Mohamed Nabeel

Data 205

2019 Montgomery County Arrests

**II.**

The dataset I looked at first was Daily\_Arrests.csv. It was retrieved from dataMontgomery and updated almost daily. It contains 666 rows of data and 6 variables that I could use and analyze. The next dataset I looked at was Crime\_Year\_To\_Date\_-\_2019.csv and was also updated daily. This dataset had over 40,000 rows of data with 30 different variables that I could use. I decided to use the Crime\_Year\_To\_Date\_-\_2019.csv because it was much more appealing because it had more information I could utilize. The Crime\_Year\_To\_Date\_-\_2019.csv had crimes that were reported by the National Incident-Based Reporting System of the Criminal Justice Information Services.

The variables I focused on from the Crime\_Year\_To\_Date\_-\_2019.csv dataset was crime\_name2, victims, zip code, police district and the exact location. There is crime\_name1 and crime\_name3 as well. Crime\_name1 was too broad for me to use while crime\_name3 was too specific for me to use. Crime\_name2 was right in the middle, so it was the best one for me to analyze. The victims showed the number of victims affected by each crime. The corresponding zip code and police district were reported with each arrest. The location had the longitude and latitude coordinates, putting this on a map will give you the exact location of the arrest.

**III.**

One of the first things I looked at intensively was the location. I used a map to look at the different areas that contained the arrests. I saw that areas like Clarksburg, Poolesville, Laytonsville and Darnestown had the least amount of reported crimes, all under 1000 arrests in all of 2019. The most affected areas were Germantown, Bethesda, Gaithersburg and Rockville. All these areas had over 3000 rows of data. The most likely reason for this is because of the denser areas, with a higher population count. The color correlation is the greener areas having the lesser amount of arrests, while the redder areas were the higher amount of arrests.

A close up of a map

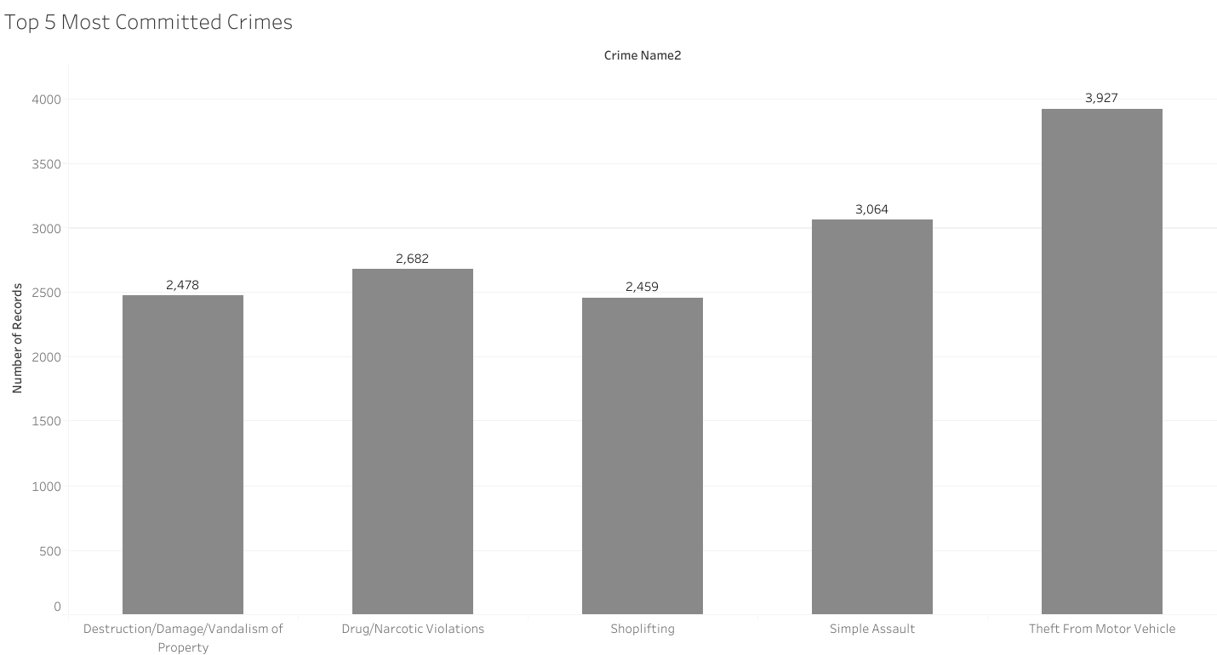
Description automatically generated

Also, regarding the location, I looked at the zip code of the reported arrests and sorted the data from the zip codes with the most amount of arrests to the zip codes with the least amount of arrests. I made a chart the two displaying the 5 zip codes with the most occurring arrests. 4 of the top 5 zip codes all occurred in Silver Spring.

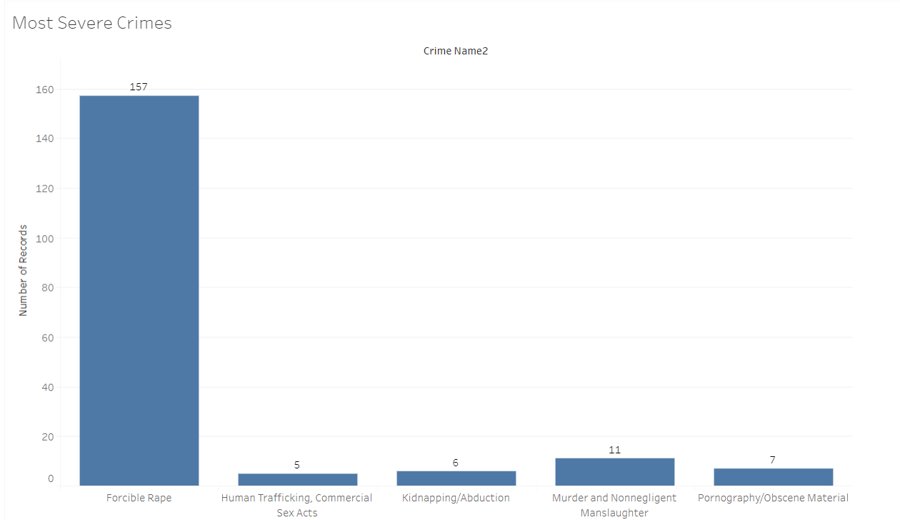


The location and areas that are heavily affected is critical to know. If an area with a low crime rate is having more personnel, resources and time than the need, then that will take away from the more higher crime rate areas.

The next variable that I analyzed was the cime\_name2. I first set the variable to sort by the number of occurrences in descending order. I wanted to see the crimes that caused the most arrests. After doing this I learned that the top 5 crimes occurred was theft from motor vehicle, which had 3,927 arrests, simple assault with 3,064 arrests, drug/narcotic violations with 2,682 arrests, damage to property to 2,478 arrests and finally shoplifting with 2,459 arrests. It is important to know this because we can see which crimes we needed to combat against the most, which crimes we needed to be ready for and which crimes affected people the most. I was able to get this visualization through the sorting. It was important because it shows us which areas had the most arrests and needed more patrolling cars and police officers in the area.

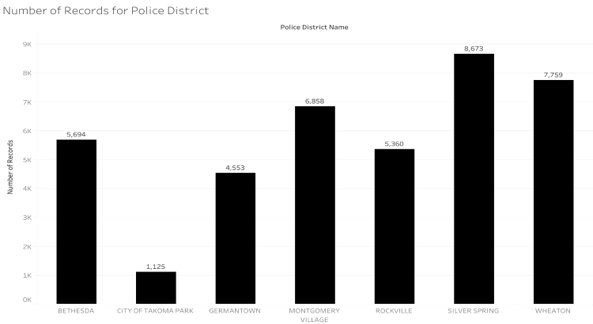


I investigated crime\_name2 even further and looked at the different crime names. I made another visualization that looked at, what I thought, were the most severe crimes. The reported crimes were forcible rape, with 157 arrests, human trafficking with 5 arrests, kidnapping with 6 arrests, murder with 11 arrests and child pornography with 7 arrests. It is important to know this because we need to know about which arrests, we should be aware of the most. I made another visualization representing this.



In my opinion it is more important for police, people making budgets and even regular citizens to look at preventing these more severe crimes rather than the more occurring crimes. If you hear a crime about a murder or kidnapping in your area it was affect you much more than a shoplifting or theft from motor vehicle crime.

Another variable I wanted to investigate more were the number of arrests that each police district made throughout 2019. I wanted to look at the police districts of the most populated cities in Montgomery County. As a result, this visualization was developed.



Looking at the visual, the Police Districts in Silver Spring, Wheaton and Montgomery Village handled most of the arrests, each over 6,800 arrests, while the City of Takoma Park had the least amount of Arrests, only around 1,125. The more populated areas will have more crime, so it is crucial to have the necessary amount of police enforcement in those areas. It is important to know this information because we can see which districts handled majority of the arrests in Montgomery County, so we are able to provide them more funding, more equipment and even more personnel, if necessary.

**IV.**

The dataMontgomery website was a great website to use with the different filters and visualizations available. One thing I would recommend is to take out some of the repetitive columns. The dataset had a latitude column, longitude column and a latitude and longitude coordinate column. Since there is a column with both the latitude and longitude coordinates, there is no need to have columns with them as separates. Another recommendation is having the missing values taken out before the dataset is officially published. The most important feature a dataset can have is accuracy. The more accurate a dataset is, the better than dataset is. It is simple for coders to remove the NA values from a dataset, but not everybody will know how to do that. A final recommendation is having the dataset contain separate columns for the time and date. Because the time and date columns were fused together into one column, to be able to extract one from the other is extremely complicated. I wanted to work with the time separately and see what information I could find by looking into that. I tried researching online and tried some recommendations that other peers gave me, but I was unsuccessful in separating the two, so I missed out on looking into the time within this dataset. Although I do have these recommendations that will improve the datasets, I found dataMontgomery a great site to use! There are over 40,000 lines of data in the Crime\_Year\_To\_Date\_-\_2019.csv dataset, which is an enormous amount of data. The dataset was also updated daily, with arrests being added and even having arrests being removed, I am assuming because the potential suspect was acquitted of the crime. I will be looking forward to using dataMontgomery again in the future.

**V.**

Acknowledgments: I want to thank Dennis Linders, the Montgomery County Government, Victoria Lewis and dataMontgomery for giving me access to these datasets and information. I want to thank Professor Linehan, Professor Mohamed, Professor Saidi and Professor Hamman for guiding me to this point in my educational career.