2019

**Vehicle Speed Information**

**Data analysis report**

**Shabari nath k (shabari.nath.kk@gmail.com)**

## Table of Contents

[Summary](#_Summary)

[Introduction](#_Introduction)

[Business objectives and Data analytics objectives](#_Business_objective_and)

[Tools Used](#_/Tools_used)

[Analysis and Modeling](#_Data_analysis_and)

[Conclusion and Products](#_Conclusion)

## Summary

## Background

Over the last 5 decades, road traffic has increased along with number of roads. Due to ever-changing supply-demand trends, there is an inclination towards need for speed, to reach on time, to save time and eventually save money.

## Objective

Core Objective here is to make roads safer, change traffic rules if possible and train drivers to follow traffic laws and probably award the drivers who follow the law and punish those who don’t.

## Methodology

Given traffic data was analyzed and explored to find out best and worst drivers. A monthly speed trend was found. This also includes other outlier detections and statistics

## Conclusion

Tableau Worksheets was successful in finding out interesting trends. Hence the end users were able to make informed decisions

## Introduction

A lot of variables decide the variations in traffic speed. Different regions have different speed limits. Also, a dire need to reach the destination on time pushes the driver to hit the pedal to the metal, whose consequences are speed limit violations. Even though most states have highway access, not all routes are completely well built and have access to highways. Some vehicles have the ability to reach above speed limit, but the driver might not hit the speed in fear of breaking the speed limit.

Data analytics:-

In the era of data analytics and machine learning, we can leverage data and analytics to identify instances of speed violations and other trends to help build a better road and traffic system. Traffic data can be harnessed to improve the same

## Author

Shabari Nath k(Anti)

Shabari Nath graduated from Amrita University in 2015 with an Engineering degree in Computer Science .Then he did PGDM (Post graduate Diploma in Management) in Supply Chain and Logistics. He enjoys spending solitary nights in the dungeons of PyCharm Python Coding to become a master of the language while listening to Enigma or PinkFloyd

## Business objective and Data analysis objective

**Business Objective:**

In order to be proactive, it is imperative to get information on traffic that includes any details about speed violations. The government agencies of each state would like to be well informed about traffic information, in order to come up with better solutions to solve traffic problems, which places need better roads and infrastructure. Building a driver profile can help Insurance agencies separate the good drivers from the bad. State motor department can punish or eventually cancel the driver license of regular traffic law offenders

**Data analytics Objective:**

Exploratory analysis to determine the underlying characteristics of information that is potentially useful

Understanding characteristics of drivers and vehicles

Visualization of filtered data that would summarize traffic information

Addition of interactive visualization such as line graphs, box plots etc

## python-logo-with_list.pngTools used

**Python 3.6**

Packages used

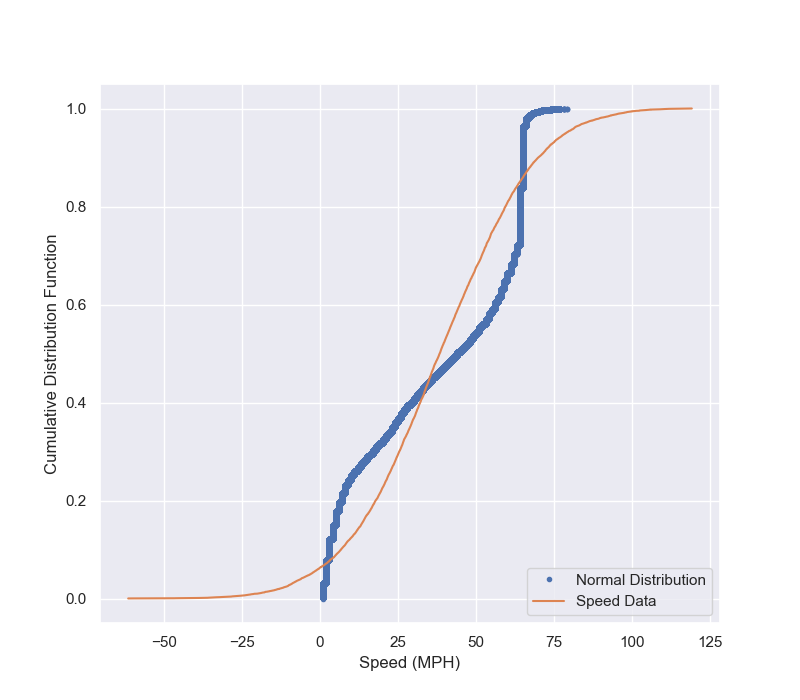
* Pandas to handle data frame
* Csv : to export data frame to csv files
* Seaborn,Matplotlib to visualize data
* PandasSql to handle Sql Queries



