```
import pandas as pd
    df = pd.read_csv(r'C:\Users\PC-chetan\Downloads\3. Police Data.csv')
    df

Out[1]:
    stop_date stop_time country_name driver_gender driver_age_raw driver_age driver_race
```

]:		stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
	0	1/2/2005	1:55	NaN	М	1985.0	20.0	White
	1	1/18/2005	8:15	NaN	М	1965.0	40.0	White
	2	1/23/2005	23:15	NaN	М	1972.0	33.0	White
	3	2/20/2005	17:15	NaN	M	1986.0	19.0	White
	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	М	1985.0	27.0	Black
	65533	12/7/2012	0:23	NaN	NaN	NaN	NaN	NaN
	65534	12/7/2012	0:30	NaN	F	1985.0	27.0	White

 $65535 \text{ rows} \times 15 \text{ columns}$ 

## **Data Exploration**

```
In [2]:
              df.head()
                              stop_time country_name driver_gender driver_age_raw driver_age driver_race
   Out[2]:
                 stop_date
              0
                   1/2/2005
                                    1:55
                                                      NaN
                                                                          Μ
                                                                                        1985.0
                                                                                                        20.0
                                                                                                                     White
                 1/18/2005
                                    8:15
                                                      NaN
                                                                                        1965.0
                                                                                                        40.0
                                                                                                                     White
                 1/23/2005
                                                      NaN
                                                                                        1972.0
                                                                                                        33.0
                                                                                                                     White
              2
                                   23:15
                                                                          Μ
                                                                                                        19.0
                                                                                                                     White Call
              3
                  2/20/2005
                                   17:15
                                                      NaN
                                                                          Μ
                                                                                        1986.0
                  3/14/2005
                                                      NaN
                                                                          F
                                                                                        1984.0
                                                                                                        21.0
                                                                                                                     White
                                   10:00
   In [3]:
              df.columns
             Index(['stop_date', 'stop_time', 'country_name', 'driver_gender',
   Out[3]:
                       'driver_age_raw', 'driver_age', 'driver_race', 'violation_raw', 'violation', 'search_conducted', 'search_type', 'stop_outcome', 'is_arrested', 'stop_duration', 'drugs_related_stop'],
                     dtype='object')
   In [4]:
              df.info()
             <class 'pandas.core.frame.DataFrame'>
             RangeIndex: 65535 entries, 0 to 65534
             Data columns (total 15 columns):
                   Column
                                             Non-Null Count Dtype
Loading [MathJax]/extensions/Safe.js
                                             65535 non-null object
```

```
2
                                  0 non-null
             country name
                                                   float64
         3
             driver_gender
                                  61474 non-null object
         4
                                  61481 non-null
             driver_age_raw
                                                  float64
         5
             driver age
                                  61228 non-null
                                                  float64
         6
                                  61475 non-null
                                                   object
             driver_race
         7
                                  61475 non-null
                                                   object
             violation raw
         8
             violation
                                  61475 non-null
                                                   object
         9
             search conducted
                                  65535 non-null
                                                   bool
         10 search_type
                                  2479 non-null
                                                   object
         11 stop outcome
                                  61475 non-null
                                                   object
         12
             is arrested
                                  61475 non-null
                                                   object
         13
             stop duration
                                  61475 non-null
                                                   object
             drugs_related_stop 65535 non-null
                                                   bool
        dtypes: bool(2), float64(3), object(10)
        memory usage: 6.6+ MB
In [5]:
         df.isnull().sum()
                                   0
        stop date
Out[5]:
                                   0
        stop time
                               65535
        country_name
                                4061
        driver gender
                                4054
        driver_age_raw
        driver age
                                4307
        driver race
                                4060
        violation raw
                                4060
                                4060
        violation
        search conducted
                                   0
        search type
                               63056
                                4060
        stop_outcome
        is_arrested
                                4060
                                4060
        stop duration
        drugs related stop
                                   0
        dtype: int64
In [6]:
         df.notnull().sum()
                               65535
        stop date
Out[6]:
        stop time
                               65535
        country name
                                   0
        driver_gender
                               61474
        driver_age_raw
                               61481
                               61228
        driver age
                               61475
        driver race
        violation raw
                               61475
        violation
                               61475
        search conducted
                               65535
        search type
                                2479
        stop_outcome
                               61475
        is arrested
                               61475
                               61475
        stop_duration
        drugs_related_stop
                               65535
        dtype: int64
In [7]:
         df.describe()
Out[7]:
               country_name driver_age_raw
                                               driver_age
                         0.0
                                             61228.000000
         count
                                61481.000000
         mean
                         NaN
                                 1967.791106
                                                34.148984
```

65535 non-null

object

Loading [MathJax]/extensions/Safe.js

1

stop\_time

	Sta	IVAIV	121.050100	12.700710	
	min	NaN	0.000000	15.000000	
	25%	NaN	1965.000000	23.000000	
	50%	NaN	1978.000000	31.000000	
	<b>75</b> %	NaN	1985.000000	43.000000	
	max	NaN	8801.000000	88.000000	
In [8]:	df.shape				
Out[8]:	(65535, 15)				
In [9]:	df.index				
Out[9]:	RangeIndex(sta	rt=0, s	top=65535, step	=1)	
In [10]:	df.count()				
Out[10]:	stop_date stop_time country_name driver_gender driver_age_raw driver_age driver_race violation_raw violation search_conduct search_type stop_outcome is_arrested stop_duration drugs_related_s dtype: int64		65535 65535 0 61474 61481 61228 61475 61475 65535 2479 61475 61475 61475 61475		
In [11]:	df.violation.v	/alue_co	ounts()		
Out[11]:	Speeding Moving violation Equipment Other Registration/poseat belt Name: violation	lates	37204 11926 6516 3583 2243 3 e: int64		
In [12]:	df.violation.u	unique(			
Out[12]:	array(['Speedi	ng', '0 ration/	ther', 'Equipme plates', 'Seat	nt', 'Moving ' belt'], dtype	violation', nan, e=object)
In [13]:	df.violation.r	nunique	()		
Out[13]:	6				
	lax1/extensions/Safe	is			

**driver\_age** 12.760710

 ${\color{red} \textbf{country\_name}} \quad {\color{red} \textbf{driver\_age\_raw}}$ 

