https://doi.org/10.51153/kjcis.v5i2.135

Statistical Analysis for the Traffic Police Activity: Nashville, Tennessee, USA

M. Y. Tufail[1*] S. Gul[2]

Abstract

Data Science is one of the fastest growing interdisciplinary field and has many applications in various disciplines. The actual motivation of data science came from John Tukey. In his seminal paper, in 1962, he presented the idea of data analysis which is now the field of data science. Several algorithms for data science related to statistical analysis have been developed and applied over variety of datasets since 1962. In this field, the significant development began with the aid of high performance computers that help to analyse a massive datasets. In this paper, we study the statistical analysis of the traffic stops in Nashville, Tennessee, USA for the year 2011–2021. Data is taken from the Stanford open policing project. Analysis is based on total number of 3071706 traffic stops. In this paper, we consider and investigate various aspects. This study comprises gender comparison (male vs female) and race comparison (black vs white) for different traffic offences. Complete findings and possible gaps are discussed in the conclusion.

Keyword: Data analysis, Statistical analysis, Traffic stops analysis, Traffic related social issues.

1. Introduction

Data Science is one of the fastest growing interdisciplinary field and has many applications[1–3]. John Tukey can be considered as the pioneer of Data analysis. In 1962, nearly 60 years ago, in his seminal paper [4] he published the idea of data analysis, that is now a field of data science [4, 5].

There have been many developments in data science since 1962. Data science is widely been used in various disciplines such as, social sciences [6–8], data engineering [9, 10], data mining [11, 12], predictive analytics [13, 14], machine learning [15, 16], image processing [17–20], data visualization [21, 22] and many more [23, 24].

[1]*Department of Mathematics, NED University of Engineering & Technology, University Road, Karachi, 75270, Sindh, Pakistan | tufail@neduet.edu.pk
[2]Department of Mathematics, NED University of Engineering & Technology, University Road, Karachi, 75270, Sindh, Pakistan l sagul@neduet.edu.pk

The most significant boost in the field started due to the high-performance computers and the use of statistical analysis [25–27], that make this field inter- disciplinary and ease the gigantic calculations. Statistical analysis is almost used everywhere, whenever we deal with datasets [28–33]. Conclusions made in this paper heavily relies on statistical analysis.
In this paper, we apply statistical analysis to the data (during the year 2011 to year 2020) of the traffic stops by the police officers at Nashville, Tennessee, USA. Total number of traffic stops in the considered data is 3071706. Complete details of the data set can be found in the link https://openpolicing.stanford.edu/data/ .

More than 20 million Americans are stopped each year for traffic violations [34]. Without any doubt, police stops is the most common form of interaction, the public has with the police around the world [35–38]. These activities and interactions help the police to achieve both traffic safety and crime control [39]. In recent years, the traffic stops data, is largely studied and analysed to understand the behavior of police and their interactions with the public, specially at traffic stops [40–43]. The law, prohibits, law enforcement agencies from stopping, detaining, or searching motorists when the stop is motivated solely based on the race, color, ethnicity, age, gender, or sexual orientation of drivers [6]. But incidents like George Floyd [44, 45] raise a very big question mark over police behavior, their act and their biased views towards specific race. These kind of incidents make the specific race vulnerable and cause the anger within the race towards other race. Further, we would also like to high- light the fact that (apparently) society is not gender biased. But we believe that the behavior of police towards the gender requires attention and statistical analysis.

This study comprises gender comparison (male vs female) and two race comparison (black vs white) for different traffic offences. Various aspects are considered and investigated. Complete findings and possible gaps are discussed in the conclusion.

List of desired variables (that are given in the form of columns in the dataset) are demonstrated in the Table 1 below:

Table 1: This table displays the variables and their corresponding details.

Details of the variables	
Column Names	**Remarks**
Date	Date of a traffic stop
Time	Time of a traffic stop
Subject race	The race of a subject
Subject sex	The gender of a driver
Violation	Eight different violations are presented in this column that include moving traffic violation, vehicle equipment violation, safety violation, registration, seatbelt violation, investigative stop, parking violation and child restraint.
Arrest made	It is a Boolean column [46, 47]. This columns has only two logical values, (i) 'True': this value indicates that arrest has been made and (ii) 'False': it Indicates that no arrest has been made due to traffic violation.
Outcome	This columns contains the information of the outcome corresponding to the respective traffic stop. Three distinct features for outcomes are available that include warning [48], citation [49] and arrest [50].
Contraband drugs	A Boolean columns consist upon two logical values. (i) 'True': it indicates that contraband drug is found and (ii) 'False: when contraband drug is not being found.
Contraband weapons	Analogous to the above explanation.
Frisk performed	A Boolean column analogously
Search conducted	A Boolean column analogous to other Boolean columns.

