

Chicago Crime Data Analysis

Aomkar Mathakar - 015925013

Mugdha Gumphekar - 015786238

Sandesh Gupta - 015649036

Venkata Prithvi Raj Namburi - 015944994

Dataset

Number of records: 7,424,694

Number of attributes: 22

Sample Attributes:

Crime Type	Description	Location Description	Arrest	Domestic	District
Ward	Community area	Date	Latitude	Longitude	Block

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/data>

Data mining methods used

1. Association rule mining
 - a. Apriori algorithm
 - b. Efficient Apriori
 - c. FP Growth
2. Clustering methods
 - a. A
 - b. B
 - c. C
 - d. D

Association Rule Mining - Analysis

1. Analyse 2021 crime data
 - a. Dates: 01/01/2021 - 11/10/2021
 - b. Total 158084 records

2. Analyse Pre and Post Covid data
 - a. Dates: Pre-covid (03/20/2018 - 03/20/2020), Post- covid (03/21/2020 - 10/11/2021)
 - b. Total records: Pre-covid: 528391 and Post-covid: 319531

Association Rule Mining - Preprocessing

Preprocessing methods:

1. Drop redundant or unuseful columns
2. Add new column
3. Convert the record data to transaction data

Association Rule Mining - Algorithms used

1. Apriori algorithm
2. Efficient Apriori
3. FP Growth

Concepts:

- Support
- Confidence
- Lift

Association Rule Mining - 2021 data analysis

Support	Confidence	Lift	Number of rules generated	Time taken
0.05	0.9	1	85	16s
0.05	0.8	1	110	14s
0.10	0.8	1	18	2s
0.10	0.9	1	11	2s
0.20	0.9	1	1	50ms
0.5	0.9	1	0	34ms

Association Rule Mining - Algorithm Comparison

	Apriori	Efficient Apriori	FP Growth
Support	0.05	0.05	0.05
Confidence	0.9	0.9	0.9
Lift	1	1	1
Time taken	16s	0.84s	1.8s

Association Rule Mining - Itemset Generation

Min_support = 0.05, min_confidence = 0.9, min_lift = 1

	A	B	C	D	E
1	antecedents	consequents	support	confidence	lift
20	'Description-DOMESTIC BATTERY SIMPLE', 'Arrest-False'	'Domestic-True'	0.08	0.99	4.43
21	'Description-DOMESTIC BATTERY SIMPLE', 'Primary Type-BATTERY', 'Arrest-False'	'Domestic-True'	0.08	0.99	4.43
22	'Domestic-False', 'Description-OVER \$500'	'Primary Type-THEFT', 'Arrest-False'	0.05	0.99	5.44
23	'Domestic-False', 'Description-OVER \$500'	'Arrest-False'	0.05	0.99	1.13
24	'Domestic-False', 'Primary Type-THEFT', 'Description-OVER \$500'	'Arrest-False'	0.05	0.99	1.13
25	'Description-OVER \$500'	'Primary Type-THEFT', 'Arrest-False'	0.06	0.99	5.43
26	'Description-OVER \$500'	'Arrest-False'	0.06	0.99	1.13
27	'Description-OVER \$500', 'Primary Type-THEFT'	'Arrest-False'	0.06	0.99	1.13
28	'Primary Type-DECEPTIVE PRACTICE'	'Arrest-False'	0.08	0.99	1.13
29	'Domestic-False', 'Primary Type-DECEPTIVE PRACTICE'	'Arrest-False'	0.08	0.99	1.13
30	'Domestic-False', 'Description-\$500 AND UNDER'	'Primary Type-THEFT', 'Arrest-False'	0.07	0.99	5.43
31	'Domestic-False', 'Description-\$500 AND UNDER'	'Arrest-False'	0.07	0.99	1.13
32	'Domestic-False', 'Description-\$500 AND UNDER', 'Primary Type-THEFT'	'Arrest-False'	0.07	0.99	1.13
33	'Description-\$500 AND UNDER'	'Primary Type-THEFT', 'Arrest-False'	0.07	0.99	5.42
34	'Description-\$500 AND UNDER'	'Arrest-False'	0.07	0.99	1.12
35	'Description-\$500 AND UNDER', 'Primary Type-THEFT'	'Arrest-False'	0.07	0.99	1.12

Association Rule Mining - 2021 data conclusion

- In districts 4 and 8, no arrests have been made for more than 90% of the crimes.
- The months of August and September contribute to 20% of the annual crime. And yet no arrests were made for more than 90% of the cases.
- In following crimes, arrests rates were very low:
 - Theft over \$500 - 1%
 - Theft under \$500 - 1%
 - Deceptive practice - 1%
 - Criminal Damage to vehicle - 2%
 - Simple Assault - 6%
 - Crimes committed at residences, apartments - 6%, contributes to 25% of the crimes.