Seattle

December 11, 2022

[1]:
```
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
```

[2]:
```
crimeDF= pd.read_csv("Seattle_crime.csv", low_memory=False)
weatherDF = pd.read_csv("Seattle_weather.csv")
```

```
C:\Users\dylan\anaconda3\lib\site-
packages\IPython\core\interactiveshell.py:3444: DtypeWarning: Columns (0,4) have
mixed types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

[3]:
```
crimeDF.head()
```

[3]:

	Report Number	Offense Start DateTime	Offense End DateTime	Report DateTime	Crime Against Category	Offense Parent Group	Offense
0	2020-044620	2/5/2020 10:10	NaN	2/5/2020 11:24	SOCIETY	DRUG/NARCOTIC OFFENSES	Drug/Narcotic Violations
1	2020-044452	2/3/2020 8:00	2/4/2020 8:00	2/5/2020 10:06	PROPERTY	LARCENY-THEFT	Theft of Motor Vehicle Parts or Accessories
2	2020-044465	2/2/2020 20:30	2/2/2020 21:30	2/5/2020 9:39	PROPERTY	ROBBERY	Robbery
3	2020-044225	2/5/2020 1:17	2/5/2020 2:21	2/5/2020 3:30	PROPERTY	DESTRUCTION/DAMAGE/VANDALISM OF PROPERTY	Destruction/Damage/Vandalism of Property
4	2020-044076	2/5/2020 0:51	NaN	2/5/2020 0:51	SOCIETY	DRIVING UNDER THE INFLUENCE	Driving Under the Influence

[4]:
```
weatherDF.head()
```

[4]:
```
          DATE  PRCP  TMAX  TMIN   RAIN  TAVG
0   12/14/2017   0.0  50.0  36.0  False  43.0
1   12/13/2017   0.0  48.0  34.0  False  41.0
2   12/12/2017   0.0  46.0  32.0  False  39.0
3   12/11/2017   0.0  49.0  29.0  False  39.0
4   12/10/2017   0.0  49.0  34.0  False  41.5
```

[5]:
```
df_split = crimeDF['Offense Start DateTime'].str.split(' ',expand=True)
```

[6]:
```
crimeDF['Offense Start DateTime'] = df_split[0]
crimeDF.rename(columns={"Offense Start DateTime": "DATE"},inplace = True)
```

[7]:
```
crimeDF.drop('Offense End DateTime', axis=1, inplace=True)
crimeDF.drop('Report DateTime', axis=1, inplace=True)
crimeDF.drop('Report Number', axis=1, inplace=True)
```

[8]:
```
crimeDF
```

[8]:
```
              DATE Crime Against Category  \
0         2/5/2020                SOCIETY
1         2/3/2020               PROPERTY
2         2/2/2020               PROPERTY
3         2/5/2020               PROPERTY
4         2/5/2020                SOCIETY
…              …                      …
1012484  7/13/2013               PROPERTY
1012485  6/26/2013               PROPERTY
1012486  2/14/2012               PROPERTY
1012487  9/19/2010               PROPERTY
1012488  2/25/2010               PROPERTY

                            Offense Parent Group  \
0                         DRUG/NARCOTIC OFFENSES
1                                  LARCENY-THEFT
2                                        ROBBERY
3        DESTRUCTION/DAMAGE/VANDALISM OF PROPERTY
4                    DRIVING UNDER THE INFLUENCE
…                                              …
1012484                      MOTOR VEHICLE THEFT
1012485                      MOTOR VEHICLE THEFT
1012486                            LARCENY-THEFT
1012487                            LARCENY-THEFT
1012488                      MOTOR VEHICLE THEFT

                                          Offense
0                         Drug/Narcotic Violations
1        Theft of Motor Vehicle Parts or Accessories
```

```
2                                          Robbery
3          Destruction/Damage/Vandalism of Property
4                        Driving Under the Influence
…                                                 …
1012484                        Motor Vehicle Theft
1012485                        Motor Vehicle Theft
1012486                                Shoplifting
1012487                                Shoplifting
1012488                        Motor Vehicle Theft

[1012489 rows x 4 columns]
```

[9]:
```
# weatherDF.to_csv('Seattle_weather_trimmed.csv')
# crimeDF.to_csv('Seattle_crime_trimmed.csv')
```

[10]:
```
un = crimeDF['Crime Against Category'].unique()
un
```

[10]:
```
array(['SOCIETY', 'PROPERTY', 'PERSON', 'NOT_A_CRIME'], dtype=object)
```

[20]:
```
crime_2017= pd.read_csv("Seattle_crime_2014-2017.csv", low_memory=False)
weather_2017 = pd.read_csv("Seattle_weather_2014-2017.csv")
```

[21]:
```
split = crime_2017['DATE'].str.split('/',expand=True)
crime_2017['Month'] = split[0]
crime_2017['Year'] = split[2]
crime_2017
```

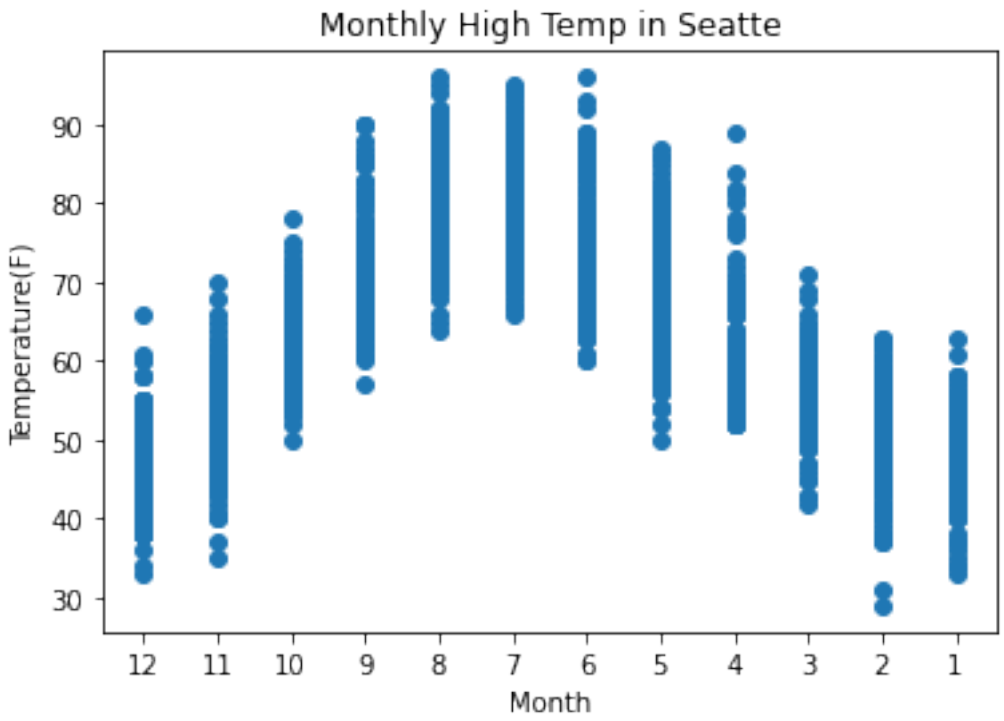
[21]:
```
             DATE Crime Against Category        Offense Parent Group  \
0        4/19/2017                PROPERTY               LARCENY-THEFT
1       12/27/2017                PROPERTY               LARCENY-THEFT
2        3/11/2016                PROPERTY               LARCENY-THEFT
3         4/7/2017                PROPERTY              FRAUD OFFENSES
4         9/6/2017                PROPERTY              FRAUD OFFENSES
…              …                       …                         …
280862   3/17/2016                PROPERTY  BURGLARY/BREAKING&ENTERING
280863   9/12/2014                PROPERTY                BAD CHECKS
280864   3/15/2016                PROPERTY  BURGLARY/BREAKING&ENTERING
280865  11/16/2014                 SOCIETY       WEAPON LAW VIOLATIONS
280866   3/20/2014                PROPERTY         MOTOR VEHICLE THEFT

                                        Offense Month  Year
0                       Theft From Motor Vehicle     4  2017
1      Theft of Motor Vehicle Parts or Accessories    12  2017
2                            All Other Larceny     3  2016
3                                Impersonation     4  2017
4                                Impersonation     9  2017
```

```
…                                          …   …   …
280862                 Burglary/Breaking & Entering     3  2016
280863                                  Bad Checks     9  2014
280864                 Burglary/Breaking & Entering     3  2016
280865                       Weapon Law Violations    11  2014
280866                         Motor Vehicle Theft     3  2014

[280867 rows x 6 columns]
```

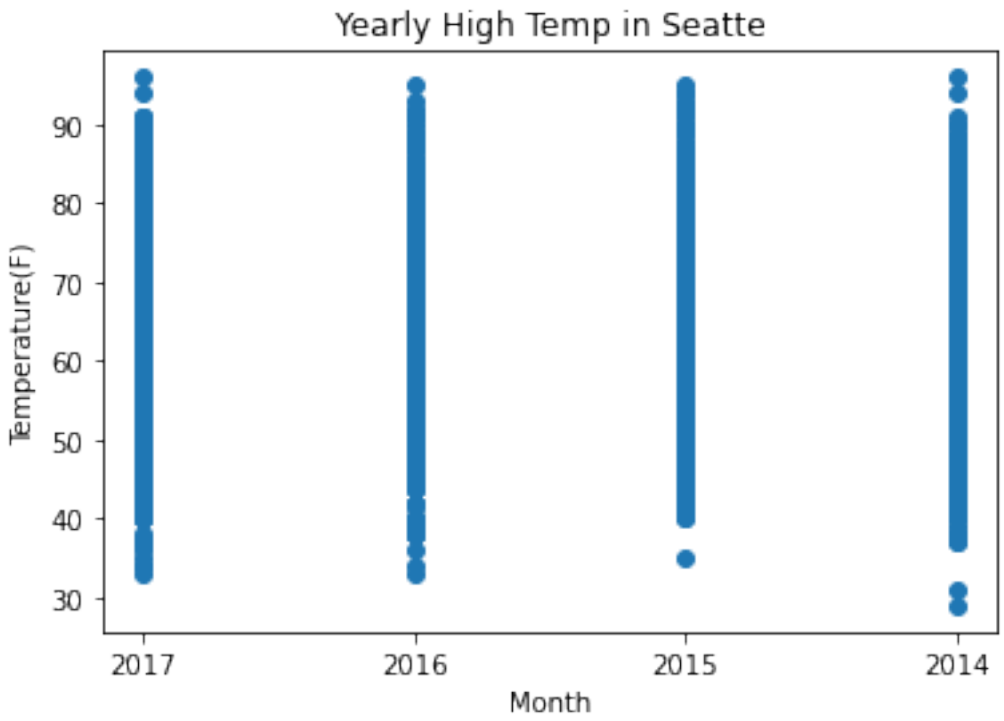
[22]:
```
splitW = weather_2017['DATE'].str.split('/',expand=True)
weather_2017['Month'] = splitW[0]
weather_2017['Year'] = splitW[2]
weather_2017
```

[22]:
```
            DATE  PRCP  TMAX  TMIN   RAIN  TAVG Month  Year
0     12/14/2017  0.00    50    36  False  43.0    12  2017
1     12/13/2017  0.00    48    34  False  41.0    12  2017
2     12/12/2017  0.00    46    32  False  39.0    12  2017
3     12/11/2017  0.00    49    29  False  39.0    12  2017
4     12/10/2017  0.00    49    34  False  41.5    12  2017
…           …   …   …   …    …   …   …   …
1439    1/5/2014  0.00    47    31  False  39.0     1  2014
1440    1/4/2014  0.00    46    33  False  39.5     1  2014
1441    1/3/2014  0.06    48    37   True  42.5     1  2014
1442    1/2/2014  0.16    51    43   True  47.0     1  2014
1443    1/1/2014  0.00    45    38  False  41.5     1  2014

[1444 rows x 8 columns]
```

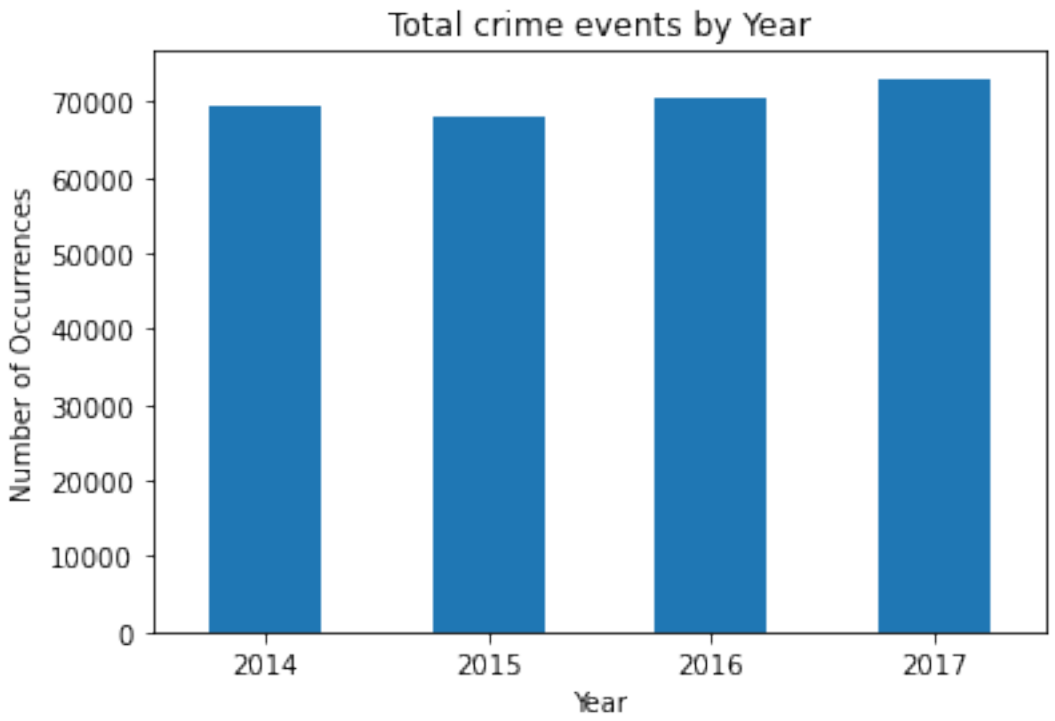
[26]:
```
plt.scatter(weather_2017["Month"], weather_2017["TMAX"])
plt.title("Monthly High Temp in Seatte")
plt.xlabel("Month")
plt.ylabel("Temperature(F)")
plt.show()
```



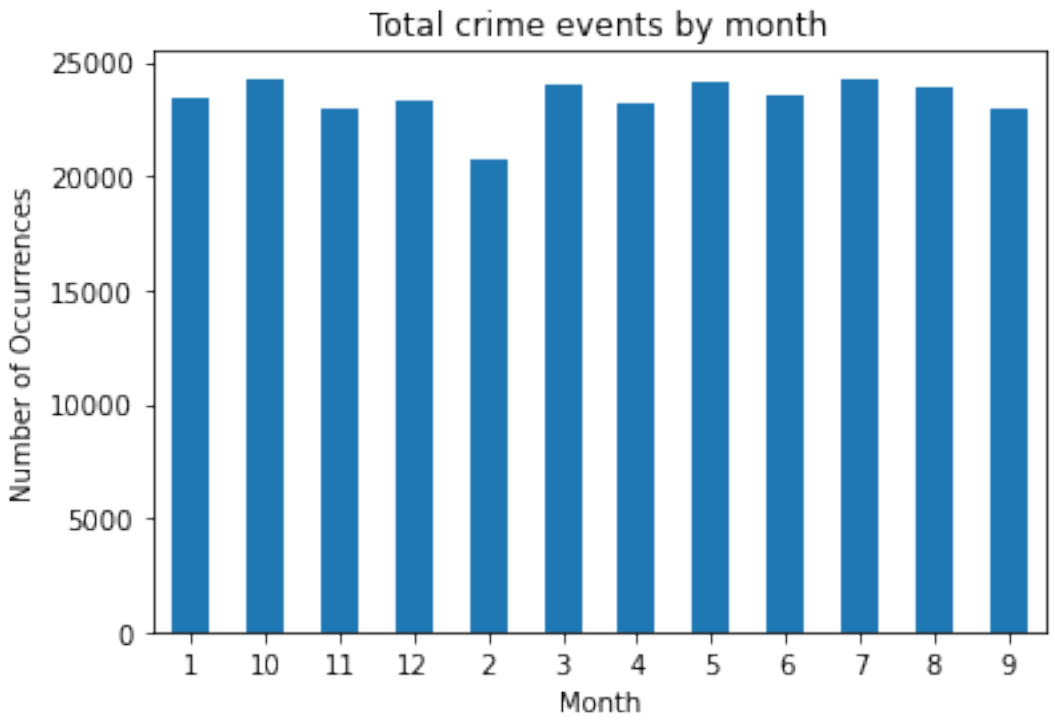
[27]:
```
plt.scatter(weather_2017["Year"], weather_2017["TMAX"])
plt.title("Yearly High Temp in Seatte")
plt.xlabel("Month")
plt.ylabel("Temperature(F)")
plt.show()
```



[28]:
```
ax = crime_2017.groupby('Year').size().plot(kind = "bar", title= "Total crime␣
↪events by Year")
ax.set_ylabel("Number of Occurrences")
plt.xticks(rotation=0)
plt.show()
```



[29]:
```
ax = crime_2017.groupby('Month').size().plot(kind = "bar", title= "Total crime␣
 ↪events by month")
ax.set_ylabel("Number of Occurrences")
plt.xticks(rotation=0)
plt.show()
```



[30]:
```
joined = pd.merge(crime_2017, weather_2017,how='inner',on='DATE')
```

[31]:
```
joined
```

[31]:

	DATE	Crime Against Category	Offense Parent Group	Offense	Month_x	Year_x	PRCP	TMAX	TMIN	RAIN
0	4/19/2017	PROPERTY	LARCENY-THEFT	Theft From Motor Vehicle	4	2017	0.35	56	46	True
1	4/19/2017	PROPERTY	LARCENY-THEFT	All Other Larceny	4	2017	0.35	56	46	True
2	4/19/2017	PROPERTY	FRAUD OFFENSES	Impersonation	4	2017	0.35	56	46	True
3	4/19/2017	PROPERTY	LARCENY-THEFT	Theft From Building	4	2017	0.35	56	46	True
4	4/19/2017	PROPERTY	FRAUD OFFENSES	Impersonation	4	2017	0.35	56	46	True
…	…	…	…	…						
277490	1/19/2014	PERSON	ASSAULT OFFENSES							
277491	1/19/2014	PROPERTY	LARCENY-THEFT							
277492	1/19/2014	PROPERTY	LARCENY-THEFT							
277493	1/19/2014	PERSON	ASSAULT OFFENSES							
277494	1/19/2014	PERSON	ASSAULT OFFENSES							

```
…                          …     …    …   …   …   …    …
277490            Simple Assault       1  2014  0.00   43   38  False
277491            Pocket-picking       1  2014  0.00   43   38  False
277492               Shoplifting       1  2014  0.00   43   38  False
277493            Simple Assault       1  2014  0.00   43   38  False
277494            Simple Assault       1  2014  0.00   43   38  False

        TAVG Month_y Year_y
0       51.0       4   2017
1       51.0       4   2017
2       51.0       4   2017
3       51.0       4   2017
4       51.0       4   2017
…       …     …    …
277490  40.5       1   2014
277491  40.5       1   2014
277492  40.5       1   2014
277493  40.5       1   2014
277494  40.5       1   2014

[277495 rows x 13 columns]
```

Probability of Offense Parent Group Given Cold, Mild, or Hot:

[32]:
```
types = {'HUMAN TRAFFICKING': 'Person', 'SEX OFFENSES, CONSENSUAL':␣
 ↪'Person','GAMBLING OFFENSES':'Person','BRIBERY': 'Society',
       'LIQUOR LAW VIOLATIONS':'Society','CURFEW/LOITERING/VAGRANCY␣
 ↪VIOLATIONS':'Society','PORNOGRAPHY/OBSCENE MATERIAL':'Society',
       'EMBEZZLEMENT':'Society','EXTORTION/BLACKMAIL':'Person','HOMICIDE␣
 ↪OFFENSES':'Person','PEEPING TOM':'Person','ARSON':'Property',
       'KIDNAPPING/ABDUCTION':'Person','COUNTERFEITING/FORGERY':
 ↪'Society','BURGLARY/BREAKING&ENTERING':'Property','ASSAULT OFFENSES':
 ↪'Person',
       'TRESPASS OF REAL PROPERTY':'Property','WEAPON LAW VIOLATIONS':
 ↪'Society','STOLEN PROPERTY OFFENSES':'Property','FAMILY OFFENSES,␣
 ↪NONVIOLENT':'Person',
       'DRUG/NARCOTIC OFFENSES':'Society','ROBBERY':'Property','PROSTITUTION␣
 ↪OFFENSES':'Society','DRIVING UNDER THE INFLUENCE':'Society',
       'DESTRUCTION/DAMAGE/VANDALISM OF PROPERTY':'Property','SEX OFFENSES':
 ↪'Person','MOTOR VEHICLE THEFT':'Property','BAD CHECKS':'Society',
       'FRAUD OFFENSES':'Society','LARCENY-THEFT':'Property'}
joined['Offense Parent Group'] = joined['Offense Parent Group'].map(types)
```

[33]:
```
offenses = joined['Offense Parent Group'].unique()
offenses
```

[33]:
```
array(['Property', 'Society', 'Person'], dtype=object)
```

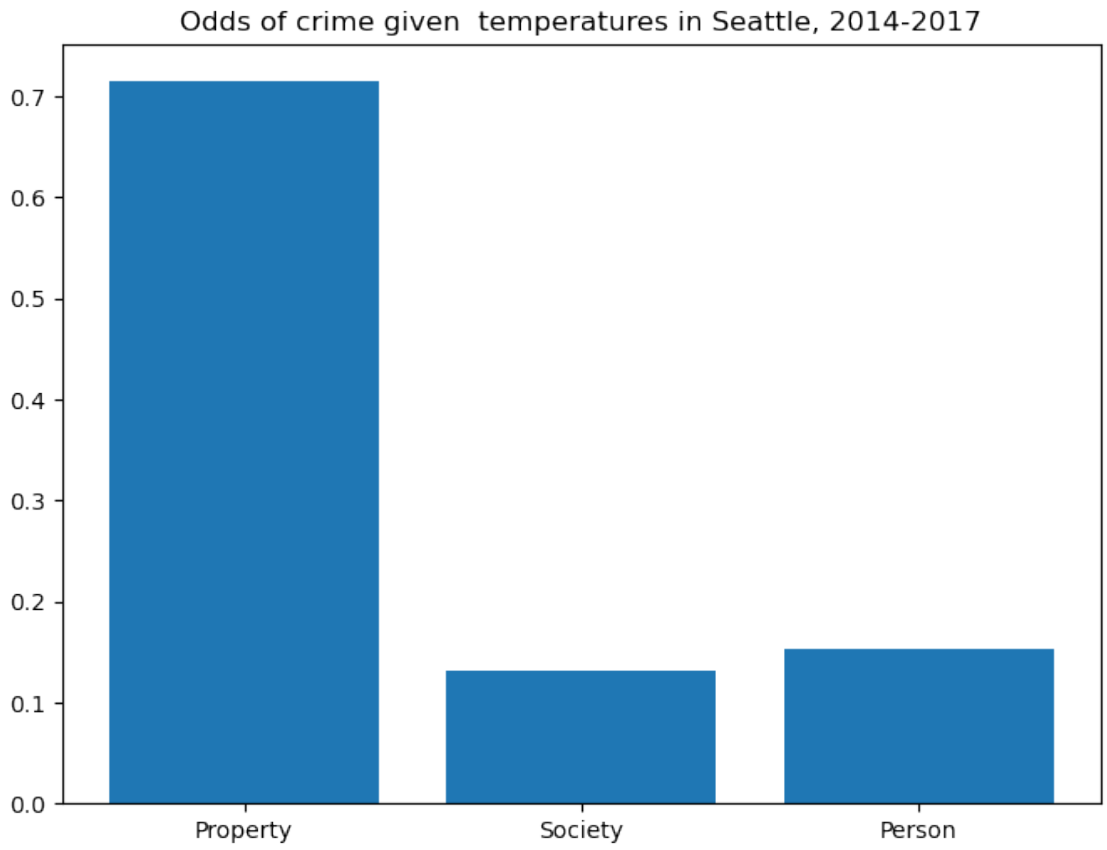
[34]:
```
temp = joined['TMAX'].astype(float)
tot_cold = temp[temp < 50.0].count()
tot_mild = temp[(temp >= 50.0) & (temp <= 80.0)].count()
tot_hot = temp[temp > 80.0].count()
total = temp.count()
cold = []
mild = []
hot = []
for offense in offenses:
    num_cold = len(joined[(joined['Offense Parent Group'] == offense) & (temp <␣
 ↪50.0)])
    num_mild = len(joined[(joined['Offense Parent Group'] == offense) & (temp␣
 ↪>= 50.0) & (temp <= 80.0)])
    num_hot = len(joined[(joined['Offense Parent Group'] == offense) & (temp >␣
 ↪80.0)])
    cold.append((offense,(num_cold/tot_cold).round(5)))
    mild.append((offense,(num_mild/tot_mild).round(5)))
    hot.append((offense,(num_hot/tot_hot).round(5)))
```

[35]:
```
cold
```

[35]:
```
[('Property', 0.71543), ('Society', 0.13151), ('Person', 0.15305)]
```

[46]:
```
off = []
perc = []
for o,p in cold:
    off.append(o)
    perc.append(p)

figure(figsize=(8, 6), dpi=100)
plt.title("Odds of crime given  temperatures in Seattle, 2014-2017")
plt.bar(off,perc)
```

[46]:
```
<BarContainer object of 3 artists>
```

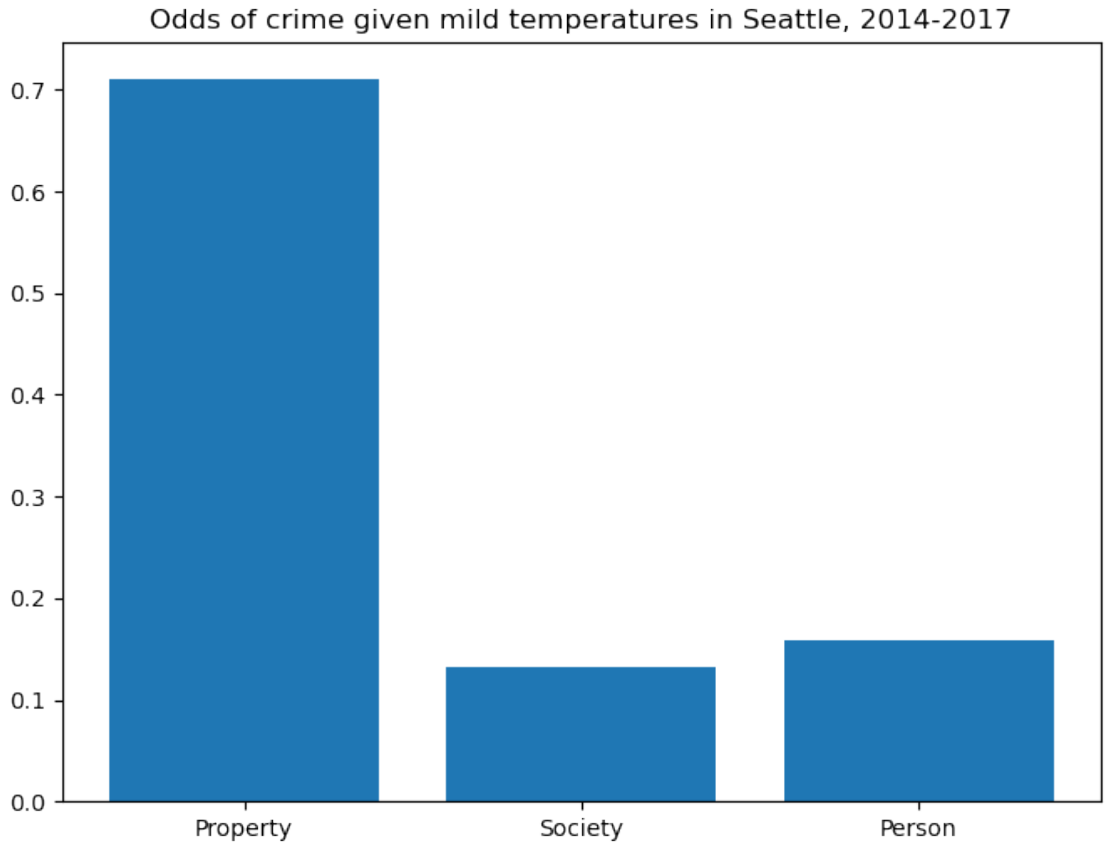


[47]:
```
mild
```

[47]:
```
[('Property', 0.71066), ('Society', 0.13157), ('Person', 0.15777)]
```

[48]:
```
off = []
perc = []
for o,p in mild:
    off.append(o)
    perc.append(p)

figure(figsize=(8, 6), dpi=100)
plt.title("Odds of crime given mild temperatures in Seattle, 2014-2017")
plt.bar(off,perc)
```

[48]:
```
<BarContainer object of 3 artists>
```

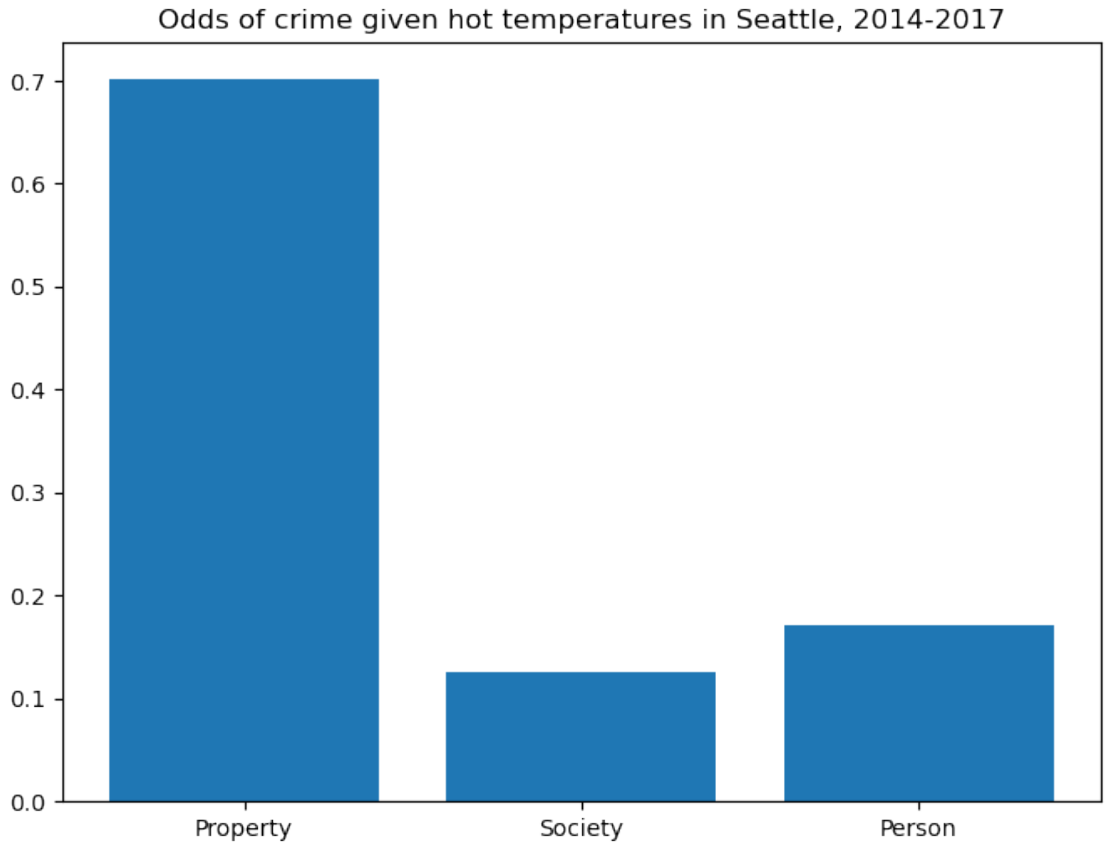


[49]:
```
hot
```

[49]:
```
[('Property', 0.70184), ('Society', 0.12648), ('Person', 0.17168)]
```

[50]:
```
off = []
perc = []
for o,p in hot:
    off.append(o)
    perc.append(p)

figure(figsize=(8, 6), dpi=100)
plt.title("Odds of crime given hot temperatures in Seattle, 2014-2017")
plt.bar(off,perc)
```

[50]:
```
<BarContainer object of 3 artists>
```



Data Spread

[51]:
```
coldDF = joined[joined['TMAX'] < 50.0]
```

[52]:
```
coldDF
```

[52]:

	DATE	Crime Against Category	Offense Parent Group	Offense	PRCP	TMAX	TMIN	RAIN	TAVG
1333	1/1/2014	PERSON	Person						
1334	1/1/2014	PROPERTY	Property						
1335	1/1/2014	PROPERTY	Society						
1336	1/1/2014	PROPERTY	Property						
1337	1/1/2014	PROPERTY	Society						
…	…	…	…						
277490	1/19/2014	PERSON	Person						
277491	1/19/2014	PROPERTY	Property						
277492	1/19/2014	PROPERTY	Property						
277493	1/19/2014	PERSON	Person						
277494	1/19/2014	PERSON	Person						

```
1333              Fondling   0.0    45    38  False  41.5
1334      All Other Larceny   0.0    45    38  False  41.5
1335          Impersonation   0.0    45    38  False  41.5
1336    Theft From Building   0.0    45    38  False  41.5
1337          Impersonation   0.0    45    38  False  41.5
…                       …   …   …   …    …   …
277490       Simple Assault   0.0    43    38  False  40.5
277491       Pocket-picking   0.0    43    38  False  40.5
277492          Shoplifting   0.0    43    38  False  40.5
277493       Simple Assault   0.0    43    38  False  40.5
277494       Simple Assault   0.0    43    38  False  40.5

[44429 rows x 9 columns]
```

[55]:
```
mildDF = joined[(joined['TMAX'] >= 50.0) & (joined['TMAX'] <= 80.0)]
mildDF
```

[55]:
```
             DATE Crime Against Category Offense Parent Group  \
0       4/19/2017                PROPERTY             Property
1       4/19/2017                PROPERTY             Property
2       4/19/2017                PROPERTY              Society
3       4/19/2017                PROPERTY             Property
4       4/19/2017                PROPERTY              Society
…           …                       …                    …
277135  1/15/2014                PROPERTY             Property
277136  1/15/2014                 SOCIETY              Society
277137  1/15/2014                PROPERTY             Property
277138  1/15/2014                 SOCIETY              Society
277139  1/15/2014                 SOCIETY              Society

                                          Offense  PRCP  TMAX  TMIN   RAIN  \
0                         Theft From Motor Vehicle  0.35    56    46   True
1                                All Other Larceny  0.35    56    46   True
2                                    Impersonation  0.35    56    46   True
3                               Theft From Building  0.35    56    46   True
4                                    Impersonation  0.35    56    46   True
…                                              …   …   …   …    …
277135                           All Other Larceny  0.00    52    42  False
277136                       Drug/Narcotic Violations  0.00    52    42  False
277137  Theft of Motor Vehicle Parts or Accessories  0.00    52    42  False
277138                       Drug/Narcotic Violations  0.00    52    42  False
277139                       Drug/Narcotic Violations  0.00    52    42  False

        TAVG
0       51.0
1       51.0
2       51.0
```

```
3       51.0
4       51.0
…       …
277135  47.0
277136  47.0
277137  47.0
277138  47.0
277139  47.0

[202113 rows x 9 columns]
```

[56]:
```
hotDF = joined[joined['TMAX'] > 80.0]
```

[57]:
```
hotDF
```

[57]:
```
             DATE Crime Against Category Offense Parent Group  \
599      9/6/2017                PROPERTY              Society
600      9/6/2017                PROPERTY             Property
601      9/6/2017                PROPERTY              Society
602      9/6/2017                PROPERTY             Property
603      9/6/2017                PROPERTY             Property
…           …                     …                    …
269047  7/13/2014                PROPERTY             Property
269048  7/13/2014                PROPERTY              Society
269049  7/13/2014                PROPERTY             Property
269050  7/13/2014                PROPERTY              Society
269051  7/13/2014                 SOCIETY              Society

                                        Offense  PRCP  TMAX  TMIN   RAIN  \
599                                Impersonation   0.0    81    63  False
600    Destruction/Damage/Vandalism of Property   0.0    81    63  False
601                                Impersonation   0.0    81    63  False
602                            All Other Larceny   0.0    81    63  False
603                            All Other Larceny   0.0    81    63  False
…                                            …   …   …   …    …
269047                     Motor Vehicle Theft   0.0    85    59  False
269048  False Pretenses/Swindle/Confidence Game   0.0    85    59  False
269049                     Motor Vehicle Theft   0.0    85    59  False
269050  False Pretenses/Swindle/Confidence Game   0.0    85    59  False
269051                 Drug/Narcotic Violations   0.0    85    59  False

        TAVG
599     72.0
600     72.0
601     72.0
602     72.0
603     72.0
```

```
…       …
269047  72.0
269048  72.0
269049  72.0
269050  72.0
269051  72.0

[30953 rows x 9 columns]
```

[58]:
```
hot_person = len(hotDF[hotDF['Offense Parent Group'] == 'Person'])
hot_property = len(hotDF[hotDF['Offense Parent Group'] == 'Property'])
hot_society = len(hotDF[hotDF['Offense Parent Group'] == 'Society'])
cold_person = len(coldDF[coldDF['Offense Parent Group'] == 'Person'])
cold_property = len(coldDF[coldDF['Offense Parent Group'] == 'Property'])
cold_society = len(coldDF[coldDF['Offense Parent Group'] == 'Society'])
mild_person = len(mildDF[mildDF['Offense Parent Group'] == 'Person'])
mild_property = len(mildDF[mildDF['Offense Parent Group'] == 'Property'])
mild_society = len(mildDF[mildDF['Offense Parent Group'] == 'Society'])
```

[59]:
```
print("Hot and Person: {}".format(hot_person / len(joined)))
print("Hot and Property: {}".format(hot_property / len(joined)))
print("Hot and Society: {}".format(hot_society / len(joined)))
print("Mild and Person: {}".format(mild_person / len(joined)))
print("Mild and Property: {}".format(mild_property / len(joined)))
print("Mild and Society: {}".format(mild_society / len(joined)))
print("Cold and Person: {}".format(cold_person / len(joined)))
print("Cold and Property: {}".format(cold_property / len(joined)))
print("Cold and Society: {}".format(cold_society / len(joined)))
```

```
Hot and Person: 0.019149894592695362
Hot and Property: 0.07828609524495937
Hot and Society: 0.014108362312834466
Mild and Person: 0.11491017856177589
Mild and Property: 0.5176093262941674
Mild and Society: 0.09582875367123732
Cold and Person: 0.024504946035063695
Cold and Property: 0.1145462080397845
Cold and Society: 0.021056235247481937
```

[ ]: