Mark Middleton

Date: 11/27/2023

In [88]:

```
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

```
In [102]: # Reads and imports the csv file as data frame, displays and analyzes informati

          df = pd.read_csv('police.csv')

          print(df)
          df.describe(include='all')
```

```
      state   stop_date stop_time  county_name driver_gender driver_race  \
0        RI    1/4/2005     12:55          NaN             M       White   
1        RI   1/23/2005     23:15          NaN             M       White   
2        RI   2/17/2005      4:15          NaN             M       White   
3        RI   2/20/2005     17:15          NaN             M       White   
4        RI   2/24/2005      1:20          NaN             F       White   
...     ...         ...       ...          ...           ...         ...   
91736    RI  12/31/2015     21:21          NaN             F       Black   
91737    RI  12/31/2015     21:59          NaN             F       White   
91738    RI  12/31/2015     22:04          NaN             M       White   
91739    RI  12/31/2015     22:09          NaN             F    Hispanic   
91740    RI  12/31/2015     22:47          NaN             M       White   

                        violation_raw            violation  search_conducted  \
0      Equipment/Inspection Violation            Equipment             False   
1                            Speeding             Speeding             False   
2                            Speeding             Speeding             False   
3                    Call for Service                Other             False   
4                            Speeding             Speeding             False   
...                               ...                  ...               ...   
91736         Other Traffic Violation     Moving violation             False   
91737                        Speeding             Speeding             False   
91738         Other Traffic Violation     Moving violation             False   
91739  Equipment/Inspection Violation            Equipment             False   
91740          Registration Violation  Registration/plates             False   

      search_type   stop_outcome is_arrested stop_duration  \
0             NaN       Citation       False      0-15 Min   
1             NaN       Citation       False      0-15 Min   
2             NaN       Citation       False      0-15 Min   
3             NaN  Arrest Driver        True     16-30 Min   
4             NaN       Citation       False      0-15 Min   
...           ...            ...         ...           ...   
91736         NaN       Citation       False      0-15 Min   
91737         NaN       Citation       False      0-15 Min   
91738         NaN       Citation       False      0-15 Min   
91739         NaN        Warning       False      0-15 Min   
91740         NaN       Citation       False      0-15 Min   

       drugs_related_stop district  
0                   False  Zone X4  
1                   False  Zone K3  
2                   False  Zone X4  
3                   False  Zone X1  
4                   False  Zone X3  
...                   ...      ...  
91736               False  Zone K2  
91737               False  Zone K3  
91738               False  Zone X3  
91739               False  Zone K3  
91740               False  Zone X4  

[91741 rows x 15 columns]
```

Out[102]:

	state	stop_date	stop_time	county_name	driver_gender	driver_race	violation_raw	vic
count	91741	91741	91741	0.0	86536	86539	86539	
unique	1	3757	1431	NaN	2	5	12	
top	RI	5/16/2007	11:00	NaN	M	White	Speeding	Sp
freq	91741	63	358	NaN	62762	61872	48424	
mean	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
std	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
min	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
25%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
50%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
75%	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
max	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

In [109]:

```
df_imputed = df.fillna(0)
df_imputed.describe(include='all')
```

Out[109]:

	state	stop_date	stop_time	county_name	driver_gender	driver_race	violation_raw	vic
count	91741	91741	91741	91741.0	91741	91741	91741	
unique	1	3757	1431	NaN	3	6	13	
top	RI	5/16/2007	11:00	NaN	M	White	Speeding	Sp
freq	91741	63	358	NaN	62762	61872	48424	
mean	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
std	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
min	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
25%	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
50%	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
75%	NaN	NaN	NaN	0.0	NaN	NaN	NaN	
max	NaN	NaN	NaN	0.0	NaN	NaN	NaN	

In [122]:

```
# Creates a new data frame organized by sex, then groups them individually

sex = df_imputed.groupby("driver_gender")
male = sex.get_group("M")
female = sex.get_group("F")
```

In [123]:
```
sex.describe(include='all')
```

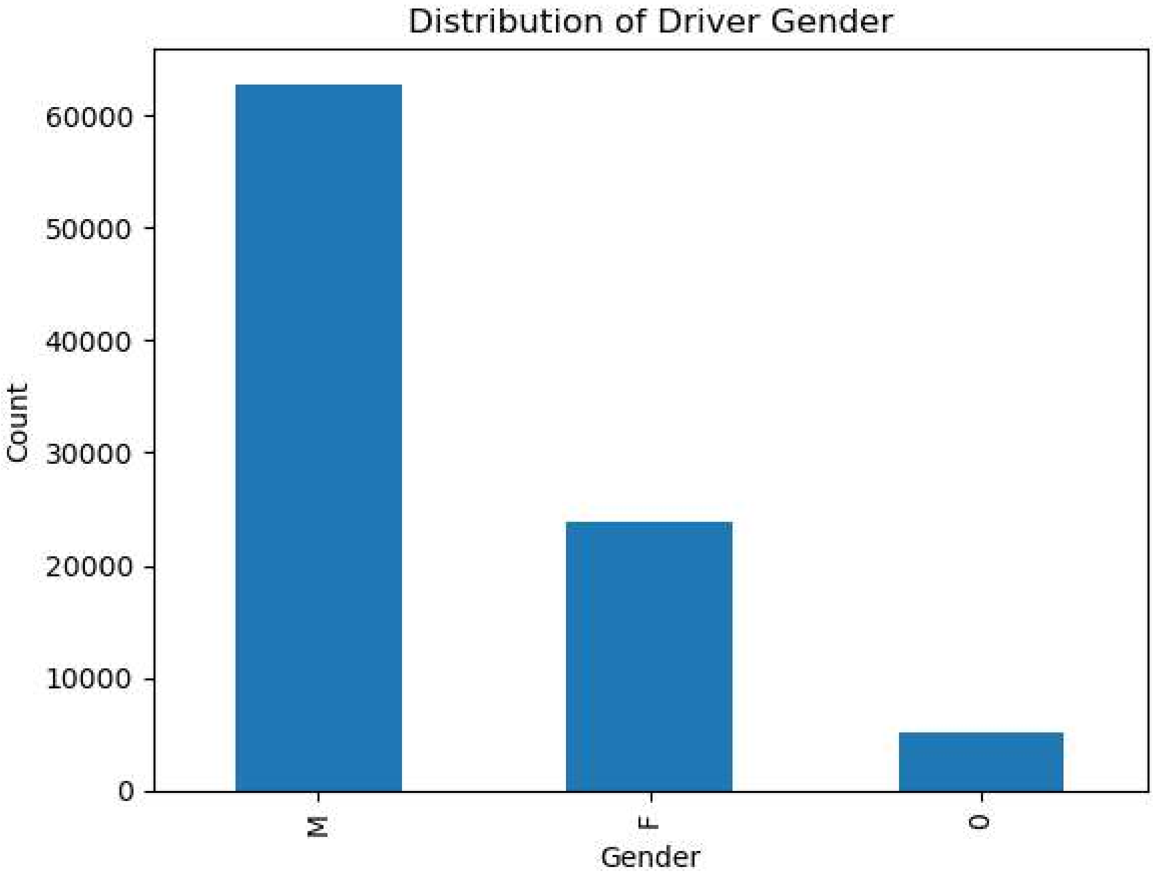
Out[123]:

	state										...	district	
	count	unique	top	freq	mean	std	min	25%	50%	75%	...	unique	top
driver_gender													
0	5205	1	RI	5205	NaN	NaN	NaN	NaN	NaN	NaN	...	6	Zone X4
F	23774	1	RI	23774	NaN	NaN	NaN	NaN	NaN	NaN	...	6	Zone X4
M	62762	1	RI	62762	NaN	NaN	NaN	NaN	NaN	NaN	...	6	Zone X4

3 rows × 154 columns

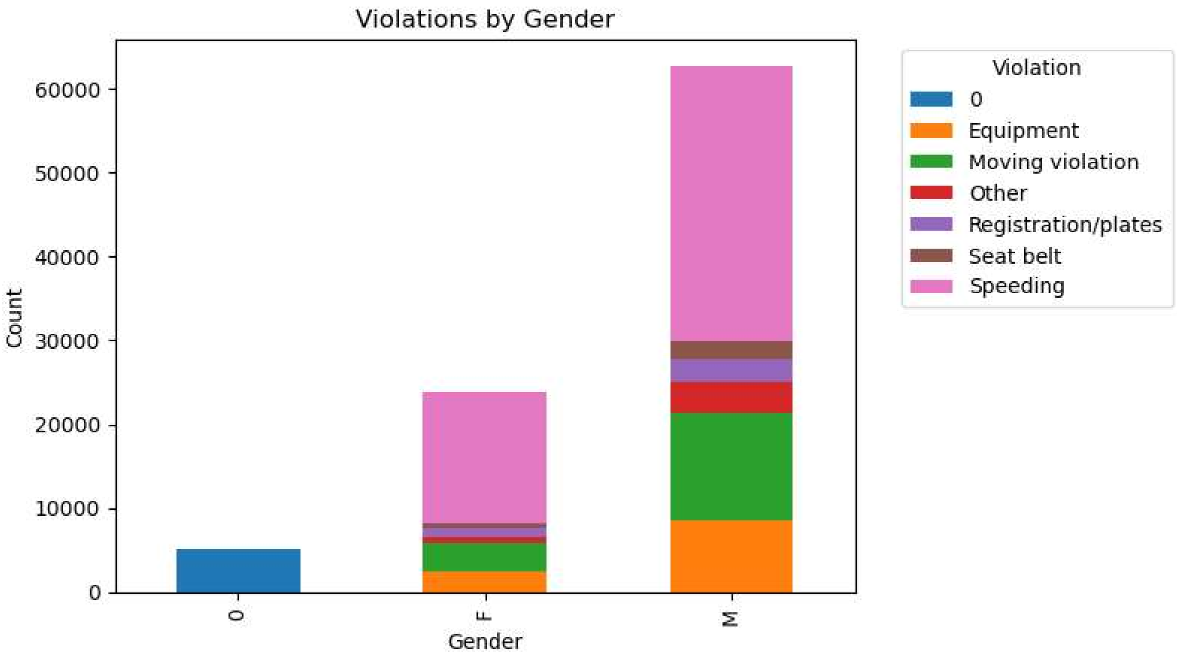
In [124]:
```
df_imputed['driver_gender'].value_counts().plot(kind='bar')
plt.title('Distribution of Driver Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
```

Out[124]:
```
Text(0, 0.5, 'Count')
```



In [128]:
```
grouped_data = df_imputed.groupby(['driver_gender', 'violation']).size().unstac
grouped_data.plot(kind='bar', stacked=True)
plt.title('Violations by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.legend(title='Violation', bbox_to_anchor=(1.05, 1), loc='upper left')

plt.show()
```



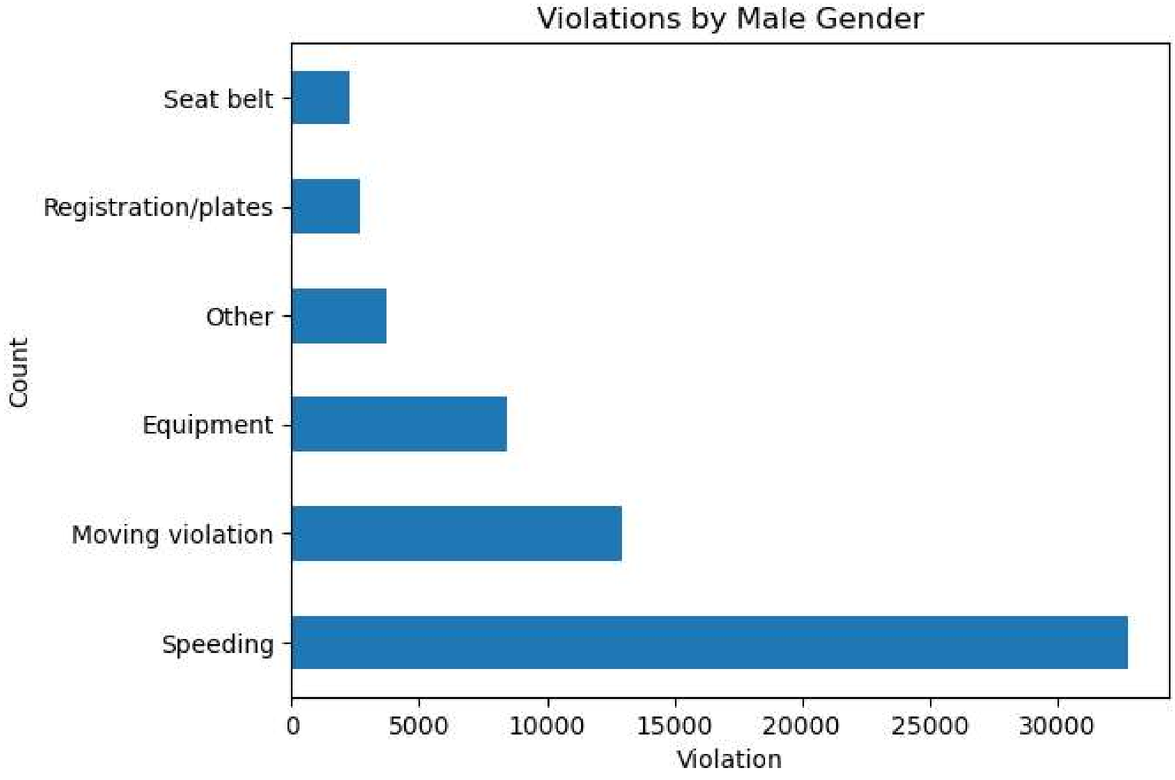
In [129]:
```
grouped_data.describe()
```

Out[129]:

violation	0	Equipment	Moving violation	Other	Registration/plates	Seat belt	S
count	1.0	3.000000	2.000000	3.000000	2.000000	2.000000	3
mean	5202.0	3640.666667	8112.000000	1470.000000	1851.500000	1428.000000	16141
std	NaN	4323.658212	6824.994652	1964.936895	1125.006889	1202.081528	16393
min	5202.0	1.000000	3286.000000	1.000000	1056.000000	578.000000	1
25%	5202.0	1251.000000	5699.000000	354.000000	1453.750000	1003.000000	7823
50%	5202.0	2501.000000	8112.000000	707.000000	1851.500000	1428.000000	15646
75%	5202.0	5460.500000	10525.000000	2204.500000	2249.250000	1853.000000	24211
max	5202.0	8420.000000	12938.000000	3702.000000	2647.000000	2278.000000	32777

In [144]:

```
male_data = df_imputed[df_imputed['driver_gender'] == 'M']
male_violations = male_data['violation'].value_counts()
male_violations.plot(kind='barh')
plt.title('Violations by Male Gender')
plt.xlabel('Violation')
plt.ylabel('Count')
plt.show()
```



In [148]:

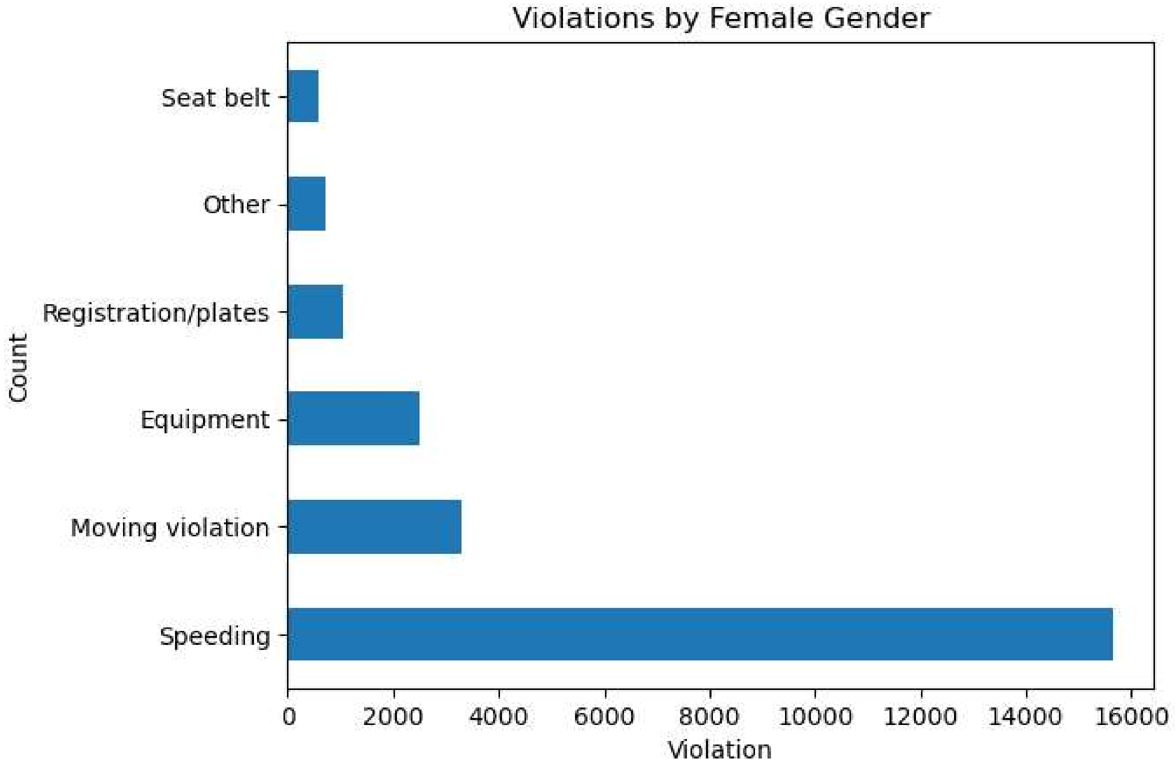
```
male_violation_counts = male_data['violation'].value_counts()
print('Traffic Offenses Committed by Males:\n')
print(male_violation_counts)
```

```
Traffic Offenses Committed by Males:

Speeding               32777
Moving violation       12938
Equipment               8420
Other                   3702
Registration/plates     2647
Seat belt               2278
Name: violation, dtype: int64
```

In [140]:

```
female_data = df_imputed[df_imputed['driver_gender'] == 'F']
female_violations = female_data['violation'].value_counts()
female_violations.plot(kind='barh')
plt.title('Violations by Female Gender')
plt.xlabel('Violation')
plt.ylabel('Count')
plt.show()
```



In [147]:

```
female_violation_counts = female_data['violation'].value_counts()
print('Traffic Offenses Committed by Females:\n')
print(female_violation_counts)
```

```
Traffic Offenses Committed by Females:

Speeding               15646
Moving violation        3286
Equipment               2501
Registration/plates     1056
Other                    707
Seat belt                578
Name: violation, dtype: int64
```

In [ ]: