

https://www.academia.edu/113294629/EXPLORING_SEASONAL_AFFECT_IN_CRIMES_CORRELATION_BETWEEN_CRIMES_AND PARTICULAR_DAYS

<https://medium.com/@yuce.ahmet.tr/exploring-seasonal-affect-in-crimes-b91beb0ea3be>

EXPLORING_SEASONAL_AFFECT_IN_CRIMES_1_ZILHIJJAH_PERIOD, Ahmet YÜCE

Table of Contents

- 1 IMPORT LIBRARIES
- 2 USER DEFINED FUNCTIONS
- 3 ZILHIJJAH FIRST TEN DAYS IN THE GREGORIAN CALENDAR
- 4 SAMPLE DATA-1: UNC-CHAPEL HILL CAMPUS POLICE CRIME LOG_2013-2018
- 5 SAMPLE DATA-2: LOS ANGELES CRIME DATA_2010-2019
- 6 SAMPLE DATA-3: KANSAS CITY CRIME DATA_2009-2016
- 7 SAMPLE DATA-4: DETROIT CRIME INCIDENTS_2009-2016
- 8 SAMPLE DATA-5: DENVER CRIME DATASET_2019-2023
- 9 SAMPLE DATA-6: VANCOUVER CRIME DATASET_2003-2017
- 10 SAMPLE DATA-7: CHICAGO CRIME DATASET_2001-2023
- 11 SAMPLE DATA-8: BALTIMORE CRIME DATASET_2011-2015
- 12 SAMPLE DATA-9: ATLANTA CRIME DATASET_2009-2017
- 13 SAMPLE DATA-10: OAKLAND CRIME STATISTICS_2011-2016
- 14 CONCLUSION

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_12th_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 12th month and day: whether it's between 1 and 10
    return month == 12 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]
    zilhijjah_df = df[df['Hijri_Date']].astype(str).apply(filter_first_ten_days_of_12th_months)
    zilhijjah_dates_list_in_gregorian = pd.to_datetime(zilhijjah_df['date']).dt.strftime('%Y-%m-%d').unique().tolist()
    zilhijjah_dates_list_in_gregorian = sorted(zilhijjah_dates_list_in_gregorian, key=lambda x:datetime.strptime(x, '%Y-%m-%d'))
    count_zilhijjah_days = df.Hijri_Year_Month[df['Hijri_Year_Month'].astype(str).str.endswith("12")].nunique() * 10
    count_other_days = count_days - count_zilhijjah_days
    return count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], zilhijjah_df,
           zilhijjah_dates_list_in_gregorian, count_zilhijjah_days
```

```
In [6]: def zilhijjah_10_days(df):
    df['date'] = pd.to_datetime(df['date'])
    count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], zilhijjah_df,
```

```

zilhijjah_dates_list_in_gregorian, count_zilhijjah_days = common_codes(df)

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:") * "-")

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 Zilhijjah months:") * "-")

# Total number of cases in the first ten days of Zilhijjah months (rows)
count_zilhijjah_rows = len(zilhijjah_df)

# Average case count in the first ten days of Zilhijjah months:
avg_count_of_zilhijjah_incidents = count_zilhijjah_rows / count_zilhijjah_days
print("Average case count in the first ten days of Zilhijjah months:", round(avg_count_of_zilhijjah_incidents,4))
print(len("Average case count in the first ten days of Zilhijjah months:") * "-")

# Total number of cases in other days (rows)
count_other_rows = len(df[~df['Hijri_Date'].apply(filter_first_ten_days_of_12th_months)])
count_other_days = count_days - count_zilhijjah_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 Zilhijjah cases to other cases:",f"{avg_count_of_zilhijjah_incidents/avg_count_of_other_incidents:.{4}f}")
print(len("Ratio of Zilhijjah cases to other cases:") * "-")

```

In [7]:

```

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

    count_days, count_other_days, df['Hijri_Date'], df['Hijri_Year_Month'], zilhijjah_df, zilhijjah_dates_list_in_gregorian, \

```

```

count_zilhijjah_days = common_codes(df)

all_incidents_count = df.incident.value_counts()

zilhijjah_incidents = df[df['date'].astype(str).isin(zilhijjah_dates_list_in_gregorian)]
zilhijjah_incidents_count = zilhijjah_incidents["incident"].value_counts()

other_days_incidents = df[~df['date'].astype(str).isin(zilhijjah_dates_list_in_gregorian)]
other_days_incidents_count = other_days_incidents["incident"].value_counts()

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

# Bar plot
plt.figure(figsize=(12, 10))
# Zilhijjah Days bar plot
plt.bar(zilhijjah_grouped.index, zilhijjah_grouped/count_zilhijjah_days, label='Zilhijjah 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
zilhijjah_incidents_desc = df[df['date'].astype(str).isin(zilhijjah_dates_list_in_gregorian)]["incident"].describe()
other_days_incidents_desc = df[~df['date'].astype(str).isin(zilhijjah_dates_list_in_gregorian)]["incident"].describe()

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

zilhijjah_dominant_incidents = incident_ratios[incident_ratios["zilhijjah incidents/total incidents"]>equal_ratio]
zilhijjah_dominant_incidents = zilhijjah_dominant_incidents.sort_values(by="zilhijjah incidents", ascending=False)

```

```
zilhijjah_dominant_incidents['zilhijjah incidents'] = zilhijjah_dominant_incidents['zilhijjah incidents'].astype(int)

return sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_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
    last_month = [i for i, month in enumerate(monthly_avg.index) if '12' in month]
    plt.scatter(monthly_avg.index[last_month], monthly_avg.values[last_month], color='red', s=100, label='12th Month:Zilhijjah')

    # 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()
```

ZILHIJJAH FIRST TEN DAYS IN THE GREGORIAN CALENDAR

```
# The days corresponding to the first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2001-02-24' - '2001-03-05' '2002-02-13' - '2002-02-22'  
'2003-02-02' - '2003-02-11' '2004-01-23' - '2004-02-01' '2005-01-12' - '2005-01-21' '2006-01-01' - '2006-01-10' '2006-12-22' - '2006-12-31' '2007-12-11' - '2007-12-20'  
'2008-11-29' - '2008-12-08' '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' -  
'2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11' '2017-08-23' - '2017-09-01' '2018-08-12' - '2018-08-21' '2019-08-02'  
- '2019-08-11' '2020-07-22' - '2020-07-31' '2021-07-11' - '2021-07-20' '2022-06-30' - '2022-07-09' '2023-06-19' - '2023-06-28'
```

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

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

This database is the UNC Police Crime Log, which contains a record of all incidents reported to the UNC Police. The part of the database I acquired is the Jan. 1, 2013 - Oct. 10, 2018 incident names, date and times reported, and locations. I acquired this data on Oct. 10, 2018. The database of incidents reported to UNC Police helps it keep a record of criminal activity throughout the years and allows it to fulfill its federally-mandated Clery Act responsibilities and produce an annual Security and Fire Safety Report. The database is updated daily. The original data included the following fields: offense, date_rept and street. Many cells in the incident field are blank: these incidents include all traffic violations, according to UNC Police, and constitute about one-fourth of incidents (3,283 of 12,817, or 25.61 percent). The data names and locations were cleaned using OpenRefine's clustering tools. While this cleaning method got rid of many of the inconsistencies in the raw data, there are still several locations and incidents that are misspelled or categorized inconsistently. In particular, small numbers of alcohol and marijuana offenses exist under a variety of individual spellings, and there are many kinds of larceny that exist under unique spellings as well (e.g. LARCENY-P2P SIGN, which is one of many larceny incidents that could be categorized under the LARCENY-ALL OTHER distinction). The names of the original fields were changed to incident, date-time, and location, and other fields were added. These fields are: -year, the year the incident took place -date-no year, which includes the day and month but not the year -res hall, which indicates if the incident took place in an on-campus residence hall, in Granville Towers, or neither -alcohol, which lists all incidents containing the word alcohol

```
In [9]: 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
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]: df = df.drop_duplicates()

In [11]: 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')
```

```
# The days corresponding to the first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04'  
'2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11' '2017-08-23' - '2017-09-01' '2018-08-12' - '2018-08-21'
```

```
In [17]: df.info()
```

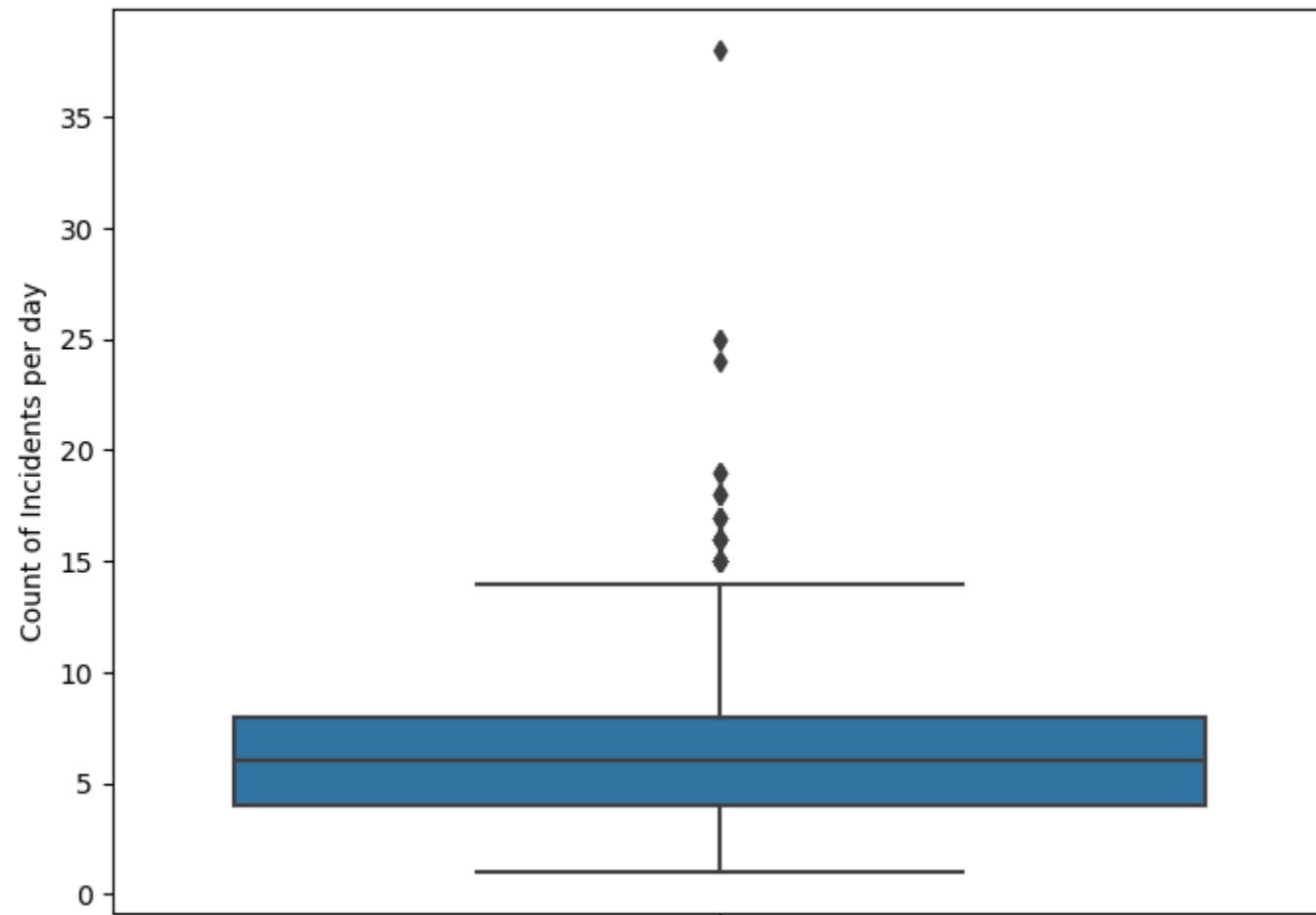
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 12714 entries, 0 to 12816
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   incident    9518 non-null    object 
 1   date        12714 non-null    object 
 2   year         12714 non-null    int64  
 3   date-no year 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 [18]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

```
Out[18]: 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 [19]: # 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



In [20]: df.date

```
Out[20]: 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 [21]: df.date.nunique()
```

```
Out[21]: 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.

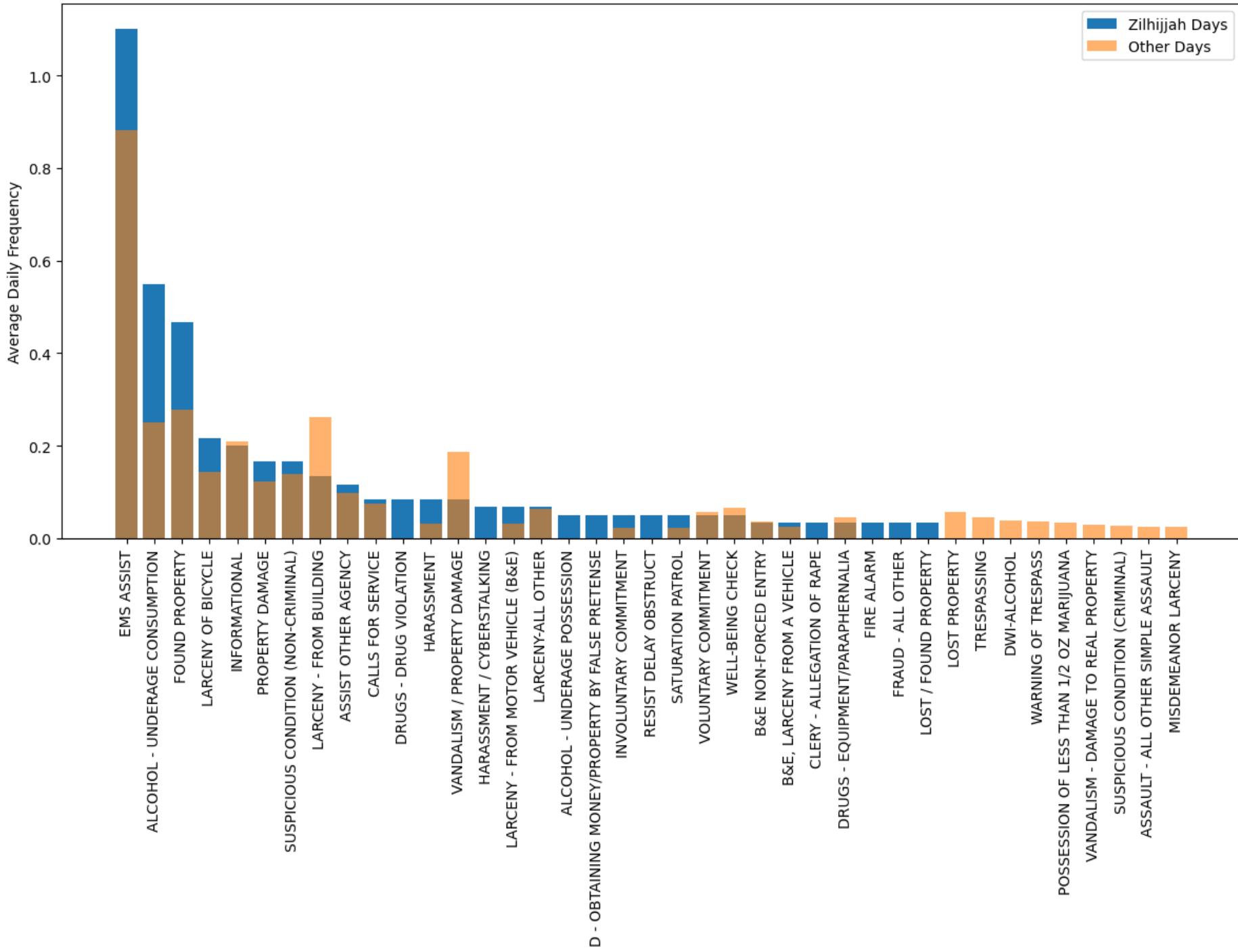
```
In [22]: zilhijjah_10_days (df)
```

```
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 Zilhijjah months: 6.8667
-----
Average case count in other days: 6.0039
-----
Ratio of Zilhijjah cases to other cases: 1.1437
```

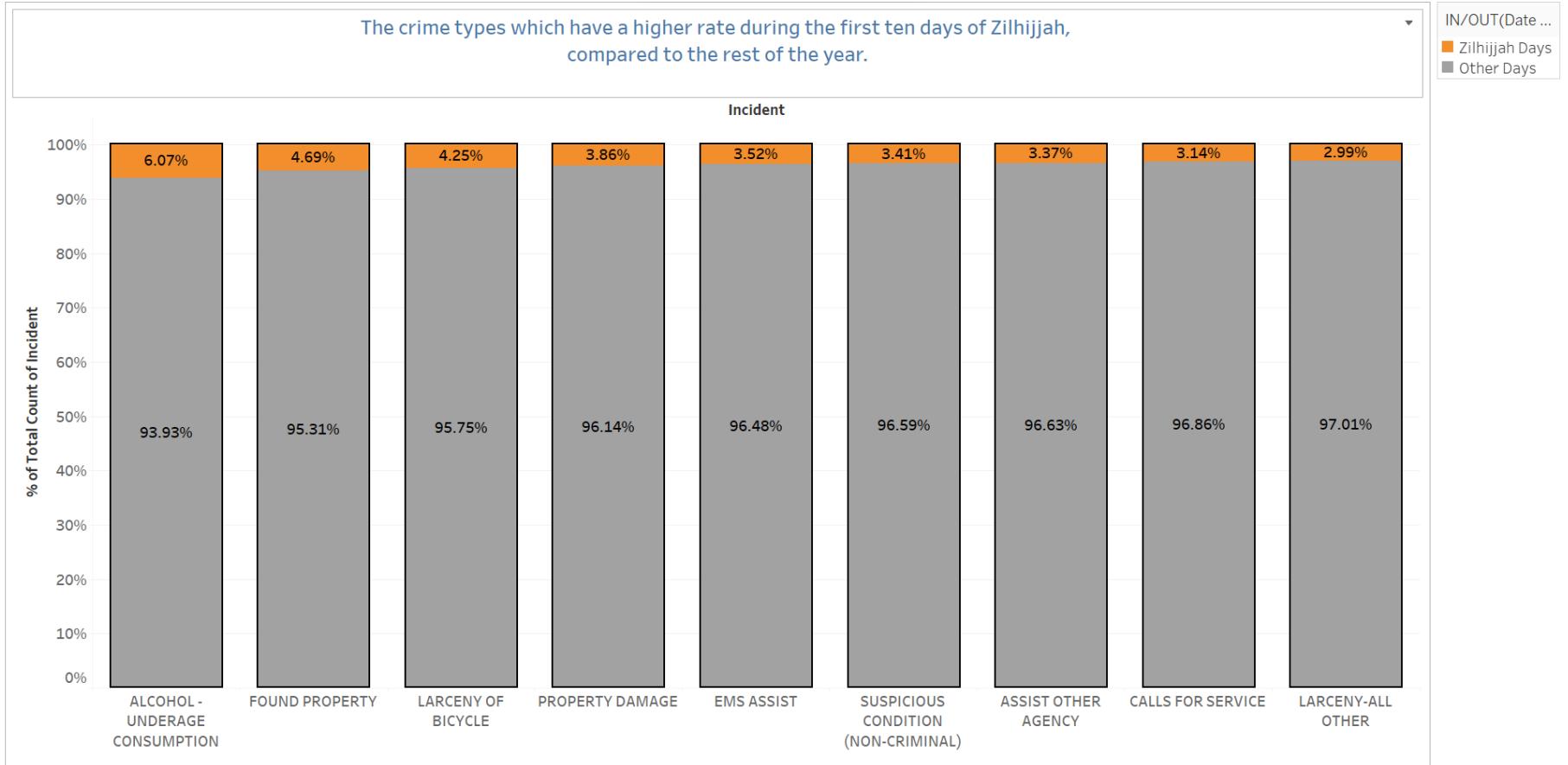
We observe a 14.37% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [23]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
```

Top 30 Incidents by Type



Incident Types



```
In [24]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio
sorted_ratios
```

Out[24]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
VIOLATION OF CAMPUS POLICY	1	1	1.000000
ASSAULT-PHYSICAL ASSAULT W/ SEXUAL MOTIVES (TOUCHING)	1	1	1.000000
REMOVAL OF SMOKE DETECTOR	1	1	1.000000
REQ. FOR SVC. REF. WELL-BEING CHK.	1	1	1.000000
SECRETLY PEEPING	1	1	1.000000
BELATED INJURY	1	1	1.000000
LARCENY OF MACBOOK AIR	1	1	1.000000
CLERY - ALLEGATION OF ROBBERY	1	1	1.000000
COUNTERFEITING - POSSESSING/CONCEALING	1	1	1.000000
SUICIDAL THOUGHTS	2	3	0.666700
FRAUD - POSSESSION OF A FAKE ID	1	2	0.500000
COMPUTER CRIMES	1	2	0.500000
DWI CHECK POINT	1	2	0.500000
SEATBELT CHECK POINT	1	3	0.333300
SUICIDE	1	3	0.333300
VANDALISM	2	7	0.285700
SMOKE INVESTIGATION	1	4	0.250000
POSSESSION OF SCHEDULE II	1	4	0.250000
CONCEALED WEAPON	1	5	0.200000
BURGLARY FORCIBLE ENTRY (STRUCTURES)	1	5	0.200000
B&E FROM BUILDING	1	5	0.200000
RAPE	1	6	0.166700
TRAFFIC (EXCEPT DWI)	1	6	0.166700
STOLEN PROPERTY - POSSESSION OF STOLEN GOODS	1	6	0.166700

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
TRESPASS WARNING	2	13	0.153800
RESIST DELAY OBSTRUCT	3	20	0.150000
FIRE	1	7	0.142900
VANDALISM DEFACING PUBLIC PROPERTY	1	8	0.125000
FRAUD - ALL OTHER	2	18	0.111100
HARASSMENT / CYBERSTALKING	4	38	0.105300

```
In [25]: # In which categories were more crimes committed during the first ten days of Zilhijjah?
zilhijjah_dominant_incidents
```

Out[25]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
EMS ASSIST	66	1875	0.0352
ALCOHOL - UNDERAGE CONSUMPTION	33	544	0.0607
FOUND PROPERTY	28	597	0.0469
LARCENY OF BICYCLE	13	306	0.0425
SUSPICIOUS CONDITION (NON-CRIMINAL)	10	293	0.0341
PROPERTY DAMAGE	10	259	0.0386
ASSIST OTHER AGENCY	7	208	0.0337
DRUGS - DRUG VIOLATION	5	50	0.1000
HARASSMENT	5	71	0.0704
CALLS FOR SERVICE	5	159	0.0314
LARCENY-ALL OTHER	4	134	0.0299
HARASSMENT / CYBERSTALKING	4	38	0.1053
LARCENY - FROM MOTOR VEHICLE (B&E)	4	69	0.0580
INVOLUNTARY COMMITMENT	3	50	0.0600
ALCOHOL - UNDERAGE POSSESSION	3	48	0.0625
SATURATION PATROL	3	50	0.0600
FRAUD - OBTAINING MONEY/PROPERTY BY FALSE PRETENSE	3	41	0.0732
RESIST DELAY OBSTRUCT	3	20	0.1500
B&E, LARCENY FROM A VEHICLE	2	50	0.0400
FRAUD - ALL OTHER	2	18	0.1111
CLERY - ALLEGATION OF RAPE	2	25	0.0800
FIRE ALARM	2	37	0.0541
LOST / FOUND PROPERTY	2	44	0.0455
SUICIDAL THOUGHTS	2	3	0.6667

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
MISDEMEANOR LARCENY	2	50	0.0400
TRESPASS WARNING	2	13	0.1538
VANDALISM	2	7	0.2857
POSSESSION OF MARIJUANA	2	33	0.0606
REMOVAL OF SMOKE DETECTOR	1	1	1.0000
RAPE - FORCIBLE	1	11	0.0909
SEATBELT CHECK POINT	1	3	0.3333
REQ. FOR SVC. REF. WELL-BEING CHK.	1	1	1.0000
SIMPLE POSSESSION OF MARIJUANA	1	28	0.0357
SECRETLY PEEPING	1	1	1.0000
SIMPLE ASSAULT	1	15	0.0667
ASSAULT-PHYSICAL ASSAULT W/ SEXUAL MOTIVES (TOUCHING)	1	1	1.0000
SMOKE INVESTIGATION	1	4	0.2500
STOLEN PROPERTY - POSSESSION OF STOLEN GOODS	1	6	0.1667
SUICIDE	1	3	0.3333
ASSAULT-PHYSICAL (NON-AGGRAVATED)	1	30	0.0333
TRAFFIC (EXCEPT DWI)	1	6	0.1667
VANDALISM DEFACING PUBLIC PROPERTY	1	8	0.1250
RAPE	1	6	0.1667
CLERY - ALLEGATION OF STALKING	1	17	0.0588
POSSESSION OF SCHEDULE II	1	4	0.2500
ORDER FOR ARREST	1	13	0.0769
COMPUTER CRIMES	1	2	0.5000
CONCEALED WEAPON	1	5	0.2000

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
COUNTERFEITING - POSSESSING/CONCEALING	1	1	1.0000
DAYTIME CHECK POINT (PART)	1	17	0.0588
CLERY - ALLEGATION OF SEX OFFENSE	1	11	0.0909
DWI CHECK POINT	1	2	0.5000
CLERY - ALLEGATION OF ROBBERY	1	1	1.0000
FIGHTING (AFFRAY)	1	10	0.1000
FIRE	1	7	0.1429
FRAUD - POSSESSION OF A FAKE ID	1	2	0.5000
BURGLARY FORCIBLE ENTRY (STRUCTURES)	1	5	0.2000
BELATED INJURY	1	1	1.0000
LARCENY - SHOPLIFTING	1	11	0.0909
LARCENY OF CELL PHONE	1	20	0.0500
LARCENY OF MACBOOK AIR	1	1	1.0000
B&E FROM BUILDING	1	5	0.2000
CLERY - REPORT RAPE	1	29	0.0345
VIOLATION OF CAMPUS POLICY	1	1	1.0000

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

```
In [26]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]
```

```
Out[26]: (730, 64)
```

```
In [27]: zilhijjah_incidents_desc
```

```
Out[27]: count      312
unique     81
top      EMS ASSIST
freq      66
Name: incident, dtype: object
```

In [28]: `other_days_incidents_desc`

Out[28]:

count	9206
unique	721
top	EMS ASSIST
freq	1809
Name:	incident, dtype: object

UNC dataset encompasses 730 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 81 incident categories, with 64 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2013	2014	2015	2016	2017	2018
Zilhijjah Days	412	58	79	71	68	93	43
Other Days	12,302	2,051	2,160	2,197	2,156	2,398	1,340

412 incidents / 60 days = 6.8667 incidents per Zilhijjah day
 12302 incidents / (2109-60) days = 6.0039 incidents per other day
 $6.8667 / 6.0039 = 14.37\%$

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% $[0.02817 = (10 \text{ days} / 355 \text{ days})]$

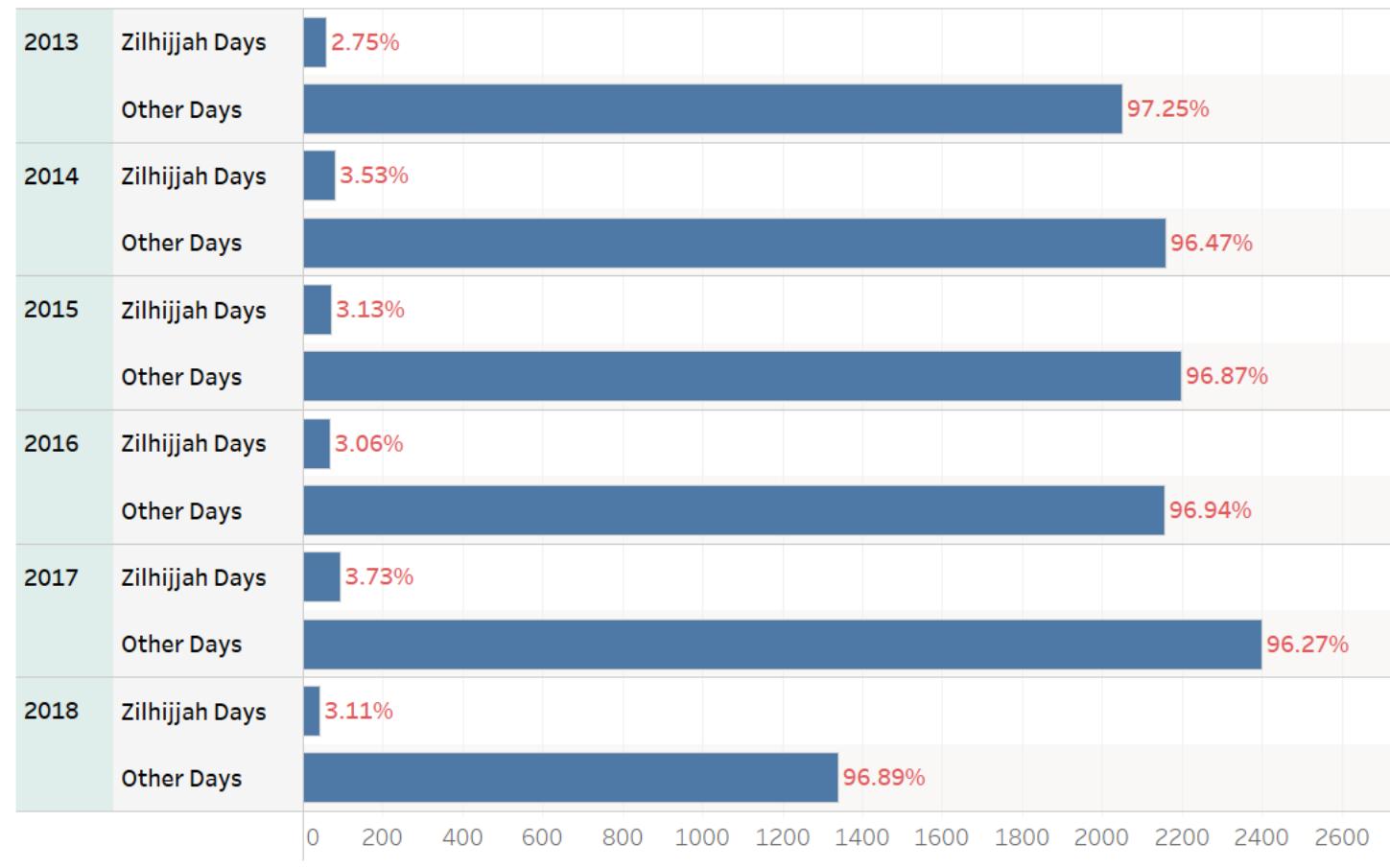
	Total Incidents	2013	2014	2015	2016	2017	2018
Zilhijjah Days	3.24%	2.75%	3.53%	3.13%	3.06%	3.73%	3.11%
Other Days	96.76%	97.25%	96.47%	96.87%	96.94%	96.27%	96.89%

Our dataset has missing days for the year 2018, hence the Equal Rate will not be 0.02817. The ratio of 0.02817 (10/355) will be used for complete years.

Equal Rate = Count Zilhijjah days / Count other days = 60 / 2049 = 0.029282
 $3.24 / 96.76 = 0.033485 > \text{Equal Rate } 0.033485 / 0.029282 = 14.35\%$

Crime Rates: Zilhijjah Days vs Other Days

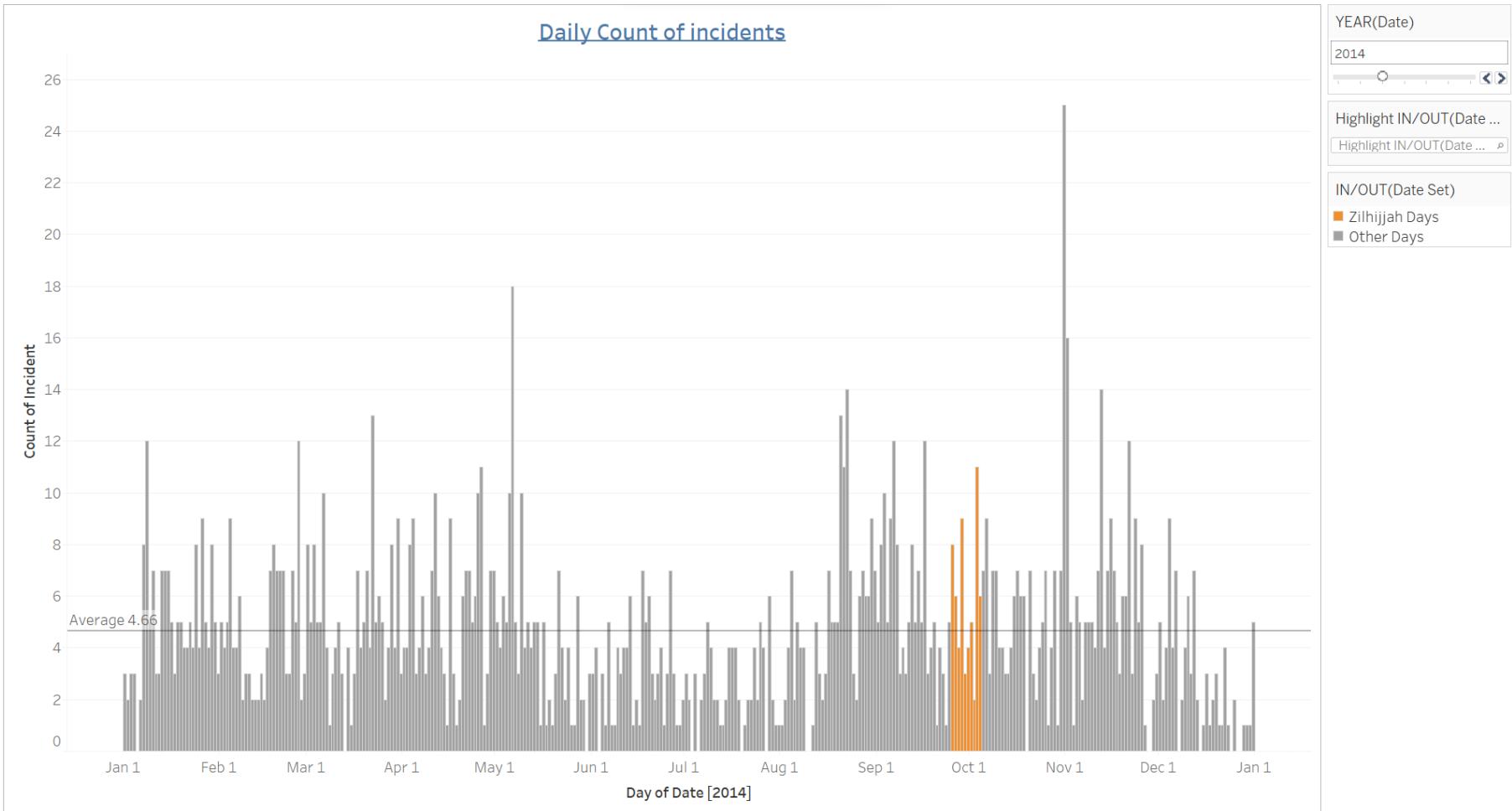
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

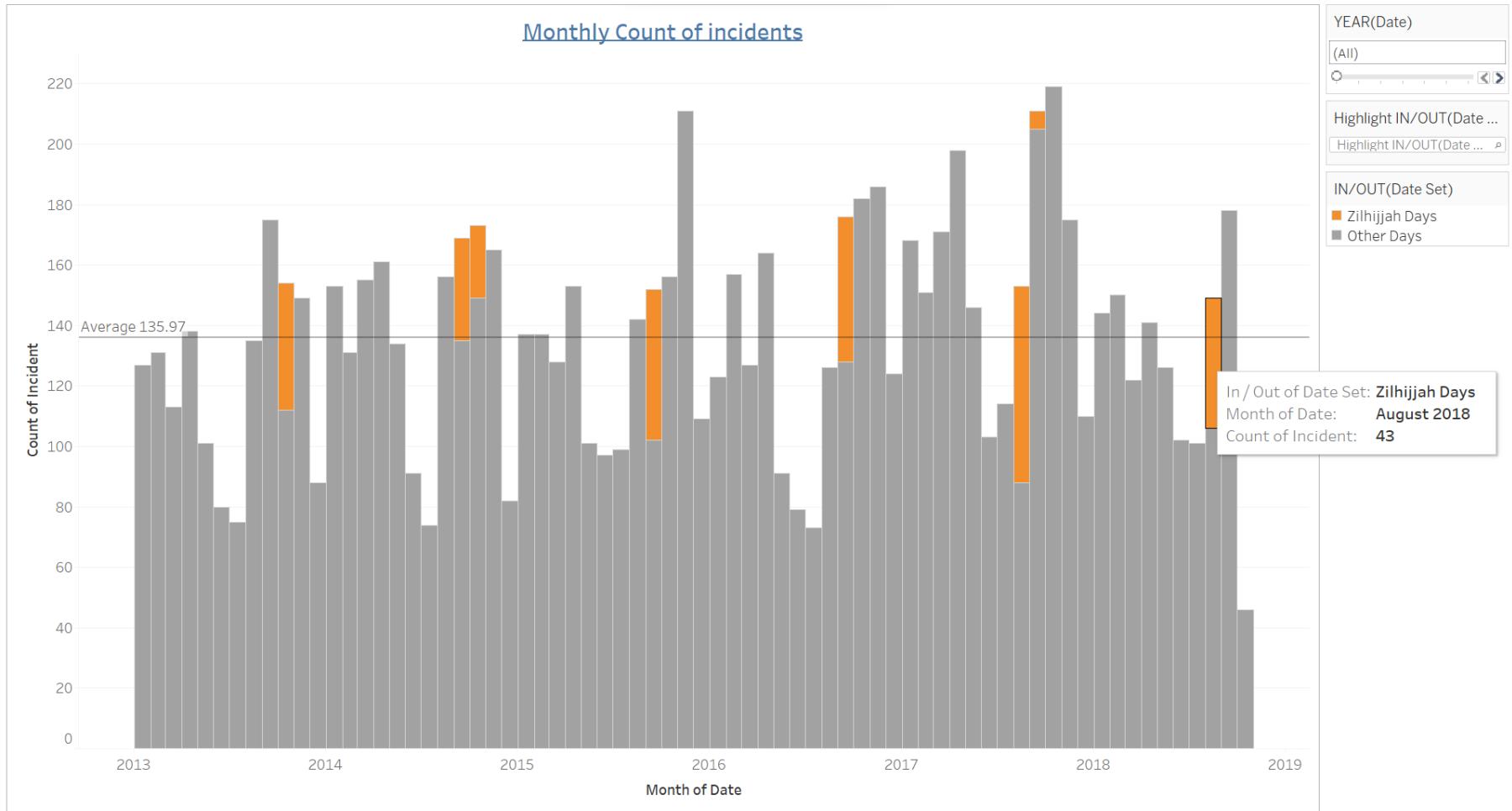


Caption

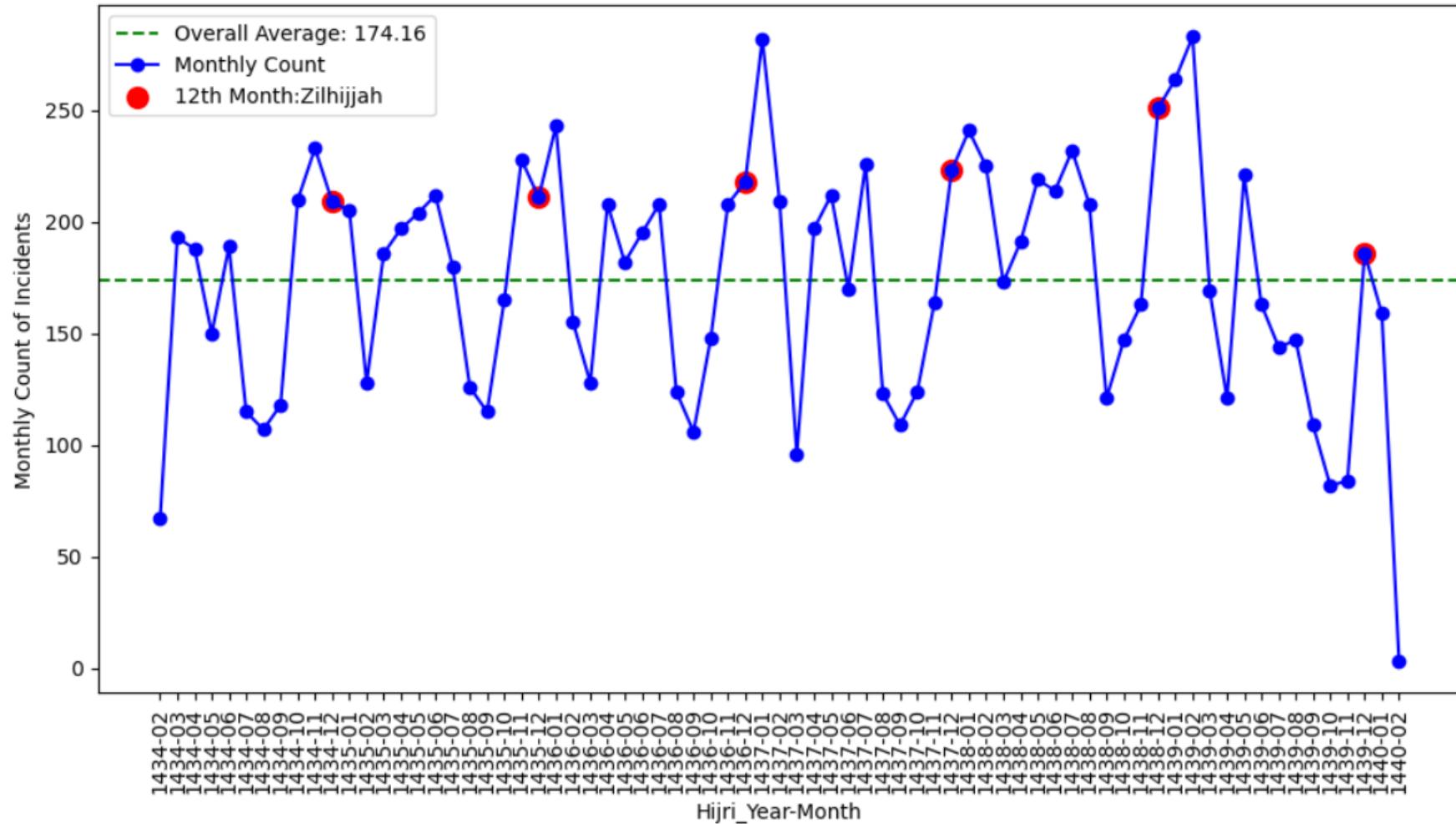
In comparison to other days of the year, the number of crimes committed during the first ten days of the month of Zilhijjah was higher in all years.

2013:
 $2.75/97.25=0.02827 > 0.02817$

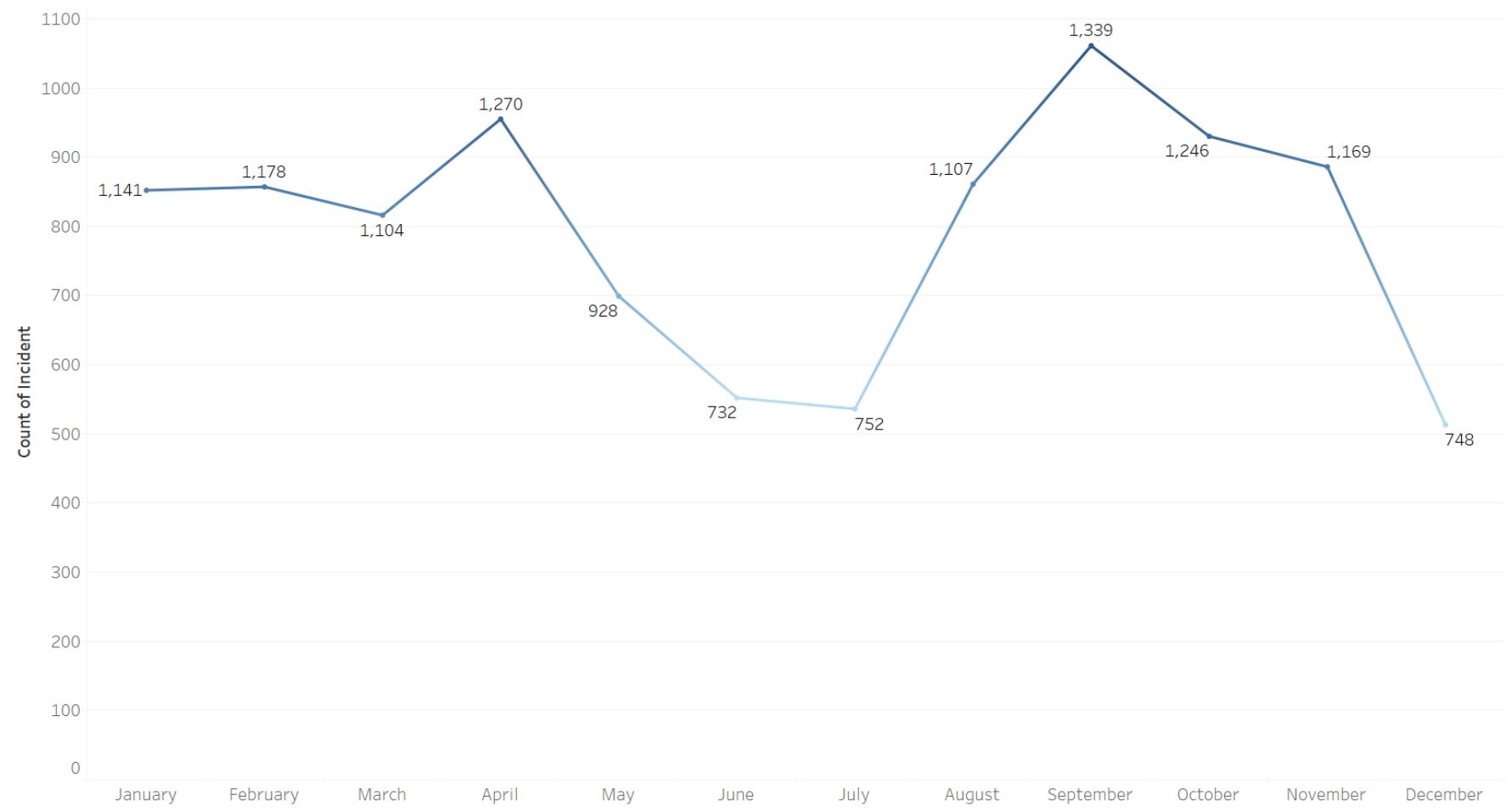


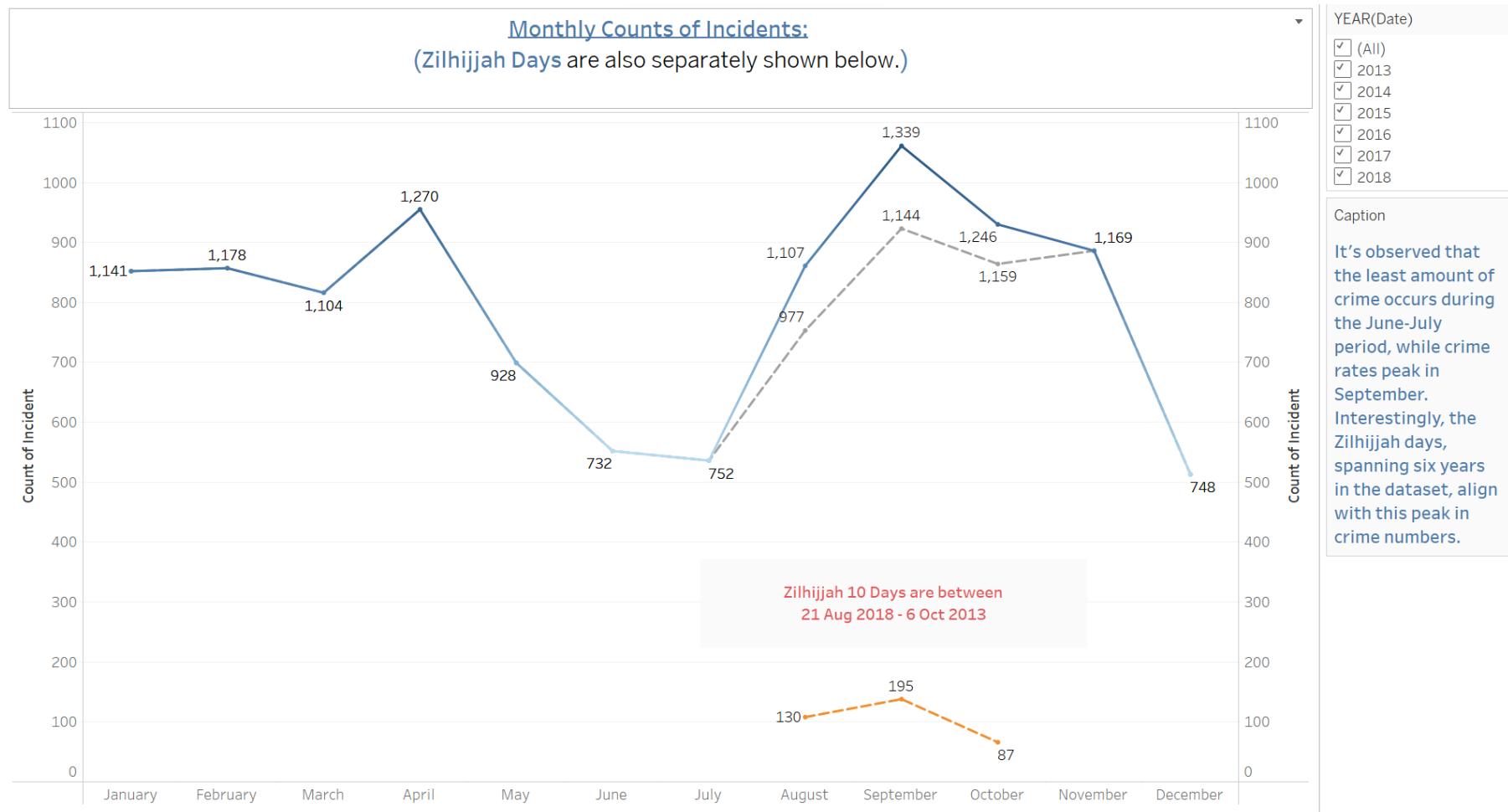


Monthly Count of Incidents on Hijri Calendar



In which month of the year are the most crimes committed?
Monthly Counts of Incidents: (Seasonality in Gregorian Calendar)





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

Out[30]:

	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	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 [31]: df.duplicated().value_counts()

Out[31]: False 1993259
dtype: int64

In [32]: 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 [33]: df = pd.read_csv("Los_Angeles_crime-data-from-2010-to-present.csv", usecols=[0,2,8])
```

```
In [34]: df["Crime Code Description"].value_counts()[:15]
```

```
Out[34]:
```

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

```
In [36]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [37]: min(df.date), max(df.date)
```

```
Out[37]: ('2010-01-01', '2019-06-22')
```

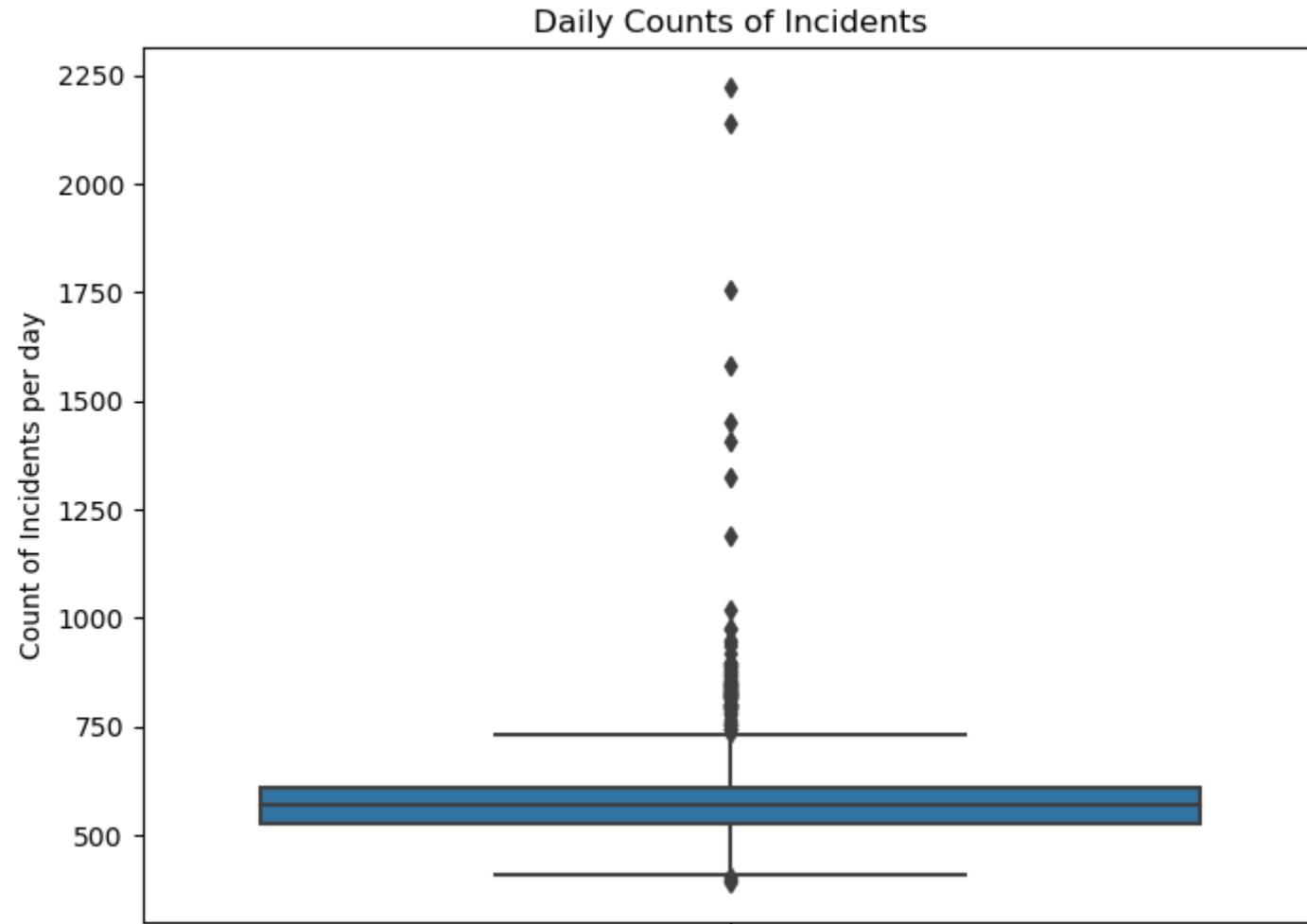
```
# The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11' '2017-08-23' - '2017-09-01' '2018-08-12' - '2018-08-21'
```

```
In [38]: # df.to_csv("Ls_Angles.csv", index=False)
```

```
In [39]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[39]: 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 [40]: # 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 [41]: df.date.unique()
```

```
Out[41]: 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 [42]: zilhijjah_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 Zilhijjah months: 580.5111

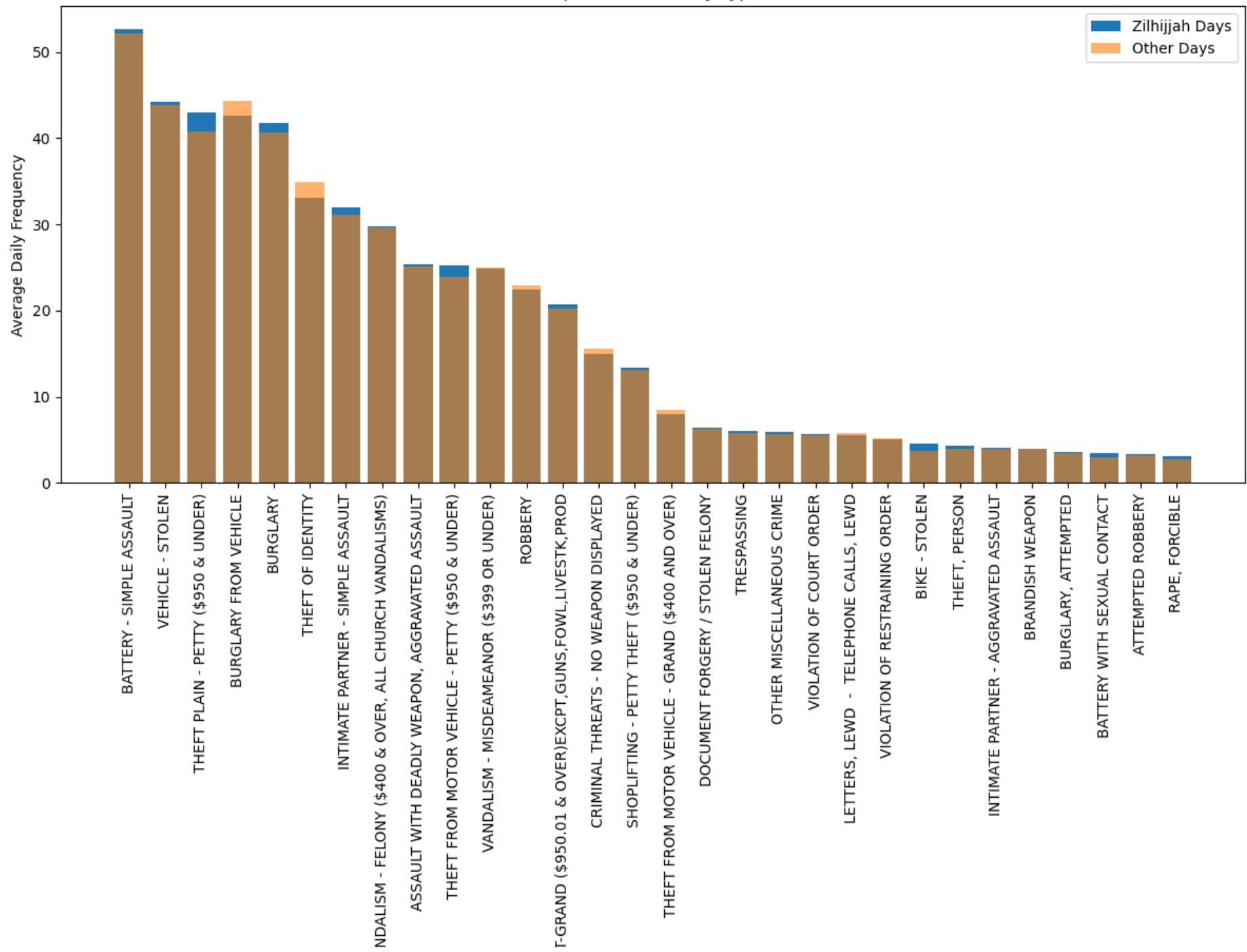
Average case count in other days: 575.9682

Ratio of Zilhijjah cases to other cases: 1.0079

We observe a 0.79% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [43]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
```

Top 30 Incidents by Type



Incident Types

```
In [44]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio  
sorted_ratios
```

Out[44]:

		zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
	BIGAMY	1	13	0.076900
	CHILD ABANDONMENT	6	97	0.061900
	TELEPHONE PROPERTY - DAMAGE	2	35	0.057100
	BIKE - ATTEMPTED STOLEN	2	37	0.054100
	FAILURE TO DISPERSE	1	20	0.050000
	DISHONEST EMPLOYEE - PETTY THEFT	6	122	0.049200
	DISRUPT SCHOOL	2	41	0.048800
	PICKPOCKET, ATTEMPT	1	22	0.045500
	PURSE SNATCHING - ATTEMPT	2	46	0.043500
	BOAT - STOLEN	11	270	0.040700
	BEASTIALITY, CRIME AGAINST NATURE SEXUAL ASSLT WITH ANIM	1	25	0.040000
	PANDERING	11	299	0.036800
	DRIVING WITHOUT OWNER CONSENT (DWOC)	16	450	0.035600
	THEFT FROM PERSON - ATTEMPT	10	282	0.035500
	LEWD CONDUCT	45	1288	0.034900
	HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE	3	87	0.034500
	DOCUMENT WORTHLESS (\$200 & UNDER)	2	60	0.033300
	KIDNAPPING	64	1929	0.033200
	INDECENT EXPOSURE	108	3307	0.032700
	TILL TAP - PETTY (\$950 & UNDER)	3	93	0.032300
	BIKE - STOLEN	414	12996	0.031900
	ORAL COPULATION	52	1641	0.031700
	PEEPING TOM	35	1113	0.031400
	ILLEGAL DUMPING	15	480	0.031200

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
SHOTS FIRED AT INHABITED DWELLING	75	2424	0.030900
DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER \$400	7	230	0.030400
CHILD STEALING	32	1053	0.030400
CRUELTY TO ANIMALS	38	1251	0.030400
BATTERY WITH SEXUAL CONTACT	315	10529	0.029900
VEHICLE - ATTEMPT STOLEN	95	3190	0.029800

In [45]: *# In which categories were more crimes committed during the first ten days of Zilhijjah?*
zilhijjah_dominant_incidents

Out[45]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
BATTERY - SIMPLE ASSAULT	4741	180434	0.0263
VEHICLE - STOLEN	3976	151622	0.0262
THEFT PLAIN - PETTY (\$950 & UNDER)	3871	141489	0.0274
BURGLARY	3756	140926	0.0267
INTIMATE PARTNER - SIMPLE ASSAULT	2874	107900	0.0266
VANDALISM - FELONY (\$400 & OVER, ALL CHURCH VANDALISMS)	2682	102589	0.0261
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	2286	86829	0.0263
THEFT FROM MOTOR VEHICLE - PETTY (\$950 & UNDER)	2276	82791	0.0275
THEFT-GRAND (\$950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	1863	70081	0.0266
SHOPLIFTING - PETTY THEFT (\$950 & UNDER)	1206	45493	0.0265
DOCUMENT FORGERY / STOLEN FELONY	576	21940	0.0263
TRESPASSING	541	19965	0.0271
OTHER MISCELLANEOUS CRIME	533	19751	0.0270
VIOLATION OF COURT ORDER	507	19097	0.0265
BIKE - STOLEN	414	12996	0.0319
THEFT, PERSON	386	13884	0.0278
INTIMATE PARTNER - AGGRAVATED ASSAULT	365	13736	0.0266
BURGLARY, ATTEMPTED	318	12010	0.0265
BATTERY WITH SEXUAL CONTACT	315	10529	0.0299
ATTEMPTED ROBBERY	297	11361	0.0261
RAPE, FORCIBLE	275	9583	0.0287
CHILD ABUSE (PHYSICAL) - SIMPLE ASSAULT	251	8640	0.0291
EMBEZZLEMENT, GRAND THEFT (\$950.01 & OVER)	212	7359	0.0288
CHILD ANNOYING (17YRS & UNDER)	140	4890	0.0286

		zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
	BUNCO, PETTY THEFT	138	4825	0.0286
	CHILD NEGLECT (SEE 300 W.I.C.)	133	5092	0.0261
	INDECENT EXPOSURE	108	3307	0.0327
	VEHICLE - ATTEMPT STOLEN	95	3190	0.0298
	DISTURBING THE PEACE	94	3483	0.0270
	CONTEMPT OF COURT	94	3587	0.0262
	RESISTING ARREST	83	3122	0.0266
	SHOTS FIRED AT INHABITED DWELLING	75	2424	0.0309
	KIDNAPPING	64	1929	0.0332
	ORAL COPULATION	52	1641	0.0317
	THEFT PLAIN - ATTEMPT	47	1589	0.0296
	LEWD CONDUCT	45	1288	0.0349
	ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER	44	1542	0.0285
	CHILD ABUSE (PHYSICAL) - AGGRAVATED ASSAULT	44	1568	0.0281
	CRUELTY TO ANIMALS	38	1251	0.0304
	PEEPING TOM	35	1113	0.0314
	CHILD STEALING	32	1053	0.0304
	PURSE SNATCHING	32	1139	0.0281
	BOMB SCARE	31	1126	0.0275
	PICKPOCKET	26	899	0.0289
	PROWLER	26	889	0.0292
	KIDNAPPING - GRAND ATTEMPT	18	689	0.0261
	DRIVING WITHOUT OWNER CONSENT (DWOC)	16	450	0.0356
	ILLEGAL DUMPING	15	480	0.0312

		zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
	PANDERING	11	299	0.0368
	BOAT - STOLEN	11	270	0.0407
	FALSE POLICE REPORT	11	381	0.0289
	THEFT FROM PERSON - ATTEMPT	10	282	0.0355
	DEFRAUDING INNKEEPER/THEFT OF SERVICES, OVER \$400	7	230	0.0304
	CHILD ABANDONMENT	6	97	0.0619
	DISHONEST EMPLOYEE - PETTY THEFT	6	122	0.0492
	TILL TAP - PETTY (\$950 & UNDER)	3	93	0.0323
	HUMAN TRAFFICKING - INVOLUNTARY SERVITUDE	3	87	0.0345
	TELEPHONE PROPERTY - DAMAGE	2	35	0.0571
	DISRUPT SCHOOL	2	41	0.0488
	DOCUMENT WORTHLESS (\$200 & UNDER)	2	60	0.0333
	PURSE SNATCHING - ATTEMPT	2	46	0.0435
	BIKE - ATTEMPTED STOLEN	2	37	0.0541
	PICKPOCKET, ATTEMPT	1	22	0.0455
	DRUNK ROLL	1	34	0.0294
	FAILURE TO DISPERSE	1	20	0.0500
	BIGAMY	1	13	0.0769
	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 Zilhijjah compared to the other days of the year.

In [46]: `df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]`

Out[46]: `(140, 67)`

In [47]: `zilhijjah_incidents_desc`

```
Out[47]: count      52246
unique      120
top    BATTERY - SIMPLE ASSAULT
freq      4741
Name: incident, dtype: object
```

```
In [48]: other_days_incidents_desc
```

```
Out[48]: count      1941013
unique      140
top    BATTERY - SIMPLE ASSAULT
freq      175693
Name: incident, dtype: object
```

Los Angeles dataset encompasses 140 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 120 incident categories, with 67 of these categories experiencing incident counts exceeding the annual averages.

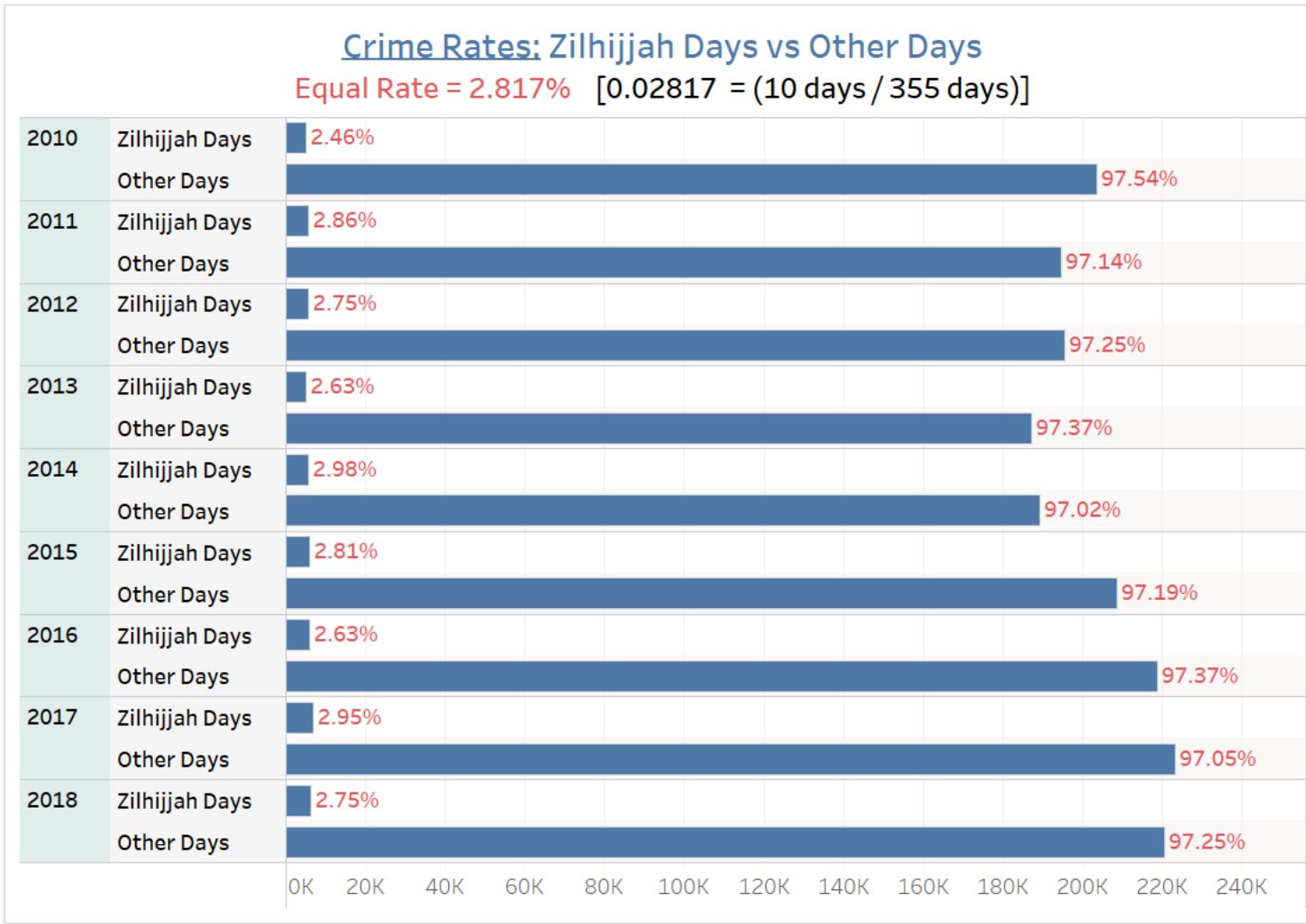
Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incide..	2010	2011	2012	2013	2014	2015	2016	2017	2018
Zilhijjah Days	52,246	5,147	5,734	5,538	5,059	5,806	6,032	5,912	6,778	6,240
Other Days	1,841,723	203,676	194,703	195,632	187,152	189,216	208,790	218,733	223,152	220,669

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incide..	2010	2011	2012	2013	2014	2015	2016	2017	2018
Zilhijjah Days	2.76%	2.46%	2.86%	2.75%	2.63%	2.98%	2.81%	2.63%	2.95%	2.75%
Other Days	97.24%	97.54%	97.14%	97.25%	97.37%	97.02%	97.19%	97.37%	97.05%	97.25%



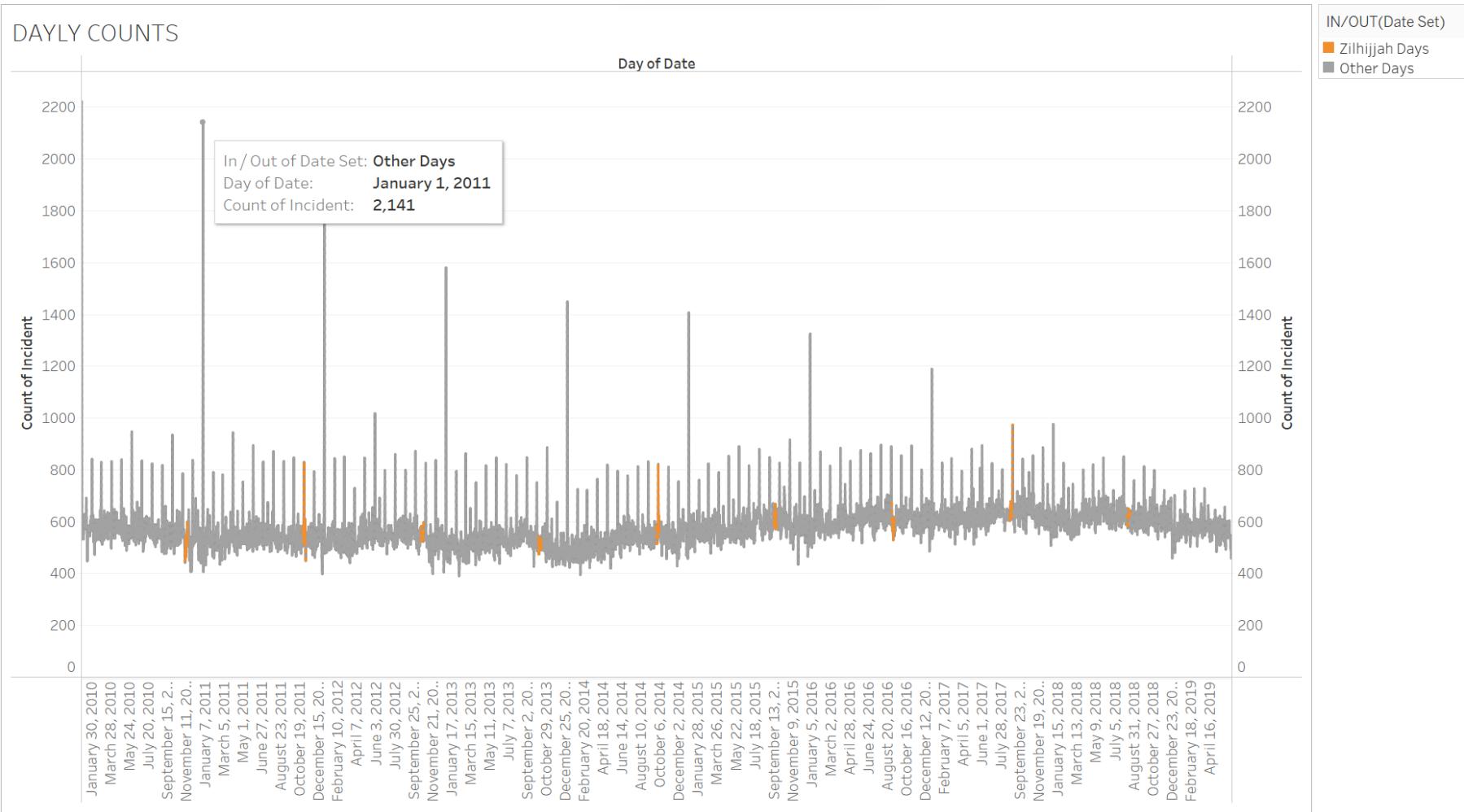
Caption

According to this graph, compared to the rest of the year, crime rates during the initial ten days of Zilhijjah were higher in all years except 2010, 2013, 2016.

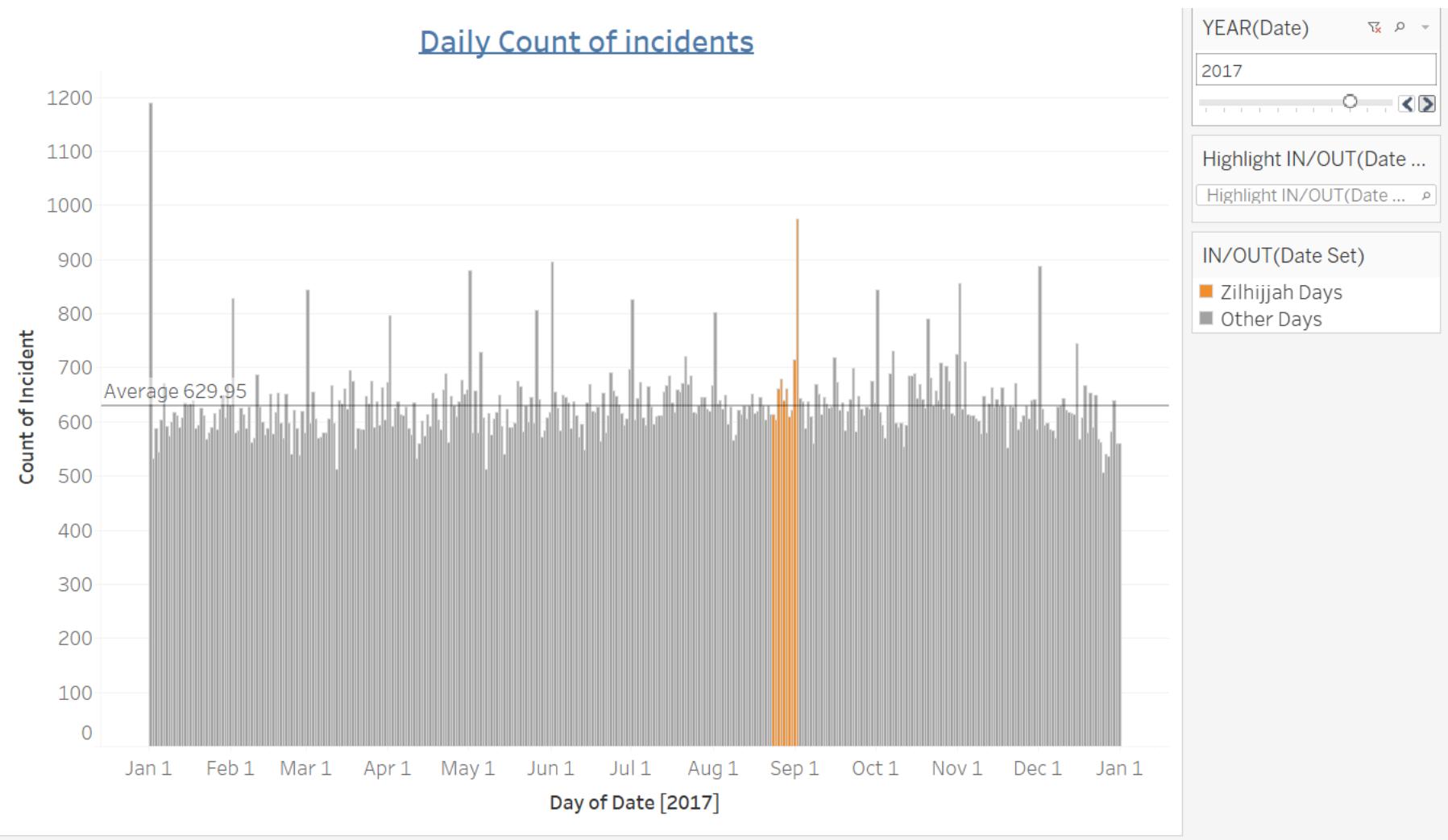
2013:
 $2.63/97.37=0.02710 < 0.02817$

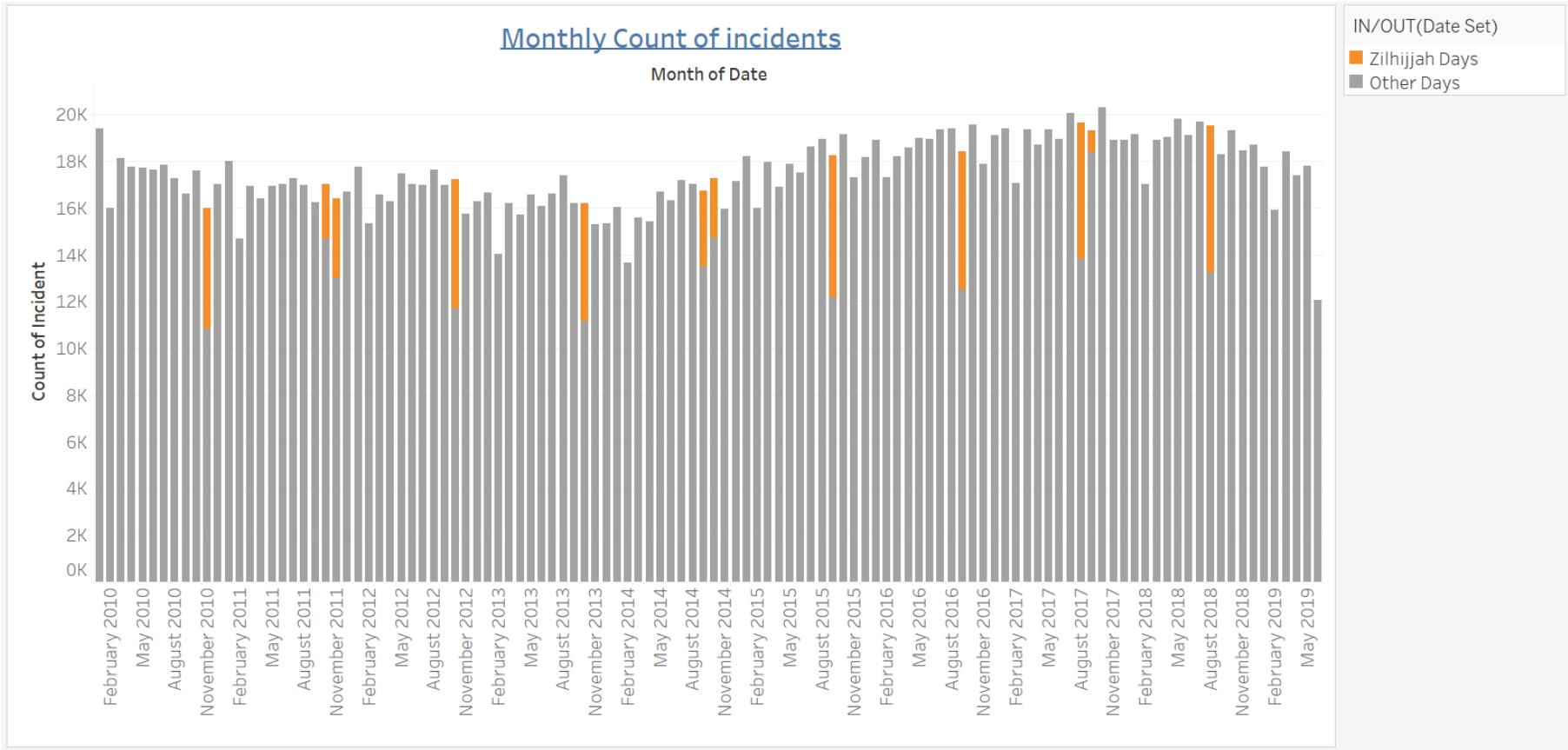
2012:
 $2.75/97.25=0.02827 > 0.02817$

This dataset indicates low disparity. In total, Zilhijjah days exhibit a marginal 0.79% increase.

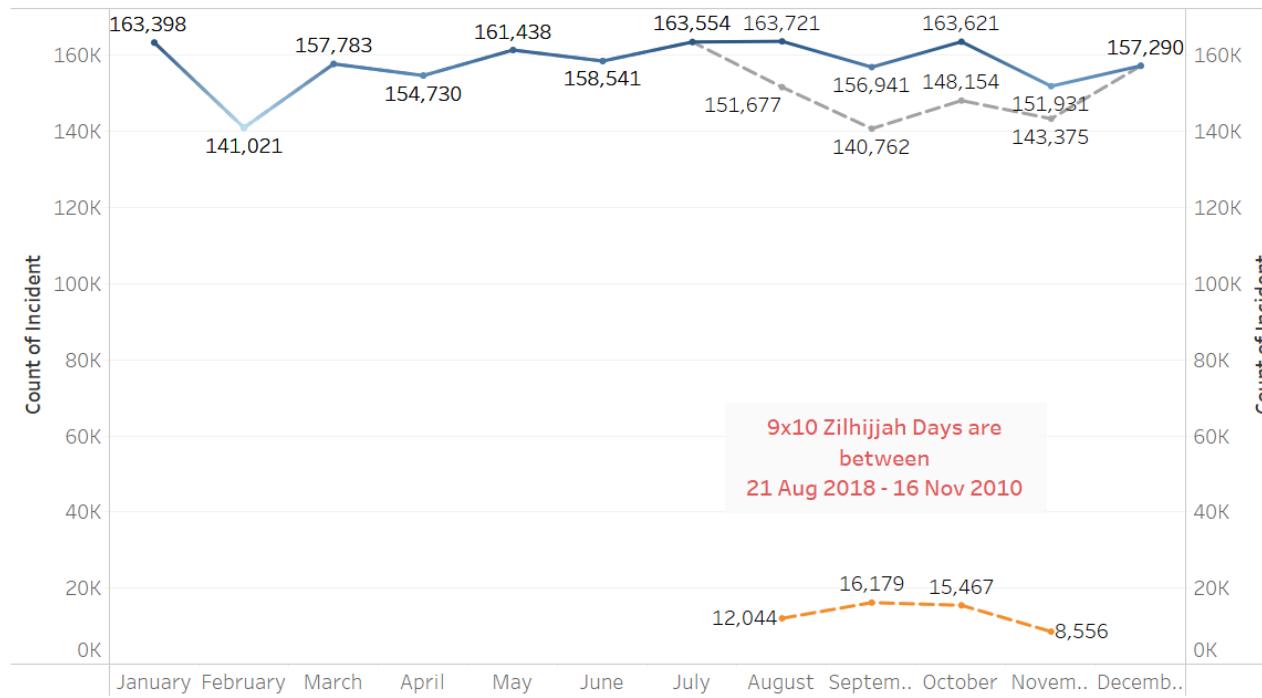


Daily Count of incidents





Monthly Counts of Incidents: (Zilhijjah Days are also separately shown below.)



Caption

During the period of August to November, Zilhijjah days mark an average 580 incidents per day.

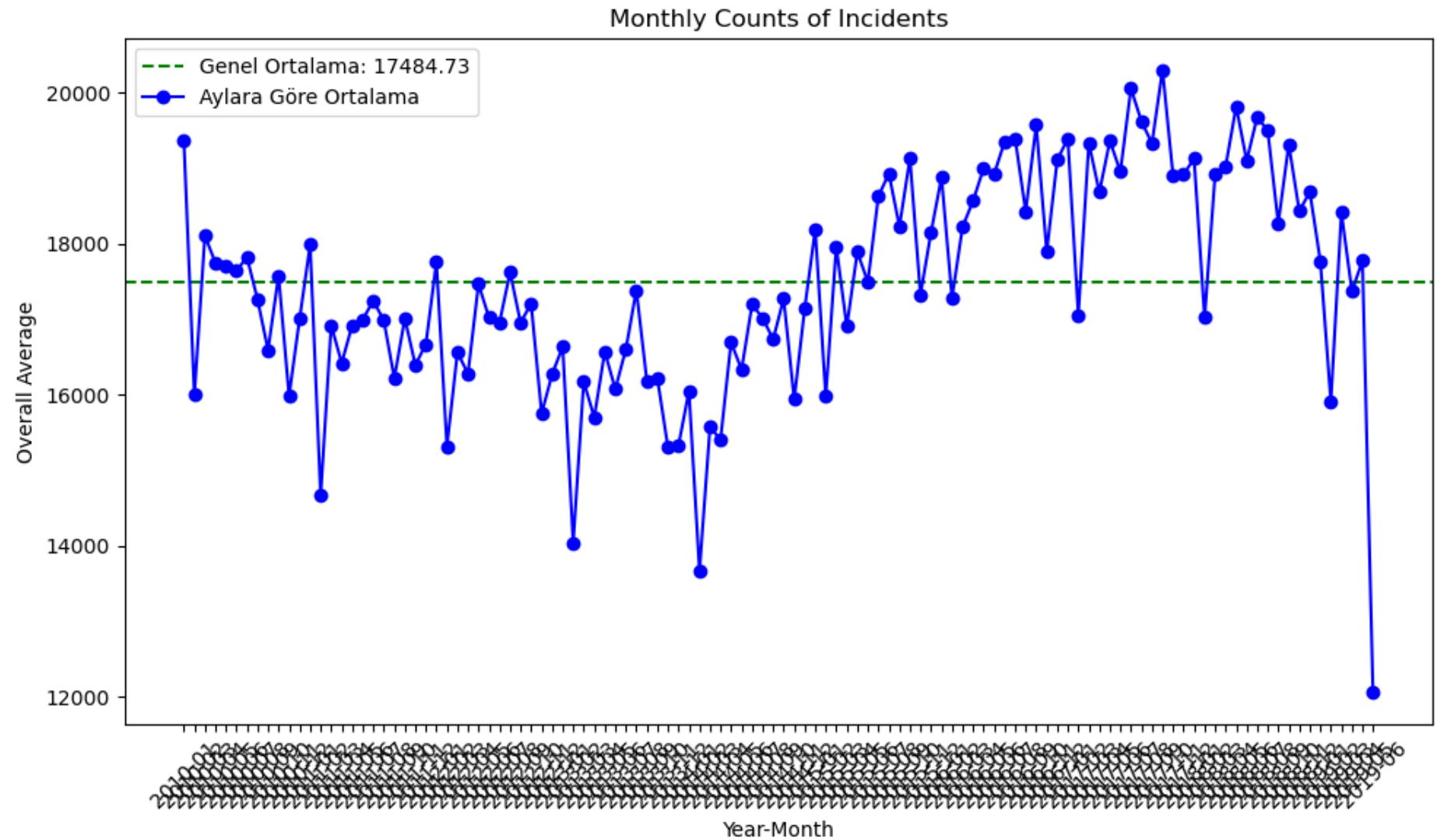
$$(12044+16179+15467+8556 \text{ incidents})/90 \text{ days} = 580.51$$

In [49]:

```
df['date'] = pd.to_datetime(df['date'])
monthly_count = df.groupby([df['date'].dt.year.rename('Year'), df['date'].dt.month.rename('Month')])['date'].count()
monthly_avg = monthly_count.groupby(['Year', 'Month']).mean()
monthly_avg.index = monthly_avg.index.map(lambda x: f"{x[0]}-{x[1]:02d}")
overall_avg = monthly_count.mean()

plt.figure(figsize=(10, 6))
plt.axhline(y=overall_avg, color='g', linestyle='--', label=f'Genel Ortalama: {overall_avg:.2f}')
plt.plot(monthly_avg.index, monthly_avg.values, marker='o', linestyle='-', color='b', label='Aylara Göre Ortalama')
plt.title('Monthly Counts of Incidents')
plt.xlabel('Year-Month')
plt.ylabel('Overall Average')
plt.xticks(rotation=45)
plt.legend()
```

```
plt.tight_layout()  
plt.show()
```



```
In [50]: monthly_count_plot()
```

Monthly Count of Incidents on Hijri Calendar



SAMPLE DATA-3: KANSAS CITY CRIME DATA_2009-2016

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

```
In [51]: 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 [52]: frames = [df2, df3, df4, df5, df6, df7, df8, df9]
df = pd.concat(frames)
```

```
In [53]: df.head()
```

Out[53]:

	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 [54]: df.duplicated().value_counts()
Out[54]: False 1007956 True 3518 dtype: int64
In [55]: df = df.drop_duplicates()
In [56]: 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 [57]: df["Description"].value_counts()[:15]
```

```
Out[57]:
```

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

```
In [59]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [60]: df.date.min(), df.date.max()
```

```
Out[60]: ('2009-01-01', '2016-11-06')
```

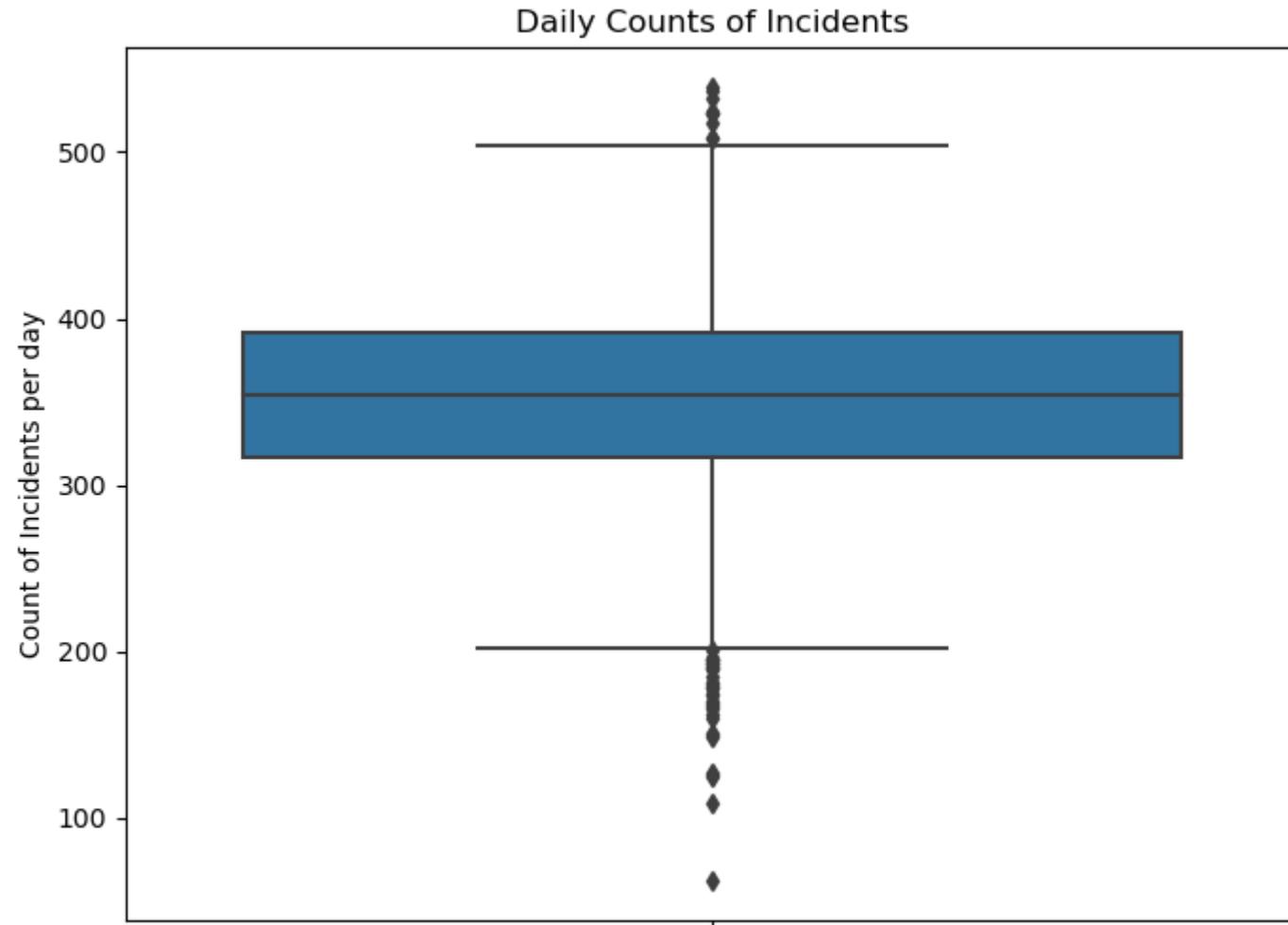
The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11'

```
In [61]: df = df.iloc[:, [1,9]]  
# df.to_csv("Kansas.csv", index=False)
```

```
In [62]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[62]: 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 [63]: # 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 [64]: df.date.unique()
```

```
Out[64]: 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 [65]: zilhijjah_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 Zilhijjah months: 372.375
```

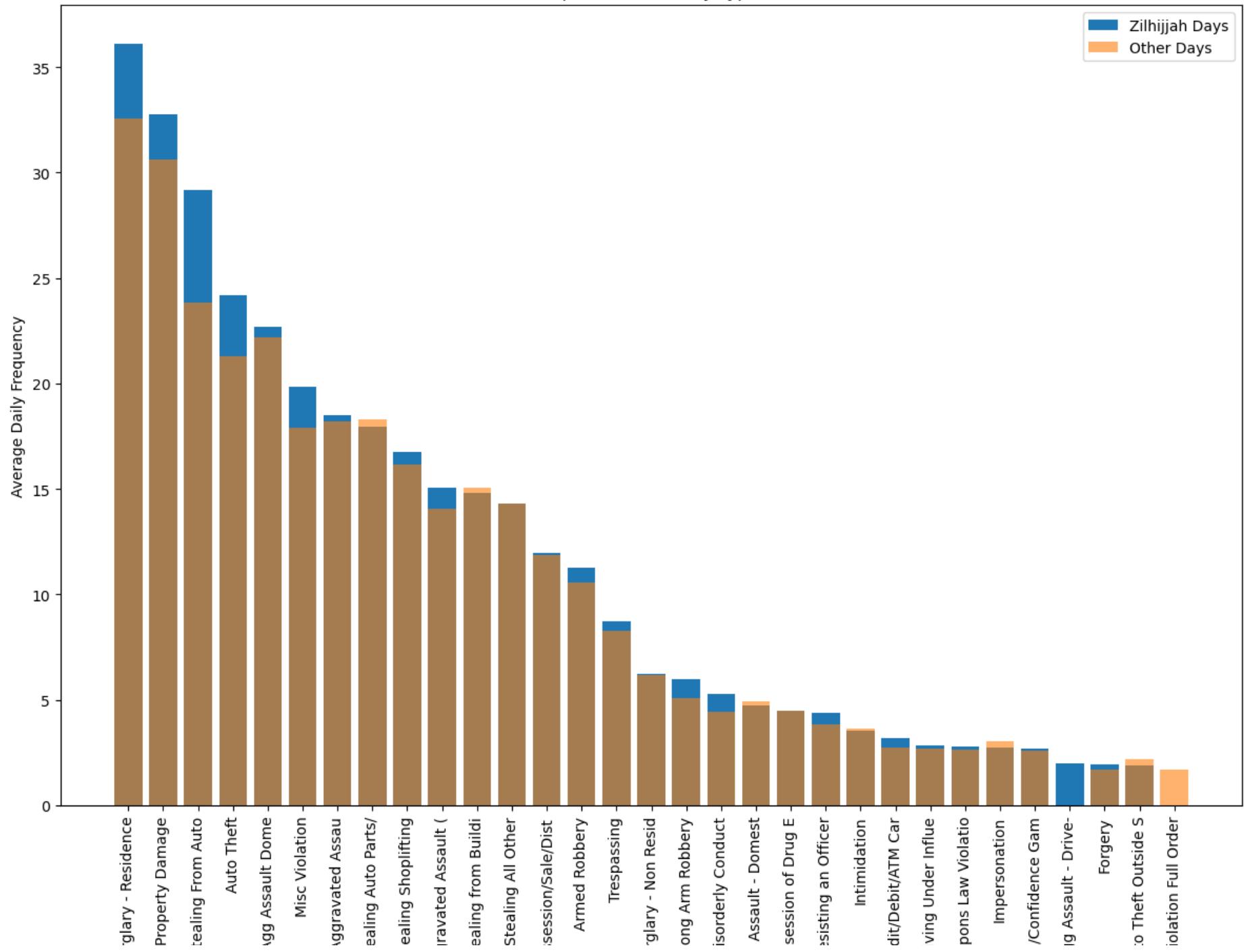
```
-----  
Average case count in other days: 350.9745
```

```
-----  
Ratio of Zilhijjah cases to other cases: 1.0610
```

We observe a 6.10% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [66]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)  
# display(sorted_ratios)  
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type





```
In [67]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio
sorted_ratios
```

Out[67]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
misc Inv	1	1	1.000000
counterfeit	2	2	1.000000
burglary	2	2	1.000000
Intimidation (DV)	2	2	1.000000
embezzlement	2	4	0.500000
armed robbery	4	11	0.363600
child endangerment	2	6	0.333300
Aggravated Assault	8	33	0.242400
Negligent Manslaughter	2	9	0.222200
Sex Off Incest	3	17	0.176500
Suicide By Cutting	1	9	0.111100
Attempt Suicide by G	1	9	0.111100
Suicide - Domestic V	1	11	0.090900
Hit and Run of a Per	9	121	0.074400
Passing Bad Checks	22	325	0.067700
Loitering	14	219	0.063900
Animal Cruelty	1	16	0.062500
Attempt Suicide by O	14	228	0.061400
Suicide by Other Mea	1	17	0.058800
Counterfeiting	43	862	0.049900
Interference with Cu	36	732	0.049200
Extortion/Blackmail	2	44	0.045500
Attempt Suicide by H	5	112	0.044600
Pornography	11	249	0.044200

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
Attempt Suicide by C	21	538	0.039000
Agg Assault - Drive-	159	4250	0.037400
Arson with Fire Bomb	5	142	0.035200
Failure to Return	24	683	0.035100
False ID	3	86	0.034900
Sex Off Indecent Exp	33	962	0.034300

```
In [68]: # In which categories were more crimes committed during the first ten days of Zilhijjah?
zilhijjah_dominant_incidents
```

Out[68]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
Burglary - Residence	2890	93609	0.0309
Property Damage	2621	87921	0.0298
Stealing From Auto	2335	68821	0.0339
Auto Theft	1935	61302	0.0316
Non Agg Assault Dome	1814	63608	0.0285
Misc Violation	1586	51407	0.0309
Non Aggravated Assau	1480	52272	0.0283
Stealing Shoplifting	1342	46401	0.0289
Aggravated Assault (1206	40332	0.0299
Possession/Sale/Dist	956	34013	0.0281
Armed Robbery	902	30324	0.0297
Trespassing	700	23800	0.0294
Burglary - Non Resid	498	17776	0.0280
Strong Arm Robbery	478	14691	0.0325
Disorderly Conduct	421	12846	0.0328
Resisting an Officer	349	11064	0.0315
Credit/Debit/ATM Car	255	7865	0.0324
Driving Under Influe	229	7747	0.0296
Weapons Law Violatio	225	7573	0.0297
Fraud/Confidence Gam	217	7380	0.0294
Agg Assault - Drive-	159	4250	0.0374
Forgery	157	4938	0.0318
Violation Full Order	144	4871	0.0296
Arson	126	4311	0.0292

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
Embezzlement	88	3100	0.0284
Wire Fraud	83	2635	0.0315
Prostitution/Patroni	55	1772	0.0310
HOMICIDE/Non Neglige	47	1558	0.0302
False Information to	46	1541	0.0299
Counterfeiting	43	862	0.0499
Interference with Cu	36	732	0.0492
Sex Off Indecent Exp	33	962	0.0343
Sex Off Misconduct	31	983	0.0315
Kidnapping/Abduction	28	874	0.0320
Failure to Return	24	683	0.0351
Passing Bad Checks	22	325	0.0677
Attempt Suicide by C	21	538	0.0390
Sex Offense -others	21	741	0.0283
Impersonation - NOT	16	540	0.0296
Loitering	14	219	0.0639
Attempt Suicide by O	14	228	0.0614
Pornography	11	249	0.0442
Prostitution/Solicit	11	384	0.0286
Hit and Run of a Per	9	121	0.0744
Aggravated Assault	8	33	0.2424
Arson with Fire Bomb	5	142	0.0352
Attempt Suicide by H	5	112	0.0446
Sex Off Follow/Entic	4	122	0.0328

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
armed robbery	4	11	0.3636
Sex Off Incest	3	17	0.1765
False ID	3	86	0.0349
Casualty (includes a	3	105	0.0286
counterfeit	2	2	1.0000
child endangerment	2	6	0.3333
Extortion/Blackmail	2	44	0.0455
embezzlement	2	4	0.5000
Negligent Manslaught	2	9	0.2222
Intimidation (DV)	2	2	1.0000
burglary	2	2	1.0000
Suicide - Domestic V	1	11	0.0909
Animal Cruelty	1	16	0.0625
Attempt Suicide by G	1	9	0.1111
Suicide by Other Mea	1	17	0.0588
Suicide By Cutting	1	9	0.1111
misc Inv	1	1	1.0000

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

In [69]: `df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]`

Out[69]: `(259, 65)`

In [70]: `zilhijjah_incidents_desc`

```
Out[70]: count          29790
unique         100
top    Burglary - Residence
freq           2890
Name: incident, dtype: object
```

```
In [71]: other_days_incidents_desc
```

```
Out[71]: count          978166
unique         255
top    Burglary - Residence
freq           90719
Name: incident, dtype: object
```

```
In [72]: zilhijjah_dominant_incidents.count()[0]
```

```
Out[72]: 65
```

Kansas dataset encompasses 259 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 100 incident categories, with 65 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	29,790	3,426	3,753	3,729	3,772	3,663	3,642	3,913	3,892
Other Days	978,166	129,109	132,303	124,343	126,518	121,069	120,590	117,629	106,605

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

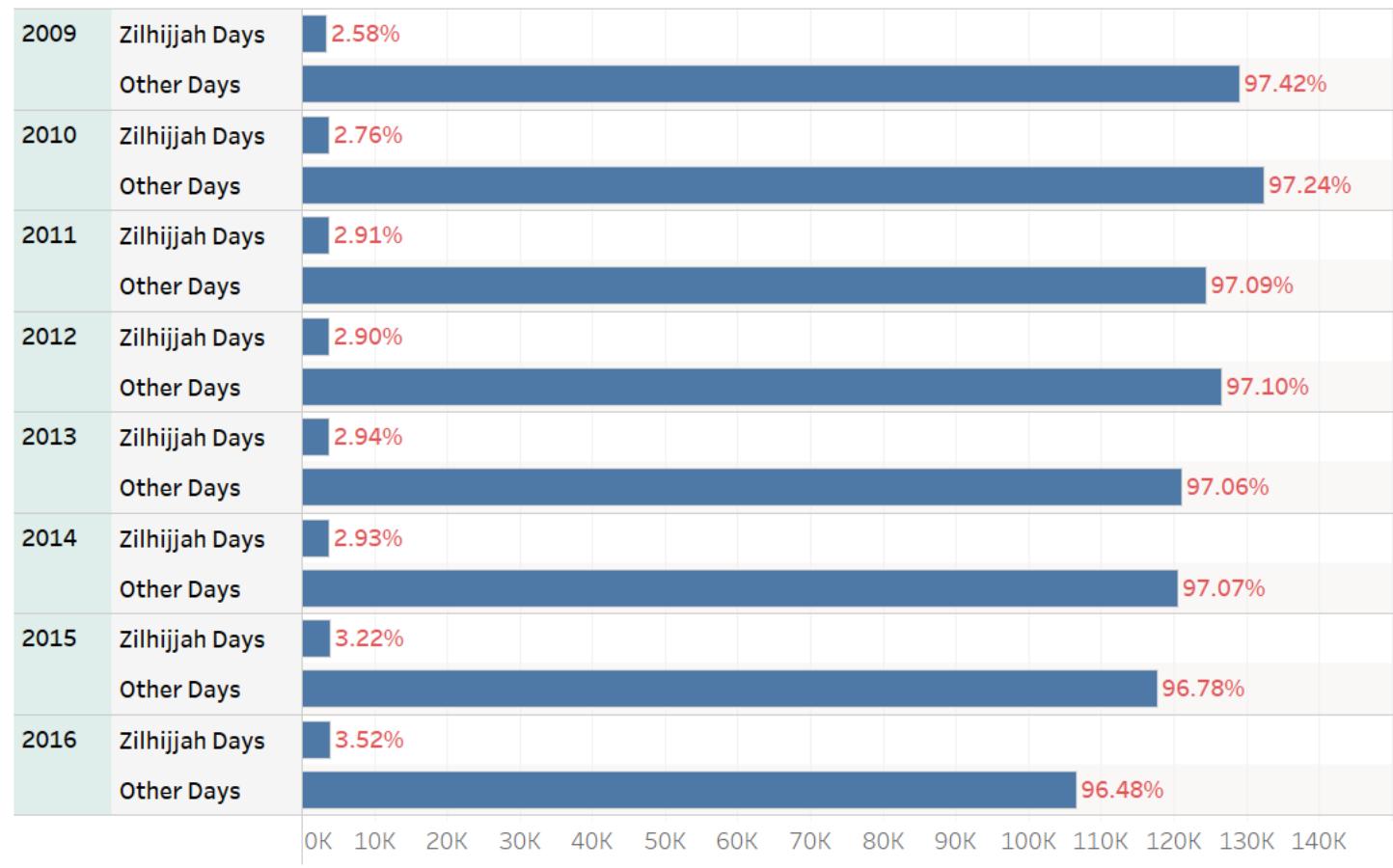
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incidents	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	2.96%	2.58%	2.76%	2.91%	2.90%	2.94%	2.93%	3.22%	3.52%
Other Days	97.04%	97.42%	97.24%	97.09%	97.10%	97.06%	97.07%	96.78%	96.48%

Days with no records are excluded in Tableau calculations.

