In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
cd=pd.read_excel('Crime data.xlsx')
cd
```

Out[1]:

	Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Type	...	Victim Ethnicity	Perpetrator Sex	Perpetrato Ag
0	1	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	NaN	Murder or Manslaughter	...	Unknown	Male	2
1	2	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	NaN	Murder or Manslaughter	...	Unknown	Male	1
2	3	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	NaN	Manslaughter by Negligence	...	Unknown	Male	1
3	4	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	February	NaN	Manslaughter by Negligence	...	Unknown	Female	3
4	5	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	March	NaN	Murder or Manslaughter	...	Unknown	Unknown	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
221686	221687	WY01501	Cody	Municipal Police	Park	Wyoming	2012	February	NaN	Murder or Manslaughter	...	Unknown	Male	2
221687	221688	WY01701	Sheridan	Municipal Police	Sheridan	Wyoming	2012	September	NaN	Murder or Manslaughter	...	Unknown	Female	5
221688	221689	WY01900	Sweetwater County	Sheriff	Sweetwater	Wyoming	2012	August	NaN	Murder or Manslaughter	...	Unknown	Male	1
221689	221690	WY01902	Rock Springs	Municipal Police	Sweetwater	Wyoming	2012	February	NaN	Murder or Manslaughter	...	Unknown	Male	3
221690	221691	WY01902	Rock Springs	Municipal Police	Sweetwater	Wyoming	2012	July	NaN	Murder or Manslaughter	...	Unknown	Male	2

221691 rows × 24 columns

In [53]:

```
cd.columns
```

Out[53]:

```
Index(['Record ID', 'Agency Code', 'Agency Name', 'Agency Type', 'City',
       'State', 'Year', 'Month', 'Incident', 'Crime Type', 'Crime Solved',
       'Victim Sex', 'Victim Age', 'Victim Race', 'Victim Ethnicity',
       'Perpetrator Sex', 'Perpetrator Age', 'Perpetrator Race',
       'Perpetrator Ethnicity', 'Relationship', 'Weapon', 'Victim Count',
       'Perpetrator Count', 'Record Source'],
      dtype='object')
```

In [2]:

```
cd.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 221691 entries, 0 to 221690
Data columns (total 24 columns):
 #   Column                 Non-Null Count   Dtype  
---  ------                 --------------   -----  
 0   Record ID              221691 non-null  int64  
 1   Agency Code            221691 non-null  object 
 2   Agency Name            221691 non-null  object 
 3   Agency Type            221691 non-null  object 
 4   City                   221691 non-null  object 
 5   State                  221691 non-null  object 
 6   Year                   221691 non-null  int64  
 7   Month                  221691 non-null  object 
 8   Incident               0 non-null       float64
 9   Crime Type             221691 non-null  object 
 10  Crime Solved           221691 non-null  object 
 11  Victim Sex             221691 non-null  object 
 12  Victim Age             221691 non-null  int64  
 13  Victim Race            221691 non-null  object 
 14  Victim Ethnicity       221691 non-null  object 
 15  Perpetrator Sex        221691 non-null  object 
 16  Perpetrator Age        221691 non-null  int64  
 17  Perpetrator Race       221691 non-null  object 
 18  Perpetrator Ethnicity  221691 non-null  object 
 19  Relationship           221691 non-null  object 
 20  Weapon                 221691 non-null  object 
 21  Victim Count           221691 non-null  int64  
 22  Perpetrator Count      221691 non-null  int64  
 23  Record Source          221691 non-null  object 
dtypes: float64(1), int64(6), object(17)
memory usage: 40.6+ MB
```

In [3]:

```
cd1=cd.copy()
```

In [7]:

```
cd.median()
```

```
C:\Users\91984\AppData\Local\Temp\ipykernel_12216\3630406611.py:1: FutureWarning: The default value of numeric_only in
DataFrame.median is deprecated. In a future version, it will default to False. In addition, specifying 'numeric_only=No
ne' is deprecated. Select only valid columns or specify the value of numeric_only to silence this warning.
  cd.median()
```

Out[7]:

```
Record ID            110846.0
Year                   2006.0
Incident                  NaN
Victim Age               29.0
Perpetrator Age          21.0
Victim Count              0.0
Perpetrator Count         0.0
dtype: float64
```

In [10]:

