Crime Analysis utilizing LLMs

A presentation by:

Siji Chen

Prateek Kumar Padhy

Jayanth Koripalli

Motivation

- Addressing Urban Crime
- Improving Public Safety
- Community Engagement and Empowerment
- Innovative Use of Technology
- Proactive Crime Prevention

Problem Statement

- Rising crime rates.
- No proper way to predict crimes.
- Limitations of Traditional Crime Analysis.

Proposed Approaches

- Crime Hotspots
- Hotspot Map
- DC Background

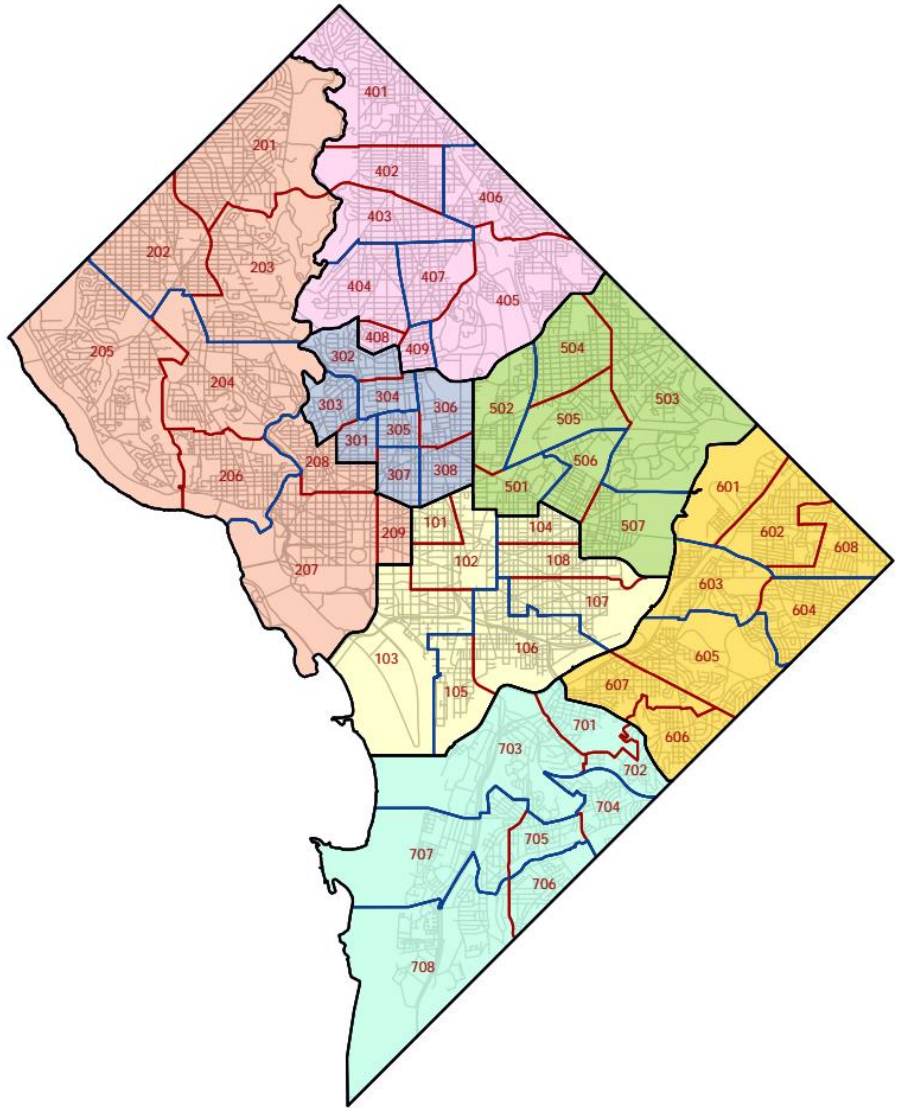
Crime Hot Spots

- Crime hot spots are crime concentrated area.
  - A hot spot is an area that has a greater than average number of criminal or disorder events
  - people have a higher than average risk of victimization
  - Cool spot
  - some hot spots may be hotter than others
- Types of hot spots
  - Repeat places hot spots
  - Repeat victimization hot spots
  - Repeat streets hot spots
  - Neighborhoods and other area hot spots

Hot Spot Map

- Dot maps
- Line maps: When the hot spots are along streets, point maps and area maps are of far less utility than line maps.
- Ellipse, choropleth, and isoline maps

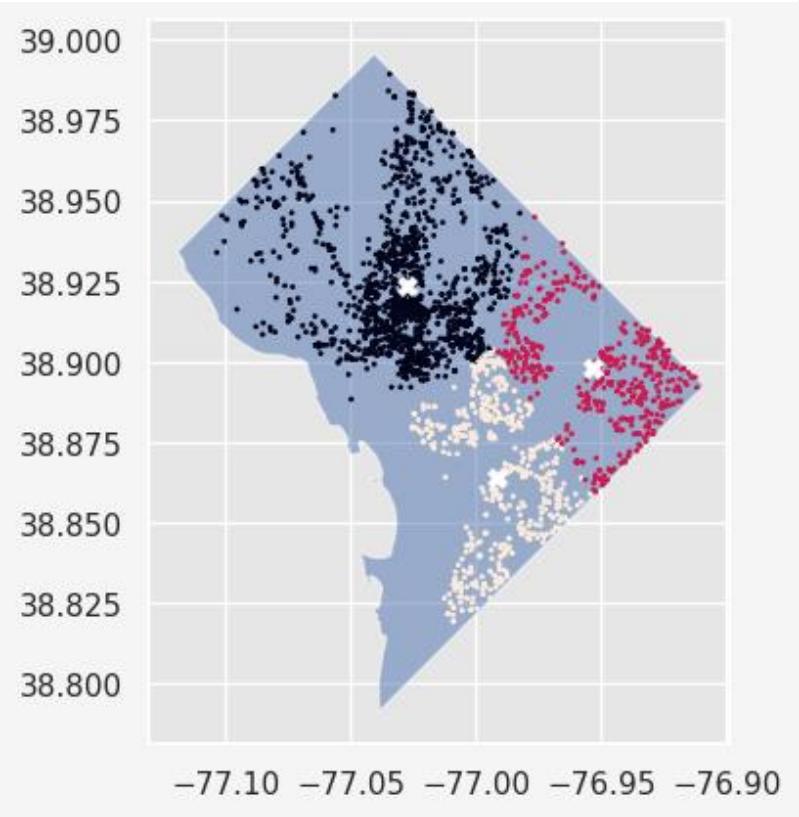
DC background



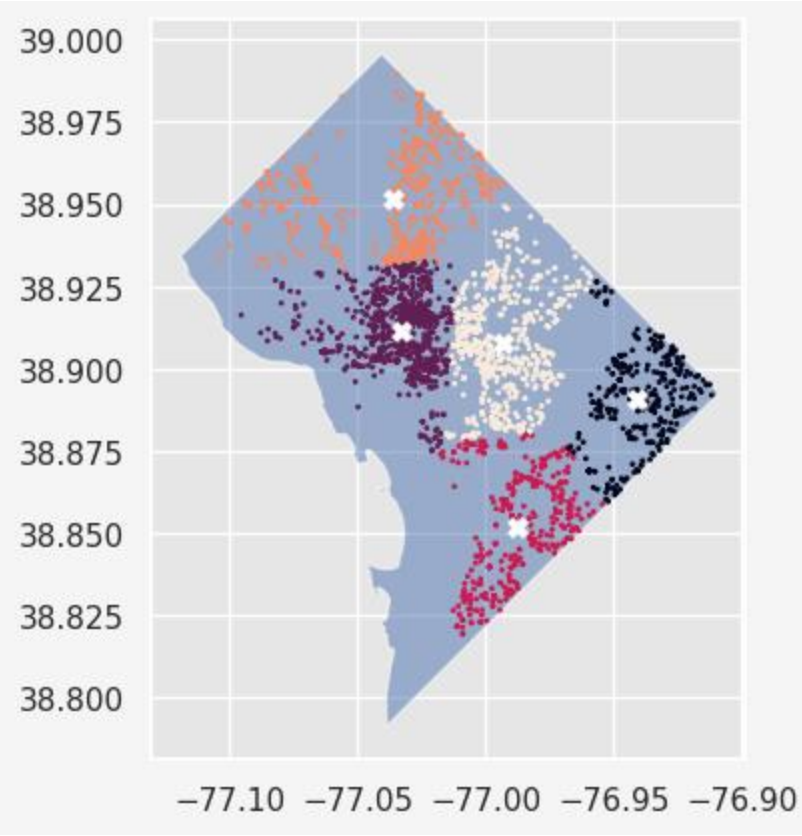
- There are **seven police districts** in Washington, DC, and each police district is divided into three sectors, which are groups of **Police Service Areas (PSAs)**. There are a total of **57** PSAs in the District of Columbia. Each police district has between seven and nine PSAs.

K-Means

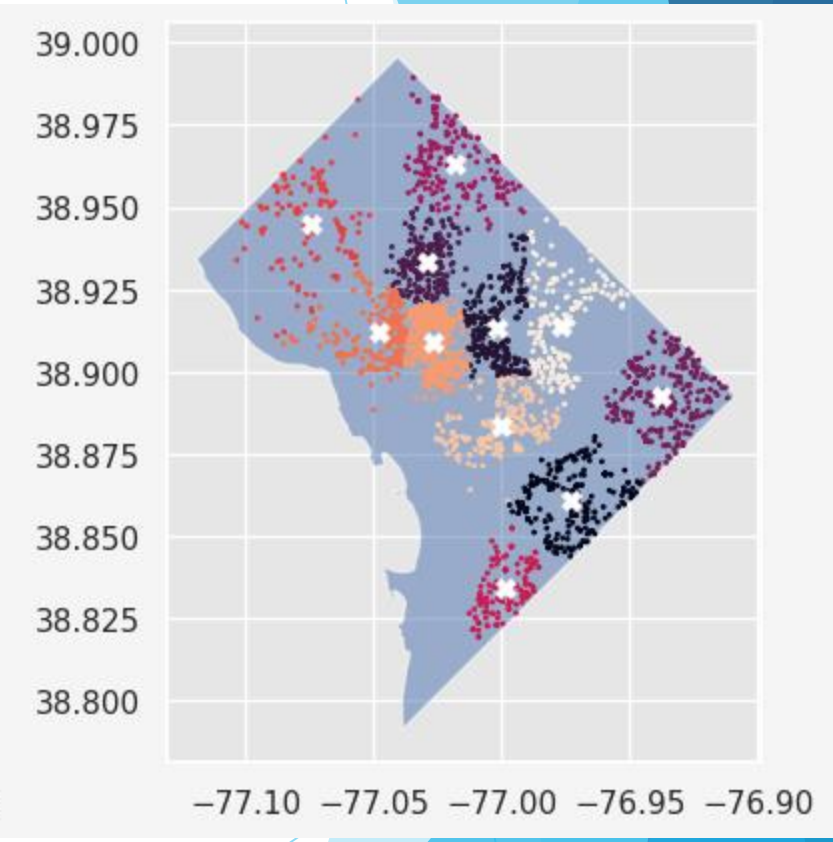
3 Clusters



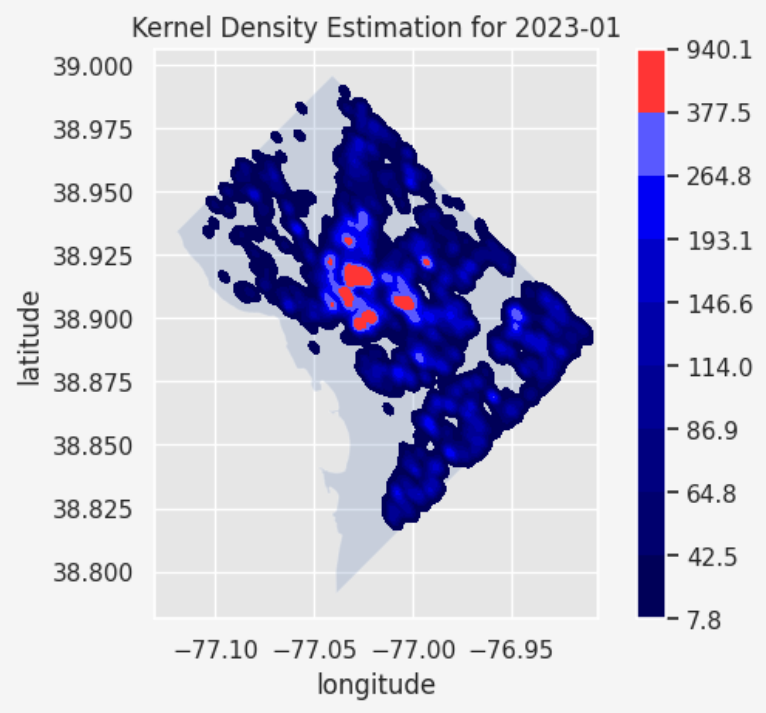
5 Clusters

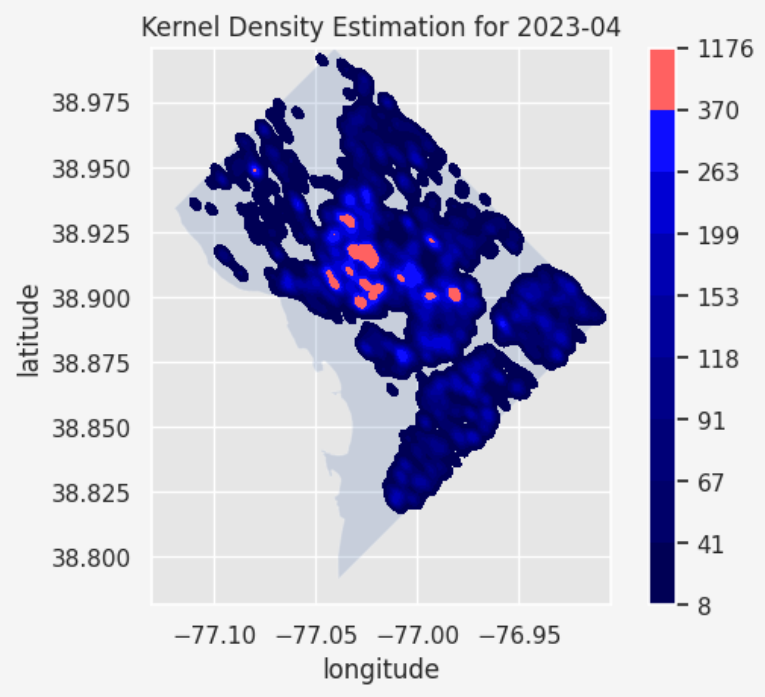


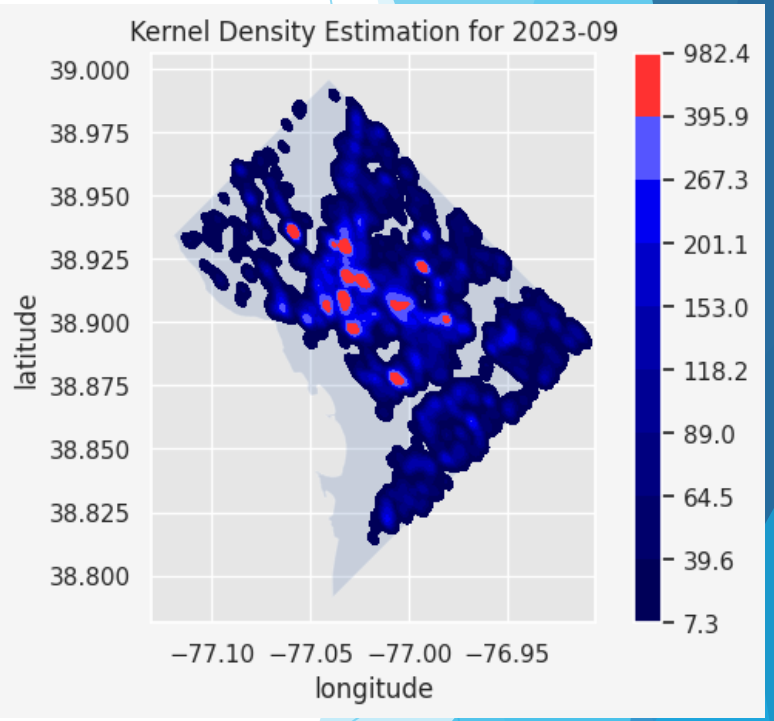
11 Clusters



Kernel Density Estimation

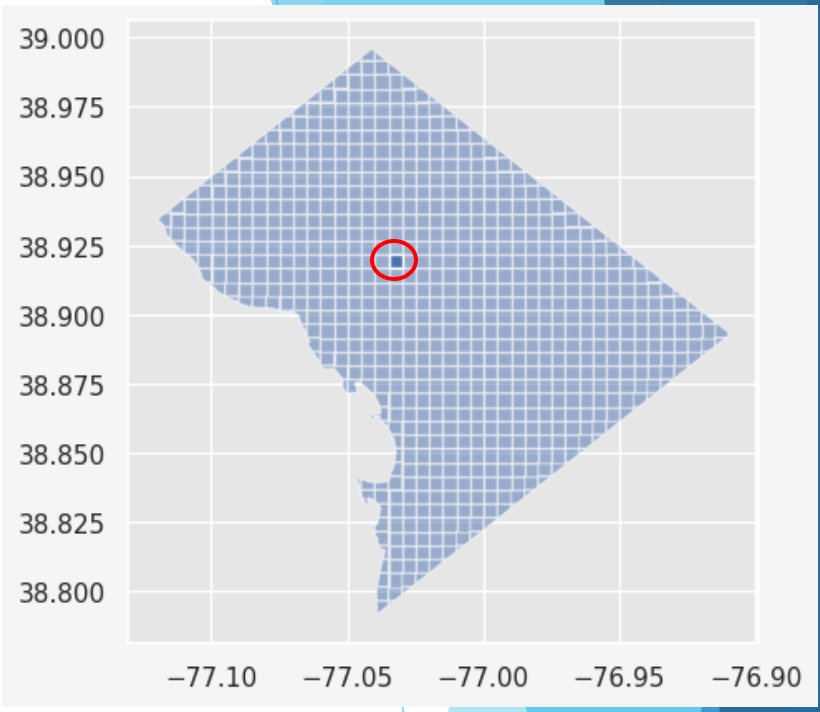




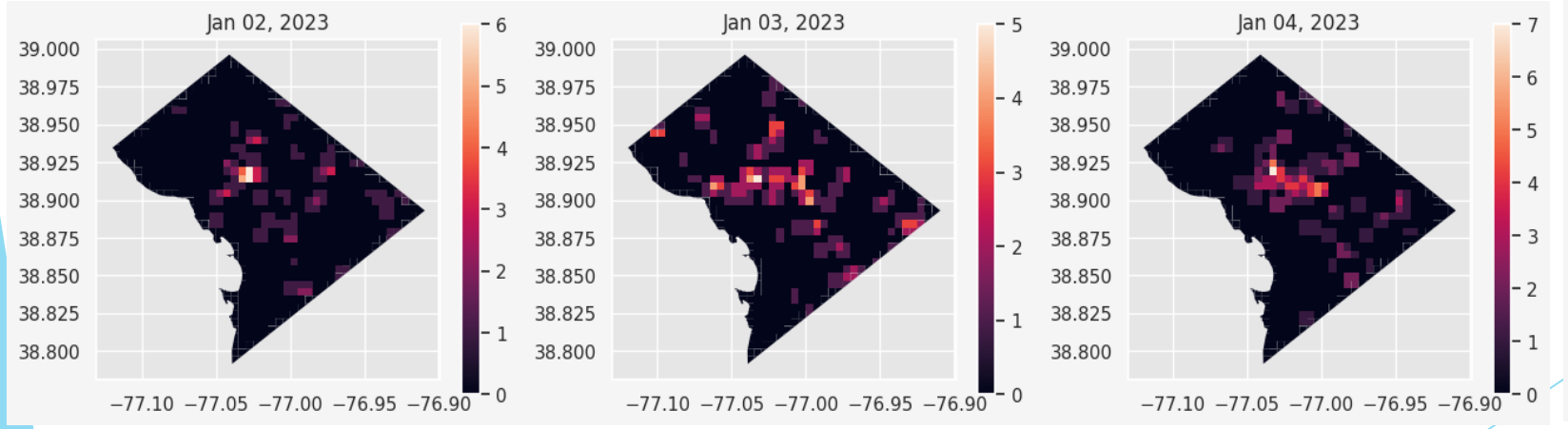


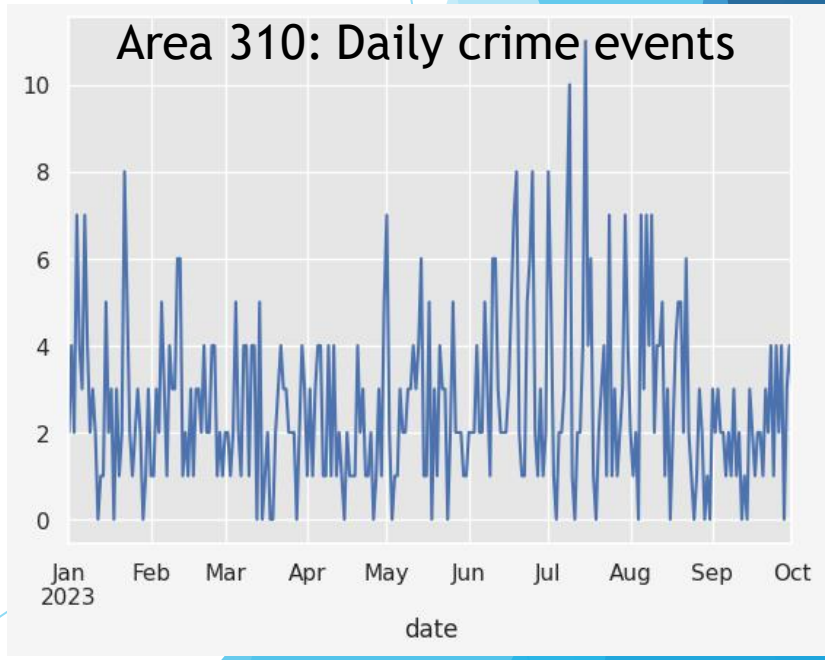
Backend model

- Partition DC map into 823 equally smaller areas(around 0.12 square miles/0.31 square km)
- Crime events for each area is a time series
- Each area can have multiple banks, ATMs, gas stations.



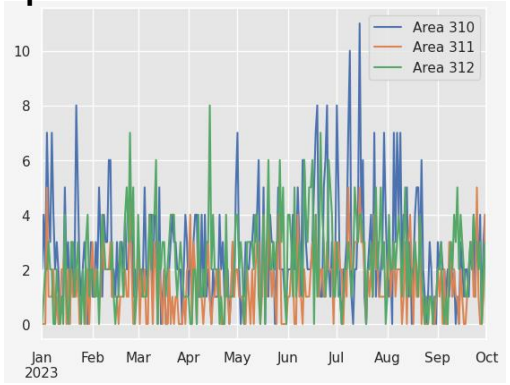
	atm	gas	banks	metro_stations	parks	museums	police_stations	post_office	shopping_centers	grocery	metro_bus	shuttle_bus
310	6.0	2.0	5.0	1.0	2.0	0.0	0.0	1.0	0.0	2.0	15.0	0.0

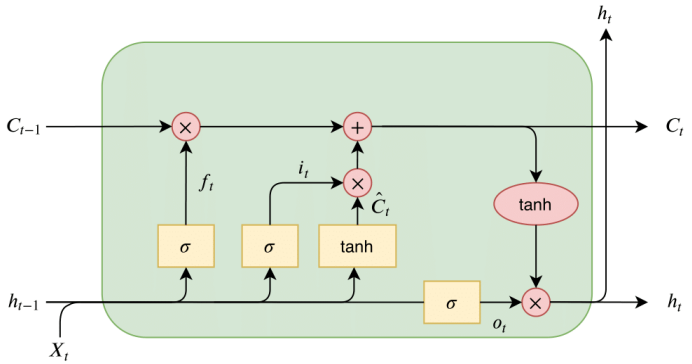


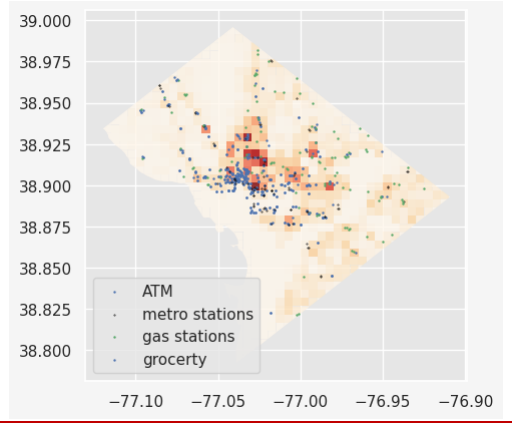


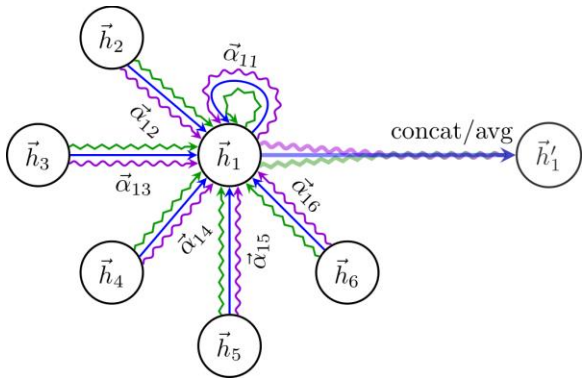
Model architecture

Input: time series for all areas



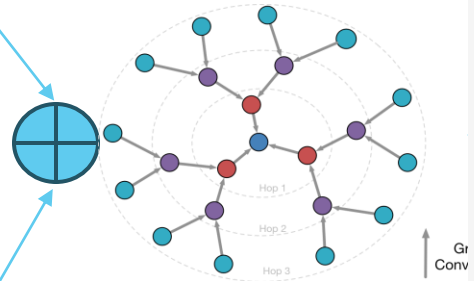


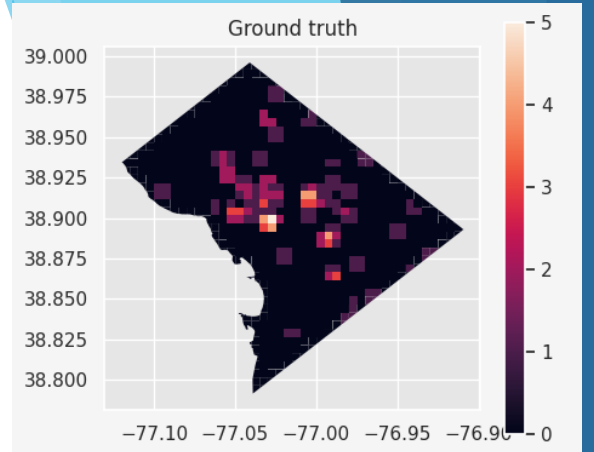


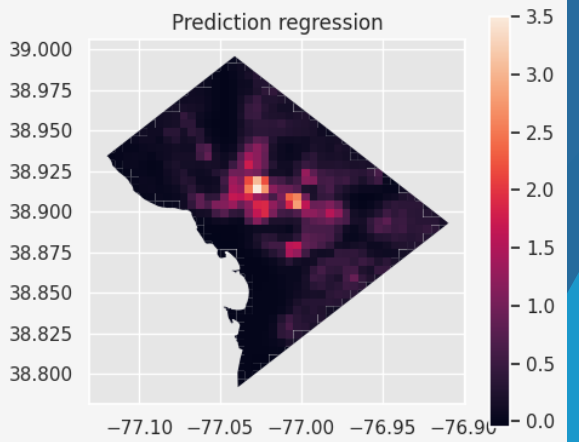


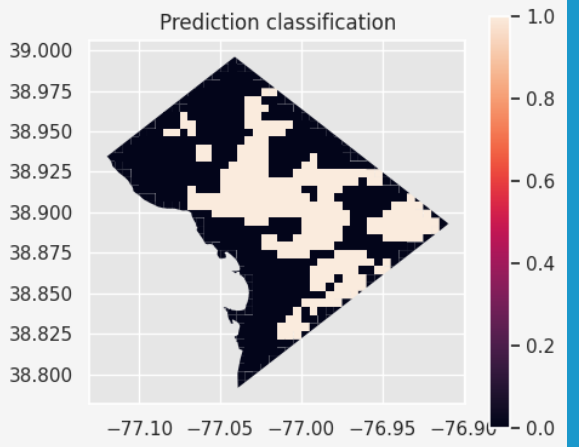
	atm	gas	banks	metro_stations	parks	museums	police_stations	post_office	shopping_centers	grocery	metro_bus	shuttle_bus
310	6.0	2.0	5.0	1.0	2.0	0.0	0.0	1.0	0.0	2.0	15.0	0.0

Input: graphs of the urban infrastructure





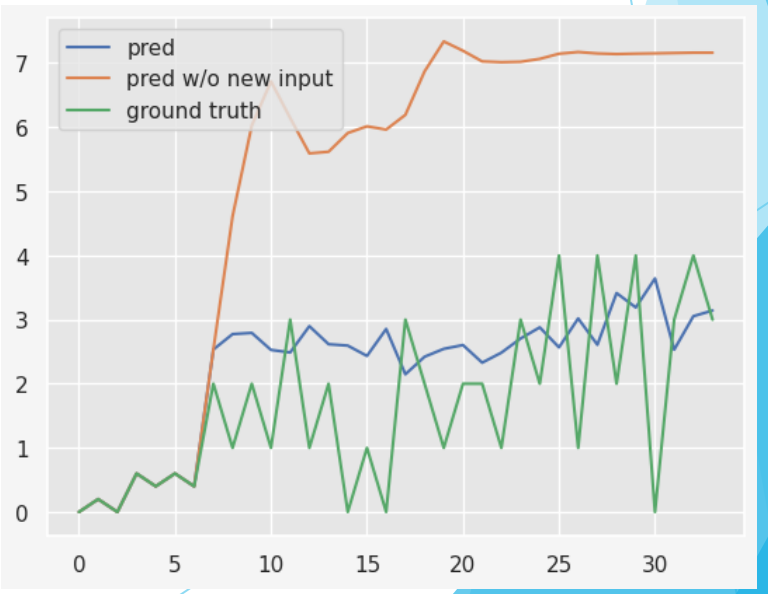




Result

	MAE	MSE	MAE%	Accuracy	Precision	Recall	F1	AUC
MLP	0.357	0.351	1.552	0.782	0.477	0.630	0.543	0.726
LSTM	0.357	0.353	1.791	0.777	0.471	0.637	0.540	0.725
GAT	0.343	0.336	1.598	0.763	0.456	0.736	0.562	0.753
Proposed	0.349	0.338	1.416	0.763	0.455	0.745	0.564	0.756

- Plot prediction for area 310.



Model Training/Validation

- Daily aggregated data for each area is used.
- With sequence length is 7 (one week), 266 data points.
- The first 215 (80%) data points are used for training model.
- The next 24 (10%) data points are used for validation/early stopping.
- The last 27 (10%) data points are reserved for model testing.

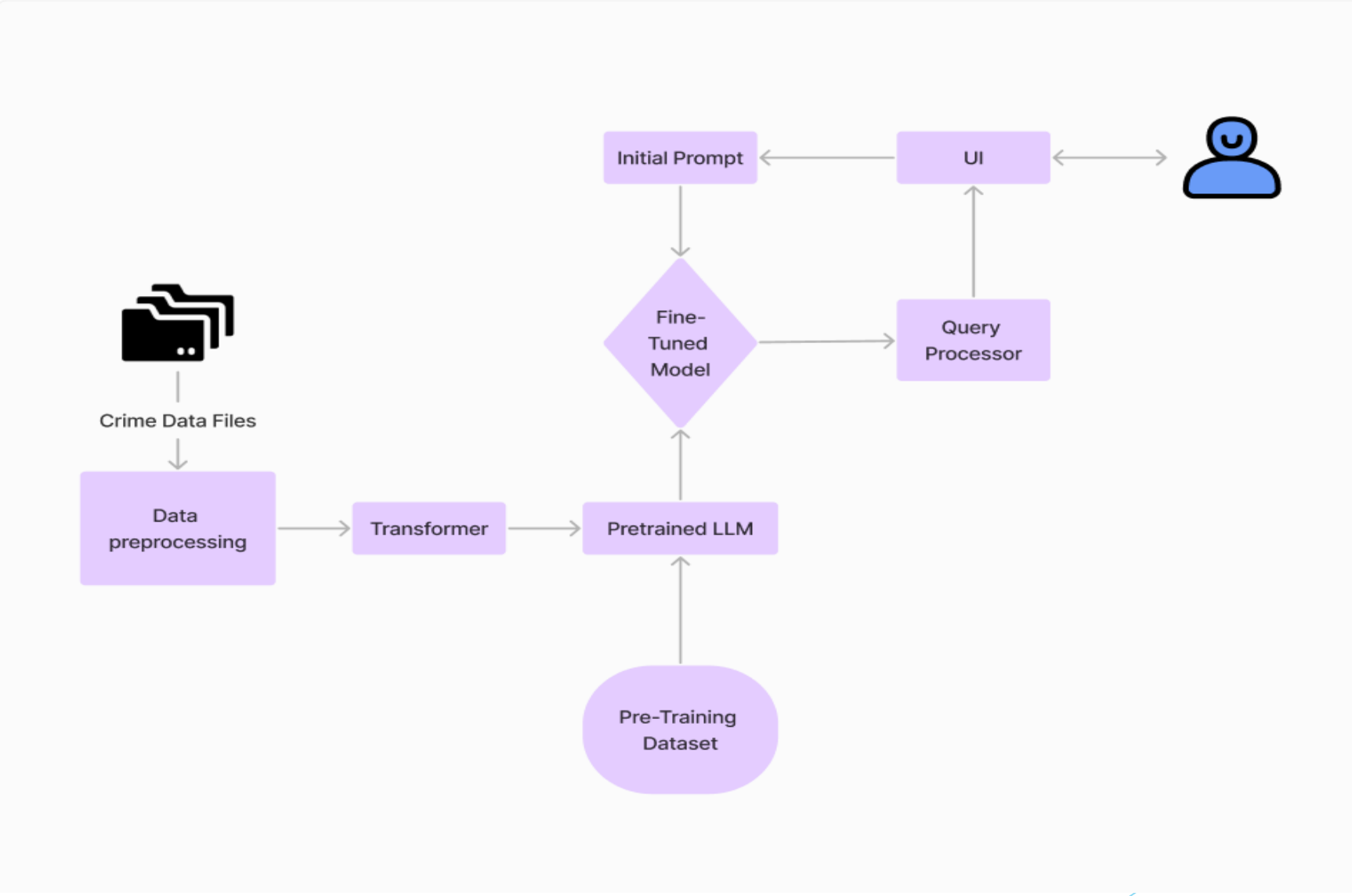
Understanding the news articles

- They are 526 articles from Washington Post, 2023 in D.C. metro area.

Search Summary

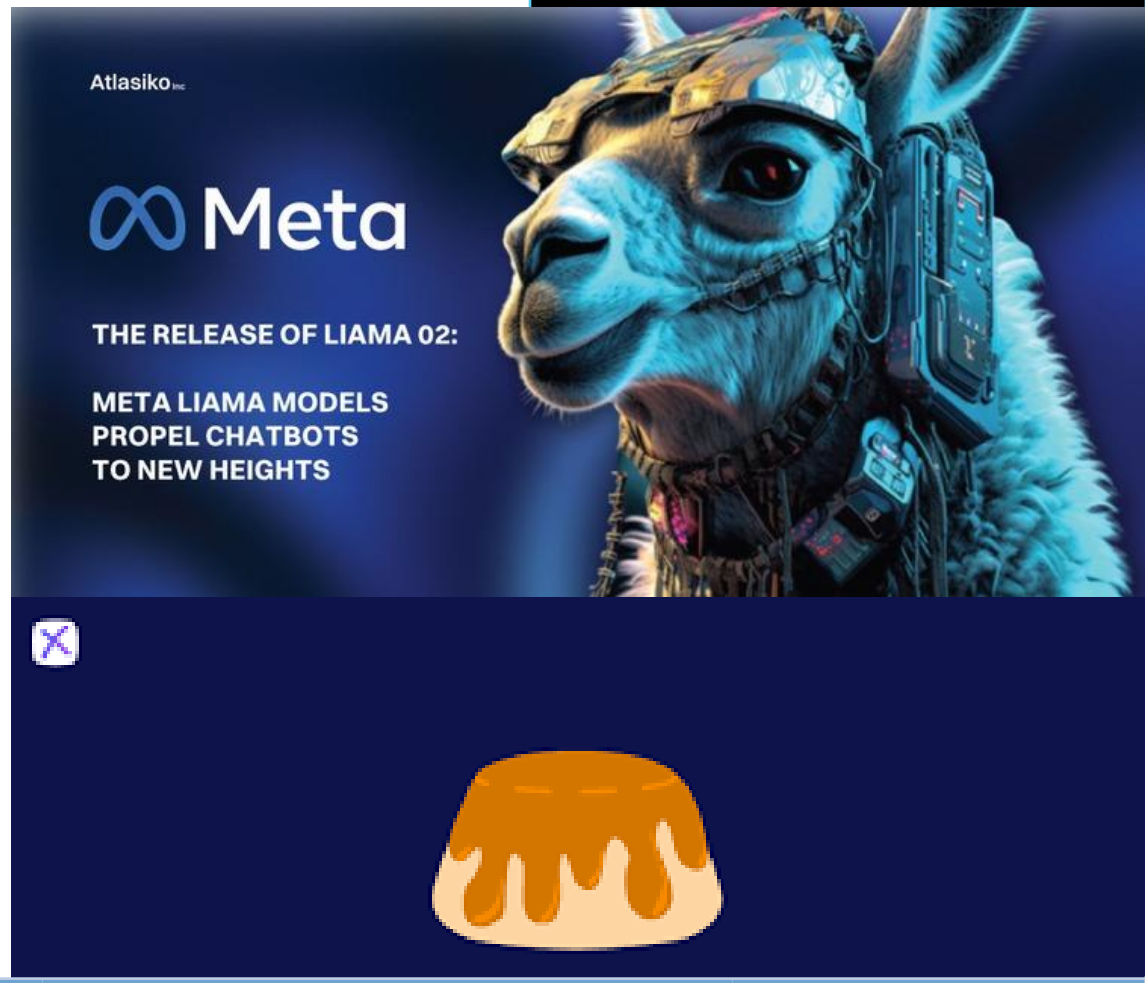
Text	
Date	01/01/2023 to 12/31/2023
Source	The Washington Post
Author	All Authors
Company	All Companies
Subject	Crime/Legal Action Or Fire/ Rescue Services Or Fake News Phenomenon
Industry	All Industries
Region	Washington DC
Language	English
News Filters	Company: Not Taliban Subject: Not Domestic Politics Region(s): Not Palestine Not Florida Not Malaysia Not California Not Massachusetts Not New York State Not Illinois Not Israel

System Architecture



LLM Models

- Llama_index
- Flan



Title	Q:Which day, month and year did the crime happen?	Q:what time of the day did the crime happen?	Q:Did the crime happen in the morning, afternoon, evening or night?	Q:Did the crime happend on a weekday, weekend or holidays?	Q:What type of crime?	Q:What is the type of the crime?
Day-care center owner gets 4 years for shooting her husband	July	morning	afternoon	Friday	aggravated assault	aggravated assault
New video shows Rand Paul staffer getting stabbed repeatedly	25-Mar	afternoon	afternoon	on a sunny afternoon	assault with intent to kill	assault with intent to kill
One critically wounded after U.S. Marshals task force opens fire	Tuesday	afternoon	afternoon	Tuesday	attempted murder	attempted murder
While you were sleeping, the acting D.C. police chief could not	9-Aug-23	night	night	weekend	armed carjackings and homicide	armed carjackings and homicide
Water utility files suit over chemicals	Friday	Friday	Friday	Friday	pollution	pollution

Llama_index

Efficiency: Enables rapid access and querying of large datasets.

User Interaction: Supports dynamic user queries for on-the-fly crime analysis.

Reusability: Index persistence reduces load times and computational overhead.

Functions Used:

VectorStoreIndex:

Function: Manages a store of numerical representations of text documents, facilitating efficient text search and retrieval.

Application: Creates an index from documents to enable quick querying capabilities for crime-related information.

SimpleDirectoryReader:

Function: Handles the settings for persisting the index on a storage medium.

Application: Saves the created index to a specified directory (PERSIST_DIR), allowing for quick reloading rather than rebuilding on each application start.

Functions Used:

load_index_from_storage:

Function: Loads an existing index from storage using the defined StorageContext.

Application: Ensures efficient reuse of the index by loading it if already present, avoiding unnecessary re-indexing.

Query Engine:

Function: Executes searches on the indexed data based on user queries.

Application: Powers the user interface allowing for real-time crime analysis based on specific queries related to time and location of incidents.

Indexing the Data

```
PERSIST_DIR = "./storage"
if not os.path.exists(PERSIST_DIR):
    documents = SimpleDirectoryReader("data
    index = VectorStoreIndex.from_documents(documents)
    index.storage_context.persist(persist_dir=PERSIST_DIR)
else:
    storage_context = StorageContext.from_defaults(persist_dir=PERSIST_DIR)
    index = load_index_from_storage(storage_context)
query_engine = index.as_query_engine()
```

Metrics

Similarity Score:

_____ Source Node 1/2 Node ID: 712aa4b5-62e6-4bc8-b790-51de415bd4fa Similarity: 0.8661705037720195 Text: Metro In D.C., another deadly Independence Day Emily Davies Peter Hermann 1,307 words 6 July 2023 The Washington Post WP FINAL B01 English Copyright 2023, The Washington Post Co. All Rights Reserved Gathered around the clear casket of an 11-year-old boy killed at a stop-the-violence cookout on the Fourth of July in 2020, D.C. ...

_____ Source Node 2/2 Node ID: 75886d97-ad45-40f1-9010-ee55f152832b Similarity: 0.8221685923824704 Text: of two men charged in illegal firearms cases. Graves said those investigations uncovered a more coordinated robbery scheme targeting the jewelry stores, including two in Falls Church and two in Springfield, Va. Prosecutors said the case then sprawled to four states, involving many local and federal law enforcement officers. Authorities said the ...

Metrics

Similarity Score:

- Function: These scores indicate how similar the source node's text is to the query or main topic being analyzed (in this case, violent incidents on the Fourth of July in Washington, D.C.).
- Usage: Higher similarity scores suggest that the text from the source nodes is highly relevant to the query. These scores help in deciding which sources are most pertinent to include or highlight in the final output.

Results

Final Response: The D.C. Metro system has been experiencing a surge in violent crime, prompting officials to implement additional security measures such as deploying D.C. police officers to patrol certain Metro stations. The increase in crime on the Metro system has raised concerns among passengers and the community, with multiple incidents of shootings and homicides occurring within the transit system. Despite efforts to enhance security, including the deployment of crisis intervention specialists and additional patrols, some riders still express feeling unsafe while using the Metro. _____
Source Node 1/2 Node ID: a3a36a1d-6b8d-4f02-9470-23da807b722f Similarity: 0.8651764244237017
Text: Metro Transit police will get an assist Emily Davies Peter Hermann 565 words 9 February 2023 The Washington Post WP FINAL B01 English Copyright 2023, The Washington Post Co. All Rights Reserved D.C. police will soon start patrolling five Metro stations in the District to augment the transit police force that is struggling to p...
_____ Source Node 2/2 Node ID: 45df98cc-48c0-4cfe-9c9b-dfac98cce7b5 Similarity: 0.8649982570978357 Text: Metro Metro violence worries deepen Emily Davies Justin George 817 words 31 May 2023 The Washington Post WP FINAL B01 English Copyright 2023, The Washington Post Co. All Rights Reserved In February, D.C. Mayor Muriel E. Bowser (D) announced plans to deploy D.C. police to accompany transit officers at Metro stations across the ...

Crime rating: low

Potential for violence: unlikely

Safety advice: It's relatively safe to go out.

Improvements from the first prototype.

- Shift from a local model to an API model.
- Response time reduced from 15 minutes to 10 seconds.
- User friendly UI.
- Gives proper guidelines based on the crime concentration.
- Metrics.

Future work

- Add other urban infrastructure data, such as street map, weather.
- Extend the work to other cities and longer timespan to find the crime cycles.
- Denoise methods.
- Get access to more publicly available crime data without hindering with public data privacy.
- Refine the model to even more accurately pinpoint the crime possibilities also keeping in mind the ethical implications.

Reference

- Eck, John E, Spencer Chainey, James G Cameron, Michael Leitner, and Ronald E Wilson. “Mapping Crime: Understanding Hot Spots,” n.d.
- Butt, Umair Muneer, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir, and Hafiz Husnain Raza Sherazi. “Spatio-Temporal Crime HotSpot Detection and Prediction: A Systematic Literature Review.” *IEEE Access* 8 (2020): 166553–74. https://doi.org/10.1109/ACCESS.2020.3022808.
- Catlett, Charlie, Eugenio Cesario, Domenico Talia, and Andrea Vinci. “Spatio-Temporal Crime Predictions in Smart Cities: A Data-Driven Approach and Experiments.” *Pervasive and Mobile Computing* 53 (February 2019): 62–74. https://doi.org/10.1016/j.pmcj.2019.01.003.

Demo