Visual Analysis of Crimes in Montgomery County in Maryland, United States of America

Deenadhayalan Ravi

Abstract — In order to support the Police Department of Montgomery County, Maryland, USA on their crime management and resourse allocation, a detailed visual analysis has been carried out on the US Crime data from 2017 to 2022. Temporal, spatial and spatiotemporal analysis has been carried out using Tableau for visvalization.

Using the line graphs for trends patterns, heatmaps for seasonality to find the trend of the crime frequency. Using density based maping for finding the cities with high crime frequency. Using K-Means clustring for clustring the cities based on the nature of the crime to allocate resources who handle similar crimes. Using density map with time series for analysing the spatiotemporal part to find which crime category has varied over the time to have a pattern shown in the line graphs of temporal analysis.

1 Problem Statement

The temporal and spatial patterns from the analysis should show the crime patterns in the county and support the decision making for the police department management and planning.
Temporal analysis:

- How is the crime pattern over the years, is there any pattern? If so, what is the reason for the same?
- Is there any seasonality in the crime pattern in months / day of month / weekdays / hours of a day.

Temporal pattern supports in the resource allocation, with more police officers potentially becoming available during the peak times, when crimes are most likely to occur.

Spatial analysis:

- Which areas of the county have the most crimes
- Cluster the areas of the county with respect to nature of the crime. This gives a better picture of areas with similar crime types and their intensities.

Identifying crime hotspots helps the police department to make a strategy to bring the numbers down. Doing spatial analysis on the given area will give a clarity.

Spatiotemporal analysis:

- Is there a spatial-temporal crime pattern in the given area? The crime pattern of the county (spatial) over the years (temporal).

The spatiotemporal analysis highlights the crime intensities of the area over the time which leads to better understanding on the features or factors which influence a change in the pattern.

2 State of the Art

From the journal Crime Analysis using k-means Clustering 2017, International Conference on Computational Intelligence and Networks.[2]. We understood how to do use K Means clustering for a crime data.
D. Gao, X. Li, C. Yang and Y. Zhang. Spatial patterns analysis of urban road traffic accidents based on GIS, International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), Xiamen, 2012, pp. 1898-1901[3] Journal has used various techniques like trend analysis using bar and line graphs and used heatmaps for analysing the seasonality and used density mapping for find the hotspots of the traffic accidents.
From all these journals and some non- Journals refences, books references and internet search I have carried out the analysis.

3 Properties of the Data

Our dataset is all about the crimes happened in United States of America over the seven years from July-2016 to August-2022. There are 306,094 observations and 36 features with Incident ID as a unique value for each crime incident, the no. of victims, category and sub category of crime, location of the crime (state, county, city, streets, latitude and longitude), police department name who handled the crime, date of crime with timestamp.
- The date and time provided for each incident supports the temporal analysis with which we can analyse the crime data in several dimensions like year, month, date, day of the week, hours of the day.
- The latitude and longitude coordinates for each incident supports the spatial analysis. In this dataset 99% of the data belongs to the Montgomery County in Maryland State. So, the remaining outliers has been removed from the dataset. The county has 42 cities with zip codes as subsets of cities.
- The dataset has been cleaned in the aspects of null values, missing values, duplicate values, correcting datetime format, etc. using Python and all the charts and graph are generated in Tableau.
- For the analysis, we will be focusing on the crime categories and subcategories which is explained in the following two figures,

From fig 1, we can understand that there are three main categories of crime, in which Crime Against Property tops the table with almost 45% and followed by Crime Against Society 22% then others – 20% and then Crime Against Person nearly 10%. The no. of victims is almost equal to the no. of crimes in all the category, in an overview this clarifies that none of the crime category is a reason for any bulk victims.

