https://www.academia.edu/113294629/EXPLORING_SEASONAL_AFFECT_IN_CRIMES_CORRELATION_BETWEEN_CRIMES_AND_PARTICULAR_DAYS

EXPLORING_SEASONAL_AFFECT_IN_CRIMES_2_MUHARRAM_PERIOD, Ahmet YÜCE

Table of Contents

IMPORT LIBRARIES

In [1]:

```
# pip install hijri_converter
from hijri_converter import convert
```

In [2]:

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

```
from datetime import datetime
pd.set_option('display.max_rows', 300)    # None
```

USER DEFINED FUNCTIONS

In [3]:
```
def highlight_greater_than(value, threshold):
    if value > threshold:
        return 'background-color: yellow'
    else:
        return ''
```

In [4]:
```
# Filter rows that fall within the first ten days of the 12th months
def filter_first_ten_days_of_1th_months(date):
    # Extract the month and day from the Hijri_Date column
    month = int(date.split('-')[1])
    day = int(date.split('-')[2])

    # Check for the month: whether it's the 1th month and day: whether it's between 1 and 10
    return month == 1 and day <= 10
```

In [5]:
```
def common_codes(df):
    count_days = (max(df.date) - min(df.date)).days + 1
    df['Hijri_Date'] = df['date'].apply(lambda x: convert.Gregorian(x.year, x.month, x.day).to_hijri())
    df['Hijri_Year_Month'] = df['Hijri_Date'].astype(str).str[:-3]
    muharrem_df = df[df['Hijri_Date'].astype(str).apply(filter_first_ten_days_of_1th_months)]
    muharrem_dates_list_in_gregorian = pd.to_datetime(muharrem_df['date']).dt.strftime('%Y-%m-%d').unique().tolist()
    muharrem_dates_list_in_gregorian = sorted(muharrem_dates_list_in_gregorian, key=lambda x: datetime.strptime(x, '%Y-%m-%d'))
    count_muharrem_days = df.Hijri_Year_Month[df['Hijri_Year_Month'].astype(str).str.endswith("01")].nunique() * 10
    count_other_days = count_days - count_muharrem_days
    return count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], muharrem_df, muharrem_dates_list_in_gregorian,
```

In [6]:
```
def muharrem_10_days(df):
    df['date'] = pd.to_datetime(df['date'])
    count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], muharrem_df, muharrem_dates_list_in_gregorian, count_

    print("Total number of days:", count_days)
    print(len("Total number of days:") * "-")
    print("Total number of cases:", len(df))
    print(len("Total number of cases:") * "-")
```

```python
    average_case_count = round((len(df) / count_days), 2)
    print("Average Daily Case Count:", average_case_count)
    print(len("Average Daily Case Count:") * "-")

    print("Yearly case counts according to the Gregorian calendar:")
    print(len("Yearly case counts according to the Gregorian calendar:") * "-")
    print(df.date.dt.year.value_counts())
    print(len("Case counts according to the Hijri calendar:") * "-")

    print("Case counts according to the Hijri calendar:")
    print(len("Case counts according to the Hijri calendar:") * "-")
    print(df['Hijri_Date'].astype(str).str[:-6].value_counts())
    print(len("Average case count in the first ten days of Muharram months:") * "-")

    # Total number of cases in the first ten days of Muharram months (rows)
    count_muharrem_rows = len(muharrem_df)

    # Average case count in the first ten days of Muharram months:
    avg_count_of_muharrem_incidents = count_muharrem_rows / count_muharrem_days
    print("Average case count in the first ten days of Muharram months:", round(avg_count_of_muharrem_incidents,4))
    print(len("Average case count in the first ten days of Muharram months:") * "-")

    # Total number of cases in other days (rows)
    count_other_rows = len(df[~df['Hijri_Date'].astype(str).apply(filter_first_ten_days_of_1th_months)])
    count_other_days = count_days - count_muharrem_days
    avg_count_of_other_incidents = count_other_rows / count_other_days
    print("Average case count in other days:", round(avg_count_of_other_incidents,4))
    print(len("Average case count in other days:") * "-")

    print("Ratio of Muharram cases to other cases:", f"{avg_count_of_muharrem_incidents / avg_count_of_other_incidents:.{4}f}")
    print(len("Ratio of Muharram cases to other cases:") * "-")
```

In [7]:

```python
def incidents_by_types(df):
    # Explore distribution of incidents by types

    count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], muharrem_df, muharrem_dates_list_in_gregorian, count_

    all_incidents_count = df.incident.value_counts()

    muharrem_incidents = df[df['date'].astype(str).isin(muharrem_dates_list_in_gregorian)]
    muharrem_incidents_count = muharrem_incidents["incident"].value_counts()

    other_days_incidents = df[~df['date'].astype(str).isin(muharrem_dates_list_in_gregorian)]
```

```
    other_days_incidents_count = other_days_incidents["incident"].value_counts()

    # Top 30 incidents by the highest frequencies
    muharrem_grouped = muharrem_incidents.groupby('incident').size().nlargest(30)
    other_days_grouped = other_days_incidents.groupby('incident').size().nlargest(30)

    # Bar plot
    plt.figure(figsize=(12, 10))
    # Muharram Days bar plot
    plt.bar(muharrem_grouped.index, muharrem_grouped/count_muharrem_days, label='Muharram Days')
    # Other Days bar plot
    plt.bar(other_days_grouped.index, other_days_grouped/count_other_days, alpha=0.6, label='Other Days')
    plt.xlabel('Incident Types')
    plt.ylabel('Average Daily Frequency')
    plt.title('Top 30 Incidents by Type')
    plt.legend()
    plt.xticks(rotation=90, ha='center')
    plt.tight_layout()
    plt.show()

    # describe() statistics
    muharrem_incidents_desc = df[df['date'].astype(str).isin(muharrem_dates_list_in_gregorian)]["incident"].describe()
    other_days_incidents_desc = df[~df['date'].astype(str).isin(muharrem_dates_list_in_gregorian)]["incident"].describe()

    incident_ratios = pd.DataFrame({'muharrem incidents': muharrem_incidents_count, 'all incidents': all_incidents_count})
    incident_ratios["muharrem incidents/total incidents"]=incident_ratios["muharrem incidents"]/incident_ratios["all incidents"]
    incident_ratios["muharrem incidents/total incidents"]=round(incident_ratios["muharrem incidents/total incidents"],4)
    sorted_ratios = incident_ratios.sort_values(by="muharrem incidents/total incidents", ascending=False)[:30]
    equal_ratio = count_muharrem_days/count_days
    sorted_ratios['muharrem incidents'] = sorted_ratios['muharrem incidents'].astype(int)
    sorted_ratios = sorted_ratios.style.applymap(lambda x: highlight_greater_than(x, equal_ratio), subset=['muharrem incidents/to

    muharrem_dominant_incidents = incident_ratios[incident_ratios["muharrem incidents/total incidents"]>equal_ratio]
    muharrem_dominant_incidents = muharrem_dominant_incidents.sort_values(by="muharrem incidents", ascending=False)
    muharrem_dominant_incidents['muharrem incidents'] = muharrem_dominant_incidents['muharrem incidents'].astype(int)

    return sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc
```

In [8]:
```
def monthly_count_plot ():

    # Extract Year-Month part of the dates
    def extract_hijri_year_month(date):
        year_month = date.split('-')[:2]
```

```
        return '-'.join(year_month)

    # Add 'Hijri_Year_Month' column
    df['Hijri_Year_Month'] = df['Hijri_Date'].astype(str).apply(extract_hijri_year_month)

    # Count monthly incidents
    monthly_count = df.groupby('Hijri_Year_Month')['Hijri_Date'].count()

    # Avg count of incidents on Hijri_Year_Month basis
    monthly_avg = monthly_count.groupby('Hijri_Year_Month').mean()

    # Overall Avg count of incidents
    overall_avg = monthly_count.mean()

    # Line plot
    plt.figure(figsize=(10, 6))

    # Line of Avg Incident counts
    plt.axhline(y=overall_avg, color='g', linestyle='--', label=f'Overall Average: {overall_avg:.2f}')

    # Line of monthly count of incidents
    plt.plot(monthly_avg.index, monthly_avg.values, marker='o', linestyle='-', color='b', label='Monthly Count')

    # Mark 9th. months
    first_month = [i for i, month in enumerate(monthly_avg.index) if '01' in month]
    plt.scatter(monthly_avg.index[first_month], monthly_avg.values[first_month], color='red', s=100, label='1th Month:Muharram')

    # Plot
    plt.title('Montly Count of Incidents on Hijri Calendar')
    plt.xlabel('Hijri_Year-Month')
    plt.ylabel('Monthly Count of Incidents')
    plt.xticks(rotation=90)
    plt.legend()
    plt.tight_layout()
    plt.show()
```

SAMPLE DATA-1: UNC-CHAPEL HILL CAMPUS POLICE CRIME LOG_2013-2018

https://data.world/skillenberg/unc-police-incidents-2013-2018

In [9]:
```python
df = pd.read_excel("unc-police-data-killenberg.xlsx", sheet_name=1)
df
```

Out[9]:

	incident	date-time	year	date-no year	location	res hall	alcohol
0	NaN	1/1/13 0:00	2013	01/01	E FRANKLIN ST	NaN	NaN
1	EMS ASSIST	1/1/13 2:25	2013	01/01	AYCOCK CIRCLE PARKING LOT UNC	NaN	NaN
2	SUSPICIOUS CONDITION (NON-CRIMINAL)	1/1/13 2:49	2013	01/01	S COLUMBIA ST/E CAMERON AVE	NaN	NaN
3	EMS ASSIST	1/1/14 12:25	2014	01/01	FINLEY CLUB HOUSE UNC	NaN	NaN
4	WELL-BEING CHECK	1/1/14 20:17	2014	01/01	CRAIGE RES HALL UNC	RES HALL	NaN
...	...	...	...	...	...	...	...
12812	EMS ASSIST	9/9/17 4:24	2017	09/09	OLD EAST RES HALL UNC	RES HALL	NaN
12813	FOUND PROPERTY	9/9/17 8:19	2017	09/09	DOGWOOD PARKING DECK UNC	NaN	NaN
12814	LARCENY - FROM BUILDING	9/9/17 9:48	2017	09/09	BERRYHILL UNC	NaN	NaN
12815	ASSIST OTHER AGENCY	9/9/18 2:40	2018	09/09	E FRANKLIN ST	NaN	NaN
12816	DWI-ALCOHOL	9/9/18 3:35	2018	09/09	COUNTRY CLUB RD	NaN	DWI-ALCOHOL

12817 rows × 7 columns

In [10]:
```python
df = df.drop_duplicates()
```

In [11]:
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12714 entries, 0 to 12816
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   incident      9518 non-null   object
 1   date-time     12714 non-null  object
 2   year          12714 non-null  int64
 3   date-no year  12714 non-null  object
 4   location      12711 non-null  object
 5   res hall      2262 non-null   object
 6   alcohol       746 non-null    object
dtypes: int64(1), object(6)
memory usage: 794.6+ KB
```

In [12]:
```
df["incident"].value_counts().count()
```

Out[12]:
```
730
```

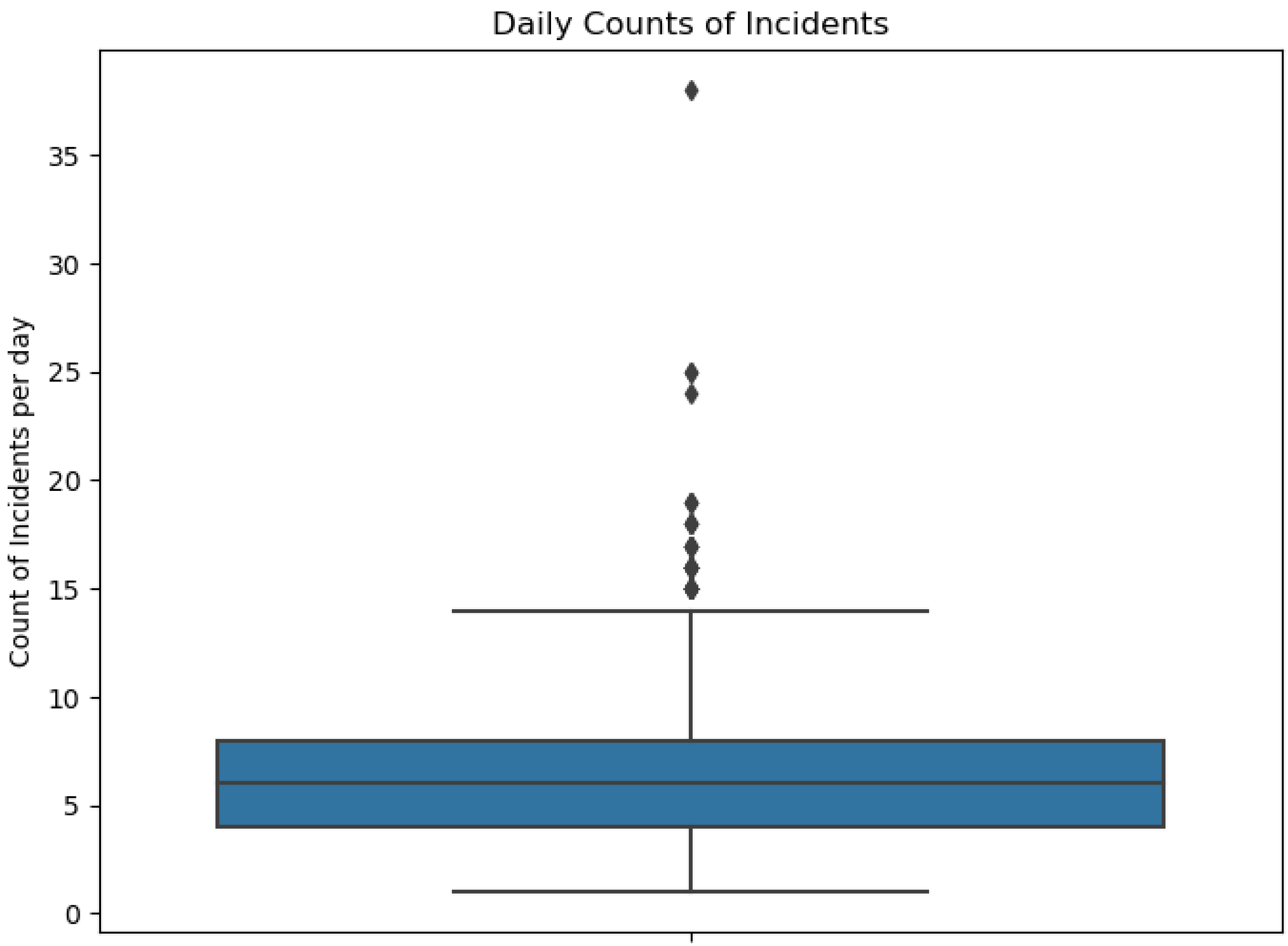
In [13]:
```
df["incident"].value_counts()[:15]
```

Out[13]:
```
EMS ASSIST                               1875
FOUND PROPERTY                            597
LARCENY - FROM BUILDING                   546
ALCOHOL - UNDERAGE CONSUMPTION            544
INFORMATIONAL                             440
VANDALISM / PROPERTY DAMAGE               388
LARCENY OF BICYCLE                        306
SUSPICIOUS CONDITION (NON-CRIMINAL)       293
PROPERTY DAMAGE                           259
ASSIST OTHER AGENCY                       208
CALLS FOR SERVICE                         159
WELL-BEING CHECK                          138
LARCENY-ALL OTHER                         134
VOLUNTARY COMMITMENT                      119
LOST PROPERTY                             118
Name: incident, dtype: int64
```

In [14]:
```
df = df.rename(columns = {'date-time':'date'})
```

In [15]:
```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [16]:
```
min(df.date), max(df.date)
```
Out[16]:
```
('2013-01-01', '2018-10-10')
```
In [17]:
```
df = df.iloc[:, [0,1]]
# df.to_csv("UNC.csv")
```
In [18]:
```
df.info()
```
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12714 entries, 0 to 12816
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   incident  9518 non-null   object
 1   date      12714 non-null  object
dtypes: object(2)
memory usage: 298.0+ KB
```
In [19]:
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```
Out[19]:
```
count    2057
mean        6
std         3
min         1
25%         4
50%         6
75%         8
95%        12
98%        15
99%        16
max        38
Name: date, dtype: int32
```
In [20]:
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



Daily Counts of Incidents
Count of Incidents per day
35
30
25
20
15
10
5
0

In [21]:
```
df.date
```

Out[21]:
```
0        2013-01-01
1        2013-01-01
2        2013-01-01
3        2014-01-01
4        2014-01-01
            ...
12812    2017-09-09
12813    2017-09-09
12814    2017-09-09
12815    2018-09-09
12816    2018-09-09
Name: date, Length: 12714, dtype: object
```

In [22]:
```
df.date.nunique()
```

Out[22]:
```
2057
```

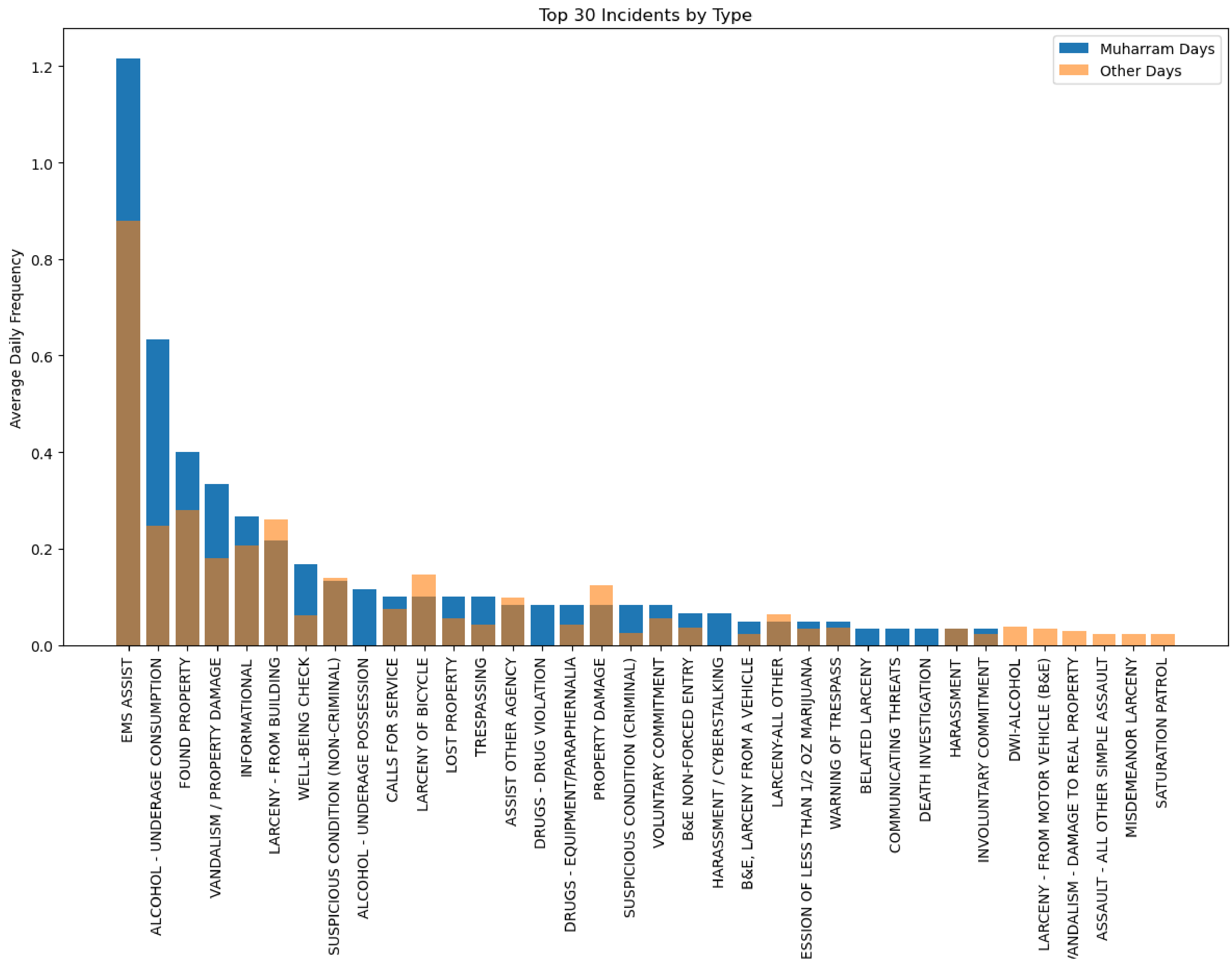
As seen below, our dataset spans a total of 2109 days. During this period, incidents occurred on 2057 days, while there were no records of incidents on the remaining 52 days.

In [23]:
```
muharrem_10_days (df)
```

```
Total number of days: 2109
---------------------
Total number of cases: 12714
----------------------
Average Daily Case Count: 6.03
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2017    2491
2015    2268
2014    2239
2016    2224
2013    2109
2018    1383
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1438    2385
1435    2157
1437    2135
1436    2123
1439    1973
1434    1779
1440     162
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 7.4833
------------------------------------------------------------
Average case count in other days: 5.9858
-------------------------------
Ratio of Muharram cases to other cases: 1.2502
---------------------------------------
```

We observe a 25.027% higher crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [24]:
```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0.0
0.2
0.4
0.6
0.8
1.0
1.2
EMS ASSIST
ALCOHOL - UNDERAGE CONSUMPTION
FOUND PROPERTY
VANDALISM / PROPERTY DAMAGE
INFORMATIONAL
LARCENY - FROM BUILDING
WELL-BEING CHECK
SUSPICIOUS CONDITION (NON-CRIMINAL)
ALCOHOL - UNDERAGE POSSESSION
CALLS FOR SERVICE
LARCENY OF BICYCLE
LOST PROPERTY
TRESPASSING
ASSIST OTHER AGENCY
DRUGS - DRUG VIOLATION
DRUGS - EQUIPMENT/PARAPHERNALIA
PROPERTY DAMAGE
SUSPICIOUS CONDITION (CRIMINAL)
VOLUNTARY COMMITMENT
B&E NON-FORCED ENTRY
HARASSMENT / CYBERSTALKING
B&E, LARCENY FROM A VEHICLE
LARCENY-ALL OTHER
ESSION OF LESS THAN 1/2 OZ MARIJUANA
WARNING OF TRESPASS
BELATED LARCENY
COMMUNICATING THREATS
DEATH INVESTIGATION
HARASSMENT
INVOLUNTARY COMMITMENT
DWI-ALCOHOL
LARCENY - FROM MOTOR VEHICLE (B&E)
/ANDALISM - DAMAGE TO REAL PROPERTY
ASSAULT - ALL OTHER SIMPLE ASSAULT
MISDEMEANOR LARCENY
SATURATION PATROL

In [25]:
```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[25]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
(NON-CRIMINAL) CIVIL MATTER	1	1	1.000000
DRUGS - LESS THEN 1/2 OZ SCH VI	1	1	1.000000
SERVICE CALL - REPAIRS	1	1	1.000000
POSSESSION OF A CONTROLLED SUBSTANCE	1	1	1.000000
MISSING PROPERTY	1	1	1.000000
LARCENY-FROM VEHICLE	1	1	1.000000
INFORMATION-ASSIST OTHER AGENCY	1	1	1.000000
FELONY POSSESSION OF COCAINE	1	1	1.000000
DRUGS -POSSESS HASHISH	1	1	1.000000
WHILE LICENSE REVOKED	1	1	1.000000
ASSISTED THE FIRE DEPARTMENT	1	1	1.000000
EMS	1	2	0.500000
FIRE DAMAGE	1	2	0.500000
ROAD RAGE	1	2	0.500000
ODOR INVESTIGATION	2	5	0.400000
DEATH INVESTIGATION	2	5	0.400000
ASSAULT-AGGRAVATED	1	3	0.333300
SEXUAL BATTERY	1	3	0.333300
HARASSMENT / STALKING	1	4	0.250000
SMOKE INVESTIGATION	1	4	0.250000
COMMUNICATING THREATS	2	8	0.250000
DRUGS - PARAPHERNALIA - POSSESSING	1	4	0.250000
PROPERTY DAMAGE (NON-CRIMINAL)	1	5	0.200000
DRIVERS LICENSE CHECKPOINT	1	5	0.200000

	muharrem incidents	all incidents	muharrem incidents/total incidents
DRUNK AND DISRUPTIVE	1	5	0.200000
FRAUD-CREDIT CARD	1	5	0.200000
COMMUNICATE THREATS	1	6	0.166700
CLERY REPORT FORCIBLE FONDLING	1	6	0.166700
ALCOHOL - UNDERAGE POSSESSION	7	48	0.145800
CLERY - ALLEGATION OF FONDLING	1	7	0.142900

In [26]:

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[26]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
EMS ASSIST	73	1875	0.0389
ALCOHOL - UNDERAGE CONSUMPTION	38	544	0.0699
FOUND PROPERTY	24	597	0.0402
VANDALISM / PROPERTY DAMAGE	20	388	0.0515
INFORMATIONAL	16	440	0.0364
WELL-BEING CHECK	10	138	0.0725
ALCOHOL - UNDERAGE POSSESSION	7	48	0.1458
TRESPASSING	6	92	0.0652
LOST PROPERTY	6	118	0.0508
CALLS FOR SERVICE	6	159	0.0377
VOLUNTARY COMMITMENT	5	119	0.0420
SUSPICIOUS CONDITION (CRIMINAL)	5	58	0.0862
DRUGS - EQUIPMENT/PARAPHERNALIA	5	94	0.0532
DRUGS - DRUG VIOLATION	5	50	0.1000
B&E NON-FORCED ENTRY	4	77	0.0519
HARASSMENT / CYBERSTALKING	4	38	0.1053
B&E, LARCENY FROM A VEHICLE	3	50	0.0600
POSSESSION OF LESS THAN 1/2 OZ MARIJUANA	3	73	0.0411
WARNING OF TRESPASS	3	76	0.0395
ODOR INVESTIGATION	2	5	0.4000
MISDEMEANOR LARCENY	2	50	0.0400
SATURATION PATROL	2	50	0.0400
SUSPICIOUS CONDITION	2	16	0.1250
DEATH INVESTIGATION	2	5	0.4000

	muharrem incidents	all incidents	muharrem incidents/total incidents
COMMUNICATING THREATS	2	8	0.2500
LOST / FOUND PROPERTY	2	44	0.0455
INVOLUNTARY COMMITMENT	2	50	0.0400
BELATED LARCENY	2	17	0.1176
MISSING PERSONS	1	23	0.0435
POSSESSION OF A CONTROLLED SUBSTANCE	1	1	1.0000
MISSING PROPERTY	1	1	1.0000
(NON-CRIMINAL) CIVIL MATTER	1	1	1.0000
SEXUAL BATTERY	1	3	0.3333
PROPERTY DAMAGE (NON-CRIMINAL)	1	5	0.2000
RAPE - FORCIBLE	1	11	0.0909
ROAD RAGE	1	2	0.5000
SERVICE CALL - REPAIRS	1	1	1.0000
LARCENY OF CELL PHONE	1	20	0.0500
SIMPLE ASSAULT	1	15	0.0667
SIMPLE POSSESSION OF MARIJUANA	1	28	0.0357
SMOKE INVESTIGATION	1	4	0.2500
SUSPICIOUS PERSON	1	24	0.0417
UNDERAGE DRINKING	1	14	0.0714
VANDALISM	1	7	0.1429
LARCENY-FROM VEHICLE	1	1	1.0000
FRAUD-CREDIT CARD	1	5	0.2000
LARCENY OF A LAPTOP	1	15	0.0667
LARCENY - AUTO PART & ACCESSORIES	1	24	0.0417

	muharrem incidents	all incidents	muharrem incidents/total incidents
ANIMAL COMPLAINT	1	14	0.0714
ASSAULT - COMMUNICATE THREATS	1	35	0.0286
ASSAULT-AGGRAVATED	1	3	0.3333
ASSISTED THE FIRE DEPARTMENT	1	1	1.0000
BURGLARY- NON-FORCED ENTRY (STRUCTURES)	1	18	0.0556
CLERY - ALLEGATION OF FONDLING	1	7	0.1429
CLERY - ALLEGATION OF RAPE	1	25	0.0400
CLERY - REPORT STALKING	1	19	0.0526
CLERY REPORT FORCIBLE FONDLING	1	6	0.1667
COMMUNICATE THREATS	1	6	0.1667
DAYTIME CHECK POINT (PART)	1	17	0.0588
DAYTIME CHECKPOINT (HOST)	1	25	0.0400
DISORDERLY CONDUCT	1	9	0.1111
DISTURBANCE	1	21	0.0476
DRIVERS LICENSE CHECKPOINT	1	5	0.2000
DRUGS - LESS THEN 1/2 OZ SCH VI	1	1	1.0000
DRUGS - PARAPHERNALIA - POSSESSING	1	4	0.2500
DRUGS -POSSESS HASHISH	1	1	1.0000
DRUNK AND DISRUPTIVE	1	5	0.2000
EMS	1	2	0.5000
EMS ASSIST - ALCOHOL	1	15	0.0667
FELONY POSSESSION OF COCAINE	1	1	1.0000
FIRE DAMAGE	1	2	0.5000
FRAUD - ALL OTHER	1	18	0.0556

	muharrem incidents	all incidents	muharrem incidents/total incidents
FRAUD - WIRE/COMPUTER/OTHER ELECTRONIC	1	11	0.0909
HARASSMENT / STALKING	1	4	0.2500
INFORMATION-ASSIST OTHER AGENCY	1	1	1.0000
WHILE LICENSE REVOKED	1	1	1.0000

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [27]:
```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[27]:
```
(730, 76)
```

In [28]:
```
muharrem_incidents_desc
```

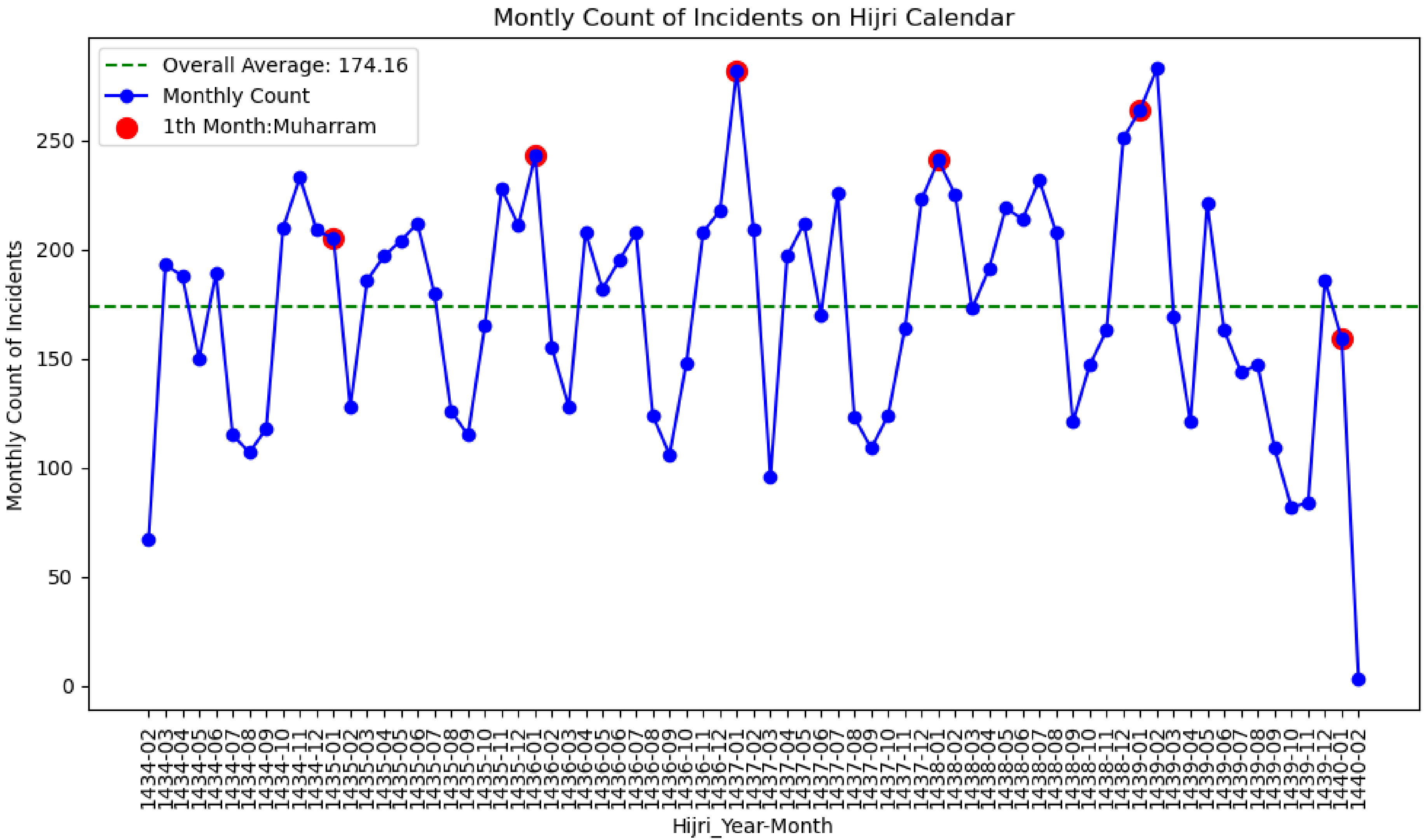
Out[28]:
```
count            358
unique            90
top       EMS ASSIST
freq              73
Name: incident, dtype: object
```

In [29]:
```
other_days_incidents_desc
```

Out[29]:
```
count           9160
unique           719
top       EMS ASSIST
freq            1802
Name: incident, dtype: object
```

UNC dataset encompasses 730 distinct incident types. During the first ten days of Muharram, crimes were committed across 90 incident categories, with 76 of these categories experiencing incident counts exceeding the annual averages.

In [30]:
```
monthly_count_plot()
```



SAMPLE DATA-2: LOS ANGELES CRIME DATA_2010-2019

https://www.kaggle.com/datasets/cityofLA/los-angeles-crime-arrest-data/?select=crime-data-from-2010-to-present.csv

In [73]:
```
df = pd.read_csv("Los_Angles_crime-data-from-2010-to-present.csv")
df.head()
```

Out[73]:

	DR Number	Date Reported	Date Occurred	Time Occurred	Area ID	Area Name	Reporting District	Crime Code	Crime Code Description	MO Codes	...	Weapon Description	Status Code	Status Description	Crime Code 1	Cr C
0	102005556	2010-01-25T00:00:00	2010-01-22T00:00:00	2300	20	Olympic	2071	510	VEHICLE - STOLEN	NaN	...	NaN	IC	Invest Cont	510.0	
1	101822289	2010-11-11T00:00:00	2010-11-10T00:00:00	1800	18	Southeast	1803	510	VEHICLE - STOLEN	NaN	...	NaN	IC	Invest Cont	510.0	
2	101105609	2010-01-28T00:00:00	2010-01-27T00:00:00	2230	11	Northeast	1125	510	VEHICLE - STOLEN	NaN	...	NaN	IC	Invest Cont	510.0	
3	101620051	2010-11-11T00:00:00	2010-11-07T00:00:00	1600	16	Foothill	1641	510	VEHICLE - STOLEN	NaN	...	NaN	IC	Invest Cont	510.0	
4	101910498	2010-04-07T00:00:00	2010-04-07T00:00:00	1600	19	Mission	1902	510	VEHICLE - STOLEN	NaN	...	NaN	IC	Invest Cont	510.0	

5 rows × 26 columns

duplicate_values = df.duplicated(subset=None, keep='first').sum() df = df.drop_duplicates()

In [74]:
```
df.duplicated().value_counts()
```

Out[74]:
```
False    1993259
dtype: int64
```

In [75]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1993259 entries, 0 to 1993258
Data columns (total 26 columns):
 #   Column                  Dtype  
---  ------                  -----  
 0   DR Number               int64  
 1   Date Reported           object 
 2   Date Occurred           object 
 3   Time Occurred           int64  
 4   Area ID                 int64  
 5   Area Name               object 
 6   Reporting District      int64  
 7   Crime Code              int64  
 8   Crime Code Description  object 
 9   MO Codes                object 
 10  Victim Age              int64  
 11  Victim Sex              object 
 12  Victim Descent          object 
 13  Premise Code            float64
 14  Premise Description     object 
 15  Weapon Used Code        float64
 16  Weapon Description      object 
 17  Status Code             object 
 18  Status Description      object 
 19  Crime Code 1            float64
 20  Crime Code 2            float64
 21  Crime Code 3            float64
 22  Crime Code 4            float64
 23  Address                 object 
 24  Cross Street            object 
 25  Location                object 
dtypes: float64(6), int64(6), object(14)
memory usage: 395.4+ MB
```

In [76]:
```
df = pd.read_csv("Los_Angles_crime-data-from-2010-to-present.csv", usecols=[0,2,8])
```

In [77]:
```
df["Crime Code Description"].value_counts()[:15]
```

Out[77]:

```
BATTERY - SIMPLE ASSAULT                                  180434
BURGLARY FROM VEHICLE                                     153451
VEHICLE - STOLEN                                          151622
THEFT PLAIN - PETTY ($950 & UNDER)                        141489
BURGLARY                                                  140926
THEFT OF IDENTITY                                         120835
INTIMATE PARTNER - SIMPLE ASSAULT                         107900
VANDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)   102589
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT             86829
VANDALISM - MISDEAMEANOR ($399 OR UNDER)                   86440
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)            82791
ROBBERY                                                    79392
THEFT-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD   70081
CRIMINAL THREATS - NO WEAPON DISPLAYED                     53959
SHOPLIFTING - PETTY THEFT ($950 & UNDER)                   45493
Name: Crime Code Description, dtype: int64
```

In [78]:

```
df = df.rename(columns = {'Date Occurred':'date'})
df = df.rename(columns = {'Crime Code Description':'incident'})
```

In [79]:

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [80]:

```
min(df.date), max(df.date)
```

Out[80]:

```
('2010-01-01', '2019-06-22')
```

In [81]:

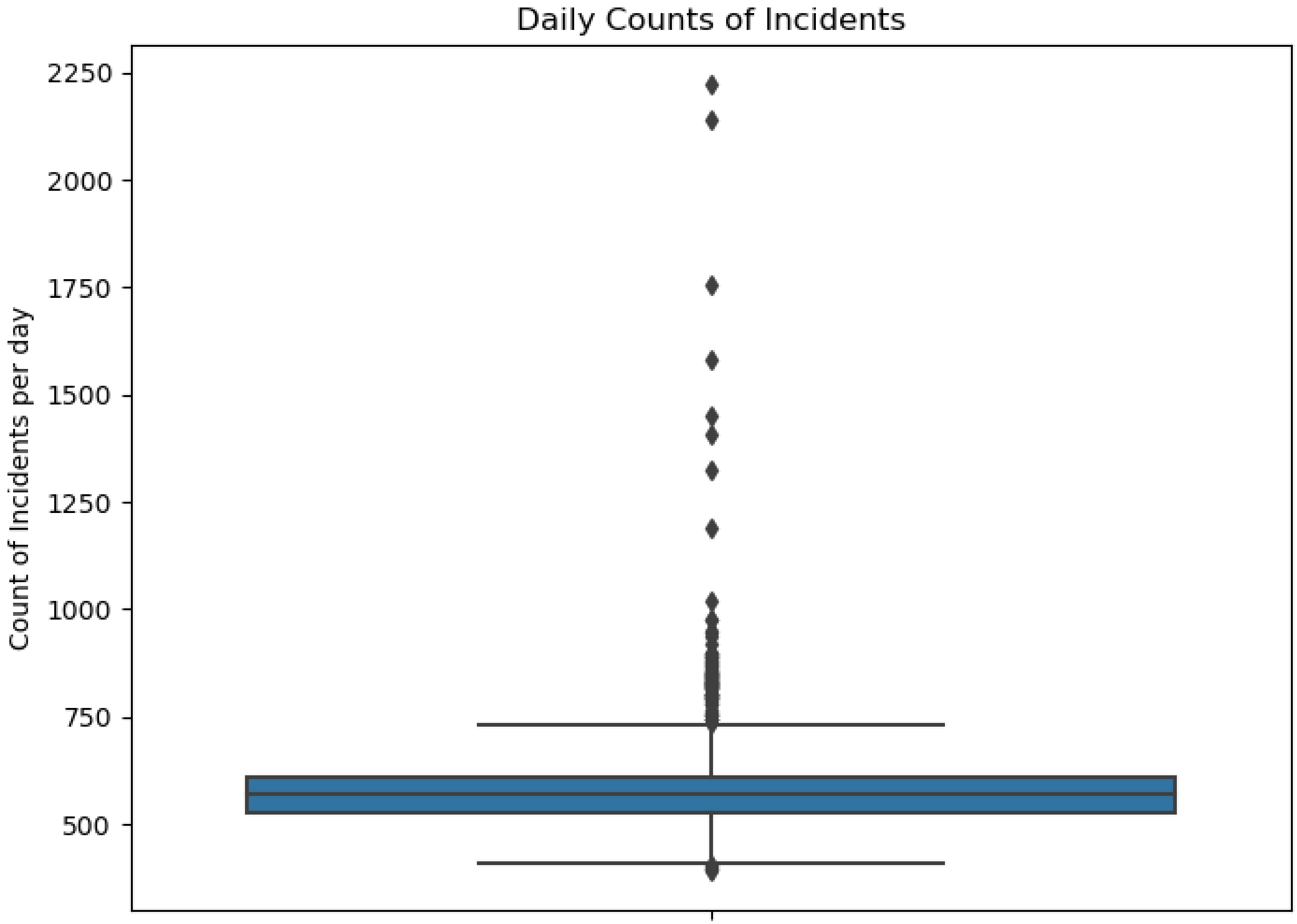
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[81]:

```
count    3460
mean      576
std        89
min       390
25%       528
50%       568
75%       610
95%       683
98%       820
99%       854
max      2222
Name: date, dtype: int32
```

In [82]:

```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [83]:

```
df.date.nunique()
```

Out[83]:

```
3460
```

As seen below, our dataset spans a total of 3460 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

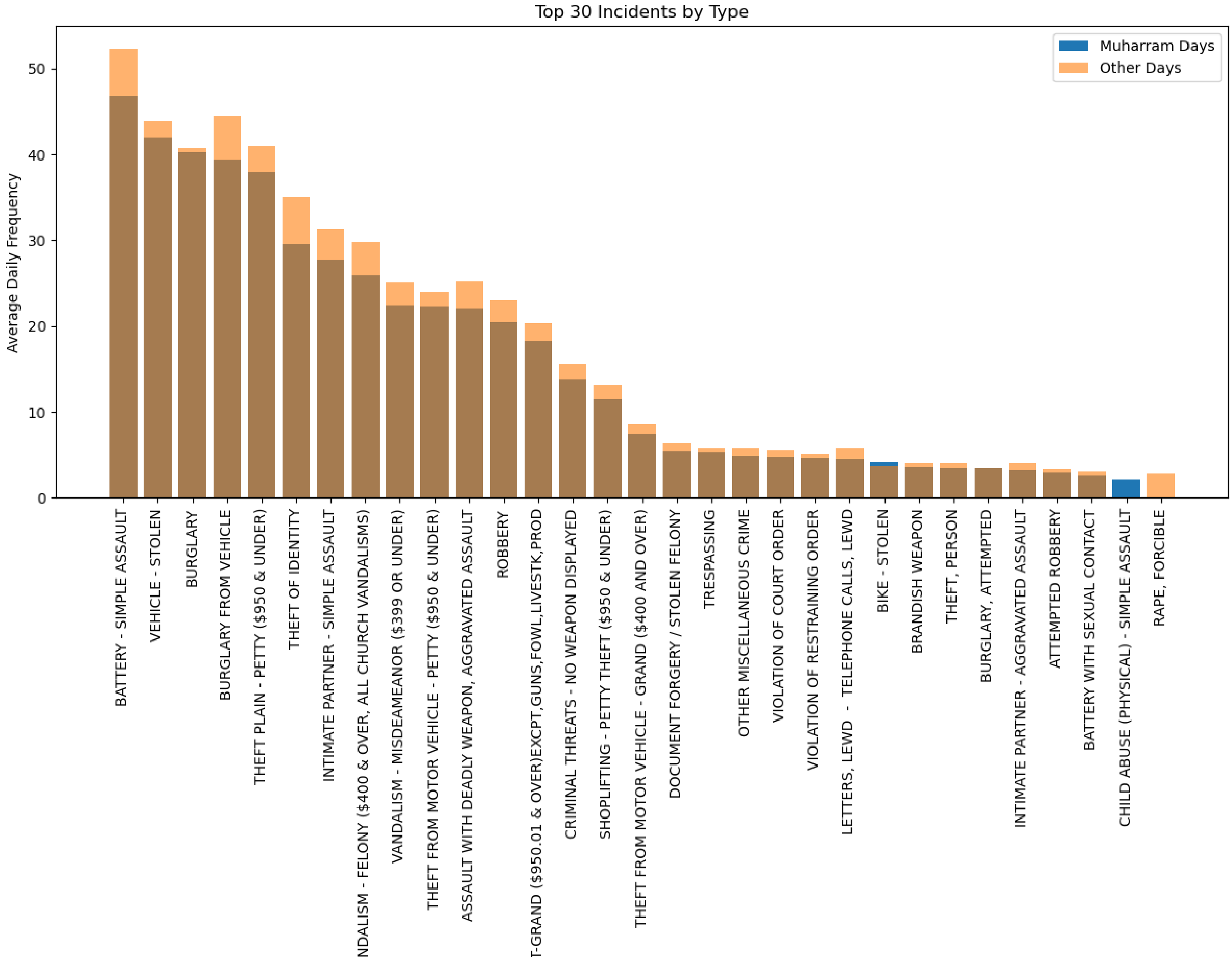
In [84]:
```
muharrem_10_days (df)
```

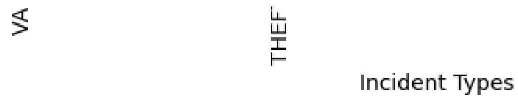
```
Total number of days: 3460
---------------------
Total number of cases: 1993259
----------------------
Average Daily Case Count: 576.09
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2017    229930
2018    226909
2016    224645
2015    214822
2010    208823
2012    201170
2011    200437
2014    195022
2013    192211
2019     99290
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1439    222769
1438    221164
1437    215516
1436    205095
1433    195939
1431    195346
1432    194561
1434    188032
1435    186913
1440    167924
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 518.65
------------------------------------------------------------
Average case count in other days: 577.7958
---------------------------------
Ratio of Muharram cases to other cases: 0.8976
---------------------------------------
```

We observe a -10.24% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [85]:

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
```


Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
10
20
30
40
50
BATTERY - SIMPLE ASSAULT
VEHICLE - STOLEN
BURGLARY
BURGLARY FROM VEHICLE
THEFT PLAIN - PETTY ($950 & UNDER)
THEFT OF IDENTITY
INTIMATE PARTNER - SIMPLE ASSAULT
NDALISM - FELONY ($400 & OVER, ALL CHURCH VANDALISMS)
VANDALISM - MISDEAMEANOR ($399 OR UNDER)
THEFT FROM MOTOR VEHICLE - PETTY ($950 & UNDER)
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT
ROBBERY
Γ-GRAND ($950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD
CRIMINAL THREATS - NO WEAPON DISPLAYED
SHOPLIFTING - PETTY THEFT ($950 & UNDER)
THEFT FROM MOTOR VEHICLE - GRAND ($400 AND OVER)
DOCUMENT FORGERY / STOLEN FELONY
TRESPASSING
OTHER MISCELLANEOUS CRIME
VIOLATION OF COURT ORDER
VIOLATION OF RESTRAINING ORDER
LETTERS, LEWD - TELEPHONE CALLS, LEWD
BIKE - STOLEN
BRANDISH WEAPON
THEFT, PERSON
BURGLARY, ATTEMPTED
INTIMATE PARTNER - AGGRAVATED ASSAULT
ATTEMPTED ROBBERY
BATTERY WITH SEXUAL CONTACT
CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT
RAPE, FORCIBLE



In [86]:

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[86]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)	2	9	0.222200
FAILURE TO DISPERSE	2	20	0.100000
TELEPHONE PROPERTY - DAMAGE	3	35	0.085700
DISRUPT SCHOOL	3	41	0.073200
THEFT, COIN MACHINE - GRAND ($950.01 & OVER)	3	43	0.069800
DRUNK ROLL	2	34	0.058800
TILL TAP - PETTY ($950 & UNDER)	5	93	0.053800
DRIVING WITHOUT OWNER CONSENT (DWOC)	21	450	0.046700
PICKPOCKET, ATTEMPT	1	22	0.045500
COUNTERFEIT	36	808	0.044600
CHILD STEALING	47	1053	0.044600
LYNCHING - ATTEMPTED	1	23	0.043500
REPLICA FIREARMS(SALE,DISPLAY,MANUFACTURE OR DISTRIBUTE)	1	24	0.041700
BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM	1	25	0.040000
LEWD/LASCIVIOUS ACTS WITH CHILD	8	213	0.037600
CHILD PORNOGRAPHY	7	189	0.037000
CONSPIRACY	2	54	0.037000
KIDNAPPING	71	1929	0.036800
WEAPONS POSSESSION/BOMBING	6	163	0.036800
EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)	20	560	0.035700
THEFT PLAIN - ATTEMPT	56	1589	0.035200
KIDNAPPING - GRAND ATTEMPT	23	689	0.033400
BRIBERY	1	30	0.033300
DOCUMENT WORTHLESS ($200 & UNDER)	2	60	0.033300

	muharrem incidents	all incidents	muharrem incidents/total incidents
CREDIT CARDS, FRAUD USE ($950 & UNDER	9	272	0.033100
BOMB SCARE	37	1126	0.032900
BIKE - STOLEN	421	12996	0.032400
SHOPLIFTING - ATTEMPT	7	216	0.032400
DEFRAUDING INNKEEPER/THEFT OF SERVICES, $400 & UNDER	64	2068	0.030900
PICKPOCKET	27	899	0.030000

In [87]:

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[87]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
BIKE - STOLEN	421	12996	0.0324
VEHICLE - ATTEMPT STOLEN	94	3190	0.0295
KIDNAPPING	71	1929	0.0368
DEFRAUDING INNKEEPER/THEFT OF SERVICES, $400 & UNDER	64	2068	0.0309
THEFT PLAIN - ATTEMPT	56	1589	0.0352
CHILD STEALING	47	1053	0.0446
THROWING OBJECT AT MOVING VEHICLE	45	1529	0.0294
BOMB SCARE	37	1126	0.0329
COUNTERFEIT	36	808	0.0446
PURSE SNATCHING	33	1139	0.0290
PEEPING TOM	33	1113	0.0296
PICKPOCKET	27	899	0.0300
KIDNAPPING - GRAND ATTEMPT	23	689	0.0334
DRIVING WITHOUT OWNER CONSENT (DWOC)	21	450	0.0467
EMBEZZLEMENT, PETTY THEFT ($950 & UNDER)	20	560	0.0357
CREDIT CARDS, FRAUD USE ($950 & UNDER	9	272	0.0331
LEWD/LASCIVIOUS ACTS WITH CHILD	8	213	0.0376
CHILD PORNOGRAPHY	7	189	0.0370
SHOPLIFTING - ATTEMPT	7	216	0.0324
WEAPONS POSSESSION/BOMBING	6	163	0.0368
TILL TAP - PETTY ($950 & UNDER)	5	93	0.0538
TELEPHONE PROPERTY - DAMAGE	3	35	0.0857
DISRUPT SCHOOL	3	41	0.0732
THEFT, COIN MACHINE - GRAND ($950.01 & OVER)	3	43	0.0698

	muharrem incidents	all incidents	muharrem incidents/total incidents
INCEST (SEXUAL ACTS BETWEEN BLOOD RELATIVES)	2	9	0.2222
FAILURE TO DISPERSE	2	20	0.1000
DRUNK ROLL	2	34	0.0588
DOCUMENT WORTHLESS ($200 & UNDER)	2	60	0.0333
CONSPIRACY	2	54	0.0370
LYNCHING - ATTEMPTED	1	23	0.0435
PICKPOCKET, ATTEMPT	1	22	0.0455
REPLICA FIREARMS(SALE,DISPLAY,MANUFACTURE OR DISTRIBUTE)	1	24	0.0417
BRIBERY	1	30	0.0333
BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM	1	25	0.0400

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [88]:
```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[88]:
```
(140, 34)
```

In [89]:
```
muharrem_incidents_desc
```

Out[89]:
```
count                        51865
unique                         122
top       BATTERY - SIMPLE ASSAULT
freq                          4677
Name: incident, dtype: object
```

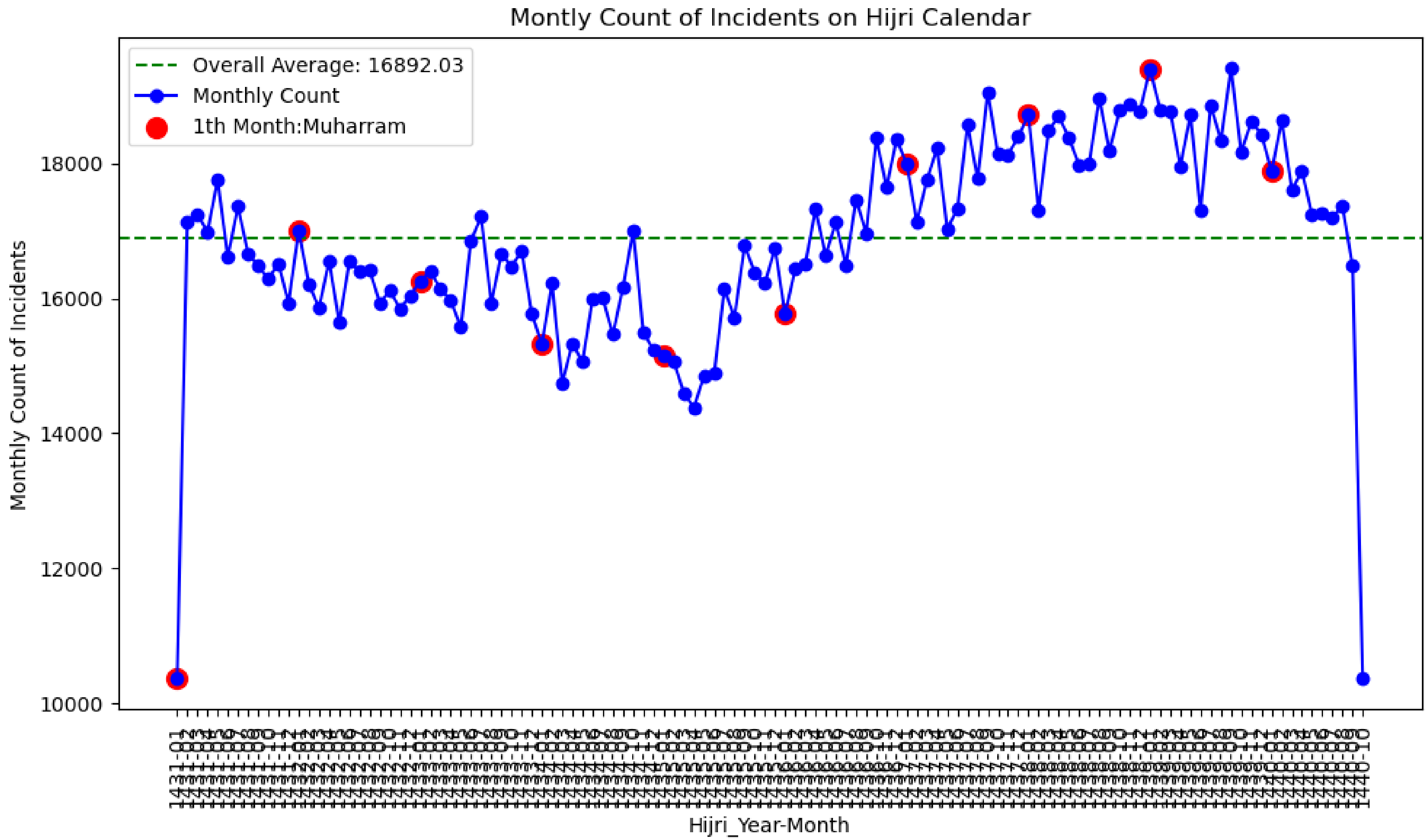
In [90]:
```
other_days_incidents_desc
```

Out[90]:
```
count                      1941394
unique                         140
top       BATTERY - SIMPLE ASSAULT
freq                        175757
Name: incident, dtype: object
```

Los Angeles dataset encompasses 140 distinct incident types. During the first ten days of Muharram, crimes were committed across 122 incident categories, with 34 of these categories experiencing incident counts exceeding the annual averages.

