**Pseudocode for Crime Analysis Using GIS for Data Science**

1. **Data Gathering Phase**

* For each year from 2019 to 2023:
  + Access and retrieve the crime dataset for the specified year.
  + Consolidate datasets for each year into a single comprehensive dataset.

1. **Data Preprocessing Phase**

* Inspect the dataset for any missing or corrupt data.
* Filter the dataset to focus on essential attributes: SHIFT, METHOD, BLOCK, XBLOCK, YBLOCK.
* Prepare and format the dataset to be compatible for visualization and machine learning.

Add incident data on the Map

Add features 1 – numeric features, categorical features,

Add features 2 – Census track with X,Y

Batch Import Data

[Feature Class To Geodatabase (Conversion)](https://pro.arcgis.com/en/pro-app/latest/tool-reference/conversion/feature-class-to-geodatabase.htm)

[Regression](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/regression-analysis-basics.htm)

[Count the number of point features](https://support.esri.com/en-us/knowledge-base/how-to-count-the-number-of-point-features-within-a-poly-000008599)

[Add XY](https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/xy-table-to-point.htm)

[KDE](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/kernel-density.htm)

Hotspot detection (KDE)

Make Environmental variables: Distance, Buffer

**Modeling**

1. Training in python and DP on ArcPro : expect clear time using Spatial feature

Sampling..

2. GeoCoded Grid : expect Grid spot..

1. **Visualization Phase using ArcGIS Pro**

* Load the pre-processed dataset into ArcGIS Pro:
  + Map – Add Data - Display XY Data – Set the coordinate system (WGS 1984)
  + Iterate the process for 5 years of Crime data in DC
* Generate visualizations based on:
  + Timing of crimes (SHIFT)
  + Method of crimes (METHOD)
  + Location of crimes (BLOCK, XBLOCK, YBLOCK)
  + Save the visualizations for reference and presentations.
* Create features: (ex. 100ft Buffer from crime incidents spots)
  + New Feature Class.
  + Make Buffer (set the distance of buffer as 100 feet).
  + Select intersection between Key Properties. (hospital, restaurant, metro)
  + Create numeric features pertaining to selections.
  + Extract generated features to link with python scripts.

1. **Machine Learning Phase**

* Split the dataset into training and test subsets.
* Choose an appropriate supervised machine learning model.
* Train the chosen model on the training subset.
* Validate the model's accuracy and effectiveness using the test subset.

1. **Prediction Phase**

* Using the trained model, make predictions on potential future crime trends or attributes.
* Store the predictions for spatial representation.

1. **Spatial Representation Phase using ArcGIS Pro and Python**

* Load the predictions into ArcGIS Pro.
* Use Python scripting to generate spatial maps based on the predictions.
* Ensure the generated maps are intuitive and informative.
* Save the spatial representations for further use.

1. **Reporting Phase**

* Compile a comprehensive report detailing:
  + Project objectives.
  + Methodologies employed.
  + Key findings and visualizations.
  + Predictions and their implications.
  + Potential areas for improvement or future exploration.