```
#cd2=cd1.fillna(cd.median())
#cd2
cd.fillna(0)
```

Out[10]:

	Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Type	...	Victim Ethnicity	Perpetrator Sex	Perpetrato Ag
0	1	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	0.0	Murder or Manslaughter	...	Unknown	Male	2
1	2	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	0.0	Murder or Manslaughter	...	Unknown	Male	1
2	3	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	January	0.0	Manslaughter by Negligence	...	Unknown	Male	1
3	4	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	February	0.0	Manslaughter by Negligence	...	Unknown	Female	3
4	5	AK00101	Anchorage	Municipal Police	Anchorage	Alaska	2000	March	0.0	Murder or Manslaughter	...	Unknown	Unknown	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
221686	221687	WY01501	Cody	Municipal Police	Park	Wyoming	2012	February	0.0	Murder or Manslaughter	...	Unknown	Male	2
221687	221688	WY01701	Sheridan	Municipal Police	Sheridan	Wyoming	2012	September	0.0	Murder or Manslaughter	...	Unknown	Female	5
221688	221689	WY01900	Sweetwater County	Sheriff	Sweetwater	Wyoming	2012	August	0.0	Murder or Manslaughter	...	Unknown	Male	1
221689	221690	WY01902	Rock Springs	Municipal Police	Sweetwater	Wyoming	2012	February	0.0	Murder or Manslaughter	...	Unknown	Male	3
221690	221691	WY01902	Rock Springs	Municipal Police	Sweetwater	Wyoming	2012	July	0.0	Murder or Manslaughter	...	Unknown	Male	2

221691 rows × 24 columns

In [11]:

```
#1. Find the total number of crimes
print("Total number of crimes:",cd1['Crime Type'].value_counts().sum())
```

```
Total number of crimes: 221691
```

In [13]:

```
#2. Find the Unique number of Investigation Agency
print("Unique number of Investigation Agency:",cd1['Agency Name'].value_counts().sum())
```

```
Unique number of Investigation Agency: 221691
```

In [28]:

```
#3.Find the total number of crimes reported by each Investigation Agency.
cd1.groupby('Agency Name').count()['Crime Type']
```

Out[28]:

```
Agency Name
Abbeville            29
Abbeville County      3
Aberdeen             27
Aberdeen Township     2
Abernathy             2
                     ..
Zeeland               3
Zephyrhills           7
Zolfo Springs         2
Zumbrota              1
Zuni Tribal           1
Name: Crime Type, Length: 7142, dtype: int64
```

In [29]:

```
#4.Find the unique number of Investigation Agency Type
cd1['Agency Type'].value_counts().sum()
```

Out[29]:

```
221691
```

In [30]:

```
#5. Find the total number of crimes by Investigation Agency Type
cd1.groupby('Agency Type').count()['Crime Type']
```

Out[30]:

```
Agency Type
County Police         9553
Municipal Police    168696
Regional Police        104
Sheriff              37442
Special Police        1244
State Police          4624
Tribal Police           28
Name: Crime Type, dtype: int64
```

In [31]:

```
#6. Find the total number of crimes reported in each state
cd1.groupby('State').count()['Crime Type']
```

Out[31]:

```
State
Alabama                 3350
Alaska                   558
Arizona                 6017
Arkansas                2320
California             32130
Colorado                2376
Connecticut             1584
Delaware                 556
District of Columbia    1734
Florida                15231
Georgia                 8168
Hawaii                   375
Idaho                    429
Illinois                7528
Indiana                 4576
Iowa                     691
Kansas                  1376
Kentucky                2206
Louisiana               7296
Maine                    341
Maryland                6947
Massachusetts           2309
Michigan                9268
Minnesota               1576
Mississippi             2473
Missouri                5532
Montana                  294
Nebraska                 495
Nevada                  2567
New Hampshire            206
New Jersey              5439
New Mexico              1857
New York               11849
North Carolina          7485
North Dakota             125
Ohio                    6834
Oklahoma                3066
Oregon                  1263
Pennsylvania            9879
Rhodes Island            443
South Carolina          4497
South Dakota             205
Tennessee               6015
Texas                  19453
Utah                     791
Vermont                  162
Virginia                5415
Washington              2807
West Virginia            939
Wisconsin               2461
Wyoming                  197
Name: Crime Type, dtype: int64
```

In [34]:

