Jason Sheu - A15536122

DSC 180

A01 - Aaron Fraenkel

3 December 2021

Is Policing in San Diego Fair?

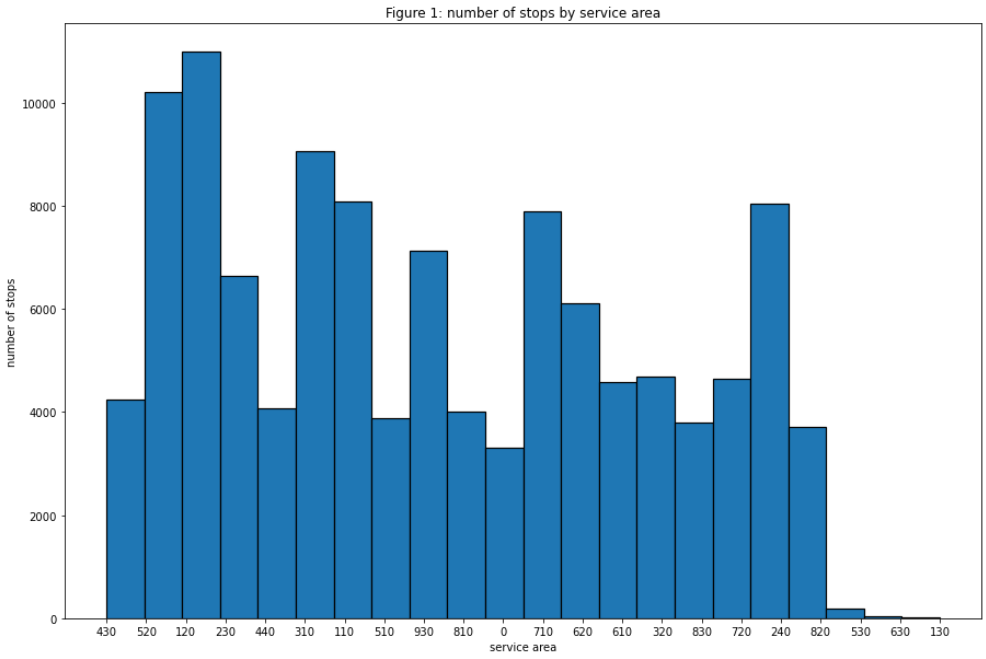
Introduction

Racism in America is a topic of widespread concern, attention, and research. In recent years, a topic of discussion and debate has been if there is any implicit bias or profiling occurring within police traffic stops. In America, police pull over 50,000 drivers a day, and over 20 million drivers per year, making it one of the primary ways that the average citizen interacts with their local police department. (Stanford Open Policing Project) In a recent Gallup poll done in 2020, a majority of Black Americans (59%) feel that their recent interactions with police have been positive, while the remainder (49%) of Black Americans feel that they have not had a positive experience. While this may seem good, 75% of adults in the United States have had positive experiences with police, compared to 25% of adults who feel they did not have a positive experience (Gallup 2020). In the same poll, it was also discovered that, among whites, hispanics, and blacks, the percentage of Black Americans that feel they were treated with respect and fairness was the lowest of those racial groups (Gallup 2020). While these results may have been influenced by the social unrest and nationwide protests related to several police brutality incidents in 2020, it points to a concerning trend that blacks, and to a lesser degree hispanics, are being profiled in their interactions with police. In this report, I will be exploring a dataset of 115422 traffic stops in San Diego from 2015 to determine if there is indeed a racial disparity in

police stops, specifically by doing several statistical tests, comparing post-stop outcomes, and by using a statistical technique called the Veil of Darkness.

Exploratory Data Analysis

The dataset I am using is found from the San Diego police dataset website, and has 115422 rows of traffic stops from 2015 recorded. To help with exploration, I am also using United States Census data to perform spatial joins using geopandas. In the stops data, there are 15 columns of information: the id of the stop, the cause of the stop, the police service area, the subject's race, sex, age, the time of the stop, date of the stop, if the subject is a resident, searched, arrested, consented to a search, if contraband waws found, and if property was seized in the search. In my analysis, I will be focusing on the time of the stop, the arrested/searched columns, the service area, and the subject's race. In the exploratory data analysis of this dataset, I first cleaned the dataset to make it easier to use and understand. First off, I joined the race code with a table of corresponding names to make it easier to read. Then, I replaced all instances of the characters 'Y', 'N', 'n', 'y', 'b', 'M' and blank characters with 0 in the 'arrested,' and 'searched' columns, to make calculations of averages and percentages easy. I then replaced any missing values in those columns with 0s in the case I missed any. To start off my exploratory data analysis, I plotted some



histograms to visualize my data, figures 1 and 2, which show the number of stops in each police service area, and the number of stops by race. From these histograms, we can see that some areas had much more stops than others, such as 520, 120, and 310, which are the service areas located towards the city. We can also see that when we look at number of stops by race, White drivers easily make up the majority of the stops in the dataset, with hispanics being the second most. After this, I first calculated a rough estimate of the overall stop rate, using the total population of San Diego County from 2019, which was the closest year the US census had data available. With the estimated population being 1,423,851, I divided the length of the dataset by the population to get an 8.1% stop rate overall. I also calculated the overall arrest and search rate by dividing the total number of

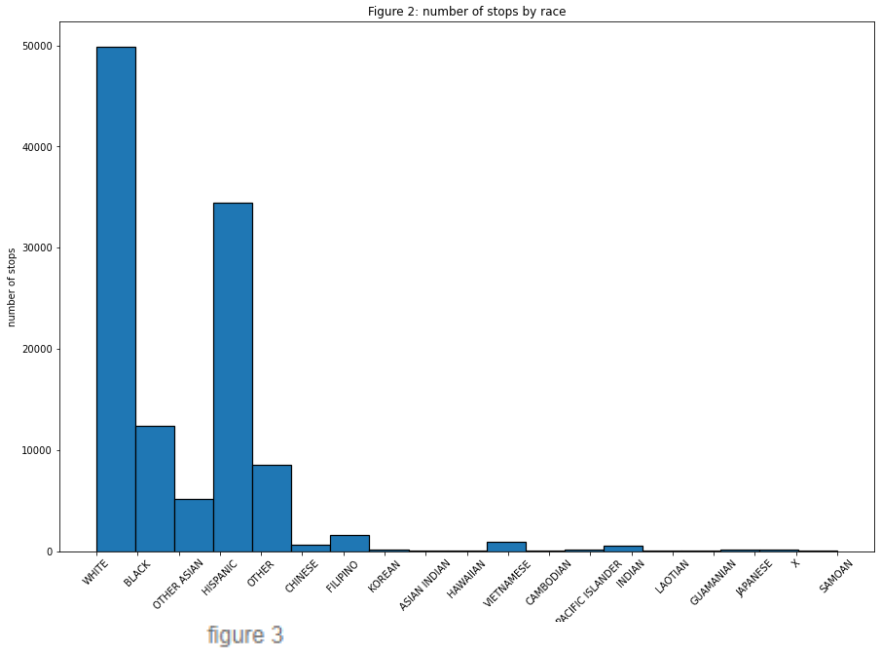


figure 3

subject_race	search_rate	stop_rate	arrest_rate
ASIAN INDIAN	2.727273	0.007726	0.909091
BLACK	7.985803	0.870667	1.742357
CAMBODIAN	16.071429	0.003933	5.357143
CHINESE	1.391036	0.045440	0.618238
FILIPINO	2.492212	0.112722	0.623053
GUAMANIAN	5.405405	0.002599	5.405405
HAWAIIAN	16.666667	0.001264	5.555556
HISPANIC	5.002177	2.419144	1.306431
INDIAN	2.033898	0.041437	1.186441
JAPANESE	1.503759	0.009341	1.503759
KOREAN	6.214689	0.012431	4.519774
LAOTIAN	20.634921	0.004425	1.587302
OTHER	1.608170	0.598307	0.504754
OTHER ASIAN	1.650672	0.365909	0.575816
PACIFIC ISLANDER	7.438017	0.008498	0.826446
SAMOAN	1.515152	0.004635	0.000000
VIETNAMESE	4.535865	0.066580	1.265823
WHITE	2.570117	3.503246	1.112648
X	0.000000	0.010324	0.000000

each by the length of the dataset, getting a 1.17% arrest rate and 3.79% search rate. To follow up on these basic statistics, I grouped the stops by race to see what the stop search and arrest rates are like per race, which are shown in figure 3 to the right. An interesting observation can be drawn from this: It looks that on average, Blacks are searched almost 8% of the time when they are stopped, which is a stark increase from the 2.57% that whites are searched. Furthermore, we can see that even though there is a large disparity between search rates in these two groups, the arrest rate is almost the same, with blacks being arrested in 1.74% of stops, and whites being arrested in 1.11% of stops. It is also of note that Cambodians, Hawaiians, and Laotians are searched quite often, with double digit search rates. However, this can be attributed to the small number of stops being composed of members of these racial groups. In the next phase in my exploratory data analysis, I grouped the stops by service area and joined the San Diego police department service area shapefiles to get the search, stop and arrest rates by each area. The results can be seen in the 3 figures below:

Figure 4: Search Rate by Area

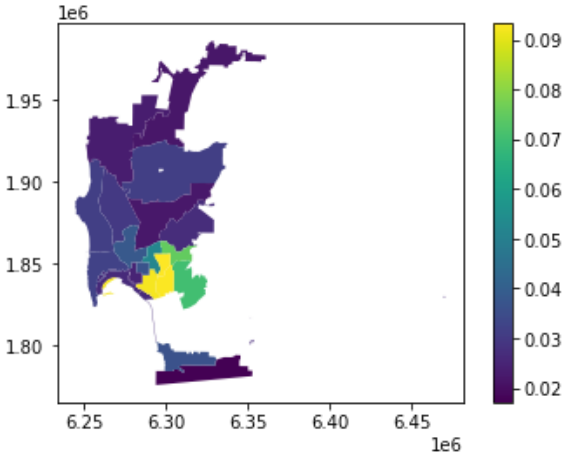


Figure 5: Arrest Rate by Area

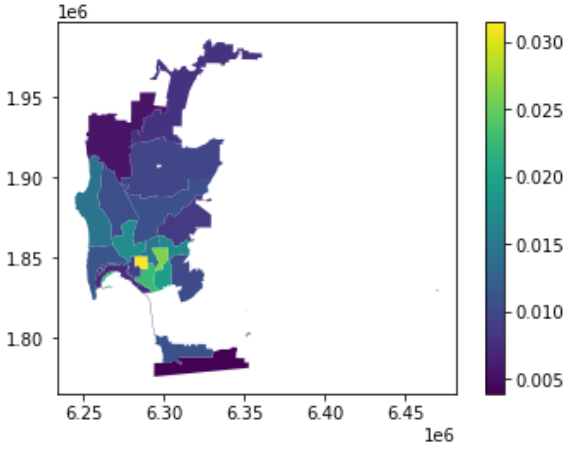
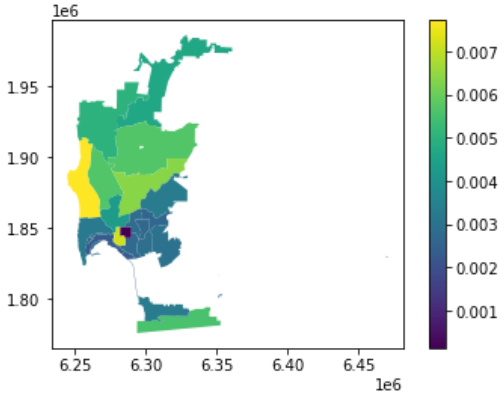


