Project: Capstone Project 1 – (Data Storytelling Exercise section 7.2)

Project name: -	Explanatory Analysis of Traffic pullover pattern for Florida v/s Vermont
Student Name: -	Jitendra Agarwal
Course: -	Springboard cohort Jan2 2018
Data Source: -	https://openpolicing.stanford.edu/data/
Data provider: -	Openpolicing project by Stanford

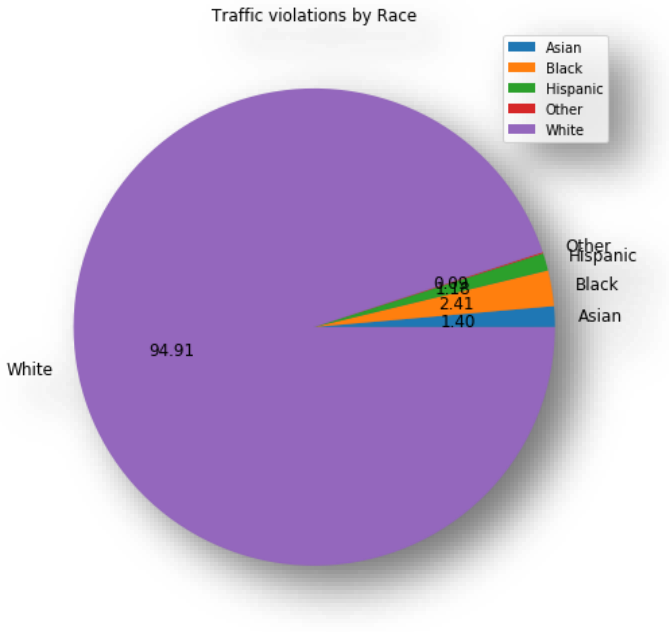
About the DATA

The Raw data for this project contains the traffic stop data collected for 30+ states for open police project by Stanford research team. Standardized stop data are available to download (by state) from the link above provided by Stanford.
The csv includes a subset of common fields for each state, and indicates whether data are available for at least 70% of records in that state. Some states have more fields.
The original, unprocessed data we collected contain even more information.

Downloaded excel sheet of raw data for VT: -
https://github.com/jiagarwa/capstone-project1-Jitendra
file name: - 'VT-clean.csv.gz'

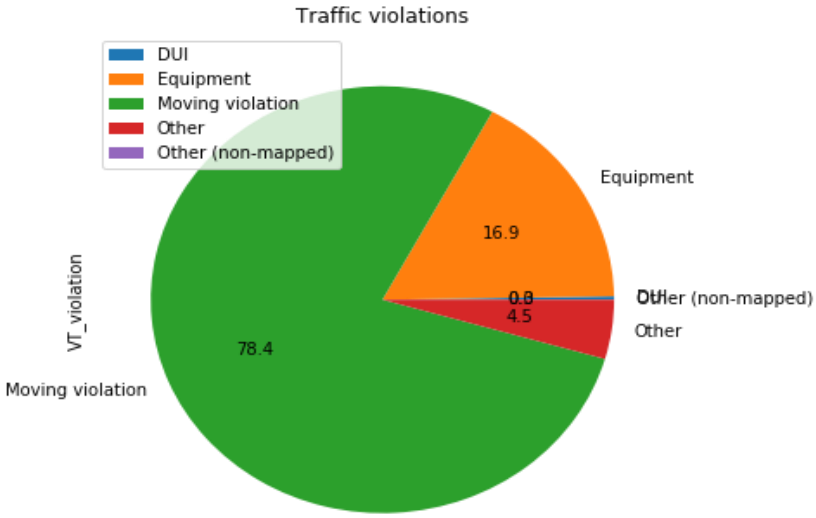
Data Story:

