In [1]:
```
import pandas as pd
```

In [6]:
```
data = pd.read_csv('C:\\Users\\UP\\Downloads\\3. Police Data.csv')
```

In [3]:
```
data
```

Out[3]:

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice	EngineSize	Cylinders	Horsepowe
0	Acura	MDX	SUV	Asia	All	$36,945	$33,337	3.5	6.0	265.(
1	Acura	RSX Type S 2dr	Sedan	Asia	Front	$23,820	$21,761	2.0	4.0	200.(
2	Acura	TSX 4dr	Sedan	Asia	Front	$26,990	$24,647	2.4	4.0	200.(
3	Acura	TL 4dr	Sedan	Asia	Front	$33,195	$30,299	3.2	6.0	270.(
4	Acura	3.5 RL 4dr	Sedan	Asia	Front	$43,755	$39,014	3.5	6.0	225.(
...	...	...	...	...	...	...	...	...	...	..
427	Volvo	C70 LPT convertible 2dr	Sedan	Europe	Front	$40,565	$38,203	2.4	5.0	197.(
428	Volvo	C70 HPT convertible 2dr	Sedan	Europe	Front	$42,565	$40,083	2.3	5.0	242.(
429	Volvo	S80 T6 4dr	Sedan	Europe	Front	$45,210	$42,573	2.9	6.0	268.(
430	Volvo	V40	Wagon	Europe	Front	$26,135	$24,641	1.9	4.0	170.(
431	Volvo	XC70	Wagon	Europe	All	$35,145	$33,112	2.5	5.0	208.(

432 rows × 15 columns

In [7]:
```
data.isnull().sum()
```

Out[7]:
```
stop_date                0
stop_time                0
country_name         65535
driver_gender         4061
driver_age_raw        4054
driver_age            4307
driver_race           4060
violation_raw         4060
violation             4060
search_conducted         0
search_type          63056
stop_outcome          4060
is_arrested           4060
stop_duration         4060
drugs_related_stop       0
dtype: int64
```

In [16]:
```
data.drop( columns = 'country_name', inplace = True)
```

In [17]:
```
data
```

Out[17]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw
0	1/2/2005	1:55	M	1985.0	20.0	White	Speeding
1	1/18/2005	8:15	M	1965.0	40.0	White	Speeding
2	1/23/2005	23:15	M	1972.0	33.0	White	Speeding
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Service
4	3/14/2005	10:00	F	1984.0	21.0	White	Speeding
...	...	...	...	...	...	...	...
65530	12/6/2012	17:54	F	1987.0	25.0	White	Speeding
65531	12/6/2012	22:22	M	1954.0	58.0	White	Speeding
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspection Violation
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	NaN
65534	12/7/2012	0:30	F	1985.0	27.0	White	Speeding

65535 rows × 14 columns

In [25]:
```
data[data.violation == 'Speeding'].driver_gender.value_counts()
```

Out[25]:
```
M    25517
F    11686
Name: driver_gender, dtype: int64
```

In [34]:
```
data.groupby('driver_gender').search_conducted.sum()
```

Out[34]:
```
driver_gender
F     366
M    2113
Name: search_conducted, dtype: int64
```

In [30]:
```

```

Out[30]:

stop_time	driver_age_raw	driver_age	search_conducted	drugs_related_stop
0:00	1976.308943	31.089431	0.116279	0.031008
0:01	1979.345455	29.709091	0.169492	0.067797
0:02	1939.823529	31.220000	0.155172	0.086207
0:03	1931.348837	32.119048	0.062500	0.000000

stop_time	driver_age_raw	driver_age	search_conducted	drugs_related_stop
0:04	1976.800000	32.345455	0.016393	0.000000
...	...	...	...	...
9:55	2037.018519	34.429907	0.033898	0.008475
9:56	1972.869565	36.934783	0.020000	0.020000
9:57	1940.285714	34.163636	0.068966	0.017241
9:58	1972.693878	36.755102	0.019231	0.000000
9:59	1972.644444	36.733333	0.042553	0.000000

1432 rows × 4 columns

In [31]:
```
data.stop_duration.value_counts()
```

Out[31]:
```
0-15 Min     47379
16-30 Min    11448
30+ Min       2647
2                1
Name: stop_duration, dtype: int64
```

In [33]:
```
data['stop_duration'].map( {'0-15 Min' : 7.5, '16-30' : 24, '30+ Min' : 45 })
```

Out[33]:
```
0        7.5
1        7.5
2        7.5
3        NaN
4        7.5
        ...
65530    7.5
65531    7.5
65532    7.5
65533    NaN
65534    7.5
Name: stop_duration, Length: 65535, dtype: float64
```

In [35]:
```
data.groupby('violation').driver_age.describe() #to remove the records satidfying a per
```

Out[35]:

violation	count	mean	std	min	25%	50%	75%	max
Equipment	6507.0	31.682957	11.380671	16.0	23.0	28.0	39.0	81.0
Moving violation	11876.0	36.736443	13.258350	15.0	25.0	35.0	47.0	86.0
Other	3477.0	40.362381	12.754423	16.0	30.0	41.0	50.0	86.0
Registration/plates	2240.0	32.656696	11.150780	16.0	24.0	30.0	40.0	74.0
Seat belt	3.0	30.333333	10.214369	23.0	24.5	26.0	34.0	42.0
Speeding	37120.0	33.262581	12.615781	15.0	23.0	30.0	42.0	88.0