CRIMES IN THE CITIES

PRESENTED BY LAYLA LIU & LEJIA HU

ABSTRACT

PLATFORM SUPPORTED BY NYU DUMBO CLUSTER

This project aimed to provide some insight on how the police department could better arrange their limited police resources in major cities in the United States, represented by New York City and Boston, using big data analytics tools.

In the project, we used MapReduce to clean the data and Impala for further analysis. We found the top 20 most frequently admitted crime in the cities. For each crime type, we further analyzed the month and exact time of the day the crime is most likely to happen.

From the analytics, results were that the most common types of crime in NYC were petit larceny and harassment with a percentage of 17% and 13% accordingly. The time these were most likely to happen is at the end of the year, from 3-8 pm in a day. The most frequent crime type in Boston was human injury. It most frequently occurs in the afternoons of July.

The conclusion using the above data analytics showed that the police department should focus more on the above type of crimes in these particular times.

MOTIVATION

- The U.S News reported that "Oakland spent 41 percent of the city's general fund on policing in Fiscal Year 2017. Chicago spent nearly 39 percent, Minneapolis almost 36 percent, Houston 35 percent" (Neuhauser).
- With these much money spent, the crime rates in big cities were still much higher than small cities or suburb areas (Glaeser).
- So we would like to examine how to use police resources in the most efficient way.
- We would like to use our knowledge as well as the big data tools to give advice to the police department and hopefully lower crime rates in big cities.

GOODNESS

The results of our analysis were trustworthy for the following reasons:

- 1. The most frequently committed crimes we found are consistent with what's on wikipedia, namely larceny, motor vehicle theft, etc.
- 2. Our conclusions were drawn from a sufficient amount of data.
- 3. There were no logistic errors in the codes we wrote.

Some of the results we found were counter-intuitive. For instance, according to our analysis, crimes like larceny, property loss and vandalism occurs most frequently during the day in Boston. This might be due to that Boston dataset records on this type of crimes are somewhat biased.

United Sta	ites
Crime rates* (2	2017)
Violent crim	nes
Homicide	5.3
Rape	41.7
Robbery	98.0
Aggravated assault	248.9
Total violent crime	382.9
Property cris	mes
Burglary	430.4
Larceny-theft	1,694.4
Motor vehicle theft	237.4
Total property crime	2,362.2
Notes	
*Number of reported crimes per	100,000 population.
Estimated total population: 3	25,719,178.
In 2013 the FBI modified the def	finition of rape.
Source: Crime in the United S	tates by Volume and

Rate per 100,000 Inhabitants, 1998–2017 (Table

1) 🚱

DATA SOURCES

TITLE: NEW YORK CITY CRIMES

SIZE: 253.42 MB

DESCRIPTION: This data set comes from NYC Open Data, and it was uploaded to Kaggle. It records crimes in NYC between 2014 and 1025. In this data set, we used month of when it happened, time in a day it happened and the crime types.

LINK:

HTTPS://WWW.KAGGLE.COM/ADAMSCHROEDER/CRIMES-NEW-YORK-CITY?SELECT=NYPD_COMPLAINT_DATA_HIST ORIC.CSV



SAMPLE DATA - NYC

CMPLNT, CMPI	NT_CMPLNT_CMPLNT_	CMPLNT_RP	PT_DT K	Y_CD OFNS_DE	PD_CD PD_DESC CRM_ATP	LAW_CAT JURIS_DE BORO_N	ADDR_PC LOC_OF_(PREM_TY PARKS_	N HADEVEL X_COORD	Y_COORD Latitude	Longitude Lat_Lon	
1.01E+06 ####	#### 23:45:00	##	######	113 FORGERY	729 FORGERY COMPLET	FELONY N.Y. POLICERONX	44 INSIDE BAR/NIGHT CLUB	1007314	241257 40.82885	-73.9167 (40.828848333, -	-73.916661142)
1.53E+08 ####	#### 23:36:00	##	######	101 MURDER 8	NON-NEGL. MANSI COMPLET	FELONY N.Y. POLICQUEENS	103 OUTSIDE	1043991	193406 40.69734	-73.7846 (40.697338138, -	-73.784556739)
5.69E+08 ####	#### 23:30:00	##	######	117 DANGERO	503 CONTROL COMPLET	FELONY N.Y. POLICMANHAT	28 OTHER	999463	231690 40.80261	-73.9451 (40.802606608, -	-73.945051911)
9.68E+08 ####	#### 23:30:00	##	######	344 ASSAULT	101 ASSAULT COMPLET	MISDEME N.Y. POLICQUEENS	105 INSIDE RESIDENCE-HOUS	E 1060183	177862 40.65455	-73.7263 (40.654549444, -	-73.726338791)
6.42E+08 ####	#### 23:25:00 #######	23:30:00 ##	######	344 ASSAULT	101 ASSAULT COMPLET	MISDEME N.Y. POLI(MANHAT	13 FRONT OF OTHER	987606	208148 40.738	-73.9879 (40.7380024, -73	.98789129)
3.66E+08 ####	#### 23:18:00 #######	23:25:00 ##	######	106 FELONY A	109 ASSAULT ATTEMPT	FELONY N.Y. POLICEROOKLY	71 FRONT OF DRUG STORE	996149	181562 40.66502	-73.9571 (40.665022689, -	73.957110763)
6.08E+08 ####	#### 23:15:00	##	######	235 DANGERO	511 CONTROL COMPLET	MISDEME N.Y. POLI(MANHAT	7 OPPOSITE STREET	987373	201662 40.7202	-73.9887 (40.720199996, -	-73.988735082)
2.65E+08 ####	#### 23:15:00 #######	23:15:00 ##	######	118 DANGERO	792 WEAPON COMPLET	FELONY N.Y. POLICERONX	46 FRONT OF STREET	1009041	247401 40.84571	-73.9104 (40.845707148, -	73.910398033)
9.89E+08 ####			######	344 ASSAULT	101 ASSAULT COMPLET	MISDEME N.Y. POLICERONX	48 INSIDE RESIDENCE - APT.	HOUSE 1014154	251416 40.85671	-73.8919 (40.856711291, -	73.891899956)
4.15E+08 ####				341 PETIT LAF	338 LARCENY COMPLET	MISDEME N.Y. POLI(MANHAT		994327	218211 40.76562	-73.9636 (40.765617688, -	73.96362342)
7.31E+08 ####			######	341 PETIT LAF		MISDEME N.Y. POLI(BRONX	41 INSIDE FAST FOOD	1014216	238784 40.82204	-73.8917 (40.822039935, -	
1.78E+08 ####		23:05:00 ##	######	341 PETIT LAF		MISDEME N.Y. POLI(MANHAT		R 988113	206263 40.73283	-73.9861 (40.732828332, -	73.986062857)
8.98E+08 ####			######	109 GRAND L		FELONY N.Y. POLI(MANHAT		987215	215403 40.75792		
5.66E+08 ####			######	109 GRAND L		FELONY N.Y. POLICMANHAT		987215	215403 40.75792	-73.9893 (40.757915693, -	73.98929902)
5.85E+08 ####	#### 23:00:00	##	######	113 FORGERY		FELONY N.Y. POLI(BROOKL)		1004325	174113 40.64456	-73.9277 (40.644562053, -	73.92766205)
7.16E+08 ####				105 ROBBERY		FELONY N.Y. POLICQUEENS	103 INSIDE BAR/NIGHT CLUB	1044662	197327 40.7081		
3.38E+08 ####				109 GRAND L		FELONY N.Y. POLICMANHAT			214414 40.7552	-73.9683 (40.755197275, -	
4.33E+08 ####				359 OFFENSE		MISDEME N.Y. POLICERONX	48 INSIDE RESIDENCE - APT.		251858 40.85792		
7.61E+08 ####				344 ASSAULT		MISDEME N.Y. POLI(BROOKL)				-73.9651 (40.606308897, -	
2.35E+08 ####				351 CRIMINAL		MISDEME N.Y. POLICQUEENS	102 FRONT OF DRY CLEANER/LAU		186863 40.6795	-73.8628 (40.679498977, -	73.862825258)
2.32E+08 ####				235 DANGERO		MISDEME N.Y. POLI(BROOKL)		995387	178144 40.65564	-73.9599 (40.655642023, -	73.959863084)
3.41E+08 4/6/2				104 RAPE		FELONY N.Y. POLICQUEENS	110 INSIDE OTHER				
7.04E+08 ####				235 DANGERO		MISDEME N.Y. POLICQUEENS	108 STREET	1012629			
5.49E+08 ####	#### 22:50:00 #######	22:58:00 ##	######	106 FELONY A	105 STRANGU COMPLET	FELONY N.Y. POLI(BROOKL)	75 INSIDE RESIDENCE - APT.	HOUSE 1018309	183847 40.67124	-73.8772 (40.67123699, -7	3.877222202)

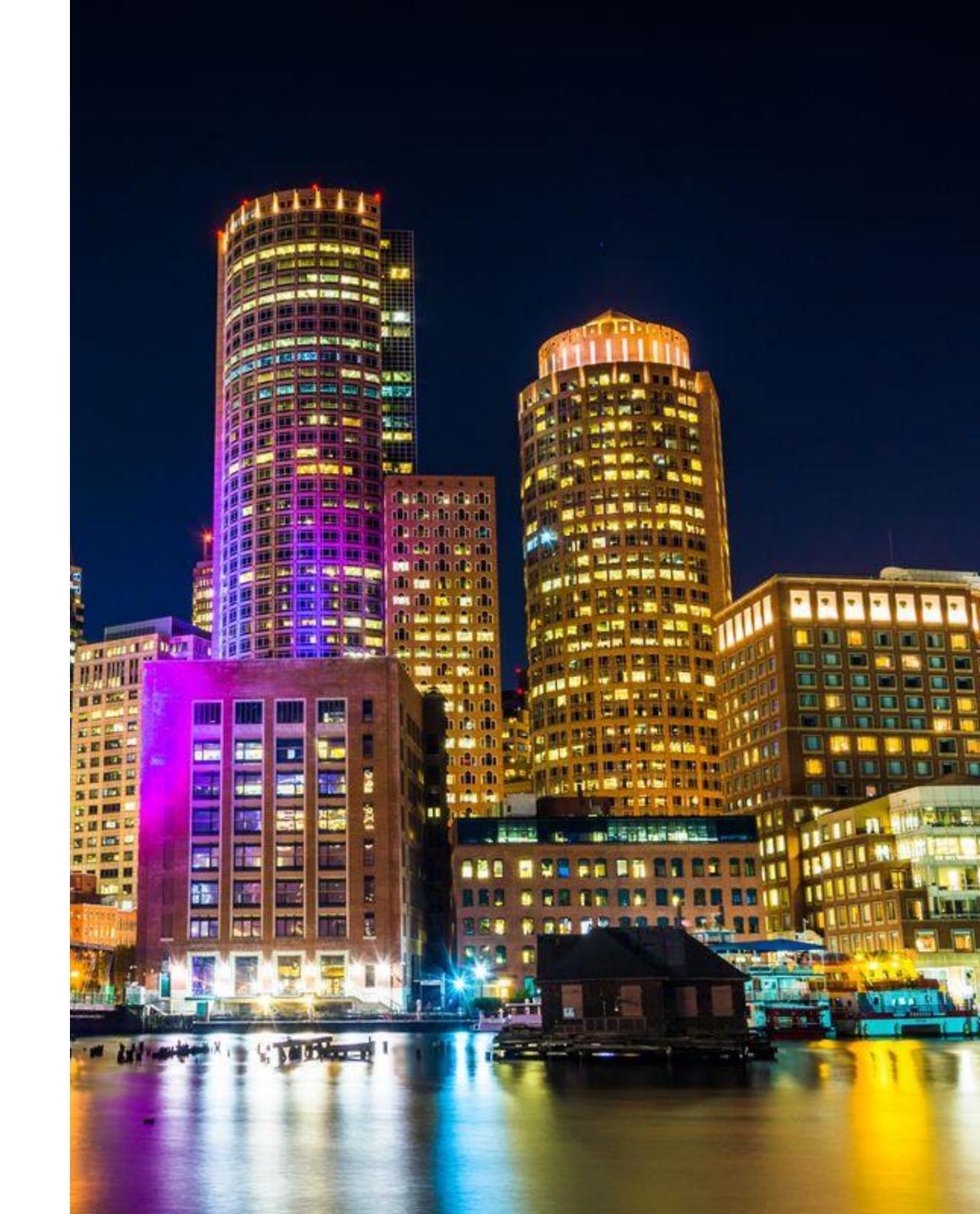
DATA SOURCES

TITLE: Crimes in Boston

SIZE: 10 MB

DESCRIPTION: Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred. Records begin in June 14, 2015 and continue to September 3, 2018.

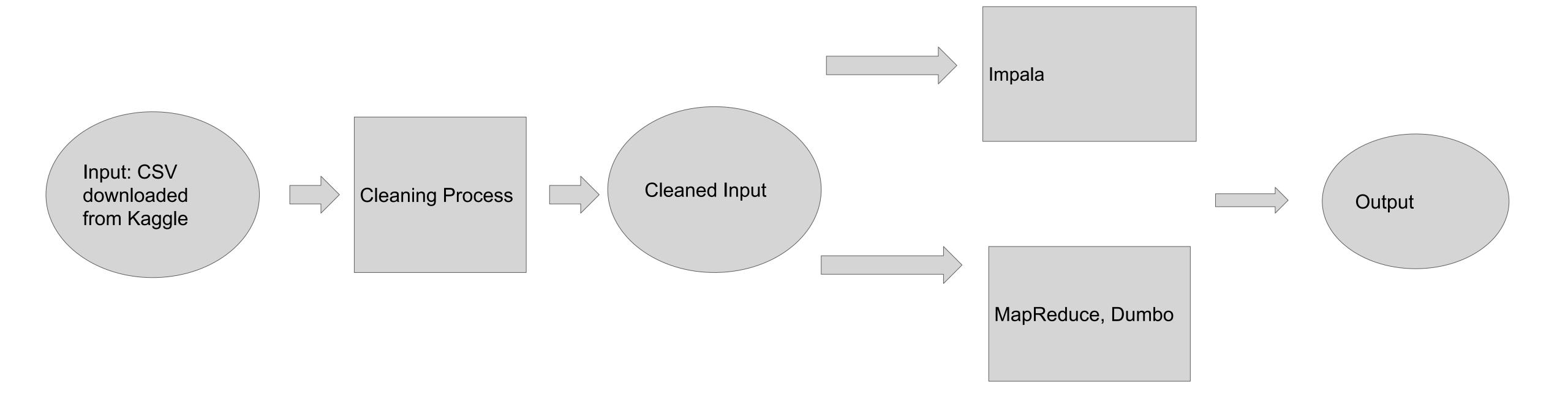
LINK: https://www.kaggle.com/AnalyzeBoston/crimes-in-boston