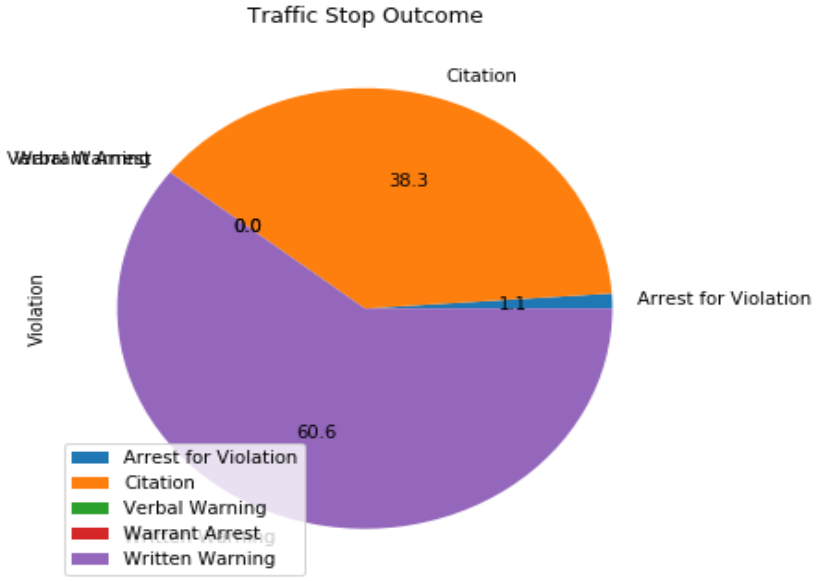
1. Exploratory Analysis to find something interesting.
    - Traffic stops by race.



- ✓ 95% Traffic stops involved driver from majority race and as per Wikipedia 2015 demographic population of Vermont state is 94.1% white and as per US census 2016 demographic population of Vermont state is 94.6% white
- ✓ So we can firmly conclude that there is no role of Race in traffic stop. The % of stops for Minority races is in line with the demographic ratio of Vermont state.

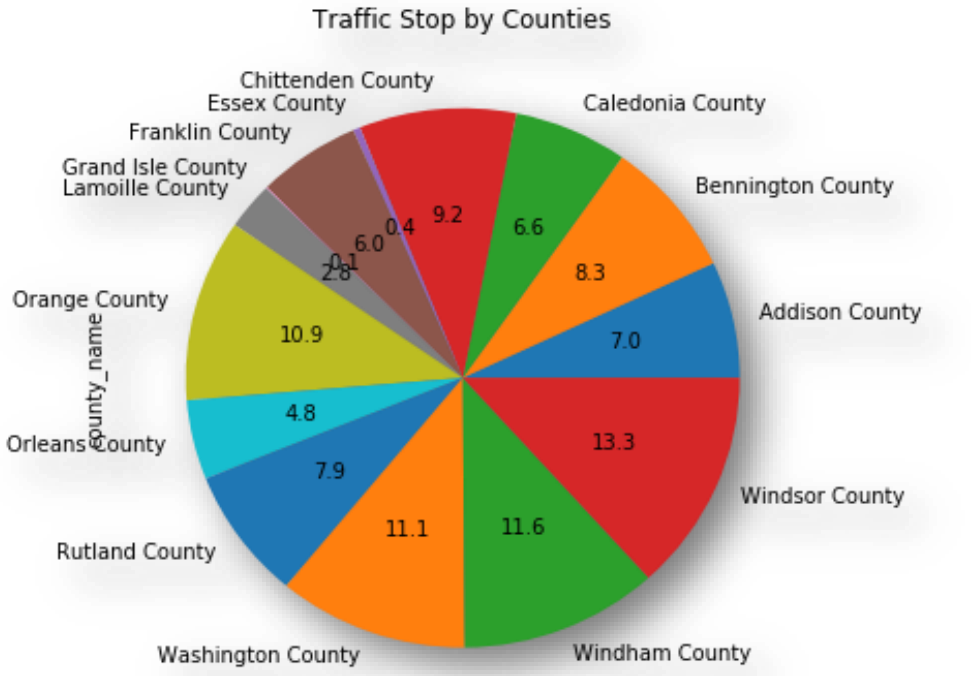
- Traffic violations and % Traffic Stop Outcome





