1 Maryland Traffic Violations Perform exploration analysis on the Kaggle Maryland Traffic Violations dataset.

Answer the following questions:

1. Which colors of the vehicles are more likely to get involved in a traffic violation?

2. Which models of the car are more likely to get involved in a traffic violation? This is an open-ended question. I encourage you to try as many data preprocessing and exploratory analysis tasks as you can possibly do. I am ready to be impressed.

2 Comments:

1. You can download the data here: https://www.kaggle.com/rounak041993/trafficviolations-in-maryland-county. It’s about 500 MB uncompressed. Kaggle Notebook has a limit of 100 GB per dataset, and Google Colab has a limit of 70 GB of storage

. 2. You may use pluto as it is a powerful server with few restrictions. To work on a data science project on pluto, the easiest way is to install an anaconda under your own directory. Then use ssh tunnel to access your Notebook from a browser at any place, such as your home. You may Google ’SSH Tunnel Jupyter Notebook’ for instructions.

3. You can also use your own computer.

4. R is also allowed for this homework. 5. The most relevant skill-set you may need for this assignment is Pandas. You may find a quick tutorial here: <https://www.kaggle.com/learn/pandas>.

On a separate file: Continue with the same data set, build a Decision Tree to tell whether a traffic violation may result in a citation or a warning.  You may use the SK-learn or other libraries for the Decision Tree classifier.