**Enhancing Predictive Policing Through Spatiotemporal Crime Forecasting**

**Progress so far**

The first part was the preprocessing part of the project in the following areas:

* Integrating the 2013, 2014 and 2015 datasets. Data was merged using concat() function.
* Converting categorical variables into numerical variables. This was applied to ‘CATEGORY’ and ‘CALL’ GROUPS variables. LabelEncoder from sklearn.preprocessing was used.
* Checking for duplicates. A total of 7890 rows were found to be duplicated. The duplicated rows were removed.
* Unnecessary columns were also removed. In this dataset, I removed the ‘final\_case\_type’ and ‘CASE DESC’ columns.

The second step was feature engineering. To achieve this, the following steps were undertaken.

* Converting categorical variables into numerical variables. This was applied to ‘CATEGORY’ and ‘CALL’ GROUPS variables. LabelEncoder from sklearn.preprocessing was used.

The third step involved creating visualizations for each variable over the period. Thus, the following visualizations were generated:

* Visualizing the ‘x\_coordinate’, ‘y\_cordinate’ and ‘CALL GROUPS’ over the three years timeframe.

**Challenges**

I had challenges in selecting the best encoding method for my categorical data. I am not sure whether I will be able to get the best model based on the encoding that I have used. If the model does not work as expected I might have to change the categorical data encoding step.

I am having challenges understanding and handling time-related aspects of my datasets. I am doing more research in this area.

Another challenge that I have encountered is visualization. Some of the visualizations that I have to be uninterpretable and I am working on improving them.

**Next steps**

The remaining part involves the following:

* Time series decomposition,
* Geospatial analysis,
* Building ARIMA, SARIMA models and testing the accuracy
* Writing a report

I have started working on the remaining steps and I hope to be done in a weeks’ time.