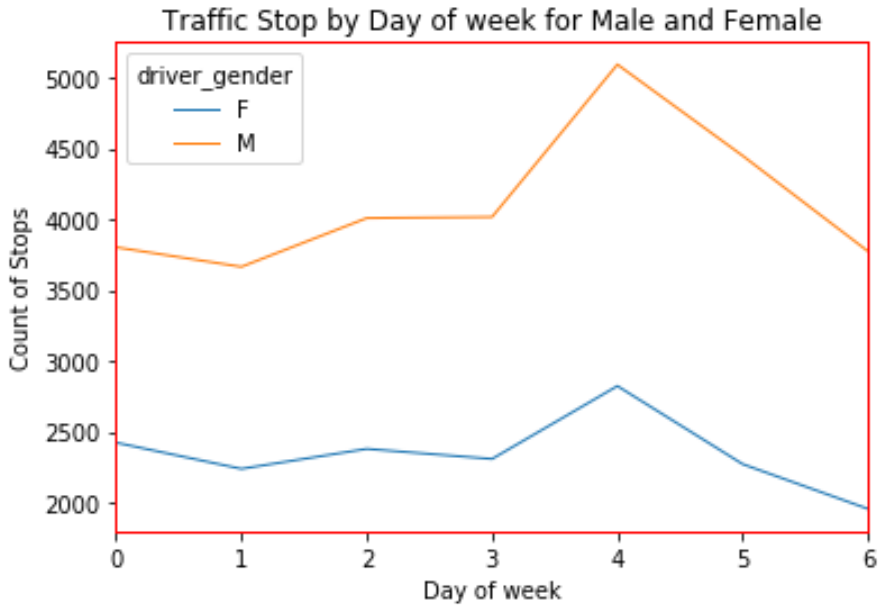
- ✓ Speeding is the most common reason for Traffic pullover.
- ✓ 60% Drivers were let go with warning and 38% received a ticket