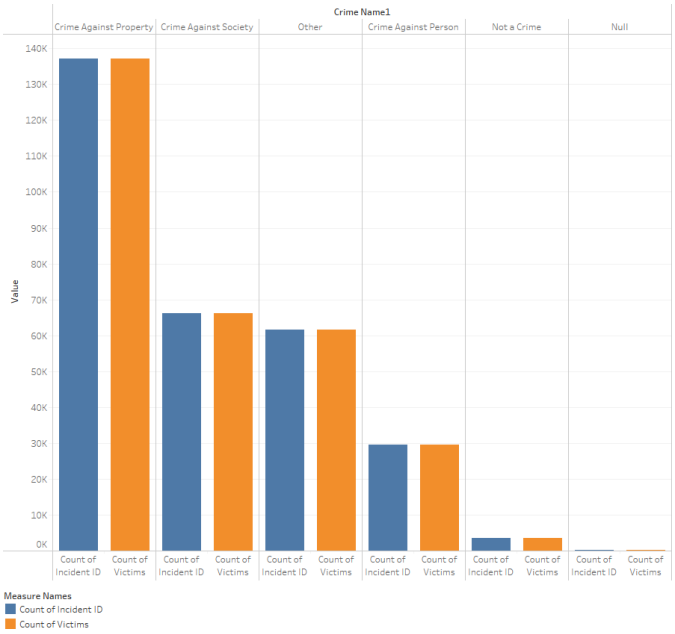

Fig 1: No. of Crimes and Victims by crime category.

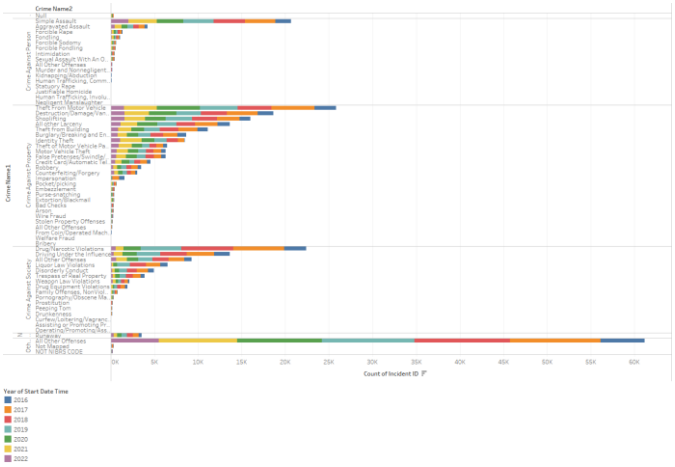

Fig 2: No. of Crimes by Crime Subcategory with yearly distribution in colour.

From fig 2, we can understand that Simple Assault is the subcategory which contributes a lot in the Crime Against Person category. Theft From Motor Vehicle, Destruction/Damage/Vandalism of Property, Shoplifting are the major subcategories of Crime Against property. Whereas, Drug/Narcotic Violations, driving Under the Influence are the major subcategories of Crime Against Society. The colours in the bars shows the distribution of crimes over the years, where the distribution seems to be same over the years but we will be analysing the same in detail in this study.

4 ANALYSIS

4.1 Approach

In this section, we discussed about the approach towards analysing the research questions and how the computational and visual techniques with human intervention are used to get better insights. We used Tableau for visualization and generating charts and graphs and used Python for data cleaning and clustering.

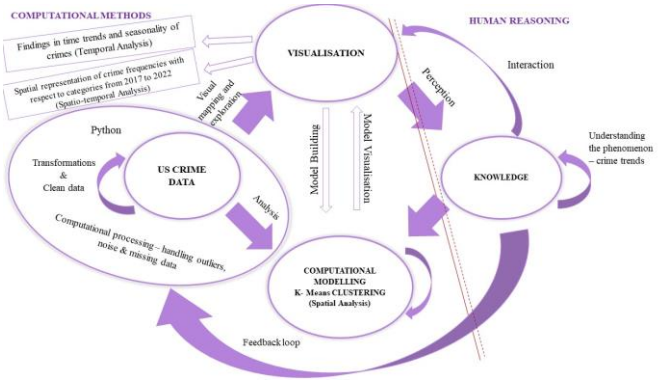

Fig 3: Analysis Workflow Plan

Temporal Approach:

For temporal data, we will be looking for trend patterns, seasonality in the crime frequency as mentioned under,

1. Represent the overall trend of crime frequency in a line graph over the years to see is there a significant temporal variation.
2. Check the trend by month-year pattern in a line graph for detailed understanding on the trend.
3. Plot a heat map with day of the week and hours of the day and check for any seasonality?
4. Plot a heat map with month of the year and day of the month and check for any seasonality?

