Darious Brown

DSC630

Project Sub 3

01Jun25

The chosen dataset is stems from the Office of Juvenile Justice and Delinquency Prevention. This government agency hosts national data of crime convictions. The data is available to the public as either of the following file types: Csv, Excel, PDF, or for print. The data is unique because of the sensitivity, but it’s public data it entails for a topic that is often stereotyped. The data will entail of accurate information that will either support or refuse such stereotypes. Its relevance will shine light on the disparities within the judicial system and how societal views may differ from actual results. The intent of this project is a call for awareness and change. It is very important to understand accurate information with so many potential risks in the world regardless of bias.

The dataset will host information comparing statistics by offense in comparison to. race (White, Black, American Indian, Asian). It takes into consideration of all ages, men and women within the 2020 year. Within the dataset crime is broken into two categories. First, violent crime index, which entitles of crimes such as: murder, rape, nonnegligent manslaughters, robbery and aggravated assault. The second category is property crime index offenses which includes: burglary. Larceny-theft, motor vehicle theft, and arson. The data will suggest disproportionate arrest patterns by comparing statistics of different offenses. The insight will allow for better understanding of first which violent crimes are more popular than others per race. This will also influence the thought of which racial groups who are assumed to be more violent than others, by proving societal norms wrong. This information is primal to approval or refusing social norms and prying disparities from judicial systems, in hopes of change and overall better fairness.

It crucial to understand today’s society. Propaganda and false statements fill ears and eyes will inaccurate information daily. Some racial groups are assumed to commit more specific criminal offenses than others. One of these stereotypes is that black people are more involves into violent offenses than other groups. Also, society also portrays American Indians to be the primary racial group involved in substance abuse related crimes. And that Asians are rarely the offender in a criminal situation which can lead to under representation of what true values there are. How do you think society reacts when they are potentially portrayed to be someone or something they are not? Many misconceptions are available for the taking in today’s world, but regardless crime is inevitable. Social norms tell us the notion that Asian communities are essentially crime free and black communities are crime striction.

The dataset appears to be a smaller dataset focusing only on the 2020 crime year and is spread over 30 different offenses. A unique property of the dataset is that the overall crime offenses committed per racial group is broken down the into each individual offense type. At a quick glance, the dataset infers which racial groups overall excel in particular crime offenses. This gives the ability to create a variety of visuals. Scatterplots built of specific crimes, would allow for a representation of the number of offenses per crime. A scatterplot will present the independent and dependent variables, relationship between each offense and racial group. This will allow for a comparison that will either support or refuse social norms and stereotypes within communities around the world. It is important to understand the significance of what the dataset provides and the importance of a outlier in terms of offenses committed.

Moreover, histograms will also be used to visualize information and better fit the data and research. Histograms will support a ranking system of which racial group is deemed more violent by statistics. This will provide a simplistic approach, that at a quick glance will shed light on the disparities in the juridical system. the information presented should spark community interest and governmental support. This will open a door for potential future programs that will promote safer spaces and rime stopping programs. Promoting advocacy can also insight on which communities need the most resources.

The results of the analysis will be compared by offense and racial group. The information will go to support or disagree with social norms and stereotypes with statical data with the intent to inform the society, policy makers, social movements and more. The results should affect personal ignorance, influence advocacy, and governmental assistance. The data will be telling of very impactful information in aims to expose and disapprove social norms, and by this very specific information will be answered. Such as, are Black people the actual leaders of violet crime? Are Asian people as innocent as perceived to be or how involved are Indians into substance abuse crimes? The ability of the data is that data will shed light on what which racial group leads in each induvial crime conviction. When comparing the racial groups, it can be identified which group is the overall more violent group.

To accurately depict the data points a histogram will make interpretation which crime and which group had the most offenders. A simplistic back color scheme would be vital as each racial group would need to be represented by their own color to ensure it pops from the background and making it easily readable. Secondly, a scatterplot will effectively show the data points as it will easily be comparable to each other. Because scatterplots are on a rising scale, it will easily identify the trend which racial group lands higher on the scale then others in each crime conviction. Pie charts and network graphs will also be used to compare what percentage of a crime is claimed by each racial group.

