By Evaristo Acosta

Car Accidents in Seattle

Deaths in -or by- automobiles are a pressing public health concern. It is the second leading cause of accidental death in the country.

In 2018, Seattle alone had over 180 traffic accidents that resulted in serious injury or death. Since 2015, Seattle has made it its goal to drop the number of accidents to 0 by 2030, using Sweden's Vision Zero plan.

What can we do to reduce the number of collisions and their severity?

The goal is to use data that can help determine which characteristics increase the severity of accidents, such as road conditions, weather conditions, the exact time and location of the accident.

The Dataset

The dataset contains a table with several caracteristics of each colission reported.

There are 194,673 colissions reported since 2004 to the present.

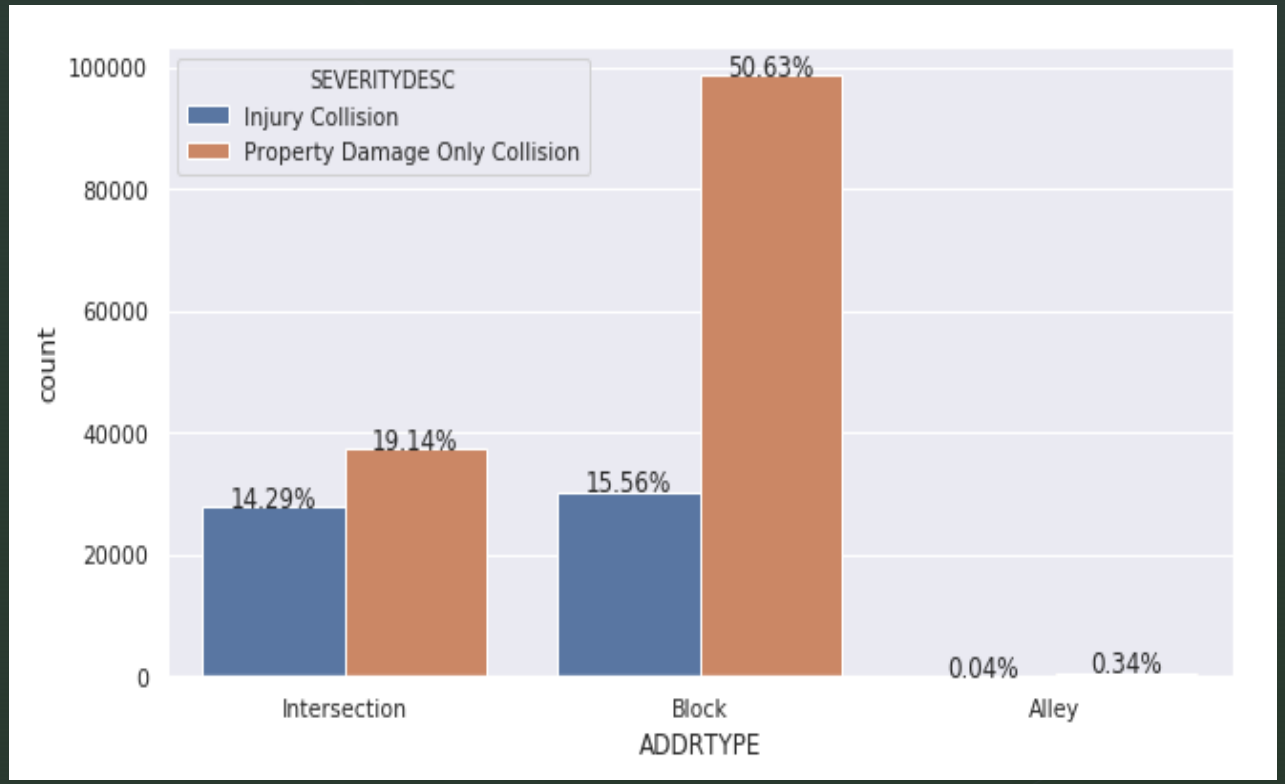
In thease accidents where involved:

```
Total number of people involved in collisions:  475864
Total number of vehicles involved in collisions:  373924
The number of bicycles involved in collisions:  5527
The number of pedestrians involved in collisions:  7230
```

