Event of Crime against Property: Robbery & Theft Prediction using Probabilistic Graphical model

♦ **Nico Duran, July 2019** ♦
nlduran@mymail.mapua.edu.ph
School of Information Technology, Mapua University, NCR, Philippines

I. Introduction

A crime is an act committed or omitted in violation of a law forbidding or commanding it and for which a punishment is imposed upon conviction.[1] Crimes are categorized into *index* and *non-index* crimes to create a standard definition of crime classification.[1] Index crimes involves *crimes against persons* and *crimes against property*; index crimes are considered to be serious and happens with regularity across cities .[1][6] Non-index crimes are violations of special laws such as local ordinances, etc..[1] The Philippine National Police (PNP) focuses on seven types of (index) crimes: (1) murder, (2) homicide, (3) physical injury, (4) rape, (5) robbery, (6) theft, (7) carnapping.[2] Robbery and theft tops the list of *crimes against property*,[3] and will be the focus on this study.

Crime pattern theory stipulates that crime does not occur randomly, it is either planned or opportunistic; the theory states that incidents are patterned with respect to time and space, occurring where the activity of the offender and victim/target intersect.[10] The current (assumed) PNP operations on patrolling is statically scheduled, routed, and resourced; it can be optimized to contribute on reducing crime index as police visibility results in lower estimates of subjective risks of victimization[12]. To establish a good public perception of safety, patrolling can be (theoretically) scheduled and routed in line with the probability of crime happening in a given area and time, and in harmony with police resources. Since crimes does not occur at random, this study believes that a pattern can be learned based from historical data of (recorded) crimes; probabilistic models can be used to represent, and infer from the given data.

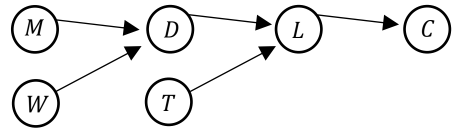
Probabilistic graphical models use a graph-based representation as the basis for compactly encoding a complex distribution over a high-dimensional space.[9] The nodes on probabilistic graph models store probabilities from where it is easy to input evidence and output updated information. Direct relationship between variables give more information about complicated relationship. A Bayesian network is a directed graphical model that is constructed based on two main components: (1) nodes/vertices, and (2) arcs/directed edges. A node can represent propositional variable of interest that possess mutually exclusive states of associated probability distribution.[4] The probability distribution is in the form of *node probability table*. Probability theory states that the sum of probabilities associated with each state of a variable must be in unity.[11]

This study aims to infer marginal/conditional probability of robbery & theft on locations within Manila given the evidence of crime happening derived from intrinsic (conditional) probability distribution with respect to month, week of month, day of week, and time of day. The study proposes a discrete Bayesian network model for representation, and inference from the given crime data. The resulting network is intended (only) to assist PNP on their operations such as: patrolling – schedule and route prioritization, and resource management – pre-emptive response strategies.

II. Methodology

The design method for modeling crime instance in a Bayesian network is described in this section. The goal of the method is to model factors that determines a crime happening in one Bayesian network. A 5-year data of robbery & theft in Manila (with conviction) was used to determine intrinsic probability distribution of factors. *Figure 1* represent the proposed model (representation) and its joint probability for a robbery & theft incident.

M = Month
W = Week of Month
D = Day of Week
T = Time of Day
L = Crime location
C = Crime



$$P(M, W, D, T, L, C) = P(M)\, P(W)\, P(D \mid M, W)\, P(T)\, P(L \mid T, D)\, P(C \mid L)$$

Figure 1

Each factor in a Bayesian network represents a conditional probability distribution (CPD) that can be derived from the given data. *Table 1- 6* shows CPD for each factor. *Table 6* is derived with respect to the population of the given location (*Table 7*), instead of its intrinsic value as it is more logical to model the probability of a crime (rate) happening in a location given its population.

Table 1 - CPD for M:
Crime happening on a Month (ϕ)

January	0.086168
February	0.106576
March	0.074830
April	0.068027
May	0.070295
June	0.083900
July	0.106576
August	0.081633
September	0.072562
October	0.063492
November	0.086168
December	0.099773

$P(M)$

Table 2- CPD for W:
Crime happening on a Week of Month(ϕ)

1st Week	0.185941
2nd Week	0.283447
3rd Week	0.226757
4th Week	0.235828
5th Week	0.068027

$P(W)$

Table 3 - CPD for T
Crime happening on a Time of Day(ϕ)

12 AM – 3 AM	0.095238
3 AM – 6 AM	0.099773
6 AM – 9 AM	0.070295
9 AM – 12 PM	0.156463
12 PM – 3 PM	0.163265
3 PM – 6 PM	0.170068
6 PM – 9 PM	0.136054
9 PM – 12 AM	0.108844

$P(T)$

Table 4 - CPD for D:
Crime happening on a Day on a Week of Month (ϕ)

	Sunday	Monday	Tuesday	Wednesday	...	Saturday
1st Week of January	0.125000	0.000000	0.250000	0.125000	...	0.250000
2nd Week of January	0.000000	0.000000	0.500000	0.000000	...	0.250000
3rd Week of January	0.076923	0.076923	0.153846	0.153846	...	0.153846
4th Week of January	0.090909	0.000000	0.090909	0.272727	...	0.181818
5th Week of January	0.000000	0.000000	0.000000	0.000000	...	0.000000
1st Week of February	0.200000	0.000000	0.000000	0.300000	...	0.400000
2nd Week of February	0.142857	0.214286	0.071429	0.285714	...	0.000000
3rd Week of February	0.090909	0.000000	0.090909	0.272727	...	0.090909
4th Week of February	0.083333	0.083333	0.083333	0.416667	...	0.166667
5th Week of February	0.142857	0.142857	0.142857	0.142857	...	0.142857
1st Week of March	0.250000	0.250000	0.000000	0.000000	...	0.000000
2nd Week of March	0.066667	0.133333	0.333333	0.200000	...	0.000000
3rd Week of March	0.000000	0.200000	0.200000	0.100000	...	0.300000
4th Week of March	0.250000	0.000000	0.000000	0.250000	...	0.500000
5th Week of March	0.142857	0.142857	0.142857	0.142857	...	0.142857
1st Week of April	0.000000	0.000000	0.666667	0.333333	...	0.000000
2nd Week of April	0.000000	0.083333	0.083333	0.166667	...	0.333333
3rd Week of April	0.166667	0.166667	0.166667	0.166667	...	0.166667
...	...	...	...	...	...	...
5th Week of December	0.125000	0.375000	0.000000	0.125000	...	0.250000

