

Richa Patel

Summary

Rhode Island Police data are a very important and distinct segment of the Research analysis. They vary greatly in sizes, and general

structure. This report uses a collection of state, stop date, stop time, county name, driver gender, driver race, violation raw, violation, search conducted, search type, stop outcome, is arrested, stop duration, drugs_related_stop, district.

Section 1. Introduction

To utilize the Rhode Island police data set to Analyzing the effect of weather on policing and Topmost reason for violation and *Comparing violations by gender and Comparing stop outcomes by gender*

.91,741 records from Rhode Island Police Department's traffic stops between January 4th, 2005 and December 31st, 2015.

The analysis assumed that all the data is gathered with the same method and naming is consistent.

Read data into Data Frame and separated records based on gender, and then used to create visuals to better analyze the data. Reviewed data in .csv file, used Python to import into Data Frame, and worked with Data Frame to answer questions.

After error handling, no notable errors were encountered.

The conclusion I reached was that Topmost reason for violation is Speeding and least reason is Seat Belt. Citation has high value. and There's no bias here for pulling over non-white driver's vs white drivers but other race is less stopped by. Male is stopped than Female.

In the remainder of the report, final conclusions reached. Section 2 outlines the data we gathered. Section 3 provides information about the police and weather that were developed. The results of the

report, as well as possible limitations and improvements, can be found in Section 4.

Section 2. Data characteristics

In order to get a data read file for police and weather CSV. we can see that we have 91741 rows and 15 columns for police and [4017 rows x 27 columns] for weather. Topmost reason for violation is Speeding and least reason is Seat Belt. Citation has high value. and There's no bias here for pulling over non-white driver's vs white drivers but other race is less stopped by. Male is stopped than Female.

```
police['violation_raw'].value_counts()
Speeding                                48424
Other Traffic Violation                 16224
Equipment/Inspection Violation         10922
Registration Violation                  3703
Seatbelt Violation                     2856
Special Detail/Directed Patrol         2467
Call for Service                       1392
Motorist Assist/Courtesy                205
Violation of City/Town Ordinance       181
APB                                     91
Suspicious Person                      56
Warrant                                18
Name: violation_raw, dtype: int64
```

```
police['stop_outcome'].value_counts()
```

```
Citation          77092
Warning           5137
Arrest Driver     2735
No Action         625
N/D              607
Arrest Passenger  343
Name: stop_outcome, dtype: int64
```

We can see citation has high value.

```
police['stop_outcome'].value_counts()
```

```
Citation          77092
Warning           5137
Arrest Driver     2735
No Action         625
N/D              607
Arrest Passenger  343
Name: stop_outcome, dtype: int64
```

We can see citation has high value.

```
: police['driver_race'].value_counts()
```

```
: White          61872
  Black          12285
  Hispanic        9727
  Asian           2390
  Other           265
Name: driver_race, dtype: int64
```

There's no bias here for pulling over non-white drivers vs white drivers but other race is less stopped by.

```
: police[["stop_date", "stop_time"]].head()
```

```
:
   stop_date stop_time
0  2005-01-04    12:55
1  2005-01-23    23:15
2  2005-02-17     04:15
3  2005-02-20    17:15
4  2005-02-24     01:20
```

Section 3. Weather and Police

We can see TMIN TAVG and TMAX . Create a 'TDIFF' column that represents temperature difference weather ['TDIFF'] = weather.TMAX - weather.TMIN

```
# Describe the 'TDIFF' column using print method
```

```
print(weather.TDIFF.describe())
```

```
# Create a histogram with 20 bins to visualize 'TDIFF'
```

```
weather.TDIFF.plot(kind = 'hist', bins = 25)
```

```
# Create a box plot of the temperature columns
```

```
weather[['TDIFF']].plot(kind = 'box')
```

```
# Display the plot using show method
```

```
plt.show()
```

```
# Describe the temperature columns
```

```
print(weather[['TMIN', 'TAVG', 'TMAX']].describe())
```

```
# Create a box plot of the temperature columns
```

```
weather[['TMIN', 'TAVG', 'TMAX']].plot(kind = 'box')
```

```
# Display the plot
```

```
plt.show()
```

	TMIN	TAVG	TMAX
count	4017.000000	1217.000000	4017.000000
mean	43.484441	52.493016	61.268608
std	17.020298	17.830714	18.199517
min	-5.000000	6.000000	15.000000
25%	30.000000	39.000000	47.000000
50%	44.000000	54.000000	62.000000
75%	58.000000	68.000000	77.000000
max	77.000000	86.000000	102.000000

Section 4. Summary and concluding remarks

I came into this blind, but I assumed that both Genders would have roughly the same chance at getting cited

Across the sample data, the most-common types of traffic violation were speeding violations, moving violations, and equipment violations; we found that males are the drivers in about 72.5270% of the sampled traffic stops; and that the Rhode Island Police Department Districts ranked by total sample traffic stops are the “Zone X4” district, the “Zone K3” district, the “Zone K2” district, the “Zone X3” district, the “Zone K1” district, and the “Zone X1” district.

A.2 summary data about police and weather

police.csv

summary(data)

	state	stop_date	stop_time	county_name	driver_gender	driver_race	\
0	RI	2005-01-04	12:55	NaN	M	White	
1	RI	2005-01-23	23:15	NaN	M	White	
2	RI	2005-02-17	04:15	NaN	M	White	
3	RI	2005-02-20	17:15	NaN	M	White	
4	RI	2005-02-24	01:20	NaN	F	White	
...
91736	RI	2015-12-31	21:21	NaN	F	Black	
91737	RI	2015-12-31	21:59	NaN	F	White	
91738	RI	2015-12-31	22:04	NaN	M	White	
91739	RI	2015-12-31	22:09	NaN	F	Hispanic	
91740	RI	2015-12-31	22:47	NaN	M	White	

	violation_raw	violation	search_conducted	\
0	Equipment/Inspection Violation	Equipment	False	
1	Speeding	Speeding	False	
2	Speeding	Speeding	False	
3	Call for Service	Other	False	
4	Speeding	Speeding	False	
...
91736	Other Traffic Violation	Moving violation	False	
91737	Speeding	Speeding	False	
91738	Other Traffic Violation	Moving violation	False	
91739	Equipment/Inspection Violation	Equipment	False	
91740	Registration Violation	Registration/plates	False	

	search_type	stop_outcome	is_arrested	stop_duration	\
0	NaN	Citation	False	0-15 Min	
1	NaN	Citation	False	0-15 Min	
2	NaN	Citation	False	0-15 Min	
3	NaN	Arrest Driver	True	16-30 Min	
4	NaN	Citation	False	0-15 Min	
...
91736	NaN	Citation	False	0-15 Min	
91737	NaN	Citation	False	0-15 Min	

books/Downloads/PYTHON_ASSIGNMENTS/Rhode%20Island%20Police%20and%20local%20Weather%20data.ipynb#List-of-all-county-name

ocal Weather data - Jupyter Notebook

91738	NaN	Citation	False	0-15 Min
91739	NaN	Warning	False	0-15 Min
91740	NaN	Citation	False	0-15 Min

	drugs_related_stop	district
0	False	Zone X4
1	False	Zone K3
2	False	Zone X4
3	False	Zone X1
4	False	Zone X3
...
91736	False	Zone K2
91737	False	Zone K3
91738	False	Zone X3
91739	False	Zone K3
91740	False	Zone X4

[91741 rows x 15 columns]

Weather csv

	STATION	DATE	TAVG	TMIN	TMAX	AWND	WSF2	WT01	WT02	\
0	USW00014765	2005-01-01	44.0	35	53	8.95	25.1	1.0	NaN	
1	USW00014765	2005-01-02	36.0	28	44	9.40	14.1	NaN	NaN	
2	USW00014765	2005-01-03	49.0	44	53	6.93	17.0	1.0	NaN	
3	USW00014765	2005-01-04	42.0	39	45	6.93	16.1	1.0	NaN	
4	USW00014765	2005-01-05	36.0	28	43	7.83	17.0	1.0	NaN	
...	
4012	USW00014765	2015-12-27	51.0	44	61	9.17	28.0	1.0	NaN	
4013	USW00014765	2015-12-28	40.0	30	44	12.30	23.0	NaN	NaN	
4014	USW00014765	2015-12-29	33.0	28	40	12.53	18.1	1.0	NaN	
4015	USW00014765	2015-12-30	30.0	27	35	6.93	15.0	1.0	NaN	
4016	USW00014765	2015-12-31	39.0	35	50	8.05	18.1	1.0	NaN	

	WT03	...	WT11	WT13	WT14	WT15	WT16	WT17	WT18	WT19	WT21	WT22
0	NaN	...	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	...	NaN	NaN	NaN	NaN	1.0	NaN	1.0	NaN	NaN	NaN
2	NaN	...	NaN	1.0	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN
3	NaN	...	NaN	1.0	1.0	NaN	1.0	NaN	NaN	NaN	NaN	NaN
4	NaN	...	NaN	1.0	NaN	NaN	1.0	NaN	1.0	NaN	NaN	NaN
...
4012	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4013	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4014	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4015	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4016	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

[4017 rows x 27 columns]

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	state	91741 non-null	object
1	stop_date	91741 non-null	object
2	stop_time	91741 non-null	object
3	county_name	0 non-null	float64
4	driver_gender	86536 non-null	object
5	driver_race	86539 non-null	object
6	violation_raw	86539 non-null	object
7	violation	86539 non-null	object
8	search_conducted	91741 non-null	bool
9	search_type	3307 non-null	object
10	stop_outcome	86539 non-null	object
11	is_arrested	86539 non-null	object
12	stop_duration	86539 non-null	object
13	drugs_related_stop	91741 non-null	bool
14	district	91741 non-null	object

A.3 model and diagnostics

Common search types with Four races



