

CRIMES in BOSTON

Exploratory Data Analysis
on Crimes in Boston

TABLE OF CONTENTS

01

BOSTON

A bit about Boston & Crime.

03

OVERVIEW

General analysis on crimes in Boston.

02

WRANGLING

Brief walkthrough on data cleaning and manipulation.

04

DIGGING DEEPER

Further insights & patterns.



01

BOSTON

A little about Boston
and Crime.

Boston...

is the most populous city of the Commonwealth of Massachusetts and the 24th most populous city in the US.

It was founded in September 7, 1630.

It plays host to one of the oldest marathons in the world, The Boston Marathon, it is the city where Alexander Graham Bell made his first phone call, and is where event that led to the American Revolution war, the Boston Tea Party.

Boston...



is the most populous city of the Commonwealth of Massachusetts and the 24th most populous city in the US.

It was founded in September 7, 1630.

It plays host to one of the oldest marathons in the world, The Boston Marathon, it is the city where Alexander Graham Bell made his first phone call, and is where event that led to the American Revolution war, the Boston Tea Party.

Boston...

just as any other major city it
faces a persistent major problem: **Crime**

Through this presentation I will tell you my data analysis exploration of
the crimes in Boston.

Things we're trying to produce:



General Overview



Insights



Patterns

Defining Crime

Crime

/krīm/

noun

an action or omission that constitutes an offense that may be prosecuted by the state and is punishable by law.

- *Oxford Languages*

Why does crime matter?



Crime and bad lives are the measure of a State's failure, all crime in the end is the crime of the community.

— H.G. Wells —

AZ QUOTES

In modern criminal law, There is no one simple and universal definition of crime, although one proposed definition states that crime is a harmful act not only to certain individuals but also towards the community and society.

In short, crime isn't the problem of an individual, it's the problem of society.

Defining Crime

Classifications of Crime



UCR Part One

Serious crimes covering violent and property crimes.

Includes severe crimes such as murder, rape, and robbery.

UCR Part Two

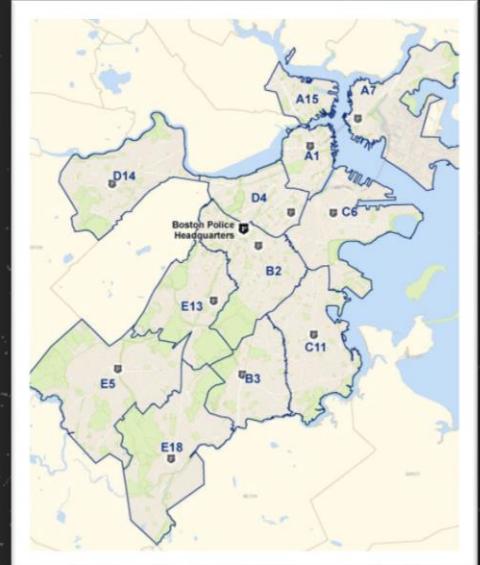
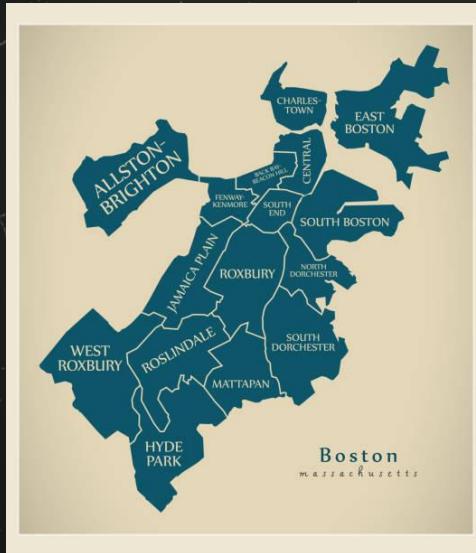
Category of less serious, non-violent crimes.

Only arrest data are collected.

Vandalism, fraud, drug abuse, prostitution are some crime categories within.

DISTRICTS

City of Neighborhoods





02

WRANGLING

Cleaning Data & Features



Cleaning

DUPLICATES

Out of **319,073** CASE ID

282,517 were unique

36,556 data were duplicates

and **DELETED**

Cleaning

MISSING YEARS

Latest Entry

2018-09-03 21:25:00

Earliest Entry

2015-06-15 00:00:00

- ⌚ 2015 is missing 5 months worth of entries - 42% missing.
- ⌚ 2018 missing 3 months worth of entries - 25% missing.
- ⌚ EDA will focus on complete years (2016 - 2017)

UNUSED COLUMNS

- ⌚ Dropped “OFFENSE_CODE_GROUP”
- ⌚ “LOCATION” column removed.
- ⌚ “STREET” and “REPORTING_AREA” column dropped.
- ⌚ Dropped “SHOOTING”.

Cleaning

SHOOTING



⌚ There are 281,891 null values / 282,517 entries.

⌚ No way of distinguishing between “No shooting” and “Missing Data”.

⌚ In order to prevent confusion, the column is removed.

Cleaning

COLUMNS RENAME

Before

INCIDENT_NUMBER	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	_OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	Lat	Long	
569	I182070304	Fraud	FRAUD - IMPERSONATION	D4	2017-11-01 00:00:00	2017	11	Wednesday	0	Part Two	42.339268	-71.072088
744	I182070115	Investigate Property	INVESTIGATE PROPERTY	D14	2017-10-11 17:43:00	2017	10	Wednesday	17	Part Three	42.346501	-71.136128
923	I182069920	Property Lost	PROPERTY - LOST	B2	2017-08-31 12:08:00	2017	8	Thursday	12	Part Three	42.331925	-71.084112
1162	I182069657	Fraud	FRAUD - FALSE PRETENSE / SCHEME	D14	2017-06-27 07:00:00	2017	6	Tuesday	7	Part Two	42.338482	-71.152890
1444	I182069334	Property Lost	PROPERTY - LOST	B3	2017-12-08 22:00:00	2017	12	Friday	22	Part Three	42.279591	-71.081323
...	

After

Case ID	Offense Type	Description	District	Date	Year	Month	Day	Hour	UCR Part	Latitude	Longitude	
569	I182070304	Fraud	Fraud - impersonation	D4	2017-11-01 00:00:00	2017	11	Wednesday	0	Part Two	42.339268	-71.072088
1162	I182069657	Fraud	Fraud - false pretense / scheme	D14	2017-06-27 07:00:00	2017	6	Tuesday	7	Part Two	42.338482	-71.152890
1813	I182068927	Fraud	Fraud - impersonation	C11	2016-01-10 00:00:00	2016	1	Sunday	0	Part Two	42.280498	-71.048885
2302	I182068383	Other	Trespassing	C6	2017-09-01 08:00:00	2017	9	Friday	8	Part Two	42.335468	-71.044891
2974	I182067634	Confidence games	Fraud - credit card / atm fraud	C11	2017-10-11 00:00:00	2017	10	Wednesday	0	Part Two	42.318532	-71.057790
...	

All of the columns were renamed for better comprehension, conciseness, and uniform capitalization for aesthetics.

Entries in Description column are now also in uniform capitalization.

Cleaning

THE “OTHERS”

```
1 df.loc[(df["Offense Type"]=="Other")].head()
```

✓ 0.7s

Case ID	Offense Type	Description	District	Date	Year	Month	Season	Day	Hour	UCR Part	Latitude	Longitude	
2302	I182068383	Other	Trespassing	C6	2017-09-01 08:00:00	2017	9	Autumn	Friday	8	Part Two	42.335468	-71.044891
9799	I182060154	Other	Obscene materials - pornography	C11	2017-07-07 23:00:00	2017	7	Summer	Friday	23	Part Two	42.281586	-71.060334
16134	I182053399	Other	Kidnapping/custodial kidnapping	C11	2017-07-01 12:00:00	2017	7	Summer	Saturday	12	Part Two	42.280917	-71.065654
21441	I182047771	Other	Threats to do bodily harm	C11	2016-01-01 12:00:00	2016	1	Winter	Friday	12	Part Two	42.299909	-71.064248
25309	I182043696	Other	Threats to do bodily harm	D14	2017-11-14 13:38:00	2017	11	Autumn	Tuesday	13	Part Two	42.359236	-71.128540

This Offense Type category is a combination of very different types of crimes, for example it has Kidnapping, which is a rather serious crime, and also pornography, which is trivial if compared with kidnapping.

