Car accidents in Seattle

by Carlos Ríos

Introduction

This work is aimed at society in general and the objective is to generate social awareness regarding trac accidents. Taking into account that traffic accidents represent the death of 1.3 million people around the world each year (World Health Organization WHO).

In this work it is intended to better understand this type of accidents, and evaluate which are the most relevant factors for their occurrence According to the WHO, the victims (worldwide) correspond to passers-by in 50 % of the cases, it is also explained that the majority of accidents in general occur due to human errors, however there are also environmental factors, such as automobile, or roads.

Data Source

The dataset used was the Collisions - All Years database provided by the Seattle SDOT

Trac Management Division.

The data is made up of 194673 cases

Data Cleaning

The data was classified into 3 groups

a) Not Relevant, for example those that correspond to identification codes (id), there are also repeated features or features with little information.

b) Relevant to understand the data, for example: date, location, coordinates.

c) Relevant to apply machine learning. For example: Severity (is the classification label). Number of pedestrians, type of collision, etc.

Methodology and results

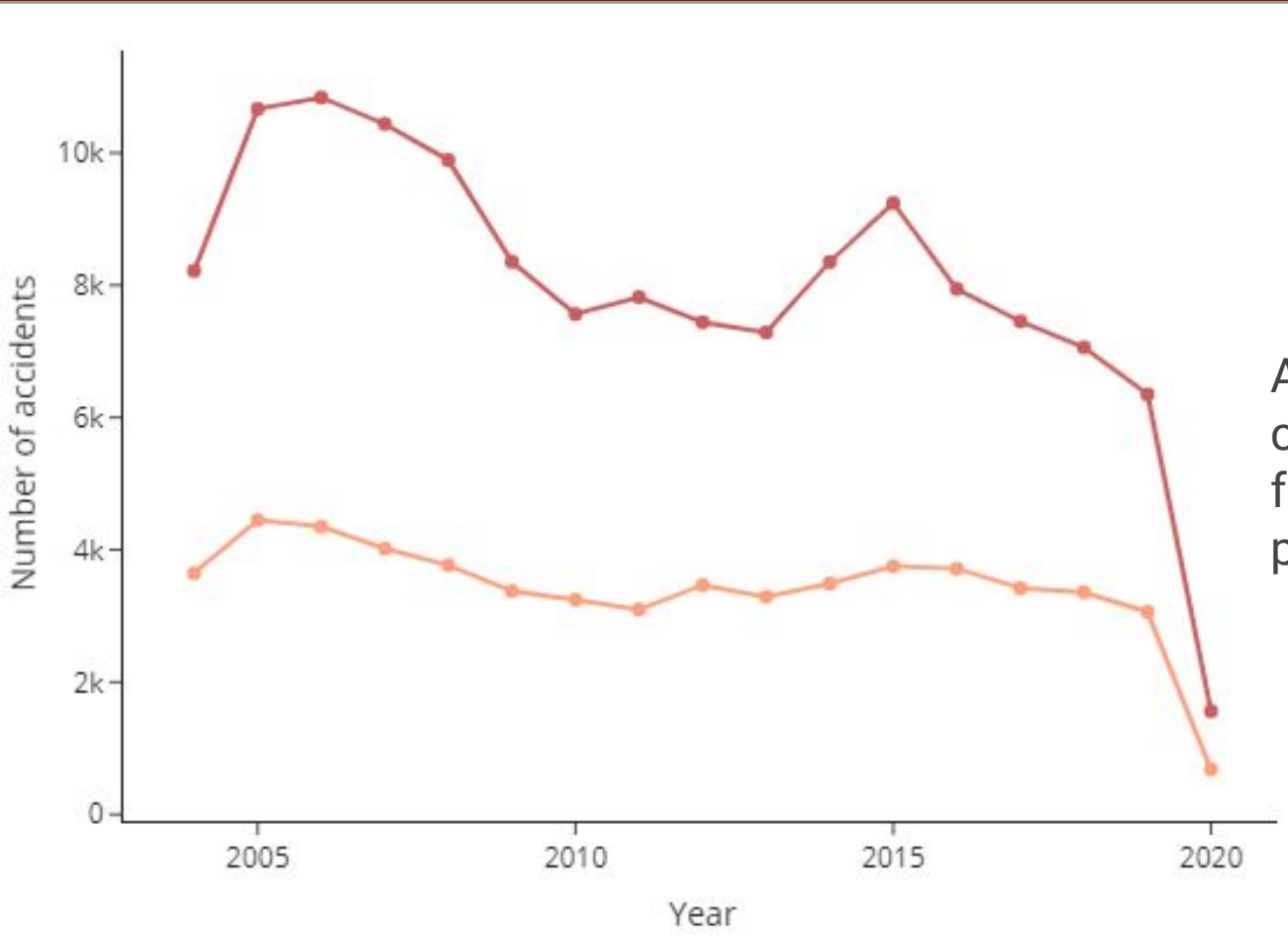
Methods

Count and explain

In order to better understand the accidents, methods of counting and graphing histograms were used mainly.

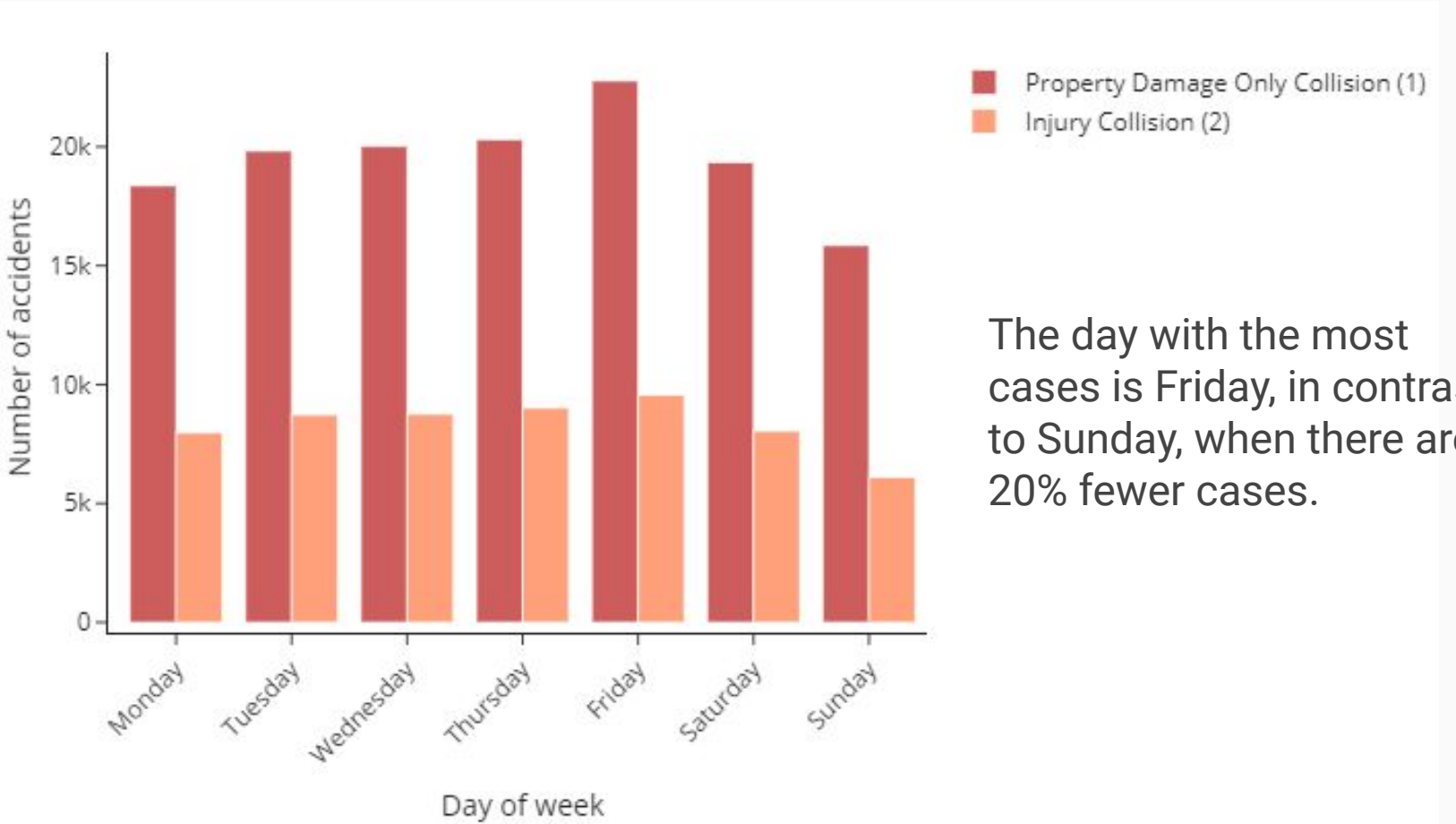
And in order to explain the principal factors a decision tree was used

Number of accidents year by year



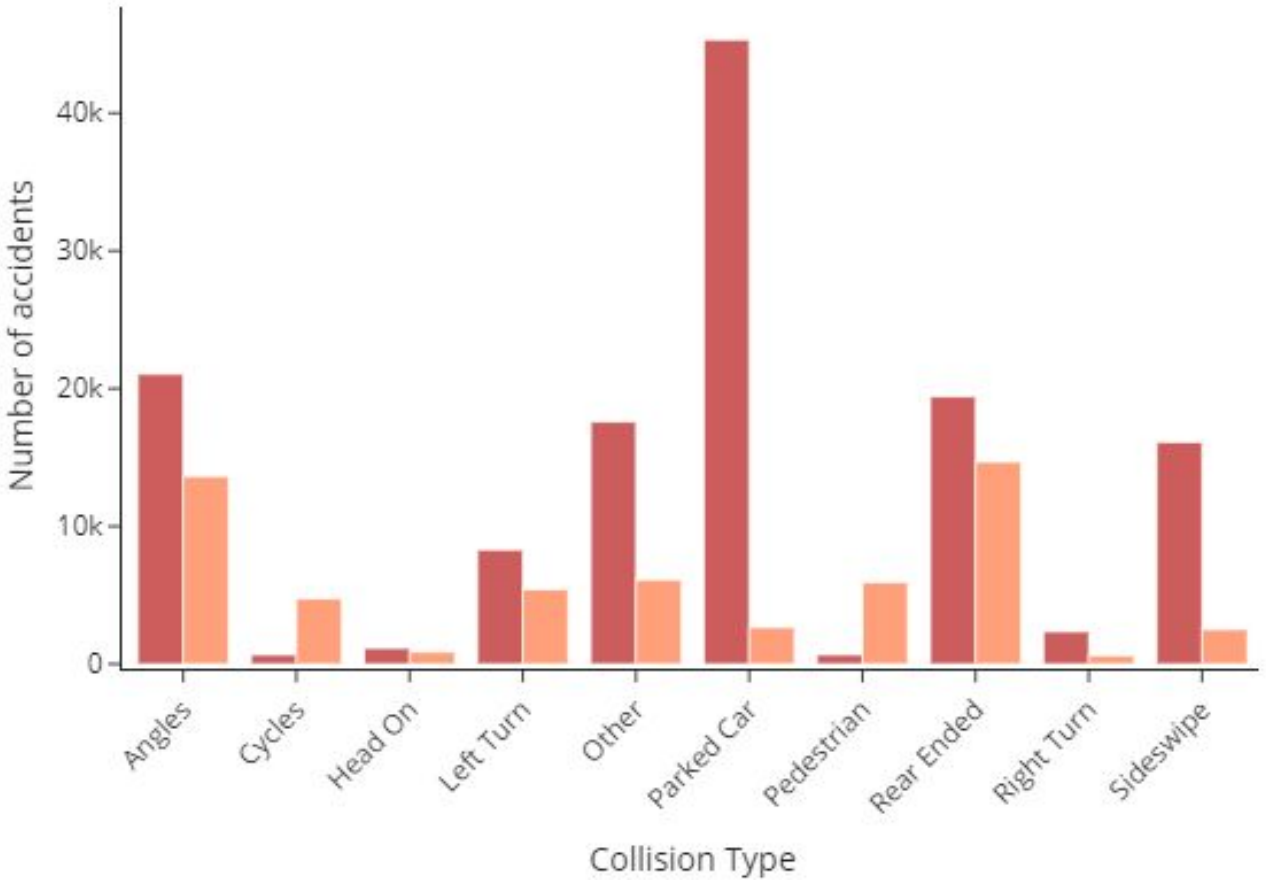
A decrease in accidents can be observed, except for 2015 where there is a peak

Number of accidents every day of week (from 2004 to 2020)

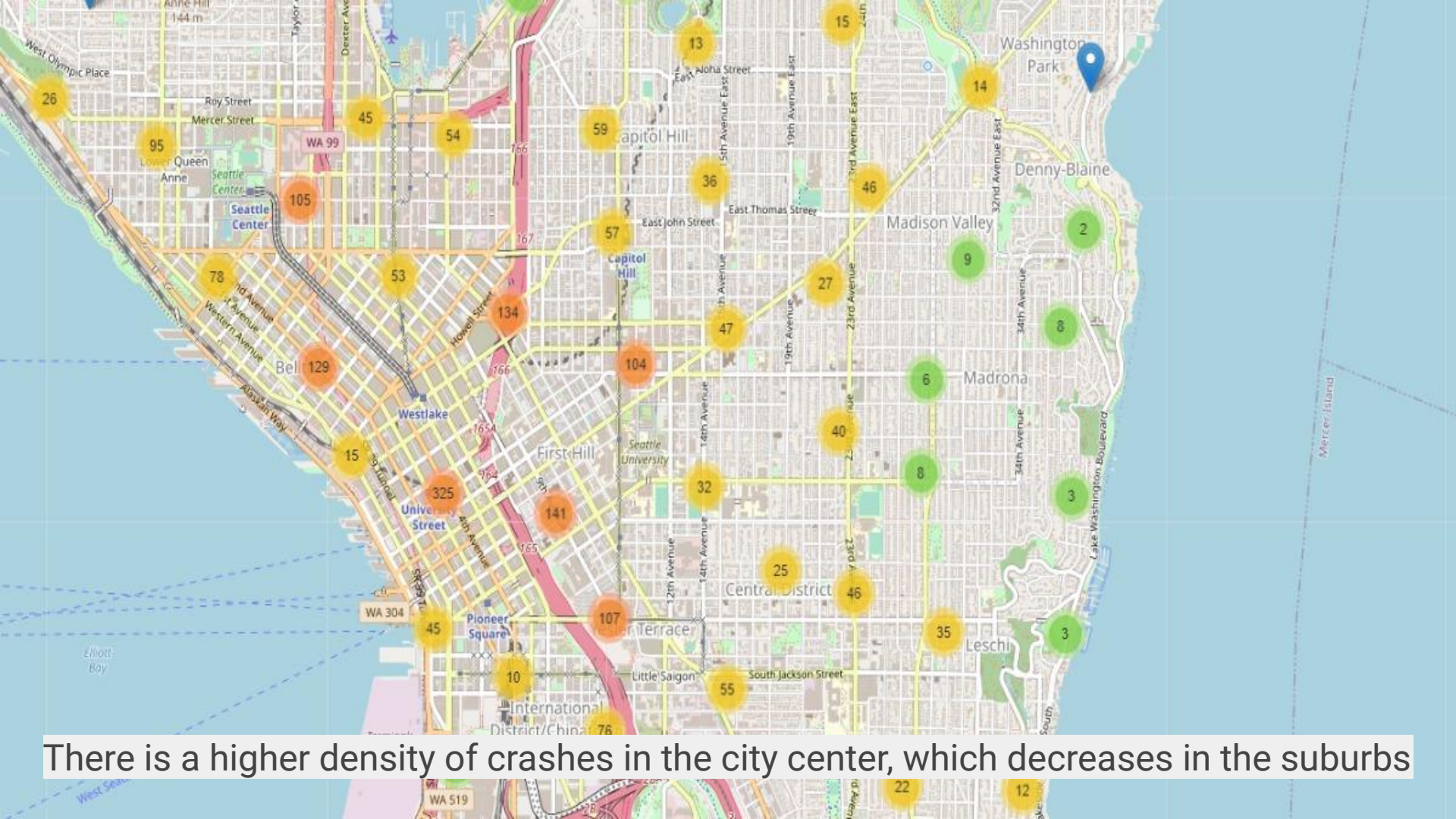


The day with the most cases is Friday, in contrast to Sunday, when there are 20% fewer cases.

Number of accidents by collision type



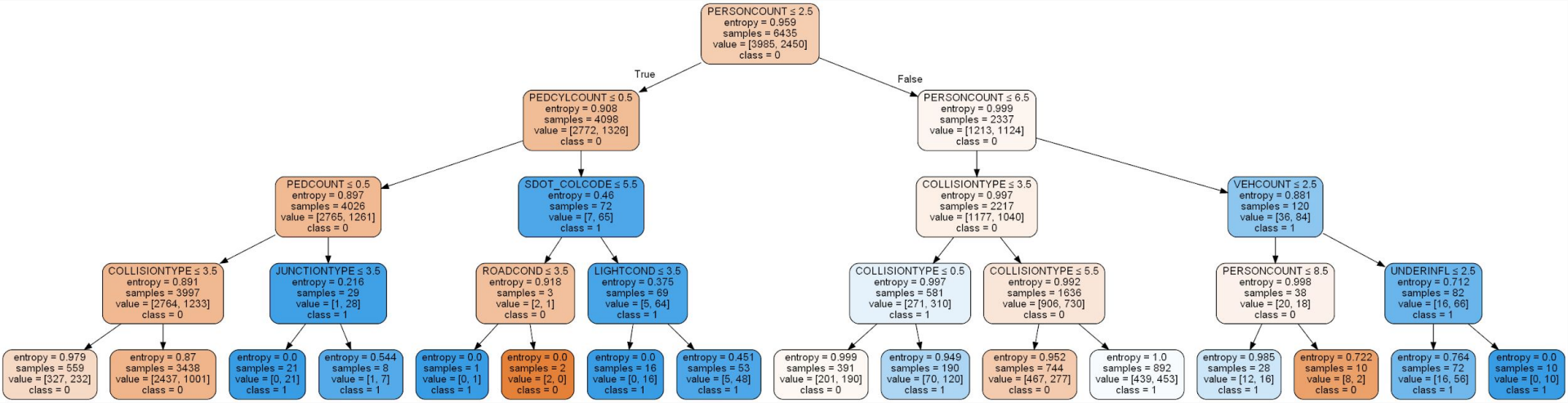
Most property-only crashes occur with parked cars. However, most injury accidents occur when the collision is at an angle or when it hits the car from behind. There is a curious asymmetry between the number of crashes when turning left compared to turning right.



There is a higher density of crashes in the city center, which decreases in the suburbs

Decision tree

A label encoder was used to encode non-numerical data, the "entropy" criterion was used and an accuracy of 0.64 was obtained.



Conclusions

Decision tree

Without a doubt, the tree could be improved with a better understanding of the features so we could have a more consistent label encoder.

General Conclusions

My suggestion is that you be more cautious on Fridays, do not leave your car parked anywhere, be more cautious when driving through the center and nally pay attention when you turn left!

Acknowledgment: Victor Once - Comercial Driver.
Photos: Unsplash