



SAPIENZA
UNIVERSITÀ DI ROMA



POLE CRIMES

Objective:

We will employ [Neo4J](#) to query a comprehensive crime dataset, aiming to extract both quantitative and qualitative insights. These insights will assist investigative units, law enforcement agencies, and any institutions or companies interested in studying social phenomena for forensic sciences.

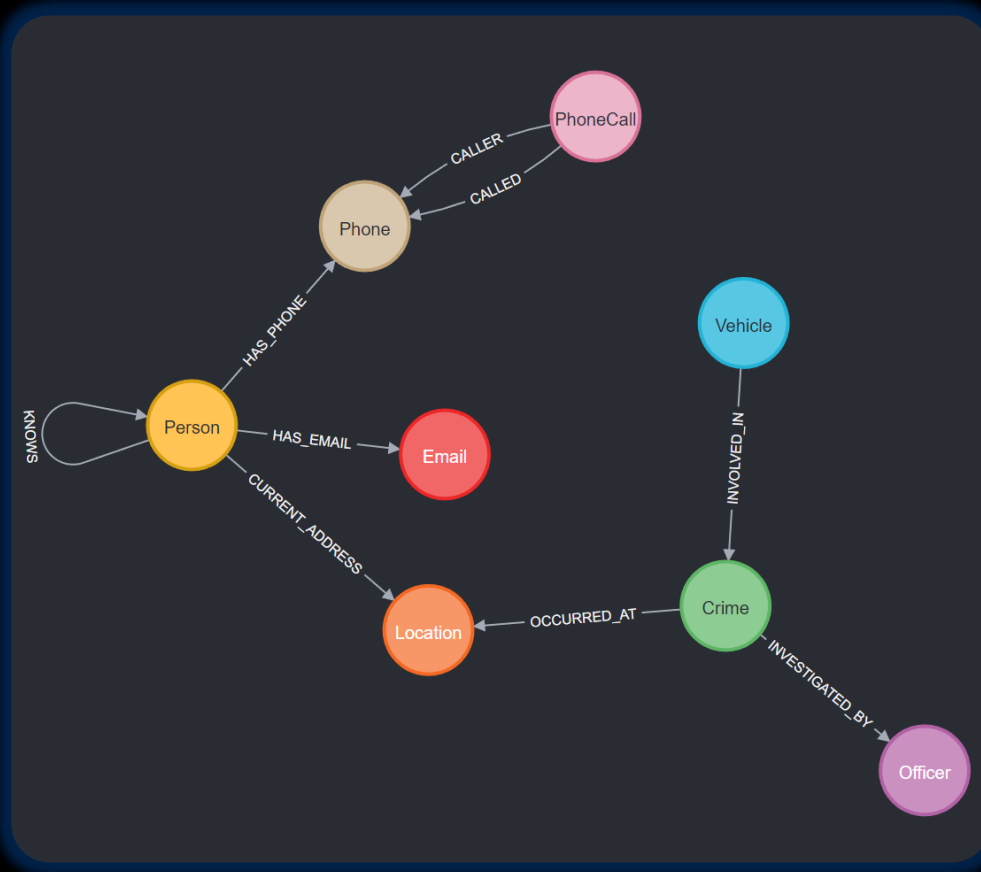
Potential Benefits:

- Enhance forensic analysis capabilities.
- Support crime prevention and resolution strategies.
- Facilitate the understanding of complex social dynamics in urban environments.