Overall, the dataset will effectually attack social norms and shine light on injustice within the judicial system. Although, it may become important later to add data from other years, enabling a comparison from different decades. This would answer how consistent the data is and if specific groups have a history of committing crimes. By making such prominent information feasible many naysayers’ ignorance will be addresses directly. Future research it would be significant to further invest time into comparing crime ratio and understanding that there is only one majority group. It may also shed light onto where social norms and injustice stem from.

Due to the nature of the topic the ethical considerations are of the upmost importance, but the delivery will be direct. The intent is to attack social ignorance through facts regardless anyone personal emotions. The intention is not to put a spotlight on any group, but to effectively advocate for misrepresented groups in aims to take change into policy and injustice in the Judicial system.

A graph of a bar graph

Description automatically generated with medium confidence

This visualization shows the total number of drug abuse arrests broken down by race for the year 2020. White individuals accounted for most of these arrests, with approximately 837,000, followed by Black individuals with roughly 281,000 arrests. Arrests for American Indians and Asians were substantially lower, at around 20,000 and 15,000 respectively. The combined arrests across all racial groups totaled nearly 1.2 million. This considerable volume reveals how drug-related offenses dominate law enforcement activities. Notably, Black individuals are overrepresented in these arrests when compared to their national population share of approximately 13%. In contrast, although White individuals had the highest raw numbers, their arrest rate may underrepresent their population share. The disparity seen here suggests potential racial bias or targeting in drug enforcement practices. While the chart reflects only a single offense category, it effectively illustrates racial imbalances. The data underscores systemic trends in drug policing and invites further exploration of contributing policies or socioeconomic factors.

A graph of a person with a white background

Description automatically generated with medium confidence

This area chart provides a comparative view of arrest counts across all offenses and racial groups. White individuals dominated in overall arrests, especially in high-frequency crime categories such as larceny-theft and fraud. Black individuals consistently ranked second in each category, while American Indians and Asians had significantly lower arrest numbers across the board. The “All offenses” category alone represented over 7.5 million arrests, with the racial hierarchy evident across each subcategory. Property-related crimes revealed broad racial participation, while certain categories like curfew and loitering showed disproportionate targeting of Black individuals. The shape of each racial area mirrored the others, suggesting structural patterns in law enforcement rather than random variance. Notably, Black individuals appear overrepresented in several offense types relative to their national demographic percentage. Asians and American Indians, while underrepresented numerically, may still be affected by localized policing strategies. The visualization makes evident that arrest disparities span across crime types.

A graph of a person with a number of fingers

Description automatically generated with medium confidence

The stacked area chart focuses on murder and nonnegligent manslaughter offenses, totaling just over 12,000 arrests. Black individuals made up the largest share with approximately 6,500 arrests, while White individuals followed with around 5,700. Arrest counts for American Indians and Asians remained below 500 each. Though the overall numbers are low compared to other offenses, the disparity is statistically and socially significant. The data reveals that Black individuals are considerably overrepresented in murder arrests compared to their population size. Conversely, White individuals, despite being the most populous group, had fewer murder-related arrests than expected. The visualization starkly highlights the racial distribution of the most severe criminal offenses. This disproportion may result from systemic factors such as over-policing in predominantly Black neighborhoods or socioeconomic disparities. The narrow focus allows for clear interpretation without the noise of multiple variables.

A graph of a bar graph

Description automatically generated with medium confidence

This histogram illustrates the number of aggravated assault arrests by race. White individuals had the highest number of arrests, exceeding 230,000, while Black individuals followed with approximately 127,000. American Indian and Asian groups had far fewer arrests, totaling under 15,000 combined. The gap between White and Black arrests stands at around 100,000. However, when adjusted for population, Black individuals remain overrepresented. The chart shows that most arrests for aggravated assault are concentrated in just two racial groups. The high frequency of these arrests supports the narrative of aggressive policing of violent offenses. Although raw numbers place White individuals at the top, proportional analysis points to more significant enforcement pressure on Black individuals. The histogram allows for immediate visual comparison, making disparities easily discernible. It also suggests aggravated assault is among the most heavily patrolled and reported offenses.

