```
import pandas as pd
import numpy as np

df = pd.read_csv(r"C:\Users\Ashish\OneDrive\Pyhton Project file ALL\
Police Data.csv")

df.head(2)
```

```
   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

   driver_age driver_race violation_raw violation  search_conducted  \
0        20.0       White      Speeding  Speeding             False
1        40.0       White      Speeding  Speeding             False

  search_type stop_outcome is_arrested stop_duration
drugs_related_stop
0         NaN     Citation       False      0-15 Min
False
1         NaN     Citation       False      0-15 Min
False
```

Remove coloumn that contain missing values?

df.isnull().sum()

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

For speeding, were men and women stooped more often?

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

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

  violation_raw violation  search_conducted search_type
stop_outcome  \
0      Speeding  Speeding             False         NaN     Citation

1      Speeding  Speeding             False         NaN     Citation


  is_arrested stop_duration  drugs_related_stop
```

```
0       False      0-15 Min               False
1       False      0-15 Min               False
```

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

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

Dose gender affect who gets searched during a stop?

```
df.groupby("driver_gender").search_conducted.sum()
```

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

What is the mean Stop_duration?

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

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

  violation_raw violation  search_conducted search_type
stop_outcome  \
0      Speeding  Speeding             False         NaN     Citation

1      Speeding  Speeding             False         NaN     Citation


  is_arrested stop_duration  drugs_related_stop
0       False      0-15 Min               False
1       False      0-15 Min               False
```

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

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

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

df["stop_duration"].mean().round(2)

12.19
```

Compare the age diffrence for each voilation?

```
df.head(2)

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

  violation_raw violation  search_conducted search_type
stop_outcome  \
0      Speeding  Speeding             False         NaN     Citation

1      Speeding  Speeding             False         NaN     Citation


  is_arrested  stop_duration  drugs_related_stop
0       False            7.5               False
1       False            7.5               False

df.groupby("violation").driver_age.describe().round(2)

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

Equipment             6507.0  31.68  11.38  16.0  23.0  28.0  39.0
81.0
Moving violation     11876.0  36.74  13.26  15.0  25.0  35.0  47.0
86.0
Other                 3477.0  40.36  12.75  16.0  30.0  41.0  50.0
86.0
Registration/plates   2240.0  32.66  11.15  16.0  24.0  30.0  40.0
74.0
Seat belt                3.0  30.33  10.21  23.0  24.5  26.0  34.0
42.0
Speeding             37120.0  33.26  12.62  15.0  23.0  30.0  42.0
88.0
```