# Introduction

## Charts and Tables

## Machine Learning

## Software Languages and Packages

Python Pandas

Python Scikit-Learn

Postgres DB and SQL

AWS S3 and RDS

Tableau

## About the Data

The data is available for download from a FTP site and is in csv format. Date for years 2012 to 2018 was downloaded (223Mb). The data was well formatted and documented. Differences in data format versions were resolved and all column heading named consistently , 35 files (81.5Mb) was loaded into the AWS hosted Postgres Database.

## Data Sources

National Highway Traffic Safety Administration

<https://www.nhtsa.gov/> (Home)

<ftp://nhtsa.gov/FARS/> (Data Repository)

U.S. General Services Administration

<https://www.gsa.gov/reference/geographic-locator-codes/glcs-for-the-us-and-us-territories>

## References

National Conference of State Legislatures

<https://www.ncsl.org/research/transportation/cellular-phone-use-and-texting-while-driving-laws.aspx>

National Center for Biotechnology Information

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001667/>

## Submitted Files

HighwaySafetyMunging.ipynb – The development version of the Jupyter Notebook for munging the NHTSA FARS csv data.

HighwaySafetyMungingSingleCell.ipynb – The final implementation for data munging. Same as the above and has final changes and merged to a single cell for better automation.

hwy\_safety\_helpers.py – Supplies helper methods and dictionaries for categorizing distract and impairment numbers.

SQL\_queries.ipynb

Saving\_Models.ipynb

FARSmunge2018.ipynb

Project3Startup\TrafficData\ExtractionZone\Cleaned\\* The output files from data munging.

Project3Startup\TrafficData\ExtractionZone\\* The working area for zipped and extracted folders with the downloaded files.

acc\_dis\_to\_cvs\_query.sql – Script to merge accident and distract data.