Table 2: Comparison of traffic violation between the male and female for Nashville, Tennessee, USA is given.

Male vs Female comparison for eight different categories of traffic violation				
Violations	**Female**	**Female relevant ratio (individual values divided by total sum**	**Male**	**Male relevant ratio**
Moving traffic violation	633001	0.510	907881	0.50
Vehicle equipment violation	407379	0.330	585433	0.321
Safety violation	75168	0.060	110037	0.060
Registration	76994	0.062	107948	0.059
Seatbelt violation	33915	0.027	68932	0.038
Investigative stop	19275	0.015	36879	0.020
Parking violation	2992	0.002	4757	0.003
Child restraint	725	5.8×10^{-4}	390	2.14×10^{-4}
Total	1249449		1822257	

2. Statistical analysis

In this section we provide statistical analysis over five different examples using above mentioned dataset. The details are given below.

Example 1 (Traffic violation (Male vs Female)) In this warm-up example, we com- pare the count of eight different traffic violation committed by male and female for Nashville, Tennessee, USA during the year 2011–2020. We have found that nearly 50% of traffic violations are related to moving traffic violations (MTV) and this per- centage is nearly the same for both male and female drivers. MTV are those traffic violations that occur when vehicle is in motion such as over speeding, stop sign violation, give way violation, driving under the influence of alcohol or drugs, hit and runs etc [51–55]. We have also found that the over all rate of traffic violation is higher in males than females. According to 2010 census, composition of male population is ≈ 48.5%
whereas; ≈ 51.5% of female population [56, 57]. But whether or not the data is biased (with respect to number of drivers) would be an interesting future problem to address. Results are presented in Table 2 and graphical representation can be seen in Figure 1.

Table 2 suggests that most of the violations are related to moving traffic violations for both the genders. Nearly 50% violation involve MTV. Males are involved in 59% of the total violation whereas, females involvement is 41%. Both male and female are conscious when it comes to the child safety. Apparently, Table 2 indicates that males seems to be more concerned than female regarding child safety.

