Predicting Accident Severity

Alexa Boncimino

September 2020

1. **Introduction**

This project aims to predict accident severity. It is targeted at drivers who may wish to know the likelihood of an accident’s severity based on road/weather condition attributes when making travel plans. It is meant to help drivers solve a lack of knowledge problem surrounding road safety.

With access to the machine learning information in this report they can plan the safest route of travel that takes into consideration the likelihood of accident’s severity based on weather, road conditions, and light conditions.

1. **Data Description**

I will be using the shared data in part one of this capstone project to predict accident severity. It is supervised data meaning that the machine learning output will learn based on inputs with known outputs.

The target variable is “Severity Code” which measures the severity of an accident as either 1 or 2.

Attributes or dependent variables used to calculate the severity of an accident are weather, road conditions, and light conditions.

1. **Methodology and Data Processing**
2. *Data Cleaning and Manipulation*

The original data set had 36 attributes. I started my analysis by limiting the data set to the dependent variable (Severity Code) and my three independent variables weather, road conditions, and light conditions.

Because my three independent variables were all categorical, I needed to transform them into a numeric scale.

For Weather

|  |  |
| --- | --- |
| 0 | Unknown |
| 1 | Clear |
| 2 | Overcast |
| 3 | Raining |
| 4 | Snowing |
| 5 | Fog/Smog/Smoke |
| 6 | Sleet/Hail/Freezing Rain |
| 7 | Blowing Sand/Dirt |
| 8 | Other |

For Road Conditions:

|  |  |
| --- | --- |
| 0 | Other |
| 1 | Dry |
| 2 | Wet |
| 3 | Snow/Slush |
| 4 | Ice |
| 5 | Standing Water |
| 6 | Sand/mud/dirt |

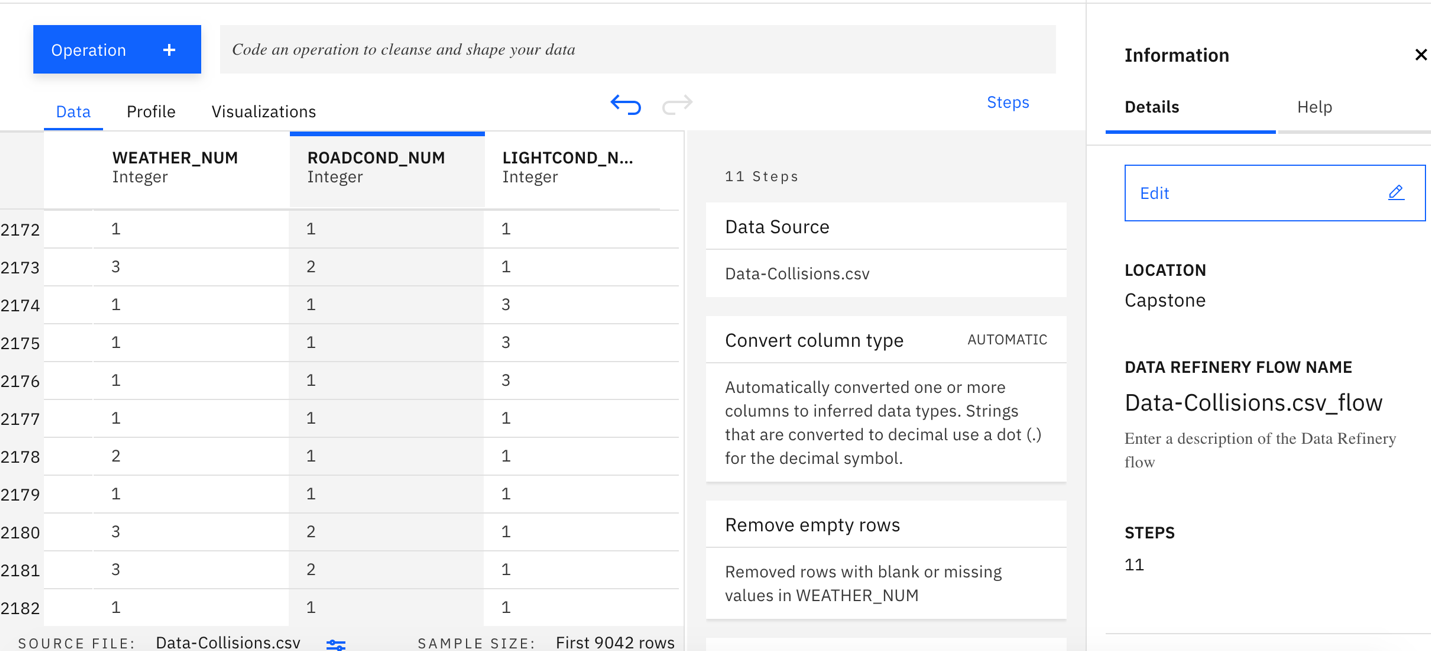
For Light Conditions:

|  |  |
| --- | --- |
| 0 | Other |
| 1 | Daylight |
| 2 | Dark (no streetlights/streetlights off) |
| 3 | Dark (streetlights |
| 4 | Dawn |
| 5 | Dusk |

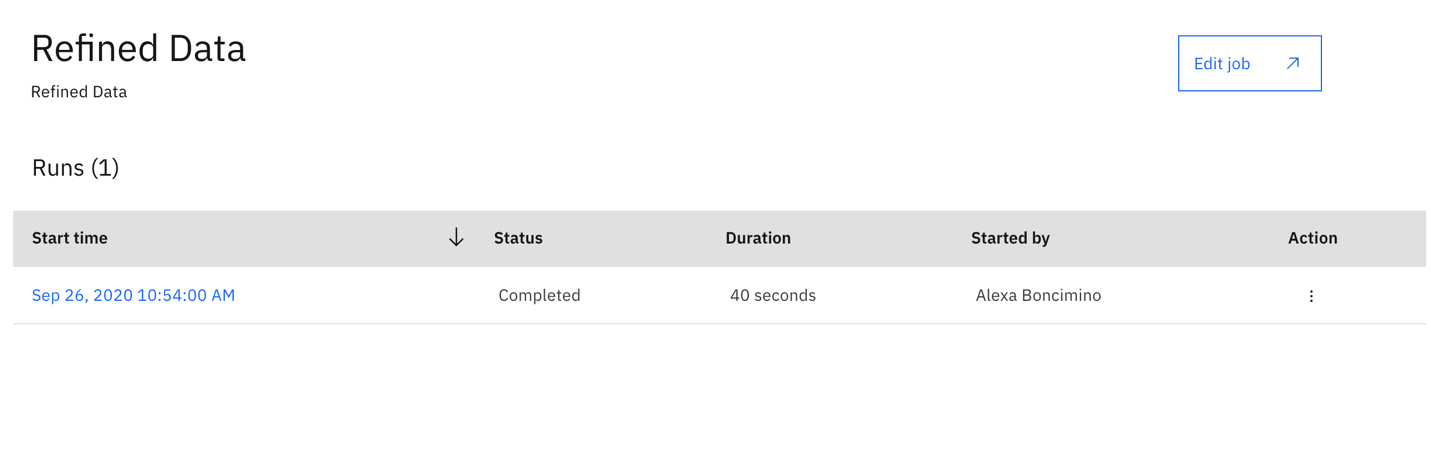
1. **Dropping unknown data columns = 0 from columns WEATHER\_NUM, ROADCOND\_NUM, LIGHTCOND\_NUM**

This was done through data refinery flow tab in Watson Studio

1. **Converting Columns Weather, ROADCOND\_NUM, and LIGHTCOND\_NUM to integer from string**

****

Completed data refinery run job:



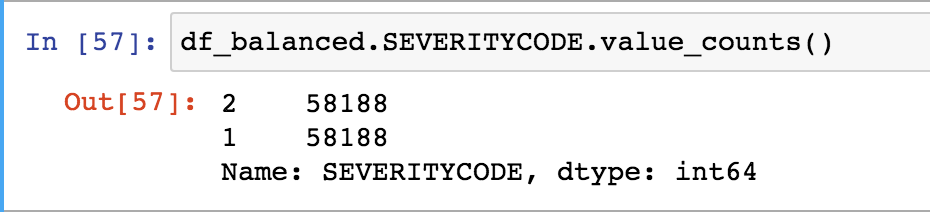
1. ***Balancing the data***

The original data was unbalanced.

Count: Severity code= 2 : 58188

Count: Severity code =1 : 136485

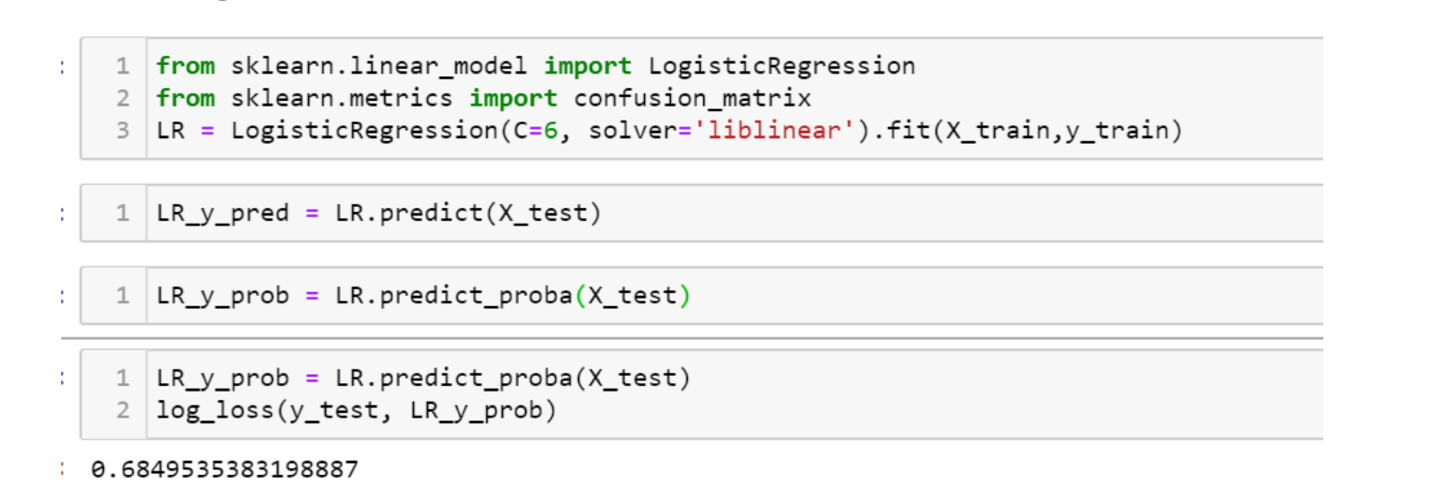
In order to balance the data I down-sampled the majority class so that both samples were equal to 58,188.



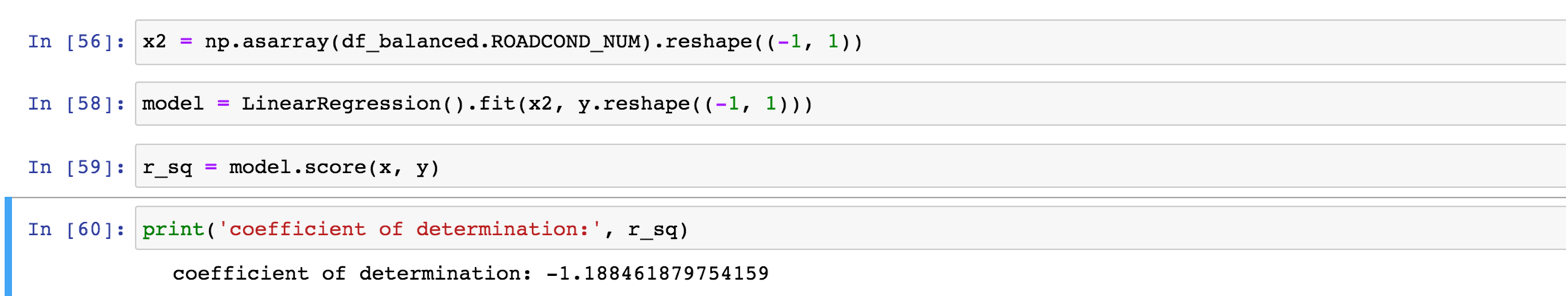
**Methodology Linear Regression:**

In the methodology section of my report I developed three linear regression models.

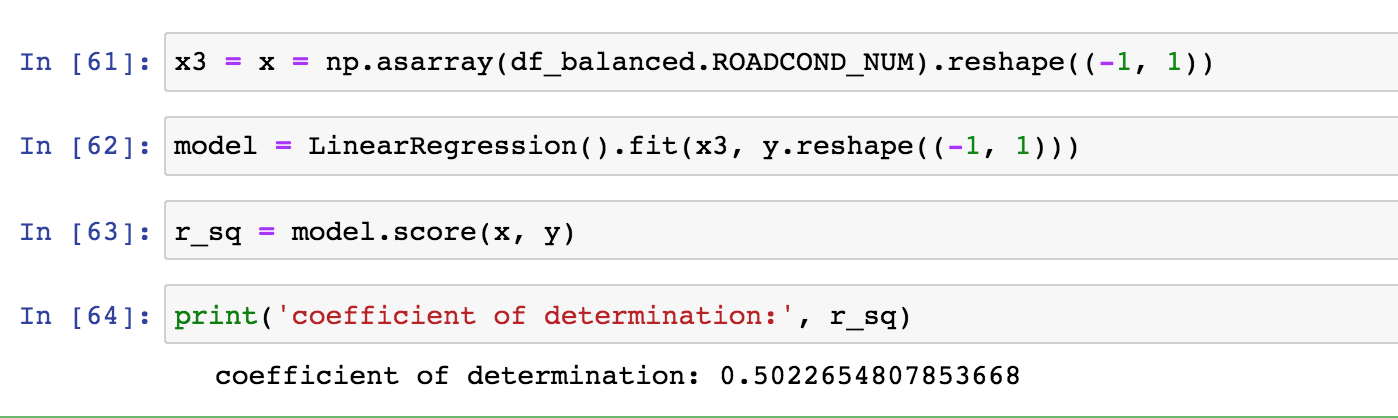
Model 1: Examined the linear relationship between a change in Weather Conditions and a change in the severity of the accident



Model 2: Examined the linear relationship between a change in light conditions and the severity score.



Model 3: Examined the linear relationship between a change in road conditions and the severity score.

**

1. **Results and Evaluation**

These three models represent the change in severity depending on weather conditions, light conditions, and road conditions.

As weather conditions become more sever there is a 68% change in the severity of the accident.

As light conditions increase there is a -1.18% decrease in the severity of the accident.

As road conditions became less favorable there was a .50% increase in the severity score.

1. **Discussion**

As weather conditions, light conditions, and road conditions become more dangerous the severity of the accident also rises.

We know this by creating an index of these three conditions and determining their linear relationship with the severity score.

1. **Conclusion**

We can conclude that the three independent conditions have an impact on travel safety. As these conditions become more unstable drivers may wish to take a different route.