**ADTA 5340: Group Project Proposal**

**Group Name:** Brainy Bots

**Team Members:** Tejamanikanta Gudla, Sai Preethi Jiddigam, Harish Chinnakadiri, Indusree Terala, Sai Swetha Puli

**Name of the Dataset: NYPD Crime Data**

**Dataset Source:** <https://data.world/data-society/nyc-crime-data>

**Dataset Description:**

1. **CMPLNT\_NUM**: Complaint Number – A unique number for each complaint.
2. **CMPLNT\_FR\_DT**: Complaint From Date – The starting date of the incident.
3. **CMPLNT\_FR\_TM**: Complaint From Time – The starting time of the incident.
4. **CMPLNT\_TO\_DT**: Complaint To Date – The ending date of the incident (if applicable).
5. **CMPLNT\_TO\_TM**: Complaint To Time – The ending time of the incident (if applicable).
6. **RPT\_DT**: Report Date – The date on which the incident was reported to the police.
7. **KY\_CD**: Key Code – A numeric code representing the offense category.
8. **OFNS\_DESC**: Offense Description – The description of the offense.
9. **PD\_CD**: Police Department Code – A numeric code representing the specific offense.
10. **PD\_DESC**: Police Department Description – The description of the specific offense according to the police department's coding.
11. **CRM\_ATPT\_CPTD\_CD**: Crime Attempted/Completed Code – Indicates whether the crime was attempted or completed.
12. **LAW\_CAT\_CD**: Law Category Code – Classification of the offense (e.g., felony, misdemeanor, violation).
13. **JURIS\_DESC**: Jurisdiction Description – Description of the jurisdiction where the incident occurred.
14. **BORO\_NM**: Borough Name – The name of the borough in which the incident occurred.
15. **ADDR\_PCT\_CD**: Address Precinct Code – The precinct in which the incident occurred.
16. **LOC\_OF\_OCCUR\_DESC**: Location of Occurrence Description – Description of where the incident occurred (e.g., inside, outside).
17. **PREM\_TYP\_DESC**: Premises Type Description – Type of location where the incident occurred (e.g., apartment, street).
18. **PARKS\_NM**: Parks Name – Name of the park where the incident occurred (if applicable).
19. **HADEVELOPT**: Housing Development – Name of the NYCHA housing development where the incident occurred (if applicable).
20. **X\_COORD\_CD**: X Coordinate – The X coordinate for the location of the incident in a New York State Plane Coordinate System.
21. **Y\_COORD\_CD**: Y Coordinate – The Y coordinate for the location of the incident in a New York State Plane Coordinate System.
22. **Latitude**: Latitude – The latitude of the incident location.
23. **Longitude**: Longitude – The longitude of the incident location.
24. **Lat\_Lon**: Latitude and Longitude – The combined latitude and longitude of the incident location.

From the above columns we will pick required columns to build models.

**Proposal Title:**

"Predictive Analytics for Crime Occurrence and Classification in New York City: A Machine Learning Approach"

**Proposal:** Examine the NYC crime dataset thoroughly to identify patterns in crime rates across different neighborhoods and crime types, focusing on temporal trends and spatial distributions. Explore variations over time, including daily, weekly, and seasonal fluctuations, to gain insights into the dynamics of crime in New York City.

**Methodology:**

* **Data Preparation:** Clean and preprocess the NYPD Complaint Data, including handling missing values, encoding categorical variables, and feature engineering.
* **Exploratory Data Analysis (EDA):** Conduct EDA to uncover patterns and trends in the data, and to identify significant predictors for crime occurrences and categories.
* **Model Development:**
  + **Prediction Models:** Experiment with different algorithms, including Random Forest, Gradient Boosting Machines, and LSTM networks, to forecast crime occurrences.
  + **Classification Models:** Employ algorithms like Support Vector Machines, Decision Trees, and Random Forest to classify crimes into their respective categories.
* **Evaluation and Optimization:** Use cross-validation and hyperparameter tuning to optimize model performance. Evaluate models using metrics such as accuracy, precision, recall, F1 score, and AUC-ROC.