Data Analysis

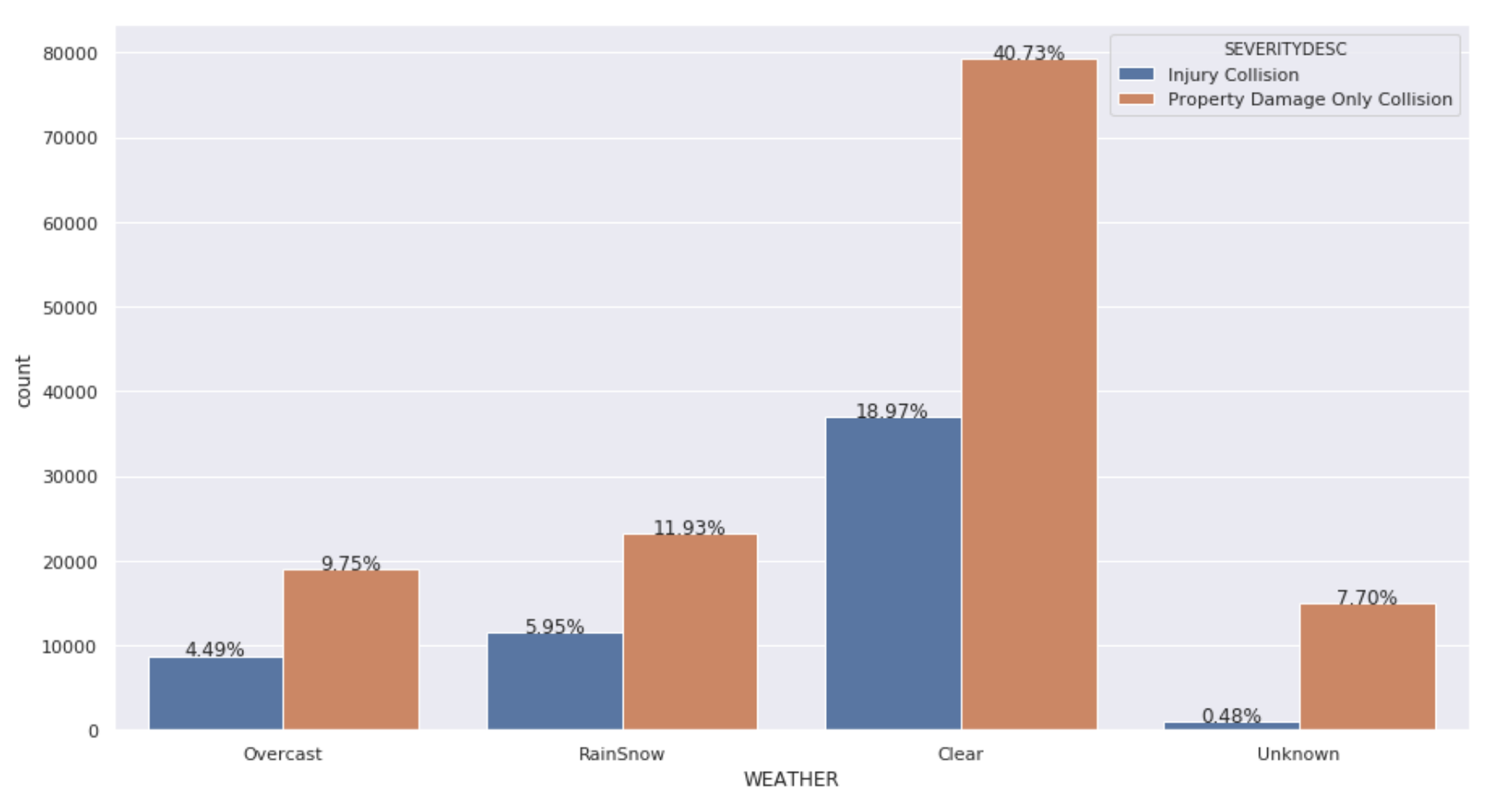
ANALYZING USING VISUALIZATION

- Location
- "At the intersection" is an important factor, where accidents are more likely to involve injuries
- Most of these collisions occur while parking, that also explains the high percentage of property damage on “Block”



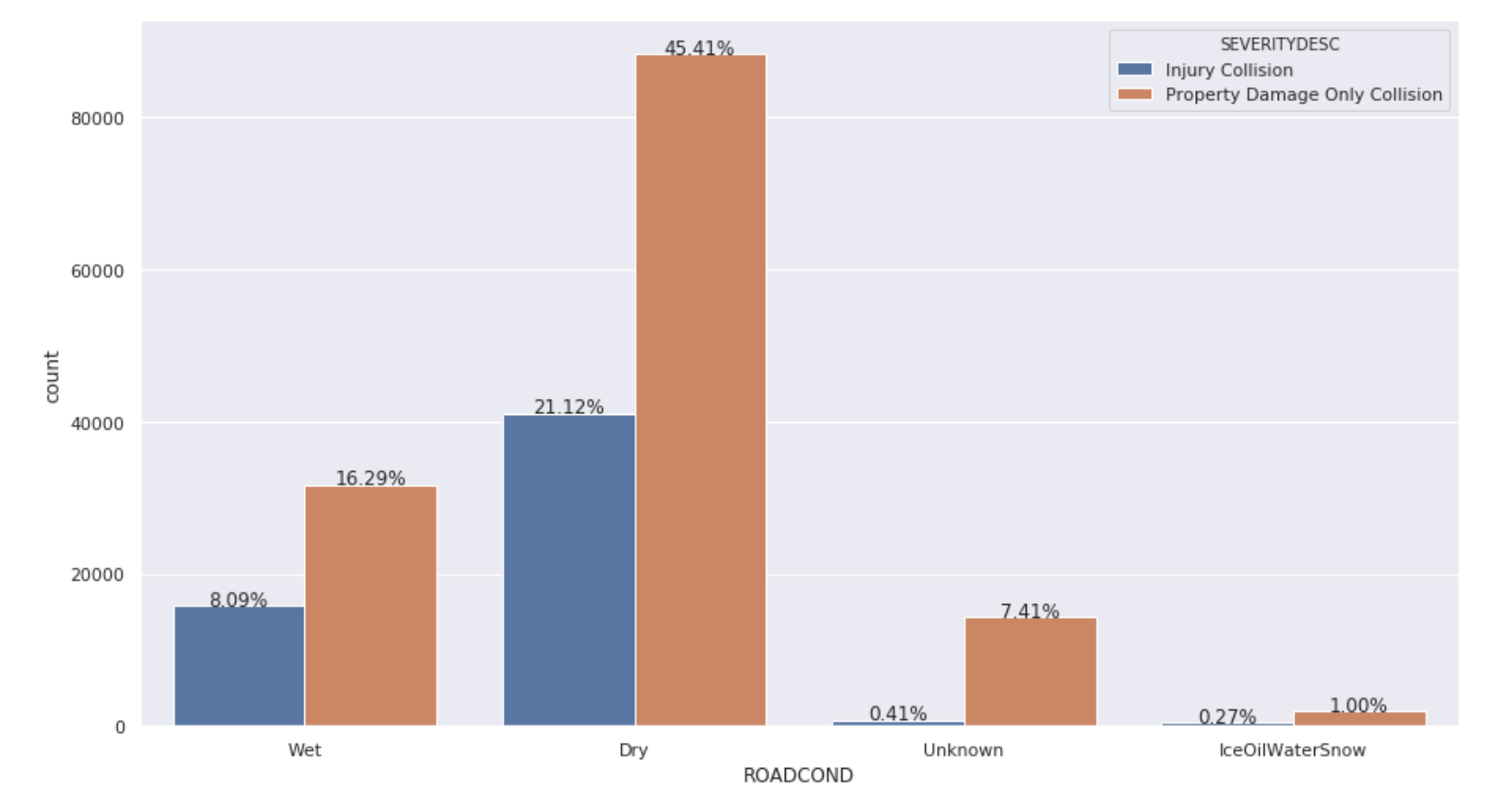
- **Weather**

- The weather does not seem to change the relationship in the severity of accidents between injuries and property damage. As can be seen in the graph below, the relationship remains constant.

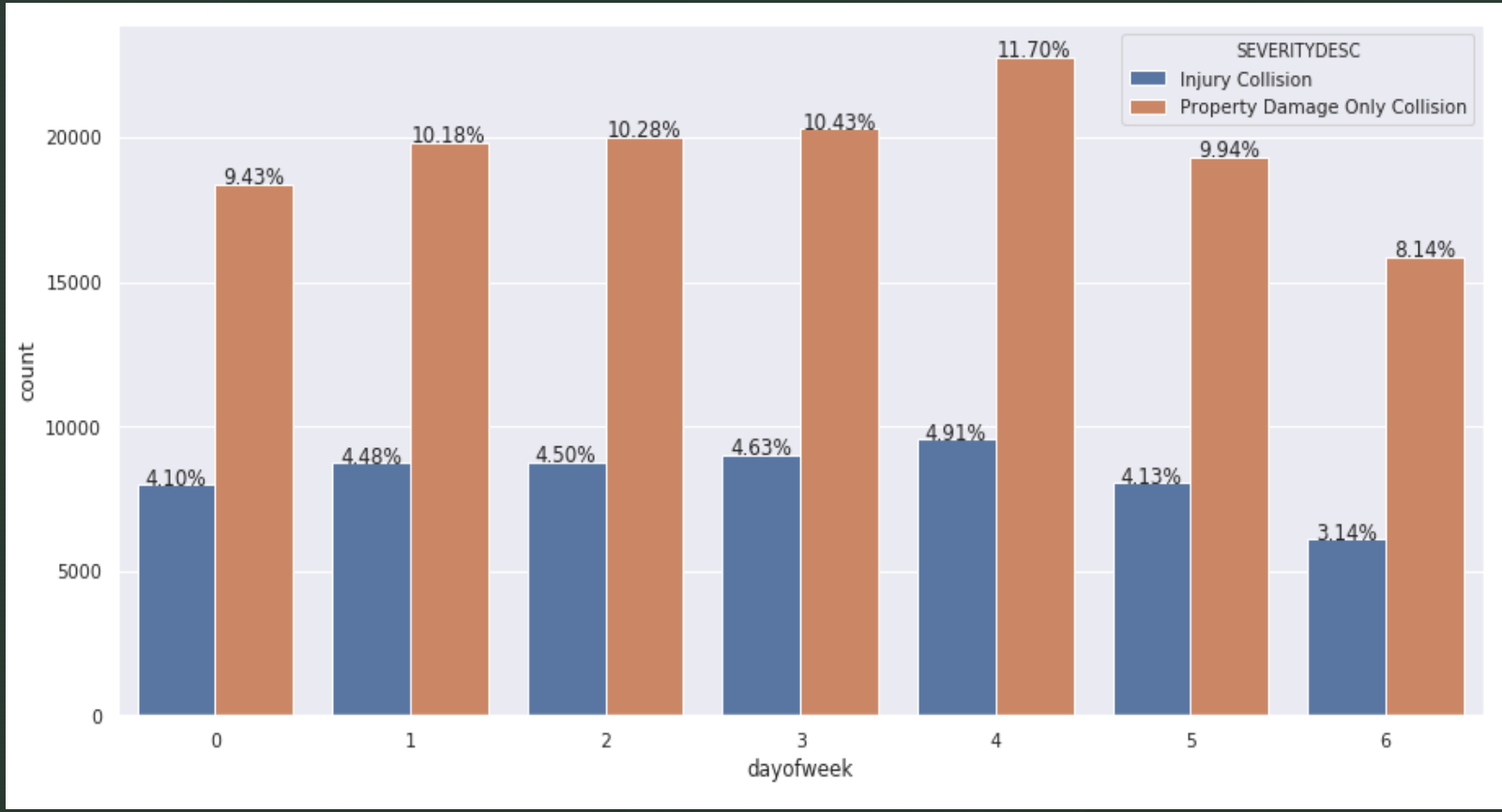


Road condition

- Regarding the condition of the road, the condition of the same does not seem to significantly modify the percentage of accidents with injuries.