121.050106

NaN

std

Loading [MathJax]/extensions/Safe.js

```
In [ ]:
```

Loading [MathJax]/extensions/Safe.js

## **Data Analisys**

```
In [14]:
            df.head()
                          stop_time country_name driver_gender driver_age_raw driver_age driver_race
Out[14]:
              stop_date
           0
                1/2/2005
                                1:55
                                                NaN
                                                                               1985.0
                                                                                              20.0
                                                                                                          White
               1/18/2005
                                8:15
                                                NaN
                                                                   Μ
                                                                               1965.0
                                                                                              40.0
                                                                                                          White
               1/23/2005
                               23:15
                                                NaN
                                                                   Μ
                                                                               1972.0
                                                                                              33.0
                                                                                                          White
           3
                               17:15
                                                NaN
                                                                   Μ
                                                                               1986.0
                                                                                              19.0
                                                                                                          White
                                                                                                                 Call
               2/20/2005
                                                                   F
               3/14/2005
                               10:00
                                                NaN
                                                                               1984.0
                                                                                              21.0
                                                                                                          White
In [15]:
            df.violation.unique()
           array(['Speeding', 'Other', 'Equipment', 'Moving violation', nan,
Out[15]:
                   'Registration/plates', 'Seat belt'], dtype=object)
In [16]:
            df[df['driver race'] == 'White']
Out[16]:
                   stop_date stop_time country_name driver_gender driver_age_raw driver_age driver_race
                     1/2/2005
                0
                                     1:55
                                                     NaN
                                                                       Μ
                                                                                    1985.0
                                                                                                   20.0
                                                                                                               White
                1
                    1/18/2005
                                     8:15
                                                     NaN
                                                                                    1965.0
                                                                                                   40.0
                                                                                                               White
                    1/23/2005
                                    23:15
                                                     NaN
                                                                       Μ
                                                                                    1972.0
                                                                                                   33.0
                                                                                                               White
                                                                                                   19.0
                                                                                                               White
                3
                    2/20/2005
                                    17:15
                                                     NaN
                                                                       Μ
                                                                                    1986.0
                                                                        F
                    3/14/2005
                                    10:00
                                                     NaN
                                                                                    1984.0
                                                                                                   21.0
                                                                                                               White
                                                                                                   25.0
                    12/6/2012
                                                                       Μ
                                                                                                               White
           65528
                                    15:39
                                                     NaN
                                                                                    1987.0
           65529
                    12/6/2012
                                    16:00
                                                     NaN
                                                                                    1994.0
                                                                                                   18.0
                                                                                                               White
           65530
                    12/6/2012
                                    17:54
                                                     NaN
                                                                        F
                                                                                    1987.0
                                                                                                   25.0
                                                                                                               White
                                                                                    1954.0
                                                                                                               White
           65531
                    12/6/2012
                                    22:22
                                                     NaN
                                                                       Μ
                                                                                                   58.0
                    12/7/2012
                                     0:30
                                                     NaN
                                                                        F
                                                                                                   27.0
                                                                                                               White
           65534
                                                                                    1985.0
          45747 \text{ rows} \times 15 \text{ columns}
In [17]:
            df[df['driver age'] == 20]
                              stop_time
                                           country_name
                                                           driver_gender
                                                                           driver_age_raw
                                                                                            driver_age
                                                                                                         driver_race
Out[17]:
                   stop_date
                0
                     1/2/2005
                                     1:55
                                                                        Μ
                                                                                                    20.0
                                                                                                               White
                                                      NaN
                                                                                     1985.0
               43
                    10/1/2005
                                     1:00
                                                      NaN
                                                                                     1985.0
                                                                                                    20.0
                                                                                                                White
               73
                    10/2/2005
                                     6:10
                                                      NaN
                                                                        M
                                                                                     1985.0
                                                                                                    20.0
                                                                                                                White
             125
                    10/4/2005
                                    11:50
                                                                                     1985.0
                                                                                                    20.0
                                                                                                                White
                                                      NaN
```

	65389	11/30/2012	15:54	NaN	F	1992.0	20.0	White
	65472	12/4/2012	11:20	NaN	М	1992.0	20.0	White
	65485	12/5/2012	1:26	NaN	М	1992.0	20.0	Black
	65513	12/6/2012	10:10	NaN	F	1992.0	20.0	White
	65521	12/6/2012	14:00	NaN	F	1992.0	20.0	White
	2783 ro	ws × 15 co	umns					
In [18]:	df.isr	null().sum(	)					
Out[18]:	driver_driver_violat: search_search_stop_ouis_arrestop_du	ime y_name y_name _gender _age_raw _race ion_raw ion _conducted _type utcome ested uration related_sto	65535 4063 4054 4307 4066 4066 4066 4066 4066 4066	L 4 7 9 9 9 9 9 9 9				
In [19]:	df.col	umns						
Out[19]:		'driver_aq' 'violation	ge_raw', 'd n', 'search ced', 'stop	driver_age', ' n_conducted',		<pre>'violation_raw 'stop_outcome</pre>		
In [20]:	df.rer	name(columr	s= {'viola	ntion': 'vio'}	, inplace =Tr	ue)		
In [21]:	df							
Out[21]:		stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
	0	1/2/2005	1:55	NaN	М	1985.0	20.0	White
	1	1/18/2005	8:15	NaN	М	1965.0	40.0	White
	2	1/23/2005	23:15	NaN	М	1972.0	33.0	White
	3	2/20/2005	17:15	NaN	М	1986.0	19.0	White
	4	3/14/2005	10:00	NaN	F	1984.0	21.0	White
Loading [Math	Jax]/exten	sions/Safe.js						

 $stop\_date \quad stop\_time \quad country\_name \quad driver\_gender \quad driver\_age\_raw \quad driver\_age \quad driver\_race$ 

NaN

1985.0

20.0

White

10/4/2005

137

23:10

		stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
	65530	12/6/2012	17:54	NaN	F	1987.0	25.0	White
	65531	12/6/2012	22:22	NaN	М	1954.0	58.0	White
	65532	12/6/2012	23:20	NaN	М	1985.0	27.0	Black
	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 r	rows × 15 c	olumns					
In [22]:	df.dr:	iver_age.me	ean()					
Out[22]:	34.148	9841249101	7					
In [23]:	df.dr	iver_age.va	ar()					
Out [23] :	162.83	5725526525	38					

Out[23]:

In [24]: df[df['vio'] == 'Equipment']

ut[24]:		stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
	5	3/23/2005	9:45	NaN	М	1982.0	23.0	Black
	17	7/19/2005	23:30	NaN	М	1979.0	26.0	White
	35	9/30/2005	0:15	NaN	М	1978.0	27.0	White
	40	10/1/2005	0:00	NaN	М	1988.0	17.0	White
	41	10/1/2005	0:00	NaN	М	1988.0	17.0	White
	65520	12/6/2012	12:43	NaN	М	1953.0	59.0	White
	65525	12/6/2012	15:09	NaN	М	1980.0	32.0	White
	65526	12/6/2012	15:22	NaN	М	1982.0	30.0	White
	65528	12/6/2012	15:39	NaN	М	1987.0	25.0	White
	65532	12/6/2012	23:20	NaN	М	1985.0	27.0	Black