In [91]:

```
monthly_count_plot()
```



SAMPLE DATA-3: KANSAS CITY CRIME DATA_2009-2016

https://data.world/data-society/kansas-city-crime-data

In [92]:
```
df2 = pd.read_csv("KCPD_Crime_Data_2009.csv")
df3 = pd.read_csv("KCPD_Crime_Data_2010.csv")
df4 = pd.read_csv("KCPD_Crime_Data_2011.csv")
df5 = pd.read_csv("KCPD_Crime_Data_2012.csv")
df6 = pd.read_csv("KCPD_Crime_Data_2013.csv")
df7 = pd.read_csv("KCPD_Crime_Data_2014.csv")
df8 = pd.read_csv("KCPD_Crime_Data_2015.csv")
df9 = pd.read_csv("KCPD_Crime_Data_2016.csv")
```

In [93]:
```
frames = [df2, df3, df4, df5, df6, df7, df8, df9]
df = pd.concat(frames)
```

In [94]:
```
df.head()
```

Out[94]:

	Report_No	Reported_Date	Reported Time	From_Date	From Time	To_Date	To Time	Offense	IBRS	Description	...	Involvement	Race	Sex	Age
0	70059279	10/06/2009 12:00:00 AM	3:24	10/05/2009 12:00:00 AM	22:56	10/05/2009 12:00:00 AM	23:10	1850	35B	Possession of Drug E	...	ARR	B	F	28.0
1	80005443	02/05/2009 12:00:00 AM	11:45	01/22/2008 12:00:00 AM	12:00	NaN	NaN	121	09C	Justifiable Homicide	...	SUS	W	M	27.0
2	80019629	06/18/2009 12:00:00 AM	22:50	06/18/2009 12:00:00 AM	21:15	NaN	NaN	1849	35A	Possession/Sale/Dist	...	ARR	W	M	22.0
3	70060962	01/28/2009 12:00:00 AM	18:44	01/28/2009 12:00:00 AM	18:44	NaN	NaN	1352	280	Stolen Property OFFE	...	VIC	U	U	NaN
4	80005443	02/05/2009 12:00:00 AM	11:45	01/22/2008 12:00:00 AM	12:00	NaN	NaN	121	09C	Justifiable Homicide	...	SUS	W	M	28.0

5 rows × 28 columns

In [95]:
```
df.duplicated().value_counts()
```

Out[95]:
```
False    1007956
True        3518
dtype: int64
```

In [96]:
```
df = df.drop_duplicates()
```

In [97]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1007956 entries, 0 to 110891
Data columns (total 28 columns):
 #   Column              Non-Null Count    Dtype
---  ------              --------------    -----
 0   Report_No           1007956 non-null  int64
 1   Reported_Date       1007956 non-null  object
 2   Reported Time       897459 non-null   object
 3   From_Date           1006558 non-null  object
 4   From Time           895625 non-null   object
 5   To_Date             422549 non-null   object
 6   To Time             378834 non-null   object
 7   Offense             1007956 non-null  int64
 8   IBRS                999468 non-null   object
 9   Description         1007956 non-null  object
 10  Beat                1006287 non-null  object
 11  Address             883108 non-null   object
 12  City                883108 non-null   object
 13  Zip Code            976536 non-null   float64
 14  Rep_Dist            1005055 non-null  object
 15  Area                1005028 non-null  object
 16  DVFlag              1007956 non-null  object
 17  Invl_No             1007956 non-null  int64
 18  Involvement         1007956 non-null  object
 19  Race                875255 non-null   object
 20  Sex                 875255 non-null   object
 21  Age                 582586 non-null   float64
 22  Location 1          1007262 non-null  object
 23  Firearm Used Flag   775917 non-null   object
 24  Firearm Used Flag   232039 non-null   object
 25  Reported_Time       110497 non-null   object
 26  From_Time           110180 non-null   object
 27  To_Time             39925 non-null    object
dtypes: float64(2), int64(3), object(23)
memory usage: 223.0+ MB
```

In [98]:

```
df["Description"].value_counts()[:15]
```

Out[98]:
```
Burglary - Residence    93609
Property Damage         87921
Stealing From Auto      68821
Non Agg Assault Dome    63608
Auto Theft              61302
Stealing Auto Parts/    52498
Non Aggravated Assau    52272
Misc Violation          51407
Stealing Shoplifting    46401
Stealing from Buildi    43149
Stealing All Other      41074
Aggravated Assault (    40332
Possession/Sale/Dist    34013
Armed Robbery           30324
Trespassing             23800
Name: Description, dtype: int64
```

In [99]:
```
df = df.rename(columns = {'Reported_Date':'date'})
df = df.rename(columns = {'Description':'incident'})
```

In [100…
```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [101…
```
df.date.min(), df.date.max()
```

Out[101]:
```
('2009-01-01', '2016-11-06')
```

In [102…
```
df = df.iloc[:, [1,9]]
# df.to_csv("Kansas.csv", index=False)
```

In [103…
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[103]:
```
count    2858
mean      352
std        58
min        62
25%       316
50%       354
75%       392
95%       443
98%       468
99%       484
max       539
Name: date, dtype: int32
```

In [104…
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```

In [105…
```
df.date.nunique()
```
Out[105]:
```
2858
```

As seen below, our dataset spans a total of 2867 days. During this period, incidents occurred on 2858 days, while there were no records of incidents on the remaining 9 days.

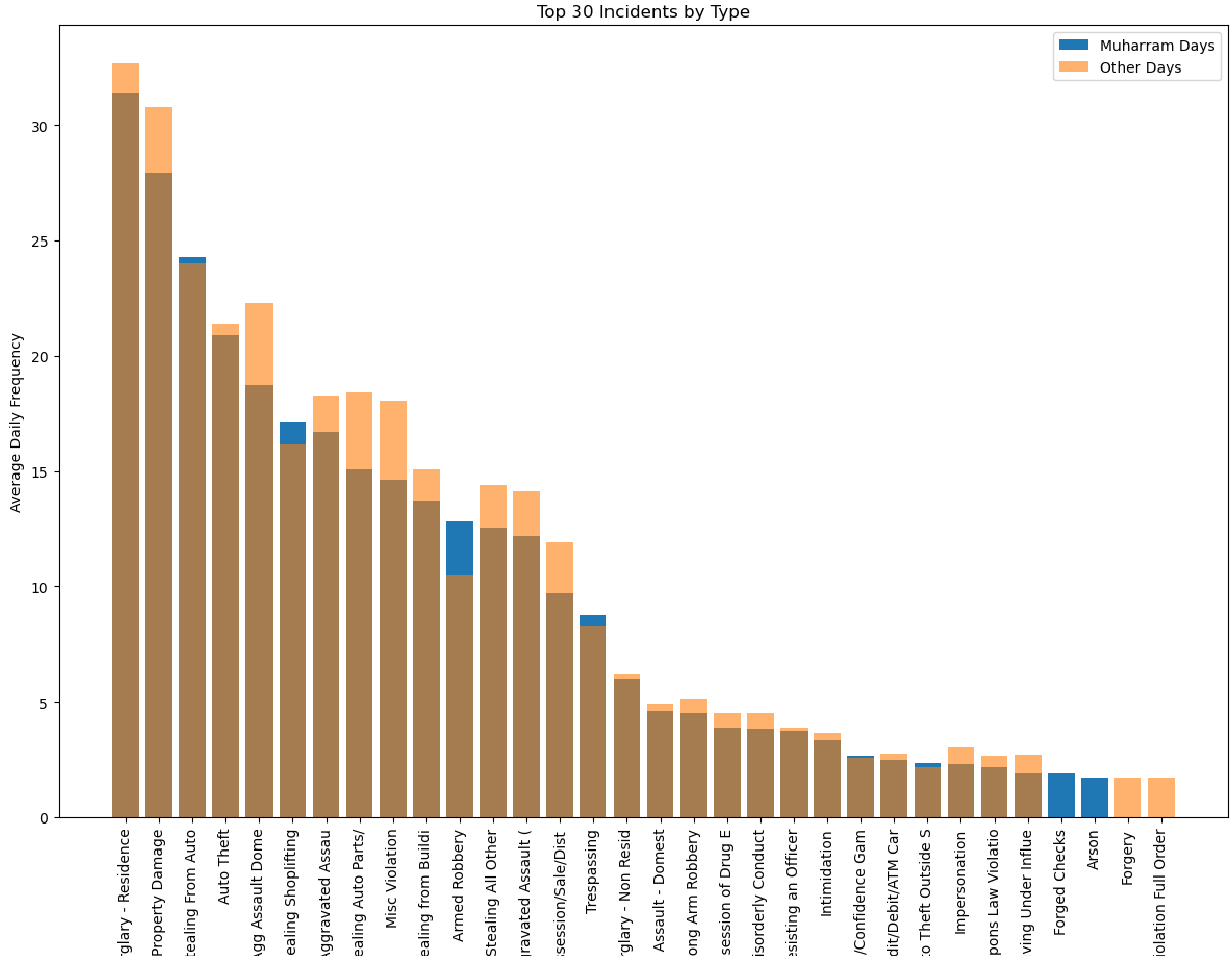
In [106…
```
muharrem_10_days (df)
```

```
Total number of days: 2867
---------------------
Total number of cases: 1007956
----------------------
Average Daily Case Count: 351.57
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2010    136056
2009    132535
2012    130290
2011    128072
2013    124732
2014    124232
2015    121542
2016    110497
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1431    131791
1430    128579
1433    126531
1434    126102
1432    124047
1437    123207
1435    117771
1436    117002
1438     12926
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 324.0444
------------------------------------------------------------
Average case count in other days: 352.4638
---------------------------------
Ratio of Muharram cases to other cases: 0.9194
---------------------------------------
```

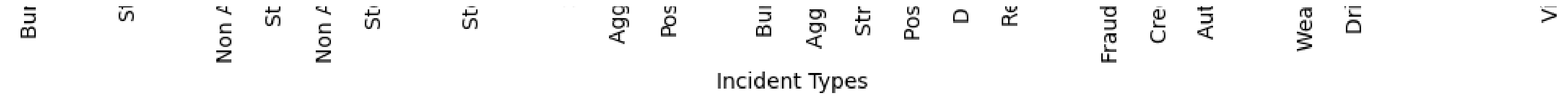
We observe a -8.06% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [107…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
5
10
15
20
25
30
rglary - Residence
Property Damage
tealing From Auto
Auto Theft
Agg Assault Dome
ealing Shoplifting
Aggravated Assau
ealing Auto Parts/
Misc Violation
ealing from Buildi
Armed Robbery
Stealing All Other
ravated Assault (
ssession/Sale/Dist
Trespassing
rglary - Non Resid
Assault - Domest
ong Arm Robbery
session of Drug E
isorderly Conduct
esisting an Officer
Intimidation
/Confidence Gam
dit/Debit/ATM Car
o Theft Outside S
Impersonation
pons Law Violatio
ving Under Influe
Forged Checks
Arson
Forgery
iolation Full Order



In [108…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[108]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
stolen/recoverd auto	4	4	1.000000
Harassment	2	4	0.500000
auto theft	2	4	0.500000
aggravated assault	2	4	0.500000
Agg Assault Dome	4	10	0.400000
AUTO THEFT	2	6	0.333300
Alien	3	10	0.300000
c	2	8	0.250000
Attempt Suicide by G	1	9	0.111100
False Bomb Report/Bo	2	29	0.069000
Suicide by Other Mea	1	17	0.058800
Stalking	4	80	0.050000
Forged Checks	173	3494	0.049500
Impersonation - NOT	26	540	0.048100
Kidnapping/Abduction	40	874	0.045800
Sex Off Indecent Con	31	682	0.045500
Stealing Purse Snatc	50	1154	0.043300
Passing Bad Checks	14	325	0.043100
Dangerous and Concea	8	189	0.042300
Stealing Pickpocket	45	1074	0.041900
Statutory Rape	29	703	0.041300
Pornography	10	249	0.040200
HOMICIDE	2	52	0.038500
Counterfeiting	33	862	0.038300

	muharrem incidents	all incidents	muharrem incidents/total incidents
Armed Robbery	1157	30324	0.038200
Interference with Cu	27	732	0.036900
Arson	155	4311	0.036000
Attempt Suicide by H	4	112	0.035700
Arson with Fire Bomb	5	142	0.035200
Attempt Suicide by O	8	228	0.035100

In [109…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[109]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
Stealing From Auto	2184	68821	0.0317
Stealing Shoplifting	1545	46401	0.0333
Armed Robbery	1157	30324	0.0382
Trespassing	787	23800	0.0331
Fraud/Confidence Gam	241	7380	0.0327
Auto Theft Outside S	212	6255	0.0339
Forged Checks	173	3494	0.0495
Arson	155	4311	0.0360
Liquor Law Violaton	71	2130	0.0333
Stealing Purse Snatc	50	1154	0.0433
Stealing Pickpocket	45	1074	0.0419
Kidnapping/Abduction	40	874	0.0458
Counterfeiting	33	862	0.0383
Sex Off Indecent Con	31	682	0.0455
Statutory Rape	29	703	0.0413
Interference with Cu	27	732	0.0369
Impersonation - NOT	26	540	0.0481
Passing Bad Checks	14	325	0.0431
Pornography	10	249	0.0402
Dangerous and Concea	8	189	0.0423
Attempt Suicide by O	8	228	0.0351
Loitering	7	219	0.0320
Arson with Fire Bomb	5	142	0.0352
Stalking	4	80	0.0500

	muharrem incidents	all incidents	muharrem incidents/total incidents
stolen/recoverd auto	4	4	1.0000
Agg Assault Dome	4	10	0.4000
Hit and Run of a Per	4	121	0.0331
Attempt Suicide by H	4	112	0.0357
Alien	3	10	0.3000
Harassment	2	4	0.5000
HOMICIDE	2	52	0.0385
False Bomb Report/Bo	2	29	0.0690
aggravated assault	2	4	0.5000
auto theft	2	4	0.5000
c	2	8	0.2500
AUTO THEFT	2	6	0.3333
Attempt Suicide by G	1	9	0.1111
Suicide by Other Mea	1	17	0.0588

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [110…
```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```
Out[110]:
```
(259, 38)
```

In [111…
```
muharrem_incidents_desc
```
Out[111]:
```
count                  29164
unique                    96
top       Burglary - Residence
freq                    2827
Name: incident, dtype: object
```

In [112…
```
other_days_incidents_desc
```

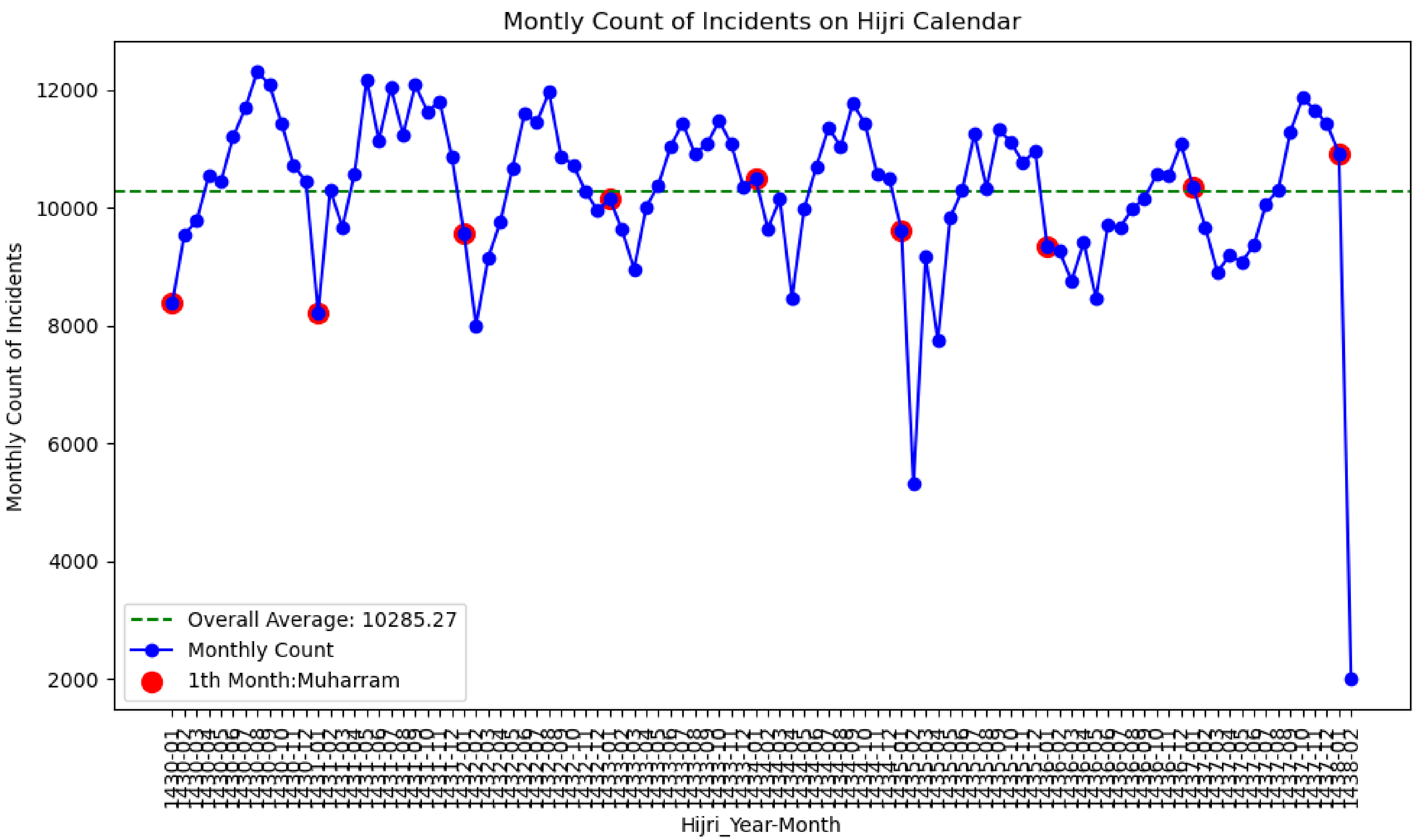
Out[112]:
```
count                     978792
unique                       258
top       Burglary - Residence
freq                       90782
Name: incident, dtype: object
```

In [113…
```
muharrem_dominant_incidents.count()[0]
```

Out[113]:
```
38
```

Kansas dataset encompasses 259 distinct incident types. During the first ten days of Muharram, crimes were committed across 96 incident categories, with 38 of these categories experiencing incident counts exceeding the annual averages.

In [114…
```
monthly_count_plot()
```



SAMPLE DATA-4: DETROIT CRIME INCIDENTS_2009-2016

https://data.world/detroit/dpd-crime-incidents-2009-2016

In [115…

```
df = pd.read_csv("Detroit_crime-incidents-2009-2016.csv", index_col=0, low_memory=False)
df
```

Out[115]:

ROWNUM	CASEID	CRIMEID	CRNO	ADDRESS	CATEGORY	OFFENSEDESCRIPTION	STATEOFFENSEFILECLASS	INCIDENTDATE	HOUR	S
1	1099487	1321797	0910020373.1	18000 WEXFORD	MISCELLANEOUS	MISCELLANEOUS - GENERAL NON-CRIMINAL	99009.0	01/01/2009	0	110
2	1117507	1344185	0911060289.1	00 UNKNOWN	MISCELLANEOUS	MISCELLANEOUS - GENERAL NON-CRIMINAL	99009.0	01/01/2009	0	N
3	985415	1181882	0902190512.1	02000 CALVERT	MISCELLANEOUS	MISCELLANEOUS - ABANDONED VEHICLE	99009.0	01/01/2009	0	100
4	986019	1182632	0902200294.1	00 W GRAND BLVD AND W FORT	MISCELLANEOUS	MISCELLANEOUS - GENERAL NON-CRIMINAL	99009.0	01/01/2009	0	41
5	996883	1195867	0903170149.1	12500 CONNER	LARCENY	LARCENY - FROM BUILDING (INCLUDES LIBRARY, OFF...	23003.0	01/01/2009	0	90
...	...	...	...	...	...	...	...	...	...	
1150376	2104068	2593922	1606250233.1	00 WOODWARD AND WEBB	DANGEROUS DRUGS	COCAINE -POSSESS	35001.0	06/25/2016	17	30
1150499	2104173	2594059	1606260030.1	20100 STANSBURY	MISCELLANEOUS	MISCELLANEOUS - GENERAL ASSISTANCE	99008.0	06/26/2016	1	120
1150527	2104193	2594076	1606260049.1	12100 FORRER	BURGLARY	BURGLARY - BURGLARY - FORCED ENTRY - RESIDENCE	22001.0	06/26/2016	3	60
1150614	2104329	2594236	1606260182.1	00 PURITAN GREENFIELD	MISCELLANEOUS	MISCELLANEOUS - IMPOUNDED VEHICLE	99009.0	06/26/2016	12	20
1150712	2104430	2594364	1606260281.1	15100 LAHSER	ASSAULT	ASSAULT AND BATTERY/SIMPLE ASSAULT	13001.0	06/26/2016	20	60

In [116…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1151019 entries, 1 to 1150712
Data columns (total 17 columns):
 #   Column                Non-Null Count    Dtype
---  ------                --------------    -----
 0   CASEID                1151019 non-null  int64
 1   CRIMEID               1151019 non-null  int64
 2   CRNO                  1151019 non-null  object
 3   ADDRESS               1151019 non-null  object
 4   CATEGORY              1151019 non-null  object
 5   OFFENSEDESCRIPTION    1150959 non-null  object
 6   STATEOFFENSEFILECLASS 1150959 non-null  float64
 7   INCIDENTDATE          1151019 non-null  object
 8   HOUR                  1151019 non-null  int64
 9   SCA                   1142973 non-null  float64
 10  PRECINCT              1142973 non-null  float64
 11  COUNCIL               1127234 non-null  object
 12  NEIGHBORHOOD          1136581 non-null  object
 13  CENSUSTRACT           1072395 non-null  float64
 14  LON                   1151018 non-null  float64
 15  LAT                   1151018 non-null  float64
 16  LOCATION              1151019 non-null  object
dtypes: float64(6), int64(3), object(8)
memory usage: 158.1+ MB
```

In [117…

```
df.duplicated().value_counts()
```

Out[117]:

```
False    1151019
dtype: int64
```

In [118…

```
df["CATEGORY"].value_counts()
```

Out[118]:

```
MISCELLANEOUS              174164
ASSAULT                    141608
LARCENY                    133666
BURGLARY                   109073
DAMAGE TO PROPERTY          93781
STOLEN VEHICLE              89445
AGGRAVATED ASSAULT          72577
MURDER/INFORMATION          71694
TRAFFIC                     63890
FRAUD                       40357
ROBBERY                     39292
DANGEROUS DRUGS             28081
ESCAPE                      13617
WEAPONS OFFENSES            13268
OBSTRUCTING JUDICIARY       11345
DISORDERLY CONDUCT           9252
OUIL                         8557
ARSON                        6891
SOLICITATION                 3872
OBSTRUCTING THE POLICE       3268
STOLEN PROPERTY              3249
OTHER                        3135
OTHER BURGLARY               3065
HOMICIDE                     2541
FAMILY OFFENSE               1997
FORGERY                      1609
KIDNAPING                    1588
RUNAWAY                      1457
VAGRANCY (OTHER)             1138
EXTORTION                    1099
LIQUOR                        782
ENVIRONMENT                   563
EMBEZZLEMENT                  346
CIVIL                         176
IMMIGRATION                   159
JUSTIFIABLE HOMICIDE          136
NEGLIGENT HOMICIDE             67
KIDNAPPING                     64
OBSCENITY                      46
GAMBLING                       39
BRIBERY                        19
DRUNKENNESS                    19
MISCELLANEOUS ARREST           17
TRAFFIC OFFENSES                5
```

```
MILITARY                       4
ABORTION                       1
Name: CATEGORY, dtype: int64
```

In [119…

```
df["OFFENSEDESCRIPTION"].value_counts()
```

Out[119]:

```
ASSAULT AND BATTERY/SIMPLE ASSAULT                            98560
VEHICLE THEFT                                                 82957
INFORMATION                                                   71694
DAMAGE TO PROPERTY - PRIVATE PROPERTY                         68202
BURGLARY - BURGLARY - FORCED ENTRY - RESIDENCE                67025
                                                              ...
ARSON - BUSINESS - DEFRAUD INSURER                                1
TRAFFIC VIOLATIONS - ILLEGAL TOWING EQUIPMENT                     1
PERMITTED PERSON UNDER THE INFLUENCE OF DRUGS TO OPERATE          1
TRAFFIC VIOLATIONS - DEFECTIVE OR IMPROPER BRAKES                 1
ACCIDENTS, ALL OTHER NON-CRIMINAL - AIRCRAFT                      1
Name: OFFENSEDESCRIPTION, Length: 584, dtype: int64
```

In [120…

```
df = df.rename(columns = {'INCIDENTDATE':'date'})
df = df.rename(columns = {'CATEGORY':'incident'})
```

In [121…

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [122…

```
df.date.min(), df.date.max()
```

Out[122]:

```
('2009-01-01', '2016-06-28')
```

In [123…

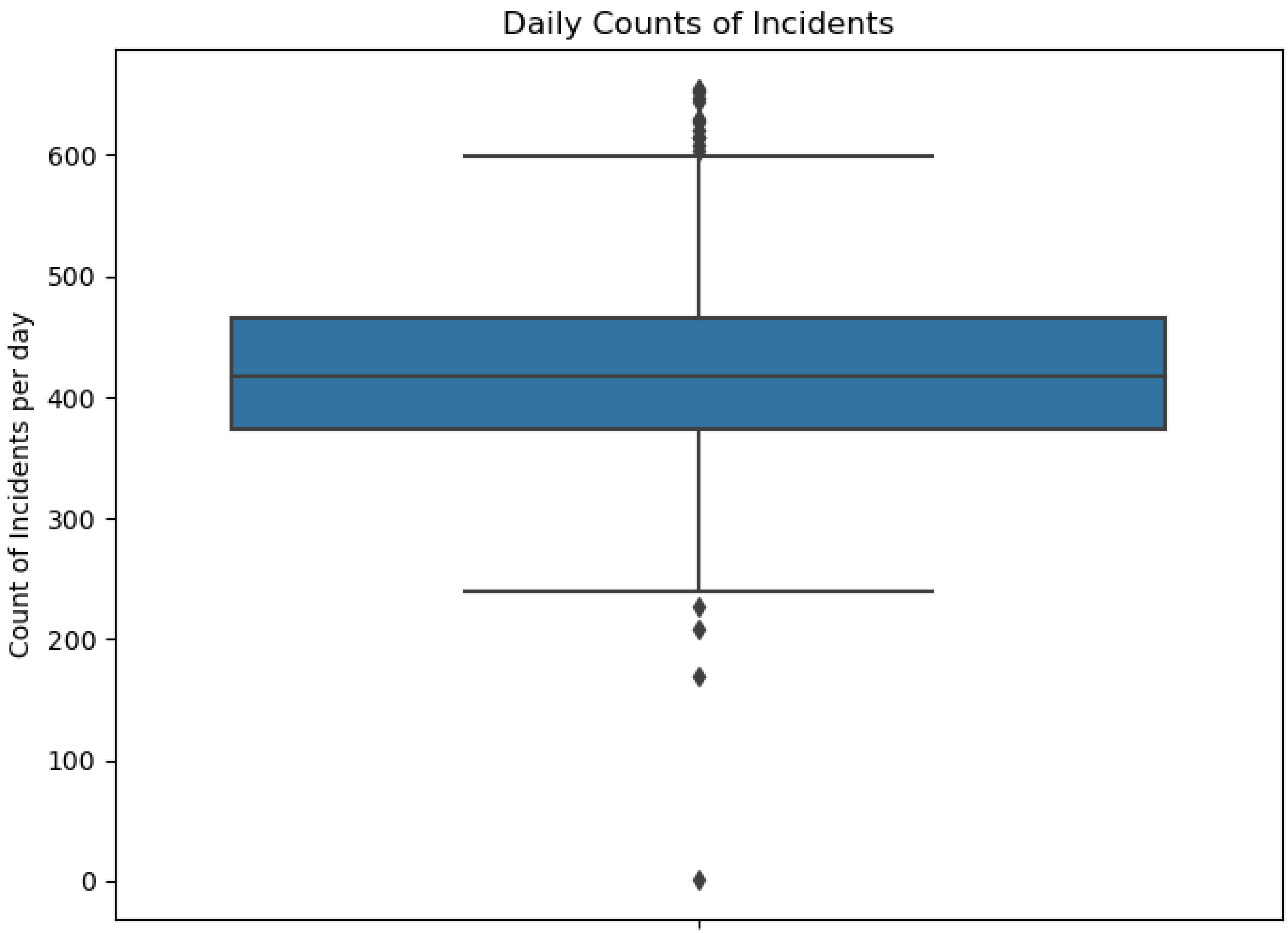
```
df = df.iloc[:, [4,7]]
# df.to_csv("Detroit.csv", index=False)
```

In [124…

```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[124]:
```
count    2736
mean      420
std        68
min         1
25%       374
50%       417
75%       465
95%       539
98%       565
99%       584
max       655
Name: date, dtype: int32
```

In [125…
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [126…

```
df.date.nunique()
```

Out[126]:

```
2736
```

As seen below, our dataset spans a total of 2736 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

In [127…

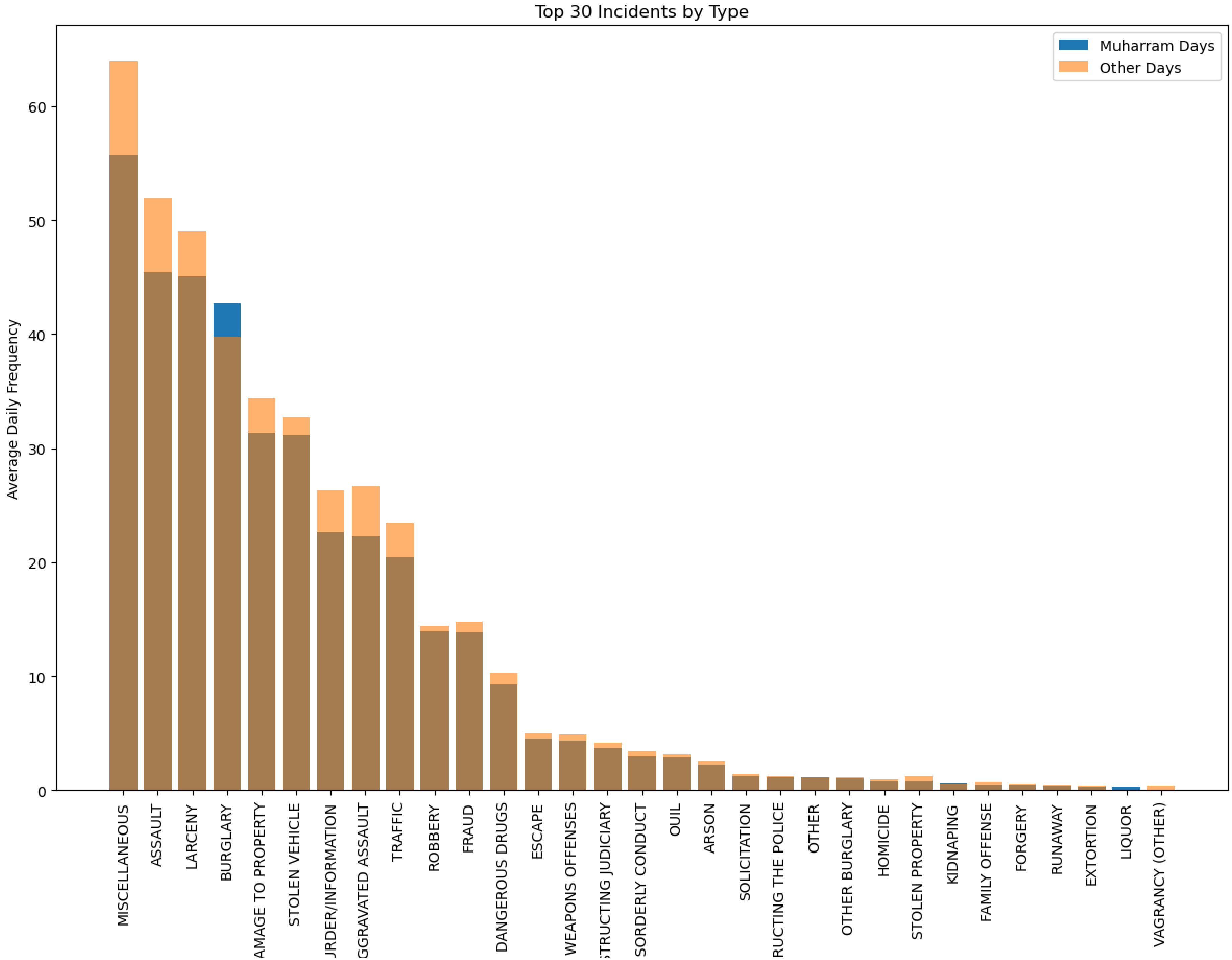
```
muharrem_10_days(df)
```

```
Total number of days: 2736
---------------------
Total number of cases: 1151019
----------------------
Average Daily Case Count: 420.69
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2009    181681
2010    170599
2011    157248
2012    156133
2013    146908
2015    137044
2014    136628
2016     64778
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1430    175663
1431    166733
1433    153006
1432    152251
1434    144344
1435    132697
1436    132393
1437     93932
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 384.4375
------------------------------------------------------------
Average case count in other days: 421.7861
-----------------------------------
Ratio of Muharram cases to other cases: 0.9115
---------------------------------------
```

We observe a -8.85% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [128…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
10
20
30
40
50
60
MISCELLANEOUS
ASSAULT
LARCENY
BURGLARY
AMAGE TO PROPERTY
STOLEN VEHICLE
JRDER/INFORMATION
GGRAVATED ASSAULT
TRAFFIC
ROBBERY
FRAUD
DANGEROUS DRUGS
ESCAPE
WEAPONS OFFENSES
STRUCTING JUDICIARY
SORDERLY CONDUCT
OUIL
ARSON
SOLICITATION
RUCTING THE POLICE
OTHER
OTHER BURGLARY
HOMICIDE
STOLEN PROPERTY
KIDNAPING
FAMILY OFFENSE
FORGERY
RUNAWAY
EXTORTION
LIQUOR
VAGRANCY (OTHER)

In [129…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[129]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
MISCELLANEOUS ARREST	1	17	0.058800
EMBEZZLEMENT	18	346	0.052000
KIDNAPING	53	1588	0.033400
LIQUOR	25	782	0.032000
BURGLARY	3417	109073	0.031300
HOMICIDE	73	2541	0.028700
ROBBERY	1115	39292	0.028400
OTHER	89	3135	0.028400
STOLEN VEHICLE	2495	89445	0.027900
OBSTRUCTING THE POLICE	91	3268	0.027800
OTHER BURGLARY	85	3065	0.027700
FRAUD	1111	40357	0.027500
LARCENY	3607	133666	0.027000
OUIL	231	8557	0.027000
DAMAGE TO PROPERTY	2505	93781	0.026700
DANGEROUS DRUGS	746	28081	0.026600
ESCAPE	361	13617	0.026500
ARSON	181	6891	0.026300
OBSTRUCTING JUDICIARY	295	11345	0.026000
WEAPONS OFFENSES	344	13268	0.025900
ASSAULT	3635	141608	0.025700
MISCELLANEOUS	4456	174164	0.025600
GAMBLING	1	39	0.025600
TRAFFIC	1633	63890	0.025600

	muharrem incidents	all incidents	muharrem incidents/total incidents
EXTORTION	28	1099	0.025500
DISORDERLY CONDUCT	235	9252	0.025400
MURDER/INFORMATION	1812	71694	0.025300
SOLICITATION	97	3872	0.025100
ENVIRONMENT	14	563	0.024900
AGGRAVATED ASSAULT	1782	72577	0.024600

In [130…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[130]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
BURGLARY	3417	109073	0.0313
KIDNAPING	53	1588	0.0334
LIQUOR	25	782	0.0320
EMBEZZLEMENT	18	346	0.0520
MISCELLANEOUS ARREST	1	17	0.0588

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [131…

```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[131]:

```
(46, 5)
```

In [132…

```
muharrem_incidents_desc
```

Out[132]:

```
count             30755
unique               40
top       MISCELLANEOUS
freq               4456
Name: incident, dtype: object
```

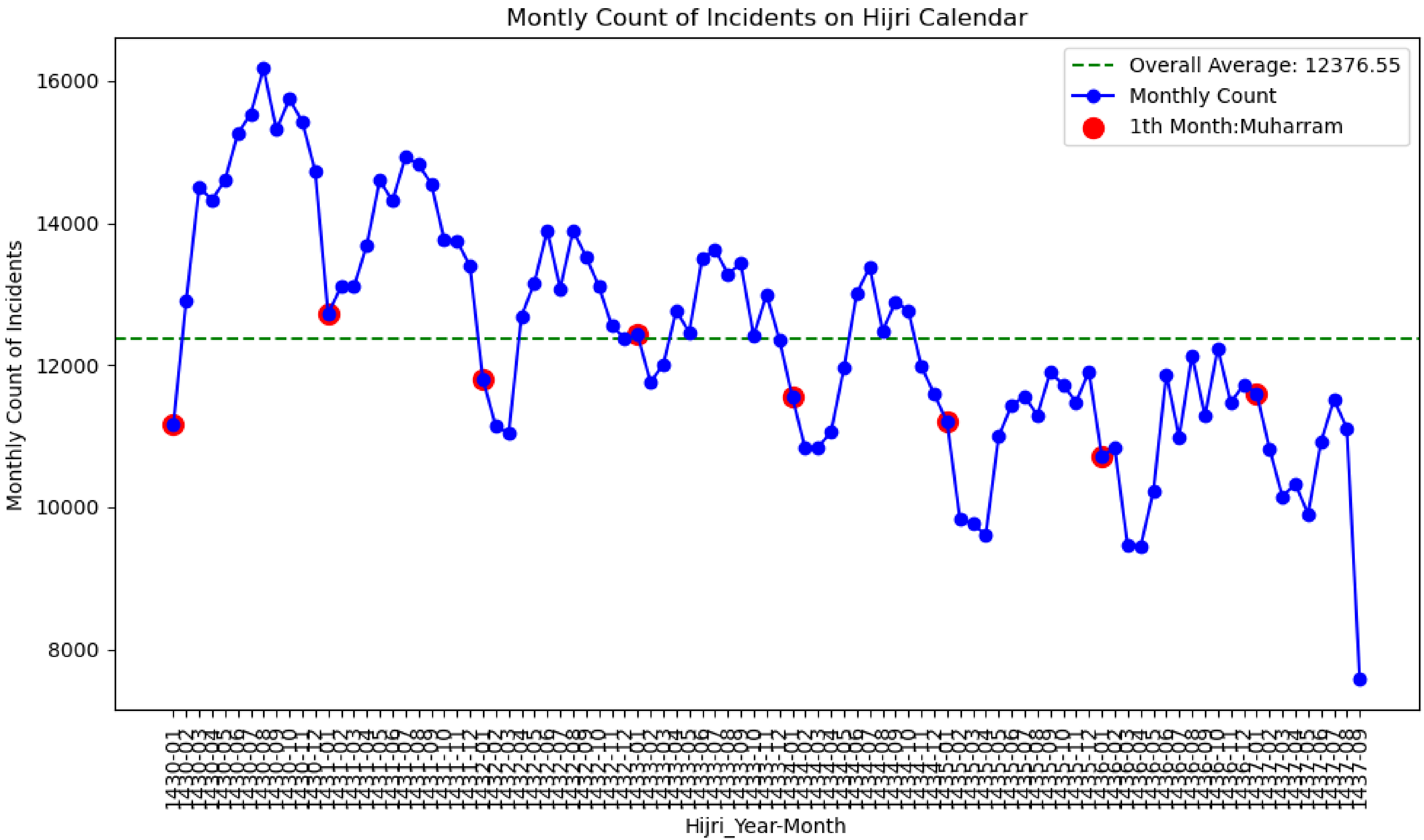
In [133…

```
other_days_incidents_desc
```

Out[133]:
```
count            1120264
unique                46
top        MISCELLANEOUS
freq              169708
Name: incident, dtype: object
```

Detroit dataset encompasses 46 distinct incident types. During the first ten days of Muharram, crimes were committed across 40 incident categories, with 5 of these categories experiencing incident counts exceeding the annual averages.

In [134…
```
monthly_count_plot()
```



SAMPLE DATA-5: DENVER CRIME DATASET_2019-2023

https://www.denvergov.org/opendata/dataset/city-and-county-of-denver-crime

In [135…

```
df = pd.read_csv("denver_crime.csv", index_col=0, encoding="Latin-1", low_memory=False)
df
```

Out[135]:

incident_id	offense_id	offense_code	offense_code_extension	offense_type_id	offense_category_id	first_occurrence_date	last_occurrence_date	re
202268791	202268791299900	2999	0	criminal-mischief-other	public-disorder	2/10/2022 2:50:00 AM	NaN	
2021387586	2021387586299900	2999	0	criminal-mischief-other	public-disorder	7/7/2021 9:02:00 PM	NaN	
2020641486	2020641486299900	2999	0	criminal-mischief-other	public-disorder	10/29/2020 1:30:00 AM	NaN	
2018612468	2018612468299900	2999	0	criminal-mischief-other	public-disorder	9/6/2018 5:00:00 PM	9/6/2018 11:00:00 PM	
2020293614	2020293614299900	2999	0	criminal-mischief-other	public-disorder	5/8/2020 5:00:00 AM	5/8/2020 6:30:00 PM	
...	...	...	...	...	...	...	...	
2023654815	2023654815260500	2605	0	theft-unauth-use-of-ftd	white-collar-crime	12/7/2023 4:45:00 PM	NaN	
2023652916	2023652916260900	2609	0	fraud-by-use-of-computer	white-collar-crime	12/4/2023 3:00:00 PM	12/5/2023 3:00:00 PM	
2023652471	2023652471260900	2609	0	fraud-by-use-of-computer	white-collar-crime	12/5/2023 4:30:00 PM	12/6/2023 8:00:00 AM	
2023652591	2023652591269903	2699	3	theft-of-services	larceny	11/6/2023 10:00:00 AM	12/2/2023 5:00:00 PM	
2023654247	2023654247269905	2699	5	pawn-broker-viol	all-other-crimes	12/7/2023 10:00:00 AM	12/7/2023 11:00:00 AM	

398091 rows × 19 columns

In [136…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 398091 entries, 202268791 to 2023654247
Data columns (total 19 columns):
 #   Column                  Non-Null Count   Dtype  
---  ------                  --------------   -----  
 0   offense_id              398091 non-null  int64  
 1   offense_code            398091 non-null  int64  
 2   offense_code_extension  398091 non-null  int64  
 3   offense_type_id         398091 non-null  object 
 4   offense_category_id     398091 non-null  object 
 5   first_occurrence_date   398091 non-null  object 
 6   last_occurrence_date    217265 non-null  object 
 7   reported_date           398091 non-null  object 
 8   incident_address        382236 non-null  object 
 9   geo_x                   382236 non-null  float64
 10  geo_y                   382236 non-null  float64
 11  geo_lon                 381975 non-null  float64
 12  geo_lat                 381975 non-null  float64
 13  district_id             398035 non-null  object 
 14  precinct_id             398091 non-null  int64  
 15  neighborhood_id         397405 non-null  object 
 16  is_crime                398091 non-null  int64  
 17  is_traffic              398091 non-null  int64  
 18  victim_count            398091 non-null  int64  
dtypes: float64(4), int64(7), object(8)
memory usage: 60.7+ MB
```

In [137…

```
df.duplicated().value_counts()
```

Out[137]:

```
False    398091
dtype: int64
```

In [138…

```
df["offense_type_id"].value_counts()
```

```
Out[138]: theft-of-motor-vehicle           56767
theft-items-from-vehicle         40388
theft-parts-from-vehicle         26244
criminal-mischief-mtr-veh        25687
theft-other                      23994
criminal-mischief-other          16485
criminal-trespassing             15035
assault-simple                   14280
theft-shoplift                   13096
theft-bicycle                     9889
burglary-residence-no-force       9057
burglary-business-by-force        8624
theft-from-bldg                   8149
weapon-unlawful-discharge-of      7881
aggravated-assault                7145
burglary-residence-by-force       6426
threats-to-injure                 5297
assault-dv                        5163
menacing-felony-w-weap            4962
drug-poss-paraphernalia           4533
disturbing-the-peace              4196
drug-methampetamine-possess       3874
public-order-crimes-other         3799
violation-of-restraining-order    3555
drug-pcs-other-drug               3390
robbery-street                    3344
aggravated-assault-dv             3223
violation-of-court-order          2686
liquor-possession                 2600
criminal-mischief-graffiti        2587
burglary-business-no-force        2495
sex-aslt-rape                     2465
robbery-business                  1999
police-false-information          1817
weapon-by-prev-offender-powpo     1770
fraud-by-telephone                1493
harassment                        1381
fraud-by-use-of-computer          1374
theft-of-services                 1362
sex-off-fail-to-register          1357
drug-heroin-possess               1317
drug-cocaine-possess              1293
police-interference               1292
drug-cocaine-sell                 1237
```

weapon-fire-into-occ-bldg	1227
drug-methampetamine-sell	1130
theft-stln-vehicle-trailer	1067
robbery-car-jacking	993
forgery-checks	895
theft-unauth-use-of-ftd	841
sex-aslt-fondle-adult-victim	794
agg-aslt-shoot	790
police-resisting-arrest	742
weapon-poss-illegal-dangerous	737
assault-police-simple	730
theft-fail-return-rent-veh	726
sex-aslt-non-rape	707
indecent-exposure	697
drug-opium-or-deriv-sell	683
burg-auto-theft-resd-no-force	673
weapon-other-viol	663
theft-from-mails	662
curfew	618
fraud-criminal-impersonation	615
harassment-dv	565
false-imprisonment	558
prostitution-engaging-in	538
public-fighting	518
harassment-sexual-in-nature	460
burglary-poss-of-tools	447
weapon-carrying-concealed	415
public-peace-other	412
agg-aslt-police-weapon	395
weapon-flourishing	394
homicide-other	387
contraband-into-prison	386
drug-heroin-sell	381
robbery-residence	375
harassment-stalking-dv	371
weapon-fire-into-occ-veh	356
burg-auto-theft-busn-w-force	319
weapon-carrying-prohibited	313
arson-other	309
theft-purse-snatch-no-force	300
extortion	297
drug-marijuana-possess	293
police-disobey-lawful-order	289
liquor-sell	285

drug-opium-or-deriv-possess	274
forgery-other	260
kidnap-dv	254
arson-vehicle	242
kidnap-adult-victim	242
stolen-property-buy-sell-rec	239
robbery-purse-snatch-w-force	239
sex-off-registration-viol	235
obstructing-govt-operation	228
robbery-bank	228
bomb-threat	197
drug-marijuana-sell	196
sex-aslt-rape-pot	181
drug-make-sell-other-drug	173
drug-synth-narcotic-sell	163
burg-auto-theft-resd-w-force	157
animal-cruelty-to	157
sex-aslt-non-rape-pot	156
property-crimes-other	155
arson-residence	151
theft-pick-pocket	151
illegal-dumping	150
burglary-vending-machine	150
fraud-nsf-closed-account	143
arson-business	130
burglary-safe	123
intimidation-of-a-witness	123
drug-fraud-to-obtain	122
reckless-endangerment	116
theft-embezzle	111
burg-auto-theft-busn-no-force	111
drug-marijuana-cultivation	110
other-enviornment-animal-viol	103
fireworks-possession	98
forgery-poss-of-forged-inst	97
window-peeping	93
drug-synth-narcotic-possess	92
police-obstruct-investigation	90
accessory-conspiracy-to-crime	78
drug-hallucinogen-possess	76
theft-stln-veh-const-eqpt	71
escape	69
violation-of-custody-order	65
drug-hallucinogen-sell	65

theft-of-rental-property	61
littering	61
forgery-counterfeit-of-obj	54
sex-asslt-sodomy-man-strng-arm	49
sex-aslt-w-object	46
bribery	45
disarming-a-peace-officer	43
drug-forgery-to-obtain	42
impersonation-of-police	40
theft-confidence-game	39
contraband-possession	39
police-making-a-false-rpt	38
drug-methamphetamine-mfr	38
pawn-broker-viol	38
fraud-identity-theft	38
theft-gas-drive-off	34
forgery-posses-forge-device	33
obscene-material-possess	32
forgery-poss-of-forged-ftd	31
homicide-family	30
explosive-incendiary-dev-pos	25
weapon-altering-serial-number	22
explosive-incendiary-dev-use	20
arson-public-building	20
aslt-agg-police-gun	19
sex-aslt-w-object-pot	13
drug-barbiturate-possess	11
prostitution-pimping	9
animal-poss-of-dangerous	9
probation-violation	8
parole-violation	7
drug-barbiturate-sell	7
drug-hallucinogen-mfr	6
wiretapping	5
obscene-material-mfr	5
weapon-unlawful-sale	5
loitering	5
homicide-police-by-gun	4
prostitution-aiding	4
altering-vin-number	4
money-laundering	4
gambling-gaming-operation	4
liquor-manufacturing	3
eavesdropping	2

```
homicide-conspiracy                 2
homicide-negligent                  2
escape-aiding                       2
riot-incite                         2
gambling-betting-wagering           1
bigamy                              1
liquor-other-viol                   1
theft-of-cable-services             1
drug-barbiturate-mfr                1
homicide-accessory-to               1
Name: offense_type_id, dtype: int64
```

In [139…

```
df["offense_category_id"].value_counts()
```

Out[139]:

```
theft-from-motor-vehicle        66632
public-disorder                 58617
auto-theft                      57905
larceny                         57788
all-other-crimes                47898
burglary                        28432
drug-alcohol                    22354
other-crimes-against-persons    21000
aggravated-assault              17703
robbery                          7178
white-collar-crime               6898
sexual-assault                   4411
arson                             852
murder                            423
Name: offense_category_id, dtype: int64
```

In [140…

```
df = df.rename(columns = {'first_occurrence_date':'date'})
df = df.rename(columns = {'offense_category_id':'incident'})
```

In [141…

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [142…

```
df.date.min(), df.date.max()
```

Out[142]:

```
('2018-01-02', '2023-12-07')
```

In [143…

```
df = df.iloc[:, [4,5]]
# df.to_csv("Denver.csv", index=False)
```

In [144…

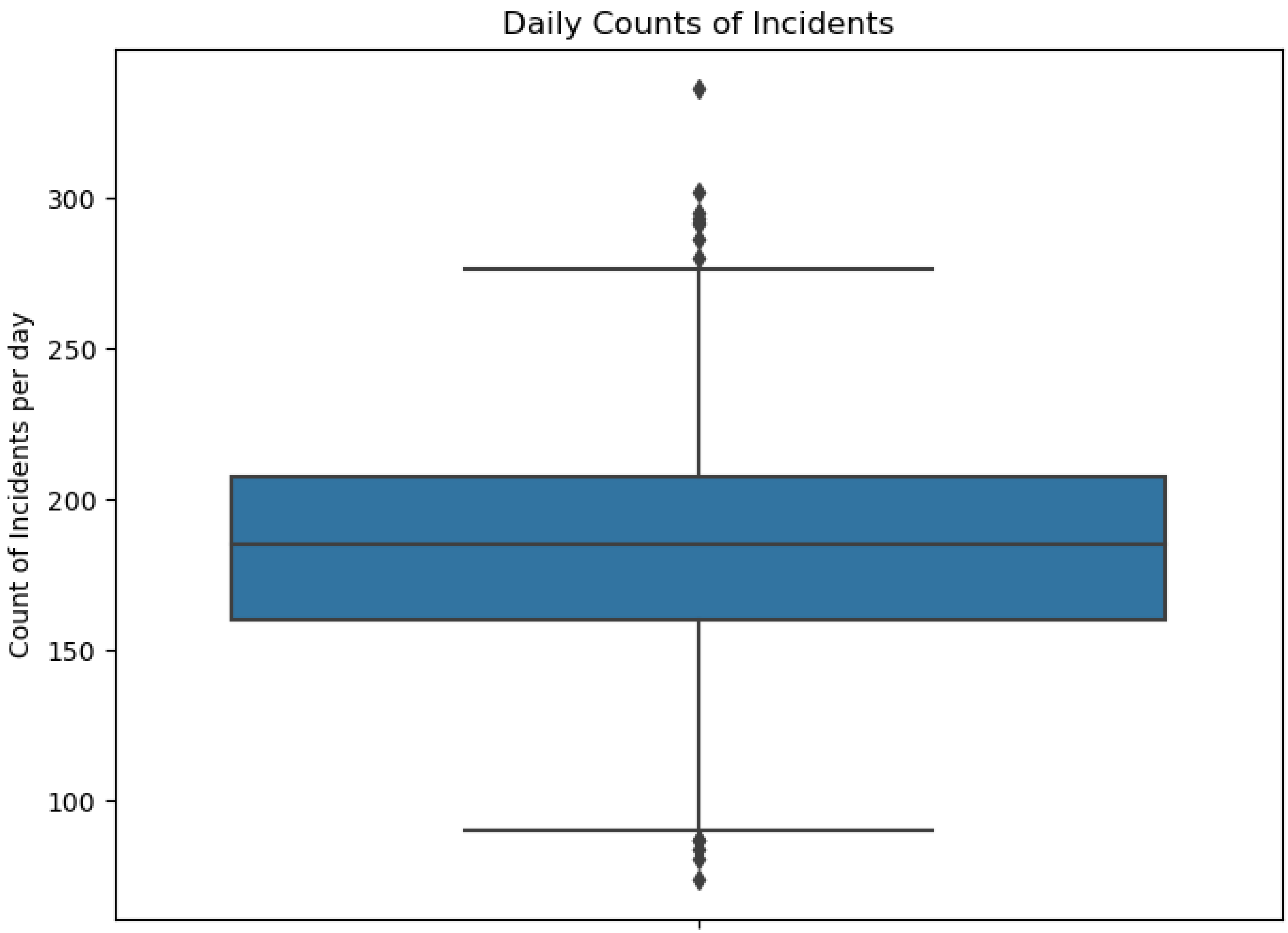
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[144]:

```
count    2166
mean      183
std        33
min        74
25%       160
50%       185
75%       207
95%       237
98%       250
99%       259
max       336
Name: date, dtype: int32
```

In [145…

```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [146…
```
df.date.nunique()
```

Out[146]:
```
2166
```

As seen below, our dataset spans a total of 2166 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

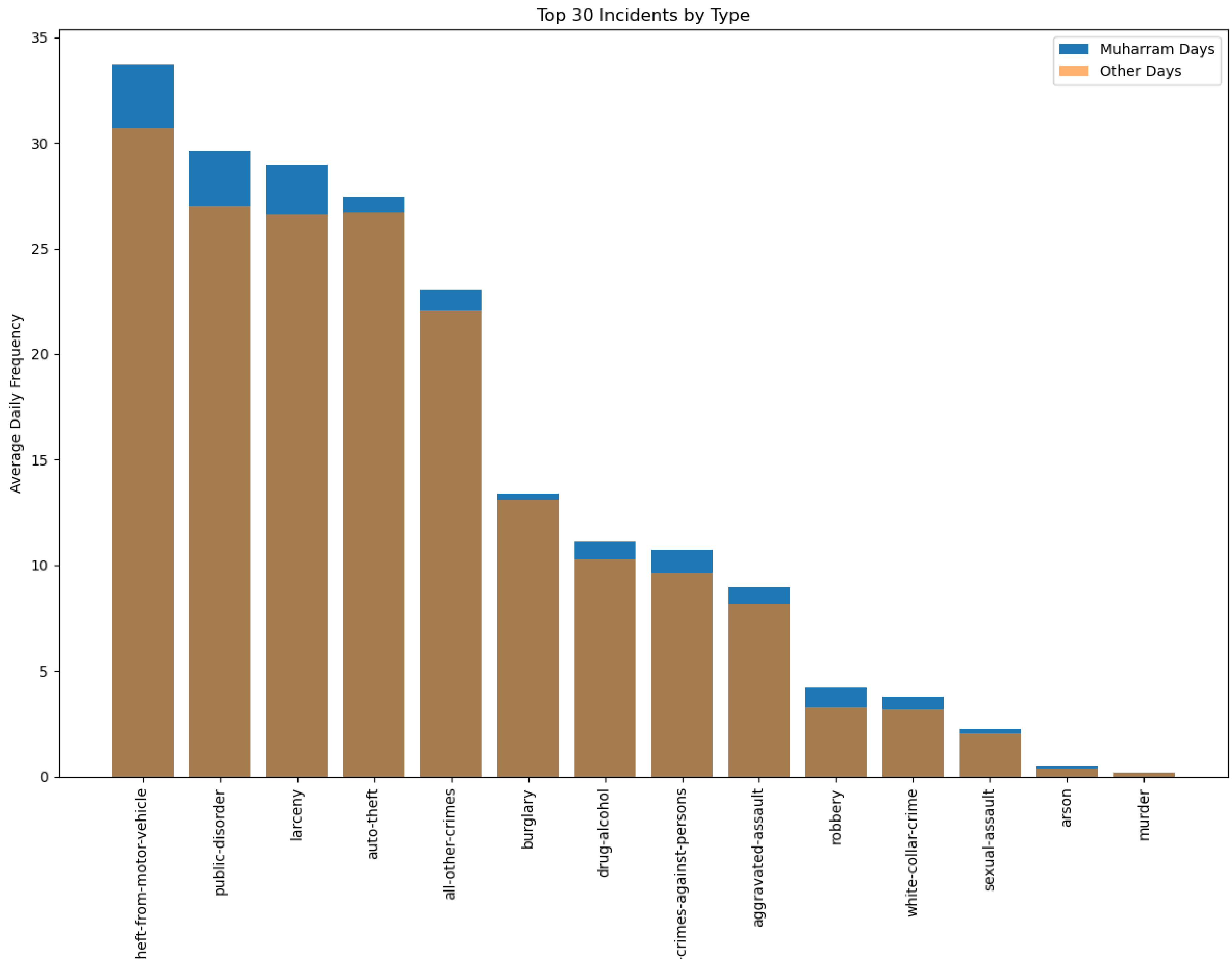
In [147…
```
muharrem_10_days(df)
```

```
Total number of days: 2166
---------------------
Total number of cases: 398091
----------------------
Average Daily Case Count: 183.79
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2022    76854
2021    72561
2023    71028
2020    63754
2019    57625
2018    56269
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1444    74148
1443    73197
1442    68855
1441    58223
1440    55247
1439    39212
1445    29209
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 197.8833
------------------------------------------------------------
Average case count in other days: 183.3894
-----------------------------------
Ratio of Muharram cases to other cases: 1.0790
---------------------------------------
```

We observe a 7.90% higher crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [148…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
35
30
25
20
15
10
5
0
heft-from-motor-vehicle
public-disorder
larceny
auto-theft
all-other-crimes
burglary
drug-alcohol
crimes-against-persons
aggravated-assault
robbery
white-collar-crime
sexual-assault
arson
murder

In [149…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[149]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
robbery	253	7178	0.035200
arson	29	852	0.034000
white-collar-crime	226	6898	0.032800
other-crimes-against-persons	643	21000	0.030600
sexual-assault	135	4411	0.030600
aggravated-assault	536	17703	0.030300
public-disorder	1777	58617	0.030300
theft-from-motor-vehicle	2021	66632	0.030300
larceny	1738	57788	0.030100
drug-alcohol	667	22354	0.029800
all-other-crimes	1384	47898	0.028900
auto-theft	1647	57905	0.028400
murder	12	423	0.028400
burglary	805	28432	0.028300

In [150…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[150]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
theft-from-motor-vehicle	2021	66632	0.0303
public-disorder	1777	58617	0.0303
larceny	1738	57788	0.0301
auto-theft	1647	57905	0.0284
all-other-crimes	1384	47898	0.0289
burglary	805	28432	0.0283
drug-alcohol	667	22354	0.0298
other-crimes-against-persons	643	21000	0.0306
aggravated-assault	536	17703	0.0303
robbery	253	7178	0.0352
white-collar-crime	226	6898	0.0328
sexual-assault	135	4411	0.0306
arson	29	852	0.0340
murder	12	423	0.0284

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [151…

```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[151]:

```
(14, 14)
```

In [152…

```
muharrem_incidents_desc
```

Out[152]:

```
count                         11873
unique                           14
top        theft-from-motor-vehicle
freq                           2021
Name: incident, dtype: object
```

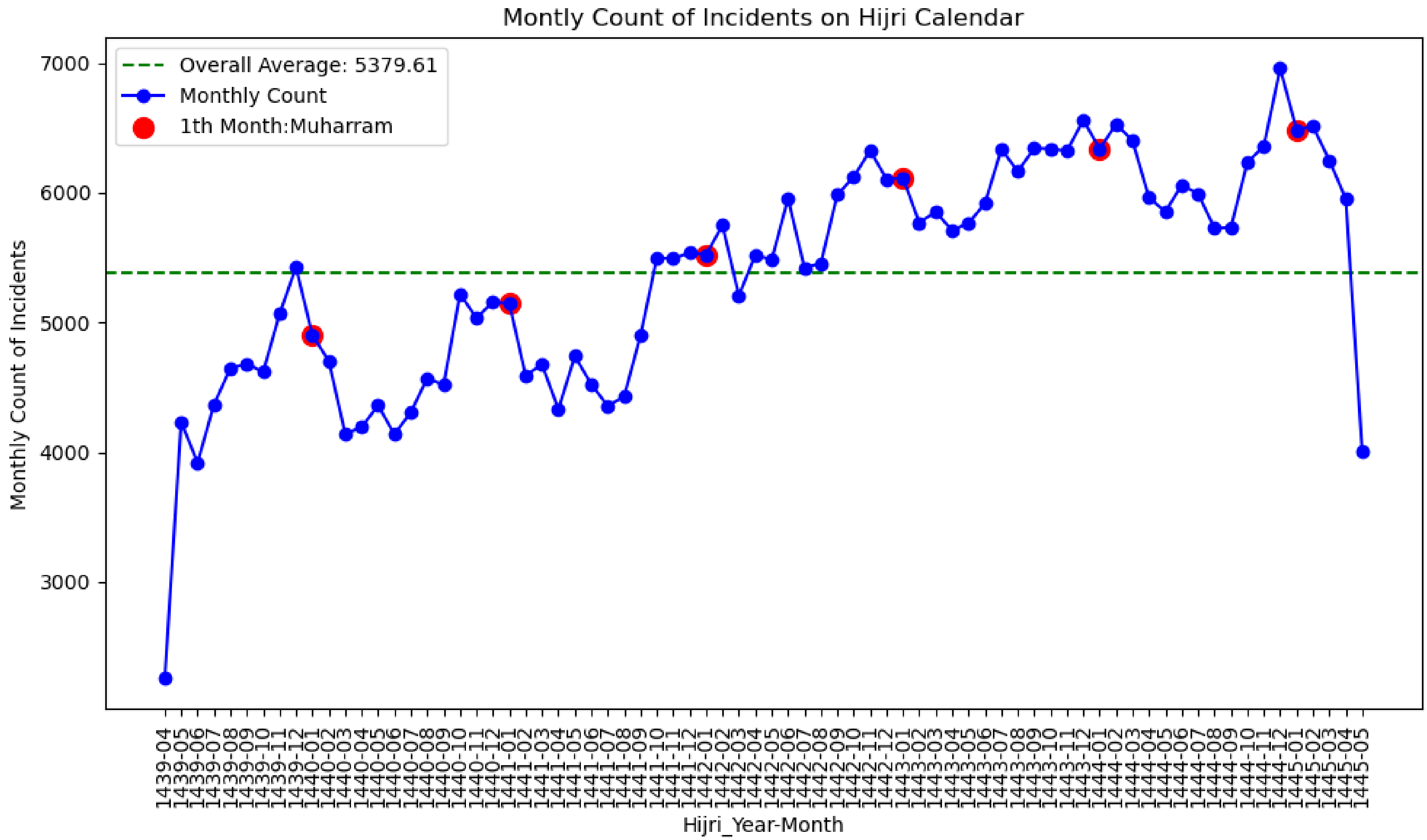
In [153…

```
other_days_incidents_desc
```

Out[153]:
```
count                       386218
unique                          14
top       theft-from-motor-vehicle
freq                         64611
Name: incident, dtype: object
```

Denver dataset encompasses 14 distinct incident types. During the first ten days of Muharram, crimes were committed across 14 incident categories, with 14 of these categories experiencing incident counts exceeding the annual averages.

In [154…
```
monthly_count_plot()
```



SAMPLE DATA-6: VANCOUVER CRIME DATASET_2003-2017

https://www.kaggle.com/datasets/wosaku/crime-in-vancouver

In [155…

```
df = pd.read_csv("crime_vancouver.csv", low_memory=False)
df
```

Out[155]:

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	40.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	15.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	45.0	9XX TERMINAL AVE	Strathcona	493906.50	5457452.47	49.269802	-123.083763
...	...	...	...	...	...	...	...	...	...	...	...	...
530647	Break and Enter Residential/Other	2017	3	3	9.0	16.0	31XX ADANAC ST	Hastings-Sunrise	497265.49	5458296.71	49.277420	-123.037595
530648	Mischief	2017	5	29	22.0	30.0	14XX E 7TH AVE	Grandview-Woodland	494533.97	5456824.97	49.264163	-123.075129
530649	Offence Against a Person	2017	4	13	NaN	NaN	OFFSET TO PROTECT PRIVACY	NaN	0.00	0.00	0.000000	0.000000
530650	Theft from Vehicle	2017	6	5	17.0	0.0	8XX HAMILTON ST	Central Business District	491487.85	5458385.78	49.278168	-123.117031
530651	Vehicle Collision or Pedestrian Struck (with I...	2017	6	6	17.0	38.0	13XX BLOCK PARK DR	Marpole	490204.00	5451444.00	49.215706	-123.134512

530652 rows × 12 columns

In [156…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 530652 entries, 0 to 530651
Data columns (total 12 columns):
 #   Column         Non-Null Count   Dtype  
---  ------         --------------   -----  
 0   TYPE           530652 non-null  object 
 1   YEAR           530652 non-null  int64  
 2   MONTH          530652 non-null  int64  
 3   DAY            530652 non-null  int64  
 4   HOUR           476290 non-null  float64
 5   MINUTE         476290 non-null  float64
 6   HUNDRED_BLOCK  530639 non-null  object 
 7   NEIGHBOURHOOD  474028 non-null  object 
 8   X              530652 non-null  float64
 9   Y              530652 non-null  float64
 10  Latitude       530652 non-null  float64
 11  Longitude      530652 non-null  float64
dtypes: float64(6), int64(3), object(3)
memory usage: 48.6+ MB
```

In [157…

```
df.duplicated().value_counts()
```

Out[157]:

```
False    481814
True      48838
dtype: int64
```

In [158…

```
df=df.drop_duplicates()
```

In [159…

```
df["TYPE"].value_counts()
```

Out[159]:

```
Theft from Vehicle                                        172699
Mischief                                                   70413
Break and Enter Residential/Other                          60862
Other Theft                                                52167
Theft of Vehicle                                           38418
Break and Enter Commercial                                 33845
Theft of Bicycle                                           25730
Vehicle Collision or Pedestrian Struck (with Injury)       21901
Offence Against a Person                                    5307
Vehicle Collision or Pedestrian Struck (with Fatality)       254
Homicide                                                     218
Name: TYPE, dtype: int64
```

In [ ]:
```
df['date'] = pd.to_datetime(df[['YEAR', 'MONTH', 'DAY']])
```

In [ ]:
```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [162…
```
df = df.rename(columns={'TYPE':'incident'})
```

In [163…
```
df.date.min(),df.date.max()
```

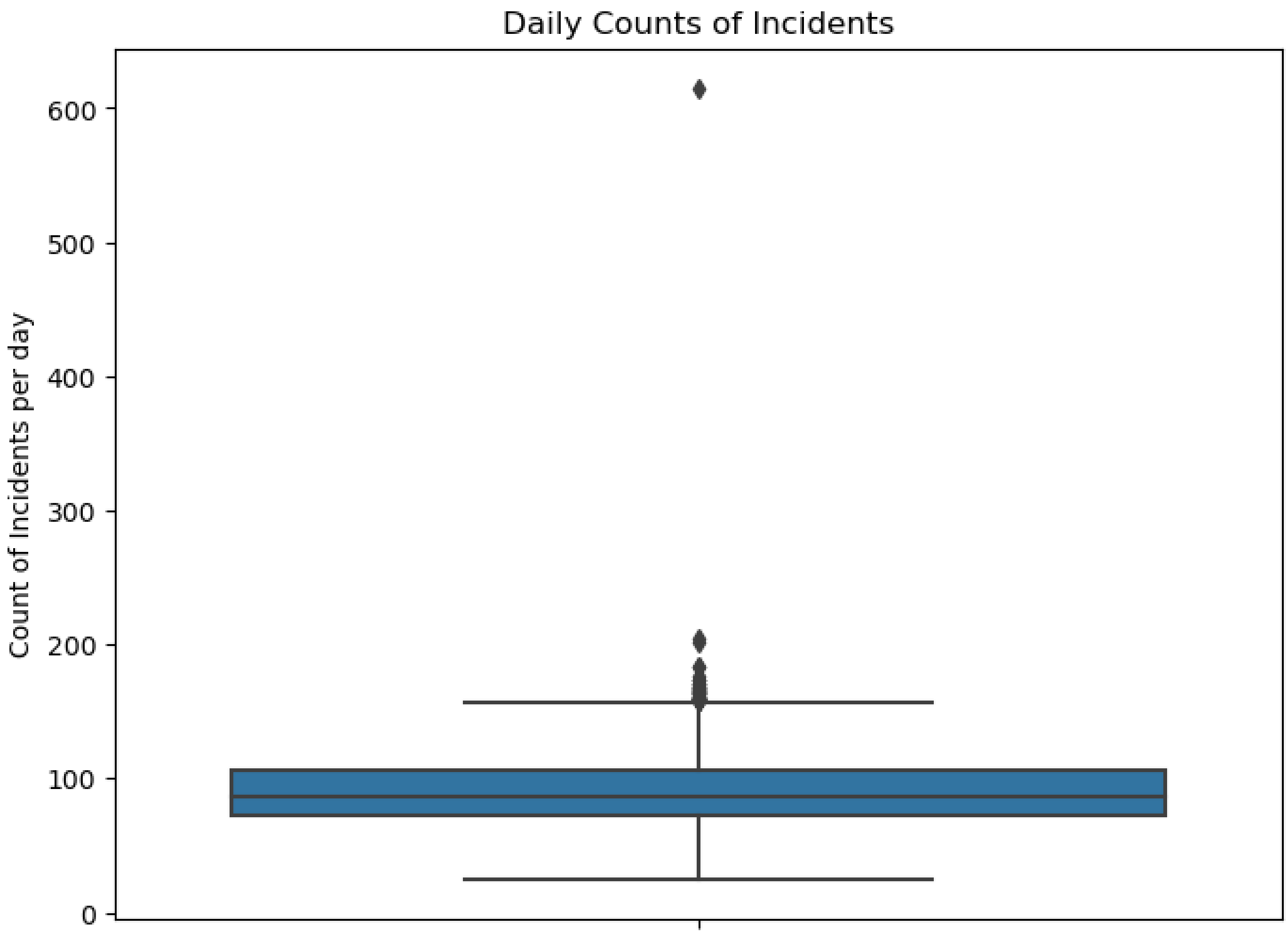
Out[163]:
```
('2003-01-01', '2017-07-13')
```

In [164…
```
df = df.iloc[:, [0,12]]
# df.to_csv("Vancouver.csv", index=False)
```

In [165…
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[165]:
```
count    5308
mean       90
std        25
min        25
25%        72
50%        87
75%       106
95%       137
98%       149
99%       157
max       615
Name: date, dtype: int32
```

In [166…
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```




```
In [167…: df.date.nunique()
```

Out[167]: 5308

As seen below, our dataset spans a total of 5308 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

```
In [168…: muharrem_10_days(df)
```

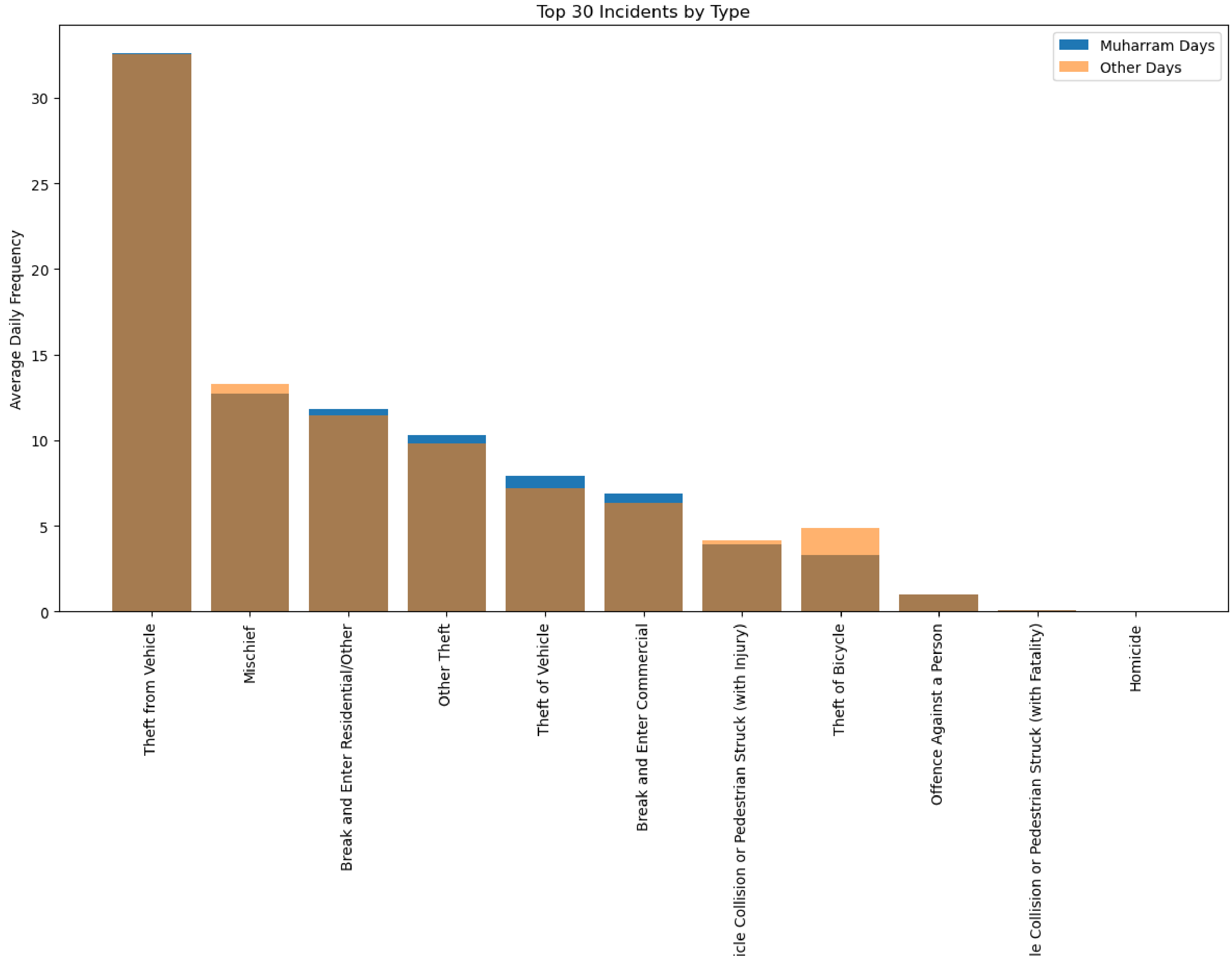
```
Total number of days: 5308
---------------------
Total number of cases: 481814
----------------------
Average Daily Case Count: 90.77
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2003    46785
2004    45841
2005    41257
2006    38336
2016    34991
2007    33645
2008    31554
2015    31491
2014    29866
2009    28652
2010    26322
2012    25802
2013    25758
2011    25068
2017    16446
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1424    44964
1425    44520
1426    40126
1427    36870
1437    34998
1428    32436
1429    30522
1436    29213
1435    28641
1430    27944
1431    25668
1433    24866
1434    24799
1432    24377
1438    24138
1423     7732
Name: Hijri_Date, dtype: int64
```

```
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 90.5267
------------------------------------------------------------
Average case count in other days: 90.7784
---------------------------------
Ratio of Muharram cases to other cases: 0.9972
---------------------------------------
```

We observe a -0.28% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [169…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```


Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
5
10
15
20
25
30
Theft from Vehicle
Mischief
Break and Enter Residential/Other
Other Theft
Theft of Vehicle
Break and Enter Commercial
icle Collision or Pedestrian Struck (with Injury)
Theft of Bicycle
Offence Against a Person
le Collision or Pedestrian Struck (with Fatality)
Homicide


```
In [170…]: # Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[170]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
Vehicle Collision or Pedestrian Struck (with Fatality)	8	254	0.031500
Theft of Vehicle	1186	38418	0.030900
Break and Enter Commercial	1032	33845	0.030500
Other Theft	1541	52167	0.029500
Break and Enter Residential/Other	1771	60862	0.029100
Offence Against a Person	150	5307	0.028300
Theft from Vehicle	4891	172699	0.028300
Mischief	1910	70413	0.027100
Vehicle Collision or Pedestrian Struck (with Injury)	589	21901	0.026900
Theft of Bicycle	499	25730	0.019400
Homicide	2	218	0.009200

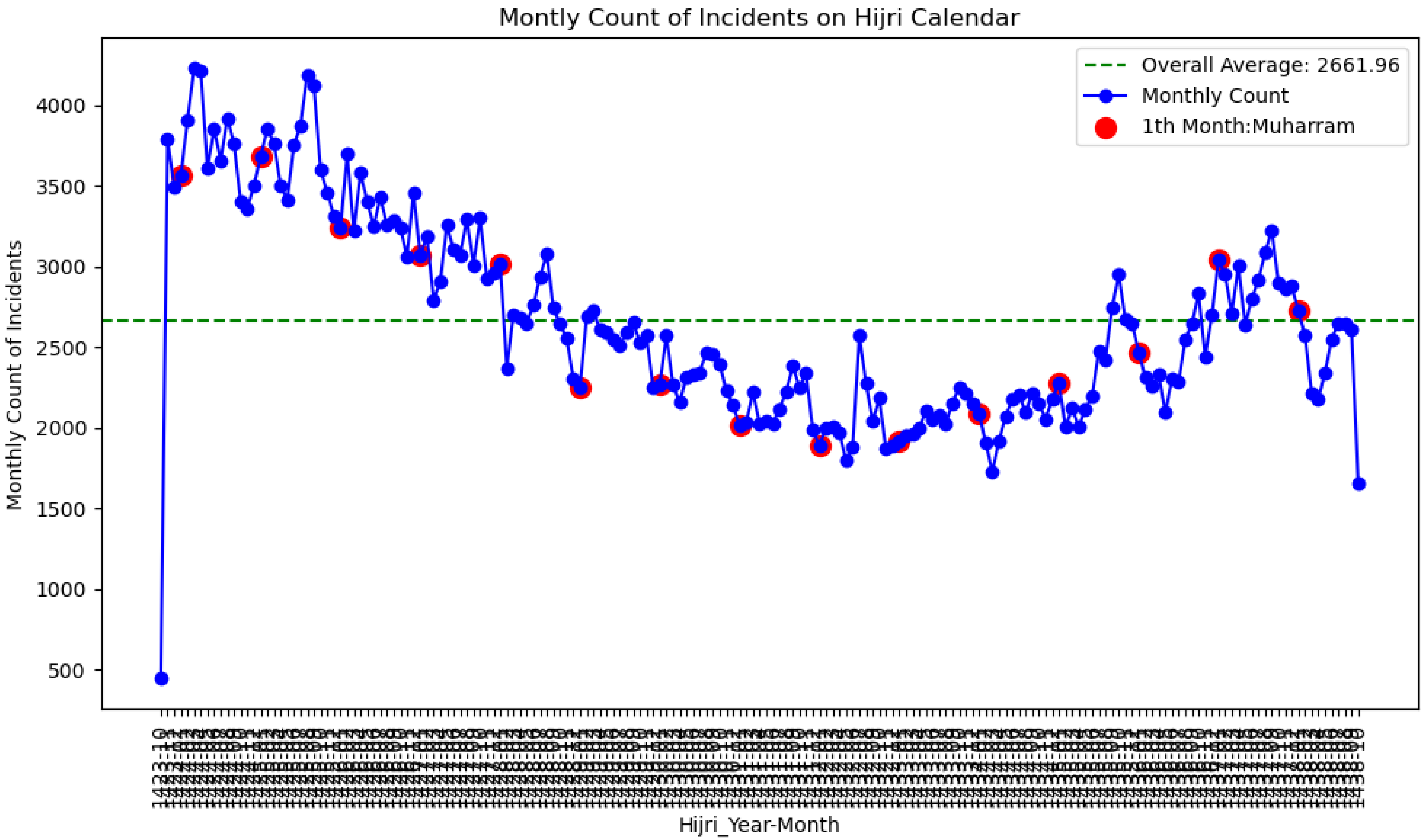
```
In [171…]: # In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[171]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
Theft from Vehicle	4891	172699	0.0283
Break and Enter Residential/Other	1771	60862	0.0291
Other Theft	1541	52167	0.0295
Theft of Vehicle	1186	38418	0.0309
Break and Enter Commercial	1032	33845	0.0305
Offence Against a Person	150	5307	0.0283
Vehicle Collision or Pedestrian Struck (with Fatality)	8	254	0.0315

In [172…

```
monthly_count_plot()
```



More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [173…

```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[173]:

```
(11, 7)
```

In [174…

```
muharrem_incidents_desc
```

Out[174]:
```
count                    13579
unique                      11
top         Theft from Vehicle
freq                      4891
Name: incident, dtype: object
```

In [175…
```
other_days_incidents_desc
```

Out[175]:
```
count                   468235
unique                      11
top         Theft from Vehicle
freq                    167808
Name: incident, dtype: object
```

Vancouver dataset encompasses 11 distinct incident types. During the first ten days of Muharram, crimes were committed across 11 incident categories, with 7 of these categories experiencing incident counts exceeding the annual averages.

SAMPLE DATA-7: CHICAGO CRIME DATASET_2001-2023

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2/data_preview

In [176…
```
df = pd.read_csv("Chicago_Crimes_2001_to_Present.csv", index_col=0,)
# to drop the duplicated rows with the same index numbers use index_col=0
df.head()
```

Out[176]:

ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	...	Ward	Community Area	FBI Code	Coordi
11037294	JA371270	03/18/2015 12:00:00 PM	0000X W WACKER DR	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	BANK	False	False	111	...	42.0	32.0	11	I
11646293	JC213749	12/20/2018 03:00:00 PM	023XX N LOCKWOOD AVE	1154	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT $300 AND UNDER	APARTMENT	False	False	2515	...	36.0	19.0	11	I
11645836	JC212333	05/01/2016 12:25:00 AM	055XX S ROCKWELL ST	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	NaN	False	False	824	...	15.0	63.0	11	I
11645959	JC211511	12/20/2018 04:00:00 PM	045XX N ALBANY AVE	2820	OTHER OFFENSE	TELEPHONE THREAT	RESIDENCE	False	False	1724	...	33.0	14.0	08A	I
11645601	JC212935	06/01/2014 12:01:00 AM	087XX S SANGAMON ST	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	RESIDENCE	False	False	2222	...	21.0	71.0	11	I

5 rows × 21 columns

In [177…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7953830 entries, 11037294 to 12002171
Data columns (total 21 columns):
 #   Column                Dtype  
---  ------                -----  
 0   Case Number           object 
 1   Date                  object 
 2   Block                 object 
 3   IUCR                  object 
 4   Primary Type          object 
 5   Description           object 
 6   Location Description  object 
 7   Arrest                bool   
 8   Domestic              bool   
 9   Beat                  int64  
 10  District              float64
 11  Ward                  float64
 12  Community Area        float64
 13  FBI Code              object 
 14  X Coordinate          float64
 15  Y Coordinate          float64
 16  Year                  int64  
 17  Updated On            object 
 18  Latitude              float64
 19  Longitude             float64
 20  Location              object 
dtypes: bool(2), float64(7), int64(2), object(10)
memory usage: 1.2+ GB
```

In [178…

```
# df = pd.read_csv("Chicago_Crimes_2001_to_Present.csv", usecols=[0, 2, 5, 6])
```

In [179…

```
df.duplicated().value_counts()
```

Out[179]:

```
False    7953669
True         161
dtype: int64
```

In [180…

```
df = df.drop_duplicates()
```

In [181…

```
df = df.iloc[:, [1, 4, 5]]
```

In [182…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7953669 entries, 11037294 to 12002171
Data columns (total 3 columns):
 #   Column        Dtype 
---  ------        ----- 
 0   Date          object
 1   Primary Type  object
 2   Description   object
dtypes: object(3)
memory usage: 242.7+ MB
```

In [183…

```
df["Primary Type"].value_counts()
```

Out[183]:

```
THEFT                                1679300
BATTERY                              1451708
CRIMINAL DAMAGE                       906654
NARCOTICS                             750859
ASSAULT                               521978
OTHER OFFENSE                         493465
BURGLARY                              429109
MOTOR VEHICLE THEFT                   393673
DECEPTIVE PRACTICE                    356129
ROBBERY                               299815
CRIMINAL TRESPASS                     217211
WEAPONS VIOLATION                     112009
PROSTITUTION                           69958
OFFENSE INVOLVING CHILDREN             56968
PUBLIC PEACE VIOLATION                 52892
SEX OFFENSE                            31684
CRIM SEXUAL ASSAULT                    27555
INTERFERENCE WITH PUBLIC OFFICER       18746
LIQUOR LAW VIOLATION                   15014
GAMBLING                               14630
ARSON                                  13586
HOMICIDE                               12820
CRIMINAL SEXUAL ASSAULT                 7881
KIDNAPPING                              7320
STALKING                                5136
INTIMIDATION                            4817
CONCEALED CARRY LICENSE VIOLATION       1205
OBSCENITY                                840
PUBLIC INDECENCY                         197
NON-CRIMINAL                             184
OTHER NARCOTIC VIOLATION                 149
HUMAN TRAFFICKING                        105
NON - CRIMINAL                            38
RITUALISM                                 24
NON-CRIMINAL (SUBJECT SPECIFIED)           9
DOMESTIC VIOLENCE                          1
Name: Primary Type, dtype: int64
```

In [184…

```
df = df.rename(columns = {'Date':'date'})
df = df.rename(columns = {'Primary Type':'incident'})
```

In [185…

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [186…

```
df.date.min(), df.date.max()
```

Out[186]:

```
('2001-01-01', '2023-12-03')
```

In [187…

```
# df.to_csv("Chicago.csv", index=False)
```

In [188…

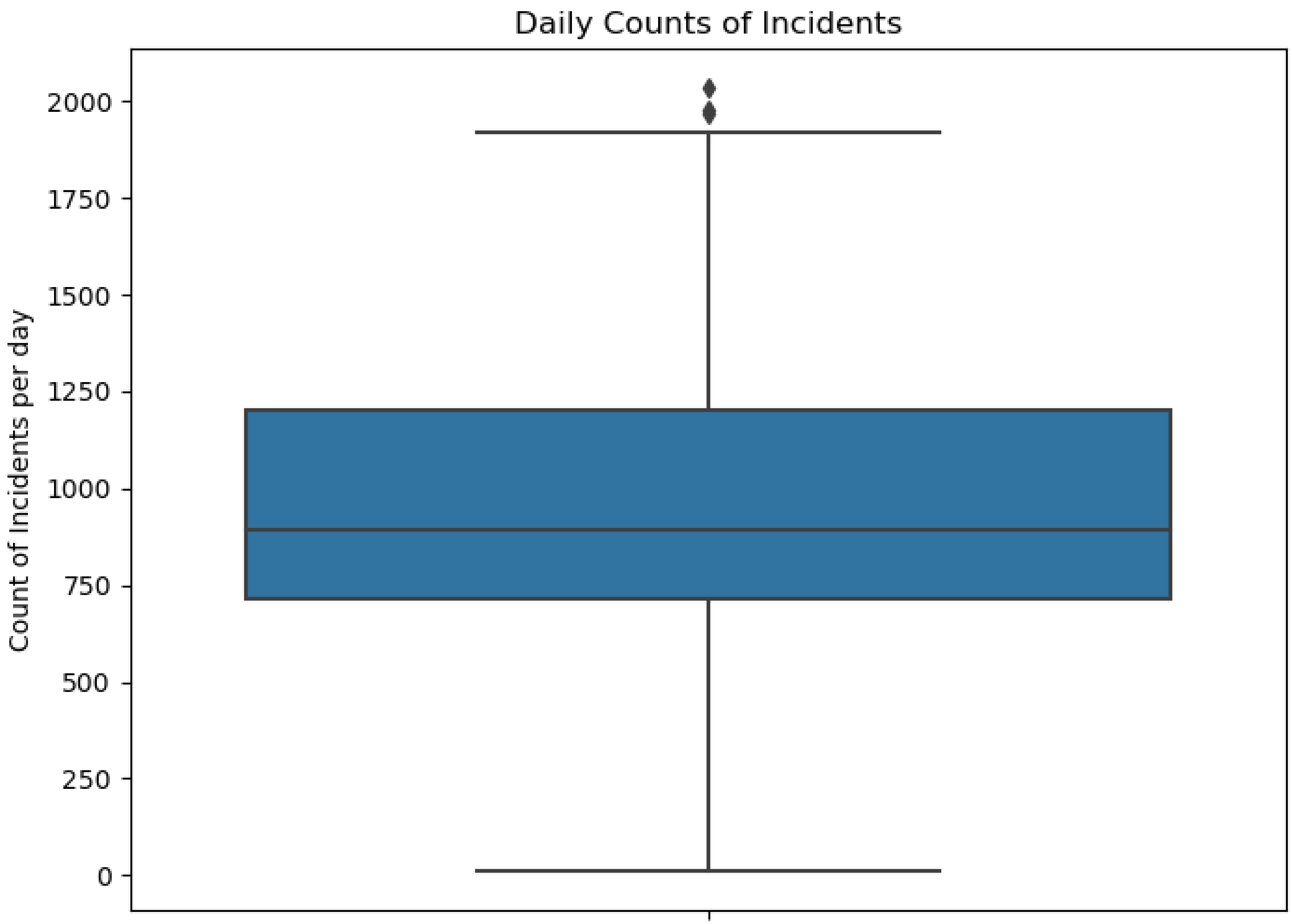
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[188]:

```
count    8372
mean      950
std       283
min        11
25%       716
50%       894
75%      1199
95%      1406
98%      1470
99%      1523
max      2033
Name: date, dtype: int32
```

In [189…

```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [190…
```
df.date.nunique()
```

Out[190]:
```
8372
```

As seen below, our dataset spans a total of 8372 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

In [191…
```
muharrem_10_days(df)
```

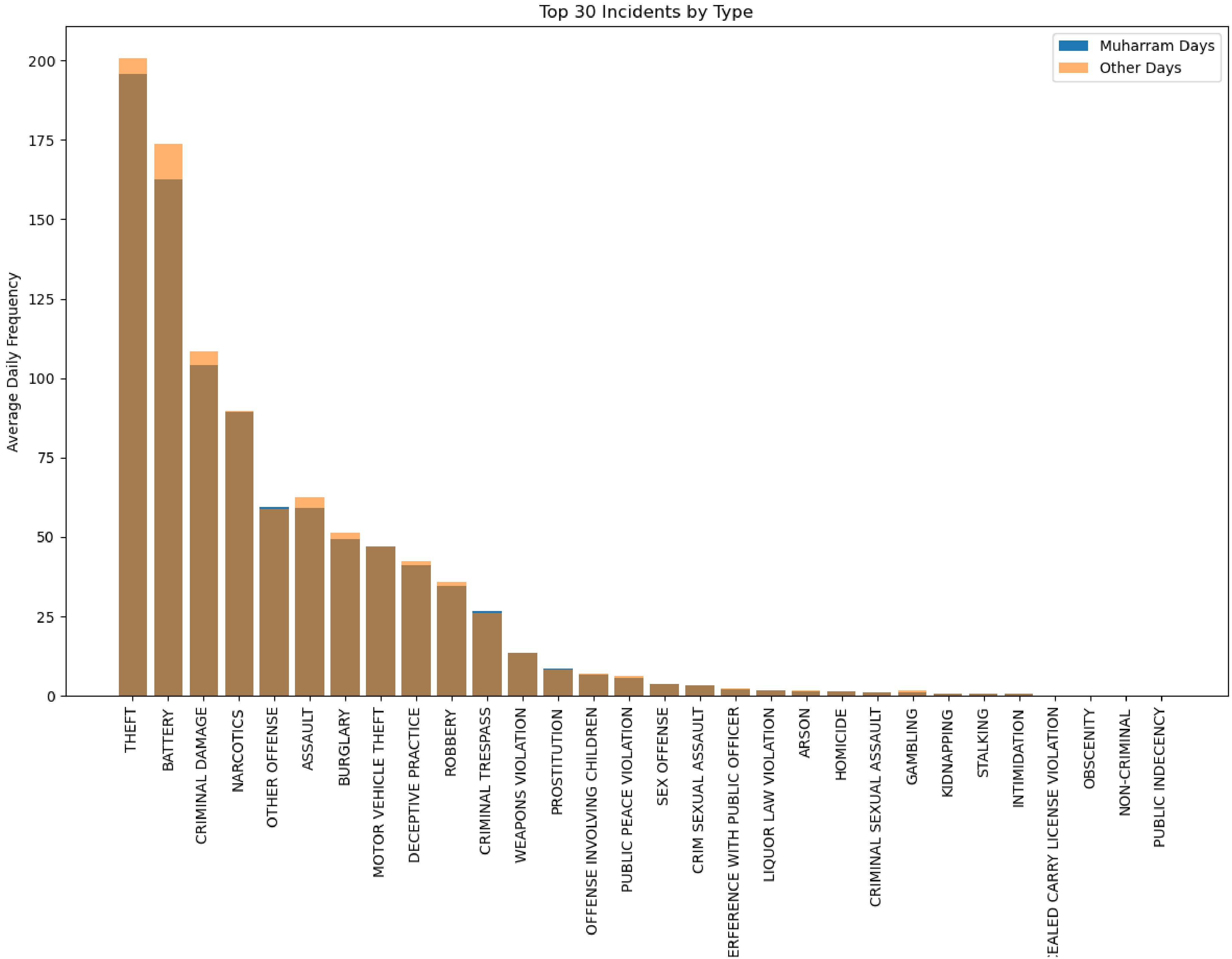
```
Total number of days: 8372
---------------------
Total number of cases: 7953669
----------------------
Average Daily Case Count: 950.03
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2002    486807
2001    485896
2003    475976
2004    469423
2005    453774
2006    448176
2007    437082
2008    427177
2009    392824
2010    370507
2011    351990
2012    336322
2013    307541
2014    275801
2016    269840
2017    269108
2018    268927
2015    264807
2019    261391
2023    240008
2022    239057
2020    212263
2021    208972
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1422    470958
1423    470195
1424    461123
1425    455377
1426    445279
1427    434421
1428    425843
1429    413583
1430    383982
```

```
1431    362344
1432    340567
1433    328725
1434    303696
1435    271358
1439    261354
1438    260778
1437    259935
1436    256965
1440    256539
1444    249476
1441    220744
1443    215897
1442    199257
1421    104539
1445    100734
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 923.0042
------------------------------------------------------------
Average case count in other days: 950.8298
---------------------------------
Ratio of Muharram cases to other cases: 0.9707
---------------------------------------
```

We observe a -2.93% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [192…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
25
50
75
100
125
150
175
200
THEFT
BATTERY
CRIMINAL DAMAGE
NARCOTICS
OTHER OFFENSE
ASSAULT
BURGLARY
MOTOR VEHICLE THEFT
DECEPTIVE PRACTICE
ROBBERY
CRIMINAL TRESPASS
WEAPONS VIOLATION
PROSTITUTION
OFFENSE INVOLVING CHILDREN
PUBLIC PEACE VIOLATION
SEX OFFENSE
CRIM SEXUAL ASSAULT
ERFERENCE WITH PUBLIC OFFICER
LIQUOR LAW VIOLATION
ARSON
HOMICIDE
CRIMINAL SEXUAL ASSAULT
GAMBLING
KIDNAPPING
STALKING
INTIMIDATION
EALED CARRY LICENSE VIOLATION
OBSCENITY
NON-CRIMINAL
PUBLIC INDECENCY

In [193…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[193]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
RITUALISM	1	24	0.041700
NON-CRIMINAL	7	184	0.038000
PUBLIC INDECENCY	7	197	0.035500
OTHER NARCOTIC VIOLATION	5	149	0.033600
CONCEALED CARRY LICENSE VIOLATION	40	1205	0.033200
STALKING	165	5136	0.032100
INTIMIDATION	147	4817	0.030500
OBSCENITY	25	840	0.029800
CRIMINAL SEXUAL ASSAULT	233	7881	0.029600
PROSTITUTION	2066	69958	0.029500
CRIMINAL TRESPASS	6386	217211	0.029400
WEAPONS VIOLATION	3261	112009	0.029100
CRIM SEXUAL ASSAULT	802	27555	0.029100
OTHER OFFENSE	14319	493465	0.029000
MOTOR VEHICLE THEFT	11313	393673	0.028700
NARCOTICS	21484	750859	0.028600
SEX OFFENSE	892	31684	0.028200
LIQUOR LAW VIOLATION	423	15014	0.028200
KIDNAPPING	205	7320	0.028000
THEFT	46956	1679300	0.028000
OFFENSE INVOLVING CHILDREN	1588	56968	0.027900
ROBBERY	8324	299815	0.027800
DECEPTIVE PRACTICE	9885	356129	0.027800
BURGLARY	11865	429109	0.027700

	muharrem incidents	all incidents	muharrem incidents/total incidents
HOMICIDE	355	12820	0.027700
CRIMINAL DAMAGE	25014	906654	0.027600
ASSAULT	14214	521978	0.027200
ARSON	368	13586	0.027100
BATTERY	39063	1451708	0.026900
INTERFERENCE WITH PUBLIC OFFICER	494	18746	0.026400

In [194…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[194]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
OTHER OFFENSE	14319	493465	0.0290
MOTOR VEHICLE THEFT	11313	393673	0.0287
CRIMINAL TRESPASS	6386	217211	0.0294
WEAPONS VIOLATION	3261	112009	0.0291
PROSTITUTION	2066	69958	0.0295
CRIM SEXUAL ASSAULT	802	27555	0.0291
CRIMINAL SEXUAL ASSAULT	233	7881	0.0296
STALKING	165	5136	0.0321
INTIMIDATION	147	4817	0.0305
CONCEALED CARRY LICENSE VIOLATION	40	1205	0.0332
OBSCENITY	25	840	0.0298
NON-CRIMINAL	7	184	0.0380
PUBLIC INDECENCY	7	197	0.0355
OTHER NARCOTIC VIOLATION	5	149	0.0336
RITUALISM	1	24	0.0417

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [195…

```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[195]:

```
(36, 15)
```

In [196…

```
muharrem_incidents_desc
```

Out[196]:

```
count     221521
unique        34
top        THEFT
freq       46956
Name: incident, dtype: object
```

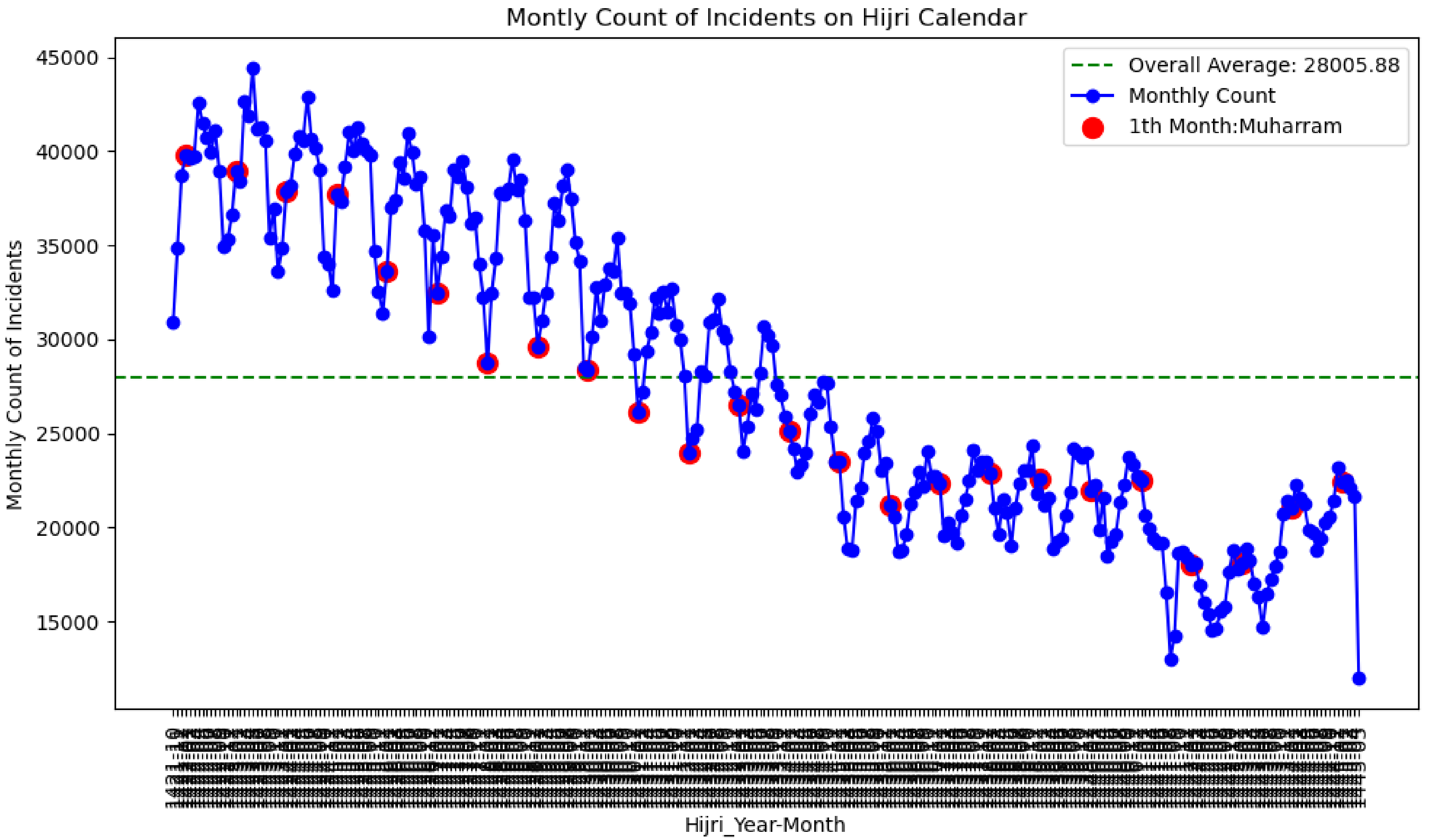
In [197…

```
other_days_incidents_desc
```

Out[197]:
```
count     7732148
unique         36
top         THEFT
freq      1632344
Name: incident, dtype: object
```

Chicago dataset encompasses 36 distinct incident types. During the first ten days of Muharram, crimes were committed across 34 incident categories, with 15 of these categories experiencing incident counts exceeding the annual averages.

In [198…
```
monthly_count_plot()
```



SAMPLE DATA-8: BALTIMORE CRIME DATASET_2011-2015

https://data.world/baltimore/baltimore-crime-data/workspace/file?filename=BPD_Part_1_Victim_Based_Crime_Data.csv

In [199…

```
df = pd.read_csv("BPD_Part_1_Victim_Based_Crime_Data.csv", low_memory=False)
df.head()
```

Out[199]:

	CrimeDate	CrimeTime	CrimeCode	Location	Description	Weapon	Post	District	Neighborhood	Location 1	Total Incidents
0	06/18/2016	00:33:00	4E	2700 CHESLEY AVE	I	HANDS	424.0	NORTHEASTERN	North Harford Road	(39.3679000000, -76.5555900000)	1
1	06/18/2016	00:39:00	4B	2700 FAIT AVE	O	KNIFE	232.0	SOUTHEASTERN	Canton	(39.2831500000, -76.5783400000)	1
2	06/18/2016	0015	9S	2400 CYLBURN AV	Outside	FIREARM	532.0	NORTHERN	Levindale	(39.3510400000, -76.6597600000)	1
3	06/18/2016	01:53:00	3AF	2300 ORLEANS ST	O	FIREARM	221.0	SOUTHEASTERN	McElderry Park	(39.2955600000, -76.5844600000)	1
4	06/18/2016	02:05:00	6C	800 N WOLFE ST	I	NaN	321.0	EASTERN	Middle East	(39.3002700000, -76.5909700000)	1

In [200…

```
df.duplicated().value_counts()
```

Out[200]:

```
False    253900
True      10596
dtype: int64
```

In [201…

```
df = df.drop_duplicates()
```

In [202…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 253900 entries, 0 to 264495
Data columns (total 11 columns):
 #   Column           Non-Null Count   Dtype  
---  ------           --------------   -----  
 0   CrimeDate        253900 non-null  object 
 1   CrimeTime        253900 non-null  object 
 2   CrimeCode        253900 non-null  object 
 3   Location         252173 non-null  object 
 4   Description      250114 non-null  object 
 5   Weapon           82019 non-null   object 
 6   Post             253701 non-null  float64
 7   District         253842 non-null  object 
 8   Neighborhood     252105 non-null  object 
 9   Location 1       252175 non-null  object 
 10  Total Incidents  253900 non-null  int64  
dtypes: float64(1), int64(1), object(9)
memory usage: 23.2+ MB
```

In [203…

```
df["CrimeCode"].value_counts()
```

```
Out[203]: 4E      41377
6D      35539
5A      25352
7A      22986
6G      15212
6J      11980
6C      11939
6E      11702
4C       9312
5D       7677
4B       6393
3AF      5660
3B       5296
4A       4019
4D       3482
5B       3367
6B       3345
5C       3187
4F       3167
6F       2436
9S       2118
3CF      1697
7C       1597
3K       1393
3AK      1381
2A       1354
1F       1047
3AO       949
5F        798
3AJF      742
5E        687
8H        652
3JF       645
3D        625
3P        544
6A        439
6L        373
3CK       259
8AO       243
2B        223
3GF       222
3BJ       221
3CO       194
3JK       185
```

```
3NF       179
3JO       151
1K        146
8J        145
6H        138
8FO        95
3H         92
1O         88
3NK        80
7B         80
8AV        68
8BO        67
3AJK       65
3EF        60
3M         58
3AJO       52
8EO        50
3NO        43
3F         40
3GK        35
3LF        21
3GO        20
3LO        17
8GO        15
3EK        13
8BV        12
8CO        10
8EV         8
3EO         7
8I          7
8GV         5
3N          5
8CV         4
6K          3
8FV         3
3LK         1
8DO         1
Name: CrimeCode, dtype: int64
```

In [204…

```
df = df.rename(columns = {'CrimeDate':'date'})
df = df.rename(columns= {'CrimeCode' : 'incident'})
```

In [205…

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [206…

```
df.date.min(), df.date.max()
```

Out[206]:

```
('2011-01-01', '2016-06-18')
```

In [207…

```
df = df.iloc[:, [0,2]]
# df.to_csv("Baltimore.csv", index=False)
```

In [208…

```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[208]:

```
count    1996
mean      127
std        21
min        31
25%       114
50%       128
75%       142
95%       160
98%       167
99%       171
max       342
Name: date, dtype: int32
```

In [209…

```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```

In [210…

```
df.date.nunique()
```

Out[210]:

```
1996
```

As seen below, our dataset spans a total of 1996 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

In [211…

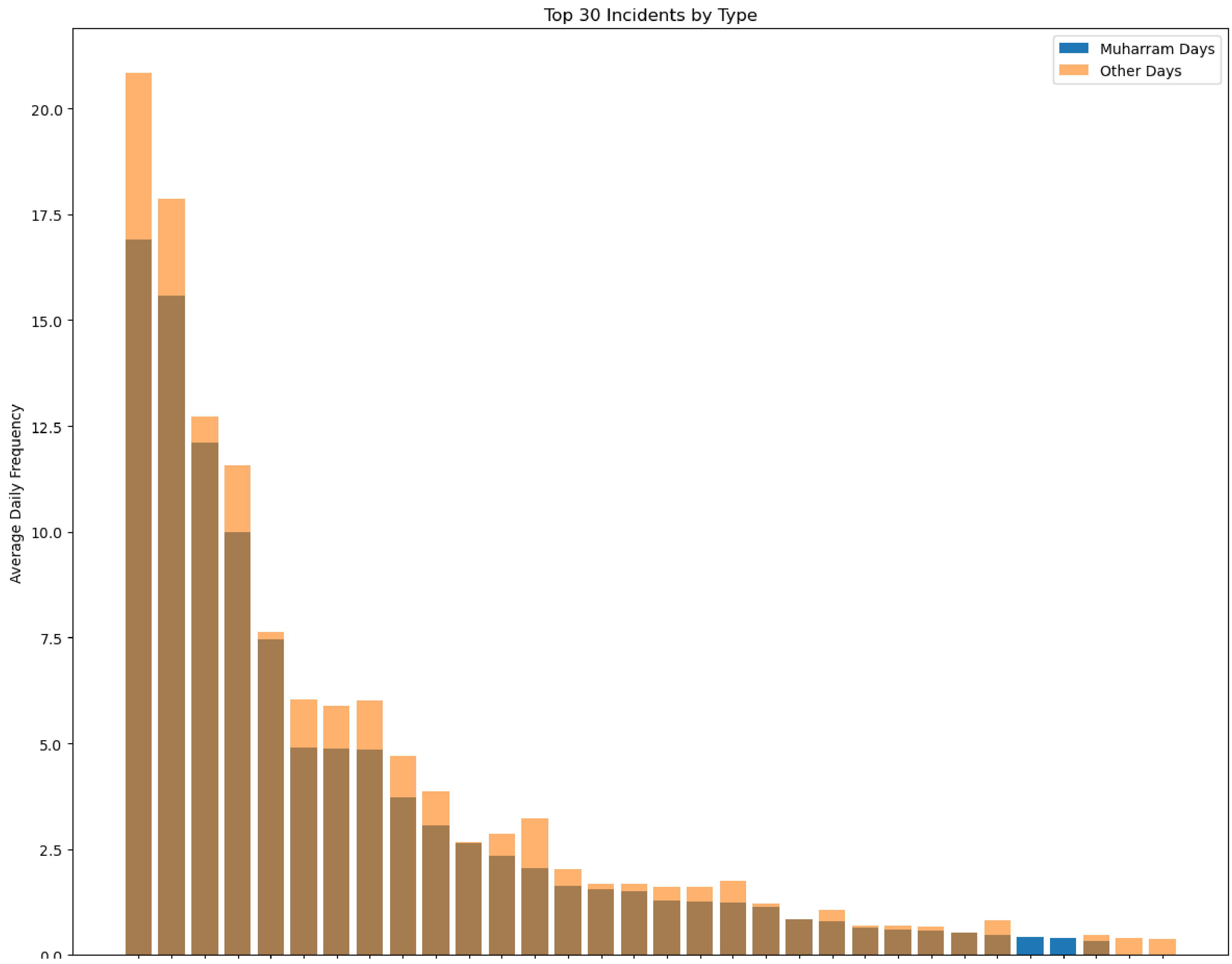
```
muharrem_10_days(df)
```

```
Total number of days: 1996
---------------------
Total number of cases: 253900
----------------------
Average Daily Case Count: 127.2
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2011    48555
2012    47643
2013    47569
2015    46643
2014    44343
2016    19147
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1433    46536
1434    45861
1436    44720
1432    43976
1435    43592
1437    29215
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 108.6167
------------------------------------------------------------
Average case count in other days: 127.7805
--------------------------------
Ratio of Muharram cases to other cases: 0.8500
---------------------------------------
```

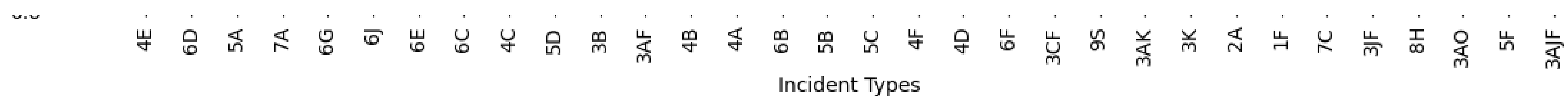
We observe a -15% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [212…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0



In [213…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[213]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
3N	1	5	0.200000
3GK	3	35	0.085700
8FO	7	95	0.073700
7B	4	80	0.050000
3NO	2	43	0.046500
8BO	3	67	0.044800
3JF	25	645	0.038800
3NK	3	80	0.037500
8H	24	652	0.036800
3P	19	544	0.034900
6L	13	373	0.034900
3NF	6	179	0.033500
3JO	5	151	0.033100
3H	3	92	0.032600
1F	32	1047	0.030600
3CF	51	1697	0.030100
3B	159	5296	0.030000
6G	448	15212	0.029500
5A	726	25352	0.028600
6F	69	2436	0.028300
3AK	39	1381	0.028200
6B	94	3345	0.028100
5E	19	687	0.027700
5B	90	3367	0.026700

	muharrem incidents	all incidents	muharrem incidents/total incidents
6D	935	35539	0.026300
7A	599	22986	0.026100
3CO	5	194	0.025800
3K	36	1393	0.025800
2A	34	1354	0.025100
6E	292	11702	0.025000

In [214…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[214]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
3CF	51	1697	0.0301
1F	32	1047	0.0306
3JF	25	645	0.0388
8H	24	652	0.0368
3P	19	544	0.0349
6L	13	373	0.0349
8FO	7	95	0.0737
3NF	6	179	0.0335
3JO	5	151	0.0331
7B	4	80	0.0500
3GK	3	35	0.0857
3H	3	92	0.0326
3NK	3	80	0.0375
8BO	3	67	0.0448
3NO	2	43	0.0465
3N	1	5	0.2000

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [215…

```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[215]:

```
(81, 16)
```

In [216…

```
muharrem_incidents_desc
```

Out[216]:

```
count     6517
unique      60
top         4E
freq      1015
Name: incident, dtype: object
```

In [217…

```
other_days_incidents_desc
```

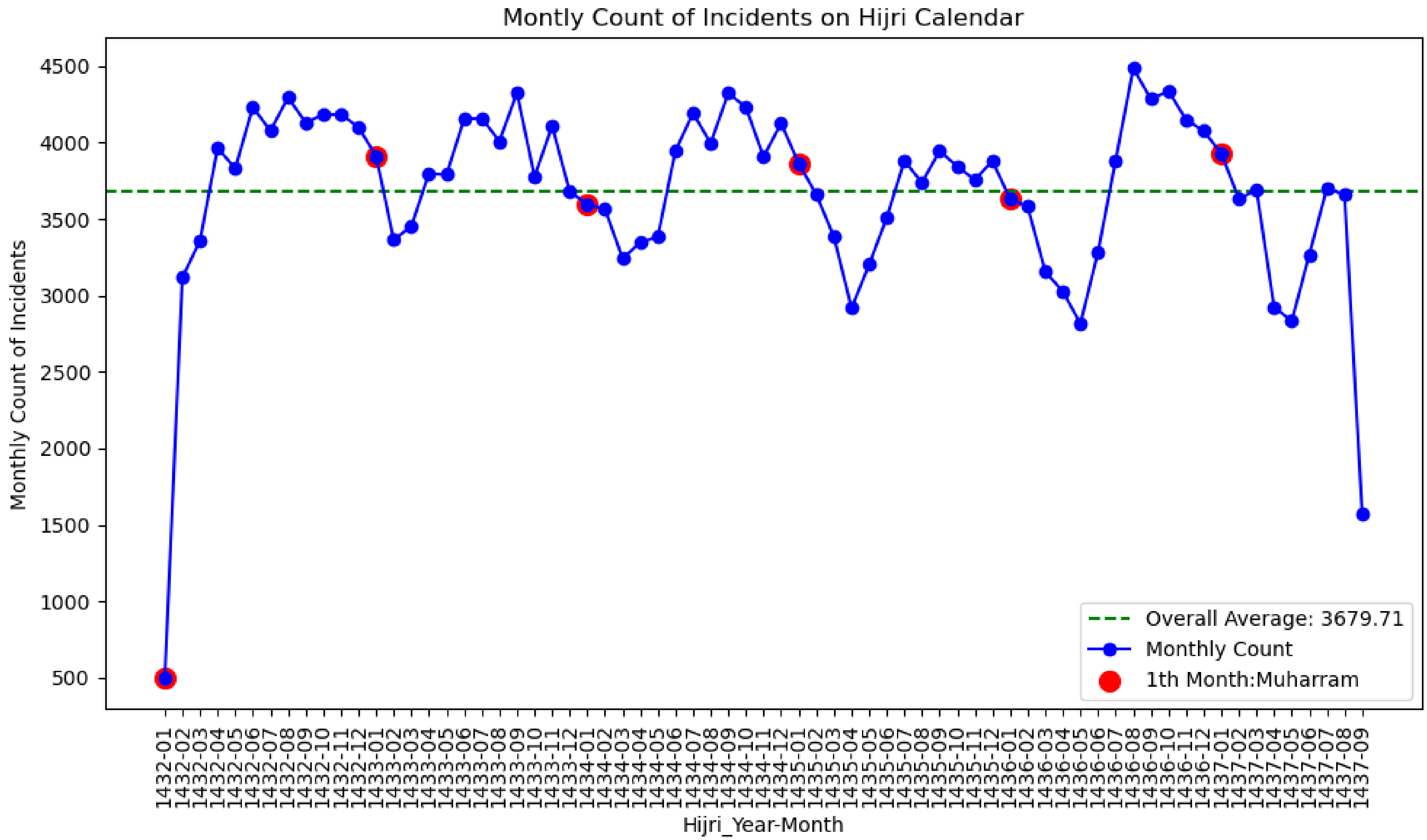
Out[217]:

```
count     247383
unique        81
top           4E
freq       40362
Name: incident, dtype: object
```

Baltimore dataset encompasses 81 distinct incident types. During the first ten days of Muharram, crimes were committed across 60 incident categories, with 16 of these categories experiencing incident counts exceeding the annual averages.

In [218…

```
monthly_count_plot()
```



SAMPLE DATA-9: ATLANTA CRIME DATASET_2009-2017

https://data.world/bryantahb/crime-in-atlanta-2009-2017

In [219…

```
df = pd.read_csv("atlcrime.csv", low_memory=False)
df.drop(columns="Unnamed: 0", inplace=True)
```

```
df
```

Out[219]:

	crime	number	date	location	beat	neighborhood	npu	lat	long
0	LARCENY-NON VEHICLE	103040029	10/31/2010	610 SPRING ST NW	509	Downtown	M	33.77101	-84.38895
1	AUTO THEFT	103040061	10/31/2010	850 OAK ST SW	401	West End	T	33.74057	-84.41680
2	LARCENY-FROM VEHICLE	103040169	10/31/2010	1344 METROPOLITAN PKWY SW	301	Capitol View Manor	X	33.71803	-84.40774
3	AUTO THEFT	103040174	10/31/2010	1752 PRYOR RD SW	307	Betmar LaVilla	Y	33.70731	-84.39674
4	LARCENY-NON VEHICLE	103040301	10/31/2010	JOHN WESLEY DOBBS AVE NE / CORLEY ST NE	604	Old Fourth Ward	M	33.75947	-84.36626
...	...	...	...	...	...	...	...	...	...
270683	BURGLARY-RESIDENCE	92442142	09/01/2009	1226 PORTLAND AVE SE	612	East Atlanta	W	33.73927	-84.34741
270684	LARCENY-FROM VEHICLE	92442164	09/01/2009	317 PICKFAIR WAY SW	307	Lakewood Heights	Y	33.70436	-84.40013
270685	LARCENY-NON VEHICLE	92448045	09/01/2009	6234 SPINE RD @ATRIUM	50	NaN	NaN	33.64068	-84.44204
270686	LARCENY-NON VEHICLE	92440866	09/01/2009	30 WARREN ST	610	Kirkwood	O	33.75374	-84.32600
270687	HOMICIDE	92440372058	09/01/2009	2860 MARTIN L KING JR DR SW	405	Harland Terrace	I	33.75399	-84.48138

270688 rows × 9 columns

In [220…

```
df.duplicated().value_counts()
```

Out[220]:

```
False    270688
dtype: int64
```

In [221…

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270688 entries, 0 to 270687
Data columns (total 9 columns):
 #   Column        Non-Null Count   Dtype  
---  ------        --------------   -----  
 0   crime         270688 non-null  object 
 1   number        270688 non-null  int64  
 2   date          270688 non-null  object 
 3   location      270686 non-null  object 
 4   beat          270688 non-null  object 
 5   neighborhood  258928 non-null  object 
 6   npu           268592 non-null  object 
 7   lat           270688 non-null  float64
 8   long          270688 non-null  float64
dtypes: float64(2), int64(1), object(6)
memory usage: 18.6+ MB
```

In [222…

```
df["crime"].value_counts()
```

Out[222]:

```
LARCENY-FROM VEHICLE    77345
LARCENY-NON VEHICLE     64697
BURGLARY-RESIDENCE      42941
AUTO THEFT              38168
AGG ASSAULT             19133
ROBBERY-PEDESTRIAN      14446
BURGLARY-NONRES          8505
ROBBERY-RESIDENCE        1880
ROBBERY-COMMERCIAL       1855
RAPE                      990
HOMICIDE                  728
Name: crime, dtype: int64
```

In [223…

```
df = df.rename(columns= {'crime' : 'incident'})
```

In [224…

```
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [225…

```
df.date.min(), df.date.max()
```

Out[225]:

```
('2009-01-01', '2017-02-28')
```

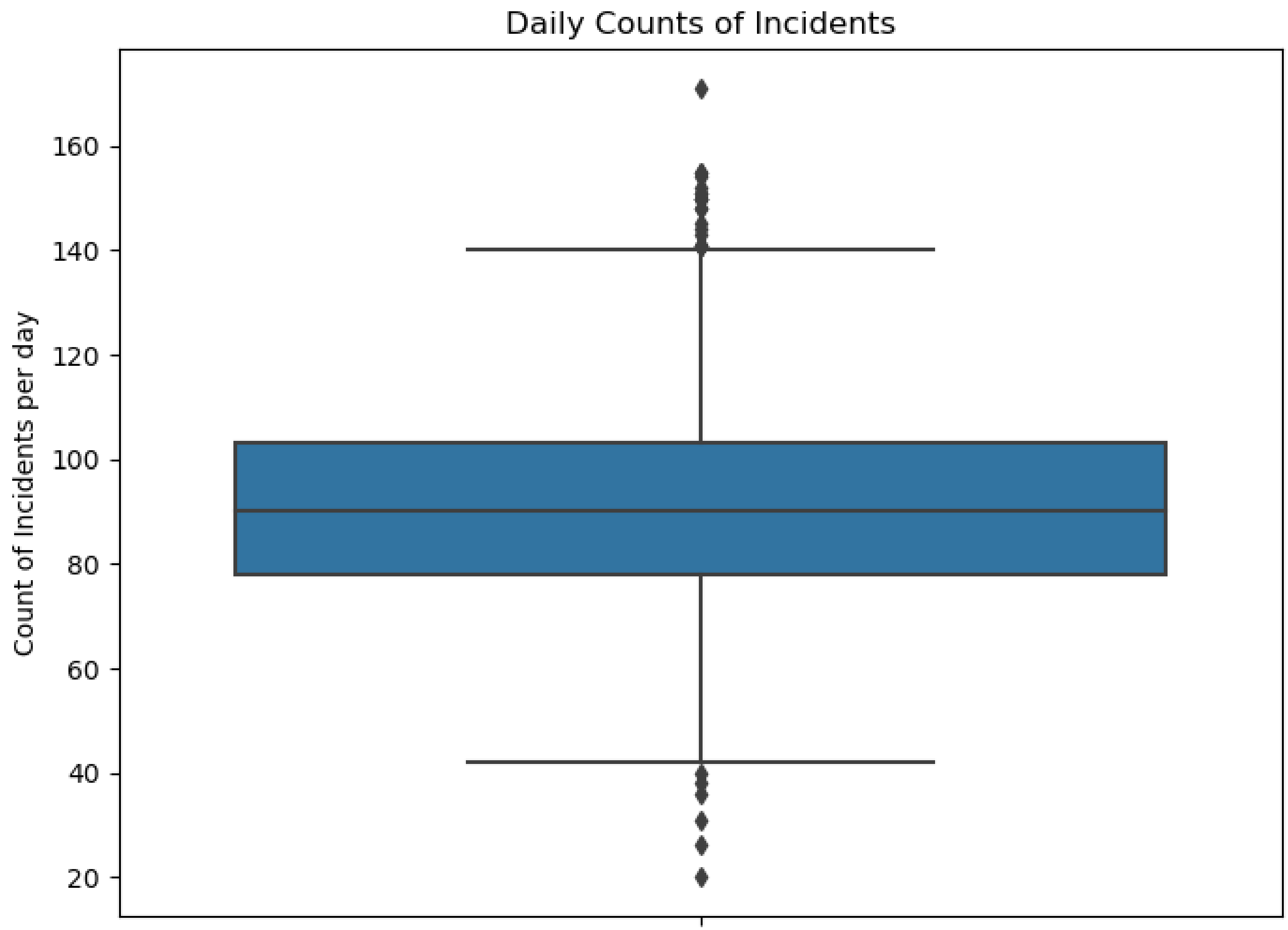
In [226…

```
df = df.iloc[:, [0,2]]
# df.to_csv("Atlanta.csv", index=False)
```

In [227…
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[227]:
```
count    2981
mean       90
std        18
min        20
25%        78
50%        90
75%       103
95%       121
98%       130
99%       137
max       171
Name: date, dtype: int32
```

In [228…
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [229…
```
df.date.nunique()
```
Out[229]:
```
2981
```

As seen below, our dataset spans a total of 2981 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

In [230…
```
muharrem_10_days(df)
```

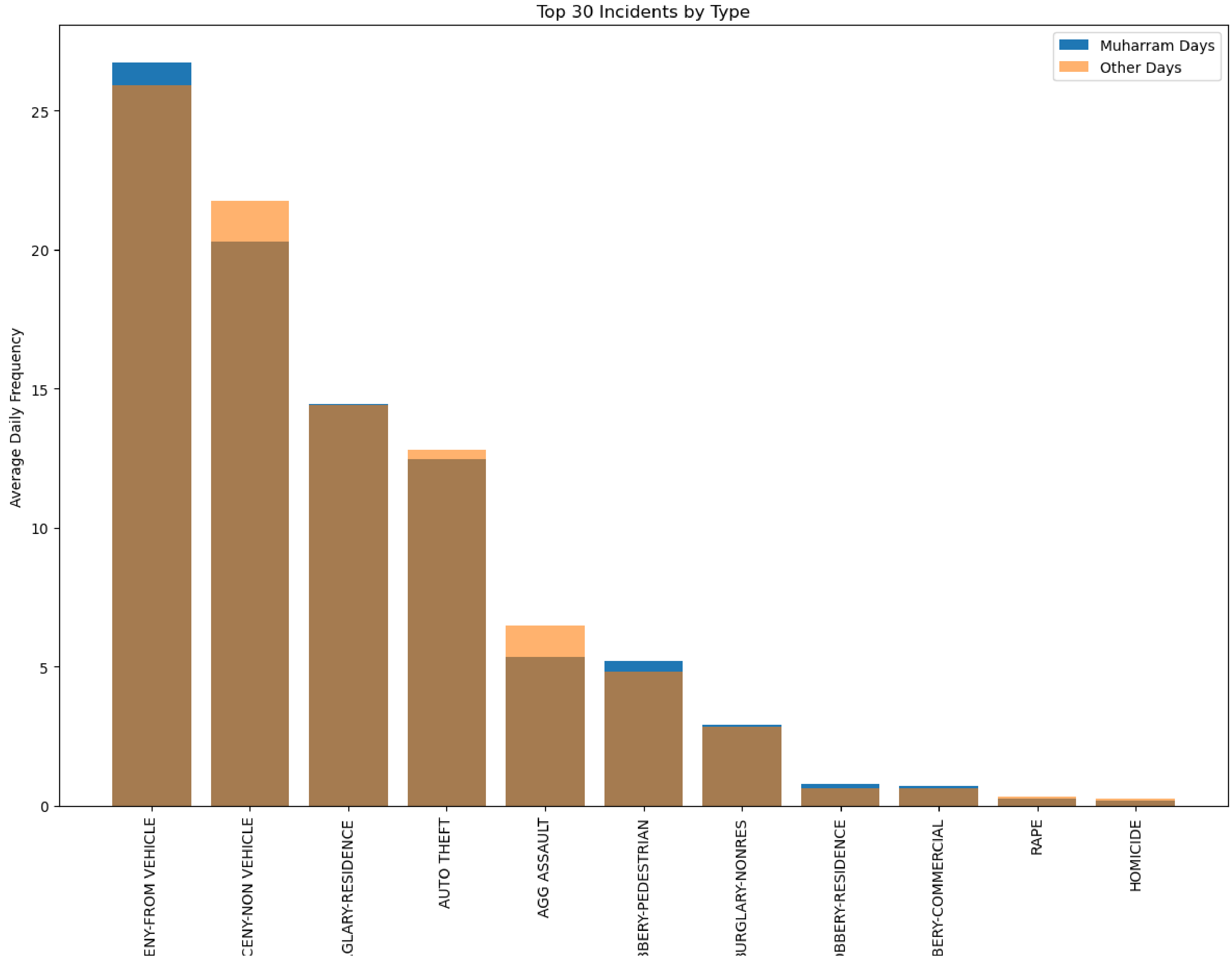
```
Total number of days: 2981
---------------------
Total number of cases: 270688
----------------------
Average Daily Case Count: 90.8
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2009    39626
2010    35770
2011    35174
2012    33394
2013    32303
2014    31166
2015    30117
2016    29131
2017     4007
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1430    38259
1431    34884
1432    34032
1433    32539
1434    31776
1436    30133
1435    29559
1437    28162
1438    11344
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 89.3333
------------------------------------------------------------
Average case count in other days: 90.8502
---------------------------------
Ratio of Muharram cases to other cases: 0.9833
---------------------------------------
```

We observe a -1.67% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [231…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
```

```python
# display(muharrem_dominant_incidents)
```



Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
5
10
15
20
25
ENY-FROM VEHICLE
CENY-NON VEHICLE
GLARY-RESIDENCE
AUTO THEFT
AGG ASSAULT
BBERY-PEDESTRIAN
BURGLARY-NONRES
BBERY-RESIDENCE
BERY-COMMERCIAL
RAPE
HOMICIDE

In [232…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[232]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
ROBBERY-RESIDENCE	70	1880	0.037200
ROBBERY-COMMERCIAL	63	1855	0.034000
ROBBERY-PEDESTRIAN	466	14446	0.032300
LARCENY-FROM VEHICLE	2406	77345	0.031100
BURGLARY-NONRES	263	8505	0.030900
BURGLARY-RESIDENCE	1300	42941	0.030300
AUTO THEFT	1123	38168	0.029400
LARCENY-NON VEHICLE	1827	64697	0.028200
AGG ASSAULT	482	19133	0.025200
HOMICIDE	17	728	0.023400
RAPE	23	990	0.023200

In [233…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[233]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
LARCENY-FROM VEHICLE	2406	77345	0.0311
BURGLARY-RESIDENCE	1300	42941	0.0303
ROBBERY-PEDESTRIAN	466	14446	0.0323
BURGLARY-NONRES	263	8505	0.0309
ROBBERY-RESIDENCE	70	1880	0.0372
ROBBERY-COMMERCIAL	63	1855	0.0340

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

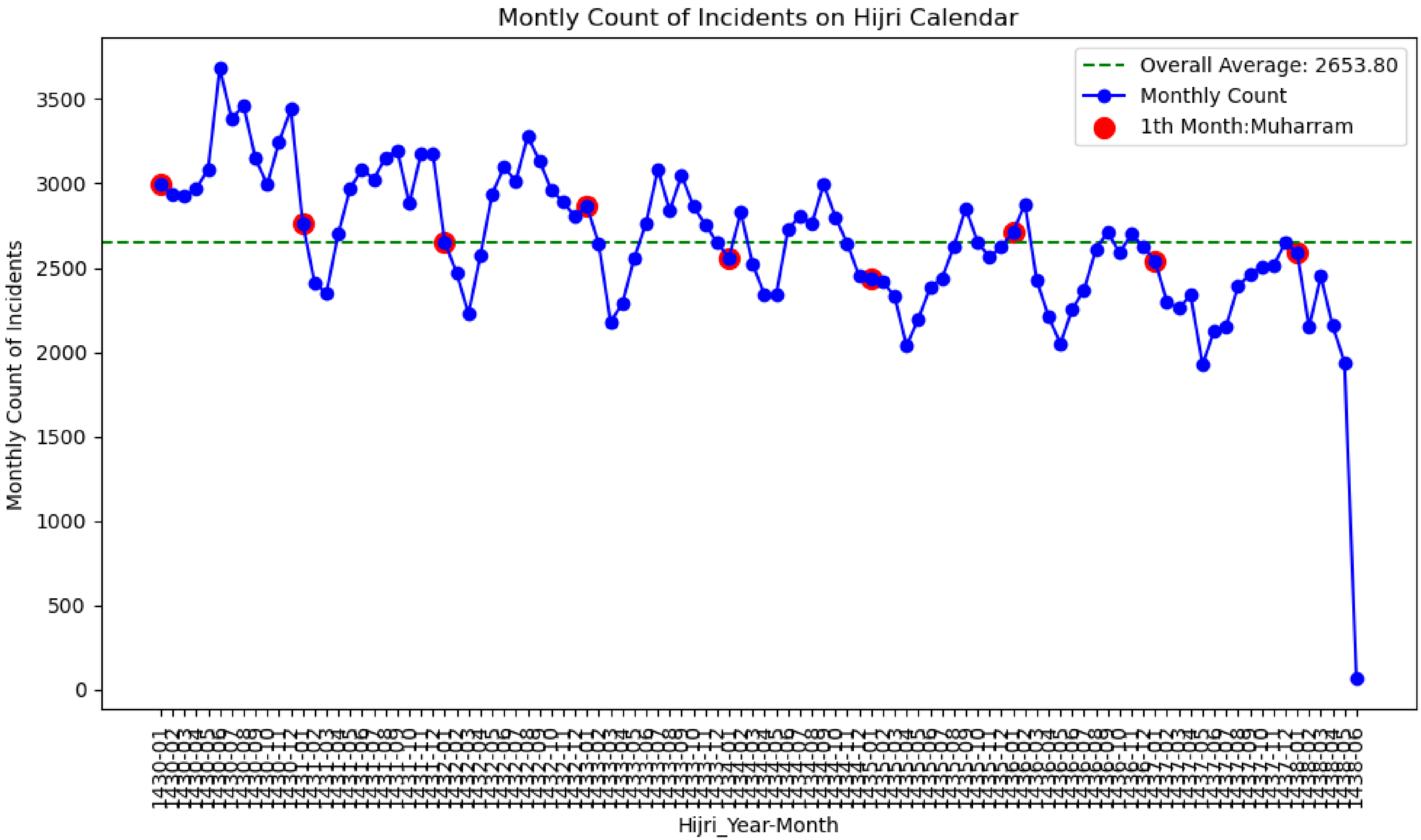
In [234…
```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```
Out[234]:
```
(11, 6)
```

In [235…
```
muharrem_incidents_desc
```
Out[235]:
```
count                     8040
unique                      11
top       LARCENY-FROM VEHICLE
freq                      2406
Name: incident, dtype: object
```

In [236…
```
other_days_incidents_desc
```
Out[236]:
```
count                   262648
unique                      11
top       LARCENY-FROM VEHICLE
freq                     74939
Name: incident, dtype: object
```

Atlanta dataset encompasses 11 distinct incident types. During the first ten days of Muharram, crimes were committed across 11 incident categories, with 6 of these categories experiencing incident counts exceeding the annual averages.

In [237…
```
monthly_count_plot()
```



SAMPLE DATA-10: OAKLAND CRIME STATISTICS_2011-2016

https://www.kaggle.com/datasets/cityofoakland/oakland-crime-statistics-2011-to-2016/

In [238…

```
df2 = pd.read_csv("records-for-2011.csv")
df3 = pd.read_csv("records-for-2012.csv")
```

```python
df4 = pd.read_csv("records-for-2013.csv")
df5 = pd.read_csv("records-for-2014.csv")
df6 = pd.read_csv("records-for-2015.csv")
df7 = pd.read_csv("records-for-2016.csv")
```

In [239…

```python
frames = [df2, df3, df4, df5, df6, df7]
df = pd.concat(frames)
df.head()
```

Out[239]:

	Agency	Create Time	Location	Area Id	Beat	Priority	Incident Type Id	Incident Type Description	Event Number	Closed Time	Location 1	Zip Codes	Location
0	OP	2011-01-01T00:00:00.000	ST&SAN PABLO AV	1.0	06X	1.0	PDOA	POSSIBLE DEAD PERSON	LOP110101000001	2011-01-01T00:28:17.000	NaN	NaN	NaN
1	OP	2011-01-01T00:01:11.000	ST&HANNAH ST	1.0	07X	1.0	415GS	415 GUNSHOTS	LOP110101000002	2011-01-01T01:12:56.000	NaN	NaN	NaN
2	OP	2011-01-01T00:01:25.000	ST&MARKET ST	1.0	10Y	2.0	415GS	415 GUNSHOTS	LOP110101000003	2011-01-01T00:07:20.000	NaN	NaN	NaN
3	OP	2011-01-01T00:01:35.000	PRENTISS ST	2.0	21Y	2.0	415GS	415 GUNSHOTS	LOP110101000005	2011-01-01T00:02:28.000	NaN	NaN	NaN
4	OP	2011-01-01T00:02:10.000	AV&FOOTHILL BLVD	2.0	20X	1.0	415GS	415 GUNSHOTS	LOP110101000004	2011-01-01T00:50:04.000	NaN	NaN	NaN

In [240…

```python
df.duplicated().value_counts()
```

Out[240]:

```
False    1046388
dtype: int64
```

In [241…

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1046388 entries, 0 to 110827
Data columns (total 13 columns):
 #   Column                     Non-Null Count    Dtype  
---  ------                     --------------    -----  
 0   Agency                     1046384 non-null  object 
 1   Create Time                1046384 non-null  object 
 2   Location                   483425 non-null   object 
 3   Area Id                    864023 non-null   object 
 4   Beat                       1040583 non-null  object 
 5   Priority                   1046384 non-null  float64
 6   Incident Type Id           1046384 non-null  object 
 7   Incident Type Description  1045996 non-null  object 
 8   Event Number               1046384 non-null  object 
 9   Closed Time                1046359 non-null  object 
 10  Location 1                 374799 non-null   object 
 11  Zip Codes                  352 non-null      float64
 12  Location                   188052 non-null   object 
dtypes: float64(2), object(11)
memory usage: 111.8+ MB
```

In [242…

```
df["Incident Type Description"].value_counts()[:30]
```

Out[242]:

```
ALARM-RINGER            98497
SECURITY CHECK          70965
911 HANG-UP             54935
STOLEN VEHICLE          47958
DISTURBING THE PEACE    38257
MENTALLY ILL            37218
415 UNKNOWN             33470
BATTERY                 30636
SUSPICIOUS PERSON       26984
BATTERY ON CO-HABITA    23964
415 GUNSHOTS            21520
415 FAMILY              21372
SUSPICIOUS VEHICLE      20781
ROBBERY                 19452
HAZARDOUS SITUATION/    18948
WELFARE CHECK -- CHE    17450
TRESPASS:               14819
VEHICLE COLLISION-PE    13782
ASSAULT W/DEADLY WEA    12982
415 THREATS             12819
HIT & RUN (PROPERTY     12417
FIGHT                   12062
STAND BY AND PRESERV    10584
DISTURBANCE-NEIGHBOR    10553
SUBJECT ARMED WITH W    10376
BURGLARY                10001
OBSTRUCT PERSON'S MO     9782
RUNAWAY                  9557
ALARM-MANUALLY ACTIV     9356
DISTURBANCE-CUSTOMER     8543
Name: Incident Type Description, dtype: int64
```

In [243…

```python
df = df.rename(columns = {'Create Time':'date'})
df = df.rename(columns = {'Incident Type Description':'incident'})
```

In [244…

```python
# df[df['date'].isnull()]
df = df.dropna(subset=['date'])
```

In [245…

```python
df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

In [246…

```python
df.date.min(), df.date.max()
```

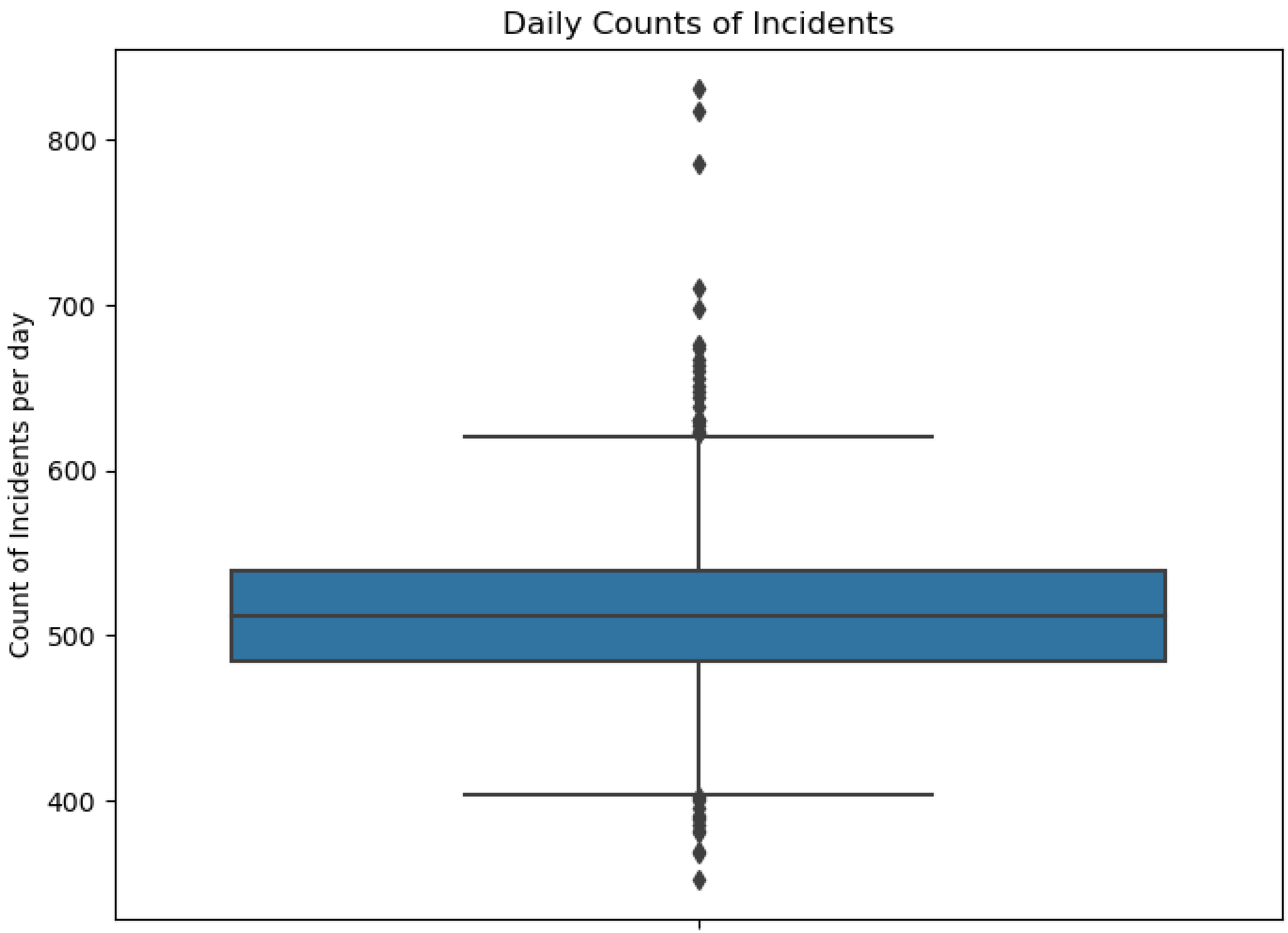
Out[246]: ('2011-01-01', '2016-07-31')

In [247…
```
df = df.iloc[:, [1,7]]
# df.to_csv("Oakland.csv", index=False)
```

In [248…
```
daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

Out[248]:
```
count    2039
mean      513
std        44
min       352
25%       485
50%       512
75%       539
95%       584
98%       607
99%       623
max       831
Name: date, dtype: int32
```

In [249…
```
# Display the days with high incident numbers
plt.figure(figsize=(8, 6))
sns.boxplot(y=df.groupby("date")['date'].value_counts())
plt.title('Daily Counts of Incidents')
plt.ylabel('Count of Incidents per day')
plt.show()
```



In [250…
```
df.date.nunique()
```
Out[250]:
```
2039
```

As seen below, our dataset spans a total of 2039 days. Every day in the dataset contains a record of an incident. In other words, there are no days without any recorded incidents.

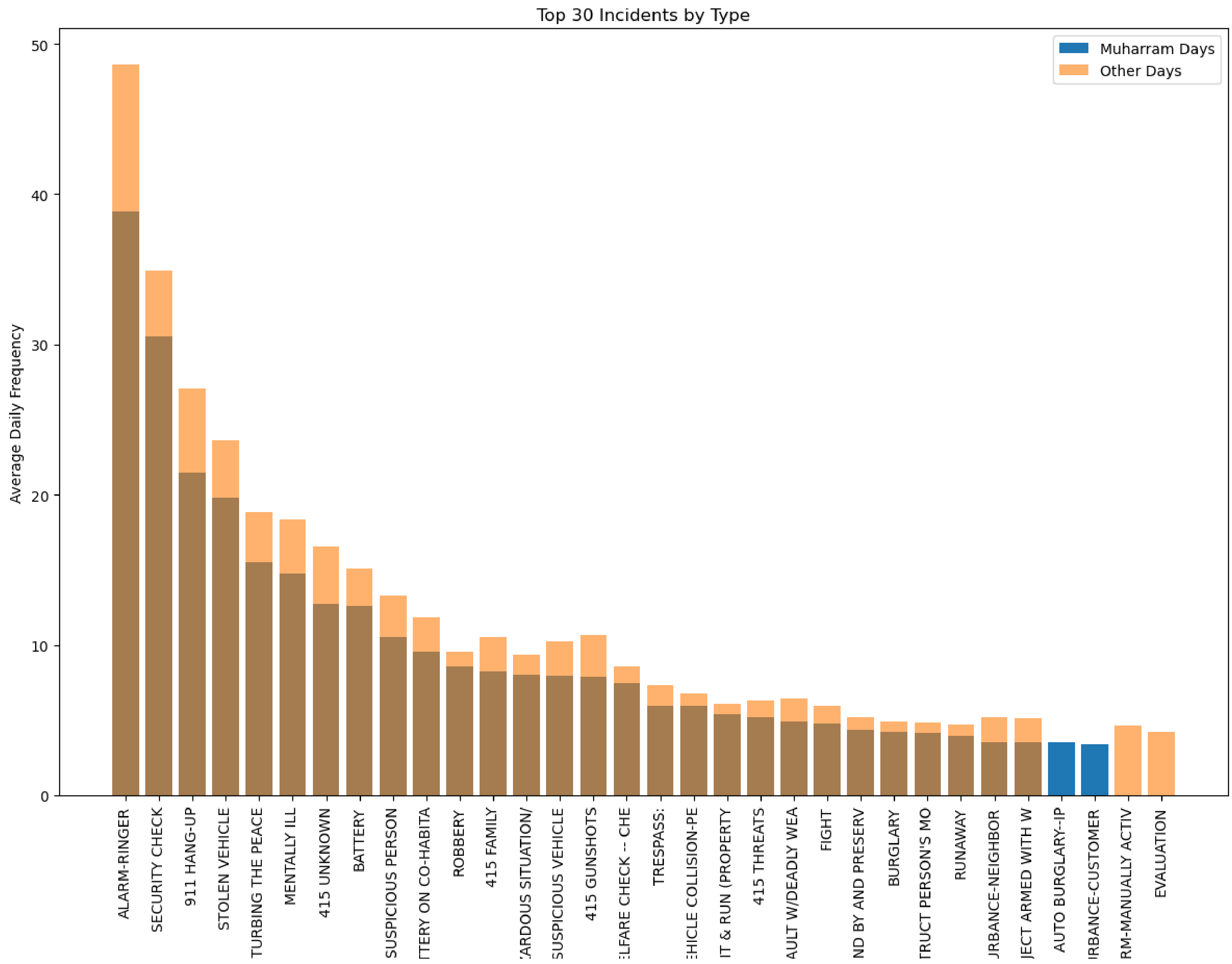
In [251…
```
muharrem_10_days(df)
```

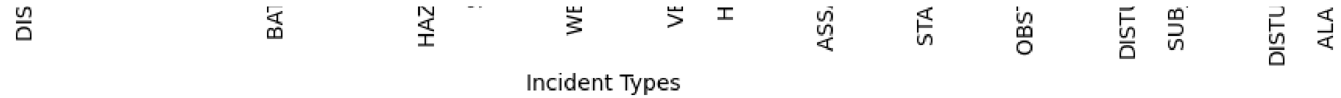
```
Total number of days: 2039
---------------------
Total number of cases: 1046384
----------------------
Average Daily Case Count: 513.18
-------------------------
Yearly case counts according to the Gregorian calendar:
-------------------------------------------------------
2015    192581
2013    188051
2014    187480
2012    187430
2011    180015
2016    110827
Name: date, dtype: int64
--------------------------------------------
Case counts according to the Hijri calendar:
--------------------------------------------
1436    188396
1434    184302
1435    180329
1433    180248
1432    163146
1437    149963
Name: Hijri_Date, dtype: int64
------------------------------------------------------------
Average case count in the first ten days of Muharram months: 415.0167
------------------------------------------------------------
Average case count in other days: 516.1612
---------------------------------
Ratio of Muharram cases to other cases: 0.8040
---------------------------------------
```

We observe a -19.60% lower crime rate during the initial 10 days of the Muharram month compared to the annual average.

In [252…

```
sorted_ratios, muharrem_dominant_incidents, muharrem_incidents_desc, other_days_incidents_desc  = incidents_by_types(df)
# display(sorted_ratios)
# display(muharrem_dominant_incidents)
```


Top 30 Incidents by Type
Muharram Days
Other Days
Average Daily Frequency
0
10
20
30
40
50
ALARM-RINGER
SECURITY CHECK
911 HANG-UP
STOLEN VEHICLE
TURBING THE PEACE
MENTALLY ILL
415 UNKNOWN
BATTERY
SUSPICIOUS PERSON
TTERY ON CO-HABITA
ROBBERY
415 FAMILY
ARDOUS SITUATION/
SUSPICIOUS VEHICLE
415 GUNSHOTS
LFARE CHECK -- CHE
TRESPASS:
HICLE COLLISION-PE
IT & RUN (PROPERTY
415 THREATS
AULT W/DEADLY WEA
FIGHT
ND BY AND PRESERV
BURGLARY
TRUCT PERSON'S MO
RUNAWAY
JRBANCE-NEIGHBOR
JECT ARMED WITH W
AUTO BURGLARY--IP
IRBANCE-CUSTOMER
RM-MANUALLY ACTIV
EVALUATION



In [253…

```
# Top 30 incident types sorted by "muharrem incidents / total incidents" ratio
sorted_ratios
```

Out[253]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
VICE	1	1	1.000000
ESCAPE DETENTION	2	7	0.285700
INTERFERE WITH POWER	2	8	0.250000
PROTECTIVE CUSTODY-N	1	8	0.125000
POSSESS FORGED NOTES	1	9	0.111100
VEHICLE PARKED ON SI	1	10	0.100000
FALSE REPORT OF CRIM	1	12	0.083300
SEWER PROBLEMS	2	27	0.074100
OBSTRUCTING JUSTICE-	2	29	0.069000
LOCATED MISSING PERS	3	49	0.061200
MAYHEM	2	33	0.060600
GAS LEAK	30	507	0.059200
FAILURE TO PROVIDE F	3	55	0.054500
CHILD TAKEN INTO PRO	1	19	0.052600
AGGRAVATED ASSAULT	5	98	0.051000
SURRENDER OF GUN OR	2	43	0.046500
HATE CRIME	2	43	0.046500
AUTO IMPROPERLY PARK	2	44	0.045500
ELECTRICITY	2	46	0.043500
INTIMIDATION OF A WI	4	92	0.043500
PICKETERS/PROTESTERS	30	713	0.042100
FOUND PROPERTY	9	215	0.041900
ANIMAL-STRAYING	4	96	0.041700
LOST PERSON	6	147	0.040800

	muharrem incidents	all incidents	muharrem incidents/total incidents
TARASOFF	4	98	0.040800
ATTEMPTED RAPE	4	100	0.040000
CONTRIBUTING TO THE	4	100	0.040000
PASS FICTITIOUS CHEC	3	77	0.039000
ABSENT WITHOUT LEAVE	16	419	0.038200
PURSUIT	26	688	0.037800

In [254…

```
# In which categories were more crimes committed during the first ten days of Muharram?
muharrem_dominant_incidents
```

Out[254]:

	muharrem incidents	all incidents	muharrem incidents/total incidents
INDECENT EXPOSURE	76	2559	0.0297
ATTEMPTED BURGLARY	44	1190	0.0370
CRUELTY TO DEPENDENT	35	1017	0.0344
AUTO ALARM	32	1010	0.0317
PICKETERS/PROTESTERS	30	713	0.0421
GAS LEAK	30	507	0.0592
PURSUIT	26	688	0.0378
SMOKE	25	711	0.0352
CHECK VEHICLE	25	804	0.0311
ASSAULT WITH CAUSTIC	19	580	0.0328
FOUND GUN	19	622	0.0305
ABSENT WITHOUT LEAVE	16	419	0.0382
FOUND PROPERTY	9	215	0.0419
ATTEMPTED AUTO THEFT	7	222	0.0315
LOST PERSON	6	147	0.0408
AGGRAVATED ASSAULT	5	98	0.0510
TARASOFF	4	98	0.0408
CONTRIBUTING TO THE	4	100	0.0400
ATTEMPTED RAPE	4	100	0.0400
ANIMAL-STRAYING	4	96	0.0417
INTIMIDATION OF A WI	4	92	0.0435
FAILURE TO PROVIDE F	3	55	0.0545
PASS FICTITIOUS CHEC	3	77	0.0390
LOCATED MISSING PERS	3	49	0.0612

	muharrem incidents	all incidents	muharrem incidents/total incidents
SEWER PROBLEMS	2	27	0.0741
SCOOTERS INVOLVED IN	2	62	0.0323
SURRENDER OF GUN OR	2	43	0.0465
OBSTRUCTING JUSTICE-	2	29	0.0690
SODOMY	2	62	0.0323
INTERFERE WITH POWER	2	8	0.2500
ELECTRICITY	2	46	0.0435
HATE CRIME	2	43	0.0465
AUTO IMPROPERLY PARK	2	44	0.0455
ESCAPE DETENTION	2	7	0.2857
MAYHEM	2	33	0.0606
CONTEMPT OF COURT/DI	2	67	0.0299
CHOP SHOP OWNERSHIP/	1	27	0.0370
VEHICLE PARKED ON SI	1	10	0.1000
SUSPECTS	1	28	0.0357
CHILD TAKEN INTO PRO	1	19	0.0526
SPOUSAL RAPE	1	31	0.0323
OAKLAND TRAFFIC CODE	1	32	0.0312
FALSE REPORT OF CRIM	1	12	0.0833
PROTECTIVE CUSTODY-N	1	8	0.1250
POSSESSION OR PURCHA	1	33	0.0303
POSSESS FORGED NOTES	1	9	0.1111
KIDNAPPING FOR RANSO	1	30	0.0333
VICE	1	1	1.0000

More crimes were committed in the above-mentioned crime categories during the first ten days of Muharram compared to the other days of the year.

In [255…
```
df.incident.nunique(), muharrem_dominant_incidents.count()[0]
```

Out[255]:
```
(288, 48)
```

In [256…
```
muharrem_incidents_desc
```

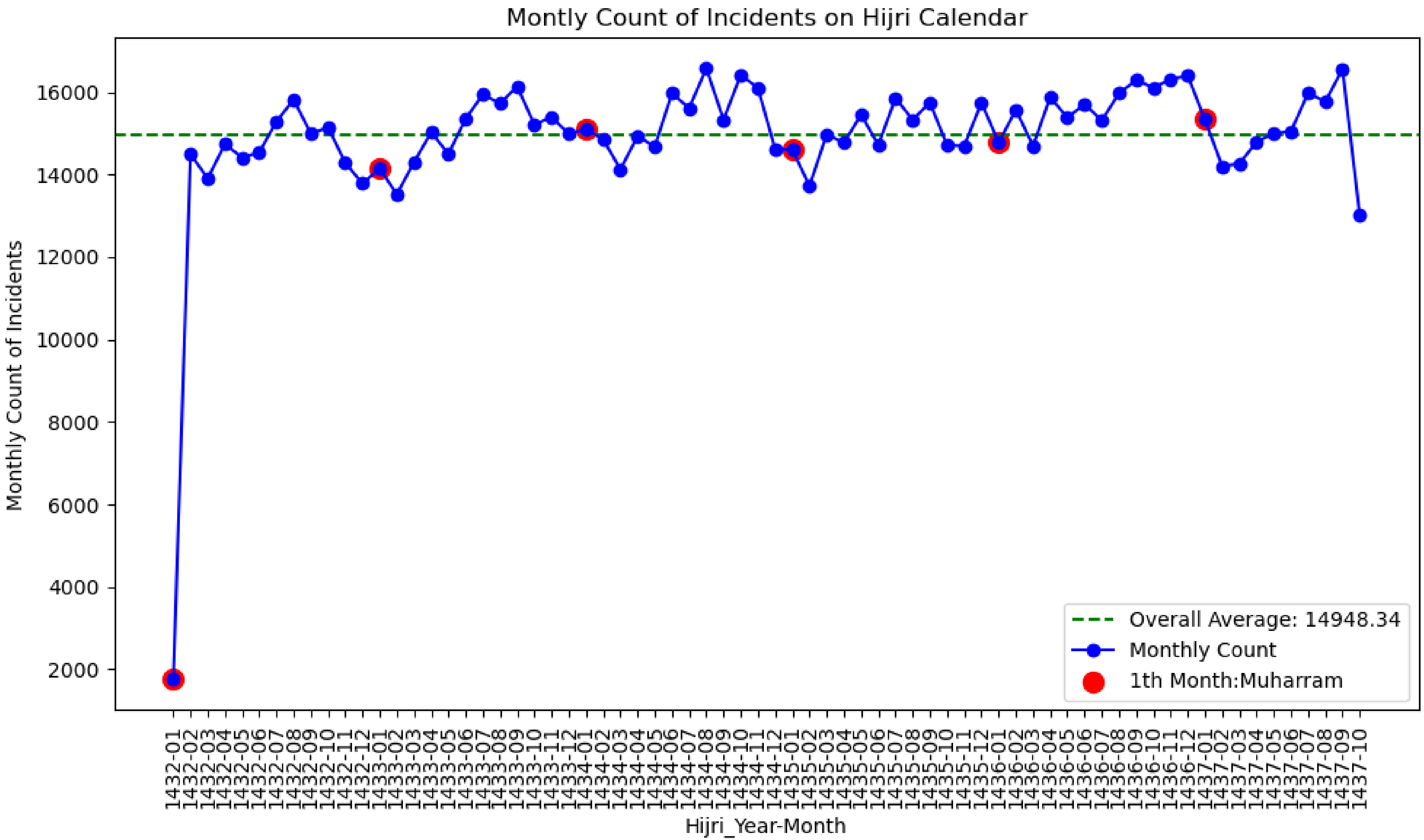
Out[256]:
```
count            24889
unique             221
top       ALARM-RINGER
freq              2333
Name: incident, dtype: object
```

In [257…
```
other_days_incidents_desc
```

Out[257]:
```
count          1021107
unique             287
top       ALARM-RINGER
freq             96164
Name: incident, dtype: object
```

Oakland dataset encompasses 288 distinct incident types. During the first ten days of Muharram, crimes were committed across 221 incident categories, with 48 of these categories experiencing incident counts exceeding the annual averages.

In [258…
```
monthly_count_plot()
```



CONCLUSION

Up to this point, we've examined whether there is an increase or decrease in the first ten days of Muharram compared to the other days of the year in ten different datasets. By making slight adaptations in our code, we also specifically analyzed the 10th Muharram (Ashura Day), and compared them to the remaining days separately. The results obtained are displayed in the table below. The table below displays the percentage ratio of the daily average number of crimes committed during the first ten days of Muharram to the daily average of crimes committed during the rest of the year. Positive percentages indicate a higher incidence of crimes during the Muharram days, while negative values signify a lower incidence during these days compared to the rest of the year.

In [259…
```
results = pd.DataFrame({'UNC CHAPEL HILL': [25, 7.85],
                   'LOS ANGELES': [-10.24, -8.75],
                   'KANSAS': [-8.06, -7.01],
                   'DETROIT': [-8.85,-3.83],
                   'DENVER': [7.90,-0.25],
                   'VANCOUVER': [-0.28,-3.43],
                   'CHICAGO': [-2.93,-4.12],
                   'BALTIMORE': [-15,-14.09],
                   'ATLANTA': [1.67,2.06],
                   'OAKLAND': [-19.60,-19.99]
                   })
```

In [260…
```
results.index = ['Muharram First 10 Days / Other Days Ratio(%)', 'Only Ashura(10th Muharram) Day / Other Days Ratio(%)']
daily_crime_ratios = results.T
```

In [261…
```
daily_crime_ratios
```

Out[261]:

	Muharram First 10 Days / Other Days Ratio(%)	Only Ashura(10th Muharram) Day / Other Days Ratio(%)
UNC CHAPEL HILL	25.00	7.85
LOS ANGELES	-10.24	-8.75
KANSAS	-8.06	-7.01
DETROIT	-8.85	-3.83
DENVER	7.90	-0.25
VANCOUVER	-0.28	-3.43
CHICAGO	-2.93	-4.12
BALTIMORE	-15.00	-14.09
ATLANTA	1.67	2.06
OAKLAND	-19.60	-19.99

Interestingly, except for UNC, Denver, and Atlanta seven data sets/cities showed lower crime rates during the Muharram period. The fact that only 10 days of the year showcased a substantial difference, up to -20% in crime rates compared to other days, presents a significant finding.

In this analysis, the results have shown an opposite effect compared to our examination of Zilhijjah days. It appears that while crime rates increase during Zilhijjah days, there is a decrease during Muharram days. In seven out of ten datasets, fewer crimes were committed during the first ten days of Muharram. Even on the tenth day, the Day of Ashura, the results remain consistent, with fewer crimes observed in eight out of ten datasets. The occurrence of fewer crimes during Muharram days aligns with the Islamic context as well. It's worth recalling that the month of Muharram is regarded as a time of deliverance from calamities, seeking healing from illnesses, attaining peace and forgiveness, emphasizing not wrath but mercy and pardon from Allah, signifying days abundant in tranquility and well-being. Interestingly, the isolated Ashura Day alone displays crime rates well below the yearly averages, with decreases reaching up to 20% in certain datasets. These findings suggest a tendency toward decreased criminal activity during the initial ten days of the Muharram month.
- Contrary to Zilhijjah, seven out of ten datasets reveal lower crime rates during the initial ten days of Muharram, with reductions of up to -19.60% compared to annual averages. - Crimes against individuals decrease during this period, aligning with Islamic contexts that emphasize peace, forgiveness, and deliverance from calamities during Muharram.