- Traffic stops distribution by Various counties

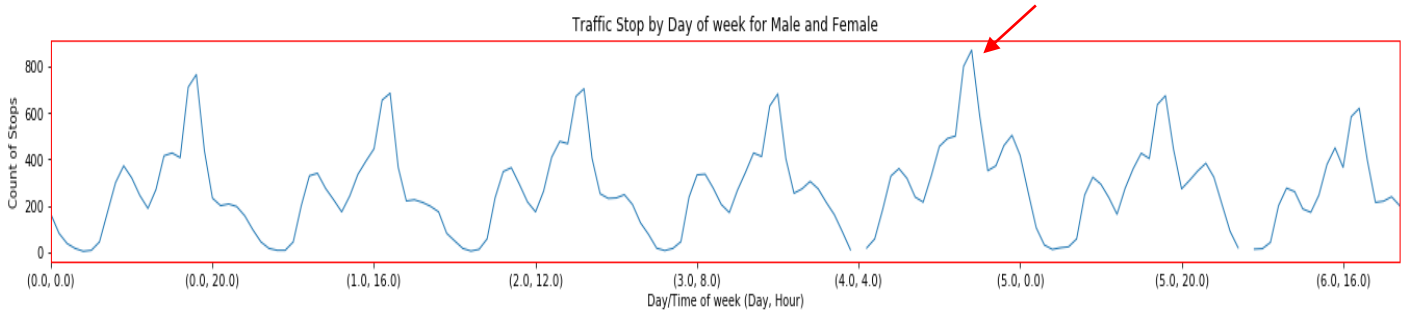


2. Finding trends

- Comparison of Male v/s Female in traffic stops over weekdays

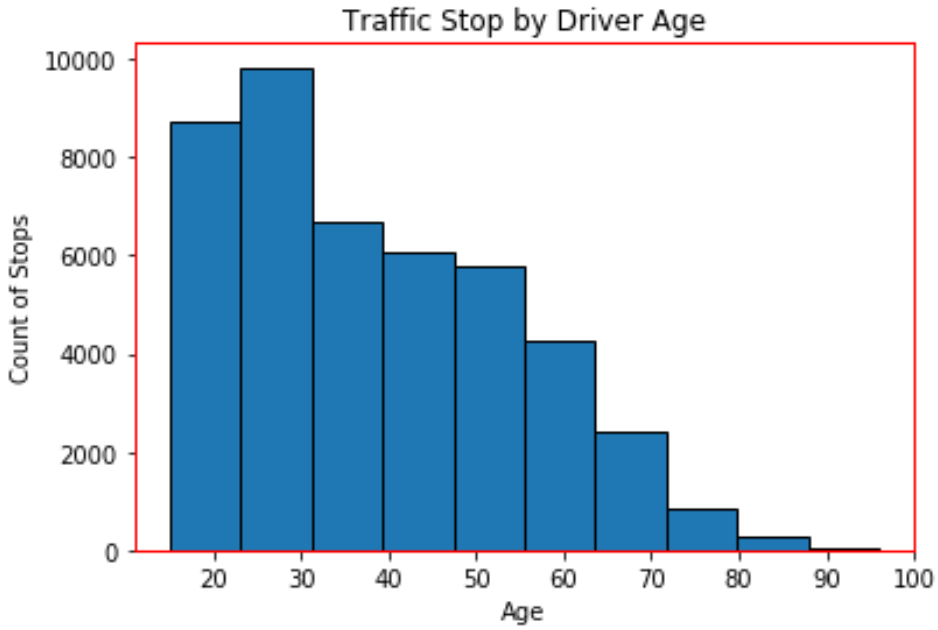


- Traffic stops over the time for entire week



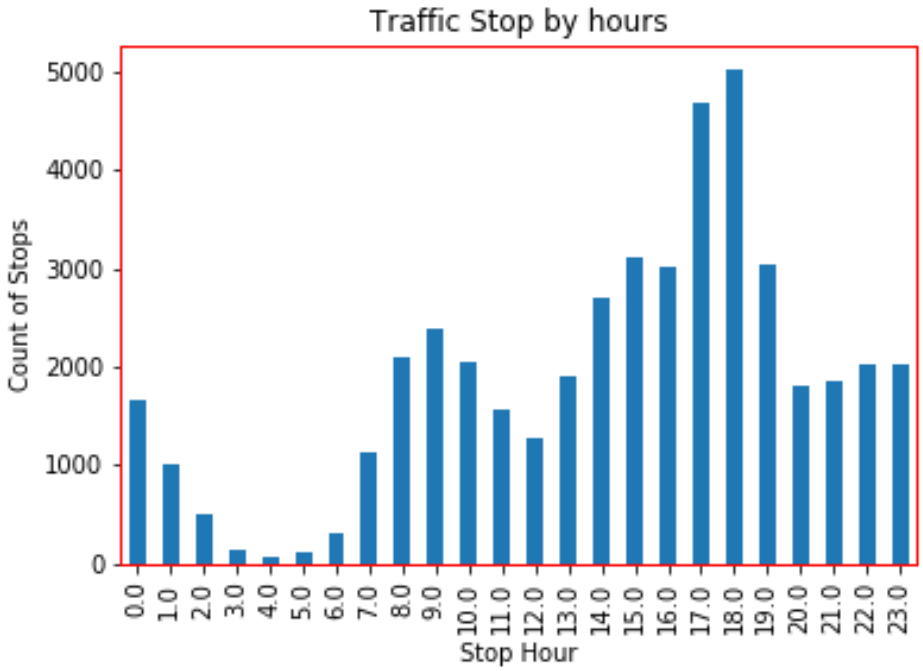
- ✓ As we can see that most busy day of week for pullover is not Friday or Saturday. It is actually Thursday. But when we put the time and date factor combined the trend is little different and interesting.
- ✓ It is not true that traffic police are most active on Friday or Saturday evening.

