**Predicting the Severity of Car Accidents in Seattle**

**Yurgenis Rincón**

**October, 2020**

**1. Introduction**

**1.1 Background**

Seattle is the largest city of Washington state, its metropolitan area has a population of 3.98 million people, all this according to U.S Census data released in 2019, the city itself is hilly, though not uniformly so, it lies on seven hills like Capitol Hill, First Hill and more. Due to its location in the Pacific Ring of Fire, Seattle is in a major earthquake zone, on February 28, 2001, the magnitude 6.8 Nisqually earthquake did significant architectural damage. Seattle has a temperate climate, classified in the Mediterranean Zone by the main climatic classification*,* but some sources put the city in the Oceanic Zone. The Seattle area is the cloudiest region of the United States, due in part to frequent storms and lows moving in from the adjacent Pacific Ocean. With many more "rain days" than other major American cities.

**1.2 Problem**

Seattle residents get around by car, trolley, streetcar, public bus, bicycle, on foot, and by rail. With such bustling streets, it’s no surprise that Seattle sees car accidents every day. In 2015, a crash occurred in Washington [every 4.5 minutes](https://www.wsdot.wa.gov/mapsdata/crash/pdf/2015_Annual_Collision_Summary.pdf). Seattle recorded the highest number of car accidents in the state that year. There are many causes that provoke car accidents, and those car accidents have different kinds and levels of injuries, due to that, this project will predict the possibility of getting into a car accident and the severity of them, taking in mind the climate and the roads condition in Seattle.

**2. Data Acquisition and Cleaning**

**2.1 Data Sources**

The data to use in this project was obtained by SDOT Traffic Management Division, Traffic Records Group. The SDOT Traffic Management Division develops, maintains, and operates a transportation system that promotes the mobility of people and goods, and enhances the quality of life, environment, and economy of Seattle. The information of the data was collected between January, 2004 and May, 2020.

The data contain 194.673 rows and 38 columns. However there is a lot of lack of information, for example the column named *"PEDROWNOTGRNT*" has only 4667 rows of total data rows, and the data shows all necessary information related with collisions, car accidents, fatalities, injuries, weather and more details about those accidents.

**2.2 Data Wrangling**

As the first step, I visualized the data using the pandas’ command **.head**, **.shape**, **.columns**, to know the names of the attributes, what information there is in those attributes and to figure out what attributes will help me to achieve the objectives. In this step I noticed that there are several attributes that won’t fit with the purpose of this project, eventually I dropped those attributes, staying with only 19 attributes.

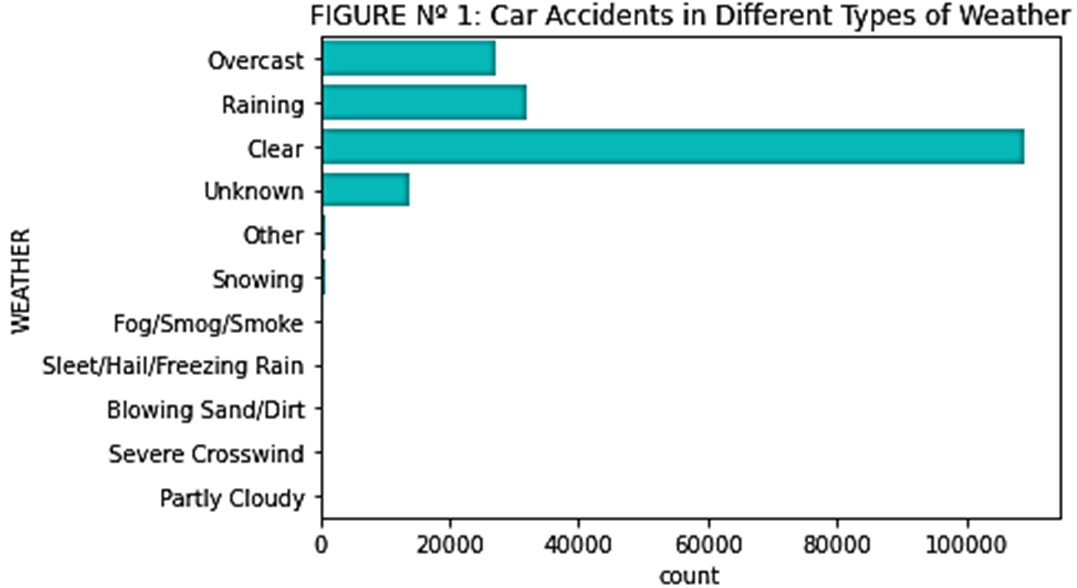
After that I looked for the missing values in the rows. Only 10 attributes had missing values, the most of them were attributes of *object* type, what I did was replace with the function **.replace** the missing values with the most common argument. However, there were other attributes that needed other types of treatment, for instance, the attributes **X, Y** and **LOCATION,** those attributes are geographical coordinates. The best way to take care of the missing values in those attributes is use the function **.dropna,** to eliminate all missing values of those specific attributes.

**3. Data Exploratory**

In this section of the project I will study several attributes, in order to know what aspects make possible car accidents or what kind of characteristics are involved in car accidents, all this taking in mind the objectives. The features that I will study are weather, road condition, light condition, if there are people under influence of alcohol or drugs involved in a car accidents, the type of collision related with other aspects.

**3.1 Weather Attribute Study**

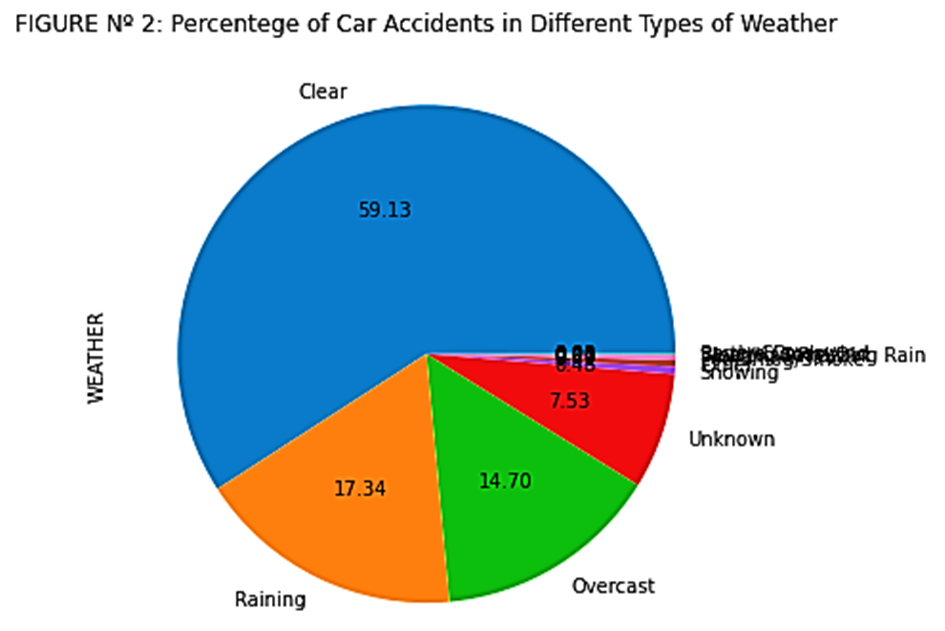
Before to start to analyze this attribute, it is necessary to know what is the amount of car accidents in the different kinds of weather. In this case there are 11 kinds of weathers, from *Clear* to *Snowing*, and after plotting these attributes (FIGURE Nº 1) I can see that there are 4 weathers who have the most quantity of car accidents between the 11 weathers.



As you can see in the FIGURE Nº 1, *Overcast* has 27.130 car accidents, *Raining* has 32.008 car accidents, *Clear* has 109.142 car accidents and *Unknown* weather has 13.891 car accidents. The other weathers don’t overpass the 1.000 car accidents, (See TABLE Nº 1). In the FIGURE Nº 2, you can see the percentage of car accidents in those kinds of weathers.

