Police Dataset Analysis

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
```

```
from google.colab import files


uploaded = files.upload()
```

Choose Files 3. Police Data.csv

- **3. Police Data.csv**(text/csv) - 6215282 bytes, last modified: 6/5/2022 - 100% done

```
Saving 3. Police Data.csv to 3. Police Data (3).csv
```

```
df=pd.read_csv("3. Police Data.csv")
```

```
df
```

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	d
0	1/2/2005	1:55	NaN	M	1985.0	20.0	

```
df.isnull()
```

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	d
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

```
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
```

```
df['country_name']
```

```
0       NaN
1       NaN
2       NaN
3       NaN
4       NaN
         ..
65530   NaN
65531   NaN
65532   NaN
65533   NaN
65534   NaN
Name: country_name, Length: 65535, dtype: float64
```

```
df
```

	stop_date	stop_time	country_name	driver_gender	driver_age_raw	driver_age	d
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	
4	3/14/2005	10:00	NaN	F	1984.0	21.0	
...	...	...	...	...	...	...	
65530	12/6/2012	17:54	NaN	F	1987.0	25.0	
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	

65535 rows × 15 columns

```
df.drop(columns="country_name",inplace=True)
```

```
df.head()
```

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

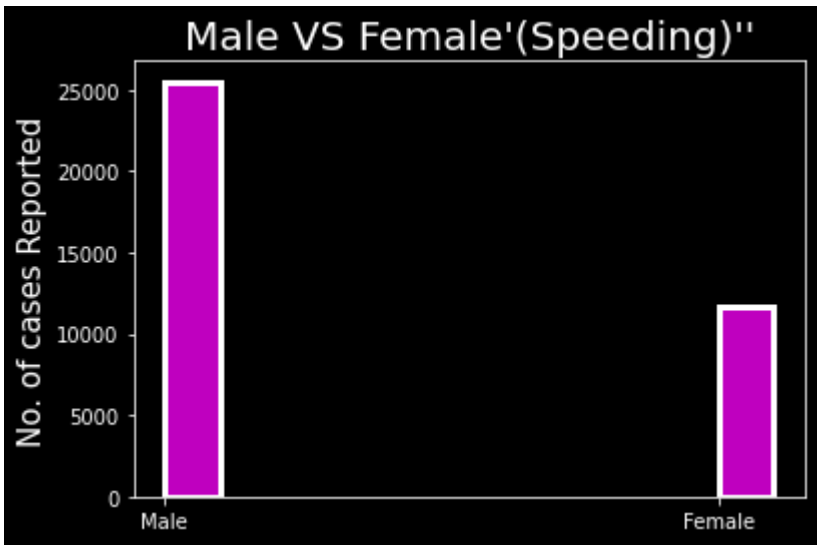
Q. Who Caught more often for Speeding?

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

```
speeding
```

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

```
from matplotlib import style
%matplotlib inline
gender=["Male","Female"]
plt.style.use("dark_background")
plt.title("Male VS Female'(Speeding)''",fontsize=20)
plt.bar(gender,speeding,width=0.1,color="m",align="edge",edgecolor="w",linewidth=3)
plt.ylabel("No. of cases Reported",fontsize=15,)
plt.show()
```



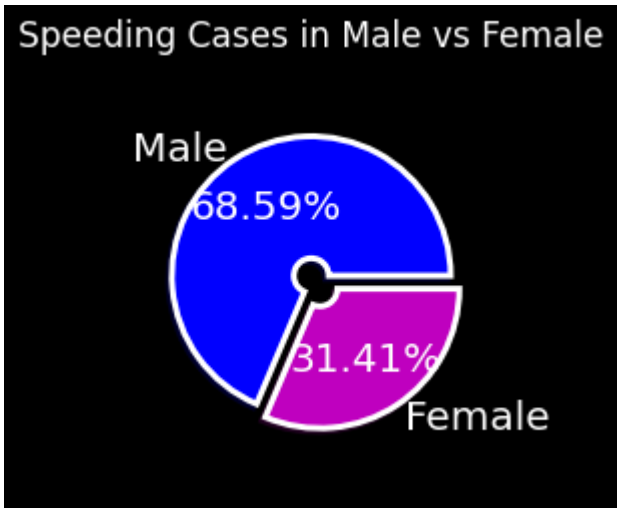
Insights

Male were caught more often then Female.

