Tips

The data you have received is real data coming from various sources. It comes from:

* SAFER (Government web site)
* Point-of-Sale System
* Workday (Human Resource Management System)
* eDelivery (Delivery tracking system)

As you will find throughout your careers, data is never perfect. We’ve left some challenges in the data to keep it more ‘real life’. You may need to change data types; you may need to impute NA fields.

Date Range of files are different so keep this in mind as you join tables and conduct your analysis.

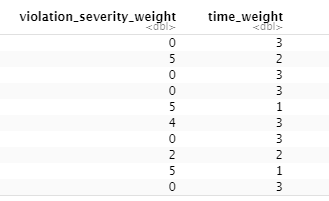
Drivers can move from Branch to Branch and can have other status changes. Effective date will be important as you join this table (drivers) to other tables, like inspections and crashes.

There are many inspections that don’t result in violations, however there are some inspections that result in more than one violation.



NA’s in basic, violation group, and code means that no citation was given.

When points are assigned you need to consider two items, the violation severity weight and the time weight. To calculate points, you MULTIPLY these two elements together for each incident.



Your focus is on Violations and Crashes

There are two main types of Violations:

1. Vehicle Maintenance
2. Driver Related

Joining tables will have some challenges. If you have limited memory to work with you may need to reduce the table sizes before joining them.

Performing thorough EDA will help you uncover areas that ABC should focus their efforts on. Make sure you present your plots/graphs/heatmaps and tell a story from your analysis. We know one day is not sufficient to build a complete solution. Demonstrate that you understand the problem at hand. Show your creativity in possible solutions.

The Judges will be Business leaders, Product Owners, Software Engineers, and other Data Scientists.

Time management is critical. Your presentation is all the judges will see so make sure you build it while you are conducting your analysis.

Feel free to use any tools that you’ve become accustomed to in class: R, Python, Tableau, etc.