- Histogram of age range and frequency of stops



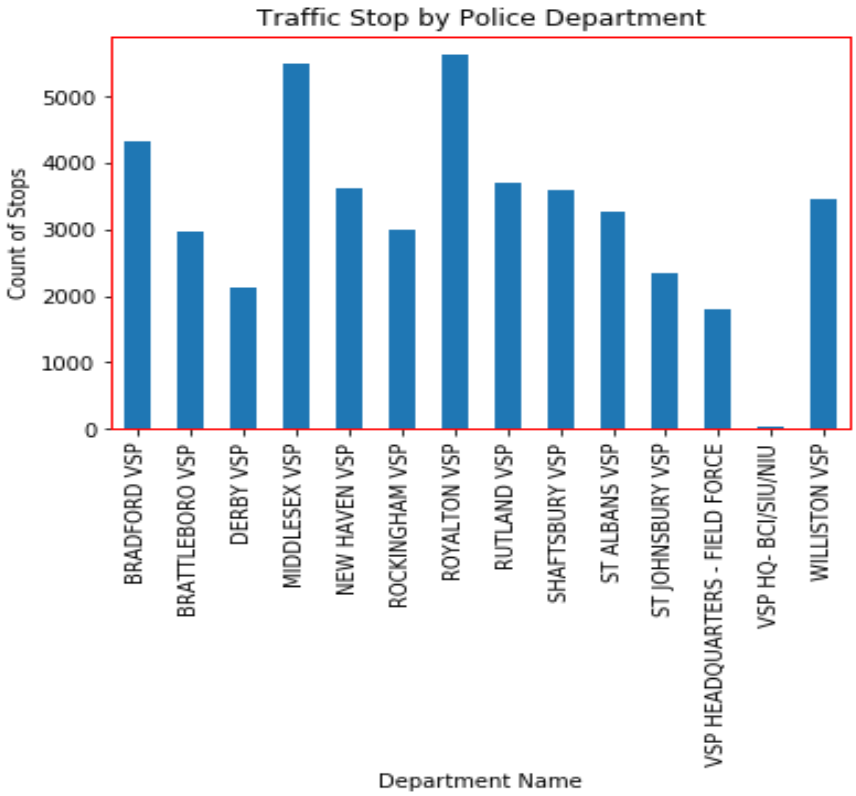
- ✓ 20-25 is the most common age of Drivers involved in the traffic stops.
- ✓ Young or New drivers are more likely to make mistakes and get pulled over.

- Bar plot of Stop hour and count of stop at each hour.



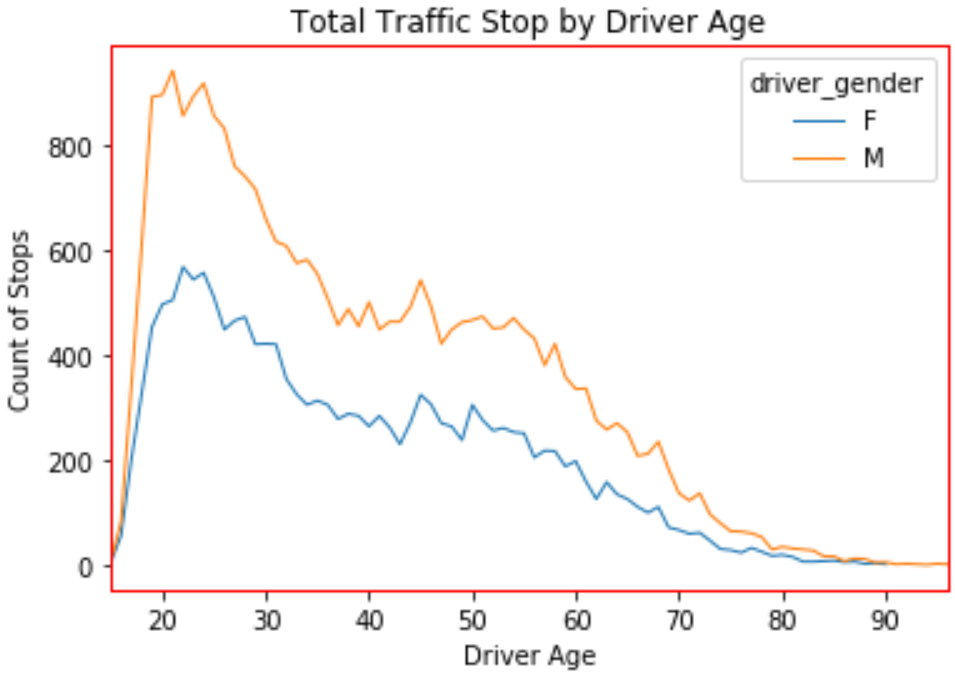
  - ✓ Traffic stops increases during evening rush hours 5-7 PM.

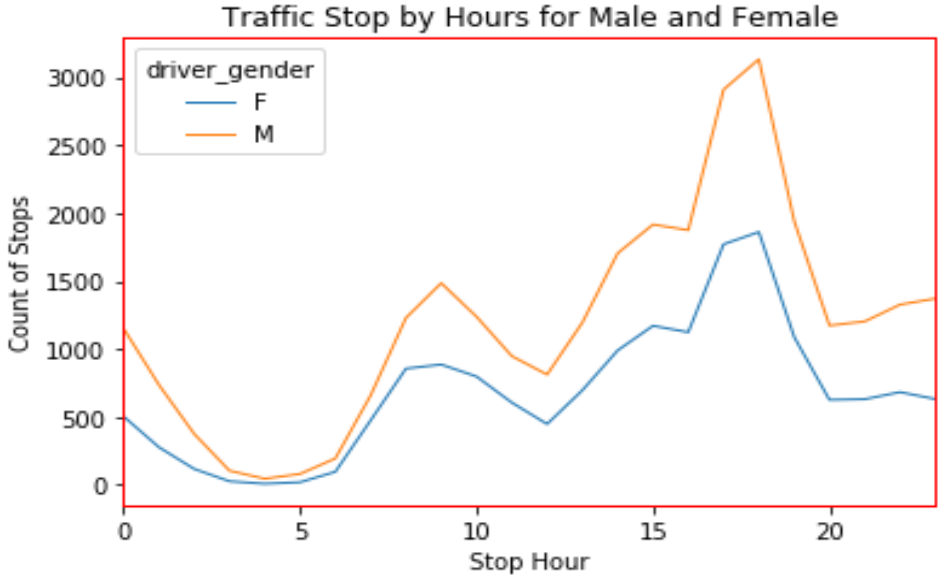
- Bar plot of police departments and count of stop by each Department



3. Compare two related quantities.

- Compare by trend chart of Male V/s female and Driver age



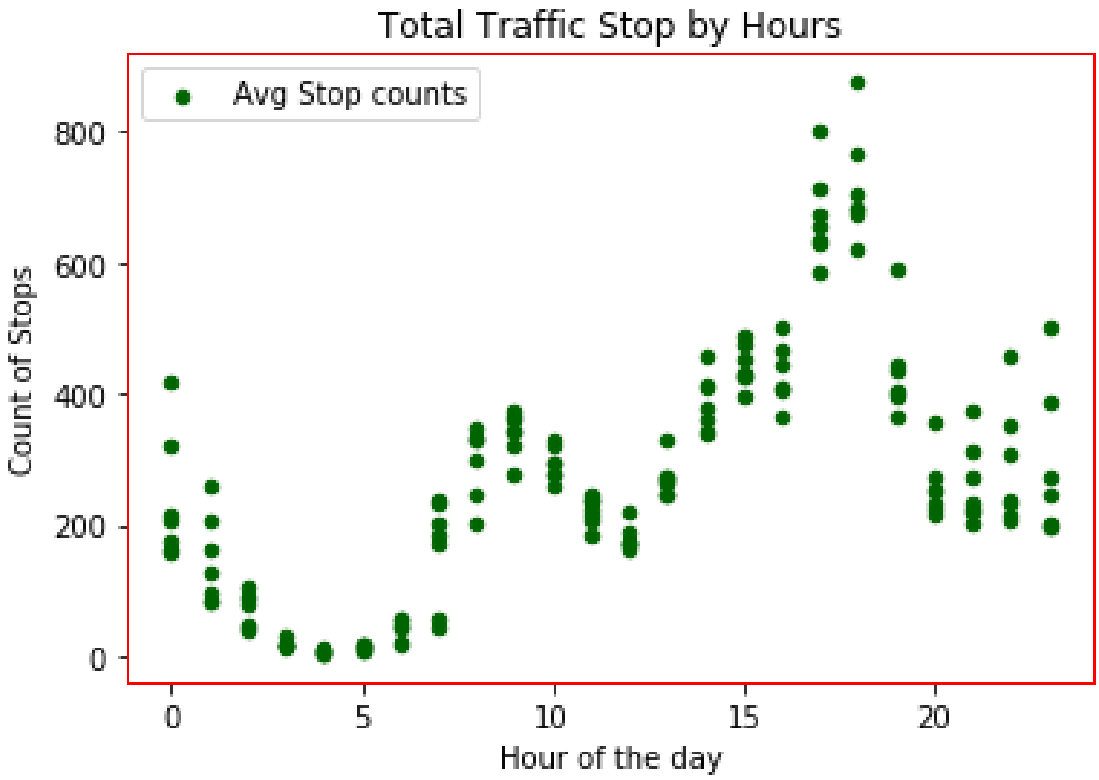


✓ There is no difference in the Male and Female traffic pattern except that female driver frequency of stop is less because there are simply less % of female drivers out there.