Because this Offense Type is not well specified, it will not be included in the analysis.

Cleaning

INCONSISTENCIES

Burglary - other - attempt	13
Obscene phone calls	13
Weapon - firearm - sale / trafficking	10
Larceny theft from coin-op machine	8
Fraud - welfare	7
Explosives - possession or use	6
Violation - hawker and peddler	6
Custodial kidnapping	6
Abduction - enticing	4
Prisoner escape / escape & recapture	3
Gambling - betting / wagering	3
Drugs - class d trafficking over 50 grams	2
Biological threats	2
Prostitution - common nightwalker	2
Prostitution - assisting or promoting	2
Contributing to delinquency of minor	1
Conspiracy except drug law	1
Drugs - glue inhalation	1
Auto theft - recovered in by police	1
Drugs - poss class a - heroin, etc.	1
Drugs - poss class e intent to mf dist disp	1
Auto theft other	1
Larceny in a building \$200 & over	1
Prostitute - common nightwalker	1
Assault & battery d/w - other	1
Larceny other \$50 to \$199	1
Name - Description - d/voe - infid	1

There were several crime descriptions that were not inputted with the correct template.

Entries were scanned and put together with the corresponding categories.

Some typos including white space.

Features

SEASONS



Added season feature
based on Months.

```
def season(month):
    if 3 <= month < 6 : return "Spring"
    elif 6 <= month < 9 : return "Summer"
    elif 9 <= month < 12: return "Autumn"
    else : return "Winter"
```

0.6s

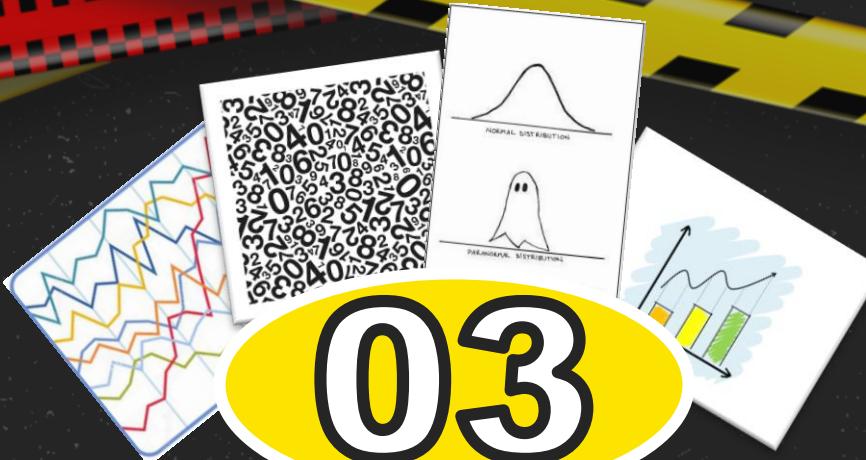
```
1 df["Season"] = df["Month"].apply(season)
2 df
✓ 0.7s
```

How I did it.

03

OVERVIEW

Statistics and General
Analysis



BASIC STATISTICS

Crime per Year:

2016 - 38,882 cases
2017 - 36,793 cases

Friday, 11,568, 15.29% | **Wednesday**, 11,186, 14.78% | **Tuesday**, 10,998, 14.53% | **Thursday**, 10,856, 14.35% | **Monday**, 10,824, 14.30% | **Saturday**, 10,583, 13.98% | **Sunday**, 9,660, 12.77%

Avg / Day:
104 crimes

Top 5 Crimes:

Larceny	14,831 cases
Vandalism	8810 cases
Simple Assault	8365 cases
Drug Violation	6,887 cases
Larceny from MV	6,435 cases

Top 5 Time with Most Crime

18:00	- 5,283
17:00	- 5,273
16:00	- 4,958
12:00	- 4,622
19:00	- 4,594

Avg / Month:

6306.25 incidents

Avg / District:

6278.25 incidents

BASIC STATISTICS IN DETAIL

Per Year

Crime per year in Boston:

Year

2016 38882

2017 36793

Name: Case ID, dtype: int64



Top Crimes

Top 5 Crimes in Boston in 2016-2017:

Offense Type

Larceny 14831

Vandalism 8810

Simple assault 8365

Drug violation 6887

Larceny from motor vehicle 6435

dtype: int64

Peak Hours

Top 5 time of day with most crime reports:

Hour

18 5283

17 5273

16 4958

12 4622

19 4594

Name: Case ID, dtype: int64

Per District

District Counts Percent

0 D4 11892 15.78

1 B2 11477 15.23

2 C11 9434 12.52

3 A1 9189 12.20

4 B3 7458 9.90

5 C6 5503 7.30

6 D14 4861 6.45

7 E13 4361 5.79

8 E18 3656 4.85

9 A7 3189 4.23

10 E5 2743 3.64

11 A15 1576 2.09

Per Month

Counts Percent

Month 1 6251 8.26

2 5531 7.31

3 6112 8.08

4 6325 8.36

5 6627 8.76

6 6674 8.82

7 6744 8.91

8 6948 9.18

9 6447 8.52

10 6501 8.59

11 5881 7.77

12 5634 7.44

Per Day

Crime frequency per day of week:

Counts Percent

Day

Friday 11568 15.29

Wednesday 11186 14.78

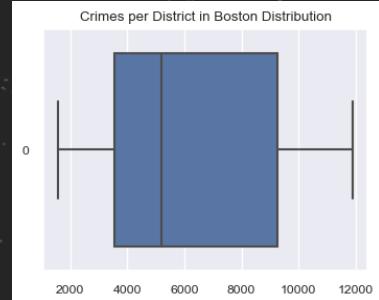
Tuesday 10998 14.53

Thursday 10856 14.35

Monday 10824 14.30

Saturday 10583 13.98

Sunday 9660 12.77



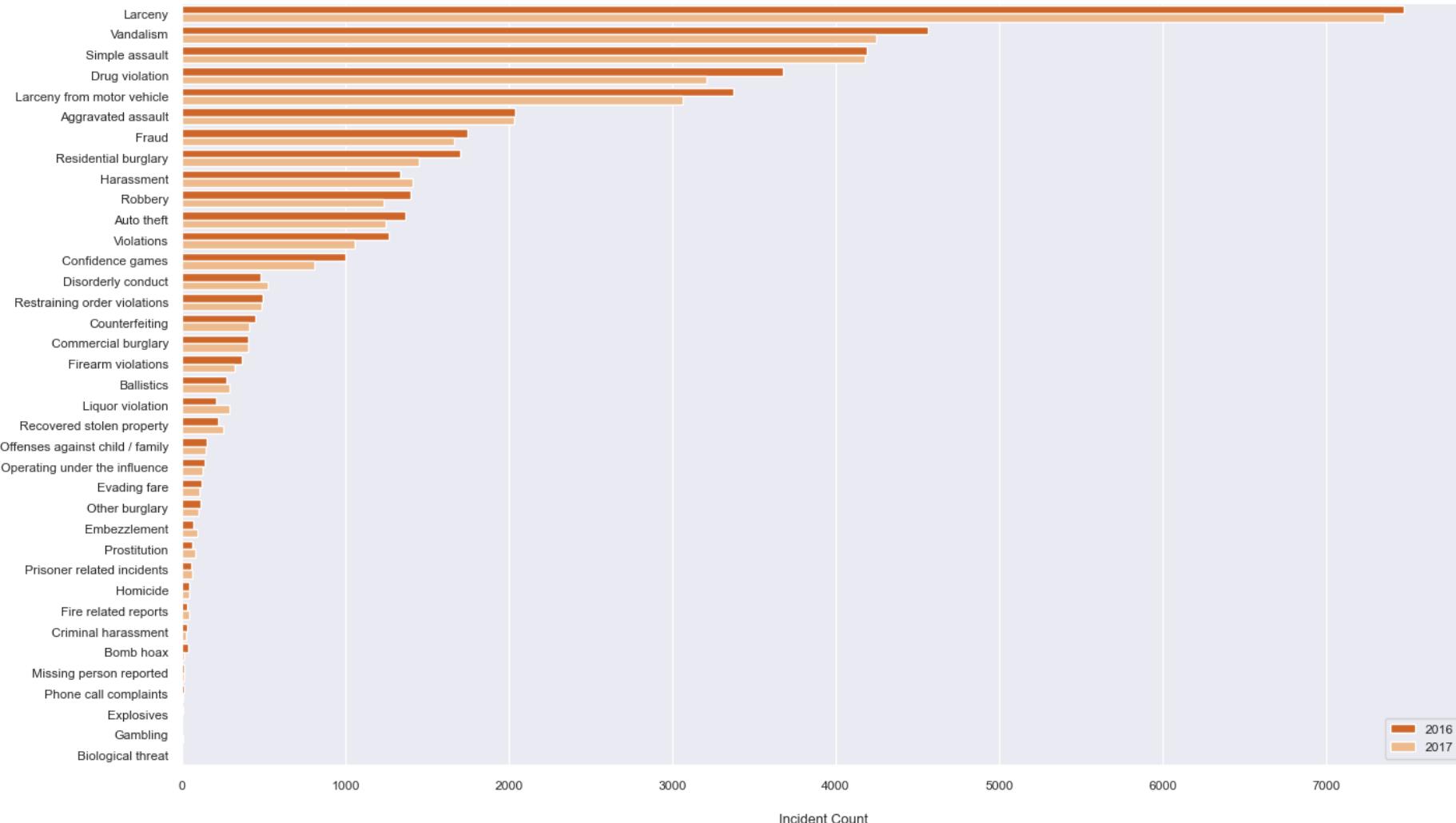
Just making
sure there
aren't any
outliers so I
can use mean.

VISUALIZATION TIME

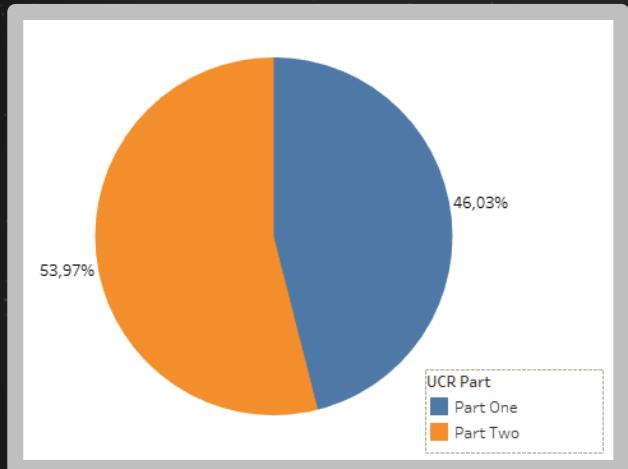
General Analysis



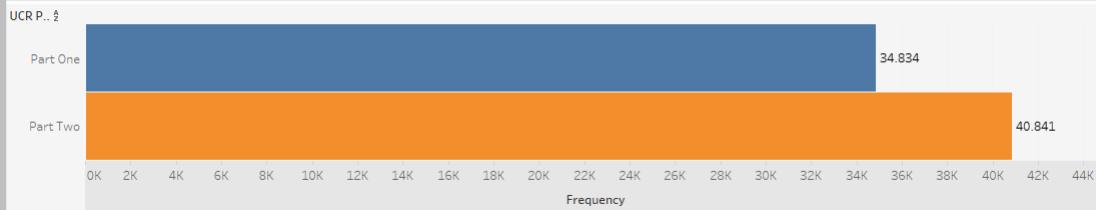
Crimes in Boston 2016-2017



UCR PART Distribution



Crime count in 2016 and 2017 with UCR Part



UCR PART Distribution



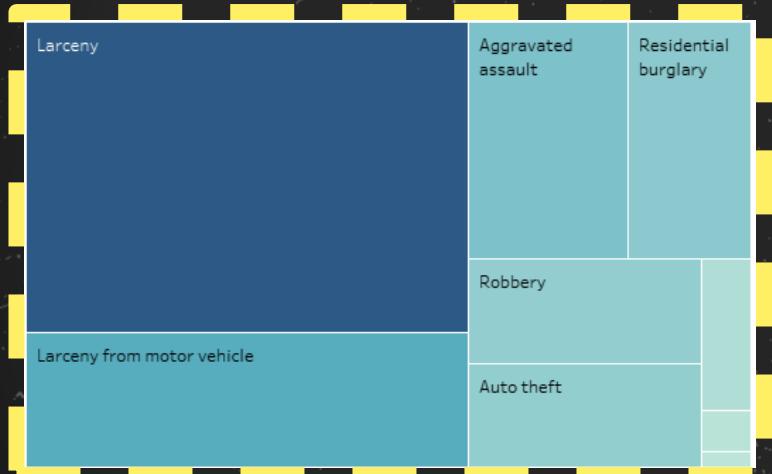
UCR Part

Part One

Part Two

UCR PARTS

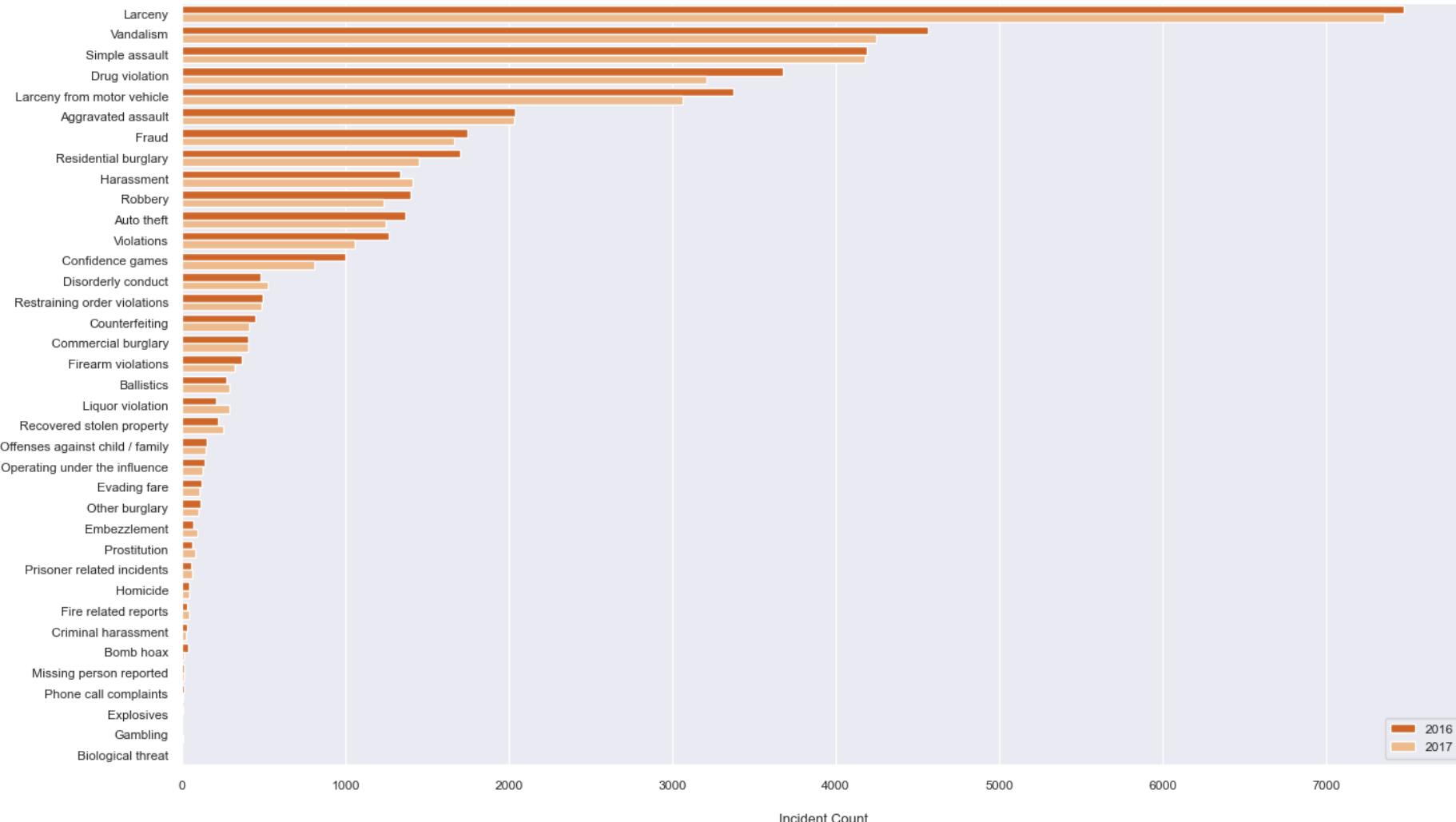
UCR Part One



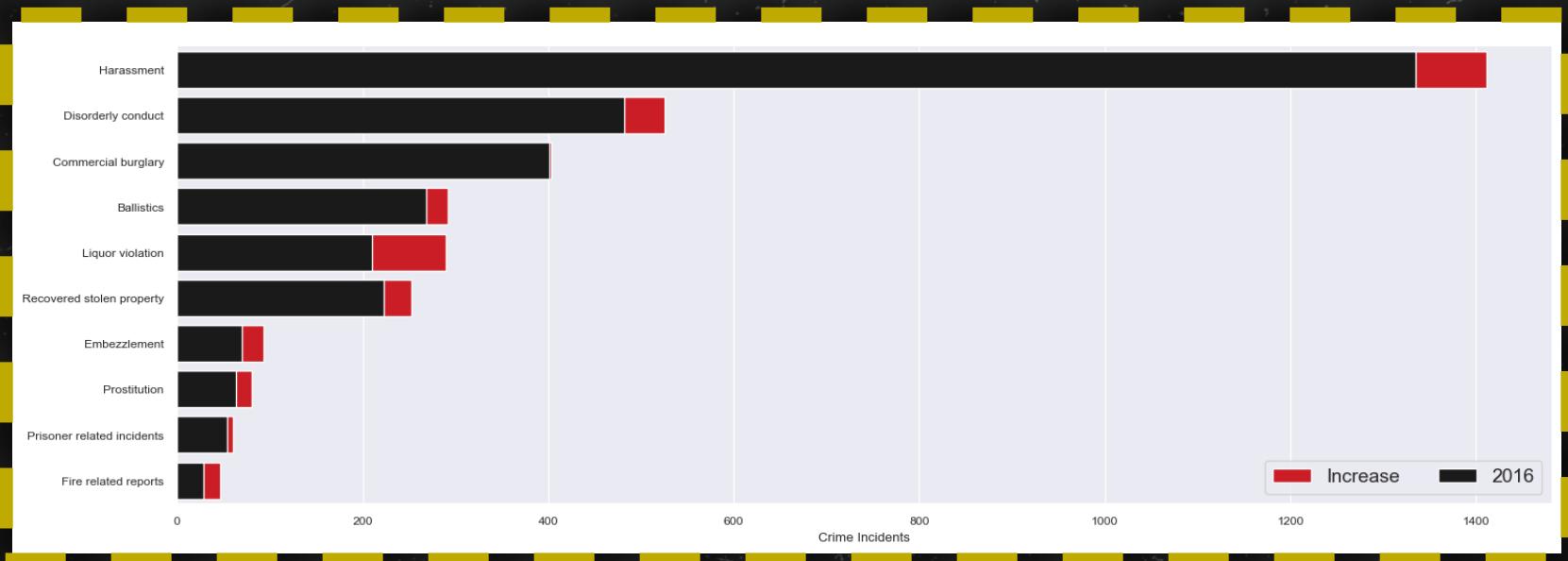
UCR Part Two



Crimes in Boston 2016-2017

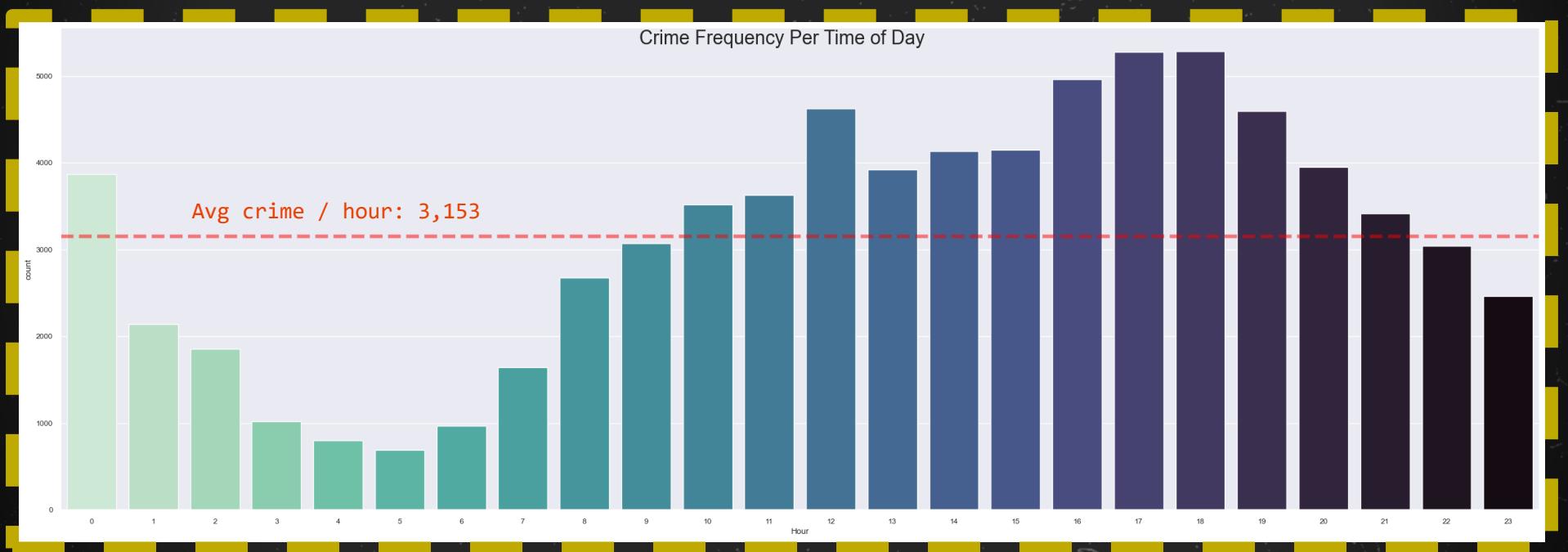


CRIME INCREASE



CRIME FREQUENCY

Per Hour



CRIME FREQUENCY

Per Hour – Line Chart



Crime activity starts to be climb and exceeds the average line at 9AM – 10 AM and peaks at 6PM. Police presence could be increased during these hours.

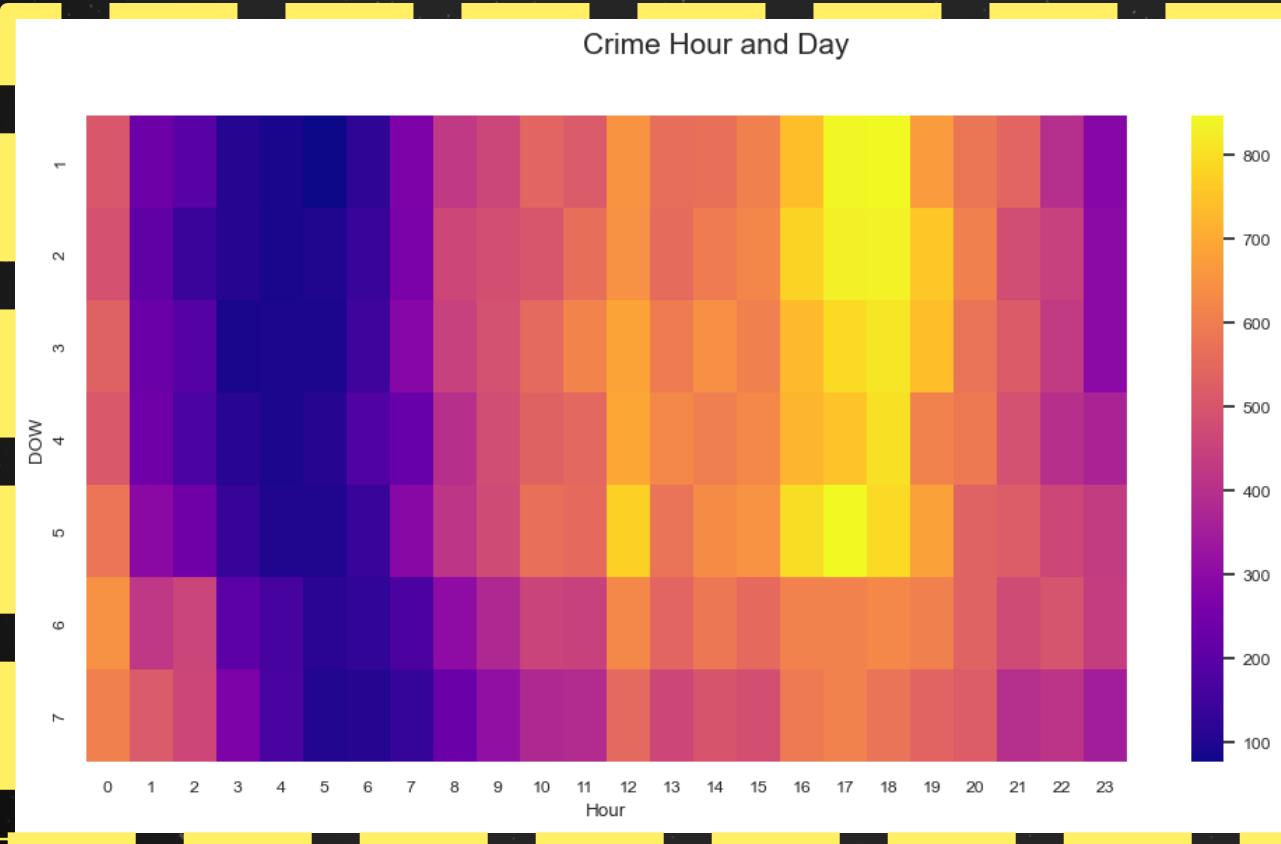
CRIME FREQUENCY

Per Hour & Day



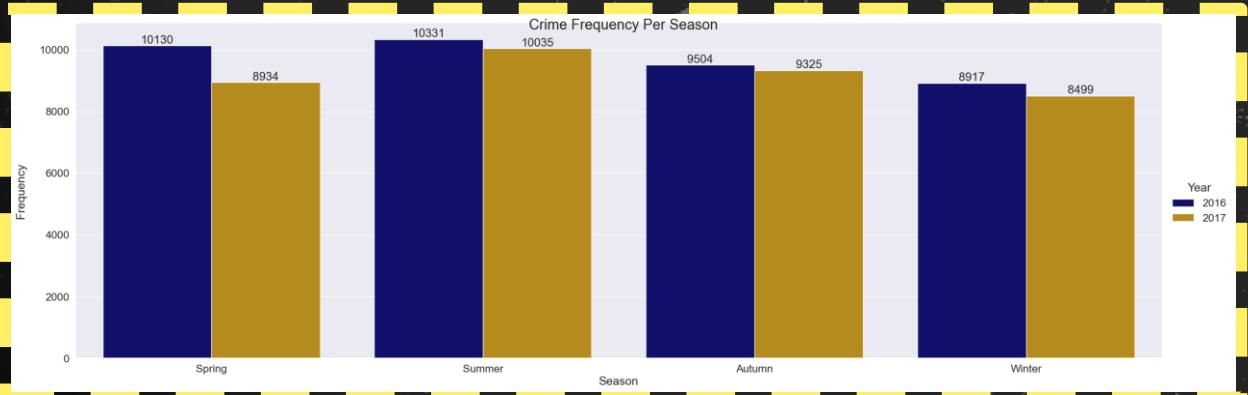
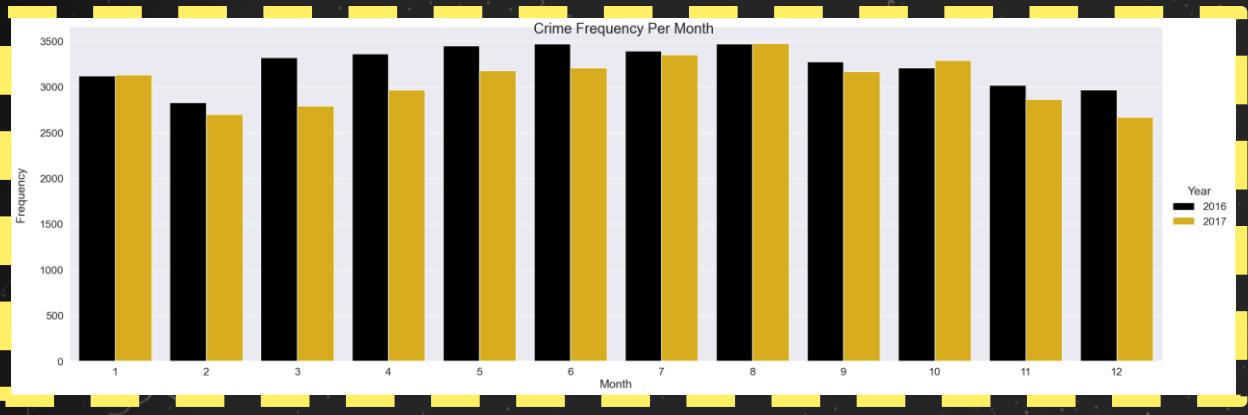
CRIME FREQUENCY

Per Hour & Day



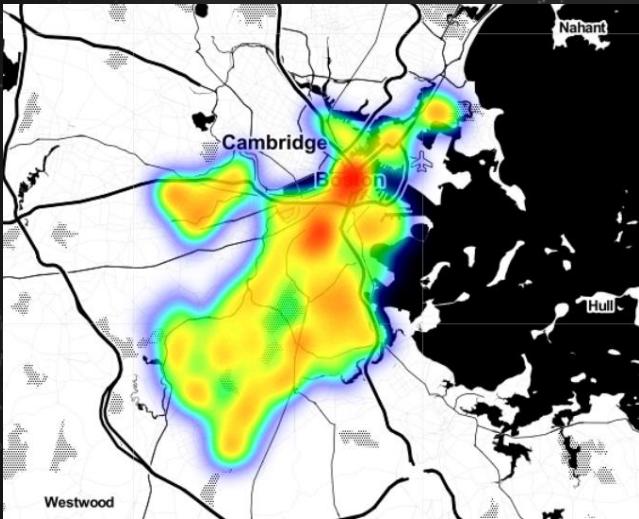
CRIME FREQUENCY

Month and Season

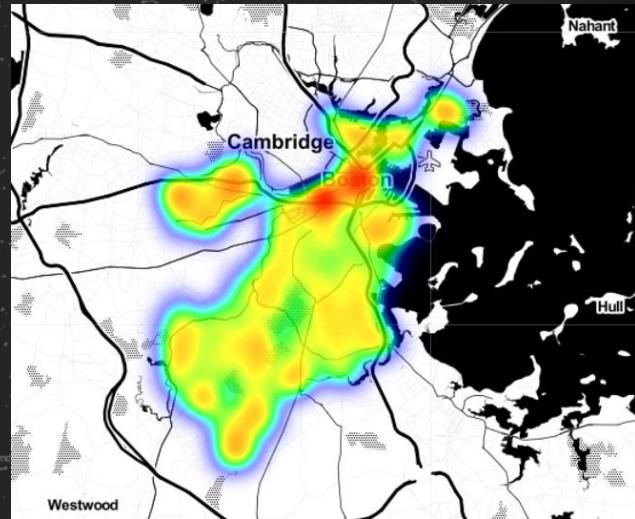


CRIME MAP

UCR Part Two



UCR Part One

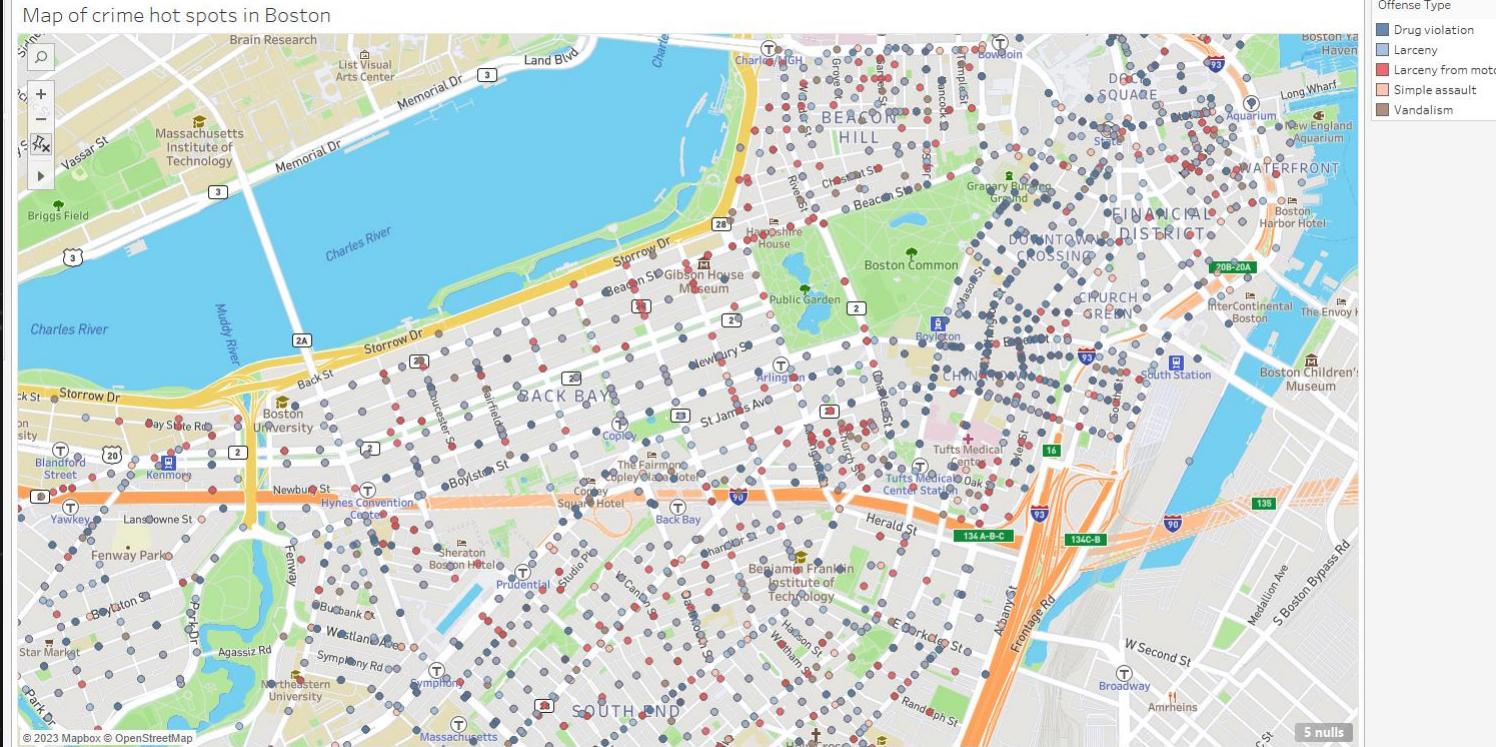


Based on the geo heat map visualization above, we can see that UCR Part One crimes are more concentrated in the central area in Boston, particularly starting from the downtown area and gradually dissipating as it spreads to the outer areas of Boston

UCR Part Two crimes are less centralized and has three noticeable heat spots: Downtown Boston, Harrington Avenue, Blue Hill Avenue

CRIME MAP

Map of crime hot spots in Boston



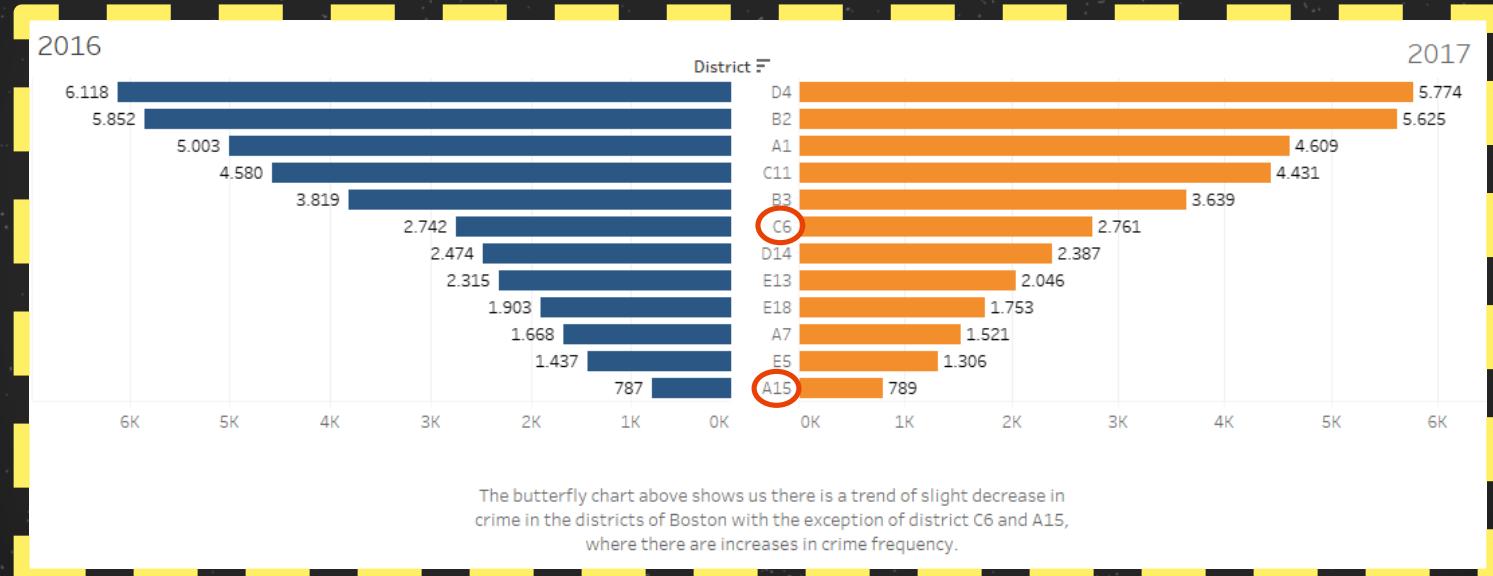
There's literally a crime on every block in South End, Back Bay, and Downtown.

CRIME MAP

An idea of how district D4 looks like.

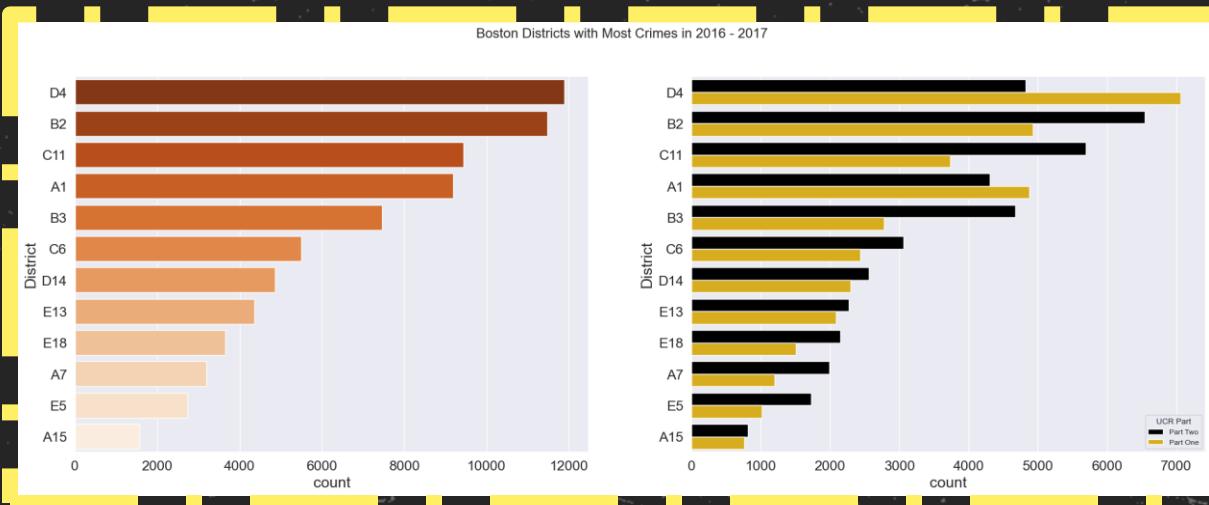


CRIME PER DISTRICT



CRIME PER DISTRICT

Another Angle.



Generally most districts have more UCR Part Two crimes than Part One but not only does district D4 have the most crimes it also has most violent crimes as well.

It is highly recommended to focus efforts on District D4.

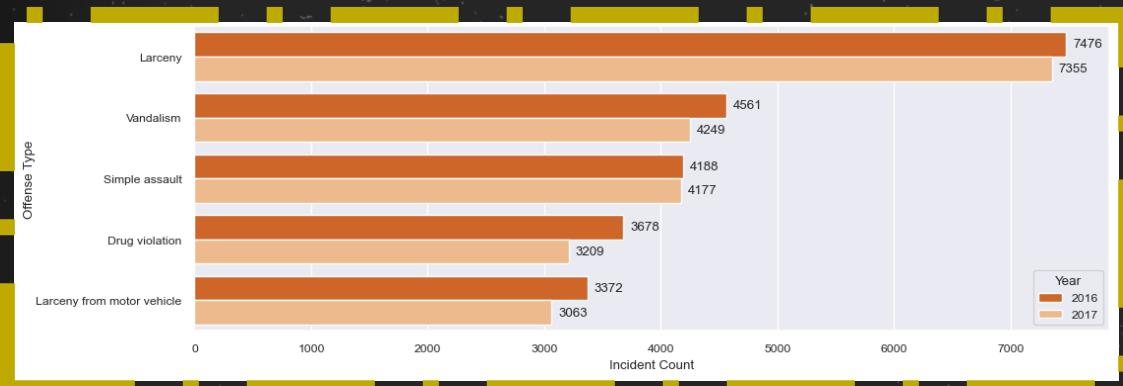


04

DIGGING DEEPER

Analysis on Boston's
Most Prominent Crimes

Boston Top 5 Crimes 2016 -2017



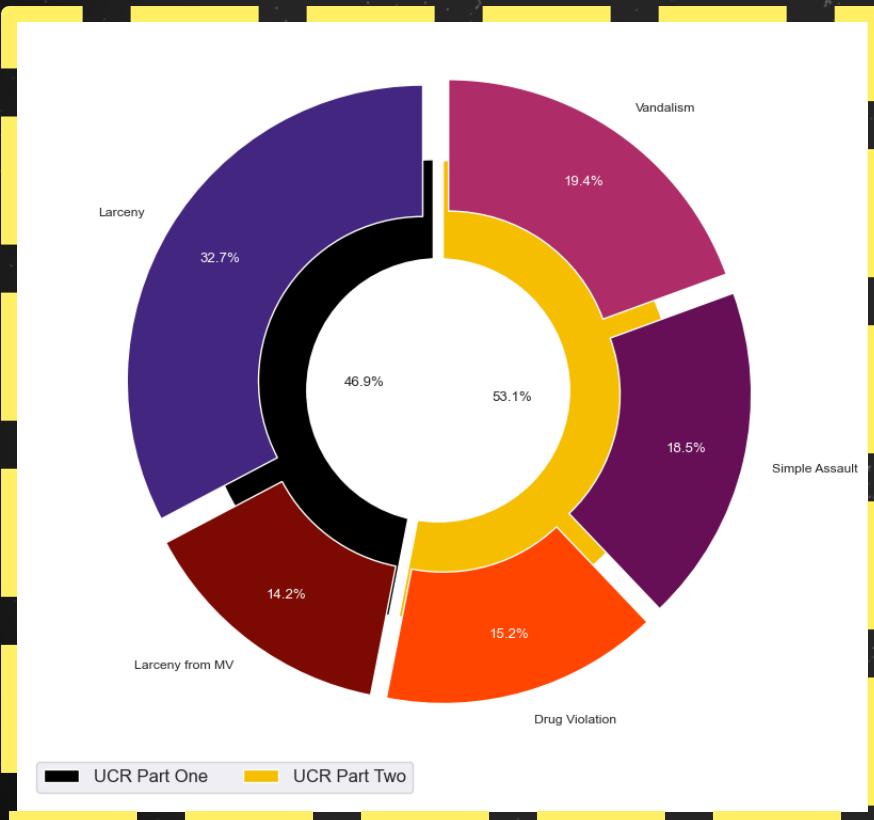
Fun fact about larceny is, excluding “Others” category, there are 75,675 incidents in 2016 & 2017 combined. In those two years there are 14.831 larceny cases, that’s 19.6% of all the total cases in Boston in 2016-2017, that’s how serious this crime is.



Boston population in 2017 is 688,239, larceny case in 2017 is 7,355 if we do the math that means 1 in 94 people is a victim of larceny in Boston.

The Crime Pie

Distribution of top 5 crimes with UCR



TOP 5 CRIMES

BREAKDOWN

Crimes Breakdown

Part Two

Vandalism

Vandalism

Simple assault

Assault simple - battery

Drug violation

Poss class b - Intent

Sale / manufacturing

Poss class b - cocaine, etc.

Sick assist - heroin

Poss class a - Intent

Poss class a - heroin, etc.

Other

Poss class d - Intent

Poss class d

Poss class d

Poss class c

Poss class b

Poss class a

Poss class b

Poss class a

Poss class b

Poss class a

Part One

Larceny

From building

Shoplifting

All others

Of bicycle

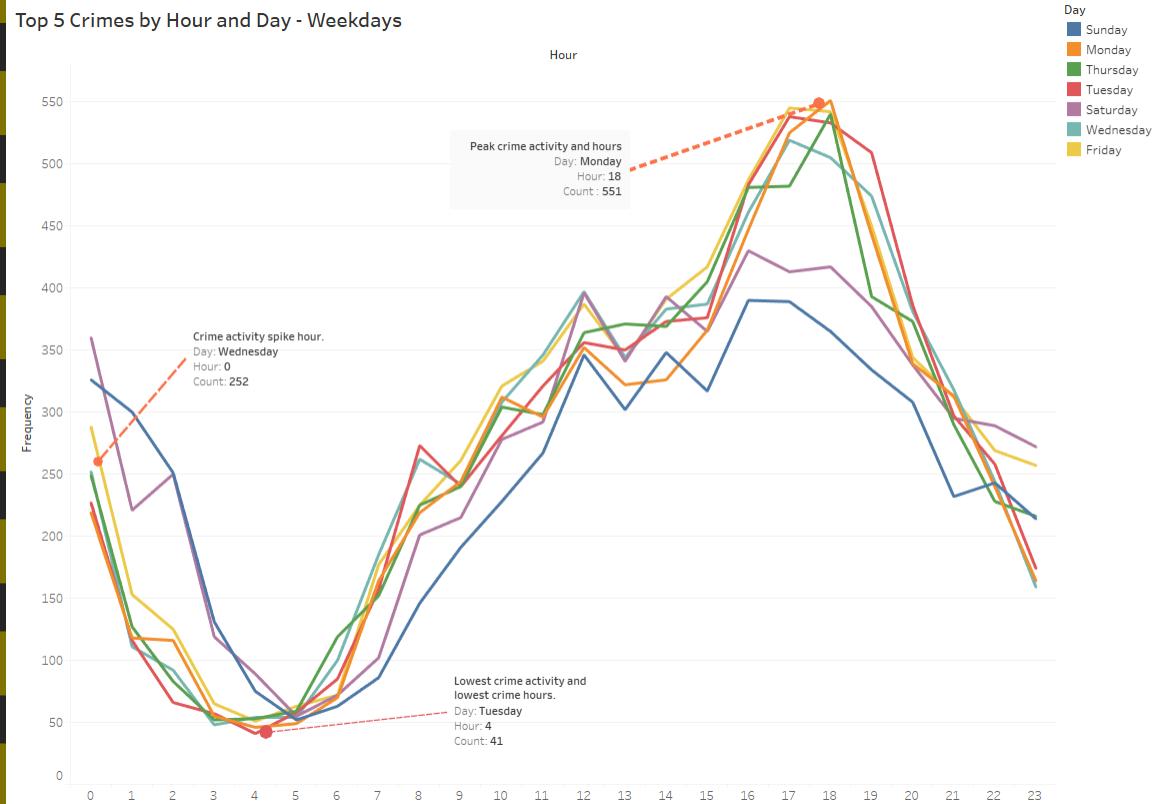
Pick-pocket

Larceny from motor vehicle

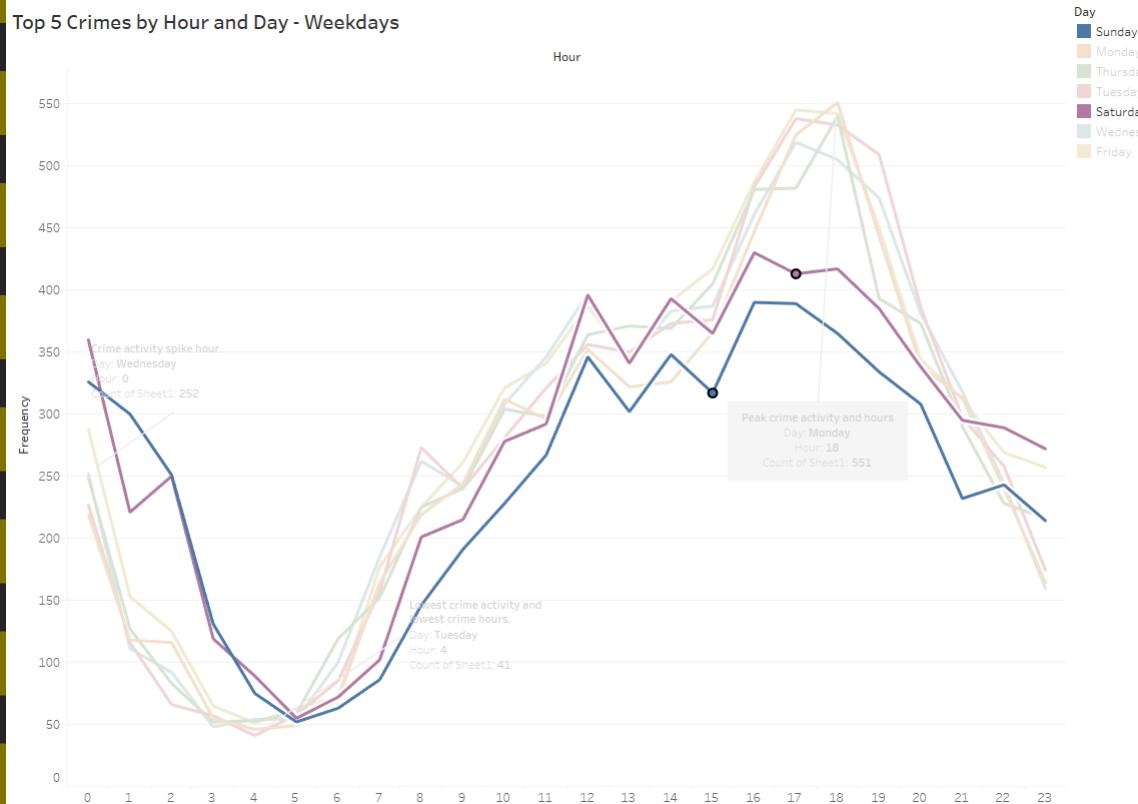
From mv - non-accessory

Of mv parts & accessories

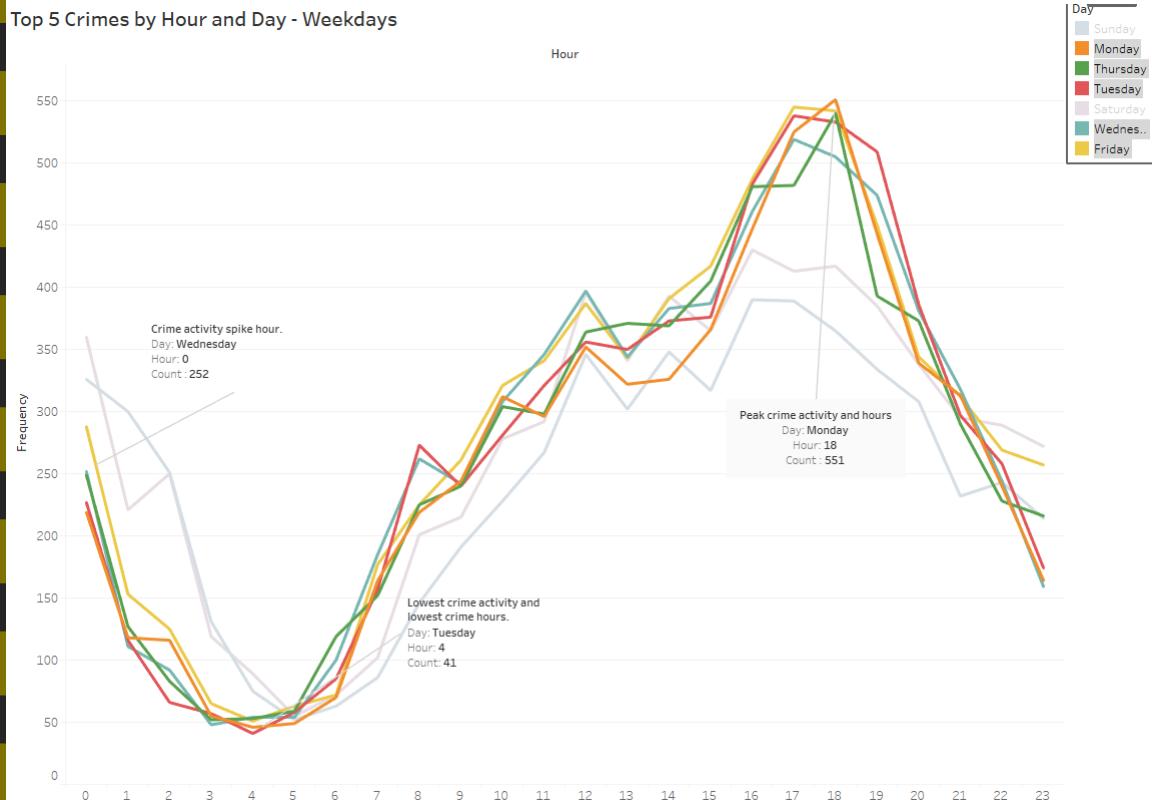
Top 5 Crimes Per Hour and Day



Top 5 Crimes Per Hour and Day

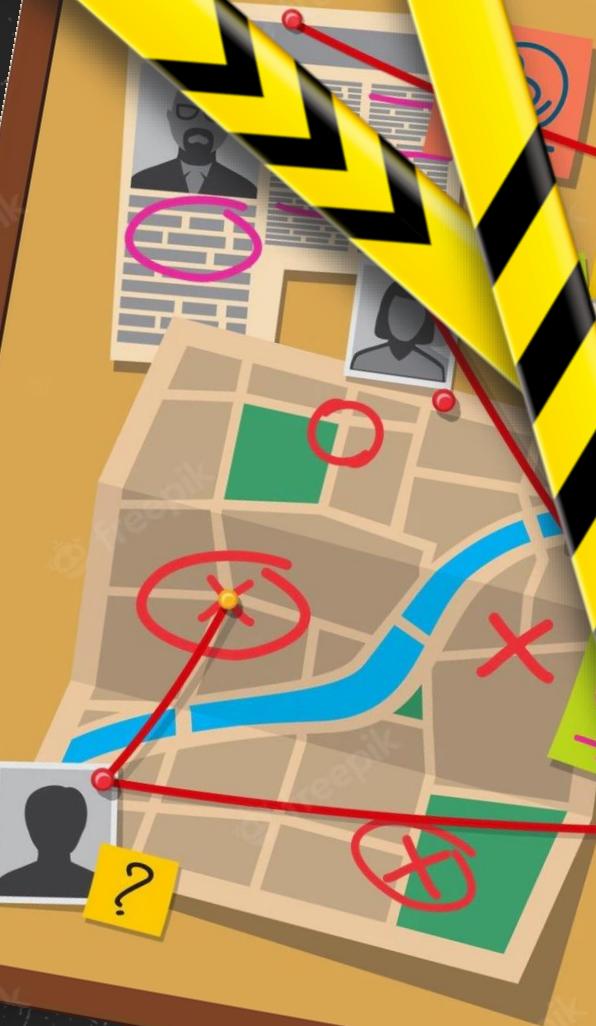


Top 5 Crimes Per Hour and Day



ACTIONABLE SUGGESTIONS

Highlights



Focus

- ❖ Larceny is the most common crime in Boston.
- ❖ Reducing larceny crimes means reducing overall crime in Boston.
- ❖ D4 district should be top priority in reducing larceny and also other crimes.
- ❖ Larceny from MV should be second priority.

Time

- ❖ 12 PM
- ❖ 12 AM
- ❖ Between 10 AM – 9 PM

Method

- ❖ Most of the top common crime in Boston could be prevented.
- ❖ Social media awareness campaigns could be done before peak hours of crimes to increase awareness and vigilance.
- ❖ Educating owners of establishments that are targets of crime on how to prevent and stop larceny.
- ❖ Training and education on the importance of surveillance / CCTV.
- ❖ Strategic police assignment in crime hotspots.



END

Thank You

Kindly find the complete
detailed analysis in my IPYNB
and Tableau story.