Working on Real Project wih Python

Police Dataset

In this, I have taken dataset of Police check post

The dataset is available in csv format. We are going to analyse it using pandas

```
#importing pandas
import pandas as pd
#reading data from file
police=pd.read_csv('file.csv')
#checking data
police
       stop date stop time country name driver gender driver age raw
0
                                                                    1985.0
        1/2/2005
                       1:55
                                       NaN
       1/18/2005
                       8:15
                                       NaN
                                                                    1965.0
1
                                                         М
2
       1/23/2005
                      23:15
                                       NaN
                                                         М
                                                                    1972.0
       2/20/2005
                                       NaN
                                                         М
                                                                    1986.0
                      17:15
       3/14/2005
                                       NaN
                      10:00
                                                                    1984.0
65530
       12/6/2012
                      17:54
                                       NaN
                                                                    1987.0
                                       NaN
65531
       12/6/2012
                      22:22
                                                         М
                                                                    1954.0
                                       NaN
65532
       12/6/2012
                      23:20
                                                         М
                                                                    1985.0
65533
       12/7/2012
                       0:23
                                       NaN
                                                      NaN
                                                                        NaN
65534
       12/7/2012
                       0:30
                                       NaN
                                                         F
                                                                    1985.0
       driver_age driver_race
                                                   violation raw
violation
                         White
                                                         Speeding
              20.0
Speeding
              40.0
                         White
                                                         Speeding
Speeding
```

2	33.0	White			Speeding	
Speeding						
3	19.0	White		Call f	or Service	
Other 4	21 0	Whi+o			Coooding	
4 Speeding	21.0	White			Speeding	
						•
65530	25.0	White			Speeding	
Speeding					- p 5	
65531	58.0	White			Speeding	
Speeding						
65532	27.0	Black	Equip	ment/Inspection	Violation	
Equipment						
65533	NaN	NaN			NaN	
NaN 65534	27.0	White			Speeding	
Speeding	27.0	MILLICE			Speeding	
Specuring						
sear	ch_conducte	d search	_type	stop_outcome	is_arrested	
stop_durati				-	_	
0	Fals	e	NaN	Citation	False	0 -
15 Min				6' 1 1'	- 1	0
1 15 Min	Fals	е	NaN	Citation	False	0 -
15 Min 2	Fals	Δ	NaN	Citation	False	0 -
15 Min	1 a c s	C	IVAIV	CITACION	Tacse	0-
3	Fals	e	NaN	Arrest Driver	True	16-
30 Min		-				
4	Fals	е	NaN	Citation	False	0 -
15 Min						
		•				
	F-1-	_	NI - NI	C:+-+:	F-1	0
65530	Fals	e	NaN	Citation	False	0 -
15 Min 65531	Fals	Δ	NaN	Warning	False	0 -
15 Min	1403	C	Nan	Warning	racsc	U
65532	Fals	е	NaN	Citation	False	0 -
15 Min						
65533	Fals	e	NaN	NaN	NaN	
NaN						
65534	Fals	е	NaN	Citation	False	0 -
15 Min						
drug	s related s	ton				
0		lse				
1		lse				
2	Fa	lse				
3	Fa	lse				

```
4 False
...
65530 False
65531 False
65532 False
65533 False
65534 False
[65535 rows x 15 columns]
```

Q-> Remove the column that only contains missing values

```
#checking the count of null values in each column
police.isnull().sum()
stop date
                           0
stop time
                           0
                      65535
country_name
driver gender
                       4061
driver_age_raw
                       4054
                       4307
driver age
driver race
                       4060
violation raw
                       4060
                       4060
violation
search conducted
                      63056
search type
stop outcome
                       4060
is arrested
                       4060
stop duration
                       4060
drugs related stop
dtype: int64
#so from here, our targetted column is country name
#so we have to remove this column
police.drop(columns='country name',inplace=True)
#checking
police
       stop_date stop_time driver_gender driver_age_raw
driver_age \
       1/2/2005
                                                                  20.0
                      1:55
                                                   1985.0
                                                   1965.0
                                                                  40.0
       1/18/2005
                      8:15
       1/23/2005
                     23:15
                                                   1972.0
                                                                  33.0
                                                                  19.0
3
       2/20/2005
                     17:15
                                                   1986.0
       3/14/2005
                     10:00
                                                   1984.0
                                                                  21.0
```

65530	12/6/2012	17:	54	F	:	1987.0	25.0
65531	12/6/2012	22:	22	М	:	1954.0	58.0
65532	12/6/2012	23:	20	М		1985.0	27.0
65533	12/7/2012	0:	23	NaN		NaN	NaN
65534	12/7/2012	0:	30	F	:	1985.0	27.0
0 1 2 3 4	driver_race White White White White White			Spee Spee l for Ser	ding ding ding	Speeding Speeding Speeding Other Speeding	\
65530 65531 65532 65533 65534	White White Black NaN White	Equip	ment/Inspect	Spee ion Viola	eding eding eding NaN NaN	Speeding Speeding Equipment NaN Speeding	
	search cond	ducted	search type	·		is arrested	
stop_d 0	uration \	False	NaN	· -	ation	– False	Θ-
15 Min 1		False	NaN	Cit	ation	False	Θ-
15 Min 2		False	NaN	Cit	ation	False	Θ-
15 Min 3		False	NaN	Arrest D		True	16-
30 Min 4		False	NaN		ation	False	
15 Min							
65530		False	NaN	Ci+	ation	False	0 -
15 Min							
65531 15 Min		False	NaN		rning	False	0 -
65532 15 Min		False	NaN	Cit	ation	False	0 -
65533 NaN		False	NaN		NaN	NaN	
65534		False	NaN	Cit	ation	False	0 -

```
15 Min
       drugs_related_stop
0
                     False
1
                     False
                     False
3
                     False
                     False
                     False
65530
65531
                     False
65532
                     False
65533
                     False
65534
                     False
[65535 rows x 14 columns]
```

Q-> For speeding, men or women who stopped more often?

police					
driver	stop_date age \	stop_time	driver_gender	driver_age_raw	
0	1/2/2005	1:55	М	1985.0	20.0
1	1/18/2005	8:15	М	1965.0	40.0
2	1/23/2005	23:15	М	1972.0	33.0
3	2/20/2005	17:15	М	1986.0	19.0
4	3/14/2005	10:00	F	1984.0	21.0
65530	12/6/2012	17:54	F	1987.0	25.0
65531	12/6/2012	22:22	М	1954.0	58.0
65532	12/6/2012	23:20	М	1985.0	27.0
65533	12/7/2012	0:23	NaN	NaN	NaN
65534	12/7/2012	0:30	F	1985.0	27.0
0 1	driver_race White White	9		tion_raw violation Speeding Speeding Speeding Speeding	-

2 3 4	White White White		Call	Speeding for Service Speeding	Speeding Other Speeding	
65530 65531 65532 65533 65534	White White Black NaN White	Equipment	/Inspecti	Speeding Speeding on Violation NaN Speeding	Speeding Speeding Equipment NaN Speeding	
sear stop durati	_	ucted sear	ch_type	stop_outcome		
o 15 Min		False	NaN	Citation	False	0 -
1 1 15 Min		False	NaN	Citation	False	0 -
2 15 Min		False	NaN	Citation	False	0 -
3 30 Min		False	NaN	Arrest Driver	True	16-
4 15 Min		False	NaN	Citation	False	0 -
65530 15 Min		False	NaN	Citation	False	0 -
65531 15 Min		False	NaN	Warning	False	Θ-
65532 15 Min		False	NaN	Citation	False	0 -
65533 NaN		False	NaN	NaN	NaN	
65534 15 Min		False	NaN	Citation	False	0 -
drug 0 1 2 3 4 65530 65531 65532 65533 65534	s_relat x 14 c	False				