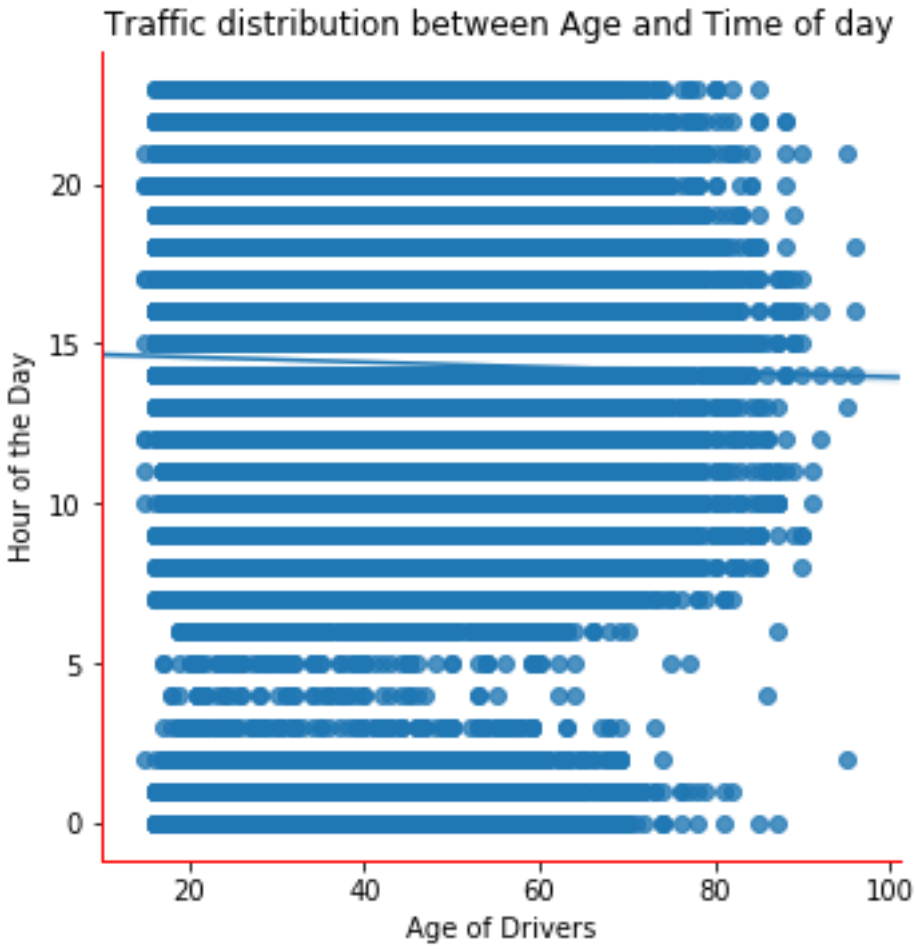
4. Show distribution using Scatter plot

   There is not much numeric data in the data set for this project so we are using a partially aggregated dataset for scatter plot.

- The average stop counts during the Day by different hours

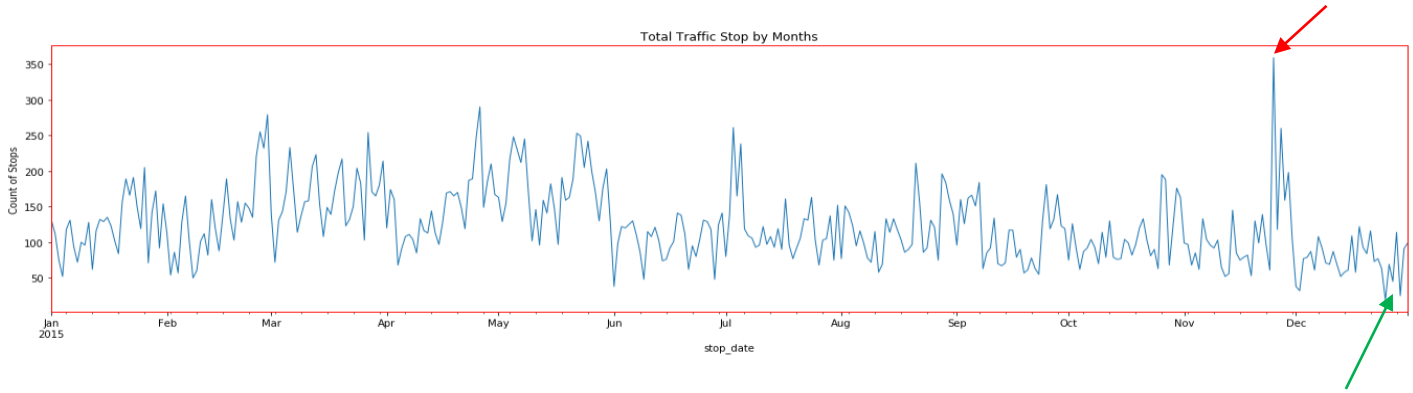


- Plotting the Traffic stop distribution between Age and Time of day.



5. Time-series plot

- Compare with trend chart of traffic stops over the dates of the entire year– using time series plot method.



- ✓ it is interesting to see there was a spike in march first week and a big spike in Thanksgiving weekend traffic stops.
- ✓ Last week of year has less frequent stops compared to entire year.
- ✓ Holidays plays a major role in increasing or decreasing traffic stop mainly due to the no of troops deployed on road by police department.

Other considerations: -

- We did not have the geolocation data available for traffic stops hence couldn't do much on find the hotspot of traffic stops. This can be achieved by combining latitude and longitude data with the google map api and show the hotspots in colors.