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

1. **Data Gathering Phase (Completed)**

* 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.

**Result (also updated on Data & Code Book)**

***[DATASET] CRIME INCIDENTS*** *(recent 5 years: 2019~2023)*

*- Source:* [*Open Data DC*](https://opendata.dc.gov/datasets/DCGIS::crime-incidents-in-2023/about)*, the crime DB by the DC Metropolitan Police Department (MPD).*

*- Format: SHP, KML, CSV (same data with three formats)*

*- Location:* [*Team3 Github repository*](https://github.com/Abhi050/GIS_Project/tree/main/Crim%20Raw%20data%20(openDC)_KML)

*- Description: Spatiotemporal Information of crime events in Washington, DC*

***[CODE BOOK] Economic Characteristics DC***

*- Source:* [*CENSUS Tract*](https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-geodatabase-file.2023.html#list-tab-1258746043)

*- Location:* [*Team3 Github repository*](https://github.com/Abhi050/GIS_Project/tree/main/DC_Census_Track(openDC))

*- Description: 5-year estimates shown by 2020 Census Tract boundaries sourced from American Community Survey (ACS), which represent about 200 economic characteristics- Employment, Commuting, Occupation, Income, Health Insurance, Poverty, etc.*

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.

1. **Visualization Phase using ArcGIS Pro (~ing)**

* Load the pre-processed dataset into ArcGIS Pro:
  + Batch Import Data – Add Crime incident file(\*.SHP) to the current GDB
  + (Skipped) Display XY Data Set the coordinate system (WGS 1984)
  + Add feature class (incident spots) on the current Map
  + Iterate the process for 5 years of Crime data in DC
* Create KDE area into ArcGIS Pro:
  + Connect to the GDB through Python
  + Set Environment – import Spatial Analyst Module(from arcpy.sa import \*)
  + Set checkout Extension: arcpy.CheckOutExtension("Spatial")
  + Create KernelDensity feature class with defining cell size,etc.

from arcpy.sa import \*

arcpy.CheckOutExtension("Spatial")

arcpy.env.overwriteOutput=True

with arcpy.EnvManager( extent=arcpy.Extent(-77.112318,38.814667,-76.910140,38.993573)):

    outKernelDensity2 = KernelDensity(in\_features='Crime\_Incidents\_in\_2023\_shp',

                                     population\_field='',

                                     cell\_size=0.005,

                                    search\_radius=0.01)  *#"PLANAR"*

    outKernelDensity2.save("Crime\_Incidents\_in\_2023\_KDE")

**Result (Crime\_incidents\_2023: Screenshots on the ArcPro)**

\*Purple color: KED density area

A map of a city

Description automatically generated

* Generate visualizations based on:
  + Yearly variation by Years (2019~2023)
* Define hypothetical phases

Ex. Before Pandemic – Pandemic – After Pandemic

* + Crime type (OFFENSE): figure out patterns by selecting by attribute

A screenshot of a computer

Description automatically generated

* + Timing of crimes (SHIFT)
  + Method of crimes (METHOD)
  + Location of crimes (BLOCK, XBLOCK, YBLOCK)
  + Save the visualizations for reference and presentations.
* Create Property features and add visualizations

* + Add property feature classes (school, liquor shops, hospital, metro etc)
  + Make Buffer features
* Set the distance of buffer as N feet
* Find properties by using intersection
* Calculate number of properties within buffer (ex. Buffer\_Eat\_drinking)
  + Make Distance features
* Find the closest property from a crime spot
* Calculate a distance (ex. Dist\_Eat\_drinking)

* After we have completed the visualisation for all the 5 years, we going to compare them based on the following measure:
  + Develop a unified map to simultaneously depict crime incidents and census data boundaries, highlighting significant factors associated with crime in different areas.
  + Produce these maps over a five-year timeline to analyse changes in neighbourhood demographics and crime rates, offering insights into the development and safety of specific areas.

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.