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

In [2]:
```
police=pd.read_csv(r'E:\new download\3. Police Data.csv')
```

In [3]:
```
police.head(7)
```

Out[3]:

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race	violation_
0	1/2/2005	1:55	NaN	M	1985.0	20.0	White	Spee
1	1/18/2005	8:15	NaN	M	1965.0	40.0	White	Spee
2	1/23/2005	23:15	NaN	M	1972.0	33.0	White	Spee
3	2/20/2005	17:15	NaN	M	1986.0	19.0	White	Call for Se
4	3/14/2005	10:00	NaN	F	1984.0	21.0	White	Spee
5	3/23/2005	9:45	NaN	M	1982.0	23.0	Black	Equipment/Inspe Viol
6	4/1/2005	17:30	NaN	M	1969.0	36.0	White	Spee

Q.1) Remove the column that only contains missing values.

In [4]:
```
police.isnull().sum()
```

Out[4]:
```
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 [5]:
```
police.shape
```

Out[5]:
```
(65535, 15)
```

In [6]:
```
police.drop(columns= 'country_name', inplace= True)
```

In [7]:
```
police
```

Out[7]:

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

65535 rows × 14 columns

Q.2) For speeding, were Men or women stoped more often?

In [8]:
```
police[police['violation']=='Speeding'].driver_gender.value_counts()
```

Out[8]:
```
M    25517
F    11686
Name: driver_gender, dtype: int64
```

Q.3) Does gender affect who gets searched during the stop?

In [9]:
```
police.head()
```

Out[9]:

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search
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	
2	1/23/2005	23:15	M	1972.0	33.0	White	Speeding	Speeding	
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for Service	Other	
4	3/14/2005	10:00	F	1984.0	21.0	White	Speeding	Speeding	

In [10]:
```
police.groupby('driver_gender').search_conducted.sum()
```

Out[10]:
```
driver_gender
F     366
M    2113
Name: search_conducted, dtype: int64
```

Q.4) What is the mean stop_duration?

In [11]:
```
police.dtypes
```

Out[11]:
```
stop_date             object
stop_time             object
driver_gender         object
driver_age_raw       float64
driver_age           float64
driver_race           object
violation_raw         object
violation             object
search_conducted        bool
search_type           object
stop_outcome          object
is_arrested           object
stop_duration         object
drugs_related_stop      bool
dtype: object
```

In [12]:
```
police.stop_duration.value_counts()
```

Out[12]:
```
0-15 Min     47379
16-30 Min    11448
30+ Min       2647
2                1
Name: stop_duration, dtype: int64
```

In [13]:
```
police['stop_duration']=police['stop_duration'].map({'0-15 Min': 7.5, '16-30 Min': 25, '
```

In [14]:
```
police
```

Out[14]:

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

65535 rows × 14 columns

In [15]:
```
police.stop_duration.mean()
```

Out[15]:
```
12.373645768942968
```

Q.5) Compare the age distributions for each violation

In [24]:
```
police.groupby('violation').driver_age.describe()
```

Out[24]:

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 [ ]:
```

```