Crime and Victimhood in Los Angeles

Madison Razook & David Stone

Research Questions

- Which demographics in LA are most at risk to be a victim of crime?
- Does this change depending on the type of crime (violent or property?)

Data Preparation

- Eliminated attributes that were redundant or irrelevant to our question
- Put attribute values into groups with string names
- Took out null values

	DR Number	Date Occurred	Time Occurred	Area ID	Crime Code	Victim Age	Victim Sex	Victim Descent	Year Occurred	Month Occurred	Day of Week
0	1208575	2013-03-11	5pm to 8:59pm	South	Violent	Age 30-39	Female	White	2013	Q1	Sunday
4	42104479	2014-01-04	9pm to 12:59am	Valley	Property	Age 80-89	Male	White	2014	Q1	Friday
5	120125367	2013-01-08	1pm to 4:59pm	Central	Violent	Age 40-49	Female	White	2013	Q1	Monday
9	120908292	2013-01-15	5am to 8:59am	Valley	Property	Age 20-29	Female	Other	2013	Q1	Monday
12	121207315	2013-02-13	9am to 12:59pm	South	Property	Age 40-49	Male	Hispanic/Latinx	2013	Q1	Tuesday

Data Preparation

- Changed dataset to a table of boolean values for frequent pattern mining

Age 20-29	Age 30-39	Age 40-49	Age 50-59	Age 60-69	Age 70-79	Age 80-89	Age 90-99	AmerIndian/Alaska Native	...	Central	Female	Hispanic/Latinx	Male	Other	Pacific Islander	South	Valley	West	White
False	True	False	False	False	False	False	False	False	...	False	True	False	False	False	False	True	False	False	True
False	False	True	False	False	False	False	False	False	...	True	True	False	False	False	False	False	False	False	True
True	False	False	False	False	False	False	False	False	...	False	False	True	True	False	False	True	False	False	False
False	False	False	False	False	False	False	False	False	...	False	True	True	False	False	False	True	False	False	False
True	False	False	False	False	False	False	False	False	...	True	False	True	True	False	False	False	False	False	False

Tools Used

- Jupyter Notebook
- NumPy, Matplotlib, pandas
- Scikit-learn, mlxtend

Data Mining

- Association rule mining with apriori
  - Split data into Violent and Property
  - Tried to find frequent patterns in victim information: age, sex, descent, area
  - min_support = 0.1, lift > 1, confidence > 0.5, leverage > 0.01
- Random forest classification
  - Multiple decision trees
  - Better accuracy

```
Number training: 1048801
Number test: 349527
```

Predicted Violent Crime	False	True
Actual Violent Crime		
False	18605	210842
True	20575	99505

Knowledge Gained

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
9	(Central)	(Hispanic/Latinx)	0.194174	0.362665	0.100627	0.51823	1.428949	0.030207	1.322902

	antecedants	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
11	(South)	(Black)	0.273208	0.253128	0.142881	0.522974	2.066043	0.073724	1.565684
10	(Black)	(South)	0.253128	0.273208	0.142881	0.564460	2.066043	0.073724	1.668714
13	(Central)	(Hispanic/Latinx)	0.244794	0.473884	0.153514	0.627116	1.323354	0.037510	1.410938
19	(South)	(Female)	0.273208	0.587924	0.177047	0.648030	1.102234	0.016421	1.170769
9	(Black)	(Female)	0.253128	0.587924	0.163200	0.644733	1.096627	0.014380	1.159906
1	(Age 20-29)	(Female)	0.292520	0.587924	0.187839	0.642141	1.092218	0.015860	1.151503
3	(Age 20-29)	(Hispanic/Latinx)	0.292520	0.473884	0.150139	0.513260	1.083092	0.011518	1.080897
21	(Valley)	(Hispanic/Latinx)	0.292076	0.473884	0.148697	0.509105	1.074325	0.010287	1.071749

Applications

- Victim advocacy groups
- Crime matching
- Law enforcement
- Educating the public