Figure 6: Stop Rate by Area



It seems that arrests and searches increase towards the downtown areas of San Diego, and would decrease as it heads into the more suburban areas, such as La Jolla. This is probably due to downtown having a larger police presence, which results in having more stops, which raises the number of arrests and searches as well. Another reason that could be causing an increase in searches and arrests in the downtown area is the demographics of the area. Generally speaking, large urban areas, such as downtown San Diego, have a higher proportion of blacks, hispanics, and other minorities than the suburbs do, and if police are indeed racially profiling in their traffic stops, this would be reflected in high search, stop and arrest rates in areas with higher proportions of blacks, hispanics, and minorities.

Statistical Analysis

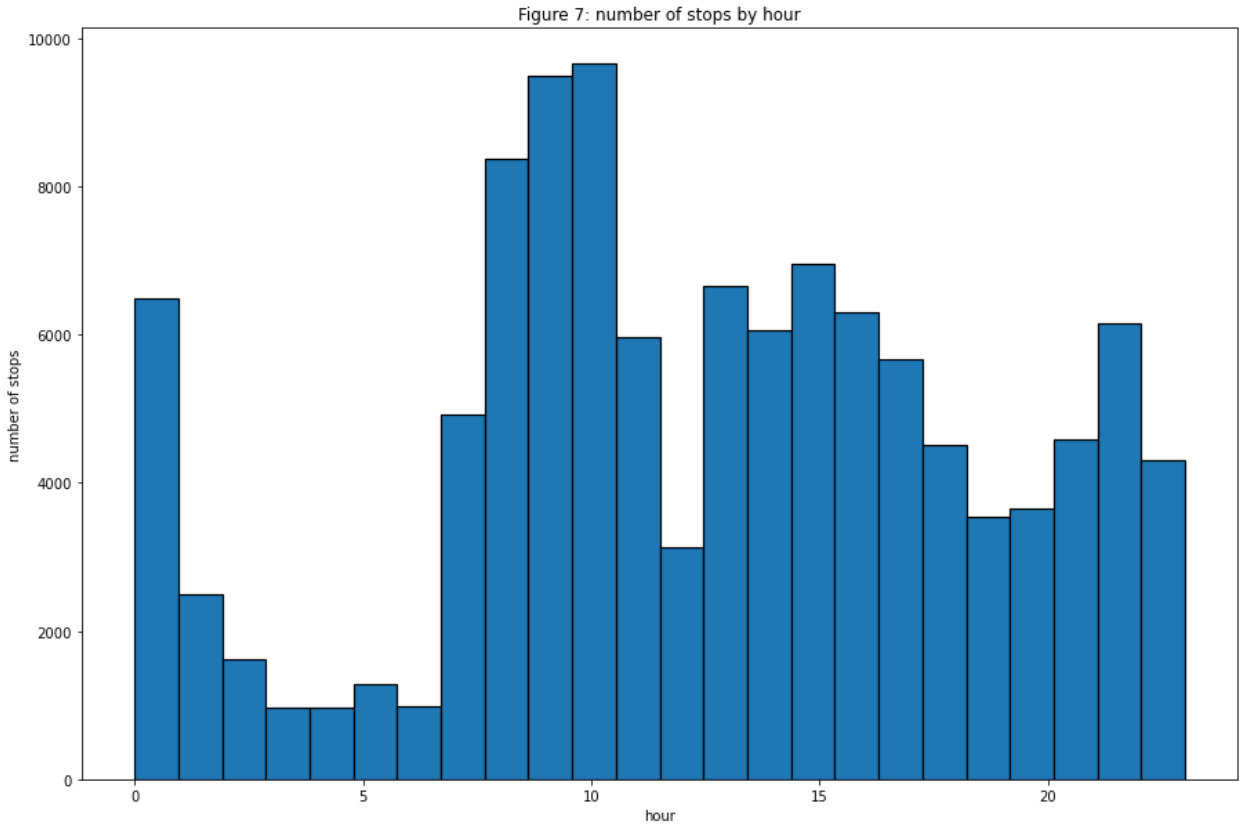
Upon seeing the disparities between the search rates of blacks and whites in San Diego, I suspected that it was due to racial profiling done by the San Diego police department. To test this, I performed several t tests on the search rates, grouped by race. The first t test was done on all of the stops where the race was black and white, and compared the number of searches. The result of this first test was a p value of 1.11x10^99, which is almost zero. However, I realized that there were many many more stops with a white driver than a black driver. In the 2015 stops dataset, there are 49881 stops with white drivers, and 12397 stops with black drivers, which is a 400% increase. To rectify this, I randomly sampled 10000 stops with black drivers, and 10000 stops with white drivers to do a second t test. The p value from this t test was 1.639x10^-66, which is also pretty close to 0. Based on these results, I can conclude that there is definitely some sort of racial bias or profiling in the search rate involving black Americans in San Diego. When conducting the same t-test on the number of arrests, I got a p-value of 0.00018, which also leads me to reject the null hypothesis that there is no racial bias in arrests involving black Americans.

The reason that the p-value for searches is so low is caused by police doing what is commonly called “fishing” for evidence. Because some police officers are implicitly biased towards black people having a higher chance of having drugs or other contraband on them, they are more likely to search a pulled over black driver to look for any evidence of contraband. This bias is being represented in the data, and we can see that black drivers do have a much higher search rate than white drivers do, despite having similar arrest rates. The similar arrest rates means that police find a reason to arrest the driver at roughly the same rate, despite searching black drivers eight times more than white drivers. Due to the incredibly small p-values that the t-tests I conducted produced, I used a method called the “veil of darkness” to verify that there is indeed racial profiling involved in the traffic stops being conducted in San Diego.

The Veil of Darkness

The Veil of Darkness test is an idea conceived by Jeffery Grogger and Greg Ridgeway in 2006, to test for racial profiling. The idea is quite simple: at night time, police are less likely to racially profile because they have a hard time seeing the color of someone’s skin due to the darkness. However, the validity of this could be put into question, since the groups of people who drive at night are not necessarily the same groups that drive in the daytime. To counteract this, the variation in the hours of darkness throughout the year is used. The inter-twilight period is the range of time starting from the earliest dusk time in the year and ending at the latest dusk time in the year. By carrying out analysis in this period, we are able to minimize the variance that can be caused by different groups of people travelling at different times (Stanford Open Policing Project). By using the astral and datetime python libraries, I was able to filter my dataset to only include traffic stops from between 5:09 PM and 8:29PM, which is the inter-twilight period of 2015 in the San Diego area. I also graphed the number of stops at each point of the day to

visualize the data, which can be found in figure 7.



To start off my analysis with the veil of darkness in play, I conducted the same t-test I did above on the veil of darkness stops, and found that for searches, there was a p-value of 0.00058, meaning that police are still racially profiling black drivers with searches. The p-value for the t-test comparing arrests was 0.327, meaning that police officers are not racially profiling in terms of arrests. Even though it is dark out, if the police officer is racially profiling the people they stop, they will see the race of the driver when they walk up to gather information, leading them to be more likely to search the car if they are biased that way towards black drivers. The p-value being larger than the prior t-test done on the whole dataset is indicative of the probability of being profiled being smaller as well, which lines up with the hypothesis posed by the veil of darkness test. To see if black drivers were being stopped less, I calculated the proportions of

driver races at different time periods: darkness (8pm-5am), morning rush hour (6am-9am), daytime (10am-2pm), and evening rush hour(3pm-7pm), and found something quite interesting. The darkness time period had the highest proportion of black drivers, at 13.39%, while the daytime and morning rush hour had 8-9% of drivers being black. Evening rush hour was also more than morning rush hour, at 11.2% of drivers being black. Even though this goes against the hypothesis stated by the veil of darkness test, a cause for this could be that black drivers are more likely to drive at night than during the day. This could be due to many black commuters working double shifts, or night shift jobs that require them to be out later than other commuters. To verify this, the veil of darkness test could be used in a large context, maybe across the whole state of California or across the country to see if the results are different than San Diego.

