**Increment - 2**

**Team(10)** Members and Contribution:

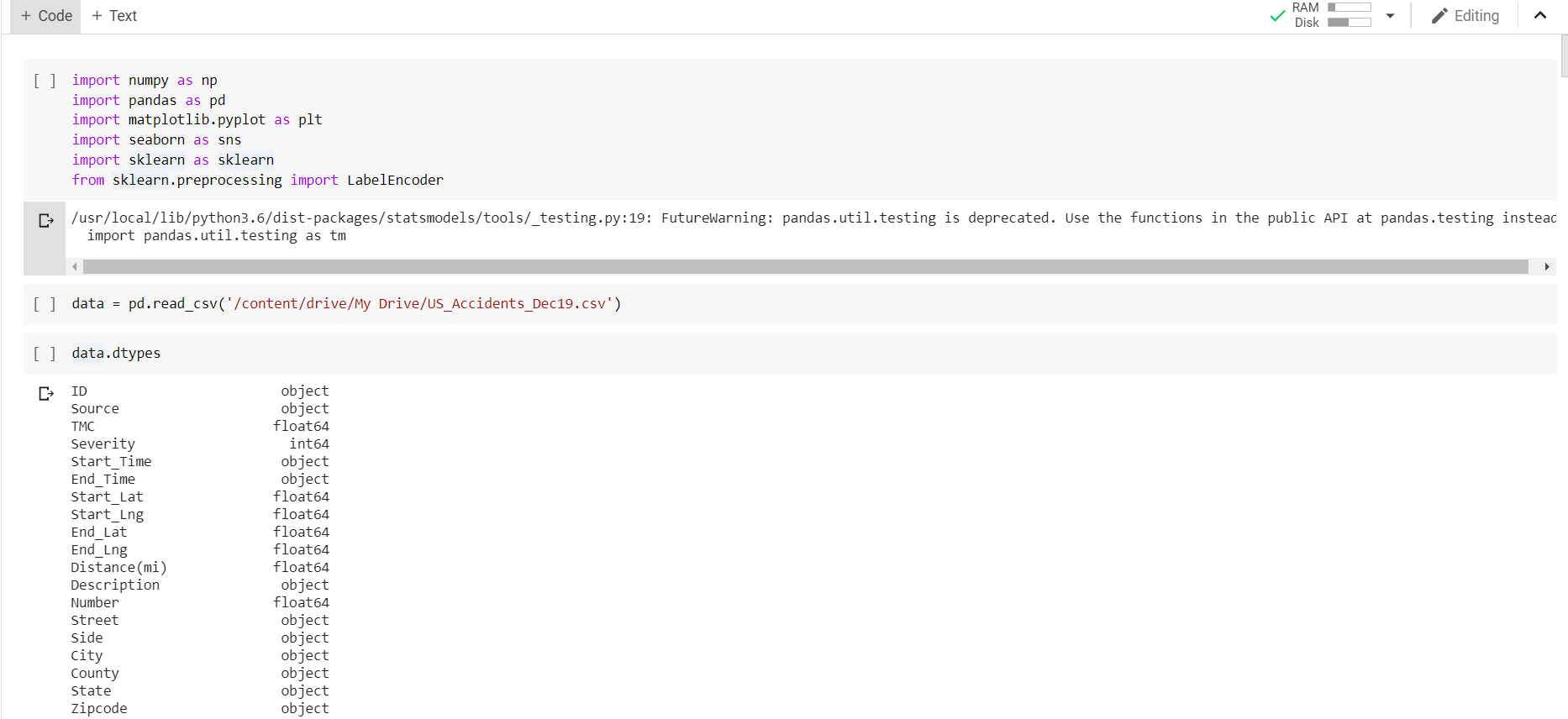
Sai Rohith Guntupally – 8 (30%)

Chaitanya Mallepudi – 14 (40%)

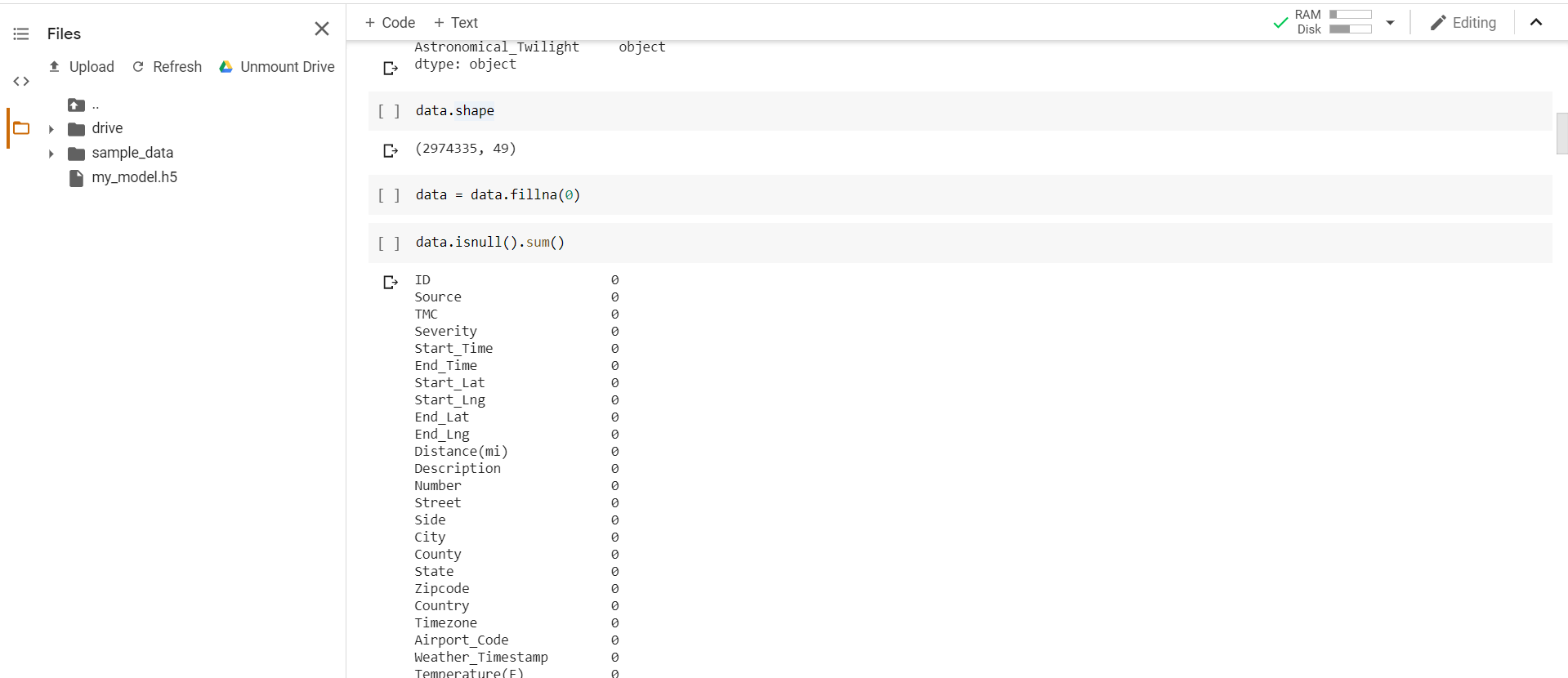
Madhu Varma Rudraraju – 20 (30%)

We have taken the accidents data set from kaggle , Done pre-processing, predict the accident probability using regression, done binomial classification, passed the model to neural network to improve the accuracy of the model and then build the model using keras and saved it, then predicted the probability using the model.

Import the accident dataset and see the dtypes present in the dataset.



Then as a part of pre-processing remove the null values in the dataset.