Crime Rates: Zilhijjah Days vs Other Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

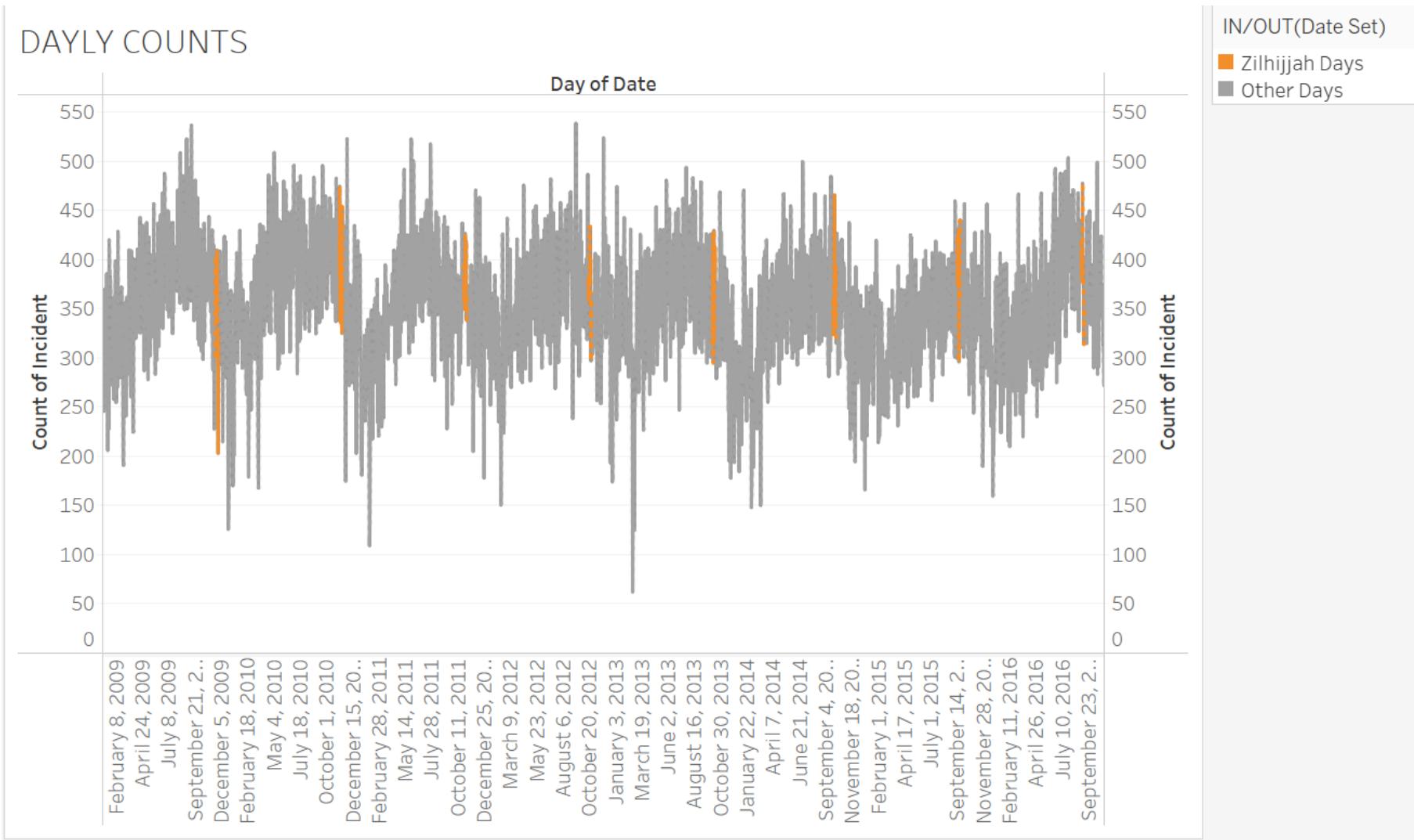


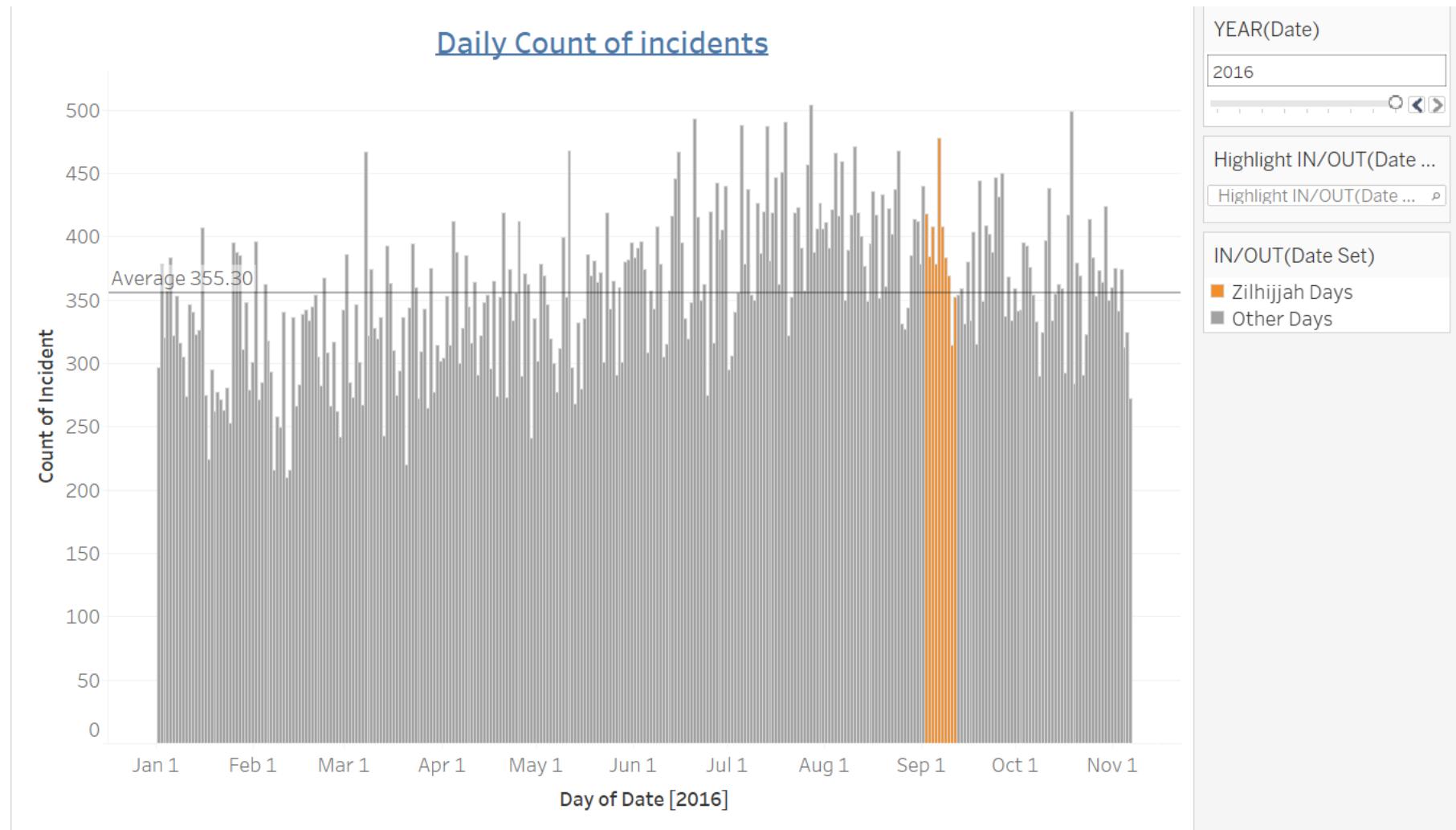
Caption

According to this graph, compared to the rest of the year, crime rates during the initial ten days of Zilhijjah were higher in all years except 2009.
 $(2.58/97.42=0.02648 < 0.02817)$

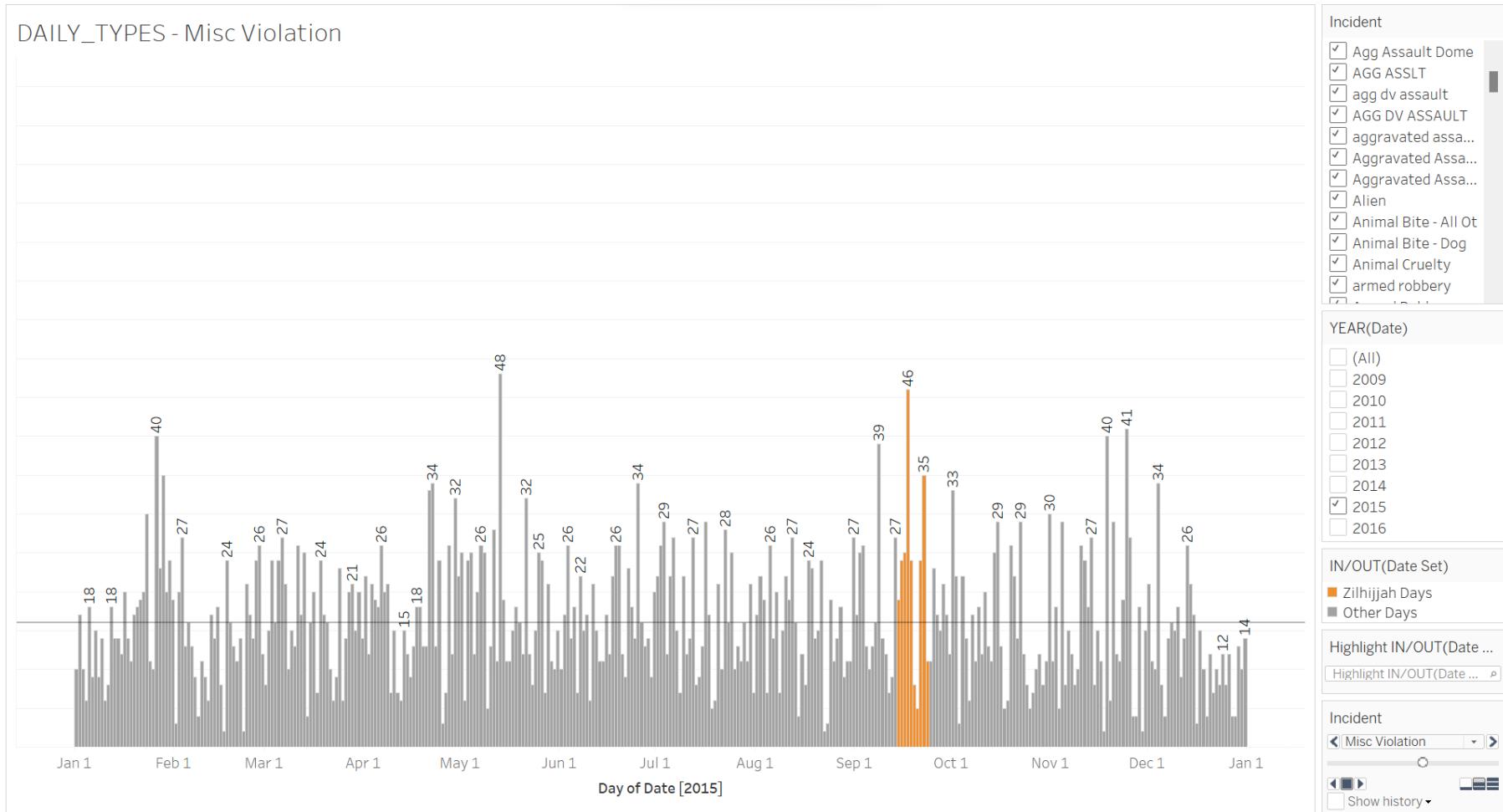
2010:
 $0.02838(2.76/97.24) > 0.02817$

DAYLY COUNTS



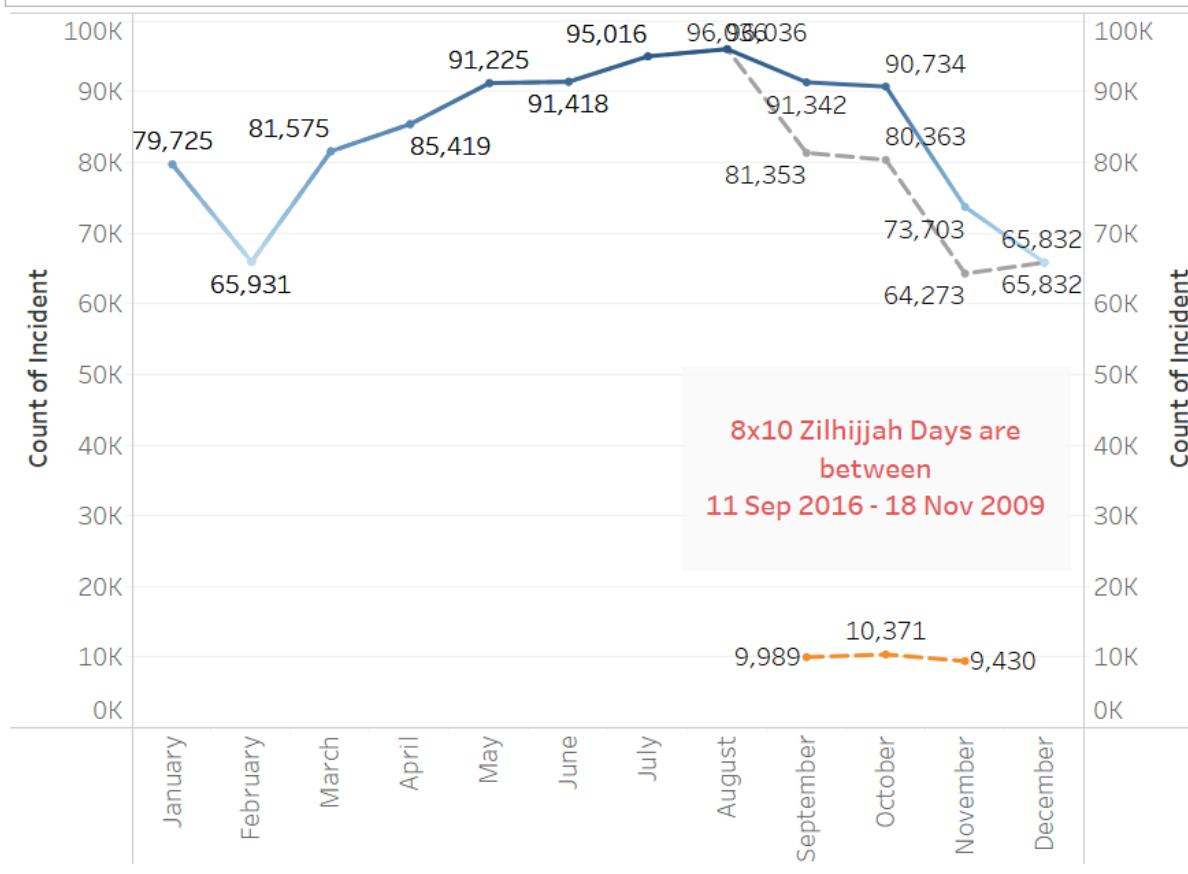


DAILY_TYPES - Misc Violation



Monthly Counts of Incidents:

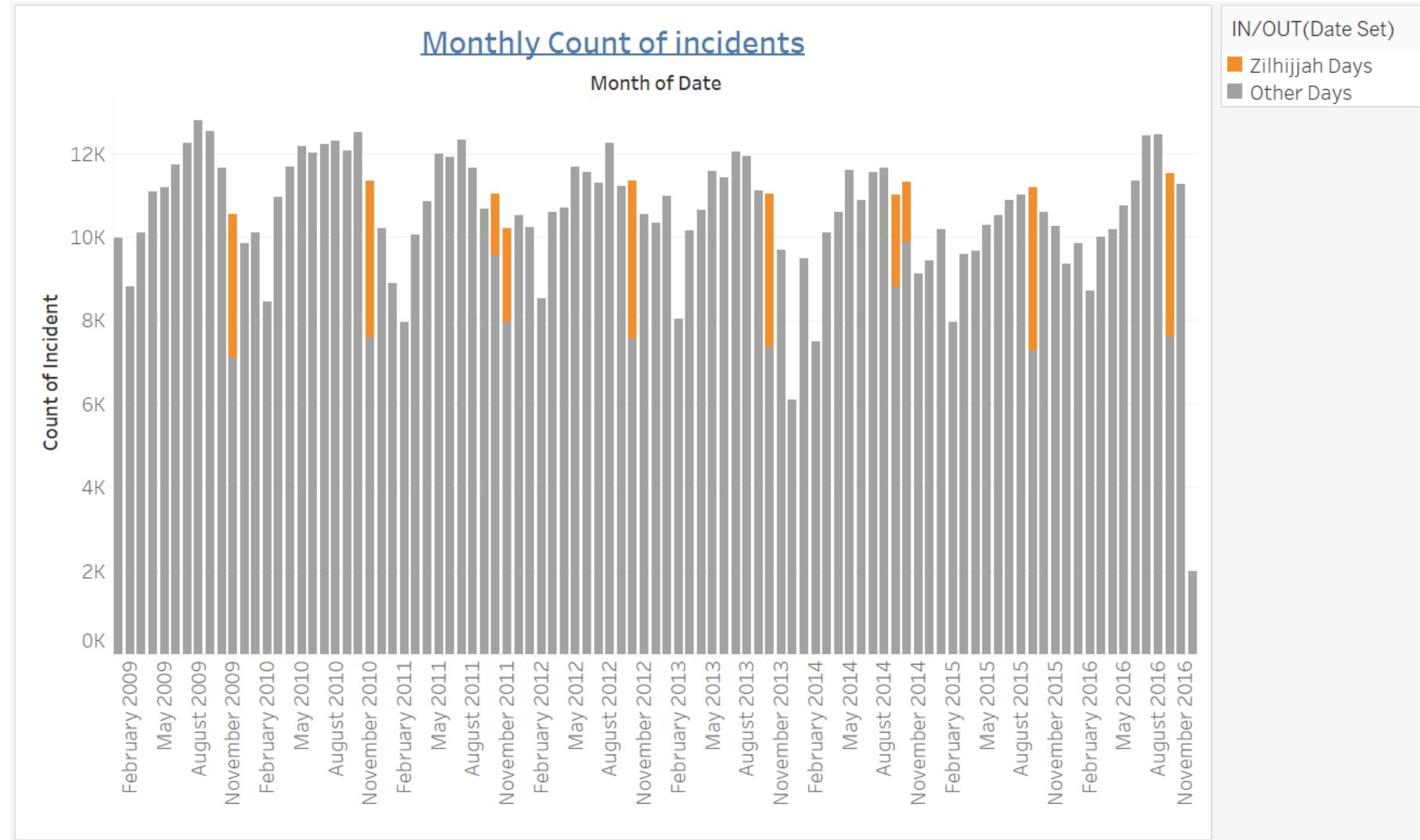
(Zilhijjah Days are also separately shown below.)



Caption

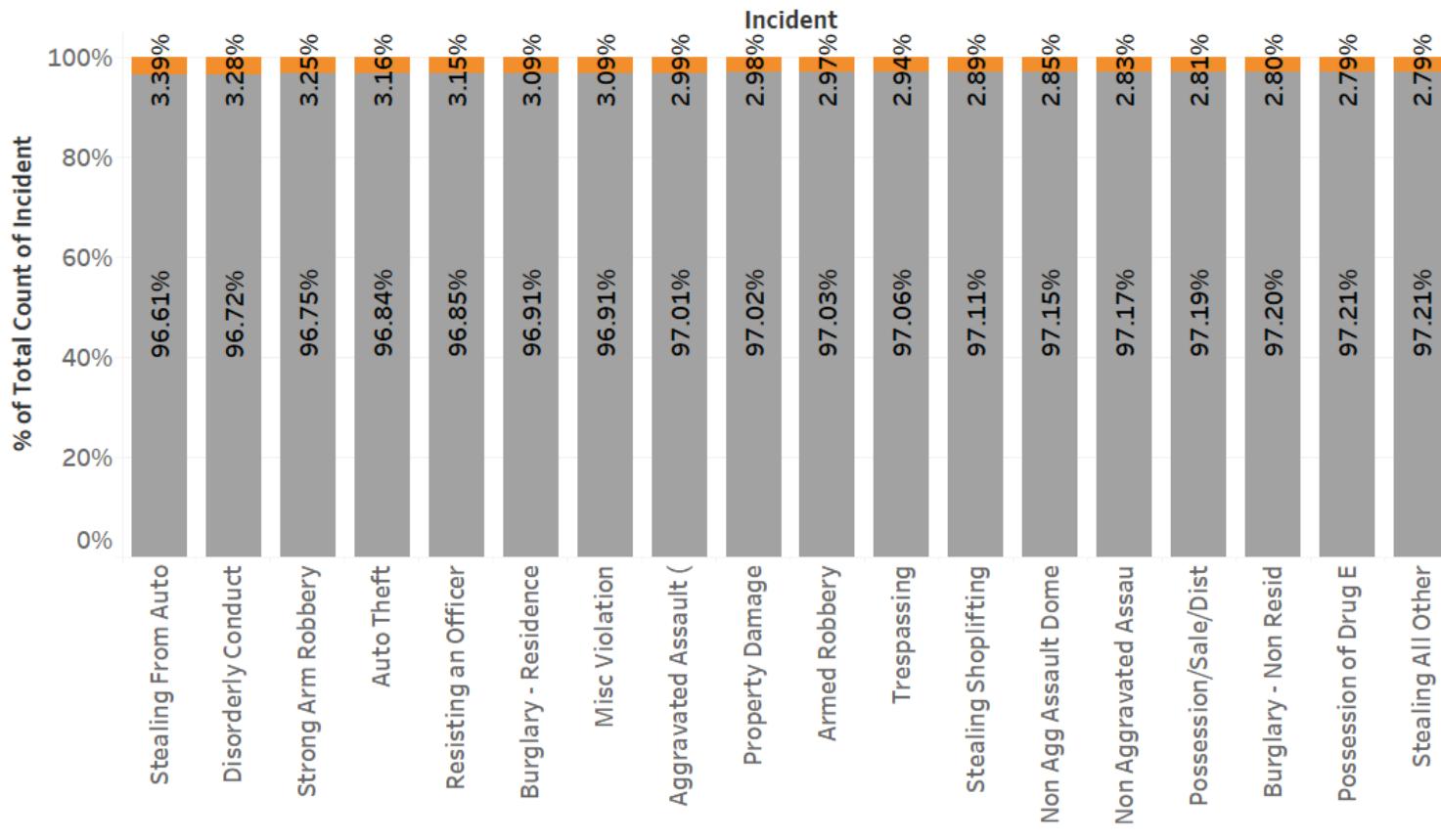
During the period of September to November, Zilhijjah days mark an average 371 incidents per day, higher than the overall annual average 351.

$$(9989+10371+9340)/80 \text{ days} = 371$$



The crime types which have a higher rate during the first ten days of Zilhijjah, compared to the rest of the year.

IN/OUT(Date Set)
 Zilhijjah Days
 Other Days



SAMPLE DATA-4: DETROIT CRIME INCIDENTS_2009-2016

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

This data reflects reported criminal offenses that have occurred in the City of Detroit since January 1, 2009. Offense data was extracted from the Detroit Police Department's CrisNet/NetRMS records management system. This data set contains the most recent data available and is updated hourly. It should be noted that some incidents involve the commission of multiple offenses, such as a domestic assault where property was also vandalized. Accordingly, the data describe all

offenses associated with all reported incidents, excluding sexual assaults. In order to protect the privacy of crime victims, addresses are shown at the block level only and the geographic coordinates (longitude and latitude) have been randomly adjusted to provide accuracy only to the street block and do not identify precise crime locations. Field descriptions: ROWNUM - Dataset Row number CASEID - Numeric Caseid Identifier CRIMEID - Numeric Offense Identifier CRNO - Detroit Police Incident Number (YYYYMMDD Report Number) ADDRESS - Location Where Incident Occurred (block or intersection) CATEGORY - Crime Category OFFENSEDESCRIPTION - Description of Offense STATEOFFENSECODE - Michigan State Police Offense Code INCIDENTDATE - Date of Crime Incident HOUR - Earliest Estimated Hour of Crime Occurrence SCA - Police Scout Car Area PRECINCT - Police Precinct COUNCIL - City Council District NEIGHBORHOOD - Neighborhood Name CENSUSTRACT - 2010 US Census Tract LON - longitude WGS 1984 LAT - latitude WGS 1984 LOCATION - Socrata's Hybrid Location Field for Mapping

```
In [150]: df = pd.read_csv("Detroit_crime-incidents-2009-2016.csv", index_col=0, low_memory=False)
df
```

Out[150]:

	CASEID	CRIMEID	CRNO	ADDRESS	CATEGORY	OFFENSEDESCRIPTION	STATEOFFENSEFILECLASS	INCIDENTDATE	HOUR	\$
ROWNUM										
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 [151...]: 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 [152...]: df.duplicated().value_counts()
```

```
Out[152]: False    1151019
dtype: int64
```

```
In [153...]: df["CATEGORY"].value_counts()
```

Out[153]:

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

```
Out[154]: 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 [155]: df = df.rename(columns = {'INCIDENTDATE':'date'})  
df = df.rename(columns = {'CATEGORY':'incident'})
```

```
In [156]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [157]: df.date.min(), df.date.max()
```

```
Out[157]: ('2009-01-01', '2016-06-28')
```

The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23'

```
In [158]: df = df.iloc[:, [4,7]]  
# df.to_csv("Detroit.csv", index=False)
```

```
In [159]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[159]: 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 [160... # 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 [161]: df.date.unique()
```

```
Out[161]: 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 [162]: zilhijjah_10_days(df)
```

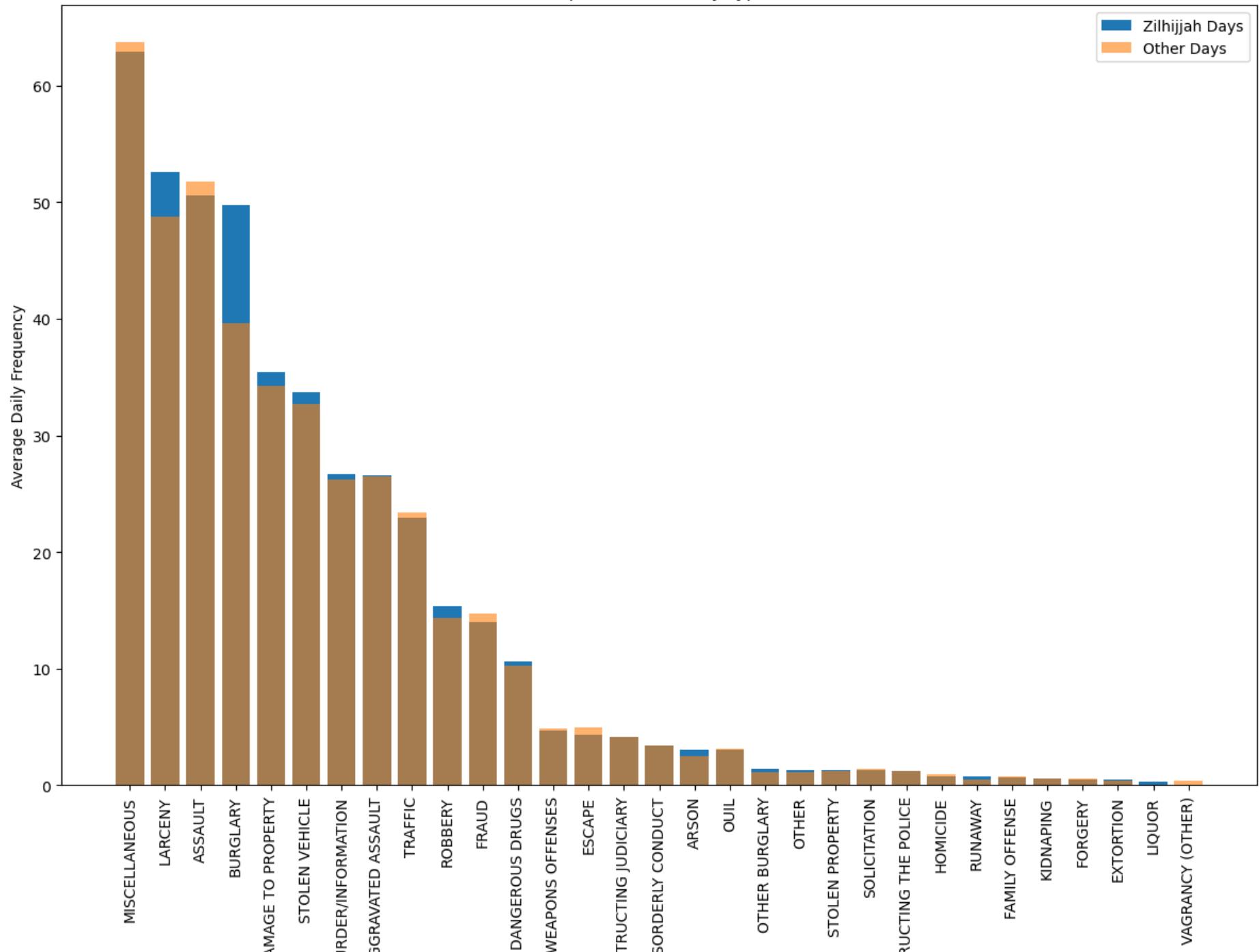
```
Total number of days: 2736
-----
Total number of cases: 1151019
-----
Average Daily Case Count: 420.69
-----
Yearly case counts according to the Gregorian calendar:
-----
2009    181681
2010    170599
2011    157248
2012    156133
2013    146908
2015    137044
2014    136628
2016    64778
Name: date, 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 Zilhijjah months: 435.4857
-----
Average case count in other days: 420.3057
-----
Ratio of Zilhijjah cases to other cases: 1.0361
```

We observe a 3.61% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

In [163..

```
sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type



D.
M
A

OBS
DI

OBST

Incident Types

```
In [164...]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio
sorted_ratios
```

Out[164]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
DRUNKENNESS	2	19	0.105300
BRIBERY	1	19	0.052600
CIVIL	7	176	0.039800
RUNAWAY	53	1457	0.036400
EMBEZZLEMENT	12	346	0.034700
ENVIRONMENT	19	563	0.033700
OTHER BURGLARY	99	3065	0.032300
BURGLARY	3482	109073	0.031900
ARSON	215	6891	0.031200
EXTORTION	34	1099	0.030900
OTHER	91	3135	0.029000
STOLEN PROPERTY	91	3249	0.028000
KIDNAPING	44	1588	0.027700
LARCENY	3682	133666	0.027500
ROBBERY	1076	39292	0.027400
LIQUOR	21	782	0.026900
DANGEROUS DRUGS	741	28081	0.026400
DAMAGE TO PROPERTY	2479	93781	0.026400
STOLEN VEHICLE	2359	89445	0.026400
MURDER/INFORMATION	1870	71694	0.026100
DISORDERLY CONDUCT	239	9252	0.025800
AGGRAVATED ASSAULT	1862	72577	0.025700
OBSTRUCTING JUDICIARY	290	11345	0.025600
OBSTRUCTING THE POLICE	83	3268	0.025400

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
MISCELLANEOUS	4403	174164	0.025300
TRAFFIC	1603	63890	0.025100
ASSAULT	3541	141608	0.025000
OUIL	213	8557	0.024900
WEAPONS OFFENSES	330	13268	0.024900
FRAUD	981	40357	0.024300

In [165]:

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

Out[165]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
LARCENY	3682	133666	0.0275
BURGLARY	3482	109073	0.0319
DAMAGE TO PROPERTY	2479	93781	0.0264
STOLEN VEHICLE	2359	89445	0.0264
MURDER/INFORMATION	1870	71694	0.0261
AGGRAVATED ASSAULT	1862	72577	0.0257
ROBBERY	1076	39292	0.0274
DANGEROUS DRUGS	741	28081	0.0264
OBSTRUCTING JUDICIARY	290	11345	0.0256
DISORDERLY CONDUCT	239	9252	0.0258
ARSON	215	6891	0.0312
OTHER BURGLARY	99	3065	0.0323
STOLEN PROPERTY	91	3249	0.0280
OTHER	91	3135	0.0290
RUNAWAY	53	1457	0.0364
KIDNAPING	44	1588	0.0277
EXTORTION	34	1099	0.0309
LIQUOR	21	782	0.0269
ENVIRONMENT	19	563	0.0337
EMBEZZLEMENT	12	346	0.0347
CIVIL	7	176	0.0398
DRUNKENNESS	2	19	0.1053
BRIBERY	1	19	0.0526

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

```
In [166]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]
```

```
Out[166]: (46, 23)
```

```
In [167]: zilhijjah_incidents_desc
```

```
Out[167]: count      30484  
unique       39  
top    MISCELLANEOUS  
freq      4403  
Name: incident, dtype: object
```

```
In [168]: other_days_incidents_desc
```

```
Out[168]: count      1120535  
unique       46  
top    MISCELLANEOUS  
freq      169761  
Name: incident, dtype: object
```

Detroit dataset encompasses 46 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 39 incident categories, with 23 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2009	2010	2011	2012	2013	2014	2015
Zilhijjah Days	30,484	4,974	4,586	4,458	4,523	4,011	4,106	3,826
Other Days	1,055,757	176,707	166,013	152,790	151,610	142,897	132,522	133,218

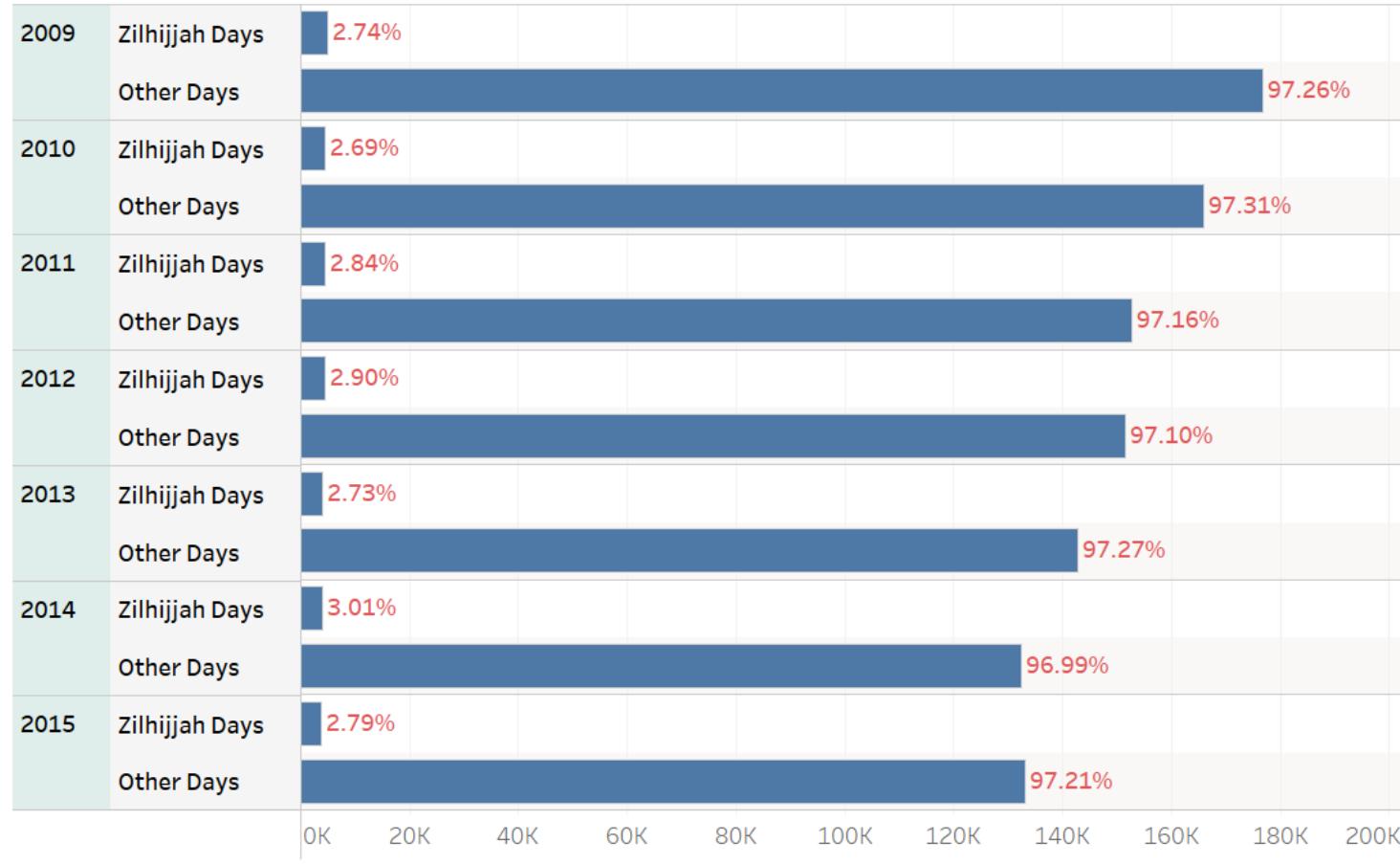
Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incidents	2009	2010	2011	2012	2013	2014	2015
Zilhijjah Days	2.81%	2.74%	2.69%	2.84%	2.90%	2.73%	3.01%	2.79%
Other Days	97.19%	97.26%	97.31%	97.16%	97.10%	97.27%	96.99%	97.21%

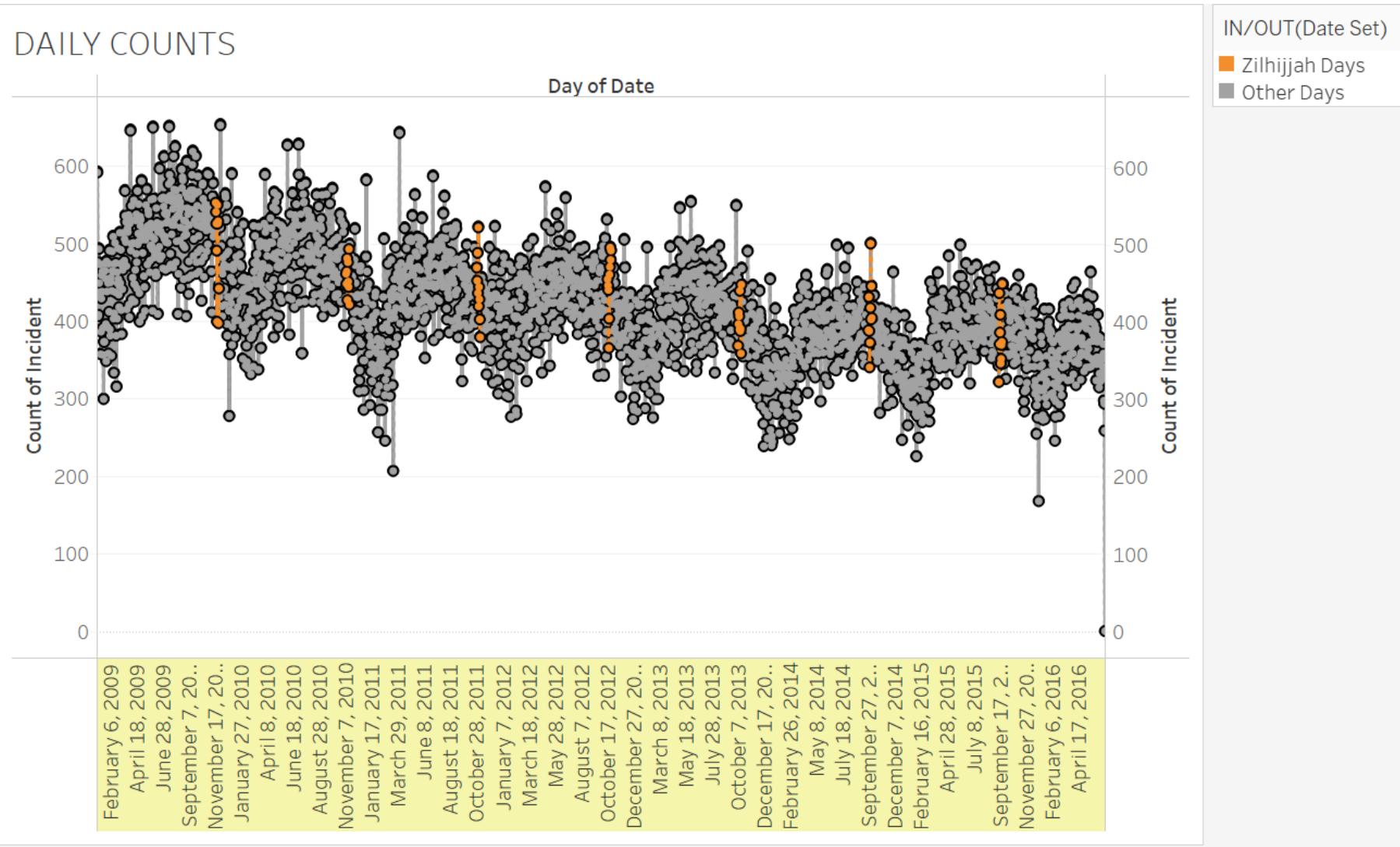
Crime Rates: Zilhijjah Days vs Other Days

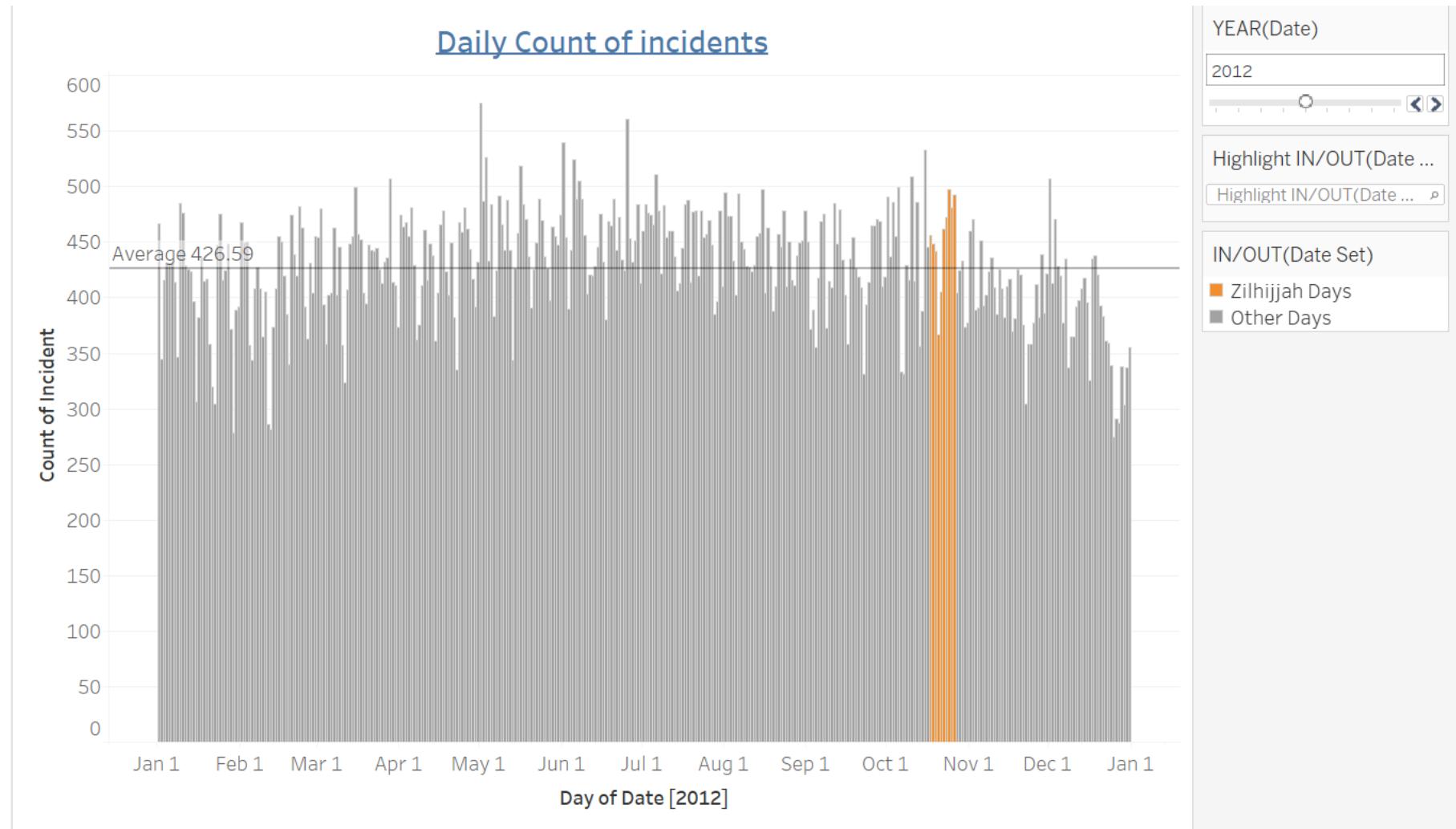
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]



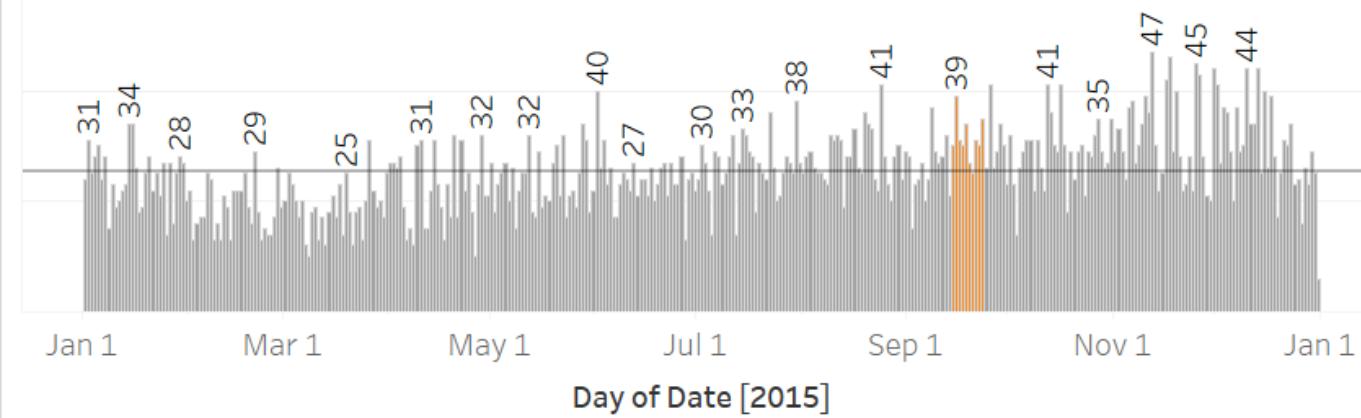
Caption

Compared to the rest of the year, crime rates during the initial ten days of Zilhijjah were higher in all years except 2010, and 2013.





DAILY_TYPES - BURGLARY



Incident

- ABORTION
- AGGRAVATED AS...
- ARSON
- ASSAULT
- BURGLARY

YEAR(Date)

- 2012
- 2013
- 2014
- 2015
- 2016

IN/OUT(Date Set)

- Zilhijjah Days
- Other Days

Highlight IN/OUT(Date ...)

Highlight IN/OUT(Date ...)

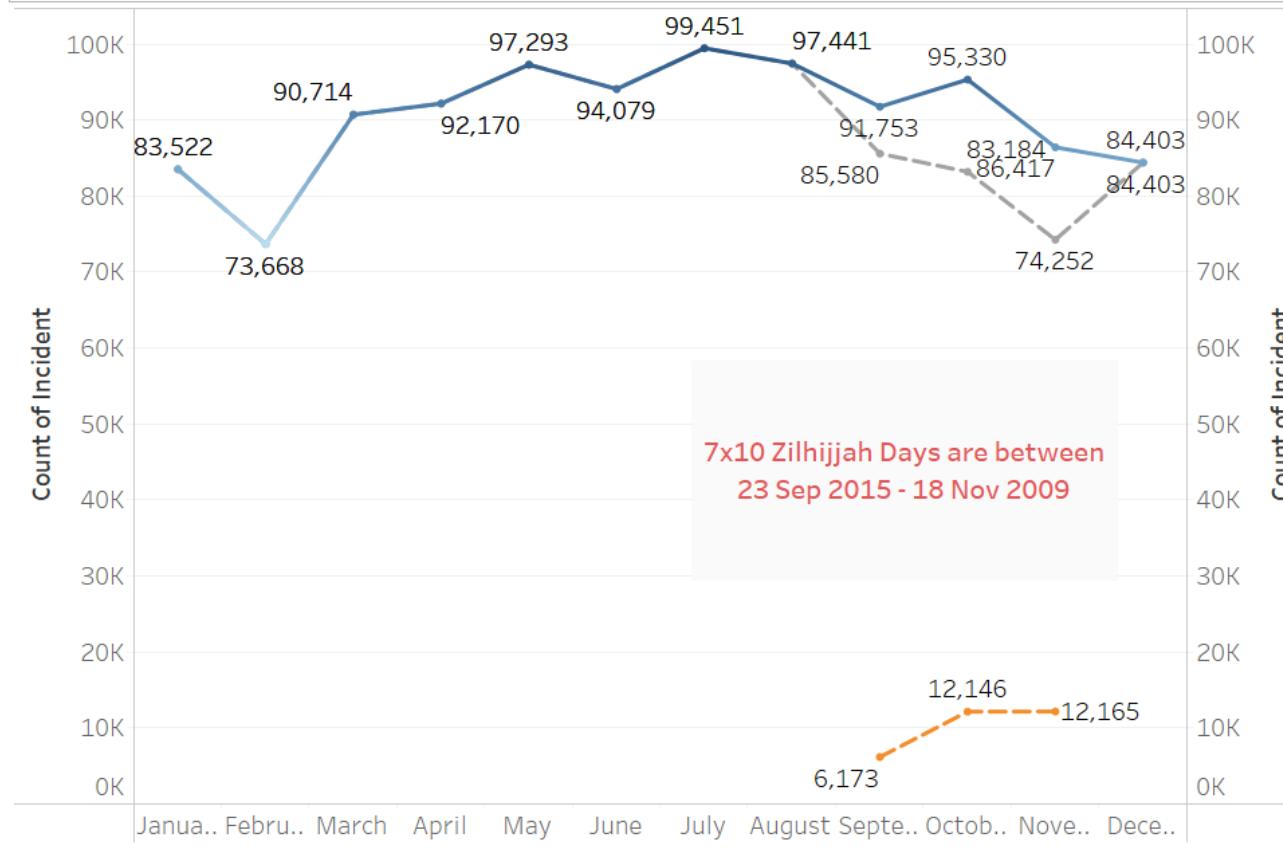
Incident

BURGLARY



Show history

Monthly Counts of Incidents: (Zilhijjah Days are also separately shown below.)

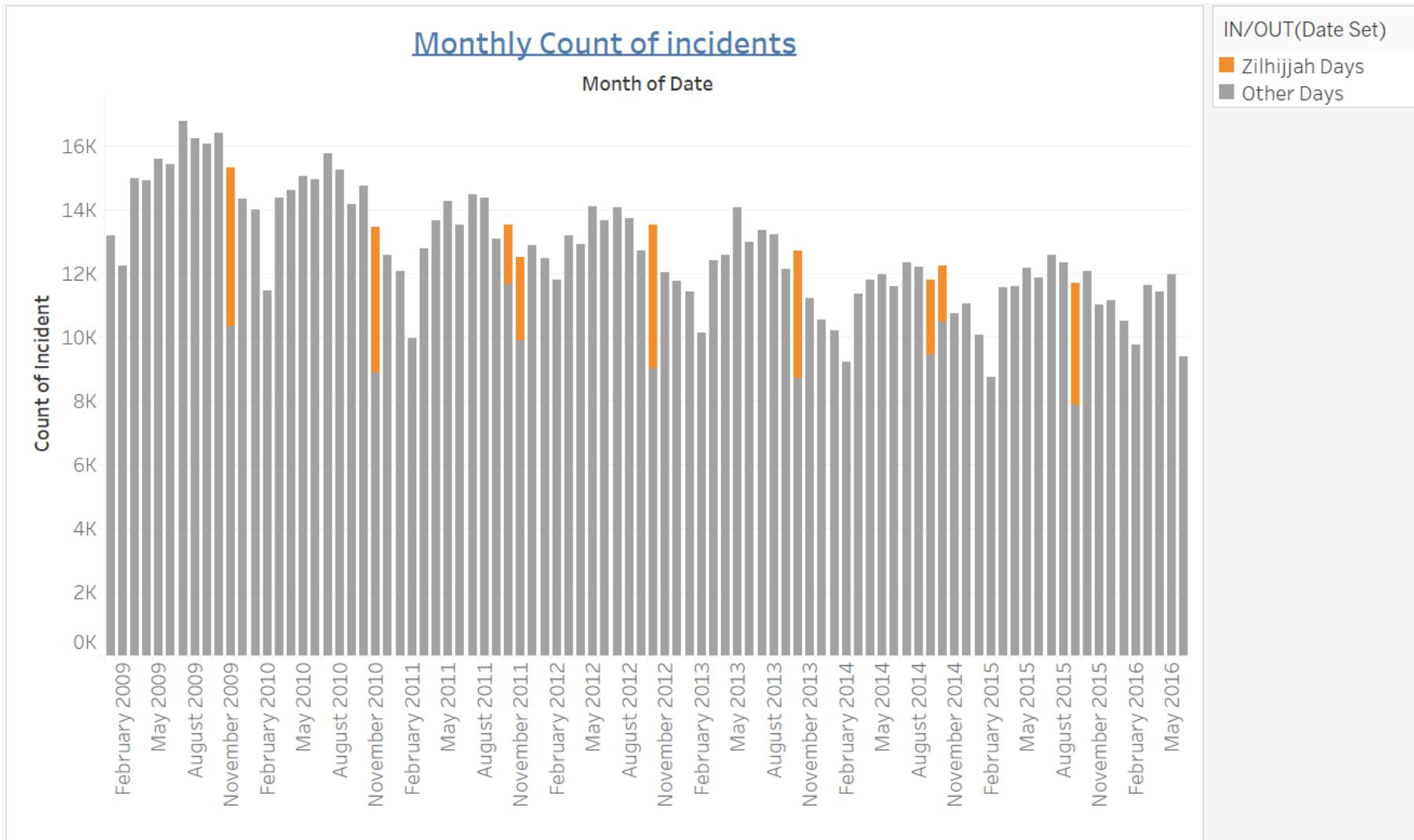


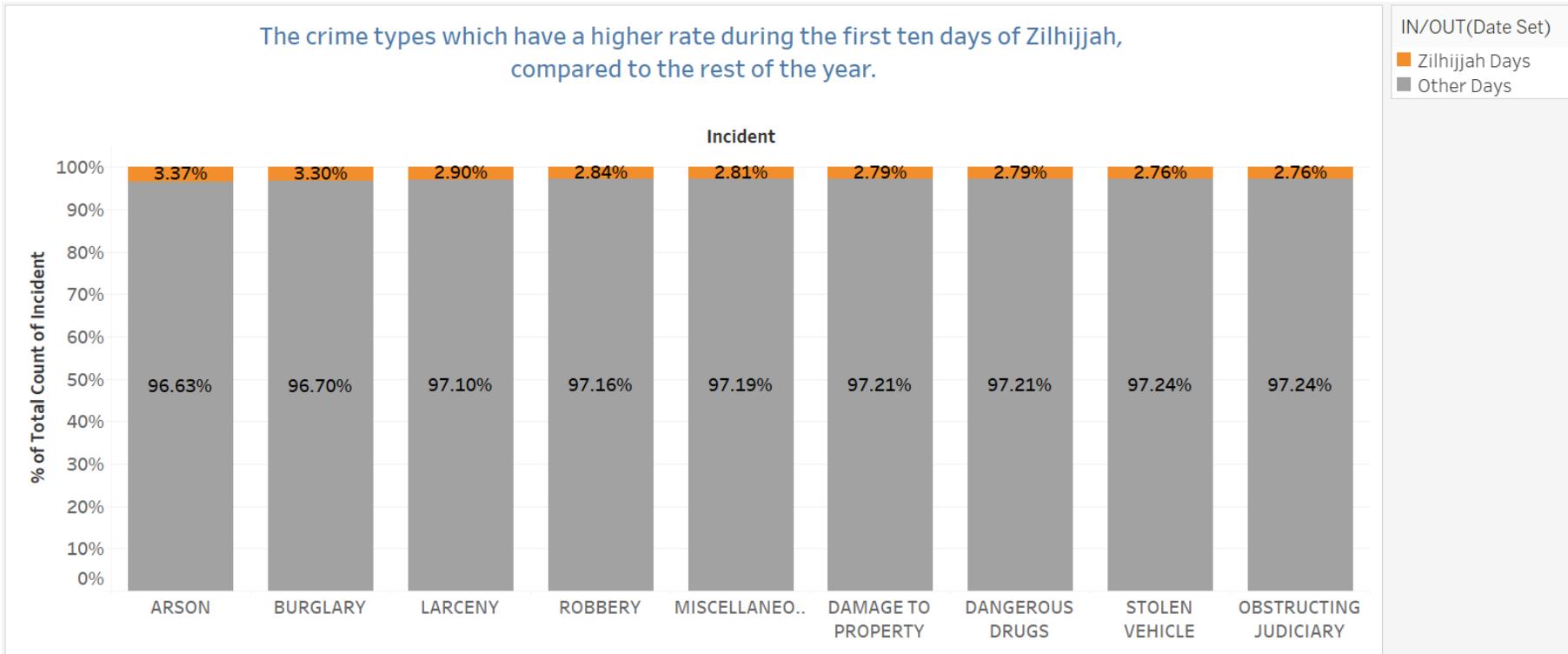
Caption

During the period of September to November, while the average of other days shows a decreasing trend, the first ten days of Zilhijjah have remained above the annual average.

Zilhijjah days mark an average 435 incidents per day while other days mark 420.

$$(6173+12146+12165 \text{ incidents})/70 \text{ days} = 435.48$$





SAMPLE DATA-5: DENVER CRIME DATASET_2019-2023

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

```
In [87]: df = pd.read_csv("denver_crime.csv", index_col=0, encoding="Latin-1", low_memory=False)
df
```

Out[87]:

	offense_id	offense_code	offense_code_extension	offense_type_id	offense_category_id	first_occurrence_date	last_occurrence_date	re
incident_id								
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 [88]: 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 [89]: df.duplicated().value_counts()
```

```
Out[89]: False    398091
dtype: int64
```

```
In [90]: df["offense_type_id"].value_counts()
```

```
Out[90]:
```

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-methamphetamine-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-methamphetamine-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 [91]: df["offense_category_id"].value_counts()
```

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

```
In [93]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [94]: df.date.min(), df.date.max()
```

```
Out[94]: ('2018-01-02', '2023-12-07')
```

```
# The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2018-08-12' - '2018-08-21' '2019-08-02' - '2019-08-11' '2020-07-22' - '2020-07-31' '2021-07-11' - '2021-07-20' '2022-06-30' - '2022-07-09' '2023-06-19' - '2023-06-28'
```

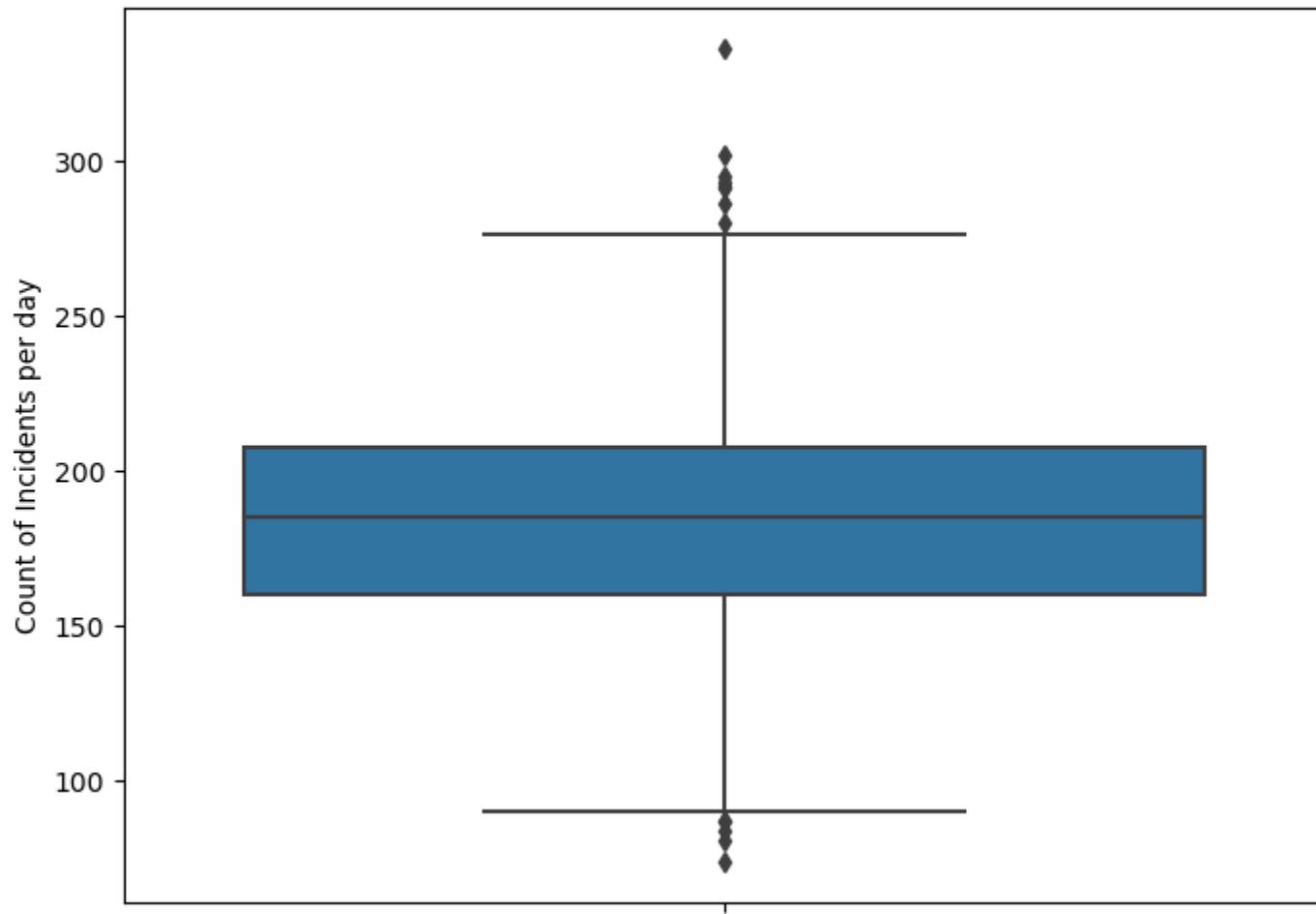
```
In [95]: df = df.iloc[:, [4,5]]
# df.to_csv("Denver.csv", index=False)
```

```
In [96]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
daily_incident_counts_stats
```

```
Out[96]: 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 [97]: # 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



```
In [98]: df.date.nunique()
```

```
Out[98]: 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.

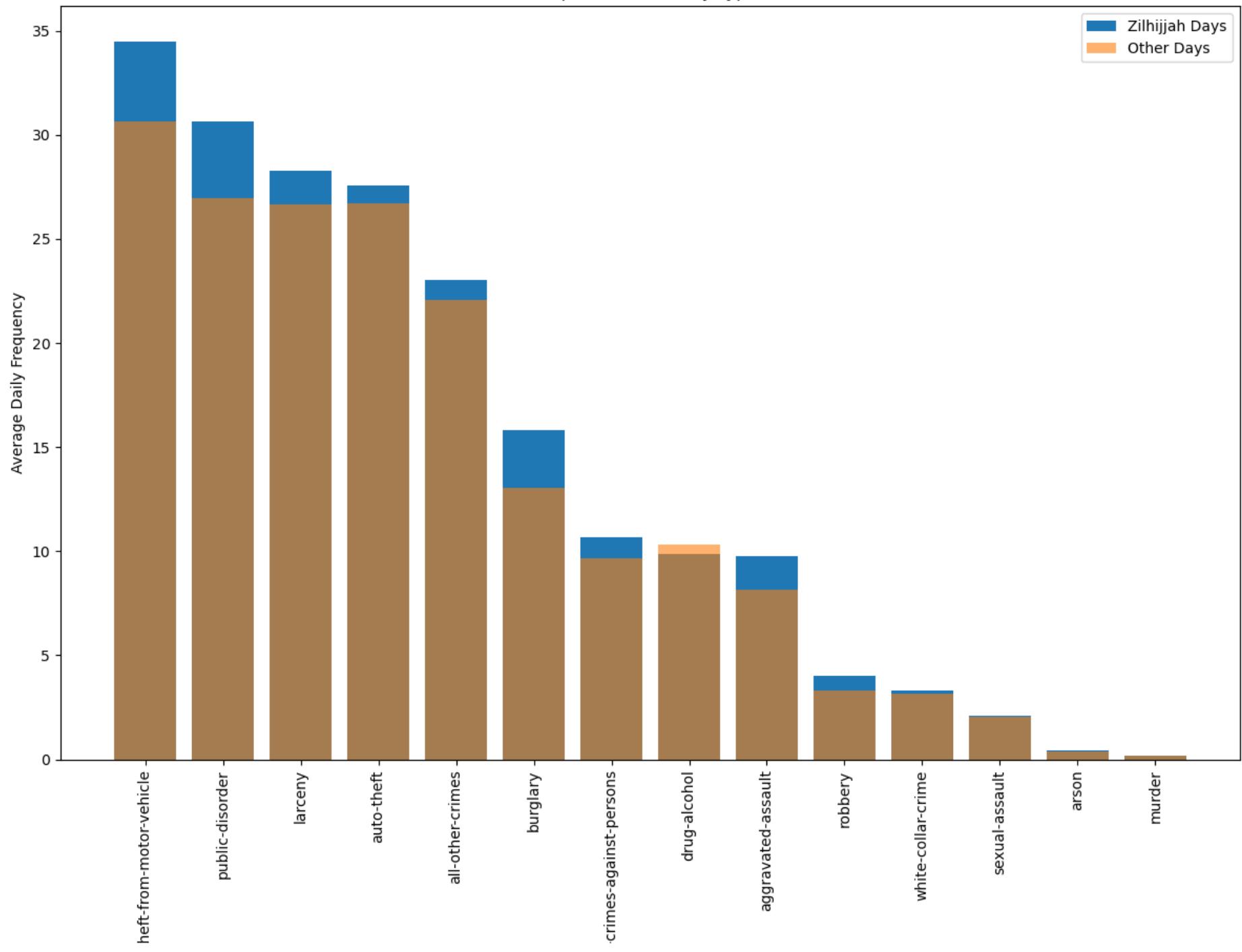
```
In [99]: zilhijjah_10_days(df)
```

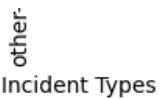
```
Total number of days: 2166
-----
Total number of cases: 398091
-----
Average Daily Case Count: 183.79
-----
Yearly case counts according to the Gregorian calendar:
-----
2022    76854
2021    72561
2023    71028
2020    63754
2019    57625
2018    56269
Name: date, dtype: int64
-----
Case counts 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 Zilhijjah months: 200.1333
-----
Average case count in other days: 183.3253
-----
Ratio of Zilhijjah cases to other cases: 1.0917
```

We observe a 9.17% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [100]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type





```
In [101]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio  
sorted_ratios
```

Out[101]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
robbery	242	7178	0.033700
burglary	949	28432	0.033400
aggravated-assault	587	17703	0.033200
arson	27	852	0.031700
public-disorder	1837	58617	0.031300
theft-from-motor-vehicle	2067	66632	0.031000
other-crimes-against-persons	639	21000	0.030400
larceny	1696	57788	0.029300
all-other-crimes	1383	47898	0.028900
white-collar-crime	198	6898	0.028700
auto-theft	1653	57905	0.028500
murder	12	423	0.028400
sexual-assault	125	4411	0.028300
drug-alcohol	593	22354	0.026500

```
In [102]: # In which categories were more crimes committed during the first ten days of Zilhijjah?  
zilhijjah_dominant_incidents
```

Out[102]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
theft-from-motor-vehicle	2067	66632	0.0310
public-disorder	1837	58617	0.0313
larceny	1696	57788	0.0293
auto-theft	1653	57905	0.0285
all-other-crimes	1383	47898	0.0289
burglary	949	28432	0.0334
other-crimes-against-persons	639	21000	0.0304
aggravated-assault	587	17703	0.0332
robbery	242	7178	0.0337
white-collar-crime	198	6898	0.0287
sexual-assault	125	4411	0.0283
arson	27	852	0.0317
murder	12	423	0.0284

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

In [103...]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]

Out[103]: (14, 13)

In [104...]: zilhijjah_incidents_desc

Out[104]:

count	12008
unique	14
top	theft-from-motor-vehicle
freq	2067
Name: incident, dtype: object	

In [105...]: other_days_incidents_desc

```

Out[105]: count          386083
unique           14
top    theft-from-motor-vehicle
freq            64565
Name: incident, dtype: object

```

Denver dataset encompasses 14 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 14 incident categories, with 13 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2018	2019	2020	2021	2022	2023
Zilhijjah Days	12,008	1,802	1,728	1,861	2,133	2,134	2,350
Other Days	386,083	54,467	55,897	61,893	70,428	74,720	68,678

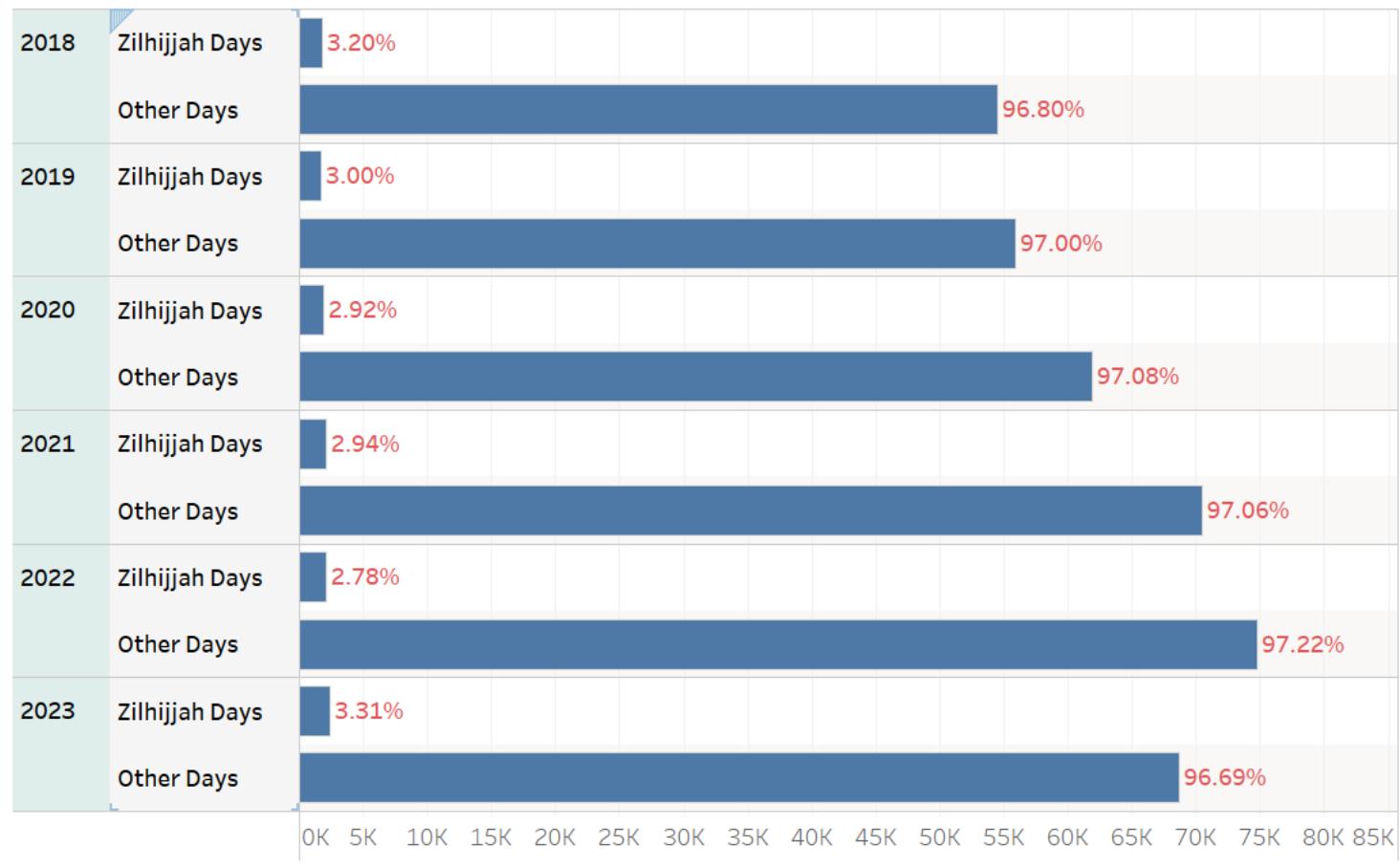
Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incidents	2018	2019	2020	2021	2022	2023
Zilhijjah Days	3.02%	3.20%	3.00%	2.92%	2.94%	2.78%	3.31%
Other Days	96.98%	96.80%	97.00%	97.08%	97.06%	97.22%	96.69%

Crime Rates: Zilhijjah Days vs Other Days

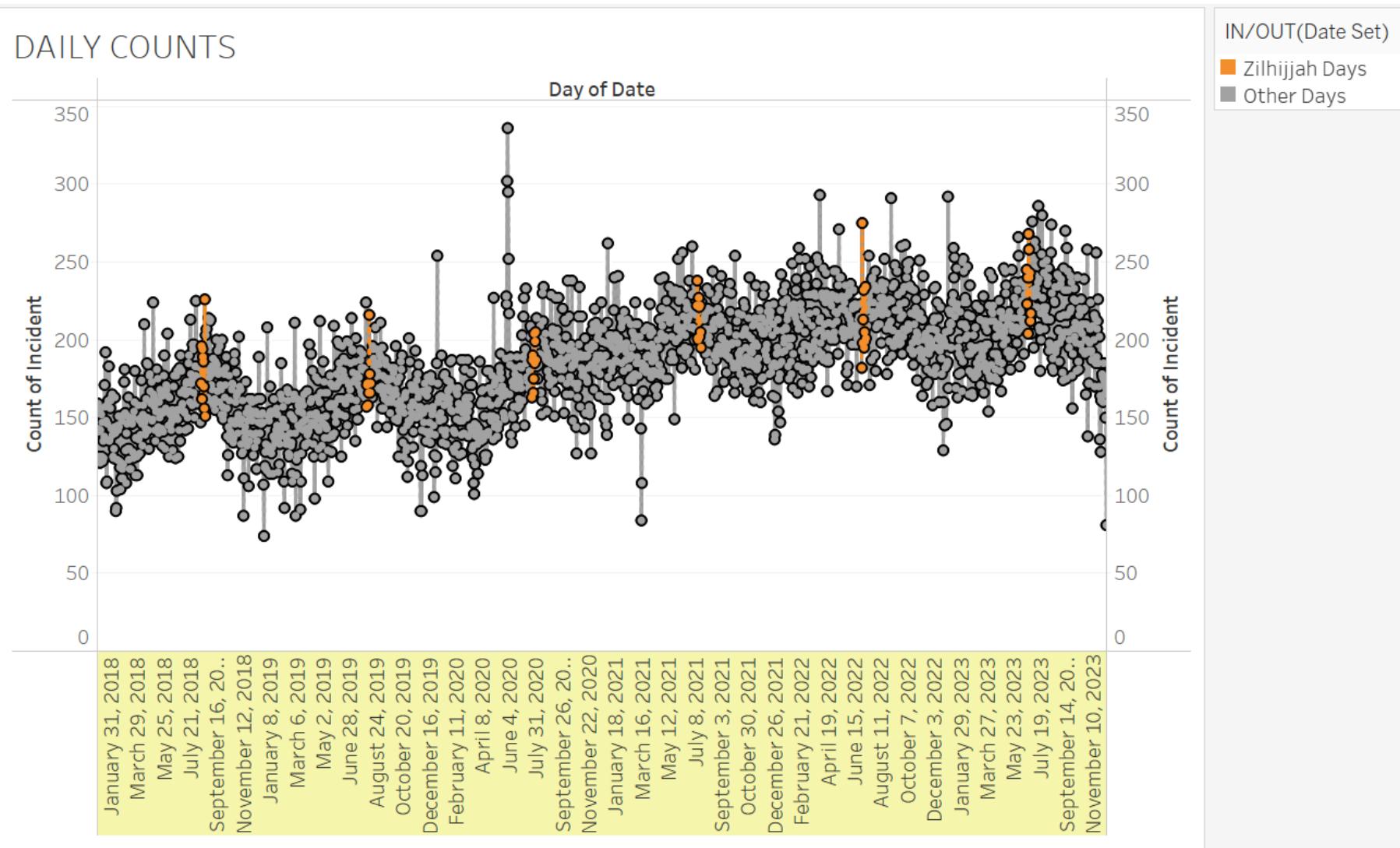
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]



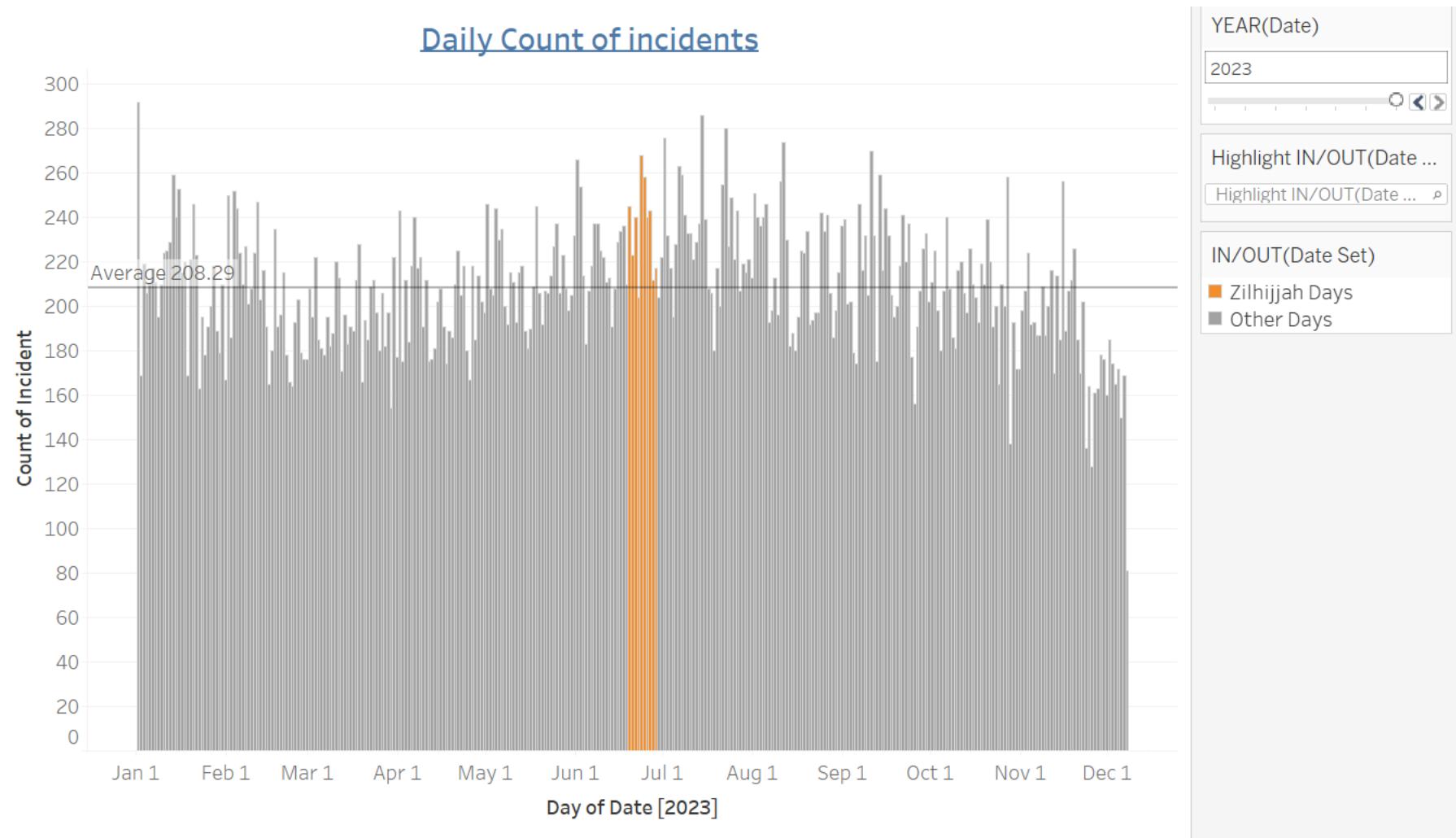
Caption

Compared to the rest of the year, crime rates during the initial ten days of Zilhijjah were higher in all years.

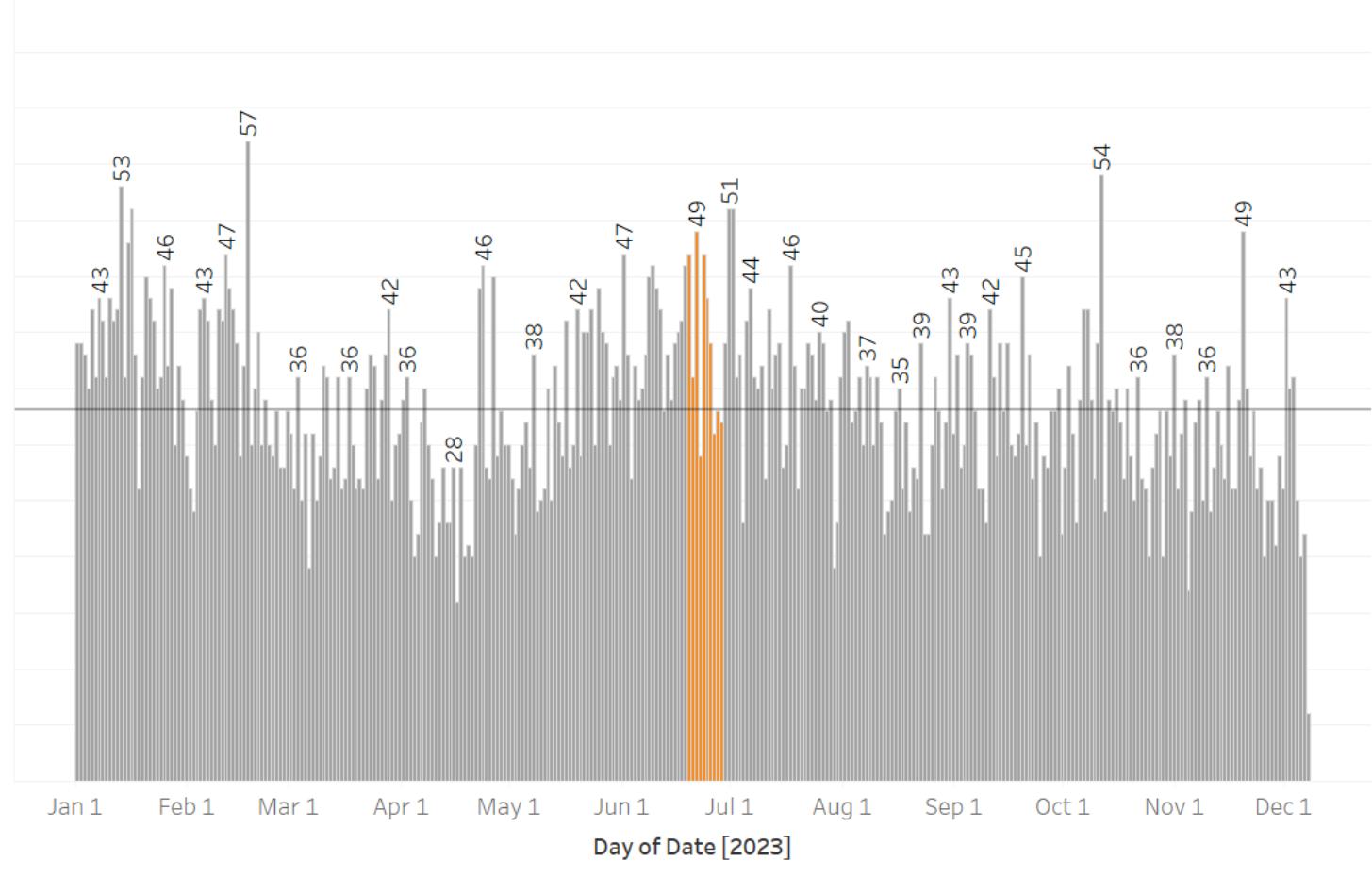
DAILY COUNTS



Daily Count of incidents



DAILY_TYPES - auto-theft



Incident

- (All)
- aggravated-assa...
- all-other-crimes
- arson
- auto-theft
- burglary

YEAR(Date)

- 2018
- 2019
- 2020
- 2021
- 2022
- 2023

IN/OUT(Date Set)

- Zilhijjah Days
- Other Days

Highlight IN/OUT(Date ...)

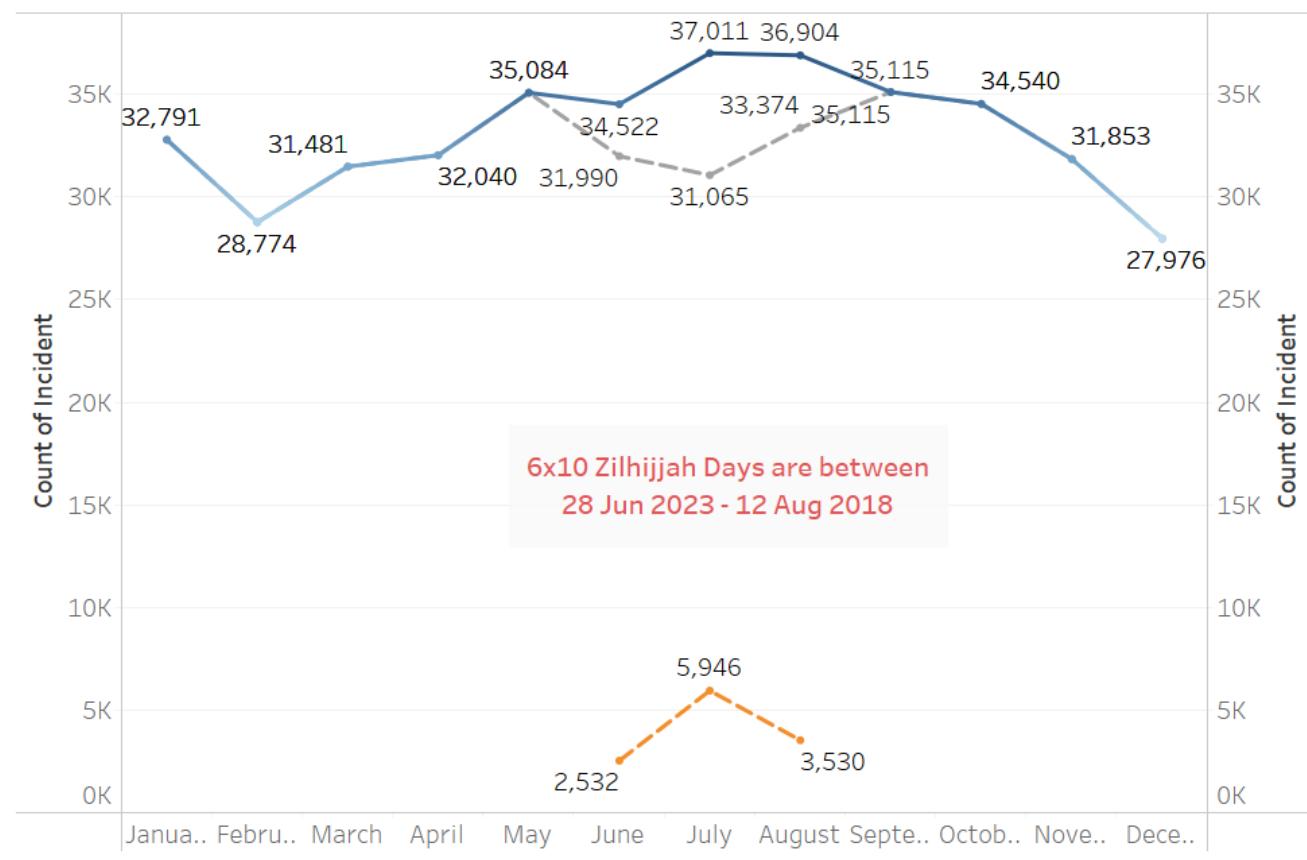
Incident

auto-theft

Show history ▾

Monthly Counts of Incidents:

(Zilhijjah Days are also separately shown below.)

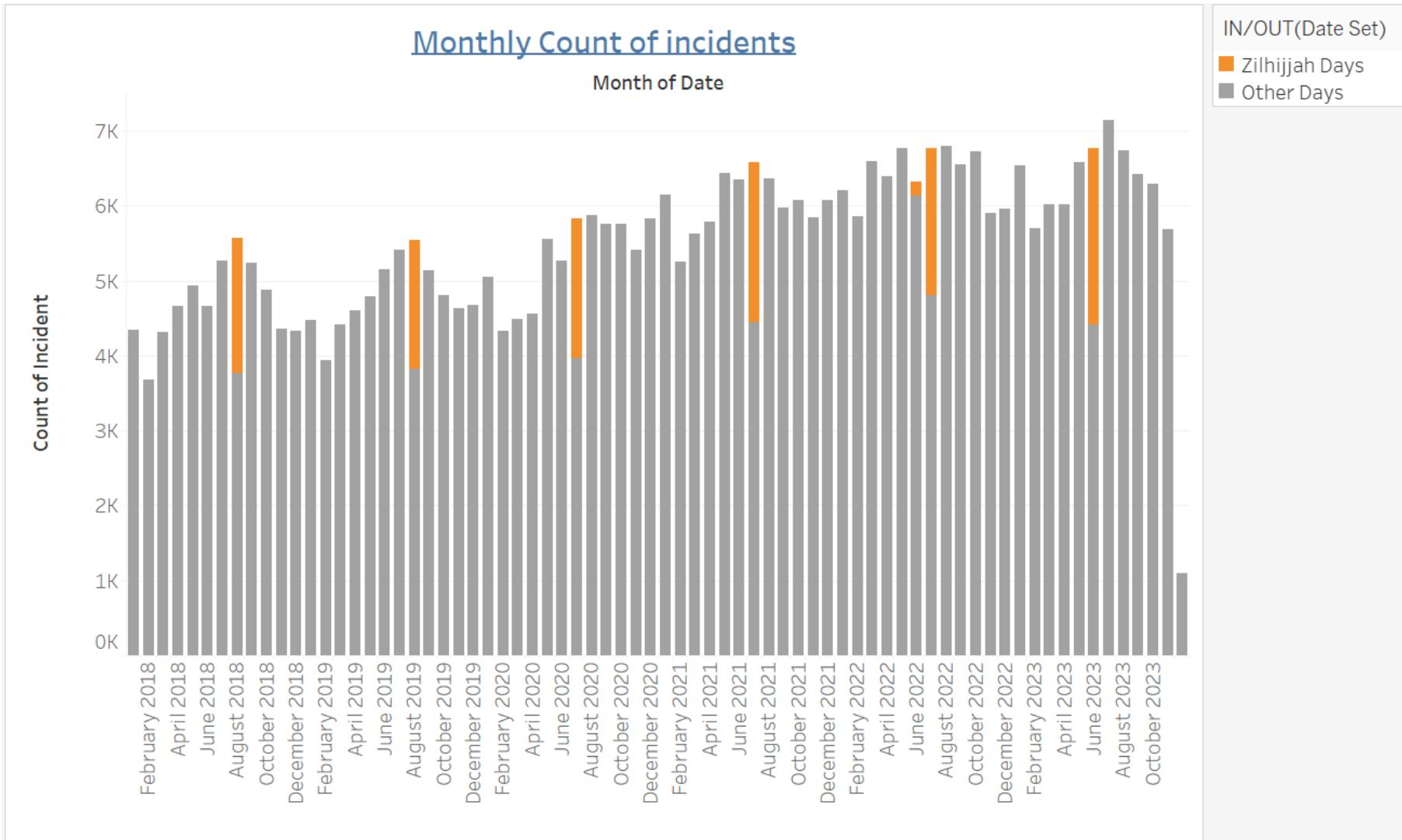


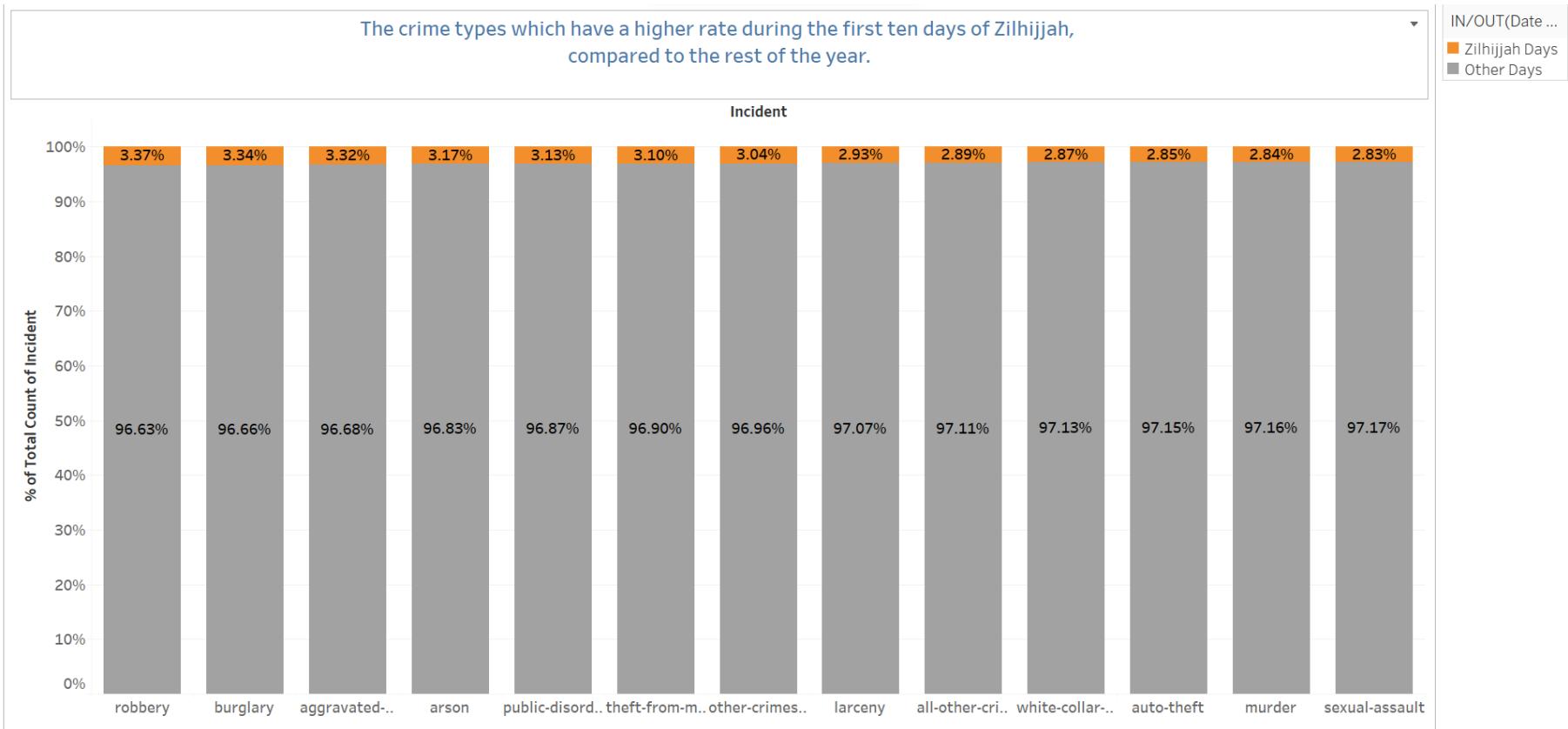
Caption

Between June and August, there is a peak in crime rates, coinciding with the first ten days of Zilhijjah that surpass the annual average.

Zilhijjah days mark an average 200 incidents per day while other days mark 183.

$$(2532+5946+3530 \text{ incidents}) / 60 \text{ days} = 200.13$$





SAMPLE DATA-6: VANCOUVER CRIME DATASET_2003-2017

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

```
In [106]: df = pd.read_csv("crime_vancouver.csv", low_memory=False)
df
```

Out[106]:

	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 [107... 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 [108]: df.duplicated().value_counts()
```

```
Out[108]: False    481814
           True     48838
           dtype: int64
```

```
In [109]: df=df.drop_duplicates()
```

```
In [110]: df["TYPE"].value_counts()
```

```
Out[110]: 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 [111]: df['date'] = pd.to_datetime(df[['YEAR', 'MONTH', 'DAY']])
```

```
C:\Users\yucea\AppData\Local\Temp\ipykernel_21140\1722800228.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

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

In [112...]

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

```
C:\Users\yucea\AppData\Local\Temp\ipykernel_21140\4031785424.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

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

In [113...]

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

In [114...]

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

Out[114]:

```
('2003-01-01', '2017-07-13')
```

```
# The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2003-02-02' - '2003-02-11' '2004-01-23' - '2004-02-01' '2005-01-12' - '2005-01-21' '2006-01-01' - '2006-01-10' '2006-12-22' - '2006-12-31' '2007-12-11' - '2007-12-20' '2008-11-29' - '2008-12-08' '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11'
```

In [115...]

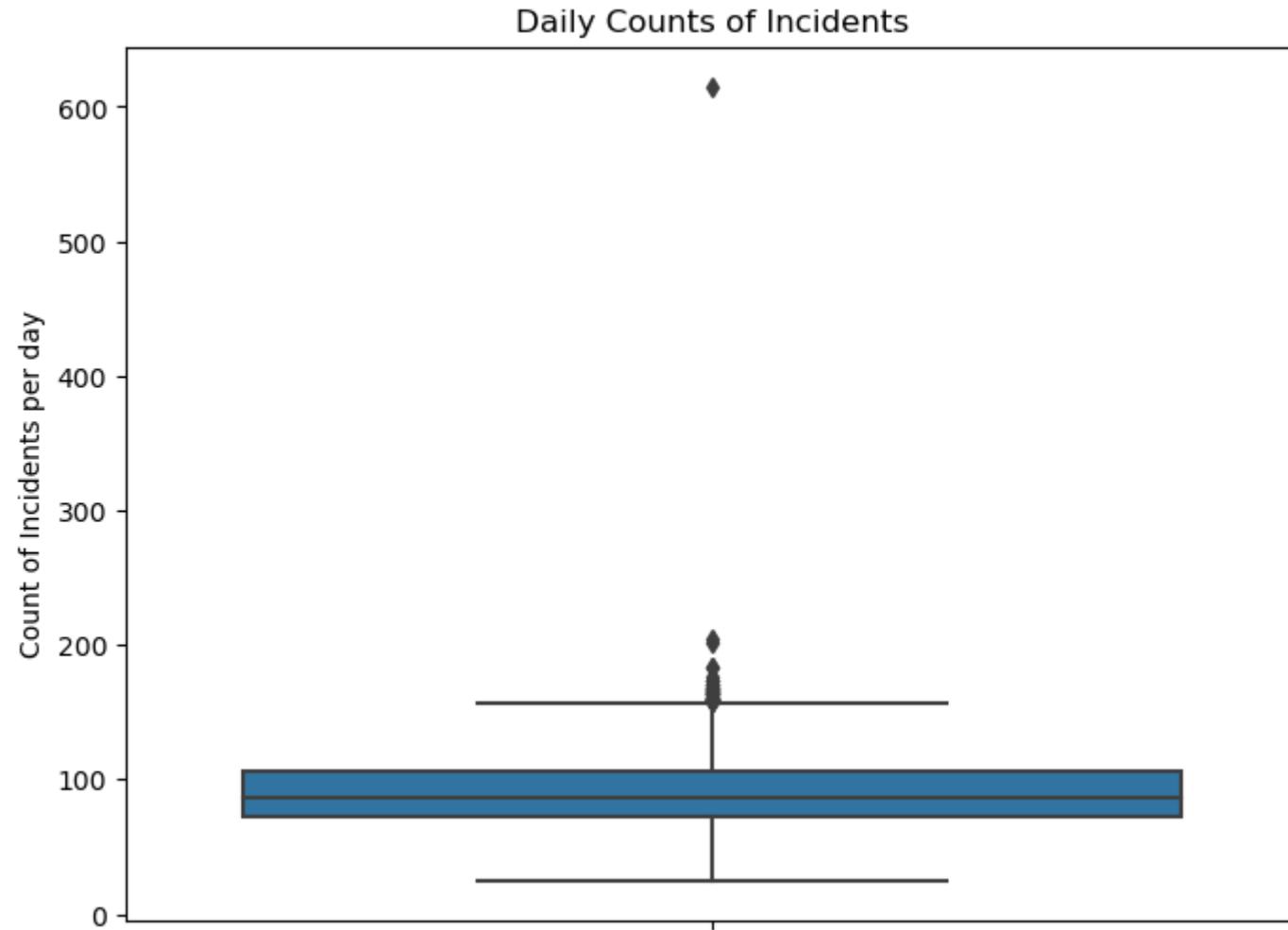
```
df = df.iloc[:, [0,12]]  
# df.to_csv("Vancouver.csv", index=False)
```

In [116...]

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

```
Out[116]: 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 [117...]: # 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 [118]: df.date.nunique()
```

```
Out[118]: 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 [119]: zilhijjah_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 Zilhijjah months: 91.34

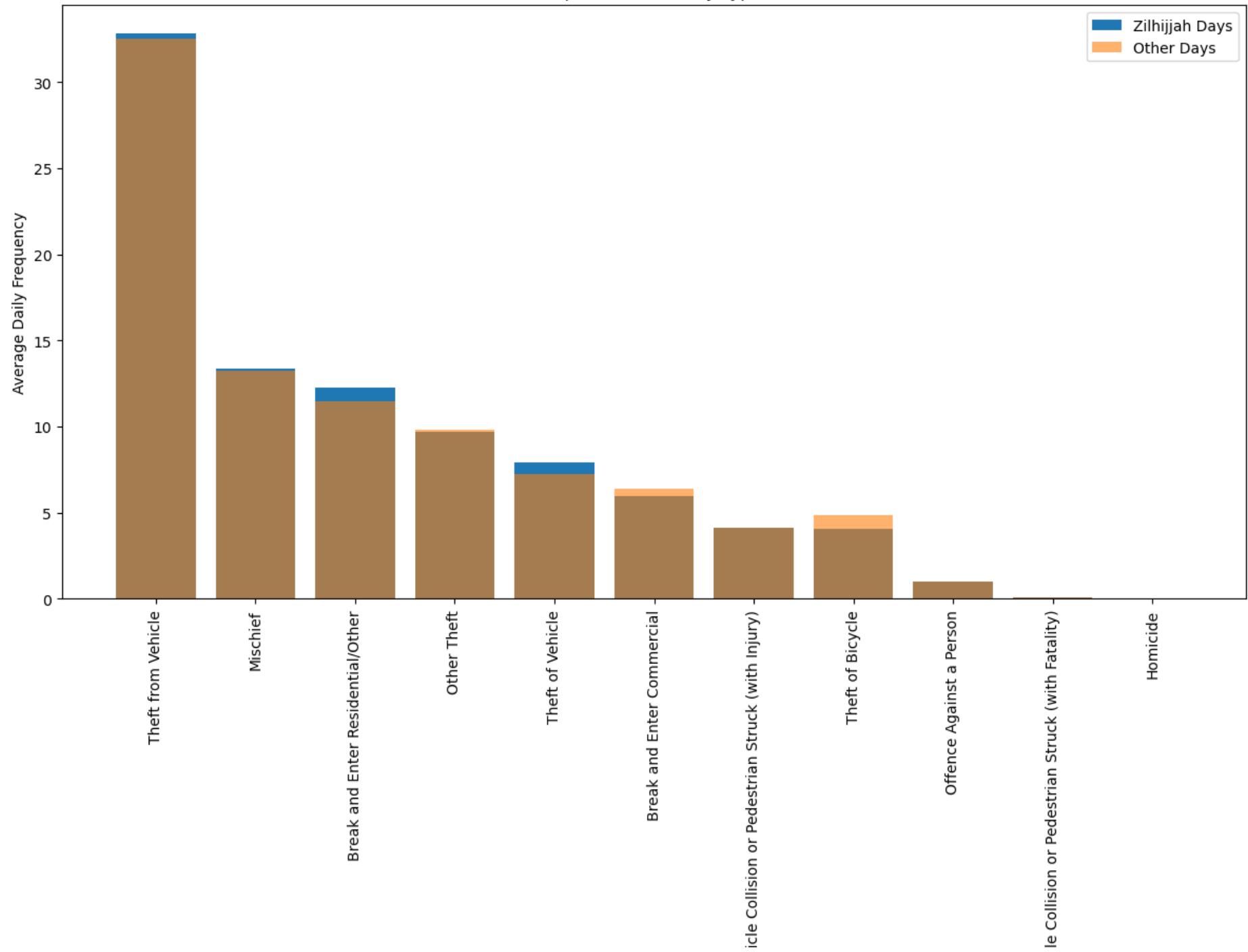
Average case count in other days: 90.7547

Ratio of Zilhijjah cases to other cases: 1.0064

We observe a 0.64% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [120...]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type



In [121...]

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

Out[121]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
Vehicle Collision or Pedestrian Struck (with Fatality)	8	254	0.031500
Theft of Vehicle	1190	38418	0.031000
Break and Enter Residential/Other	1838	60862	0.030200
Mischief	2004	70413	0.028500
Theft from Vehicle	4923	172699	0.028500
Vehicle Collision or Pedestrian Struck (with Injury)	621	21901	0.028400
Offence Against a Person	150	5307	0.028300
Other Theft	1453	52167	0.027900
Break and Enter Commercial	897	33845	0.026500
Theft of Bicycle	612	25730	0.023800
Homicide	5	218	0.022900

In [122...]