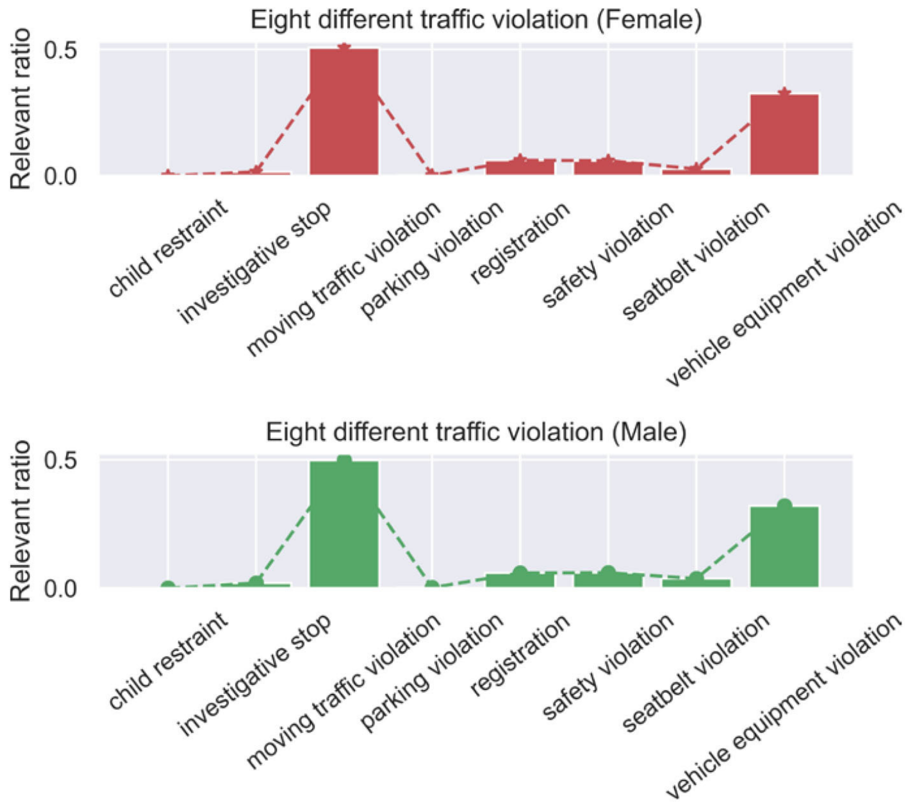


Figure 1: This graph represents the comparison of relevant ratios between male and female drivers among eight different categories of traffic violation whose details are given in Table 2. In all eight different categories, violation rate is nearly the same for both the genders. But for seatbelt violation, male ratio is slightly higher than female.

But, due to the limitation of current dataset, we have not investigated which gender transport their children mostly. It would be an interesting research problem and we aim to address it in future.

As the rate of moving traffic violation is higher than the rest of traffic offences and it covers nearly 50% of the total traffic violations. Therefore, we further explored this traffic offence in three different categories (with respect to the outcome/result of these traffic stops), i.e., warning, citation, arrest. Details are given in Table 3.

We further investigated the moving traffic violation among the different age groups. The frequency distribution for this classification is given in Table 4.

Table 4 indicates that the involvement of young driver in moving traffic violation is higher than the older people. This may be because most drivers are under 40.

Moving traffic violation: comparison between male and female		
Categories	**Female**	**Male**
Warning	0.6859	0.6712
Citation	0.3048	0.3103
Arrest	0.0093	0.0185

Table 3: Data represents the relevant ratios for the outcome due to moving traffic violations (MTV) for three different categories for female (total=632856) and male (total=907752) drivers out of 1540608 total MTV. The ratios are nearly similar for males and females. About 68% of stops for MTV result in a warning. This investigation also indicates that for MTV, the outcome is unbiased for gender.

Table 4: The frequency distribution for moving traffic violation corresponding to different age group.

Moving traffic violation VS Age group	
Age groups	**Number of observations (x)**
$10 \le x \le 20$	119391
$20 < x \le 30$	485533
$30 < x \le 40$	356318
$40 < x \le 50$	269579
$50 < x \le 60$	191360
$60 < x \le 70$	87795
$70 < x \le 80$	25097
$80 < x \le 90$	5185
$90 < x \le 100$	624
	Total=$\sum x$ = 1540882

It is still an open problem because we are not too sure whether or not population has a similar number of people in all age groups (which is highly unlikely in unbiased data). Approximately 62.4% offences are caused by under 40 age group. The violations are keep decreasing with the maturity of a driver. Figure 2 illustrates this fact.

Example 2 (Search rate corresponding to each violation (Male vs Female)) In this example, we have calculated the mean search rate† among male and female drivers for

† Table 1 indicates that the 'search conducted' is a Boolean column (0: no search, 1: search is done). Mean search rate is simply an arithmetic mean ($\sum_{i=1}^{N} x_i/N$) for search corresponding to each violation.

eight different categories of traffic violations. We have found that the mean search rate for males are higher than female in all eight offences. Limitation of the data stops us to investigate the fact whether is data is being biased corresponding to specific gender. Investigative stops caused the highest mean search rate among all the violations for both the genders. Complete details are given in Table 5.

Figure 2: Comparison of moving traffic violations among different age groups. Young drivers are involved in more offences than the older drivers.

Table 5: This table shows the average values for search rate corresponding to each traffic violation for male and female. For all types of violations, it seems that the search rate is higher for males than for females.

Mean values for search rate corresponding to each traffic violation: Female vs Male		
Gender	**Violations**	**Mean values**
Female	Child restraint	0.030345
	Investigative stop	0.094319
	Moving traffic violation	0.019689
	Parking violation	0.024733
	Registration	0.025768
	Safety violation	0.023334
	Seatbelt violation	0.031903
	Vehicle equipment violation	0.024815
Male	Child restraint	0.071795
	Investigative stop	0.184089
	Moving traffic violation	0.046675
	Parking violation	0.047719
	Registration	0.055277
	Safety violation	0.048620
	Seatbelt violation	0.060277
	Vehicle equipment	
	Violation	0.055152

Example 3 (Overall Search rate and Frisk rate (Male vs Female)) This example gives the details of overall search rate, frisk rate and arrest rate for male and female drivers. It is found that males are leading in all three categories in a given dataset (keeping the limitation of the data in account). Complete details are provided in Table 6.

Table 6: Male drivers are searched and arrested more than twice (approximately) as often as female drivers. Whereas, frisk rate among males are nearly three times higher than female drivers.

Overall mean Search rate, frisk rate and arrest rate between Male and Female		
Categories	Female	Male
Search rate	0.0235	0.0533
Frisk rate	0.008595	0.027683
Arrest rate	0.010633	0.019976

We have further investigated the overall all mean arrest rate for a period of 2011–2020 for a given 24 hours a day. ‘0’ indicates the midnight, ‘12’ represents the noon and ‘23’ states the 11:00 PM. Results are shown in Figure 3. It is found that the arrest rate is higher in overnight than any other hour of a given day.

Example 4 (Contraband Drugs and Contraband Weapons) In this example, we are trying to investigate whether or not the rate of contraband drugs and contraband weapons are increased over the past 10 years. Drug related stops are shown in Figure 4. Whereas, weapon related stops can be found in the Figure 5.

It is found that the drug related stops have kept increasing every year for the course of past ten years. However, weapon related stops have (continuously) declined during this period.

The comparison of drug related stops with the search rate for the period of past ten years is presented in Figure 6.
In this example, we have found that the drug related stops are increasing whereas, weapon related stops and search rate are decreasing. To check the validity of this claim, we further explore these trends. Non-parametric regressions [58, 59] results are given in the Figure 7 and the results of Modified Mann–Kendall [60] test are presented in Table 7.

Example 5 (Black race vs White race comparison corresponding to each violation) The motivation for this example came after the murder of George Floyd [44, 45, 63, 64]. This incident happened in Minneapolis, Minnesota, USA, dated: May 25, 2020. We investigate whether the police officers are biased for any race or they are neutral. Table 8 displays the comparison of black and white race drivers for eight different traffic violations.

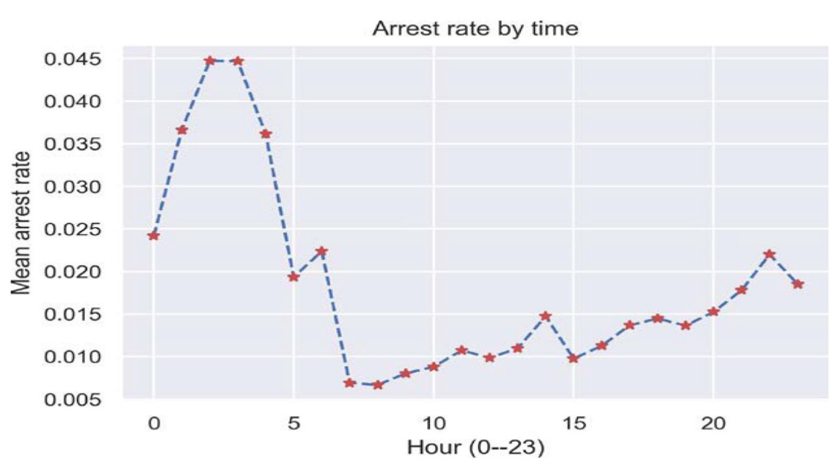


Figure 3: Graph indicates the hourly mean arrest rate for ten years period (2011–2020) for Nashville, Tennessee, USA. 0, 12, 23 indicate the midnight, noon and 11:00 PM respectively in a given day. The arrest rate has a significant spike overnight, and then dips in the early morning hours.

Table 7: This table confirms the actual trend with their respective parameters for Example 4. See [60–62] for more details.

Modified Mann-Kendall Test (at 5% level)		
Search rate	**Drug rate**	**Weapon rate**
0.0318	0.00067	0.1524
Decreasing	Increasing	No trend
-0.555	0.866	-0.377
-0.00068	0.018	-0.001
0.043	0.126	0.0196

Their respective relevant ratio can also be seen. Majority of the violations involve moving traffic violations (MTV) for both the races. Data suggests that there are total 2819799 traffic offences among black and white race. Out of which black people have committed 1158721
(≈ 41%) offences. Whereas, white people are involved in 1661078
(≈ 59%) offences. Apparently, this implies that white people committed more traffic offences than black race. But it is due to the