Conclusions

In this report, I explored the 2015 San Diego traffic stops dataset to determine if there is racial profiling occuring. In my exploratory data analysis, I found that blacks were being searched at much higher rates when compared to other racial groups, leading me to believe that some form of profiling was occuring. I tested my hypotheses with a t-test on 10,000 stops involving blacks and 10,000 stops involving whites. As a result, I was able to determine that blacks are being searched and arrested at disproportionately high rates, due to the officer profiling black drivers as being higher risk for carrying contraband. I followed this up with the veil of darkness test, which is a tool that helps determine if blacks are being stopped at proportionally high rates. However, I found that even during the inter-twilight period, blacks were being stopped at higher rates in night time hours when compared to daylight hours, leading me to believe that the amount of blacks commuters on the road are higher at night, potentially due to them working double or night shifts. In conclusion, I found evidence supporting racial

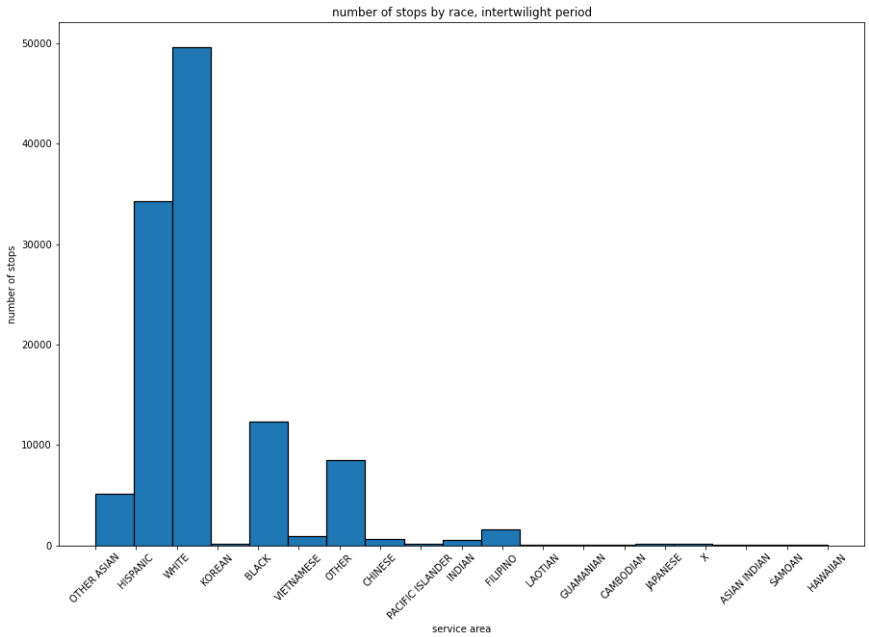
profiling in black arrests and searches, but was unable to prove any discrimination occuring in the amount of stops. In further research, I aim to expand the veil of darkness to include a wider area, as well as work on a fair and bias correcting predictive policing algorithm.

Appendix

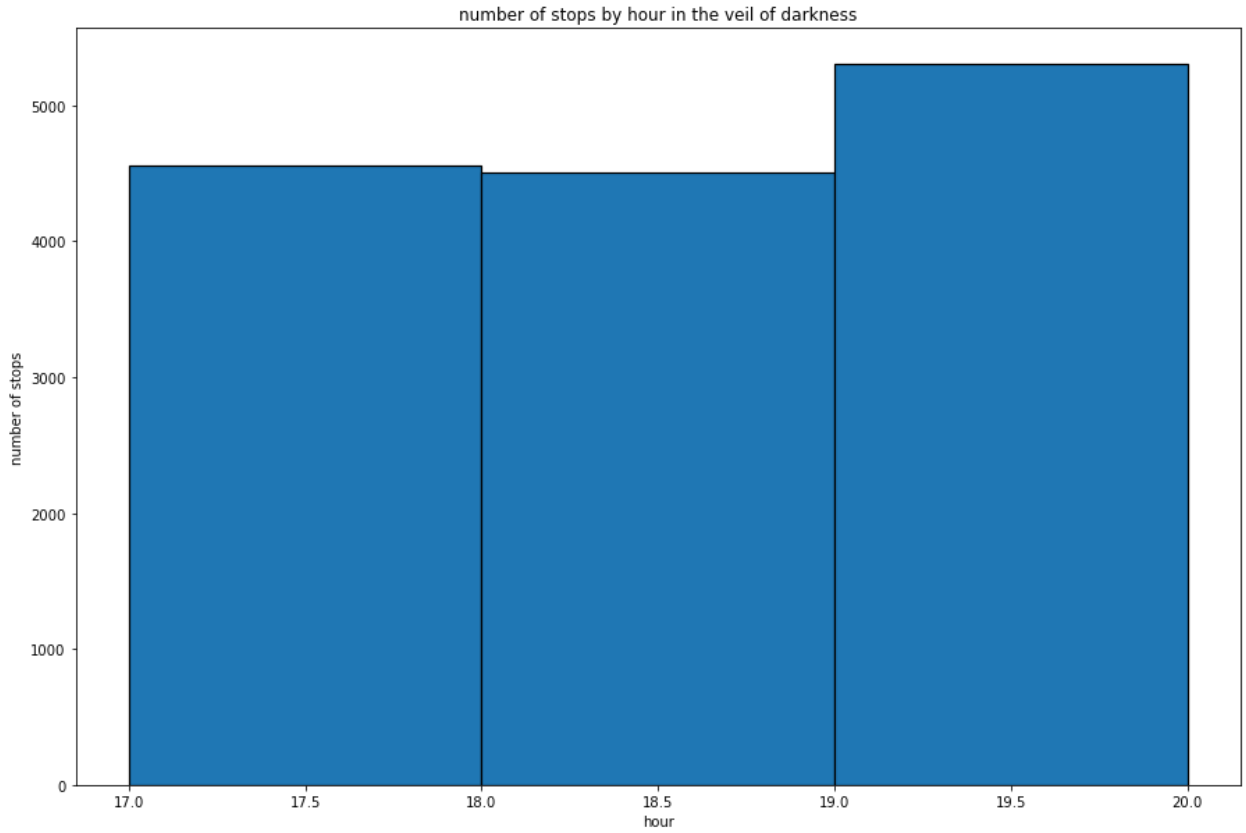
Breakdown of Stops by Race:

```
WHITE                49881
HISPANIC             34445
BLACK                12397
OTHER                 8519
OTHER ASIAN           5210
FILIPINO              1605
VIETNAMESE             948
CHINESE                647
INDIAN                 590
KOREAN                 177
X                      147
JAPANESE               133
PACIFIC ISLANDER       121
ASIAN INDIAN           110
SAMOAN                  66
LAOTIAN                 63
CAMBODIAN               56
GUAMANIAN               37
HAWAIIAN                18
Name: subject_race, dtype: int64
```

Number of stops by race during the inter twilight period



Number of stops by hour in the veil of darkness intertwilight period



References

Chanin, Joshua, et al. *Traffic Enforcement through the Lens of Race: A Sequential Analysis of Post-Stop Outcomes in San Diego, California*. 2018, https://spa.sdsu.edu/documents/Traffic_enforcement.pdf.

Grogger, Jeffery, and Greg Ridgeway. "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness." *RAND*, 2006.

Lloyd, Camille. "For Black Americans, 41% of Police Encounters Not Positive." *Gallup.com*, Gallup, 2 Nov. 2021, https://news.gallup.com/poll/316247/black-americans-police-encounters-not-positive.aspx.

"The Stanford Open Policing Project." *Openpolicing.stanford.edu*, https://openpolicing.stanford.edu/findings/.

U.S. Census Bureau Quickfacts: United States. https://www.census.gov/quickfacts/fact/table/US/PST045219.