 $6516 \text{ rows} \times 15 \text{ columns}$ 

```
In [25]:
          df[(df['stop\_time'] < '23:20') & (df['driver\_age'] > 23)]
```

stop date stop\_time country\_name driver\_gender driver\_age\_raw driver\_age driver\_race Out[25]: Loading [MathJax]/extensions/Safe.js

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race
2	1/23/2005	23:15	NaN	М	1972.0	33.0	White
6	4/1/2005	17:30	NaN	М	1969.0	36.0	White
8	7/13/2005	10:15	NaN	М	1970.0	35.0	Black
9	7/13/2005	15:45	NaN	M	1970.0	35.0	White
10	7/13/2005	16:20	NaN	M	1979.0	26.0	Asian
65527	12/6/2012	15:26	NaN	F	1981.0	31.0	White
65528	12/6/2012	15:39	NaN	М	1987.0	25.0	White
65530	12/6/2012	17:54	NaN	F	1987.0	25.0	White
65531	12/6/2012	22:22	NaN	М	1954.0	58.0	White
65534	12/7/2012	0:30	NaN	F	1985.0	27.0	White

30739 rows  $\times$  15 columns

In [26]:

df.head()

Out[26]:

:	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	driver_race	vio
0	1/2/2005	1:55	NaN	М	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	

In [27]:

df.groupby('driver\_age').mean()

Out[27]:

	country_name	driver_age_raw	search_conducted	drugs_related_stop
driver_age				
15.0	NaN	1992.600000	0.200000	0.000000
16.0	NaN	1992.088235	0.088235	0.029412
17.0	NaN	1991.144766	0.031180	0.008909
18.0	NaN	1990.211896	0.044610	0.013383
19.0	NaN	1989.463568	0.049414	0.016332
83.0	NaN	1926.000000	0.000000	0.000000
84.0	NaN	1925.666667	0.000000	0.000000
85.0	NaN	1927.000000	0.000000	0.000000
86.0	NaN	1923.166667	0.000000	0.000000
88.0	NaN	1920.500000	0.000000	0.000000

73 rows × 4 columns

Loading [MathJax]/extensions/Safe.js

In [28]: df.groupby('driver age').min()

C:\Users\PC-CHE~1\AppData\Local\Temp/ipykernel\_6736/3430904074.py:1: FutureWarning: Droppi ng invalid columns in DataFrameGroupBy.min is deprecated. In a future version, a TypeError will be raised. Before calling .min, select only columns which should be valid for the fun ction.

df.groupby('driver\_age').min()

Out[28]: stop\_date stop\_time country\_name driver\_age\_raw search\_conducted is\_arrested dr

driver_age						
15.0	11/10/2010	0:40	NaN	1991.0	False	False
16.0	1/10/2006	0:01	NaN	1989.0	False	False
17.0	1/1/2008	0:00	NaN	1988.0	False	False
18.0	1/10/2007	0:00	NaN	1987.0	False	False
19.0	1/1/2007	0:00	NaN	1986.0	False	False
83.0	3/19/2007	10:50	NaN	1924.0	False	False
84.0	1/24/2009	11:14	NaN	1925.0	False	False
85.0	2/22/2012	13:58	NaN	1927.0	False	False
86.0	11/17/2009	10:07	NaN	1922.0	False	False
88.0	6/20/2011	12:05	NaN	1918.0	False	False

 $73 \text{ rows} \times 7 \text{ columns}$ 

In [29]: df.groupby('driver\_age').max()

C:\Users\PC-CHE~1\AppData\Local\Temp/ipykernel\_6736/1046300893.py:1: FutureWarning: Droppi ng invalid columns in DataFrameGroupBy.max is deprecated. In a future version, a TypeError will be raised. Before calling .max, select only columns which should be valid for the function.

df.groupby('driver age').max()

Out[29]: stop\_date stop\_time country\_name driver\_age\_raw search\_conducted is\_arrested dri

driver_age						
15.0	9/30/2006	1:30	NaN	1995.0	True	True
16.0	9/16/2008	7:20	NaN	1996.0	True	True
17.0	9/9/2007	9:51	NaN	1995.0	True	True
18.0	9/9/2012	9:59	NaN	1994.0	True	True
19.0	9/9/2012	9:58	NaN	1993.0	True	True
83.0	9/28/2011	8:20	NaN	1928.0	False	False
84.0	9/21/2011	7:04	NaN	1927.0	False	False
85.0	2/22/2012	13:58	NaN	1927.0	False	False
86.0	4/5/2008	18:32	NaN	1925.0	False	False
88.0	9/6/2006	12:15	NaN	1923.0	False	False

```
In [30]:
          df.isnull().sum()
         stop_date
                                     0
Out[30]:
                                     0
         stop time
                                 65535
         country_name
         driver_gender
                                  4061
                                  4054
         driver age raw
         driver_age
                                  4307
         driver race
                                  4060
                                  4060
         violation raw
                                  4060
                                     0
         search conducted
         search_type
                                 63056
         stop outcome
                                  4060
                                  4060
         is arrested
                                  4060
         stop duration
         drugs related stop
                                     0
         dtype: int64
In [31]:
          df.drop(columns= ['country name'], inplace =True)
In [32]:
          df.columns
         Index(['stop_date', 'stop_time', 'driver_gender', 'driver_age_raw',
Out[321:
                 'driver_age', 'driver_race', 'violation_raw', 'vio', 'search_conducted',
                 'search_type', 'stop_outcome', 'is_arrested', 'stop_duration',
                 'drugs_related_stop'],
                dtype='object')
In [33]:
          df.isnull().sum()
         stop_date
Out[33]:
                                     0
         stop time
                                  4061
         driver_gender
         driver_age_raw
                                  4054
                                  4307
         driver age
         driver_race
                                  4060
                                  4060
         violation raw
                                  4060
         vio
         {\sf search\_conducted}
                                     0
                                 63056
         search type
                                  4060
         stop_outcome
         is arrested
                                  4060
                                  4060
         stop duration
         drugs_related_stop
         dtype: int64
In [34]:
          df.driver gender.value counts()
               45164
Out[34]:
               16310
         Name: driver gender, dtype: int64
In [35]:
          df.driver_gender.isnull().sum()
         4061
Out[35]:
```

```
In [37]:
           df.driver gender.isnull().sum()
Out[37]:
In [38]:
           df.isnull().sum()
                                        0
          stop date
Out[38]:
                                        0
          stop time
                                        0
          driver_gender
                                     4054
          driver_age_raw
          driver age
                                     4307
                                     4060
          driver race
          violation raw
                                     4060
                                     4060
          search_conducted
                                        0
                                    63056
          search type
                                     4060
          stop outcome
                                     4060
          is arrested
          stop_duration
                                     4060
                                        0
          drugs_related_stop
          dtype: int64
In [39]:
           df['driver age raw'][df.driver age raw.min()]
          1985.0
Out[39]:
In [40]:
           df[df['driver age raw'] == 1985]
                                                                                                         violation
Out[40]:
                  stop_date stop_time driver_gender driver_age_raw driver_age driver_race
               0
                    1/2/2005
                                    1:55
                                                      Μ
                                                                  1985.0
                                                                                 20.0
                                                                                            White
                                                                                                              Spe
              43
                   10/1/2005
                                    1:00
                                                      Μ
                                                                  1985.0
                                                                                 20.0
                                                                                            White
                                                                                                              Spe
              73
                   10/2/2005
                                    6:10
                                                      Μ
                                                                  1985.0
                                                                                 20.0
                                                                                            White
                                                                                                              Spe
             125
                   10/4/2005
                                   11:50
                                                                                 20.0
                                                                                            White
                                                      Μ
                                                                  1985.0
                                                                                                              Spe
             137
                   10/4/2005
                                  23:10
                                                      Μ
                                                                  1985.0
                                                                                 20.0
                                                                                            White
                                                                                                              Spe
                                      ...
                                                      ...
                                                                                                   Equipment/Inspe
          65499
                   12/5/2012
                                  23:13
                                                      Μ
                                                                  1985.0
                                                                                 27.0
                                                                                             Black
                                                                                                               Vio
          65509
                   12/6/2012
                                    7:13
                                                      Μ
                                                                  1985.0
                                                                                 27.0
                                                                                            White
                                                                                                              Spe
          65519
                   12/6/2012
                                   11:40
                                                      F
                                                                  1985.0
                                                                                 27.0
                                                                                          Hispanic
                                                                                                              Spe
                                                                                                   Equipment/Inspe
          65532
                   12/6/2012
                                  23:20
                                                      Μ
                                                                  1985.0
                                                                                 27.0
                                                                                            Black
                                                                                                               Vio
                                                      F
          65534
                   12/7/2012
                                    0:30
                                                                  1985.0
                                                                                 27.0
                                                                                            White
                                                                                                              Spe
         2630 rows \times 14 columns
```

