

## **Week 1**

Gopinadh Batchu

Dundi Vivek Reddy

Data Science, Saint Peter's University

DS-670-HYB2-25SPTR: Capstone: Big Data & Data Sci

Professor: Reda Mastouri

March 9, 2025

## **Literature Review, Analysis, Project Plan**

**Topic: NYC Crime and Motor vehicle Collisions Economical and Social Study.**

**Using Quantitative and Qualitative Approaches**

### **Introduction**

Urban metropolitans, huge cities like New York City (NYC), face multiple issues that significantly impact public safety, economic well-being, and community welfare. Crime and motor vehicle collisions (MVCs) are some of the problems that occur frequently and have severe consequences. Historically, these incidents have been examined individually, sometimes disregarding potential connections spurred by socioeconomic factors. This literature review synthesizes recent studies in urban crime, MVCs, and socioeconomic determinants—demographic, social, economic, and housing determinants highlighting the potential of an integrated analytical framework for enhancing policy-making and urban management.

### **Abstract**

This literature review examines integrating New York City (NYC) crime and motor vehicle collision (MVC) data with broader socioeconomic indicators, including demographic, social, economic, and housing factors. Historically, crime and MVCs have been analyzed independently, and they significantly impact urban life, the

economy, and public health. Current methodologies rarely explore their intersection, neglecting crucial interactions shaped by socioeconomic contexts. Through the use of advanced machine learning, big data analysis, and natural language processing techniques, the review emphasizes the need to utilize standardized analytical methods. Integrating these heterodox data sets provides comprehensive insight, enabling policymakers to enact targeted, high-impact urban interventions. This socio-economically contextualized analysis aligns with today's government, NGO, and industry standards, providing public safety, urban planning, and resource allocation improvement opportunities in NYC.

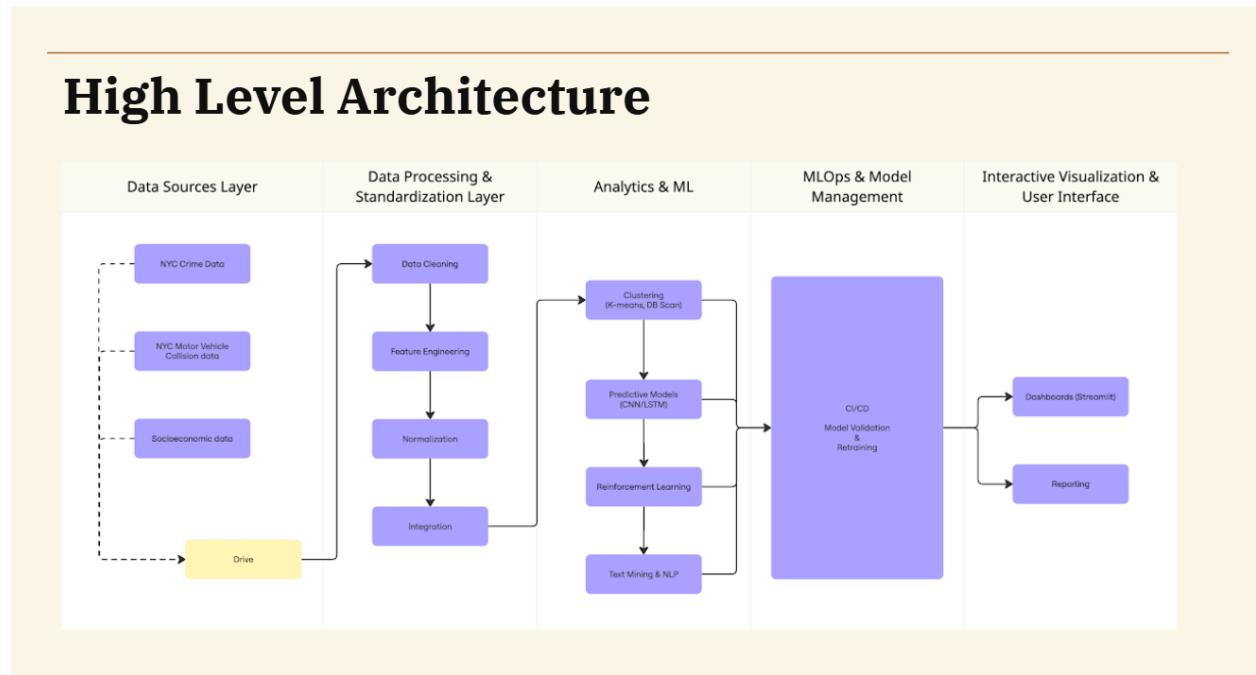
### **Business Need**

Crime and motor vehicle collisions impact the socioeconomic security and public health of cities like New York City (NYC). Analyzing these problems separately tends to create piecemeal knowledge and less efficient interventions. An urgent demand exists among government organizations, city planners, and community-oriented NGOs for an integrated analysis approach. By integrating crime and collision data with high socioeconomic variables—demographic, social, economic, and housing variables—this study provides a rich foundation for making informed, effective, and comprehensive policy decisions and resource allocation strategies that more effectively address community well-being.

## **Problem Statement**

Currently, NYC crime and motor vehicle collisions are predominantly analyzed independently, which results in missing potential interdependencies and underlying socioeconomic influences. Independent processing limits the range of options and dilutes the efficacy of urban policymaking and crime policies. Our project fills this space through an integrated, socio-economically enhanced analysis framework. This approach aims to uncover intersectional patterns and associations between crime incidents and motor vehicle accidents, thus facilitating more effective public safety interventions, optimized resource allocation, and policymaking in New York City.

## High level Architecture :



## Minimum Viable Product (MVP)

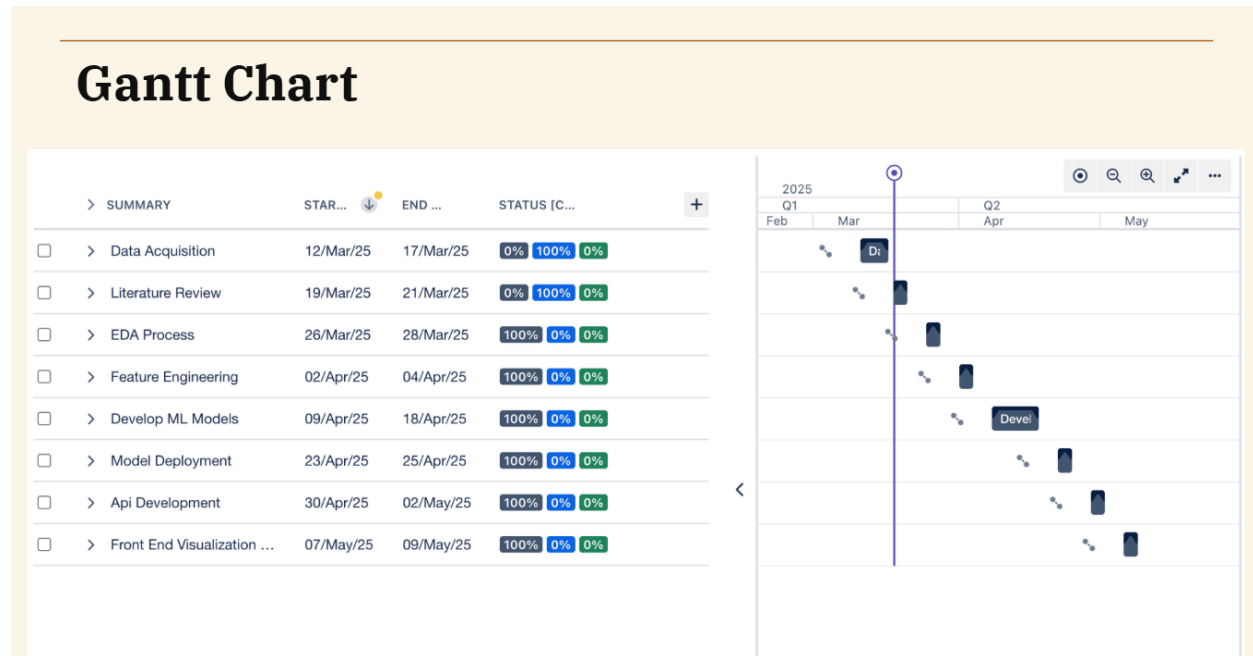
This project's Minimum Viable Product (MVP) involves developing an initial analytic pipeline designed to integrate NYC crime data, motor vehicle collisions, and socio-economic datasets. The MVP includes:

- Establish standardized preprocessing and normalization for crime, collision, and socio-economic data.
- Utilize clustering algorithms (K-means, DBSCAN) to identify overlapping hotspots influenced by socio-economic variables.

- Apply preliminary machine learning models (CNN/LSTM) to demonstrate the ability to forecast collision and crime patterns based on historical and socio-economic data.
- Deploy basic natural language processing (NLP) techniques to analyze unstructured textual data from incident reports, enhancing feature extraction for predictive modelling.
- Develop a streamlined visualization interface for stakeholder feedback, allowing policymakers to interact with predictive insights and validate model outcomes.

This MVP serves as a foundational prototype to validate the feasibility of integrative machine learning methodologies. Early-stage models allow rapid hypothesis testing, highlight data quality issues, and assess initial predictive performance. Feedback from this MVP phase will guide iterative improvements, ensure alignment with stakeholder needs, and optimize the deployment of robust, scalable machine-learning solutions for urban safety and policymaking.

## Gantt Chart:



## Team Members:

- Gopinadh Batchu
- Vivek Reddy Dundi

## References:

- Barton, M. S. (2014). Gentrification and violent crime in New York City. *Crime & Delinquency*, 62(9), 1180–1202. <https://doi.org/10.1177/0011128714549652>
- Helland, E., & Tabarrok, A. (2004). The Fugitive: Evidence on Public versus Private Law Enforcement from Bail Jumping. *The Journal of Law and Economics*, 47(1), 93–122. <https://doi.org/10.1086/378694>
- Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “Neighborhood Effects”: Social processes and new directions in research.

*Annual Review of Sociology*, 28(1), 443–478.  
<https://doi.org/10.1146/annurev.soc.28.110601.141114>

- White, M. D. (2012). The New York City Police Department, its Crime Control Strategies and Organizational Changes, 1970-2009. *Justice Quarterly*, 31(1), 74–95. <https://doi.org/10.1080/07418825.2012.723032>
- Hall, J. J., & Ratcliffe, J. H. (2024). Assessing the impact of safe consumption sites on neighborhood crime in New York City: A synthetic Control approach. *Journal of Experimental Criminology*. <https://doi.org/10.1007/s11292-024-09630-z>
- Singh, A., Kumar, S., & Bansal, S. (2024). NYC CrimeWatch: Crime Analysis Tool. *2024 IEEE 18th International Conference on Semantic Computing (ICSC)*, 311–316.  
<https://doi.org/10.1109/icsc59802.2024.00056>