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       1/2/2005      1:55           NaN             M          1985.0

1      1/18/2005      8:15           NaN             M          1965.0

2      1/23/2005     23:15           NaN             M          1972.0

3      2/20/2005     17:15           NaN             M          1986.0

4      3/14/2005     10:00           NaN             F          1984.0

...          ...       ...           ...           ...             ...

65530  12/6/2012     17:54           NaN             F          1987.0

65531  12/6/2012     22:22           NaN             M          1954.0

65532  12/6/2012     23:20           NaN             M          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  \
0            20.0       White                       Speeding
Speeding
1            40.0       White                       Speeding
Speeding
```

```
2            33.0       White                          Speeding
Speeding
3            19.0       White                 Call for Service
Other
4            21.0       White                          Speeding
Speeding
...           ...         ...                               ...         .
..
65530        25.0       White                          Speeding
Speeding
65531        58.0       White                          Speeding
Speeding
65532        27.0       Black  Equipment/Inspection Violation
Equipment
65533         NaN         NaN                               NaN
NaN
65534        27.0       White                          Speeding
Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False      0-
15 Min
1                 False         NaN       Citation       False      0-
15 Min
2                 False         NaN       Citation       False      0-
15 Min
3                 False         NaN  Arrest Driver        True     16-
30 Min
4                 False         NaN       Citation       False      0-
15 Min
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False      0-
15 Min
65531             False         NaN        Warning       False      0-
15 Min
65532             False         NaN       Citation       False      0-
15 Min
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False      0-
15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
```

```
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
country_name          65535
driver_gender          4061
driver_age_raw         4054
driver_age             4307
driver_race            4060
violation_raw          4060
violation              4060
search_conducted          0
search_type           63056
stop_outcome           4060
is_arrested            4060
stop_duration          4060
drugs_related_stop        0
dtype: int64
```

```
#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  \
0       1/2/2005      1:55             M          1985.0        20.0

1      1/18/2005      8:15             M          1965.0        40.0

2      1/23/2005     23:15             M          1972.0        33.0

3      2/20/2005     17:15             M          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             M          1954.0         58.0

65532  12/6/2012     23:20             M          1985.0         27.0

65533  12/7/2012      0:23           NaN             NaN          NaN

65534  12/7/2012      0:30             F          1985.0         27.0


      driver_race                   violation_raw  violation  \
0           White                        Speeding   Speeding
1           White                        Speeding   Speeding
2           White                        Speeding   Speeding
3           White               Call for Service      Other
4           White                        Speeding   Speeding
...           ...                             ...        ...
65530       White                        Speeding   Speeding
65531       White                        Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                             NaN        NaN
65534       White                        Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False      0-
15 Min
1                 False         NaN       Citation       False      0-
15 Min
2                 False         NaN       Citation       False      0-
15 Min
3                 False         NaN  Arrest Driver        True     16-
30 Min
4                 False         NaN       Citation       False      0-
15 Min
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False      0-
15 Min
65531             False         NaN        Warning       False      0-
15 Min
65532             False         NaN       Citation       False      0-
15 Min
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False      0-
```

```
15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

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

```
police

       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0         20.0

1       1/18/2005      8:15             M          1965.0         40.0

2       1/23/2005     23:15             M          1972.0         33.0

3       2/20/2005     17:15             M          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             M          1954.0         58.0

65532   12/6/2012     23:20             M          1985.0         27.0

65533   12/7/2012      0:23           NaN             NaN          NaN

65534   12/7/2012      0:30             F          1985.0         27.0


      driver_race                   violation_raw  violation  \
0           White                        Speeding   Speeding
1           White                        Speeding   Speeding
```

```
2           White                          Speeding   Speeding
3           White                 Call for Service      Other
4           White                          Speeding   Speeding
...           ...                                ...        ...
65530       White                          Speeding   Speeding
65531       White                          Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                                NaN        NaN
65534       White                          Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False      0-
15 Min
1                 False         NaN       Citation       False      0-
15 Min
2                 False         NaN       Citation       False      0-
15 Min
3                 False         NaN  Arrest Driver        True     16-
30 Min
4                 False         NaN       Citation       False      0-
15 Min
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False      0-
15 Min
65531             False         NaN        Warning       False      0-
15 Min
65532             False         NaN       Citation       False      0-
15 Min
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False      0-
15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

```
police[police['violation']=='Speeding']  #this gives all columns
containing speeding
```

```
       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0        20.0

1      1/18/2005      8:15             M          1965.0        40.0

2      1/23/2005     23:15             M          1972.0        33.0

4      3/14/2005     10:00             F          1984.0        21.0

6       4/1/2005     17:30             M          1969.0        36.0

...           ...       ...           ...             ...         ...

65527  12/6/2012     15:26             F          1981.0        31.0

65529  12/6/2012     16:00             M          1994.0        18.0

65530  12/6/2012     17:54             F          1987.0        25.0

65531  12/6/2012     22:22             M          1954.0        58.0

65534  12/7/2012      0:30             F          1985.0        27.0


      driver_race violation_raw violation  search_conducted
search_type  \
0           White      Speeding  Speeding            False
NaN
1           White      Speeding  Speeding            False
NaN
2           White      Speeding  Speeding            False
NaN
4           White      Speeding  Speeding            False
NaN
6           White      Speeding  Speeding            False
NaN
...           ...           ...       ...              ...         ..
.
65527       White      Speeding  Speeding            False
NaN
65529       White      Speeding  Speeding            False
NaN
65530       White      Speeding  Speeding            False
NaN
65531       White      Speeding  Speeding            False
NaN
65534       White      Speeding  Speeding            False
```

```
NaN

      stop_outcome is_arrested stop_duration  drugs_related_stop
0         Citation       False      0-15 Min               False
1         Citation       False      0-15 Min               False
2         Citation       False      0-15 Min               False
4         Citation       False      0-15 Min               False
6         Citation       False      0-15 Min               False
...            ...         ...           ...                 ...
65527     Citation       False      0-15 Min               False
65529     Citation       False      0-15 Min               False
65530     Citation       False      0-15 Min               False
65531      Warning       False      0-15 Min               False
65534     Citation       False      0-15 Min               False

[37204 rows x 14 columns]
```

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

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

```
#Thus male is the required answer
```

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

```
police
```

```
       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0        20.0

1      1/18/2005      8:15             M          1965.0        40.0

2      1/23/2005     23:15             M          1972.0        33.0

3      2/20/2005     17:15             M          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             M          1954.0        58.0

65532  12/6/2012     23:20             M          1985.0        27.0

65533  12/7/2012      0:23           NaN             NaN         NaN
```

```
65534  12/7/2012      0:30              F          1985.0        27.0


      driver_race                    violation_raw  violation  \
0           White                          Speeding   Speeding
1           White                          Speeding   Speeding
2           White                          Speeding   Speeding
3           White                 Call for Service      Other
4           White                          Speeding   Speeding
...           ...                               ...        ...
65530       White                          Speeding   Speeding
65531       White                          Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                               NaN        NaN
65534       White                          Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False      0-
15 Min
1                 False         NaN       Citation       False      0-
15 Min
2                 False         NaN       Citation       False      0-
15 Min
3                 False         NaN  Arrest Driver        True     16-
30 Min
4                 False         NaN       Citation       False      0-
15 Min
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False      0-
15 Min
65531             False         NaN        Warning       False      0-
15 Min
65532             False         NaN       Citation       False      0-
15 Min
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False      0-
15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
```

```
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

```
police.groupby('driver_gender')
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x0000017B2CF8B5E0>
```

```
police.groupby('driver_gender').search_conducted.sum()
```

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

Q-> What is the mean_stop duration?

```
police
```

```
       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0         20.0

1      1/18/2005      8:15             M          1965.0         40.0

2      1/23/2005     23:15             M          1972.0         33.0

3      2/20/2005     17:15             M          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             M          1954.0         58.0

65532  12/6/2012     23:20             M          1985.0         27.0

65533  12/7/2012      0:23           NaN             NaN          NaN

65534  12/7/2012      0:30             F          1985.0         27.0


      driver_race                    violation_raw  violation  \
0           White                         Speeding   Speeding
1           White                         Speeding   Speeding
2           White                         Speeding   Speeding
```

```
3           White                Call for Service      Other
4           White                        Speeding   Speeding
...           ...                             ...        ...
65530       White                        Speeding   Speeding
65531       White                        Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                             NaN        NaN
65534       White                        Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False      0-
15 Min
1                 False         NaN       Citation       False      0-
15 Min
2                 False         NaN       Citation       False      0-
15 Min
3                 False         NaN  Arrest Driver        True     16-
30 Min
4                 False         NaN       Citation       False      0-
15 Min
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False      0-
15 Min
65531             False         NaN        Warning       False      0-
15 Min
65532             False         NaN       Citation       False      0-
15 Min
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False      0-
15 Min

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

```
police['stop_duration'].value_counts()
```

```
0-15 Min     47379
16-30 Min    11448
30+ Min       2647
2                1
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 ,'16-30 Min':24,'30+ Min':45})

police

       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0         20.0

1       1/18/2005      8:15             M          1965.0         40.0

2       1/23/2005     23:15             M          1972.0         33.0

3       2/20/2005     17:15             M          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             M          1954.0         58.0

65532   12/6/2012     23:20             M          1985.0         27.0

65533   12/7/2012      0:23           NaN             NaN          NaN

65534   12/7/2012      0:30             F          1985.0         27.0


      driver_race                   violation_raw  violation  \
0           White                        Speeding   Speeding
1           White                        Speeding   Speeding
2           White                        Speeding   Speeding
3           White               Call for Service      Other
4           White                        Speeding   Speeding
...           ...                             ...        ...
65530       White                        Speeding   Speeding
65531       White                        Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                             NaN        NaN
65534       White                        Speeding   Speeding
```

```
       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False
7.5
1                 False         NaN       Citation       False
7.5
2                 False         NaN       Citation       False
7.5
3                 False         NaN  Arrest Driver        True
24.0
4                 False         NaN       Citation       False
7.5
...                 ...         ...            ...         ...
...
65530             False         NaN       Citation       False
7.5
65531             False         NaN        Warning       False
7.5
65532             False         NaN       Citation       False
7.5
65533             False         NaN            NaN         NaN
NaN
65534             False         NaN       Citation       False
7.5

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

```
#now we can find the mean
police['stop_duration'].mean()
```

```
12.187420698181345
```

Q-> Compare the age distribution for each violation.

```
police
```

```
       stop_date stop_time driver_gender  driver_age_raw
driver_age  \
0       1/2/2005      1:55             M          1985.0        20.0

1       1/18/2005      8:15             M          1965.0        40.0

2       1/23/2005     23:15             M          1972.0        33.0

3       2/20/2005     17:15             M          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             M          1954.0        58.0

65532   12/6/2012     23:20             M          1985.0        27.0

65533   12/7/2012      0:23           NaN             NaN         NaN

65534   12/7/2012      0:30             F          1985.0        27.0


      driver_race                    violation_raw  violation  \
0           White                         Speeding   Speeding
1           White                         Speeding   Speeding
2           White                         Speeding   Speeding
3           White                Call for Service      Other
4           White                         Speeding   Speeding
...           ...                              ...        ...
65530       White                         Speeding   Speeding
65531       White                         Speeding   Speeding
65532       Black  Equipment/Inspection Violation  Equipment
65533         NaN                              NaN        NaN
65534       White                         Speeding   Speeding

       search_conducted search_type   stop_outcome is_arrested
stop_duration  \
0                 False         NaN       Citation       False
7.5
1                 False         NaN       Citation       False
7.5
2                 False         NaN       Citation       False
7.5
3                 False         NaN  Arrest Driver        True
24.0
4                 False         NaN       Citation       False
7.5
...                 ...         ...            ...         ...
```

```
...
65530              False         NaN       Citation       False
7.5
65531              False         NaN        Warning       False
7.5
65532              False         NaN       Citation       False
7.5
65533              False         NaN            NaN         NaN
NaN
65534              False         NaN       Citation       False
7.5

       drugs_related_stop
0                   False
1                   False
2                   False
3                   False
4                   False
...                   ...
65530               False
65531               False
65532               False
65533               False
65534               False

[65535 rows x 14 columns]
```

```
police.groupby('violation')
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at
0x0000017B2E156680>
```

```
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
Other                 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                  86.0
Registration/plates    74.0
Seat belt              42.0
Speeding               88.0
```

Thanks !!