

Analyzing Arrest Patterns and Racial Representation in Toronto*

Akshat Aneja

September 26, 2024

This paper analyzes arrest data from the Toronto Police Service for the years 2020 and 2021, focusing on the distribution of arrests by ethnic identity, age, and booking status to assess whether systemic bias exists in policing practices. The analysis reveals that, while arrests appear to be driven by objective factors, disparities related to socio-economic status may still exist. These findings provide important insights that can help improve transparency and accountability in policing, contributing to more equitable law enforcement practices.

Table of contents

1	Introduction	1
2	Data	2
2.1	Measurement	3
2.2	Total number of arrests by age group for the year 2020 and 2021	3
2.3	Arrests by Ethnic and Identity-Based Groups for Year 2020 and 2021	4
2.4	Total Arrests in 2020-2021	5
2.5	Number of Arrests by Age Group for the year 2020-2021	6
3	Discussion	7
	References	9

*Code and data supporting this analysis is available at: <https://github.com/Akshat211202/Analysis-of-bias-in-arrests/>

1 Introduction

Racial bias in policing has become a central issue in public discourse, particularly regarding its disproportionate impact on marginalized communities. This issue is of special significance in cities like Toronto, known for its cultural and ethnic diversity. Amid calls for police reform, historical ties to slavery and contemporary racial disparities often dominate discussions, prompting concerns that marginalized groups may experience discriminatory treatment, including racial profiling, in interactions with law enforcement.

A report from the Toronto Police Services highlights that individuals from Black, East/Southeast Asian, Middle Eastern, and Latino communities are overrepresented in use-of-force incidents compared to their presence in enforcement action populations. Despite efforts by the Toronto Police to promote fairness and equity, these findings suggest that racial bias may still persist in policing practices. However, a thorough, data-driven analysis of arrest patterns is necessary to determine the extent and nature of this bias. This paper addresses the gap by analyzing arrest data from 2020 and 2021 to examine how interactions between police and different ethnic groups are shaped by broader socio-economic factors such as poverty and unemployment. These factors, which disproportionately affect marginalized communities, may contribute to higher involvement in criminal activities, influencing arrest rates.

The analysis centers on arrest data, which provides an objective and quantifiable view of law enforcement activities, focusing on the correlation between arrests, ethnicity, and age. The data was processed and cleaned using R, followed by an analysis of the distribution of arrests across various age groups to identify potential trends and disparities in Section 2. Age is considered a critical factor in understanding patterns of arrests, as it may reveal systemic biases in policing practices. The findings, discussed in Section 3, offer valuable insights into arrest patterns and their implications.

This study fills a crucial gap in the understanding of racial bias in policing in Toronto by providing an empirical assessment of arrest patterns across ethnic groups. It seeks to contribute to ongoing discussions of police reform by offering data-driven insights that can inform policy changes and promote community engagement. Furthermore, this analysis highlights the importance of transparency and accountability in policing, ensuring that law enforcement actions are equitable and just for all communities. engagement efforts.

2 Data

This data is sourced from the Toronto Open Data Portal [opendatatoronto](#) (Gelfand 2022). The following data is used for an in-depth analysis of arrests and strip searches which involve various ethnic and identity-based groups in the city of Toronto covers the period from 2020 to 2021. The information is gathered by the Toronto Police Service (TPS) (Data 2022) and

is collected under the authority of the Police Services Act. Instrumentation for this process involved the direct input of data by law enforcement officers during or after arrest events, ensuring the capture of relevant variables tied to each arrest.

The raw dataset comprised 32,000 arrest records from 2020 and 2021, with 26 variables available for analysis which reflects real-time operational data collected through TPS’s digital reporting systems, designed to improve transparency and accountability in policing practices. For this study, we focused on specific variables, including Arrest Year, Perceived Race, Age Group, Youth at Arrest, Booked, and the various Search Reasons (Cause Injury, Assist Escape, Possess Weapons, and Possess Evidence). To simplify the analysis, we consolidated these search-related columns into a single “Search Reason” column using the `any()` function (Wickham et al. 2022).

The data underwent cleaning and analysis using the R programming language (R Core Team 2024). Cleaning was performed with the `tidyverse` package (Wickham et al. 2019). Subsequently, analysis was conducted utilizing the `dplyr` package (Wickham et al. 2022), visualization was done using the `ggplot2` package (Wickham 2016), depicting the number of arrests and strip searches across various ethnic and identity-based groups.

2.1 Measurement

This study aims to estimate the extent of racial bias in arrest patterns across different ethnic groups in Toronto for the years 2020 and 2021. The key variables of interest include Perceived Race, Age Group, Youth at Arrest, and Booking Status. The primary estimand is the relationship between perceived race and arrest outcomes, specifically whether individuals from marginalized communities are disproportionately arrested or booked relative to their representation in the population.

The analysis is conducted using the arrest data as the primary estimator. Arrest records are grouped by ethnic identity, age, and booking status to calculate the distribution of arrests across these categories. The study also considers broader socio-economic factors such as poverty and unemployment, which may affect arrest rates, as potential confounders. By quantifying these relationships through statistical analysis and data visualization, the study seeks to determine whether arrest practices reflect systemic bias or are driven by objective factors.

2.2 Total number of arrests by age group for the year 2020 and 2021

It is crucial to analyze the distribution of arrests across different age groups to understand potential patterns and disparities. Table 1 presents a summary of arrest counts for various ethnic groups across all age categories in 2020 and 2021. The age groups included in the dataset are: Under 17, 18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, and 65 years and over.

Given the complexity and extensive nature of the dataset, which involves nine age groups across eight ethnicities, summarizing the entire dataset in a comprehensible manner is challenging. To address this and ensure clarity, we have selected a subset of the data focusing specifically on individuals aged 25 to 34 years. This group was chosen as it represents a significant proportion of the data and allows for a more straightforward comparison of arrest trends across ethnic groups while avoiding information overload. By narrowing the scope, the analysis becomes more manageable and the insights more focused.

Table 1: Arrests Booked by ethnic background for the year 2020, and 2021

Ethnic and Identity Based Groups	Age Group	2020	2021
White	Aged 25 to 34 years	4179	4112
Black	Aged 25 to 34 years	3234	3098
Unknown or Legacy	Aged 25 to 34 years	804	842
East/Southeast Asian	Aged 25 to 34 years	714	497
South Asian	Aged 25 to 34 years	614	512
Middle-Eastern	Aged 25 to 34 years	553	439
Indigenous	Aged 25 to 34 years	338	384
Latino	Aged 25 to 34 years	356	272
None	Aged 25 to 34 years	0	1

2.3 Arrests by Ethnic and Identity-Based Groups for Year 2020 and 2021

Our analysis primarily focuses on the distribution of arrests across different ethnic identities. We extracted data for the following ethnic groups: Black, White, South Asian, East Asian, Middle Eastern, Latin American, Southeast Asian, and Indigenous.

The data for 2020 and 2021 is summarized in Table 2 and Table 3. Table 2 presents the number of individuals arrested from each ethnic group during these years, while Table 3 details how many of those arrested were subsequently booked under criminal charges.

Table 2: Year 2020 vs Year 2021 Total number of arrests

Race	2020	2021
White	14116	13607
Black	8878	8648
Unknown or Legacy	2444	2612
East/Southeast Asian	2361	2054
South Asian	1871	1742
Middle-Eastern	1730	1507
Indigenous	935	999

Table 2: Year 2020 vs Year 2021 Total number of arrests

Race	2020	2021
Latino	960	808

Table 3: Year 2020 vs Year 2021 Total number of Booked

Race	2020	2021
White	7185	7075
Black	4885	4787
Unknown or Legacy	1288	1149
East/Southeast Asian	1037	1057
South Asian	901	876
Middle-Eastern	741	891
Indigenous	583	491
Latino	476	489

It is important to note that the OpenDataToronto (Gelfand 2022) provides population distribution data by ethnicity. However, due to the size and diversity of Toronto and its numerous wards, it has been challenging to cross-verify this data with crime rate statistics in a comprehensive manner. Further, a more detailed analysis could be conducted by comparing specific ethnic groups directly.

Now that the data has been cleaned, we can begin to analyze the data. The following section will provide a summary of the data, and the trends that emerge from the data. Figure 1, Figure 2, and Figure 3, was created using `ggplot` (Wickham 2016), displays the information that can be used as compare the trends as described above.

2.4 Total Arrests in 2020-2021

Referring to Table 2, Table 3 and Figure 1 the data reveals a non-uniform distribution of arrests across ethnic groups, with the Black and White populations showing the highest numbers of arrests in both 2020 and 2021. Additionally, the data indicates a decline in arrests for these two ethnic groups from 2020 to 2021. In contrast, arrests for other ethnic groups, such as South Asian, East Asian, and Indigenous populations, increased over the same period. These opposing trends highlight the divergent patterns of arrest across different ethnic groups during this time.

Referring to Table 1 for the proportion of arrests that resulted in bookings, Figure 2 was plotted with the y-axis representing the percentage of individuals booked after arrest. The data

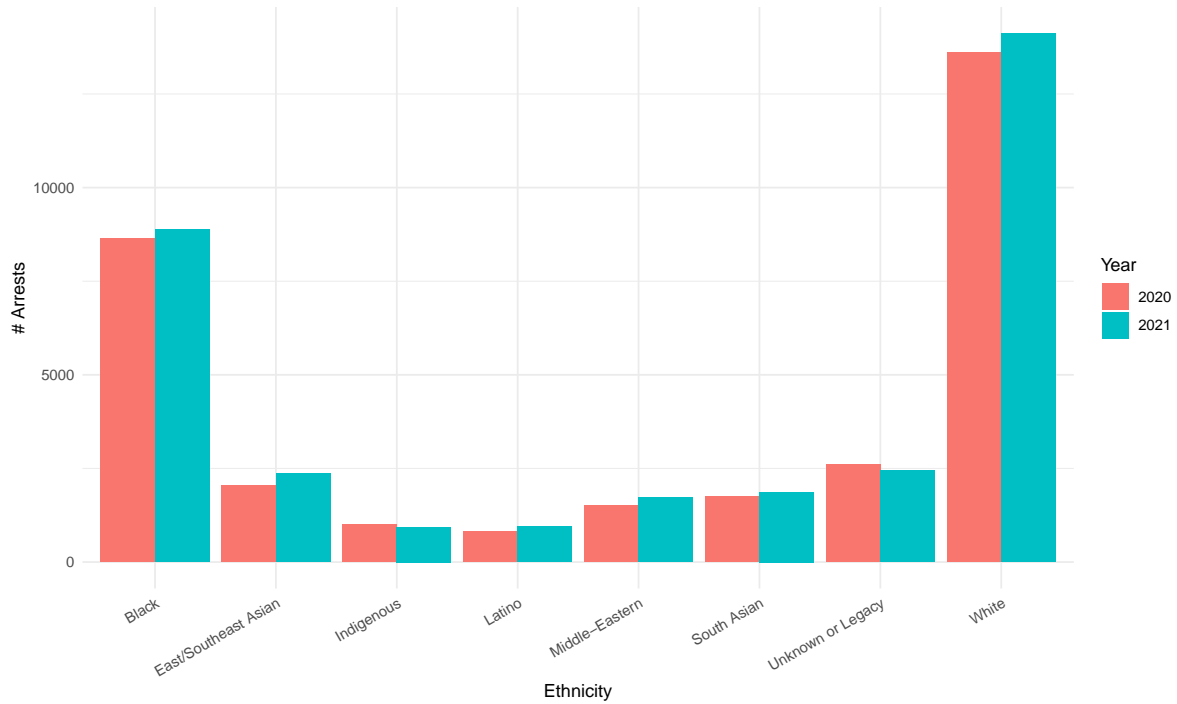


Figure 1: # arrests by Ethnicity in Year 2020 and 2021

demonstrates a consistent distribution across various ethnic groups, with significantly more individuals being booked following arrest than those who were not. Notably, the percentage of individuals booked ranges between 50% and 55% across all ethnic groups, a pattern that remains consistent in 2021.

However, it is worth highlighting that the Black population had the highest number of individuals who were booked after arrest in both 2020 and 2021.

2.5 Number of Arrests by Age Group for the year 2020-2021

An analysis of arrests by age groups is essential to understanding the demographic patterns of law enforcement activity. Figure 3 illustrates the distribution of arrests across different age groups for the years 2020 and 2021. The age groups included in the dataset are Under 17, 18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, and 65 years and over.

The data reveals a non-uniform distribution of arrests across age groups, with 25 to 34 and 35 to 44 age brackets accounting for the highest number of arrests in both 2020 and 2021. However, there is a notable decrease in the number of arrests for these age groups of 2020 to 2021 which indicates that there may be a possible shift in law enforcement activity.

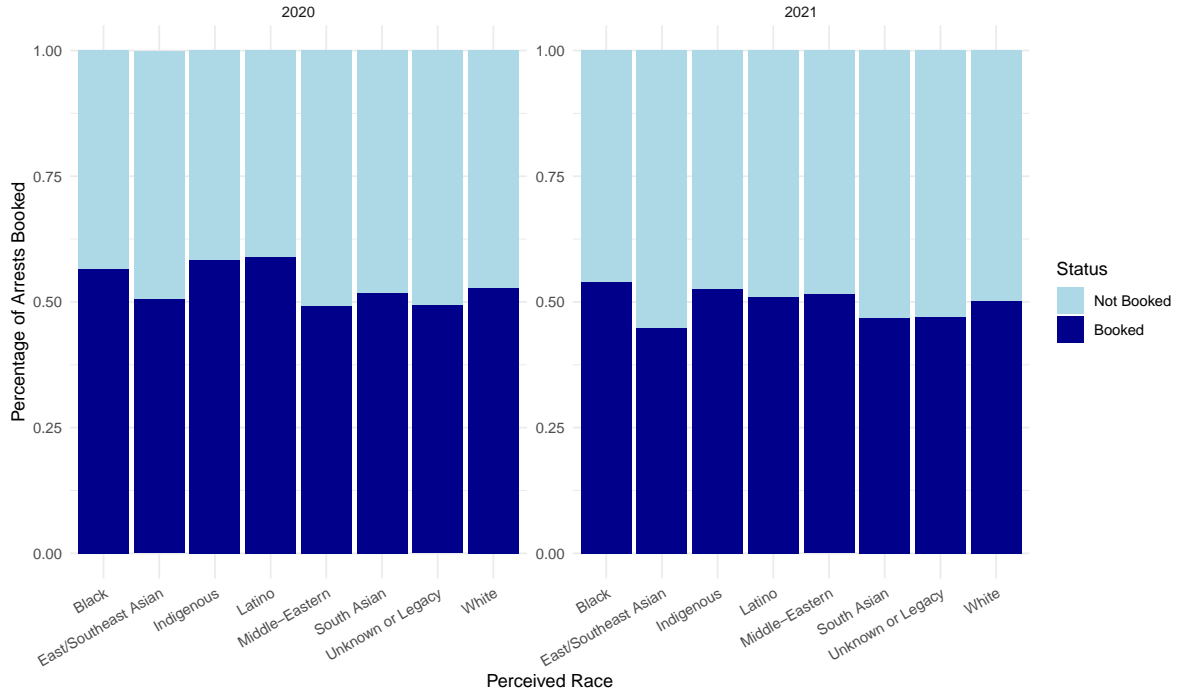


Figure 2: Booked after Arrest in the Year 2020 and 2021

In contrast, the other age groups experienced an increase in arrests during the same period. In particular, the number of arrests for individuals aged 18 to 24 and 45 to 54 saw an upward trend from 2020 to 2021.

3 Discussion

In Section 2.2, the analysis highlights that arrests are not evenly distributed across age groups, with certain age ranges, particularly individuals aged 25 to 44, being more likely to be arrested during the study period. These findings suggest that certain demographic factors related to age may influence interactions with law enforcement. Further exploration is required to understand the socio-economic, behavioral, and legal factors that contribute to the higher arrest rates among individuals in their mid-20s to mid-40s.

In Section 2.3 the data suggests that there are notable disparities in arrest rates across different ethnic groups, with Black and Indigenous populations being of particular concern.

In Section 2.5 the findings suggest that law enforcement practices or external factors influencing crime rates may vary significantly across age groups.

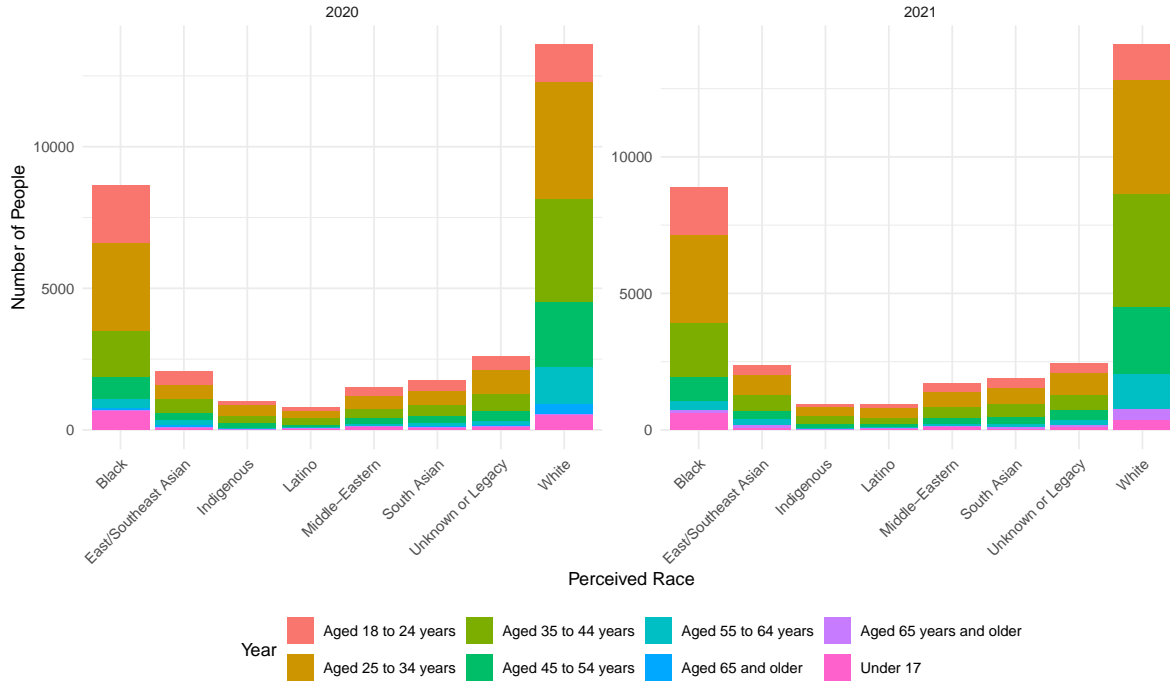


Figure 3: Number of arrests by Age-Group for the year 2020, and 2021

However, none of these findings can definitively point to systemic bias without further context, particularly around socio-economic factors and community-specific conditions.

While this paper offers several important insights, it is important to acknowledge its limitations. A more detailed comparison of specific ethnic groups and further research are necessary for a deeper understanding of the data. Law enforcement agencies and community advocates can use these findings to develop targeted interventions and enhance community engagement efforts. Additionally, policymakers may consider using this analysis to inform necessary policy adjustments. The need for transparency and accountability in policing remains critical, as emphasized by the Toronto Police Services Board (Toronto Police Services Board 2019).

In conclusion, the analysis of Toronto's arrest data suggests that, despite concerns about racial bias, arrests appear to be influenced by objective factors rather than racial profiling. The reforms implemented by the Toronto Police Service have contributed to increased transparency and fairness within the police force. Although disparities may exist due to socio-economic factors, the data does not indicate systemic bias.

References

- Data, Toronto Open. 2022. “POLICE RACE AND IDENTITY BASED DATA - ARRESTS AND STRIP SEARCHES.” <https://open.toronto.ca/dataset/police-race-and-identity-based-data-collection-arrests-strip-searches/>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Police Services Board. 2019. “Toronto Police Services Board’s Race-Based Data Collection Policy.” Report. Toronto Police Services Board. <https://www.tpsb.ca/policies-by-laws/board-policies/177race-based-data-collection-analysis-and-public-reporting>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.