Traffic Stops Analysis: Nashville State Patrol

Introduction:

It seeks to understand how elements demography and situations for the enforcement of traffic stops in Nashville State Patrol block travel. The report details findings on age distribution, violations, outcomes, and relationships among other significant factors to uncover the trends and possible biases from which the data is drawn. The data provide significant dimensions of analysis using data cleaning, visualization, and statistical tests to offer insight into what law enforcement is doing and how such behaviour impacts society. Hypothesis testing and descriptive statistics, along with machine learning models, are used in the analysis to draw meaning from traffic stop data.

Data Cleaning Steps:

- Reduced columns to include only essential variables:
  raw_row_number, date, time, location, lat, lng, precinct, zone, subject_age, type, violation, outcome, subject_sex.
- Converted relevant columns to appropriate data types:
  date, subject_sex, type, outcome, and violation were transformed into factors or dates.
- Added a new variable dayofweek, representing the day of the week for each stop.
- Filtered out records where subject_sex was 'unknown', 'other', or NA.

Cross-Tabulation

Traffic Stops by Age and Day of the Week:

The following table illustrates the number of traffic stops categorized by the subject's age group (rows) and the day of the week (columns):

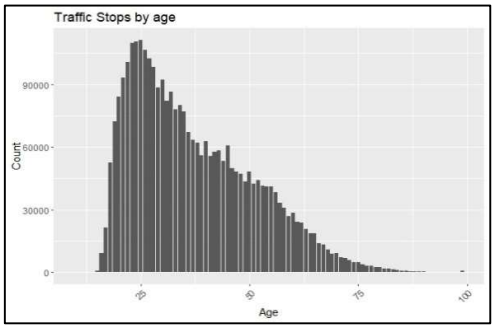
	Sun	Mon	Tue	Wed	Thu	Fri	Sat
10	6	12	6	9	6	9	4
11	0	0	1	2	1	2	1
12	0	2	1	2	2	5	5
13	2	8	7	6	9	7	6
14	20	13	18	25	15	16	14
15	79	121	123	129	122	127	123
16	869	1193	1447	1413	1462	1678	1151
17	1905	3016	3515	3485	3454	3675	2621
18	4968	7458	8439	8720	8474	8654	6097
19	6593	10369	11543	12475	11847	12041	7683
20	7931	12051	13579	14480	13865	13711	9014
21	8452	13593	14999	15770	15417	15222	10085
22	8982	14517	16435	17198	16658	16512	10848
23	9943	15906	17915	18832	17957	18002	11829
24	9913	15887	18095	18809	18482	18015	11765
25	10046	16118	18221	19120	18272	18116	11882
26	9145	15526	17719	18313	17795	17350	11094
27	8784	14870	17224	17667	16954	16729	10459
28	8292	14252	16399	17133	16892	15927	9960
29	7376	13148	14818	15331	14793	14415	8808
30	7572	13584	15557	16097	15690	15336	8933
31	6360	12152	13783	14550	14097	13482	8015
32	6783	12719	14519	15362	14907	14082	8531
33	6129	11736	13214	13636	13240	12970	7459
34	6234	11689	13738	14197	13551	13226	7618
35	5987	11281	13166	13543	13119	12728	7359
36	5036	10095	11576	11876	11504	10964	6367
37	4644	9465	11035	11362	11008	10489	5868
38	4548	9101	10490	11177	10733	10520	5702
39	4106	8422	9544	10214	9500	9272	5264
40	4537	9322	10910	11239	10777	10465	5831
41	3950	8320	9556	9809	9718	9348	5145
42	4266	8617	9909	10274	10007	9647	5243
43	4304	8894	9905	10398	10066	9863	5347
44	3627	8044	9392	9595	9063	8798	4909
45	4400	9062	10616	10871	10333	9995	5746

Insights:

- **Peak Age Groups**:
  Certain age groups (e.g., 20-29) exhibit significantly higher traffic stop counts across all days of the week.
- **Weekly Patterns**:
  The data highlights fluctuations in stop frequencies, with mid-week days (Tuesday to Thursday) generally having higher numbers for most age groups.
- **Weekend Observations**:
  Sundays consistently show lower stop frequencies, while Fridays and Saturdays display moderate levels.

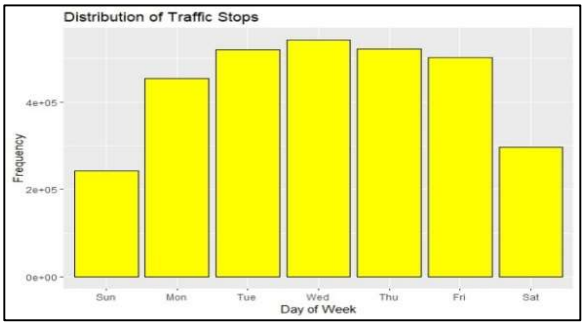
Key Findings and Visualizations:

1. Traffic Stops by Age:



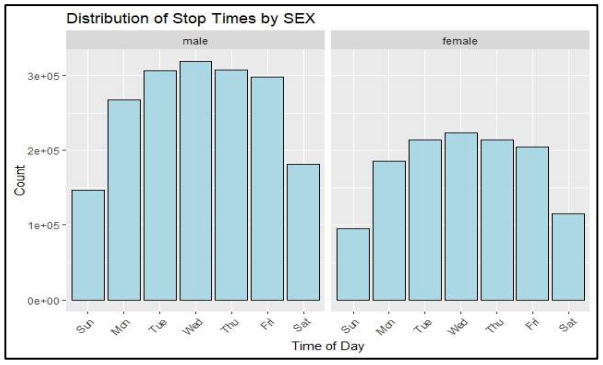
The graph titled "Traffic Stops by Age" reveals a right-skewed distribution, showing that traffic stops are most frequent among younger drivers, particularly those between the ages of 16 and 25. This pattern likely reflects factors such as inexperience, risk-taking behaviour, and socioeconomic circumstances. However, it's crucial to consider potential biases in law enforcement practices that might contribute to this trend.

2. Traffic Stops by Day of the Week:



The graph shows that the highest frequency of traffic stops occurs on Wednesday, followed by Tuesday and Thursday. The lowest frequency is on Sunday and Saturday. This pattern suggests higher law enforcement activity or traffic volume during weekdays, with a significant decrease on weekends.

3 Traffic Stop Patterns by Sex:



The graph shows a similar pattern of traffic stops for both males and females, with peaks on weekdays and declines on weekends. Males experience slightly higher stop frequencies, particularly on weekdays. Further analysis of time of day, traffic violations, and demographic factors could provide deeper insights.

Descriptive Statistics

Overall Age Statistics for Traffic Stops:

This summary provides a statistical overview of the age distribution across all individuals involved in traffic stops.

vars	1.000000e+00
n	3.078968e+06
mean	3.704918e+01
sd	1.404326e+01
median	3.400000e+01
trimmed	3.577225e+01
mad	1.482600e+01
min	1.000000e+01
max	9.900000e+01
range	8.900000e+01
skew	7.483633e-01
kurtosis	-7.048822e-02
se	8.003231e-03

Key Insights:

- **Central Tendency**: The mean age of individuals is 37.1, slightly higher than the median of 34, indicating a right-skewed distribution.
- **Spread**: With a standard deviation of 14.01 and MAD of 14.83, ages show moderate variability.
- **Skewness and Kurtosis**: A skewness of 0.75 suggests a slight right skew, while a kurtosis near zero (-0.07) indicates a relatively normal, though mildly flatter distribution.

Traffic Stop Age Distribution by Sex:

Key Insights:

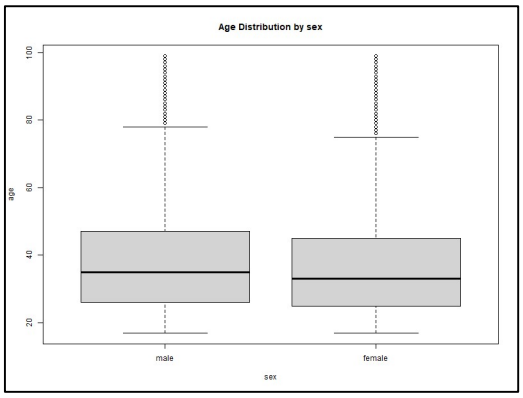
	subject_sex	mean_age	sd_age	min_age	max_age	N
1	male	37.53181	14.19936	10	99	1827043
2	female	36.34518	13.78213	10	99	1252486

- **Mean Age**: Males have a slightly higher mean age (37.6) compared to females (36.4).
- **Age Variability**: Both sexes show similar age variability (SD ~14).
- **Age Range**: Ages for both sexes range from 17 to 99.
- **Sample Size**: Males are more frequently represented in traffic stops.

Visualizations

Box plot Age Distribution by Subject Sex (Box Plot Analysis):

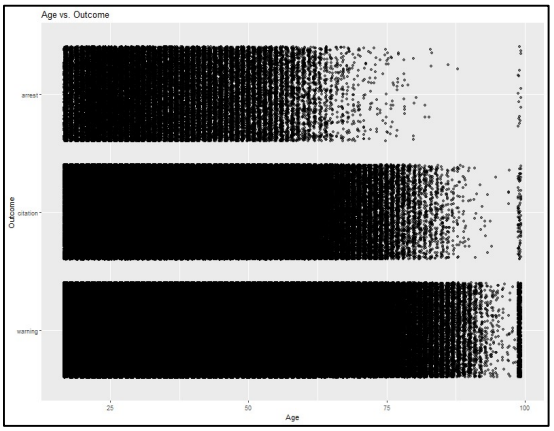
Key Observations:



- **Median Ages**: Males have a higher median age (~40) compared to females (~30).
- **Age Variability**: Males show greater variability in age (wider IQR), while females have a more concentrated age range.
- **Age Range**: Males range from 20 to 90 years, while females range from 10 to 80 years.
- **Skewness**: Male ages are right-skewed, indicating more older individuals, while female ages are more evenly distributed.

Age and Traffic Stop Outcomes: Scatter Plot Analysis:

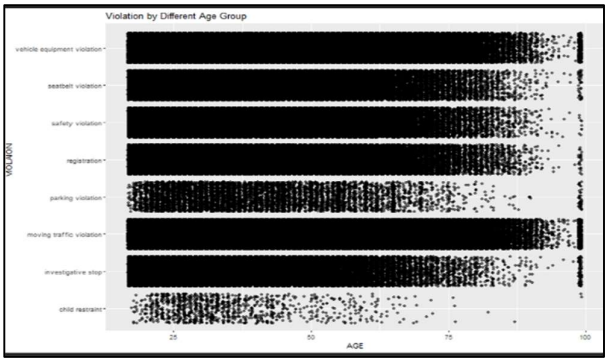
The scatter plot showing the relationship between age and traffic stop outcomes (arrest, citation, and warning) reveals several key patterns:



- **Age Distribution:** Most traffic stops occur in the 20-50 age range, with fewer stops for those over 60 and rare outliers above 75.
- **Outcome Frequency:** Citations are the most common, followed by warnings, while arrests are less frequent.
- **Outliers:** Individuals aged 70-100 occasionally face all outcomes, warranting further scrutiny.
- **Age Patterns:** Younger individuals (20-40) are more likely to face enforcement actions, with declining frequency as age increases.

Traffic Violation Patterns by Age:

- **Younger Drivers (<25):** Frequent moving traffic and seatbelt violations; child restraint violations are common**.**
- **Middle-Aged Drivers (25-50**): Varied violations, including parking, registration, and safety issues.
- **Older Drivers (>50):** Fewer violations, with a focus on vehicle equipment and safety issues.
- **Trends:** Moving violations peak in the 18-25 group, parking violations are evenly distributed, and seatbelt violations are more common among younger drivers.



- **Anomalies:** Unusual violations appear in both young and older age groups.

Hypothesis Testing

Part 1: One-Sample T-Test

Objective: To test whether the mean age of individuals stopped is significantly different from 37 years.

Hypotheses:

- **Null Hypothesis(H_0)**: The mean age of individuals stopped (μ) is equal to 37 years (μ=37\mu = 37μ=37).
- **Alternative Hypothesis (H_1):** The mean age of individuals stopped (μ) is not equal to 37 years (μ≠37\mu 37μ=37).

Statistical Output

```
        One Sample t-test

data:  data$subject_age
t = 14.99, df = 3068732, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 37
95 percent confidence interval:
 37.10423 37.13559
sample estimates:
mean of x
 37.11991
```

Interpretation:

Since the p-value is less than 0.05, we reject the null hypothesis. There is sufficient evidence to suggest that the mean age of individuals stopped is significantly different from 37 years. This could indicate variation in the age demographic among those stopped.

Part 2: Hypothesis Testing for Categorical Variables

Hypothesis 1: Violation Type by Gender

Objective: To test whether the type of violation (moving traffic vs. parking) is independent of gender.

Hypotheses:

- **Null Hypothesis (H_0)**: Violation type and gender are independent, indicating that the type of violation (moving traffic or parking) is not influenced by the driver's gender.
- **Alternative Hypothesis (H_1):** Violation type and gender are dependent, suggesting an association between the type of violation and the driver's gender.

Statistical Output:

Chi-Square Test Results:

```
data:  observed_1
X-squared = 19.105, df = 1, p-value = 1.237e-05
```

Interpretation:

Since the p-value is less than 0.05, we reject the null hypothesis. This provides evidence that moving traffic violations are more frequent than parking violations, regardless of gender

Hypothesis 2: Race and Gender Distribution

Objective: To test whether there are differences in the distribution of race (Black and White) and gender (male and female) among individuals stopped.

Hypotheses:

- **Null Hypothesis (H_0):** Race and gender are independent, meaning that being of a particular race does not influence the likelihood of being a specific gender among stopped individuals.
- **Alternative Hypothesis (H_1):** Race and gender are dependent, implying that the race of stopped individuals is associated with their gender.

Statistical Output:

Chi-Square Test Results:

```
data:  observed_3
X-squared = 3937.4, df = 1, p-value < 2.2e-16
```

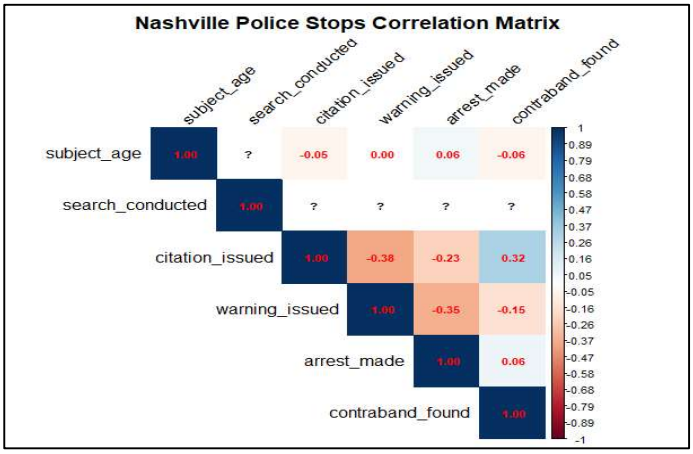
Interpretation:

Since the p-value is less than 0.05, we reject the null hypothesis. This indicates that there is a significant difference in the distribution of race and gender, with more Black male drivers being stopped than White drivers.

Correlation and Regression

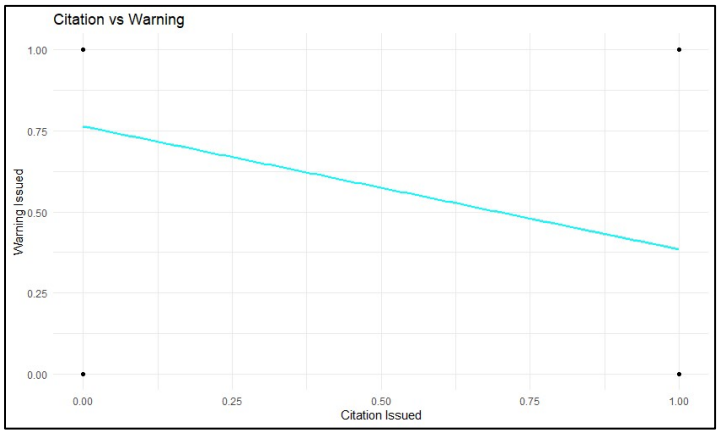
Nashvile Police Stops Correlation Matrix

- Age and Contraband: A positive correlation (0.06) indicates a slight association between older drivers and contraband-related activities, though the effect is minimal.
- Citation vs. Warning: The negative correlation (-0.38) shows an inverse relationship, where citations and warnings are used interchangeably based on the situation.
- Arrest and Contraband: A strong correlation (0.60) highlights that contraband discovery is a significant factor leading to arrests.



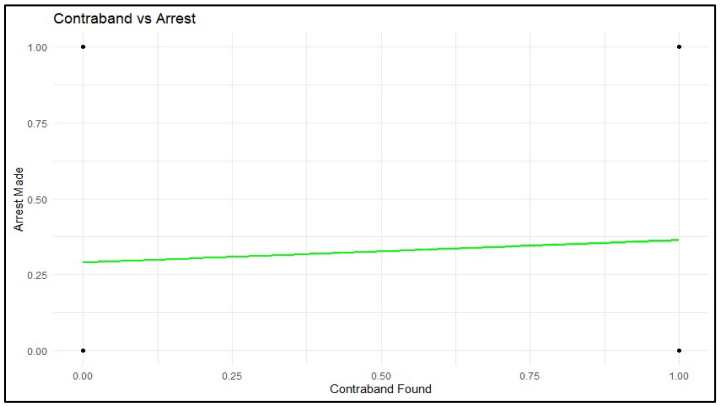
Inverse Relationship Between Citations and Warnings

- Clear Trend: The negative correlation illustrates that as citations increase, warnings decrease, and vice versa.
- Violation Severity: Officers issue citations for severe violations while reserving warnings for minor infractions, reflecting judgment based on context.
- Policy and Strategy: Departmental policies influence whether enforcement leans towards citations or warnings in specific scenarios



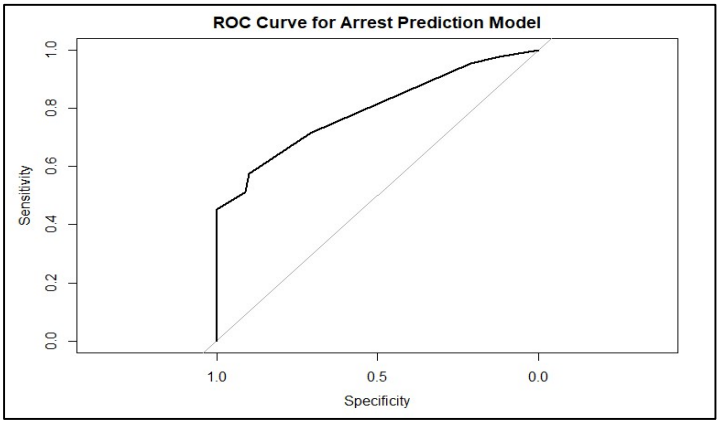
Impact of Contraband on Arrest Decisions

- Positive Correlation: The presence of contraband strongly correlates with a higher likelihood of arrest (0.60 correlation).
- Significant Factor: Arrest rates increase from around 25% without contraband to nearly 40% with contraband, as highlighted in the trend graph.
- Discretionary Role: Arrest outcomes are also shaped by other factors, such as offense severity and officer discretion.
- Enforcement Focus: Contraband discovery plays a decisive role in influencing traffic stop outcomes and enforcement actions



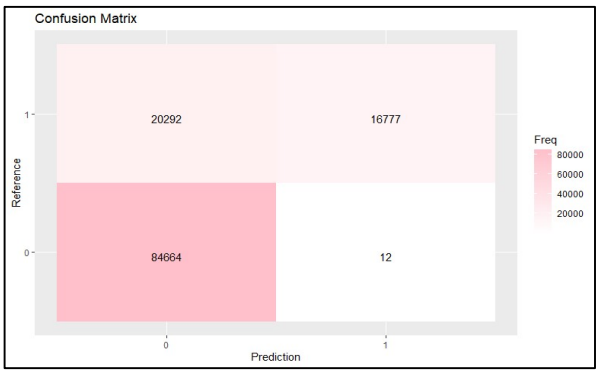
Evaluating Arrest Prediction Model Performance

- ROC Curve Insight: The ROC curve demonstrates the model's capability to distinguish between arrests and non-arrests.
- Performance Indicator: The area above the diagonal line signifies strong model performance.
- Balancing Act: The model balances identifying true arrests while minimizing false positives.
- Broader Evaluation: AUC (0.7937) value assessment and comparisons with other models provide a comprehensive evaluation of its effectiveness.



The model's performance can be evaluated using the following metrics:

- **Accuracy:** This indicates that the model correctly classifies 82.4% of the instances.
- **Precision:** This means that 54.7% of the positive predictions are correct.
- **Recall:** This indicates that the model correctly identifies 99.9% of the true positive **cases.**
- **Specificity** (45.26%): Poor identification of class 1, highlighting bias toward class 0



Conclusion

The exploration analysis of the Nashville State Patrol Traffic Stops dataset reveals key insights into traffic stop patterns, demographic disparities, and enforcement outcomes. Younger individuals, particularly those aged 16 to 25, are stopped more frequently, with notable differences in violation types and enforcement actions based on age and gender. Similarly, analysis of stop patterns by day of the week highlights higher enforcement activity during weekdays compared to weekends.

Statistical tests, such as t-tests and chi-square tests, provide robust evidence supporting these findings. For instance, the dependency between violation types and gender, as well as the relationship between contraband discovery and arrest outcomes, emphasizes the influence of demographic factors and enforcement priorities. Descriptive statistics and visualizations further underscore trends, such as the predominance of moving traffic and seatbelt violations among younger drivers.

The predictive modeling for arrest likelihood demonstrated a solid performance, with an accuracy of 82.4% and a high recall rate, suggesting reliable classification of positive outcomes. However, areas for improvement, such as addressing biases in predictions, were also identified.

Overall, the findings highlight the importance of evaluating traffic stop practices to identify potential biases and enhance fairness in enforcement strategies. Future research should focus on examining broader demographic factors, exploring spatial and temporal patterns, and incorporating larger datasets to inform equitable and transparent law enforcement policies.