SAMPLE DATA - Boston

INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long	Location
I182070945	619	Larceny	LARCENY ALL OTHERS	D14	808		2018-09-02 13:00:00	2018	9	Sunday	13	Part One	LINCOLN ST	42.35779134	-71.13937053	(42.35779134, -71.1393705
1182070943	1402	Vandalism	VANDALISM	C11	347		2018-08-21 00:00:00	2018	8	Tuesday	0	Part Two	HECLA ST	42.30682138	-71.06030035	(42.30682138, -71.0603003
I182070941	3410	Towed	TOWED MOTOR VEHICLE	D4	151		2018-09-03 19:27:00	2018	9	Monday	19	Part Three	CAZENOVE ST	42.34658879	-71.07242943	(42.34658879, -71.0724294
I182070940	3114	Investigate Property	INVESTIGATE PROPERTY	D4	272		2018-09-03 21:16:00	2018	9	Monday	21	Part Three	NEWCOMB ST	42.33418175	-71.07866441	(42.33418175, -71.0786644
I182070938	3114	Investigate Property	INVESTIGATE PROPERTY	B3	421		2018-09-03 21:05:00	2018	9	Monday	21	Part Three	DELHI ST	42.27536542	-71.09036101	(42.27536542, -71.0903610
I182070936	3820	Motor Vehicle Accident Response	M/V ACCIDENT INVOLVING PEDESTRIAN - INJURY	C11	398		2018-09-03 21:09:00	2018	9	Monday	21	Part Three	TALBOT AVE	42.29019621	-71.07159012	(42.29019621, -71.0715901
1182070933	724	Auto Theft	AUTO THEFT	B2	330		2018-09-03 21:25:00	2018	9	Monday	21	Part One	NORMANDY ST	42.30607218	-71.0827326	(42.30607218, -71.0827326
l182070932	3301	Verbal Disputes	VERBAL DISPUTE	B2	584		2018-09-03 20:39:37	2018	9	Monday	20	Part Three	LAWN ST	42.32701648	-71.10555088	(42.32701648, -71.1055508
1182070931	301	Robbery	ROBBERY - STREET	C6	177		2018-09-03 20:48:00	2018	9	Monday	20	Part One	MASSACHUSETTS AVE	42.33152148	-71.07085307	(42.33152148, -71.0708530
1182070929	3301	Verbal Disputes	VERBAL DISPUTE	C11	364		2018-09-03 20:38:00	2018	9	Monday	20	Part Three	LESLIE ST	42.29514664	-71.05860832	(42.29514664, -71.0586083
l182070928	3301	Verbal Disputes	VERBAL DISPUTE	C6	913		2018-09-03 19:55:00	2018	9	Monday	19	Part Three	OCEAN VIEW DR	42.31957856	-71.04032766	(42.31957856, -71.0403276
l182070927	3114	Investigate Property	INVESTIGATE PROPERTY	C6	936		2018-09-03 20:19:00	2018	9	Monday	20	Part Three	DALESSIO CT	42.34011469	-71.05339029	(42.34011469, -71.0533902
l182070923	3108	Fire Related Reports	FIRE REPORT - HOUSE, BUILDING, ETC.	D4	139		2018-09-03 19:58:00	2018	9	Monday	19	Part Three	MARLBOROUGH ST	42.3503876	-71.0878529	(42.35038760, -71.0878529
l182070922	2647	Other	THREATS TO DO BODILY HARM	B3	429		2018-09-03 20:39:00	2018	9	Monday	20	Part Two	WOODROW AVE	42.28647012	-71.08714661	(42.28647012, -71.0871466
l182070921	3201	Property Lost	PROPERTY - LOST	B3	469		2018-09-02 14:00:00	2018	9	Sunday	14	Part Three	MULVEY ST	42.27924052	-71.09667382	(42.27924052, -71.0966738
l182070920	3006	Medical Assistance	SICK/INJURED/MEDICAL - PERSON				2018-09-03 19:43:00	2018	9	Monday	19	Part Three		42.35287456	-71.0738297	(42.35287456, -71.0738297
I182070919	3301	Verbal Disputes	VERBAL DISPUTE	C11	341		2018-09-03 18:52:00	2018	9	Monday	18	Part Three	STONEHURST ST	42.30526428	-71.06683755	(42.30526428, -71.0668375
I182070918	3305	Assembly or Gathering Violations	DEMONSTRATIONS/RIOT	D4	130		2018-09-03 17:00:00	2018	9	Monday	17	Part Three	HUNTINGTON AVE	42.34857652	-71.07772012	(42.34857652, -71.0777201
I182070917	2647	Other	THREATS TO DO BODILY HARM	B2	901		2018-09-03 19:52:00	2018	9	Monday	19	Part Two	HORADAN WAY	42.33371742	-71.09665806	(42.33371742, -71.0966580
1182070915	614	Larceny From Motor Vehicle	LARCENY THEFT FROM MV - NON-ACCESSORY	B2	181		2018-09-02 18:00:00	2018	9	Sunday	18	Part One	SHIRLEY ST	42.3256949	-71.06816778	(42.32569490, -71.0681677

DIAGRAM



CODE CHALLENGE

Challenge: Separating the hour time from the specific time. For example: 12 out of 12:30:00 Solution:

```
dateFormat1 = new SimpleDateFormat("HH:mm:ss");
DateFormat dateFormatHour = new SimpleDateFormat("K aa", Locale.ENGLISH);
try {
    Date date = dateFormat1.parse(fields[1]);
    String hourString = dateFormatHour.format(date);
    fields[1] = hourString;
} catch (ParseException e) {
    // e.printStackTrace();
}
String newString = String.join(",", fields);
key = new Text(newString);
super.reduce(key, values, context);
```

CODE CHALLENGE

Challenge: Removing unwanted commas in strings

Solution:

```
flag = False
line_list = list(line)
for i in range(len(line_list)):
    if line_list[i] == '"' and flag == False:
        flag = True
    elif line_list[i] == '"' and flag == True:
        flag = False
    if line_list[i] == "," and flag == True:
        line_list[i] = "."
line = "".join(line_list)
```

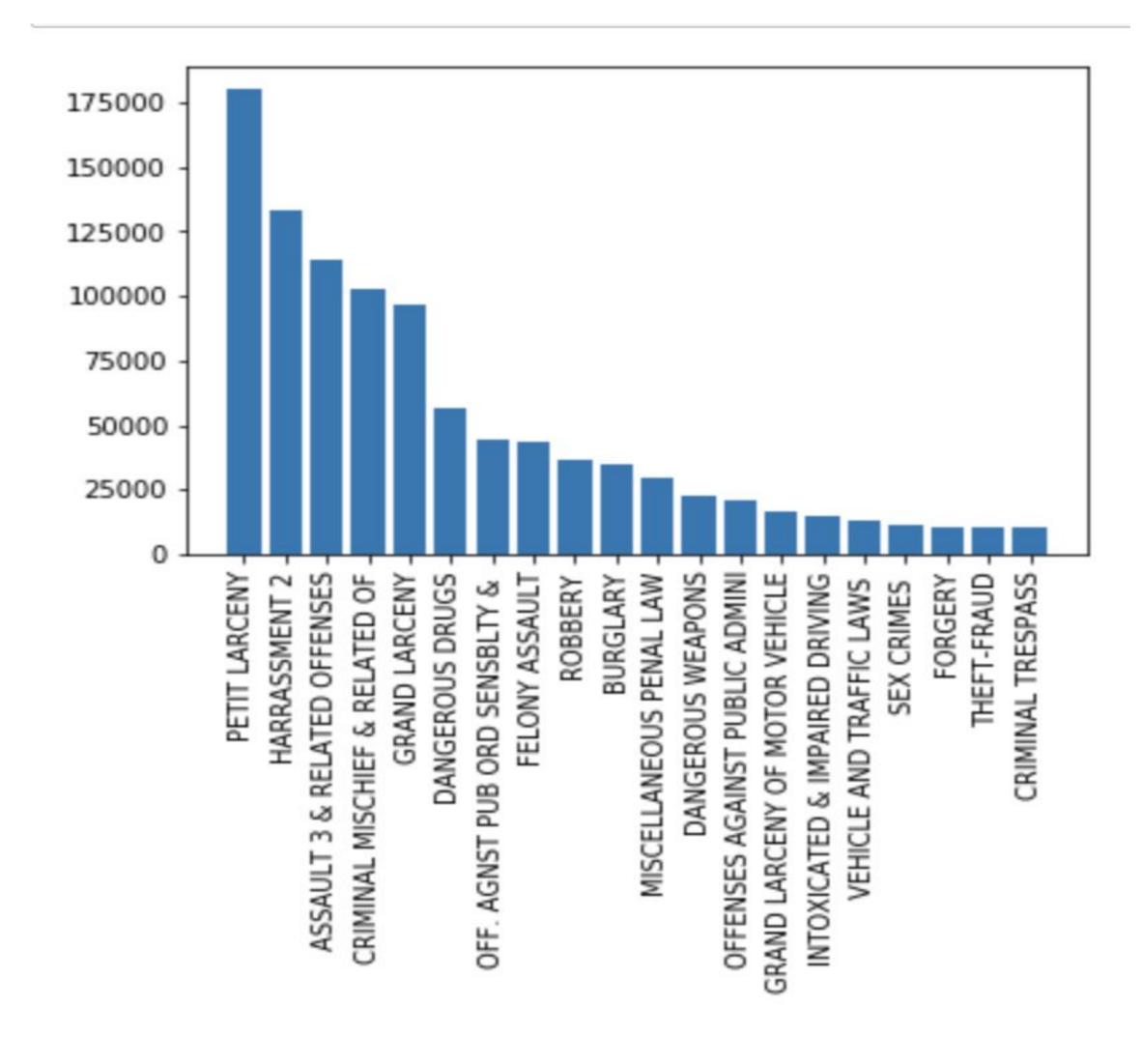
INSIGHTS - Crimes in NYC

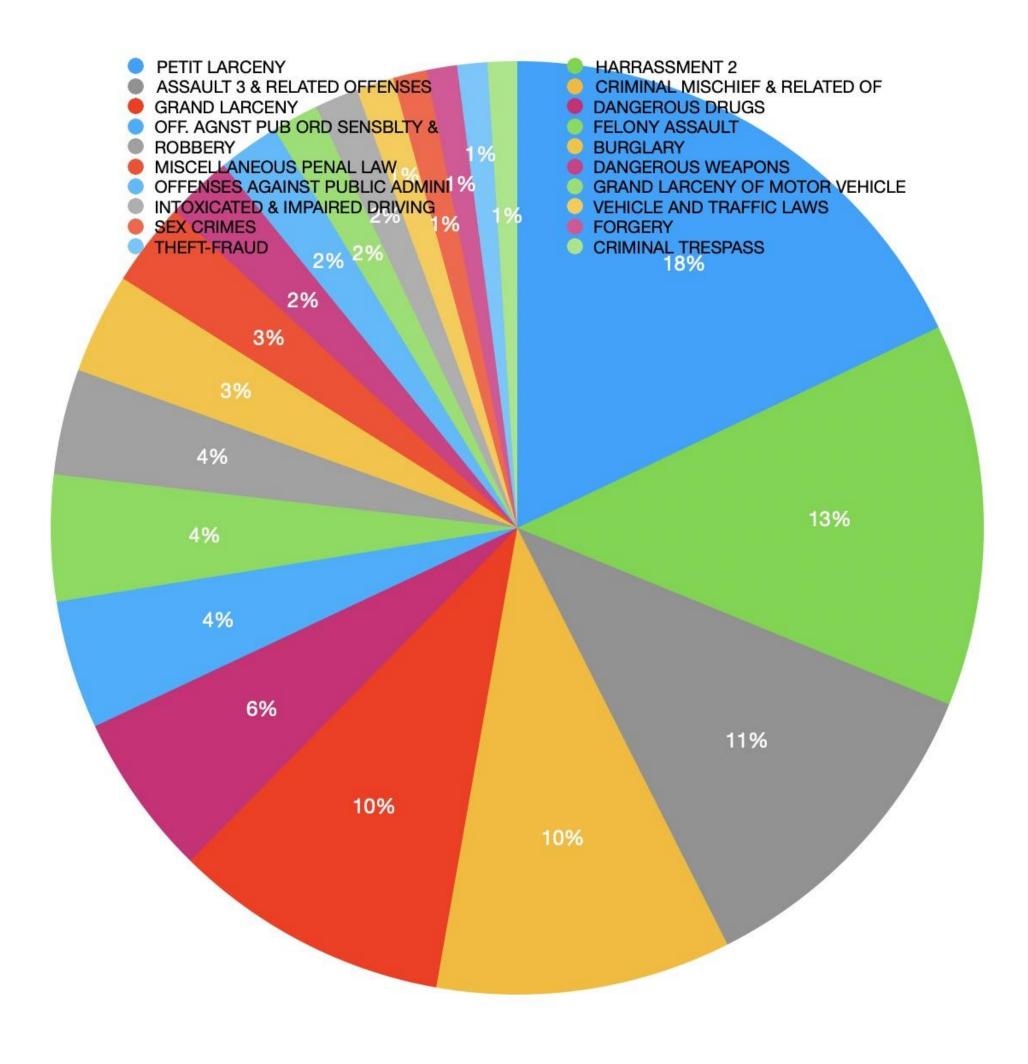
Month		Number
	11	114936
	12	110305
	8	90052
	7	89611
	10	87570
	5	87352
	9	85217
	6	84739
	4	79373
	3	78740
	1	74778
	2	65837

PETIT LARCENY 180246
HARRASSMENT 2 133179
ASSAULT 3 & RELATED OFFENSES 114430
CRIMINAL MISCHIEF & RELATED OF 102771
GRAND LARCENY 96232
DANGEROUS DRUGS 56868
OFF. AGNST PUB ORD SENSBLTY & 44772
FELONY ASSAULT 43921
ROBBERY 36801
BURGLARY 34994
MISCELLANEOUS PENAL LAW 29221
DANGEROUS WEAPONS 22953
OFFENSES AGAINST PUBLIC ADMINI 21353
GRAND LARCENY OF MOTOR VEHICLE 16223
INTOXICATED & IMPAIRED DRIVING 15169
VEHICLE AND TRAFFIC LAWS 13050
SEX CRIMES 11780
FORGERY 10591
THEFT-FRAUD 10472
CRIMINAL TRESPASS 10292