A screen shot of a graph

Description automatically generated

This scatter plot compares total arrests for simple and aggravated assault across four racial groups. White individuals had the highest totals in both categories, with approximately 580,000 simple assault arrests and 235,000 aggravated assault arrests. Black individuals followed with around 275,000 simple and 127,000 aggravated assault arrests. American Indian and Asian individuals had markedly fewer arrests, each under 50,000 across both categories. The spread of points shows a linear correlation between simple and aggravated assaults, particularly for the top two racial groups. This suggests that arrest patterns are not random but may be influenced by racial profiling or enforcement bias. The closeness of the American Indian and Asian data points underscores their statistical marginalization in this dataset. The plot implies that prosecutorial discretion or police judgment may shift arrests between simple and aggravated charges. The chart effectively visualizes dual-outcome disparities that correlate with race.

A pie chart with text

Description automatically generated

This donut chart presents the racial breakdown of simple assault arrests. White individuals represent the largest proportion, accounting for **65.2%** of all simple assault arrests. Black individuals follow with a significant **30.6%,** showing notable overrepresentation when compared to their share of the U.S. population. American Indian and Asian groups make up **2.4%** and **1.8%,** respectively, together totaling less than 5% of these arrests. The chart’s circular format enhances the contrast in racial representation and allows for easy interpretation of the relative percentages. While the White arrest rate appears high in raw percentage, it is somewhat aligned with population statistics. However, the Black arrest rate remains elevated, suggesting potential bias in enforcement or reporting practices. Asian and American Indian figures, though low, highlight the need for deeper localized data to interpret their true significance. The visualization reinforces broader patterns found in other assault-related charts, where arrests disproportionately affect Black individuals. Overall, this chart effectively illustrates racial disparity within a specific offense category.

Overall, the study was able to reveal which crime conviction each racial group lead in. African Americans were observed to be more involved in violent offenses than other groups accounting for 26.1% of convictions. These disapprovals social norms as poverty and not race is the strongest predictor as many crimes committed are intra-racial. In addition, 69.9% of crimes against other racial groups were subject to Caucasian people. This also disapproves social norms as well due to 80% of crimes being intra-racial when considering population size. American Indians were composed of only 2.4% of substance abuse related crimes and American Asians comprised of 1.6% of all crimes disapproving social norms. Lastly, this disapproves social norms as wel. In both cultures they have both consistently have one of the lowest arrest rates possibly connected to cultural emphasis on lawfulness and family honor. It is important to understand that the data is focused on the 2020 crime year, to create a bro predictability a SAMIRA model was engineered.

A graph with different colored lines

Description automatically generated

This line forecast chart displays the projected number of arrests by race from 2010 to 2030, based on a SARIMA model using simulated historical trends. White individuals show a steady and substantial increase in arrests, climbing from just over **1 million in 2010** to a forecasted **1.7 million by 2030.** Black individuals also experience a gradual rise, starting around **400,000 arrests in 2010** and approaching **500,000 by 2030**. Meanwhile, American Indian and Asian arrest numbers remain relatively stable, both projected to stay below**100,000 annually.** The widening gap between White and Black arrest lines over time suggests persistent disparities with no indication of convergence. These forecasted trends imply that without systemic change, racial disproportionality in arrests will likely endure. The flatness of the Asian and American Indian trends may mask underreporting or smaller population bases, but they also highlight which groups face heightened law enforcement focus. The model reflects real-world concerns about predictive policing reinforcing historical inequities. Overall, this visualization emphasizes the long-term consequences of current patterns in racialized law enforcement.

For future research a linear regression a random Forrest model as engineered. The liner regression method was to order the offense numerically. This analyzed a RMSE value of 2,882,332 arrest that on average predictions deviated by -2.88 Million and a R^2 score of -0.0831 disapproving the model as it performed worse than simply predicting the mean it indicated poor model fit. Essentially this meant this was a **large error**, suggesting the model lacks predictive accuracy — likely due to having only one feature (offense code). The R score model explains **none of the variability** in arrest counts and is performing worse than a naive model that predicts the average for every case. When focusing on the random Forrest model were recovered a RMSE score of 2,817,198 which was a similar large error as in the linear regression model showing of a non-linearly relationship. Also a R^2 score of -0.0347 was also recovered which was slightly better preforming then the linear regression model but still showed a week model that could not clarify variance and arrest.