$P(D \mid M, W)$

Table 5 - CPD for L:
Crime happening in a Location on a Time of Day (ϕ)

		Binondo, Manila	Ermita, Manila	Intramuros, Manila	Malate, Manila	…	Tondo, Manila
12 AM - 3 AM	Sunday	0.000000	0.166667	0.000000	0.000000	…	0.333333
3 AM - 6 AM	Sunday	0.000000	0.000000	0.000000	0.500000	…	0.250000
6 AM - 9 AM	Sunday	0.000000	0.000000	0.000000	0.000000	…	0.000000
9 AM - 12 PM	Sunday	0.000000	0.285714	0.000000	0.285714	…	0.000000
12 PM - 3 PM	Sunday	0.000000	0.000000	0.000000	0.125000	…	0.625000
3 PM - 6 PM	Sunday	0.000000	0.100000	0.100000	0.000000	…	0.300000
6 PM - 9 PM	Sunday	0.000000	0.200000	0.000000	0.100000	…	0.100000
9 PM - 12 AM	Sunday	0.000000	0.142857	0.000000	0.142857	…	0.142857
12 AM - 3 AM	Monday	0.000000	0.500000	0.000000	0.000000	…	0.000000
3 AM - 6 AM	Monday	0.000000	0.000000	0.000000	0.000000	…	0.100000
6 AM - 9 AM	Monday	0.000000	0.000000	0.500000	0.000000	…	0.000000
9 AM - 12 PM	Monday	0.000000	0.285714	0.000000	0.000000	…	0.000000
12 PM - 3 PM	Monday	0.000000	0.125000	0.000000	0.000000	…	0.500000
3 PM - 6 PM	Monday	0.000000	0.333333	0.000000	0.166667	…	0.000000
6 PM - 9 PM	Monday	0.142857	0.000000	0.000000	0.000000	…	0.285714
9 PM - 12 AM	Monday	0.000000	0.000000	0.000000	0.400000	…	0.200000
12 AM - 3 AM	Tuesday	0.000000	0.250000	0.000000	0.000000	…	0.000000
3 AM - 6 AM	Tuesday	0.000000	0.000000	0.000000	0.000000	…	0.400000
…	…	…	…	…	…	…	…
9 PM - 12 AM	Saturday	0.000000	0.375000	0.000000	0.125000	…	0.000000

$P(L \mid T, D)$

Table 6 - CPD for C:
Crime happening in a Location (ϕ)

	Robbery & Theft	No Robbery & Theft
Binondo, Manila	0.000539	0.999461
Ermita, Manila	0.006652	0.993348
Intramuros, Manila	0.000842	0.999158
Malate, Manila	0.000525	0.999475
Paco, Manila	0.000188	0.999812
Pandacan, Manila	0.000487	0.999513
Port Area, Manila	0.000075	0.999925
Quiapo, Manila	0.000864	0.999136
Sampaloc, Manila	0.000084	0.999916
San Andres, Manila	0.000103	0.999897
Sta. Ana, Manila	0.000084	0.999916
Sta. Cruz, Manila	0.000699	0.999301
Sta. Mesa, Manila	0.000045	0.999955
Tondo, Manila	0.000146	0.999854

$P(C \mid L)$

Table 7
Population in a given location

Location	Population
Binondo, Manila	12985
Ermita, Manila	10523
Intramuros, Manila	5935
Malate, Manila	78132
Paco, Manila	69300
Pandacan, Manila	82194
Port Area, Manila	66742
Quiapo, Manila	23138
Sampaloc, Manila	395111
San Andres, Manila	116998
Sta. Ana, Manila	178769
Sta. Cruz, Manila	118779
Sta. Mesa, Manila	110073
Tondo, Manila	631363

$\text{Crime\_CPD} = \frac{\textit{crime count}}{\textit{location population}}$ → **as seen on Table 6**

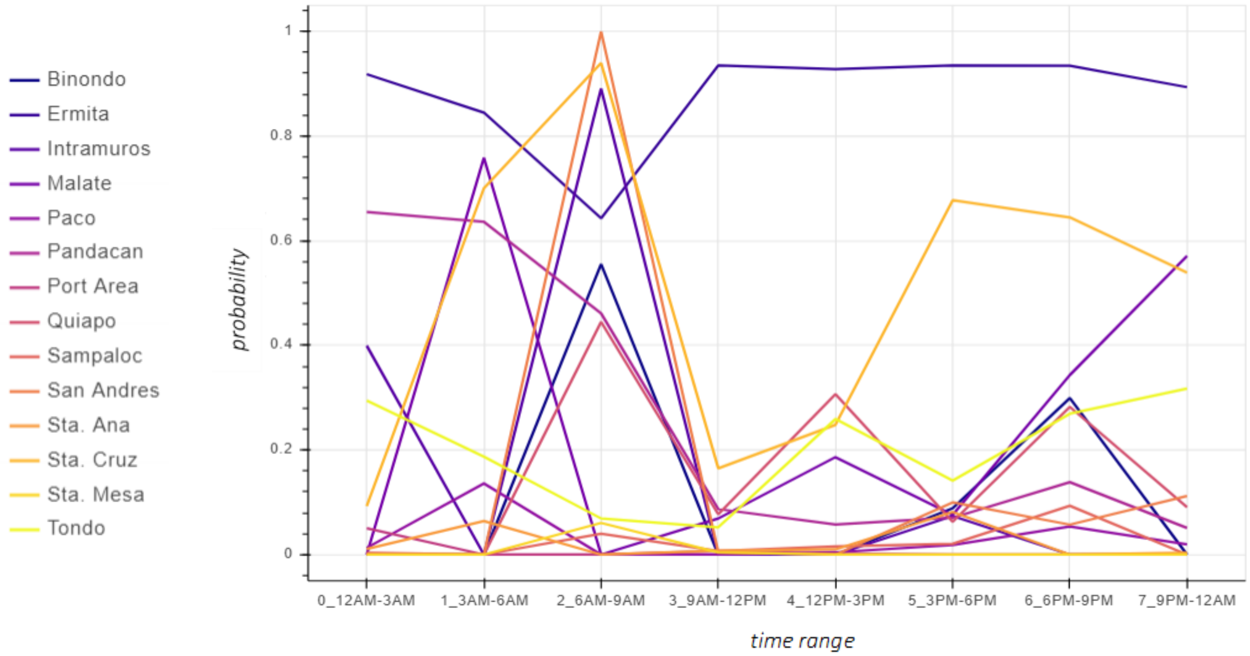
The model structure should allow the distribution to be used effectively for inference. A python implementation (leveraging on existing python package: pgmpy) of *variable elimination* will be used to compute for probability (*Equation 1*) of a crime happening at a location given the evidence of day, and time. The results generated from combining variables were achieved and can be seen on *Appendix i*.

$$P(L \mid T, D) = \Sigma_M \Sigma_W \Sigma_C\, P(M)\, P(W)\, P(D \mid M, W)\, P(T)\, P(L \mid T, D)\, P(C \mid L)$$

Equation 1: Variable elimination for marginal/conditional probability of L given T, D

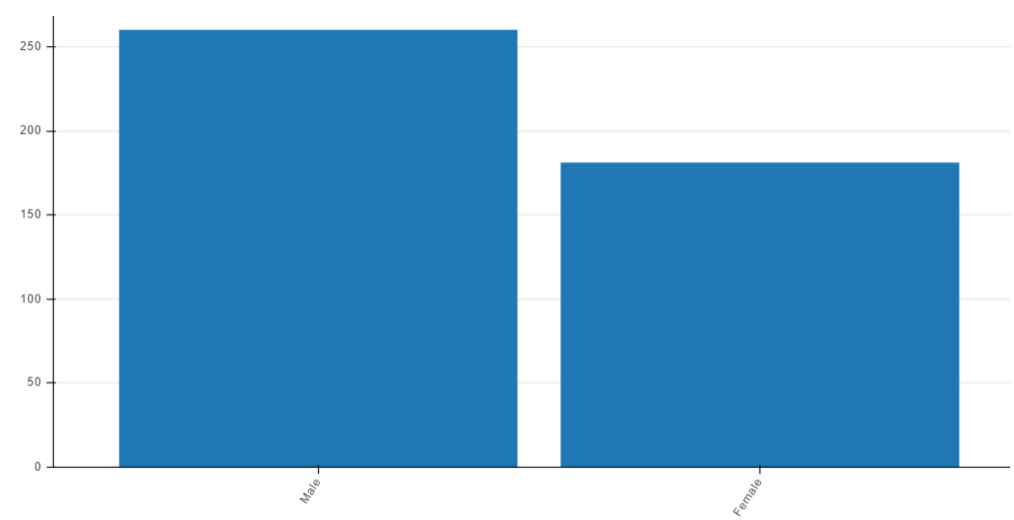
The probability distribution of robbery and theft happening on a location is plotted in a line graph as seen on *Figure 1*, having time ranges of: '12AM-3AM', '3AM-6AM', '6AM-9AM', '9AM-12PM', '12PM-3PM', '3PM-6PM', '6PM-9PM', '9PM-12AM' respectively. The line graph breakdown per day can be seen on *Appendix ii*.

Figure 1: Line graph (Grouped)



The likelihood of victim's gender can be inferred by counting the instance of a gender for every crime committed within our data (intrinsic). The results were determined in *Figure 2* considering crime total count. Victim's gender likelihood breakdown per location can be seen on *Appendix iii*.

Figure 2: Bar graph : victim gender



Understanding Bayes network have some pre-requisites. If there are confusions regarding probabilistic graph modeling in general, check out reference [9]. To see the full python implementation of this methodology, please refer to: https://github.com/goodguynic/probabilistic_graphical_model/blob/master/Crimes_PGM.ipynb

III. Results & Discussion

The prioritization of PNP operations on patrolling within Manila can be inferred from the line charts on *Appendix ii*; it is summarized in *Table 8*. The corresponding location that peaks the line graph at a given time range is the location of interest for that time range and day to be prioritized for patrolling.

The recommended locations for patrol route or police resource deployment (pre-emptive) are represented through a table with *time range* as columns, and *days of week* as index.

Table 8: Patrol priority locations

	12AM-3AM	12PM-3PM	3AM-6AM	3PM-6PM	6AM-9AM	6PM-9PM	9AM-12PM	9PM-12AM
Monday	Ermita	Ermita	Sta. Cruz	Ermita	Intramuros	Sta. Cruz	Ermita	Sta. Cruz
Tuesday	Ermita	Ermita	Sta. Cruz	Ermita	Sta. Cruz	Ermita	Ermita	Ermita
Wednesday	Pandacan	Ermita	Ermita	Ermita	Binondo	Ermita	Ermita	Ermita
Thursday	Pandacan	Ermita	Sta. Cruz	Sta. Cruz	Ermita	Malate	Ermita	Ermita
Friday	Ermita	Ermita	Pandacan	Ermita	Quiapo	Malate	Ermita	Malate
Saturday	Ermita	Ermita	Ermita	Ermita	Pandacan	Sta. Cruz	Ermita	Ermita
Sunday	Ermita	Quiapo	Malate	Ermita	San Andres	Ermita	Ermita	Ermita

The likelihood of victim's gender can be inferred from *Appendix iii.* The result is summarized in *Table 9*. Overall, the *male* population are more likely to be a victim of robbery and/or theft as observed in *Figure 2*. It would be more interesting to know the distribution of male and female per location, and profession (classification), to gain more insights of the population's (extrinsic) probabilities of being a victim of crime.

Table 9: Victim's gender likelihood per location

Location	Female	Male
Binondo, Manila	0.29	0.71
Ermita, Manila	0.37	0.63
Intramuros, Manila	0.40	0.60
Malate, Manila	0.22	0.78
Paco, Manila	0.46	0.54
Pandacan, Manila	0.53	0.48
Port Area, Manila	0.20	0.80
Quiapo, Manila	0.50	0.50
Sampaloc, Manila	0.48	0.52
San Andres, Manila	0.50	0.50
Sta. Ana, Manila	0.33	0.67
Sta. Cruz, Manila	0.41	0.59
Sta. Mesa, Manila	0.20	0.80
Tondo, Manila	0.46	0.54

IV. Conclusion

Given the assumptions on PNP operations regarding static patrolling and on-demand police resource deployment; this study would be relevant to optimize patrolling routes within Manila and enable pre-emptive police response on crimes (robbery/theft). The proposed model can be updated with concurrent conditional probability distributions of its factors to ensure relevance. This proof-of-concept may be operationalized to be evaluated, calibrated, and secure timely updating of CPDs; it is always better to have more quality data to be confident on a Bayesian network inference.

The data that was used for the Bayesian network are crime records for robbery & theft with conviction; this provided enough information to generate inference. This study could be a benchmark on queries that would involve not only crime records with conviction, but also with reported crimes that did not make it until conviction due to some legal or moral (or other) factors. Some crimes are not filed by victims, some are arranged by the offender with the victim before formal crime report, and some are just false report, etc.

More sophisticated Bayesian network are expected to handle different cases given the right data for representation, inferring, and learning. Data strategies should be implemented on future studies or operationalization, especially on what periods should be covered/considered for factor's CPDs; we may consider monthly, quarterly, or yearly data coverage for CPDs.

As stipulated from *Crime Pattern Theory*, crimes do not occur randomly but rather with pattern, it may relate geographically, socially, etc..[10] This study was able to infer a robbery and/or theft happening within Manila (location) given time; the model was able to infer a weekly pattern for robbery and/or theft incident . For future studies, if given more (quality) data, the researchers may be able to model crime (not limited to robbery & theft) incident with more factors. The researchers may be able to infer unusual crime pattern (geographic, social, etc.), and infer seasonality, trend, etc.. It would be interesting to pin down a syndicate (or other offenders) using a Bayesian network as part of PNP intelligence tool.

V. References

[1] Senate of the Philippines, Philippine Crime statistics , June 2013, AG-13-03

[2] Jeff Canoy, ABS-CBN news, online article, February 2017

[3] Erwin Colcol/KG, GMA news, online article, July 2018

[4] F. V. Jensen, T. D. Nielsen, Bayesian Networks and Decision Graphs, 2nd Edition, Springer, New York, 2007

[5] P. Juchli, A. Biedermann, F. Taroni, Graphical probabilistic analysis of evidence item combination, Lausanne-Dorigny,Switzerland,2011

[6] R. Talabong, Rappler media online article, August 2018

[7] A. Merenez, ABS-CBN News, online article, June 2018

[8] F. W. Winkel, Reducing Fear Of Crime Through Police Visibility: A Field Experiment, December 1986 (https://doi.org/10.1177/088740348600100402)

[9] D. Koller, N. Friedman, Probabilistic Graphical Models principles & tchniques, The MIT press, 2009

[10] P. J. Brantingham, P. L. Brantingham, Crime Pattern Theory, January 2013

[11] S. M. Ross, A First course in probability, Pearson Prentice Hall, 2010

VI. Appendix

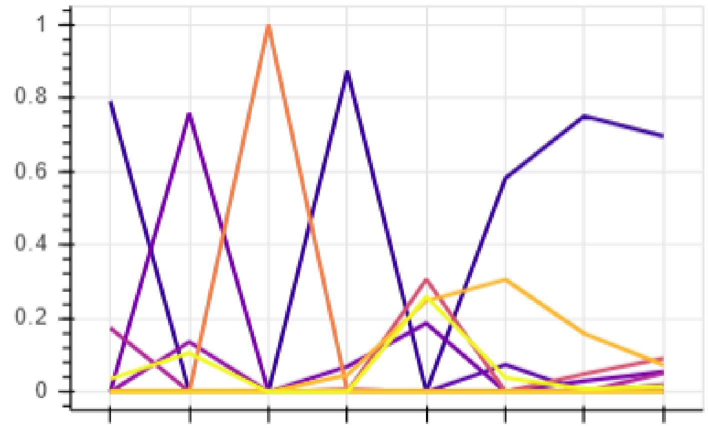
i. Probability distribution given time

month	week_ofmonth	day_of_week	time_range	Binondo	Ermita	Intramuros	Malate	...	Tondo
January	1st week	Sunday	12AM-3AM	0	0.791434	0	0	...	0.034741
January	1st week	Sunday	3AM-6AM	0	0	0	0.758671	...	0.105491
January	1st week	Sunday	6AM-9AM	0	0	0	0	...	0
January	1st week	Sunday	9AM-12PM	0	0.874573	0	0.069024	...	0
January	1st week	Sunday	12PM-3PM	0	0	0	0.186302	...	0.259049
January	1st week	Sunday	3PM-6PM	0	0.58213	0.073685	0	...	0.03833
January	1st week	Sunday	6PM-9PM	0	0.750832	0	0.029629	...	0.00824
January	1st week	Sunday	9PM-12AM	0	0.695743	0	0.054911	...	0.01527
January	1st week	Monday	12AM-3AM	0	0.91815	0	0	...	0
January	1st week	Monday	3AM-6AM	0	0	0	0	...	0.029253
January	1st week	Monday	6AM-9AM	0	0	0.891005	0	...	0
January	1st week	Monday	9AM-12PM	0	0.784249	0	0	...	0
January	1st week	Monday	12PM-3PM	0	0.71274	0	0	...	0.062574
January	1st week	Monday	3PM-6PM	0	0.895229	0	0.035327	...	0
January	1st week	Monday	6PM-9PM	0.299278	0	0	0	...	0.162132
January	1st week	Monday	9PM-12AM	0	0	0	0.40478	...	0.056284
January	1st week	Tuesday	12AM-3AM	0	0.839687	0	0	...	0
January	1st week	Tuesday	3AM-6AM	0	0	0	0	...	0.18694
January	1st week	Tuesday	6AM-9AM	0	0	0	0	...	0
January	1st week	Tuesday	9AM-12PM	0	0.93497	0	0.024597	...	0.013681
January	1st week	Tuesday	12PM-3PM	0	0.787778	0	0	...	0.051871
January	1st week	Tuesday	3PM-6PM	0	0.934828	0.029582	0	...	0.005129
January	1st week	Tuesday	6PM-9PM	0	0.808295	0	0.021265	...	0.017741
January	1st week	Tuesday	9PM-12AM	0	0.794458	0	0.188105	...	0.017437
January	1st week	Wednesday	12AM-3AM	0	0	0.399431	0	...	0.13852
January	1st week	Wednesday	3AM-6AM	0	0.76884	0	0	...	0.050624
January	1st week	Wednesday	6AM-9AM	0.555098	0	0	0	...	0
January	1st week	Wednesday	9AM-12PM	0	0.924949	0	0.05475	...	0.020301
January	1st week	Wednesday	12PM-3PM	0	0.816598	0	0.064449	...	0.035846
January	1st week	Wednesday	3PM-6PM	0	0.481821	0	0.076053	...	0.0423
January	1st week	Wednesday	6PM-9PM	0	0.934729	0	0	...	0.010258
January	1st week	Wednesday	9PM-12AM	0	0.835783	0	0.065963	...	0
January	1st week	Thursday	12AM-3AM	0	0	0	0	...	0.294553
January	1st week	Thursday	3AM-6AM	0	0	0	0.401376	...	0
January	1st week	Thursday	6AM-9AM	0	0.642892	0.081376	0	...	0
January	1st week	Thursday	9AM-12PM	0	0.870301	0	0.022896	...	0
January	1st week	Thursday	12PM-3PM	0	0.928013	0	0	...	0
January	1st week	Thursday	3PM-6PM	0	0	0	0	...	0.141473
January	1st week	Thursday	6PM-9PM	0.176201	0	0	0.343249	...	0.143184
January	1st week	Thursday	9PM-12AM	0	0.796838	0	0.062889	...	0.017489
January	1st week	Friday	12AM-3AM	0	0.882072	0	0	...	0
January	1st week	Friday	3AM-6AM	0	0	0	0.137183	...	0
January	1st week	Friday	6AM-9AM	0.188198	0	0	0	...	0.050978
January	1st week	Friday	9AM-12PM	0	0.593877	0	0.046871	...	0.052138
January	1st week	Friday	12PM-3PM	0	0.779973	0	0.061558	...	0.025679
January	1st week	Friday	3PM-6PM	0.088513	0.546188	0	0.043107	...	0.071927
January	1st week	Friday	6PM-9PM	0	0	0	0.298465	...	0.041501
January	1st week	Friday	9PM-12AM	0	0	0	0.570652	...	0.317391
January	1st week	Saturday	12AM-3AM	0	0.908867	0	0	...	0.039896
January	1st week	Saturday	3AM-6AM	0	0.844913	0	0	...	0.018544
January	1st week	Saturday	6AM-9AM	0	0	0	0	...	0.069096
...	...	...	...	...	...	...	...	...	...
December	5th week	Saturday	9PM-12AM	0	0.893486	0	0.023506	...	0

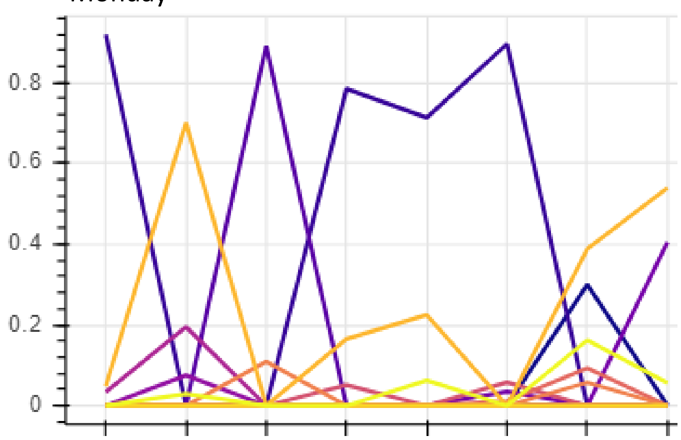
see attachments or access: https://github.com/goodguynic/probabilistic_graphical_model/blob/master/crime_location_CPD_results.csv

ii. Line graph breakdown per day

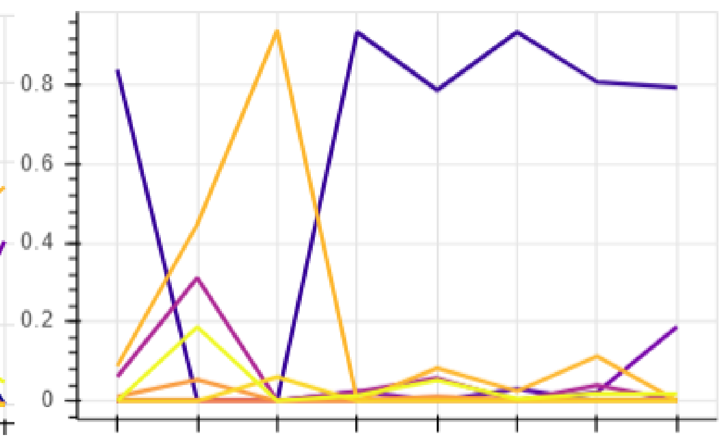
Sunday



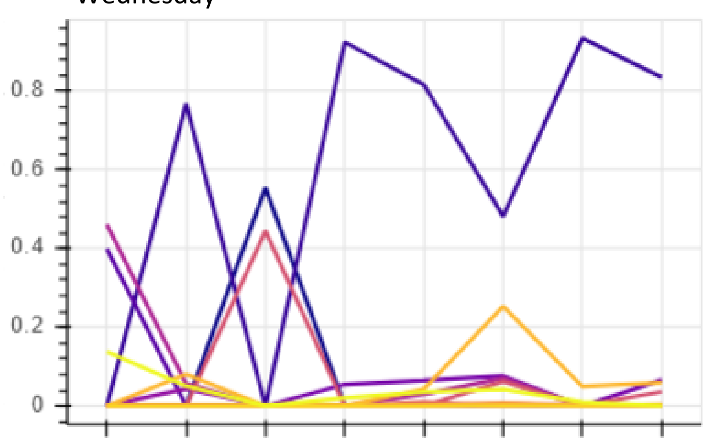
Monday



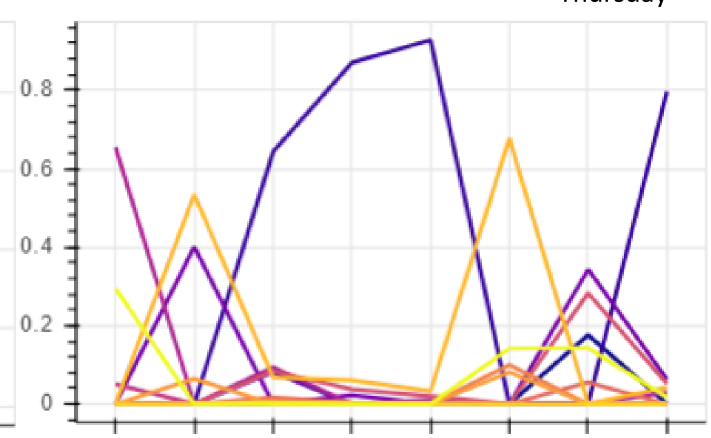
Tuesday



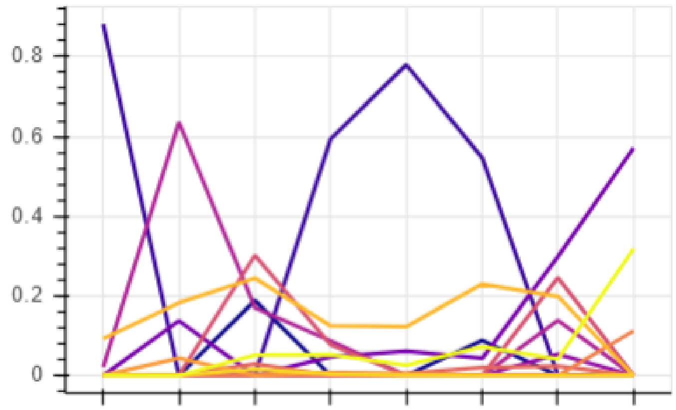
Wednesday



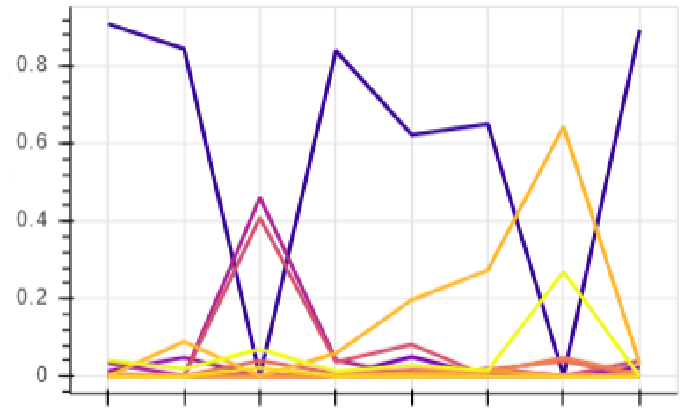
Thursday



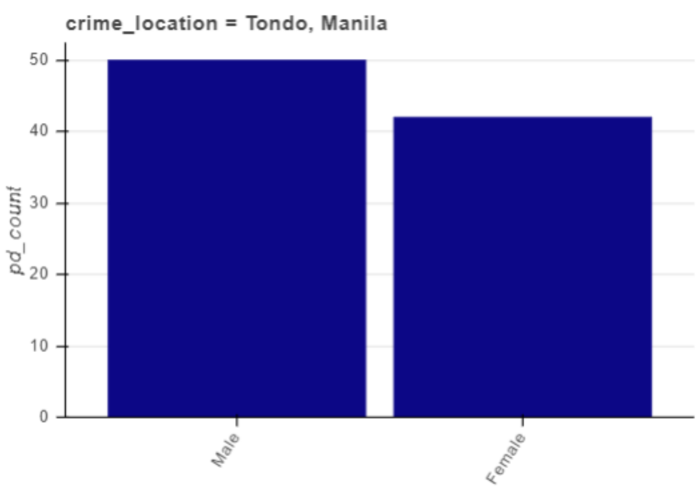
Friday

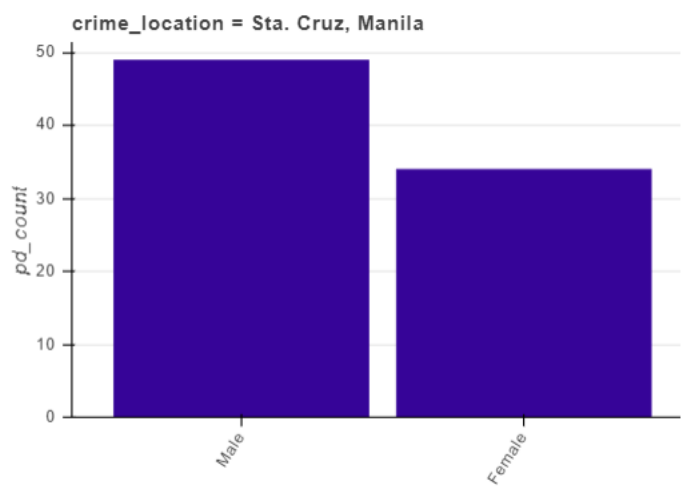


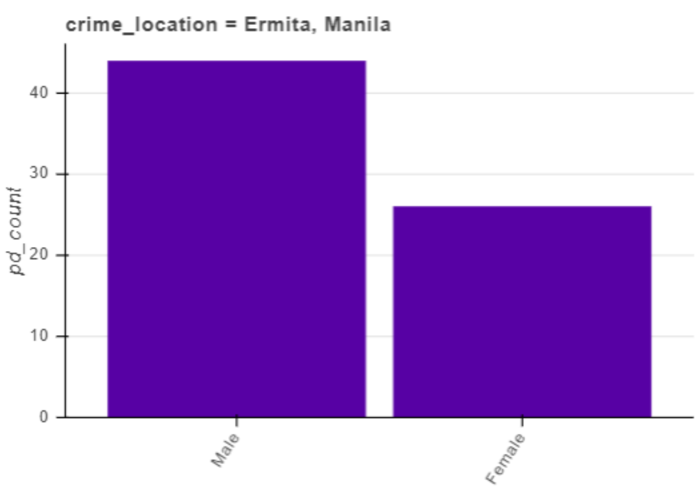
Saturday

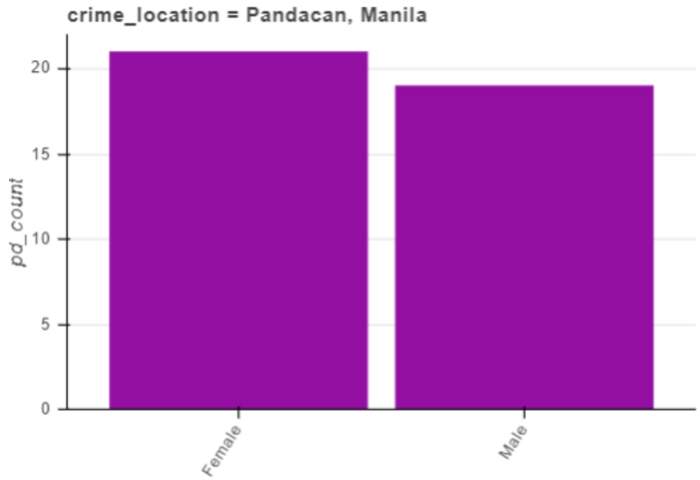


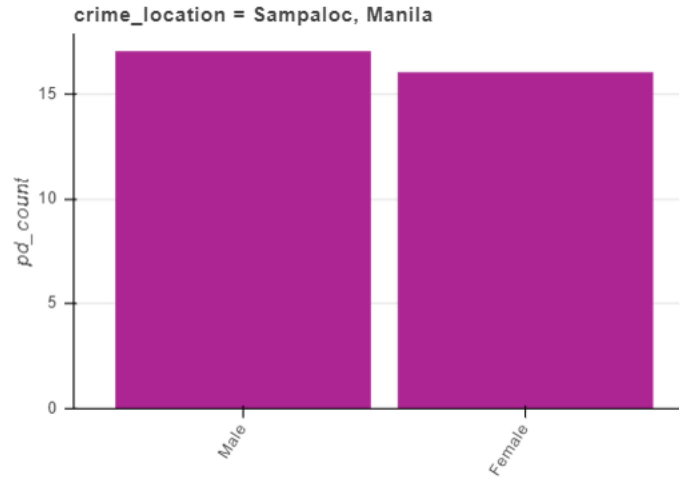
iii. Bar graph – victim gender likelihood per location