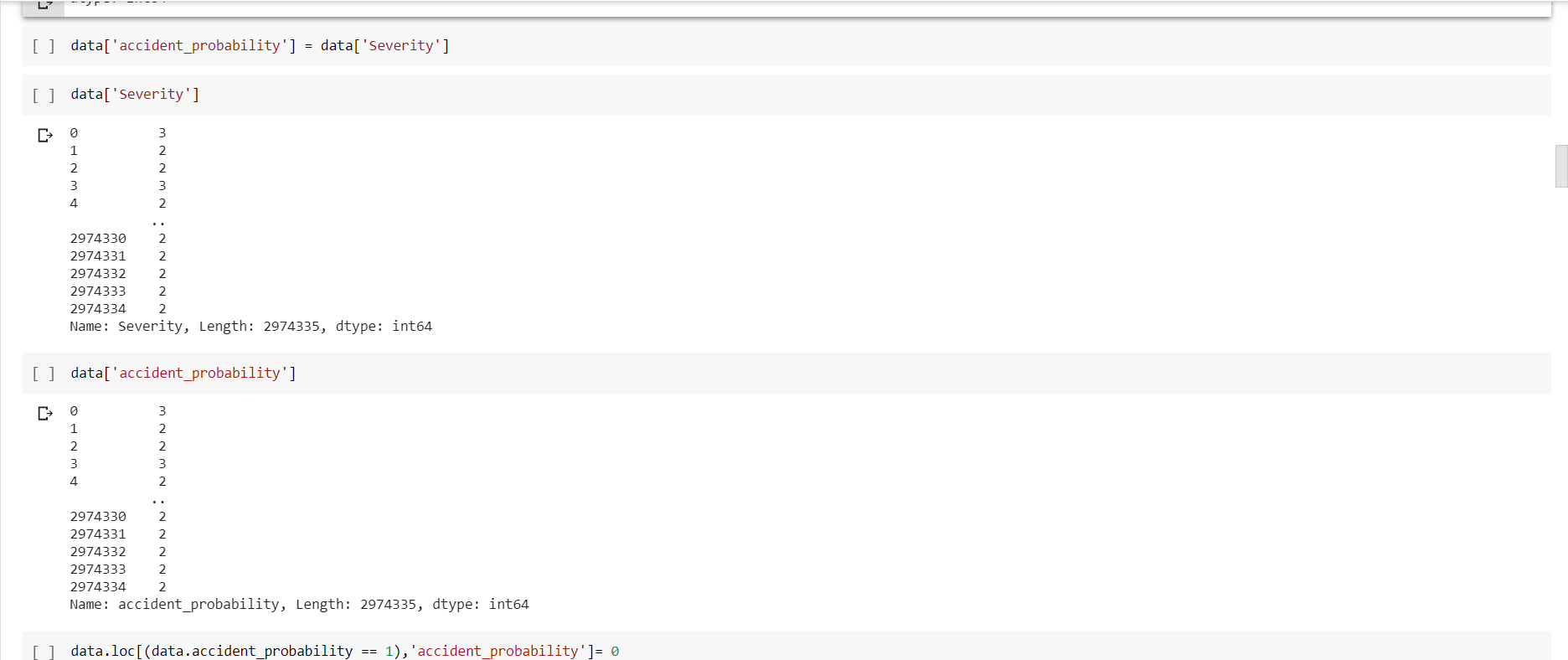


Now, let us include the accident probability variable to the dataset.

The main purpose of this variable is to predict the accident probability based on the given inputs,