To ensure that the visuals were reliable for visualization, several important steps were taken to prepare the data set. First, the data set was used from the Office of Juvenile Justice and Delinquency Prevention database that is backed by the U.S Department of Justice. Within the code a csv file was used titled “Arrests by Offense, Age, and Race – 2020” from the Office of Juvenile Justice and Delinquency Prevention. Some of the necessary cleaning’s steps of the data involved removing commas and converting numeric values to integers. In addition to dropping rows with missing and incomplete data (e.g. rape column). Then aggregated race-based arrest counts across offense categories to return as percentages. The dataset was crime was broken down categorically: violent crimes, drug-related crimes, and property crimes for clarity, in addition to deriving percentages from overall total of all arrest.

Through both storytelling and statistical evaluation several raw truths were unfolded. White people make up 70% of all overall arrest, reflecting of population dominance but also highlighting blind spots within perception. Black people are responsible for 26% of total overall arrest, which is disproportionate being as through black people account for only 13% of the population. American Indians (2.4%) and Asian American (1.6%) were relatively irrelevant in crime convictions but still were subject to common stereotypes.

So, what do we do with this information? we call for policy change and installation of crime combating programs. Models trained show that data pf 2020 is not promising as if other years will perform consistently but rather our SAMIRA model shows that overtime arrest will continue rise. This educating using the data to challenge social norms and using prediction models to understand how society influences compare to actual convictions. We most push for policy reforms by advocating justice systems suitable for policing and transparency for the future. We must also create a safe communicable space as crime affects us all one way or another. Empowering and funding organizations and community projects would be pivotal for future research as data may be altered with newly installed community recreations.

In conclusion many assumptions about racial crime roles in society are misleading and false and are most likely stemmed from systemic force rather than reality. Arrest disparities shine light on complex issues within systemic forces. Some of these factors that fuel these stereotypes are stemmed from poverty, politics, social bias and generational ignorance. This information as the ability to shake the grounds of already divided communities. The significance of. The raw data is to promote awareness and combat societal stereotypes. Secondly, to advocate for fair policing and support underrepresented communities.

In contrast the plain ignorance can be a black sheep. The will to know better must be a want for any societal movement, thus information will be presented as a need for change. Thy risk lies with the support of larger entities and an unforeseen time of society’s ability to adapt. If the risk outweighs the benefits than we shall alternatively direct our results to an open forum

Citations

1. . Hinton, Elizabeth, et al. An Unjust Burden: The Disparate Treatment of Black Americans in the Criminal Justice System. Vera Institute of Justice, 2018, [www.vera.org/downloads/publications/for-the-record-unjust-burden-racial-disparities.pdf](http://www.vera.org/downloads/publications/for-the-record-unjust-burden-racial-disparities.pdf).
2. Bureau of Justice Statistics. Criminal Victimization, 2020. U.S. Department of Justice, Office of 2. 3. Justice Programs, 2021, [www.bjs.gov/content/pub/pdf/cv20.pdf](http://www.bjs.gov/content/pub/pdf/cv20.pdf).
3. Substance Abuse and Mental Health Services Administration (SAMHSA). National Survey on Drug Use and Health: 2020 Detailed Tables. U.S. Department of Health and Human Services, 2021, [www.samhsa.gov/data/report/2020-nsduh-detailed-tables](http://www.samhsa.gov/data/report/2020-nsduh-detailed-tables).
4. Pew Research Center. Asian Americans: Diverse Outcomes, Few Arrests. 2022, [www.pewresearch.org/fact-tank/2022/05/10/asian-americans-diverse-outcomes/](http://www.pewresearch.org/fact-tank/2022/05/10/asian-americans-diverse-outcomes/).
5. Federal Bureau of Investigation. Crime in the United States, 2022: Arrests by Race. U.S. Department of Justice, 2023, <https://ucr.fbi.gov>.
6. Le, C. N. Asian-Nation: The Landscape of Asian America. 2002, www.asian-nation.org.
7. Bureau of Justice Statistics. Race and Hispanic Origin of Victims and Offenders, 2012–2015. U.S. Department of Justice, 2017, [www.bjs.gov/content/pub/pdf/rhovo1215.pdf](http://www.bjs.gov/content/pub/pdf/rhovo1215.pdf).
8. Pew Research Center. Race and Crime in America. 2021, [www.pewresearch.org](http://www.pewresearch.org)