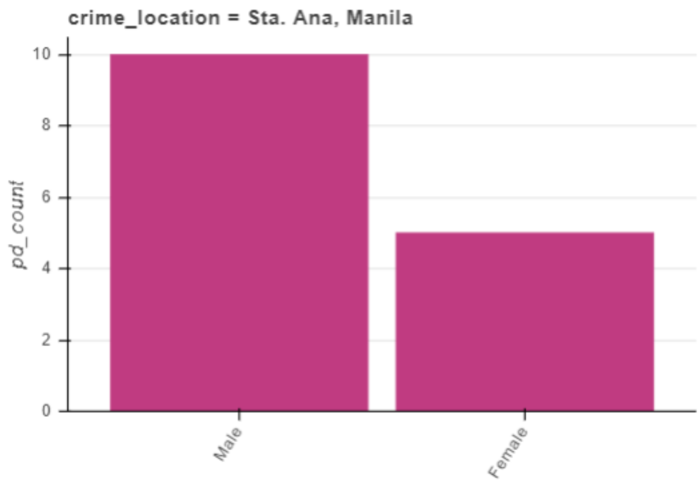


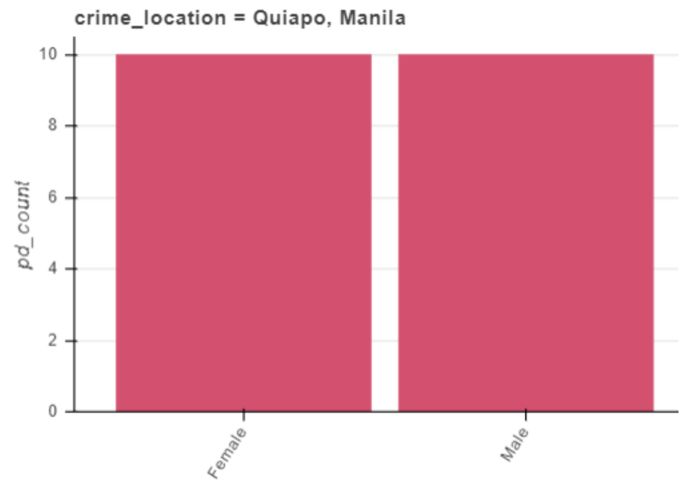


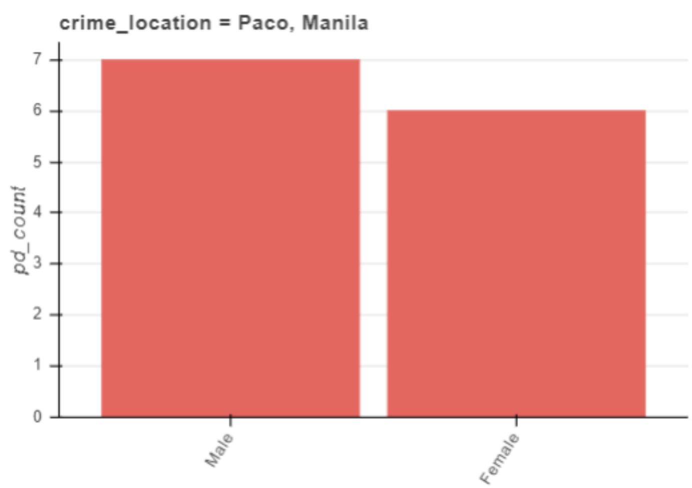


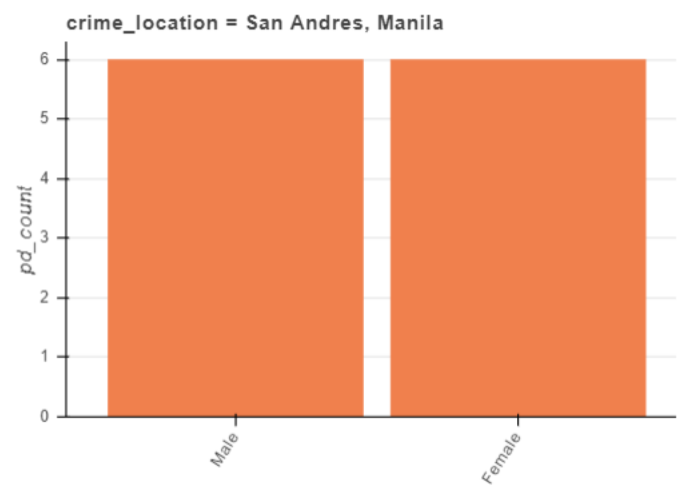

crime_location = Pandacan, Manila
pd_count
0
5
10
15
20
Female
Male

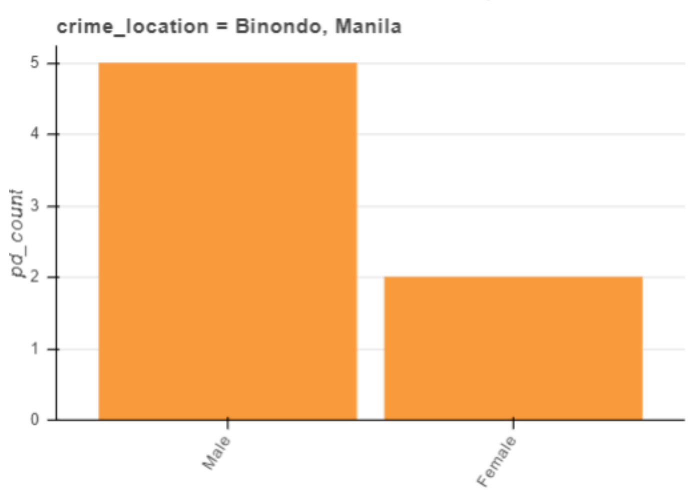

crime_location = Sampaloc, Manila
pd_count
0
5
10
15
Male
Female

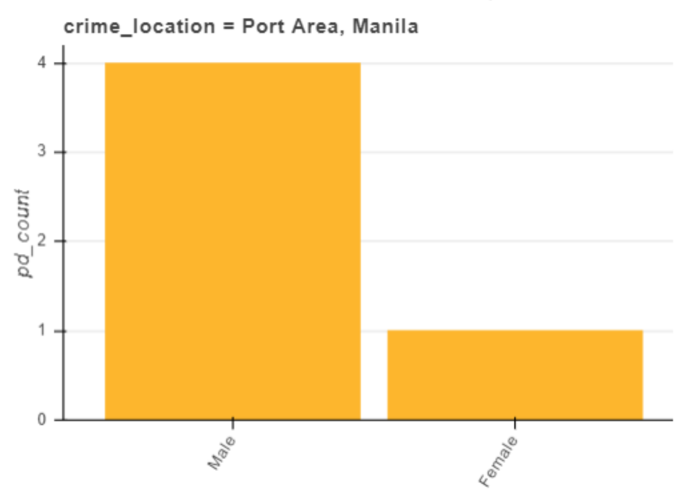

crime_location = Sta. Ana, Manila
pd_count
0
2
4
6
8
10
Male
Female

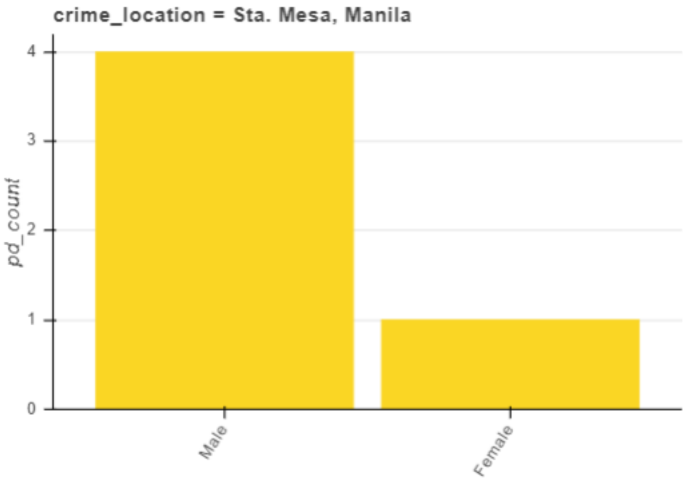

crime_location = Quiapo, Manila
pd_count
0
2
4
6
8
10
Female
Male

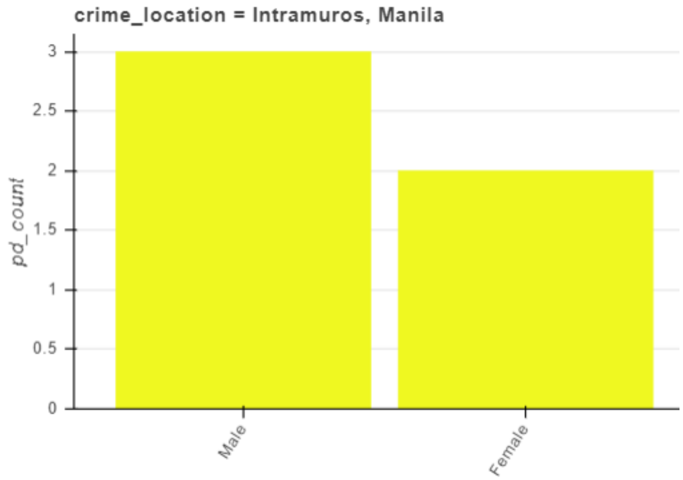

crime_location = Paco, Manila
pd_count
0
1
2
3
4
5
6
7
Male
Female


crime_location = San Andres, Manila
pd_count
0
1
2
3
4
5
6
Male
Female


crime_location = Binondo, Manila
pd_count
0
1
2
3
4
5
Male
Female


crime_location = Port Area, Manila
pd_count
0
1
2
3
4
Male
Female


crime_location = Sta. Mesa, Manila
pd_count
0
1
2
3
4
Male
Female


crime_location = Intramuros, Manila
pd_count
0
0.5
1
1.5
2
2.5
3
Male
Female