Statistical Analysis

While the source data itself illustrated some interesting relationships between a number of variables, we had to check to see if those relationships were statistically significant. With the size of the dataset that we were using, Excel provided a sufficient medium for analysis. This analysis was performed in three files. The first working with the main dataset and being used to corroborate and double-check some of the findings from in the Jupyter Notebook. It was also used for some exploratory trend highlighting, Many of which ended up being outside the scope of this particular project, but brought up interesting questions, regardless.

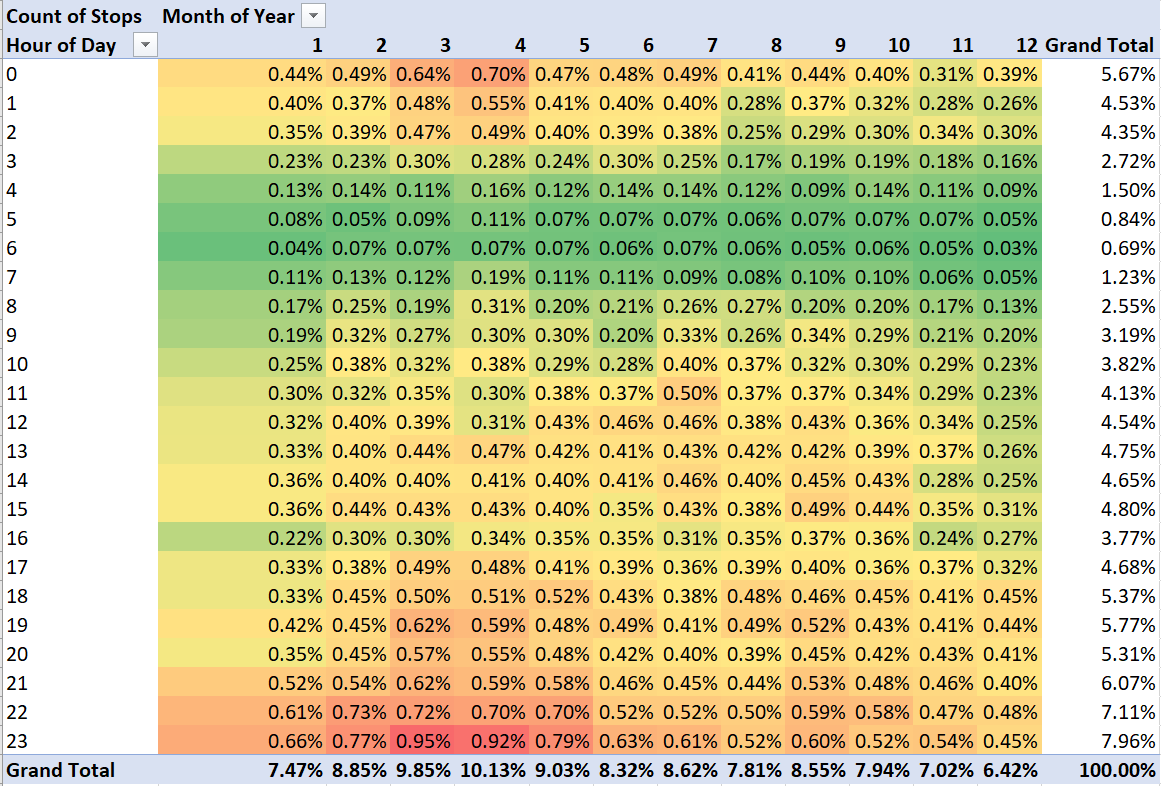


Figure 1: Why is such a high proportion of the year's police stops occurring at midnight in March & April?

The second excel file of data incorporated census data from Hennepin county from 2010. Unfortunately, the census data was in a format poorly suited for importation and recombination with the Minneapolis dataset.

We wanted to compare the racial percentage breakdown of police stops with the overall percentages of the population. This proved difficult due to the census categorizing race differently than the Police data, (For example: the 2010 census grouped “East African” in with “Black” while the Minneapolis police did not. Likewise, the census categorized people of two or more mixed races, while the police data did not.) Each combination of census race category, ager category, and sex category had a number of respondents which could be aggregated for comparison. But in order to create the comparable aggregate, the census categories had to be divided into the Police Data categories (grouping East-African with Black) and each group of respondents matching was flagged with a Boolean. This was slightly complicated by the census categorizing Latinx in a category of its own.

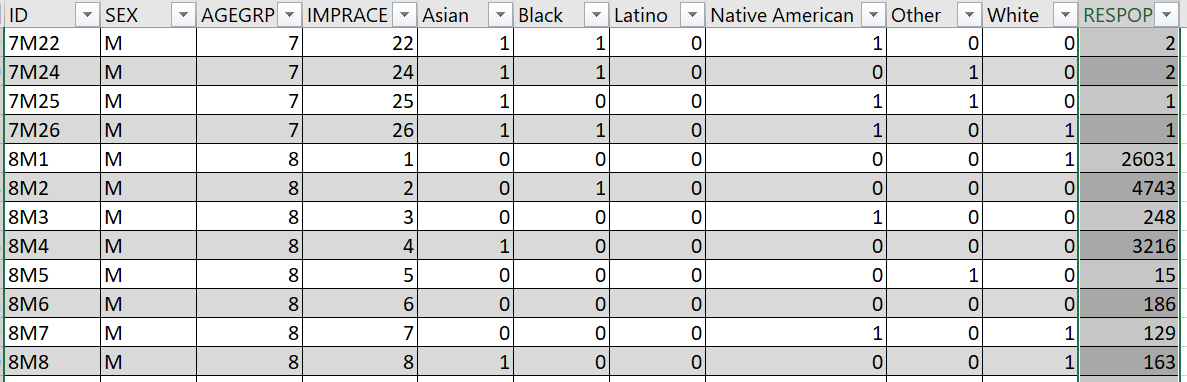


Figure 2: Sample of Census data cleaned to boolean police data race categories. For example: Census Racial category 22 is classified as a mix of Asian AND Black AND Native American. There are two men of age category 7 (age 30 -34 years) in Hennepin county matching that description.

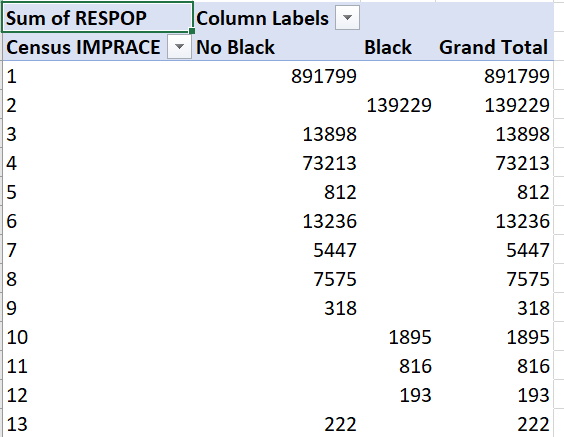


Figure 3: Census Race categories to the left, one of the selectable Police data categories at the top. The table illustrating which respondent groups match the police data designation.

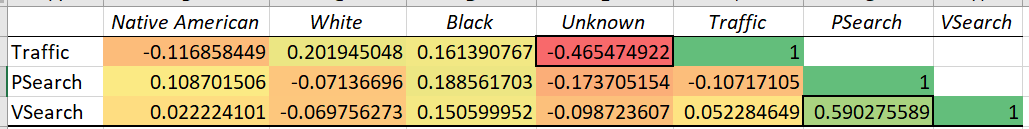
The Third excel data file used the data of the first two in order to create correlation tables and run linear regressions. While there were numerous correlations, most of them were statistically weak. An example of some of the stronger correlations were whether certain races were searched more frequently during traffic stops. While correlation is not causation, it was significantly less likely that a racial determination would NOT be made during a traffic stop. This is in contract to the possibility that people whose race could not be determined were significantly less likely to get stopped by police for traffic violations.  


Figure 4: Native Americans were slightly less likely to be involved in traffic stops. Whites were more likely to be involved in traffic stops, but less likely to be searched. Blacks were more likely to be in traffic stops and more likely to be searched.

The regressions performed were descriptive, but not in the manner we had hoped. While the T-stats and p-values were all good, the R² values were low, indicating that the variables very accurately accounted for a very small portion of the variation in the data. The rest was just random noise. Regressions were run on precinct predictability of race mismatches, precinct predictability of issuing citations, Racial prediction of traffic stops, personal searches and vehicle searches, none of them were statistically significant enough to reject the null hypothesis.