```
#7. Find the total number of crimes in each year
cd1.groupby('Year').count()['Crime Type']
```

Out[34]:

```
Year
2000    14671
2001    15803
2002    16268
2003    16512
2004    16233
2005    16836
2006    17275
2007    17303
2008    15595
2009    15840
2010    15121
2011    14756
2012    15033
2013    14445
Name: Crime Type, dtype: int64
```

In [35]:

```
#8. Find the total number of crimes in each month and year
cd1.groupby(['Month','Year']).count()['Crime Type']
```

Out[35]:

```
Month      Year
April      2000    1114
           2001    1268
           2002    1273
           2003    1375
           2004    1335
                   ...
September  2009    1266
           2010    1252
           2011    1248
           2012    1308
           2013    1169
Name: Crime Type, Length: 168, dtype: int64
```

In [39]:

```
#9. Find the total number of crimes by crime type
cd1.groupby('Crime Type').count()
```

Out[39]:

Crime Type	Record ID	Agency Code	Agency Name	Agency Type	City	State	Year	Month	Incident	Crime Solved	...	Victim Ethnicity	Perpetrator Sex	Perpetrator Age	Perpetrat Ra
Manslaughter by Negligence	3449	3449	3449	3449	3449	3449	3449	3449	0	3449	...	3449	3449	3449	34
Murder or Manslaughter	218242	218242	218242	218242	218242	218242	218242	218242	0	218242	...	218242	218242	218242	2182

2 rows × 23 columns

In [40]:

```
#10. Find the total number of crime type and Investigation agency
cd1.groupby('Agency Name').count()['Crime Type']
```

Out[40]:

```
Agency Name
Abbeville            29
Abbeville County      3
Aberdeen             27
Aberdeen Township     2
Abernathy             2
                     ..
Zeeland               3
Zephyrhills           7
Zolfo Springs         2
Zumbrota              1
Zuni Tribal           1
Name: Crime Type, Length: 7142, dtype: int64
```

In [48]:

```
#11. Find the total number of solved crimes.??
cd1['Crime Solved'].value_counts()
```

Out[48]:

```
Yes    153778
No      67913
Name: Crime Solved, dtype: int64
```

In [51]:

```
#12. Find the total number of solved crimes by investigation agency
cd1.groupby('Agency Name').count()['Crime Solved']
```

Out[51]:

```
Agency Name
Abbeville            29
Abbeville County      3
Aberdeen             27
Aberdeen Township     2
Abernathy             2
                     ..
Zeeland               3
Zephyrhills           7
Zolfo Springs         2
Zumbrota              1
Zuni Tribal           1
Name: Crime Solved, Length: 7142, dtype: int64
```

In [47]:

```
#cd1.loc[cd1['Perpetrator Sex']=='Perpetrator Sex','Perpetrator Sex']='Male'
#cd1
```

In [58]:

```
#13. Find the total number of victims by gender
cd1['Victim Sex'].value_counts()
```

Out[58]:

```
Male       172999
Female      48269
Unknown       423
Name: Victim Sex, dtype: int64
```

In [55]:

```
#14. Find the total number of crimes by victim race
cd1.groupby('Victim Race').count()['Crime Type']
```

Out[55]:

```
Victim Race
Asian/Pacific Islander             3868
Black                            106752
Native American/Alaska Native      1591
Unknown                            2767
White                            106713
Name: Crime Type, dtype: int64
```

In [54]:

```
cd1.columns
```

Out[54]:

```
Index(['Record ID', 'Agency Code', 'Agency Name', 'Agency Type', 'City',
       'State', 'Year', 'Month', 'Incident', 'Crime Type', 'Crime Solved',
       'Victim Sex', 'Victim Age', 'Victim Race', 'Victim Ethnicity',
       'Perpetrator Sex', 'Perpetrator Age', 'Perpetrator Race',
       'Perpetrator Ethnicity', 'Relationship', 'Weapon', 'Victim Count',
       'Perpetrator Count', 'Record Source'],
      dtype='object')
```

In [63]:

```
#15. Find the total number of solved crimes by victim sex and Investigation Agency
cd1.groupby(['Victim Sex','Agency Name']).count()['Crime Solved']
```

Out[63]:

```
Victim Sex  Agency Name
Female      Abbeville             4
            Aberdeen              8
            Aberdeen Township     1
            Abernathy             1
            Abilene              18
                                 ..
Unknown     Winter Garden         1
            Woodbury              1
            Wooster               1
            Wytheville            1
            Yakima                1
Name: Crime Solved, Length: 11860, dtype: int64
```

In [64]:

```
#16. Find the total number of Perpetrator by gender
cd1['Perpetrator Sex'].value_counts()
```

Out[64]:

```
Male       139446
Unknown     68041
Female      14204
Name: Perpetrator Sex, dtype: int64
```

In [71]:

```
#17. Find the total number of Perpetrator by perpetrator race
cd1.groupby(['Perpetrator Sex','Perpetrator Ethnicity']).count()['Perpetrator Race']
```

Out[71]:

```
Perpetrator Sex  Perpetrator Ethnicity
Female           Hispanic                855
                 Not Hispanic           3559
                 Unknown                9790
Male             Hispanic              14911
                 Not Hispanic          32814
                 Unknown               91721
Unknown          Hispanic                  7
                 Not Hispanic             37
                 Unknown               67997
Name: Perpetrator Race, dtype: int64
```

In [120]:

```
ind the total number of crimes by relationship between perpetrator and the victim
oupby(['Perpetrator Sex','Perpetrator Ethnicity','Perpetrator Race','Victim Sex','Victim Ethnicity','Victim Race','Relationship']).cou
```

Out[120]:

```
Perpetrator Sex  Perpetrator Ethnicity  Perpetrator Race  Victim Sex  Victim Ethnicity  Victim Race  Relationship
Female           Hispanic               Black             Female      Hispanic          Black        Acquaintance    1
                                                                                                     Friend          1
                                                                                                     Stranger        1
                                                                                          White      Daughter        1
                                                            Male        Hispanic          Black        Family          1
                                                                                                                     ..
Unknown          Unknown                White             Male        Unknown           White        Acquaintance    2
                                                                                                     Family          1
                                                                                                     Friend          1
                                                                                                     Stranger        1
                                                                                                     Unknown         6
Name: Crime Type, Length: 2441, dtype: int64
```

In [81]:

```
#19. Find the mostly use weapon
cd1['Weapon'].value_counts().max()
```

Out[81]:

```
109611
```

In [83]:

```
cd1['Weapon'].value_counts()#.max()
```

Out[83]:

```
Handgun          109611
Knife             27371
Firearm           25551
Blunt Object      22654
Unknown           15185
Shotgun            6881
Rifle              6524
Strangulation      1814
Fire               1696
Suffocation        1532
Gun                1395
Drugs               886
Drowning            299
Poison              174
Explosives           74
Fall                 44
Name: Weapon, dtype: int64
```

In [84]:

```
cd1.columns
```

Out[84]:

```
Index(['Record ID', 'Agency Code', 'Agency Name', 'Agency Type', 'City',
       'State', 'Year', 'Month', 'Incident', 'Crime Type', 'Crime Solved',
       'Victim Sex', 'Victim Age', 'Victim Race', 'Victim Ethnicity',
       'Perpetrator Sex', 'Perpetrator Age', 'Perpetrator Race',
       'Perpetrator Ethnicity', 'Relationship', 'Weapon', 'Victim Count',
       'Perpetrator Count', 'Record Source'],
      dtype='object')
```

In [86]:

```
#20. Find the total number of crimes by state and city
cd1.groupby(['State','City']).count()['Crime Type']
```

Out[86]:

```
State    City
Alabama  Autauga       13
         Baldwin       52
         Barbour       10
         Bibb           1
         Blount         6
                       ..
Wyoming  Sweetwater    19
         Teton          2
         Uinta          6
         Washakie       2
         Weston         2
Name: Crime Type, Length: 2728, dtype: int64
```

In [87]:

```
#21. Find the total number of crimes by state and city and investigation agency
cd1.groupby(['State','City','Agency Name']).count()['Crime Type']
```

Out[87]:

```
State    City      Agency Name
Alabama  Autauga   Autauga        6
                   Prattville     7
         Baldwin   Baldwin        9
                   Bay Minette    3
                   Daphne         7
                                 ..
Wyoming  Uinta     Uinta          1
         Washakie  Washakie       1
                   Worland        1
         Weston    Newcastle      1
                   Weston         1
Name: Crime Type, Length: 9728, dtype: int64
```

In [92]:

```
#22.Find the total number of crimes by stateand city by Crime Types
cd1.groupby(['State','City']).count()['Crime Type']
```

Out[92]:

```
State    City
Alabama  Autauga       13
         Baldwin       52
         Barbour       10
         Bibb           1
         Blount         6
                       ..
Wyoming  Sweetwater    19
         Teton          2
         Uinta          6
         Washakie       2
         Weston         2
Name: Crime Type, Length: 2728, dtype: int64
```

In [97]:

```
#23. Find the average age of victim in each crime type
cd1.groupby('Crime Type').mean()['Victim Age']
#cd1['Victim Age'].mean()['Crime Type']
```

```
C:\Users\91984\AppData\Local\Temp\ipykernel_12216\226043484.py:2: FutureWarning: The default value of numeric_only in D
ataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_onl
y or select only columns which should be valid for the function.
  cd1.groupby('Crime Type').mean()['Victim Age']
```

Out[97]:

```
Crime Type
Manslaughter by Negligence    27.423601
Murder or Manslaughter        33.531355
Name: Victim Age, dtype: float64
```

In [98]:

```
#24. Find the average age of victim in Manslaughter by Negligence crime.
cd1.groupby(['Crime Type','Victim Sex']).mean()['Victim Age']
```

```
C:\Users\91984\AppData\Local\Temp\ipykernel_12216\2345924892.py:2: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_on
ly or select only columns which should be valid for the function.
  cd1.groupby(['Crime Type','Victim Sex']).mean()['Victim Age']
```

Out[98]:

```
Crime Type                  Victim Sex
Manslaughter by Negligence  Female        26.486346
                            Male          27.732710
                            Unknown        6.200000
Murder or Manslaughter      Female        36.864842
                            Male          32.656578
                            Unknown       11.174641
Name: Victim Age, dtype: float64
```

In [99]:

```
#25. Find the average age of Perpetrator in each crime type
cd1.groupby('Crime Type').mean()['Perpetrator Age']
```

```
C:\Users\91984\AppData\Local\Temp\ipykernel_12216\159289448.py:2: FutureWarning: The default value of numeric_only in D
ataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_onl
y or select only columns which should be valid for the function.
  cd1.groupby('Crime Type').mean()['Perpetrator Age']
```

Out[99]:

```
Crime Type
Manslaughter by Negligence    28.013627
Murder or Manslaughter        19.933991
Name: Perpetrator Age, dtype: float64
```

In [100]:

```
#26. Find the average age of Perpetrator in Murder or Manslaughter crime
cd1.groupby(['Crime Type','Perpetrator Sex']).mean()['Perpetrator Age']
```

```
C:\Users\91984\AppData\Local\Temp\ipykernel_12216\3796029186.py:2: FutureWarning: The default value of numeric_only in
DataFrameGroupBy.mean is deprecated. In a future version, numeric_only will default to False. Either specify numeric_on
ly or select only columns which should be valid for the function.
  cd1.groupby(['Crime Type','Perpetrator Sex']).mean()['Perpetrator Age']
```

Out[100]:

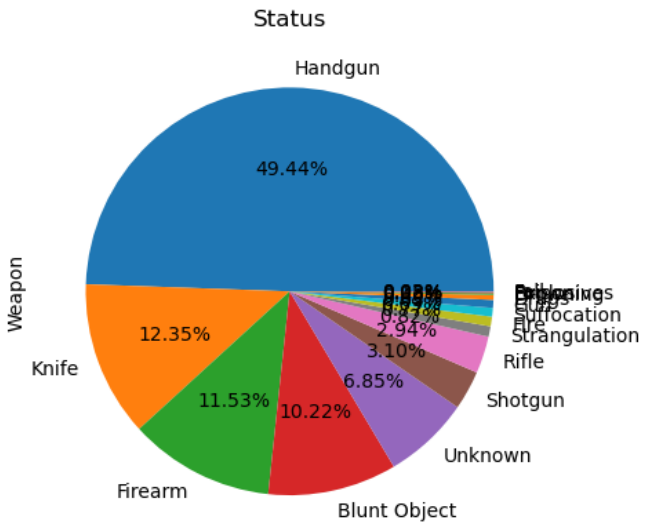
```
Crime Type                  Perpetrator Sex
Manslaughter by Negligence  Female             29.589506
                            Male               29.562309
                            Unknown             0.594595
Murder or Manslaughter      Female             32.748008
                            Male               28.431609
                            Unknown             0.238815
Name: Perpetrator Age, dtype: float64
```

In [105]:

```
#27. Create a pie chart to show most weapons used in the crime.
cd1['Weapon'].value_counts().plot(kind='pie',autopct='%1.2f%%')
cd1
plt.title('Status')
```

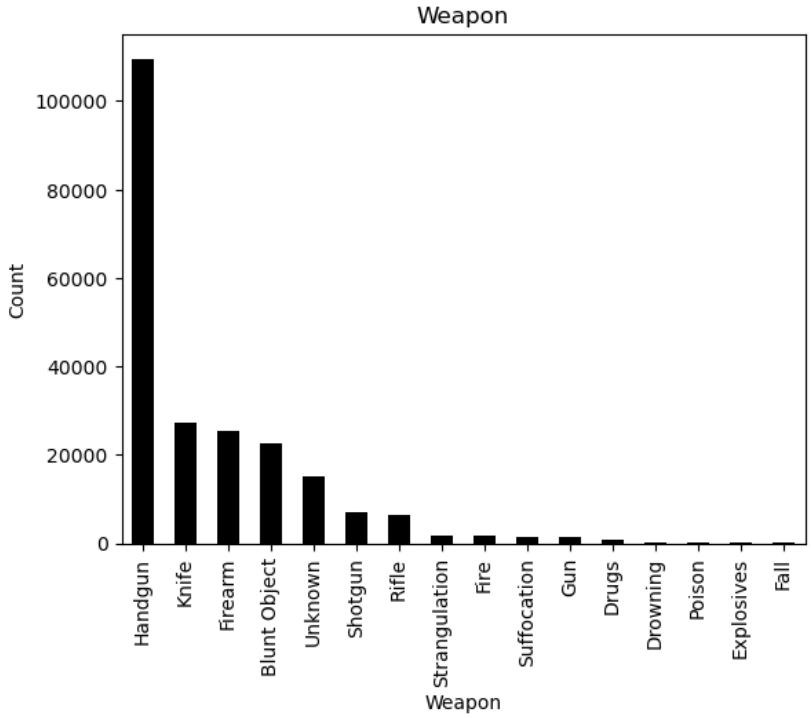
Out[105]:

```
Text(0.5, 1.0, 'Status')
```



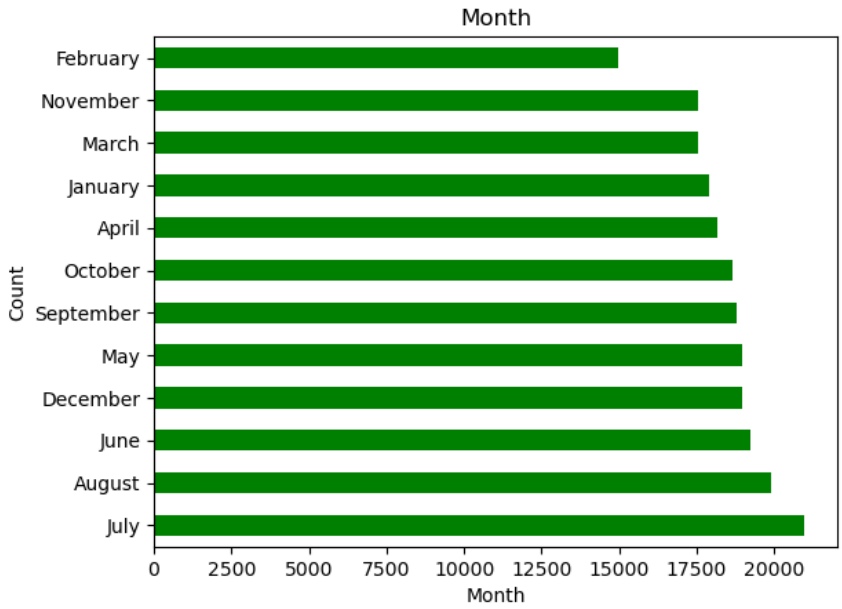
In [107]:

```
#28. Create visualization plots for all the possible questions. (bring appropriate charts, charts should be modified and it should be
cd1['Weapon'].value_counts(normalize= True)
cd1['Weapon'].value_counts(dropna= False).plot.bar(color='black')
plt.title('Weapon')
plt.xlabel('Weapon')
plt.ylabel('Count')
plt.show()
```



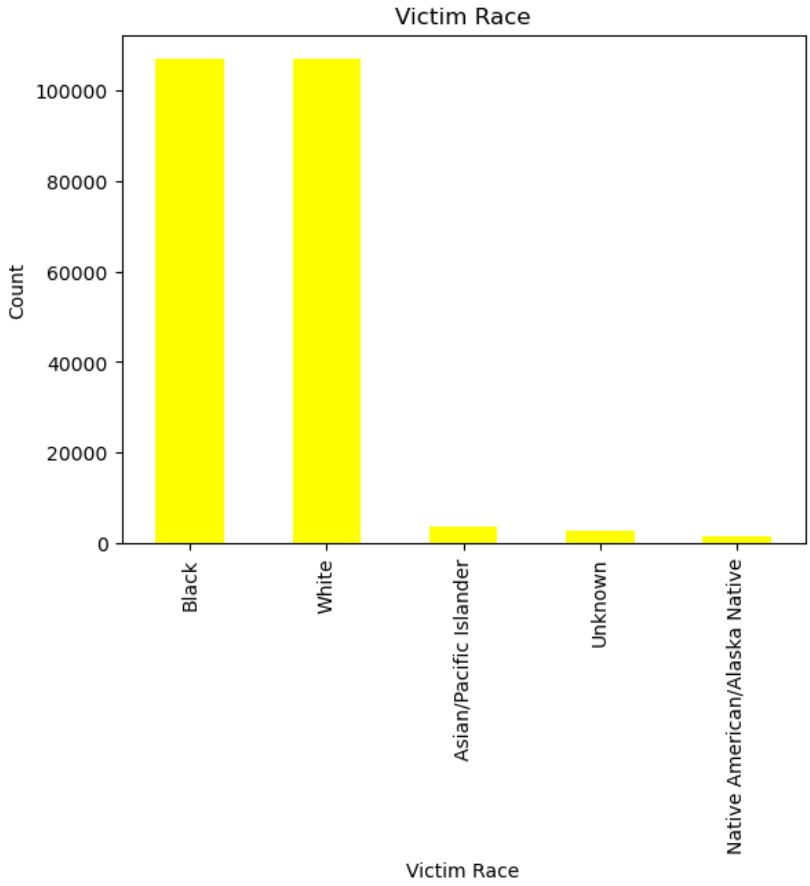
In [113]:

```
cd1['Month'].value_counts(normalize= True)
cd1['Month'].value_counts(dropna= False).plot.barh(color='green')
plt.title('Month')
plt.xlabel('Month')
plt.ylabel('Count')
plt.show()
```



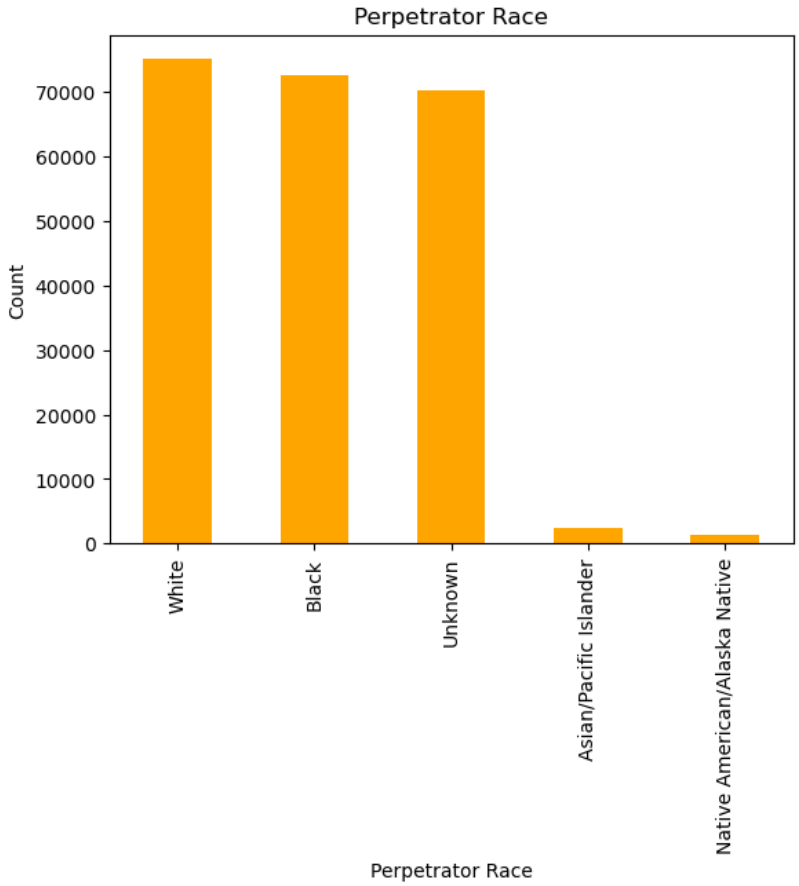
In [117]:

```
cd1['Victim Race'].value_counts(normalize= True)
cd1['Victim Race'].value_counts(dropna= False).plot.bar(color='yellow')
plt.title('Victim Race')
plt.xlabel('Victim Race')
plt.ylabel('Count')
plt.show()
```



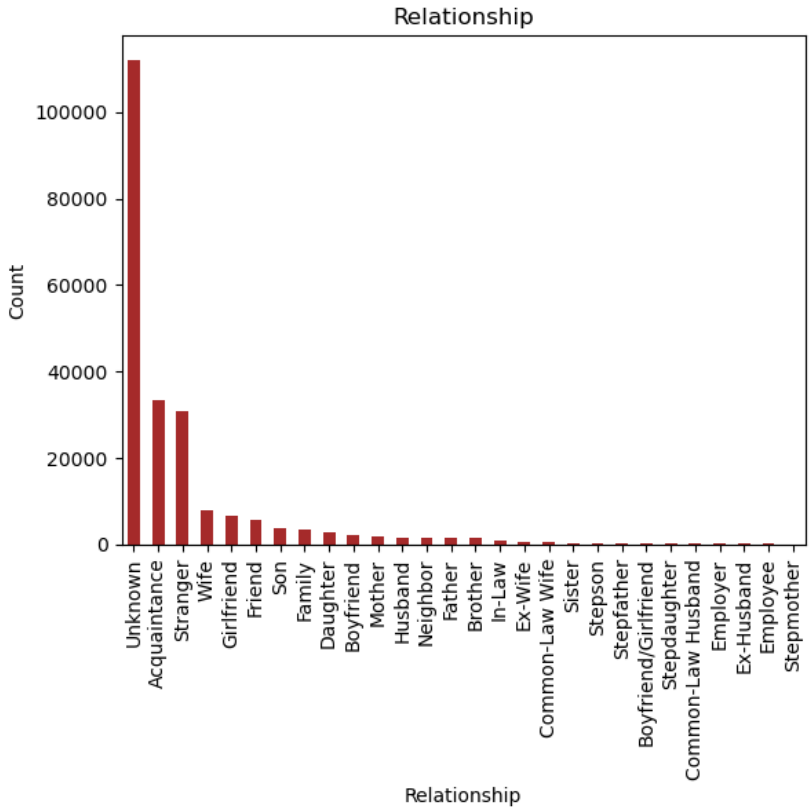
In [118]:

```
cd1['Perpetrator Race'].value_counts(normalize= True)
cd1['Perpetrator Race'].value_counts(dropna= False).plot.bar(color='orange')
plt.title('Perpetrator Race')
plt.xlabel('Perpetrator Race')
plt.ylabel('Count')
plt.show()
```



In [122]:

```
cd1['Relationship'].value_counts(normalize= True)
cd1['Relationship'].value_counts(dropna= False).plot.bar(color='brown')
plt.title('Relationship')
plt.xlabel('Relationship')
plt.ylabel('Count')
plt.show()
```



In [ ]:

```
```