Milestone 1: Data Visualization:
"NYC: Does Crime Ever Sleep in the City That Never Does?"

Mahdi Fourati 313078,
Elsa Heitz 316735,
Benoit Mathey-Doret dit Doret 324620

March 2025

1 Dataset Description

We selected the *NYC Crime Complaint Dataset* dataset, publicly available on Kaggle:[1]. This dataset is already well-maintained, containing only very few missing values, thus requiring minimal preprocessing, primarily related to date formatting.

The dataset comprises **8,914,838 data points** with **35 features** detailed in Table 1. We chose this dataset for its completeness and we think we will be able to find very interesting visualizations such as maps, demographic comparisons across gender of suspects and victims, or even temporal contextualization to help us answer our problematic.

by associating each crime incident with its corresponding community district. This was achieved by mapping the geographic coordinates of each incident to the official district boundaries provided by the NYC Department of City Planning webpage [2]. In order to analyze crime in NYC we want to study the social economic situation of the city and therefore created another dataset grouping many information from *CCC New York Data*:[3] such as the unemployment rate by district and the percentage of people under the Federal Poverty Level (FPL).

2 Problematic

New York City is one of the most dynamic and cosmopolitan city in the world. As all big cities in the world, it faces significant challenges when it comes to public safety. Crime is often deeply related to social-economic conditions, with neighborhoods affected differently based on factors such as unemployment, poverty rates, and demographic composition. Although higher poverty levels are often assumed to lead to higher criminality, this is not always the case. A neighborhood with high poverty rates on paper may experience lower crime levels if, for instance, organized crime groups exert control over the area, limiting crimes to drug-related activities. Understanding these relationships is crucial for designing effective crime prevention policies and improving urban safety.

In our study, our aim is to examine the relationship between the NYC Crime Complaint Dataset, which contains nearly 9 million crime records, with socio-economic data from CCC New York to analyze how crime trends vary across the city's districts. By doing so, we hope to uncover the root causes of crime disparities and explore the broader social factors influencing these patterns: Is there a correlation between crime rates and poverty levels across the different NYC districts? Are economically disadvantaged areas disproportionately affected by specific types of crime? Is crime more prevalent in neighborhoods adjacent to poor areas? Does unemployment contribute to increased crime within neighborhoods? Over time, does poverty evolve in the same way as crime? Are there clear trends in crime patterns based on seasons, days of the week, or hours of the day?

---

[1] https://www.kaggle.com/datasets/aniket0712/nypd-complaint-data-historic?resource=download
[2] https://www.nyc.gov/site/planning/data-maps/open-data/districts-download-metadata.page
[3] https://data.cccnewyork.org/data/download#0,8/66

Our target audience includes New York City residents seeking information about crime in their neighborhoods, as well as visitors or potential residents. Additionally, this analysis could help New York State authorities assess issues and identify areas for improvement. As first ideas of nice visualizations to show our results, we are thinking about creating a nice interactive map of the city. We could therefore visualize the districts and have an overview of the proportion of crimes by ditrict as well as social-economic information. Then when clicking on the district we could get a closer look and display specific crime locations. There, we could study different trends and uncover specific areas where crimes aggregate and link it to social-economic explanations.

3 Exploratory Data Analysis

The complete Exploratory Data Analysis can be found in the Jupyter notebook available on GitHub. However, here are some of our most interesting findings.
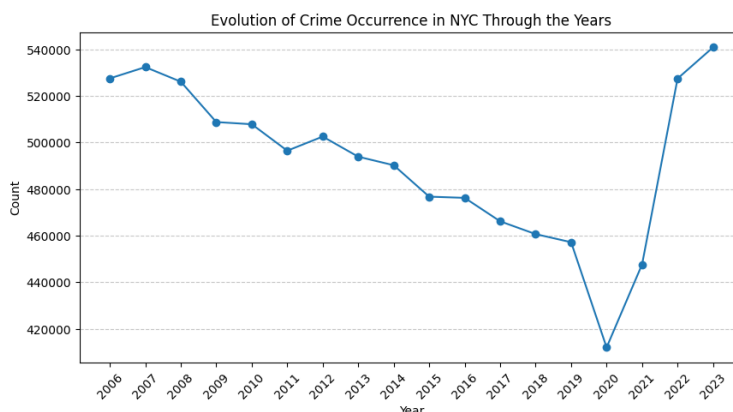

Figure 1: Evolution of crime in New York since 2006

First, by analyzing the dates of reported incidents, we observed on Figure 1 the evolution of crime complaints since 2006. Interestingly, crime steadily declined until 2020, reaching its lowest level during the COVID-19 pandemic. However, post-pandemic, crime increased past pre-COVID levels, likely due to socioeconomic problems caused by the pandemic and law enforcement challenges, such as staff shortages and NYPD controversies, warranting further investigation.

Then, observing the mean number of incidents per year across different districts, we find that crime is, as expected, is unevenly distributed across NYC, with the Midtown Business District experiencing the highest average (17,897 crimes/year) and Bayside the lowest (3,197 crimes/year). However, crime density alone does not fully reflect danger levels. One key factor to consider is the severity of offenses. A district may experience lots of minor infractions, such as petit larceny (low-value thefts), while another district with lower crime counts might face more violent offenses. Additionally, population density and foot traffic significantly influence crime counts.

Additionally, a statistically significant Spearman correlation (0.453, p = 0.000318) in Figure 2 indicates a positive association between crime and poverty levels. This suggests that higher poverty levels tend to correspond with higher crime rates, though the relationship may not be strictly linear. The significance of the correlation implies that this pattern is unlikely to be due to chance. However, correlation does not imply causation, other socioeconomic factors, such as unemployment, education, and population density, may also influence crime rates. Further analysis will be needed in order to make any conclusions.

Finally, we explored additional avenues of analysis, such as examining the distribution of crime types in relation to the race and age of suspects. These findings could be worth further investigation. In addition, we analyzed the distribution of crime types in different districts to identify potential disparities and establish connections with our previous work on disadvantaged neighborhoods. Figure 3 presents a heatmap illustrating this distribution. All additional exploring covered here can be found in our git hub repository.
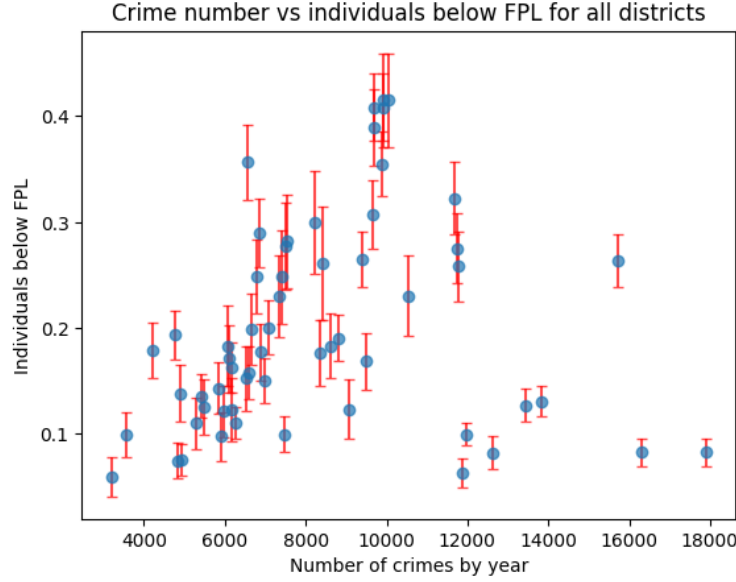
Figure 2: Scatter plot showcasing the distribution of crimes with respect to districts
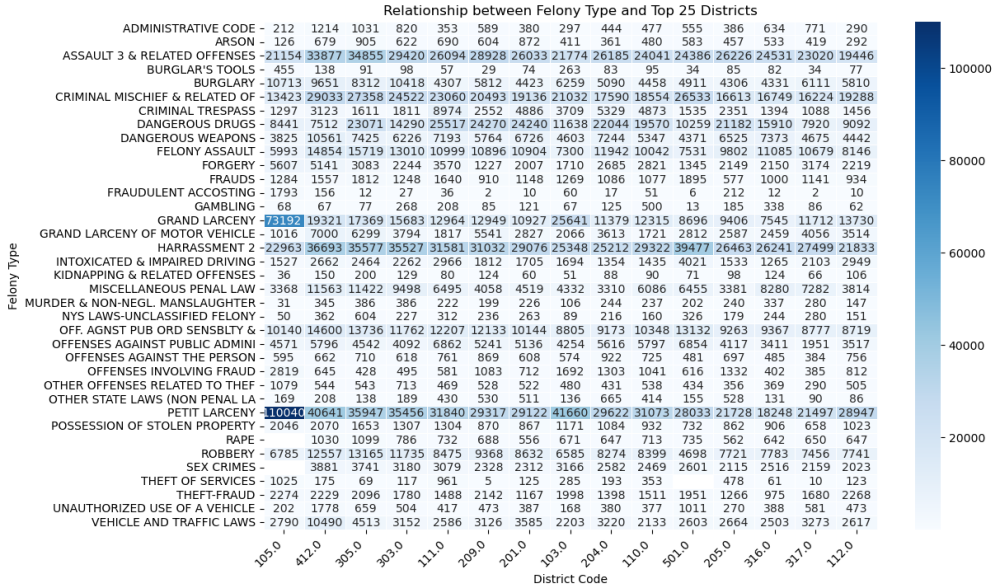

Relationship between Felony Type and Top 25 Districts

Felony Type	105.0	412.0	305.0	303.0	111.0	209.0	201.0	103.0	204.0	110.0	501.0	205.0	316.0	317.0	112.0
ADMINISTRATIVE CODE	212	1214	1031	820	353	589	380	297	444	477	555	386	634	771	290
ARSON	126	679	905	622	690	604	872	411	361	480	583	457	533	419	292
ASSAULT 3 & RELATED OFFENSES	21154	33877	34855	29420	26094	28928	26033	21774	26185	24041	24386	26226	24531	23020	19446
BURGLAR'S TOOLS	455	138	91	98	57	29	74	263	83	95	34	85	82	34	77
BURGLARY	10713	9651	8312	10418	4307	5812	4423	6259	5090	4458	4911	4306	4331	6111	5810
CRIMINAL MISCHIEF & RELATED OF	13423	29033	27358	24522	23060	20493	19136	21032	17590	18554	26533	16613	16749	16224	19288
CRIMINAL TRESPASS	1297	3123	1611	1811	8974	2552	4886	3709	5329	4873	1535	2351	1394	1088	1456
DANGEROUS DRUGS	8441	7512	23071	14290	25517	24270	24240	11638	22044	19570	10259	21182	15910	7920	9092
DANGEROUS WEAPONS	3825	10561	7425	6226	7193	5764	6726	4603	7244	5347	4371	6525	7373	4675	4442
FELONY ASSAULT	5993	14854	15719	13010	10999	10896	10904	7300	11942	10042	7531	9802	11085	10679	8146
FORGERY	5607	5141	3083	2244	3570	1227	2007	1710	2685	2821	1345	2149	2150	3174	2219
FRAUDS	1284	1557	1812	1248	1640	910	1148	1269	1086	1077	1895	577	1000	1141	934
FRAUDULENT ACCOSTING	1793	156	12	27	36	2	10	60	17	51	6	212	12	2	10
GAMBLING	68	67	77	268	208	85	121	67	125	500	13	185	338	86	62
GRAND LARCENY	73192	19321	17369	15683	12964	12949	10927	25641	11379	12315	8696	9406	7545	11712	13730
GRAND LARCENY OF MOTOR VEHICLE	1016	7000	6299	3794	1817	5541	2827	2066	3613	1721	2812	2587	2459	4056	3514
HARRASSMENT 2	22963	36693	35577	35527	31581	31032	29076	25348	25212	29322	39477	26463	26241	27499	21833
INTOXICATED & IMPAIRED DRIVING	1527	2662	2464	2262	2966	1812	1705	1694	1354	1435	4021	1533	1265	2103	2949
KIDNAPPING & RELATED OFFENSES	36	150	200	129	80	124	60	51	88	90	71	98	124	66	106
MISCELLANEOUS PENAL LAW	3368	11563	11422	9498	6495	4058	4519	4332	3310	6086	6455	3381	8280	7282	3814
MURDER & NON-NEGL. MANSLAUGHTER	31	345	386	386	222	199	226	106	244	237	202	240	337	280	147
NYS LAWS-UNCLASSIFIED FELONY	50	362	604	227	312	236	263	89	216	160	326	179	244	280	151
OFF. AGNST PUB ORD SENSBLTY &	10140	14600	13736	11762	12207	12133	10144	8805	9173	10348	13132	9263	9367	8777	8719
OFFENSES AGAINST PUBLIC ADMINI	4571	5796	4542	4092	6862	5241	5136	4254	5616	5797	6854	4117	3411	1951	3517
OFFENSES AGAINST THE PERSON	595	662	710	618	761	869	608	574	922	725	481	697	485	384	756
OFFENSES INVOLVING FRAUD	2819	645	428	495	581	1083	712	1692	1303	1041	616	1332	402	385	812
OTHER OFFENSES RELATED TO THEF	1079	544	543	713	469	528	522	480	431	538	434	356	369	290	505
OTHER STATE LAWS (NON PENAL LA	169	208	138	189	430	530	511	136	665	414	155	528	131	90	86
PETIT LARCENY	100646	40641	35947	35456	31840	29317	29122	41660	29622	31073	28033	21728	18248	21497	28947
POSSESSION OF STOLEN PROPERTY	2046	2070	1653	1307	1304	870	867	1171	1084	932	732	862	906	658	1023
RAPE		1030	1099	786	732	688	556	671	647	713	735	562	642	650	647
ROBBERY	6785	12557	13165	11735	8475	9368	8632	6585	8274	8399	4698	7721	7783	7456	7741
SEX CRIMES		3881	3741	3180	3079	2328	2312	3166	2582	2469	2601	2115	2516	2159	2023
THEFT OF SERVICES	1025	175	69	117	961	5	125	285	193	353		478	61	10	123
THEFT-FRAUD	2274	2229	2096	1780	1488	2142	1167	1998	1398	1511	1951	1266	975	1680	2268
UNAUTHORIZED USE OF A VEHICLE	202	1778	659	504	417	473	387	168	380	377	1011	270	388	581	473
VEHICLE AND TRAFFIC LAWS	2790	10490	4513	3152	2586	3126	3585	2203	3220	2133	2603	2664	2503	3273	2617

District Code

Figure 3: Heatmap representing the distribution of crime types across districts.

4 Related work

There exist some articles comparing the dangerousness of the different boroughs [3] with standard statistics and static visualizations. We can also find crime maps [2] [1] but their design is not very appealing and they do not deliver an analysis of the situation like we aim to do. They only deliver location of crimes.

This makes our approach particularly original: rather than simply displaying crime locations, we combine spatial, temporal, and socio-economic data to provide a deeper, contextualized analysis. Our goal is not only to visualize where crimes occur, but to understand why they occur — uncovering patterns, correlations, and potential causes. By integrating interactive maps with layered data and allowing users to explore trends dynamically, we offer a more insightful and user-centered experience than existing static tools.

References

[1] New york city crime data - neighborhoodscout. `https://www.neighborhoodscout.com/ny/new-york/crime`. Accessed: 2025-03-10.

[2] New york city crime map. `https://www.arcgis.com/apps/instant/sidebar/index.html?appid=8153f961507040de8dbf9a53145f18c4`. Accessed: 2025-03-10.

[3] Jingjing Ge. Analysis of crime data in new york city in 2020. `https://jingjing-ge.medium.com/analysis-of-crime-data-in-new-york-city-in-2020-e6a65edf4429`, 2020. Accessed: 2025-03-10.

A   Appendix

Table 1: Description of NYPD Complaints Dataset Features

Column Name	Description
CMPLNT_NUM	Unique identifier for each complaint
CMPLNT_FR_DT	Start date of the incident
CMPLNT_FR_TM	Start time of the incident
CMPLNT_TO_DT	End date of the incident (if applicable)
CMPLNT_TO_TM	End time of the incident (if applicable)
ADDR_PCT_CD	Precinct where complaint was reported
RPT_DT	Official date of complaint recording
KY_CD	Numeric offense code
OFNS_DESC	Description of the offense
PD_CD	Police department-specific offense code
PD_DESC	Description of police offense classification
CRM_ATPT_CPTD_CD	Indicator if crime was completed or attempted
LAW_CAT_CD	Severity of the offense (Felony, Misdemeanor, Violation)
BORO_NM	Borough where crime occurred (e.g., Manhattan, Brooklyn)
LOC_OF_OCCUR_DESC	Description of the location type (e.g., street, residence)
PREM_TYP_DESC	Type of premises (e.g., apartment, store)
JURIS_DESC	Jurisdiction responsible for the case
JURISDICTION_CODE	Numeric jurisdiction code
PARKS_NM	Park name (if applicable)
HADEVELOPT	NYC Housing Authority development name
HOUSING_PSA	Housing Police Service Area
X_COORD_CD	X-coordinate in NYC coordinate system
Y_COORD_CD	Y-coordinate in NYC coordinate system
SUSP_AGE_GROUP	Age group of suspect
SUSP_RACE	Race of suspect
SUSP_SEX	Gender of suspect
TRANSIT_DISTRICT	Transit district (if applicable)
Latitude	Latitude of crime location
Longitude	Longitude of crime location
Lat_Lon	Combined latitude and longitude
PATROL_BORO	Patrol borough where incident occurred
STATION_NAME	Subway station name (if applicable)
VIC_AGE_GROUP	Age group of victim
VIC_RACE	Race of victim
VIC_SEX	Gender of victim