Spatial Approach:

For the spatial analysis, we will use the latitude and longitude coordinates, cities, crime category and sub category.

1. In order to find the hotspot of the crime, plot the crime frequency data in the map with colour hue to show the intensity of the crime.
2. K Means Clustering:

- In order to cluster the cities with respect to the nature of the crime, do Kmeans clustering with all the types of crime subcategory with crime frequency and cities.
- Fix the no. of clusters with the help of elbow method.
- Plot the clusters in the map to view the cities which are clusters together based on the nature of the crime.
- Take 2 cities in a cluster and plot the crime subcategory distribution in a bar and compare. It is expected that the distribution of crime subcategory in the 2 cities should be similar since it is falling under the same cluster.

Spatiotemporal Analysis:

In order to analyse the crime frequency of each and every category over the years in a spatial representation, plot a crime density map with crime category in y-axis and years on the x-axis.

4.2 Process

Temporal Analysis:

Year-on-year analysis:
The first thing to see in the temporal analysis is the overall trend of the crimes over the years from 2016 to 2022.

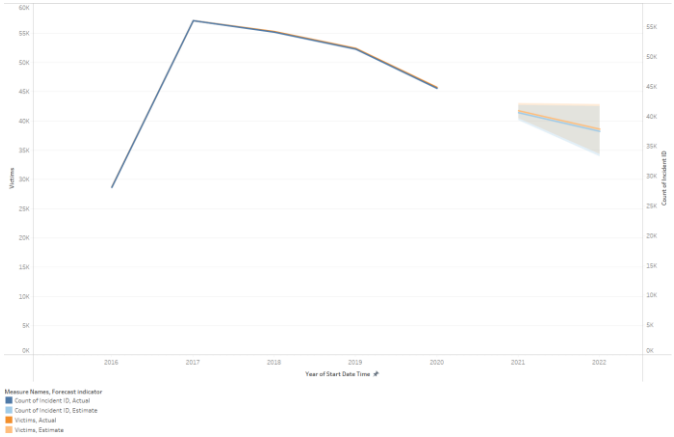

Fig 4: Year on year – Crime Frequency

- In the above graph, the frequency of crime on 2016 is lower because our data is from July-2016, so there are data missing from Jan to Jun-2016. Let us neglect that year in finding the pattern or seasonality.
- In 2017 the crimes were highest and year on year it decreased till 2021 and showing a consistent downtrend. A point to be noted and investigated further in the later stage.
- 2022 data is available till Aug-2022 and value dropped as in 2016 but we forecasted the 2022 crimes with the past data and trend and plot above. The forecast is done in Tableau and the model behind this forecast is exponential smoothing model, where the MASE (Mean Absolute Square Error) is 0.6 which is a good forecast. We are not using the forecasted data for further analysis.

Month-Year analysis:
From fig 4, it is clear that there is a down trend in crimes from 2017 which was in aggregate level. So, digging one step further to find from when (month-year) it has started falling down and is there a continuous drop or fall and rise pattern.

- From fig 5, we can understand that crime frequencies over the months are fluctuating, but we can split that into 2 parts. From start till feb-2019 the average of fluctuations was not that steep down when compared to post feb-2019 and till end. And after feb-19 the frequencies we that that high to match feb-19. This gives a hint that some crime category might be reduced drastically from feb-2019. Further analysis on crime categories gives more insight.

