

Police Dataset

Here,

The data from a Police Check Post is given.

The data is available as a CSV file. I am going to analyze this data set using the Pandas DataFrame.

In [2]:

```
import pandas as pd
```

In [3]:

```
data = pd.read_csv(r'C:\Users\harsh\Desktop\Projects\Python\Police Dataset Analysis\Police Data.csv')
```

In [4]:

```
data
```

Out[4]:

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

65535 rows × 15 columns

Q1) Remove the columns that only contain missing values.

In [5]:

```
data.isnull().sum()
```

Out[5]:

```
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 [6]:

```
data.drop(columns = 'country_name', inplace = True)
```

In [7]:

```
data.head(2)
```

Out[7]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	searched
0	1/2/2005	1:55	M	1985.0	20.0	White	Speeding	Speeding	
1	1/18/2005	8:15	M	1965.0	40.0	White	Speeding	Speeding	

Q2) For speeding, were Men or Women stopped more often?

In [8]:

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

Out[8]:

```
M    25517
F     11686
Name: driver_gender, dtype: int64
```

Q3) Does gender affect who gets searched during a stop?

In [9]:

```
data.groupby('driver_gender').search_conducted.sum()
```

Out[9]:

```
driver_gender
F      366
M     2113
Name: search_conducted, dtype: int64
```

Q4) What is the mean stop_duration?

In [10]:

```
data.head()
```

Out[10]:

r	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_conducted	search_type	stop_outco
1	1985.0	20.0	White	Speeding	Speeding	False	NaN	Citat
1	1965.0	40.0	White	Speeding	Speeding	False	NaN	Citat
1	1972.0	33.0	White	Speeding	Speeding	False	NaN	Citat
1	1986.0	19.0	White	Call for Service	Other	False	NaN	Arrest Dri
=	1984.0	21.0	White	Speeding	Speeding	False	NaN	Citat

In [11]:

```
data.stop_duration.value_counts()
```

Out[11]:

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

In [12]:

```
data['stop_duration'] = data['stop_duration'].map( {'0-15 Min': 7.5, '16-30 Min': 23, '30+ Min': 45 } )
```

In [13]:

```
data
```

Out[13]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	vi
0	1/2/2005	1:55	M	1985.0	20.0	White	Speeding	Sp
1	1/18/2005	8:15	M	1965.0	40.0	White	Speeding	Sp
2	1/23/2005	23:15	M	1972.0	33.0	White	Speeding	Sp
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	Sp
...	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Speeding	Sp
65531	12/6/2012	22:22	M	1954.0	58.0	White	Speeding	Sp
65532	12/6/2012	23:20	M	1985.0	27.0	Black	Equipment/Inspection Violation	Equ
65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	NaN	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Speeding	Sp

65535 rows × 14 columns

In [14]:

```
data['stop_duration'].mean()
```

Out[14]:

12.001195627419722

Q5) Compare the age distributions for each violation.

In [15]:

```
data.groupby('violation').driver_age.describe()
```

Out[15]:

	count	mean	std	min	25%	50%	75%	max
violation								
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