```
l=[]
for i in speeding:
    l.append(i)
male=l[0]/(l[0]+l[1])*100
print("Male : ",round(male,2))
female=l[1]/(l[0]+l[1])*100
print("Female : ",round(female,2))
```

```
Male :  68.59
Female :  31.41
```

```
explode=[0,0.09]
color=["b","m"]
plt.title('Speeding Cases in Male vs Female',fontsize=17)
wedgeprops={"linewidth":3,"width":0.7,"edgecolor":"w"}
textprops={"fontsize":20}
pie=plt.pie(speeding, labels = gender,explode=explode,colors=color,autopct="%0.2f%%",
            shadow=True,textprops=textprops,wedgeprops=wedgeprops,counterclock=True,radius=0.
plt.show()
```




```
df.violation.unique()
```

```
array(['Speeding', 'Other', 'Equipment', 'Moving violation', nan,
       'Registration/plates', 'Seat belt'], dtype=object)
```

Q. Which Violation cases reported more ?

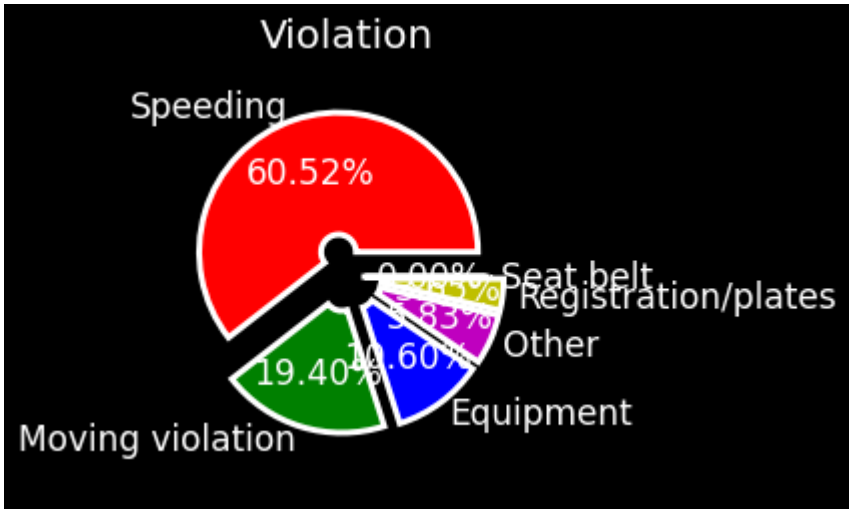
```
violation=df["violation"].value_counts()
violation
```

```
Speeding               37204
Moving violation       11926
Equipment               6516
Other                   3583
Registration/plates     2243
Seat belt                  3
Name: violation, dtype: int64
```

```
violn=['Speeding','Moving violation',"Equipment","Other","Registration/plates","Seat belt"]
```

```
explode=[0.15,0.1,0.1,0.1,0.1,0]
wedgeprops={"linewidth":3,"width":0.7,"edgecolor":"w"}

textprops={'fontsize':17}
color=['r','g','b','m','y','c']
plt.pie(violation,labels=violn,colors=color,autopct="%0.2f%%",textprops=textprops,explode=exp
        radius=0.8,wedgeprops=wedgeprops)
plt.title('Violation',fontsize=20)
plt.show()
```



Insights

Speeding Vioation cases are reported more

```
df[df.violation=='Speeding'].driver_age.mean()
```

```
33.26258081896552
```

Average Age who Do Speed Violation : 33

```
df.head()
```

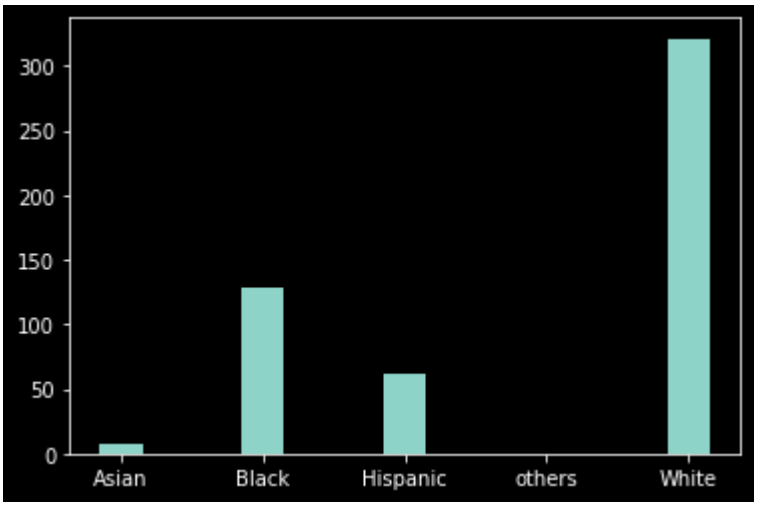
	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violat
0	1/2/2005	1:55	M	1985.0	20.0	White	S
1	1/18/2005	8:15	M	1965.0	40.0	White	S
2	1/23/2005	23:15	M	1972.0	33.0	White	S
3	2/20/2005	17:15	M	1986.0	19.0	White	Call for
4	3/14/2005	10:00	F	1984.0	21.0	White	S

Q. Which race stopeed more for Drug Checking?

```
drug_race=df.groupby("driver_race").drugs_related_stop.sum()
drug_race
```

```
driver_race
Asian          7
Black        128
Hispanic      62
Other          0
White        321
Name: drugs_related_stop, dtype: int64
```

```
race=["Asian","Black","Hispanic","others","White"]
plt.bar(race,drug_race,width=0.3)
plt.show()
```



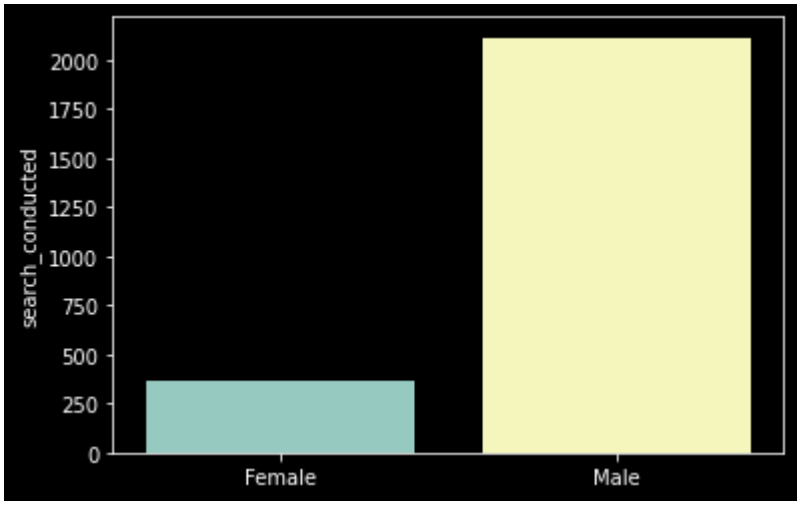
White are stooped more then any other

```
search=df.groupby("driver_gender").search_conducted.sum()
gen=["Female","Male"]
```

```
print(search)
```

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

```
sns.barplot(gen,search,data=df,)
warnings.filterwarnings('ignore')
plt.show()
```



Male are stooped more then female for searching

```
df
```

	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

```
df['stop_duration'].unique()
```

```
array(['0-15 Min', '16-30 Min', '30+ Min', nan, '2'], dtype=object)
```

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

```
from numpy import nan
df['stop_duration'].replace({'0-15 Min':7.5,'16-30 Min':23,'30+ Min':40,'nan':0})
```

```
0        7.5
1        7.5
2        7.5
3         23
4        7.5
        ...
65530    7.5
65531    7.5
65532    7.5
65533    NaN
65534    7.5
Name: stop_duration, Length: 65535, dtype: object
```

✓ 0s completed at 8:38 PM