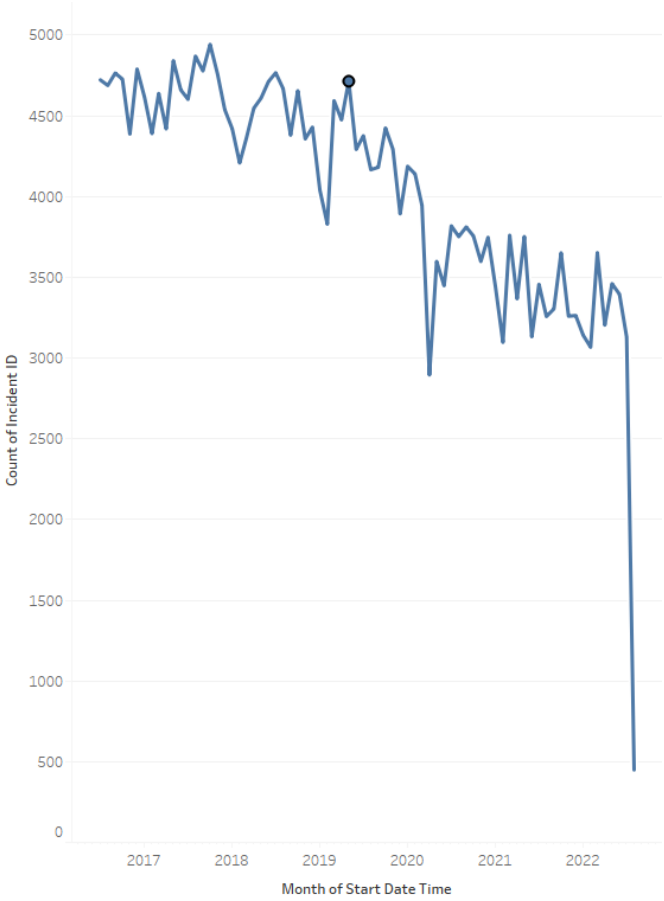

Fig 5: Month-Year – Crime Frequency

Seasonality – Time of the day by Weekday
The below chart is plotted to check for the seasonality or is there a pattern on the time of the day and on the weekday.

- From fig 6, *the seasonality on time of the day* is clearly shown, at 0 hrs (i.e. midnight 12 O'clock) the crimes are at peak.
- Then from 1 to 4 hrs, only in the weekends there is a rise when compared to the weekdays.
- Then from 4 to 8 hrs the crimes are comparatively low.
- From 9hrs it has gradually started picking and by 12hrs it has attained the 2nd peak of the table. Then from 15hrs to 23hrs there crimes are higher.

- From fig 6, *the seasonality on day of the week* is clearly shown that during Fridays there is peak and rest of the week days higher compared to the weekends which is the lowest in the lot.

- From this it is clear that the day/time people travel is directly proportional to the frequency of the crime.
- Particularly, if we note Friday from 15 hrs to 23 hrs are higher when compared to the other days. That might be, because of the people traveling for the weekends.

Start Date Time

Hour of Sta..	Sunday	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday
0	3,041	3,233	2,994	3,302	3,260	3,282	3,267
1	1,625	973	887	996	1,043	1,363	1,824
2	1,655	831	658	779	880	1,065	1,820
3	1,392	627	505	582	522	701	1,334
4	797	470	400	410	485	507	752
5	510	360	375	394	384	428	448
6	475	512	577	546	495	582	473
7	534	880	892	886	892	853	556
8	778	1,381	1,329	1,332	1,283	1,339	795
9	1,145	1,760	1,789	1,738	1,727	1,759	1,310
10	1,182	1,797	1,826	1,764	1,758	1,871	1,405
11	1,292	1,790	1,900	1,986	1,981	1,863	1,523
12	2,000	2,783	2,780	2,780	2,788	2,817	2,371
13	1,492	2,065	2,214	2,174	2,067	2,240	1,694
14	1,657	2,018	2,189	2,249	2,134	2,267	1,739
15	1,903	2,471	2,680	2,586	2,665	2,953	2,081
16	1,932	2,440	2,683	2,606	2,595	2,701	2,107
17	2,149	2,570	2,686	2,711	2,746	2,837	2,136
18	2,092	2,408	2,682	2,663	2,700	2,761	2,206
19	1,918	2,349	2,584	2,678	2,421	2,518	1,969
20	2,009	2,324	2,517	2,613	2,534	2,532	2,033
21	1,926	2,199	2,424	2,460	2,384	2,446	2,085
22	1,891	1,984	2,304	2,381	2,320	2,543	2,119
23	1,651	1,694	2,064	2,016	1,963	2,486	2,120

Count of Incident ID

360 3,302

Fig 6: Time of the day & day of the week – crimes frequency heatmap

Seasonality – Time of the day by Weekday

The below chart is plotted to check for the seasonality or is ther a pattern on the month and day of the month.

Start Date Time

Day of Star..	January	February	March	April	May	June	July	August	Septemb..	October	November	December
1	1,073	792	933	821	902	947	825	920	866	946	1,008	847
2	667	609	630	623	710	651	623	709	554	664	677	637
3	653	664	638	629	728	659	669	662	651	702	663	689
4	659	639	690	641	713	659	530	706	699	657	613	719
5	666	708	698	681	671	698	703	648	615	769	638	685
6	601	727	706	664	679	723	663	670	701	687	660	776
7	628	692	698	621	651	705	682	649	701	640	669	653
8	621	704	700	613	715	688	581	706	625	678	697	654
9	688	668	735	648	724	643	614	690	517	667	632	586
10	726	686	673	647	679	653	692	653	630	739	692	672
11	717	659	685	702	663	615	671	644	636	687	633	699
12	640	683	731	625	598	637	704	613	702	682	632	639
13	579	733	746	697	617	685	703	640	675	673	648	696
14	596	699	633	654	658	706	695	664	722	640	668	651
15	655	748	721	658	760	727	656	671	697	739	759	667
16	679	743	675	605	646	637	633	688	650	687	705	613
17	619	676	716	699	772	660	633	623	677	631	715	707
18	694	591	666	610	674	610	684	661	682	727	671	696
19	659	694	679	645	700	648	748	629	680	668	715	674
20	558	654	716	672	698	673	733	676	773	717	724	674
21	568	760	721	654	715	698	658	638	742	684	668	631
22	674	735	727	625	726	608	656	643	678	656	649	668
23	700	702	704	662	650	644	681	638	684	623	653	662
24	720	681	662	676	642	647	655	588	693	693	601	522
25	704	676	643	633	657	588	716	698	731	722	603	385
26	662	710	674	692	694	626	805	610	664	694	631	553
27	695	721	699	686	661	657	802	641	712	649	640	553
28	637	773	622	620	675	775	653	674	691	687	598	594
29	679	137	697	650	736	722	640	686	702	586	654	559
30	647		665	650	662	655	626	653	704	677	749	548
31	637		614		738		684	721		757		561

Count of Incident ID

137 1,073

Fig 7: Month and day of the month – crimes frequency heatmap

- From fig 7, *the seasonality on day of the month* is clearly shown that at 1st of every month the crimes are at peak when compared to the other days, from an article I understand that the salary credit day is last day of every month, post which there could be some seasonal crimes behind this. Further investigation on the crime subcategory level will give more clarity on the reasons.

- From fig 7, *the seasonality on month* is not that clear and crime frequency does not follow any seasonality or pattern on month. End of December is slightly lower when compared to other days/months which could be because of Christmas and new year holidays.

Spatial Analysis – Distribution of Crimes

In this part we will be analysing the total crimes (from 2016 to 2022) across the cities of the Montgomery County with the below map where the colour intensity shows the frequency of the crimes.

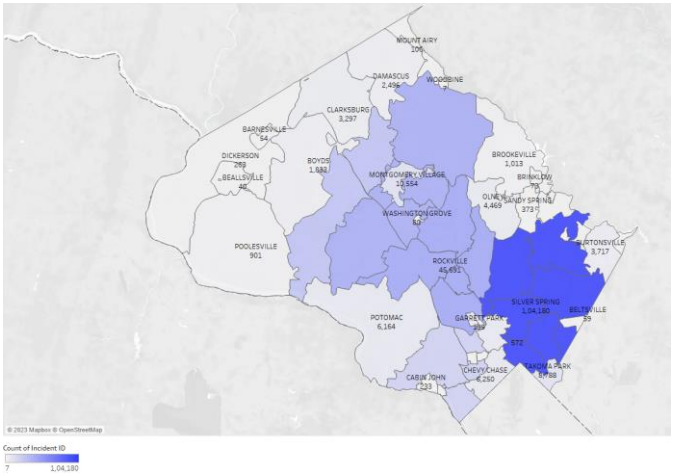

Fig 7: Crime frequency Map

It is clear from the above map that Silver Spring city has the highest crime of all the cities with values 1/3rd of the total crime valued more than 100 thousand crimes and followed by Gaithersburg, Rockville which were having more than 40 thousand crimes over the years.

Spatial Analysis – Kmeans Clustring

As you see in the fig1 and fig2 there are 3 main Crime Category and subsequent Crime Subcategories. Let us cluster the cities based on the nature of the crime (i.e. no. of crimes of each and every crime subcategory with respect to cities).

By preparing the data of no. of crimes for each and every crime subcategory with repect to cities, we have done kmeans clustring. In which with the elbow method we understood that going beyond 4 clusters does not add value. So fixed with 4 clusters and the cities were bifercated into four clusters. The same has been drawn to Tableau to visvalize in map and that is fig8.

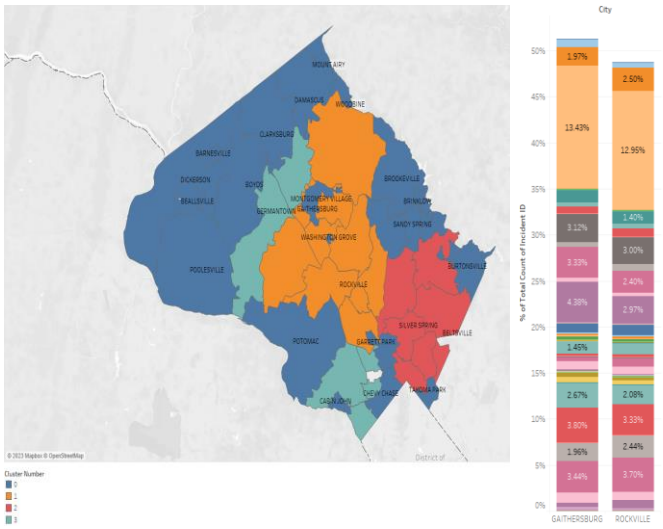

Fig 8: Cluster Map & crime distribution of cities in cluster 1

From fig 8, it is clear that Germantwn and Bethesda belong to cluster 3, Silver Spring city belongs to cluster 2, Rockville and Gaithersburg belong to cluster 1 and remaining all other cities belong to cluster 0. This clustering infers that the cities of same cluster will be having the same nature of crime. For example, the crime subcategories were ploted over the cities of cluster 1 (i.e. Rockville and Gaithersburg) which can be found in the right side of fig 8. where the distribution of crime subcategories of these 2 cities are similar.

Spatiotemporal Analysis:
So far we have analysed the temporal and spatial data separately. In this section we will be analysing the both together. We will be analysing the crime pattern with respect to crime category and years in a spatial representation. Since we are analysing the crime pattern for complete year data we removed 2016 and 2022 which are with missing months.

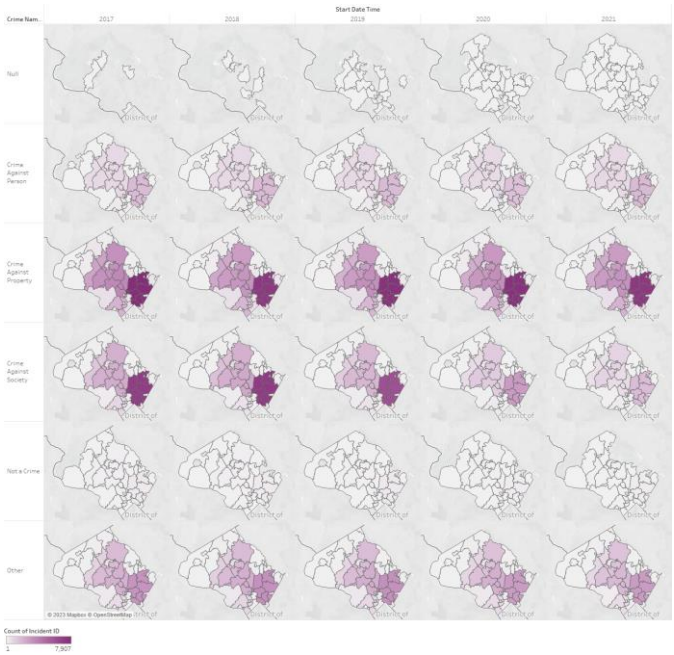

Fig 9: crime pattern with respect to crime category and years map

Let us talk about the 3 main crime categories, the Crime Against Person and Crime Against Property is not showing any evident difference over the years in the map, but if you notice the Crime Against Society there is a drastic decrease in the crime frequency of Silver Spring city on 2020 when compared the previous year and in 2021 a slight decrease when compared to 2020. So, this is could be one of the reasons for overall crime frequency drop over the years and further investigation on the Crime Against Society in Silver Spring city may give much more clarity.

4.3 Results

The frequency of crimes has a year-on-year downtrend from 2017-2021 (fig1) and by analysing one step further with month-year it is evident that there is a drastic fall from 2019-21 when compared to 2017-18 (fig2). The seasonality on the time of the day shows a pattern with peaks on 0 hrs (midnight 12:00), 12:00hrs and a gradual raise from 15:00hrs (fig3). The seasonality on the day of the week shows a clear peak on Friday (fig3). The seasonality on the day of the month shows a clear peak on the 1st of every month, whereas on the months there is not a good seasonal pattern.

Spatially Silver Spring city has the highest crime frequency followed by Gaithersburg, Rockville cities, these three cities contribute 60% of the crimes in the county which is having a same cluster in 3 cluster k means, which means their nature of crime is also the same.

The above information supports the Police Department, when and where to allocate their resources.

The spatiotemporal analysis concludes that Crime Against Society contributes the fall in the frequency of crime on 2020 and 2021 and a further investigation gives the reason which can be implemented somwhere else even.

5 Critical Reflection

At the outset, analysis started with visualizing the crime frequency with all possible features such as, crime category, subcategory, police department name, city, etc [1] to obtain an overview on the subject analysis.
- Further, analysing the temporal part of the data, the chart in the fig3 explained the downtrend of the frequency which raised a question why there is a fall? and led us to dig one step further to see month-year pattern fig4, where we understand that there is a drastic fall from a particular point which raised a question what is the reason for the fall?
- The seasonality charts in the figure 5 showed the peaks for the time of the day and day of the week. The seasonality charts in the figure 6 showed the peak pattern in day of the month. All these graphs require a human intervention to connect the peaks with some external factors or the reasons.
- Visualising hotspots on the map gives us a clear picture where the crimes are more and led us to do further analysis on the that.
- Having different crimes subcategory and cities lead us to form a cluster using the k means clustering in order to cluster the cities based on the nature of the crime.

- by analysing the hotspot and cluster, it led us to concentrate more on the particular area where 60% of the crimes happens.
- As the temporal data says that there is a downtrend in crimes over the years (fig 1 & 2) and spatial analysis giving the hotspot of the crime led us to think how is the crime pattern in spatiotemporal view, so plotted the crime pattern with respect to crime category and years in map (fig9). Where we infered that reason for fall in 2020 ans 2021 which is supporting the temporal downtrend from fig 3 & 4.
- After know that Silver Spring – Crime Against Society is the reason for the fall, we digged furhter and found that Drug/Narcotic Violation, Liquor law violations, Driving under the influence and Disorderly Conduct are the 4 mojaor subcategories contributed for the fall of Crime Against Society in 2020 and 2021.

Future Works:

- Since so many latitudes and longitudes overlaps one over the other, an OPTICS clustring on the high crime cities will give a better view on where particularly to focus in that city.
- We can anlyse the data from the Police Department aspects.

Table of word counts

Problem statement	215
State of the art	115
Properties of the data	399
Analysis: Approach	346
Analysis: Process	1174
Analysis: Results	199
Critical reflection	409

REFERENCES

[1] US Crimes Dataset: All crimes committed between July 2016 to August 2022 in the USA: https://www.kaggle.com/datasets/jgiigii/uscrimesdataset?resource=download&select=Crime.csv

[2] Crime Analysis using k-means Clustering 2017, International Conference on Computational Intelligence and Networks. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8307327

[3] D. Gao, X. Li, C. Yang and Y. Zhang. Spatial patterns analysis of urban road traffic accidents based on GIS, International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), Xiamen, 2012, pp. 1898-1901