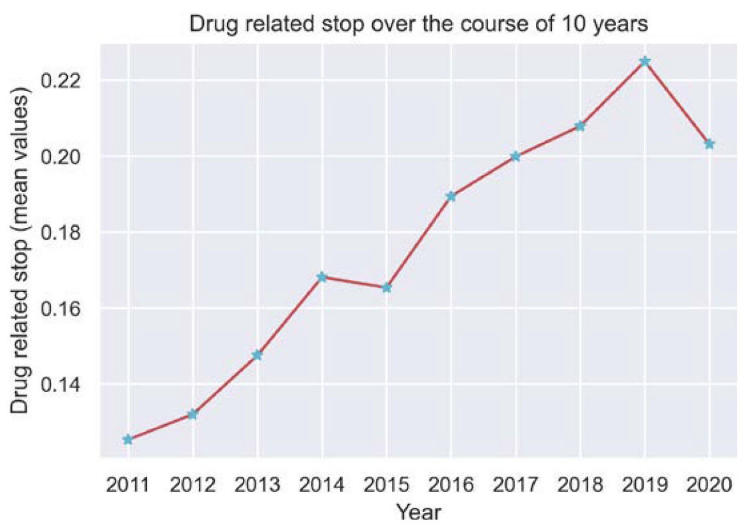


Figure 4: Graph indicates mean annual contraband drugs related traffic stops. Drug related stops are increasing every year. Surprisingly, The rate of drug related stops increased (nearly) doubled over the course of 10 years.

https://doi.org/10.51153/kjcis.v5i2.135

Figure 5: Graph indicates mean annual contraband weapons related traffic stop. Weapon related stops have decreased every year except during 2015–2017. The rate of weapon related stops decreased, and ratio is nearly one-seventh over the course of 10 years.

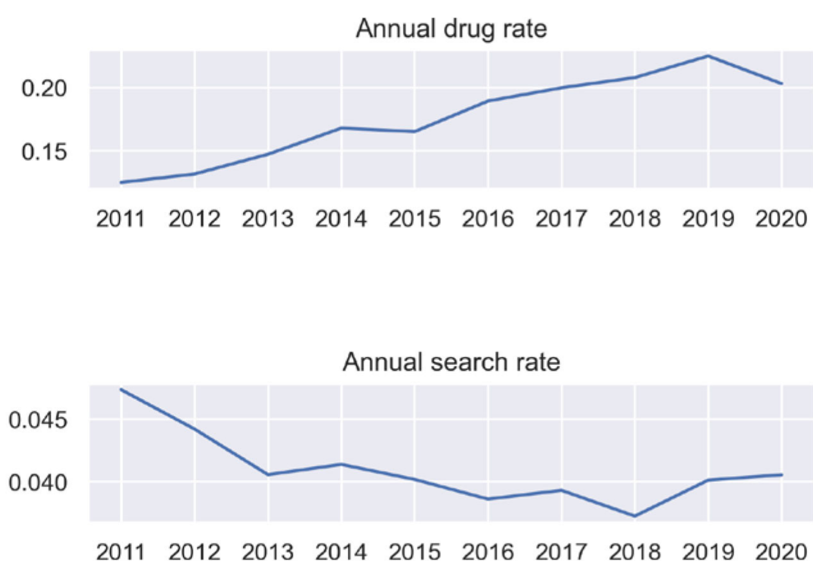


Figure 6: The rate of drug-related stops are continuously increasing. But surprisingly, the search rate is decreasing

fact that the composition of white race is higher than the black race, i.e., ≈ 56.3% versus ≈ 27.4%. (See [65, 66] for more details). The relevant ratios are plotted in Figure 8.

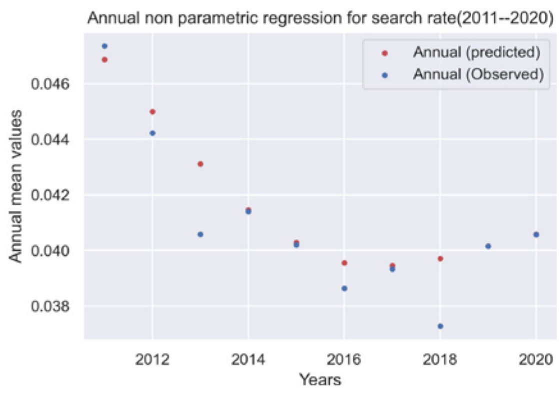
As we have found that the white race committed 59% traffic violations. Therefore we extend our investigation for three different categories that are search rate, frisk rate and arrest rate for both the races. Although (overall) white people committed more traffic offences than black but the mean search rate, frisk rate and arrest rate are higher in black race. Details can be found in Table 9.

3. Conclusion

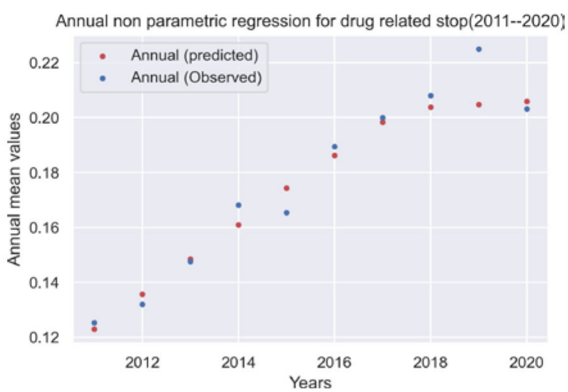
We have investigated the traffic police activity for Nashville, Tennessee, USA during 2011–2020. There are eight different traffic violations are investigated along with the outcomes of the offences (Example 1–3, 5). Example 4 contains the investigation of drug and weapon related stops. Overall findings are as below:

1. Over all males committed more traffic violations than female. But due to the limitation of given dataset, we have not investigated whether or not this dataset is biased.

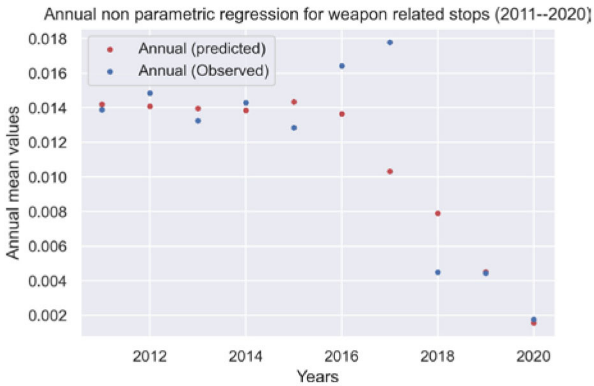
2. The rate of moving traffic violation is higher than rest of remaining traffic violations.



(a) Search rate (2011–2020)



(b) Drugs rate (2011–2020)



(c) Weapon rate (2011–2020)

Figure 7: LOWESS regression for search rate, drug rate and weapon rate in Nashville, Tennessee during traffic stops are presented.

Table 8: This table shows the comparison between two different races (black vs white) corresponding to eight different categories of traffic violation. White race committed 1661078 traffic violation. Whereas, black race committed 1158721 traffic violation. Majority of the violations are due to MTV for both the races.

Comparison between black Race and white race corresponding to each traffic violation			
Race	**Violation**	**Counts**	**Relevant ratio**
Black violation	Child restraint	598	5.2 × 10−4
	Investigative stop	25045	0.021614
	Moving traffic	532122	0.459232
	Parking violation	3719	0.003210
	Registration	74015	0.063876
	Safety violation	78708	0.067927
	Seatbelt violation	40954	0.035344
	Vehicle equipment violation	403560	0.348281
White	Child restraint	286	1.72 × 10−4
	Investigative stop	25276	0.015217
	Moving traffic	880021	0.529789
	Parking violation	3480	0.002095
	Registration	98067	0.059038
	Safety violation	89684	0.053991
	Seatbelt violation	54939	0.033074
	Vehicle equipment violation	509325	0.306623

Table 9: This table illustrates the fact that (surprisingly) mean frisk rate, arrest rate and search rate are higher in black race drivers than white.

Overall mean Search rate, frisk rate and arrest rate between Black and White race		
Categories	**Black**	**White**
Search rate	0.058130	0.028535
Frisk rate	0.029388	0.012467
Arrest rate	0.022481	0.010810

