```
1 import pandas as pd
```

```
1 police = pd.read_csv("/content/PoliceDataset.csv")
```

```
1 police.head(4)
```

| | stop_date | stop_time | country_name | driver_gender | driver_age_raw | driver_age | dr |
|---|---|---|---|---|---|---|---|
| **0** | 1/2/2005 | 1:55 | NaN | M | 1985.0 | 20.0 | |
| **1** | 1/18/2005 | 8:15 | NaN | M | 1965.0 | 40.0 | |
| **2** | 1/23/2005 | 23:15 | NaN | M | 1972.0 | 33.0 | |
| **3** | 2/20/2005 | 17:15 | NaN | M | 1986.0 | 19.0 | |

```
1 police.tail(4)
```

| | stop_date | stop_time | country_name | driver_gender | driver_age_raw | driver_age |
|---|---|---|---|---|---|---|
| **65531** | 12/6/2012 | 22:22 | NaN | M | 1954.0 | 58.0 |
| **65532** | 12/6/2012 | 23:20 | NaN | M | 1985.0 | 27.0 |
| **65533** | 12/7/2012 | 0:23 | NaN | NaN | NaN | NaN |
| **65534** | 12/7/2012 | 0:30 | NaN | F | 1985.0 | 27.0 |

```
1 police.columns
```

```
Index(['stop_date', 'stop_time', 'country_name', 'driver_gender',
       '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')
```

```
1 police.shape
```

```
(65535, 15)
```

```
1 police.isnull()
```

| | stop_date | stop_time | country_name | driver_gender | driver_age_raw | driver_age |
|---|---|---|---|---|---|---|
| **0** | False | False | True | False | False | False |
| **1** | False | False | True | False | False | False |
| **2** | False | False | True | False | False | False |
| **3** | False | False | True | False | False | False |
| **4** | False | False | True | False | False | False |
| **...** | ... | ... | ... | ... | ... | ... |
| **65530** | False | False | True | False | False | False |
| **65531** | False | False | True | False | False | False |
| **65532** | False | False | True | False | False | False |
| **65533** | False | False | True | True | True | True |
| **65534** | False | False | True | False | False | False |

65535 rows × 15 columns

```
1 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
```

```
1 police.drop(columns="country_name",inplace=True)
2 # we have dropped the country_name column permanently because it contains al
```

```
1 police.head(4)
```

| | stop_date | stop_time | driver_gender | driver_age_raw | driver_age | driver_race | vio |
|---|---|---|---|---|---|---|---|
| **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 | |

```
police[police.violation == "Speeding"]
# this returns only the violation person records
```

| stop_time | driver_gender | driver_age_raw | driver_age | driver_race | violation_raw | vi |
|---|---|---|---|---|---|---|
| 1:55 | M | 1985.0 | 20.0 | White | Speeding | S |
| 8:15 | M | 1965.0 | 40.0 | White | Speeding | S |
| 23:15 | M | 1972.0 | 33.0 | White | Speeding | S |
| 10:00 | F | 1984.0 | 21.0 | White | Speeding | S |
| 17:30 | M | 1969.0 | 36.0 | White | Speeding | S |
| ... | ... | ... | ... | ... | ... | |
| 15:26 | F | 1981.0 | 31.0 | White | Speeding | S |
| 16:00 | M | 1994.0 | 18.0 | White | Speeding | S |
| 17:54 | F | 1987.0 | 25.0 | White | Speeding | S |
| 22:22 | M | 1954.0 | 58.0 | White | Speeding | S |
| 0:30 | F | 1985.0 | 27.0 | White | Speeding | S |

ımns

```
police[police.violation == "Speeding"].driver_gender.value_counts()
# this returns the count of each gender
```

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

### GroupBy

- syntax : dataset.groupby("col1").col2

```
police.groupby("driver_gender").search_conducted.sum()
# this returns the total number of records by grouping the driver_gender and
# search conducted is true.
```

```
driver_gender
```

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

```
police.search_conducted.value_counts()
# value_counts() returns the count of each value in the column.
```

```
False    63056
True      2479
Name: search_conducted, dtype: int64
```

**Mapping** - mapping the old values to new values in a column

- syntax : dataset['column'].map({'a':b,'c':d})

```
police['stop_duration']=police['stop_duration'].map({'0-15 Min':7,'16-30 Min
# mapping the old values to new values in a column
```

```
police.head(4)
```

| | stop_date | stop_time | driver_gender | driver_age_raw | driver_age | driver_race | vio |
|---|---|---|---|---|---|---|---|
| **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 | Cal |

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

```
7.0     47379
24.0    11448
45.0     2647
Name: stop_duration, dtype: int64
```

**Mean** - used to calculate mean of the entire column in the dataset

- Syntax : dataset['column'].mean()

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

```
11.802062660637016
```

**Describe** - describe mean,sd,min, max,25%,50%,75% of the column in the dataset

- df['col'].describe()

```
1 police.describe()
```

| | driver_age_raw | driver_age | stop_duration |
|---|---|---|---|
| **count** | 61481.000000 | 61228.000000 | 61474.000000 |
| **mean** | 1967.791106 | 34.148984 | 11.802063 |
| **std** | 121.050106 | 12.760710 | 9.640422 |
| **min** | 0.000000 | 15.000000 | 7.000000 |
| **25%** | 1965.000000 | 23.000000 | 7.000000 |
| **50%** | 1978.000000 | 31.000000 | 7.000000 |
| **75%** | 1985.000000 | 43.000000 | 7.000000 |
| **max** | 8801.000000 | 88.000000 | 45.000000 |

```
1 police['stop_duration'].describe()
```

```
count    61474.000000
mean        11.802063
std          9.640422
min          7.000000
25%          7.000000
50%          7.000000
75%          7.000000
max         45.000000
Name: stop_duration, dtype: float64
```

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

| violation | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Equipment** | 6507.0 | 31.682957 | 11.380671 | 16.0 | 23.0 | 28.0 | 39.0 | 81.0 |
| **Moving violation** | 11876.0 | 36.736443 | 13.258350 | 15.0 | 25.0 | 35.0 | 47.0 | 86.0 |
| **Other** | 3477.0 | 40.362381 | 12.754423 | 16.0 | 30.0 | 41.0 | 50.0 | 86.0 |
| **Registration/plates** | 2240.0 | 32.656696 | 11.150780 | 16.0 | 24.0 | 30.0 | 40.0 | 74.0 |
| **Seat belt** | 3.0 | 30.333333 | 10.214369 | 23.0 | 24.5 | 26.0 | 34.0 | 42.0 |
| **Speeding** | 37120.0 | 33.262581 | 12.615781 | 15.0 | 23.0 | 30.0 | 42.0 | 88.0 |

✓ 0s completed at 9:11 PM