***TABLE Nº 1: Distribution of Car Accidents in Differents Kinds of Weathers***

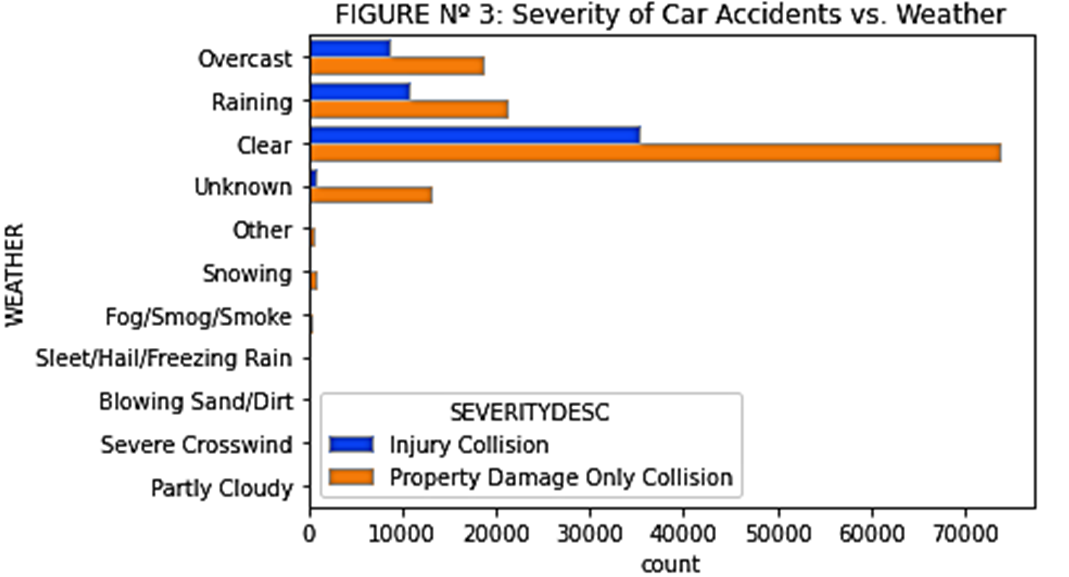
|  |  |
| --- | --- |
| **Clear** | 109.142 |
| **Raining** | 32.008 |
| **Overcast** | 27.130 |
| **Unknown** | 13.891 |
| **Snowing** | 894 |
| **Other** | 773 |
| **Fog/Smog/Smoke** | 553 |
| **Sleet/Hail/Freezing Rain** | 112 |
| **Blowing Sand/Dirt** | 50 |
| **Severe Crosswind** | 24 |
| **Partly Cloudy** | 5 |



The first three weather who have the major quantity of car accidents are *Clear* with 59,13%, *Raining* with 17,34%, *Overcast* with 14,70%.

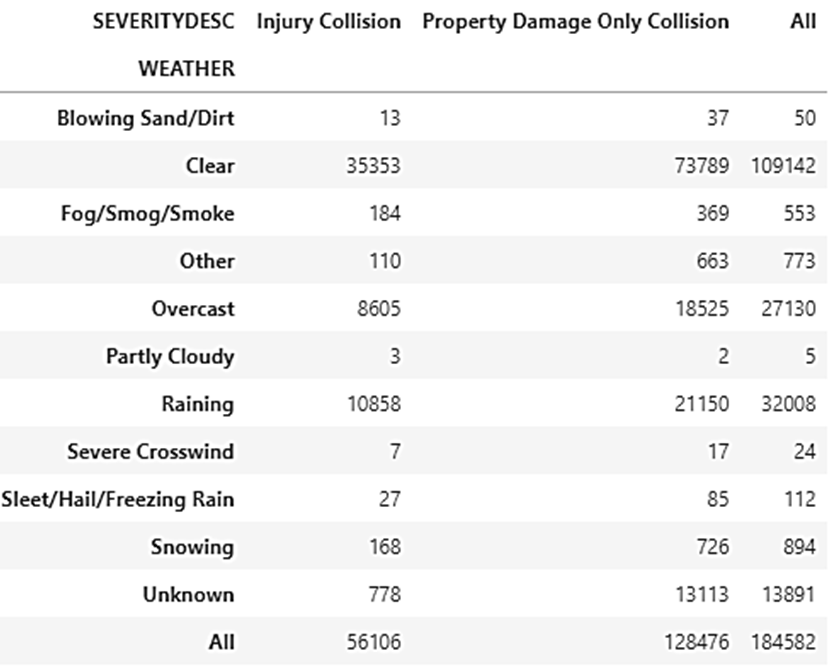
**3.1.1 Relationship Between Weather and Severity of Car Accidents**

To start to know more about car accidents, I start to analyze the behavior of weather’s attribute with other attributes, in this case is Severity of car accidents. As it is shown in the FIGURE Nº 3, *Injury Collision* and *Property Damage Only Collision* are the aspects that show the severity of car accidents. In this case the major quantity in every kind of weather is *Property Damage Only Collision.*



As I mentioned in previous section there are three main weathers where the major quantity of car accidents happen, *Clear* in this case has 35.353 car accidents with *Injury Collision* and 73.789 car accidents with *Property Damage Only Collision, Raining* has 10.858 car accidents with *Injury Collision* and 21.150 car accidents with *Property Damage Only Collision* and *Overcast* has 8.605 car accidents with *Injury Collision* and 18.525 car accidents with *Property Damage Only Collision.* (See TABLE Nº 2.)

***TABLE Nº 2: Distribution of Severity of Car Accidents in Differents Kinds of Weathers***

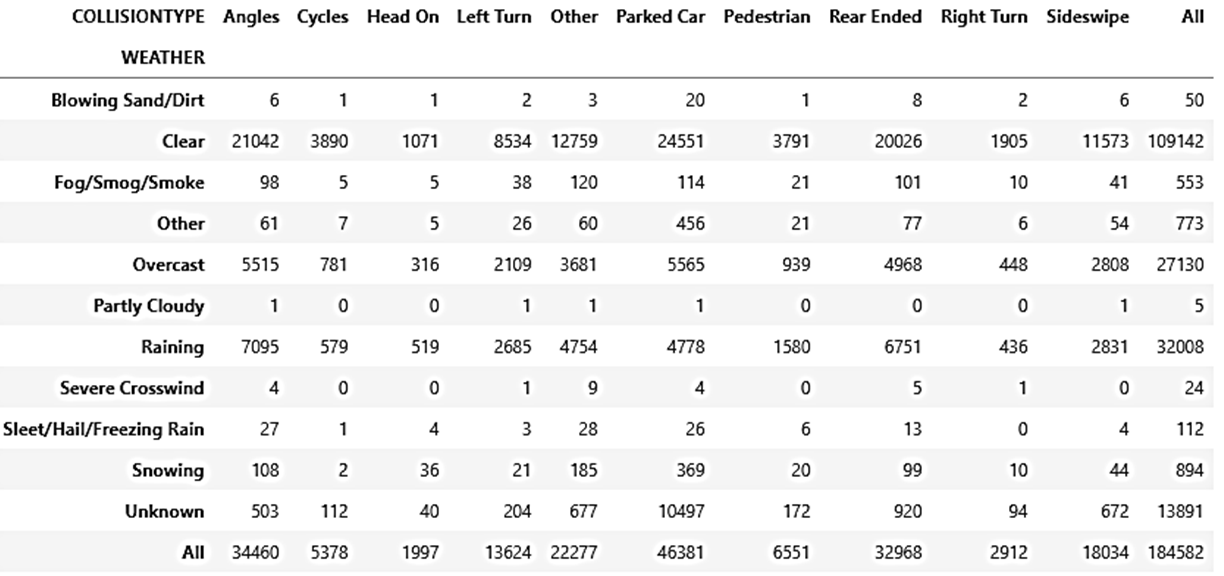


**3.1.2 Relationship Between Weather and Collision Type of Car Accidents**

In the Collision type attribute *Parked Car, Angles* and *Rear Ended* are the first three kinds of collision with the major quantity of car accidents. If I study this attribute with *Weather,* I can see that when the weather is Overcast*, Angles* and *Parked Car* are the main kinds of car accidents with almost the same quantity of accidents. When the weather is *Raining,* the statistics change a little bit, *Angles and Rear Ended* are the main kinds of car accidents and when the weather in *Clear, Parked Car* has the major quantity of car accidents. (See FIGURE Nº 4, and TABLE Nº 3)