**PyCharm IDE**

PyCharm is an integrated development environment (IDE) used in computer programming , specifically for the Python language . It is developed by the Czech company JetBrains

## Data analysis and modeling

First step was to check if the data follows normal distribution 

Speed data is orange and Normal distribution curve is the dark blue one. Since they don’t overlap, we can infer what the data doesn’t follow normal distribution because of the presence of outliers and the large sample size

# 1: Data Transformation

A new column is added to the data called violation\_speed

Violation speed= Speed-SpeedLimit (miles per hour)

Other added columns are : newtimestamp ,weekday,month,day,hour

These columns are extracted from the already existing timestamp data, in order to easily push it to the Tableau software for data visualization

With these additions , data looks something like this

|  |
| --- |
| latitude |
| Longitude |
| Timestamp |
| Speed\_mph |
| Speedlimit |
| Functionalclass |
| Controlledaccess |
| isHighway |
| State |
| VehicleID |
| DriverID |
| ViolationSpeed |
| Newtimestamp |
| Weekday |
| Month |
| Day |
| hour |

# 2: Driver analysis

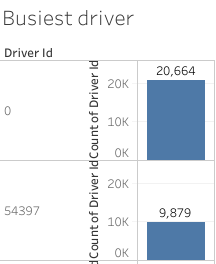
We will find the best and worst drivers according to their average speed limit violations.

We shall find other patterns like which driver works most on weekdays and weekends

We shall find drivers who take highways the most

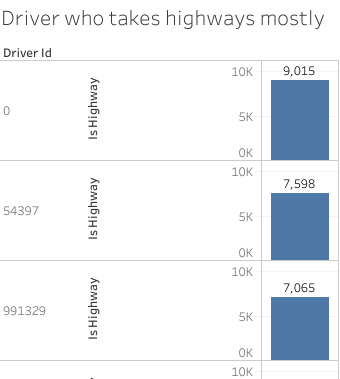
We shall find also find the busiest drivers

**Busiest driver**



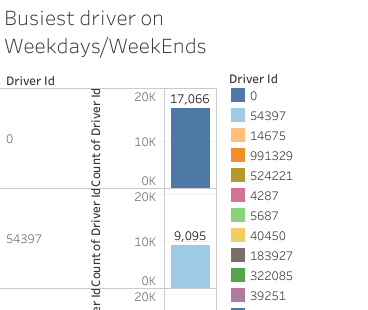
* Busiest driver is driver id : 0 with maximum count of 20,664

**Driver who takes the highways the most**

****

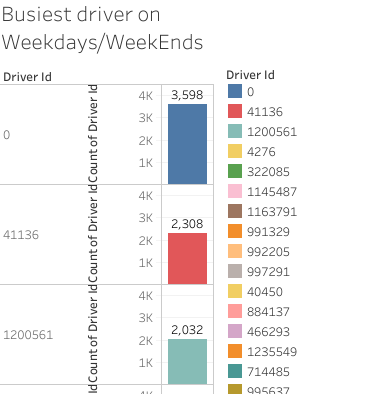
* Driver id with 0 , takes the highway the most at 9015 total is\_highway counts

**Driver who works on the weekends the most**

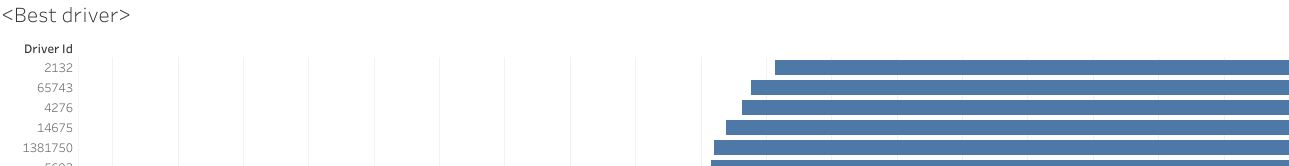


* Driver id 0 works at the weekends the most
* The above tableau sheet is filtered to show Sunday/Saturday data only

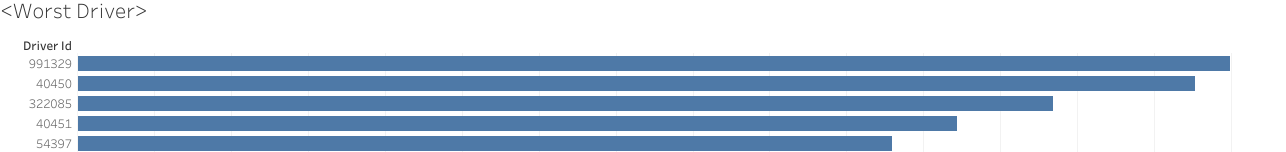
**Driver who works on weekdays the most**

****

* Driver id 0 works at the weekdays the most
* The above tableau sheet is filtered to show Weekdays data

**Best Driver **

* X axis shows the average speed violation (lesser the better )
* Y axis shows the driver id
* This shows that driver id 2132 is the best driver

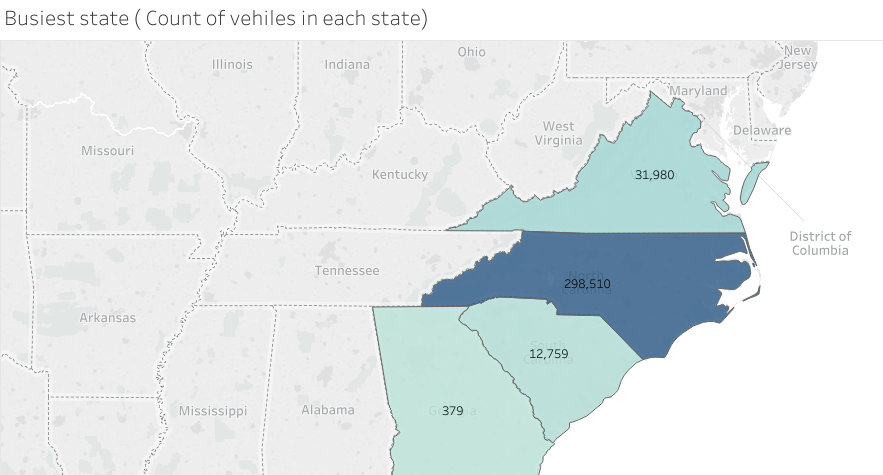
**Worst driver**

* X axis shows the average speed violation (higher the better )
* Y axis shows the driver id
* This shows that driver id 991329 is the worst driver

