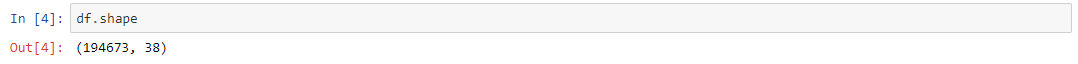
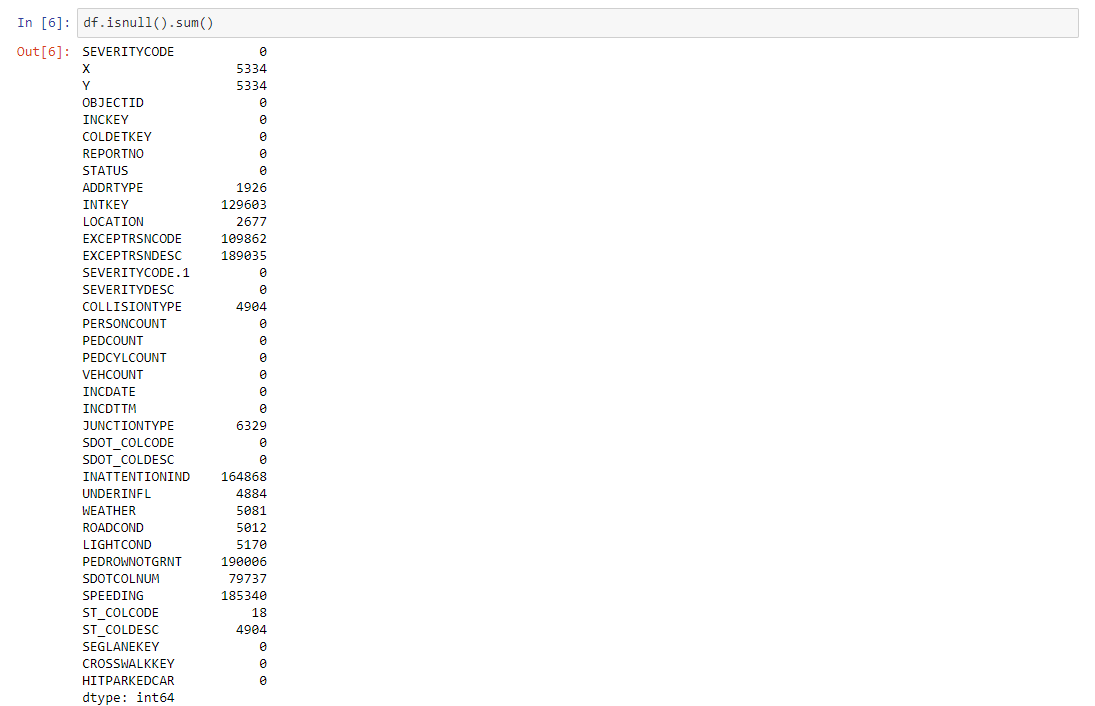
By checking the dataset, the dataset has a total of 38 attributes and 194,673 data.

At the same time, many attributes have null values.



I will process the data like this:

1. Remove text attributes like description that cannot be converted into numeric values. Then remove the unrelated latitude and longitude. Then fill in the null value (delete or mean or mode) of each attribute according to the attributes of different attributes, and then regularize the data.

2. Divide the data into training set and test set.

3. Establish the Logistics Regression Model and Random Forest model.

4. Optimize the two models and compare their accuracy.

5. Choose the most suitable model.