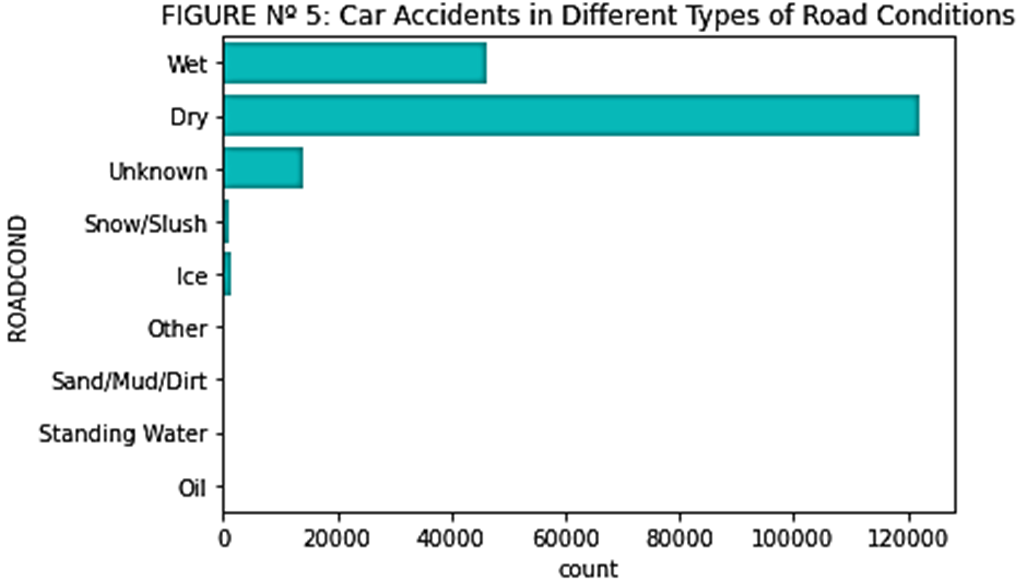


**TABLE Nº 3: *Distribution of Collision Type of Car Accidents in Differents Kinds of Weathers***



**3.2 Road Condition Attribute Study**

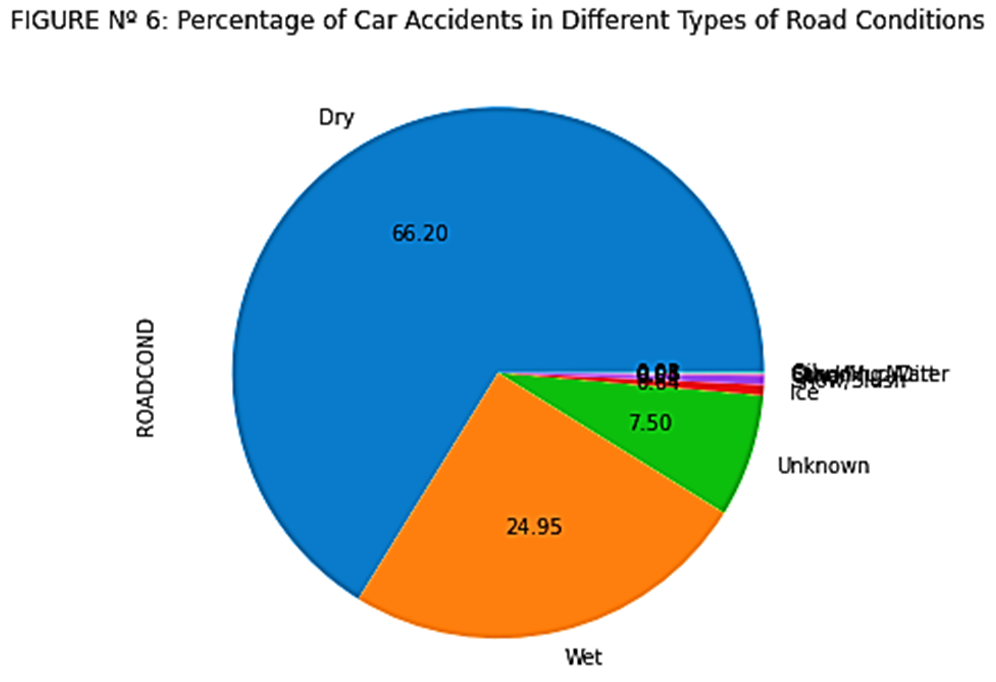
This attribute has 9 features, but there are 3 features with the major quantity of car accidents. When the road condition is dry, 122.192 car accidents occur, while 46.054 car accidents occur when the road is wet, and the other main feature of this attribute is when the road condition is unknown and the amount of car accidents is 13.835. It’s necessary to say that the rest of the conditions have car accidents but to a lesser extent. See FIGURE Nº 5 and TABLE Nº 4.



**TABLE Nº 4: *Distribution of Car Accidents in Differents Road Conditions***

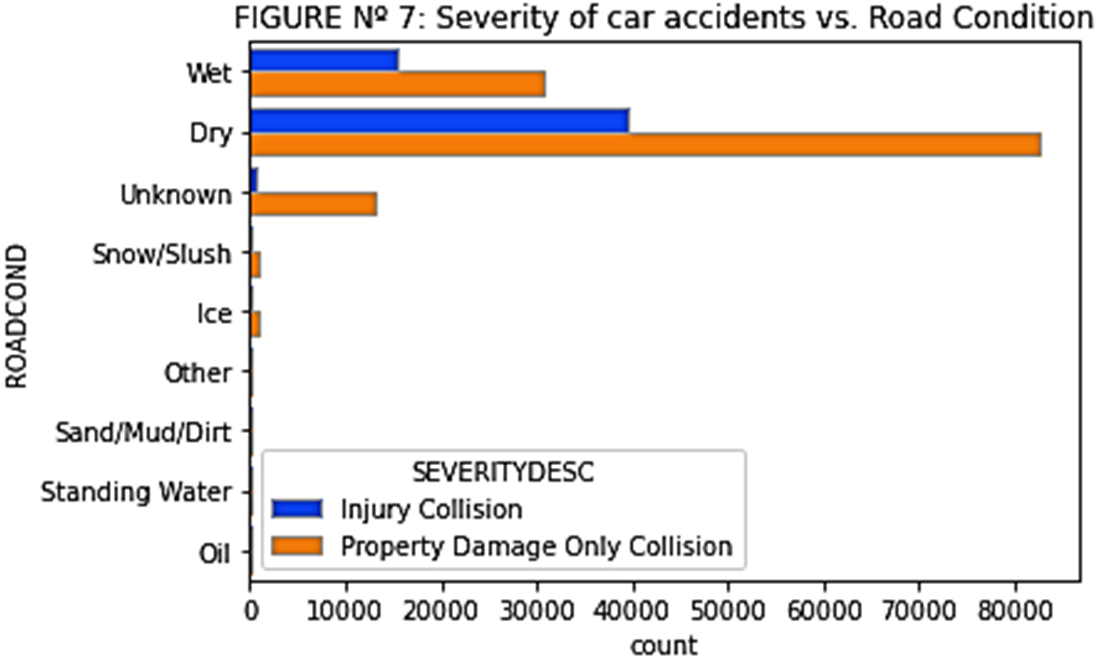
|  |  |
| --- | --- |
| **Dry** | 122.192 |
| **Wet** | 46.054 |
| **Unknown** | 13.835 |
| **Ice** | 1.176 |
| **Snow/Slush** | 989 |
| **Other** | 117 |
| **Standing Water** | 102 |
| **Sand/Mud/Dirt** | 64 |
| **Oil** | 53 |

As I mentioned before *Dry, Wet* and *Unknown* are the three features with more car accidents, putting these quantities in percentage *Dry* has 66.2% of the total car accidents, while 24.95% of car accidents occur when the road is wet and the rest of car accidents are distributed through the rest of the differents kinds of road conditions. See FIGURE Nº 6.

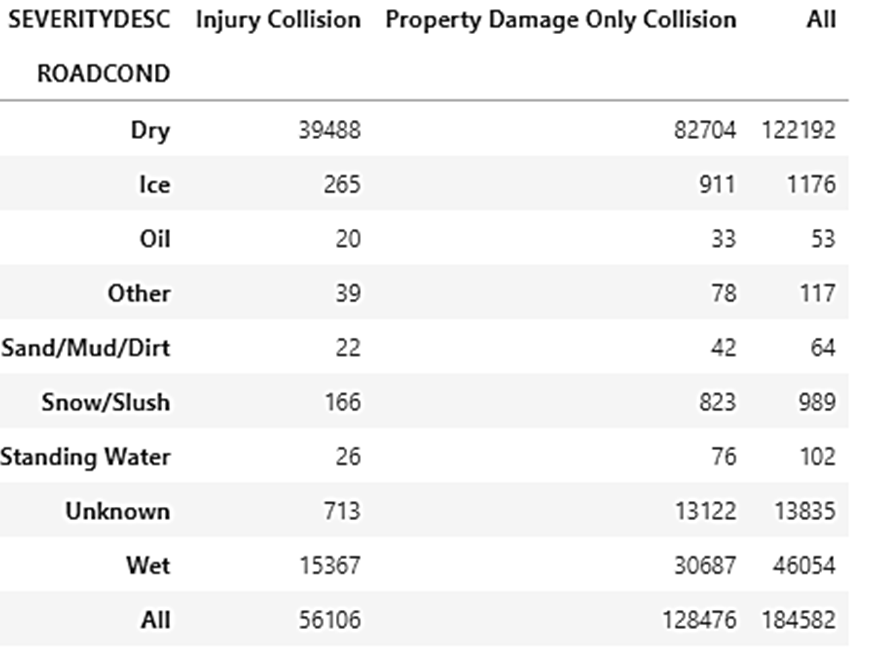


**3.2.1 Relationship Between Road Conditions and Severity of Car Accidents**

It is necessary to compare Road Conditions with other attributes, in this case the severity of car accidents. Only taking the features *Dry* and *Wet*, because they have more than 90% of car accidents, I can see in the FIGURE Nº 7, that the feature *Dry* has 82.704 car accidents with only damage to the property, and 39.488 car accidents with injury collisions. When the road condition is *Wet,* there are 30.687 accidents with only damage to the property and 15.367 accidents with injury collisions.