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

```
Out[122]:
```

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
Theft from Vehicle	4923	172699	0.0285
Mischief	2004	70413	0.0285
Break and Enter Residential/Other	1838	60862	0.0302
Theft of Vehicle	1190	38418	0.0310
Vehicle Collision or Pedestrian Struck (with Injury)	621	21901	0.0284
Offence Against a Person	150	5307	0.0283
Vehicle Collision or Pedestrian Struck (with Fatality)	8	254	0.0315

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

```
In [123...]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]
```

```
Out[123]: (11, 7)
```

```
In [124...]: zilhijjah_incidents_desc
```

```
Out[124]: count          13701
unique         11
top    Theft from Vehicle
freq           4923
Name: incident, dtype: object
```

```
In [125...]: other_days_incidents_desc
```

```
Out[125]: count          468113
unique         11
top    Theft from Vehicle
freq           167776
Name: incident, dtype: object
```

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

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total In..	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	13,701	1,259	1,192	1,102	2,193	761	875	754	738	614	647	707	898	909	1,052
Other Days	451,667	45,526	44,649	40,155	36,143	32,884	30,679	27,898	25,584	24,454	25,155	25,051	28,968	30,582	33,939

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total In..	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	2.94%	2.69%	2.60%	2.67%	5.72%	2.26%	2.77%	2.63%	2.80%	2.45%	2.51%	2.74%	3.01%	2.89%	3.01%
Other Days	97.06%	97.31%	97.40%	97.33%	94.28%	97.74%	97.23%	97.37%	97.20%	97.55%	97.49%	97.26%	96.99%	97.11%	96.99%

Caption

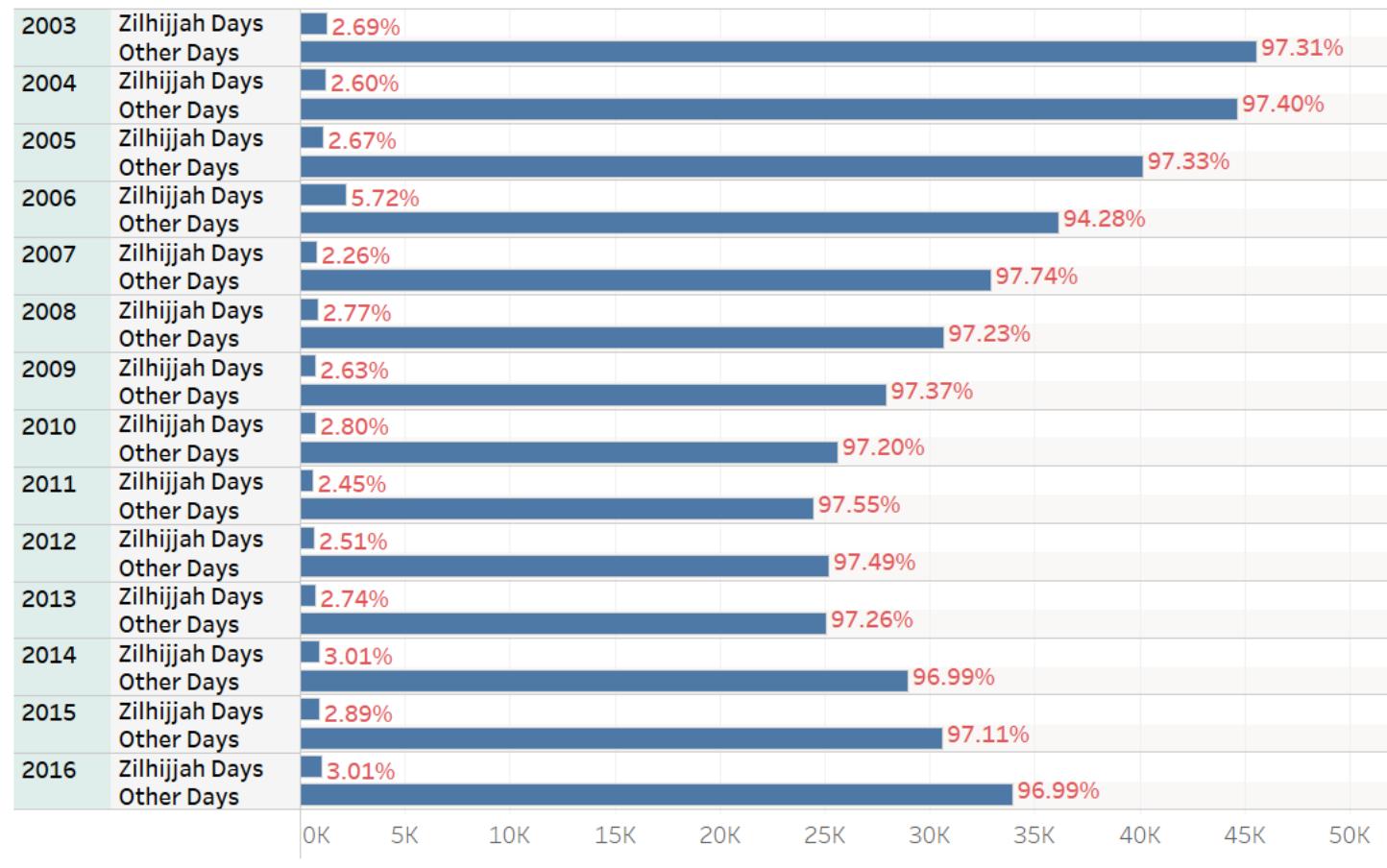
There is a 7.52% higher rate of crime committed during the Zilhijjah days compared to other days.

$$2.94/97.06=0.03029$$

$$0.03029/0.02817=7.52\%$$

Crime Rates: Zilhijjah Days vs Other Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]



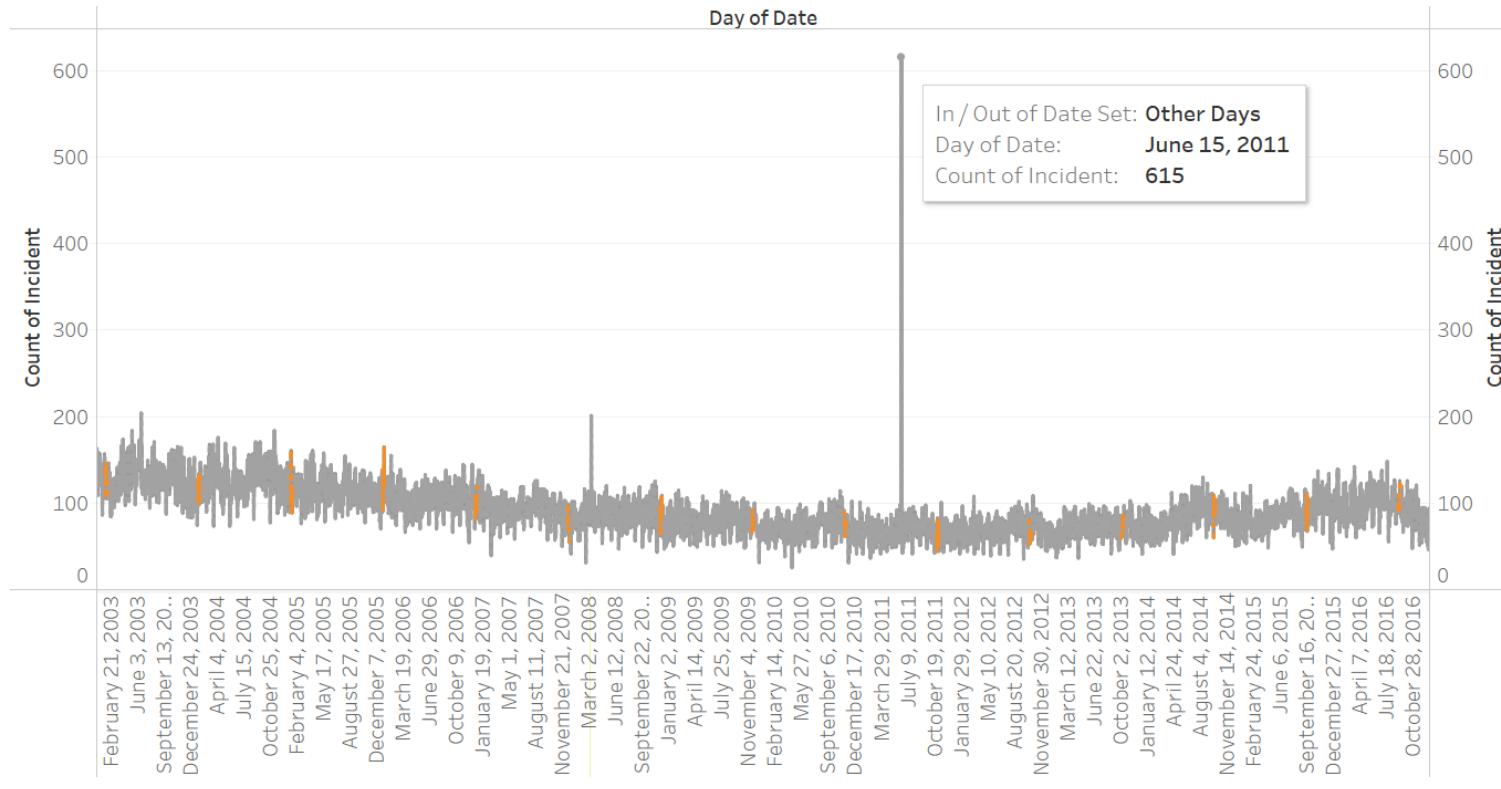
Caption

In half of the years, crime rates during the initial ten days of Zilhijjah were higher compared to the rest of the year.

In total, there was a 7.52% higher rate of crime committed during the Zilhijjah days compared to other days.

When we exclude the records of 2017, which do not include Zilhijjah days, it is observed that during the Zilhijjah days of 15 years, there was about a 7.52% higher rate of crimes compared to the other days of those years. The 2017 data spans until July and does not include the month of Zilhijjah. However, upon including the data from 2017, the excess crime rate during Zilhijjah days would decrease to around 0.6%. Another point against Zilhijjah is that calculations were made without excluding extreme events (outliers) occurring on other days. For instance, the 'Vancouver Stanley Cup riot' on June 15, 2011, alone lowers the crime rates in 2011 against Zilhijjah by about 2.25%. Considering these aspects, it would be more meaningful to accept the excess crime rate during Zilhijjah days as 7.52% rather than 0.6%.

DAYLY COUNTS



IN/OUT(Date Set)

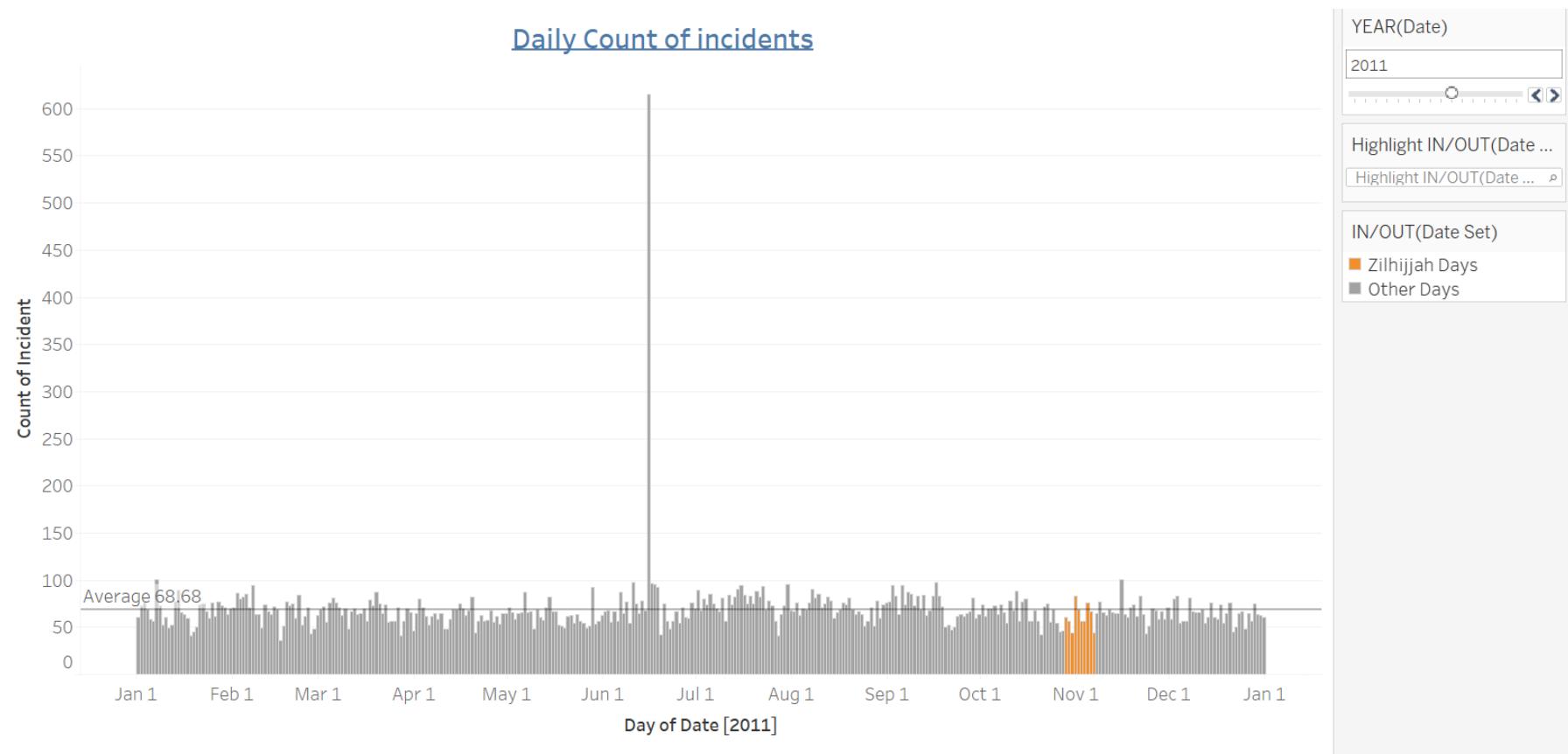
Zilhijjah Days
Other Days

Caption

Notice the point against Zilhijjah is that calculations were made without excluding extreme events (outliers) occurring on other days.

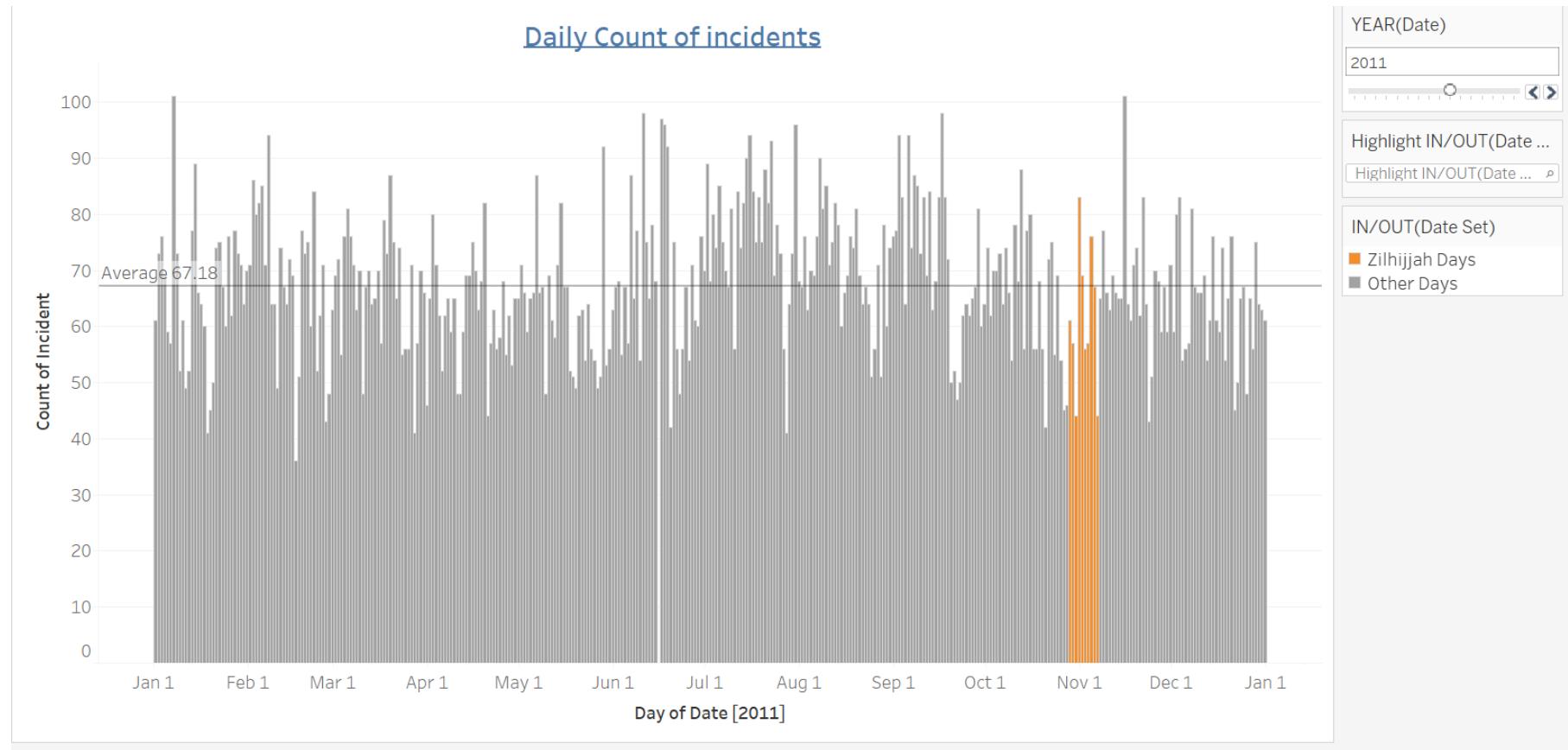
For instance, the 'Vancouver Stanley Cup riot' on June 15, 2011, alone lowers the crime rates in 2011 against Zilhijjah by about 2.25%.

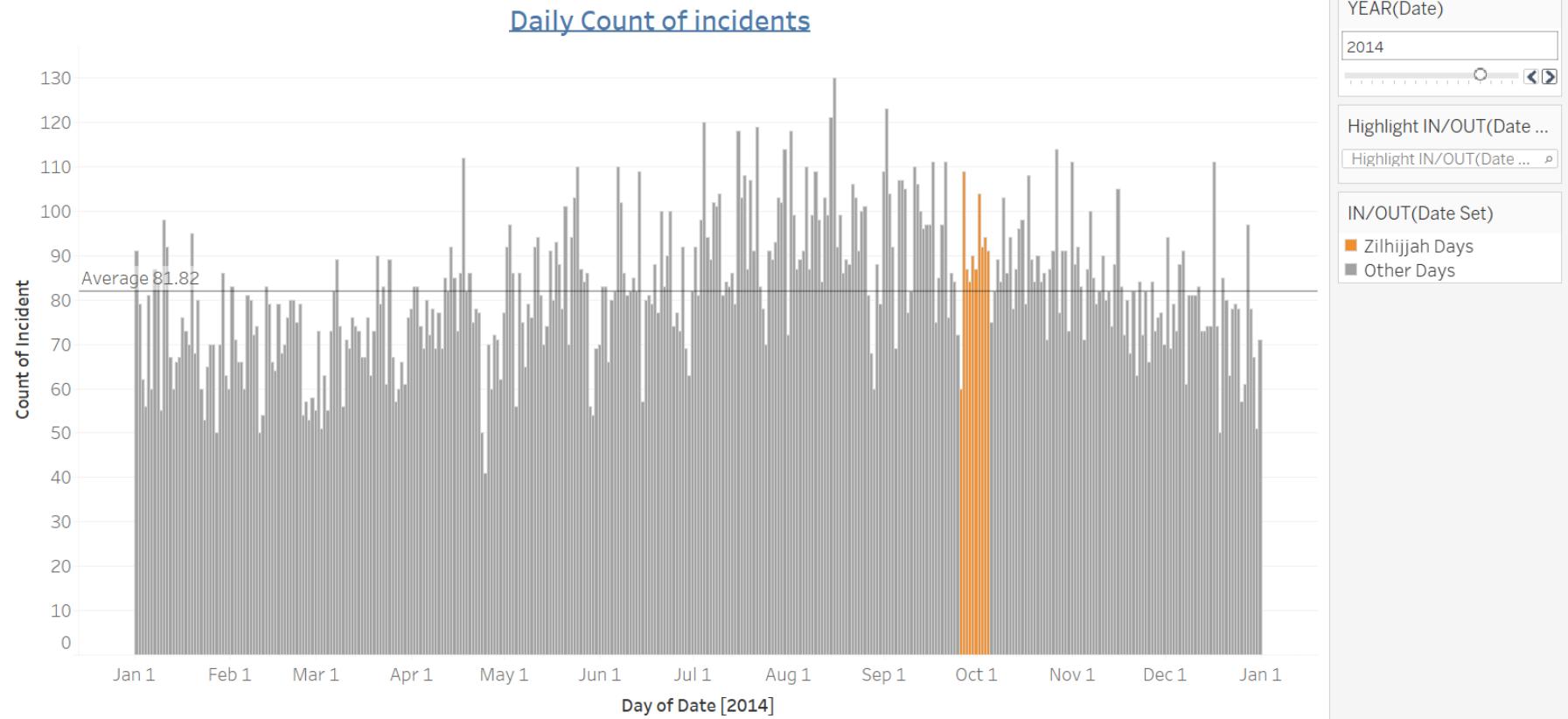
Considering these aspects, it would be more meaningful to accept the excess crime rate during Zilhijjah days as 7.52% rather than 0.6%.



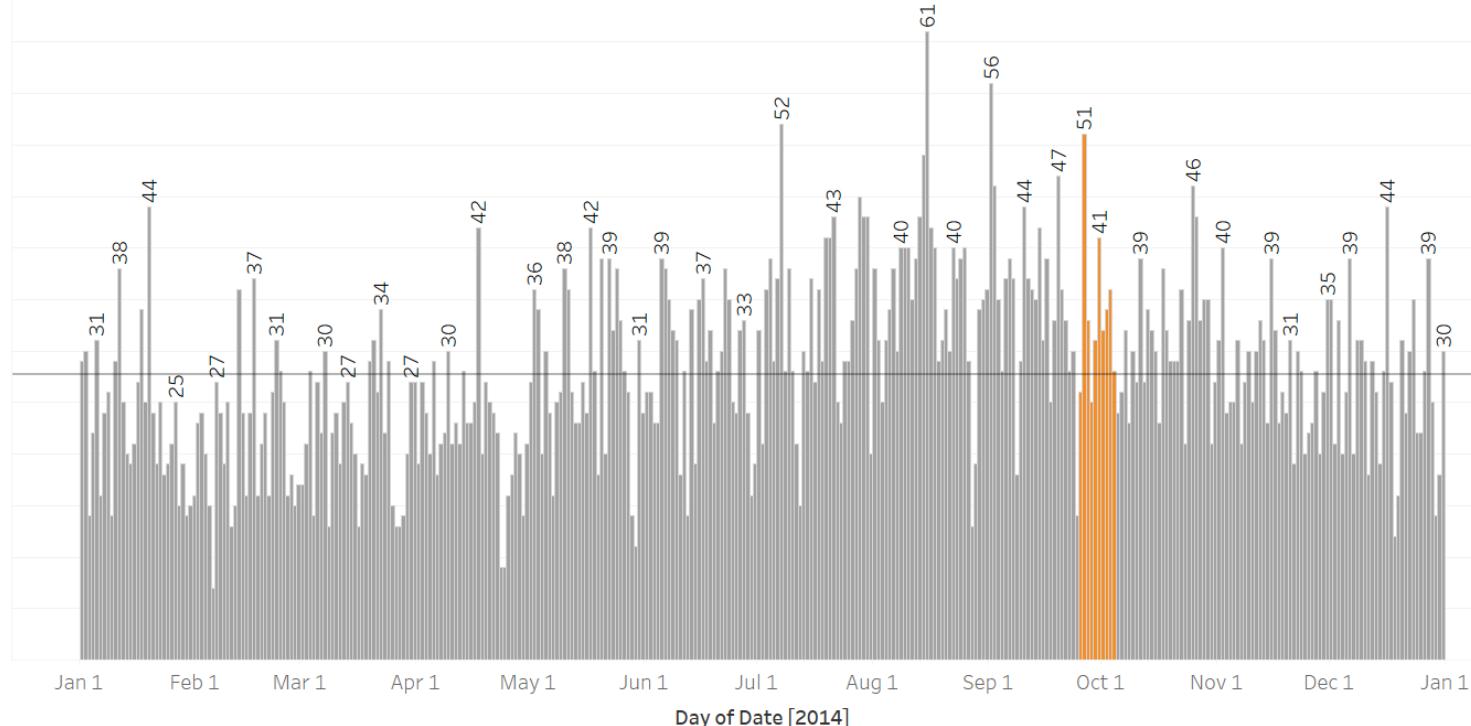
In [126]:

At the graphic below the 'Vancouver Stanley Cup riot' on June 15, 2011 is excluded.





DAILY_TYPES - Theft from Vehicle



Incident

- (All)
- Break and Enter ...
- Break and Enter ...
- Homicide
- Mischief
- Offence Against ...

YEAR(Date)

- 2010
- 2011
- 2012
- 2013
- 2014
- 2015

IN/OUT(Date Set)

- Zilhijjah Days
- Other Days

Highlight IN/OUT(Date ...

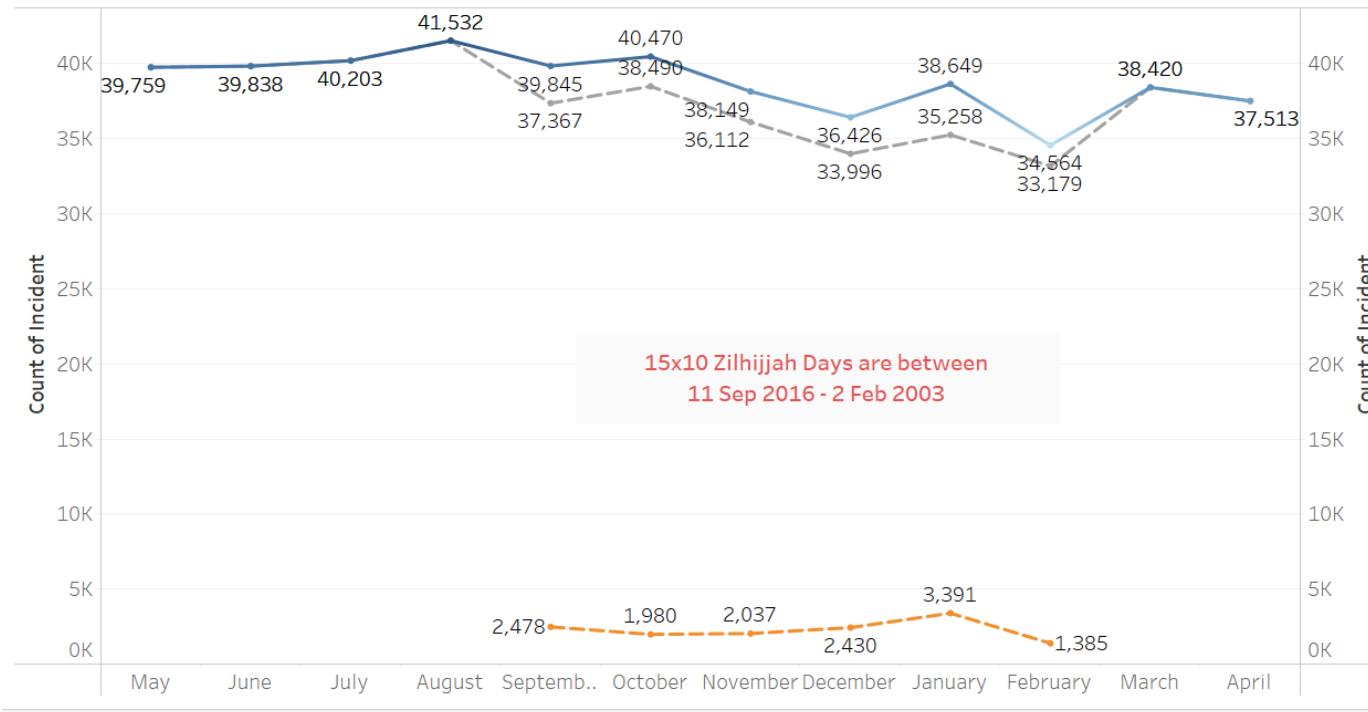
Highlight IN/OUT(Date ...

Incident

Theft from Vehicle

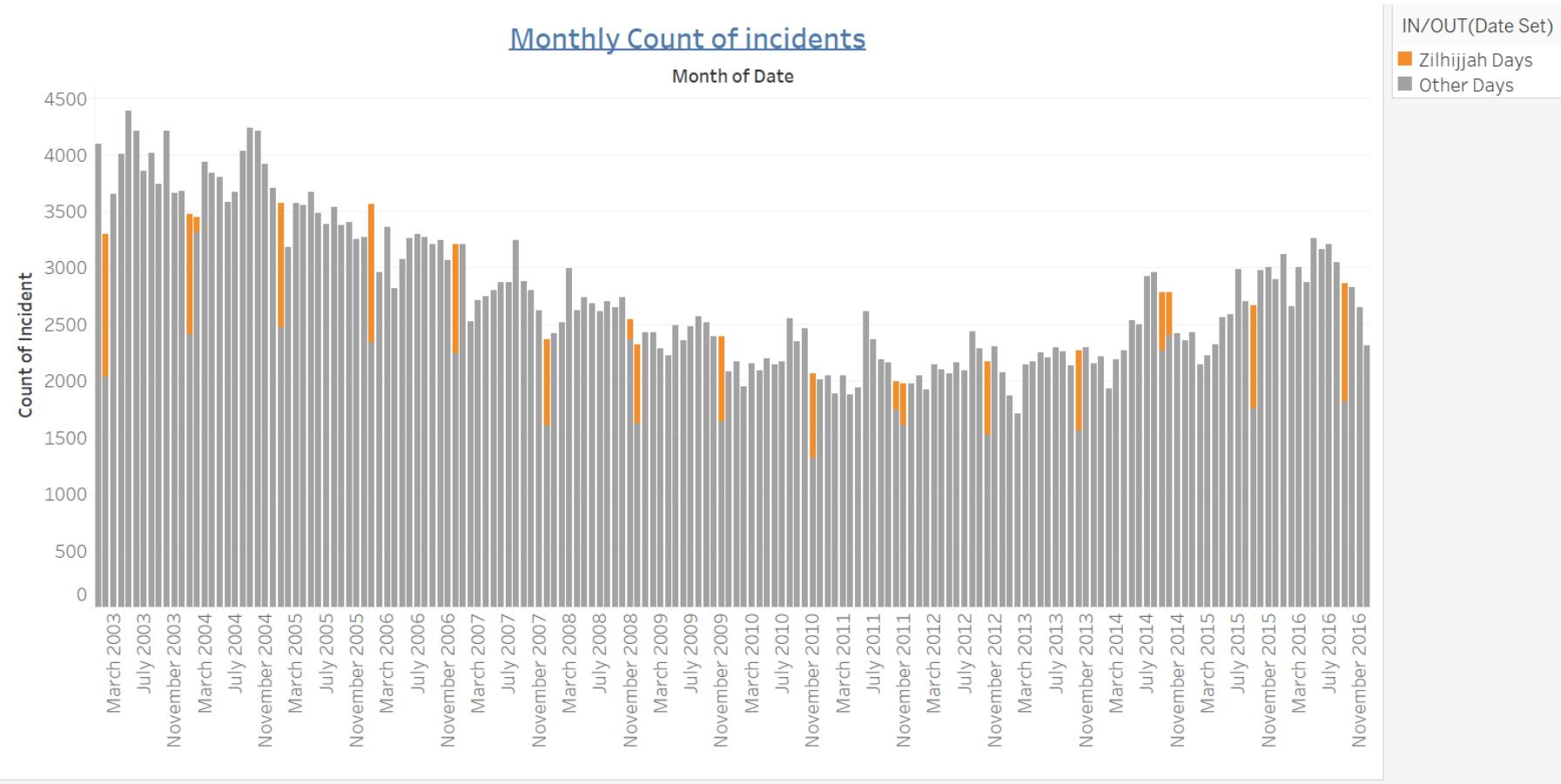
Show history

Monthly Counts of Incidents:
(Zilhijjah Days are also separately shown below.)



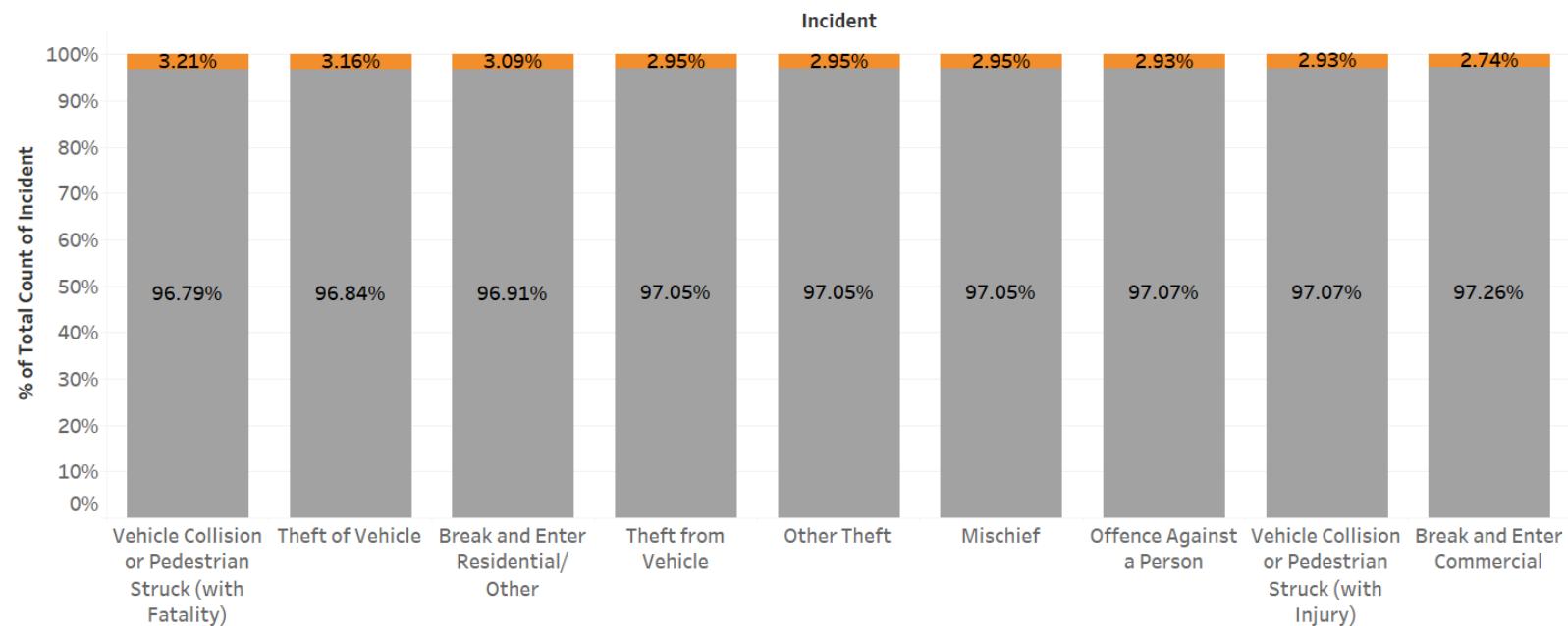
Caption

Overall, there has been a parallel trend between Zilhijjah and other days, with more crimes committed during the Zilhijjah days coinciding with the months of January.



The crime types which have a higher rate during the first ten days of Zilhijjah, compared to the rest of the year.

IN/OUT(Date ...
Zilhijjah Days
Other Days



SAMPLE DATA-7: CHICAGO CRIME DATASET_2001-2023

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

In [127...]

```
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[127]:

	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	Domestic	Beat	...	Ward	Community Area	FBI Code	Coordinates
ID															
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 [128... 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 [129]: # df = pd.read_csv("Chicago_Crimes_2001_to_Present.csv", usecols=[0, 2, 5, 6])
```

```
In [130]: df.duplicated().value_counts()
```

```
Out[130]: False    7953669
           True     161
           dtype: int64
```

```
row_values = df.iloc[154878].values # 154878 indeksine sahip satırın değerlerini al
matching_rows = df[df.apply(lambda row: row.equals(row_values), axis=1)] #
Değerlere sahip tüm satırları bul
print(matching_rows)
```

```
In [131]: df = df.drop_duplicates()
```

```
In [132]: df = df.iloc[:, [1, 4, 5]]
```

```
In [133]: 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 [134]: df["Primary Type"].value_counts()
```

```
Out[134]:
```

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 [135...]
```

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

```
In [136...]
```

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

```
In [137]: df.date.min(), df.date.max()
```

```
Out[137]: ('2001-01-01', '2023-12-03')
```

```
# first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2001-02-24' - '2001-03-05' '2002-02-13' - '2002-02-22' '2003-02-02' - '2003-02-11'  
'2004-01-23' - '2004-02-01' '2005-01-12' - '2005-01-21' '2006-01-01' - '2006-01-10' '2006-12-22' - '2006-12-31' '2007-12-11' - '2007-12-20' '2008-11-29' - '2008-12-08' '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' -  
'2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11' '2017-08-23' - '2017-09-01' '2018-08-12' - '2018-08-21' '2019-08-02' - '2019-08-11' '2020-07-22' - '2020-07-31' '2021-07-11' - '2021-07-20' '2022-06-30' - '2022-07-09' '2023-06-19' - '2023-06-28'
```

```
In [138]: # df.to_csv("Chicago.csv", index=False)
```

```
In [139]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[139]: 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 [140...]: # 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 [141]: df.date.unique()
```

```
Out[141]: 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 [142]: zilhijjah_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 Zilhijjah months: 931.9792
```

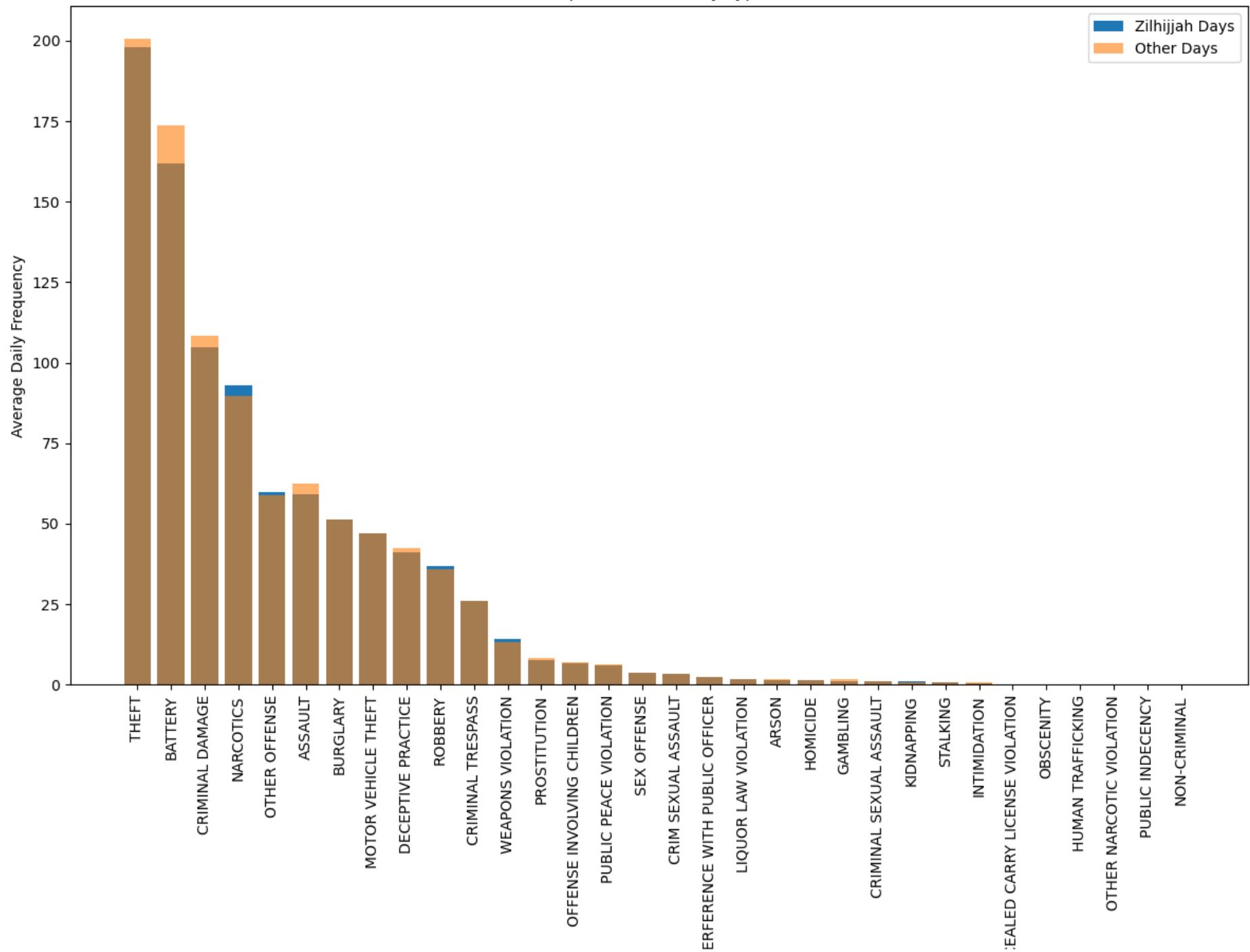
```
-----  
Average case count in other days: 950.5649
```

```
-----  
Ratio of Zilhijjah cases to other cases: 0.9804
```

We observe a 1.96% lower crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [143...]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type



CONC

INT

Incident Types

```
In [144]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio  
sorted_ratios
```

Out[144]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
RITUALISM	2	24	0.083300
HUMAN TRAFFICKING	7	105	0.066700
OTHER NARCOTIC VIOLATION	6	149	0.040300
OBSCENITY	30	840	0.035700
CRIMINAL SEXUAL ASSAULT	263	7881	0.033400
CONCEALED CARRY LICENSE VIOLATION	40	1205	0.033200
STALKING	163	5136	0.031700
WEAPONS VIOLATION	3405	112009	0.030400
NARCOTICS	22342	750859	0.029800
CRIM SEXUAL ASSAULT	816	27555	0.029600
KIDNAPPING	216	7320	0.029500
ROBBERY	8814	299815	0.029400
OTHER OFFENSE	14326	493465	0.029000
INTERFERENCE WITH PUBLIC OFFICER	539	18746	0.028800
CRIMINAL TRESPASS	6253	217211	0.028800
BURGLARY	12323	429109	0.028700
OFFENSE INVOLVING CHILDREN	1632	56968	0.028600
MOTOR VEHICLE THEFT	11246	393673	0.028600
HOMICIDE	364	12820	0.028400
SEX OFFENSE	898	31684	0.028300
THEFT	47515	1679300	0.028300
DECEPTIVE PRACTICE	9888	356129	0.027800
PUBLIC PEACE VIOLATION	1466	52892	0.027700
CRIMINAL DAMAGE	25135	906654	0.027700

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
ASSAULT	14200	521978	0.027200
NON-CRIMINAL	5	184	0.027200
ARSON	368	13586	0.027100
BATTERY	38823	1451708	0.026700
NON - CRIMINAL	1	38	0.026300
PROSTITUTION	1811	69958	0.025900

In [145]:

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

Out[145]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
NARCOTICS	22342	750859	0.0298
OTHER OFFENSE	14326	493465	0.0290
BURGLARY	12323	429109	0.0287
ROBBERY	8814	299815	0.0294
CRIMINAL TRESPASS	6253	217211	0.0288
WEAPONS VIOLATION	3405	112009	0.0304
CRIM SEXUAL ASSAULT	816	27555	0.0296
INTERFERENCE WITH PUBLIC OFFICER	539	18746	0.0288
CRIMINAL SEXUAL ASSAULT	263	7881	0.0334
KIDNAPPING	216	7320	0.0295
STALKING	163	5136	0.0317
CONCEALED CARRY LICENSE VIOLATION	40	1205	0.0332
OBSCENITY	30	840	0.0357
HUMAN TRAFFICKING	7	105	0.0667
OTHER NARCOTIC VIOLATION	6	149	0.0403
RITUALISM	2	24	0.0833

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

In [146...]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]

Out[146]: (36, 16)

In [147...]: zilhijjah_incidents_desc

Out[147]: count 223675
unique 34
top THEFT
freq 47515
Name: incident, dtype: object

In [148]: other_days_incidents_desc

Out[148]:

count	7729994
unique	36
top	THEFT
freq	1631785
Name:	incident, dtype: object

Chicago dataset encompasses 36 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 34 incident categories, with 16 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total In..	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Zilhijjah Days	223,675	12,790	12,207	11,316	10,763	10,901	22,613	10,560	10,290	9,976	10,149	9,875	9,257	8,098	8,226	7,827	7,857	7,716	8,118	8,191	6,228	5,814	7,179	7,724
Other Days	7,729,994	473,106	474,600	464,660	458,660	442,873	425,563	426,522	416,887	382,848	360,358	342,115	327,065	299,443	267,575	256,980	261,983	261,392	260,809	253,200	206,035	203,158	231,878	232,284

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total I..	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Zilhijjah Days	2.81%	2.63%	2.51%	2.38%	2.29%	2.40%	5.05%	2.42%	2.41%	2.54%	2.74%	2.81%	2.75%	2.63%	2.98%	2.96%	2.91%	2.87%	3.02%	3.13%	2.93%	2.78%	3.00%	3.22%
Other Days	97.19%	97.37%	97.49%	97.62%	97.71%	97.60%	94.95%	97.58%	97.59%	97.46%	97.26%	97.19%	97.25%	97.37%	97.02%	97.04%	97.09%	97.13%	96.98%	96.87%	97.07%	97.22%	97.00%	96.78%

Ratio of Zilhijjah/Other Days is 223,675 incidents / 7,729,994 incidents = 0.028936 Equal Ratio = 240 days / (8372-240) days = 0.029513 Ratio of Zilhijjah cases to other cases: 0.028936/0.029513 = 0.9804 = - 1.96%

Calculation Method I:

Count Zilhijjah Incidents / Other Days Incidents:

223,675 incidents / 7,729,994 incidents = 0.028936

Equal Ratio = 240 days / (8372-240) days = 0.029513

Ratio of Zilhijjah cases to other cases:

0.028936 / 0.029513 = 0.9804 = - 1.96%

Calculation Method II:

Average case count in the first ten days of Zilhijjah months: 223,675 incidents / 240 days = 932 incidents per day

Average case count in other days: 7,729,994 incidents / (8372-240) days = 951 incidents per day

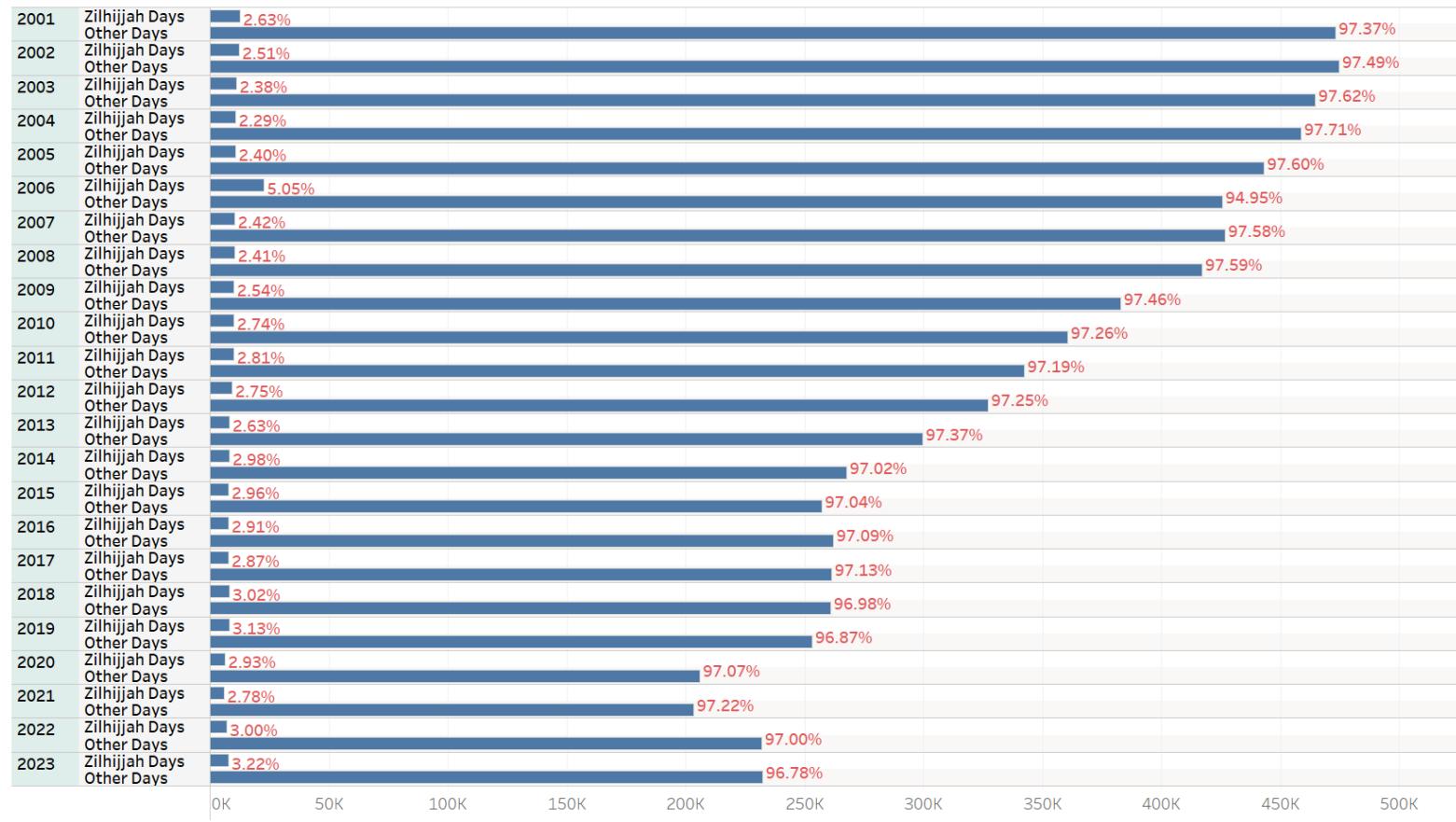
Ratio of Zilhijjah cases to other cases:

932 / 951 = 0.9804 = - 1.96%

In this analysis, we are examining whether there is a cyclical, recurring pattern in our data according to the Hijri calendar. Therefore, calculations should ideally be done based on the Hijri calendar, and the pattern should be demonstrated over the Hijri calendar. If a pattern is identified, displaying this pattern directly on the Gregorian calendar would be somewhat challenging. For instance, in the Chicago dataset, we have records spanning 23 Gregorian years. However, as the Hijri year is 10 days shorter, we have 24 Zilhijjah months in our dataset instead of 23. There are two Zilhijjah months in the year 2006. The situation here arises from the difference between the 29/30 days in the Hijri month and the 30/31 days in the Gregorian month, which becomes prominent in datasets covering extended periods. Essentially, all calculations, comparisons, and visualizations should be performed according to the Hijri months. Otherwise, while the overall difference might not be noticeable, in a year-over-year representation like the one below, we will observe that the excess of Zilhijjah in the year 2006 is significantly high, with the surrounding years being relatively lower. If this yearly representation were based on the Hijri calendar, the year averages would have been more balanced and evenly distributed. However, due to the complexities in running diverse codes related to the Hijri Calendar in the pandas library, Python, and Tableau, we often find ourselves analyzing data based on the Gregorian calendar.

Crime Rates: Zilhijjah Days vs Other Days

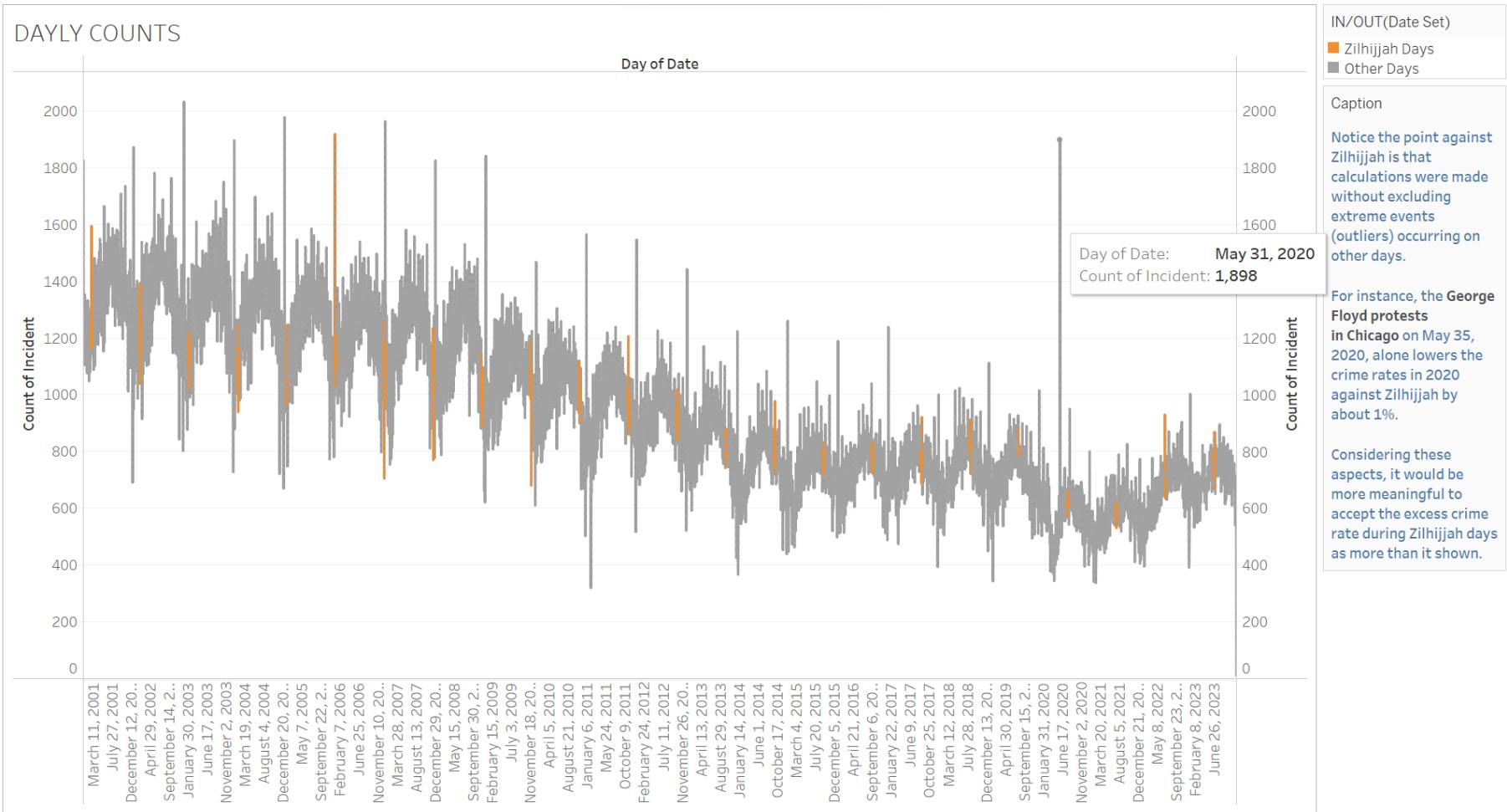
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

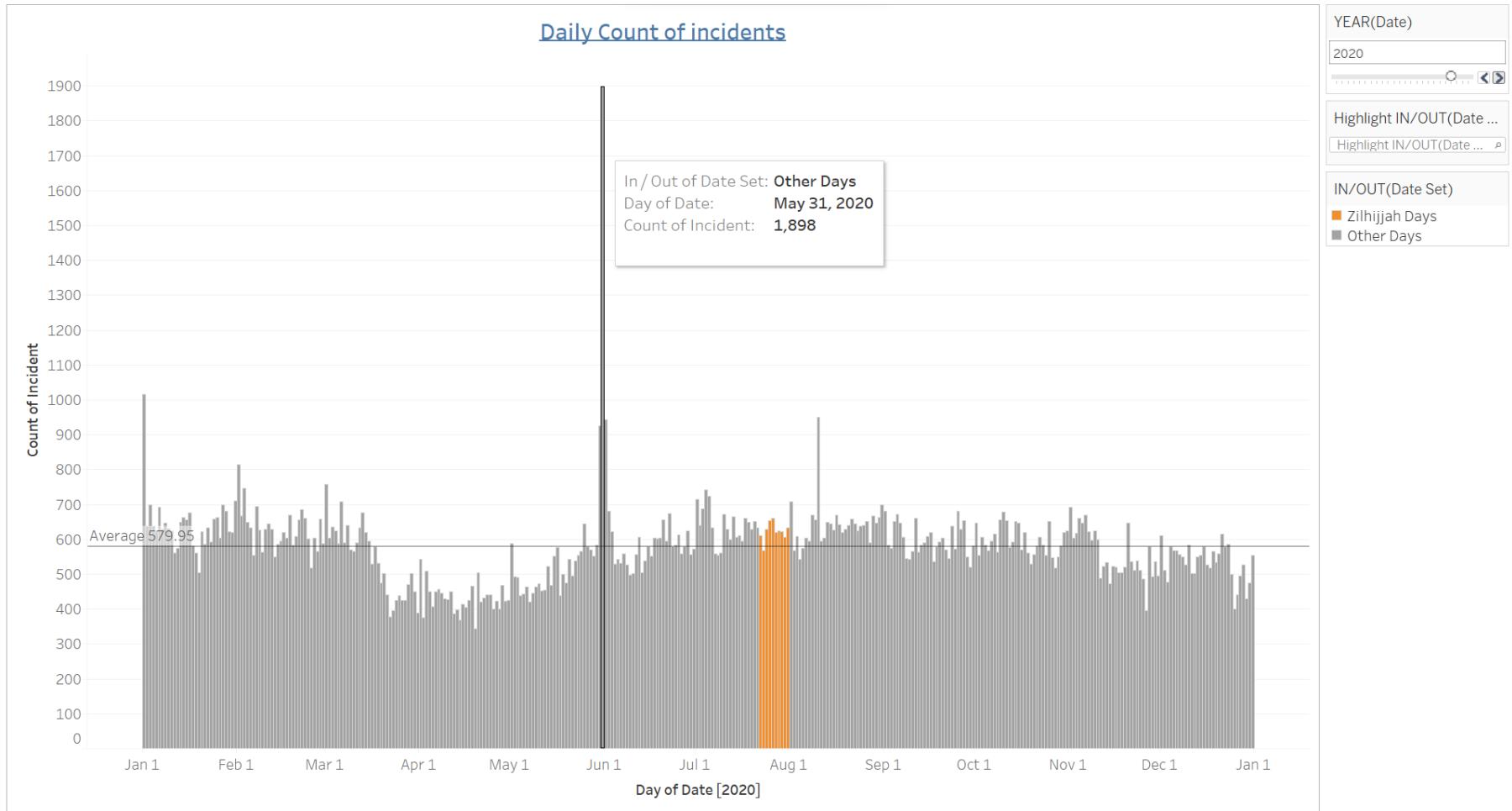


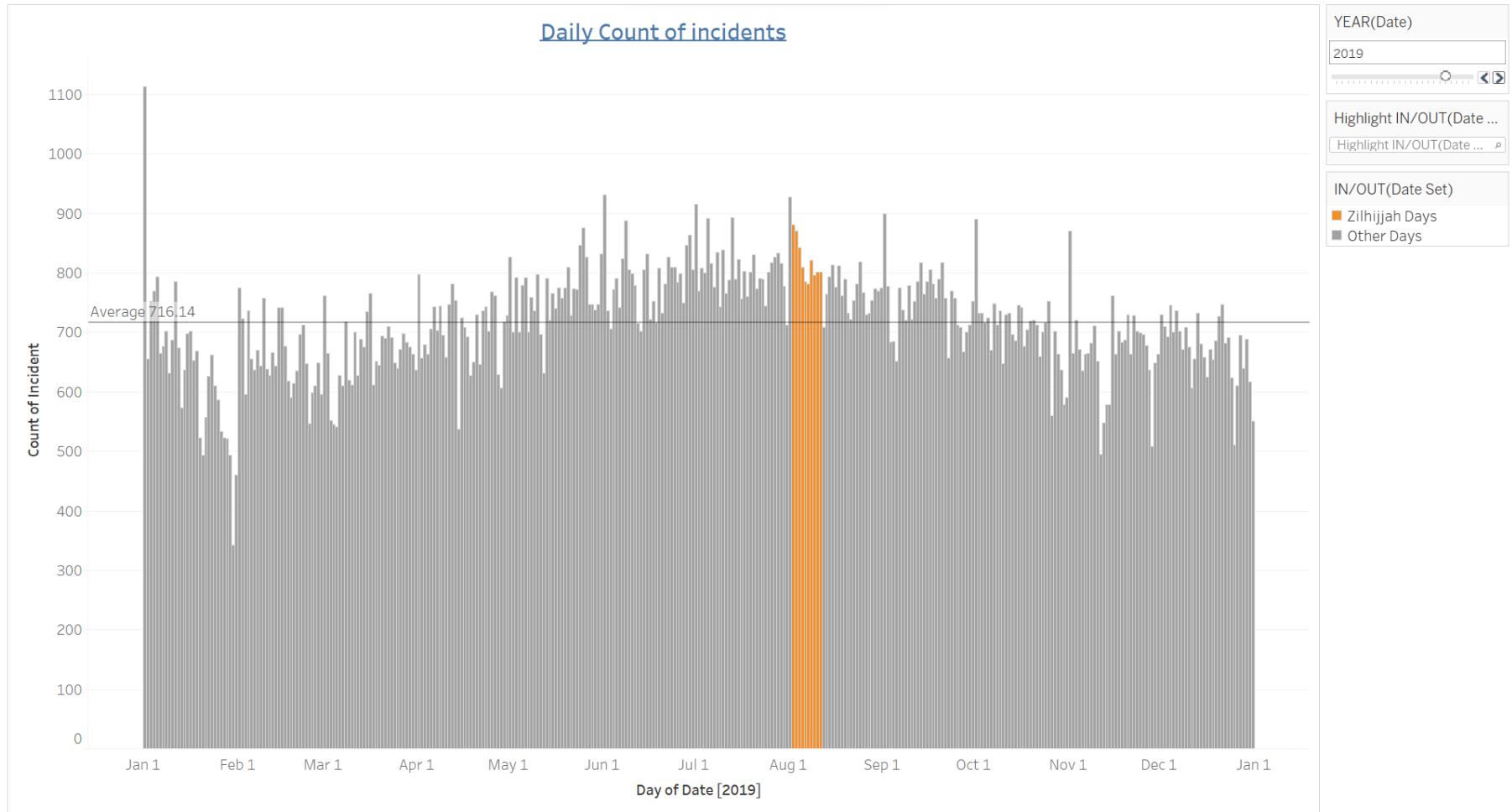
Caption

In 13 out of the 23 years, more crimes were committed during the Zilhijjah periods, in 9 years fewer crimes, and in one year (2010) an equal amount.

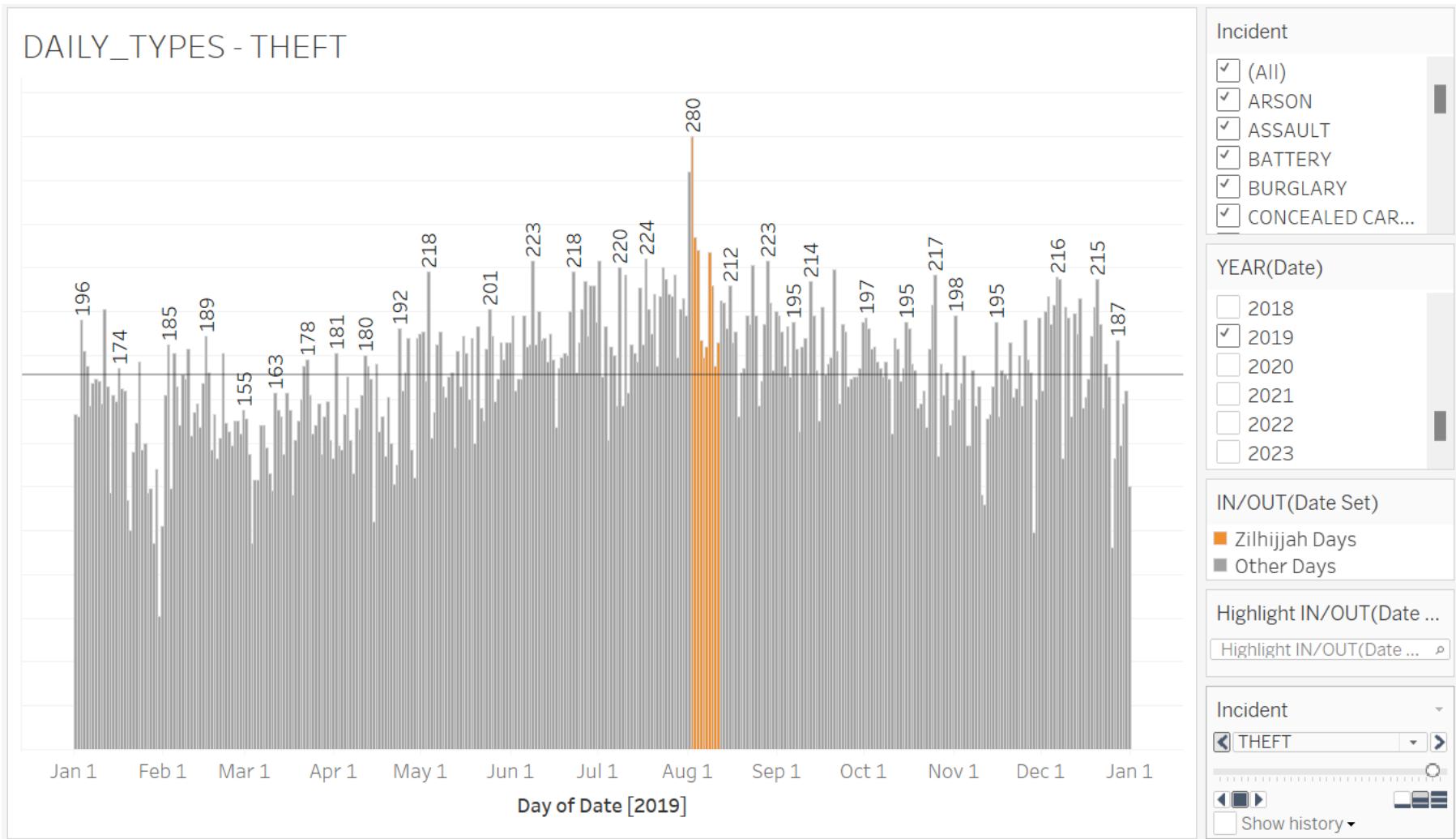
Interestingly, as the years progress, the total number of crimes decreases, yet conversely, more crimes are observed during the Zilhijjah periods.





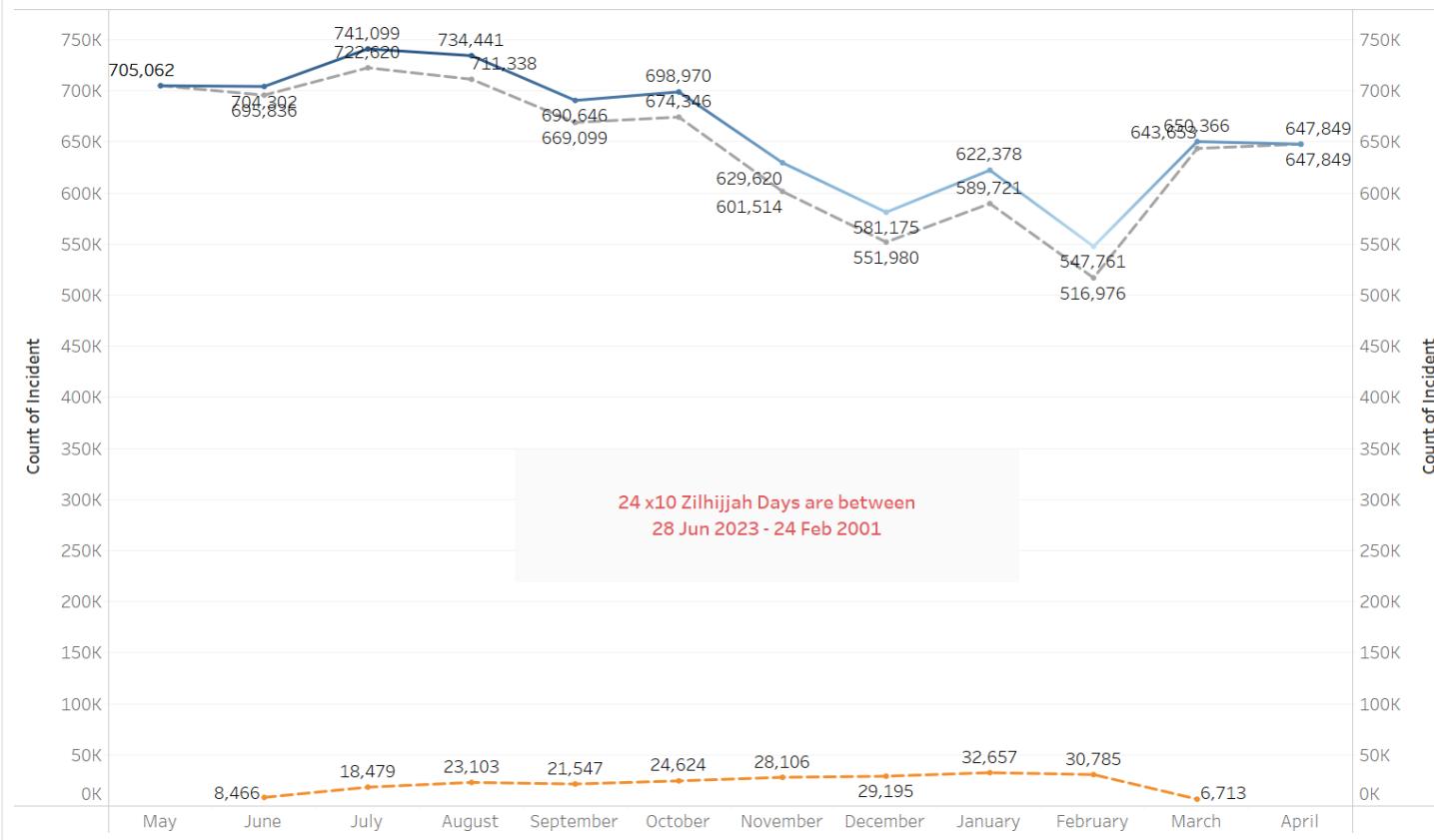


DAILY_TYPES - THEFT

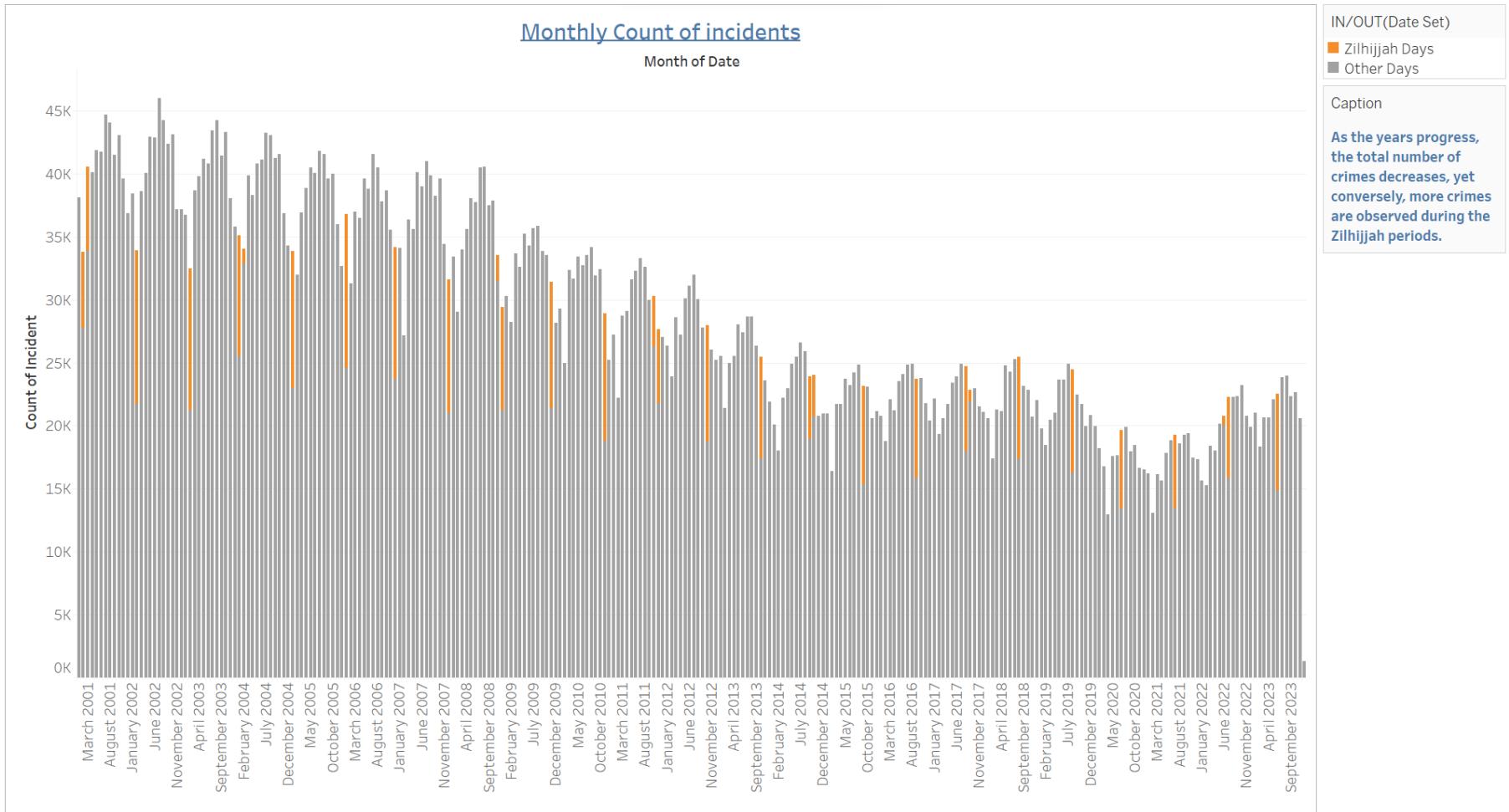


Caption

Monthly Counts of Incidents:
(Zilhijjah Days are also separately shown below.)

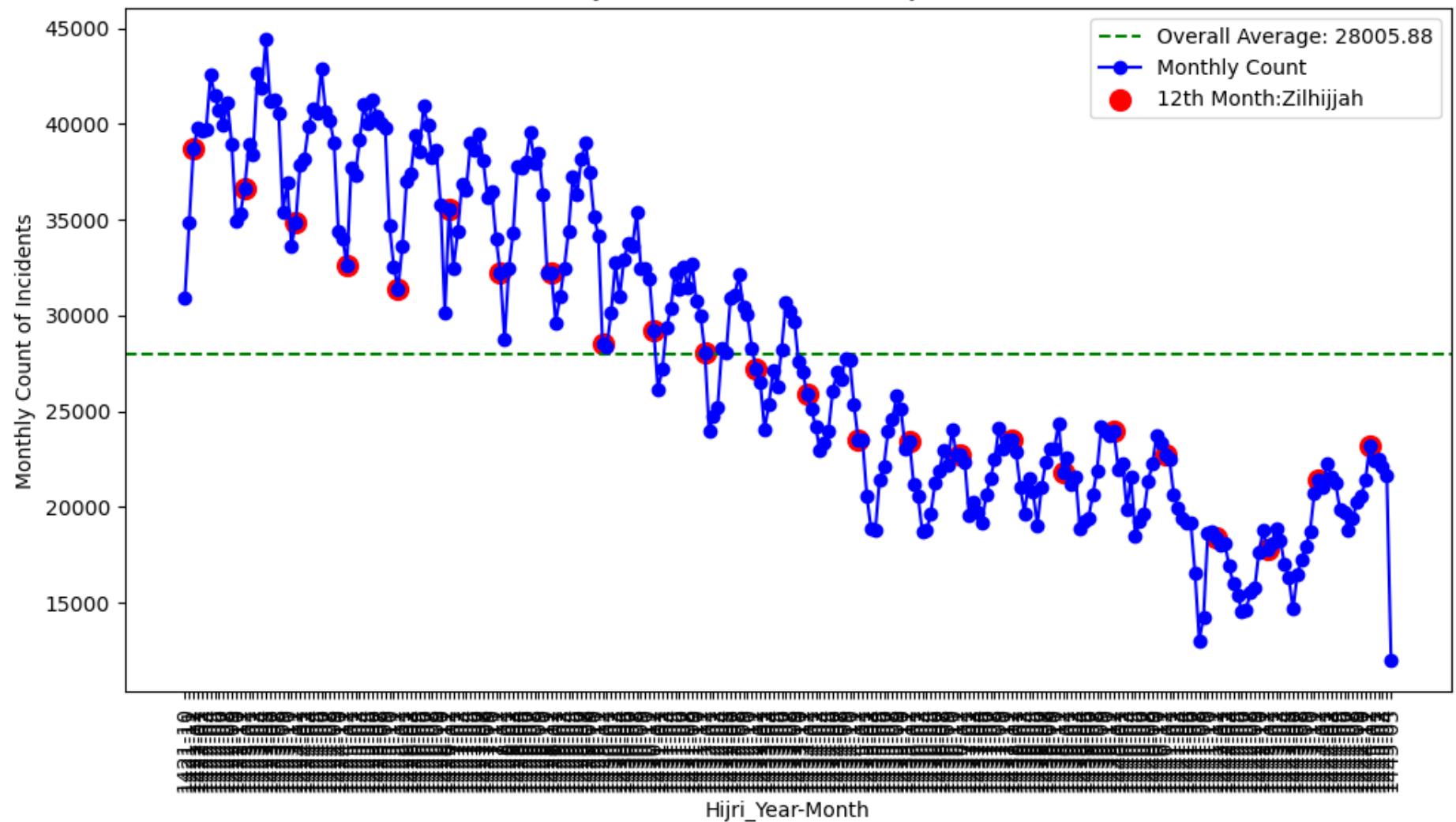


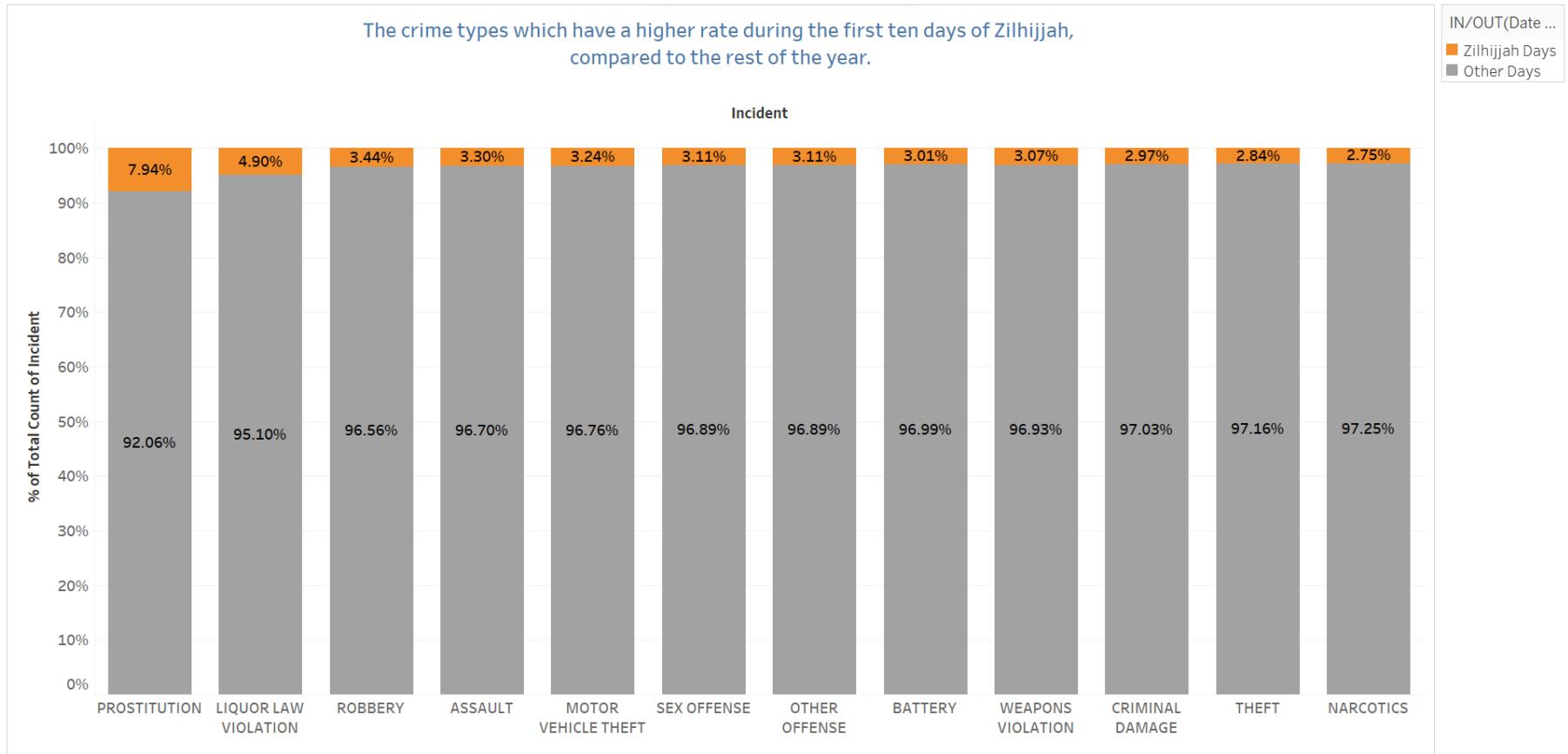
Overall, there has been a parallel trend between Zilhijjah and other days, with more crimes committed during the Zilhijjah days coinciding with the months of January.



In [149]: monthly_count_plot()

Monthly Count of Incidents on Hijri Calendar





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 [169...]

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

Out[169]:

	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	McElberry 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 [170...]: df.duplicated().value_counts()

Out[170]:

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

In [171...]: df = df.drop_duplicates()

In [172...]: 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 [173...]

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

```
Out[173]:
```

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
10      88
3NK      80
7B       80
8AV      68
8BO      67
3AJK     65
3EF      60
3M       58
3AJ0     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 [174...]: df = df.rename(columns = {'CrimeDate':'date'})
df = df.rename(columns= {'CrimeCode' : 'incident'})
```

```
In [175...]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [176]: df.date.min(), df.date.max()
```

```
Out[176]: ('2011-01-01', '2016-06-18')
```

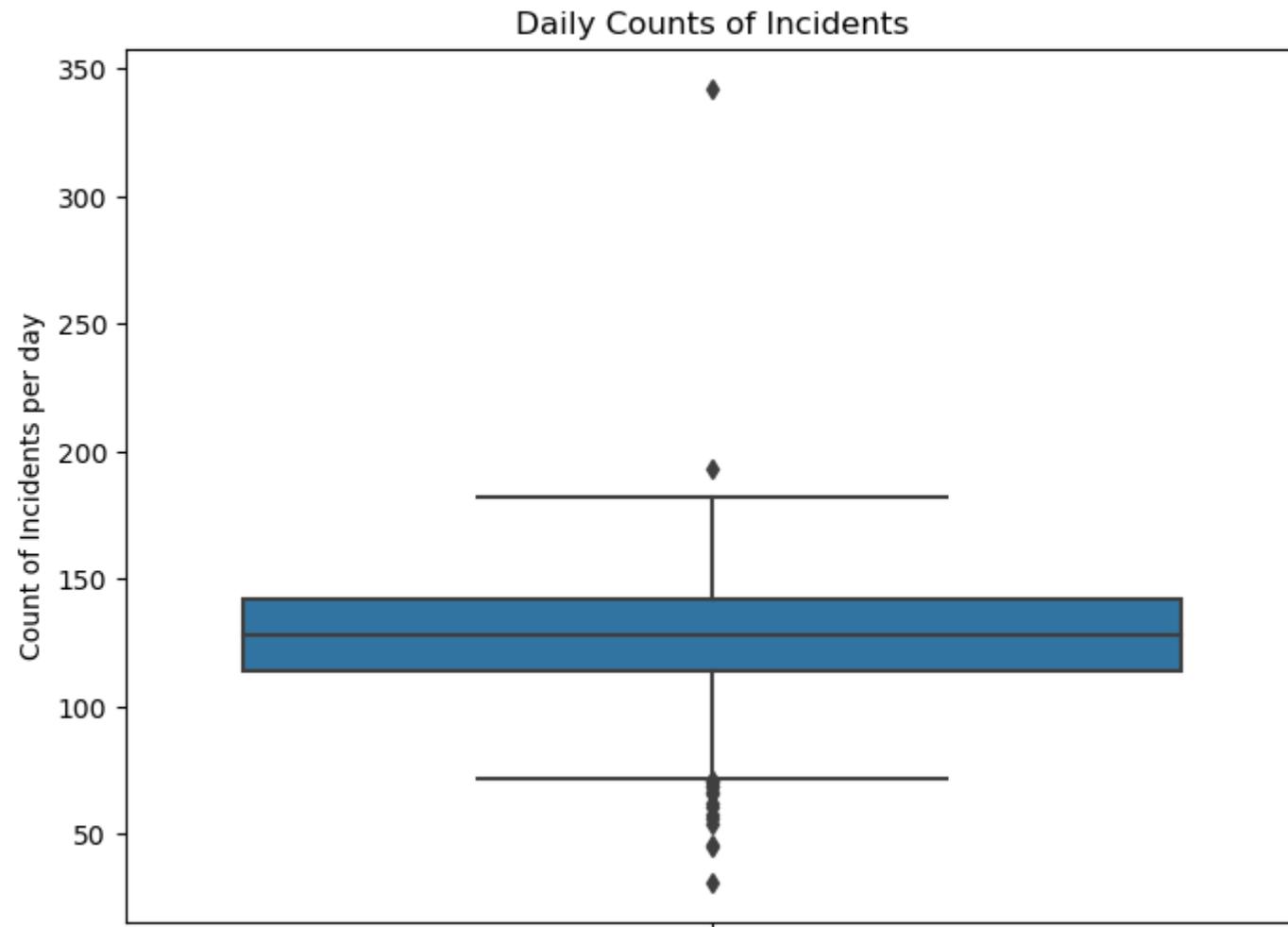
The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23'

```
In [177]: df = df.iloc[:, [0,2]]  
# df.to_csv("Baltimore.csv", index=False)
```

```
In [178]: daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[178]: 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 [179]: # 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 [180]: df.date.nunique()
```

```
Out[180]: 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.

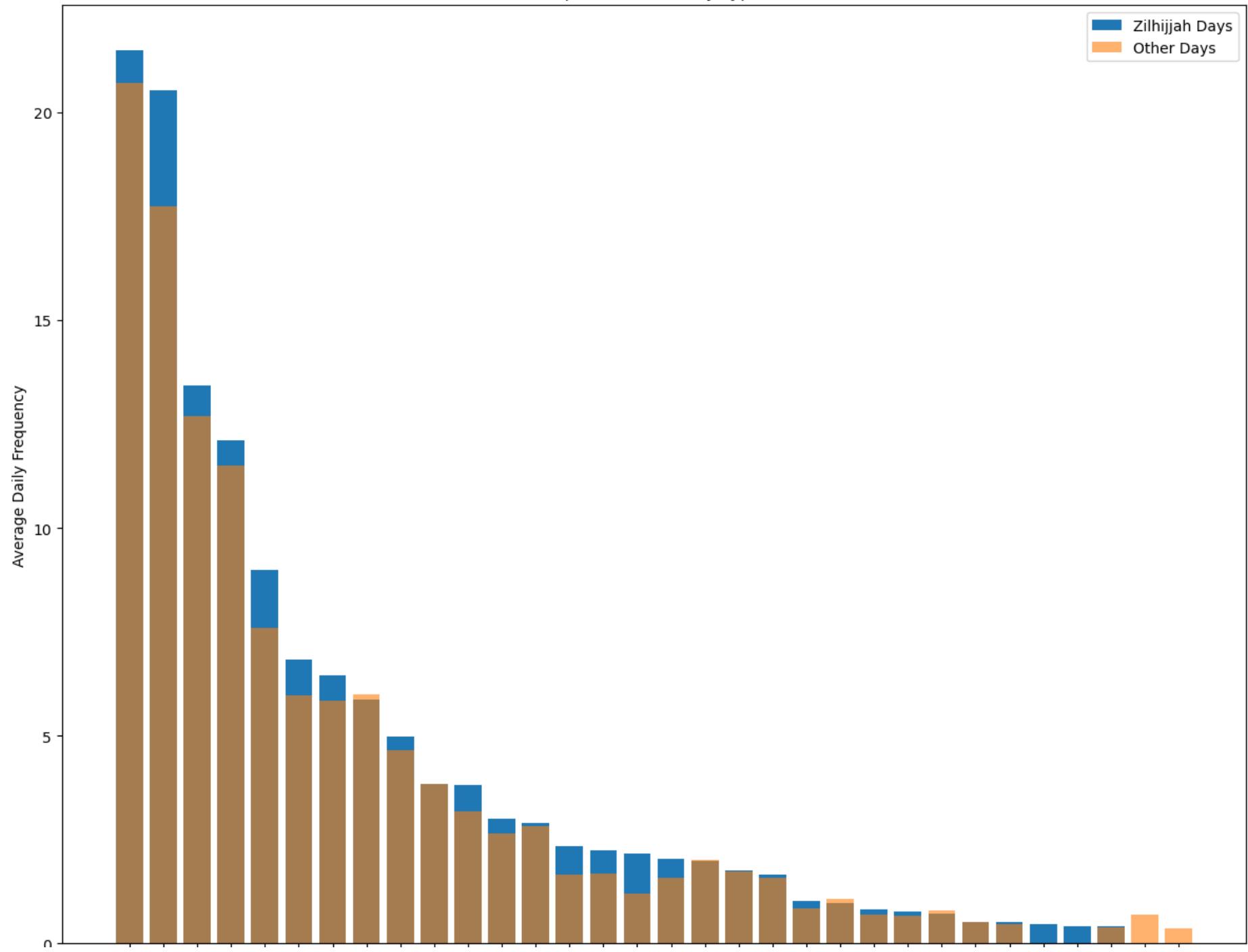
```
In [181]: zilhijjah_10_days(df)
```

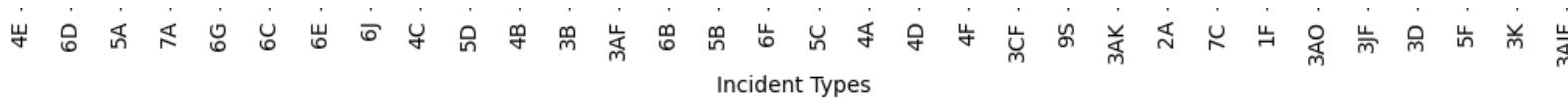
```
Total number of days: 1996
-----
Total number of cases: 253900
-----
Average Daily Case Count: 127.2
-----
Yearly case counts according to the Gregorian calendar:
-----
2011    48555
2012    47643
2013    47569
2015    46643
2014    44343
2016    19147
Name: date, dtype: int64
-----
Case counts according 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 Zilhijjah months: 138.9
-----
Average case count in other days: 126.9039
-----
Ratio of Zilhijjah cases to other cases: 1.0945
```

We observe a 9.45% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [182]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type





In [183]:
Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio
sorted_ratios

Out[183]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
8GO	1	15	0.066700
7B	5	80	0.062500
8BO	4	67	0.059700
3LO	1	17	0.058800
6F	108	2436	0.044300
8EO	2	50	0.040000
3AJO	2	52	0.038500
3JK	7	185	0.037800
6H	5	138	0.036200
3CO	7	194	0.036100
2B	8	223	0.035900
3JF	23	645	0.035700
6B	117	3345	0.035000
3M	2	58	0.034500
3D	21	625	0.033600
5B	112	3367	0.033300
3P	18	544	0.033100
5C	102	3187	0.032000
3CF	51	1697	0.030100
4B	191	6393	0.029900
3AK	41	1381	0.029700
6G	449	15212	0.029500
6D	1026	35539	0.028900
2A	39	1354	0.028800

zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
6C	342	11939 0.028600
3GK	1	35 0.028600
3B	150	5296 0.028300
6E	323	11702 0.027600
3AO	26	949 0.027400
6A	12	439 0.027300

In [184]:

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

Out[184]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
4E	1075	41377	0.0260
6D	1026	35539	0.0289
5A	671	25352	0.0265
7A	605	22986	0.0263
6G	449	15212	0.0295
6C	342	11939	0.0286
6E	323	11702	0.0276
4C	249	9312	0.0267
4B	191	6393	0.0299
3B	150	5296	0.0283
3AF	145	5660	0.0256
6B	117	3345	0.0350
5B	112	3367	0.0333
6F	108	2436	0.0443
5C	102	3187	0.0320
4D	88	3482	0.0253
4F	83	3167	0.0262
3CF	51	1697	0.0301
3AK	41	1381	0.0297
2A	39	1354	0.0288
3AO	26	949	0.0274
3JF	23	645	0.0357
3D	21	625	0.0336
5F	21	798	0.0263

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
3P	18	544	0.0331
6A	12	439	0.0273
6L	10	373	0.0268
2B	8	223	0.0359
3CK	7	259	0.0270
3CO	7	194	0.0361
3JK	7	185	0.0378
3BJ	6	221	0.0271
6H	5	138	0.0362
7B	5	80	0.0625
3JO	4	151	0.0265
8BO	4	67	0.0597
3M	2	58	0.0345
3AJO	2	52	0.0385
8EO	2	50	0.0400
3GK	1	35	0.0286
3LO	1	17	0.0588
8GO	1	15	0.0667

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

In [185...]: `df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]`

Out[185]: `(81, 42)`

In [186...]: `zilhijjah_incidents_desc`

```
Out[186]: count    6945  
unique     65  
top        4E  
freq      1075  
Name: incident, dtype: object
```

