<div align="center">

PROGRESS UPDATE REPORT

</div>

Completed Tasks

1. **Project Proposal:** We have finalized the project proposal, outlining the project's objectives, scope, and methodologies to be employed.
2. **Git Repository Setup:** We have also established a dedicated Git repository for the project on GitHub, facilitating version control and collaboration.
3. **Dataset Acquisition:** We have secured relevant datasets for the project and imported them into our python files.
4. **Data Exploration:** Comprehensive exploration of the data included:
   - Visualizing the structure of the different datasets, and initial patterns.
   - Concatenating the three datasets.
5. **Data Preprocessing:** We have executed comprehensive data preprocessing tasks successfully, encompassing:
   - Handling missing data through removal strategies.
   - Data cleaning to address inconsistencies.
   - Removing duplicates.

Data

Here-in is the link to the three datasets: - https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge-posting#data

- 2015, full year (zip, 13.8 MB)
- 2014, full year (zip, 12.6 MB)
- 2013, full year (zip, 11.7 MB)

Challenges

Encountered Difficulties:
- Class imbalance: The category for street crime is imbalanced. This can make it difficult for machine learning models to accurately predict crimes.
- Data quality: The crime incident dataset contains missing values and inconsistencies. This can make it difficult to extract meaningful patterns from the data.

How challenges were addressed:

- Class imbalance: We plan to address the class imbalance problem by using oversampling. Oversampling is a data augmentation technique that replicates minority class examples.
- Data quality: We addressed the data quality problem by cleaning the crime incident dataset. We removed missing values, duplicates, and corrected inconsistencies.

Collaboration

- Our team meets once per week to discuss progress and any challenges that we are facing.
- We have a shared repository where we store all our code and data.
- We use a communication tool to keep in touch with each other outside of our weekly meetings.
- All our team members are contributing meaningfully to the project.

Next steps

Remaining Tasks:

- Deploy the crime forecasting system to production.
- Monitor the performance of the system and adjust as needed.
- Continue with data processing.

Plan to complete remaining tasks:

- We plan to meet twice a week following the timeline to meet the deadline.
- Choose the appropriate statistical method to use.
- Allocate time for conducting statistical analysis and interpreting statistical results.

Potential challenges:

We are currently in the data preprocessing phase of our project. The primary challenge encountered is the class imbalance within the crime datasets. To address this issue, we plan to implement oversampling techniques, ensuring a more balanced representation of crime types in our training data.