Dataset Link

https://www.kaggle.com/code/rohitgrewal/police-data-analysis

Questions need to analyze in this dataset

- Instruction (For Data Cleaning) - Remove the column that only contains missing values.
- Question (Based on Filtering + Value Counts) - For Speeding , were Men or Women stopped more often ?
- Question (Groupby) - Does gender affect who gets searched during a stop ?
- Question (mapping + data-type casting) - What is the mean stop_duration ?
- Question (Groupby , Describe) - Compare the age distributions for each violation.

In [27]:

```
# Import the dataset and required libraries
```

In [28]:

```
import pandas as pd
import matplotlib.pyplot as plt
```

In [29]:

```
df=pd.read_csv("Police Dataset.csv")
```

In [30]:

```
df.head()
```

Out[30]:

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driv
0	1/2/2005	1:55	NaN	M	1985.0	20.0	
1	1/18/2005	8:15	NaN	M	1965.0	40.0	
2	1/23/2005	23:15	NaN	M	1972.0	33.0	
3	2/20/2005	17:15	NaN	M	1986.0	19.0	
4	3/14/2005	10:00	NaN	F	1984.0	21.0	

In [31]:

```
df.shape
```

Out[31]:

```
(65535, 15)
```

In [32]:

```
df.size
```

Out[32]:

```
983025
```

In [33]:

```
df.ndim
```

Out[33]:

```
2
```

In [34]:
```
df.info
```

Out[34]:
```
<bound method DataFrame.info of        stop_date stop_time  country_name driver
_gender  driver_age_raw  \
0       1/2/2005      1:55           NaN             M          1985.0
1      1/18/2005      8:15           NaN             M          1965.0
2      1/23/2005     23:15           NaN             M          1972.0
3      2/20/2005     17:15           NaN             M          1986.0
4      3/14/2005     10:00           NaN             F          1984.0
...          ...       ...           ...           ...             ...
65530  12/6/2012     17:54           NaN             F          1987.0
65531  12/6/2012     22:22           NaN             M          1954.0
65532  12/6/2012     23:20           NaN             M          1985.0
65533  12/7/2012      0:23           NaN           NaN             NaN
65534  12/7/2012      0:30           NaN             F          1985.0

       driver_age driver_race                   violation_raw  violation  \
0            20.0       White                        Speeding   Speeding
1            40.0       White                        Speeding   Speeding
2            33.0       White                        Speeding   Speeding
3            19.0       White                Call for Service      Other
4            21.0       White                        Speeding   Speeding
...           ...         ...                             ...        ...
65530        25.0       White                        Speeding   Speeding
65531        58.0       White                        Speeding   Speeding
65532        27.0       Black  Equipment/Inspection Violation  Equipment
65533         NaN         NaN                             NaN        NaN
65534        27.0       White                        Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested stop_duration  \
0                 False         NaN       Citation       False      0-15 Min
1                 False         NaN       Citation       False      0-15 Min
2                 False         NaN       Citation       False      0-15 Min
3                 False         NaN  Arrest Driver        True     16-30 Min
4                 False         NaN       Citation       False      0-15 Min
...                 ...         ...            ...         ...           ...
65530             False         NaN       Citation       False      0-15 Min
65531             False         NaN        Warning       False      0-15 Min
65532             False         NaN       Citation       False      0-15 Min
65533             False         NaN            NaN         NaN           NaN
65534             False         NaN       Citation       False      0-15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 15 columns]>
```

In [35]:
```
df.describe()
```

Out[35]:

	country_name	driver_age_raw	driver_age
count	0.0	61481.000000	61228.000000
mean	NaN	1967.791106	34.148984
std	NaN	121.050106	12.760710
min	NaN	0.000000	15.000000
25%	NaN	1965.000000	23.000000
50%	NaN	1978.000000	31.000000
75%	NaN	1985.000000	43.000000
max	NaN	8801.000000	88.000000

Data Cleaning

Check for the missing value and remove the records

In [36]:

```
df.isnull().sum()
```

Out[36]:

```
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
```

As we can see, "country_name" column is not required for analysis, so we can drop that column

In [37]:

```
df.drop(columns = "country_name",inplace=True)
```

In [38]:

```
df.head(2)
```

Out[38]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violatio
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe

- For speeding, check how many Men or Women are stopped more often?

In [39]:

```
df[df.violation == "Speeding"]. driver_gender.value_counts()
```

Out[39]:

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

- Does gender affect who gets searched during a stop?

In [40]:

```
df.groupby("driver_gender").search_conducted.sum() # groupby used to make a grou
```

Out[40]:

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

- How many times search was conducted?

In [41]:

```
df.search_conducted.value_counts()
```

Out[41]:

```
search_conducted
False    63056
True      2479
Name: count, dtype: int64
```

Mapping + Data-type Casting

What is the mean stop_duration?

- Mapping - We've to map the new values to the column
- Data - type casting -- to convert data-type of one element to another : string--> float

In [42]:

```
df.head(2)
```

Out[42]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violatio
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe

- To find how many unique values are present in stop_duration.

In [43]:

```
df.stop_duration.value_counts()
```

Out[43]:

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

Now, map new values to the column data, 0-15 Min: 7, 16-30 Min: 24,30+ Min:45

In [44]:

```
df["stop_duration"]=df["stop_duration"].map({'0-15 Min': 7, '16-30 Min': 24, '30
```

In [45]:

```
df.head()
```

Out[45]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violatio
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe
2	1/23/2005	23:15	M	1972.0	33.0	White	Spe
3	2/20/2005	17:15	M	1986.0	19.0	White	C S
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe

- Get the average stop_duration

In [46]:

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

Out[46]:

```
np.float64(11.802062660637016)
```

Compare the age distributions for each violations

In [47]:

```
df.head(2)
```

Out[47]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violatio
0	1/2/2005	1:55	M	1985.0	20.0	White	Spe
1	1/18/2005	8:15	M	1965.0	40.0	White	Spe

In [48]:

```
df.describe()
```

Out[48]:

	driver_age_raw	driver_age	stop_duration
count	61481.000000	61228.000000	61474.000000
mean	1967.791106	34.148984	11.802063
std	121.050106	12.760710	9.640422
min	0.000000	15.000000	7.000000
25%	1965.000000	23.000000	7.000000
50%	1978.000000	31.000000	7.000000
75%	1985.000000	43.000000	7.000000
max	8801.000000	88.000000	45.000000

In [49]:

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

Out[49]:

	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

In [ ]: