```python
import pandas as pd
```

```python
data=pd.read_csv(r"C:\Users\Bhavya's Surface\Documents\pandas\PoliceData.csv")
```

```python
data
```

Out[3]:

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

65535 rows × 15 columns

```python
#to clear the data which is null
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
```

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

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

Out[10]:

```
stop_date                0
stop_time                0
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 [11]:

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

In [14]:

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

Out[14]:

```
stop_date                0
stop_time                0
driver_gender         4061
driver_age_raw        4054
driver_age            4307
driver_race           4060
violation_raw         4060
violation             4060
search_conducted         0
stop_outcome          4060
is_arrested           4060
stop_duration         4060
drugs_related_stop       0
dtype: int64
```

In [15]:

```python
#for speeding
data[data.violation=='Speeding']
```

Out[15]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_conducted	stop_o
0	1/2/2005	1:55	M	1985.0	20.0	White	Speeding	Speeding	False	
1	1/18/2005	8:15	M	1965.0	40.0	White	Speeding	Speeding	False	
2	1/23/2005	23:15	M	1972.0	33.0	White	Speeding	Speeding	False	
4	3/14/2005	10:00	F	1984.0	21.0	White	Speeding	Speeding	False	
6	4/1/2005	17:30	M	1969.0	36.0	White	Speeding	Speeding	False	
...	...	...	...	...	...	...	...	...	...	
65527	12/6/2012	15:26	F	1981.0	31.0	White	Speeding	Speeding	False	
65529	12/6/2012	16:00	M	1994.0	18.0	White	Speeding	Speeding	False	
65530	12/6/2012	17:54	F	1987.0	25.0	White	Speeding	Speeding	False	
65531	12/6/2012	22:22	M	1954.0	58.0	White	Speeding	Speeding	False	
65534	12/7/2012	0:30	F	1985.0	27.0	White	Speeding	Speeding	False	

37204 rows × 13 columns

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

Out[17]:

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

In [20]:

```
#gender during the search conducted
data.groupby('driver_gender').search_conducted.sum()
```

Out[20]:

```
driver_gender
F     366.0
M    2113.0
Name: search_conducted, dtype: float64
```

In [21]:

```
#mean of stop duration
data
```

Out[21]:

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

65535 rows × 13 columns

In [22]:

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

Out[22]:

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

In [24]:

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

```
data
```

Out[25]:

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

65535 rows × 13 columns

In [26]:

```
#mean stop duration
data['stop_duration'].mean()
```

Out[26]:

12.187420698181345

In [27]:

```
#age distribution for each violation
data.groupby('violation').driver_age.describe()
```

Out[27]:

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

In [ ]: