This project focuses on analyzing crime patterns in Vancouver using weekly data from the Vancouver Police Department. The aim is to identify trends, locate crime hotspots, and understand how crime changes across the city over time.Employing Python for data analysis and creating an interactive report in Power BI, it offers valuable insights for community safety and law enforcement.

This script conducts Exploratory Data Analysis (EDA) on the cleaned crime dataset, which has been saved as a binary file. Data preparation was carried out in the Python script 'VancouverCrimesAnalysis_DataPreparation.py'.

1- Importing Libraries

In [1]:
```
import pandas as pd
import numpy as np
from IPython.display import display
import seaborn as sns
from datetime import datetime , timedelta
import matplotlib.pyplot as plt
from matplotlib.ticker import MaxNLocator
from matplotlib.colors import LinearSegmentedColormap
import warnings
from IPython.core.display import display, HTML
import folium
from folium.plugins import HeatMap
```

2- Reading the data (pickle files)

2-1- Read the Crime data

In [2]:
```
crime = pd.read_pickle('../Dataset/Prepared Data/crime_detailed_Prepared.pkl')
crime.head()
last_updatetime = crime.date.max()
crime_grouped_by_year_all = pd.read_pickle('../Dataset/Prepared Data/crime_grouped_by_
crime_grouped_by_year_all.head()
crime_grouped_by_year_all_before2023 = crime_grouped_by_year_all[crime_grouped_by_year
crime_curr = crime[crime.YEAR == last_updatetime.year]
```

2-2- Read the Population data

In [4]:
```
pop = pd.read_pickle('../Dataset/Prepared Data/VancouverPopulation_Prepared.pkl')
pop.head()
```

Out[4]:

	growth	growthRate	population	year
0	24300	0.008285	2957471	2035
1	24431	0.008399	2933171	2034
2	24582	0.008523	2908740	2033
3	24751	0.008656	2884158	2032
4	24929	0.008795	2859407	2031

2-3- Read the CPI data

In [5]:
```
cpi = pd.read_pickle('../Dataset/Prepared Data/cpi_Prepared.pkl')
cpi.head()
```

Out[5]:

	Year	All Items Index	Annual Percent Change
0	1971	21.4	NaN
1	1972	22.6	5.6
2	1973	24.2	7.1
3	1974	27.0	11.6
4	1975	30.0	11.1

2-4- Create a dataframe for the crimes occured this year

In [6]:
```
crime_curr_year = crime[crime.YEAR == last_updatetime.year]
crime_curr_year.head()
```

Out[6]:

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	
24	Break and Enter Commercial	2023	9	14	3	30	ALBERNI ST	West End	49106
101	Break and Enter Commercial	2023	4	1	4	7	BEACH AVE	West End	49019
134	Break and Enter Commercial	2023	4	3	0	50	BEACH AVE	Central Business District	49024
135	Break and Enter Commercial	2023	5	11	18	0	BEACH AVE	Central Business District	49024
184	Break and Enter Commercial	2023	8	9	4	31	BEACH AVE	Central Business District	49026

5 rows × 23 columns

3.Exploring the data

3-1- How many crimes happend this year?

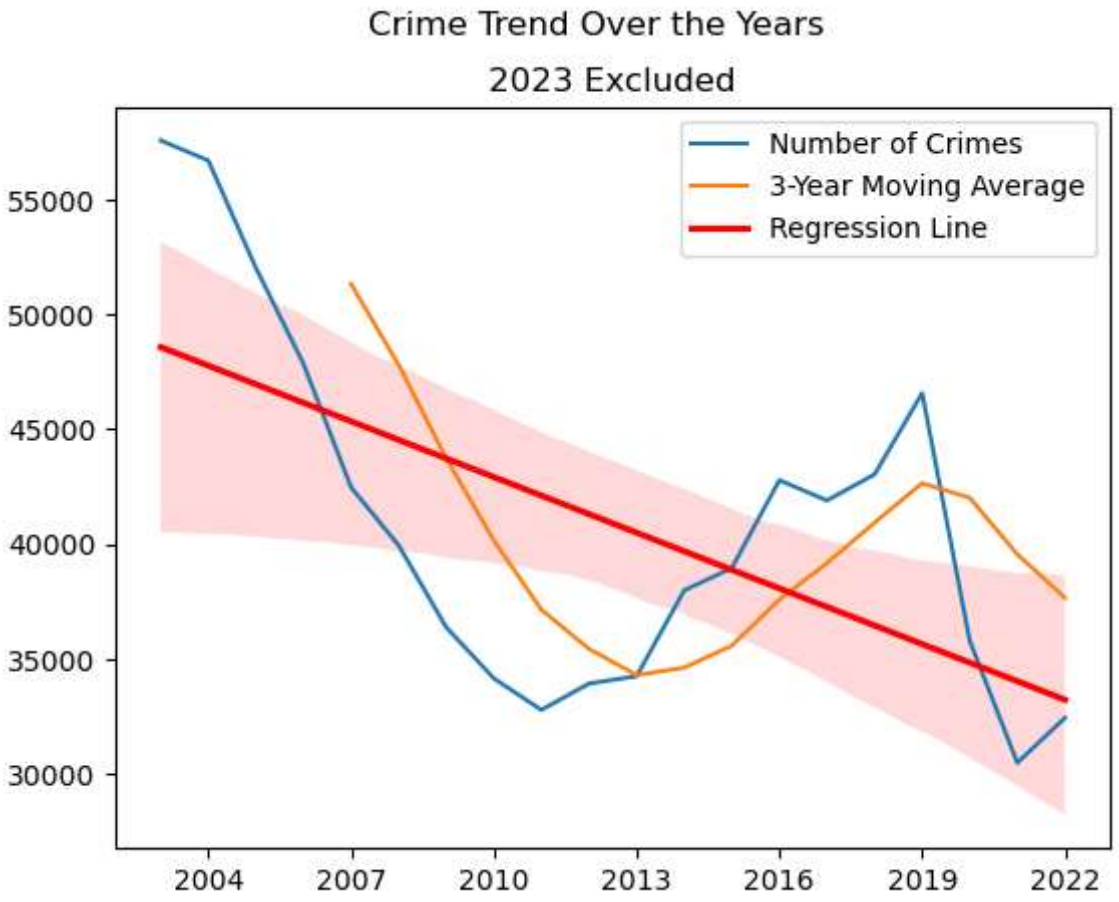
In [7]:
```
print ( "The number of crimes in 2023 : {} ".format(len(crime_curr_year)))
( print ("And the number of crimes for the same period last year was :{}"
   .format(len(crime[(crime.YEAR == last_updatetime.year -1  ) &
                     (crime.date <= last_updatetime - timedelta(days=365) ) ])) )
( print ("And the number of crimes for the same period last 2 year was :{}"
   .format(len(crime[(crime.YEAR == last_updatetime.year -2  ) &
                     (crime.date <= last_updatetime - timedelta(days=365) ) ])) )
```

```
The number of crimes in 2023 : 32189
And the number of crimes for the same period last year was :30747
And the number of crimes for the same period last 2 year was :30461
```

3-2- What Is the Trend in Crime Rates Over the Years? Is the number of crimes decreasing?

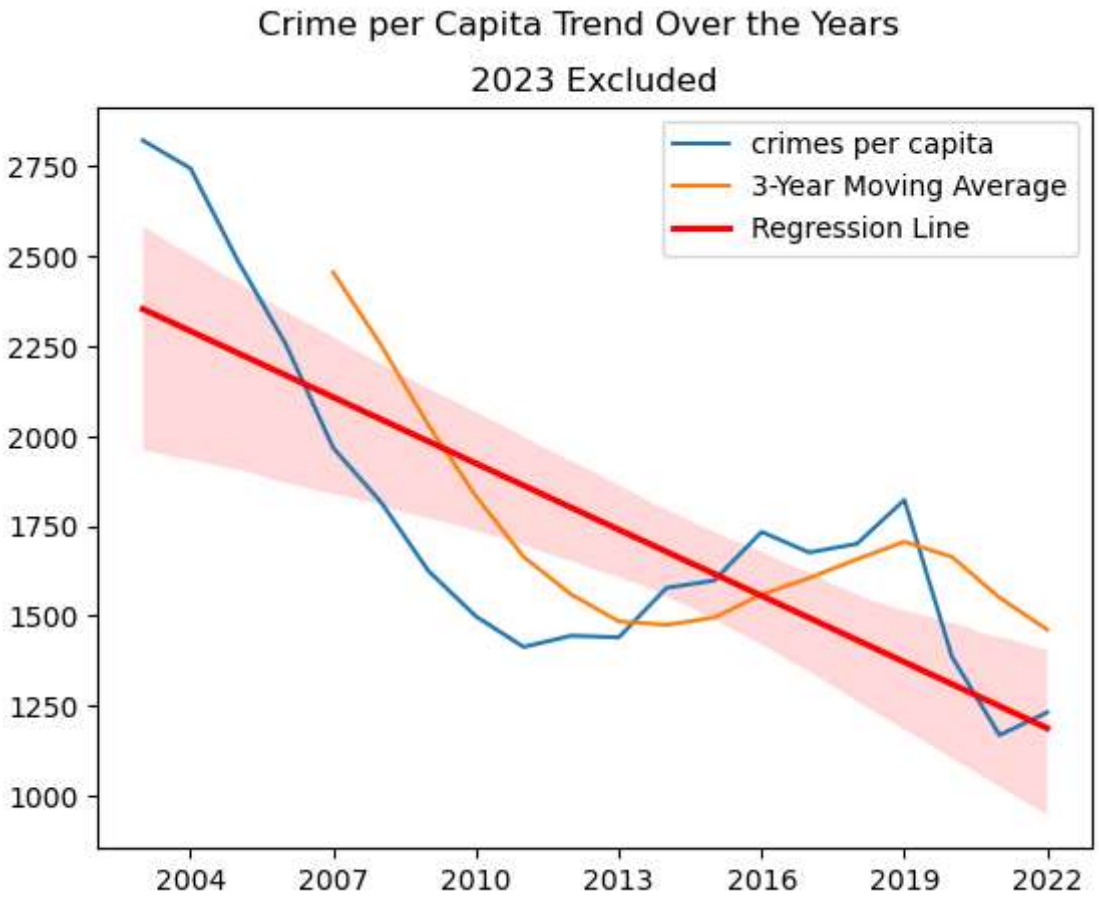
In [8]:
```
#display(crime_grouped_by_year_all_before2023[['YEAR', 'number_of_crimes']] )
print(""" \n >> The overall trend shows a decrease in the number of crimes over the ye
The highest number of crimes was recorded in 2003, and there has been a general declin
reaching its lowest in 2021. However, there are years, like between 2011 and 2019, whe
These fluctuations and potential causes need to be explored further.""")
#Adding 3-year-moving average
crime_grouped_by_year_all_before2023=crime_grouped_by_year_all_before2023.copy()
crime_grouped_by_year_all_before2023['moving_avg'] = crime_grouped_by_year_all_before2
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'number_of_
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'moving_avg
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
# Perform linear regression
sns.regplot(data=crime_grouped_by_year_all_before2023, x='YEAR', y='number_of_crimes',
plt.suptitle('Crime Trend Over the Years')
ax.set_title('2023 Excluded')
#plt.title('Crime Trend Over the Years- 2023 not included')
plt.xlabel('')
plt.ylabel('')
plt.legend(loc='best')
plt.show()
```

```
 >> The overall trend shows a decrease in the number of crimes over the years, with s
ome fluctuations in between.The highest number of crimes was recorded in 2003, and th
ere has been a general decline since then, reaching its lowest in 2021. However, ther
e are years, like between 2011 and 2019, where we see some increase before it starts
decreasing again.These fluctuations and potential causes need to be explored further.
```

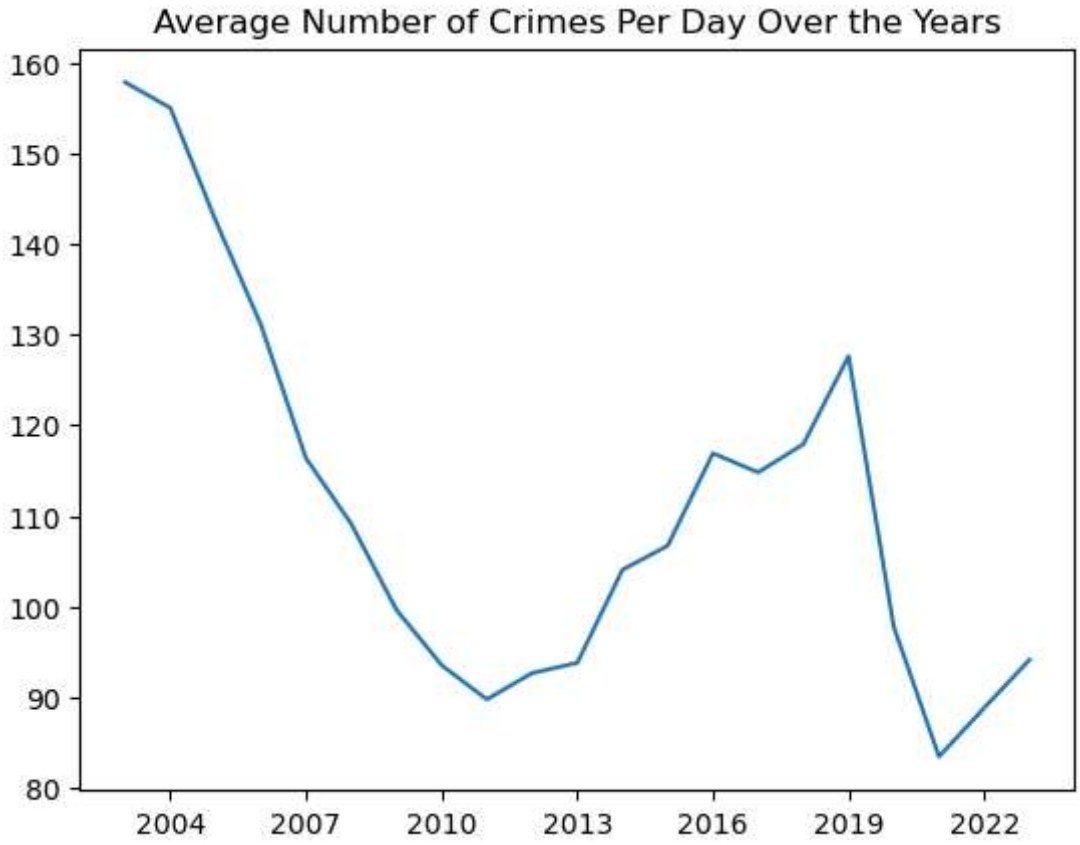


In [11]:
```
#display(crime_grouped_by_year_all_before2023[['YEAR', 'number_of_crimes']] )
print(""" \n >> Both measurements show the same patterns, showing that crime rate chan
in population. The crimes per capita metric adjusts for population size, suggesting th
has also decreased over time, particularly after 2003.""")
#Adding 3-year-moving average
crime_grouped_by_year_all_before2023=crime_grouped_by_year_all_before2023.copy()
crime_grouped_by_year_all_before2023['moving_avg'] = crime_grouped_by_year_all_before2
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'crimes_per
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'moving_avg
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
# Perform linear regression
sns.regplot(data=crime_grouped_by_year_all_before2023, x='YEAR', y='crimes_per_capita'
plt.suptitle('Crime per Capita Trend Over the Years')
ax.set_title('2023 Excluded')
#plt.title('Crime Trend Over the Years- 2023 not included')
plt.xlabel('')
plt.ylabel('')
plt.legend(loc='best')
plt.show()
```

```
 >> Both measurements show the same patterns, showing that crime rate changes aren't
just because of changesin population. The crimes per capita metric adjusts for popula
tion size, suggesting that the actual rate of crime (relative to population size)has
also decreased over time, particularly after 2003.
```



In [9]:
```
 sns.lineplot (data = crime
                  .groupby(['YEAR','date'])['TYPE']
                  .count()
                  .rename('crime_per_day')
                  .reset_index()
                  .groupby('YEAR')['crime_per_day']
                  .mean()

               )
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
plt.title('Average Number of Crimes Per Day Over the Years')
plt.xlabel('')
plt.ylabel('')
plt.show()
```



• *From 2003 to 2009*, there was a consistant decrease in the average daily crimes.

• *From 2011 to 2019*, there was a trend of increasing average daily crimes.

• *Starting in 2019* , there was another decrease in the average number of daily crimes, which continued through 2021. This decrease might be associated with the COVID-19 pandemic.

• *Rebound in 2022*, there was a slight increase in the average daily crime rate. This could indicate a potential rebound in criminal activity,

But further analysis would be needed to understand the underlying causes. During the period from 2003 to 2021, several significant events and factors potentially influencing crime rates in Vancouver include:

• 2010 Winter Olympics

• Global Economic Recession (2008)

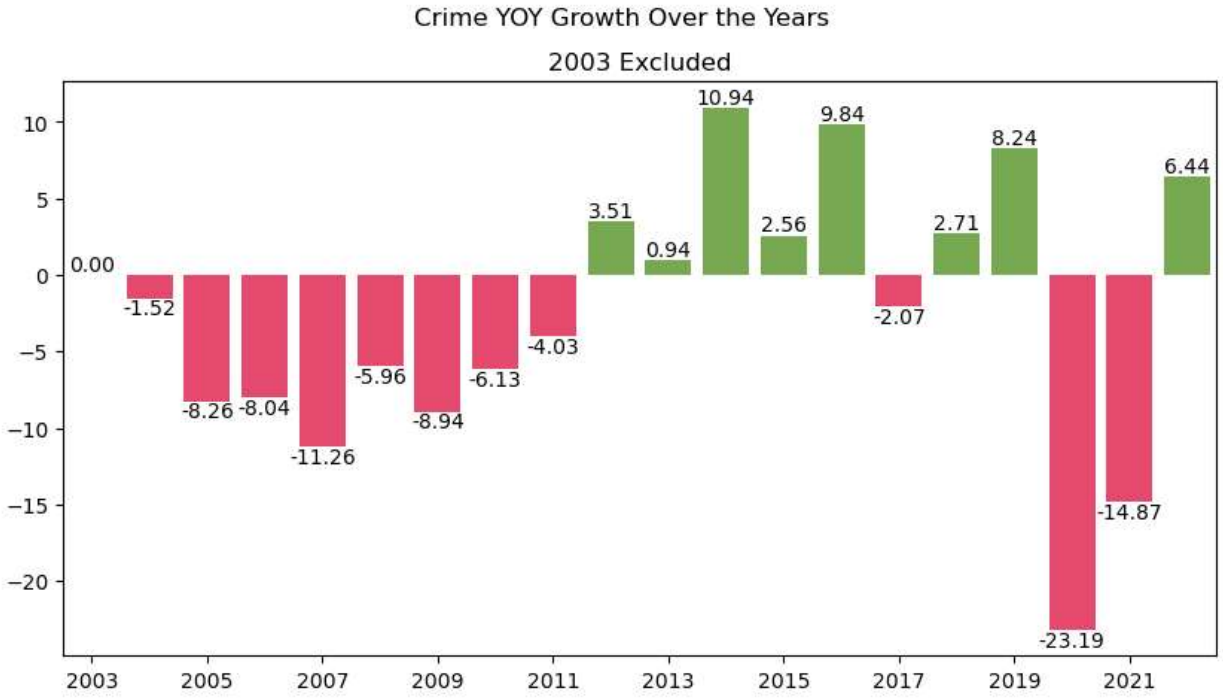
• Drug Trade and Addiction Challenges

• Social and Demographic Changes

A comprehensive analysis would require further access to data and a deeper examination

In [10]:
```
# Determine colors based on y-values
colors = ['#75B744' if value > 0 else '#FF2F5F' for value in crime_grouped_by_year_all

plt.figure(figsize=(10,5))
sns.barplot( data = crime_grouped_by_year_all_before2023, x='YEAR' , y = 'crimes_growt
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
plt.suptitle('Crime YOY Growth Over the Years')
```

```
#plt.title('Crime YOY Growth Over the Years - 2023 not included ' )
ax.set_title("2003 Excluded")
plt.xlabel('')
plt.ylabel('')

ax.bar_label(ax.containers[0],fmt="%.2f")
plt.show()
print("The crime growth rate has fluctuated over the years. The most remarkable declin

print("\nLet's see which crime types decreased more in 2020.")
df_merged_1920 = ( crime[crime.YEAR==2020].
  groupby('crime_type_category')['TYPE'].
  count().rename('no_of_crimes_2020').
  to_frame().
  reset_index()
).merge(crime[crime.YEAR==2019].
  groupby('crime_type_category')['TYPE'].
  count().rename('no_of_crimes_2019').
  to_frame().
  reset_index() )
df_merged_1920['percent_changed'] = (df_merged_1920['no_of_crimes_2020'] - df_merged_1
display(df_merged_1920.sort_values(by = 'percent_changed', ascending = True))
print('Theft and Vehicle Collision saw the largest declines in 2020, with decreases of
```



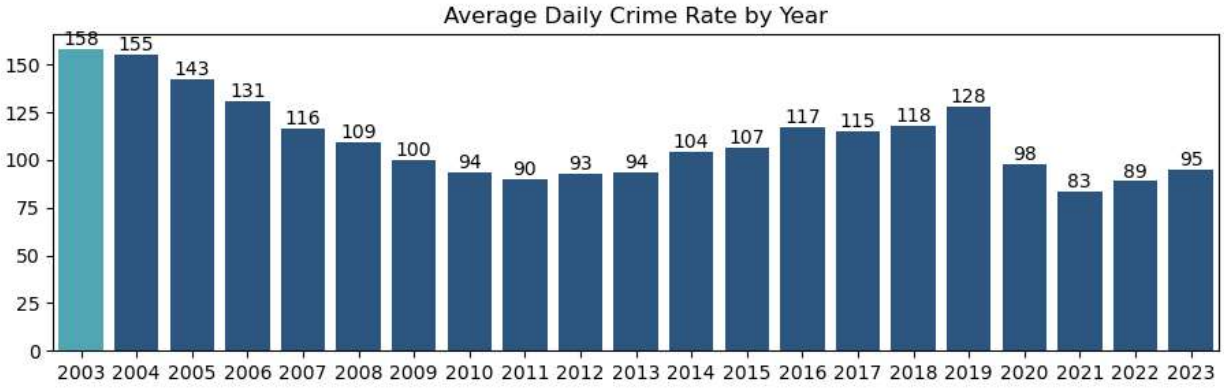
The crime growth rate has fluctuated over the years. The most remarkable decline happ
ened in 2020, witnessing a 23% drop. This significant decrease could likely be attrib
uted to the effects of the COVID-19 pandemic. Factors such as lockdowns or other rest
rictions might have played a role in deterring criminal activities. In 2022, there wa
s a 6.5% increase, possibly signaling a return to pre-pandemic crime levels or patter
ns.

Let's see which crime types decreased more in 2020.

	crime_type_category	no_of_crimes_2020	no_of_crimes_2019	percent_changed
4	Theft	21917	32927	-33.437604
5	Vehicle Collision	859	1151	-25.369244
0	Break and Enter	4871	4759	2.353436
2	Mischief	6109	5833	4.731699
3	Offence Against a Person	2006	1899	5.634544
1	Homicide	19	12	58.333333

```
Theft and Vehicle Collision saw the largest declines in 2020, with decreases of 27% a
nd 25%, respectively, compared to the previous year
```

In [19]:
```
def days_in_year(year):
    if year == last_updatetime.year:
        start_of_year = datetime(last_updatetime.year, 1, 1)
        return (last_updatetime - start_of_year).days
    if (year % 4 == 0 and year % 100 != 0) or (year % 400 == 0):
        return 366
    else:
        return 365
crime_avg_daily = (crime
                    .groupby('YEAR')['TYPE']
                    .count()
                    .rename('crime_per_day')
                    .to_frame()
                    .reset_index()
                   )

crime_avg_daily['crime_per_day'] = round(crime_avg_daily['crime_per_day']/ crime_avg_d
crime_avg_daily
plt.figure(figsize=(11,3))
max_value = crime_avg_daily['crime_per_day'].max()
colors = ['#3FB4C4' if v == max_value else '#1E558D' for v in crime_avg_daily['crime_p
sns.barplot(data = crime_avg_daily , x ='YEAR' , y = 'crime_per_day' , orient='v', pal
ax = plt.gca()
ax.set_title("Average Daily Crime Rate by Year")
plt.xlabel('')
plt.ylabel('')

ax.bar_label(ax.containers[0],fmt="%.f")
plt.show()
print("\nThe daily average of crimes reported this year, up to the time this report wa
```



```
The daily average of crimes reported this year, up to the time this report was genera
ted, has surpassed that of the prior year.
```

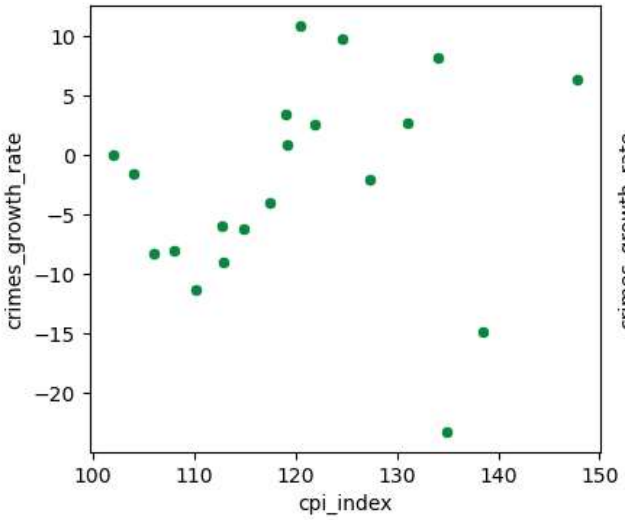
3-3- Does Population Growth Correlate with Crime Rate? And How Does CPI Influence This?

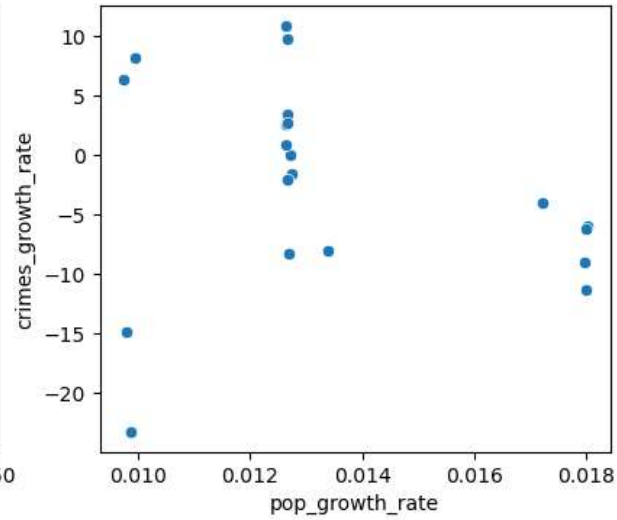
In [11]:
```
fig ,ax = plt.subplots(1,2 , figsize =(10,4))
sns.scatterplot(data = crime_grouped_by_year_all_before2023 , y='crimes_growth_rate' ,
sns.scatterplot(data = crime_grouped_by_year_all_before2023 , y='crimes_growth_rate' ,

correlation_value_pop = "{:.2f}".format(crime_grouped_by_year_all_before2023['pop_grow
correlation_value_cpi = "{:.2f}".format(crime_grouped_by_year_all_before2023['cpi_inde

print(">>Correlation between Crime Rate and CPI (Consumer Price Index): \
\nThere is a moderate to strong negative correlation between crimes per capita and the
correlation coefficient of {}. This suggests that higher economic health (reflected by
with lower crimes per capita. \
\n >>Correlation between Crime Rate and Population Growth:\
\nThe correlation between crime rate and population growth is very weak, as \
indicated by correlation coefficients of {} for crimes per capita. \
\nThese findings suggest that population growth has a minimal linear relationship with
\n In summary, the analysis indicates that economic factors (as reflected by the CPI)
\ncorrelation with crime rates compared to population growth, which shows a very weak

#print("There doesn't appear to be a strong relationship between the population growth
#nor between the change in the CPI index and the crime rate, as evidenced by correlati

#print("\nWhile our analysis reveals no strong correlation between the crime rate and

crime_grouped_by_year_all_before2023[['crimes_growth_rate','crimes_per_capita','cpi_ir
```

```
>>Correlation between Crime Rate and CPI (Consumer Price Index):
There is a moderate to strong negative correlation between crimes per capita and the
CPI index, as indicated by acorrelation coefficient of -0.75. This suggests that high
er economic health (reflected by a higher CPI index) is associated with lower crimes
per capita.
 >>Correlation between Crime Rate and Population Growth:
The correlation between crime rate and population growth is very weak, as indicated b
y correlation coefficients of 0.09 for crimes per capita.
These findings suggest that population growth has a minimal linear relationship with
crime rates.
 In summary, the analysis indicates that economic factors (as reflected by the CPI) h
ave a more noticeable
correlation with crime rates compared to population growth, which shows a very weak a
ssociation with crime rates.
```

Out[11]:

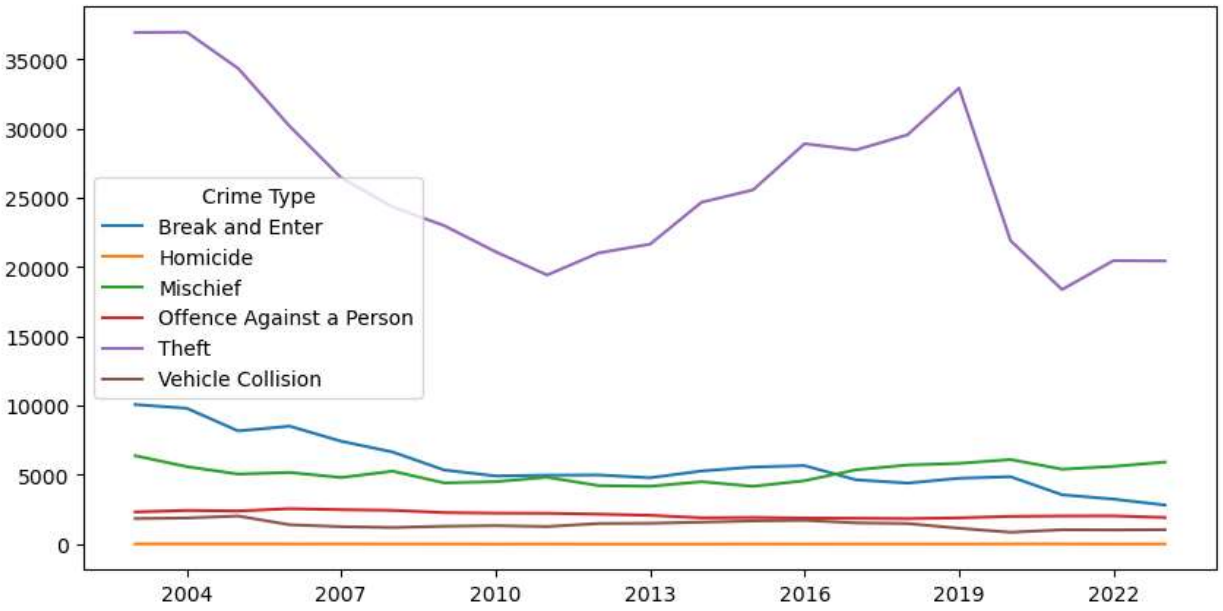
	crimes_growth_rate	crimes_per_capita	cpi_index	pop_growth_rate
crimes_growth_rate	1.000000	0.014185	0.124482	-0.196936
crimes_per_capita	0.014185	1.000000	-0.746847	0.085966
cpi_index	0.124482	-0.746847	1.000000	-0.595081
pop_growth_rate	-0.196936	0.085966	-0.595081	1.000000





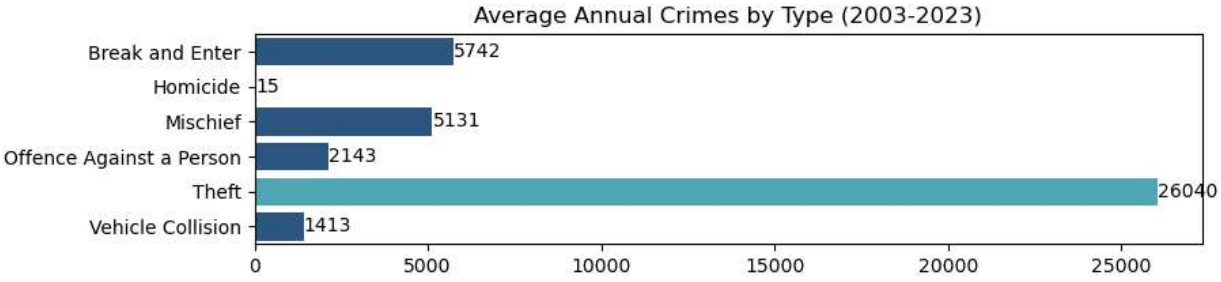
3-4- How Has Crime Trended Over the Years by Crime Type?

In [12]:
```
df = crime.groupby(['YEAR' ,'crime_type_category'])['TYPE'].count().rename('num_of_cri
plt.figure(figsize=(10,5))
sns.lineplot(data  = df , x = 'YEAR' , y = 'num_of_crimes', hue ='crime_type_category'
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
leg = plt.legend()
leg.set_title('Crime Type')

plt.title= 'Crime Trends Over the Years by Crime Type'
plt.xlabel('')
plt.ylabel('')
plt.show()
print("-Theft, despite its fluctuations, remains one of the most common crimes through
"-Homicide cases, although grave in nature, are relatively low and do not show a consi
"-The year 2020, likely influenced by the global pandemic and related restrictions, sh
```



```
-Theft, despite its fluctuations, remains one of the most common crimes throughout th
e years.
-Homicide cases, although grave in nature, are relatively low and do not show a consi
stent upward or downward trend.
-The year 2020, likely influenced by the global pandemic and related restrictions, sh
ows particular declines in several crime categories, most notably in theft and Vehicl
e Collision.
```

In [16]:
```
data = ( crime.
            groupby('crime_type_category')['TYPE'].
            count().
            rename('No_of_crimes').
            div(21).
            to_frame().
            reset_index()
       )

plt.figure(figsize=(9,2))
max_value = data.No_of_crimes.max()
colors = ['#3FB4C4' if v == max_value else "#1E558D" for v in data['No_of_crimes']]
sns.barplot(data = data   ,
            y='crime_type_category',x ='No_of_crimes' ,
            palette= colors)

ax = plt.gca()
ax.set_title("Average Annual Crimes by Type (2003-2023)")
plt.xlabel('')
plt.ylabel('')

ax.bar_label(ax.containers[0],fmt="%.f")
plt.show()
print("From 2003 to 2023, 'Theft' was the most frequent crime category averaging aroun
```




```
From 2003 to 2023, 'Theft' was the most frequent crime category averaging around 14,3
43 cases annually, while 'Homicide' was the least common with about 15 cases per yea
r.
```

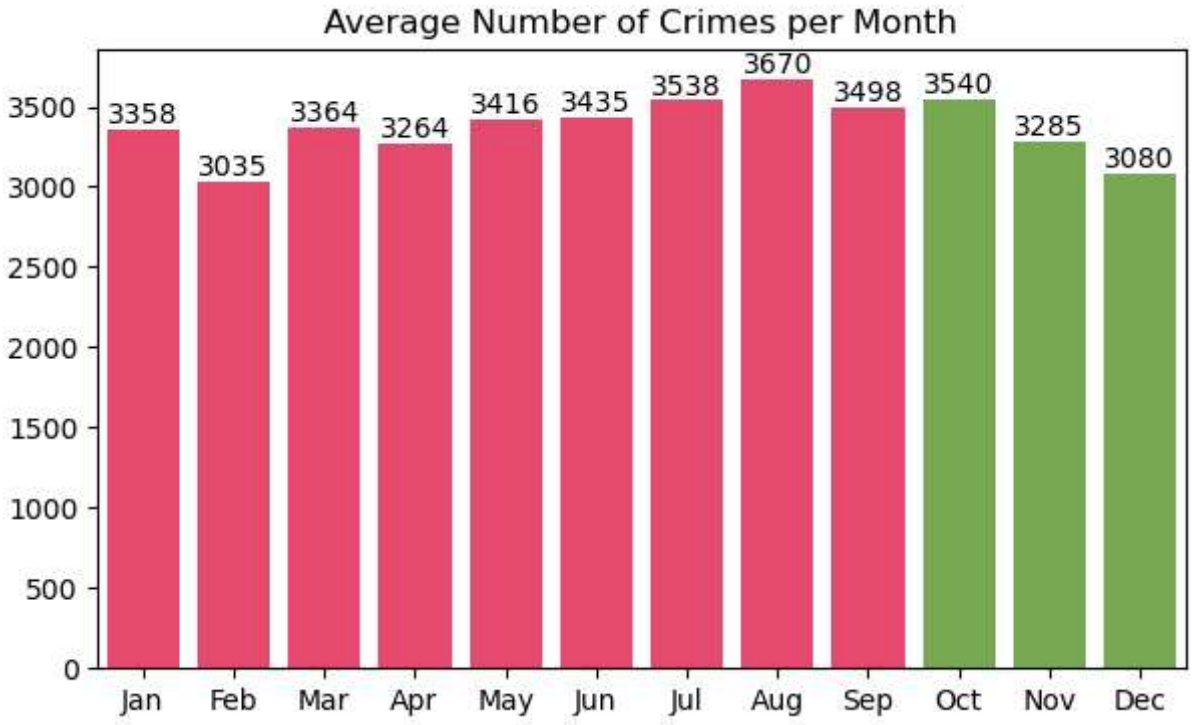
3-5- Is There any Seasonality?

3-5-1- Which Months Witness Higher/Lower Crime Rates

In [13]:
```
# First I create a dataframe for monthly crimes
print("First, let's examine the average number of crimes per month over the years.")
month_order = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul','Aug','Sep','Oct','Nov'
crime_momthly = crime.groupby(['YEAR','MONTH'])['TYPE'].count().rename('no_of_crimes')
crime_monthly_avg = (crime_momthly.
    groupby('MONTH')['no_of_crimes'].
    mean().
    rename('avg_no_of_crime').
    to_frame().
```

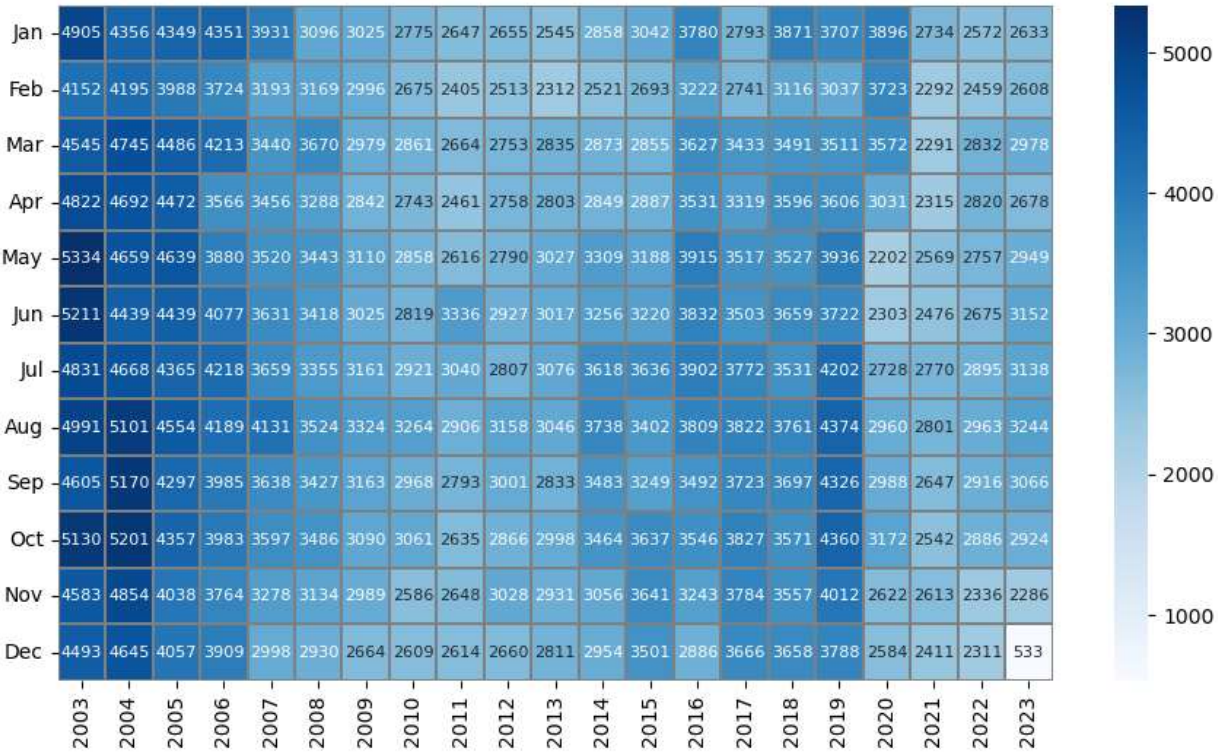
```
    reset_index()
)

plt.figure(figsize =(7,4))
sns.barplot(data =  crime_monthly_avg ,
            x ='MONTH' ,
            y = 'avg_no_of_crime',
            palette = colors,
) # lor = '#1E558D')
ax =plt.gca()
ax.set_title("Average Number of Crimes per Month")
plt.xlabel('')
plt.ylabel('')
ax.set_xticklabels(month_order)
ax.bar_label(ax.containers[0],fmt="%.f")
plt.show()
print("August appears to have the highest number of crimes, while February has the few
February typically has fewer days than other months, which can impact monthly crime st
print("\nNow, let's dig deeper to spot patterns in the monthly crime rates over the ye
plt.figure(figsize=(11,6))
sns.heatmap(data  = crime_momthly.pivot("MONTH" , "YEAR" , "no_of_crimes")  ,
            cmap='Blues' , # 'YlGnBu',
            linecolor='grey',
            linewidths=0.1,
            cbar=True,
            annot=True ,
            fmt=".0f" ,
            annot_kws={"fontsize": 8},
            yticklabels=month_order
)
ax.set_title('Number of Crime per Month and Year ', fontsize=10)
plt.xlabel('')
plt.ylabel('')
plt.yticks(rotation=0)
plt.show()
print("Crime rates generally peak during the warmer months of May to August, suggestin
rates and warmer weather and more outdoor activities. \
February often sees a decrease, possibly due to its shorter duration.\
Data for late 2023 appears incomplete.")
```

First, let's examine the average number of crimes per month over the years.



August appears to have the highest number of crimes, while February has the fewest.Fe bruary typically has fewer days than other months, which can impact monthly crime sta tistics

Now, let's dig deeper to spot patterns in the monthly crime rates over the years.

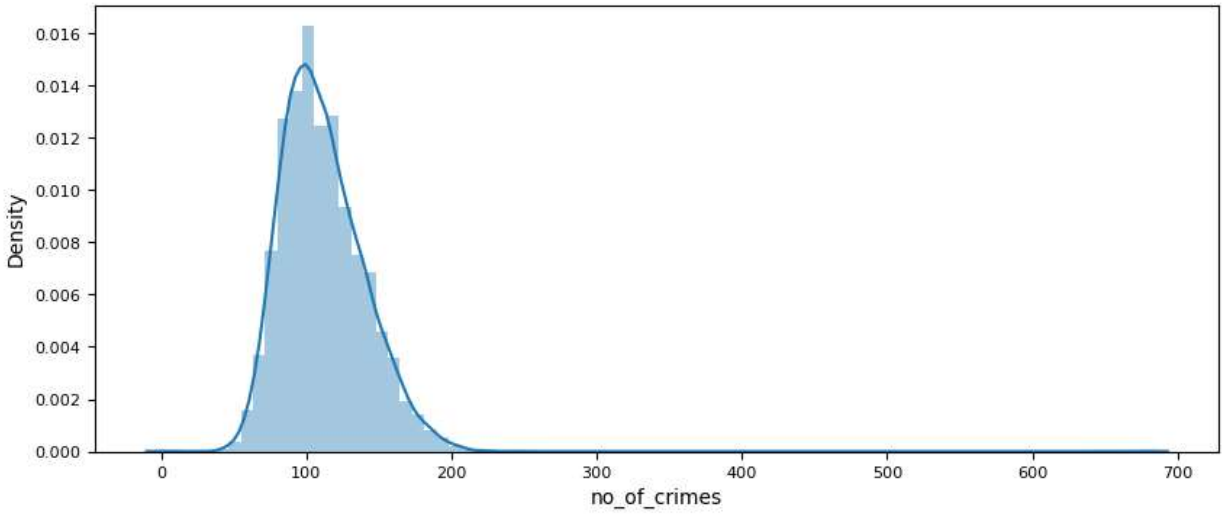


Crime rates generally peak during the warmer months of May to August, suggesting a po tential link between crimerates and warmer weather and more outdoor activities. Febru ary often sees a decrease, possibly due to its shorter duration.Data for late 2023 ap pears incomplete.

3-5-2- Which days seems to be more safe/dangerous ?

In [14]:
```
display(HTML("<style>div.output_scroll { height: auto; }</style>"))
warnings.filterwarnings('ignore')
crime_per_day  = crime.groupby('date')['TYPE'].count().rename('no_of_crimes')
display(crime_per_day.describe())
plt.figure(figsize=(10,4))
ax.set_title('Distribution of Crimes per day', fontsize=10)
plt.tick_params(labelsize=8)
#sns.distplot( crime.groupby('date')['TYPE'].count().rename('number_of_crimes_per_day'
sns.distplot( crime_per_day , bins = 80 ) ;
plt.show()
print("-The distribution appears to be normal, based on visual inspection, with a mean
      "\n-Half of the days had between 91 and 129 crimes."
      "\n-The peak day witnessed a significantly higher 678 crimes.")


print("\nLet's investigate the specific date with such a high number of crimes! \U0001
crime_per_day = crime_per_day.to_frame().reset_index()
q25 , q50 , q75 = np.percentile(crime_per_day.no_of_crimes,(25,50,75))
iqr = q75 - q25
min_crime = q25 - 1.5 * iqr
max_crime = q75 + 1.5 * iqr
q25 , q50 , q75  , min_crime , max_crime , crime_per_day.no_of_crimes.min() , crime_pe
crime_outlier = crime_per_day[  crime_per_day.no_of_crimes > max_crime ].sort_values(b
crime_outlier # .holiday.value_counts()
print(crime[crime.date=='2011-06-15'][['holiday','weekday']].drop_duplicates())

print("On June 6, 2011, we observed 678 crimes, which is 500% above the average of 112
"\nTo investigate the possible reasons for this extreme variation, I conducted a searc
"\nfactors that may have contributed to this specific day in Vancouver's history."
"\nAccording to the ")
display(HTML('<a href="https://www2.gov.bc.ca/assets/gov/law-crime-and-justice/crimina
print("in June 2011, a riot erupted in Vancouver during the Stanley Cup finals."
      "\nThe trouble began when it seemed the home team wouldn't win the championship.
and in five hours of rioting, members of the crowd committed multiple criminal offense
theft, mischief, and assault."
      "\nApproximately 55,000 people watched the game at a public viewing area on West
      in the streets surrounding these viewing areas.")
print("\nHere's a breakdown of the crime types that occurred on that day.... ")
print(crime[crime.date=='2011-06-15'].groupby('crime_type_category')['TYPE'].count().s
print("\n\n... and the neighborhoods most affected.")
print(crime[crime.date=='2011-06-15'].groupby('NEIGHBOURHOOD')['TYPE'].count().sort_va
```

```
count    7647.000000
mean      111.173794
std        28.683650
min         4.000000
25%        90.000000
50%       107.000000
75%       129.000000
max       679.000000
Name: no_of_crimes, dtype: float64
```



-The distribution appears to be normal, based on visual inspection, with a mean of 112 crimes per day.
-Half of the days had between 91 and 129 crimes.
-The peak day witnessed a significantly higher 678 crimes.

Let's investigate the specific date with such a high number of crimes!

	holiday	weekday
10991	not_holiday	2

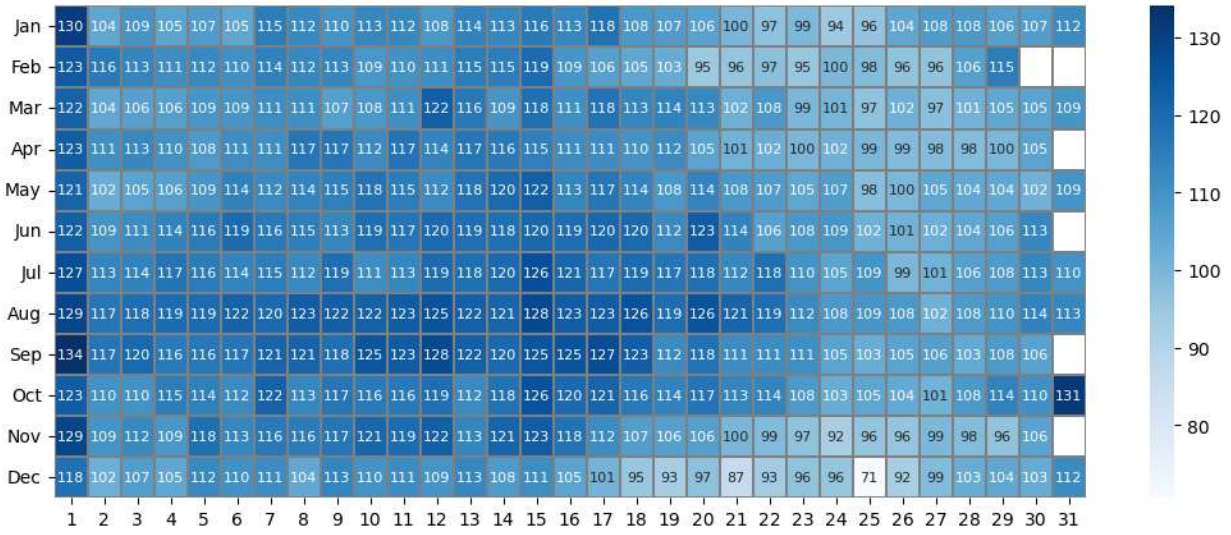
On June 6, 2011, we observed 678 crimes, which is 500% above the average of 112 crimes per day.
To investigate the possible reasons for this extreme variation, I conducted a search for significant events or
factors that may have contributed to this specific day in Vancouver's history.
According to the
Stanley Cup Riot Prosecutions Report (PDF)

```
in June 2011, a riot erupted in Vancouver during the Stanley Cup finals.
The trouble began when it seemed the home team wouldn't win the championship. Disappo
inted fans turned to rioting and in five hours of rioting, members of the crowd commi
tted multiple criminal offenses, including arson, break and enter,theft, mischief, an
d assault.
Approximately 55,000 people watched the game at a public viewing area on West Georgia
Street, with an additional 100,000 people      in the streets surrounding these viewi
ng areas.

Here's a breakdown of the crime types that occurred on that day....
crime_type_category
Mischief                    372
Break and Enter             187
Theft                       111
Offence Against a Person      7
Vehicle Collision             2
Name: TYPE, dtype: int64


... and the neighborhoods most affected.
NEIGHBOURHOOD
Central Business District    580
Strathcona                    20
Mount Pleasant                14
West End                      13
Fairview                       9
Dunbar-Southlands              6
Grandview-Woodland             4
Kitsilano                      4
Renfrew-Collingwood            4
Riley Park                     4
Hastings-Sunrise               3
Kensington-Cedar Cottage       3
Victoria-Fraserview            3
Sunset                         3
Marpole                        2
South Cambie                   2
Arbutus Ridge                  1
Oakridge                       1
Killarney                      1
Kerrisdale                     1
West Point Grey                1
Name: TYPE, dtype: int64
```

Now let's move forward and see which days seems to be more dangerous/safe! will consider June 15, 2011, as an anomaly and remove it from the analysis.

In [15]:
```
d =(crime[crime.date != '2011-06-15'].
    groupby(['MONTH','DAY','date'])['TYPE'].
    count().
    rename('no_of_crime_per_day').
    to_frame().
    reset_index().
    groupby(['MONTH','DAY'])['no_of_crime_per_day'].
    mean().
    rename('no_of_crime_per_day').
    to_frame().
```

```
    reset_index()
)

plt.figure(figsize=(13,5))
sns.heatmap(data  = d.pivot("MONTH" , "DAY" , "no_of_crime_per_day")  ,
            cmap='Blues',#'YlGnBu',
            linecolor='grey',
            linewidths=0.1,
            cbar=True,
            annot=True ,
            fmt=".0f" ,
            annot_kws={"fontsize": 8},
            yticklabels=month_order
)
ax.set_title('Number of Crime per Month and DAYXXX ', fontsize=10)
plt.xlabel('')
plt.ylabel('')
plt.yticks(rotation=0)
plt.show()
print ("Some days stand out in terms of safety and danger:"
       "\n  -Christmas Day, with only 71 crimes, which is 36% below the average of 112
       "\n  -September 1st (Labour Day), recording 134 crimes, Halloween on October 31
       "\n  with 130 crimes, are days with high crime rates. These figures are 18% abo
       "\n  -Summer seems to be more busy specially the first two weeks of the month."
       "\n  -In many months, the first half, tends to experience a higher crime rate c
       "\n   This trend might be influenced by factors like payday cycles, social even
       "\n   require further investigation to pinpoint."
       "\n  -Across all months, the first day consistently registers crime rates rangi
```

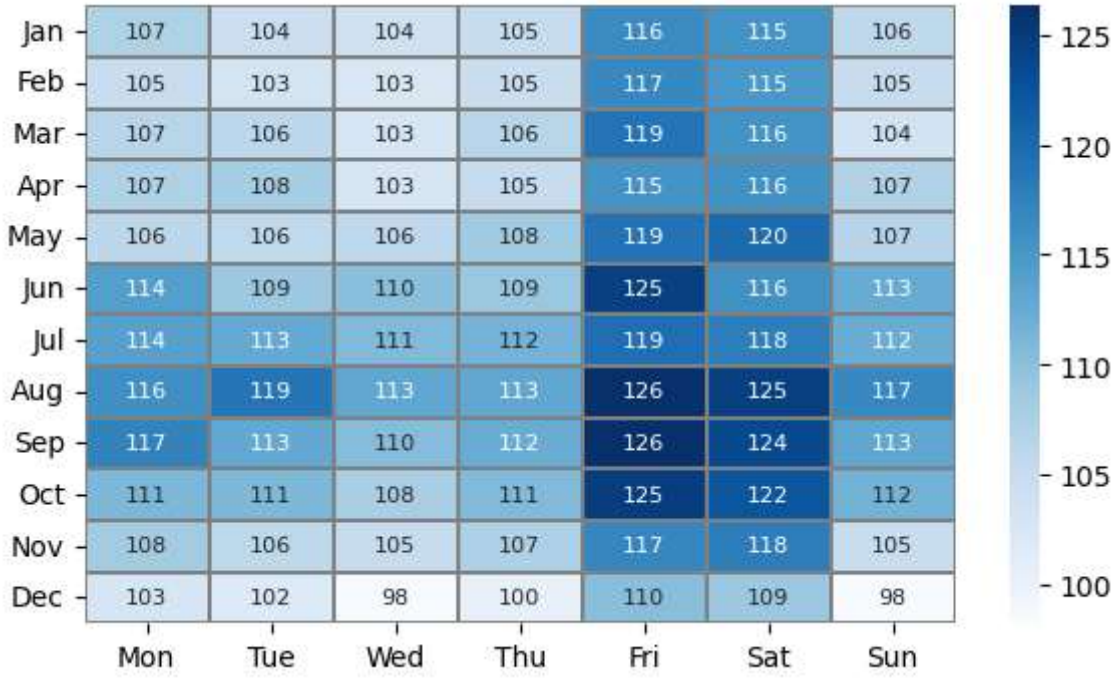


```
Some days stand out in terms of safety and danger:
  -Christmas Day, with only 71 crimes, which is 36% below the average of 112, seems t
o be the safest day in Vancouver.
  -September 1st (Labour Day), recording 134 crimes, Halloween on October 31st with 1
31 crimes, and New Year's Day
  with 130 crimes, are days with high crime rates. These figures are 18% above the av
erage.
  -Summer seems to be more busy specially the first two weeks of the month.
  -In many months, the first half, tends to experience a higher crime rate compared t
o the latter half.
   This trend might be influenced by factors like payday cycles, social events, or ot
her external influences, which would
   require further investigation to pinpoint.
  -Across all months, the first day consistently registers crime rates ranging from 1
20 to 134 incidents, which is above the average.
```

3-5-3- Is there an evident monthly crime pattern by day of the week?

In [16]:

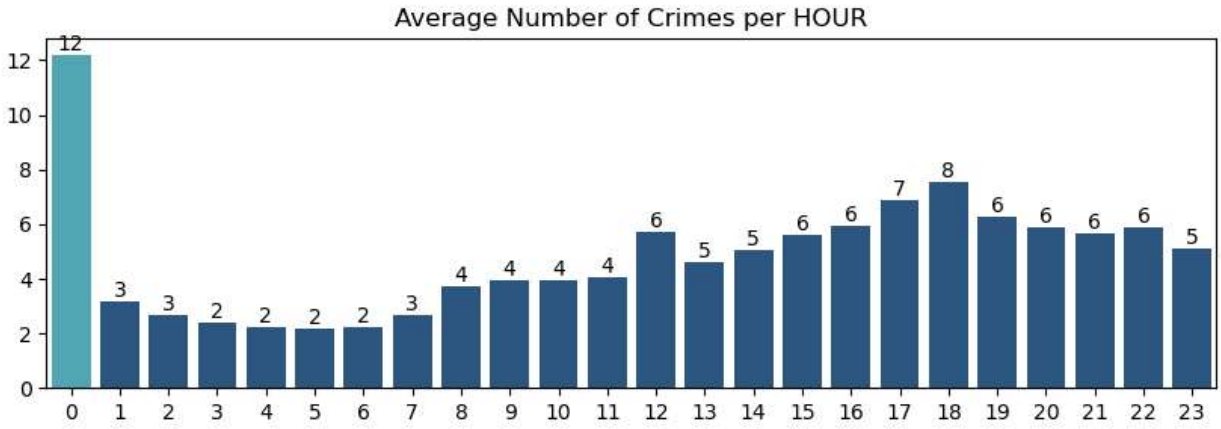
```
weekday_order = ['Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun']
d =(crime[crime.date != '2011-06-15'].
    groupby(['MONTH','weekday','date'])['TYPE'].
    count().
    rename('no_of_crime_per_weekday').
    to_frame().
    reset_index().
    groupby(['MONTH','weekday'])['no_of_crime_per_weekday'].
    mean().
    rename('no_of_crime_per_weekday').
    to_frame().
    reset_index()
)

plt.figure(figsize=(7,4))
sns.heatmap(data  = d.pivot("MONTH" , "weekday" , "no_of_crime_per_weekday")  ,
            cmap='Blues' , #'YlGnBu',
            linecolor='grey',
            linewidths=0.1,
            cbar=True,
            annot=True ,
            fmt=".0f" ,
            annot_kws={"fontsize": 8},
            yticklabels=month_order,
            xticklabels=weekday_order
)
ax.set_title('Number of Crime per Month and Weekday ', fontsize=10)
plt.xlabel('')
plt.ylabel('')
plt.yticks(rotation=0)
plt.show()
print("Crime rates peak during Fridays and Saturdays across all months, with both days
numbers. In contrast, Wednesdays and Thursdays tend to have the lowest crime rates. ")
```



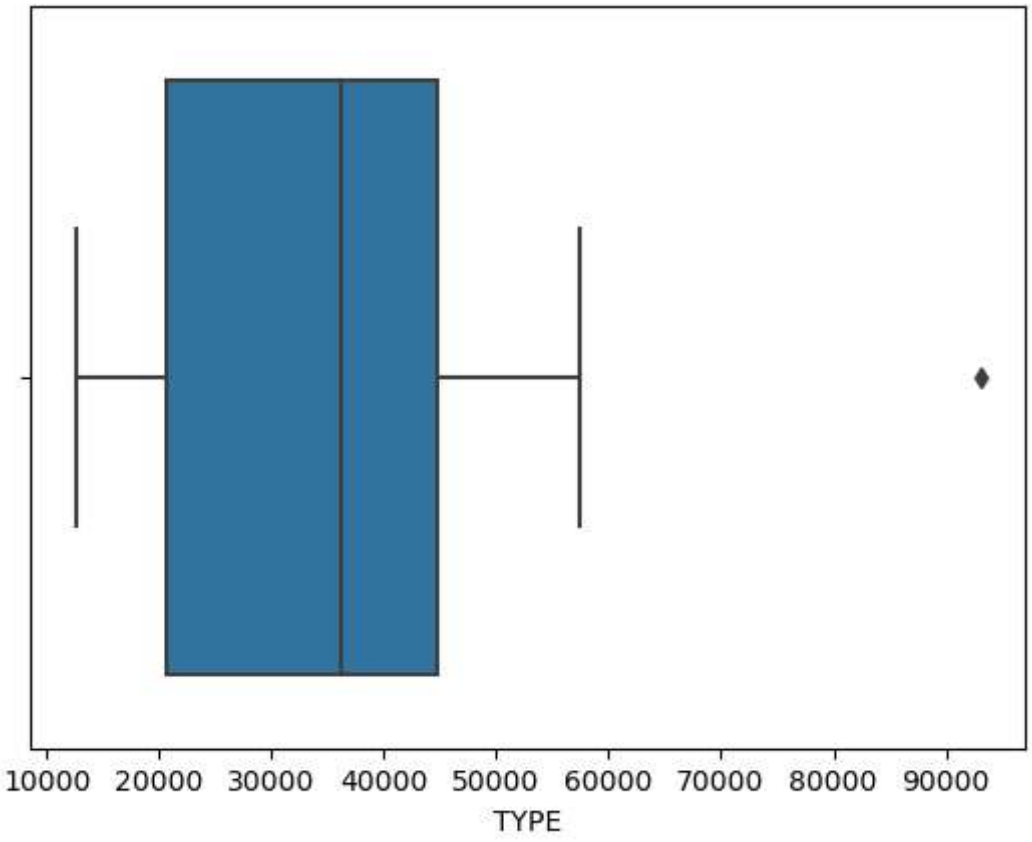
Crime rates peak during Fridays and Saturdays across all months, with both days consi stently seeing higher numbers. In contrast, Wednesdays and Thursdays tend to have the lowest crime rates.

3-5-4- What are the peak and off-peak crime periods during the week, and how do they vary by day?

In [17]:
```
data = (crime[crime.date != '2011-06-15'].
        groupby(['date','HOUR'])['TYPE'].
        count().
        rename('crime_per_hour').
        reset_index().
        groupby('HOUR')['crime_per_hour'].
        mean().
        rename('crime_per_hour').
        to_frame().
        reset_index().
        sort_values(by='HOUR',ascending = True)
)
max_value = data.crime_per_hour.max()
colors = ['#3FB4C4' if v == max_value else "#1E558D" for v in data['crime_per_hour']]
plt.figure(figsize=(10,3))
sns.barplot(data = data  , x='HOUR' , y='crime_per_hour' , palette =colors )
ax = plt.gca()
ax.set_title("Average Number of Crimes per HOUR")
plt.xlabel('')
plt.ylabel('')

ax.bar_label(ax.containers[0],fmt="%.f")
plt.show()
sns.boxplot(crime.groupby('HOUR')['TYPE'].count())

print("At midnight (0 hour), there's a notable spike in the average number of crimes.
      "\nThis surge, identified as an outlier in our boxplot analysis and could possib
      "\nwhere incidents around midnight are systematically recorded precisely at 12."
       "\nTo provide a clearer analysis and avoid potential discrepancies related to h
```



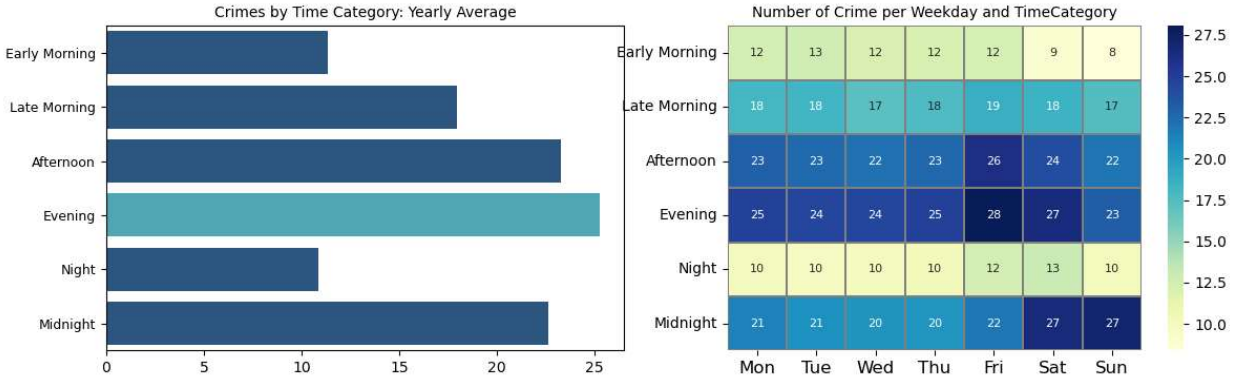
At midnight (0 hour), there's a notable spike in the average number of crimes.
This surge, identified as an outlier in our boxplot analysis and could possibly be du
e to data entry errors,
where incidents around midnight are systematically recorded precisely at 12.
To provide a clearer analysis and avoid potential discrepancies related to hourly dat
a, I've decided to use time categories instead.



Shifting the Lense: Analysis by Time Category

In [18]:
```
time_category_order = ['Early Morning', 'Late Morning', 'Afternoon', 'Evening','Night'
d1 =(crime[crime.date != '2011-06-15'].
    groupby(['weekday','time_category','date'])['TYPE'].
    count().
    rename('no_of_crime_per_timecategory').
    to_frame().
    reset_index().
    groupby(['weekday','time_category'])['no_of_crime_per_timecategory'].
    mean().
```

```
    rename('no_of_crime_per_timecategory').
    to_frame().
    reset_index()
)

d2 = ( crime.
         groupby(['date','time_category'])['TYPE'].
         count().
         rename('no_of_crime').
         to_frame().
         groupby('time_category')['no_of_crime'].
         mean().
         rename('avg_no_of_crime').
         reset_index()
     )

fig ,ax = plt.subplots(1,2 , figsize =(14,4))
#***************
max_value = d2.avg_no_of_crime.max()
colors = ['#3FB4C4' if v == max_value else "#1E558D" for v in d2['avg_no_of_crime']]
sns.barplot(data = d2 ,
            y='time_category' ,
            x ='avg_no_of_crime' ,
            palette = colors,
            ax=ax[0] )

ax[0].set_yticklabels(time_category_order)
ax[0].set_title("Crimes by Time Category: Yearly Average",fontsize=10)
ax[0].set_xlabel('')
ax[0].set_ylabel('')
ax[0].tick_params(axis='y', labelsize=9)
sns.heatmap(data  = d1.pivot("time_category" , "weekday" , "no_of_crime_per_timecategc
            cmap='YlGnBu',
            linecolor='grey',
            linewidths=0.1,
            cbar=True,
            annot=True ,
            fmt=".0f" ,
            annot_kws={"fontsize": 8},
            yticklabels=time_category_order,
            xticklabels=weekday_order,
            ax=ax[1]
)
ax[1].set_xlabel('')
ax[1].set_ylabel('')
ax[1].tick_params(axis='x', labelsize=12)
ax[1].set_title("Number of Crime per Weekday and TimeCategory",fontsize=10)
#plt.title('Number of Crime per Weekday and TimeCategory', fontsize=10)

plt.yticks(rotation=0)
plt.show()
print ("Crimes mostly occur in the evenings. On Saturdays and Sundays, crimes tend to
```




```
Crimes mostly occur in the evenings. On Saturdays and Sundays, crimes tend to occur a
t midnight,
while Friday afternoons also see a high number of crimes.
The night and early morning hours have the lowest number of crimes.
```

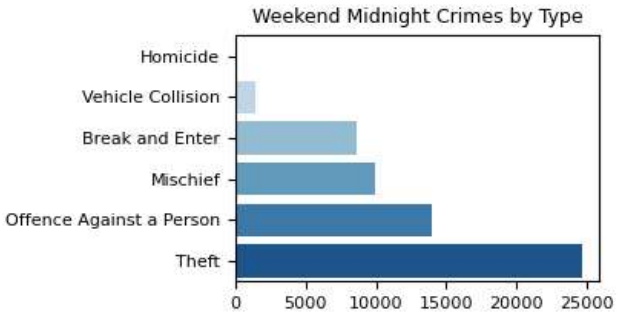
Let's see what type of crime occurs at the weekend's midnights the most?

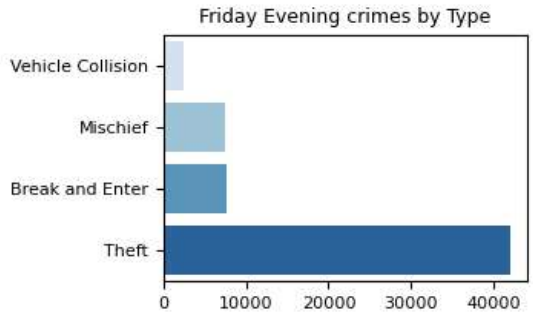
And take a look at Friday's evenings. what typically happens during Friday evenings?

In [19]:

```
    fig, axes = plt.subplots (1, 2, figsize=(8, 2) )
   # plt.figure(figsize=(6, 4))
    plt.subplots_adjust(left=0.1,
                bottom=0.1,
                    right=0.9,
                    top=0.9,
                    wspace=0.7,
                    hspace=0.4)
    sns.barplot(ax=axes[0],  data = crime[(crime.time_category == '6-Midnight') & ((cr
                        .groupby('crime_type_category')['TYPE']
                        .count()
                        .rename('number of crimes')
                        .reset_index()
                        .sort_values(by='number of crimes', ascending = True)
          ,  y='crime_type_category',x='number of crimes',palette="Blues" )


    sns.barplot(ax=axes[1],  data = crime[((crime.time_category == '4-Evening') | (cri
                        .groupby('crime_type_category')['TYPE']
                        .count()
                        .rename('number of crimes')
                        .reset_index()
                        .sort_values(by='number of crimes', ascending = True)
          ,  y='crime_type_category',x='number of crimes',palette="Blues"    )




axes[0].set_title("Weekend Midnight Crimes by Type" , fontsize = 9)
axes[1].set_title("Friday Evening crimes by Type", fontsize = 9)
axes[0].set_xlabel('')
axes[0].set_ylabel('')
axes[1].set_xlabel('')
axes[1].set_ylabel('')
axes[0].tick_params(axis='both', labelsize=8)
axes[1].tick_params(axis='both', labelsize=8)
#plt.tight_layout(w_pad=5.0)
```

```
plt.show()
print ("The most frequently occurring crime during weekend midnight hours is theft and
print("This trend may be attributed to an increase in nightlife activities, a rise in
print ("\nOn Friday afternoon and evenings, theft is the most commonly occurring crime
```






```
The most frequently occurring crime during weekend midnight hours is theft and offens
e against a person.
This trend may be attributed to an increase in nightlife activities, a rise in alcoho
l and drug consumption,
as well as reduced police presence. Further investigation is needed to better underst
and the underlying causes

On Friday afternoon and evenings, theft is the most commonly occurring crime.
```

3-6- Delving into Crime Types:

- 3-6-1-What is the top crime type for each year?

In [20]:

```
display(
            pd.DataFrame
            (
            crime
                .groupby(['YEAR','TYPE'])['TYPE']
                .count()
                .sort_values(ascending=False)
                .groupby('YEAR')
                .head(1)
                .rename('Number of crimes')
                .to_frame()
                .reset_index()
            ).sort_values('YEAR' , ascending=False)
        )
print( """\n >> The crime types Theft from Vehicle and Other Theft consistently exhibi
    """ )
```

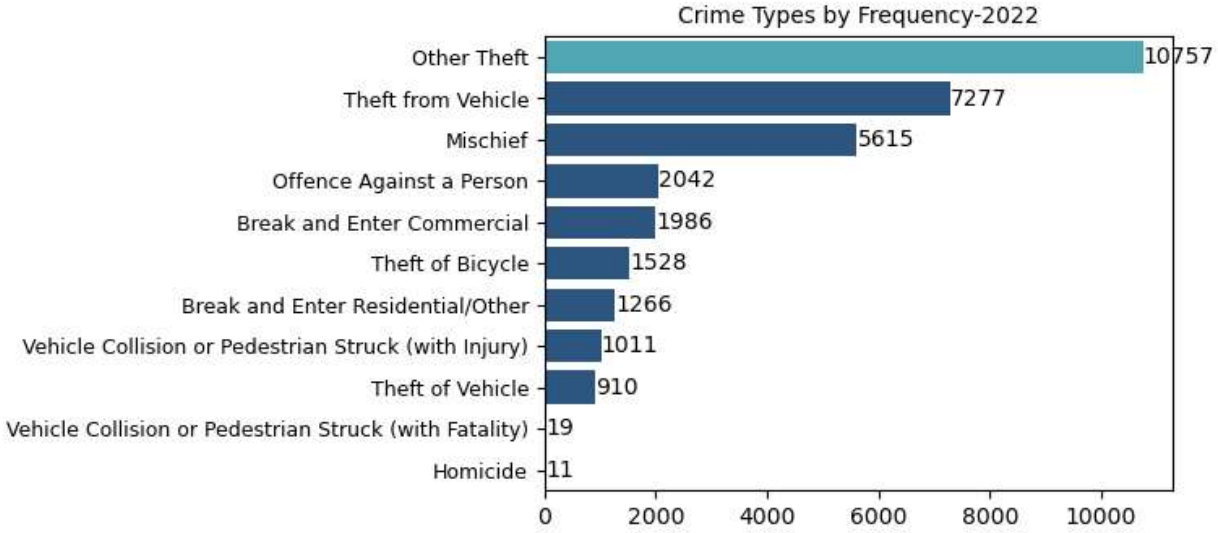
	YEAR	TYPE	Number of crimes
9	2023	Other Theft	11844
12	2022	Other Theft	10757
20	2021	Other Theft	8583
14	2020	Theft from Vehicle	10430
2	2019	Theft from Vehicle	17005
4	2018	Theft from Vehicle	14996
6	2017	Theft from Vehicle	12957
7	2016	Theft from Vehicle	12815
13	2015	Other Theft	10598
11	2014	Other Theft	10794
15	2013	Other Theft	10264
17	2012	Other Theft	9960
18	2011	Other Theft	9388
19	2010	Other Theft	9366
16	2009	Theft from Vehicle	10007
10	2008	Theft from Vehicle	11295
8	2007	Theft from Vehicle	12227
5	2006	Theft from Vehicle	14730
3	2005	Theft from Vehicle	16552
0	2004	Theft from Vehicle	18202
1	2003	Theft from Vehicle	17743

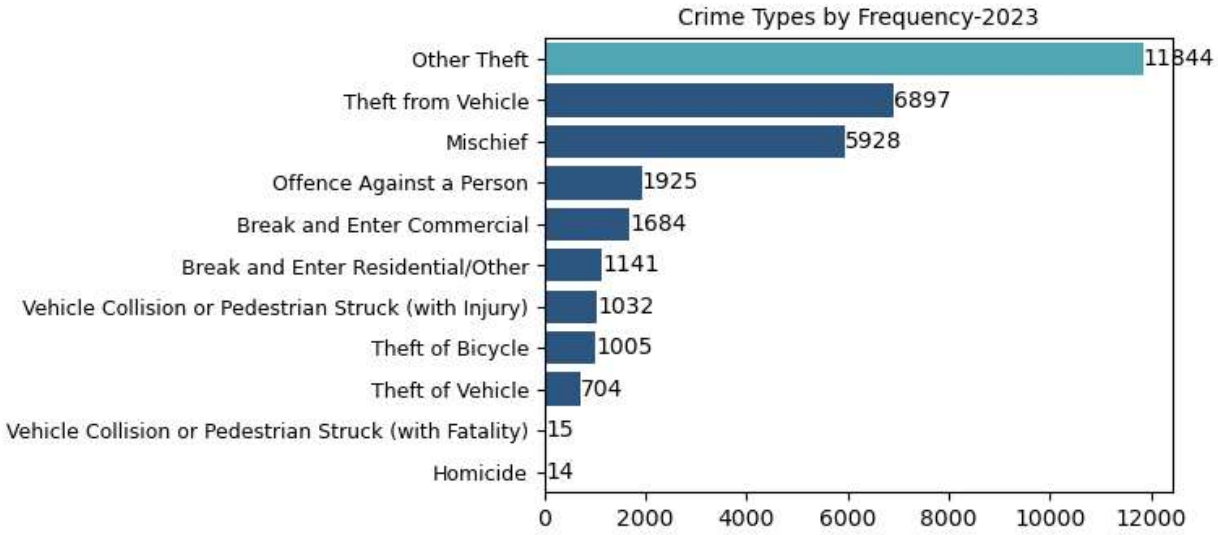
```
 >> The crime types Theft from Vehicle and Other Theft consistently exhibit the highe
st number of incidents across the years.
```

- 3-6-2-What are the top 3 crime types this year and over the years?

In [21]:
```
### ************** maybe remove
crime_this_year = ( crime[crime.YEAR == last_updatetime.year].
                        groupby('TYPE')['TYPE'].
                        count().
                        rename('no_of_crimes').
                        to_frame().
                        reset_index().
                        sort_values(by = 'no_of_crimes' , ascending =False)
             )
crime_previous_year = ( crime[crime.YEAR == last_updatetime.year-1].
                        groupby('TYPE')['TYPE'].
                        count().
                        rename('no_of_crimes').
                        to_frame().
                        reset_index().
```

```
                        sort_values(by = 'no_of_crimes' , ascending =False)
             )
fig,ax=plt.subplots(2,1 , figsize=(5,8))
max_value = crime_previous_year.no_of_crimes.max()
colors = ['#3FB4C4' if v == max_value else "#1E558D" for v in crime_previous_year['no_
sns.barplot(data = crime_previous_year ,
            x='no_of_crimes' ,
            y='TYPE',
            ax=ax[0],
            palette = colors)
max_value = crime_this_year.no_of_crimes.max()
colors = ['#3FB4C4' if v == max_value else "#1E558D" for v in crime_this_year['no_of_c
sns.barplot(data = crime_this_year ,
            x='no_of_crimes' ,
            y='TYPE',
            ax=ax[1],
            palette = colors)
ax[0].set_title("Crime Types by Frequency-2022",fontsize=10)
ax[0].set_xlabel('')
ax[0].set_ylabel('')
ax[0].tick_params(axis='y', labelsize=9)
ax[1].set_title("Crime Types by Frequency-2023",fontsize=10)
ax[1].set_xlabel('')
ax[1].set_ylabel('')
ax[0].bar_label(ax[0].containers[0],fmt="%.f")
ax[1].bar_label(ax[1].containers[0],fmt="%.f")
ax[1].tick_params(axis='y', labelsize=9)
plt.show()
```

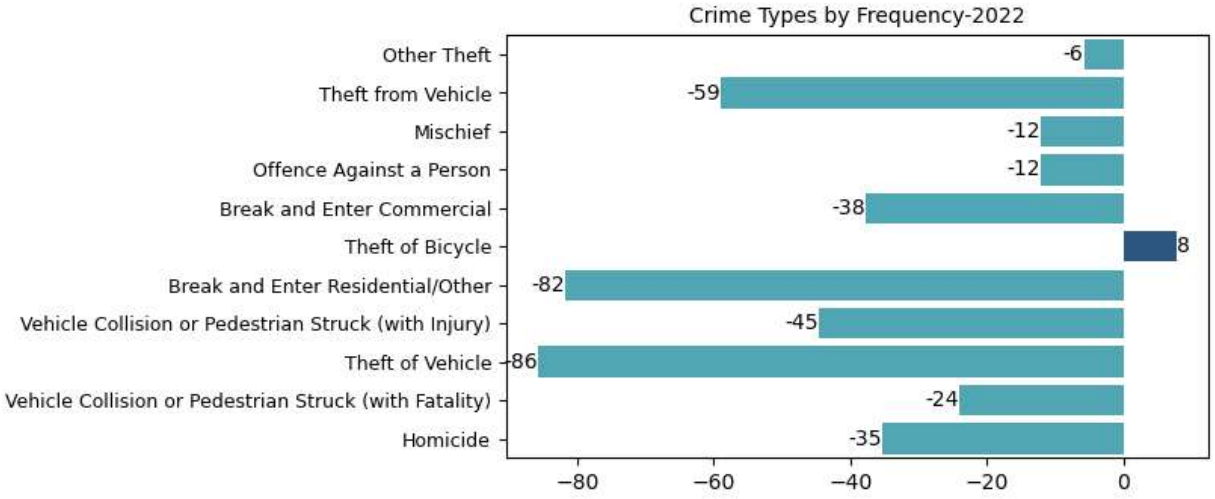


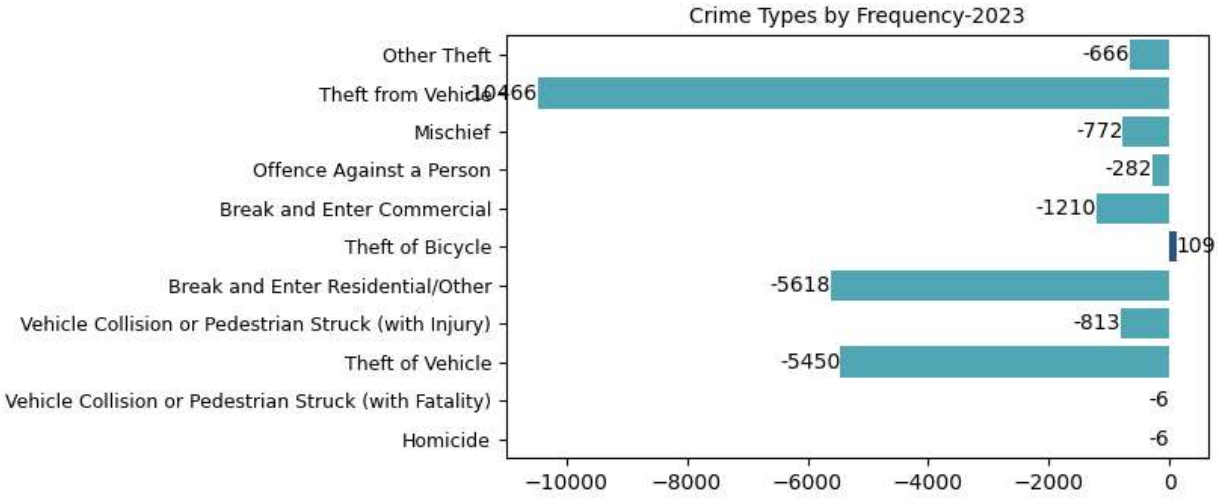


- 3-6-3-Now let's see which crime type has the most percentage changes over the years :

In [22]:
```
crime_previous_year.rename(columns ={'no_of_crimes':'no_of_crimes_pre_year'}, inplace
crime_first_year = ( crime[crime.YEAR == crime.YEAR.min()].
                        groupby('TYPE')['TYPE'].
                        count().
                        rename('no_of_crimes_first_year').
                        to_frame().
                        reset_index().
                        sort_values(by = 'no_of_crimes_first_year' , ascending =False)
            )
crime_merged = crime_previous_year.merge(crime_first_year,how='inner', on='TYPE')
crime_merged['percent_changed'] = (crime_merged.no_of_crimes_pre_year - crime_merged.r
crime_merged['absolute_changed'] = (crime_merged.no_of_crimes_pre_year - crime_merged.
###crime_merged
#***
fig,ax=plt.subplots(2,1 , figsize=(6,8))
#max_value = crime_merged.percent_changed.max()
colors = ['#3FB4C4' if v < 0  else "#1E558D" for v in crime_merged['percent_changed']]
sns.barplot(data = crime_merged ,
           x='percent_changed' ,
           y='TYPE',
           ax=ax[0],
```

```
            palette = colors)
max_value = crime_merged.absolute_changed.max()
colors = ['#3FB4C4' if v < 0  else "#1E558D" for v in crime_merged['absolute_changed']
sns.barplot(data = crime_merged ,
            x='absolute_changed' ,
            y='TYPE',
            ax=ax[1],
            palette = colors)
ax[0].set_title("Crime Types by Frequency-2022",fontsize=10)
ax[0].set_xlabel('')
ax[0].set_ylabel('')
ax[0].tick_params(axis='y', labelsize=9)
ax[1].set_title("Crime Types by Frequency-2023",fontsize=10)
ax[1].set_xlabel('')
ax[1].set_ylabel('')
ax[1].tick_params(axis='y', labelsize=9)
ax[0].bar_label(ax[0].containers[0],fmt="%.f")
ax[1].bar_label(ax[1].containers[0],fmt="%.f")
plt.show()



crime_merged = crime_merged.iloc[crime_merged['percent_changed'].abs().argsort()[::-1]
display(crime_merged)
print(" Most crime types have seen a reduction, with 'Theft of Vehicle' and 'Break and
print("Over the years, advances in technology have made it more challenging for crimin
vehicles. Modern vehicles, for instance, come equipped with advanced security systems,
smart key access, making them harder to steal. Similarly, residential areas might have
surveillance cameras, and stronger locks.\
Community Policing and Neighborhood Watch Programs: An active community policing initi
neighborhood watch programs can significantly deter crimes in residential areas. \
When residents actively report suspicious activities and collaborate with the police,
criminal activities like break-ins.")
print("\n\n It's interesting to note that while severe crimes such as homicide have se
      "\nand vehicle collisions with fatalities have decreased by 24% (from 25 cases t
      "\n'Theft of Bicycle' have seen an increase of 7.68% (from 1419 cases to 1528 ca
      "\nfor further studies to understand the varying dynamics affecting different cr
```


Crime Types by Frequency-2022
Other Theft
Theft from Vehicle
Mischief
Offence Against a Person
Break and Enter Commercial
Theft of Bicycle
Break and Enter Residential/Other
Vehicle Collision or Pedestrian Struck (with Injury)
Theft of Vehicle
Vehicle Collision or Pedestrian Struck (with Fatality)
Homicide
-6
-59
-12
-12
-38
8
-82
-45
-86
-24
-35
−80
−60
−40
−20
0


Crime Types by Frequency-2023
Other Theft
Theft from Vehicle
Mischief
Offence Against a Person
Break and Enter Commercial
Theft of Bicycle
Break and Enter Residential/Other
Vehicle Collision or Pedestrian Struck (with Injury)
Theft of Vehicle
Vehicle Collision or Pedestrian Struck (with Fatality)
Homicide
-666
-10466
-772
-282
-1210
109
-5618
-813
-5450
-6
-6
−10000
−8000
−6000
−4000
−2000
0

	TYPE	no_of_crimes_pre_year	no_of_crimes_first_year	percent_changed	absolute_change
8	Theft of Vehicle	910	6360	-85.691824	-545
6	Break and Enter Residential/Other	1266	6884	-81.609529	-561
1	Theft from Vehicle	7277	17743	-58.986643	-1046
7	Vehicle Collision or Pedestrian Struck (with I...	1011	1824	-44.572368	-81
4	Break and Enter Commercial	1986	3196	-37.859825	-121
10	Homicide	11	17	-35.294118	-
9	Vehicle Collision or Pedestrian Struck (with F...	19	25	-24.000000	-
3	Offence Against a Person	2042	2324	-12.134251	-28
2	Mischief	5615	6387	-12.087052	-77
5	Theft of Bicycle	1528	1419	7.681466	10
0	Other Theft	10757	11423	-5.830342	-66

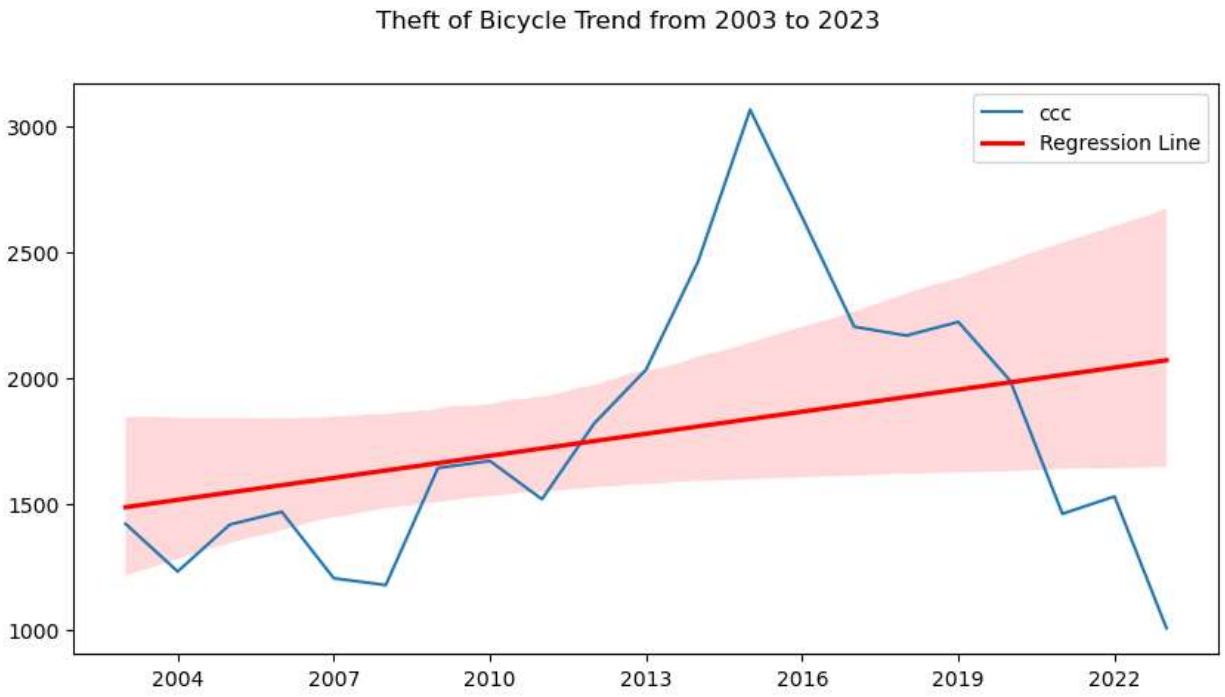
Most crime types have seen a reduction, with 'Theft of Vehicle' and 'Break and Enter Residential/Other' witnessing the most significant declines. Only 'Theft of Bicycle' registered an increase in the number of cases.
Over the years, advances in technology have made it more challenging for criminals to break into houses and steal vehicles. Modern vehicles, for instance, come equipped with advanced security systems, GPS tracking, immobilizers, and smart key access, making them harder to steal. Similarly, residential areas might have adopted better security systems,surveillance cameras, and stronger locks.Community Policing and Neighborhood Watch Programs: An active community policing initiative and the establishment ofneighborhood watch programs can significantly deter crimes in residential areas. When residents actively report suspicious activities and collaborate with the police, it can lead to a reduction in criminal activities like break-ins.

It's interesting to note that while severe crimes such as homicide have seen a reduction of 35.29% (from 17 cases to 11 cases)
and vehicle collisions with fatalities have decreased by 24% (from 25 cases to 19 cases), some less severe crimes like
'Theft of Bicycle' have seen an increase of 7.68% (from 1419 cases to 1528 cases). This contrast can be a point of interest
for further studies to understand the varying dynamics affecting different crime cate

what's the trend for this type of the crime?

In [23]:
```
TheftOfBicycle_df = crime[ (crime.TYPE == 'Theft of Bicycle')  ]
data = TheftOfBicycle_df.groupby('YEAR')['TYPE'].count().rename('NumberOfCrimes').rese
plt.figure(figsize=(10,5))
sns.lineplot( data = data , x='YEAR' , y = 'NumberOfCrimes', label='ccc')
ax = plt.gca()
```

```
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
# Perform linear regression
sns.regplot(data=data, x='YEAR', y='NumberOfCrimes', scatter=False, color='red', label
plt.suptitle('Theft of Bicycle Trend from 2003 to 2023')
#ax.set_title('2023 Excluded')
#plt.title('Crime Trend Over the Years- 2023 not included')
plt.xlabel('')
plt.ylabel('')
plt.legend(loc='best')
plt.show()
print("In summary, despite some years of decline and fluctuation, the long-term view f
positive trend in bicycle thefts, with the highest numbers recorded around 2015. Follo
decline. Let's delve deeper to see what could be the reasons behind this trend.")
TheftOfBicycle_2015 = data[data.YEAR == 2015]['NumberOfCrimes'].iloc[0]
TheftOfBicycle_2022 = data[data.YEAR == 2022]['NumberOfCrimes'].iloc[0]
#print(TheftOfBicycle_2015)
#print(TheftOfBicycle_2022)
ch =(TheftOfBicycle_2022-TheftOfBicycle_2015)/TheftOfBicycle_2015
print("Bicycle thefts in Vancouver dropped by about 50% from 3,064 cases in 2015 to 1,
I looked into this decrease and learned about Project 529, a bike registration program
This big drop shows that Project 529 might be really helping to stop bike thefts.Regis
print("\nNow let's examine which neighbourhoods were most affected in 2023?")
display( crime[ (crime.TYPE == 'Theft of Bicycle') & (crime.YEAR == 2023)].value_count
print("\nWhat times do these incidents occur most frequently?")
crime[ (crime.TYPE == 'Theft of Bicycle') & (crime.YEAR == 2023)].value_counts('time_c
```




```
In summary, despite some years of decline and fluctuation, the long-term view from 20
03 to 2023 indicates an overall positive trend in bicycle thefts, with the highest nu
mbers recorded around 2015. Following this peak, there has been adecline. Let's delve
deeper to see what could be the reasons behind this trend.
Bicycle thefts in Vancouver dropped by about 50% from 3,064 cases in 2015 to 1,528 in
2022.I looked into this decrease and learned about Project 529, a bike registration p
rogram that started in 2015.This big drop shows that Project 529 might be really help
ing to stop bike thefts.Registeratip app leads to significant decrease in bike theft

Now let's examine which neighbourhoods were most affected in 2023?
```

```
NEIGHBOURHOOD
Central Business District    232
Fairview                     144
West End                     126
Mount Pleasant               112
Kitsilano                     89
Grandview-Woodland            51
Strathcona                    49
Kensington-Cedar Cottage      39
Renfrew-Collingwood           27
Stanley Park                  20
Riley Park                    20
Marpole                       17
South Cambie                  13
West Point Grey               11
Hastings-Sunrise               9
Sunset                         9
Oakridge                       8
Killarney                      7
Arbutus Ridge                  7
Shaughnessy                    6
Kerrisdale                     4
Dunbar-Southlands              3
Victoria-Fraserview            2
dtype: int64
What times do these incidents occur most frequently?
```

Out[23]:

```
time_category
3-Afternoon        247
4-Evening          226
2-Late Morning     182
6-Midnight         152
1-Early Morning    115
5-Night             83
dtype: int64
```

In [194…

```
data
```

Out[194]:

	YEAR	NumberOfCrimes
0	2003	1419
1	2004	1230
2	2005	1416
3	2006	1467
4	2007	1203
5	2008	1176
6	2009	1641
7	2010	1669
8	2011	1517
9	2012	1817
10	2013	2032
11	2014	2461
12	2015	3064
13	2016	2637
14	2017	2202
15	2018	2167
16	2019	2221
17	2020	1987
18	2021	1460
19	2022	1528
20	2023	979

In [24]:

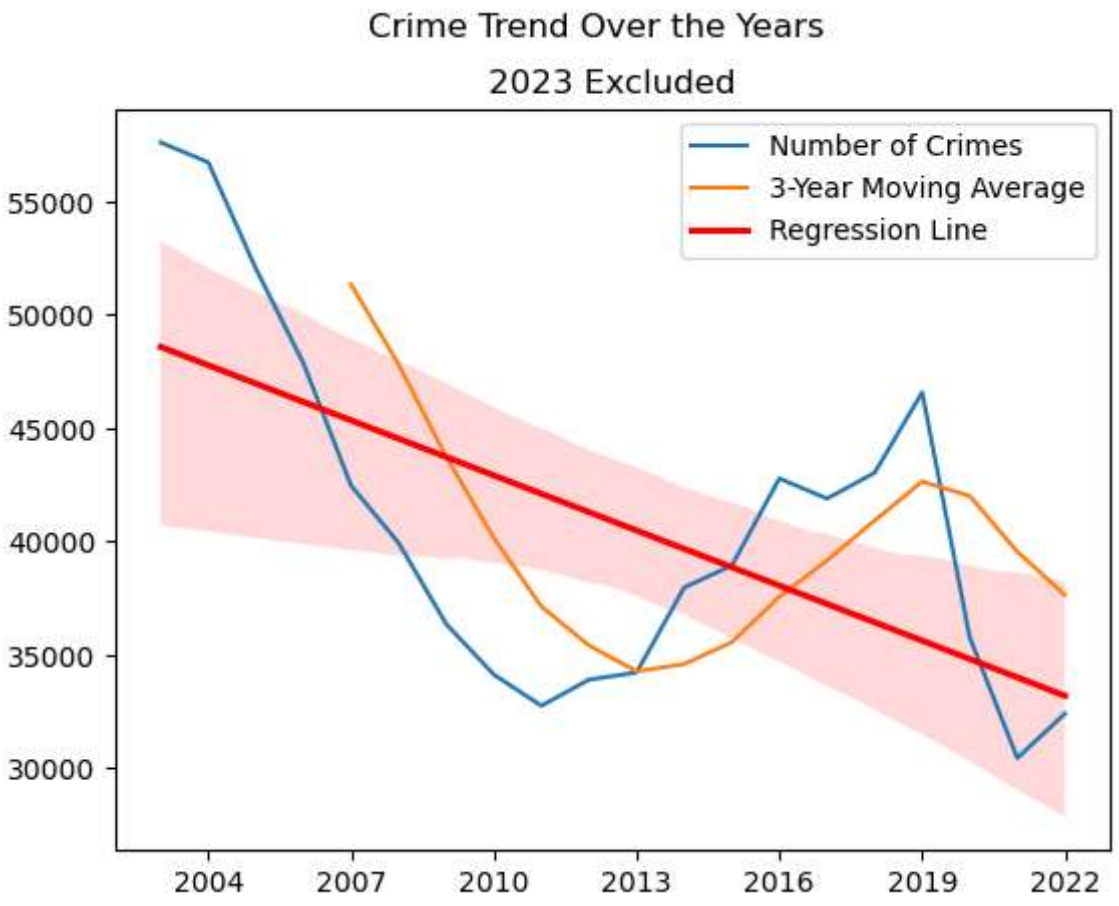
```
crime_2022_by_type = (
            crime[(crime.YEAR == last_updatetime.year - 1) & (crime.date <= last_updat
            groupby('TYPE')['TYPE'].
            count().
            rename('number_of_crimes_2022').
            reset_index()
)
crime_2023_by_type = (
            crime[(crime.YEAR == last_updatetime.year) ].
            groupby('TYPE')['TYPE'].
            count().
            rename('number_of_crimes_2023').
            reset_index()
)
crime_2022_2023= crime_2022_by_type.merge(crime_2023_by_type)
crime_2022_2023['percent_changed'] =( crime_2022_2023.number_of_crimes_2023 - crime_20
crime_2022_2023['absolute_changed'] =( crime_2022_2023.number_of_crimes_2023 - crime_2
display(crime_2022_2023.sort_values(by ='absolute_changed' , key=lambda x: x.abs(),  a
print("The comparison of crime types between 2022 and the same period in 2023 shows va
Other Theft and Mischief have notably increased, with 1,636 and 599 more incidents, re
```

```
Significant decreases are seen in Theft of Bicycle (-483), Break and Enter Commercial
Homicide incidents slightly increased, while Vehicle Collision or Pedestrian Struck ca
This data indicates a shift in crime patterns, with increases in some types of propert
```

	TYPE	number_of_crimes_2022	number_of_crimes_2023	percent_changed	absolute_chan
5	Other Theft	10208	11844	0.160266	1
3	Mischief	5329	5928	0.112404	
7	Theft of Bicycle	1488	1005	-0.324597	-
0	Break and Enter Commercial	1874	1684	-0.101387	-
8	Theft of Vehicle	874	704	-0.194508	-
10	Vehicle Collision or Pedestrian Struck (with I...	963	1032	0.071651	
1	Break and Enter Residential/Other	1180	1141	-0.033051	
6	Theft from Vehicle	6865	6897	0.004661	
4	Offence Against a Person	1936	1925	-0.005682	
9	Vehicle Collision or Pedestrian Struck (with F...	19	15	-0.210526	
2	Homicide	11	14	0.272727	

```
The comparison of crime types between 2022 and the same period in 2023 shows varied t
rends:Other Theft and Mischief have notably increased, with 1,636 and 599 more incide
nts, respectively.Significant decreases are seen in Theft of Bicycle (-483), Break an
d Enter Commercial (-190), and Theft of Vehicle (-170).Homicide incidents slightly in
creased, while Vehicle Collision or Pedestrian Struck categories show minor changes.T
his data indicates a shift in crime patterns, with increases in some types of propert
y crimes and decreases in others.
```

In [38]:
```
#Adding 3-year-moving average
crime_grouped_by_year_all_before2023=crime_grouped_by_year_all_before2023.copy()
crime_grouped_by_year_all_before2023['moving_avg'] = crime_grouped_by_year_all_before2
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'number_of_
sns.lineplot( data = crime_grouped_by_year_all_before2023 , x='YEAR' , y = 'moving_avg
ax = plt.gca()
ax.xaxis.set_major_locator(MaxNLocator(integer=True))
# Perform linear regression
sns.regplot(data=crime_grouped_by_year_all_before2023, x='YEAR', y='number_of_crimes',
plt.suptitle('Crime Trend Over the Years')
ax.set_title('2023 Excluded')
#plt.title('Crime Trend Over the Years- 2023 not included')
plt.xlabel('')
plt.ylabel('')
plt.legend(loc='best')
plt.show()
```

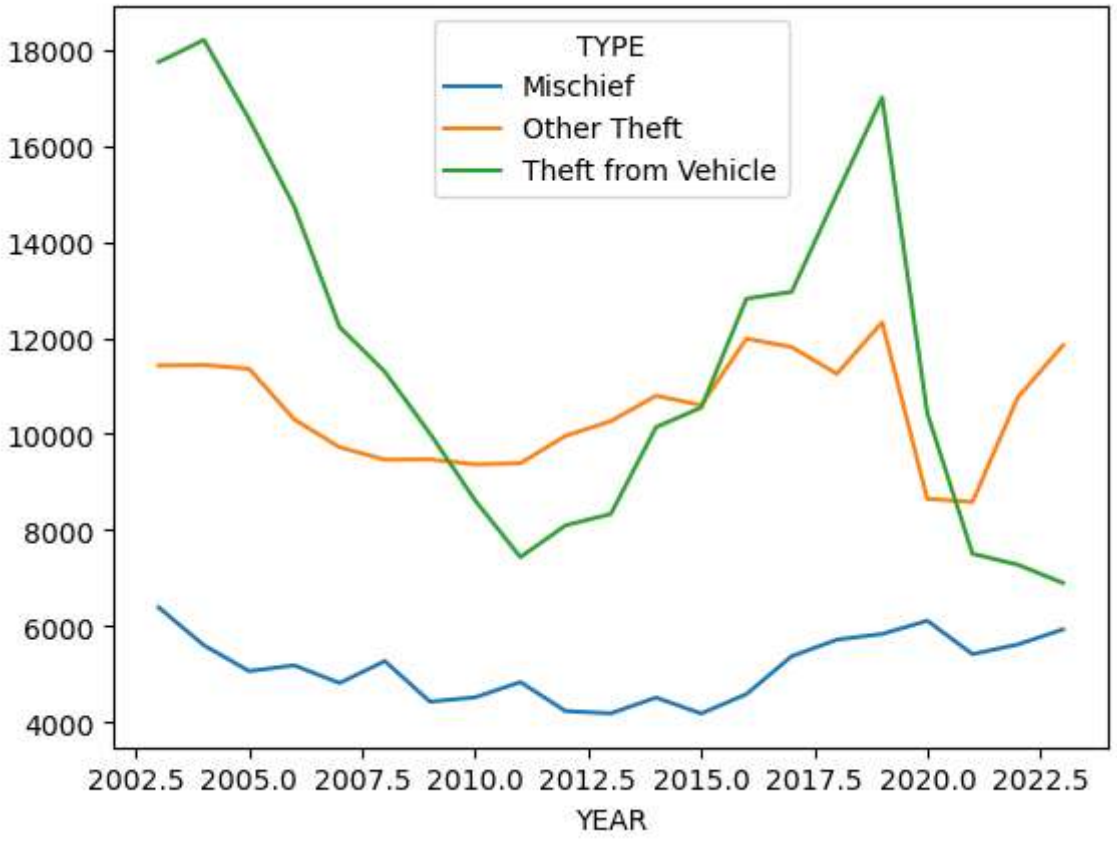


In [40]:
```
### *************** maybe remove
# To show all the items without scrolling
#display(HTML("<style>div.output_scroll { height: auto; }</style>"))
#crime_grouped_by_type = crime.groupby(['YEAR','TYPE'])['TYPE'].count().rename('no_of_
#crime_grouped_by_type['crimes_growth_rate'] = (
#     ( crime_grouped_by_type['no_of_crimes'] - crime_grouped_by_type['no_of_crimes'].
#       /       crime_grouped_by_type['no_of_crimes'].shift(1).fillna(1)
#     ) * 100
#crime_grouped_by_type[(crime_grouped_by_type.YEAR == 2022)| (crime_grouped_by_type.YE
##fig, axes = plt.subplots (len(crime_grouped_by_type.TYPE.unique()), 1, figsize=(8,10
#i=0
#for crime_type in crime_grouped_by_type.TYPE.unique():
#    df = crime_grouped_by_type[crime_grouped_by_type.TYPE  == crime_type]
#    plt.figure(figsize=(8, 2))
#    sns.barplot(data = df , x = 'YEAR' , y ='crimes_growth_rate', color = '#1E558D')
#        # Set chart title, labels, etc.
#   # plt.title(f'Percentage Change Over Years for {crime_type}')
#    plt.ylabel('Percentage Change')
#    plt.xlabel('Year')
#    plt.xticks(rotation=45)

#    plt.show()
#    i+=1
```

3-6-4-Trend in Crime Rates for the Top 3 Crimes with the Highest Frequencies

In [25]:
```
( crime[crime.TYPE.isin(['Theft from Vehicle','Other Theft','Mischief'])]
     .groupby(['YEAR','TYPE'])['TYPE']
     .count()
```

```
    .reset_index(name='num_of_crimes')
    .pivot(index='YEAR',columns='TYPE',values='num_of_crimes')
    .plot
    .line()
);
```



3-7- How Do Crime Patterns Vary Across Different Neighborhoods?

-3-7-1-Dangerous Hotspots: Average Annual Crime Rates by Neighbourhood

In [26]:
```
crime_by_neighbourhood  = crime.groupby('NEIGHBOURHOOD')['TYPE'].count().reset_index(r
crime_by_neighbourhood
# Merge the crime counts with the neighborhood polygons
#merged_data = gdf_neighborhoods.merge(crime_count, on='NEIGHBOURHOOD', how='left')
```

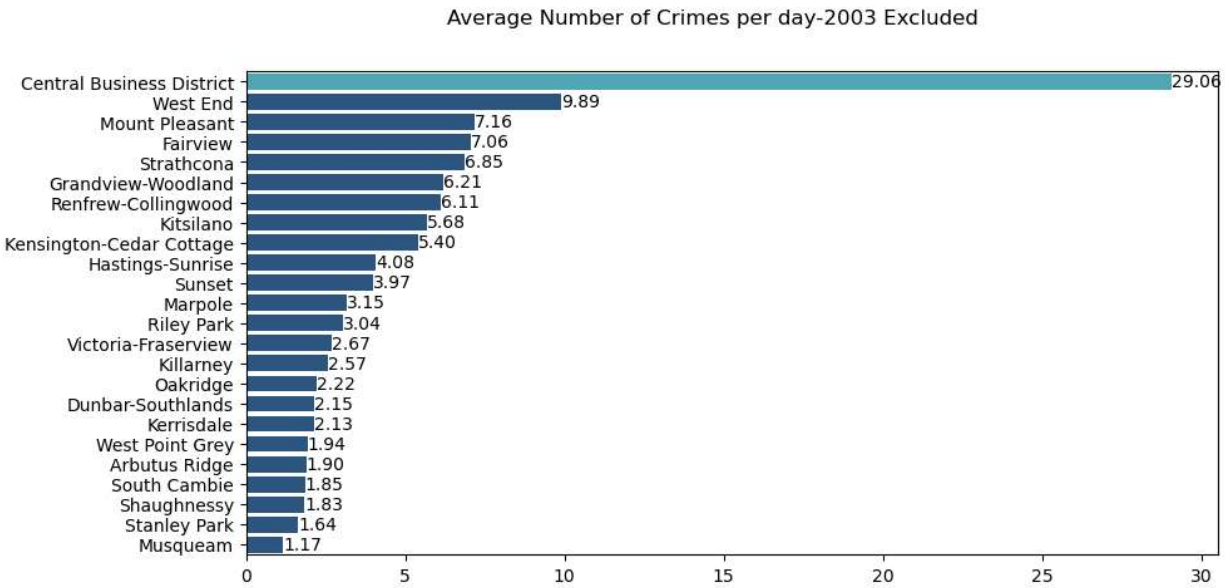
Out[26]:

	NEIGHBOURHOOD	crime_count
0	Arbutus Ridge	9397
1	Central Business District	222248
2	Dunbar-Southlands	11461
3	Fairview	53610
4	Grandview-Woodland	47143
5	Hastings-Sunrise	29884
6	Kensington-Cedar Cottage	40855
7	Kerrisdale	11530
8	Killarney	16434
9	Kitsilano	42719
10	Marpole	21869
11	Mount Pleasant	54530
12	Musqueam	972
13	Oakridge	13162
14	Renfrew-Collingwood	46402
15	Riley Park	20902
16	Shaughnessy	8661
17	South Cambie	8955
18	Stanley Park	5647
19	Strathcona	52268
20	Sunset	29009
21	Victoria-Fraserview	17396
22	West End	75524
23	West Point Grey	9568

In [27]:
```
display(HTML("<style>div.output_scroll { height: auto; }</style>"))
display(pd.DataFrame(
                crime.groupby(['YEAR' ,'NEIGHBOURHOOD'])['NEIGHBOURHOOD']
                        .count()
                        .sort_values(ascending=False)
                        .groupby('YEAR')
                        .head(1)
                        .rename('Number of crimes')
                        .to_frame()
                        .reset_index()
            ).sort_values('YEAR',ascending =False) )
### Let's see the average number of crimes per day in each neighbourhood
data = (crime.
                         groupby(['NEIGHBOURHOOD','date'])['TYPE'].
                         count().
```

```
                        rename('no_of_crimes').
                        to_frame().
                        groupby('NEIGHBOURHOOD')['no_of_crimes'].
                        mean().
                        rename('Average Number of Crimes').
                        to_frame().
                        reset_index().
                        sort_values(by = 'Average Number of Crimes' , ascending  = Fa

       )
plt.figure(figsize=(10,5))
max_value = data['Average Number of Crimes'].max()
colors = ['#3FB4C4' if v == max_value else '#1E558D' for v in data['Average Number of
sns.barplot ( data = data ,
             x ='Average Number of Crimes' ,
             y = 'NEIGHBOURHOOD',
             orient = 'h' ,
             palette = colors

            ) ;
ax.set_title("2003 Excluded")
plt.xlabel('')
plt.ylabel('')
plt.suptitle('Average Number of Crimes per day-2003 Excluded ')
ax = plt.gca()
ax.bar_label(ax.containers[0],fmt="%.2f")
plt.show()


print("""\nThe Central Business District consistently ranks as the top neighbourhood w
Given its status as a commercial and cultural hub, it's expected to witness a higher f
\nQuiet Neighborhoods: Arbutus Ridge, Musqueam, and Stanley Park have among the lowest
1 to 2 crimes per day. \
\nThese neighborhoods might be more residential or recreational, leading to fewer inci
\nActive Residential Areas: Neighborhoods like Fairview, Grandview-Woodland, Mount Ple
ranging from 6 to 7 crimes per day, indicating these could be densely populated or hav
\nWest End's Distinction: The West End stands out with an average of nearly 10 crimes
vibrant mix of residential, commercial, and recreational spots, it's a hotspot of acti
\nStable Middle Ground: Neighborhoods such as Hastings-Sunrise, Kensington-Cedar Cotta
ground with averages between 3 to 6 crimes daily.""")
```

	YEAR	NEIGHBOURHOOD	Number of crimes
11	2023	Central Business District	10345
13	2022	Central Business District	9593
19	2021	Central Business District	8215
14	2020	Central Business District	9446
0	2019	Central Business District	15045
1	2018	Central Business District	13650
2	2017	Central Business District	12746
3	2016	Central Business District	12654
9	2015	Central Business District	10622
12	2014	Central Business District	9917
16	2013	Central Business District	8849
20	2012	Central Business District	8028
18	2011	Central Business District	8242
17	2010	Central Business District	8400
15	2009	Central Business District	9261
10	2008	Central Business District	10617
8	2007	Central Business District	10644
6	2006	Central Business District	11031
7	2005	Central Business District	10724
5	2004	Central Business District	11643
4	2003	Central Business District	12576



The Central Business District consistently ranks as the top neighbourhood with the hi
ghest number of reported crimes.Given its status as a commercial and cultural hub, i
t's expected to witness a higher frequency of crimes.
Quiet Neighborhoods: Arbutus Ridge, Musqueam, and Stanley Park have among the lowest
average daily crime rates, hovering around1 to 2 crimes per day.
These neighborhoods might be more residential or recreational, leading to fewer incid
ents.
Active Residential Areas: Neighborhoods like Fairview, Grandview-Woodland, Mount Plea
sant, and Renfrew-Collingwood see higher averagesranging from 6 to 7 crimes per day,
indicating these could be densely populated or have a mix of commercialand residentia
l zones.

West End's Distinction: The West End stands out with an average of nearly 10 crimes a
day, making it second only to the CBD.Given itsvibrant mix of residential, commercia
l, and recreational spots, it's a hotspot of activity.
Stable Middle Ground: Neighborhoods such as Hastings-Sunrise, Kensington-Cedar Cottag
e, Kitsilano, and Sunset maintain a steady middleground with averages between 3 to 6
crimes daily.

3-7-2-Which Crime Type Occurs Most Frequently in Central Business District?

In [28]:
```
display( crime[crime.NEIGHBOURHOOD == 'Central Business District' ].
    groupby(['crime_type_category','YEAR'])['TYPE'].
    count().
    rename('no_of_crimes').
    to_frame().
    groupby('crime_type_category')['no_of_crimes'].
    mean().
    rename('Average Number of Crimes per YEAR').
    to_frame().
    reset_index().
    sort_values(by = 'Average Number of Crimes per YEAR' , ascending = False)
)
print("In Vancouver's Central Business District, which experiences the highest crime r
an average of 4,341 crimes per year, followed by 'Mischief' and 'Break and Enter', whi
```

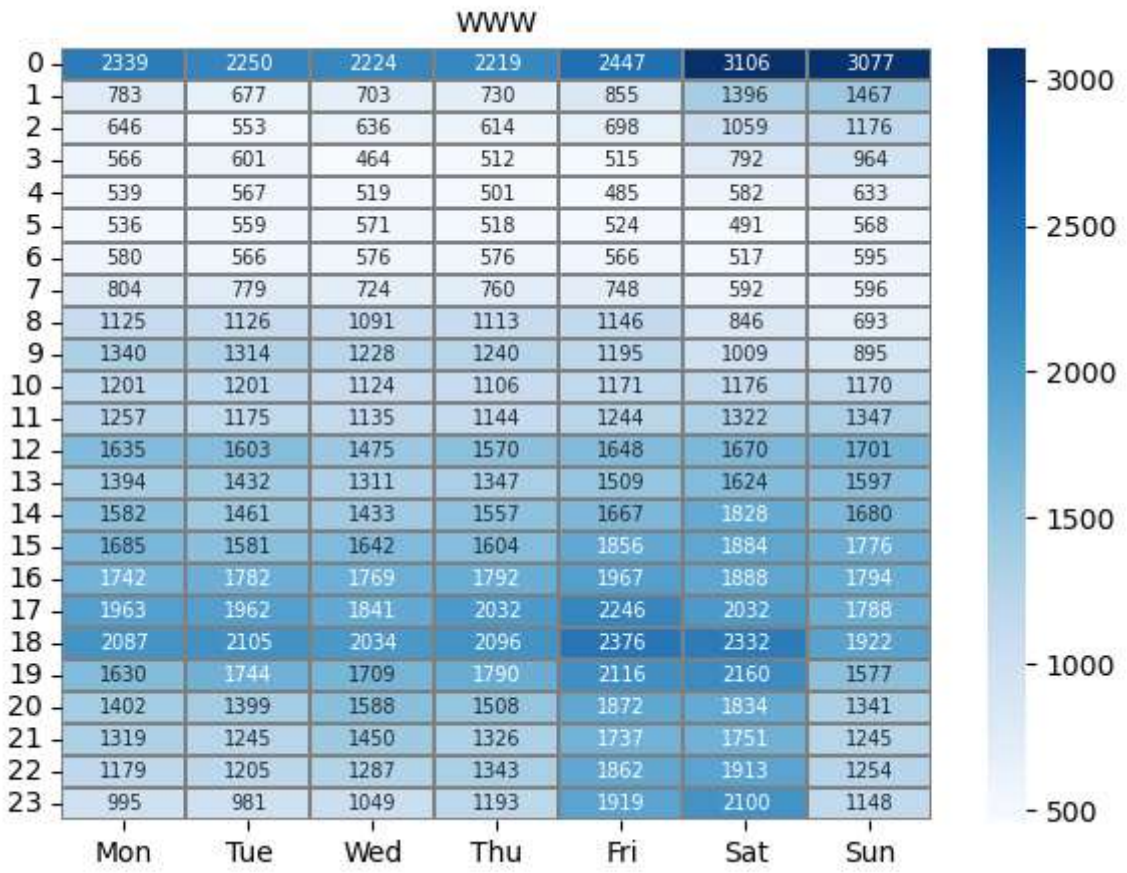
	crime_type_category	Average Number of Crimes per YEAR
4	Theft	7762.809524
2	Mischief	1416.380952
0	Break and Enter	849.428571
3	Offence Against a Person	338.666667
5	Vehicle Collision	212.285714
1	Homicide	3.850000

In Vancouver's Central Business District, which experiences the highest crime rates,
'Theft' notably leads withan average of 4,341 crimes per year, followed by 'Mischief'
and 'Break and Enter', while 'Homicide' remains relatively rare\ with an average of j
ust under 4 cases annually.

3-7-3-Which days, what times?

In [29]:
```
### what time which days
data = (
    crime[crime.NEIGHBOURHOOD == 'Central Business District' ].
```

```
    groupby(['HOUR' , 'weekday'])['TYPE'].
    count().
    rename('Number of Crimes').
    to_frame().
    reset_index()
)

fig , ax = plt.subplots(figsize=(7,5))
sns.heatmap(data  = data.pivot("HOUR" , "weekday" , "Number of Crimes")  ,
            cmap='Blues' , #YlGnBu',
            linecolor='grey',
            linewidths=0.05,
            cbar=True,
            annot=True ,
            fmt=".0f" ,
            annot_kws={"fontsize": 7},
            xticklabels=weekday_order ,
            ax = ax
)

ax.set_ylabel('')
ax.set_xlabel('')
ax.tick_params(axis='x', labelsize=10)
ax.set_title("WWW",fontsize=10)
#plt.title('Number of Crime per Weekday and TimeCategory', fontsize=10)

plt.yticks(rotation=0)
plt.show()
print("The Central Business District, has significant safety concerns, especially duri
     "\n The hours from 6 PM to 8 PM, particularly from Tuesday to Friday, register hi
```



The Central Business District, has significant safety concerns, especially during the late-night hours and on weekends.
The hours from 6 PM to 8 PM, particularly from Tuesday to Friday, register high crime rates.

In [327…
```
data.pivot("HOUR" , "weekday" , "Number of Crimes")
```

Out[327]:

HOUR	0	1	2	3	4	5	6
0	2338	2223	2219	2196	2423	3084	3050
1	777	668	695	724	844	1387	1461
2	640	545	635	611	696	1050	1170
3	562	600	460	506	511	787	955
4	536	565	516	497	478	576	632
5	531	556	566	516	523	490	567
6	575	560	571	573	562	513	588
7	802	778	721	759	744	588	591
8	1126	1120	1087	1108	1141	841	683
9	1334	1308	1216	1232	1192	1001	893
10	1193	1197	1118	1102	1169	1172	1165
11	1253	1164	1127	1140	1239	1314	1344
12	1632	1591	1470	1562	1639	1661	1689
13	1379	1425	1302	1336	1497	1613	1588
14	1566	1450	1424	1536	1655	1818	1672
15	1679	1566	1630	1587	1844	1869	1765
16	1728	1770	1752	1779	1956	1870	1782
17	1943	1950	1826	2017	2227	2018	1773
18	2080	2089	2020	2080	2365	2320	1911
19	1621	1731	1691	1774	2102	2147	1566
20	1393	1389	1578	1496	1863	1822	1335
21	1312	1239	1442	1320	1730	1740	1238
22	1175	1197	1280	1332	1857	1908	1247
23	989	976	1045	1189	1910	2091	1140

In [30]:
```
crime[crime.NEIGHBOURHOOD == 'Central Business District'].HUNDRED_BLOCK.value_counts()
```

Out[30]:
```
GRANVILLE ST                 24926
W GEORGIA ST                 14271
SEYMOUR ST                   10526
W PENDER ST                  10183
W HASTINGS ST                 9763
                             ...
BEATTY ST / CAMBIE ST            1
MAIN ST / E WATERFRONT RD        1
BEACH CRES / SEYMOUR ST          1
MAIN ST ONRP / MAIN ST           1
LENNOX PUB                       1
Name: HUNDRED_BLOCK, Length: 780, dtype: int64
```

In [31]:
```
import re
def extract_street_name(text):
    # Regular expression to find street patterns
    match = re.search(r'(\b\w+\b)\s(ST|AV|RD|DR|BLVD|LN|WAY|CT|PL|SQ|TRL|PKWY|CIR)', t
    if match:
        return match.group(0)
    else:
        return None

# Copy the HUNDRED_BLOCK column to a new DataFrame
new_df = pd.DataFrame()
new_df['HUNDRED_BLOCK'] = crime['HUNDRED_BLOCK'].copy()

# Apply the function to extract street names
new_df['Street_Name'] = new_df['HUNDRED_BLOCK'].apply(extract_street_name)

# Drop rows where Street_Name is None (if needed)
new_df = new_df.dropna(subset=['Street_Name'])
#new_df.Street_Name.unique()
#unique_streets = crimetest['Street_Name'].dropna().unique()
```

3-7-4-Which Crime Type Occurs Most Frequently in Each Neighborhood?

In [32]:
```
print(">>>>In the curr year:")
display(
            crime_curr.
            groupby(['NEIGHBOURHOOD' , 'TYPE'])['TYPE'].
            count().
            sort_values(ascending = False).
            groupby('NEIGHBOURHOOD').
            head(1).
            rename('Number of Crimes').
            to_frame().
            reset_index())


print(">>>>Over the years:")
display(
            crime.
            groupby(['NEIGHBOURHOOD' , 'TYPE'])['TYPE'].
            count().
            sort_values(ascending = False).
            groupby('NEIGHBOURHOOD').
            head(1).
            rename('Number of Crimes ').
```

```
            to_frame().
            reset_index())
```

```
>>>>In the curr year:
```

	NEIGHBOURHOOD	TYPE	Number of Crimes
0	Central Business District	Other Theft	3922
1	Renfrew-Collingwood	Other Theft	1501
2	West End	Other Theft	927
3	Mount Pleasant	Other Theft	862
4	Sunset	Other Theft	768
5	Strathcona	Other Theft	719
6	Fairview	Other Theft	667
7	Kensington-Cedar Cottage	Other Theft	400
8	Grandview-Woodland	Other Theft	393
9	Kitsilano	Other Theft	349
10	Marpole	Other Theft	259
11	Hastings-Sunrise	Theft from Vehicle	251
12	Riley Park	Theft from Vehicle	197
13	Victoria-Fraserview	Theft from Vehicle	170
14	Killarney	Theft from Vehicle	138
15	Kerrisdale	Other Theft	124
16	South Cambie	Other Theft	109
17	Dunbar-Southlands	Other Theft	102
18	West Point Grey	Theft from Vehicle	86
19	Oakridge	Theft from Vehicle	76
20	Arbutus Ridge	Theft from Vehicle	67
21	Shaughnessy	Theft from Vehicle	55
22	Stanley Park	Theft from Vehicle	52
23	Musqueam	Offence Against a Person	11

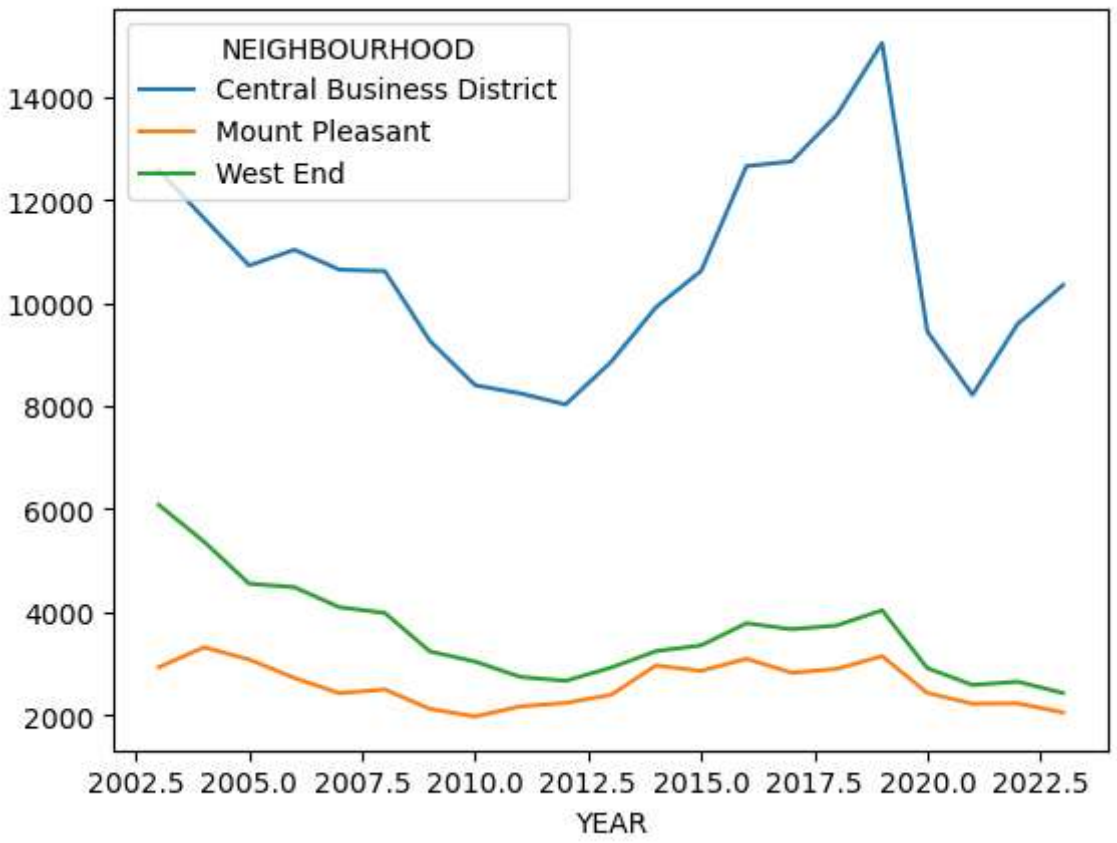
```
>>>>Over the years:
```

	NEIGHBOURHOOD	TYPE	Number of Crimes
0	Central Business District	Other Theft	76926
1	West End	Theft from Vehicle	23371
2	Fairview	Theft from Vehicle	15416
3	Mount Pleasant	Other Theft	14508
4	Strathcona	Other Theft	13615
5	Renfrew-Collingwood	Other Theft	13585
6	Kitsilano	Theft from Vehicle	12051
7	Kensington-Cedar Cottage	Theft from Vehicle	10508
8	Grandview-Woodland	Theft from Vehicle	10458
9	Hastings-Sunrise	Theft from Vehicle	8540
10	Sunset	Theft from Vehicle	7231
11	Riley Park	Theft from Vehicle	6138
12	Marpole	Theft from Vehicle	5921
13	Killarney	Theft from Vehicle	5247
14	Victoria-Fraserview	Theft from Vehicle	4823
15	Dunbar-Southlands	Theft from Vehicle	3821
16	Kerrisdale	Theft from Vehicle	3631
17	Oakridge	Other Theft	3450
18	Stanley Park	Theft from Vehicle	3302
19	West Point Grey	Theft from Vehicle	2804
20	Shaughnessy	Theft from Vehicle	2525
21	Arbutus Ridge	Theft from Vehicle	2453
22	South Cambie	Other Theft	2386
23	Musqueam	Offence Against a Person	252

3-7-5-Trend in Crime Rates for the Top 3 Nieghbourhoods with the Highest Frequencies

In [33]:
```
(crime[crime.NEIGHBOURHOOD
    .isin(['Central Business District','West End','Mount Pleasant'])]
    .groupby(['YEAR','NEIGHBOURHOOD'])['NEIGHBOURHOOD']
    .count()
    .reset_index(name='num_of_crimes')
    .pivot(index='YEAR',columns ='NEIGHBOURHOOD' , values='num_of_crimes')
    .plot
    .line()
              )
```

Out[33]: <AxesSubplot:xlabel='YEAR'>

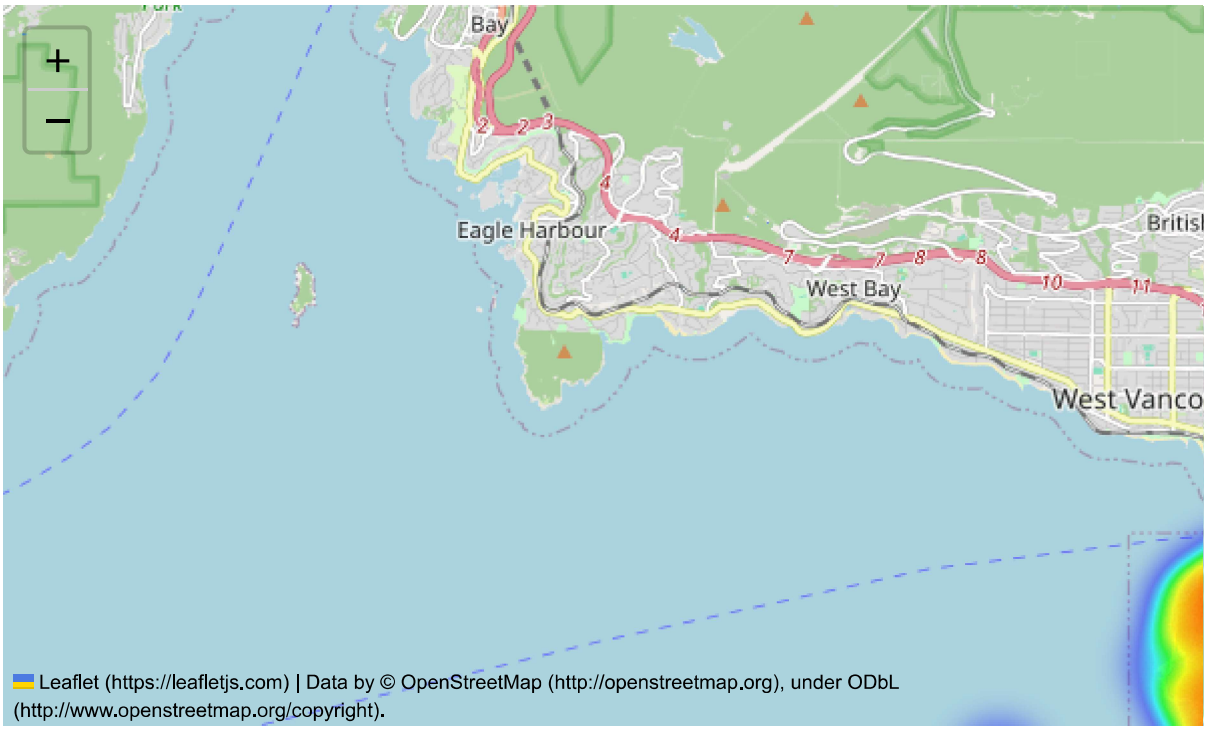


In [34]:
```
import folium
from folium.plugins import HeatMap


# Extract latitude and longitude from the dataframe
locations = crime[['lat', 'lon']].values.tolist()

m = folium.Map(location=[49.2827, -123.1207], zoom_start=12 , tiles='OpenStreetMap')
HeatMap(data=crime[['lat', 'lon']].dropna(), radius=8, max_zoom=13).add_to(m)

m.save("crime_density.html")


m = folium.Map(location=[49.285255, -123.123649], zoom_start=12, tiles='OpenStreetMap'

# Add the heatmap layer
```

I'm using Folium to map crime locations, and it's been a truly enjoyable experience – I highly recommend giving it a try!

In [35]:
```
# Base map
map_vancouver = folium.Map(location=[49.2827, -123.1207], zoom_start=12)

# Data for the heatmap (list of lat, lon points)
heat_data = [[row['lat'], row['lon']] for index, row in crime.iterrows()]

HeatMap(heat_data).add_to(map_vancouver)

map_vancouver
```

Out[35]:



Concluding this exploratory data analysis, I've also developed an **interactive Power BI report** to further visualize these insights.

You can explore the report in detail here.

I encourage you to check it out for a more dynamic view of the data.

In [ ]: