A closer look into NY parking violations

Cody Hegwer, Jeff LaPrade, Adam Grabowski

Data

Source: https://www.kaggle.com/new-york-city/ny-parking-violations-issued

- 43 attribute types
- 9.10 million objects
- Contains data ranging from years 1970 all the way to 2014.

	# Summons Number	A Plate ID	A Registration State	A Plate Type	Issue Date	# Vi
	1b 8.01b	2722215 unique values	NY 77% NJ 10% Other (67) 13%	PAS 72% COM 21% Other (86) 8%	17Dec70 22Dec69	0
1	1361929741	FCJ5493	NY	PAS	1970-12-18T00:00:00	
2	1366962000	63540MC	NY	COM	1971-02-02T00:00:00	
3	1342296187	GCY4187	NY	SRF	1971-09-18T00:00:00	
4	1342296199	95V6675	TX	PAS	1971-09-18T00:00:00	
5	1342296217	FYM5117	NY	SRF	1971-09-18T00:00:00	
6	1356906515	GFM1421	NY	PAS	1971-09-18T00:00:00	
7	1337077380	18972BB	NY	999	1971-10-10T00:00:00	
8	1364523796	WNJ4730	VA	PAS	1973-04-05T00:00:00	

Questions

What are common factors of cars getting ticketed?

Where are the most ticketed areas of NYC?

Data Preparation

- Removed holes
- Manipulated PM/AM times into a 24 hours standard time
- Errors in spelling for color classification fixed:
  - ‘BALCK 'WOOD' 'YELLL' 'BK/WH' 'GY/W' 'GY/WH' 'GY/WT' 'BLWHI' 'W/RED'

Final Dataset

	A	B	C	D	E	F	G	H	I	J	K
1	Registration State	Plate Type	Violation Code	Vehicle Body Type	Vehicle Make	Street Code1	Street Code2	Street Code3	Violation Time	Street	Vehicle Color
2	NJ	PAS	5	2 DR	BMW	6110	40404	40404	1554	NB 1ST AVE @ 86TH ST	BK
3	NJ	PAS	5	2 DR	MINI	6110	40404	40404	1751	NB 1ST AVE @ 86TH ST	BK
4	NJ	PAS	5	2 DR	BMW	6110	40404	40404	1103	NB 1ST AVE @ 86TH ST	BK
5	NJ	PAS	5	2 DR	PORSC	6110	40404	40404	2453	NB 1ST AVE @ 86TH ST	BK
6	NJ	PAS	5	2 DR	HONDA	6110	40404	40404	1405	NB 1ST AVE @ 86TH ST	BL
7	NJ	PAS	5	2 DR	INFIN	6110	40404	40404	1314	NB 1ST AVE @ 86TH ST	GY
8	NJ	PAS	5	2 DR	VOLKS	6110	40404	40404	1445	NB 1ST AVE @ 86TH ST	GY
9	NJ	PAS	5	2 DR	HONDA	6110	40404	40404	1601	NB 1ST AVE @ 86TH ST	GY
10	NJ	PAS	5	2 DR	BMW	6110	40404	40404	1619	NB 1ST AVE @ 86TH ST	GY
11	NJ	PAS	5	2 DR	NISSA	6110	40404	40404	1739	NB 1ST AVE @ 86TH ST	GY
12	NJ	PAS	5	2 DR	VOLKS	6110	40404	40404	701	NB 1ST AVE @ 86TH ST	GY
13	NJ	PAS	5	2 DR	NISSA	6110	40404	40404	1148	NB 1ST AVE @ 86TH ST	GY
14	NJ	PAS	5	2 DR	M B	6110	40404	40404	1415	NB 1ST AVE @ 86TH ST	SL
15	NJ	PAS	5	2 DR	MINI	6110	40404	40404	803	NB 1ST AVE @ 86TH ST	SL
16	NJ	PAS	5	2 DR	JEEP	6110	40404	40404	819	NB 1ST AVE @ 86TH ST	WHITE
17	NJ	PAS	5	2 DR	TOYOT	6110	40404	40404	1651	NB 1ST AVE @ 86TH ST	WT
18	CT	PAS	5	2D S	CHEVR	6110	40404	40404	1642	NB 1ST AVE @ 86TH ST	BLK
19	CT	PAS	5	2D S	BMW	6110	40404	40404	1500	NB 1ST AVE @ 86TH ST	GRY
20	CT	PAS	5	2D S	BMW	6110	40404	40404	725	NB 1ST AVE @ 86TH ST	GRY
21	CT	PAS	5	2DHT	MINI	6110	40404	40404	1645	NB 1ST AVE @ 86TH ST	RED
22	NY	PAS	21	2DSD	HONDA	19290	7240	55290	1136	100th St	BK

Tools Used

- Python
- Numpy
- Pandas
- Tableau
- Apriori Module

Classification applied

Naive Bayesian Classifier

-Classifying on Color:

1. White : (prob = 3.66e-05)
2. Grey : (prob = 2.92e-05)
3. Black : (prob = 1.12e-05)

1. White : (prob = 1.93e-05)
2. Grey : (prob = 7.61e-06)
3. Black : (prob = 3.45e-06)

1. White : (prob = 1.10e-04)
2. Grey : (prob = 5.49e-05)
3. Black : (prob = 2.10e-05)

Classification applied

Naive Bayesian Classifier

-Classifying on State:

1. New York: 739,541
2. New Jersey: 96,263
3. Pennsylvania: 23,089
4. Connecticut: 14,966
5. Florida: 12,567

Classification applied

Naive Bayesian Classifier

-Violation Code:

21: Street Cleaning

38: Failing to show a receipt on windshield

14: General No Standing (Parking)

Classification applied

Naive Bayesian Classifier

-Time:

Penn plates

Code 38:

(No receipt)

1. Morning : (2.676e-05)
2. Noon : (2.333e-05)
3. Night : (6.163e-07)

Code 20:

1. Morning : (1.701e-05)
2. Noon : (8.286e-06)

Usually tie between Morning and Afternoon, Pennsylvania had double the probability for Morning.

Classification applied

Bayesian Code Example:

```
def Bayes(state, time, color):

    totalN = 0
    rows = 100
    columns = 7
    violArray = [[0 for x in range(columns)] for y in range(rows)]

    for i in range(rows):
        violArray[i][0] = i


    for ind, row in df.iterrows():


        totalN += 1

        colortest = row['Color']
        statetest = row['State']
        vcode = row['ViolCode']
        timetest = row['Violation Time']
```

Apriori Analysis

Apriori Code Example:

In [2]:
```
## creating function to turn CSV file into a list of lists for the Apriori algorithm to use

def data_cleaner(filename):
  def data_clean():
    with open(filename) as file:
      for line in file:
        yield tuple(k.strip() for k in line.split(','))

  return data_clean

transactions = data_cleaner('/Users/adamgrabowski/Documents/Colorado Classes/Third Semester/CSCI 4502/Semester Project/
```

In [8]:
```
## Intially run with full data

itemsets, rules = apriori(transactions, min_support=0.2, min_confidence=0.5)
```

In [9]:
```
print(rules)
```

```
[{4DSD} -> {NY}, {4DSD} -> {PAS}, {AFTERNOON} -> {NY}, {AFTERNOON} -> {PAS}, {COM} -> {NY}, {GREY} -> {PAS}, {MORNIN
G} -> {NY}, {MORNING} -> {PAS}, {PAS} -> {NY}, {NY} -> {PAS}, {SUBN} -> {NY}, {WHITE} -> {NY}, {SUBN} -> {PAS}, {MORN
ING, PAS} -> {NY}, {MORNING, NY} -> {PAS}, {PAS, SUBN} -> {NY}, {NY, SUBN} -> {PAS}, {SUBN} -> {NY, PAS}]
```

In [10]:
```
transactions_NoPlate = data_cleaner('/Users/adamgrabowski/Documents/Colorado Classes/Third Semester/CSCI 4502/Semester
```

In [11]:
```
## Run again with same confidence and support, but removing plate type

itemsets, rules = apriori(transactions_NoPlate, min_support=0.2, min_confidence=0.5)
```

Apriori Analysis

Medium Confidence, High Support

- Data is to be widely distributed, no association rules for 50% support

Apriori Analysis

Medium Confidence, Medium Support

- {21} -> {MORNING}
- {37} -> {AFTERNOON}
- {38} -> {AFTERNOON}
- {DELV} -> {WHITE}
- {VAN} -> {WHITE}

Apriori Analysis

Medium Confidence, Low Support

- Similar rules, but much wider results

- {BROWN} -> {AFTERNOON},
- {21, 4DSD} -> {MORNING},
- {21, BLACK} -> {MORNING},
- {37, SUBN} -> {AFTERNOON},
- {DELV, WHITE} -> {MORNING},
- {DELV, MORNING} -> {WHITE},
- {FORD, MORNING, VAN} -> {WHITE}

Apriori Analysis

High Confidence, Low Support

- Many less rules than medium confidence, results considered more conclusive

- {21} -> {MORNING},
- {7, AFTERNOON} -> {0},
- {7, SUBN} -> {0},
- {21, 4DSD} -> {MORNING},
- {37, SUBN} -> {AFTERNOON},
- {MORNING, VAN} -> {WHITE},
- {FORD, MORNING, VAN} -> {WHITE}

Knowledge Gained

- Color, model, etc. attributes were not very useful. Was hoping for more street related patterns.

- Time and violation type strongly connected, however

- Street Cleaning: Morning
- Parking in Excess of Allowed Time: Afternoon
- Failing to Show Receipt: Afternoon
- General Parking Violation: Morning

How to Apply that Knowledge

Recommended City Traversal:

General parking violation, morning and parking in excess of allowed time violation, afternoon = **legitimately pay for parking in the morning.**

Afternoon, low chance of general violation = **park where you want?**

Evening, night low chance of excess allowed time violation = **meters set and forget in afternoon**

Tableau: Distribution of car bodies

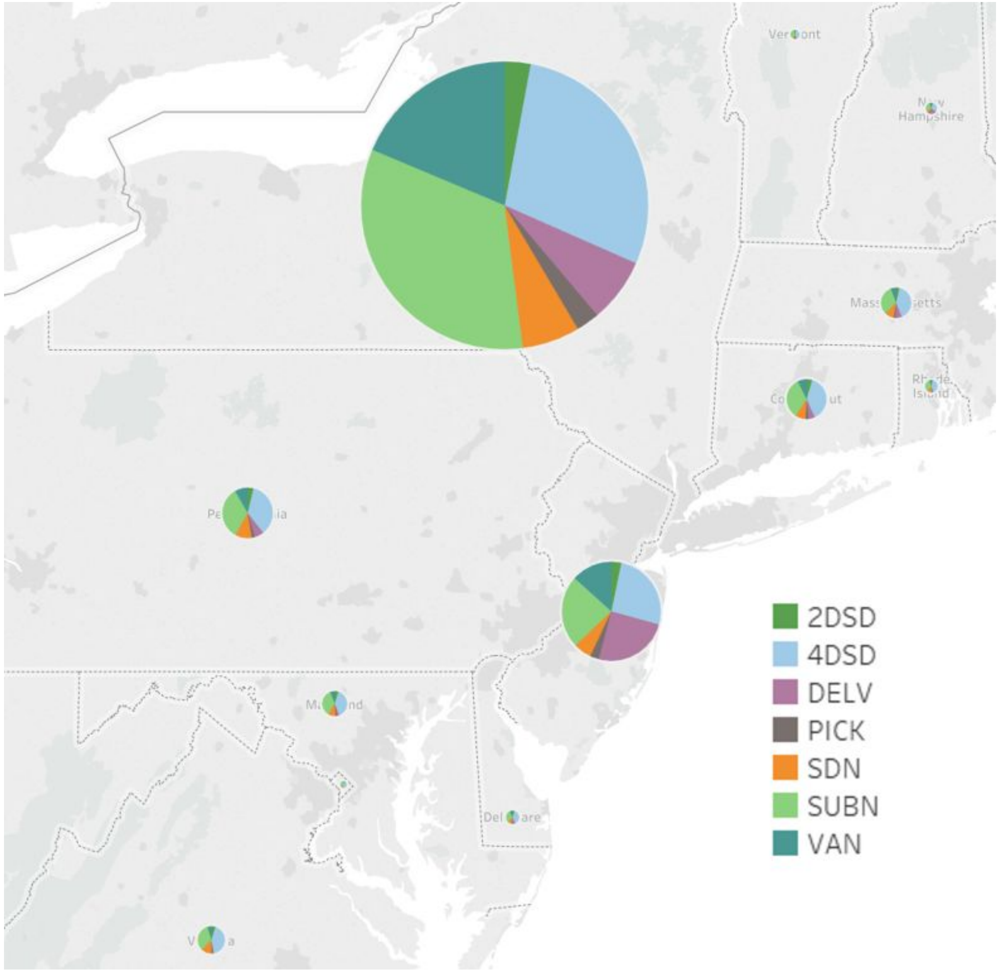

Tableau: Tree map