Project-3 Police data analysis

importing pandas library for analysing the data

```
import pandas as pd
```

loading the data

```
df=pd.read_csv("Police Data.csv")
```

to view top 5 rows

```
df.head()
```

```
   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

   driver_age driver_race     violation_raw violation
search_conducted  \
0        20.0       White          Speeding  Speeding
False
1        40.0       White          Speeding  Speeding
False
2        33.0       White          Speeding  Speeding
False
3        19.0       White  Call for Service     Other
False
4        21.0       White          Speeding  Speeding
False

  search_type   stop_outcome is_arrested stop_duration
drugs_related_stop
0         NaN       Citation       False      0-15 Min
False
1         NaN       Citation       False      0-15 Min
False
2         NaN       Citation       False      0-15 Min
False
3         NaN  Arrest Driver        True     16-30 Min
False
4         NaN       Citation       False      0-15 Min
False
```

shape of the data set

```
df.shape
```

```
(65535, 15)
```

find null values

```
df.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
```

we are going to drop "country_name","search_type"

becauseof there is a enter columns are null values present

```
df.drop(["search_type","country_name"],axis=1,inplace=True)

df.head()
```

```
   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   stop_outcome
is_arrested  \
```

```
0          Speeding  Speeding             False       Citation
False
1          Speeding  Speeding             False       Citation
False
2          Speeding  Speeding             False       Citation
False
3  Call for Service     Other             False  Arrest Driver
True
4          Speeding  Speeding             False       Citation
False

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

Demographic Insights

What is the distribution of driver genders across all stops?

What age group has the highest number of stops?

How does the stop frequency vary across different driver races?

What is the distribution of driver genders across all stops?

```
df["driver_gender"].value_counts(dropna=False).reset_index()

  driver_gender  count
0             M  45164
1             F  16310
2           NaN   4061
```

What age group has the highest number of stops?

```
df.head(2)

   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 stop_outcome
is_arrested  \
0      Speeding  Speeding             False     Citation       False

1      Speeding  Speeding             False     Citation       False
```

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

```
df["driver_age"].value_counts().sort_values(ascending=False).reset_ind
ex().head(1)
```

```
   driver_age  count
0        22.0   2912
```

the driver who having a age 22 get more stops

How does the stop frequency vary across different driver races?

```
df["driver_race"].value_counts().reset_index().head(1)
```

```
  driver_race  count
0       White  45747
```

This means White drivers had the highest number of police stops in this dataset.

Violation-Based Analysis

What are the top 5 most common violations?

Which violations are more likely to result in an arrest?

Is there a relationship between violation type and whether a search was conducted?

What are the top 5 most common violations?

```
df.head(2)
```

```
   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 stop_outcome
is_arrested  \
0      Speeding  Speeding             False     Citation       False

1      Speeding  Speeding             False     Citation       False


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

```
df["violation"].value_counts().sort_values(ascending=False).reset_inde
x().head()
```

```
              violation  count
0             Speeding  37204
1     Moving violation  11926
2            Equipment   6516
3                Other   3583
4  Registration/plates   2243
```

Is there a relationship between violation type and whether a search was conducted?

```
violation_search_relation = pd.crosstab(df["violation"],
df["search_conducted"])

# Rename the columns for better understanding
violation_search_relation.columns = ["No Search", "Search Conducted"]

violation_search_relation.reset_index()
```

```
              violation  No Search  Search Conducted
0            Equipment       5977               539
1     Moving violation      11236               690
2                Other       3418               165
3  Registration/plates       1954               289
4            Seat belt          3                 0
5             Speeding      36408               796
```

Time-Based Analysis

On which days of the week do most stops occur?

At what time of day are most stops conducted? (Morning, Afternoon, Evening, Night)

How does the number of stops change month over month?

On which days of the week do most stops occur?

```
df.head(2)
```

```
   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 stop_outcome
is_arrested  \
0      Speeding  Speeding             False     Citation       False
```

```
1      Speeding  Speeding              False     Citation       False


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

```
df["stop_date"]=pd.to_datetime(df["stop_date"])
```

```
df["week_day"]=df["stop_date"].dt.day_name()
```

```
df.head()
```

```
   stop_date stop_time driver_gender  driver_age_raw  driver_age
driver_race  \
0 2005-01-02      1:55             M          1985.0        20.0
White
1 2005-01-18      8:15             M          1965.0        40.0
White
2 2005-01-23     23:15             M          1972.0        33.0
White
3 2005-02-20     17:15             M          1986.0        19.0
White
4 2005-03-14     10:00             F          1984.0        21.0
White

      violation_raw violation  search_conducted   stop_outcome
is_arrested  \
0          Speeding  Speeding             False       Citation
False
1          Speeding  Speeding             False       Citation
False
2          Speeding  Speeding             False       Citation
False
3  Call for Service     Other             False  Arrest Driver
True
4          Speeding  Speeding             False       Citation
False

  stop_duration  drugs_related_stop week_day
0      0-15 Min               False   Sunday
1      0-15 Min               False  Tuesday
2      0-15 Min               False   Sunday
3     16-30 Min               False   Sunday
4      0-15 Min               False   Monday
```

```
df["week_day"].value_counts().sort_values(ascending=False).reset_index
()
```

```
    week_day  count
0     Monday   9637
1    Tuesday   9567
2   Saturday   9504
3  Wednesday   9376
4     Friday   9335
5     Sunday   9075
6   Thursday   9041
```

How does the number of stops change month over month?

```
df["month"]=df["stop_date"].dt.month_name()

df.head()
```

```
   stop_date stop_time driver_gender  driver_age_raw  driver_age
driver_race  \
0 2005-01-02      1:55             M          1985.0        20.0
White
1 2005-01-18      8:15             M          1965.0        40.0
White
2 2005-01-23     23:15             M          1972.0        33.0
White
3 2005-02-20     17:15             M          1986.0        19.0
White
4 2005-03-14     10:00             F          1984.0        21.0
White

      violation_raw violation  search_conducted   stop_outcome
is_arrested  \
0          Speeding  Speeding             False       Citation
False
1          Speeding  Speeding             False       Citation
False
2          Speeding  Speeding             False       Citation
False
3  Call for Service     Other             False  Arrest Driver
True
4          Speeding  Speeding             False       Citation
False

  stop_duration  drugs_related_stop week_day     month
0      0-15 Min               False   Sunday   January
1      0-15 Min               False  Tuesday   January
2      0-15 Min               False   Sunday   January
3     16-30 Min               False   Sunday  February
4      0-15 Min               False   Monday     March
```

```
df["month"].value_counts().reset_index()
```

```
        month  count
0     January   6129
1    November   5951
2         May   5661
3     October   5584
4       March   5554
5        June   5523
6       April   5424
7        July   5344
8      August   5261
9    February   5252
10  September   4990
11   December   4862
```

Drugs and Duration

What proportion of stops are drug-related?

What is the average stop duration for drug-related vs non-drug-related stops?

Which violation types are most commonly associated with drug-related stops?

```
df.head(2)
```

```
   stop_date stop_time driver_gender  driver_age_raw  driver_age
driver_race  \
0 2005-01-02      1:55             M          1985.0        20.0
White
1 2005-01-18      8:15             M          1965.0        40.0
White

  violation_raw violation  search_conducted stop_outcome
is_arrested  \
0      Speeding  Speeding             False     Citation       False

1      Speeding  Speeding             False     Citation       False


  stop_duration  drugs_related_stop week_day    month
0      0-15 Min               False   Sunday  January
1      0-15 Min               False  Tuesday  January
```

What is the average stop duration for drug-related vs non-drug-related stops?

```
df["stop_duration"]=df["stop_duration"].replace({"0-15 Min":10})
df["stop_duration"]=df["stop_duration"].replace({'16-30 Min':23})
```

```
df["stop_duration"]=df["stop_duration"].replace({"30+ Min":35})
df["stop_duration"]=df["stop_duration"].replace({"2":2})

df.head(2)
```

```
   stop_date stop_time driver_gender  driver_age_raw  driver_age
driver_race  \
0 2005-01-02      1:55             M          1985.0        20.0
White
1 2005-01-18      8:15             M          1965.0        40.0
White

  violation_raw violation  search_conducted stop_outcome
is_arrested  \
0      Speeding  Speeding             False     Citation       False

1      Speeding  Speeding             False     Citation       False


   stop_duration  drugs_related_stop week_day    month
0           10.0               False   Sunday  January
1           10.0               False  Tuesday  January
```

```
df.groupby("drugs_related_stop")["stop_duration"].mean().reset_index()
```

```
   drugs_related_stop  stop_duration
0               False      13.407648
1                True      24.036680
```

Which violation types are most commonly associated with drug-related stops?

```
violation=df[df["drugs_related_stop"]==True]
violation["violation"].value_counts().reset_index()
```

```
              violation  count
0              Speeding    179
1     Moving violation    156
2             Equipment    132
3  Registration/plates     31
4                 Other     20
```