df.driver\_gender.fillna('F', inplace = True)

Loading [MathJax]/extensions/Safe.js

df.driver age raw.fillna(1985, inplace= True)

In [41]:

```
stop_date
                                       0
Out[42]:
                                       0
          stop time
          driver gender
                                       0
          driver_age_raw
                                       0
                                    4307
          driver age
                                    4060
          driver_race
                                    4060
          violation raw
                                    4060
          vio
          search conducted
                                       0
                                   63056
          search type
          stop_outcome
                                    4060
                                    4060
          is arrested
                                    4060
          stop duration
          drugs related stop
          dtype: int64
In [43]:
           df1 = df.groupby('driver age').min()
          C:\Users\PC-CHE~1\AppData\Local\Temp/ipykernel_6736/1525749074.py:1: FutureWarning: Droppi
          ng invalid columns in DataFrameGroupBy.min is deprecated. In a future version, a TypeError
          will be raised. Before calling .min, select only columns which should be valid for the fun
          ction.
            df1 = df.groupby('driver age').min()
In [44]:
           df1
Out[44]:
                       stop_date stop_time driver_gender driver_age_raw search_conducted is_arrested dr
          driver_age
                15.0 11/10/2010
                                       0:40
                                                         F
                                                                     1991.0
                                                                                          False
                                                                                                      False
                                                          F
                16.0
                                       0:01
                                                                                                      False
                       1/10/2006
                                                                     1989.0
                                                                                          False
                17.0
                                                         F
                        1/1/2008
                                       0:00
                                                                     1988.0
                                                                                          False
                                                                                                      False
                18.0
                        1/10/2007
                                       0:00
                                                          F
                                                                     1987.0
                                                                                          False
                                                                                                      False
                                                         F
                19.0
                        1/1/2007
                                       0:00
                                                                     1986.0
                                                                                          False
                                                                                                      False
                       3/19/2007
                83.0
                                      10:50
                                                         Μ
                                                                     1924.0
                                                                                          False
                                                                                                      False
                84.0
                       1/24/2009
                                      11:14
                                                                     1925.0
                                                                                          False
                                                                                                      False
                85.0
                       2/22/2012
                                      13:58
                                                         Μ
                                                                     1927.0
                                                                                          False
                                                                                                      False
                86.0
                      11/17/2009
                                      10:07
                                                         M
                                                                     1922.0
                                                                                          False
                                                                                                      False
                88.0
                       6/20/2011
                                      12:05
                                                                     1918.0
                                                                                          False
                                                                                                      False
                                                         М
         73 rows \times 7 columns
In [45]:
           df.driver age.fillna(15, inplace =True)
In [46]:
           df.driver age.isnull().sum()
Out[46]:
```

df.isnull().sum()

Loading [MathJax]/extensions/Safe.js )

In [42]:

```
0
          stop date
Out[47]:
          stop time
                                        0
                                        0
          driver_gender
                                        0
          driver age raw
                                        0
          driver age
          driver race
                                     4060
                                     4060
          violation raw
          vio
                                     4060
                                        0
          search conducted
          search type
                                    63056
                                     4060
          stop outcome
                                     4060
          is arrested
          stop_duration
                                     4060
          drugs related stop
                                        0
          dtype: int64
In [48]:
           df.groupby('driver race').count()
                        stop_date stop_time driver_gender driver_age_raw driver_age violation_raw
                                                                                                              vio
Out[48]:
           driver_race
                             1639
                                         1639
                                                                         1639
                                                                                      1639
                                                                                                            1639
                Asian
                                                        1639
                                                                                                     1639
                Black
                             8267
                                         8267
                                                        8267
                                                                         8267
                                                                                      8267
                                                                                                     8267
                                                                                                             8267
             Hispanic
                             5611
                                         5611
                                                        5611
                                                                         5611
                                                                                      5611
                                                                                                     5611
                                                                                                            5611
                Other
                              211
                                          211
                                                          211
                                                                           211
                                                                                       211
                                                                                                      211
                                                                                                              211
                White
                            45747
                                       45747
                                                       45747
                                                                        45747
                                                                                     45747
                                                                                                    45747 45747
In [49]:
           df.driver race.fillna('Other', inplace = True)
In [50]:
           df
Out[50]:
                   stop_date stop_time driver_gender driver_age_raw driver_age driver_race
                                                                                                          violation
               0
                    1/2/2005
                                    1:55
                                                                   1985.0
                                                                                 20.0
                                                                                             White
                                                                                                               Spe
                                                      Μ
                   1/18/2005
                                    8:15
                                                      Μ
                                                                   1965.0
                                                                                 40.0
                                                                                             White
                                                                                                               Spe
               2
                   1/23/2005
                                   23:15
                                                      Μ
                                                                   1972.0
                                                                                 33.0
                                                                                             White
                                                                                                               Spe
                   2/20/2005
               3
                                   17:15
                                                      Μ
                                                                   1986.0
                                                                                 19.0
                                                                                             White
                                                                                                          Call for Se
                                   10:00
                                                       F
                                                                   1984.0
                                                                                 21.0
                                                                                             White
                                                                                                               Spe
               4
                   3/14/2005
                                                       F
           65530
                   12/6/2012
                                   17:54
                                                                   1987.0
                                                                                 25.0
                                                                                             White
                                                                                                               Spe
           65531
                   12/6/2012
                                   22:22
                                                      Μ
                                                                   1954.0
                                                                                 58.0
                                                                                             White
                                                                                                               Spe
                                                                                                    Equipment/Inspe
                                                                                              Black
           65532
                   12/6/2012
                                   23:20
                                                      Μ
                                                                   1985.0
                                                                                 27.0
           65533
                   12/7/2012
                                    0:23
                                                       F
                                                                   1985.0
                                                                                 15.0
                                                                                             Other
                                    0:30
                                                       F
                                                                                 27.0
           65534
                   12/7/2012
                                                                   1985.0
                                                                                             White
                                                                                                               Spe
          65535 \text{ rows} \times 14 \text{ columns}
```