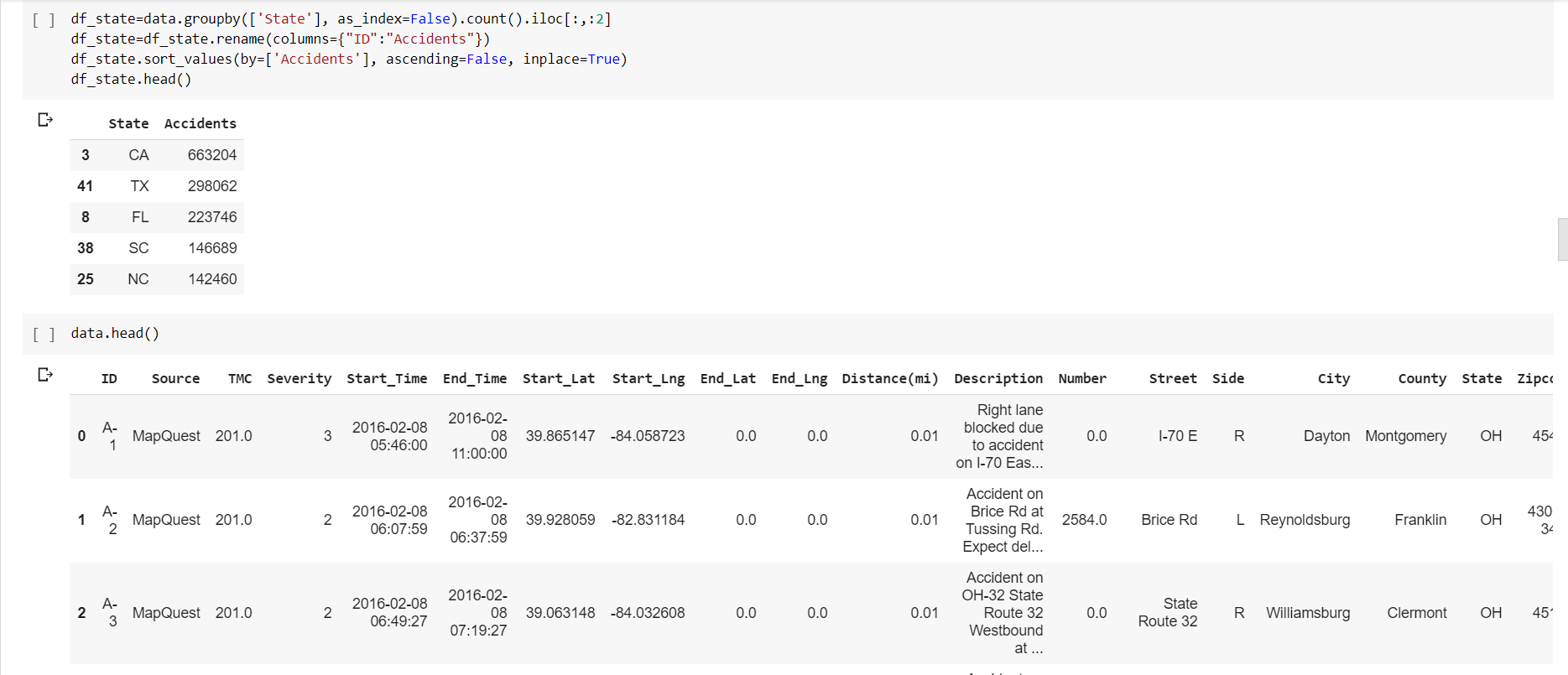
We have assigned this variable value by taking it as dependent variable on Severity.

If the severity value is very low (1) the accident probability is 0 else 1.