# 3: State wise analysis

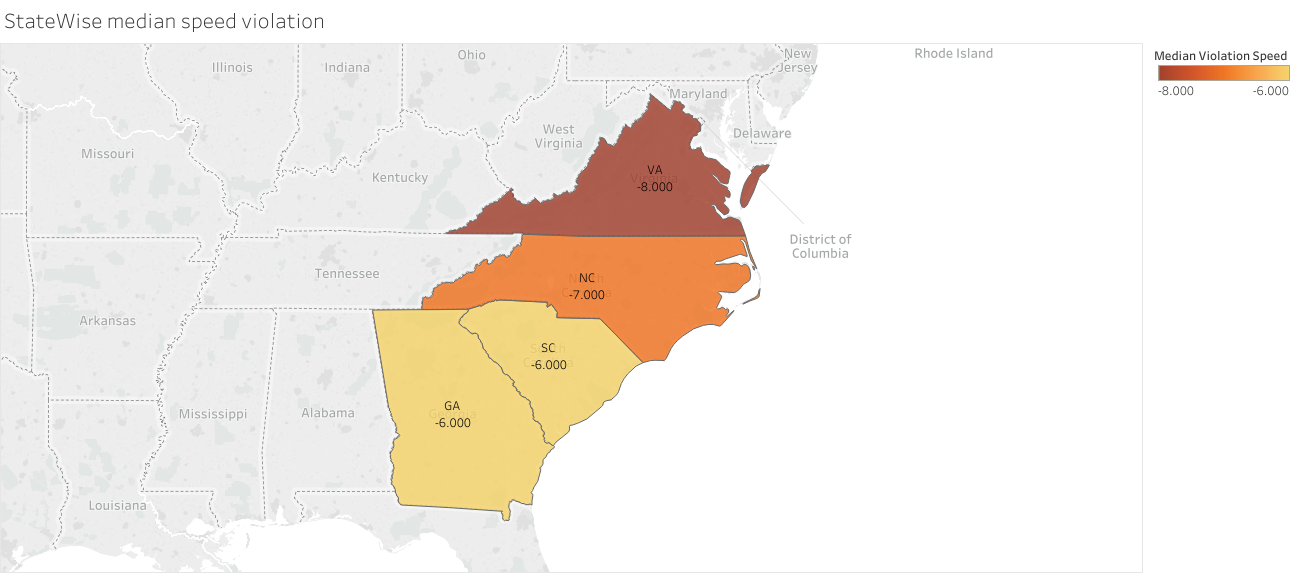
* We shall find the busiest state, number of highways in each state, speed limit violations in each state , state wise relation between average speed and average speed limit

**Busiest state**

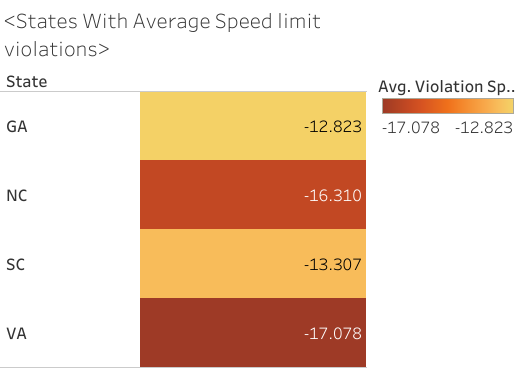


* Busiest state is North Carolina with 298510 count of vehicle ids’

**Speed violations per state**

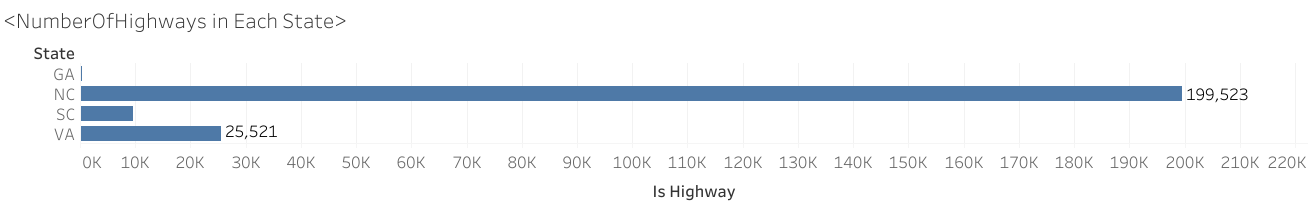


* Map shows median violation speed
* Georgia and South Carolina has the highest median
* Hence this analysis is not enough to make a conclusion



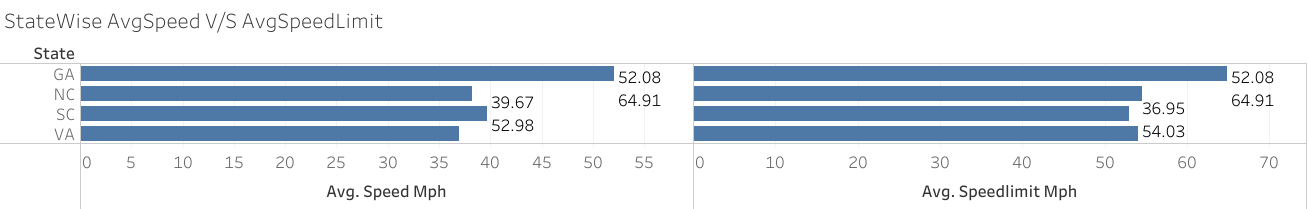
* Here we can see Gerogia has the highest number
* Hence we can conclude that the state of Georgia has the highest number of speed limit violations

**Number of highways in each state**

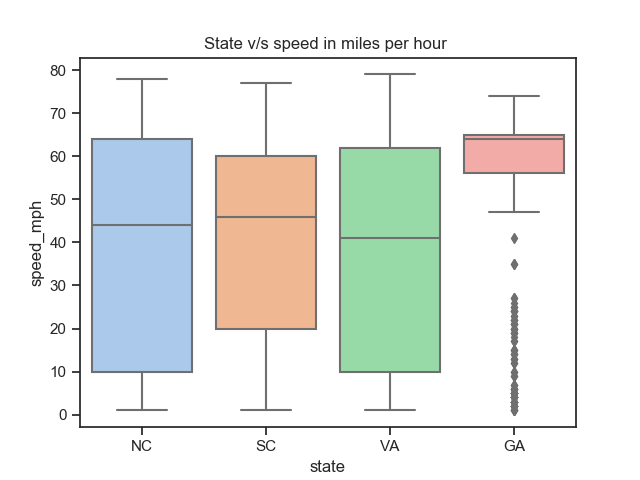
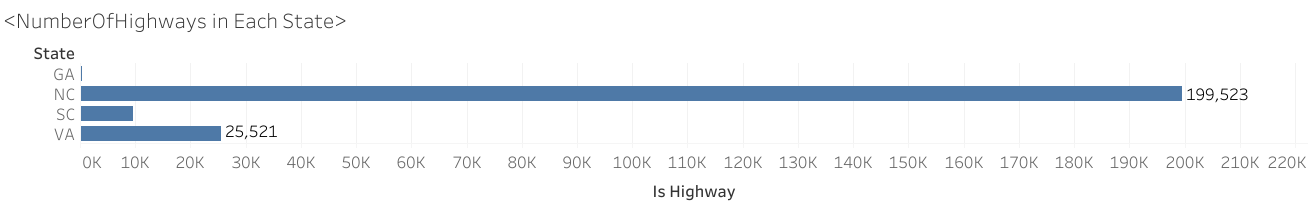
****

* X axis shows the is\_highway data, Y axis shows the states
* Georgia has the highest number of highways

**State Wise average speed/limit analysis**

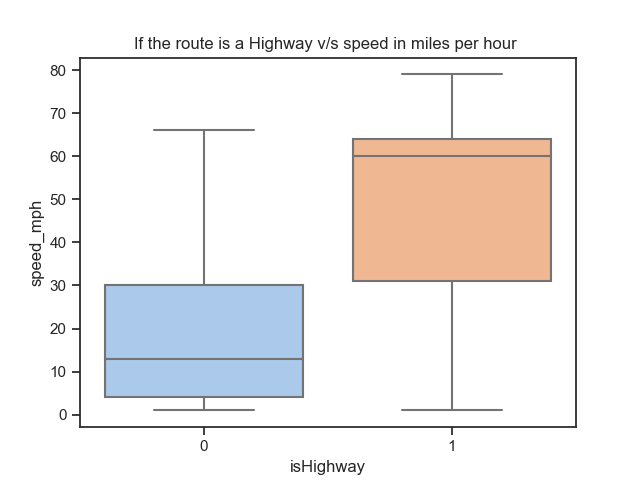
****

* Here we can see that in the state of South Carolina (SC), the average speed of the traffic is higher than that of the average speed limit
* Traffic authorities should be kept on alert

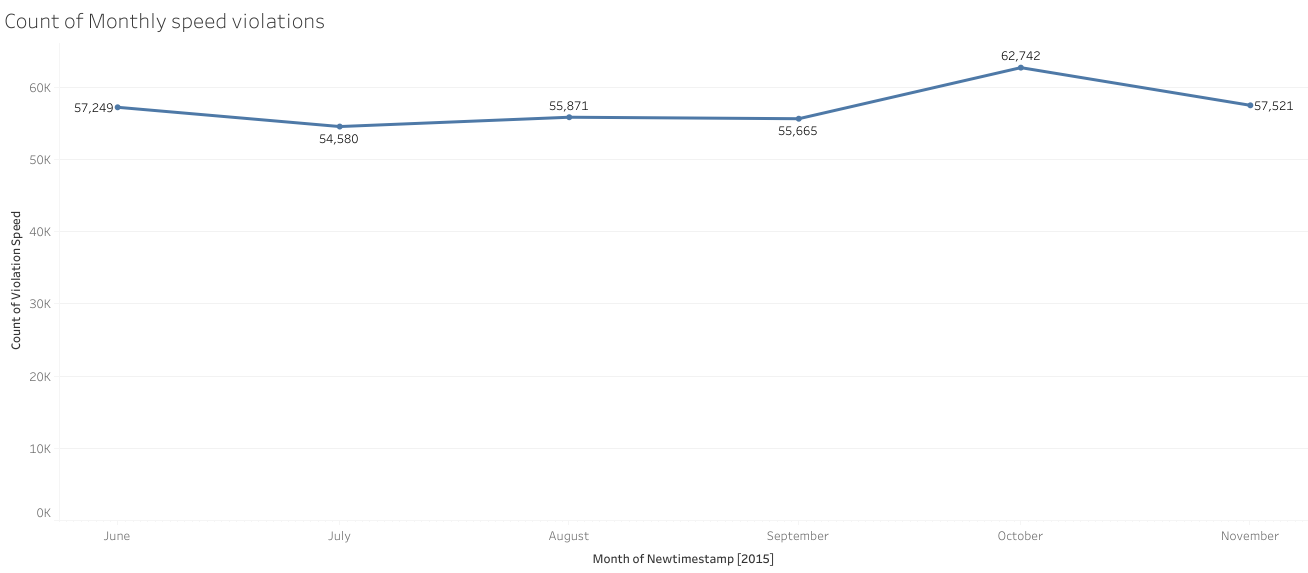


* From these two figures, we can see that most of the traffic in Georgia is over 50mph while the number of highways in Georgia is the least
* Here we can infer that the state of Georgia needs better infrastructure to accommodate the fast moving traffic

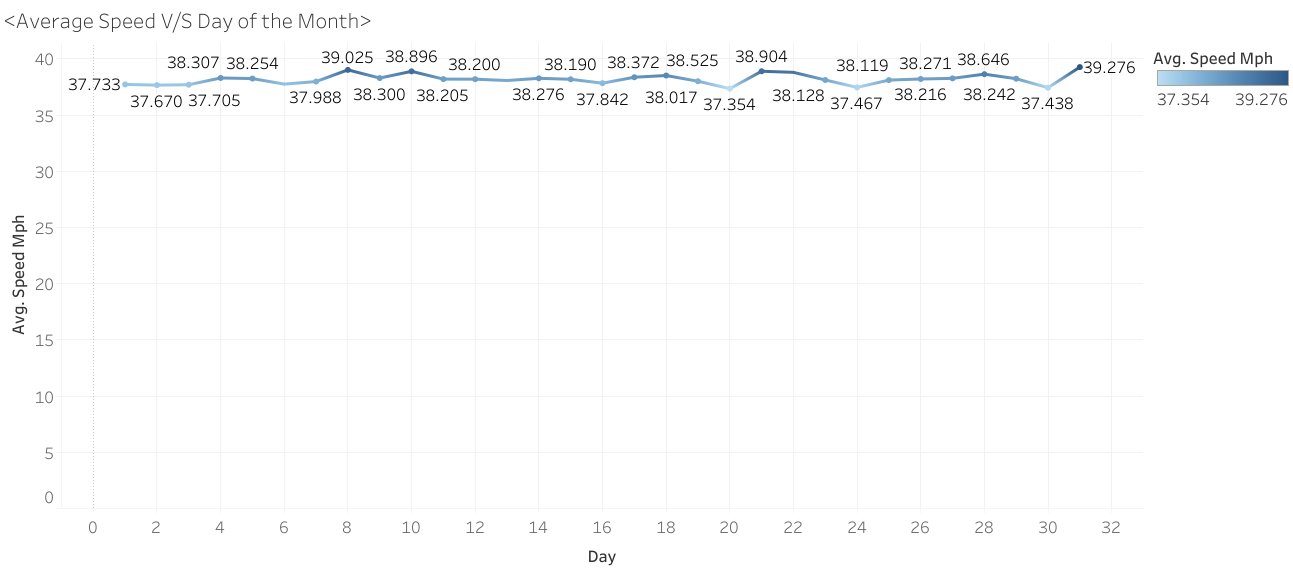
# 3: Speed analysis



* In highways, less than 50% traffic doesn’t cross over 60 mph
* Where as if its not a highway, over 50% traffic crosses over 10 mph

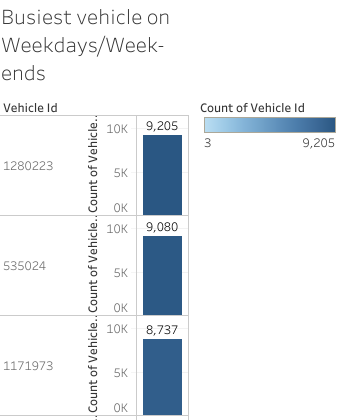


* Here , we can see that maximum speed violations happens in the month of October and November respectively
* A quick American Holiday analysis showed that these are the months where they celebrate Columbus day, Veteran day and Thanks giving
* It can be inferred that maximum speed violations happen here so that the drivers could finish their jobs and reach home quickly to spend time with their families

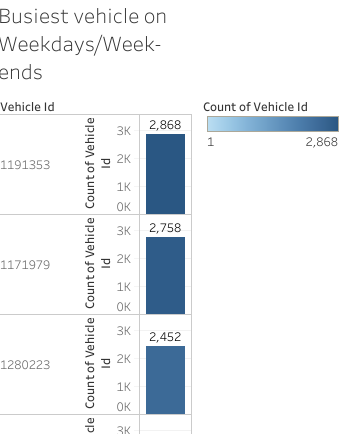


* Here, we can see that the average speed reaches the maximum on the month last
* Average speed reaches the second maximum on the 8th
* We can infer that it’s the pay roll week
* Traffic on roads might increase since people have their salaries in their account, and probably hit the road to get monthly supplies
* Traffic authorities can take a note here to implement smarter traffic management techniques during the first week of every month

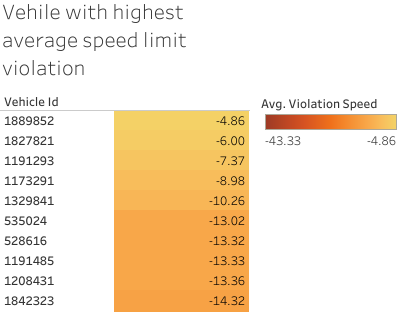
# 4: Vehicle analysis



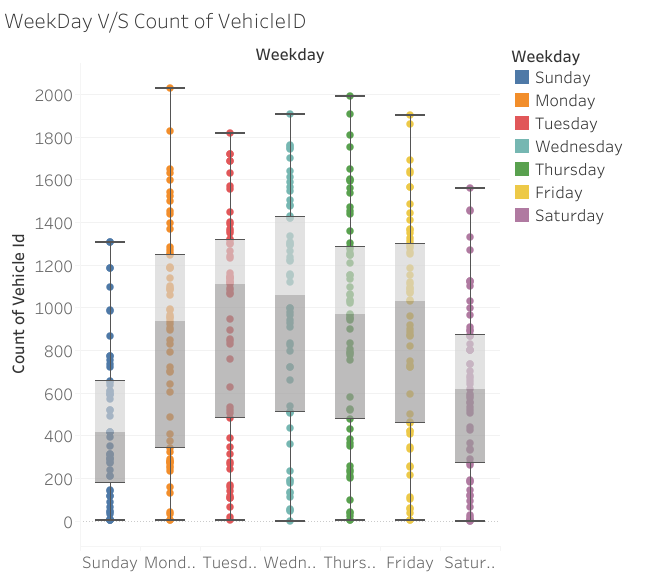
* Busiest vehicle on weekdays is 1280223



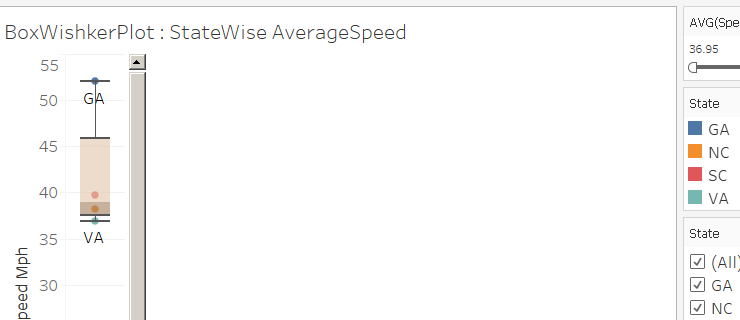
* Busiest vehicle on weekends is 1191353



* Vehicle with highest speed limit violation is 1889852



* Here , we can see that maximum vehicles hits the road on Wednesday
* Least number of vehicles are on Sunday



* Here , we can see that the highest average speed lies mostly on North Carolina and Georgia
* More than 50% of traffic hits over 40 mph in North Carolina and Georgia

## Conclusion

* Best driver is 2132 and worst driver is 991329
* Traffic authorities can take necessary actions here
* Busiest drivers is Driver ID 0
* Since this driver hits the road the most, Car insurance agency can keep a note on this
* State of Georgia has the maximum number of speed violations and the least number of Highways
* Hence the government can do something about the necessary infrastructure development
* State of South Carolina has the maximum Average speed limit violations
* Maximum speed violations happens during American holidays that falls on October and November

## Appendix

Tableau Public online workbook link :

<https://public.tableau.com/profile/shabari.nath.k3262#!/>

**Folder Description**

Main Folder : Driver

**Files :**

FileConfig\_driver.cfg

* Can be parsed to identify input data source and export data files

Driver.py

* Python source code that takes the data from the csv file and pushes to a pandas dataframe , filters the data and exports to another csv file

Hour.csv

* Final filtered dataframe ready to be pushed to Tableau software for data visualization

Viz Folder : Contains screenshots of all Tableau visualization