3. Approximately 68% MTV results in a warning.

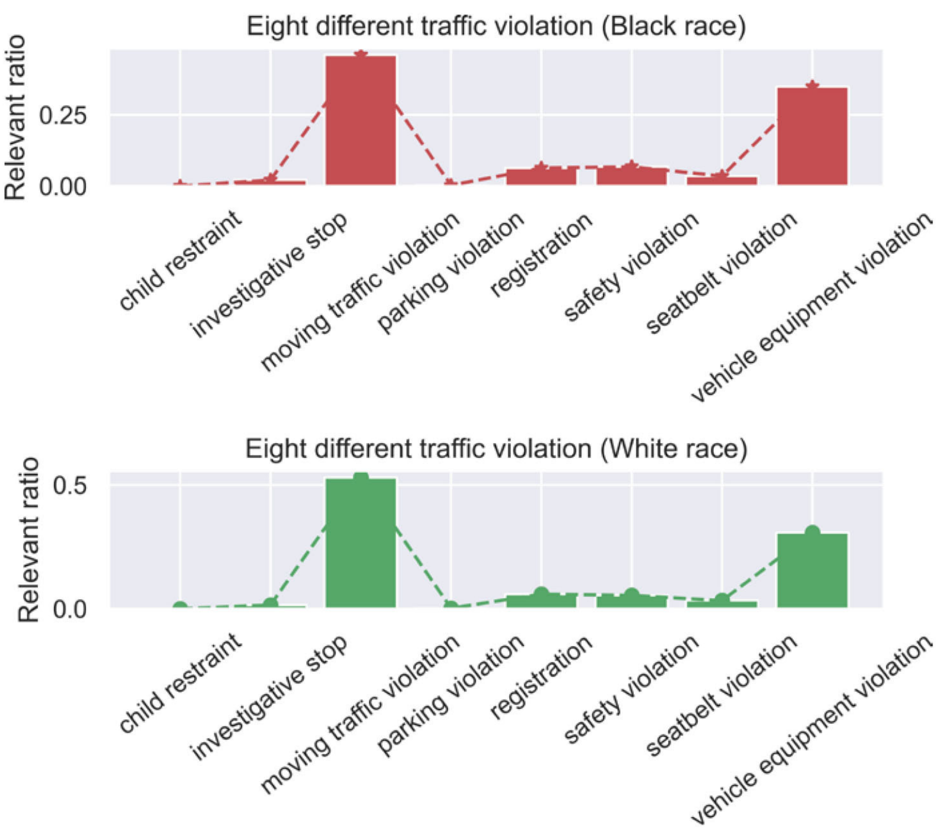


Figure 8: This graph represents the comparison between black and white race drivers among eight different categories of traffic violation. Majority of the violations are related to MTV for both the races. The relevant ratio for MTV is higher in white race than black. Whereas, vehicle equipment violation and safety violation are higher in black race. Overall white race committed more offences than black due to the fact that the composition of white race is higher than black.

4. Outcome of moving traffic violations do not indicate any kind of favoritism. Thus, the investigation concludes that for MTV, the outcome is unbiased for gender.
5. Overall search rate, frisk rate and arrest rate are higher in male's drivers than females.
6. We have found that the drug related stops are continuously increasing for the past ten years.
7. Weapon related stops indicated decreasing trend initially. But Mann-Kendal test confirms that there is no definite trend during these years.
8. We have also found that white race committed more traffic offences than black. But it is due to the fact that the composition of white race is higher than the black race. Nevertheless, the mean search rate, frisk rate and arrest rate are higher in black race.

Although we have found that males drivers committed more traffic violations than female drivers. But we have not investigated the gender of police officer. This could be our next task for the exploration. We would also like to explore the race of police officer and its impact over the decision. We have not investigated the effect of weather over a police officer. It would be very interesting question if we analyse the effect of weather over a decision.

References

[1] Van Der Aalst, W.: Data Science in Action, pp. 3–23. Springer, London (2016)

[2] Provost, F., Fawcett, T.: Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc., Sebastopol (2013)

[3] Peyr´e, G., Cuturi, M., et al.: Computational optimal transport: With applications to data science. Foundations and Trends@ in Machine Learning 11(5-6), 355–607 (2019)

[4] Tukey, J.W.: The future of data analysis. The annals of mathematical statistics 33(1), 1–67 (1962)

[5] Tukey, J.W.: Available online. https://en.wikipedia.org/wiki/Data science

[6] McCabe, J.E., Kaminski, R.J., Boehme, H.M.: Racial profiling and ct motor vehicle stops: an observational study in three towns. Police Practice and Research 22(6), 1567–1584 (2021)

[7] Scheb, J.M., Lyons, W., Wagers, K.A.: Race, gender, and age discrep- ancies in police motor vehicle stops in Knoxville, Tennessee: evidence of racially biased policing? Police Practice and Research: An International Journal 10(1), 75–87 (2009)

[8] Renauer, B.C.: Neighborhood variation in police stops and searches: A test of consensus and conflict perspectives. Police quarterly 15(3), 219–240 (2012)

[9] Gray, J., Shenoy, P.: Rules of thumb in data engineering. In: Proceedings of 16th International Conference on Data Engineering (Cat. No. 00CB37073), pp. 3–10 (2000). IEEE

[10] Birnholtz, J.P., Bietz, M.J.: Data at work: supporting sharing in science and engineering. In: Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, pp. 339–348 (2003)

[11] Tan, P.-N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Education India, Noida (2016)

[12] Grossman, R.L., Kamath, C., Kegelmeyer, P., Kumar, V., Namburu, R.: Data Mining for Scientific and Engineering Applications vol. 2. Springer, Dordrecht (2013)

[13] Eckerson, W.W.: Predictive analytics. Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report 1, 1–36 (2007)

[14] Waller, M.A., Fawcett, S.E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. Wiley Online Library (2013)

[15] Jordan, M.I., Mitchell, T.M.: Machine learning: Trends, perspectives, and prospects. Science 349(6245), 255–260 (2015)

[16] Goodfellow, I., Bengio, Y., Courville, A.: Machine learning basics. Deep learning 1(7), 98–164 (2016)

[17] Murtagh, F., Starck, J.-L.: Image processing through multiscale analysis and measurement noise modeling. Statistics and Computing 10(2), 95–103 (2000)

[18] Hill, D.L., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Physics in medicine & biology 46(3), 1 (2001)

[19] Tufail, M.Y.: Image registration under conformal diffeomorphisms: a thesis presented in partial fulfilment of the requirements for the degree of doctor of philosophy in mathematics at massey university, palmerston north, New Zealand. PhD thesis, Massey University (2017)

[20] Marsland, S., McLachlan, R., Tufail, M.: Conformal image registration based on constrained optimization. The ANZIAM Journal 62(3), 235–255 (2020)

[21] Chen, C.-h., H¨ardle, W.K., Unwin, A.: Handbook of Data Visualization. Springer, Berlin Heidelberg (2007)

[22] Waskom, M.L.: Seaborn: statistical data visualization. Journal of Open Source Software 6(60), 3021 (2021)

[23] Stedman, C.: Available online. https://searchenterpriseai.techtarget.com/definition/data-science

[24] Hilbe, J.M.: A review of stata 9. 0. The American Statistician 59(4), 335–348 (2005)

[25] Rzide-Gothenburg, L.: Statistics and the computer. ICOTS, 435–439 (1990)

[26] Lunn, D.: Computer animation: A powerful way of teaching concept of probability and statistics. In: Teaching of Statistics in the Computer Age. Proceedings of the 6th ISI Round Table Conference on Teaching of Statistics, Chartwell Bratt Ltd., Bromley, pp. 114–125 (1985)

[27] Diaconis, P., Efron, B.: Computer-intensive methods in statistics. Scien- tific American 248(5), 116–131 (1983)

[28] Zammit-Mangion, A., Rougier, J.: Multi-scale process modelling and dis- tributed computation for spatial data. Statistics and Computing 30(6), 1609–1627 (2020)

[29] Kuttikkad, S., Chellappa, R.: Statistical modeling and analysis of high- resolution synthetic aperture radar images. Statistics and Computing 10(2), 133–145 (2000)

[30] Schouten, B., Cigrang, M.: Remote access systems for statistical analysis of microdata. Statistics and Computing 13(4), 381–389 (2003)

[31] Rafanelli, M.: Aggregate statistical data: models for their representation. Statistics and Computing 5(1), 3–24 (1995)

[32] Takai, K.: Incomplete-data fisher scoring method with steplength adjustment. Statistics and Computing 30(4), 871–886 (2020)

[33] Lyu, J., Gunasekaran, A.: Statistical analysis of a portable parallel non- linear programming algorithm. Statistics and Computing 4(4), 253–258 (1994)

[34] Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., et al.: A large-scale analysis of racial disparities in police stops across the united states. Nature human behaviour 4(7), 736–745 (2020)

[35] Davis, E., Whyde, A., Langton, L.: Contacts between police and the pub- lic, 2015. US Department of Justice Office of Justice Programs Bureau of Justice Statistics Special Report, 1–33 (2018)

[36] Langton, L., Durose, M.R., et al.: Police Behavior During Traffic and Street Stops, 2011. US Department of Justice, Office of Justice Programs, Bureau of Justice , USA (2013)