Michael Armesto, David Doman, & N'Dea Jackson
December 4, 2020
IST 652 – Final Project Report

Montgomery County Crime Analysis

Data Description:

The data used for this project was obtained from dataMontgomery, an online platform where branches of the Maryland's Montgomery County Government share public data on a number of topics, such as education, housing, health, transportation, and public safety. These datasets contain information from all the different cities that make up Montgomery County, including Bethesda, Chevy Chase, Clarksburg, Gaithersburg, Germantown, Montgomery Village, Potomac, Rockville, and Silver Spring.

The Montgomery County Police share several datasets detailing police and criminal activity in the county. This report analyzes two key datasets, one detailing each dispatch call for police, and another detailing each crime reported.

The first dataset, Police Dispatched Incidents, is comprised of all police dispatched incident records that took place in Montgomery County. The earliest date within the data set is April 2, 2017 and the most recent date recorded is November 26, 2020. Some of the attributes that are used to describe the incidents include:

- **Incident ID**
- **Start Time**
- **End Time**
- **Priority**
- **Incident Type**
- **Close Type**
- **Address**
- **City**
- **State**
- **Zip**
- **Latitude**
- **Longitude**
- **Police District Number**
- **Beat**
- **PRA**
- **CallTime CallRoute**
- **CallTime Dispatch**
- **CallTime Arrive**
- **CallTime Cleared**
- **CallRoute Dispatch**
- **Dispatch Arrive**
- **Arrive Cleared**
- **Disposition Desc**
- **Location**

The second dataset, Crime, details every crime report in the County since July 1st, 2016. The dataset is updated daily. Not every incident a police officer is summoned for are crimes, and this smaller dataset provides a more detailed look at crimes committed. Some of the key attributes in this dataset are:

- **Incident ID**
- **Offence Code**
- **Crime Report Number**
- **Dispatch Date/Time**
- **NIBRS Code**
- **Number of Victims**
- **Crime Name 1**
- **Crime Name 2**
- **Crime Name 3**
- **Police District**
- **Block Address**
- **City**
- **Sector**
- **Beat**
- **PRA**
- **Start Time**
- **End Time**
- **Location**

Team Member Roles:

N'Dea: Focused mainly on the call time arrive variable and understanding what factors effected the time it takes police to respond. Also focused on creating the shell of the report.

David: Primarily focused on time/when crimes occur. Also compared violent vs. Non-violent crimes and how they differ in time of occurrence.

Michael: Focused on incorporating generalized categories into the very specific crimes from the dataset by pulling in external data, and using them to generate choropleth maps to create overlays of crimes committed by location.

Preprocessing:

Police Dispatched Incidents

To begin pre-processing, the dataset was cut down to only contain dates from the years 2019 and 2020. This allowed the data set to be cut down from 770K rows to 377K rows. Next, rows that were deemed as unnecessary for our analysis were removed, including the *Crime Reports* and *Crash Reports* column. These were deemed unnecessary because they car crashes were considered outside the focus of analysis, and crime reports were provided in greater detail in the Crime dataset.

Crime by Time

In order to perform analysis based on when crimes were committed the "Dispatch Date / Time" column needed to be converted to the datetime format. The code used is shown below:

```
#Import datetime
import datetime as dt

#Convert 'Dispatch Date / Time' to datetime format
mocoCrime2['Dispatch Date / Time']=pd.to_datetime(mocoCrime2['Dispatch Date / Time'])
```

Types of Crime

In order to investigate the types of crime committed in the Montgomery County area, the Crime dataset was used. To look into the types of crime committed, the columns detailing the criminal activity were explored. It was found that Crime Name1 was the most general field, consisting of only 6 unique types. This field, however, was considered to be too general, as the types were only as descriptive as "Crime Against Society" or "Crime Against Property", etc. The other fields were found to be too descriptive, having too many unique values to work with.

```
for col in crimeTypes:
    print(col, ': ', len(crimeTypes[col].unique()))
```

```
Offence Code :  603
NIBRS Code :  55
Crime Name1 :  6
Crime Name2 :  57
Crime Name3 :  326
```

In order to categorize the crimes, an additional dataset containing a categorized index was brought in. This data proved troublesome to find, as the Montgomery County police only provides such information in PDF form . Initially, a document from the Denver police department that contained an index of offence codes and offence categories was brought in. While the offence codes used by the Denver police did mostly match those of Montgomery County, it seemed many relevant codes were missing, leading to a high number of NA values. NIBRS codes, which are a part of the FBI's Uniform Crime Reporting program, were used instead as they are standardized nationally. The FBI does not have a publicly available index table of these codes, but a suitable excel table provided by the Texas Department of Public Safety was obtained. After some light cleaning in Excel, this table offered all the necessary NIBRS codes and their categories.

```
datapath = 'C:/Users/marmesto/Documents/Cuse/scripting/project/texascrime.csv'
crimeCategories = pd.read_csv(datapath)
crimeCategories.head()
```

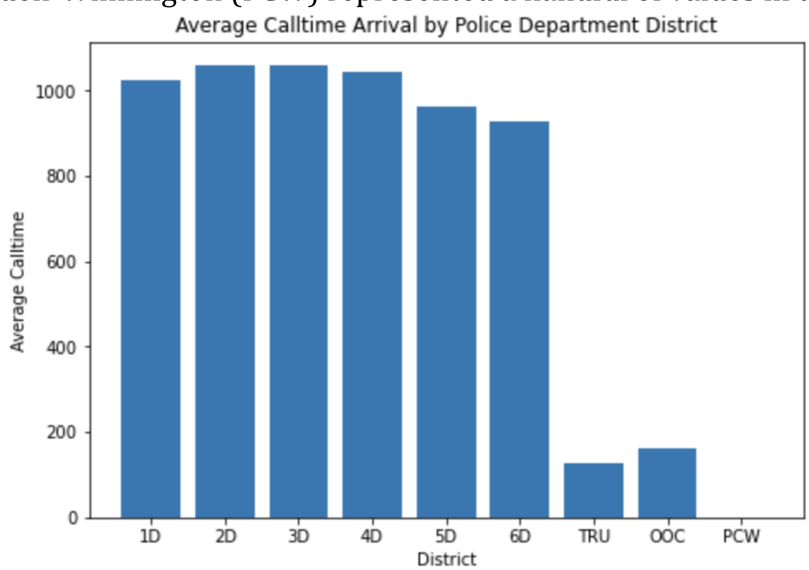
	Offense	IBR Code	Offense Description	Crime Against
0	Animal Cruelty Offenses	720	Animal Cruelty	Society
1	Arson	200	Arson	Property
2	Assault Offenses	13A	Aggravated Assault	Person
3	Assault Offenses	13B	Simple Assault	Person
4	Assault Offenses	13C	Intimidation	Person

Note the IBR codes correspond to the NIBRS codes of the crime dataset. The “Offense” column contains the desired crime categories. After incorporating the Offense column, the dataset was ready for analysis.

Methodology and Data Questions

What are the factors that affect the dispatch times of Montgomery County Police Officers?

- **Fields to be used:** police district number, priority, season of the year, time of the day.
- **Resulting Output:**
    - Police district was a factor that effected the average calltime arrival of Montgomery County Police officers. The average call time was very low in the TRU and OOC districts. The district with the highest call times were districts 2D and 3D.
    - Telephone Reporting Unit (TRU), Out of County (OOC), and Philadelphia-Camden-Wilmington (PCW) represented a handful of values in the dataset.



Incident Count by Police District

```
3D     73421
4D     73311
6D     64314
1D     60955
2D     57217
5D     48101
TRU       56
OOC       24
PCW        1
Name: Police District Number, dtype
```

Incidents by Priority

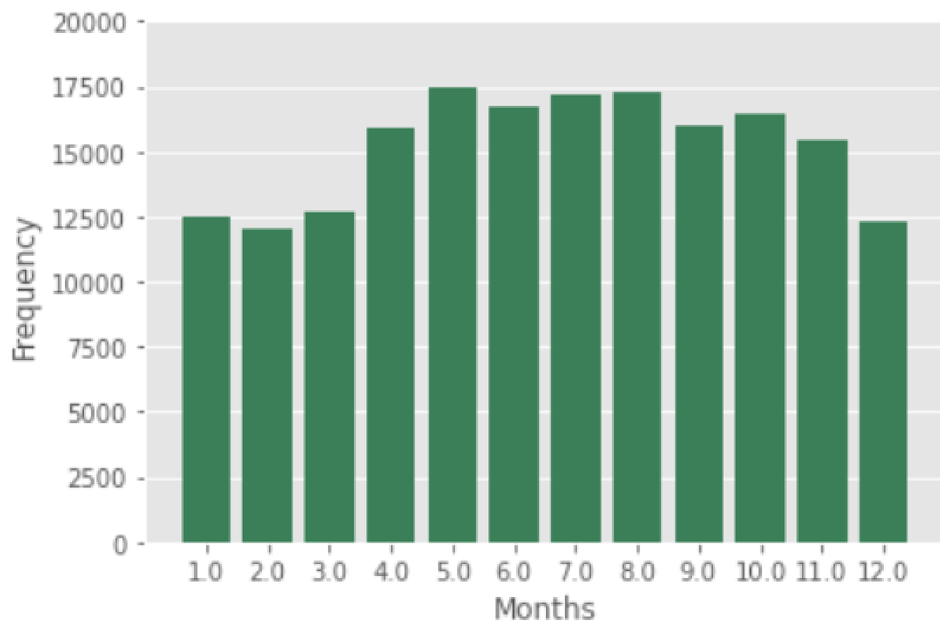
```
1    160608
2     81740
4     71281
0     35099
3     28672
Name: Priority, dtype: int64
```

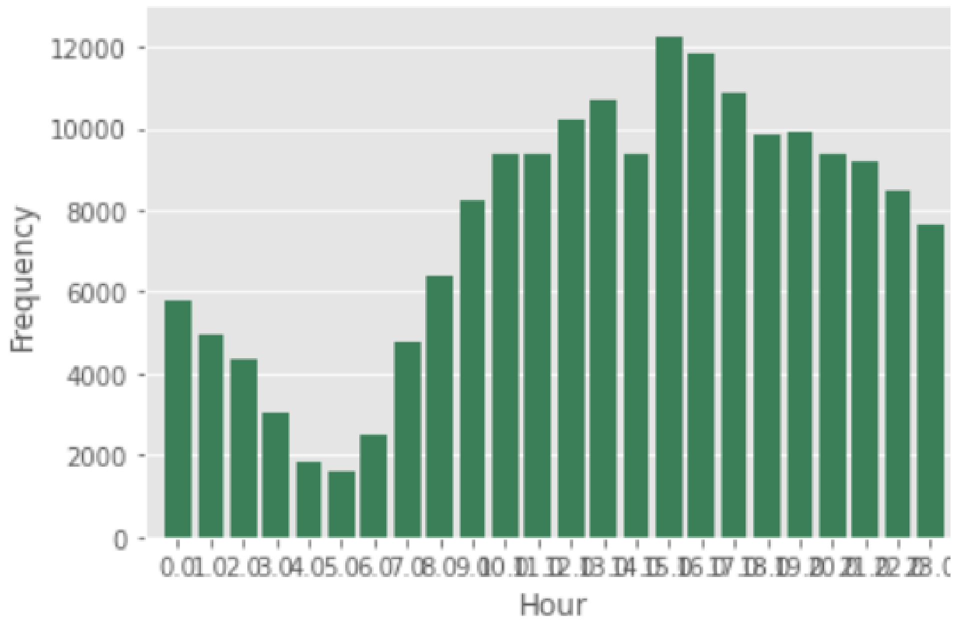
When do crimes occur most and do certain crimes occur more frequently at different times in the day?

- ***Fields to be used:*** Dispatch Date / Time, part_of_day,
- ***Resulting Output:***

In order to answer the question of when crimes occur was to look at which months are crimes most frequent. As shown below, it is clear to see that the warmer months have the most frequent crime count. Next, was to look at the hour of the day that crimes occur. It can be shown below that early in the day there are less crimes and as the day goes on crime frequency increases.

Crimes by Month & Hour





After looking at months and hour of the day for crime frequency, a field was created to break the day into 4 sections: Morning, Afternoon, Evening & Night. As shown by the resulting output below, Night is by far when the most crimes occurred followed by Evening.

```
print('Morning')
print(len(mocoCrime2[mocoCrime2["part_of_day"] == 'Morning']))

print('Afternoon')
print(len(mocoCrime2[mocoCrime2["part_of_day"] == 'Afternoon']))

print('Evening')
print(len(mocoCrime2[mocoCrime2["part_of_day"] == 'Evening']))

print('Night')
print(len(mocoCrime2[mocoCrime2["part_of_day"] == 'Night']))
```

```
Morning
44174
Afternoon
42569
Evening
51830
Night
91310
```

Crimes by Place and Part of the Day

Using the created "part_of_day "field discussed previously, the type of crime by place was analyzed. As shown below, most crimes occurred during the night and the most frequent place of crime was Street – In vehicle. Residence – Single Family was second with Residence – Apartment/Condo following.

Place	part_of_day	Count
Street - In vehicle	Night	20223
Residence - Single Family	Night	9030
Residence - Apartment/Condo	Night	8130
Street - Residential	Night	5647
Other/Unknown	Afternoon	4798
Residence -Townhouse/Duplex	Night	4279
Parking Lot - Residential	Night	3641
Parking Lot - Commercial	Night	3436
Residence - Driveway	Morning	3408
Street - Commercial	Night	3258
Retail - Department/Discount Store	Evening	2202
School/College	Morning	1986

Do violent crimes occur at different times compared to non-violent crimes?

- ***Fields to be used:*** Dispatch Date / Time, Crime Name2, Incident ID, part_of_day, Violent
- ***Resulting Output:***

To answer the question of top violent crimes and top non-violent crimes, the crime names were examined and were ultimately determined to be violent vs. non-Violent. There were 57 crime descriptions, so this took some time to go through and make the necessary designation. As shown below, Aggravated Assault was the most frequent violent crime followed by Forcible Rape. As for non-violent crimes, All Other Offenses was the leading crime, which doesn't provide much useful information followed by Drug/Narcotic Violations.

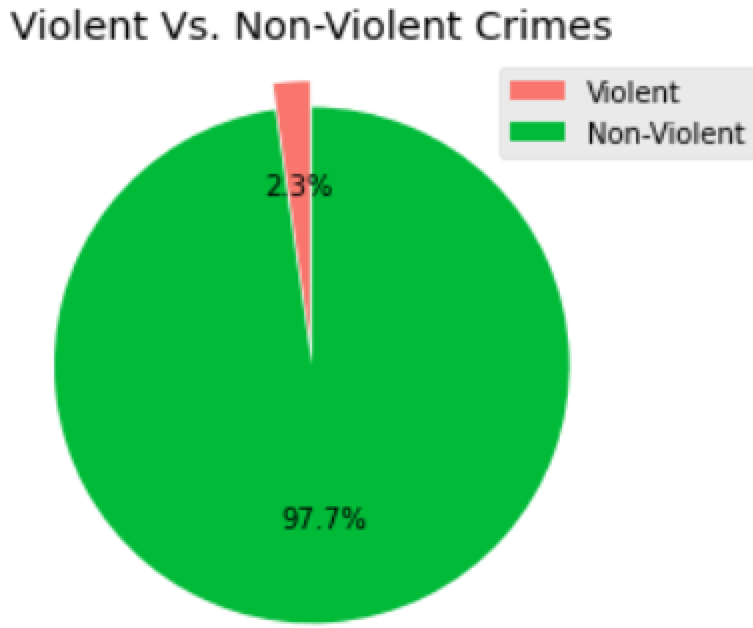
Top 5 Violent Crimes

	Violent	Crime Name2	Count
45	1	Aggravated Assault	2925
48	1	Forcible Rape	940
49	1	Forcible Sodomy	374
47	1	Forcible Fondling	341
55	1	Sexual Assault With An Object	248

Top 5 Non-Violent Crimes

	Violent	Crime Name2	Count
0	0	All Other Offenses	53195
12	0	Drug/Narcotic Violations	21014
38	0	Theft From Motor Vehicle	20175
36	0	Simple Assault	15155
8	0	Destruction/Damage/Vandalism of Property	13988

To gain an understanding of the distribution between violent and non-violent crimes, a pie graph was produced. It is clear to see that overwhelmingly non-violent crimes are much more frequent, which is expected.

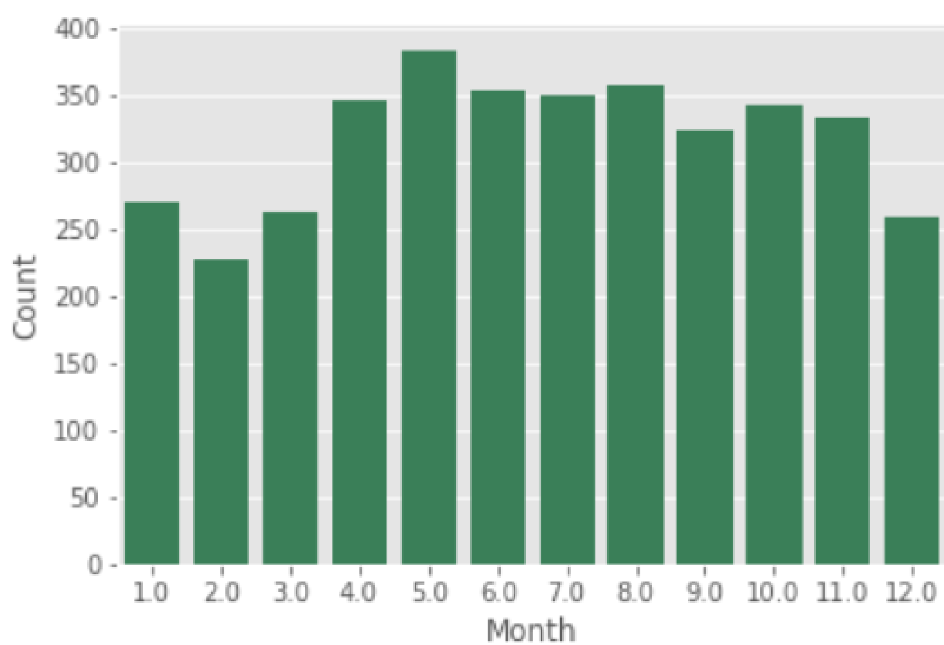


-

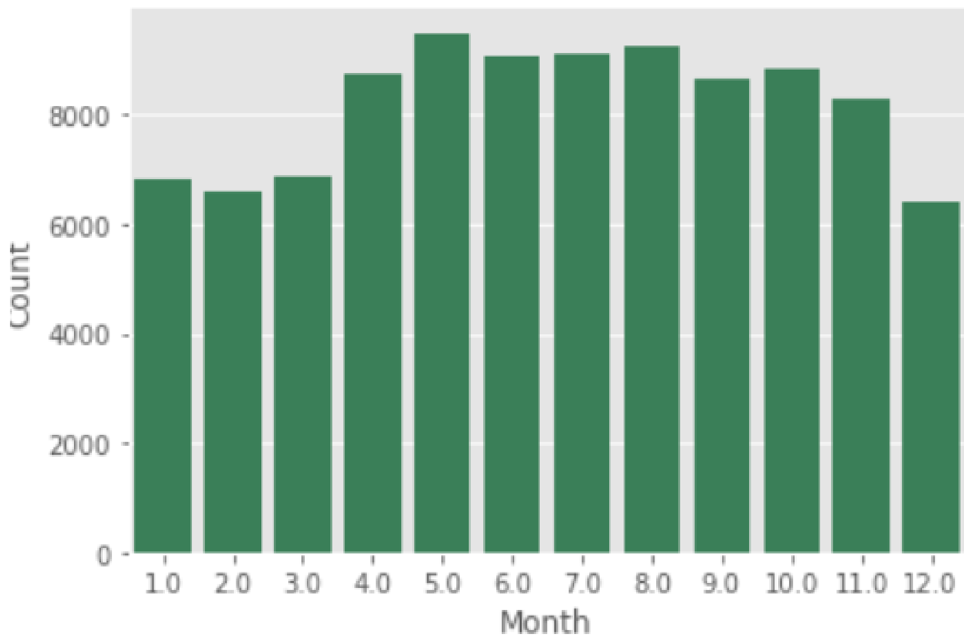
Next was to see if there was a difference between when violent and non-violent crimes occur and if there were any differences. There does not seem to be any distinct difference in the month for which violent and non-violent crimes occur. Although, when looking at the hour of the day, violent crimes definitely seem to be much more frequent in the later hours of the day when compared to non-violent crimes.

Violent Crimes by Month

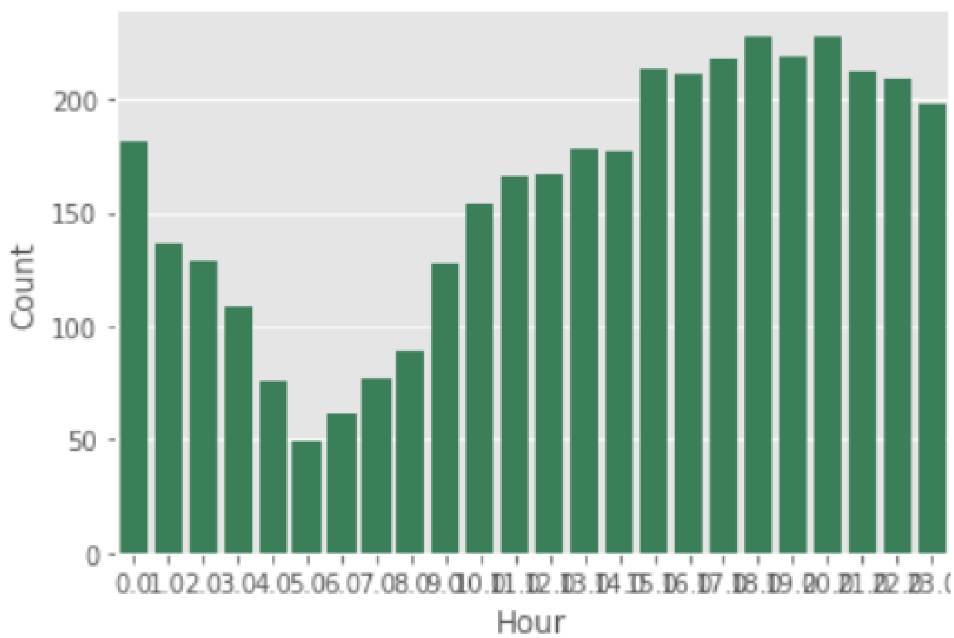
-



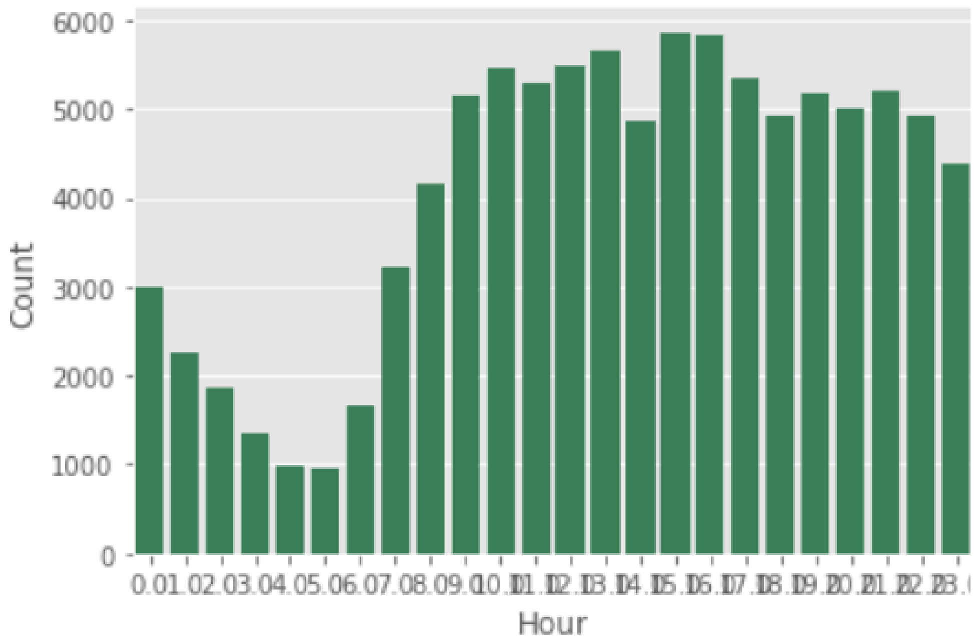
Non-Violent Crimes by Month



Violent Crimes by Hour

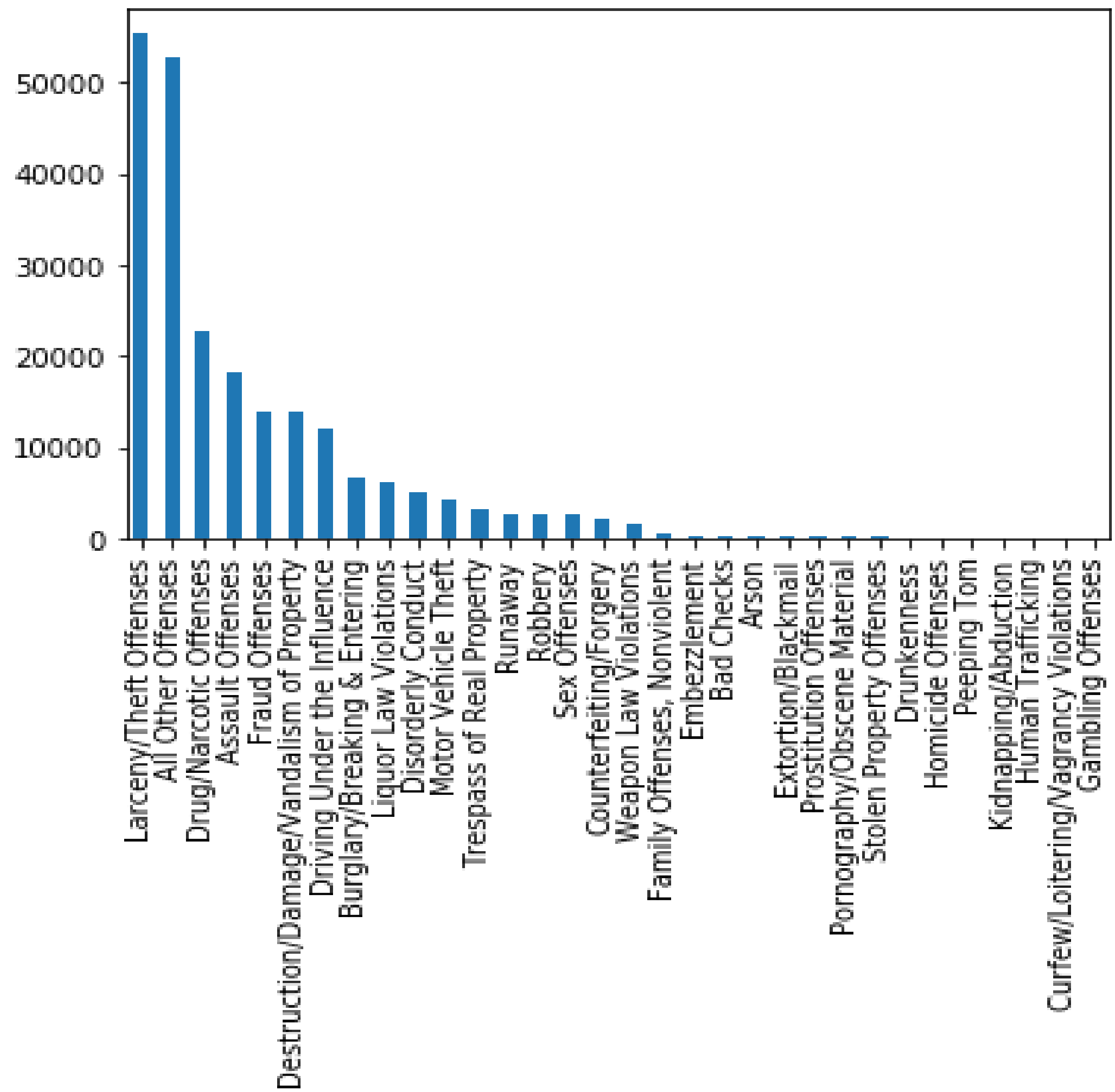


Non-Violent Crimes by Hour



What kinds of crimes were committed in Montgomery county?

After incorporating generalized categories based on the NIBRS codes of each crime report, the types of crimes committed could be more easily processed. 31 unique types of crime were identified. The vast majority of these crimes appeared to be Larceny/Theft Offenses and another category simply called All Other Offenses.



Initially, 2764 NA values were discovered amongst the Offense column, which stored the categories. By searching with the isnull() function, it was found that all 2764 of these instances were labeled as “Runaway” in Crime Name2. These values were then assigned the Offense category “Runaway”.

Because more than 50,000 instances of All Other Offenses were found, it seemed necessary to break them out further. It was found there were 85 unique values with the Crime Name3 corresponding to this category.

```
## check out "all other offenses"
othercrime = crimeReports2[crimeReports2['Offense'] == "All Other Offenses"]
print('Unique crimes classified as Other Offenses:', len(othercrime['Crime Name3'].unique()))
othercrime['Crime Name3'].value_counts()
```

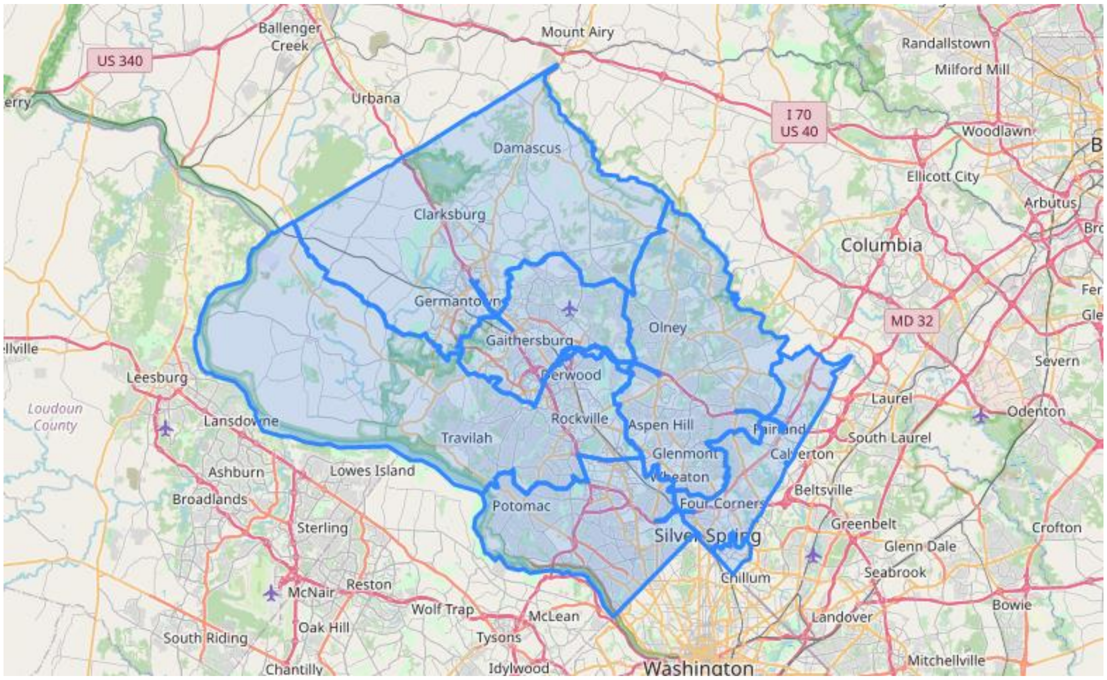
```
Unique crimes classified as Other Offenses: 85


POLICE INFORMATION                            12996
MENTAL ILLNESS - EMERGENCY PETITION            8244
LOST PROPERTY                                  8029
SUDDEN DEATH                                   4197
MISSING PERSON                                 3523
                                               ...
FAILURE REPORT CRIME                              1
PUBLIC PEACE - ASSEMBLY - UNLAWFUL                1
CONSERVATION - ANIMALS (DESCRIBE OFFENSE)         1
CONDIT RELEASE VIOLATION                          1
TRANSPORTING DANGEROUS MATERIALS                  1
Name: Crime Name3, Length: 84, dtype: int64
```

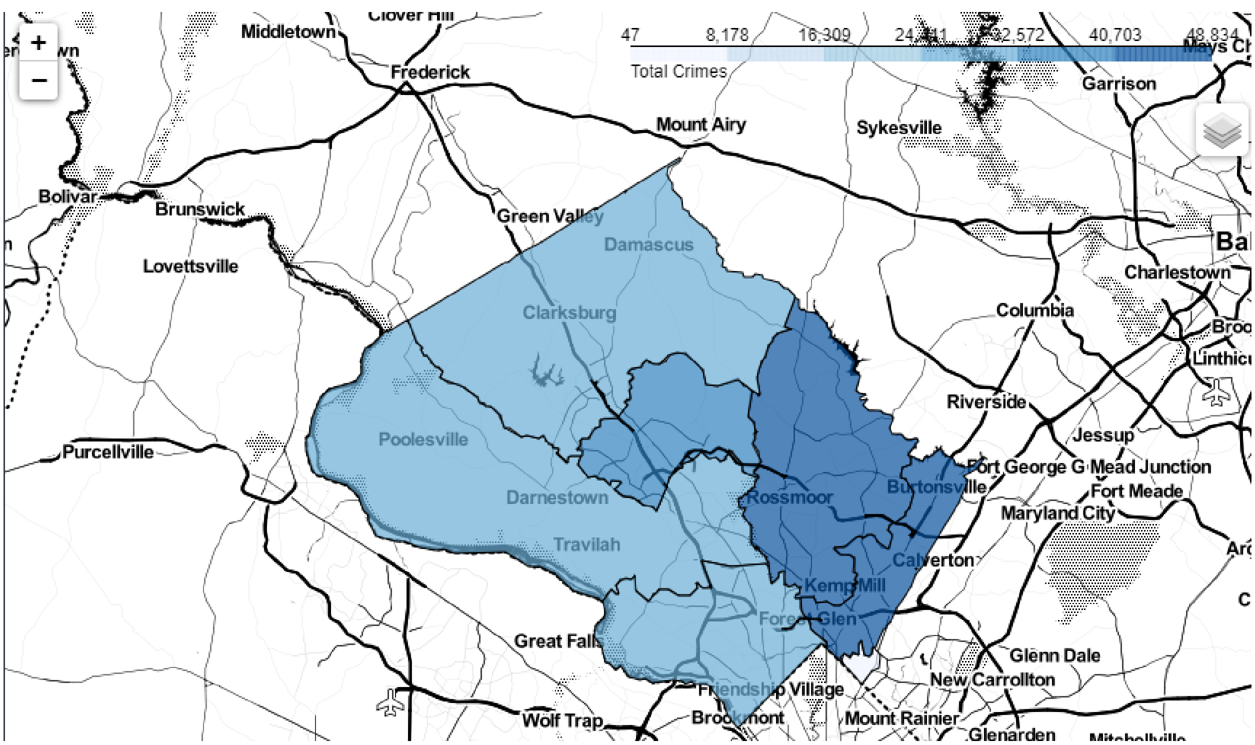
The bulk of these values contained non-crime instances like police information, mental illness, and missing persons. After those major 5 sub-categories, there was a steep drop-off to very specific and infrequent events such as transporting dangerous materials. Because of the nature of these instances, it seemed fitting to simply leave them be as "other offenses".

Where are crimes likely to occur?

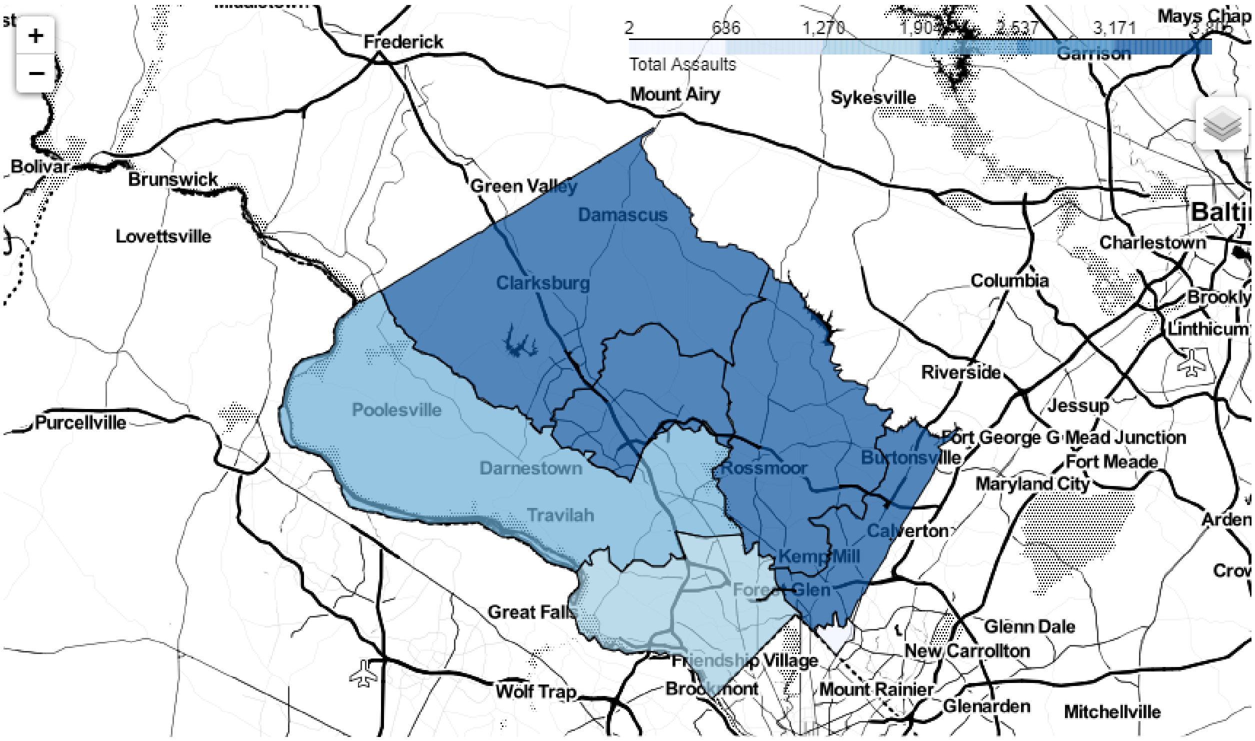
Because so many unique cities, zip codes, beats, and other location attributes existed in Montgomery County, it was determined that focusing this investigation on the 8 Police Districts would be the best option. To visualize these districts, choropleth maps were created using [GeoJSON data] hosted by the dataMontgomery site, along with the Folium package.



Using the pandas groupby() function, an aggregated dataframe containing the number of offenses by police district was created. After renaming these police districts to match the data in the geoJSON file, and a great deal of trial and error, a Folium choropleth was created. Note these maps are interactive within the Jupyter notebook.



The above map indicates total instances of crimes committed (excluding the other crimes) in each police district. To further investigate, another aggregated dataframe was created, this time consisting of only assault offenses.



We see a great deal of change in this map, and suddenly the districts containing Clarksburg and Damascus do not seem so peaceful.

Conclusion:

In regard to the 'Calltime Arrive' data, the highest average call times spanned across districts 2D and 3D. Because the data used was observed over 2019 and majority of 2020, the yearly data was not calculated. The TRU, OCC, and PCW districts seemed to represent a smaller portion of the data. It was not determined what these abbreviations stood for, but it was speculated that this district could represent special branches of the Montgomery County Police Department.

The number of crimes committed were analyzed by month and by hour of the day. It was clear that most crimes occurred during the warmer months and faded off during the colder months. This was an expected outcome. When looking at the hour of the day that most crimes were committed it was highly skewed towards hours later in the day. Times between 5PM – 9PM seemed to have the highest number of crimes committed. Again, this was an expected outcome.

Looking at violent crimes vs. non-violent crimes it showed that the time of year didn't make much of a difference. Regardless if it was a violent or non-violent crime, the most crimes occurred during the warmer times of the year. Although, when looking at the hour of the day when violent crimes occur, it is clear that violent crimes occur much later in the day compared to non-violent crimes.

When analyzing the types of crimes committed in Montgomery County, it can certainly be said that the local police officers see all types. Crime Name3, which holds the datasets most specific categorization of crimes, contains 326 unique values. Even after grouping them based on the FBI's NIBRS codes, there were still 32 unique offense categories. Everything from assaults and carjacksings to bad checks, illegal pornography and peeping toms (which is evidently the technical term) were found in the dataset. In the dataset's mere 3.5 years, a shocking 55,496 incidents of larceny/theft were documented, along with 22,765 drug/narcotic offenses, and 18,367 assault offenses. The dataset even contains 2,765 runaways, and a disturbing 36 incidents of human trafficking.

When analyzing the crimes committed in Montgomery County by police district, one can see a distinct pattern. Two neighboring districts containing Silver Spring and Wheaton contain nearly half of the county's crimes. Somewhat surprisingly, the populous, but wealthy district containing Bethesda and Chevy Chase (furthest south on the map) manages to stay relatively peaceful. When looking specifically at assaults, the Northern two districts containing Gaithersburg, Damascus and Germantown appear more menacing, hosting very similar rates to their troublesome neighbors to the east. In the assault regard, the aforementioned southern-most district is even safer, and, if you can afford it, certainly seems like the best part of Montgomery County to settle down.