<pre>police[police['violation']=='Speeding']</pre>	#this	gives	all	columns
containing speeding				

contair	ning speed.	ing			
driver		stop_time dri	ver_gender	driver_age_raw	
0	1/2/2005	1:55	М	1985.0	20.0
1	1/18/2005	8:15	М	1965.0	40.0
2	1/23/2005	23:15	М	1972.0	33.0
4	3/14/2005	10:00	F	1984.0	21.0
6	4/1/2005	17:30	М	1969.0	36.0
65527	12/6/2012	15:26	F	1981.0	31.0
65529	12/6/2012	16:00	М	1994.0	18.0
65530	12/6/2012	17:54	F	1987.0	25.0
65531	12/6/2012	22:22	М	1954.0	58.0
65534	12/7/2012	0:30	F	1985.0	27.0
search	-	e violation_ra	w violation	search_conducted	
3Cai Cii	Lype \				
0	_type \ White	e Speedin	g Speeding	False	
0 NaN 1		•		False False	
0 NaN 1 NaN 2	White	e Speedin	g Speeding		
0 NaN 1 NaN 2 NaN 4	White	e Speedin e Speedin	g Speeding g Speeding	False	
0 NaN 1 NaN 2 NaN 4 NaN 6	White White White	e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding	False False	
0 NaN 1 NaN 2 NaN 4 NaN	White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False	
0 NaN 1 NaN 2 NaN 4 NaN 6 NaN	White White White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False False	
0 NaN 1 NaN 2 NaN 4 NaN 6 NaN 65527 NaN	White White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False False	
0 NaN 1 NaN 2 NaN 4 NaN 6 NaN 65527 NaN 65529 NaN	White White White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False False False False False False	
0 NaN 1 NaN 2 NaN 4 NaN 6 NaN 65527 NaN 65529 NaN 65530 NaN	White White White White White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False False False False False False False	
0 NaN 1 NaN 2 NaN 4 NaN 6 NaN 6 NaN 65527 NaN 65529 NaN 65530	White White White White White White	e Speedin e Speedin e Speedin e Speedin	g Speeding g Speeding g Speeding g Speeding	False False False False False False False False	

NaN				
0 1	Citation		0-15 Min	drugs_related_stop False False
2 4 6	Citation	False False False	0-15 Min	False False False
65527 65529 65530	Citation		0-15 Min	False False False
65531 65534	Warning		0-15 Min	False False
[37204	4 rows x 14 co	lumns]		
police	e[police['viol	ation']=='Sp	peeding'].drive	r_gender.value_counts()
F :	25517 11686 driver_gender	, dtype: int	t64	
#Thus	male is the r	required answ	ver	

Q-> Does gender affect who gets searched during stop?

police					
driver	stop_date age \	stop_time	driver_gender	driver_age_raw	
0	1/2/2005	1:55	М	1985.0	20.0
1	1/18/2005	8:15	M	1965.0	40.0
2	1/23/2005	23:15	М	1972.0	33.0
3	2/20/2005	17:15	М	1986.0	19.0
4	3/14/2005	10:00	F	1984.0	21.0
65530	12/6/2012	17:54	F	1987.0	25.0
65531	12/6/2012	22:22	М	1954.0	58.0
65532	12/6/2012	23:20	М	1985.0	27.0
65533	12/7/2012	0:23	NaN	NaN	NaN

65534	12/7/2012	0:30		F :	1985.0	27.0
0 1 2 3 4 65530 65531 65532 65533 65534	driver_race White White White White White White White Black NaN White	Equipment	Call 1	Speeding Speeding Speeding Speeding For Service Speeding Speeding Speeding Violation NaN Speeding	violation Speeding Speeding Other Speeding Speeding Speeding Equipment NaN Speeding	
stop_d	search_conduration \			stop_outcome	_	
0 15 Min		False	NaN	Citation	False	0 -
1		False	NaN	Citation	False	0 -
15 Min 2 15 Min		False	NaN	Citation	False	0 -
3		False	NaN A	rrest Driver	True	16-
30 Min 4 15 Min		False	NaN	Citation	False	0 -
65530 15 Min		False	NaN	Citation	False	0 -
65531		False	NaN	Warning	False	0 -
15 Min 65532 15 Min		False	NaN	Citation	False	0 -
65533		False	NaN	NaN	NaN	
NaN 65534 15 Min		False	NaN	Citation	False	0 -
0 1 2 3 4 65530 65531	drugs_relat	red_stop False False False False False False				

```
65532
                     False
65533
                     False
65534
                    False
[65535 rows x 14 columns]
police.groupby('driver_gender')
<pandas.core.groupby.generic.DataFrameGroupBy object at</pre>
0x0000017B2CF8B5E0>
police.groupby('driver_gender').search_conducted.sum()
driver_gender
      366
Μ
     2113
Name: search_conducted, dtype: int64
```

Q-> What is the mean_stop duration?

police					
driver	stop_date _age \	stop_time	driver_gender	driver_age_raw	
0	1/2/2005	1:55	М	1985.0	20.0
1	1/18/2005	8:15	М	1965.0	40.0
2	1/23/2005	23:15	М	1972.0	33.0
3	2/20/2005	17:15	М	1986.0	19.0
4	3/14/2005	10:00	F	1984.0	21.0
65530	12/6/2012	17:54	F	1987.0	25.0
65531	12/6/2012	22:22	М	1954.0	58.0
65532	12/6/2012	23:20	М	1985.0	27.0
65533	12/7/2012	0:23	NaN	NaN	NaN
65534	12/7/2012	0:30	F	1985.0	27.0
0 1	driver_race White White			tion_raw violatio Speeding Speedin Speeding Speedin	g g
2	White	9		Speeding Speedin	g

3 4	White White		Cal	l for Service Speeding	Other Speeding	
65530 65531 65532 65533 65534	White White White Black NaN White	Equipment/I	nspect	Speeding Speeding Speeding ion Violation NaN Speeding	Speeding Speeding Equipment NaN Speeding	
stop dur		ucted search	_type	stop_outcome	is_arrested	
0 15 Min	uc1011 (False	NaN	Citation	False	0 -
1 15 Min		False	NaN	Citation	False	0 -
2 15 Min		False	NaN	Citation	False	0 -
3 30 Min		False	NaN	Arrest Driver	True	16-
4 15 Min		False	NaN	Citation	False	0 -
65530 15 Min		False	NaN	Citation	False	0 -
65531 15 Min		False	NaN	Warning	False	0 -
65532 15 Min		False	NaN	Citation	False	0 -
65533 NaN		False	NaN	NaN	NaN	
65534 15 Min		False	NaN	Citation	False	0 -
0 1 2 3 4 65530 65531 65532 65533 65534	rows x 14 c	False	counts	()		

```
0-15 Min
             47379
16-30 Min
             11448
30+ Min
              2647
Name: stop duration, dtype: int64
#but problem is that dtype of column is objet(str) and to get mean we
have to type cast into int
police['stop duration']=police['stop duration'].map({'0-15}
Min': 7.5, 1\overline{6}-30 Min': 24, 30+ Min': 4\overline{5})
police
       stop_date stop_time driver_gender driver_age_raw
driver age \
                                                                   20.0
       1/2/2005
                       1:55
                                                     1985.0
                                                                   40.0
       1/18/2005
                       8:15
                                                     1965.0
2
       1/23/2005
                      23:15
                                                     1972.0
                                                                   33.0
       2/20/2005
                      17:15
                                                     1986.0
                                                                   19.0
                                                                   21.0
       3/14/2005
                      10:00
                                                     1984.0
                      17:54
                                                                   25.0
65530 12/6/2012
                                                    1987.0
                                                                   58.0
65531 12/6/2012
                      22:22
                                                     1954.0
65532 12/6/2012
                      23:20
                                                     1985.0
                                                                   27.0
65533 12/7/2012
                       0:23
                                       NaN
                                                        NaN
                                                                    NaN
                                                                   27.0
65534 12/7/2012
                       0:30
                                                     1985.0
      driver race
                                      violation raw
                                                     violation \
0
            White
                                           Speeding
                                                      Speeding
1
            White
                                           Speeding
                                                      Speeding
2
                                           Speeding
                                                      Speeding
            White
3
                                  Call for Service
            White
                                                          0ther
4
            White
                                           Speeding
                                                      Speeding
65530
            White
                                           Speeding
                                                       Speeding
            White
65531
                                           Speeding
                                                      Speeding
                    Equipment/Inspection Violation
65532
            Black
                                                      Equipment
65533
              NaN
                                                NaN
                                                            NaN
65534
            White
                                           Speeding
                                                      Speeding
```

ston d	<pre>search_conducted uration \</pre>	search_type	stop_outcome	is_arrested
0	False	NaN	Citation	False
7.5 1	False	NaN	Citation	False
7.5 2	Ealco	NaN	Citation	Falso
7.5	False	NaN	Citation	False
3 24.0	False	NaN	Arrest Driver	True
4	False	NaN	Citation	False
7.5				
65530 7.5	False	NaN	Citation	False
65531	False	NaN	Warning	False
7.5 65532	False	NaN	Citation	False
7.5				
65533 NaN	False	NaN	NaN	NaN
65534	False	NaN	Citation	False
7.5				
	drugs_related_sto			
0	False			
1	False			
2	False False			
1 2 3 4	False			
65530	False			
65531	False	e		
65532	False	e		
65533	False	-		
65534	Fals	e		
[65535	rows x 14 columns]		
	e can find the mean ['stop_duration'].	· -		
12.187	420698181345			

Q-> Compare the age distribution for each violation.

police

stop_date stop_time driver_age \ 0 1/2/2005 1:55 1 1/18/2005 8:15	ver_gender driv M	ver_age_raw 1985.0	
0 1/2/2005 1:55	М	1985 Ռ	
1 1/18/2005 8:15		1303.0	20.0
	М	1965.0	40.0
2 1/23/2005 23:15	М	1972.0	33.0
3 2/20/2005 17:15	М	1986.0	19.0
4 3/14/2005 10:00	F	1984.0	21.0
65530 12/6/2012 17:54	F	1987.0	25.0
65531 12/6/2012 22:22	М	1954.0	58.0
65532 12/6/2012 23:20	М	1985.0	27.0
65533 12/7/2012 0:23	NaN	NaN	NaN
65534 12/7/2012 0:30	F	1985.0	27.0
0.00		-505.0	
driver_race White White White White White White White	violation_ Speed Speed Speed Call for Serv Speed	Jing Speeding Jing Speeding Jing Speeding Jice Other	\
65530 White 65531 White 65532 Black Equipment/In 65533 NaN 65534 White	Speed Speed Ispection Violat Speed	ding Speeding tion Equipment NaN NaN	
search_conducted search_	_type stop_out	come is_arrested	t
<pre>stop_duration \ 0</pre>	NaN Cita	ation False	2
7.5 1 False		ation False	
7.5			
False		ation False 	
False 24.0	NaN Arrest Dr	river True	į
4 False 7.5	NaN Cita	ation False	;

65530 7.5	False	NaN	Citation	False
65531	False	NaN	Warning	False
7.5	F.1	NI - NI	Citatia	E-1
65532 7.5	False	NaN	Citation	False
65533	False	NaN	NaN	NaN
NaN 65534	False	NaN	Citation	False
7.5	racse	itait	CICACION	racse
0 1 2 3 4 65530 65531 65532 65533 65534	drugs_related_stop False			
[65535	rows x 14 columns]			
police	.groupby('violation')		
	s.core.groupby.gener: 017B2E156680>	ic.DataFrameG	roupBy object a	at

police.groupby('violation').driver_age.describe()

	count	mean	std	min	25%	50%
75% \						
violation						
Equipment	6507.0	31.682957	11.380671	16.0	23.0	28.0
39.0						
Moving violation	11876.0	36.736443	13.258350	15.0	25.0	35.0
47.0						
0ther	3477.0	40.362381	12.754423	16.0	30.0	41.0
50.0						
Registration/plates	2240.0	32.656696	11.150780	16.0	24.0	30.0
40.0						
Seat belt	3.0	30.333333	10.214369	23.0	24.5	26.0
34.0						
Speeding	37120.0	33.262581	12.615781	15.0	23.0	30.0
42.0						

	max
violation	
Equipment	81.0
Moving violation	86.0
Other other	86.0
Registration/plates	74.0
Seat belt	42.0
Speeding	88.0