In [51]: df ispull() sum()
Loading [MathJax]/extensions/Safe.js

Out[51]:  In [52]:	driver_driver_violativio search_stop_ouis_arrestop_dudrugs_udtype:	ime _gender _age_raw _age _race ion_raw _conducted _type utcome ested uration related_sto	4060 4060 4060 4060 4060 4060	9 9 5 9 9				
In [52]:  Out[52]:  In [53]:  In [54]:	Speeding Other Traffic Violation Equipment/Inspection Violation Special Detail/Directed Patrol Registration Violation Call for Service Violation of City/Town Ordinance Motorist Assist/Courtesy APB Suspicious Person Warrant Seatbelt Violation Name: violation_raw, dtype: int64  df.violation_raw.fillna('Seatbelt				4 6 6 3 3 3 0 8 1 8 0 3 <b>tion'</b> , inplace=	True)		
Ou+[E/1].		stop data	stan tima	driver gender	driver age raw	driver age	drivor raco	violation
Out[54]:					driver_age_raw			violation
Out[54]:	0	1/2/2005	1:55	М	1985.0	20.0	White	Spe
Out[54]:	0 1 2	1/2/2005	1:55 8:15	M M	1985.0 1965.0	20.0	White White	Spe Spe
Out[54]:	1	1/2/2005	1:55	М	1985.0	20.0	White	Spe
Out[54]:	1 2	1/2/2005 1/18/2005 1/23/2005	1:55 8:15 23:15	M M M	1985.0 1965.0 1972.0	20.0 40.0 33.0	White White White	Spe Spe Spe
Out[54]:	1 2 3	1/2/2005 1/18/2005 1/23/2005 2/20/2005	1:55 8:15 23:15 17:15	M M M	1985.0 1965.0 1972.0 1986.0	20.0 40.0 33.0 19.0	White White White White	Spe Spe Spe Call for Se
Out[54]:	1 2 3 4	1/2/2005 1/18/2005 1/23/2005 2/20/2005 3/14/2005	1:55 8:15 23:15 17:15 10:00	M M M M	1985.0 1965.0 1972.0 1986.0 1984.0	20.0 40.0 33.0 19.0 21.0	White White White White	Spe Spe Spe Call for Se
Out[54]:	1 2 3 4	1/2/2005 1/18/2005 1/23/2005 2/20/2005 3/14/2005	1:55 8:15 23:15 17:15 10:00	M M M M F	1985.0 1965.0 1972.0 1986.0 1984.0	20.0 40.0 33.0 19.0 21.0	White White White White White	Spe Spe Spe Call for Se Spe
Out[54]:	1 2 3 4  65530	1/2/2005 1/18/2005 1/23/2005 2/20/2005 3/14/2005  12/6/2012	1:55 8:15 23:15 17:15 10:00 	M M M M F	1985.0 1965.0 1972.0 1986.0 1984.0 	20.0 40.0 33.0 19.0 21.0 	White White White White White White White	Spe Spe Spe Call for Se Spe
Out[54]:	1 2 3 4  65530 65531	1/2/2005 1/18/2005 1/23/2005 2/20/2005 3/14/2005  12/6/2012 12/6/2012	1:55 8:15 23:15 17:15 10:00  17:54 22:22	M M M F  F	1985.0 1965.0 1972.0 1986.0 1984.0  1987.0	20.0 40.0 33.0 19.0 21.0  25.0 58.0	White White White White White White White	Spe Spe Call for Se Spe Spe Spe Spe Equipment/Inspe
Out[54]:	1 2 3 4  65530 65531	1/2/2005 1/18/2005 1/23/2005 2/20/2005 3/14/2005  12/6/2012 12/6/2012	1:55 8:15 23:15 17:15 10:00  17:54 22:22 23:20	M M M M F  F M	1985.0 1965.0 1972.0 1986.0 1984.0  1987.0 1954.0	20.0 40.0 33.0 19.0 21.0  25.0 58.0 27.0	White White White White White White White Black	Spe Spe Call for Se Spe Spe Spe Equipment/Inspe Vio

Loading [MathJax]/extensions/Safe.js )

```
stop date
                                          0
 Out[55]:
                                          0
            stop time
                                          0
            driver gender
            driver age raw
                                          0
            driver_age
                                          0
                                          0
            driver race
                                          0
            violation raw
                                       4060
            vio
                                          0
            search conducted
                                      63056
            search type
                                       4060
            stop_outcome
                                       4060
            is arrested
            stop_duration
                                       4060
            drugs related stop
            dtype: int64
 In [56]:
             df.vio.value counts()
            Speeding
                                       37204
 Out[56]:
            Moving violation
                                       11926
            Equipment
                                        6516
            0ther
                                        3583
            Registration/plates
                                        2243
            Seat belt
                                           3
            Name: vio, dtype: int64
 In [57]:
             df.vio.fillna('Seat belt', inplace=True)
 In [58]:
             df
                    stop_date stop_time driver_gender driver_age_raw driver_age driver_race
                                                                                                            violation
 Out[58]:
                 0
                      1/2/2005
                                      1:55
                                                        Μ
                                                                    1985.0
                                                                                   20.0
                                                                                               White
                                                                                                                 Spe
                     1/18/2005
                                      8:15
                                                        Μ
                                                                    1965.0
                                                                                   40.0
                                                                                               White
                                                                                                                 Spe
                 2
                     1/23/2005
                                     23:15
                                                        Μ
                                                                    1972.0
                                                                                   33.0
                                                                                               White
                                                                                                                 Spe
                     2/20/2005
                                     17:15
                                                                    1986.0
                                                                                   19.0
                                                                                               White
                                                                                                            Call for Se
                                                        F
                     3/14/2005
                                     10:00
                                                                    1984.0
                                                                                   21.0
                                                                                               White
                                                                                                                 Spe
            65530
                     12/6/2012
                                     17:54
                                                        F
                                                                    1987.0
                                                                                   25.0
                                                                                               White
                                                                                                                 Spe
            65531
                     12/6/2012
                                     22:22
                                                                                   58.0
                                                                                               White
                                                        Μ
                                                                    1954.0
                                                                                                                 Spe
                                                                                                      Equipment/Inspe
            65532
                                                                                               Black
                     12/6/2012
                                     23:20
                                                        Μ
                                                                    1985.0
                                                                                   27.0
                                                        F
                                                                                   15.0
            65533
                     12/7/2012
                                      0:23
                                                                    1985.0
                                                                                               Other
                                                                                                         Seatbelt Vio
                                                        F
                                      0:30
                                                                                   27.0
                                                                                               White
            65534
                     12/7/2012
                                                                    1985.0
                                                                                                                 Spe
            65535 \text{ rows} \times 14 \text{ columns}
 In [59]:
             df.isnull().sum()
            stop date
                                          0
 Out[59]:
                                          0
            stop time
            driver_gender
                                          0
            driver and raw
                                          0
Loading [MathJax]/extensions/Safe.js
```

```
0
           driver race
                                      0
           violation raw
                                      0
           search conducted
                                      0
                                  63056
           search type
                                   4060
           stop outcome
                                   4060
           is arrested
                                   4060
           stop duration
                                      0
           drugs related stop
           dtype: int64
 In [60]:
            df.search type.value counts()
           Incident to Arrest
                                                                          1074
 Out[60]:
           Probable Cause
                                                                           579
           Inventorv
                                                                           173
           Reasonable Suspicion
                                                                           149
           Protective Frisk
                                                                           126
                                                                           104
           Incident to Arrest, Inventory
           Incident to Arrest, Probable Cause
                                                                            79
           Probable Cause, Reasonable Suspicion
                                                                            38
           Incident to Arrest, Protective Frisk
                                                                            31
                                                                            26
           Incident to Arrest, Inventory, Probable Cause
           Probable Cause, Protective Frisk
                                                                            23
           Inventory, Probable Cause
                                                                            14
           Inventory, Protective Frisk
                                                                            11
           Incident to Arrest, Inventory, Protective Frisk
                                                                            11
           Incident to Arrest, Probable Cause, Protective Frisk
                                                                            10
                                                                            10
           Protective Frisk, Reasonable Suspicion
           Incident to Arrest, Reasonable Suspicion
                                                                             8
                                                                             4
           Inventory, Reasonable Suspicion
                                                                             3
           Incident to Arrest, Probable Cause, Reasonable Suspicion
           Incident to Arrest, Inventory, Reasonable Suspicion
                                                                             2
           Inventory, Probable Cause, Protective Frisk
                                                                             2
           Incident to Arrest, Protective Frisk, Reasonable Suspicion
                                                                             1
           Inventory, Probable Cause, Reasonable Suspicion
                                                                             1
           Name: search type, dtype: int64
 In [61]:
            df.drop(columns = ['search type'], inplace = True)
 In [62]:
            df.columns
           Index(['stop date', 'stop time', 'driver gender', 'driver age raw',
 Out[62]:
                   'driver age', 'driver race', 'violation raw', 'vio', 'search conducted',
                   'stop_outcome', 'is_arrested', 'stop_duration', 'drugs_related_stop'],
                 dtype='object')
 In [63]:
            df.isnull().sum()
           stop_date
                                     0
 Out[63]:
                                     0
           stop time
                                     0
           driver gender
           driver_age_raw
                                     0
                                     0
           driver age
                                     0
           driver race
                                     0
           violation raw
                                     0
           vio
           search conducted
                                     0
           stop outcome
                                  4060
           ic arracted
                                  4060
Loading [MathJax]/extensions/Safe.js
```

0

driver\_age

```
drugs related stop
                                      0
          dtype: int64
In [64]:
           df.stop outcome.value counts()
                                 56564
          Citation
Out[64]:
          Arrest Driver
                                  2009
          Warning
                                  1889
          N/D
                                   423
          No Action
                                   339
          Arrest Passenger
                                   251
          Name: stop_outcome, dtype: int64
In [65]:
           df.stop duration.value counts()
                         47379
          0-15 Min
Out[65]:
          16-30 Min
                         11448
                          2647
          30+ Min
          Name: stop duration, dtype: int64
In [66]:
           df.is arrested.value counts()
          False
                    59215
Out[66]:
          True
                     2260
          Name: is arrested, dtype: int64
In [67]:
           df.stop outcome.fillna('Arrest Passenger', inplace=True)
In [68]:
           df.stop_duration.fillna(2, inplace=True)
In [69]:
           df.is arrested.fillna(True , inplace=True)
In [70]:
           df
Out[70]:
                  stop_date stop_time driver_gender driver_age_raw driver_age driver_race
                                                                                                       violation
               0
                    1/2/2005
                                   1:55
                                                     Μ
                                                                 1985.0
                                                                               20.0
                                                                                           White
                                                                                                            Spe
               1
                   1/18/2005
                                   8:15
                                                     Μ
                                                                 1965.0
                                                                               40.0
                                                                                           White
                                                                                                            Spe
               2
                   1/23/2005
                                  23:15
                                                     Μ
                                                                 1972.0
                                                                               33.0
                                                                                           White
                                                                                                            Spe
               3
                   2/20/2005
                                  17:15
                                                     Μ
                                                                 1986.0
                                                                               19.0
                                                                                           White
                                                                                                       Call for Se
               4
                   3/14/2005
                                  10:00
                                                     F
                                                                 1984.0
                                                                               21.0
                                                                                           White
                                                                                                            Spe
          65530
                   12/6/2012
                                  17:54
                                                     F
                                                                 1987.0
                                                                               25.0
                                                                                           White
                                                                                                            Spe
          65531
                   12/6/2012
                                  22:22
                                                     Μ
                                                                 1954.0
                                                                               58.0
                                                                                           White
                                                                                                            Spe
                                                                                                  Equipment/Inspe
          65532
                   12/6/2012
                                  23:20
                                                     Μ
                                                                 1985.0
                                                                               27.0
                                                                                           Black
                   12/7/2012
                                   0:23
                                                     F
                                                                 1985.0
                                                                               15.0
                                                                                           Other
                                                                                                     Seatbelt Vio
          65533
          65534
                   12/7/2012
                                   0:30
                                                     F
                                                                 1985.0
                                                                               27.0
                                                                                           White
                                                                                                            Spe
```

stop\_duration

Loading [MathJax]/extensions/Safe.js

4060

```
In [71]:
             df.isnull().sum()
            stop_date
                                      0
 Out[71]:
             stop time
                                      0
            driver gender
                                      0
                                      0
            driver_age_raw
            driver age
                                      0
            driver race
                                      0
            violation raw
                                      0
                                      0
            vio
            search_conducted
                                      0
                                      0
             stop outcome
            is arrested
                                      0
                                      0
            stop duration
                                      0
            drugs related stop
            dtype: int64
 In [72]:
             df
 Out[72]:
                     stop_date stop_time driver_gender driver_age_raw
                                                                              driver_age driver_race
                                                                                                              violation
                       1/2/2005
                                                                      1985.0
                                                                                                 White
                  0
                                       1:55
                                                         Μ
                                                                                     20.0
                                                                                                                   Spe
                                                                                     40.0
                                                                                                 White
                  1
                      1/18/2005
                                       8:15
                                                         Μ
                                                                      1965.0
                                                                                                                   Spe
                  2
                      1/23/2005
                                     23:15
                                                         Μ
                                                                      1972.0
                                                                                     33.0
                                                                                                 White
                                                                                                                   Spe
                  3
                      2/20/2005
                                     17:15
                                                         Μ
                                                                      1986.0
                                                                                     19.0
                                                                                                 White
                                                                                                              Call for Se
                                                          F
                  4
                      3/14/2005
                                     10:00
                                                                      1984.0
                                                                                     21.0
                                                                                                 White
                                                                                                                   Spe
             65530
                      12/6/2012
                                     17:54
                                                          F
                                                                      1987.0
                                                                                     25.0
                                                                                                 White
                                                                                                                   Spe
             65531
                      12/6/2012
                                     22:22
                                                         Μ
                                                                      1954.0
                                                                                     58.0
                                                                                                 White
                                                                                                                   Spe
                                                                                                        Equipment/Inspe
                                                                                                 Black
             65532
                      12/6/2012
                                     23:20
                                                         Μ
                                                                      1985.0
                                                                                     27.0
                                                                                     15.0
                                                                                                 Other
             65533
                     12/7/2012
                                       0:23
                                                          F
                                                                      1985.0
                                                                                                           Seatbelt Vio
             65534
                     12/7/2012
                                       0:30
                                                          F
                                                                      1985.0
                                                                                     27.0
                                                                                                 White
                                                                                                                   Spe
            65535 rows × 13 columns
 In [73]:
             df[df['vio'].isin(['Speeding', 'seat belt'])]
 Out[73]:
                     stop_date
                                stop_time driver_gender driver_age_raw
                                                                              driver_age driver_race
                                                                                                       violation_raw
                  0
                       1/2/2005
                                       1:55
                                                         Μ
                                                                      1985.0
                                                                                     20.0
                                                                                                 White
                                                                                                             Speeding
                      1/18/2005
                                                                                     40.0
                                                                                                 White
                  1
                                       8:15
                                                         Μ
                                                                      1965.0
                                                                                                             Speeding
                      1/23/2005
                                     23:15
                                                                                     33.0
                                                                                                 White
                  2
                                                         Μ
                                                                      1972.0
                                                                                                             Speeding
                                                          F
                     3/14/2005
                                     10:00
                                                                      1984.0
                                                                                     21.0
                                                                                                 White
                                                                                                             Speeding
                  6
                       4/1/2005
                                     17:30
                                                         Μ
                                                                      1969.0
                                                                                     36.0
                                                                                                 White
                                                                                                             Speeding
                                                          F
             65527
                     12/6/2012
                                     15:26
                                                                      1981.0
                                                                                     31.0
                                                                                                 White
                                                                                                             Speeding
Loading [MathJax]/extensions/Safe.js
```

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw
65529	12/6/2012	16:00	M	1994.0	18.0	White	Speeding
65530	12/6/2012	17:54	F	1987.0	25.0	White	Speeding
65531	12/6/2012	22:22	M	1954.0	58.0	White	Speeding
65534	12/7/2012	0:30	F	1985.0	27.0	White	Speeding

37204 rows × 13 columns

In [74]:

 $df[\sim(df['driver_age'] == 15)]$ 

Out[74]

:		stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
	0	1/2/2005	1:55	М	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	Call for Se
	4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
	65529	12/6/2012	16:00	M	1994.0	18.0	White	Spe
	65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
	65531	12/6/2012	22:22	М	1954.0	58.0	White	Spe
	65532	12/6/2012	23:20	М	1985.0	27.0	Black	Equipment/Inspe Vio
	65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

61223 rows × 13 columns

In [75]:

0

Out[75]:		stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
	0	1/2/2005	1:55	М	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	Call for Se
	4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
	65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
	65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
	65532	12/6/2012	23:20	М	1985.0	27.0	Black	Equipment/Inspe Vio
	65533	12/7/2012	0:23	F	1985.0	15.0	Other	Seatbelt Vio
	65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe
	CEESE	12 .	. 1					

```
In [76]:
            df['driver_age']=df['driver_age'].apply(lambda x:x+3)
In [77]:
            df
                                                                                                                violation
                    stop_date stop_time driver_gender driver_age_raw driver_age driver_race
Out[77]:
                0
                     1/2/2005
                                      1:55
                                                         Μ
                                                                      1985.0
                                                                                      23.0
                                                                                                  White
                                                                                                                     Spe
                    1/18/2005
                                      8:15
                                                         Μ
                                                                      1965.0
                                                                                     43.0
                                                                                                  White
                                                                                                                     Spe
                    1/23/2005
                                     23:15
                                                         Μ
                                                                      1972.0
                                                                                      36.0
                                                                                                  White
                                                                                                                     Spe
                3
                    2/20/2005
                                     17:15
                                                         Μ
                                                                      1986.0
                                                                                      22.0
                                                                                                  White
                                                                                                                Call for Se
                                                          F
                    3/14/2005
                                     10:00
                                                                      1984.0
                                                                                      24.0
                                                                                                  White
                                                                                                                     Spe
           65530
                    12/6/2012
                                     17:54
                                                          F
                                                                      1987.0
                                                                                      28.0
                                                                                                  White
                                                                                                                     Spe
           65531
                    12/6/2012
                                     22:22
                                                         Μ
                                                                      1954.0
                                                                                      61.0
                                                                                                  White
                                                                                                                     Spe
                                                                                                         Equipment/Inspe
                    12/6/2012
                                                                                      30.0
                                                                                                  Black
           65532
                                     23:20
                                                         Μ
                                                                      1985.0
                                                          F
           65533
                    12/7/2012
                                      0:23
                                                                      1985.0
                                                                                      18.0
                                                                                                  Other
                                                                                                             Seatbelt Vio
                                                          F
           65534
                    12/7/2012
                                      0:30
                                                                      1985.0
                                                                                      30.0
                                                                                                  White
                                                                                                                     Spe
          65535 rows × 13 columns
In [78]:
            df
Out[78]:
                    stop_date stop_time driver_gender driver_age_raw driver_age driver_race
                                                                                                                violation
                     1/2/2005
                                      1:55
                                                                      1985.0
                                                                                     23.0
                                                                                                                     Spe
                0
                                                         Μ
                                                                                                  White
                    1/18/2005
                                      8:15
                                                         Μ
                                                                      1965.0
                                                                                     43.0
                                                                                                  White
                                                                                                                     Spe
                2
                    1/23/2005
                                     23:15
                                                         Μ
                                                                      1972.0
                                                                                      36.0
                                                                                                  White
                                                                                                                     Spe
                                                                                                               Call for Se
                    2/20/2005
                                     17:15
                                                         Μ
                                                                      1986.0
                                                                                      22.0
                                                                                                  White
                                                          F
                    3/14/2005
                                     10:00
                                                                      1984.0
                                                                                      24.0
                                                                                                  White
                                                                                                                     Spe
                                        ...
                                                         ...
           65530
                    12/6/2012
                                     17:54
                                                                      1987.0
                                                                                      28.0
                                                                                                  White
                                                                                                                     Spe
           65531
                    12/6/2012
                                     22:22
                                                         Μ
                                                                      1954.0
                                                                                      61.0
                                                                                                  White
                                                                                                                     Spe
                                                                                                         Equipment/Inspe
                                                                                                  Black
           65532
                    12/6/2012
                                     23:20
                                                                                      30.0
                                                         Μ
                                                                      1985.0
                                      0:23
                                                                                                  Other
                                                                                                             Seatbelt Vio
           65533
                    12/7/2012
                                                                      1985.0
                                                                                      18.0
                                      0:30
                                                          F
                                                                                      30.0
                                                                                                  White
                                                                                                                     Spe
           65534
                    12/7/2012
                                                                      1985.0
          65535 \text{ rows} \times 13 \text{ columns}
```

Loading [MathJax]/extensions/Safe.js

df.driver age = df.driver age.apply(lambda x: x-3)

In [79]:

Out[80]:		stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
	0	1/2/2005	1:55	М	1985.0	20.0	White	Spe
	1	1/18/2005	8:15	М	1965.0	40.0	White	Spe
	2	1/23/2005	23:15	М	1972.0	33.0	White	Spe
	3	2/20/2005	17:15	М	1986.0	19.0	White	Call for Se
	4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
	•••							
	65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
	65531	12/6/2012	22:22	М	1954.0	58.0	White	Spe
	65532	12/6/2012	23:20	М	1985.0	27.0	Black	Equipment/Inspe Vio
	65533	12/7/2012	0:23	F	1985.0	15.0	Other	Seatbelt Vio
	65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe
	65535 r	ows × 13 c	olumns					
In [81]:	df.gr	oupby(' <mark>dri</mark>	ver_gender'	).count()				
Out[81]:	driver_		p_date sto	p_time driver_	age_raw driver_	age driver_	race violatio	on_raw vio
	unver_	F	20371	20371	20371 20	371 20	0371	20371 20371
		г М	45164	45164			5164	45164 45164
		IVI	45104	43104	45104 45	104 4.	5104	43104 43104
In [82]:	df[df	['vio'] ==	Speeding']	.driver_gende	r.value_counts(	)		
Out[82]:	F 1	5517 1687 driver_gen	der, dtype	: int64				
In [83]:	df.gr	oupby(' <mark>dri</mark> v	ver_gender'	).search_cond	ucted.sum()			
Out[83]:	driver_gender F 366 M 2113 Name: search_conducted, dtype: int64							
In [84]:	df.sto	p_duration	n.unique()					
Out[84]:	array(	['0-15 Min	', '16-30 N	Min', '30+ Min	', 2, '2'], dty	pe=object)		
In [85]:	df['st	top_duratio	on'] = df['	stop_duration	'].map({'0-15 M	in': 7.5, '	16-30 Min':	24, '30+ Min
In [86]:	df							
Out[86]: Loading [Math	Jax]/exten	stop_date sions/Safe.js	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation
0	1/2/2005	1:55	М	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	Call for Se
4	3/14/2005	10:00	F	1984.0	21.0	White	Spe
65530	12/6/2012	17:54	F	1987.0	25.0	White	Spe
65531	12/6/2012	22:22	M	1954.0	58.0	White	Spe
65532	12/6/2012	23:20	М	1985.0	27.0	Black	Equipment/Inspe Vio
65533	12/7/2012	0:23	F	1985.0	15.0	Other	Seatbelt Vio
65534	12/7/2012	0:30	F	1985.0	27.0	White	Spe

 $65535 \text{ rows} \times 13 \text{ columns}$ 

```
In [87]:
          df.stop_duration.mean()
```

12.187420698181345 Out[87]:

In [88]: df.groupby('driver\_age').vio.describe()

Out[88]:		count	unique	top	freq
	driver_age				
	15.0	4312	6	Seat belt	4055
	16.0	34	5	Speeding	18
	17.0	449	5	Speeding	338
	18.0	1345	6	Speeding	980
	19.0	2388	5	Speeding	1655
	83.0	2	2	Speeding	1
	84.0	3	1	Speeding	3
	85.0	1	1	Moving violation	1
	86.0	6	3	Speeding	3
	88.0	2	1	Speeding	2

73 rows  $\times$  4 columns

In [90]:	<pre>df.sort_values(by= ['driver_age'])</pre>								
Out[90]:		stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	
Loading [Math	<b>23620</b> Jax]/extens	2/5/2008 sions/Safe.js	17:43	F	1985.0	15.0	Other	Seatbelt Violation	

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw
36431	7/28/2009	0:24	F	1985.0	15.0	Other	Seatbelt Violation
36426	7/27/2009	21:53	F	1985.0	15.0	Other	Seatbelt Violation
7190	6/5/2006	20:46	F	1985.0	15.0	Other	Seatbelt Violation
36413	7/27/2009	3:20	F	1985.0	15.0	Other	Seatbelt Violation
23847	2/15/2008	14:34	M	1922.0	86.0	White	Speeding
38468	11/17/2009	15:33	М	1923.0	86.0	White	Other Traffic Violation
48329	3/20/2011	18:32	M	1925.0	86.0	White	Call for Service
9829	9/6/2006	12:15	М	1918.0	88.0	White	Speeding
50377	6/20/2011	12:05	М	1923.0	88.0	White	Speeding

65535 rows  $\times$  13 columns

In [ ]: