Introduction/Business Plan:

The purpose of this project is to create a machine learning model that predicts the severity of road accidents based on the road conditions and lighting conditions.

Very often, accidents are caused due to bad road conditions. If it were possible to predict the safety of travelling based on historical data, then a lot of accidents could be prevented.

Data:

The data that will be used is the shared dataset from Coursera’s IBM Data Science Certificate-Applied Data Science Capstone. The shared dataset is from Seattle city.

The attributes are as follows:

SEVERITYCODE

X

Y

OBJECTID

INCKEY

COLDETKEY REPORTNO

STATUS

ADDRTYPE

INTKEY ...

ROADCOND

LIGHTCOND

PEDROWNOTGRNT SDOTCOLNUM

SPEEDING

ST\_COLCODE

ST\_COLDESC

SEGLANEKEY

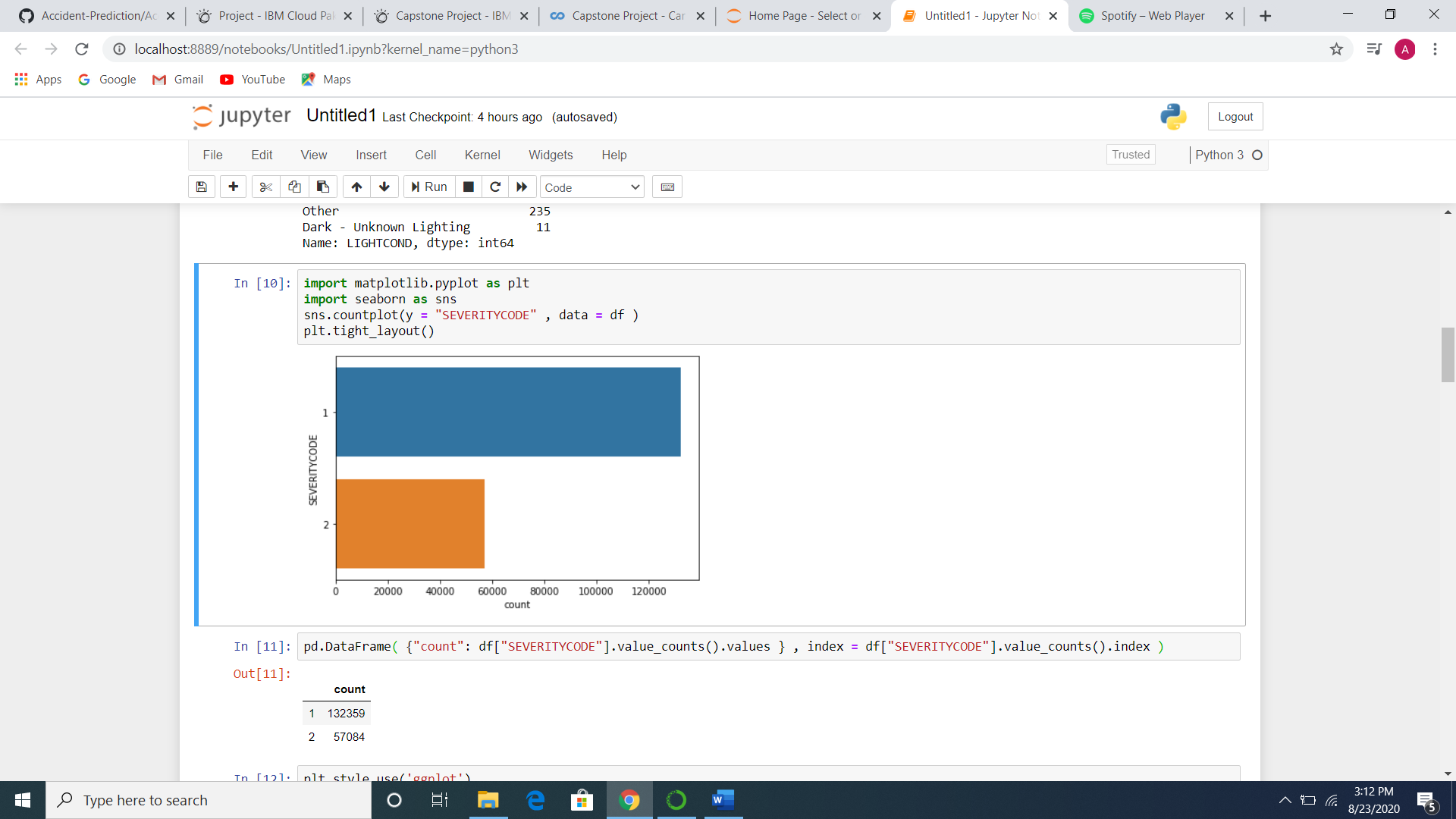
CROSSWALKKEY

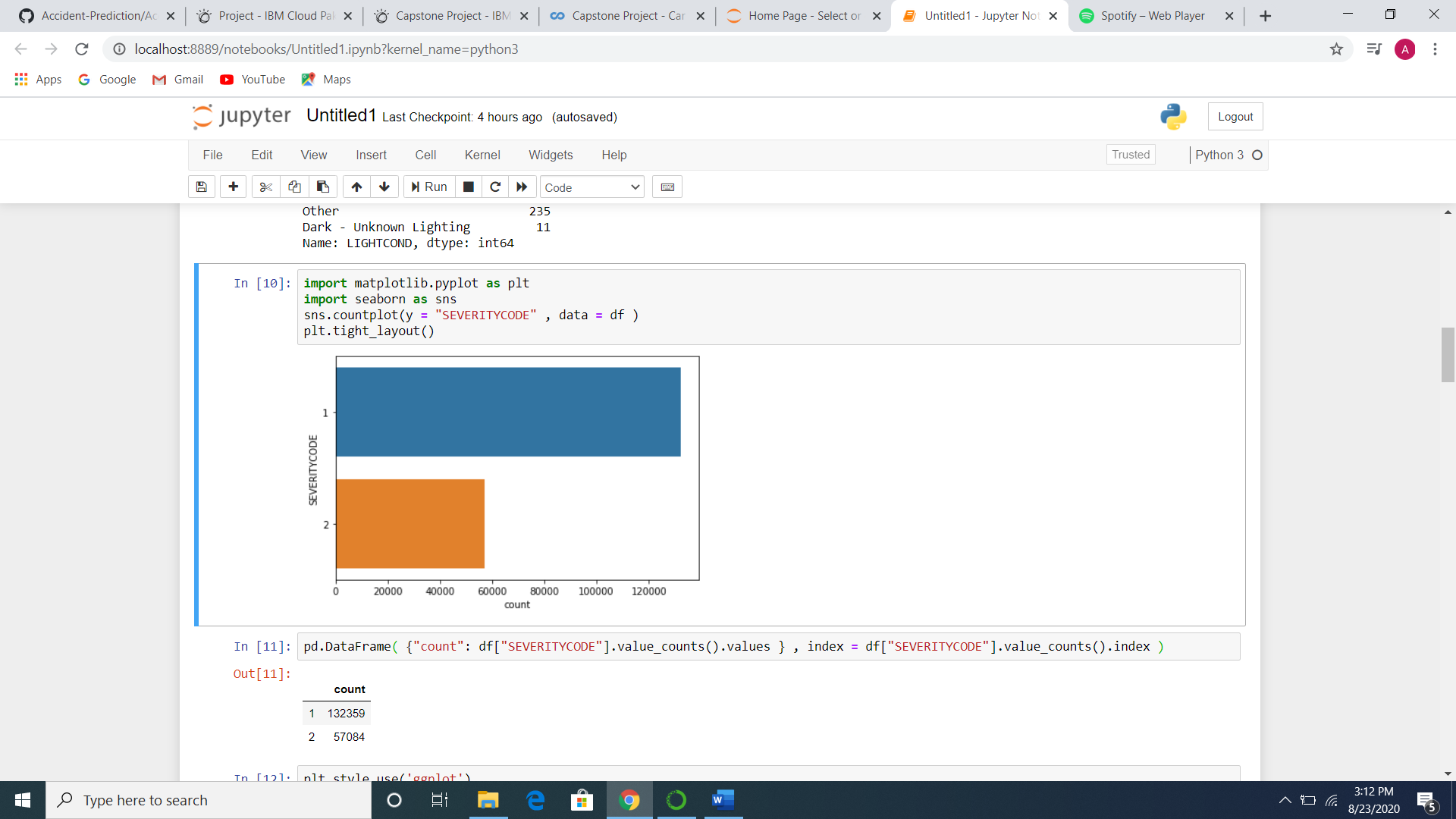
HITPARKEDCAR

Methodology:

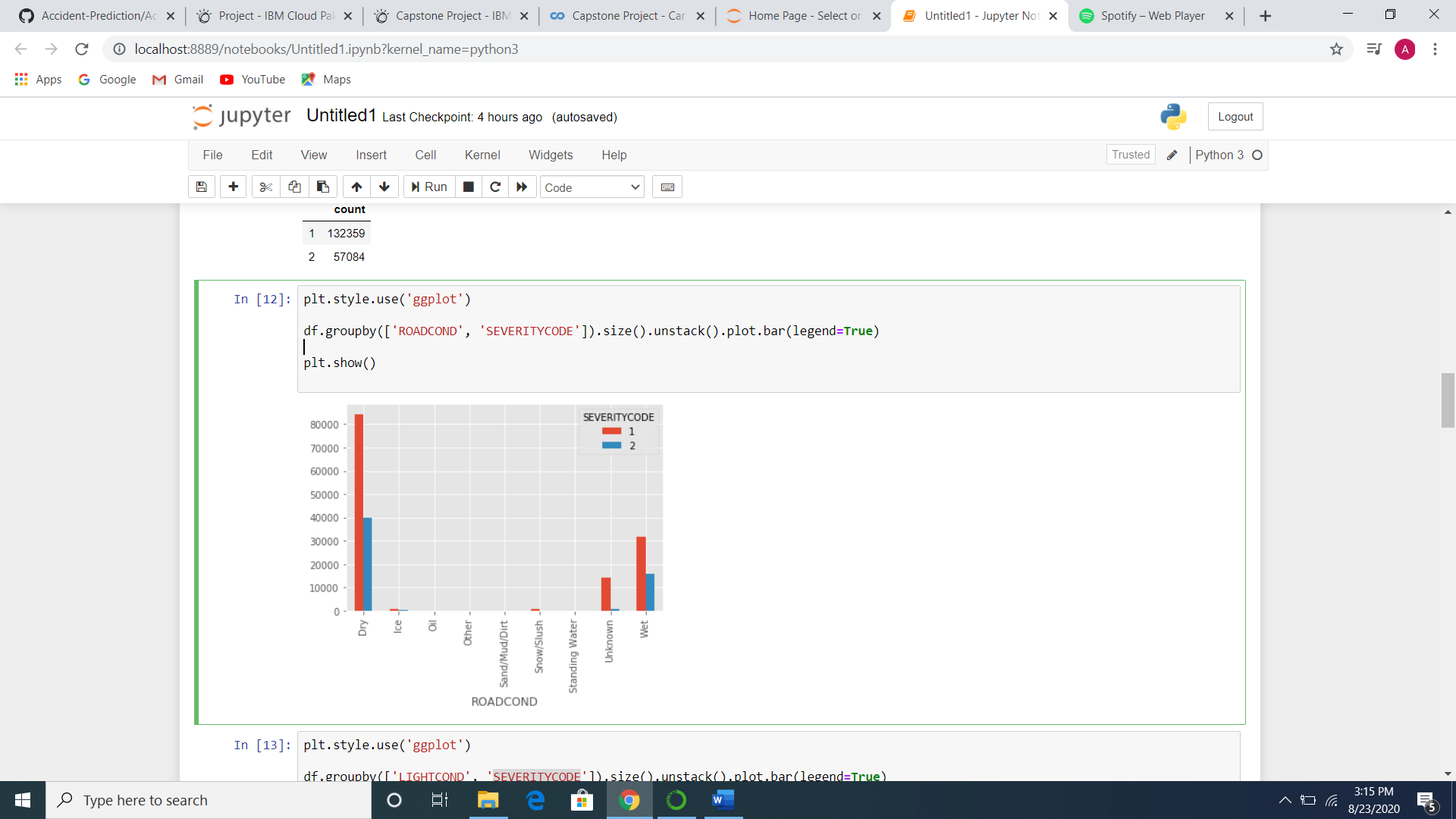
Following inferences were made from the data:

1. There were a greater number of accidents of severity code 1.

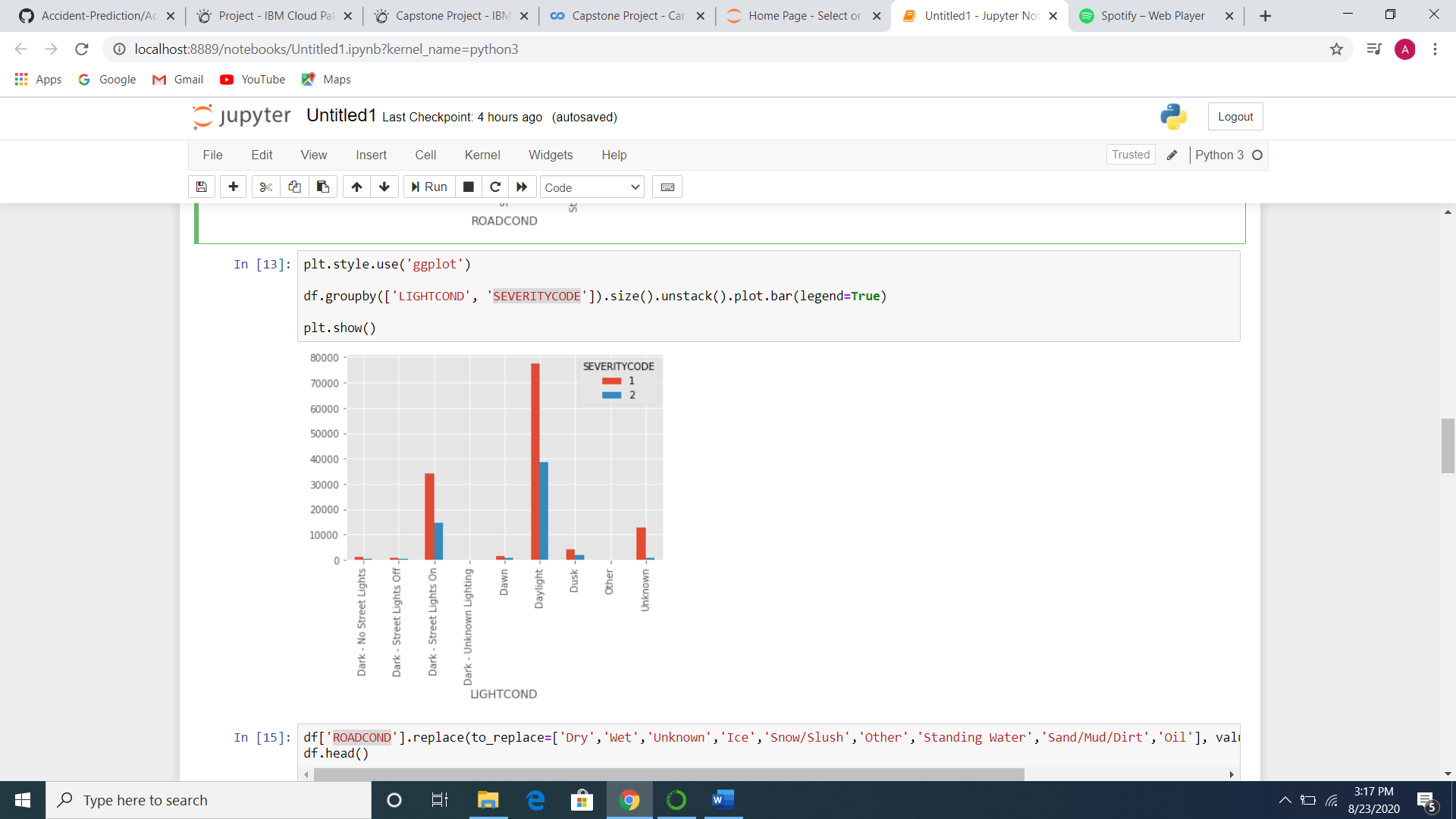




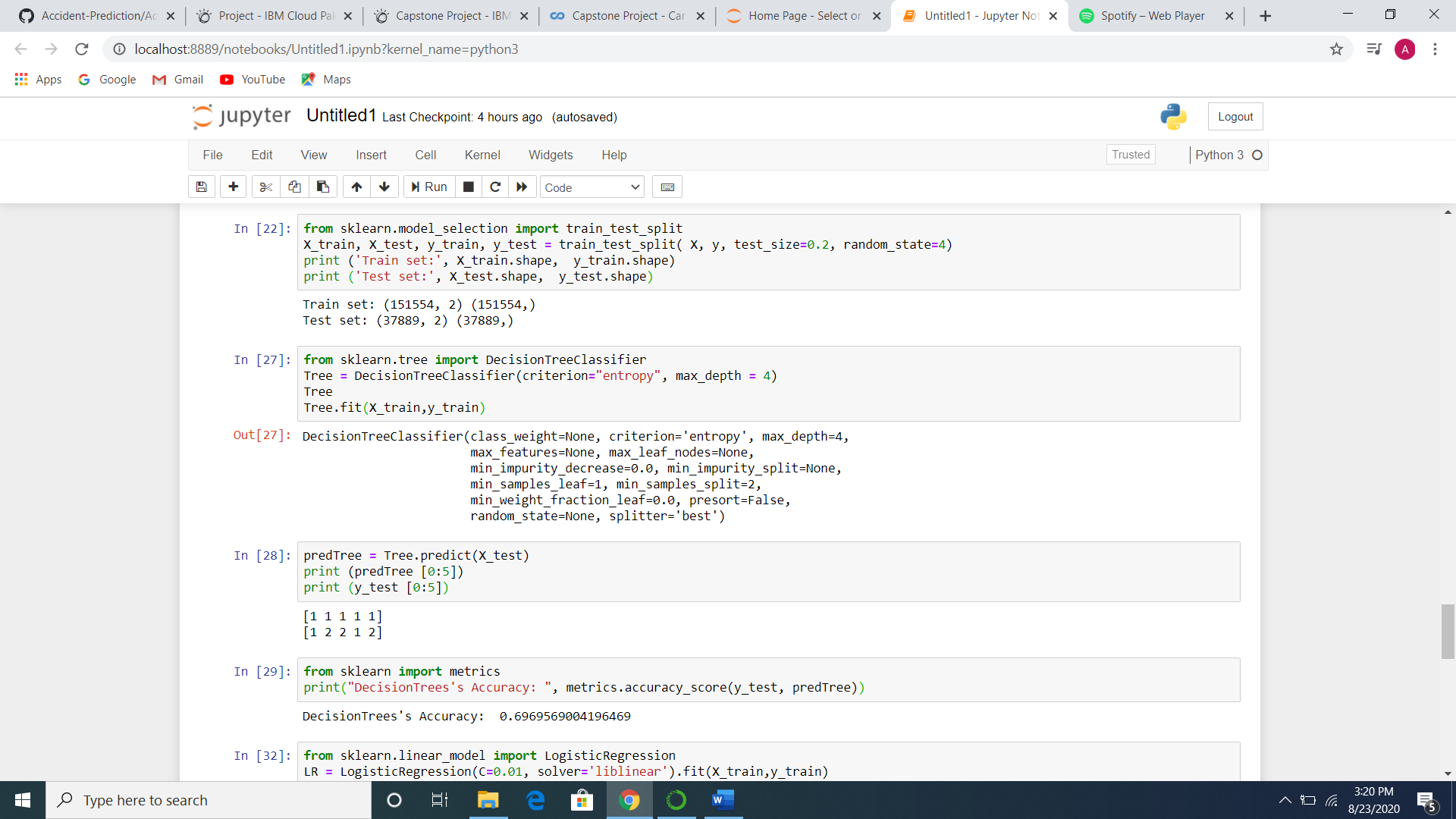
1. Dry and Wet road conditions caused the greatest number of accidents.



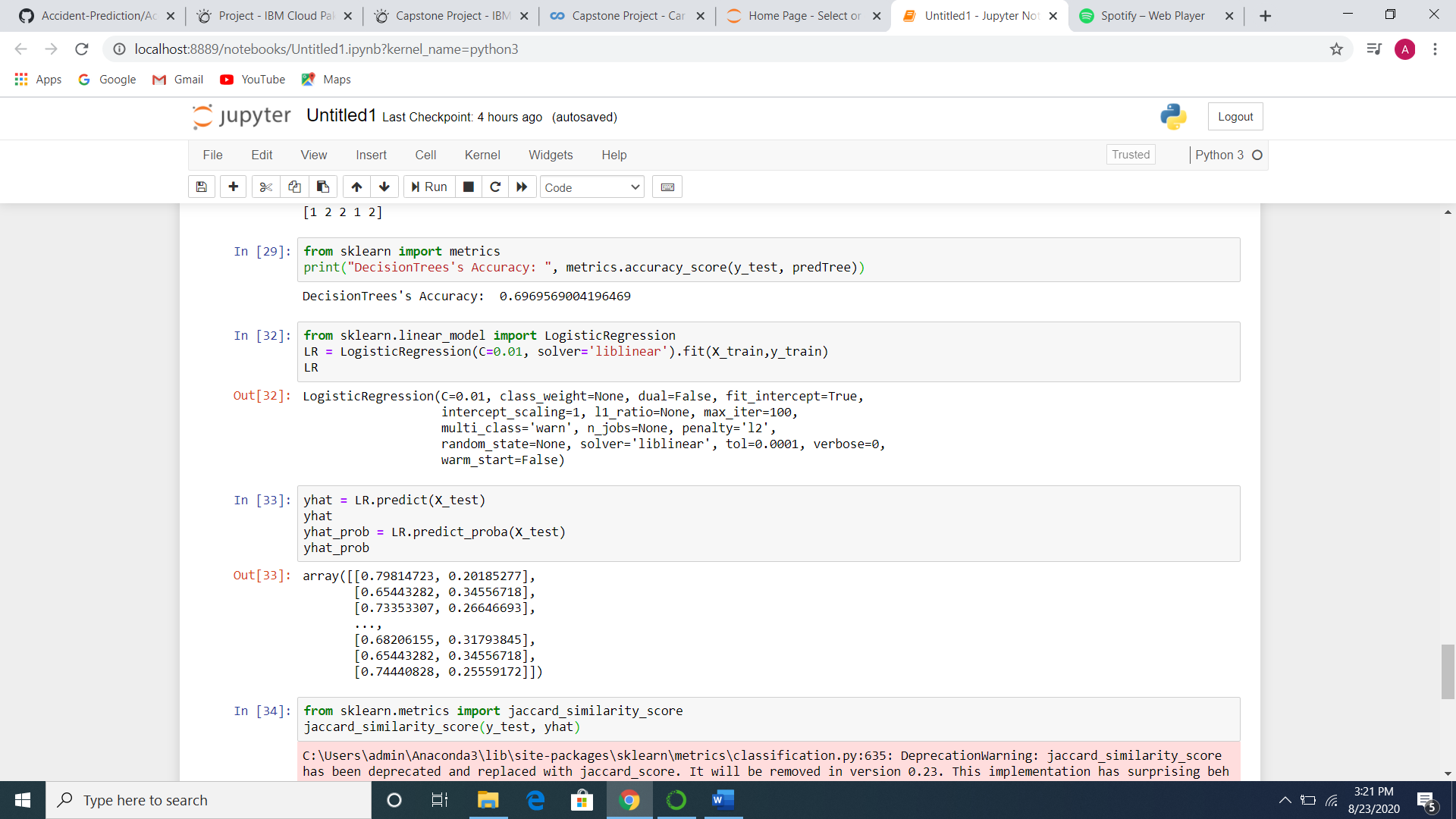
1. Most accidents were caused during Daylight and in the Dark-Street Lights On



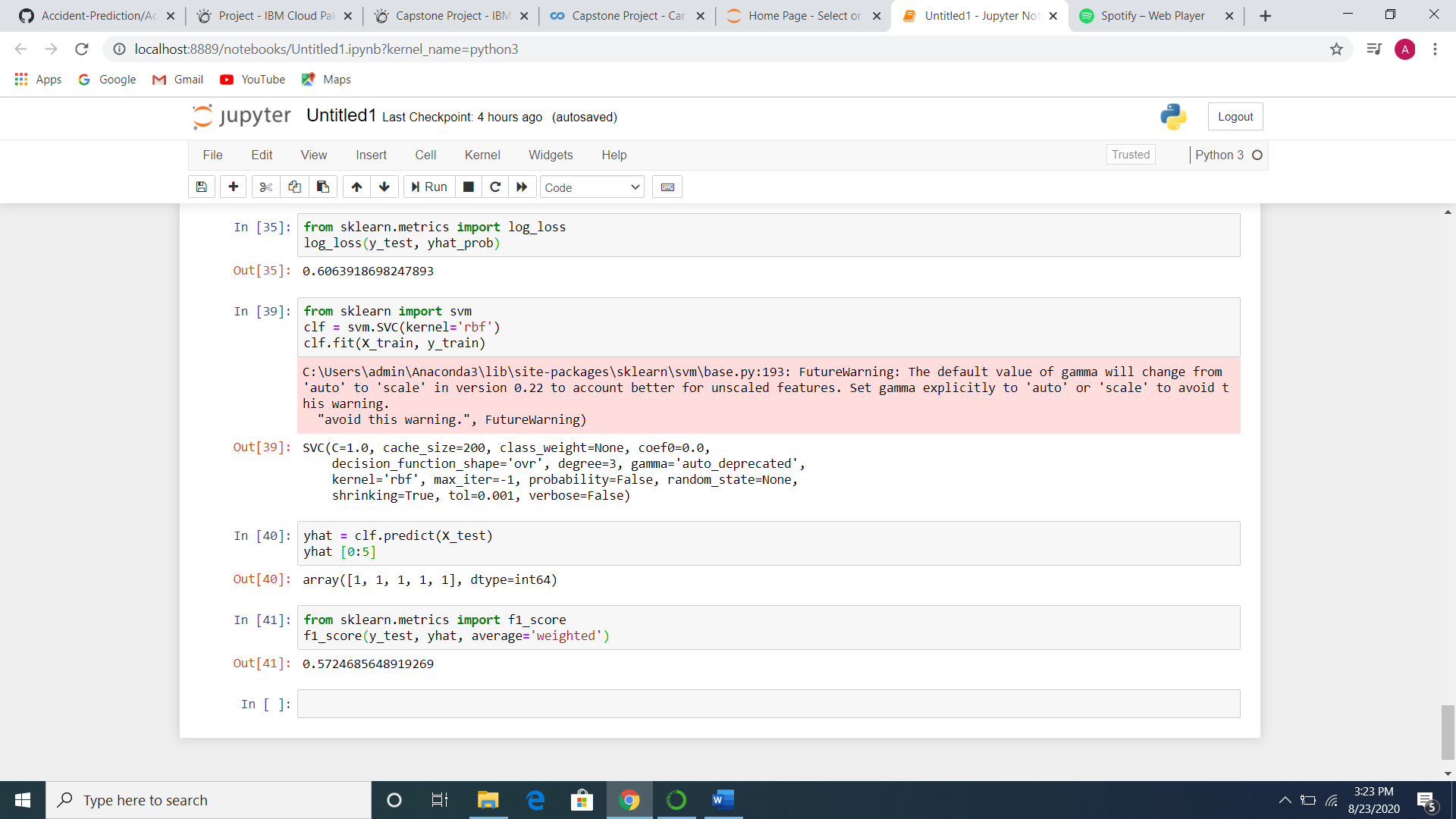
1. The following Machine Learning Models were used:
2. Decision Tree:



1. Logistic Regression:



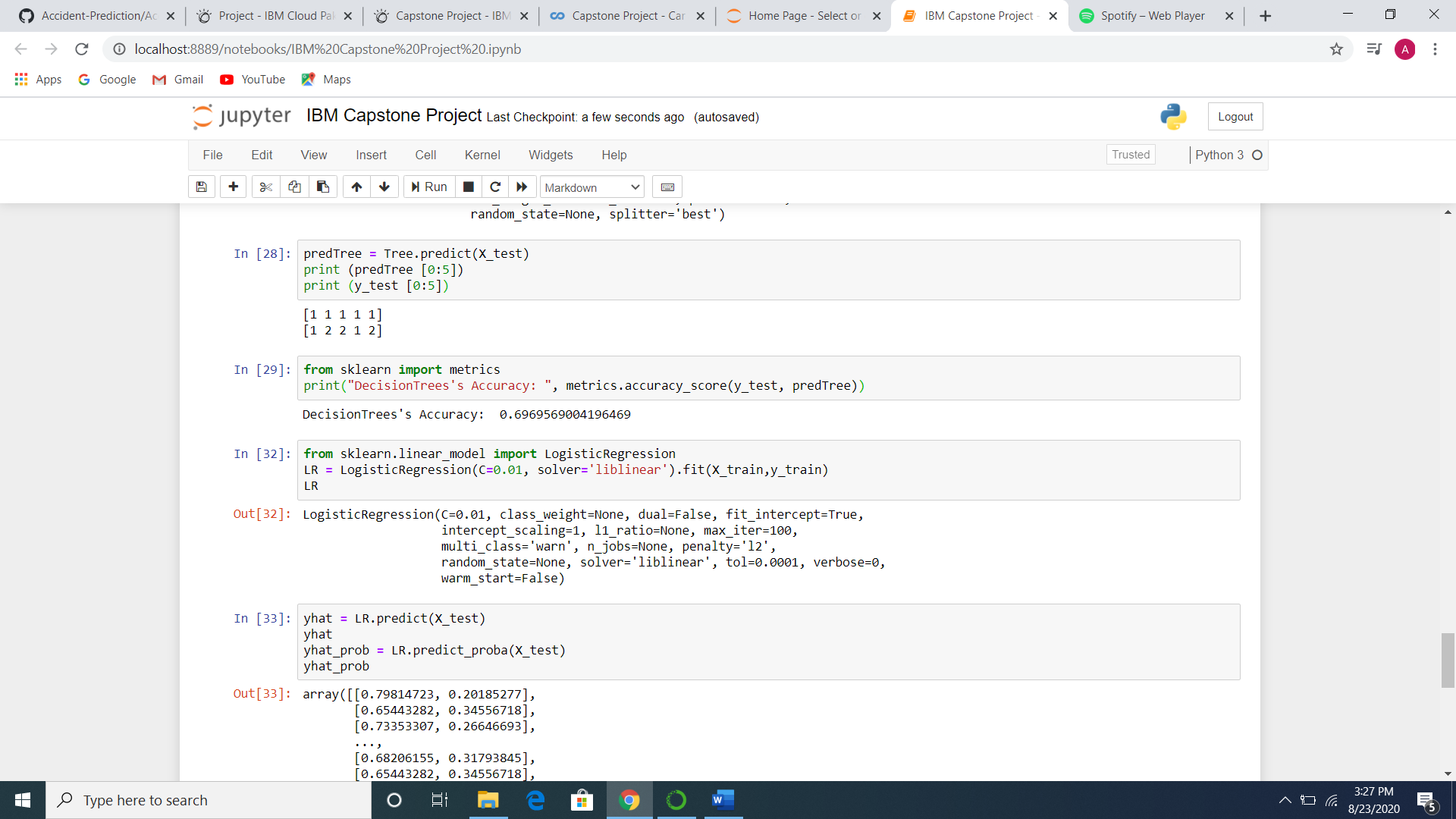
1. Support Vector Machine:



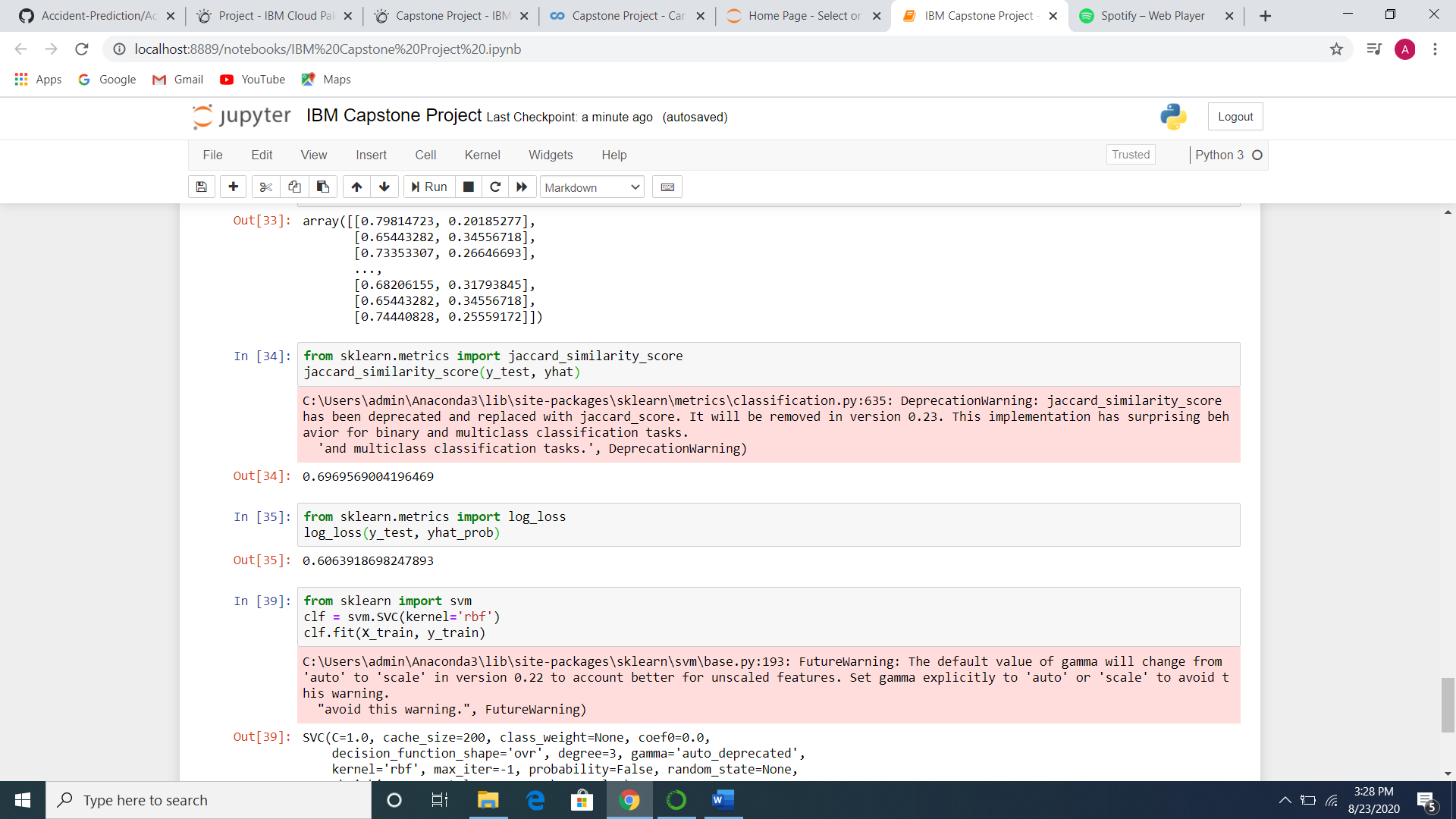
Results:

The models had the following accuracies:

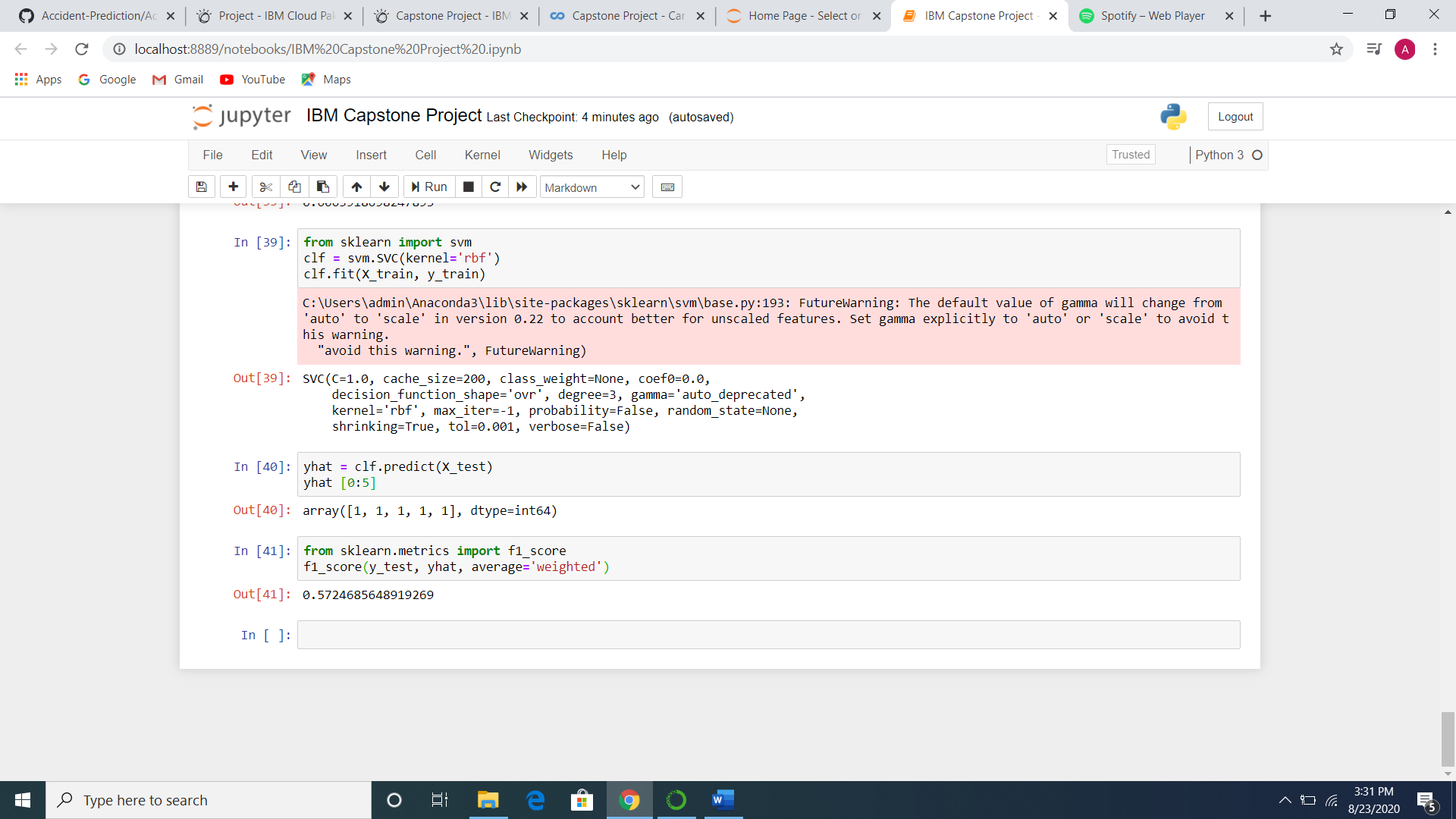
1. Decision Tree



1. Logistic Regression



1. SVM:



Recommendations:

It is recommended that car travel be avoided during wet and snowy weather conditions.

Conclusion:

Decision Tree and Logistic Regression models gave an accuracy of ~69%. These models maybe used to predict the severity of the road accidents based on road conditions.

