police-vehicle-data-analysis

October 30, 2024

1. IMPORTING LIBRARIES

[74]:
```
import pandas as pd
```

2. IMPORTING THE DATASET.

[75]:
```
police_data = r"H:\DA. Python\7 . Police Vehicle Checking Analysis\Police␣
↪Dataset.csv"
data = pd.read_csv(police_data)
```

3. CHECKING THE DATASET

[76]:
```
data.shape
```

[76]:
```
(65535, 15)
```

DATA CLEANING

1. Remove the column that only contains missing values

[77]:
```
data.isnull().sum()
```

[77]:
```
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
```

[78]:
```
data.drop(columns = ['country_name'] , inplace = True)
```

[79]:
```
data.head(2)
```

[79]:
```
   stop_date stop_time driver_gender  driver_age_raw  driver_age driver_race  \
0   1/2/2005      1:55             M          1985.0        20.0       White
1  1/18/2005      8:15             M          1965.0        40.0       White

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

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

[80]:
```
data.isnull()
```

[80]:
```
       stop_date  stop_time  driver_gender  driver_age_raw  driver_age  \
0          False      False          False           False       False
1          False      False          False           False       False
2          False      False          False           False       False
3          False      False          False           False       False
4          False      False          False           False       False
…              …          …              …               …           …
65530      False      False          False           False       False
65531      False      False          False           False       False
65532      False      False          False           False       False
65533      False      False           True            True        True
65534      False      False          False           False       False

       driver_race  violation_raw  violation  search_conducted  search_type  \
0            False          False      False             False         True
1            False          False      False             False         True
2            False          False      False             False         True
3            False          False      False             False         True
4            False          False      False             False         True
…                …              …          …                 …           …
65530        False          False      False             False         True
65531        False          False      False             False         True
65532        False          False      False             False         True
65533         True           True       True             False         True
65534        False          False      False             False         True

       stop_outcome  is_arrested  stop_duration  drugs_related_stop
0             False        False          False               False
1             False        False          False               False
2             False        False          False               False
3             False        False          False               False
```

```
4             False        False          False                False
…             …          …            …                  …
65530         False        False          False                False
65531         False        False          False                False
65532         False        False          False                False
65533          True         True           True                False
65534         False        False          False                False

[65535 rows x 14 columns]
```

[81]:
```
data.isnull().sum()
```

[81]:
```
stop_date                0
stop_time                0
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
```

FILTERING + VALUE COUNTS

2. For speeding, were Men or Women stopped more often ?

[82]:
```
data.head()
```

[82]:
```
   stop_date stop_time driver_gender  driver_age_raw  driver_age driver_race  \
0   1/2/2005      1:55             M          1985.0        20.0       White
1  1/18/2005      8:15             M          1965.0        40.0       White
2  1/23/2005     23:15             M          1972.0        33.0       White
3  2/20/2005     17:15             M          1986.0        19.0       White
4  3/14/2005     10:00             F          1984.0        21.0       White

      violation_raw violation  search_conducted search_type   stop_outcome  \
0          Speeding  Speeding             False         NaN       Citation
1          Speeding  Speeding             False         NaN       Citation
2          Speeding  Speeding             False         NaN       Citation
3  Call for Service     Other             False         NaN  Arrest Driver
4          Speeding  Speeding             False         NaN       Citation
```

```
  is_arrested stop_duration  drugs_related_stop
0       False      0-15 Min               False
1       False      0-15 Min               False
2       False      0-15 Min               False
3        True     16-30 Min               False
4       False      0-15 Min               False
```

[83]:
```
data[data.violation == 'Speeding'].driver_gender.value_counts()
```

[83]:
```
driver_gender
M    25517
F    11686
Name: count, dtype: int64
```

GROUPBY

3. Does gender affect who gets searched during a stop?

[84]:
```
data.head(2)
```

[84]:
```
    stop_date stop_time driver_gender  driver_age_raw  driver_age driver_race  \
0   1/2/2005      1:55             M          1985.0        20.0       White
1  1/18/2005      8:15             M          1965.0        40.0       White

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

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

[85]:
```
data.groupby('driver_gender').search_conducted.sum()
```

[85]:
```
driver_gender
F     366
M    2113
Name: search_conducted, dtype: int64
```

MAPPING + TYPE CASTING

4. What is the mean stop_duration ?

[86]:
```
data.head(2)
```

[86]:
```
    stop_date stop_time driver_gender  driver_age_raw  driver_age driver_race  \
0   1/2/2005      1:55             M          1985.0        20.0       White
1  1/18/2005      8:15             M          1965.0        40.0       White
```

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

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

[87]:
```
#Finding out the unique values

data.stop_duration.value_counts()
```

[87]:
```
stop_duration
0-15 Min     47379
16-30 Min    11448
30+ Min       2647
2                1
Name: count, dtype: int64
```

[88]:
```
data['stop_duration'] = data['stop_duration'].map( {'0-15 Min' :7, '16-30 Min' :
 ↪24, '30+ Min':45})
```

[89]:
```
data
```

[89]:
```
       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            7.0
1                 False         NaN       Citation       False            7.0
2                 False         NaN       Citation       False            7.0
3                 False         NaN  Arrest Driver        True           24.0
4                 False         NaN       Citation       False            7.0
…                 …           …             …          …              …
65530             False         NaN       Citation       False            7.0
65531             False         NaN        Warning       False            7.0
65532             False         NaN       Citation       False            7.0
65533             False         NaN            NaN         NaN            NaN
65534             False         NaN       Citation       False            7.0

       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]
```

[90]:
```
data['stop_duration'].mean()
```

[90]:
```
np.float64(11.802062660637016)
```

GROUPBY–DESCRIBE

5. Compare the age distributions for each violations

[92]:
```
data.head(5)
```

[92]:
```
   stop_date stop_time driver_gender  driver_age_raw  driver_age driver_race  \
0   1/2/2005      1:55             M          1985.0        20.0       White
1  1/18/2005      8:15             M          1965.0        40.0       White
2  1/23/2005     23:15             M          1972.0        33.0       White
3  2/20/2005     17:15             M          1986.0        19.0       White
4  3/14/2005     10:00             F          1984.0        21.0       White
```

```
      violation_raw violation  search_conducted search_type   stop_outcome  \
0          Speeding  Speeding             False         NaN       Citation
1          Speeding  Speeding             False         NaN       Citation
2          Speeding  Speeding             False         NaN       Citation
3  Call for Service     Other             False         NaN  Arrest Driver
4          Speeding  Speeding             False         NaN       Citation

  is_arrested  stop_duration  drugs_related_stop
0       False           7.0               False
1       False           7.0               False
2       False           7.0               False
3        True          24.0               False
4       False           7.0               False
```

[99]:
```
data.groupby('violation').driver_age.describe()
```

[99]:
```
                      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
```