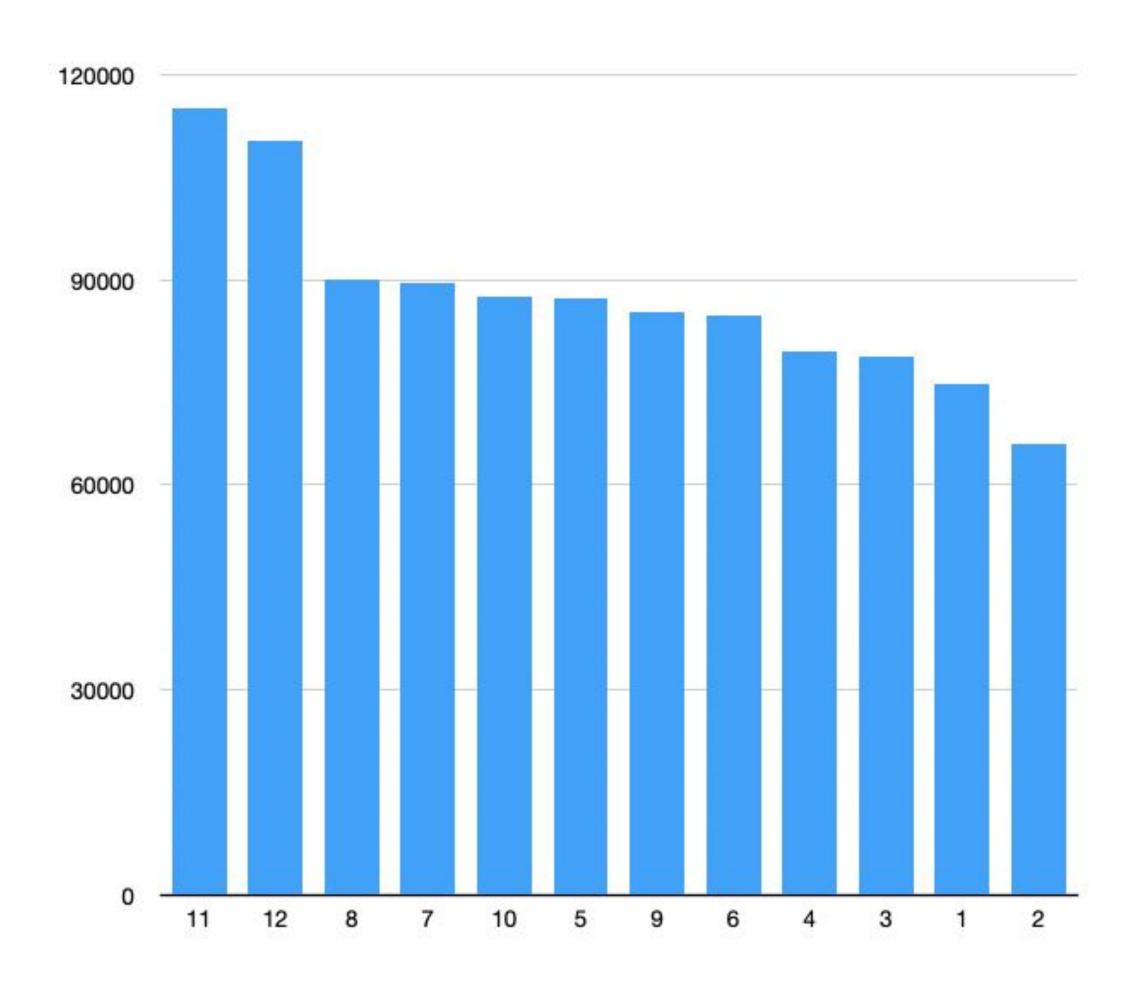
Time	Number
3 PM	62273
6 PM	62200
5 PM	61047
4 PM	60321
7 PM	59530
8 PM	58779
0 PM	58017
2 PM	56111
9 PM	53248
10 PM	50302
0 AM	49875
1 PM	49550
11 PM	45635
11 AM	41351
10 AM	40103
9 AM	38377
1 AM	35901
8 AM	34923
2 AM	29934
3 AM	25246
7 AM	22496
4 AM	21995
6 AM	15702
5 AM	15659

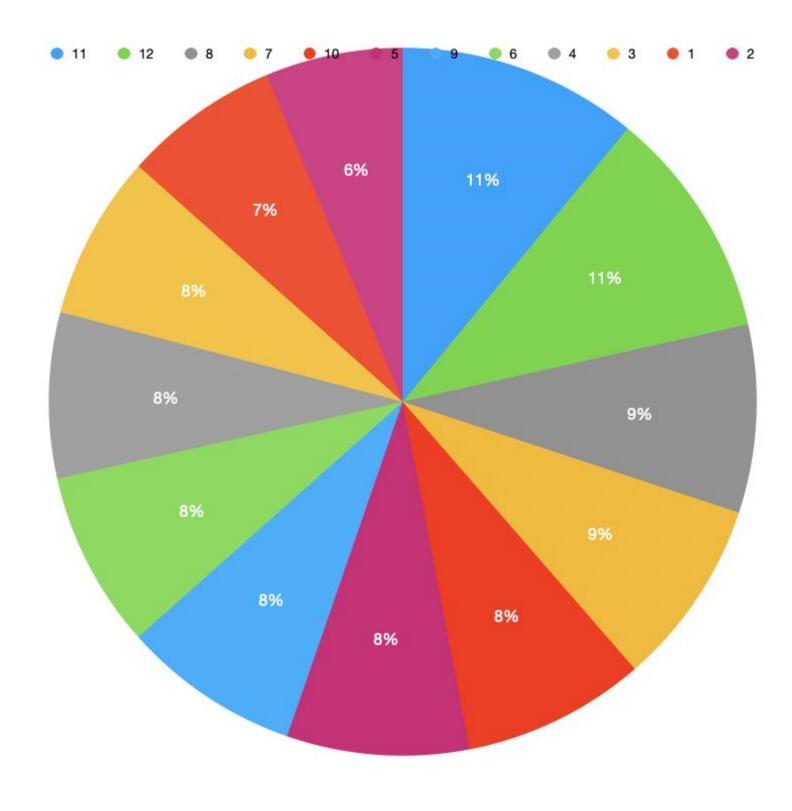
Crime Types Ranking



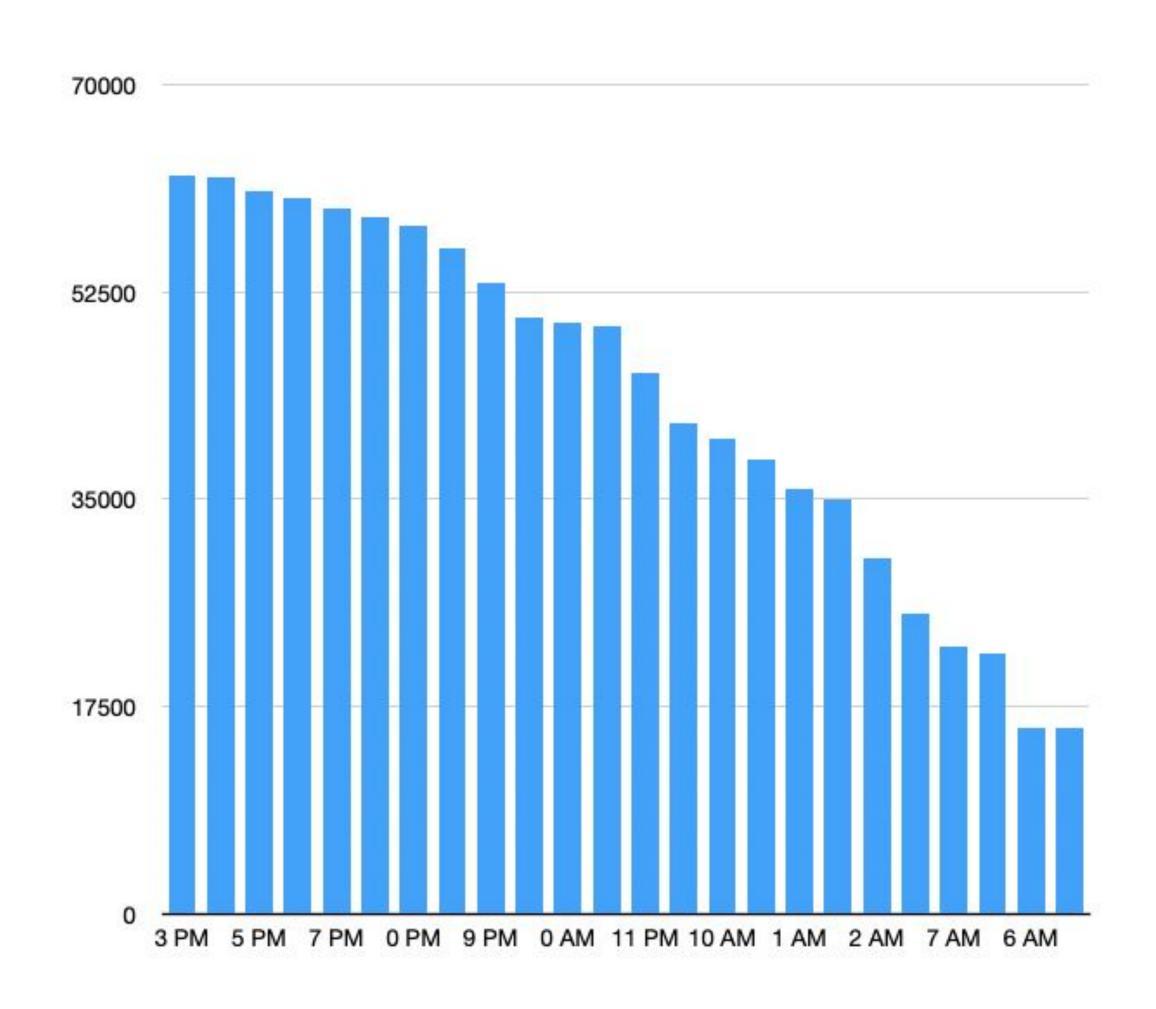


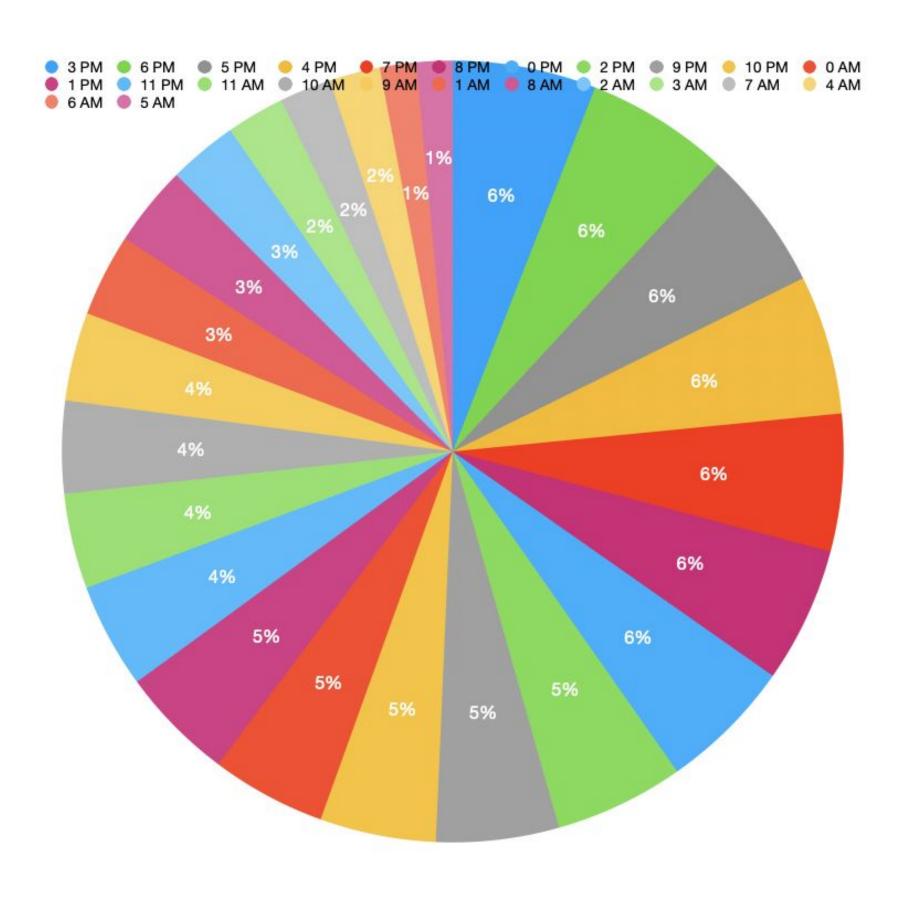
Month Ranking





Time in a day Ranking





INSIGHTS - Crimes in NYC

Crime	More likely to happen in	More likely to happen in
PETIT LARCENY 180246	NOV	4PM
HARRASSMENT 2 133179	NOV	3PM
ASSAULT 3 & RELATED OFFENSES 114430	NOV	9PM
CRIMINAL MISCHIEF & RELATED OF 102771	NOV	12AM
GRAND LARCENY 96232	DEC	12PM
DANGEROUS DRUGS 56868	NOV	8PM
OFF. AGNST PUB ORD SENSBLTY & 44772	NOV	12PM
FELONY ASSAULT 43921	NOV	10PM
ROBBERY 36801	DEC	3PM
BURGLARY 34994	DEC	8AM
MISCELLANEOUS PENAL LAW 29221	NOV	6PM
DANGEROUS WEAPONS 22953	NOV	10PM
OFFENSES AGAINST PUBLIC ADMINI 21353	NOV	8PM
GRAND LARCENY OF MOTOR VEHICLE 16223	NOV	10PM
INTOXICATED & IMPAIRED DRIVING 15169	NOV	3AM
VEHICLE AND TRAFFIC LAWS 13050	NOV	6PM
SEX CRIMES 11780	NOV	12AM
FORGERY 10591	NOV	4PM
THEFT-FRAUD 10472	MAR	12PM
CRIMINAL TRESPASS 10292	NOV	8PM

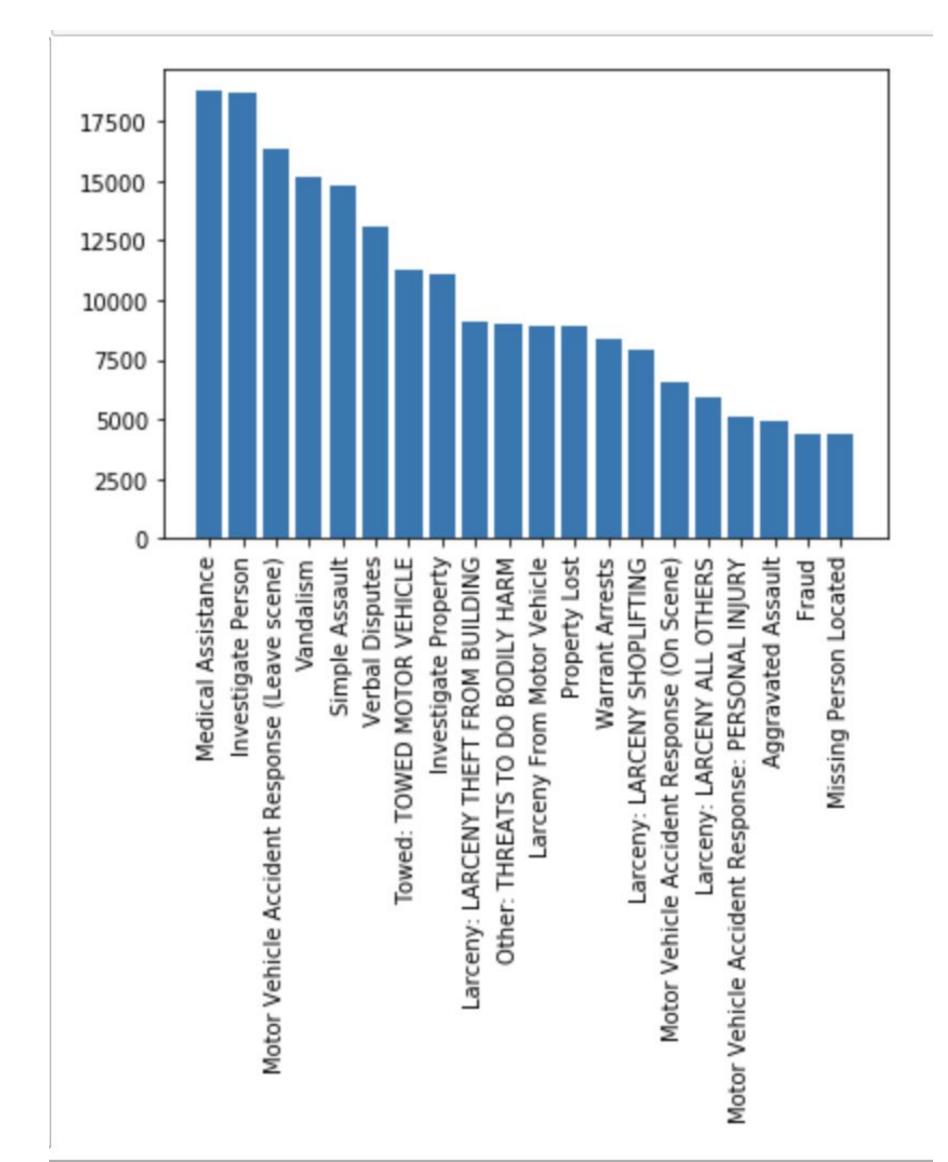
Highlights:

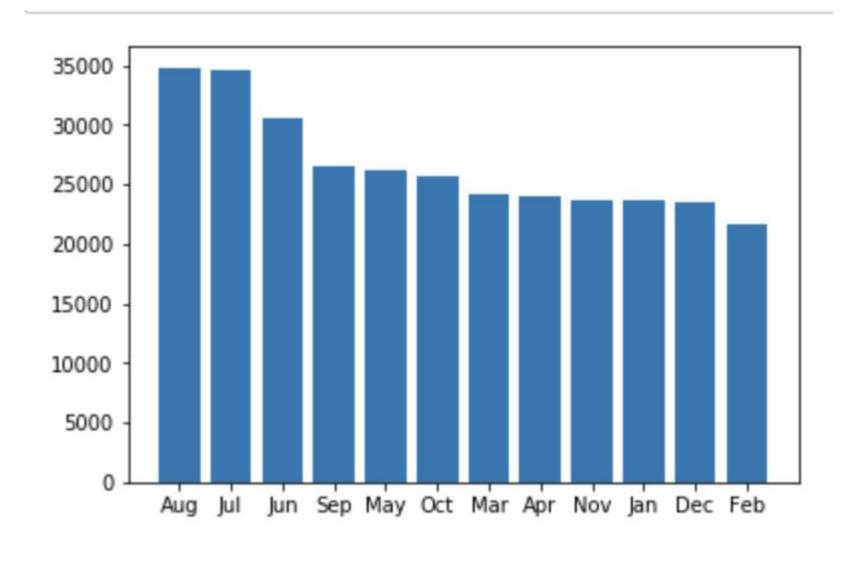
- 1. Police should focus more on Theft-fraud in March.
- 2. Other than petit larceny and harassment that would most likely to happen in the afternoon, other crimes were more likely to happen at night.
- 3. The previous result "highest crime rate at 3 pm" was due to the large base number of petit larceny and harassment.

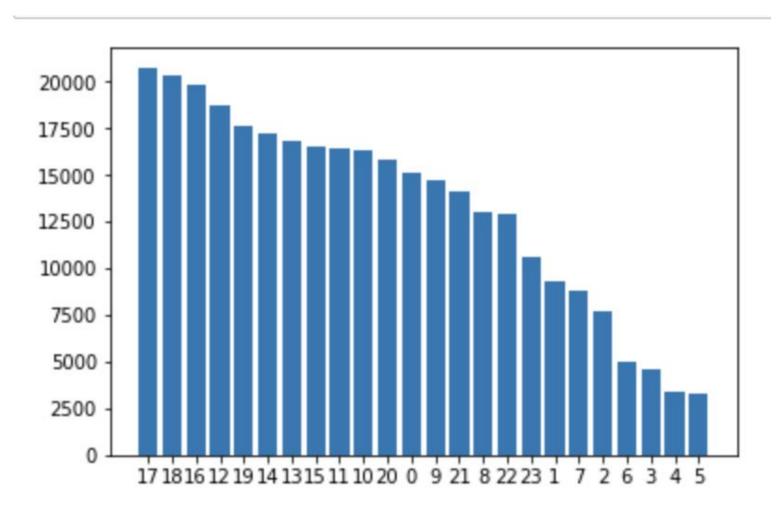
OFNS	FREQ
Medical Assistance: SICK/INJURED/MEDICAL - PERSON	18783
Investigate Person: INVESTIGATE PERSON	18750
Motor Vehicle Accident Response: M/V - LEAVING SCENE - PROPERTY DAMAGE	16323
Vandalism: VANDALISM	15154
Simple Assault: ASSAULT SIMPLE - BATTERY	14791
Verbal Disputes: VERBAL DISPUTE	13099
Towed: TOWED MOTOR VEHICLE	11287
Investigate Property: INVESTIGATE PROPERTY	11124
Larceny: LARCENY THEFT FROM BUILDING	9069
Other: THREATS TO DO BODILY HARM	9042
Larceny From Motor Vehicle: LARCENY THEFT FROM MV - NON-ACCESSORY	8893
Property Lost: PROPERTY - LOST	8892
Warrant Arrests: WARRANT ARREST	8407
Larceny: LARCENY SHOPLIFTING	7949
Motor Vehicle Accident Response: M/V ACCIDENT - PROPERTY DAMAGE	6557
Larceny: LARCENY ALL OTHERS	5963
Motor Vehicle Accident Response: M/V ACCIDENT - PERSONAL INJURY	5131
Aggravated Assault: ASSAULT - AGGRAVATED - BATTERY	4886
Fraud: FRAUD - FALSE PRETENSE / SCHEME	4413
Missing Person Located: MISSING PERSON - LOCATED	4365

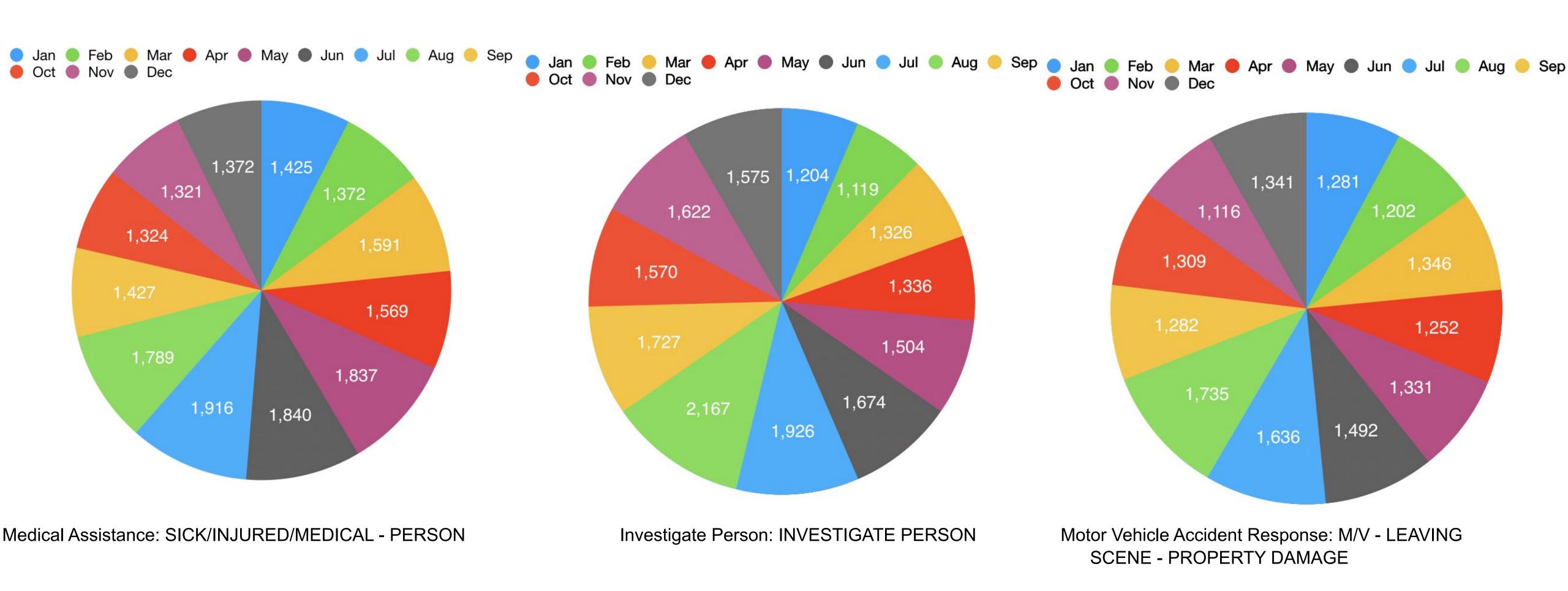
	Time	Freq	
	17	20763	
	18	20302	
	16	19871	
	12	18679	
	19	17588	
	14	17189	
	13	16846	
	15	16522	
	11	16448	
	10	16347	
	20	15850	
	0	15105	
	9	14740	
	21	14111	
	8	13043	
	22	12926	
	23	10596	
	1	9269	
	7	8850	
	2	7693	
	6	5026	
	3	4589	
	4	3408	
	5	3311	
-		The second secon	

Month	Freq
8	34823
7	34556
6	30568
9	26543
5	26199
10	25737
3	24146
4	24086
11	23675
1	23609
12	23477
2	21653









time	freq
12	1087
16	1078
13	1068
14	1057
17	1052
18	1036
19	1030
10	1012
11	998
15	991
21	985
20	981
22	869
9	795
0	772
8	672
23	636
1	635
2	495
7	397
3	313
4	278
6	274
5	272

time	timenum
16	1257
18	1185
17	1178
12	1170
19	1145
14	1119
13	1100
15	1086
11	1022
20	1020
10	954
21	894
9	818
0	766
22	763
23	597
8	592
1	447
7	402
2	391
3	262
6	213
4	200
5	169

time	timenum
17	1180
18	1082
16	1080
19	959
20	882
15	830
21	827
14	801
12	796
13	773
22	771
10	749
9	726
11	713
8	700
23	666
0	552
7	530
2	415
1	392
6	323
3	255
5	179
4	142

OBSTACLES

1. Dataset selection

Kaggle had a considerable amount of datasets on the topic of crimes in cities in the US. When choosing datasets for this project, we went through a lot of test runs to see which datasets best fit the subject we'll be studying, namely enough data on crime type and time.

2. Communication between team members

Due to different time zone of our team members, communication was not as easy. We managed to solve this situation by working independently during day time and exchanging our ideas in morning and evening of when we were both available.

SUMMARY

- The results of our research were that: New York City and Boston had different results in terms of crime types, months of when crimes happen, and time in a day of when crimes happen. These differences proved that police force should be arranged differently due to various characteristics of the cities.
- In New York, crimes happened the most in November and December, whereas in Boston they were July and August.
- In New York the most common crimes were petit larceny and harassment, and in Boston they were medical assistance and investigate person.
- In New York, 3pm was the time with the highest crime rate, and the corresponding time in Boston is 5 pm.
- Conclusion: This project had been designed to provide insight on how to better allocate police resources for most efficient use. A significant amount of city's general fund had been spent on policing. However, the crime rates in big cities were still considerably higher than those of smaller cities. As contributors, it is our ardent hope that the conclusions from this project might act as a guide on police resources allocation optimization.

ACKNOWLEDGEMENTS

- We would like to thank kaggle uploader def love(x) and analyze boston.
- Also, we would like to show our appreciation to the Boston Police Department and the NYC Open Data for the original data source.

REFERENCES

Papers:

https://www.usnews.com/news/national-news/articles/2017-07-07/cities-spend-more-and-more-on-police-is-it-working

https://www.jstor.org/stable/10.1086/250109?seq=1#metadata_info_tab_contents

Data:

https://www.kaggle.com/ADAMSCHROEDER/CRIMES-NEW-YORK-CITY?SELECT=NYPD_COMPLAINT_DATA_HISTORIC.

https://www.kaggle.com/AnalyzeBoston/crimes-in-boston

Image:

https://boston.eater.com/2020/3/15/21180625/covid-19-boston-restaurants-regulations-reduced-capacity

https://ny.curbed.com/2019/12/19/21024869/new-york-city-neighborhoods-photos-before-after

THANK YOU FOR WATCHING!