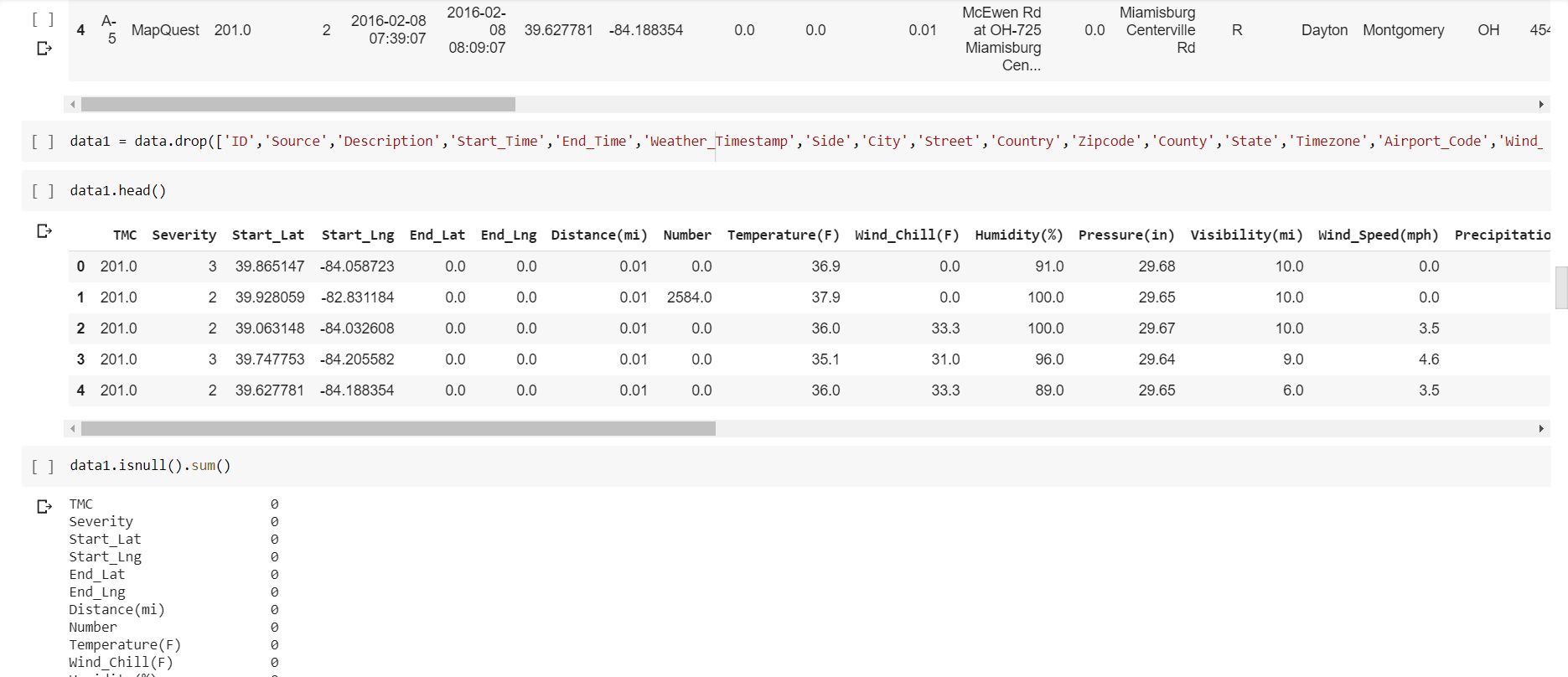




Below are the top 5 states which has most number of accidents.



We are dropping, many of the features which are of no correlation from which we need to predict.



We had mapped the Boolean values or binomial values to 0 or 1.



We had classified the model using Naïve-Bayes classification model since we need to assume the features are to be independent.



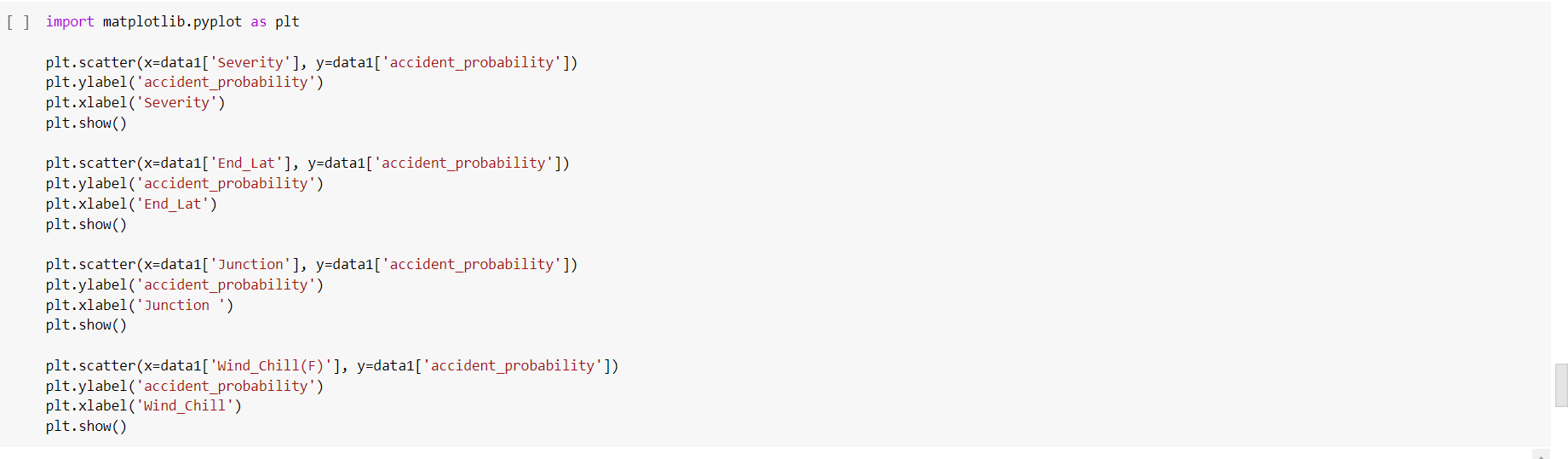
We got the accuracy of 60.72%. Is this enough? Surely, we need to increase the accuracy of the model.

We have taken the top 5 correlated variables that are effecting the accident probability.



We can infer that Severity, End\_Lat, Junction, Wind\_Chill, Distance are those features.

We had plotted those factors and represented graphically how they are related using matplotlib.

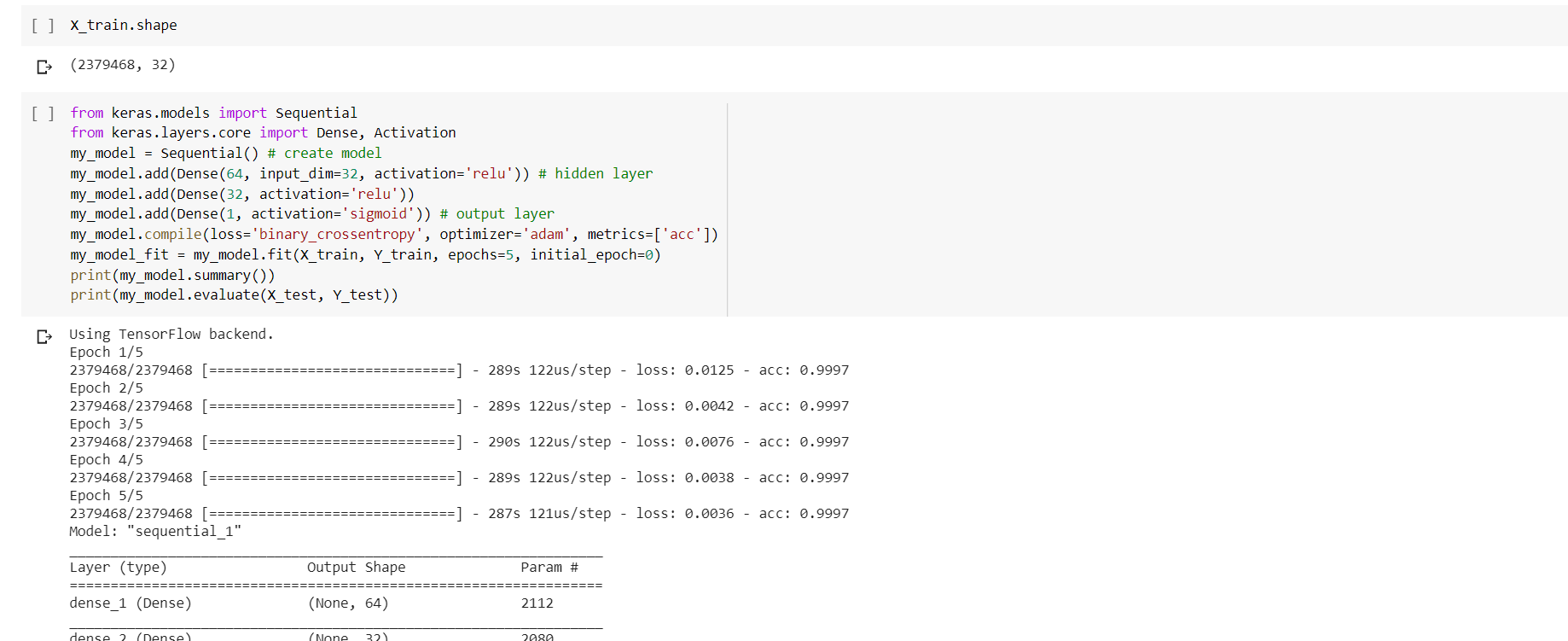


Below are those visualizations.



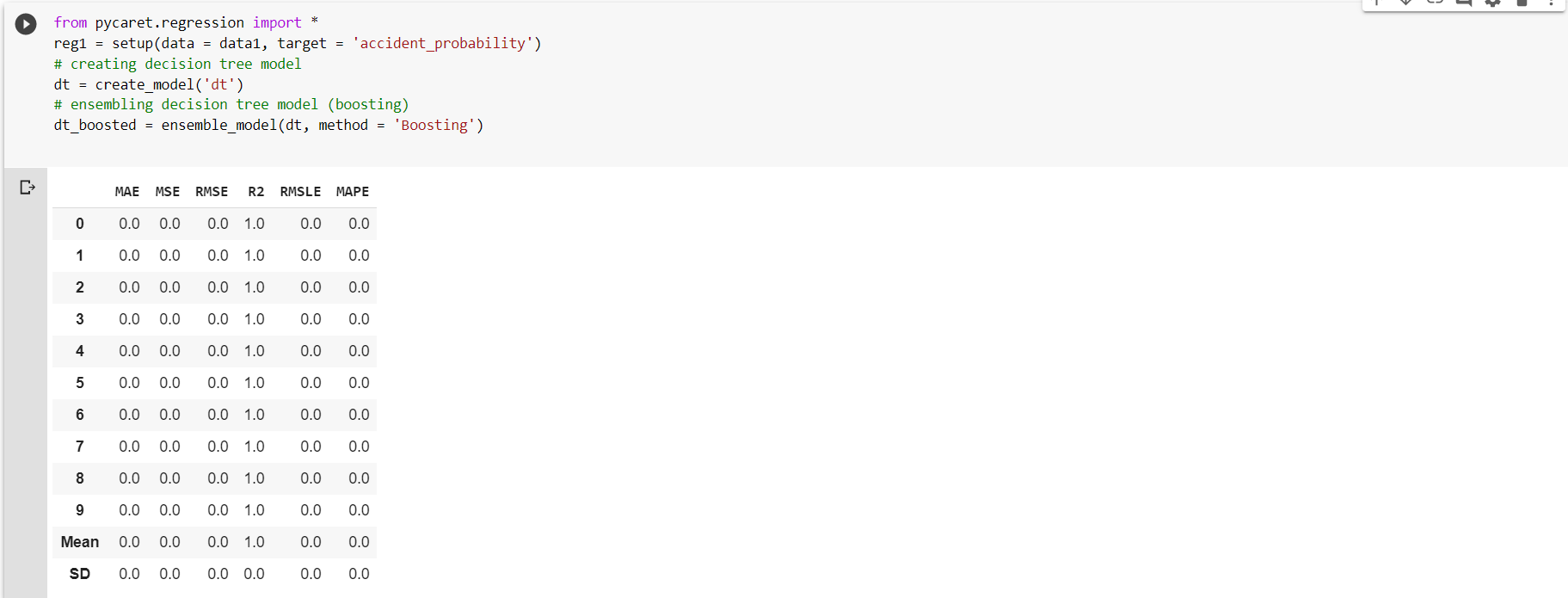


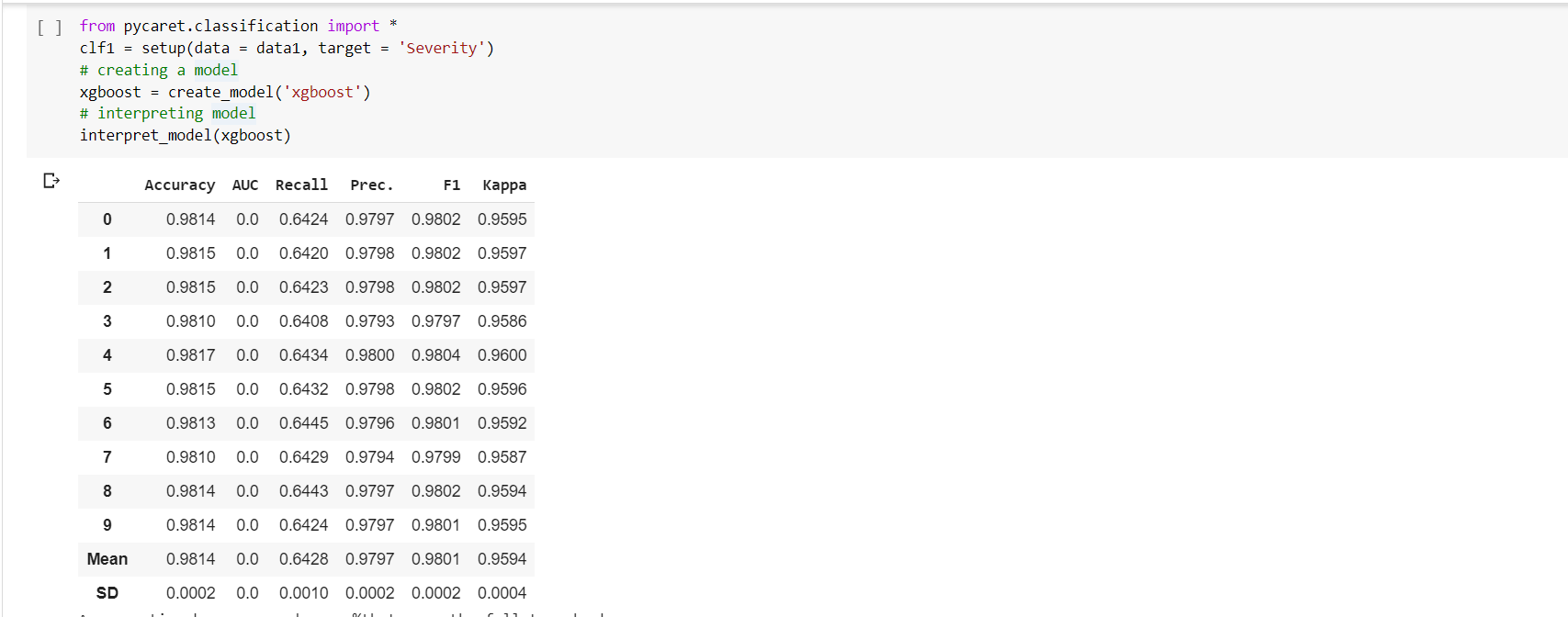
Now the next thing we need to do is to increase the accuracy of the model. So let us pass the model to neural network using sequential model.



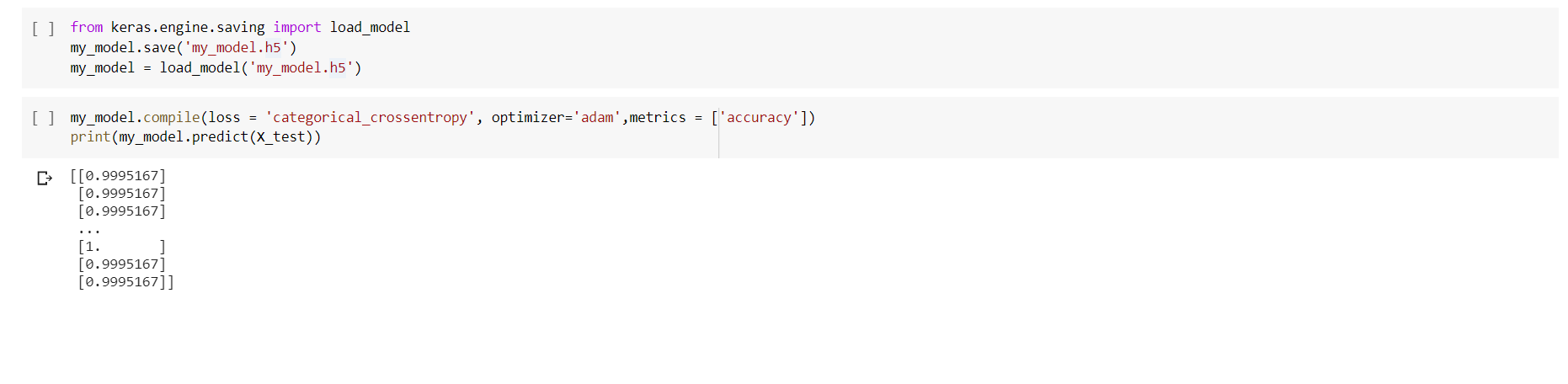
We can see the accuracy increased to 99%.

We have tried to boost the model using the boosting techniques using the pycaret which is a open source module and improved the mean and SD using predefined automated functions.





Then we had built the model using the keras library and predicted the output and saved the model.



**Future Work:**

We are trying to predict the accident probability by manually entering the features through website. As of now are we are facing server error whie trying with flask. We will try to fix the error or find an alternative and build a website by next increment.

**Languages and Libraries:**

* Python
* Matplotib
* Ski-kit learn
* Numpy
* Pandas.
* Tensorflow, Keras

**Software and platforms:**

* Google colab.

**References**:

<https://www.kaggle.com/us-accidents>

<https://pycaret.org/>

<https://stackoverflow.com/>

<https://www.tensorflow.org/api_docs/python/tf/keras/Model>