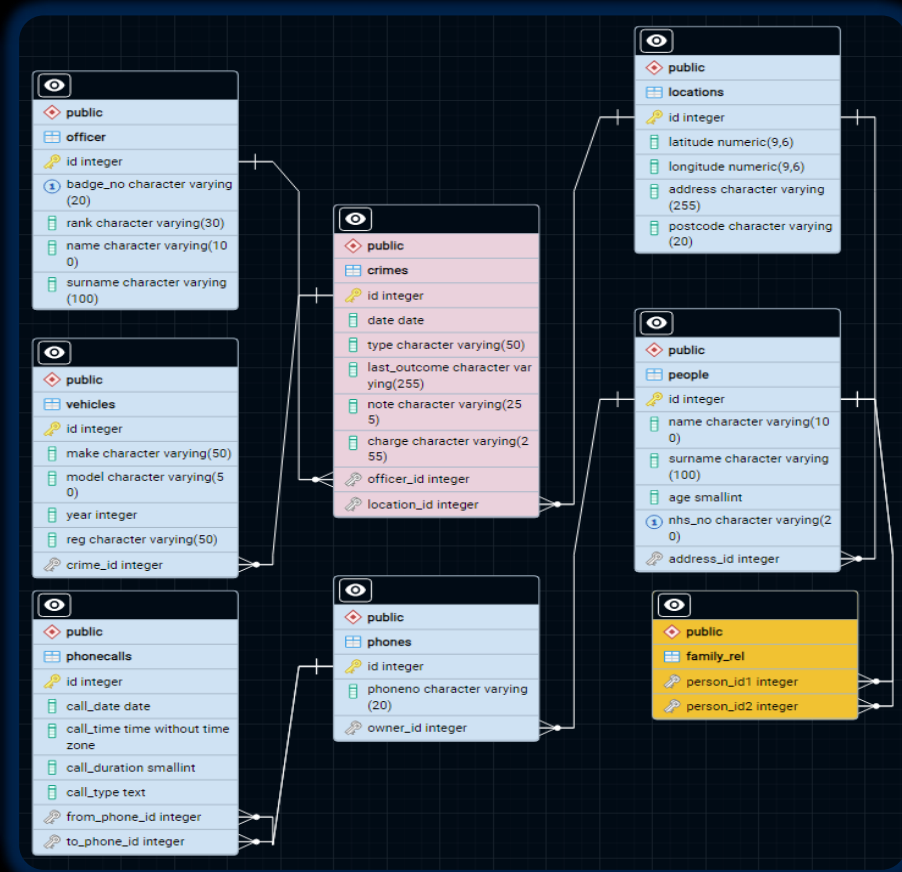
Dataset Overview:

- **Source:**
The dataset is provided by the British government.
- **Contents:**
It includes data related to crimes, phone records, individuals, and geographic areas specific to the city of Manchester during August 2017.
- **Analytical Framework:**
The data will be analyzed using the [POLE](#) (Person, Object, Location, Event) framework. This approach is intended to identify recurring patterns and potentially provide predictive deductions.

GRAPH SCHEMA



RELATIONAL SCHEMA



Why a Graph DB

- **Effective Handling of Relationships**
 - Directly references adjacent nodes for fast traversal.
 - Perfect for the analysis of criminal data in the POLE framework (it is naturally a network).
 - **Flexible Data Model**
 - Easy adaptation and expansion.
 - Dynamic relationship creation and querying.
 - **Advanced Query Capabilities**
 - Cypher query language for graph traversal and pattern matching.
 - Simplifies complex queries to uncover insights (readability).
 - **Performance**
 - Scales well for large datasets.
 - Supports real-time analysis and querying.
 - **Visualization and Insights**
 - Integrates with graph visualization tools.
 - Intuitive representation of entities and interactions.
-

Data analysis – Query 1

Task:

Find police officers linked to over 30 unique crime investigations, and list these officers' surnames, badge numbers, and ranks, sorting the results by badge number.

Interpretation:

The intent is to find the busiest officers (who have several cases to solve) and identify any causal relationships with their seniority, sorting the results by badge number.

surname	badge_no	rank
Bagby	00-6011813	Sergeant
Loveridge	01-0363085	Inspector
Lumsdaine	01-5764299	Police Constable
Reilinger	02-3468866	Sergeant
Chesnut	02-6192850	Police Constable
Fonte	03-1601856	Sergeant
Assard	03-2149839	Police Constable
Tams	03-5878006	Inspector
Aykroyd	03-6336692	Police Constable
Ellwell	03-6377521	Police Constable

Query complete: ER 2.769 seconds

Data analysis – Query 2

Task:

Identify crime hotspots by postcode, aggregating the number of crimes.

Interpretation:

By considering each area identified by a postcode, we pinpoint the most critical areas in terms of recorded criminal activity in the city, in order to then be able to propose an appropriate strategy for the deployment of anti-crime resources.

postcode	crimecount
M1 1LU	166
M60 1TA	111
M60 9AH	48
M4 3AL	46
M1 3LZ	41
M90 2AY	38
WN7 5SJ	38
M4 2BS	36
M3 1DA	35
OL1 1QN	35

Query complete: ER 0.146 seconds

Data analysis – Query 3

Task:

Analyze the evolution of crime types over different areas it calculates the absolute daily change in crime counts from one day to the next instead of growth rates.

This change measures the volatility without considering the direction of the change (increase or decrease).

area	avg_daily_change
M1 1LU	2.71
M60 1TA	1.87
M1 3EZ	1.4
WN7 5SJ	1.35
M1 3LZ	1.32
OL1 1QN	1.26
OL6 8RT	1.2
OL12 6TS	1.13
M60 9AH	1.08
M3 7JY	1

Interpretation:

By analyzing the volatility of the areas crime rate, we can identify areas for which we have a poor ability to predict criminal activity. Conversely, it is possible to analyze those areas for which the crime rate remains rather unchanged over time (whether low or high).

Data analysis – Query 4

Task:

Find the usage frequency for each vehicle type for the "vehicle crimes".

Interpretation:

We try to establish greater control over those types of vehicles that are generally more involved in crimes, perhaps because they are of particular interest to thieves or because they are suitable for carrying out criminal tasks such as drug dealing or robbery.

vehiclebrand	vehiclemodel	vehiclecount	crimecount
Ford	Mustang	10	10
Pontiac	Bonneville	9	9
Volkswagen	Passat	8	8
Hyundai	Sonata	8	8
Ford	E-Series	7	7
Mercury	Grand Marquis	7	7
Ford	Expedition	7	7
Chrysler	Town & Country	7	7
Pontiac	Grand Prix	7	7
Nissan	Pathfinder	6	6

Data analysis – Query 5

Task:

For each type of crime type evaluate the standard deviation of the distance between the locations of the related crimes in order to understand if crimes are pretty concentrated or spread out in different areas of the city. Also report the total number of areas involved in each type of crime.

Interpretation:

For each type of crime, it may be interesting to understand how uniformly the number of crimes is distributed over the urban area. We also report the number of areas involved so that the geographical extent (in terms of square footage) can be normalised by the number of areas over which it is defined.

crimetype	centroidlatitude	centroidlongitude	stdlatitude_km	stdlongitude_km	differentarea
Violence and sexual offences	53.50694	-2.27424	7.03	16.66	5515
Public order	53.50586	-2.27196	6.75	15.66	3427
Criminal damage and arson	53.50518	-2.27893	7.04	17.04	2864
Burglary	53.50044	-2.27709	7.47	17.31	2404
Vehicle crime	53.49939	-2.28067	7	17.19	2200
Other theft	53.50037	-2.27841	7	16.36	1694
Shoplifting	53.502	-2.27594	7.23	16.41	641
Other crime	53.5132	-2.29014	6.54	17.57	578
Robbery	53.48915	-2.2455	5.96	12.34	447
Bicycle theft	53.4787	-2.25895	6.11	12.43	355

Query complete: ER 0.363 seconds

Data analysis – Query 6

Task:

Identify people that possibly had crime related phone calls.

This query, for a specific query crime, returns people that have had (as a caller or a called) a lot of calls with people that do not belong to their family (order by "*No_fam_calls*" counter), in the period of 10 days before the crime and that live in the same area ("*post_code*") where the crime occurred.

Interpretation:

For a specific crime, we report a forensic analysis of the telephone traffic for people geographically linked to the crime.

In particular, we focus on calls in the days preceding the crime, made between individuals not belonging to the same family.

person	nhs_number	phone_number	no_fam_calls
Phillip Perry	884-33-9676	0-(106)592-7420	4
Deborah Ford	838-45-9343	5-(607)134-1574	3
Harry Lopez	915-75-5600	1-(776)710-5989	2
Jonathan Hunt	298-36-6290	2-(527)185-1019	2
Peter Bryant	245-63-7539	6-(184)813-2011	1

Data analysis – Query 7

Task:

For each officer, return the number of cases to which it has been assigned which last_outcome="Under investigation" as "Num_unresolved", and the most frequent type of crime to which it has been assigned. Order the result by "Num_unresolved".

Interpretation:

When considering the assignment of new cases, it may be useful to analyze the officers by number of unsolved cases currently assigned and 'area of specialisation' (i.e. the most frequent type of crime they have worked on).

officerbadge	officerspecialization	totalnumunresolved
55-5623906	Violence and sexual offences	11
21-8583779	Violence and sexual offences	10
60-2236534	Violence and sexual offences	10
09-6971653	Violence and sexual offences	10
90-5774272	Violence and sexual offences	10
30-4157186	Violence and sexual offences	9
21-9479736	Violence and sexual offences	9
32-8274663	Violence and sexual offences	9
21-3341439	Violence and sexual offences	9
48-0216838	Violence and sexual offences	9

Data analysis – Query 8

Task:

For a given query date, return the total number of crimes occurred on that day for each area (post_code), the total number of people living in each area and the total number of calls made by the inhabitants on that day.

Interpretation:

For each day in each area, we find patterns between crime levels, population density of the area and telephone traffic between inhabitants of the area.

area	totalcrimes	totalpeople	totaltraffic
M12 4GJ	9	2	0
BL1 1LE	8	1	4
SK5 8BZ	6	3	18
SK1 4HU	6	1	8
SK3 9EU	6	1	8
BL3 5NJ	6	2	4
OL8 2RB	6	1	0
BL6 6RL	5	1	8
M4 7EU	5	2	4
M8 8YT	5	2	4

Data analysis – Query 9

Task:

For each Crime Type return the most common outcome, its frequency, the total number of cases and the percentage of the most common outcome over all the total cases.

Interpretation:

It might be interesting to analyze the legal outcome for each type of crime in order to identify any bias in the legal system itself.

crimetype	mostcommonoutcome	frequency	totalcases	percentage
Violence and sexual offences	Unable to prosecute suspect	3263	8765	37.23
Public order	Investigation complete; no suspect identified	2773	4839	57.31
Criminal damage and arson	Investigation complete; no suspect identified	2681	3587	74.74
Burglary	Investigation complete; no suspect identified	2468	2807	87.92
Vehicle crime	Investigation complete; no suspect identified	2388	2598	91.92
Other theft	Investigation complete; no suspect identified	1693	2140	79.11
Shoplifting	Investigation complete; no suspect identified	873	1427	61.18
Other crime	Investigation complete; no suspect identified	275	651	42.24

Data analysis – Query 10

Task:

This query aggregates phone communication data for individuals, calculating metrics like the number of distinct phones per person, total and average call durations, total different calls, and the number of unique call days.

Interpretation:

The intent is to provide support to law enforcement agencies in analyzing telephone traffic for each individual, reporting on key factors such as the distribution of telephone traffic.

personid	numdistinctphones	totalcalltime	avgcallduration	totcalls	distinctcalldate
2	1	190	47.5	4	3
9	1	182	91	2	2
13	1	168	84	2	2
16	1	234	58.5	4	4
20	1	132	66	2	2
27	1	98	49	2	2
35	1	116	58	2	2
39	1	272	68	4	4
43	1	158	79	2	2
47	1	250	62.5	4	4

Thanks!



Enrico Grimaldi - 1884443



[Engrima18](#)



Mario Edoardo Pandolfo - 1835189



[JRhin](#)



[Check-out our code!](#)