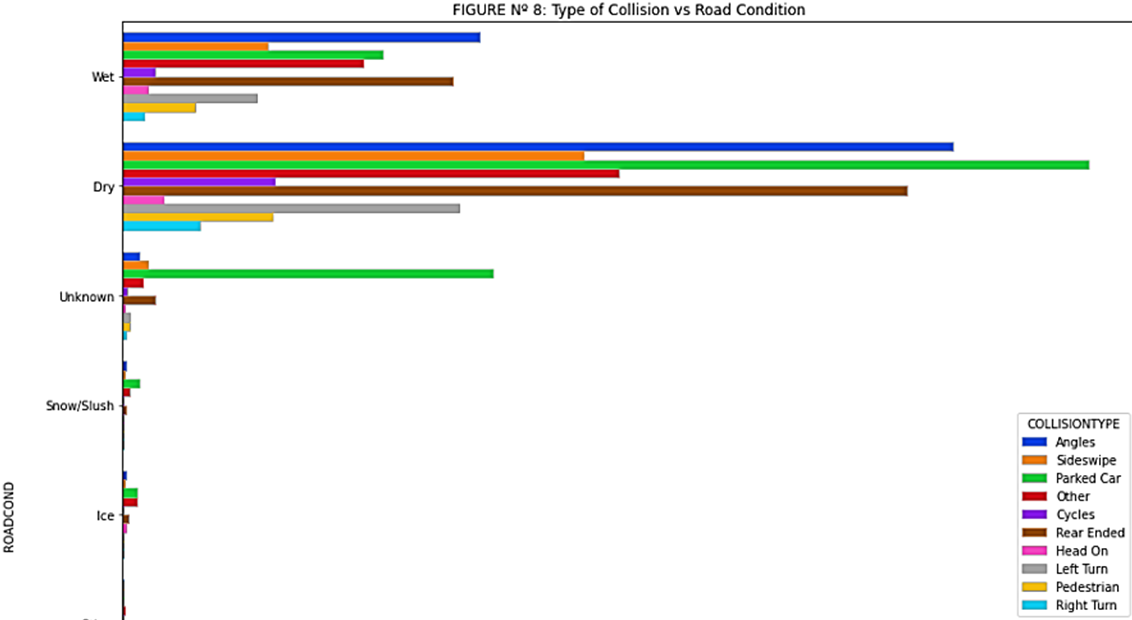


**TABLE Nº 5: *Distribution of Severity of Car Accidents in Differents Kinds of Road Conditions***



**3.2.2 Relationship Between Road Conditions and Collision Type of Car Accidents**

Comparing road condition with collision type and only taking the features *Dry* and *Wet*, I can see in the FIGURE Nº 8, that when the road is dry the major quantity of car accidents occur when the cars are parked, in this dataset there are 27.501 parked car accidents, the second type of collision is *Angles,* with 23.643 accidents occurred and in the third place 22.334 car accidents provoke by *Rear Ended.* When the road condition is *Wet*, the statistics change a little bit, in this case the major quantity of car accidents occur in *Angles,* with 10.150 car accidents, the next type of collision is *Rear Ended,* with 9.399 accidents and *Parked Car* has 7.413 car accidents. However, there is one statistic that calls my interest, when the road condition is *Unknown* there are 10.543 accidents when the cars are parked.

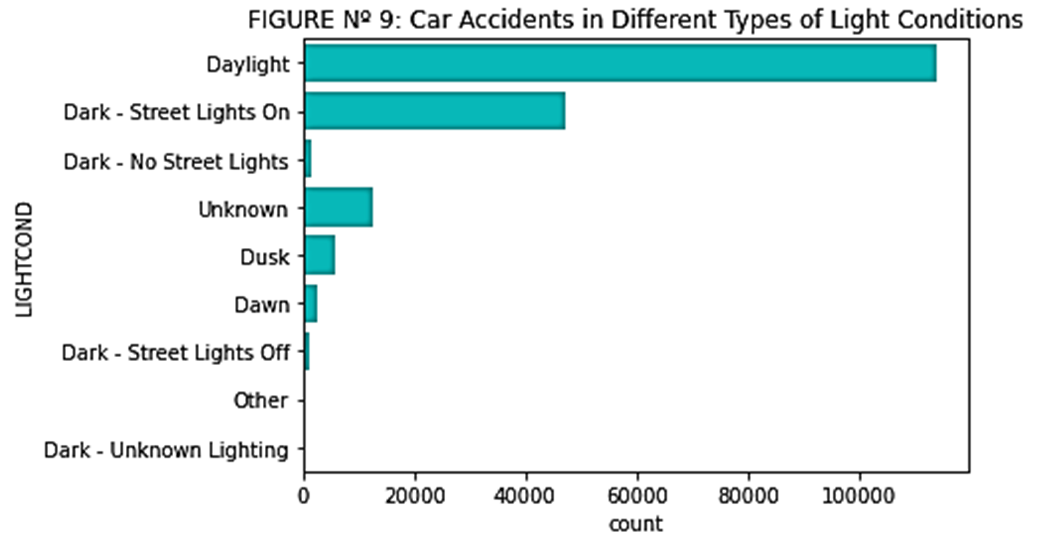


**TABLE Nº 6: *Distribution of Collision Types of Car Accidents in Differents Kinds of Road Conditions***

**

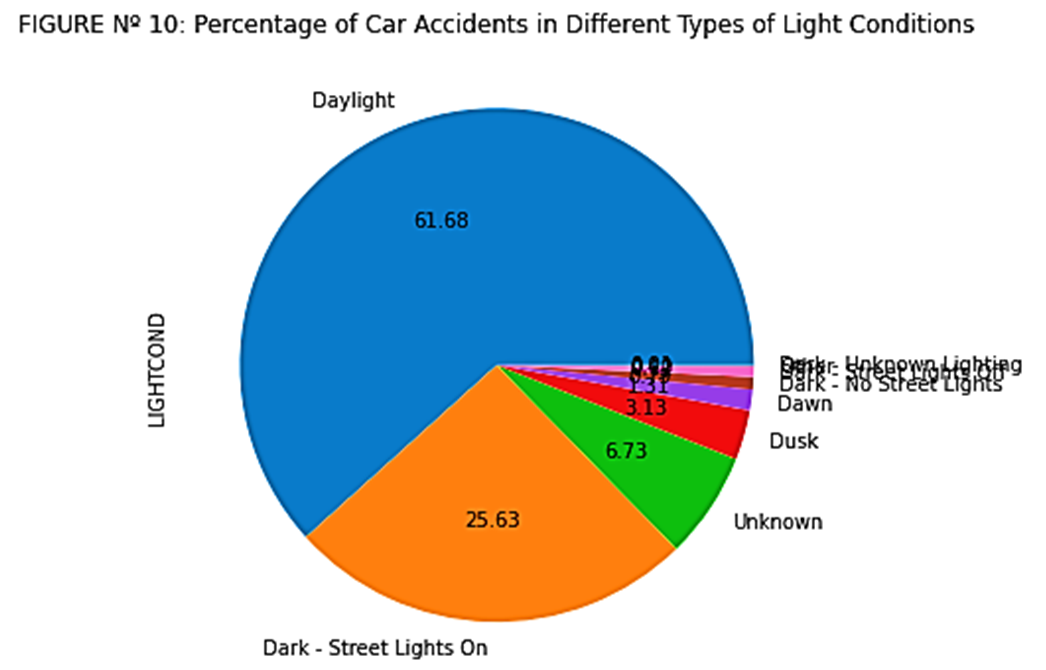
**3.3 Light Condition Attribute Study**

This attribute has 9 light conditions, from daylight to dark. Due to the nature of the results, I will only take three conditions *Daylight, Dark - Street Lights On* and *Unknown.* After plotting the data I see (FIGURE Nº 9) that the major quantity of car accidents occur in perfect light conditions, in this case *Daylight* has 113.845 car accidents, what represents the 61.68% of the total car accidents, the second conditions with the major quantity of car accidents is when the light condition is *Dark - Street Light on,* it has 47.311 car accidents, what represents the 25.63% of total car accidents and in the third position is when doesn’t know the condition of lights, what has 12.430 car accidents with a quote of 6,73% of total car accidents.



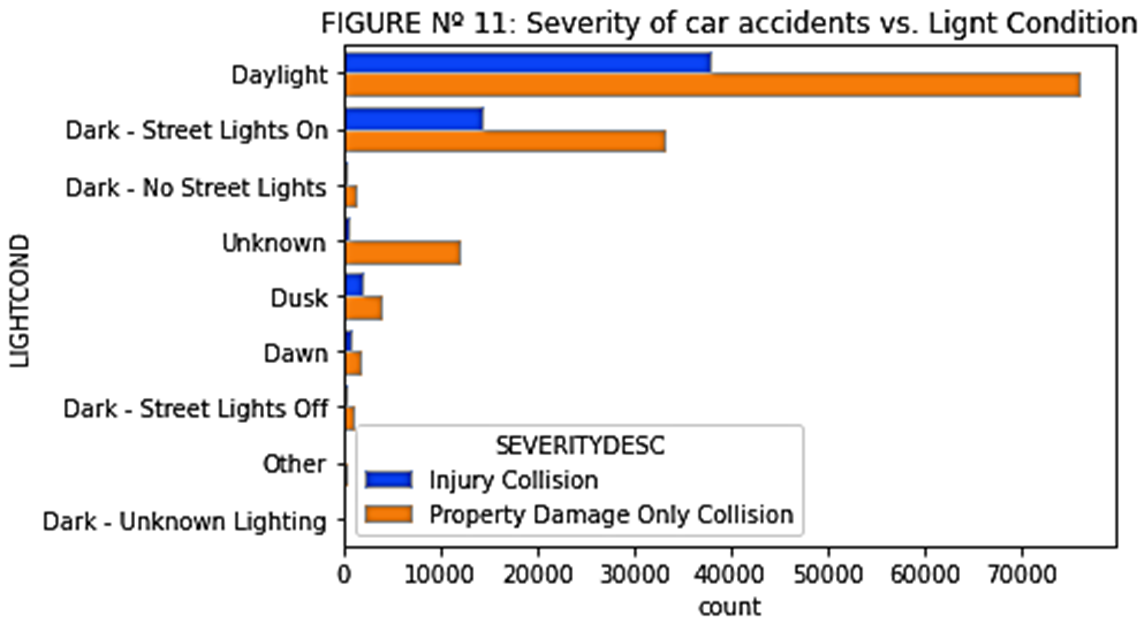
**TABLE Nº 7: *Distribution of Car Accidents in Differents Light Conditions***

|  |  |
| --- | --- |
| **Daylight** | 113.845 |
| **Dark - Street Lights On** | 47.311 |
| **Unknown** | 12.430 |
| **Dusk** | 5.774 |
| **Dawn** | 2.421 |
| **Dark - No Street Lights** | 1.450 |
| **Dark - Street Lights Off** | 1.152 |
| **Other** | 188 |
| **Dark - Unknown Lighting** | 11 |

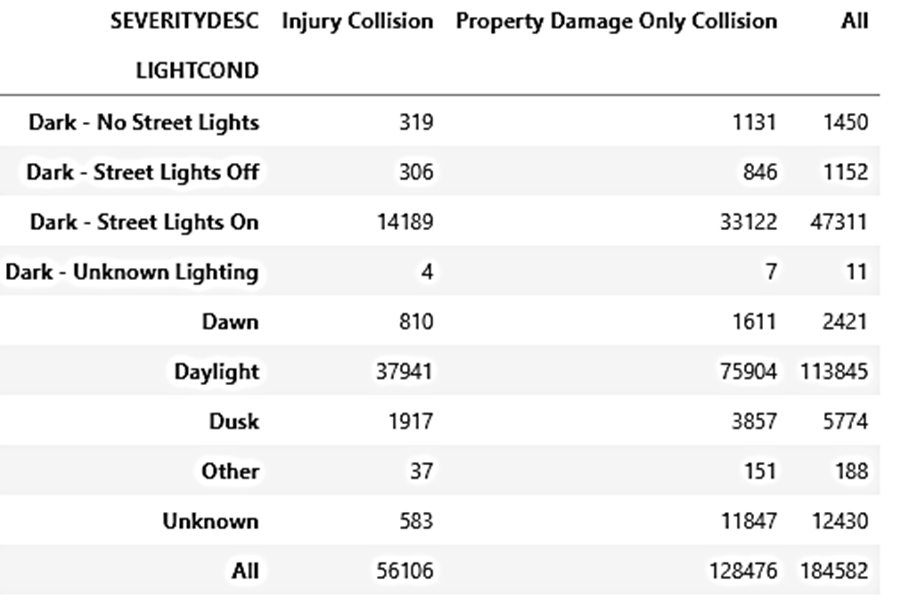


**3.3.1 Relationship Between Light Conditions and Severity of Car Accidents**

Comparing light conditions with the severity of car accidents I can see (FIGURE Nº 11) that *Daylight* has 75.904 accidents with only damage to the property and 37.941 accidents with injury collisions. When we look for the second condition with the major quantity of car accidents *Dark - Street Lights On*  appear with 33.122 car accidents with Damage to the property and 14.189 car accidents which resulted in injury lesions. And when the light condition doesn’t know, the major quantity of car accidents are only damage to the property with 11.847 car accidents.

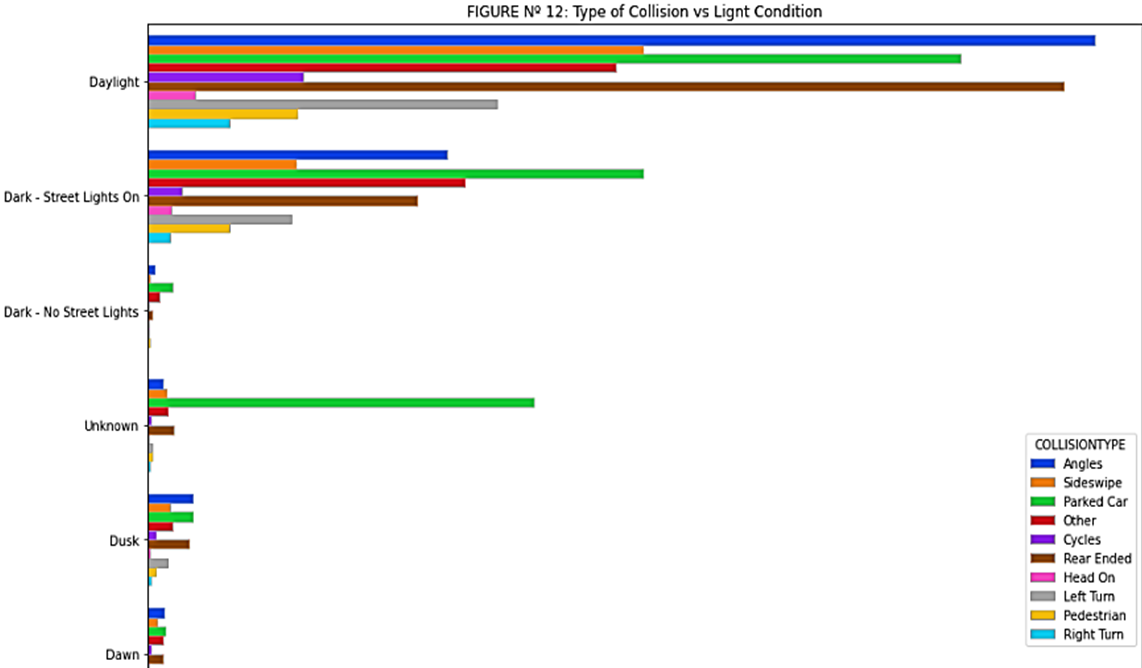


**TABLE Nº 8: *Distribution of Severity of Car Accidents in Differents Kinds of Light Conditions***

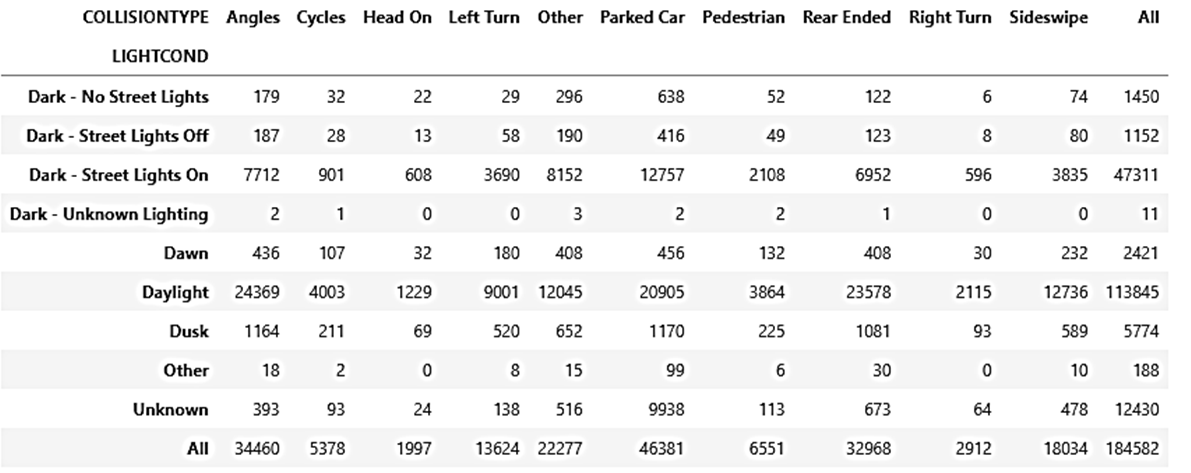


**3.3.2 Relationship Between Light Conditions and Collision Type of Car Accidents**

Studying the relation between light condition and the type of collision, I can see (FIGURE Nº 12) that when the light condition is *Daylight*, there are 3 main collisions. The collision with more car accidents is *Angles,* it has 24.369, the next one is *Rear Ended,* this type of collisionhas 23.578 car accidents and there are 20.905 accidents which involved parked cars. The other light condition with the major quantity of car accidents is *Dark -Street Lights On,* in this light condition the statistics change from *Daylight,* the most common collision type is *Parked Car,* it has 12.757 car accidents, the second one is *Other* type of collision not specifying in this dataset, it contains 8.152 cas accidents, It is follow by *Angles,* which has 7.712 cas accidentes and the other type of collision with a higher number of car accidents is *Rear Ended* with 6.952 car accidents. It is necessary to notice that when the light condition is *Unknown*, there are 9.938 accidents when the cars are parked.

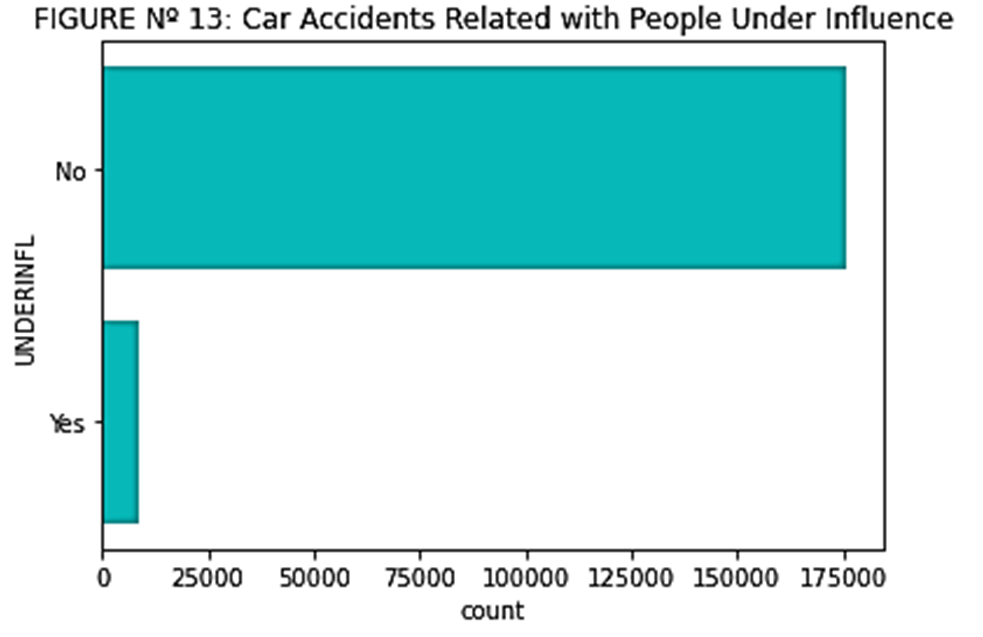


**TABLE Nº 9: *Distribution of Collision Types of Car Accidents in Differents Kinds of Light Conditions***

**

**3.4 Under Influence Attribute Study**

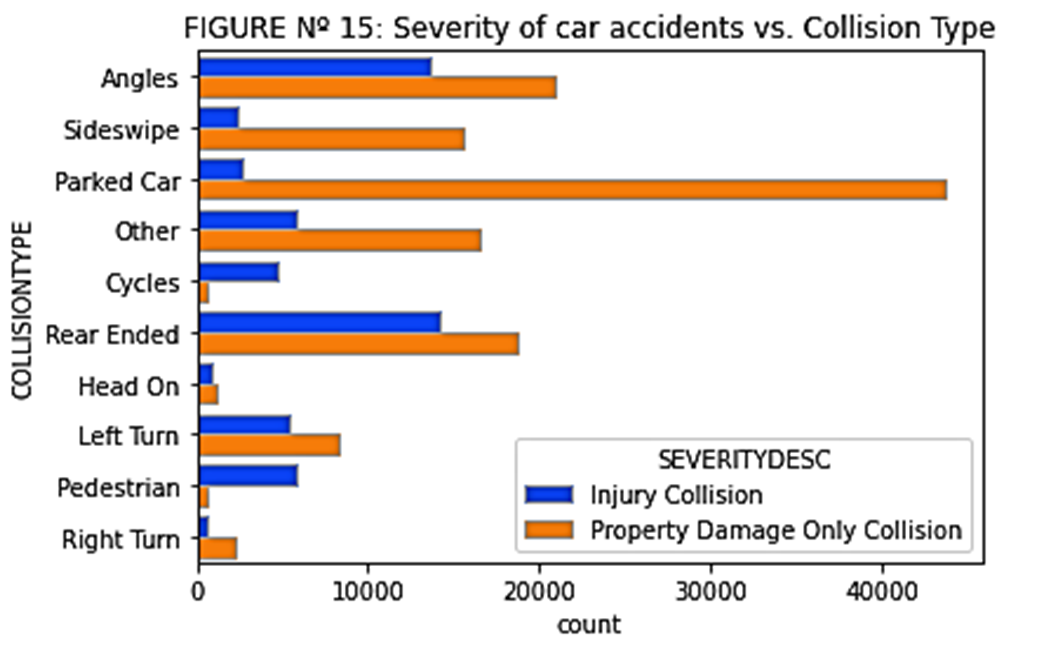
To know if the different car accidents occurred in perfect weather, road and light conditions, I decided to study if the people were under the influence of alcohol or any drugs when they got in those car accidents. I can see (FIGURE Nº 14) that major quantities of car accidents were not provoked by people which were under the influence of alcohol or drugs. As it is shown in the FIGURE Nº 13, the feature *NO,* who represent to the people which weren’t under any influence has 175.702 car accidents and the feature *YES,* who represent to the people which were under any influence has only 8.880 car accidents.



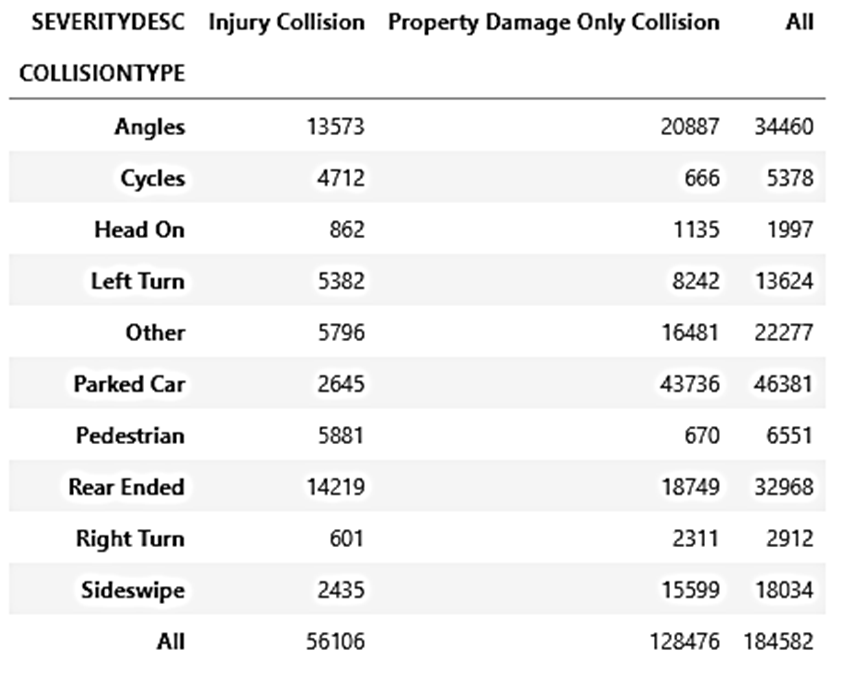


**3.5 Relationship Between Collision Type and the Severity of Car Accidents**

The attribute collision type has 10 different types of collision, however in this section I only want to know which collisions are more likely to be injury collisions. I can see in the FIGURE Nº 15, that there are 6 collisions with a high number of injury collisions, they are *Angles, Cycles, Rear Ended, Head On, Left Turn* and *Pedestrian.*

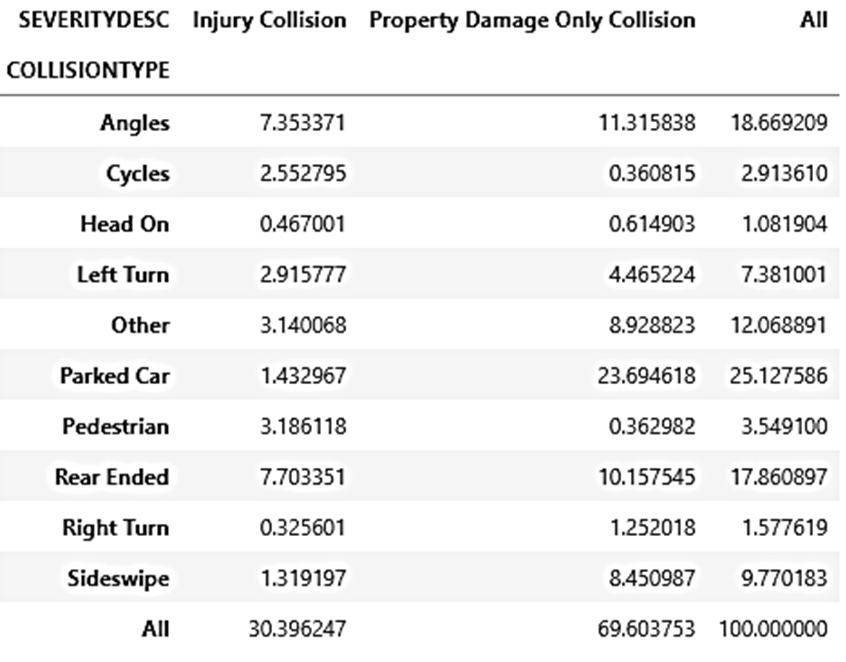
**

**TABLE Nº 10: *Distribution of Collision Types and the Severity of Car Accidents***

**

In the total car accidents independently of the type of collision, the 30.4% of accidents are injury collision and the 69.6% are accidents which only damage to the property, I can see in the TABLE Nº 11, that *Angles* with 7.35%of accidents which result in injury collisions and *Rear Ended* with 7.70% in the same type of collision, are the main collisions types that end in injury results.

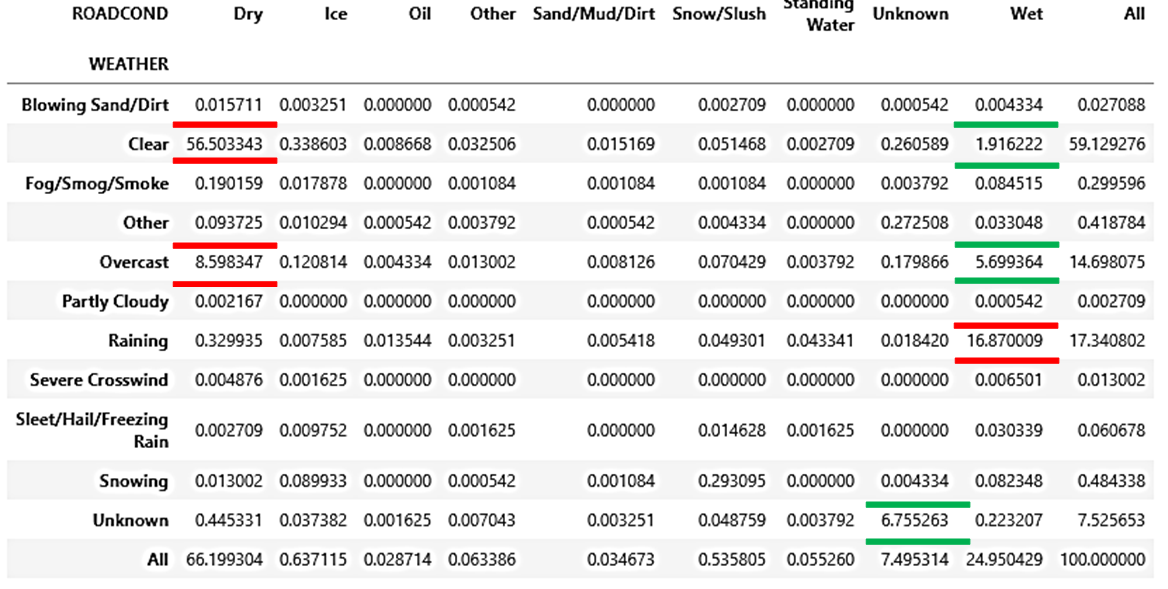
**TABLE Nº 11: *Distribution of Percentage of Collision Types and the Severity of Car Accidents***

**

**3.6 Relationship Between Weather and Road Condition**

In this segment, I will only explain the results in percentage, because I want to try to answer one of the main questions established at the beginning of this project, the possibility of getting in a car accident. As I mentioned before when I was explaining the results of weather, I said that there is three weather which have the major quantity of car accidents, those are *Clear, Overcast* and *Raining,* and combining them with the differents road conditions, I can see (TABLE Nº 12) that 56,50% of total car accidents occur when the weather is clear and the road is dry, it means that the majority of car accidents occur in safe conditions. The 16,87% of car accidents occur when the weather is raining and the road is wet, and the 8,59% of accidents occur when the weather is overcast and the road is dry. Furthermore, there are 6,75% of car accidents in unknown conditions.

**TABLE Nº 12: *Distribution of Percentage of Car Accidents Between Weather and Road Condition***

******

**3.7 Relationship Between Weather and Light Condition**

As I mentioned before, *Weather* has three main kinds of weathers with the major quantity of car accidents, and seeing the statistics (TABLE Nº 13) between this attribute with light condition, I can say that 41,68% of car accidents occur when the weather is clear and the light condition is *Daylight,* however, when is dark with street lights on and the weather is clear the percentage of car accidents is 13,57%.

The 9,31% accidents occur when it is overcast and there is daylight, and when it is dark with street lights on, there are 3,99% of car accidents. When it is raining, 8,88% of car accidents occur when it is daylight, and 6,87% when it is dark with street lights on. And finally, 5,29% of car accidents occur in unknown conditions.

**TABLE Nº 13: *Distribution of Percentage of Car Accidents Between Weather and Light Condition***

****

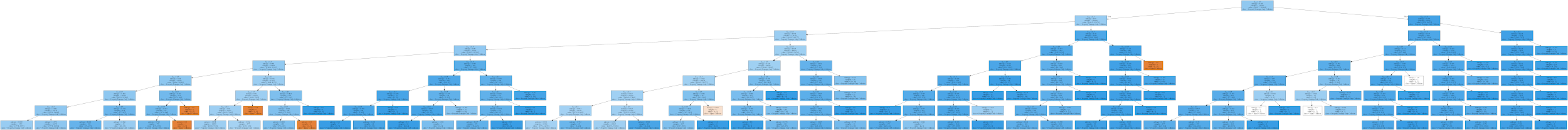
**4. Modeling**

Due to the nature of the data, I decided to use classification models, being specific *Decision Tree*. This algorithm is useful to this situation because: 1. The data is categorical, what means that it’s necessary to be careful with the data and how I process it. 2. The objectives of this project are knowing the possibility of getting in a car accident and the severity of them, which means that there are more than one statistic, there are several possibilities as a result.

**4.1 Decision Tree Algorithm with Weather, Road and Light Condition vs. Severity of Car Accidents**

To start to build the model, I set up the parameters of the Decision Tree algorithm, putting a **criterion = ‘entropy’** and a **max\_depth = 8,** to train it and test it, I set the text size in 0.2 and a random state in 1, to get better results. After training it and testing it, I look for the error in the model, using the R2 algorithm, and the accuracy of the model using three attributes (*Weather, Road Condition, Light Condition)* is 0.6994.

**PLOT Nº 1: *Decision Tree Between Weather, Road Condition and Light Condition vs. Severity of Car Accidents***



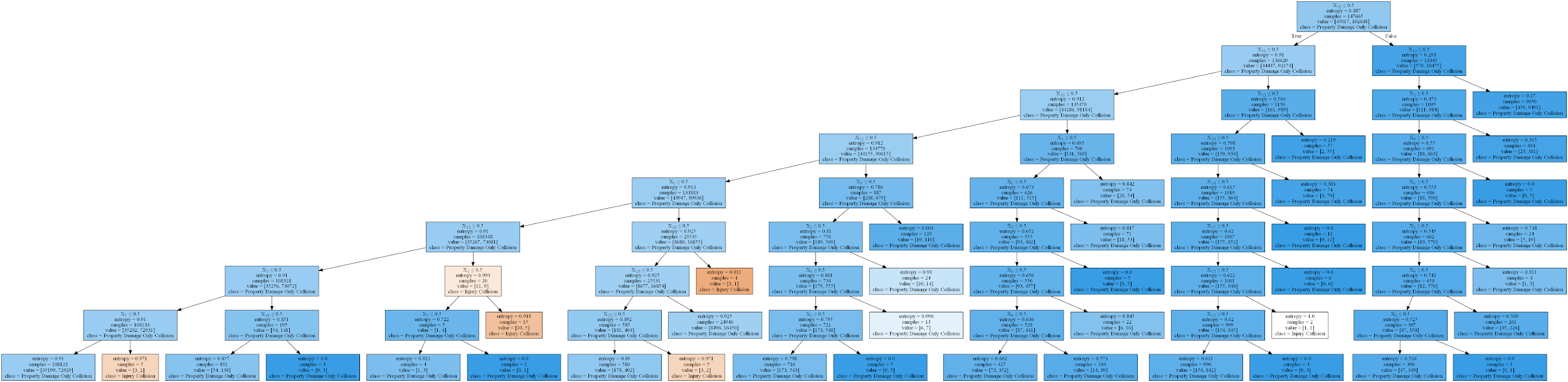
[*Link to a better quality*](https://drive.google.com/file/d/16EEchdZ5cTC0OyDDT9-Naeux94Omp3jO/view?usp=sharing)

As is shown in the PLOT Nº 1, the majority steps in the plot are *Property Damage Only Collision*, which reinforce the accuracy of the model.

**4.2 Decision Tree Algorithm with Weather and Road Condition vs. Severity of Car Accidents**

To look for other types of result, I put only two attributes (*Weather and Road Condition)* to studying them with the severity of car accidents, in order to see if any change. In this case I follow the same steps that I used in the first Decision Tree algorithm. The accuracy of this algorithm is 0.6992, this result is very similar to the algorithm above.

**PLOT Nº 2: *Decision Tree Between Weather, Road Condition vs. Severity of Car Accidents***



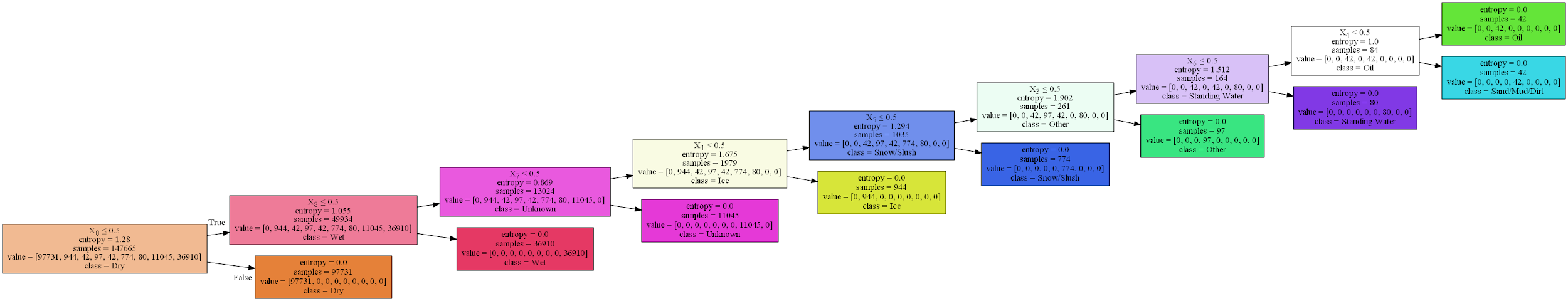
[*Link to a better quality*](https://drive.google.com/file/d/1rY4Ttd6ZCkAVMsjDTP5mVdNVKi6Mxd0U/view?usp=sharing)

In this case, the PLOT Nº 2, shows more *Injury Collision* than the PLOT Nº 1, but I can say that is normal, because there is less data to compare.

**4.3 Decision Tree Algorithm with Weather vs. Road Condition**

In this algorithm I will test the explanation in the TABLE Nº 12, all this in order to get more information to answer the established questions. After training and testing it, the accuracy of the mode is 1.0 using a **max\_depth = 8.**

**PLOT Nº 3: *Decision Tree Between Weather vs. Road Condition***

******

[*Link to a better quality*](https://drive.google.com/file/d/1ul-WWfbczrRpN-z8eYRAmqE-6JiQ5ukY/view?usp=sharing)

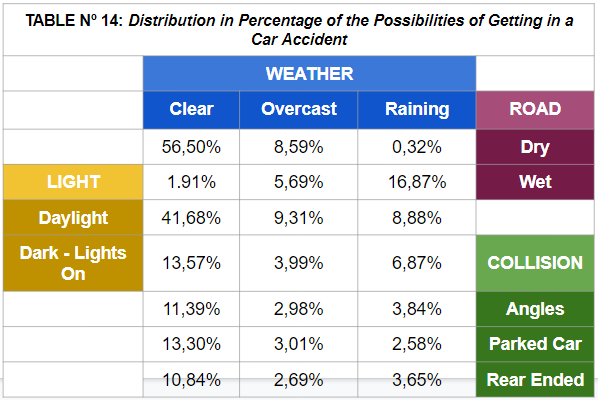
As it’s shown in the PLOT Nº 3, I can see that there is a hierarchy in the road conditions from the feature with more possibilities of getting in a car accidents to features with less possibilities of getting in a car accidents. This result is similar to the statistics in the TABLE Nº 12.

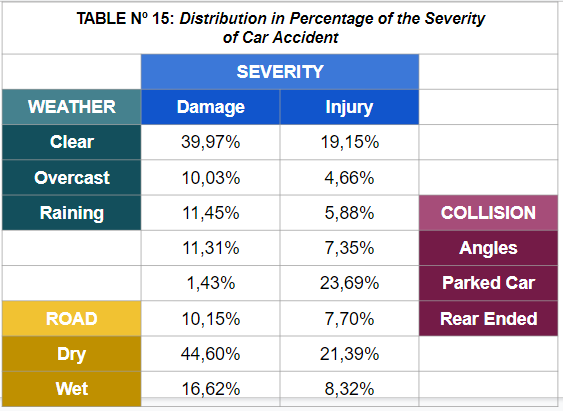
**5. Results**

After study the data and making several inferentials, I can say:

1. There are three main weathers where the majority of car accidents occur, being *Clear* with 59,13%, *Raining* with 17,34 % and *Overcast* with 14,70%. Comparing those weathers with the severity of car accidents, the 69,60% of car accidents have only damage to the property and the other 30,40% suffer injury collisions, being *Clear* with the major quantity of injury lesions, in total 19.15% of total car accidents. If I compare the same weathers with the different road conditions, I can see, when the road is dry 56,50% of total accidents occur when it is clear and 16,87% occur when the weather is raining and the road is wet. And 61.67% of car accidents occur when it is daylight, and the 41,68% of accidents happen when it is clear, and if the weather is overcast and it’s daylight the percentage of car accidents is 9,31%.
2. The most popular collision types between the different weathers, road and light conditions are three, being *Angles* with 18,66%, *Parked Car* with 25,12% and *Rear Ended* with 17,86% of total car accidents (See TABLE Nº 11). The collision types with more injury collisions are *Angles* with 7,35% and *Rear Ended* with 7,70% of total car accidents.

The next tables show some statistics related to the probability of getting in a car accident and the severity of them.





**6. Conclusion**

Giving a brief conclusion, the major quantity of car accidents occur when the weather is clear, it happens because maybe when the weather is nice, we as humans think that nothing bad will happen us, because the weather is good, we don’t take precautions, and in this statistics I can see when the weather is not good, the quantity of car accidents drop significantly, and this happens because, in this kinds of weather we take precautions while we are driving.

Other aspect that I saw was the severity and the type of collision, in every kind of weather the 69.6% of car accidents resulted in only damage to the property while the 30,4% of car accidents resulted in injury to the people involved, but when the weather is clear the 19,15% of total car accidents resulted in injury collision, being *Angles, Parked Car* and *Rear Ended* the most common type collision with variance in the differents kinds of weather, it’s necessary to clarify that other types of collision occur but in a lesser extent.

The same happen when the road and light condition are safe, the majority of car accidents occur in this perfect conditions, what is really strange because, when I started this project I could imagined that the majority of car accidents occurred in unsafe road, light and weather conditions, but as I said, in the first paragraph, when there’s no apparent danger, we down our alerts and thought that all will be right, and we get in bad situations like car accidents.

And to finish, I only recommend to the car drivers, it is only to take care and think that in nice weather, drive in good light conditions and perfect roads, car accidents could happen.