```
In [187... other_days_incidents_desc
```

```
Out[187]: count    246955  
unique     81  
top        4E  
freq      40302  
Name: incident, dtype: object
```

Baltimore dataset encompasses 81 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 65 incident categories, with 42 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2011	2012	2013	2014	2015	2016
Zilhijjah Days	6,945	1,431	1,354	1,366	1,312	1,482	
Other Days	246,955	47,124	46,289	46,203	43,031	45,161	19,147

Caption

2016 is included.

Calculation Method :

Average case count in the first ten days of Zilhijjah months: 6,945 incidents / 50 days = 138.9 incidents per day

Average case count in other days: 246,955 incidents / (1996-50) days = 126.9 incidents per day

Ratio of Zilhijjah cases to other cases:

138.9 / 126.9 = 1.0945 (9.45%)

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2011	2012	2013	2014	2015
Zilhijjah Days	6,945	1,431	1,354	1,366	1,312	1,482
Other Days	227,808	47,124	46,289	46,203	43,031	45,161

Caption

2016 is NOT included.

Calculation Method :

Count Zilhijjah Incidents / Other Days Incidents:

6,945 incidents / 227,808 incidents = **0.03048**

Zilhijjah / Other Days Ratio for a year: 10 days / 355 days = **0.02817**

0.03048 > 0.02817

Overall Zilhijjah / Others Ratio: 0.03048 / 0.02817 = **1.0820 (8.20%)**

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incidents	2011	2012	2013	2014	2015
Zilhijjah Days	2.958%	2.95%	2.84%	2.87%	2.96%	3.18%
Other Days	97.042%	97.05%	97.16%	97.13%	97.04%	96.82%

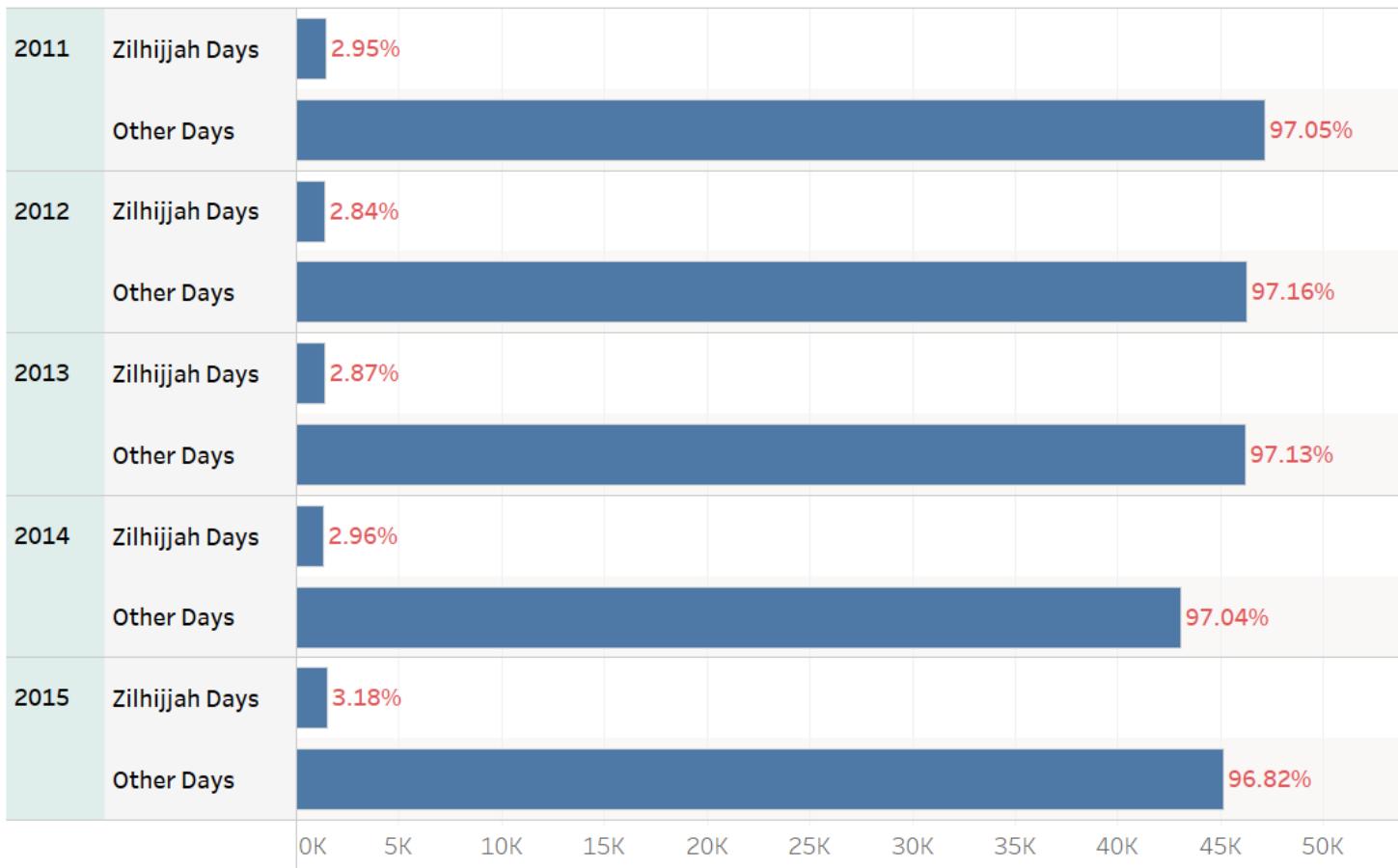
Caption

Zilhijjah / Others = 2.958 / 97.042 = 0.03048

0.03048 / 0.02817 = 8.20%

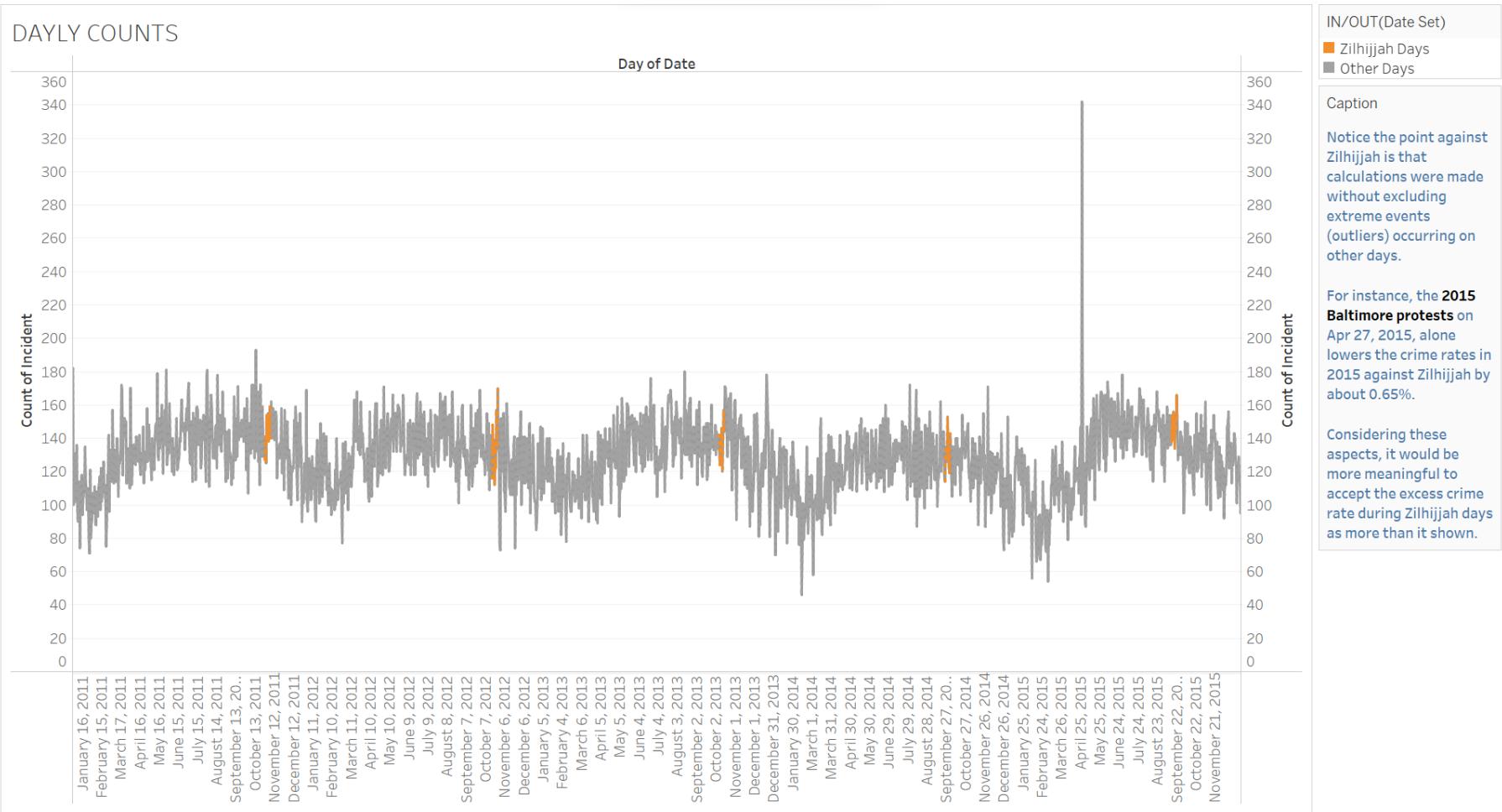
Crime Rates: Zilhijjah Days vs Other Days

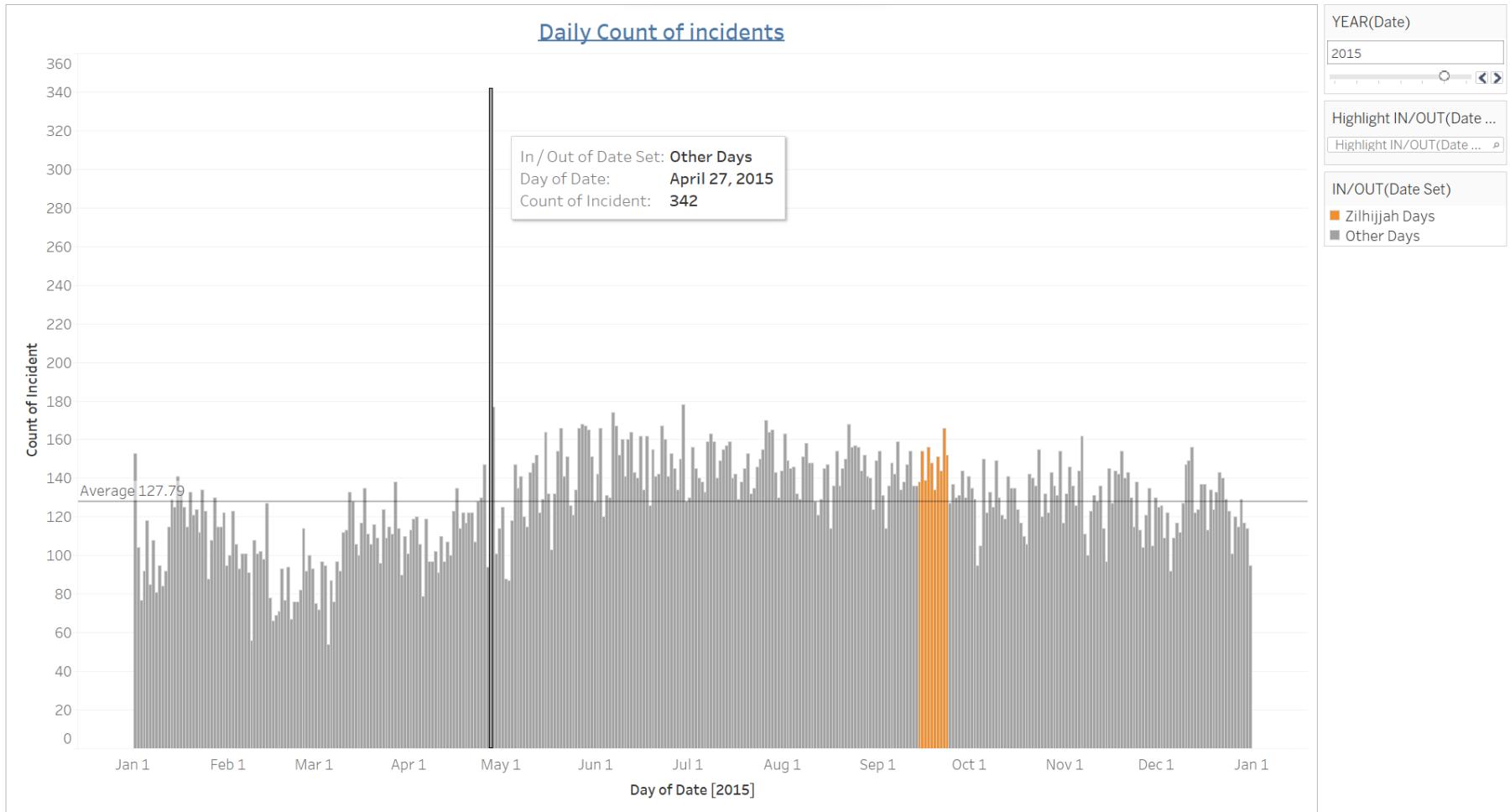
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]



Caption

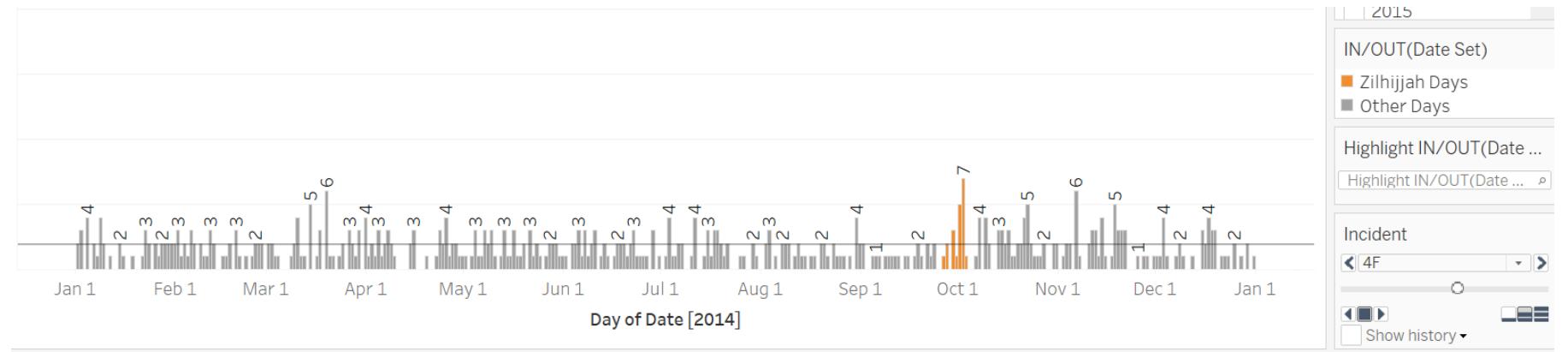
In all years, more crimes were committed during the Zilhijjah periods.

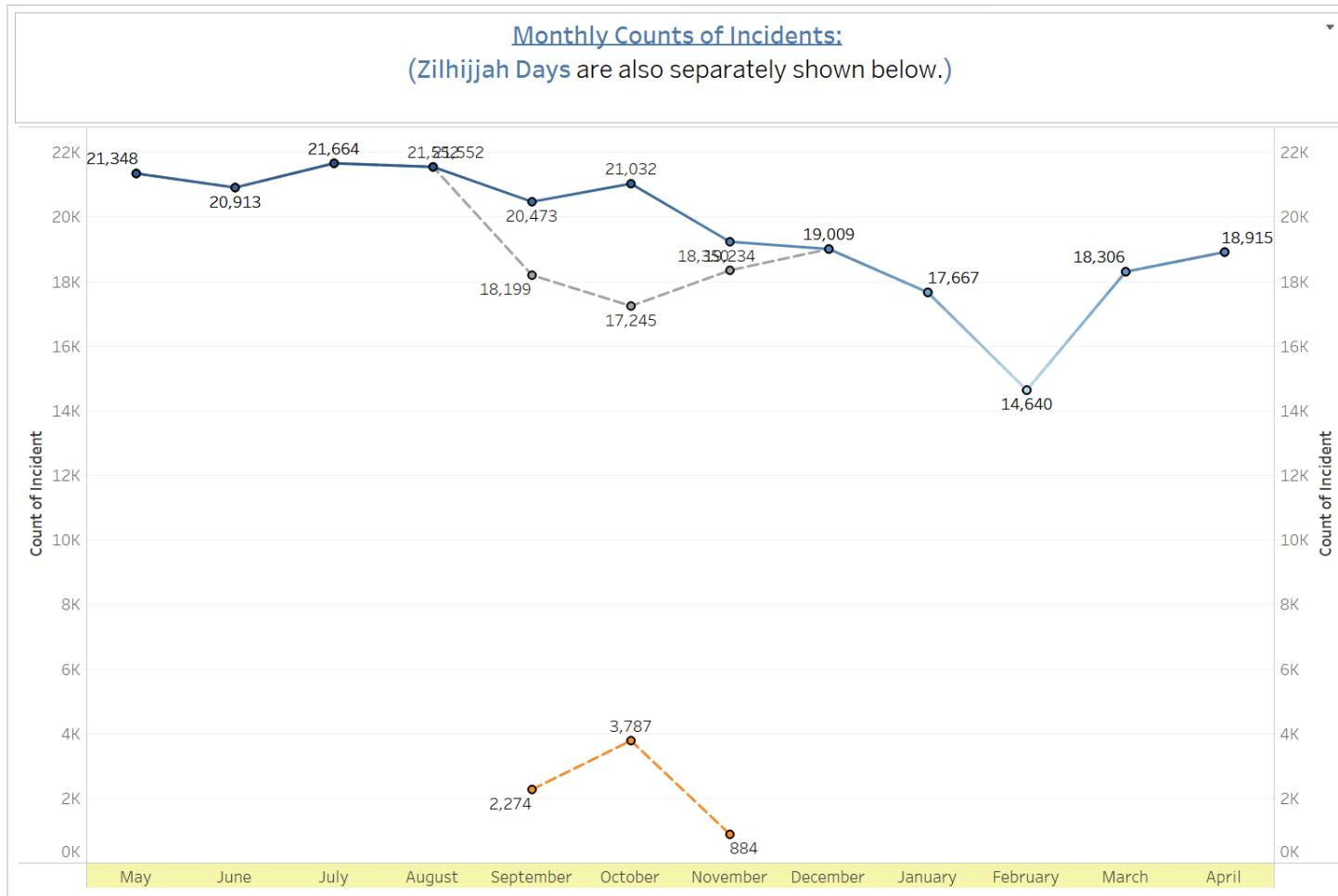




In [188...]

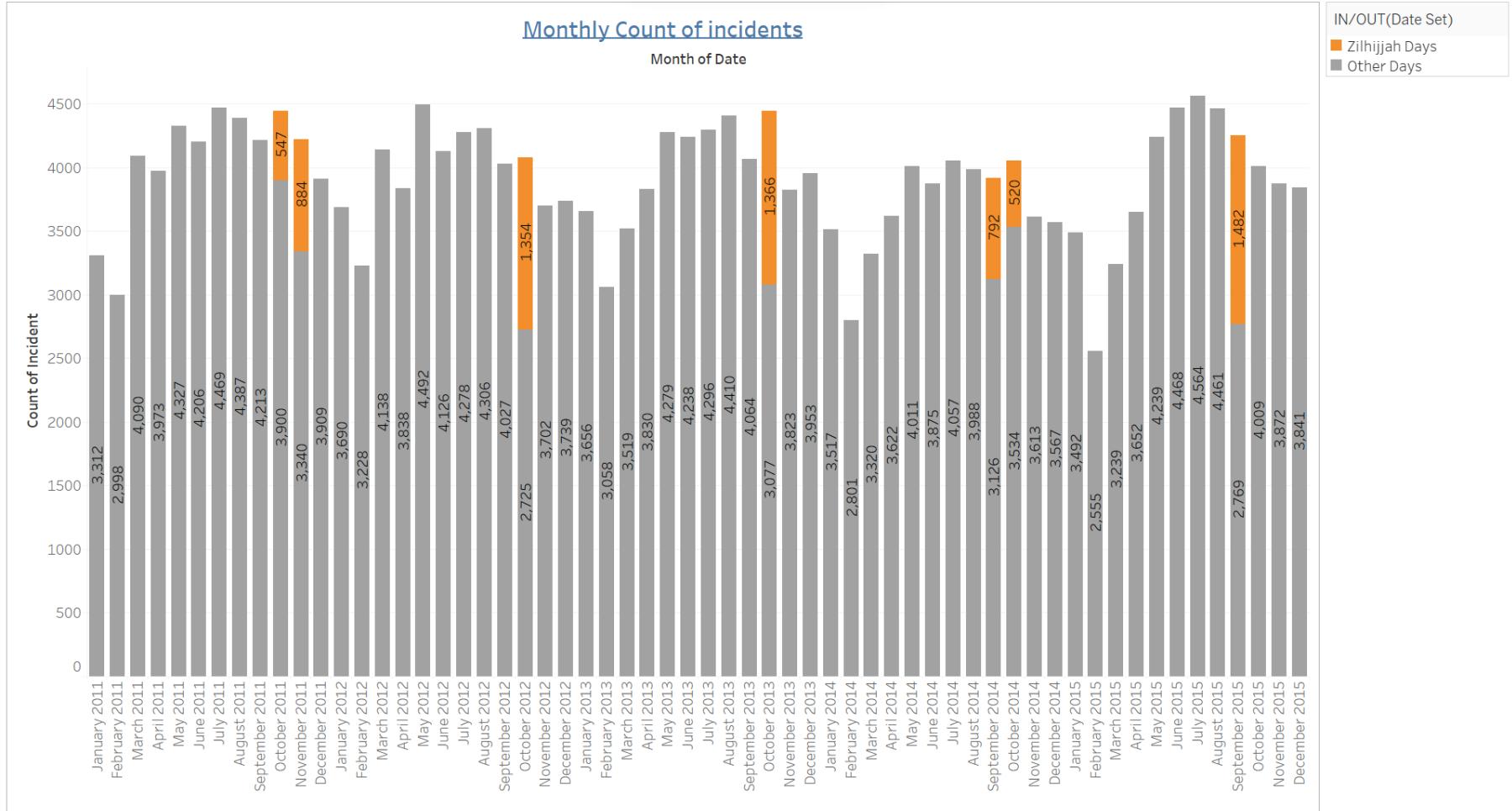
Daily Types: 4F





Caption

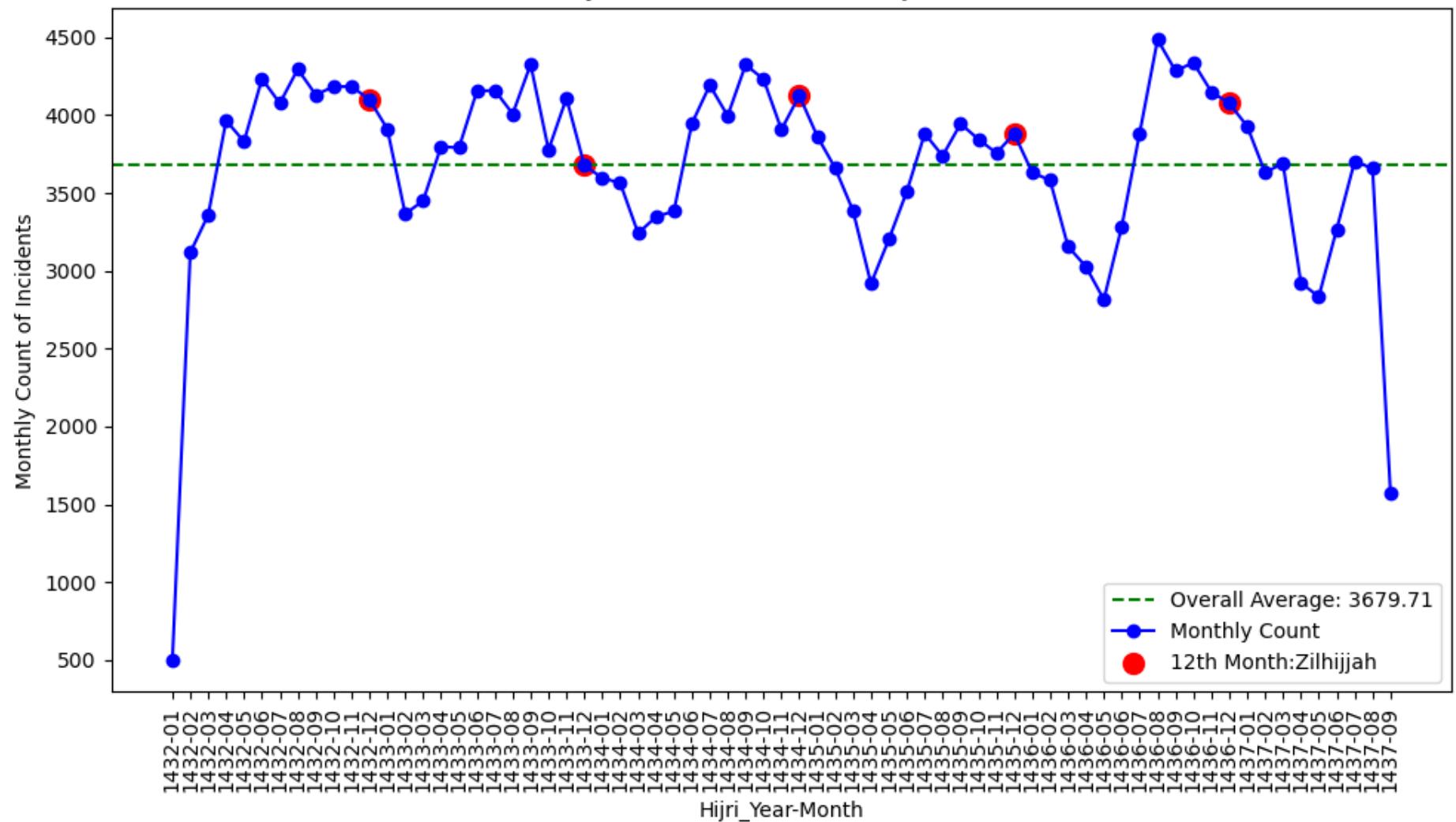
Contrary to the general decreasing trend, there is an increase from September to November, coinciding with the first ten days of Zilhijjah that surpass the annual average.

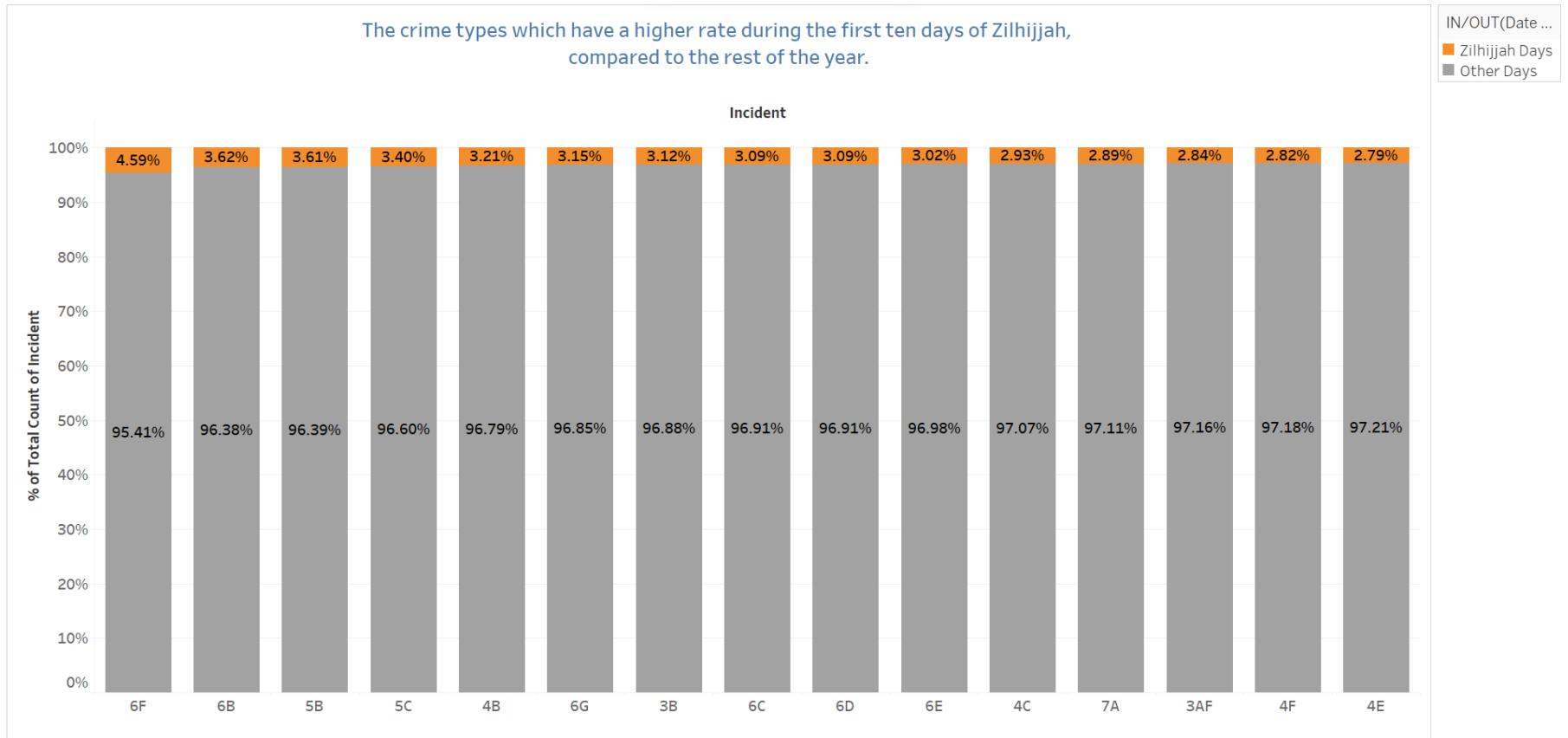


In [189]:

monthly_count_plot()

Monthly Count of Incidents on Hijri Calendar





SAMPLE DATA-9: ATLANTA CRIME DATASET_2009-2017

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

In [190]:

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

Out[190]:

		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 [191...]: df.duplicated().value_counts()

Out[191]: False 270688
dtype: int64

In [192...]: 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 [193]: df["crime"].value_counts()
```

```
Out[193]:
```

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

```
In [195]: df['date'] = pd.to_datetime(df['date']).dt.strftime('%Y-%m-%d')
```

```
In [196]: df.date.min(), df.date.max()
```

```
Out[196]: ('2009-01-01', '2017-02-28')
```

```
# The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2009-11-18' - '2009-11-27' '2010-11-07' - '2010-11-16' '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23' '2016-09-02' - '2016-09-11'
```

```
In [197... df = df.iloc[:, [0,2]]
```

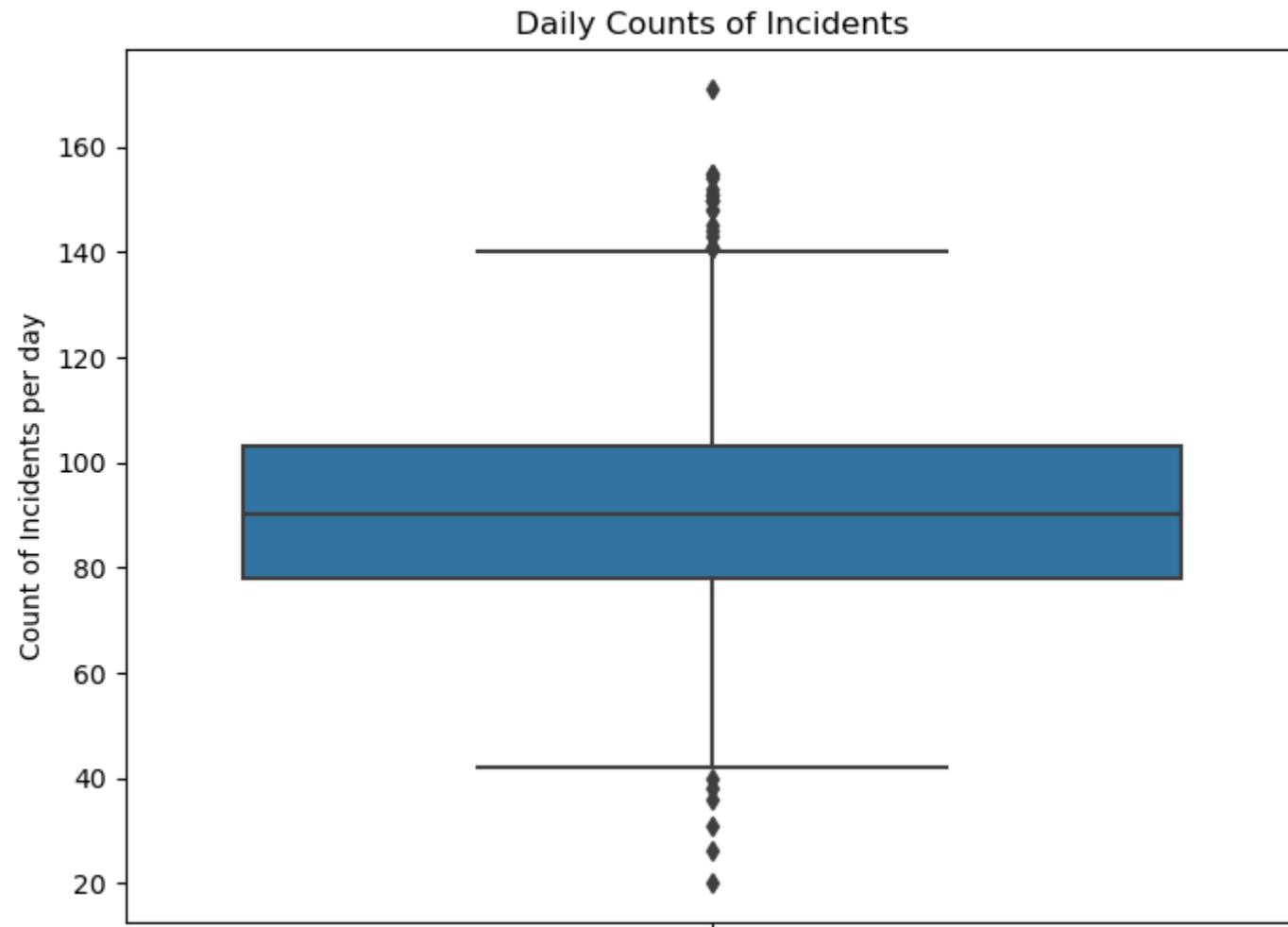
```
# df.to_csv("Atlanta.csv", index=False)
```

```
In [198... daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)
```

```
daily_incident_counts_stats
```

```
Out[198]: 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 [199... # 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 [200]: df.date.unique()
```

```
Out[200]: 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 [201]: zilhijjah_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 Zilhijjah months: 92.625
-----
Average case count in other days: 90.7542
-----
Ratio of Zilhijjah cases to other cases: 1.0206
```

We observe a 2.06% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

We can deepen our analysis by focusing on crime categories rather than solely comparing the number of crimes committed. In the study conducted on the Atlanta dataset using Tableau Desktop, it stands out that the category of rape generally experiences a higher incidence of crimes during the Zilhijjah period. This

observation might have been derived using Tableau Desktop's visualization and analysis tool, utilizing the in/out functionality of sets to analyze interactions within the Atlanta dataset.

```
In [202...]: # Crime Rates in Atlanta: Zilhijjah Days vs Other Days
# Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]
# Notice the Rape crimes in Zilhijjah
```

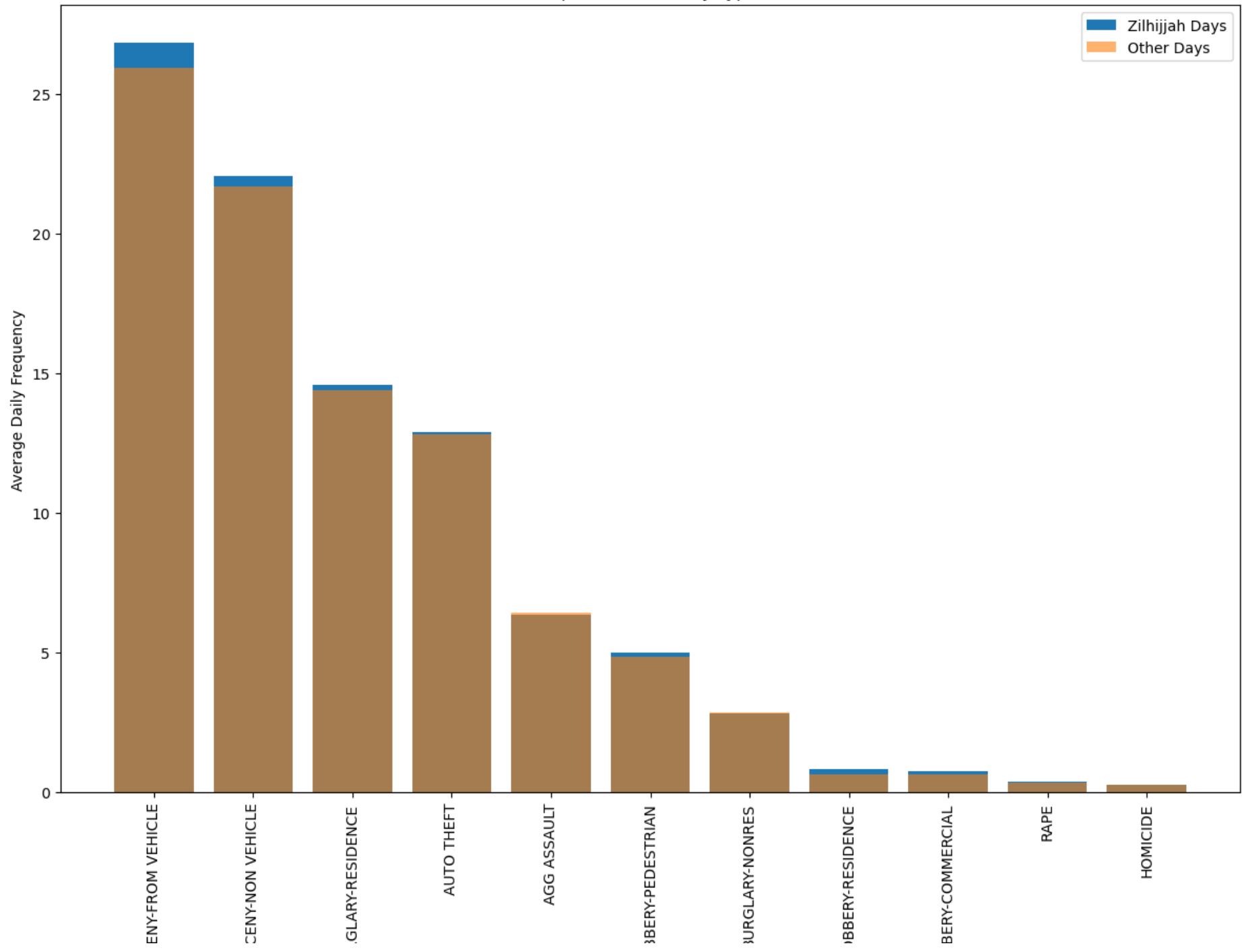
The crime rates during the first ten days of Zilhijjah, depicted by the blue color code, are higher compared to the rest of the year.

Crime	In/Out of Da...	Grand Total	Date							
			2009	2010	2011	2012	2013	2014	2015	2016
AGG ASSAULT	Other Days	97.31%	97.53%	97.18%	97.50%	97.68%	97.94%	96.80%	97.16%	96.54%
	Zilhijjah Days	2.69%	2.47%	2.82%	2.50%	2.32%	2.06%	3.20%	2.84%	3.46%
AUTO THEFT	Other Days	97.26%	97.35%	96.70%	96.97%	97.65%	97.80%	97.24%	97.52%	96.89%
	Zilhijjah Days	2.74%	2.65%	3.30%	3.03%	2.35%	2.20%	2.76%	2.48%	3.11%
BURGLARY-NONRES	Other Days	97.32%	97.20%	97.04%	98.48%	97.68%	98.08%	97.81%	95.23%	97.00%
	Zilhijjah Days	2.68%	2.80%	2.96%	1.52%	2.32%	1.92%	2.19%	4.77%	3.00%
BURGLARY-RESIDENCE	Other Days	97.26%	97.25%	97.41%	97.41%	96.81%	97.19%	96.76%	97.58%	97.77%
	Zilhijjah Days	2.74%	2.75%	2.59%	2.59%	3.19%	2.81%	3.24%	2.42%	2.23%
HOMICIDE	Other Days	97.34%	100.00%	98.86%	95.40%	97.59%	96.39%	97.85%	96.74%	96.26%
	Zilhijjah Days	2.66%		1.14%	4.60%	2.41%	3.61%	2.15%	3.26%	3.74%
LARCENY-FROM VEHICLE	Other Days	97.17%	97.27%	96.78%	97.36%	97.78%	97.09%	97.43%	96.95%	96.79%
	Zilhijjah Days	2.83%	2.73%	3.22%	2.64%	2.22%	2.91%	2.57%	3.05%	3.21%
LARCENY-NON VEHICLE	Other Days	97.23%	97.56%	97.15%	97.13%	96.97%	97.03%	97.53%	97.30%	97.23%
	Zilhijjah Days	2.77%	2.44%	2.85%	2.87%	3.03%	2.97%	2.47%	2.70%	2.77%
RAPE	Other Days	96.87%	99.00%	97.56%	96.95%	97.06%	97.03%	97.92%	95.48%	95.07%
	Zilhijjah Days	3.13%	1.00%	2.44%	3.05%	2.94%	2.97%	2.08%	4.52%	4.93%
ROBBERY-COMMERCIAL	Other Days	96.77%	95.04%	97.10%	98.60%	92.18%	97.54%	98.64%	97.02%	97.51%
	Zilhijjah Days	3.23%	4.96%	2.90%	1.40%	7.82%	2.46%	1.36%	2.98%	2.49%
ROBBERY-PEDESTRIAN	Other Days	97.21%	97.33%	97.71%	97.39%	97.45%	97.48%	96.87%	96.69%	96.67%
	Zilhijjah Days	2.79%	2.67%	2.29%	2.61%	2.55%	2.52%	3.13%	3.31%	3.33%
ROBBERY-RESIDENCE	Other Days	96.50%	97.41%	97.41%	95.91%	96.71%	95.02%	96.70%	96.79%	95.61%
	Zilhijjah Days	3.50%	2.59%	2.59%	4.09%	3.29%	4.98%	3.30%	3.21%	4.39%

In comparison to other days of the year, the ratio of RAPE crimes committed during the first ten days of the month of Zilhijjah was 11.38% higher. (3.13/2.81)

```
In [203...]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type



LARC

LAR

BUR

ROF

E

RC

ROB

Incident Types

In [204]: # Top 30 incident types sorted by "zilhijjah incidents / total incidents" ratio
sorted_ratios

Out[204]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
ROBBERY-RESIDENCE	65	1880	0.034600
ROBBERY-COMMERCIAL	59	1855	0.031800
RAPE	30	990	0.030300
LARCENY-FROM VEHICLE	2146	77345	0.027700
ROBBERY-PEDESTRIAN	398	14446	0.027600
LARCENY-NON VEHICLE	1765	64697	0.027300
BURGLARY-RESIDENCE	1166	42941	0.027200
AUTO THEFT	1030	38168	0.027000
AGG ASSAULT	508	19133	0.026600
BURGLARY-NONRES	224	8505	0.026300
HOMICIDE	19	728	0.026100

In [205]: # In which categories were more crimes committed during the first ten days of Zilhijjah?
zilhijjah_dominant_incidents

Out[205]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
LARCENY-FROM VEHICLE	2146	77345	0.0277
LARCENY-NON VEHICLE	1765	64697	0.0273
BURGLARY-RESIDENCE	1166	42941	0.0272
AUTO THEFT	1030	38168	0.0270
ROBBERY-PEDESTRIAN	398	14446	0.0276
ROBBERY-RESIDENCE	65	1880	0.0346
ROBBERY-COMMERCIAL	59	1855	0.0318
RAPE	30	990	0.0303

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

In [206...]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]

Out[206]: (11, 8)

In [207...]: zilhijjah_incidents_desc

Out[207]:

count	7410
unique	11
top	LARCENY-FROM VEHICLE
freq	2146
Name: incident, dtype: object	

In [208...]: other_days_incidents_desc

Out[208]:

count	263278
unique	11
top	LARCENY-FROM VEHICLE
freq	75199
Name: incident, dtype: object	

Atlanta dataset encompasses 11 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 11 incident categories, with 8 of these categories experiencing incident counts exceeding the annual averages.

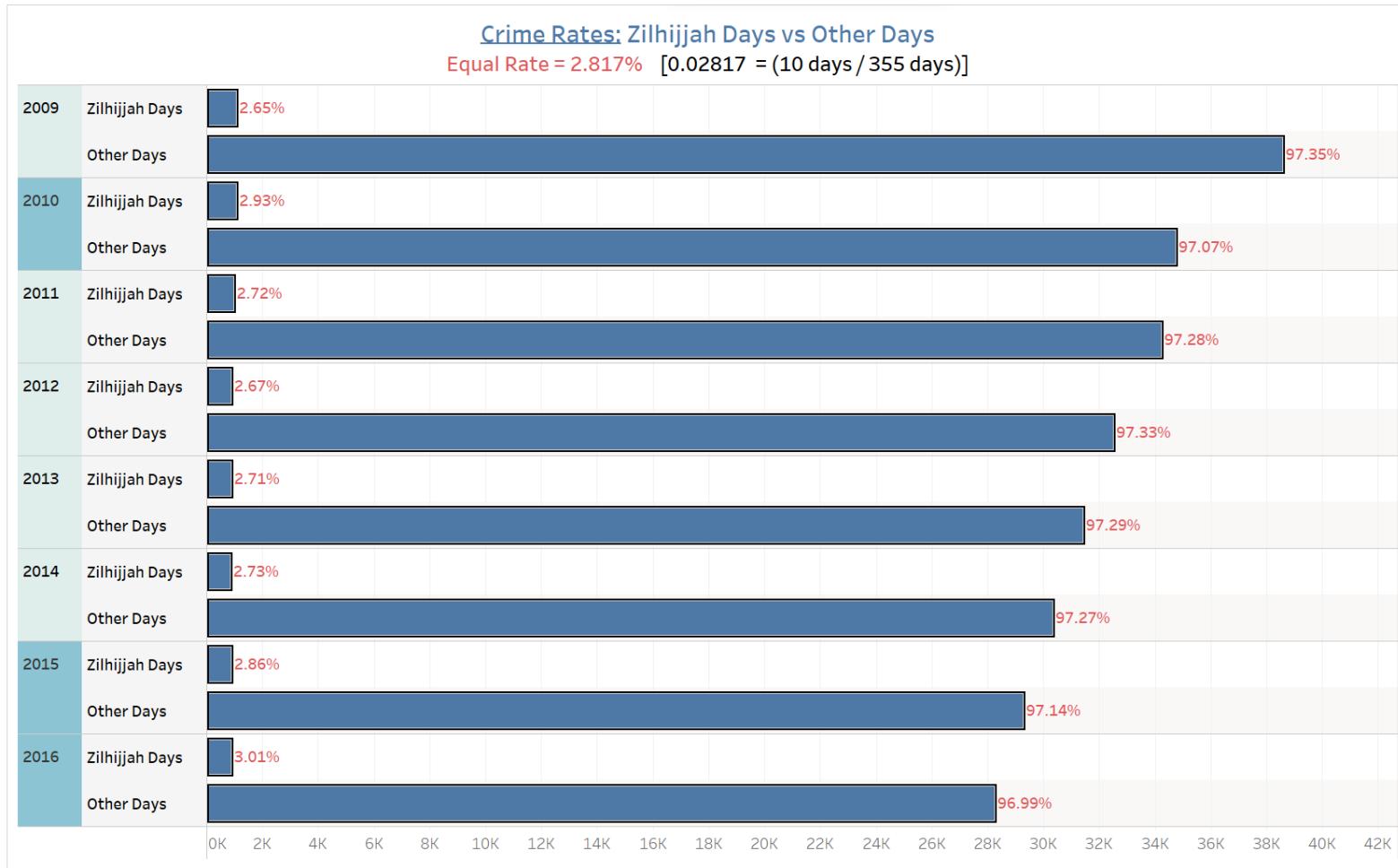
Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	7,410	1,050	1,048	955	891	877	850	862	877
Other Days	259,271	38,576	34,722	34,219	32,503	31,426	30,316	29,255	28,254

Crime Rates: Zilhijjah (10) Days vs Other (355) Days

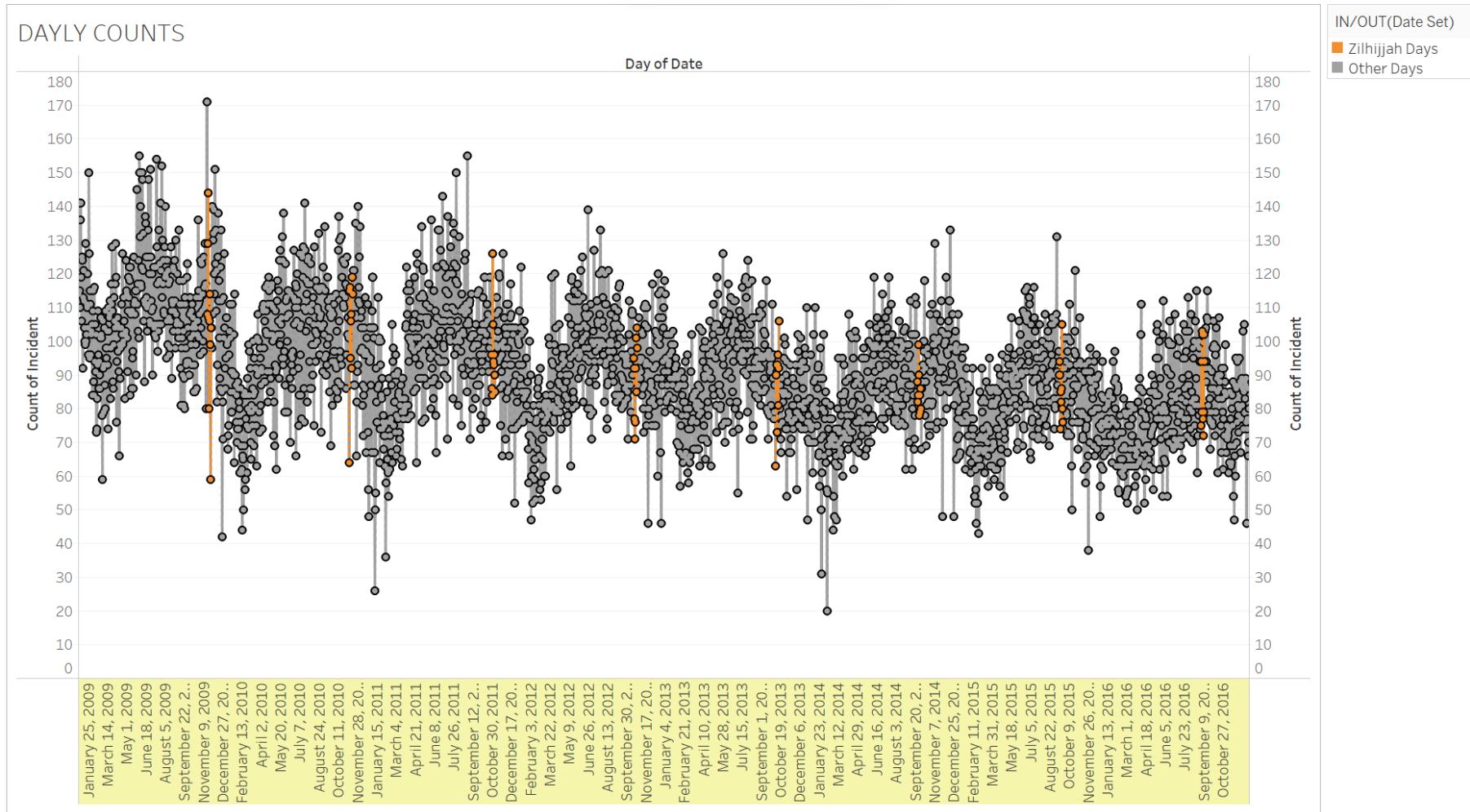
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

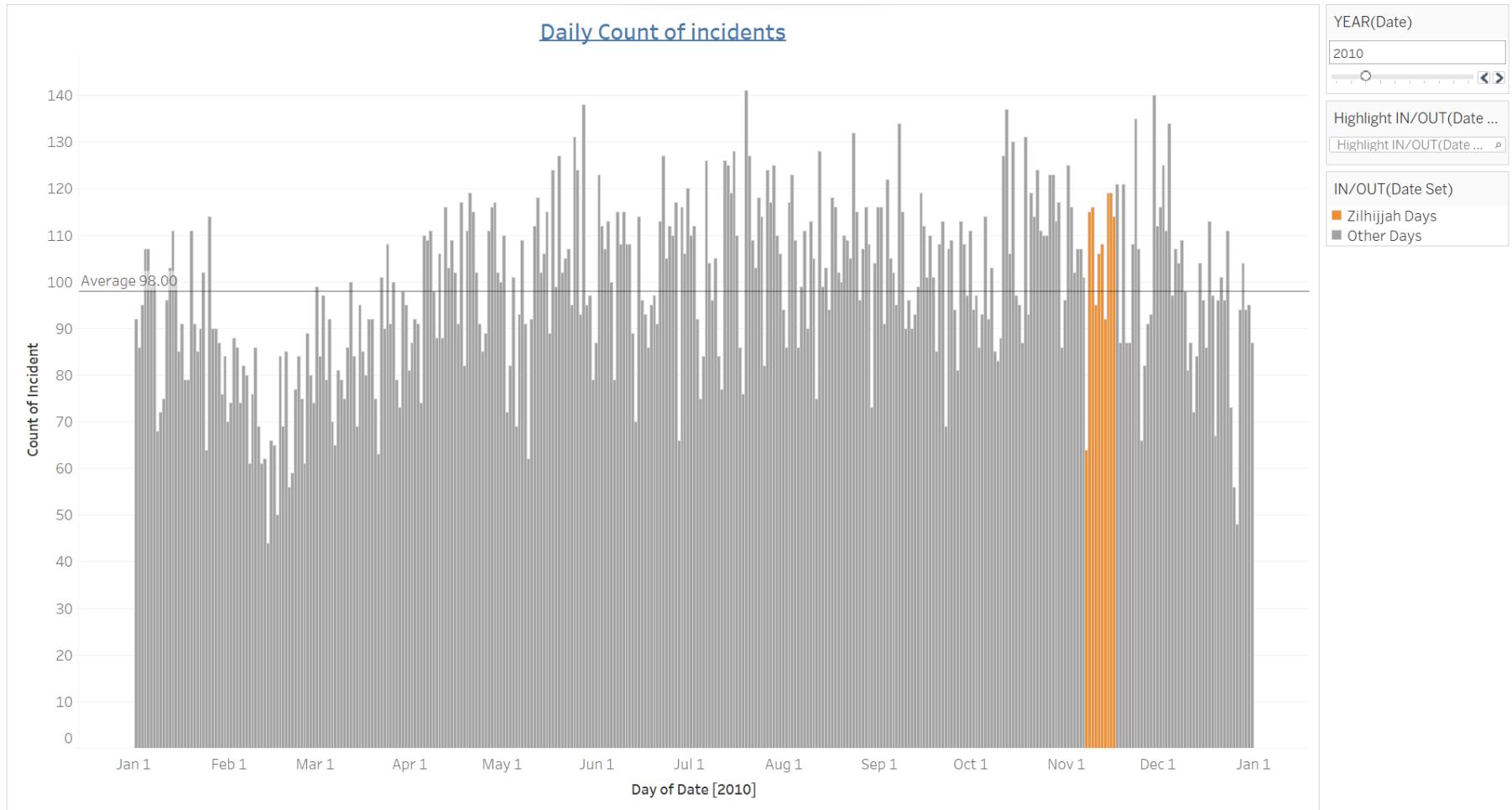
	Total Incidents	2009	2010	2011	2012	2013	2014	2015	2016
Zilhijjah Days	2.78%	2.65%	2.93%	2.72%	2.67%	2.71%	2.73%	2.86%	3.01%
Other Days	97.22%	97.35%	97.07%	97.28%	97.33%	97.29%	97.27%	97.14%	96.99%



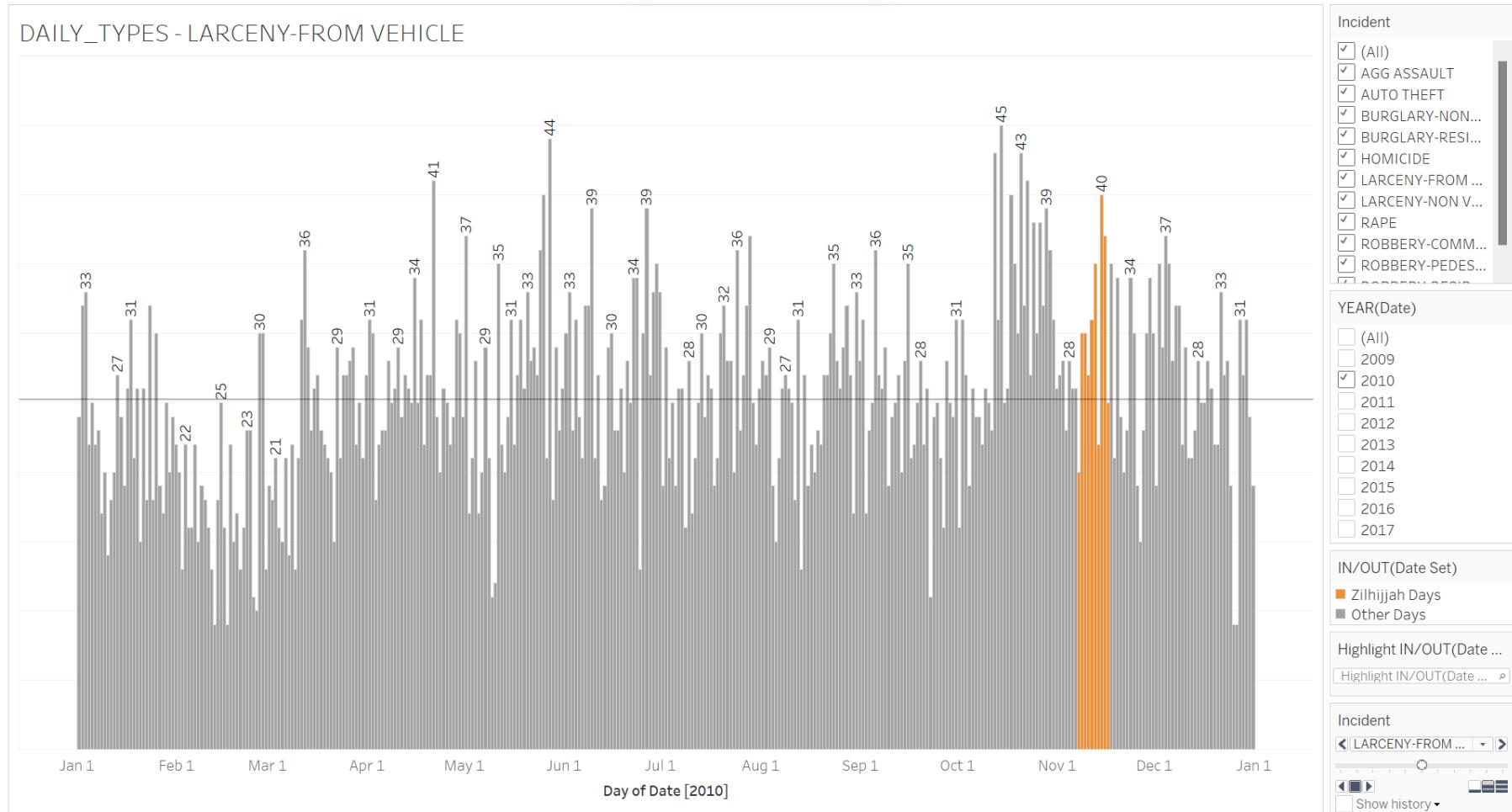
Caption

In comparison to other days of the year, the number of crimes committed during the first ten days of the month of Zilhijjah was higher in 2010, 2015, and 2016. Conversely, in other years, fewer crimes were committed during Zilhijjah.



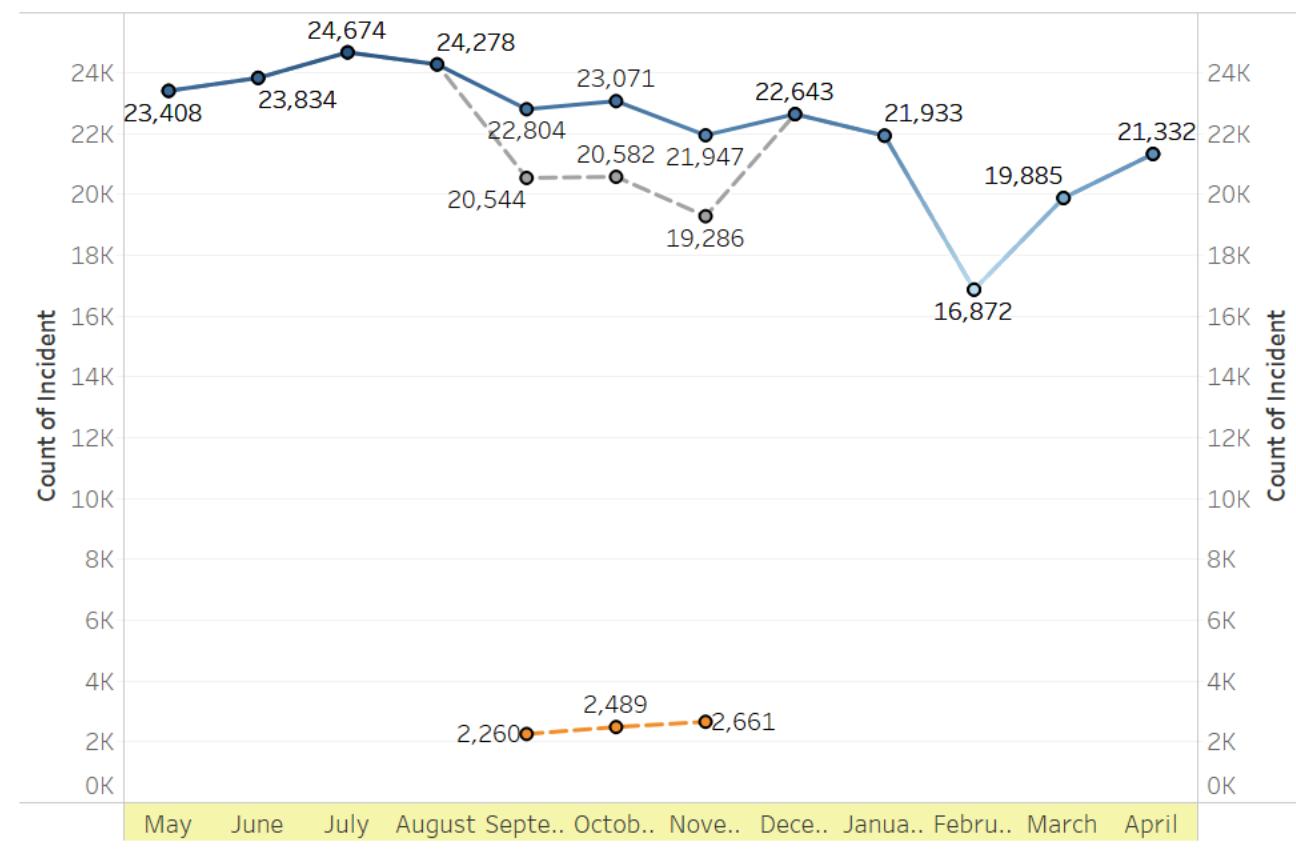


DAILY_TYPES - LARCENY-FROM VEHICLE



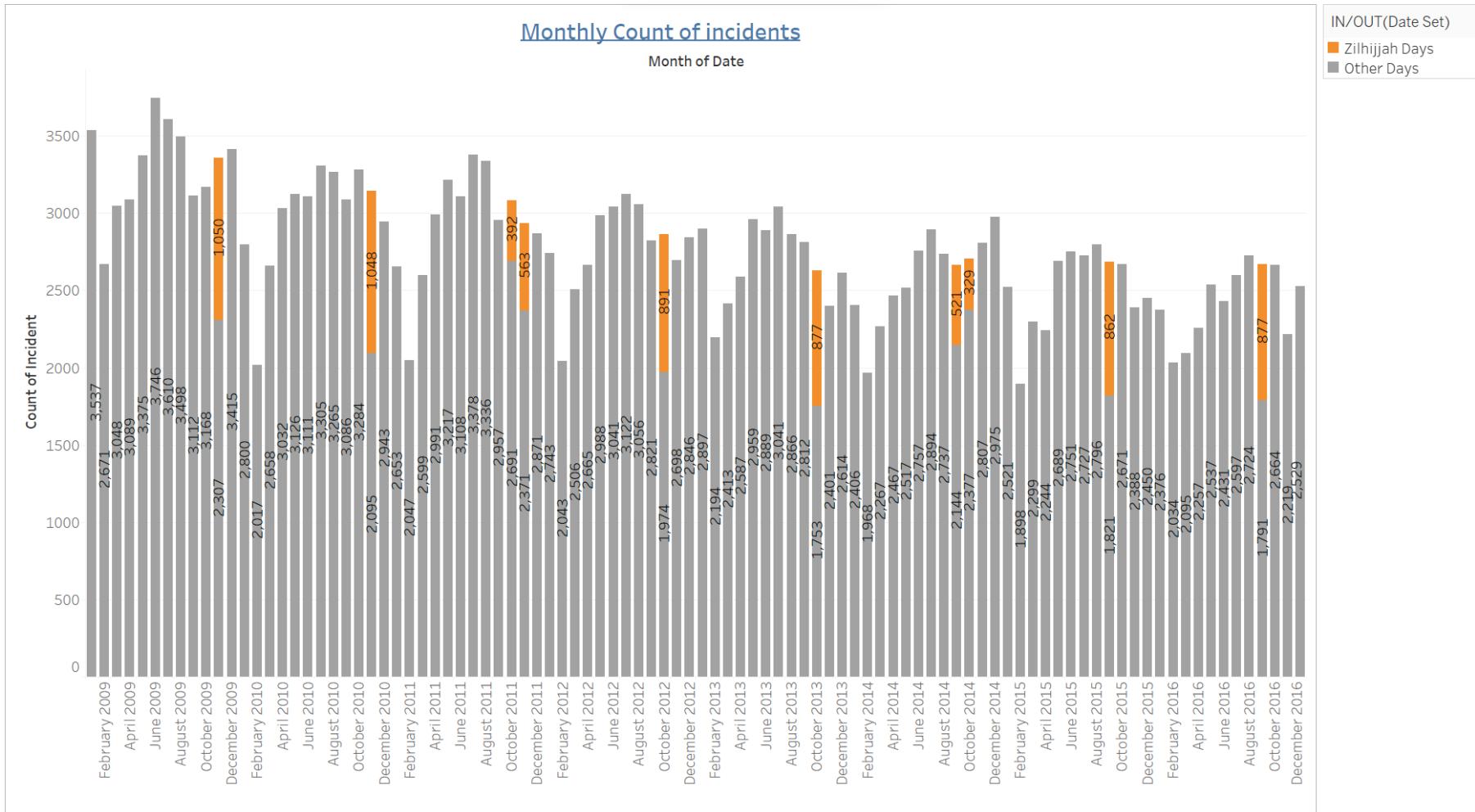
Monthly Counts of Incidents:

(Zilhijjah Days are also separately shown below.)

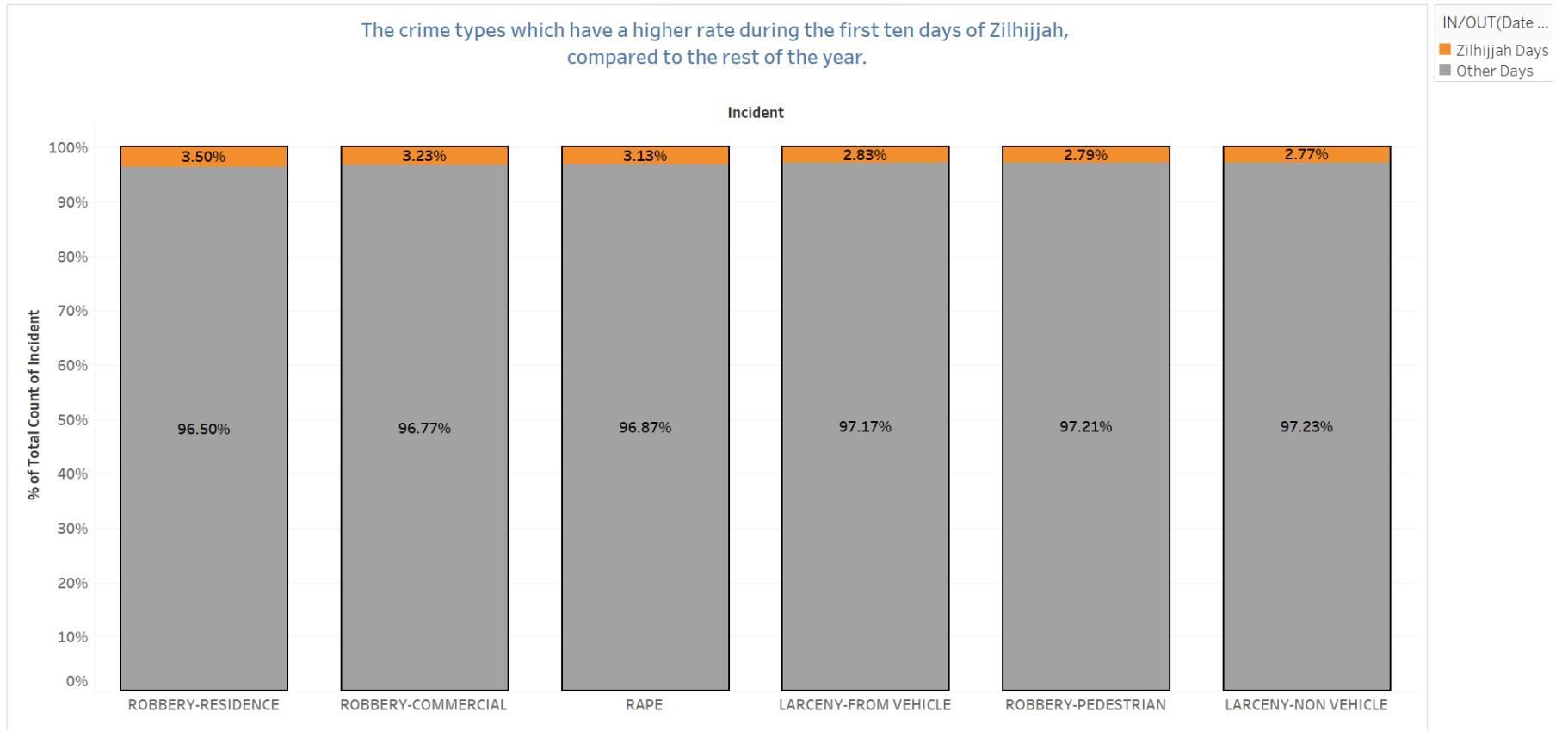


Caption

Contrary to the general decreasing trend, there is an increase from September to November, coinciding with the first ten days of Zilhijjah that surpass the annual average.



IN/OUT(Date Set)
Zilhijjah Days
Other Days



SAMPLE DATA-10: OAKLAND CRIME STATISTICS_2011-2016

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

In [209...]

```
df2 = pd.read_csv("records-for-2011.csv")
df3 = pd.read_csv("records-for-2012.csv")
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 [210...]

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

```
df.head()
```

Out[210]:

	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 [211...]

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

Out[211]:

```
False    1046388  
dtype: int64
```

In [212...]

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

```
Out[213]:
```

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 [214...]
```

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

```
In [215...]
```

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

```
In [216...]
```

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

```
In [217...]
```

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

```
Out[217]: ('2011-01-01', '2016-07-31')
```

```
# The first ten days of the month of Zilhijjah in the Gregorian calendar years are: '2011-10-28' - '2011-11-06' '2012-10-17' - '2012-10-26' '2013-10-06' - '2013-10-15' '2014-09-25' - '2014-10-04' '2015-09-14' - '2015-09-23'
```

```
In [218... df = df.iloc[:, [1,7]]  
# df.to_csv("Oakland.csv", index=False)
```

```
In [219... daily_incident_counts_stats = df.groupby("date")['date'].value_counts().describe([.25, .5, .75, .95, .98, .99]).astype(int)  
daily_incident_counts_stats
```

```
Out[219]: 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 [220... # 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 [221]: df.date.nunique()
```

```
Out[221]: 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.

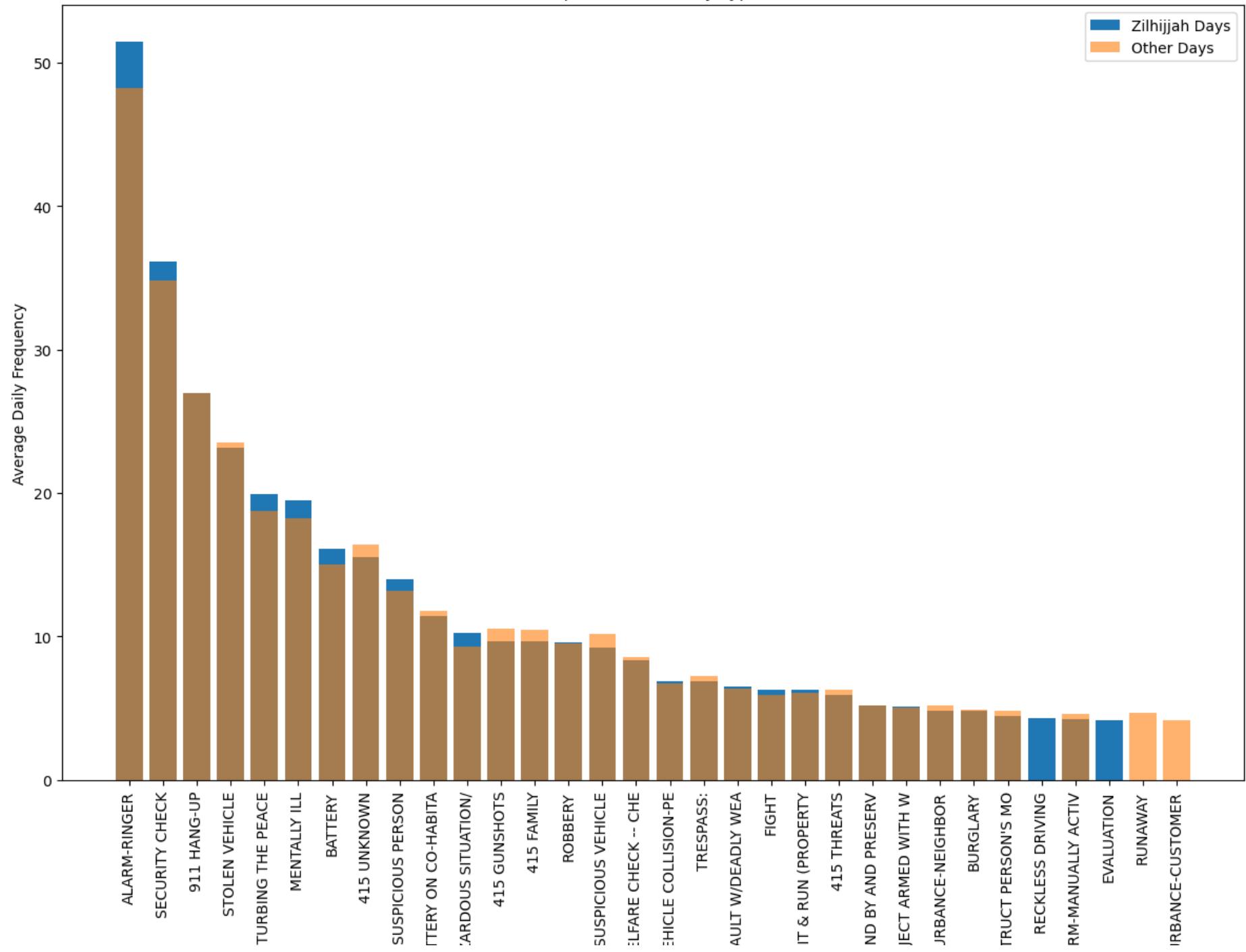
```
In [222]: zilhijjah_10_days(df)
```

```
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 Zilhijjah months: 521.06
-----
Average case count in other days: 512.9869
-----
Ratio of Zilhijjah cases to other cases: 1.0157
```

We observe a 1.57% higher crime rate during the initial 10 days of the Zilhijjah month compared to the annual average.

```
In [223...]: sorted_ratios, zilhijjah_dominant_incidents, zilhijjah_incidents_desc, other_days_incidents_desc = incidents_by_types(df)
# display(sorted_ratios)
# display(zilhijjah_dominant_incidents)
```

Top 30 Incidents by Type



In [224...]

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



Out[224]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
CIVIL	87	487	0.178600
IDENTITY THEFT	1	7	0.142900
ESCAPEE	1	8	0.125000
LOST PROPERTY	1	10	0.100000
OFFICER NEEDS IMMEDI	2	21	0.095200
CHOP SHOP OWNERSHIP/	2	27	0.074100
KIDNAPPING FOR RANSO	2	30	0.066700
TOW REQUESTED	4	66	0.060600
TILL TAP	5	84	0.059500
AUTO IN RESTRICED ZO	3	53	0.056600
CHILD TAKEN INTO PRO	1	19	0.052600
FRAUDULENT USE OF AC	2	39	0.051300
AGGRAVATED ASSAULT	5	98	0.051000
SCOOTERS INVOLVED IN	3	62	0.048400
WIRES	24	525	0.045700
UNLAWFUL DETAINER	3	69	0.043500
AUTO BLOCKING DRIVEW	13	300	0.043300
ATTEMPTED RAPE-SEXUA	5	116	0.043100
THROWING SUBSTANCE A	13	312	0.041700
MISSING JUVENILE	63	1535	0.041000
TARASOFF	4	98	0.040800
POSSESSION OF CONCEN	5	123	0.040700
DEFRAUDING AN INNKEE	17	419	0.040600
EASTBAY MUD	1	26	0.038500

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
TRUANT	10	266	0.037600
MEET WITH THE CITIZE	2	54	0.037000
SMOKE	26	711	0.036600
WILLFUL CRUELTY TO C	18	500	0.036000
SUSPECTS	1	28	0.035700
INFILCT INJURY ON CH	61	1734	0.035200

In [225]:

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

Out[225]:

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
ALARM-RINGER	2570	98497	0.0261
SECURITY CHECK	1806	70965	0.0254
911 HANG-UP	1349	54935	0.0246
DISTURBING THE PEACE	995	38257	0.0260
MENTALLY ILL	975	37218	0.0262
BATTERY	806	30636	0.0263
SUSPICIOUS PERSON	700	26984	0.0259
HAZARDOUS SITUATION/	514	18948	0.0271
ROBBERY	480	19452	0.0247
VEHICLE COLLISION-PE	346	13782	0.0251
ASSAULT W/DEADLY WEA	326	12982	0.0251
HIT & RUN (PROPERTY)	315	12417	0.0254
FIGHT	315	12062	0.0261
STAND BY AND PRESERV	261	10584	0.0247
SUBJECT ARMED WITH W	257	10376	0.0248
RECKLESS DRIVING	215	7945	0.0271
EVALUATION	210	8535	0.0246
AUTO BURGLARY--IP	202	7791	0.0259
415 BOYFRIEND	193	7623	0.0253
DISTURBANCE-DRINKERS	190	7091	0.0268
INJURED ANIMAL	184	6494	0.0283
CRUELTY TO ANIMAL	182	6811	0.0267
VEHICLE COLLISION/PR	178	7072	0.0252
SUSPECT	169	6575	0.0257

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
FIRE	160	5734	0.0279
PERSON SCREAMING	157	5233	0.0300
DISTURBANCE-LANDLORD	132	4318	0.0306
VEHICLE COLLISION-UN	130	5100	0.0255
MISSING PERSON	123	4881	0.0252
INSANE PERSON-VIOLEN	108	3519	0.0307
MISCELLANEOUS/UNKNOWN	100	3390	0.0295
HOMELESS ENCAMPMENT	97	3197	0.0303
SLEEPER	96	3652	0.0263
AMBULANCE FOLLOW UP	91	3230	0.0282
PROSTITUTION/DISORDE	90	3588	0.0251
CIVIL	87	487	0.1786
DISORDERLY CONDUCT	78	2824	0.0276
DISTURBANCE-JUVENILE	76	3050	0.0249
THEFT	74	2688	0.0275
COVER THE OFFICER	72	2579	0.0279
DISTURBING BETWEEN R	70	2439	0.0287
VANDALISM	69	2538	0.0272
INDECENT EXPOSURE	67	2559	0.0262
MISSING JUVENILE	63	1535	0.0410
INFILCT INJURY ON CH	61	1734	0.0352
SEXUAL ASSAULT; RAPE	57	2144	0.0266
CHILD ENDANGERMENT O	56	2075	0.0270
CARJACKING	53	1723	0.0308

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
DISORDERLY CONDUCT:	49	1939	0.0253
ALARM-SILENT	48	1940	0.0247
GRAND THEFT	47	1801	0.0261
OVERDOSE	44	1418	0.0310
PICK UP	43	1687	0.0255
ANIMAL BITE	43	1607	0.0268
ATTEMPT SUICIDE	42	1563	0.0269
MISSING PERSON AT RI	42	1431	0.0294
VEH PARKED IN SIGNED	40	1437	0.0278
SIGNALS-MALFUNCTIONI	35	1324	0.0264
POSSESS NARCOTIC CON	34	1131	0.0301
STALLED VEH	33	1156	0.0285
AUTO ON FIRE	29	832	0.0349
VEHICLE COLLISION-DR	29	1084	0.0268
AUTO ALARM	28	1010	0.0277
ROBBERY: FIRST DEGRE	27	954	0.0283
SMOKE	26	711	0.0366
FORCED ENTRY IN PROG	26	783	0.0332
CHECK VEHICLE	26	804	0.0323
GAMBLING	25	822	0.0304
WIRES	24	525	0.0457
PICKETERS/PROTESTERS	20	713	0.0281
SENILE PERSON	19	615	0.0309
WILLFUL CRUELTY TO C	18	500	0.0360

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
DEFRAUDING AN INNKEE	17	419	0.0406
FOUND GUN	16	622	0.0257
GAS LEAK	16	507	0.0316
WATER	15	484	0.0310
ELECTRONIC DEVICE	14	411	0.0341
FOUND SENILE	13	409	0.0318
THROWING SUBSTANCE A	13	312	0.0417
AUTO BLOCKING DRIVEW	13	300	0.0433
DRUNK INSIDE BUILDIN	12	480	0.0250
CHILD STEALING	12	428	0.0280
INFILCT INJURY ON SP	12	423	0.0284
ABSENT WITHOUT LEAVE	12	419	0.0286
RECEIVE KNOWN STOLEN	11	377	0.0292
THEFT COIN OPERATED	10	301	0.0332
TRUANT	10	266	0.0376
SEXUAL BATTERY	9	286	0.0315
OAKLAND MUNICIPLE CO	9	314	0.0287
EVIDENCE	6	242	0.0248
DUMP WASTE OR OFFENS	6	207	0.0290
ATTEMPTED AUTO THEFT	6	222	0.0270
ATTEMPTED RAPE-SEXUA	5	116	0.0431
TILL TAP	5	84	0.0595
LOST PERSON	5	147	0.0340
AGGRAVATED ASSAULT	5	98	0.0510

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
POSSESSION OF CONCEN	5	123	0.0407
TOW REQUESTED	4	66	0.0606
TARASOFF	4	98	0.0408
ATTEMPTED RAPE	3	100	0.0300
UNLAWFUL DETAINER	3	69	0.0435
SCOOTERS INVOLVED IN	3	62	0.0484
AUTO IN RESTRICED ZO	3	53	0.0566
OFFICER NEEDS IMMEDI	2	21	0.0952
FRAUDULENT USE OF AC	2	39	0.0513
STOLEN LICENSE PLATE	2	68	0.0294
CONTEMPT OF COURT/DI	2	67	0.0299
CHOP SHOP OWNERSHIP/	2	27	0.0741
PASS FICTITIOUS CHEC	2	77	0.0260
PEEPING TOM	2	77	0.0260
EMERGENCY PROTECTIVE	2	61	0.0328
KIDNAPPING FOR RANSO	2	30	0.0667
CROSSING ARMS DOWN	2	58	0.0345
MEET WITH THE CITIZE	2	54	0.0370
ILLEGAL EXHIBITION O	2	80	0.0250
ESCAPEE	1	8	0.1250
LOST PROPERTY	1	10	0.1000
EASTBAY MUD	1	26	0.0385
SUSPECTS	1	28	0.0357
ATTEMPTED GRAND THEF	1	33	0.0303

	zilhijjah incidents	all incidents	zilhijjah incidents/total incidents
MAYHEM	1	33	0.0303
IDENTITY THEFT	1	7	0.1429
OAKLAND TRAFFIC CODE	1	32	0.0312
CHILD TAKEN INTO PRO	1	19	0.0526
EMBEZZLEMENT	1	36	0.0278

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

```
In [226]: df.incident.nunique(), zilhijjah_dominant_incidents.count()[0]
```

```
Out[226]: (288, 125)
```

```
In [227]: zilhijjah_incidents_desc
```

```
Out[227]:
```

count	26036
unique	217
top	ALARM-RINGER
freq	2570
Name: incident, dtype: object	

```
In [228]: other_days_incidents_desc
```

```
Out[228]:
```

count	1019960
unique	288
top	ALARM-RINGER
freq	95927
Name: incident, dtype: object	

Oakland dataset encompasses 288 distinct incident types. During the first ten days of Zilhijjah, crimes were committed across 217 incident categories, with 125 of these categories experiencing incident counts exceeding the annual averages.

Count of incidents: Zilhijjah (10) Days vs Other (355) Days

	Total Incidents	2011	2012	2013	2014	2015
Zilhijjah Days	26,053	4,938	5,213	4,986	5,286	5,630
Other Days	909,504	175,077	182,217	183,065	182,194	186,951

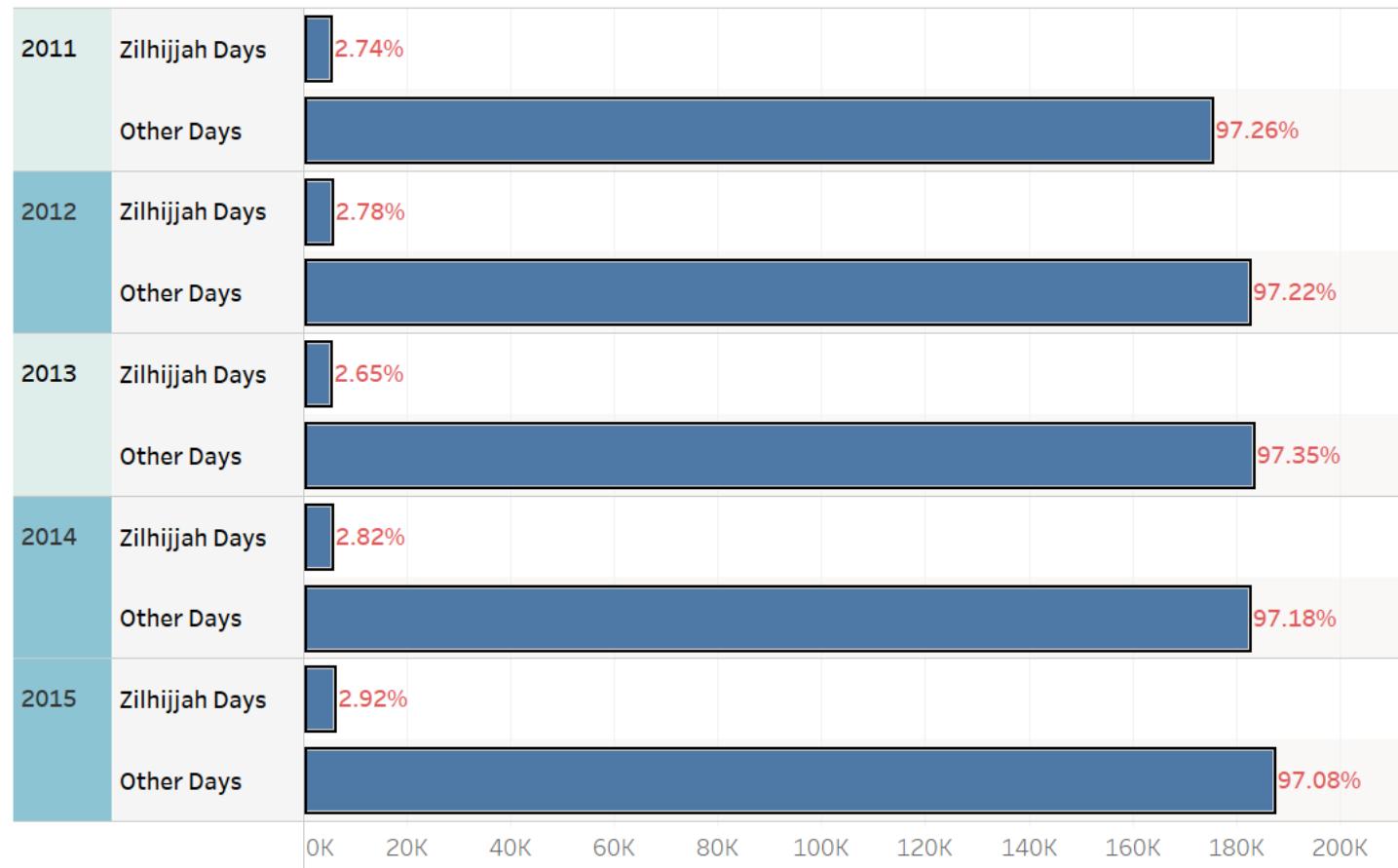
Crime Rates: Zilhijjah (10) Days vs Other (355) Days

Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

	Total Incidents	2011	2012	2013	2014	2015
Zilhijjah Days	2.78%	2.74%	2.78%	2.65%	2.82%	2.92%
Other Days	97.22%	97.26%	97.22%	97.35%	97.18%	97.08%

Crime Rates: Zilhijjah Days vs Other Days

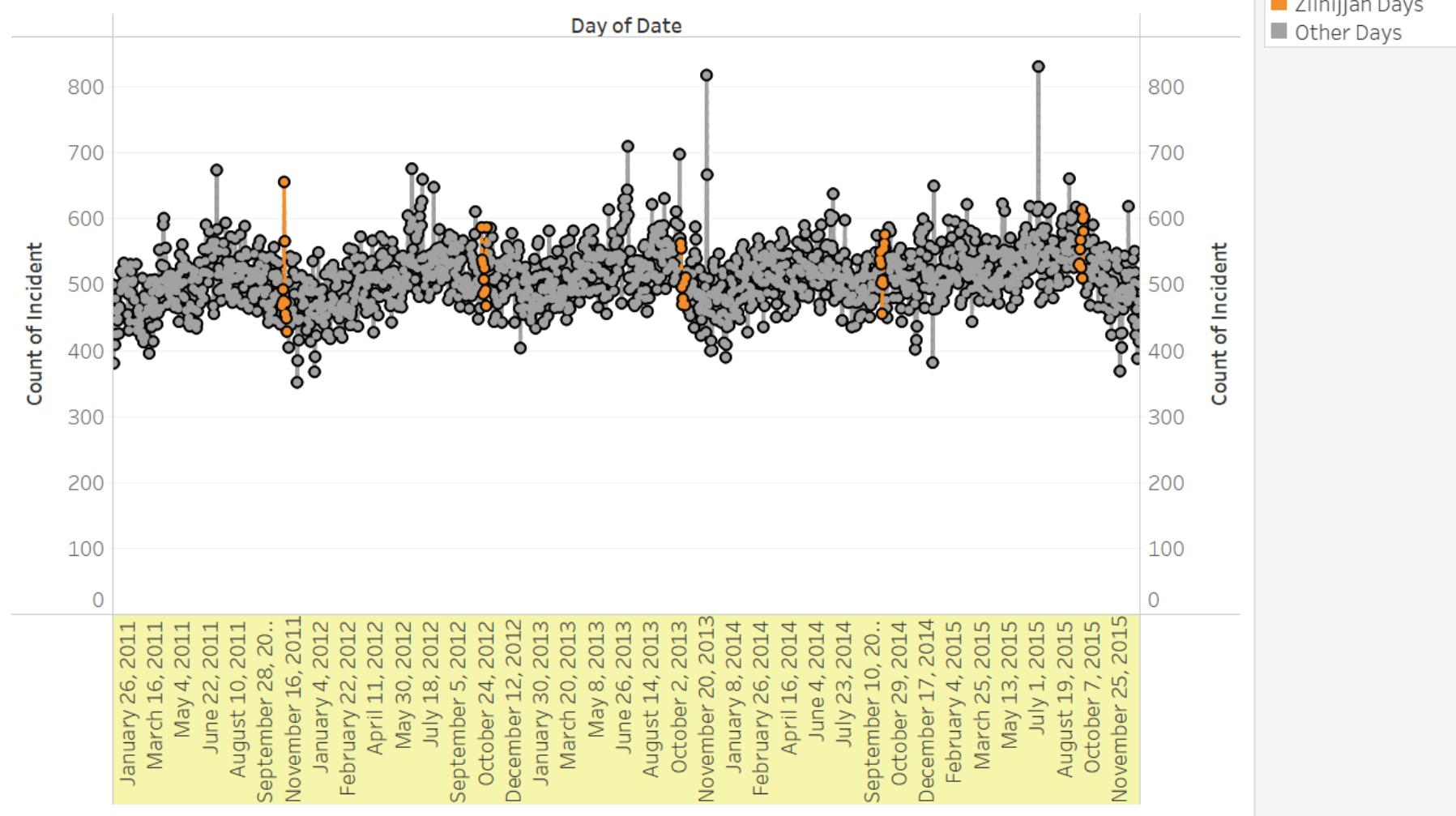
Equal Rate = 2.817% [0.02817 = (10 days / 355 days)]

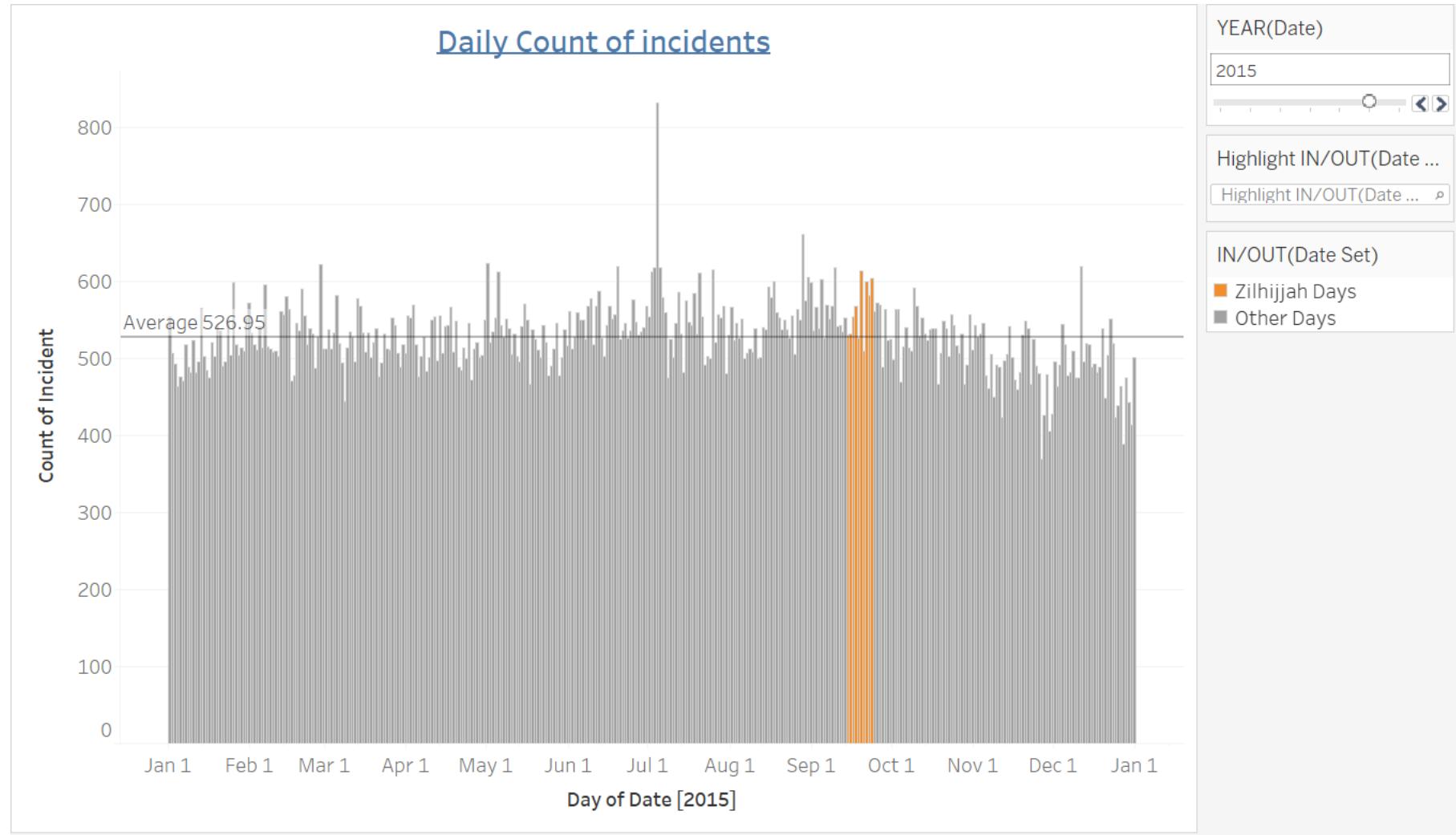


Caption

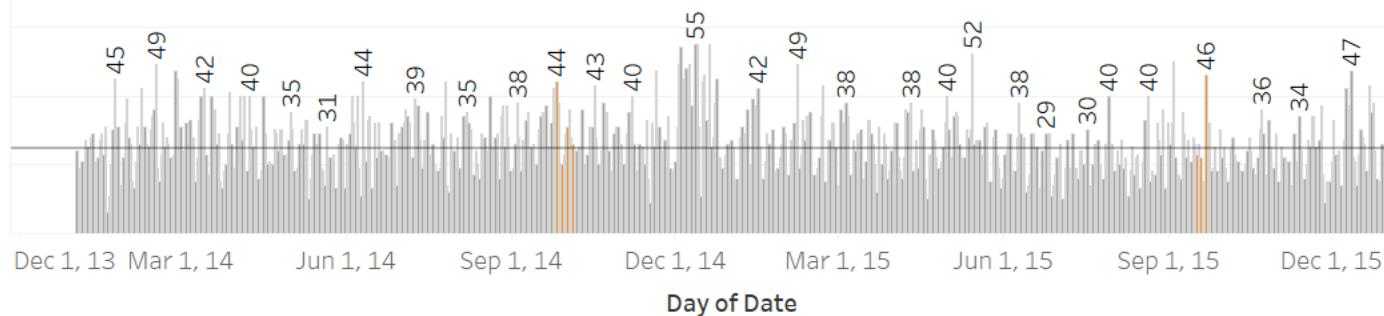
In comparison to other days of the year, the number of crimes committed during the first ten days of the month of Zilhijjah was higher in 2012, 2014, and 2015 and equal in 2011. Conversely, in 2013, fewer crimes were committed during Zilhijjah.

DAILY COUNTS





DAILY_TYPES - 911 HANG-UP



Incident

- WATER
- WELFARE CHECK ...
- WILLFUL CRUELTY...
- WIRES
- WITNESS
- YELLOW ALERT A...

YEAR(Date)

- (All)
- 2011
- 2012
- 2013
- 2014
- 2015

IN/OUT(Date Set)

- Zilhijjah Days
- Other Days

Highlight IN/OUT(Date ...)

Highlight IN/OUT(Date ...)

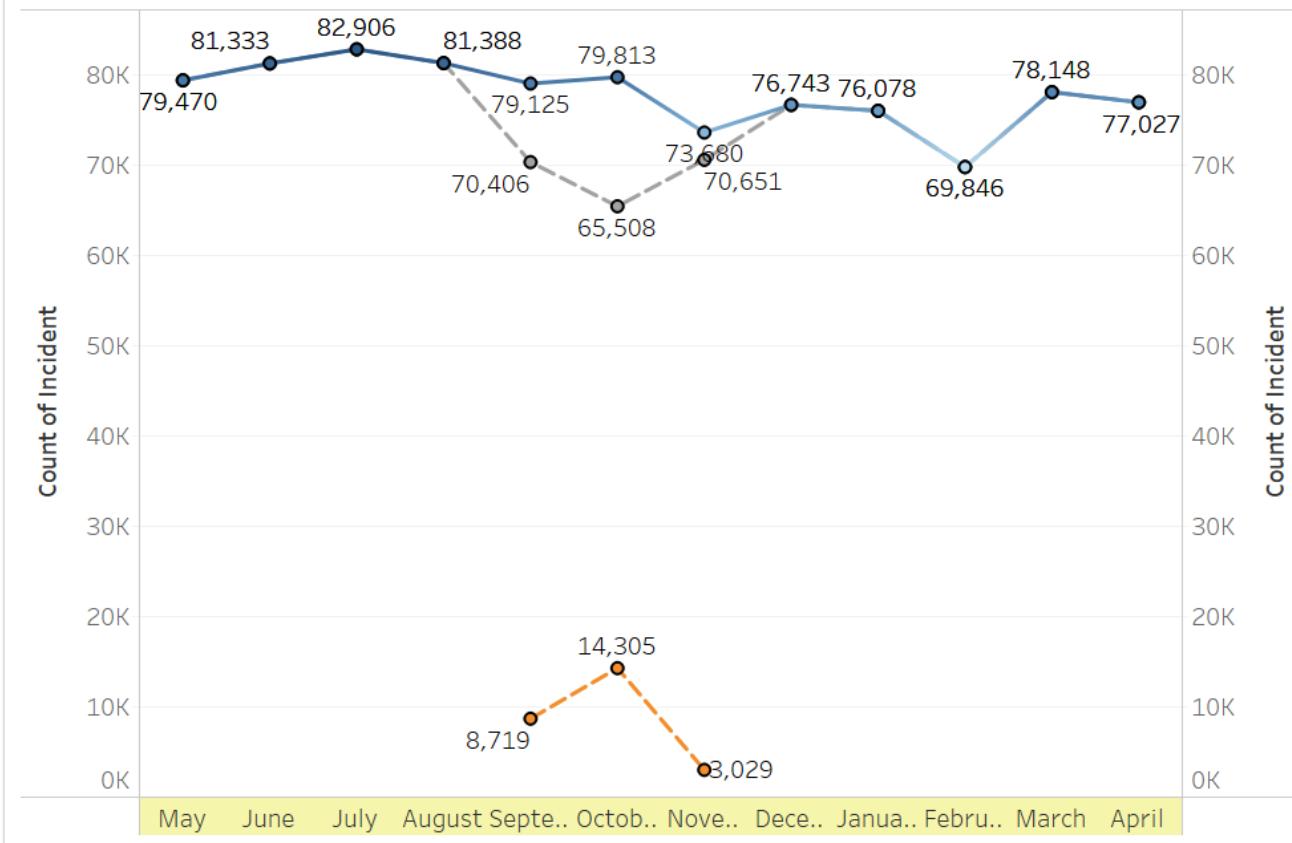
Incident

911 HANG-UP

Show history

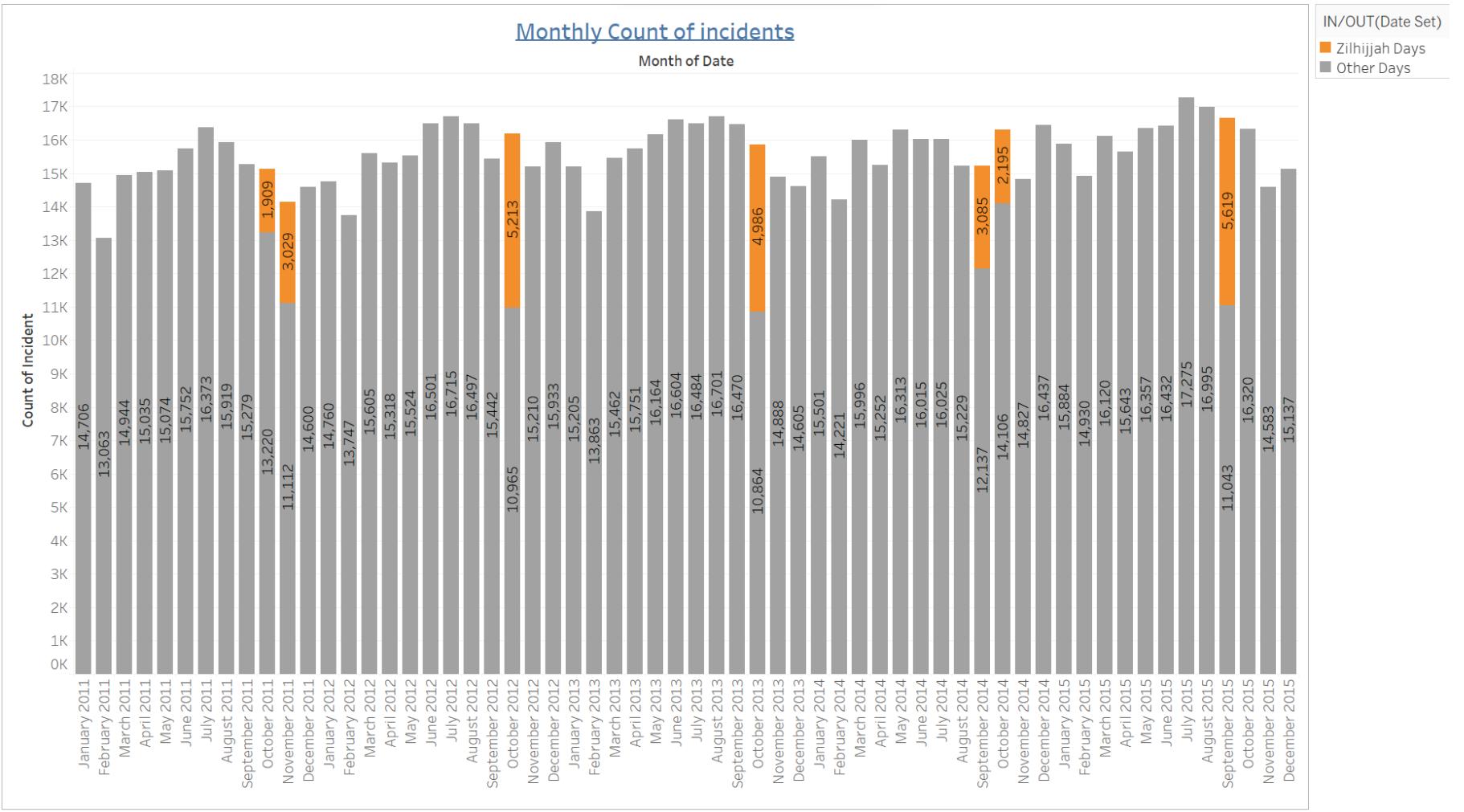
Monthly Counts of Incidents:

(Zilhijjah Days are also separately shown below.)



Caption

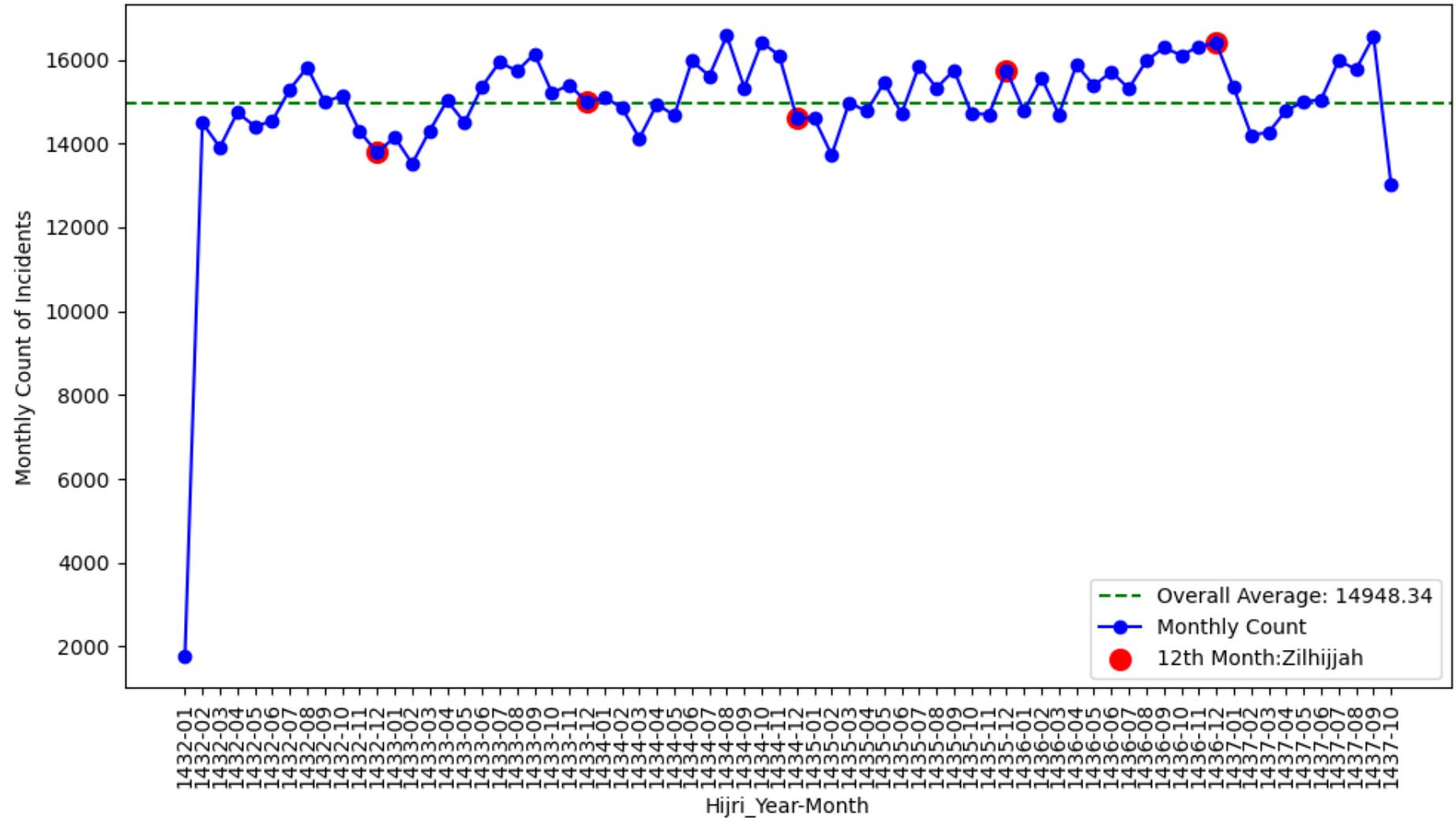
Contrary to the general decreasing trend, there is an increase from September to November, coinciding with the first ten days of Zilhijjah that surpass the annual average.

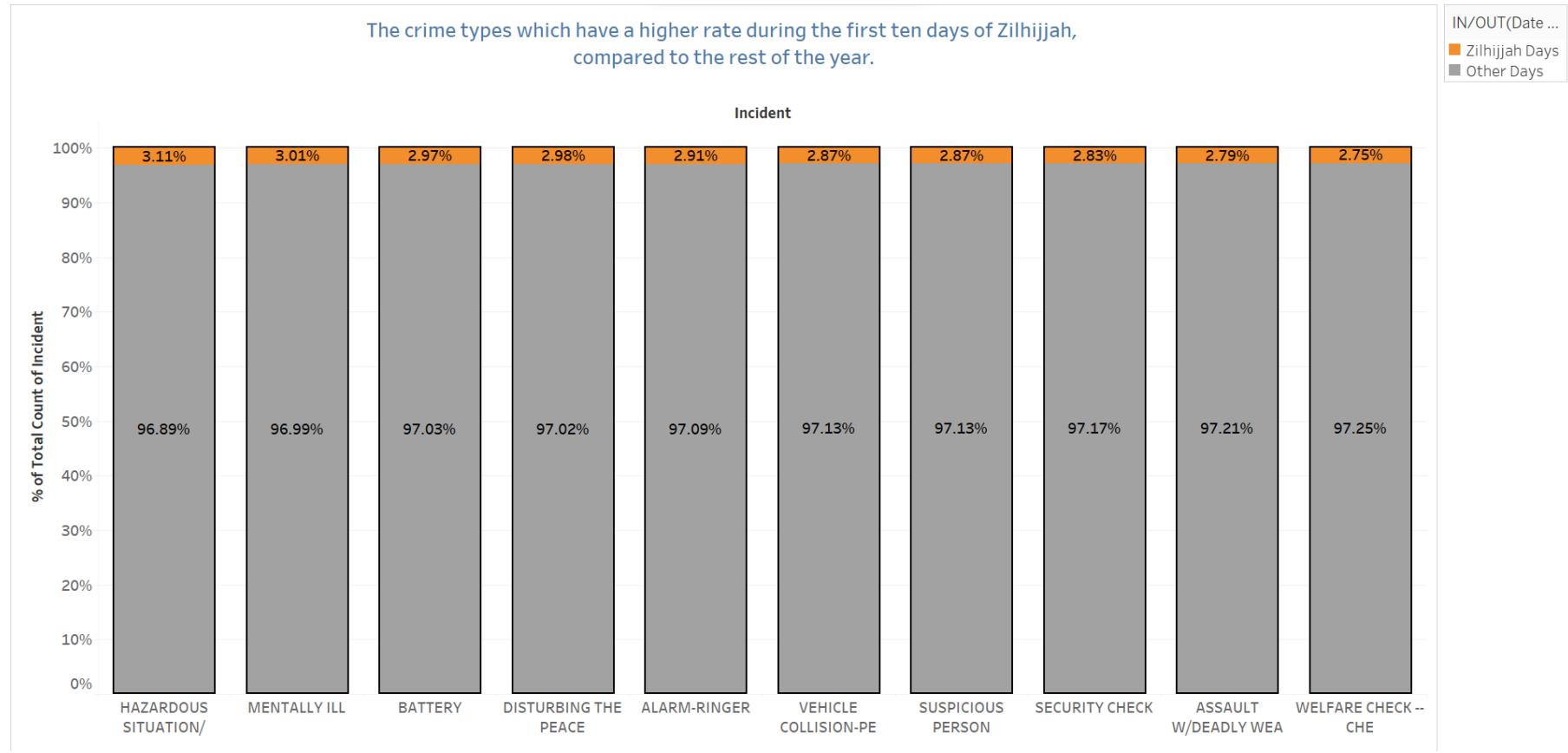


In [229]:

monthly_count_plot()

Monthly Count of Incidents on Hijri Calendar





CONCLUSION

The table below displays the percentage ratio of the daily average number of crimes committed during the first ten days of Zilhijjah to the daily average of crimes committed during the rest of the year. Positive percentages indicate a higher incidence of crimes during the Zilhijjah days, while negative values signify a lower incidence during these days compared to the rest of the year.

	Zilhijjah 10 Days/Other Days Ratio(%)	Zilhijjah 9 Days/Other Days Ratio(%)	Only Arafa(9th) Day/Other Days Ratio(%)	Only 10th Day/Other Days Ratio(%)
UNC CHAPEL HILL	14.37	14.70	57.85	10.62
BALTIMORE	9.45	8.59	19.39	16.55
DENVER	9.17	9.95	13.21	1.84
KANSAS	6.10	6.80	1.48	-0.38
DETROIT	3.61	3.61	0.55	3.44
ATLANTA	2.06	1.80	0.77	4.22
OAKLAND	1.57	1.12	-1.36	5.51
LOS ANGELES	0.79	0.39	2.38	4.24
VANCOUVER	0.64	0.44	1.28	2.39
CHICAGO	-1.96	-1.81	-6.15	-3.16

In [230...]

```
results = pd.DataFrame({'UNC CHAPEL HILL': [14.37, 14.70, 57.85, 10.62],
                       'LOS ANGELES': [0.79, 0.39, 2.38, 4.24],
                       'KANSAS': [6.10, 6.80, 1.48, -0.38],
                       'DETROIT': [3.61, 3.61, 0.55, 3.44],
                       'DENVER': [9.17, 9.95, 13.21, 1.84],
                       'VANCOUVER': [0.64, 0.44, 1.28, 2.39],
                       'CHICAGO': [-1.96, -1.81, -6.15, -3.16],
                       'BALTIMORE': [9.45, 8.59, 19.39, 16.55],
                       'ATLANTA': [2.06, 1.80, 0.77, 4.22],
                       'OAKLAND': [1.57, 1.12, -1.36, 5.51]
                      })
```

In [231...]

```
results.index = ['Zilhijjah 10 Days/Other Days Ratio(%)', 'Zilhijjah 9 Days/Other Days Ratio(%)', \
                 'Only Arafa(9th) Day/Other Days Ratio(%)', 'Only 10th Day/Other Days Ratio(%)']
daily_crime_ratios = results.T
```

In [232...]

```
daily_crime_ratios
```

Out[232] :

	Zilhijjah 10 Days/Other Days Ratio(%)	Zilhijjah 9 Days/Other Days Ratio(%)	Only Arafa(9th) Day/Other Days Ratio(%)	Only 10th Day/Other Days Ratio(%)
UNC CHAPEL HILL	14.37	14.70	57.85	10.62
LOS ANGELES	0.79	0.39	2.38	4.24
KANSAS	6.10	6.80	1.48	-0.38
DETROIT	3.61	3.61	0.55	3.44
DENVER	9.17	9.95	13.21	1.84
VANCOUVER	0.64	0.44	1.28	2.39
CHICAGO	-1.96	-1.81	-6.15	-3.16
BALTIMORE	9.45	8.59	19.39	16.55
ATLANTA	2.06	1.80	0.77	4.22
OAKLAND	1.57	1.12	-1.36	5.51

Interestingly, except for Chicago, nine data sets/cities showed higher crime rates during the Zilhijjah period. The fact that only 10 days of the year showcased a substantial difference, up to 10% in crime rates compared to other days, presents a significant finding. When categorizing the types of crimes more prevalent during the Zilhijjah period, we notice a prominence of offenses against individuals such as assault-violent, sexual offenses, and kidnapping. Even within the Chicago dataset, where the Zilhijjah period remains below the annual average, there are 16 categories, including offenses like sexual assault, kidnapping, and human trafficking, that exhibit above-average crime rates during the Zilhijjah period.

By excluding the tenth day and examining the first nine days of Zilhijjah, we observe that the outcomes remain largely unchanged. Once again, across nine datasets, we witness a higher incidence of crimes during the Zilhijjah period. When comparing the annual averages with just one specific day from the Zilhijjah month, either the 9th (Arafah Day) or the 10th day, we notice varying ratios. Nevertheless, across 10 datasets, in 8 instances, we still observe a higher incidence of crimes during the 9th or 10th day of Zilhijjah. Interestingly, these isolated days exhibit crime rates surpassing the yearly averages, with increases reaching up to 58% and 20% in certain datasets. These findings suggest a tendency toward increased criminal activity during the initial ten days of the Zilhijjah month.

- Across nine datasets, there's a consistent trend of higher crime rates during the initial ten days of Zilhijjah, with peaks of up to 14.37% above annual averages.
- Crimes against individuals like assault, sexual offenses, and kidnapping show prominence during this period, marking a significant deviation from yearly crime rates.
- Interestingly, a reversal of trends is noticed in some datasets between the first and second halves, suggesting a possible cyclic influence over the years, possibly related to astronomical cycles like the sunspot cycle.