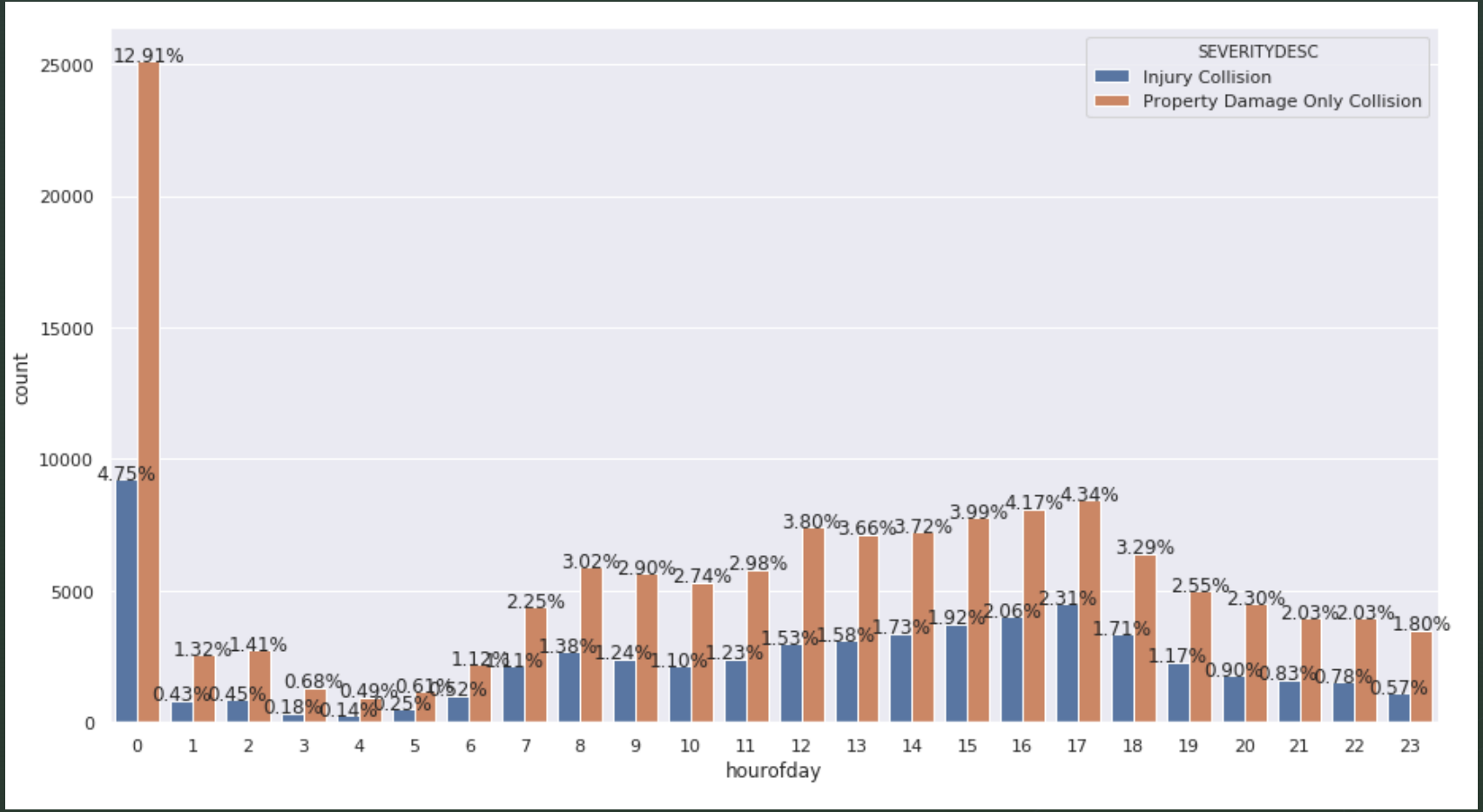
- **Days of the week**
- Here we can see and increse in the accidents on Friday but not in weekend.



- **Hours of the day**

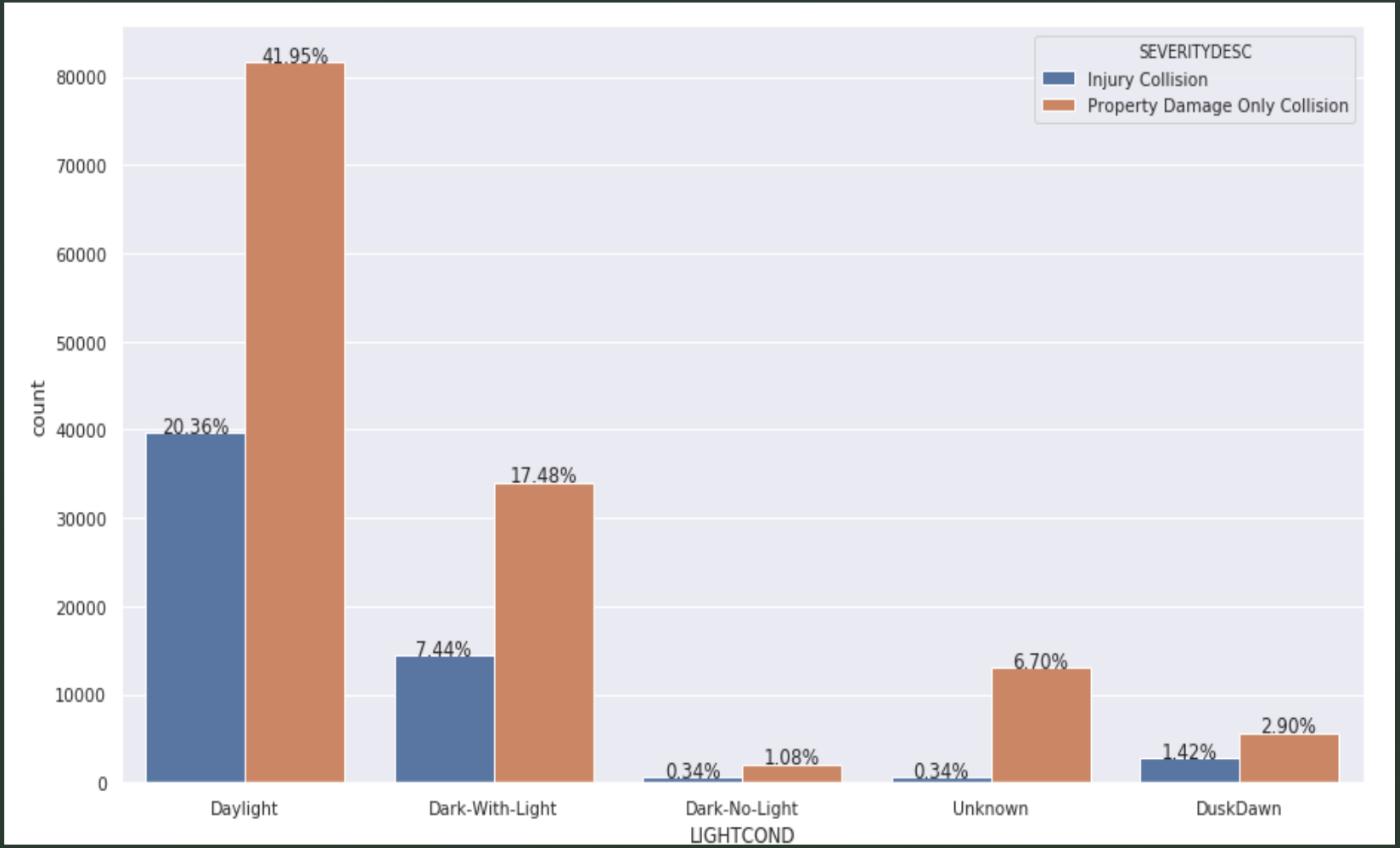
- There is a pronounced reduction in accidents during the early morning hours and an increase in working hours.

- A peak is observed at midnight, but this an error in the data collection.



- **Light Condition**

- The light does not seem to change the relationship in the severity of accidents between injuries and property damage.



Machine Learning Analysis

What do we expect from the model?

- We want the model that can best predict "Injury collisions".
- That is why the value that we are going to compare in returns is "recall" of Injury collisions. the model with the highest recall heat will be the one we must choose.

- For this Project we used three Machine Learning Model:

- **K- Nearest Neighbors**

- **Decision Tree.**

- **Logistic Regression**

Results and comparisions

Alogorithm	Average F-1 Score	Type	Precision	Recall
Decision Tree	0.61	Property collision	0.80	0.56
		Injury Collision	0.39	**0.67**
k-Nearest Neighbor	0.60	Property collision	0.74	0.65
		Injury Collision	0.36	**0.46**
Logistic Regression	0.61	Property collision	0.71	0.97
		Injury Collision	0.45	**0.06**

Based on the highest value of "recall" in "Injury collisions" we can say that the decision tree model is the one that best suits our requirements.

Conclusions

- As we see before, the Decision Tree model was the better predictor of the Injury Collisions, but the model has a large error margin. And the improvement requires large and more exhaustive datasets.

- Greater caution when driving during peak hours and on Fridays, special care at street crossings, and caution when driving when the road is wet are some of the actions that drivers can implement.

- The government can improve the drainage of the streets so that they dry faster after the rains, signal the crossings of the streets and place speed bumps in the areas of greater traffic, as well as promoting the use of public transport , reducing the circulation of vehicles during peak hours.