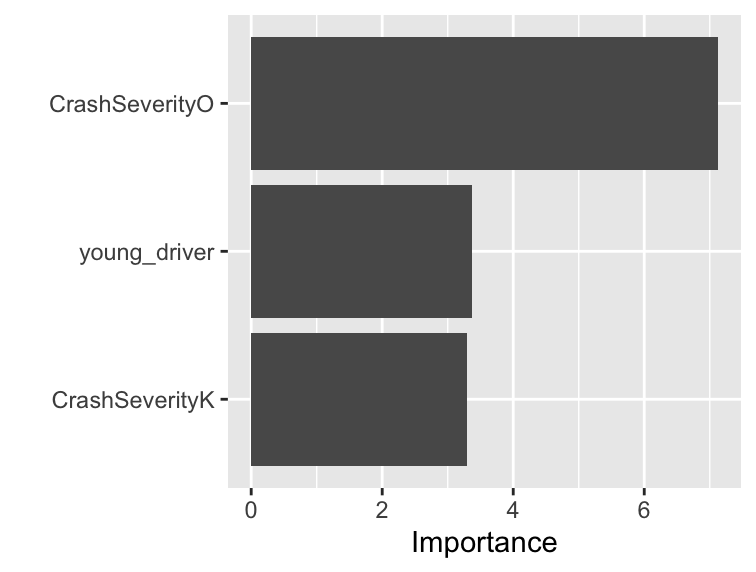
**Graph III. Accidents in Stamford, CT: weather and crash severity**



Source: own elaboration based on official data.

Finally, to dig a little further in the context of the most severe accidents and to the phenomena related to them, it was found that in 2 out of 10 accidents in which there were human losses, the use of alcohol was involved. In the same thread, the driver’s profile with DUIs is much more frequent among younger people. In this sense, it is important that preventive actions for road safety focus on the responsible use of alcoholic beverages among people who start their lives as drivers.

**Graph III. Characteristics in accidents associated with Drivers with DUIs, in Stamford, CT**



Source: Own elaboration based on official data.

**Improvements to flexdashboard & initial EDA**

In our initial EDA, we received feedback with emphasis on the following:

* Weather Conditions and the Base Rate Fallacy
* Gradient theme on bar chart
* Insights on crashes by Day and Month (Data Table)

We made revisions (particularly to these factors of the dashboard) to make our analysis more insightful, as well as more visually appealing to make it more easily interpretable by an audience.

To show that the inflated number of crashes on clear days on the Crashes by Weather Condition chart was a result of Base Rate Fallacy (more clear days than bad weather days), we created a data table showing the association of crashes by weather condition on a given day per year. The table shows that for every year of available data, clear weather crashes were most prevalent. While this does not finitely show total numbers of clear days, it does give a clearer inference on the sheer amount of clear days on a temporal level.

The gradience on our first chart: “Accidents per Year”, did not clearly separate the discrete variables that were present in the analysis. We revised for each bar to use their own respective color, instead of a gradience over a continuous spectrum.

In our first data table, “Quantity of Crashes by Day and Hour” was not very insightful, as it left out other useful temporal measures such as year and month. A new data table was created which now includes recommended variables.